Imitation et contrôle prosodique dans l'entraînement à la remédiation phonétique : évaluation, mesure et applications pour l'enseignant de langues étrangères

A ceux qui sèment

Remerciements

« *La route est droite mais la pente est forte* ». Je souhaite remercier ici ceux qui m'ont mis sur cette route, ceux qui m'ont empêché d'en sortir, ceux qui m'ont aidé à gravir la pente et enfin, ceux qui m'accueilleront à l'arrivée.

Mes premiers remerciements sont adressés à Lorraine Baqué et Daniel Hirst qui ont accepté d'être les rapporteurs de ce travail de thèse, ainsi qu'à Bernard Harmegnies et Albert Rilliard qui ont bien voulu participer à mon jury de thèse.

Naturellement, je souhaite remercier spécialement mes directeurs de thèse : Corine Astésano qui m'a invité à débuter cette recherche (qui aurait pensé qu'une proposition jetée en l'air à l'issue d'un partiel m'aurait mené ici…) et Michel Billières, qui, avec curiosité, a accepté de m'encadrer. Les discussions en votre compagnie ont toujours été riches, bienveillantes et réjouissantes et vous vous êtes efforcés de m'aider à arriver au bout. Encore merci !

Je tiens également à remercier l'ensemble des membres du laboratoire Octogone-Lordat. Je pense particulièrement à Barbara Köpke, avec qui j'ai pu travailler sur un projet ANR avec l'Autriche pendant un an et à Vanda Marijanovic avec qui nous avons pu imaginer un cours de TICE attrayant pour les Master FLE et qui m'a en plus offert de le donner. Je souhaite aussi remercier chaudement Halima, Chris, les deux Charlotte, Stéphane, Encarni et bien sûr Evelyne… Ceux qui ne sont pas cités n'en sont pas moins remerciés.

J'adresse également des remerciements aux gens du DEFLE, à Michèle Bourdeau, à Corinne Rivière, à Marie ainsi qu'à Miberg qui m'ont toujours accueilli de manière inopinée, pour un café, une cigarette, et quelques rires.

Un immense merci aux ultras incroyables Benjamin Boulbène et Julien Dupouy pour leur aide salvatrice.

Un grand merci aussi à Kévin Lepan d'avoir accepté de faire son stage d'informatique avec moi. Ton travail a été précieux !

Je souhaite aussi remercier les personnes qui ont bien voulu participer à mes expériences, locuteurs et auditeurs, sans qui la part expérimentale de ce travail n'existerait pas.

Il m’est aussi très important de remercier mes coreligionnaires, doctorant(es) et jeunes chercheurs, avec des pensées toutes particulière à Claire Del Olmo et Emilie Massa, avec qui nous usons nos fonds de culotte sur les bancs de l’université depuis un temps certain, à Laury Garnier « l’âme des doctorants de Lordat » dont la fraîcheur est irremplaçable, à Noémie Te Rietmolen pour les discussions de *nerds* passionnés et les quelques bières éclusées dans la chaleur de l’été dernier et enfin Francesca Cortelazzo et Clara Solier, fidèles au poste pour les pauses et les discussions autour d’un café. Merci aussi à Lyanne, Ajhar, Kleopatra, Aurélie, Marie Mandarine, Rogelio, Raféu, Kamran, Lucille qui peuplent ou ont peuplé les bureaux des doctorants et ont aussi apporté leur contribution.

Merci aux collègues des autres labos, Maxime, Luce, Sophie entre autres, pour les organisations de journées et de colloque, et le travail lié aux charges de cours. Beaucoup de fun malgré les galères.

Je remercie évidemment ma famille, qui me regarde maintenant depuis longtemps faire des études : mes parents, mes frères, leurs compagnes. Je vous remercie et vous l’annonce, *a priori*, j’en suis arrivé au bout.

Enfin, merci aux amis, qui le restent bien que je n’ai pas eu le temps de beaucoup les voir ces derniers mois : Lydie, Pierre, Martin, Nicolas, Valentin, Louise (bien qu’on habite ensemble…) Rachel, Elise, Jojo, Gran Ju, les membres de l’association Un Archet Dans le Yucca, dont l’irremplaçable Manu, Brendan, les Sec de la ville du rock et encore, les amis des sphères virtuelles.

J’en oublie sûrement.

Si je vous ai croisé entre 2012 et 2016, soyez remerciés ici.

Table des matières

Table des Tableaux

Table des Figures

Abréviations

ASL	Associative Sequence Learning
C+/C-	Trop clair/Trop sombre
CAT	Communication Accomodation Theory
CE	Corpus d'Edimbourg
CECR	Cadre Européen Commun de Référence pour les Langues
CI	Corpus Imitation
CL	Corpus Logatome
DTW	Dynamic Time Warping
L1	Langue maternelle
L2	Langue étrangère
MA	Méthode articulatoire
MVT	Méthode Verbo Tonale
NLMe	Native Language Magnet expanded
T+/T-	Trop tendu/Trop relâché
T-Function	Turning Function ou fonction de courbure
TSP	Turning Function Stylisation Phonologique
TSR	Turning Function Stylisation Rythmique
VTM-T	Verbo Tonal Method Trainer

Introduction

Les prémices de cette recherche sont d'abord issues de notre étonnement.

Lorsque nous étions étudiants en Master de didactique du Français Langue Etrangère, nous rendant à reculons à nos premiers cours de « phonétique corrective par la Méthode Verbo-Tonale » (MVT), nous nous sommes surpris à découvrir que la pratique de remédiation phonétique pouvait être autre chose que la matière aride, ennuyeuse et frustrante dont nous gardions le souvenir depuis notre apprentissage de l'allemand et de l'anglais.

Objet et contexte de notre recherche, la MVT postule que l'erreur phonétique en langue étrangère (L2) est le résultat d'un mauvais traitement phonologique de la matière sonore perçue, ce qui conduirait les apprenants de L2 à se comporter comme des « *durs d'oreille* » (Guberina, 1978, p. 286).

Les apprenants présenteraient donc un biais perceptif. Ce déficit perceptif des apprenants en L2 est décrit dans la littérature sous diverses appellations qui réfèrent à la notion de perception catégorielle, par exemple, le crible phonologique (Polivanov, 1931; Troubetzkoy, 1939) ou les aimants perceptifs (Kuhl *et al.*, 2008; Nguyen, 2005).

Considérant que l'erreur phonétique se situe à un niveau perceptif, la MVT préconise des pratiques de remédiation visant à rééduquer le système perceptif de l'apprenant. Ainsi, plutôt que de redire à l'identique le modèle que l'apprenant prononce de manière erronée, le praticien verbo-tonaliste le reproduit en modulant ses paramètres phonétiques et prosodiques afin de donner quelque chose de nouveau à percevoir à l'apprenant, qui se présente sous une forme optimale pour l'apprenant (Billières, 2000; Renard, 2002b).

La correction phonétique présente les caractéristiques et les enjeux des situations d'imitation. En effet, le but avoué de l'interaction est de faire imiter un nouveau comportement parolier à l'apprenant. Ainsi, nous pourrions dire que cette situation constitue un jeu d'imitation entre apprenant et enseignant. Ceci dit, l'interaction imitative en MVT présente une certaine étrangeté.

Concrètement, observer une séquence de MVT revient à regarder des individus qui s'imitent l'un et l'autre mais qui ne font pas la même chose. L'aspect paradoxal que revêt ce type d'interaction nous a donc conduit à interroger la notion d'imitation en parole et la compétence d'imitation prosodique des locuteurs, et plus particulièrement dans le contexte de la MVT.

Avant d'associer imitation et parole, il nous a semblé nécessaire de démêler la seule notion d'imitation. Notre chapitre 1 propose en premier lieu une analyse des éléments constitutifs de la situation d'imitation. Intrinsèquement, l'imitation a une dimension interactive puisqu'elle implique *a minima* deux individus : un modèle qui produit un comportement et un imitateur qui observe puis reproduit tout ou partie de ce qui a été observé. Après une description factuelle des différentes facettes de la triade imitative (imitateur, comportement, individu imité), nous proposons une définition générale des comportements imitatifs, une définition qui sera dissociée de leurs fonctions. Nous soulignons en suivant que l'imitation, ou plutôt, les imitations ne sont pas monolithiques. Il sera en effet proposé de discriminer différents types d'imitations en fonction :

- Du critère de l'intentionnalité du sujet imitant, du mimétisme à la mimésis (Baudonnière, 1997; Donald, 1993).
- De l'observation des buts du modèle et de leur compréhension par l'imitateur, afin de distinguer absence d'imitation, mimique et émulation (Call & Carpenter, 2002)

Suite à ces premières approches, les fonctions d'adaptation, de communication et d'apprentissage des imitations seront mises en exergue par des exemples concrets.

Le chapitre 2 de notre ouvrage traite des structures neuronales et cognitives de l'imitation. Imiter suppose que nous parvenions à établir une correspondance entre ce que nous observons et ce que nous faisons. En d'autres termes, il s'agit ici d'un problème de correspondance perceptuomotrice (Brass & Heyes, 2005) qui pose la question de savoir comment un sujet parvient à déterminer quels muscles activer lorsqu'il observe un comportement qu'il souhaite reproduire. Nous proposerons une brève revue de neurologie à propos des structures cérébrales qui seraient impliquées dans l'imitation gestuelle (et son inhibition) d'une part, et dans l'imitation de la parole d'autre part. Au terme de ce chapitre, nous intégrerons des aspects d'imitation à la perception et à la production de la parole, défendant l'idée que les individus ajustent de manière automatique ou consciente leur manière d'utiliser le langage, en fonction de leur environnement immédiat et de leurs expériences

sensorimotrices passées. Les comportements imitatifs influenceraient de manière dynamique notre façon d'agir.

C'est au chapitre 3 de ce travail que nous ancrerons véritablement l'imitation dans la pratique de la MVT et que nous envisagerons les enjeux de la correction phonétique (pour l'enseignant et l'apprenant) au travers du prisme des comportements imitatifs. Enseignant comme apprenants sont en effet dans la position de l'imitateur ou de l'être imité. Tandis que l'apprenant cherche à reproduire les modèles de l'enseignant, ce dernier est amené à s'imiter lui-même en modulant ses productions. Il doit alors faire preuve d'un contrôle phonatoire particulier pour appliquer avec succès les techniques vocales de la MVT.

Dans ce travail, nous nous intéresserons tout particulièrement aux procédés de la MVT manipulant la prosodie des énoncés. En d'autres termes, nous essaierons d'estimer la précision avec laquelle des locuteurs francophones, naïfs ou experts, parviennent à reproduire la prosodie des énoncés qu'ils entendent. D'un point de vue méthodologique, ceci suppose d'abord que nous puissions évaluer ou mesurer leur performance. Il s'agit là d'un enjeu fondamental de notre thèse.

Au chapitre 4, nous passerons en revue les aspects méthodologiques de l'étude de l'imitation en parole. Nous décrirons dans un premier temps la manière dont l'expérimentateur peut chercher à faire émerger deux comportements imitatifs différents :

- La convergence phonétique, imitation parolière, (Giles, Coupland, & Coupland, 1991) au moyen de tâches conversationnelles avec des locuteurs naïfs (Nathalie Lewandowski, 2012; Pardo, 2006).
- L'impersonnation, imitation vocale, (Révis, De Looze, & Giovanni, 2013; Zetterholm, 2009a) dans des tâches non-interactives mettant en scènes des locuteurs experts.

Dans la suite de notre propos, nous étudierons les moyens par lesquels les chercheurs répondent à la question : « y a-t-il imitation ? »; *i.e.* nous nous intéresserons à l'évaluation et à la mesure de l'imitation en parole. Dans ce domaine, deux approches complémentaires sont généralement associées afin de comparer les modèles et les imitations :

- Les tests d'évaluations perceptives reposant sur l'expertise de sujets humains.
- Le mesurage des paramètres acoustiques.

Bien que les deux approches présentent des avantages certains, il sera souligné que ni l'une, ni l'autre, ni leur conjonction ne permettent d'évaluer avec stabilité la réussite d'une imitation. Les évaluations perceptives sont holistiques tandis que les mesures acoustiques classiques (locales) ont souvent une focale trop grossissante qui ne permet pas de rendre compte du mouvement prosodique dans sa globalité, le niveau qui nous intéresse.

En réponse à ces constats, nous envisagerons de mesurer la similarité des formes des courbes de *f0* entières d'un modèle et de son imitation au moyen de différentes mesures. Il se pose alors la question de la validité perceptive de ces mesures : tant dans les domaines de la parole, que dans le domaine de la reconnaissance des formes géométrique, une mesure de similarité doit donner un résultat approchant de celui que donnerait une évaluation visuelle ou auditive de la similarité (Arkin, Chew, Huttenlocher, Kedem, & Mitchel, 1991; Hermes, 1998b). Ainsi, avant de pouvoir évaluer la performance d'imitation prosodique de locuteurs produisant des procédés de la MVT, il est nécessaire de résoudre le problème de mesure de la similarité.

Le chapitre 5 de ce travail vise apporter des réponses quant à ce problème. Dans ce chapitre expérimental, nous soumettons un corpus d'imitations au crible de 2 tests d'évaluation perceptive et de 4 mesures de la similarité prosodique. Le corpus d'imitation a été enregistré par 4 locuteurs qui ont imité des phrases syntaxiquement ambiguës au cours de 3 tâches proposant une gradation du degré d'imitation, d'une simple répétition à une exagération de l'imitation. De ce corpus, nous avons sélectionné 72 imitations qui ont été soumises à deux tests perceptifs :

- Un test AX (Mary, Anish Babu, & Joseph, 2012) de jugement perceptif de la similarité, proposant d'évaluer la similarité du modèle et de son imitation sur une échelle graduée de 1 à 5.
- Un test AXB (Pardo, 2013a) de jugement à choix forcé de la similarité entre un modèle et deux imitations de ce modèle.

Les résultats obtenus au moyen de ces tests sont présentés avant d'être mis en relation pour comparer leur apport respectif. Nous montrerons que le test AXB permet seulement de classer les éléments dans différentes catégories en donnant la seule information qu'il existe au moins un degré de différence entre ces catégories. Le test AX, qui propose une notation absolue, permet lui de connaître le degré de réussite de l'imitation.

Dans la suite de ce chapitre, les 72 imitations sont ensuite passées au crible de 4 mesures de la similarité prosodique :

- 2 mesures issues de la comparaison de la forme de *f0* brute (Hermes, 1998b; Rilliard, Allauzen, & de Mareüil, 2011)
- 2 mesures issues de la transformation des formes de *f0* stylisées de manière rectilinéaire, la fonction de courbure ou *Turning Function* (Arkin et al., 1991; Veltkamp, 2001)

Les résultats obtenus au moyen de ces différentes mesures sont systématiquement croisés avec les résultats des évaluations perceptives du test AX. Les mesures données par les fonctions de courbure se révèleront être les meilleurs candidats pour la mesure de la similarité prosodique, du fait de leur bonne corrélation avec les scores des évaluations perceptives.

Le chapitre 6 illustre l'application de la fonction de courbure à un corpus d'imitations lexicalisées et délexicalisées (dites sur « dada »), recueillies auprès de 2 locuteurs experts en MVT et de 2 locuteurs naïfs. La MVT propose en effet de produire des énoncés délexicalisés pour présenter l'intonation aux apprenants, il est donc intéressant dans notre cadre de recueillir un corpus de ce type.

Après écoute d'un énoncé tiré d'un dialogue, ces locuteurs ont produit des séries de 3 imitations, en alternant phrases, logatomes et phrase ou bien logatome, phrase et logatome. Nos analyses de ce corpus visent à répondre à deux types de questionnement. En premier lieu, nous questionnons encore la mesure de similarité que nous appliquons, afin d'appréhender au mieux son comportement. En parallèle, nous testons des hypothèses ayant trait :

- A l'expertise des locuteurs dans l'accomplissement des tâches d'imitations prosodiques.
- A l'effet de la longueur du modèle sur la réussite de l'imitation délexicalisée
- A la conservation du pattern prosodique original dans les imitations successives d'une même phrase

Ainsi, ce chapitre vise à obtenir une meilleure connaissance de l'outil de mesure de la similarité développé durant notre travail tout en illustrant ses applications à nos questions de recherches. Nous terminons ce chapitre en ouvrant des perspectives d'application pratique des mesures de la similarité à destination des enseignants se formant à la MVT. Nos analyses du corpus de ce chapitre ayant révélé que la production de logatomes pouvait être compliquée pour des locuteurs naïfs, il semble pertinent de proposer un outil d'entraînement aux

techniques vocales. Nous décrirons pour finir un logiciel qui mesure de manière automatique la similarité prosodique au moyen d'une fonction de courbure afin de fournir cet outil aux enseignants.

Chapitre 1 : Définition du spectre des imitations et de ses fonctions

Le copiste peignant une reproduction de tableau de maître et l'élève de karaté apprenant les *katas*[1] de son art martial représentent deux cas qui peuvent évoquer des actions imitatives dans le sens donné par Baudonnière (1997) à l'imitation *stricto senso* : « *refaire ce qu'un autre être humain a déjà fait devant (avant) soi* ». Les *katas* représentent une série de gestes codifiés, produits dans un ordre précis face à un adversaire imaginaire. Ils sont transmis de génération en génération au sein des écoles d'arts martiaux où les élèves les répètent jusqu'à ce qu'ils soient intégrés –au point de devenir des réflexes- dans les situations réelles de combat. Ce premier exemple correspond tout à fait à la définition que nous venons de citer : les élèves refont les gestes qui ont été faits avant eux et cherchent à les reproduire avec excellence. Le copiste, quant à lui, se basant sur l'observation du produit fini, peut tenter de reproduire les techniques picturales du maître, imaginer les étapes successives nécessaires à l'élaboration du tableau ainsi que les gestes utilisés pour déposer la peinture sur la toile pour atteindre un résultat le plus proche possible de l'original. Le produit fini du copiste sera désigné sous le terme d'imitation ou de copie. Ces deux exemples diffèrent par la finalité de l'action imitative (un objet pour le copiste, une séquence de gestes pour le karatéka) mais ils sont liés par le processus de reproduction, qui présuppose, comme le souligne Baudonnière, la perception et probablement la compréhension d'une action qui va être reproduite par la suite, soit une intégration des schémas moteurs par l'expérience de la répétition.

De tels exemples sont intéressants pour introduire la notion centrale de notre travail de thèse, l'imitation, notion à laquelle nous consacrerons ce premier chapitre. Dans le langage courant, l'imitation présente une apparente trivialité car le concept est exprimable très simplement, comme nous venons de le voir. En outre, l'imitation comme manière d'agir chez l'être humain est très souvent dépréciée car elle ne demande pas de créativité de la part de la personne qui la produit. Dans certains cas, elle est assimilée à un vol, comme en classe quand

[1] En japonais, « kata » peut être décrit par plusieurs kanjis issus du chinois :
- **形** signifie la forme, étymologiquement, sa reproduction scripturale
- **型** a le sens de moule, de forme idéale, ou encore de trace laissée
- **方** désigne la personne ou la façon

un élève copie son voisin ou dans les domaines culturels et intellectuels comme l'art ou la littérature. L'imitation est donc souvent un comportement répréhensible par l'autorité. Pourtant, si nous reprenons le cas de l'élève, le comportement de « copier le voisin » pourrait également être vu comme une stratégie d'adaptation pour répondre à une pression de l'environnement. En effet, il est attendu de l'élève qu'il obtienne de bonnes notes, or, si ce dernier n'a pas suffisamment travaillé sa leçon et qu'il pense ne pas obtenir une bonne note, il peut être poussé à copier les réponses de son voisin pour pallier les manquements de sa connaissance et atteindre l'objectif fixé en faisant croire qu'il a appris sa leçon. Ce faisant, l'élève imite le résultat uniquement de l'action observée, il produirait alors de l'émulation.

Dans d'autres domaines, dont le divertissement, l'imitation peut au contraire se trouver valorisée : l'imitation, gestuelle ou vocale, peut intervenir à divers degrés dans les tentatives de faire rire. Certains humoristes célèbres jouissent d'ailleurs de tribunes à des heures de grande écoute (par exemple, les « voix » des Guignols de l'Info, dont la plus fameuse est celle d'Yves Lecoq) ou arrivent à remplir les salles de spectacle parisiennes (comme Laurent Gerra). Plutôt que de mettre en question la motivation de l'imitateur, ce nouvel exemple met en lumière la possibilité pour un être humain X d'arriver à imiter la **voix** d'un autre être humain Y avec une fidélité telle que la personne X pourrait arriver à se faire passer pour Y auprès d'une personne Z, ou plutôt que Z reconnaisse la voix de Y dans la voix de X. Un tel constat pose de nombreuses questions sur la perception et le traitement des sons de parole des auditeurs, ainsi que de leur capacité à les reproduire ; soit, à les imiter.

Ces quatre exemples très divers varient en termes d'agent –le copiste, le karatéka, l'élève et le comédien-, d'objet modèle –la reproduction d'un objet, de gestes corporels, d'une information ou de sons de parole-, de temporalité de la mise en œuvre –en présence ou en l'absence de l'autre ou de la trace que son comportement a laissée, de manière différée ou simultanée-, de finalité et d'intentionnalité supposée de celui qui produit l'imitation. Ils soulignent donc les différentes facettes des comportements imitatifs et laissent entrevoir certaines questions posées par l'étude de l'imitation.

Dans un premier temps, l'objet de ce chapitre sera de discuter la notion d'imitation ainsi que des notions connexes comme le mimétisme ou l'émulation, telle qu'elles sont présentées dans la littérature de différents domaines scientifiques. Comme l'indiquait déjà Paul Guillaume, à une époque où certains champs scientifiques n'étaient pas encore très développés, voire inexistants : « *On a beaucoup étudié l'imitation : psychologues, médecins,*

naturalistes, sociologues s'y sont intéressés » (1925, p. 5). Nous ne saurions aborder l'étude de l'imitation parolière sans avoir dégagé au préalable un socle terminologique et conceptuel commun nous servant à la remettre en perspective dans son actualité. Nous nous intéresserons alors aux formes et fonctions de l'imitation chez l'être humain, que nous n'hésiterons pas à qualifier « d'animal[2] mimétique ».

Le spectre des comportements imitatifs nous permettrait de nous adapter à notre milieu, dans certains cas de communiquer avec nos pairs et serait un médiateur de certains de nos apprentissages, qu'ils soient gestuels ou linguistiques.

1 Définir l'imitation

Depuis plusieurs siècles, l'activité humaine a pu être analysée comme une activité reproduisant des aspects de son environnement. Dans le domaine des Arts, les exégètes aiment à faire référence à Aristote, pour qui, les Arts (comme le théâtre, la sculpture et la musique) ainsi que d'autres activités humaines seraient des manières de reproduire, d'imiter ou de faire référence à des sons, des schémas sociaux ou des situations préalablement vécues. Pour ce faire, l'humain utiliserait notamment « [*le] rythme, [le] langage [et la] mélodie* » Aristote, trad. de Gernez, B. (1997, p. 3–5, 1447a). C'est le concept de *mimêsis* qui apparaît central dans sa pensée : le philosophe grec estimait, d'après Gernez[3] (1997, p. XV), que la notion d'imitation peut être assimilée à « *une tendance naturelle favorisant l'apprentissage* » ; en prenant l'exemple de la tragédie antique, il s'agit de l'imitation d'une action noble ayant pour but l'expurgation des émotions. La notion d'imitation envisagée ainsi dans le domaine des Arts antiques semble bien loin de nos préoccupations à venir : nous ne nous préoccuperons pas des aspects liés à l'émotion.

[2] D'un point de vue biologique d'une part, car nous appartenons au règne animal ; d'un point de vue philosophique d'autre part, en nous référant à Derrida (2006) (plus particulièrement, pp 18, 54 & 185)

[3] Nous indiquons Gernez (1997), il s'agit de ses commentaires de la Poétique d'Aristote. Etrangement, nous devrions indiquer Aristote (1997), pour être fidèle avec la bibliographie.

1.1 Imitation(s) : une triade à la description complexe

Définie dès la fin du XIX^{ème} siècle par Thorndike (1898) comme « *learning to do something from seeing it done* », l'imitation, ainsi que d'autre comportements approchants, reçoivent des définitions différentes en fonction des chercheurs et des disciplines...

La définition de Thorndike trouve son contexte dans le domaine éthologique. Comme le notent Galef & Benett (1988, p. 55), Thorndike cherchait à observer les capacités d'apprentissage par observation des animaux, au moyen de dispositifs méthodologiques novateurs pour l'époque (Wozniak, 1999). L'échec à obtenir une preuve de l'imitation chez l'animal non-humain nous a cependant laissé cette première nuance de l'imitation. Celle-ci est considérée dans ce cadre comme un médium d'apprentissage par observation ; or, cette notion n'est pas toujours incluse dans d'autres définitions de l'imitation.

Il faudrait d'abord envisager la définition de l'imitation sous une forme fragmentée : le modèle, l'imitateur, et les comportements qui les lient sont de différentes natures selon les situations. Brass & Heyes (2005) nous donnent une première définition a minima « *Imitation – copying body movements* » qui se concentre sur le corps comme modèle, excluant des pratiques comme la pantomime. Cette dernière vise à symboliser des caractéristiques de l'environnement en les imitant, c'est à dire, en les évoquant avec son propre corps. Leur définition paraît donc restreindre l'imitation à des actions corporelles, où on suppose que l'objet imité est produit par un membre de l'espèce de l'imitateur qui peut alors s'identifier au modèle pour tout ou partie. Cette vue est cependant une réduction de la définition donnée par Heyes (2001) indiquant « *In this article, 'imitation' refers to copying by an observer of a feature of the body movement of a model* ». En réalité, la définition de Brass & Heyes (2005) semble restrictive par économie.

Dans l'étude de l'imitation, le modèle impulsant le geste étudié n'est souvent évoqué qu'implicitement dans la définition du comportement imitatif.

1.1.1 Le modèle comme source de l'imitation

Chez Baudonnière (1997), le modèle est défini comme « *un autre être vivant* » alors que dans les expériences de Press *et al*. (2005) sur la préhension de la main, le modèle est tour à tour humain ou robotique. Dans ce dernier cas, les résultats indiquent que les impulsions de la main humaine permettent d'obtenir une meilleure performance des sujets, mais que la main

robotique (*i.e.* une pince, pour être précis) provoque également des réponses d'imitation. Un modèle vivant semble donc améliorer la fréquence de réponses imitatives de la part des sujets humains. Cependant, l'action du modèle robotique semble reproductible malgré sa physionomie non humaine : qu'elle impulse également une réponse imitative n'est donc pas à exclure.

La question du modèle varie dans la définition de l'imitation, bien que sa présence ne soit pas négociable : la nature du modèle est souvent liée à la situation observée. Cette dernière est dictée par la discipline et la perspective de recherche par lesquelles on en arrive à étudier l'imitation :

- En éthologie, le modèle est en général un membre de la même espèce que l'imitateur, par exemple deux cétacés (Abramson, Hernández-Lloreda, Call, & Colmenares, 2013), mais leur âge (un diamant mandarin et son géniteur, Tchernichovski & Nottebohm, 1998) ou leur origine, peuvent différer (un éléphant d'Afrique et des éléphants d'Asie, Poole, Tyack, Stoeger-Horwath, & Watwood, 2005). Parfois l'expérimentateur humain est aussi pris comme modèle dans la résolution de problèmes (Tomasello, Savage-Rumbaugh, & Kruger, 1993)
- En psychologie du développement, le modèle peut être un pair (Nadel, 1986; Nadel & Potier, 2002b), un objet inanimé (Jacobson, 1979)[4], un *care giver* (Guillaume, 1925; Kuhl & Meltzoff, 1996)
- Les études sociales s'intéresseront souvent aux pairs, c'est-à-dire, à ce qui les inclut dans un groupe et les différencie des autres groupes (Goudailler, 2002) ou bien, aux réactions de l'individu face au groupe (Asch, 1955).

Dans le cadre expérimental, le modèle se retrouve souvent dématérialisé afin que les sujets reçoivent un même input. C'est le cas dans de nombreuses études en parole que nous évoquerons plus loin. Un exemple toutefois : dans ce contexte, différents sujets adultes écoutent tous la même voix à reproduire. Ce paradigme incluant un modèle désincarné est cependant largement répandu, ciblant un public varié d'imitateurs, par exemple les nouveaux nés (Coulon, Hemimou, & Streri, 2012) ou les oiseaux (Tchernichovski, Mitra, Lints, & Nottebohm, 2001).

[4] Ici, le stimulus était un crayon. Comme le note Baudonnière (1997), ce type de stimulus a été utilisé ici pour observer si la protrusion de la langue et l'action d'un crayon au niveau du visage avaient le même effet sur l'enfant.

Il a également été envisagé d'observer le sujet qui se prend lui-même pour modèle, produisant alors ce que Guillaume (1925, p. 84) considère chez le jeune enfant comme une sorte de réaction circulaire. Les exemples donnés par Guillaume « *pendant plus d'un quart d'heure, elle agrippe à pleines mains l'étoffe d'un coussin, […] elle joue longuement avec sa robe (opposition du pouce) sans regarder la main* » sont interprétés comme une recherche d'expérience, ou, dans ses termes « *la répétition des mêmes expériences [qui] permettra la sélection des modes d'activation efficaces* ». L'exemple du karatéka reproduisant ses katas pourrait peut-être s'assimiler à ce genre de comportement. Ici, le psychologue envisage que l'imitation sert un processus d'apprentissage et de mise en place de nouvelles stratégies cognitives. La fonction et la forme des comportements imitatifs seront débattues ultérieurement.

Même s'il est parfois sous-entendu dans la définition de l'imitation, le modèle reste le déclencheur d'un comportement imitatif, un élément définitoire primordial. Il convient alors de se demander son rôle, au-delà de véhiculer le comportement à reproduire. La relation du modèle vis-à-vis de l'objet imité et de l'imitateur peut être mise en question. La relation entre le modèle et l'imitateur pourrait peut-être nous indiquer la forme ou la fonction du comportement imitatif. Avant d'évoquer l'imitateur, nous nous attacherons à définir « l'objet » imité.

1.1.2 Objets d'imitation(s) et objets de recherches

Que ce soit une copie de maître, une série de gestes martiaux ou une voix imitée, l'action nécessaire pour produire l'imitation doit correspondre –théoriquement– à la même action qui a permis de réaliser le tableau de maître, le kata ou la voix originale. Définir alors l'objet de l'imitation comme « *mouvements corporels* » comme le font Brass & Heyes (2001, 2005) devrait pouvoir s'appliquer à toute activité humaine.

Revenons à la peinture. Le maître a utilisé d'innombrables gestes pour peindre une toile, il a appliqué des couches de couleurs successives par petites touches. Un copiste contemporain du maître aurait théoriquement pu, en même temps que le maître, faire la série de gestes qui aurait abouti au tableau. Un copiste de nos jours tenterait de reproduire uniquement le résultat des gestes : le tableau serait reproduit, certes, mais peut-être au moyen de techniques différentes. La nature de l'objet de l'imitation, ainsi que son contexte de

production peuvent amener à des processus différents : faire la même chose (l'imitation) ou atteindre le même résultat (l'émulation).

De plus, la fabrication d'objets est un cadre particulier de l'imitation car les objets persistent : le modèle et la copie. Par opposition, les *gestes* de production des objets sont perdus, sauf s'ils sont enregistrés ou consignés. C'est cette dernière solution que de nombreux psychologues comme Guillaume (1925) ou Zazzo (1957) ont choisie en tenant des carnets d'observations. Quand les moyens techniques sont devenus disponibles, les chercheurs ont parfois choisi d'enregistrer leurs protocoles expérimentaux (Nadel, Guérini, Pezé, & Rivet, 1999)[5].

Les possibilités d'enregistrement et de conservation des données ont eu un impact relativement fort sur de nombreux domaines des sciences humaines, notamment les sciences du langage qui ont pu s'ouvrir à l'oral. Cette technologie a aussi eu un impact important sur la recherche en imitation : on a maintenant accès au geste, qui dans le cas de l'activité parolière par exemple, provoque la production de la voix. Une question que peut donc se poser le chercheur en imitation est : « S'intéresse-t-on au geste, à son résultat, ou à la stratégie mise en œuvre pour l'atteindre ? Ou à autre chose, ce que signifie le fait de reproduire le geste ? ». Dans le cas de la parole, on peut par exemple se poser la question de savoir si l'activité motrice d'un imitateur change entre sa parole naturelle et sa parole imitée, s'il y a des mécanismes cognitifs spécifiques à l'imitation parolière ou bien des zones cérébrales plus/moins actives durant la production d'imitation. De même, on peut aussi aborder la question par un angle sociolinguistique en se demandant si les imitations de paramètres lexicaux ou phonétiques sont liées à la situation diaphasique des locuteurs.

Pour en revenir au geste en tant qu'objet à imiter, on peut se demander :

- Si le modèle est constitutif de l'espèce (si c'est un réflexe, par exemple). Dans certains cas, comme l'imitation en éthologie ou en psychologie du développement (Anisfeld, 1996), il faudra se poser la question d'exclure des observations les gestes qui appartiennent au répertoire normal de l'espèce. Une autre possibilité est de considérer un autre mécanisme que l'imitation (ce que propose Messum, (2007c; 2007b) en attribuant un plus grand rôle au *care giver* qu'aux capacités imitatives de l'enfant).

[5] L'exemple donné est assez récent, on peut trouver des traces d'enregistrements d'expériences datant des années 60-70, notamment dans la vulgarisation pour le grand public.

- S'il est dans l'habitude de l'imitateur. On ne pourra alors pas le considérer comme nouveau, et une production imitative de ce type d'objet pourra être analysée comme imitation dans certains contextes précis [par exemple : la préhension de la main dans les expériences de Press *et al*. (2005)]. D'autres chercheurs estimeront qu'un comportement connu ne constitue pas d'imitation...
- S'il a un caractère nouveau ou différent du geste habituel, il peut correspondre au type d'objet à imiter qui entre dans le cadre des définitions de l'imitation mettant en avant sa valeur d'apprentissage.

A la vue de ces différents cas, on pourrait supposer que le troisième est le plus représentatif des objets d'imitation dans la littérature. En effet, l'idée selon laquelle l'imitation joue un rôle prépondérant dans l'apprentissage est très ancrée dans les esprits et le caractère novateur d'un mouvement peut alors servir à dire : « Ce mouvement de Y est nouveau, il l'a vu chez X, donc c'est une imitation ».

Pourtant, lorsque nous imitons, il est rare que nous fassions des gestes absolument inconnus : cela peut être leur caractère symbolique ou leur enchaînement qui a une valeur suffisante pour que nous les produisions consciemment. Par ailleurs, dans le cas de nombreuses imitations, les sujets reproduisent des comportements sans même s'en rendre compte, que l'objet soit connu ou non. Enfin, donner une valeur d'apprentissage à l'imitation dépasse la simple définition du phénomène et devrait plutôt se rapporter à la fonction de certaines formes d'imitation.

Définir avec une telle rigueur le modèle, puis l'objet de l'imitation peut apparaître fastidieux, voire inutile. Cependant, modèle et objet ont un impact sur le comportement que l'on décrit. Prenons le temps d'imaginer les situations suivantes où un sujet va reproduire :

- Un geste connu fait par un modèle équivalent (un enfant qui imite un autre enfant lever le bras) (1)
- Un geste connu d'un modèle différent (un enfant qui imite les ailes d'un oiseau) (2)

Les situations (1) et (2) correspondent à une définition simplement motrice de l'imitation, équivalente à celle de Heyes (2001) ou Baudonnière (1997) « *Refaire ce qu'un autre être vivant a fait devant (avant) soi* ». Toutes ces situations peuvent également se rapporter à ce que Guillaume (1925, p. 133) appelle « *l'imitation symbolique, l'imitation d'un geste à vide* » quand l'imitation se produit en l'absence de l'objet lié au comportement à produire (par exemple, reproduire le geste d'allumer une cigarette sans briquet pour demander

du feu à un passant). On se trouve alors dans le domaine de la représentation. Pour le psychologue français, c'est d'ailleurs une spécificité de l'espèce humaine qui n'intervient qu'assez tard dans le développement de l'enfant, aux environs de la troisième année (1925).

Questionner l'objet de l'imitation revient donc à questionner les productions du modèle et de son imitateur. En effet, sans trace de l'objet qui sera imité, rien ne permet d'attester du comportement imitatif. Souvent, ce sera la similarité entre l'objet (soit, le comportement du modèle) et son imitation qui indiquera le degré de réussite du comportement imitatif[6]. L'objet imité servira alors de référence pour expliquer la différence (ou l'absence de différence[7]) entre plusieurs comportements produits par un imitateur. C'est ce que note Baudonnière (1997) en incluant une potentielle tierce personne pour juger que les comportements produits sont bien des imitations :

> Dans tous les cas [l'imitation] suppose que le comportement modèle soit décodé et interprété de façon suffisamment correcte pour que production et reproduction puissent être perçues comme semblables[8].

Cette dernière remarque souligne que l'imitation est souvent jugée *a posteriori* (que ce soit par l'imitateur même, par son modèle ou par un tiers) ; cela devrait avoir un impact notable sur nos pratiques expérimentales. D'autre part, l'imitateur, qui va nous intéresser ci-après, doit pouvoir faire preuve de plusieurs capacités pour produire des comportements imitatifs : décoder, interpréter et se mouvoir.

1.1.3 L'imitateur

La situation initiale de l'imitation –soit, avant sa production– nécessite donc un modèle, un comportement à reproduire et un imitateur potentiel. Ce dernier, au même titre que le modèle, varie grandement selon les situations : animal en éthologie, enfant dans la psychologie du développement, *etc*. Nous ne reviendrons pas sur la diversité des situations que la littérature offre et nous nous concentrerons pour le moment sur l'être humain. En

[6] C'est d'ailleurs un problème fondamental de l'imitation parolière : comment estimer la similarité (ou la différence) de plusieurs sons ? Voir chapitre 4

[7] Différence/absence de différence *vs*. similarité/absence de similarité, le paradigme est le même. Nous reviendrons sur la question au chapitre 4

[8] Nous soulignons

fonction des caractéristiques de l'objet à imiter des questions peuvent se poser sur les capacités intrinsèques de l'imitateur : celui-ci aura-il le dispositif nécessaire pour produire le comportement demandé ?

Dans certains cas, il paraît saugrenu d'imaginer certaines situations. Nous n'aurions pas l'idée de tenter de faire reproduire un kata à un enfant de neuf mois car il faut pour cela pouvoir se tenir de debout, synchroniser les mouvements des bras, des jambes et des autres parties du corps, ce qu'un enfant de cet âge n'est pas en mesure de faire. En plus d'avoir une compétence de décodage et d'interprétation du comportement cible, un sujet, pour réussir à imiter, doit donc disposer des capacités motrices nécessaires à la production du comportement. Ces points en particulier constituent des arguments essentiels de la thèse de Messum (2007a). Ce dernier défend l'idée que les enfants n'apprendraient pas à prononcer leurs premiers mots par imitation car leur capacité de contrôle du tractus vocal serait insuffisante à ce moment de leur développement. Par ailleurs, Messum indique également que les tout jeunes enfants seraient simplement dans l'incapacité d'imiter, car ils n'auraient pas encore appris à le faire (nous évoquerons à nouveau l'acquisition du langage par l'enfant en discutant la fonction d'apprentissage de l'imitation, *cf*. p. 70-73)

Au-delà des possibilités physiques de l'imitateur, la recherche sur l'imitation investigue donc la capacité de l'être humain à établir des correspondances entre ce qui est observé et ce qui est produit. Chez l'être humain, cette capacité semble fortement ancrée, comme le notent Chartrand & Bargh (1999, p.900) :

> The perception-behavior link posits the existence of a natural and non-conscious connection between the act of perceiving and the act of behaving.

D'ailleurs, l'espèce humaine passe pour une espèce particulièrement douée dans ce type de comportement : pour Carpenter & Call (2009), les humains seraient des imitateurs formidables, capables non seulement d'imiter les membre de leur propre espèce, mais également ceux des autres espèces. Brass & Heyes (2005, p. 489) s'accordent à cette vue en estimant qu'imiter semble un comportement simple pour l'être humain.. D'un point de vue scientifique, il est nécessaire de pondérer leur propos avec la suite de leur assertion :

> Imitation […] appears to be simple. However, the ease with which humans imitate raises a question, sometimes known as the correspondence problem that is proving difficult to answer. When we observe another person moving we do not see the muscle activation underlying their movement but rather

> the external consequences of that activation. So how does the observer's motor system 'know' which muscle activations will lead to the observed movement?

La résolution de ce problème qui pose la question de savoir comment est faite la correspondance entre observation d'un comportement et activation des schémas moteurs n'est pas triviale. Ce problème suppose de nombreux paramètres pour en comprendre le fonctionnement : des structures neurologiques et des mécanismes cognitifs qui doivent permettre au sujet d'articuler sa perception de l'environnement et sa propre production. Nous renvoyons la question au chapitre second de ce travail, où nous présenterons le substrat neurologique probablement mis en œuvre durant l'imitation, ainsi que des éléments de théories cognitives de l'imitation.

1.1.4 Synthèse : la triade imitative et ses complexités

En décrivant les forces en présence dans le comportement d'imitation –le modèle, l'objet et l'imitateur–, nous souhaitions faire émerger plusieurs idées clefs.

En premier lieu, nous avons souligné que l'imitation constitue un phénomène largement répandu dans le vivant, puisqu'il a été détecté dans le comportement de plusieurs espèces animales, que ce soient des mammifères comme les primates non-humains, les cétacés, les éléphants et les êtres humains, ou bien des oiseaux comme les diamants mandarins. Evidemment, notre inventaire des espèces qui ont été observées pour leur capacité d'imitation n'était pas exhaustif et d'autres espèces semblent encore pouvoir faire preuve de ce type de comportement. Poole *et al*. (2005) évoquent par exemple des études portant sur les chauves-souris tandis que Zentall (2006) dans une revue assez complète donne encore d'autres exemples. Une telle diversité d'imitateurs laisse penser que l'imitation n'est pas monolithique en fonction des espèces. Il est donc probable qu'il faille penser aux imitations en termes de continuum comme le note d'ailleurs Nadel (2005). Par exemple, chez l'humain, des comportements imitatifs du plus simple (un geste comme un coup de poing) au plus complexe (une série de gestes aboutissant à la réalisation d'une intention, comme le cas d'une peinture) pourraient indiquer deux extrémités d'un tel continuum.

Par ailleurs, en plus de présenter une multiplicité d'imitateurs et de modèles potentiels, nous avons fait émerger la problématique des objets imités. Ceux-ci sont des mouvements corporels bien que souvent l'imitateur tente d'en copier le produit avec ses propres organes :

ce constat aboutit au problème de correspondance (*i.e.* comment est fait le lien entre observation d'un comportement et activation musculaire), que Brass & Heyes (2005) résument de manière très claire. Ce problème est encore amplifié si on distingue les objets d'imitation en fonction de leur opacité (ou de leur transparence) perceptive. Par exemple :

- Reproduire le mouvement de préhension de la main d'un modèle, pourrait être qualifié de transparent, car l'imitateur peut voir sa main ainsi que celle du modèle
- Reproduire une expression faciale pourrait au contraire être considéré comme un mouvement opaque perceptivement, car l'imitateur ne voit pas sa propre figure.

A propos des objets, nous nous intéresserons particulièrement aux sons de parole qui résultent de l'activité de production du langage, une activité complexe :

> Speech production is a complex multistage motor process that requires phonetic encoding, intiation and coordination of sequences of supra-laryngeal and laryngeal movements produced by the combined actions of the pulmonary/respiratory system, the larynx and the vocal tract. (Sato *et al*., 2013, p.1)

Nous postulerons que tout locuteur sain est, dans une certaine mesure, apte à imiter ce qu'un autre a dit, ou plutôt, capable d'imiter comment un autre parle malgré :

- Des différences physiologiques entre locuteurs
- L'opacité de l'objet parole.

Nos premières considérations ont donc consisté à montrer que l'imitation était un phénomène complexe puisque la situation favorisant son apparition est constituée d'une triade où un observateur (1) perçoit les mouvements (2) produits par un modèle (3) ; chaque élément de cette triade étant potentiellement un objet complexe. Ce n'est qu'à partir de cette situation initiale où (1) (2) & (3) sont (ou ont été) présents qu'est susceptible d'émerger une production imitative chez (1). En effet, comme l'indique Nadel (2005, p. 342) :

> L'imitation n'est pas une réponse obligée, comme l'est le fait de fermer les yeux lorsque la lumière est trop intense, ou de tourner la tête dans la direction d'un son.

Jusqu'à maintenant, nous n'avons pas réellement tenté de donner de définition précise de l'imitation : la tâche est ardue à cause de la variété que nous n'avons eu de cesse d'évoquer. De plus, il se greffe souvent dans les descriptions de l'imitation sa forme (par exemple, en termes de temporalité) ou sa fonction (par exemple, apprentissage ou communication). Aussi sommes-nous assez tentés de nous rallier, provisoirement, à une définition très ouverte de

l'imitation[9] qui devra toujours pouvoir être incluse dans la définition des différents types d'imitation :

Un sujet **produit de l'imitation** quand il reproduit, à dessein ou inconsciemment, tout ou partie d'un comportement (et/ou de ses caractéristiques) perçu chez un autre sujet, de manière à ce qu'un tiers puisse percevoir la production de l'imitateur comme ressemblant à celle du modèle.

Nous excluons volontairement de cette définition, toute notion de forme ou de fonction que nous réserverons à la description d'utilisations particulières de l'imitation. Le terme de « comportement » est non restrictif : il désigne toute activité humaine.

1.2 Contrastes d'intentions et définitions de l'imitation

Notre propos précédent s'attachait à définir la situation nécessaire à l'émergence des comportements imitatifs dans leur plus simple manifestation. Il s'agit à présent de mettre en perspective la triade imitative –modèle, objet, imitateur– dans divers phénomènes afin de définir au moyen de contrastes successifs différents degrés d'imitations. Ces processus incluent, dans leurs descriptions, des comportements imitatifs dans le sens que nous venons de définir.

Nous faisons ce choix pour contrebalancer un effet pervers de la multiplicité de définitions des comportements imitatifs. Comme le note Mitchell (2002, p. 442) :

> Reproduction of another's facial gestures by infants using familiar actions cannot be imitations, as these are novel or learned behaviors; […] replication of a care giver's eye blink by an orangutan using its finger to push down its eyelid cannot be imitation as the same motor patterns are not used; yet both are so obviously imitations that something must be wrong with these definitions.

Ainsi, plutôt que de chercher à définir ce que pourrait être la quintessence de l'imitation –une imitation « vraie »[10]–, il nous semble plus raisonnable d'intégrer ce qui semble relever de la sphère de l'imitation à des phénomènes mieux circonscrits ou mieux décrits comme le mimétisme ou la mimésis. Certains de ces phénomènes feront écho à des comportements d'imitation en parole auxquels nous ferons référence ultérieurement.

[9] Une conception assez proche de celle de Brass & Heyes (2005)
[10] Pour Guillaume (1925), l'imitation vraie est l'imitation différée, révélatrice d'un apprentissage réussi.

1.2.1 Le Mimétisme ou l'absence d'intention imitative

De la notion de mimétisme nous retiendrons trois acceptions :

- le mimétisme batésien (ou müllesien[11]) qui résulte de l'évolution des espèces
- le mimétisme « passif[12] », comme capacité à se fondre dans le milieu au moyen de traits biologiques
- le mimétisme « actif », comme imitation machinale (inconsciente) des pairs de notre environnement

Le premier (qui ne nous intéresse pas particulièrement), a trait aux stratégies d'adaptation et de survie de certaines espèces animales qui en viennent à ressembler aux membres d'une autre espèce animale pour échapper à leurs prédateurs :

> Conventionally speaking, a Batesian mimic is a palatable prey that gains protection through its resemblance to an unpalatable species (the model). (Speed, 1993, p. 471)

Le second correspond à l'adaptation de certaines espèces animales à leur milieu au moyen de leur pelage, par exemple afin d'être moins visible de leur proie (ou de leur prédateur, selon les points de vue) :

> Ce phénomène est la faculté <u>inconsciente</u> [13]que possèdent les divers êtres de s'accommoder et d'imiter le milieu dans lequel ils sont appelés à vivre. (Morau, 1893)

Le troisième serait le phénomène/processus par lequel des sujets qui se côtoient en viennent à se ressembler[14], c'est-à-dire, à agir de la même manière, à avoir les mêmes tics, les mêmes vêtements, *etc*.

Parmi ces trois exemples, le point commun réside dans la notion de ressemblance (à une autre espèce, au milieu, à ses congénères). Cependant, seul le dernier type « actif » va plus particulièrement nous intéresser car il met en jeu des comportements imitatifs au sens où nous les avons définis précédemment. Cela dit, il semble que le second type de mimétisme évoqué puisse –métaphoriquement– être considéré comme une conséquence de ce

[11] Le phénomène touche surtout les insectes : une espèce comestible exploite sa ressemblance avec une autre espèce non comestible pour échapper à ses prédateurs.

[12] Nous pensons aux espèces qui se fondent dans leur milieu naturel, certaines espèce comme les caméléons peuvent se fondre dans plusieurs milieux de manière active.

[13] Nous soulignons : il s'agit du terme clé concernant le mimétisme, il sera opposé à l'intentionnalité pour d'autres formes d'imitation

[14] « Tel père, tel fils », « Les chiens ne font pas des chats ». Deux dictons qui illustrent le mimétisme dans la famille, à notre sens

« mimétisme actif » pour l'être humain : à mesure des ajustements inconscients de comportement, le sujet se trouverait mieux adapté à son milieu. C'est là une fonction importante du mimétisme, une réponse –parmi d'autres– à la question de l'adaptation (Baudonnière, 1997, pp. 7–14)

Quand il est observé chez l'être humain, le mimétisme est parfois associé aux comportements de foule (Baudonnière, 1997, pp. 37–41) comme la propagation des mouvements de panique, les hochements de tête des fans de rock ainsi que le partage des émotions. Par ailleurs, Baudonnière note une nette tendance des êtres humains à manifester leur appartenance à un groupe au moyen de leur ressemblance, que ce soit le cas des adolescents :

> Il est parfois surprenant de voir jusqu'à quel degré de précision vont les adolescents dans leurs imitations posturales et vestimentaires. Les mêmes mimiques, la même casquette portée de la même manière […] tout est respecté dans les moindres détails. (1997, pp. 67-68)

ou bien le cas des adultes :

> Dans de nombreux partis politiques, on assiste à de véritables imprégnations de la part du leader conduisant ses collaborateurs à utiliser les mêmes tics de langage, les mêmes intonations[15] au cours de leurs interventions publiques. Pendant toute une période au cours des années 80, ce fut particulièrement net pour les membres du Parti communiste français, vis-à-vis de leur secrétaire général, Georges Marchais. (*Ibid*, pp 69-70)

Ces derniers exemples donnés par Baudonnière[16] indiquent que les êtres humains, de manière non contrôlée, tendent à se conformer les uns aux autres. Selon les circonstances, cette conformation semble pouvoir avoir lieu également au niveau de leur comportement parolier. Par ailleurs, certains chercheurs émettent l'hypothèse que cette tendance serait naturelle et incontrôlable : c'est ce que mettent en exergue Chartrand & Bargh (1999, p.893) en parlant d'effet « caméléon » .

> The *chameleon effect* refers to non-conscious mimicry of the postures, mannerisms, facial expressions, and other behaviors of one's interaction partners, such that one's behavior passively and unintentionally changes to match that of others in one's current social environment.

[15] Nous soulignons, cela a un écho favorable pour notre matière à venir.
[16] Dont nous n'avons pu personnellement nous rendre compte, bien que nous puissions observer dans les médias actuels, le même type de comportement

La manière dont cet effet est défini s'accorde avec l'absence d'intention que nous soulignions ci-avant ; cet effet pourrait alors –avec précaution– être rangé sous l'étiquette du mimétisme.

En d'autres termes, le mimétisme pourrait être décrit comme la production non contrôlée d'imitations de la part d'un sujet, en réponse aux spécificités de son environnement interactif, afin de s'y adapter au mieux. En ce sens, le mimétisme se distingue d'autres productions imitatives par l'absence de préméditation du sujet dans la production de ses réponses. *A priori*, le mimétisme servirait donc un processus d'adaptations continuelles au milieu.

Appliqué au comportement parolier du sujet, la convergence phonétique paraît être une manifestation de comportement mimétique. Cette dernière a été définie par Giles *et al*. (1991, p. 7) comme

> A strategy whereby individuals adapt to each other's behaviors in terms of a wide range of linguistic-prosodic-nonverbal features including speech rate, pausal phenomena and utterance length, phonological variants, smiling, gaze and so on.

Si la définition de Giles et collègues n'inclut pas l'absence d'intention de la part du sujet (la dimension sociale de la communication pouvant être révélatrice d'une intention implicite), il semble toutefois que la convergence phonétique –ou un phénomène approchant– puisse apparaître dans des situations naturelles, sans motivation sociale de la part du locuteur. C'est ce que notent Sancier & Fowler (1997, p. 422) en rapportant des exemples initiateurs de leur recherche : le cas d'une étudiante brésilienne aux USA dont le père se plaint que sa manière de parler est trop « explosive » quand elle rentre en vacances ; ou le cas d'un Britannique vivant depuis longtemps aux Etats-Unis chez qui les amis du Royaume Uni trouvent un accent américain. Ce genre d'effet peut être révélé par les paradigmes expérimentaux utilisés pour faire émerger le phénomène de convergence phonétique, *i.e.* en mettant les interlocuteurs dans des situations naturelles de communication.

En étant plus prudent, nous devrions dire que la convergence phonétique semble relever du mimétisme car elle ne paraît pas être le résultat d'une intention maîtrisée du sujet. En effet, nous pensons que parler n'est jamais un acte gratuit : il devrait toujours y avoir une intentionnalité sous-jacente définissant la position du locuteur ainsi que son comportement. Comme l'imitation (et nous détournons ci-après la phrase de Nadel, 2005), la convergence phonétique n'est pas un phénomène obligé : Giles et collègues décrivent également des effets

de divergence (1991, pp. 8-9). Ils ont été provoqués expérimentalement dans des situations de communication interculturelle, où un interlocuteur se déclarait ostensiblement négatif envers un des groupes représentés ; cela pouvait être un Anglais dénigrant les Gallois (Bourhis & Giles, 1977), ou un Wallon mésestimant des membres de la communauté Flamande (Bourhis, Giles, Leyens, & Tajfel, 1979). Dans ces situations, la divergence se manifestait par des effets inverses à la convergence : amplification des marques d'appartenance au groupe (soit, une absence de partage interactif du comportement) et rupture de la communication avec l'autre. Ainsi, les effets de convergence/divergence pourraient être le résultat d'intentions sous-jacentes à la situation de communication et sortir alors du cadre de la définition du mimétisme. Pourtant, l'apparition de ces effets étant collatérale à la communication parolière, nous sommes tentés –conceptuellement– de les y rattacher.

Finalement, le mimétisme pourrait se situer à l'extrémité d'un continuum des comportements imitatifs, continuum défini par le degré d'intentionnalité du sujet dans sa production. En effet, parmi les différentes définitions que nous avons données, les mots clefs étaient « inconscient » et « non intentionnel ». Bien que le mimétisme « actif » implique des actions imitatives au sens où nous les avons définies précédemment, il est fréquent que ces types de comportements ne soient pas assimilés –dans la littérature– à de l'imitation. Pour de nombreux chercheurs, le critère d'intentionnalité apparaît comme majeur dans la définition de l'imitation. Ce critère d'intentionnalité offre donc un premier pivot nécessaire pour articuler différents types de comportements imitatifs.

1.2.2 La Mimésis ou l'acte imitatif comme intention

A l'absence d'intentionnalité contenue dans le mimétisme nous opposerons la Mimésis proposée par Donald (1993, p. 168) :

> Mimetic skill or mimesis rests on the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic. These mimetic acts are defined primarily in terms of their representational function. Therefore, reflexive, instinctual, and routine locomotor acts are excluded from this definition […]

Le concept de mimésis, dans la définition de Donald, se situerait à l'opposé du mimétisme sur un continuum d'intentionnalité. En effet, les actions mimésiennes[17] dépassent largement la

[17] Qui relève de la Mimésis. Nous utilisons ce mot par opposition à mimétique, qui doit relever du mimétisme

simple réitération machinale d'un comportement car elles sont supposées induire –dans les termes de Donald– « *l'*invention[18] *de représentations intentionnelles* » (1993, p. 169) ; alors que les actes mimétiques, qui sont inconscients et provoqués de manière fortuite (ils dépendent de l'environnement) n'ont pas de rôle de représentation mais servent l'adaptation (Baudonnière, 1997, pp. 7–41). Ainsi, la définition de la mimésis se dégage principalement du mimétisme par :

- le déclenchement maîtrisé de l'acte, qui n'est alors plus une réponse inconsciente à la pression de l'environnement
- le fait que l'acte mimésien peut être spontanément produit, comme la représentation qu'un sujet se fait de l'objet reproduit
- le caractère figuratif de l'acte, soit le fait qu'il a pour but de représenter un événement ou de faire référence à une caractéristique de l'environnement

En ce sens, une reproduction mimésienne d'un objet ajoute une signification à la simple reproduction motrice qu'en constituerait une imitation basique. Delvaux *et al*. (2005, p. 1) estiment d'ailleurs que « *l'acte* [mimésien[19]] *par excellence est le mime* » : un mime sait qu'il mime un objet, en propose à l'audience sa propre représentation[20], qui a pour but de figurer un objet connu de l'audience et/ou de le lui faire deviner. En outre, le mime n'est pas conçu comme un geste linguistique : les éléments linguistiques sont volontairement exclus –par jeu ou contrainte– de la production du mime.

Bien que Donald (1993, p 168) exclue le linguistique et les gestes linguistiques de la mimésis, cette dernière peut être envisagée par le point de vue des sciences du langage, plus particulièrement, du point de vue phonologique :

> Dans le contexte de la phonologie, la mimésis est considérée comme la capacité à développer, à amplifier, voire à réguler, la variation phonétique. Les représentations continuellement renouvelées par l'action mimétique sont autant de patrons phonétiques pour la réalisation des représentations phonologiques. (Delvaux, Demolin, & Soquet, 2004, p.1)

Dans cette optique, la mimésis phonétique se distinguerait de la convergence phonétique par le choix délibéré de la variation phonétique produite. Dans le cas de la mimésis, en fonction du contexte et des représentations phonétiques à disposition, le locuteur reproduirait

[18] Souligné par l'auteur : « the *invention* of intentional representations ».
[19] Delvaux et collègues utilisent le terme « mimétique »,
[20] Et selon les contextes, l'audience la reconnaît ou non, ce sur quoi repose certains jeux autour du mime.

délibérément ce qui lui semble adéquat pour interagir au mieux, alors que dans le cas de la convergence phonétique, le changement comportemental ne serait pas ou peu maîtrisé.

La notion de représentation, très prégnante dans la définition de la mimésis et absente dans la définition du mimétisme, appelle naturellement à faire référence à la cognition. Cette dimension est un caractère définitoire de la Mimésis, aussi bien pour Donald (1993) qui estime que la mimésis a été le premier niveau de représentation cognitive de l'espèce humaine, que pour Delvaux *et al*. (2005, p.126) qui s'appuient sur la conception de Donald :

> La mimésis est une aptitude cognitive de modélisation et de coordination des schémas moteurs qui s'appuie sur la faculté d'imitation en y ajoutant une dimension représentationnelle.

Ces derniers font d'ailleurs de cette aptitude cognitive « *une des bases essentielles du développement des espèces évoluées* » [*i.e.* les primates et les êtres humains] (Harmegnies, Delvaux, Huet, & Piccaluga, 2005, p. 312) qu'ils nomment alors « *compétence mimétique*[21] ». Outre la fonction d'apprentissage de cette compétence qui permettrait au sujet de reproduire les modèles, ces chercheurs la distinguent d'une simple capacité imitative par la possibilité du sujet de se représenter les événements et de les transposer de manière adéquate aux différentes situations qu'il rencontre. Comme ils le notent :

> Ramenée à l'oral, cette dichotomie [entre simple imitation et mimésis] oppose, schématiquement, le psittacisme -producteur de sons calqués sur la parole humaine- à l'usage du langage qui conduit à l'émission fonctionnelle des sons de parole dans une visée communicative. (Harmegnies *et al.*, 2005., p. 313)

Enfin, il nous semble intéressant de noter –à l'opposé du mimétisme et de la convergence phonétique– que la manière dont procèdent certains imitateurs professionnels semble relever de la mimésis : les travaux de J. Révis (2013) que nous évoquerons plus largement dans les chapitres ultérieurs de ce document, semblent aller en ce sens. Il semblerait en effet que des sujets puissent à loisir imiter des personnages célèbres à partir de la représentation qu'ils se font de leur voix.

Envisager l'imitation sous l'angle de la mimésis revient à considérer un phénomène plus large que la simple réitération d'un comportement, qui, en elle-même est déjà un sujet complexe. Cela se retrouve dans la formulation du problème de correspondance. Une fois la

[21] De manière récurrente, nous aimerions utiliser « mimésienne »

capacité du sujet à imiter établie (pour l'espèce humaine, on ne se pose plus la question), ce problème peut –dans le contexte linguistique– être posé de différentes manières selon le point de vue qu'on adopte :

- quelles sont les structures physiologiques qui permettraient d'établir cette correspondance ?
- comment se développe la capacité à estimer que deux objets partagent les mêmes caractéristiques ?
- comment se manifeste aux niveaux phonologique et phonétique la correspondance ?

1.2.3 Synthèse : du mimétisme à la mimésis

En présentant le mimétisme d'une part et la mimésis d'autre part, nous avons souhaité apporter des éléments de définition des imitations. En effet, l'apparente uniformité du concept « imitation » est trompeuse et nous espérons avoir montré que la position du sujet vis-à-vis de l'objet imité, en termes d'intentionnalité, doit permettre de distinguer différents types de comportements imitatifs. L'imitation produite dépend-elle d'une réaction « instinctive[22] », d'une réponse d'adaptation (intentionnelle ou non), ou bien d'une volonté maîtrisée ? L'intention (ou son absence) que nous prêtons au sujet est un élément fondamental dont nous devrons tenir compte en abordant plus finement les comportements imitatifs en parole.

Parallèlement, nous considérerons que mimétisme et mimésis occupent respectivement les extrémités d'un continuum des imitations, basé sur la notion d'intention. Nous tenterons par la suite de classifier les différents types de comportements paroliers relevant de l'imitation en fonction de l'intention supposée du sujet, ce que chaque chercheur propose à sa manière. Par exemple, Merlin Donald note :

> A distinction can be made between, mimicry, imitation and mimesis. Mimicry is literal, an attempt to render as exact a duplicate as possible. Thus exact reproduction of a facial expression or exact reproduction of the sound of another bird by a parrot, would constitute mimicry. […] Imitation is not so literal as mimicry; the offspring copying its parent's behavior imitates, but does not mimic, the parent's way of doing things. […] Mimesis adds a representational dimension to imitation. It usually incorporates both mimicry and imitation. (Donald, 1993, p168-169)

Cette distinction, que Harmegnies *et al*. (2005, p.313) reprennent également, ne devrait pas être appliquée uniformément aux comportements paroliers. Dans certains contextes, il peut

[22] Nous mettons des guillemets, car le mot est sujet à controverse, nous ne souhaitons pas y entrer.

être malaisé de déterminer l'intention du sujet quand il imite une caractéristique parolière de son interlocuteur. Il est légitime de se demander si, chez un individu, un changement phonétique est dû à un effet de mimétisme (adaptation à l'environnement), un effet de convergence (une adaptation à l'interlocuteur) ou à une imitation mimésienne (une réitération pensée, expressément voulue, d'un comportement).

Une distinction de la sorte nous semble également pertinente pour essayer de définir *a priori* les comportements que nous pourrions essayer de provoquer expérimentalement. Nous avions probablement une telle distinction en tête lorsque les consignes destinées au recueil de notre premier corpus d'imitation ont été rédigées (Nocaudie, 2012). Il s'agissait alors d'impliquer de plus en plus explicitement les sujets expérimentaux dans un comportement imitatif de la parole au fil de trois tâches de perception et production :

- Une simple réitération de l'énoncé entendu (1)
- Une imitation explicitement demandée de l'énoncé entendu (2)
- Une exagération de l'énoncé entendu. (3)

Au cours de ces trois tâches, doit-on dire qu'il y a eu un glissement d'un usage potentiellement mimétique du langage (1) à un usage éventuellement mimésien du langage (3) ?

Parallèlement, cette distinction nous semble importante pour classer *a posteriori* des comportements paroliers relevés dans la littérature en fonction des caractéristiques de leurs situations d'émergence rapportées dans la littérature. Notre première réflexion sur le statut de la convergence phonétique (entre mimétisme et imitation) reflète ce genre d'approche : il semble compliqué de la classer dans les comportements relevant du mimétisme car un but sous-jacent semble motiver le changement du comportement ; cependant, la convergence peut sembler relever du mimétisme car elle est produite sans contrôle explicite du locuteur, probablement pour s'adapter à la situation. *De facto*, faire ce genre de distinction doit nous permettre d'éviter l'écueil des descriptions uniformes des phénomènes, trop globalisantes.

Par ailleurs, il nous faut faire preuve de prudence : ces considérations théoriques sur la nature des comportements imitatifs ne doit pas occulter ce qui nous intéresse fondamentalement dans les comportements imitatifs en parole : la capacité des locuteurs à percevoir des énoncés et à en reproduire –au moins partiellement– des éléments. Toutefois, cette réflexion est nécessaire pour prendre du recul sur les différents phénomènes que nous considérons et ceux que nous aimerions provoquer dans nos pratiques expérimentales. Dans

l'absolu, nous pourrions nous satisfaire d'une définition *a minima* de l'imitation, comme la simple réitération d'un geste, cependant nous pensons que l'étude de l'imitation en parole doit aussi prendre en compte la dynamique imitative de nos comportements (voir chapitre 2).

En guise de conclusion sur ce point, la notion d'intention nous a permis de proposer quelques distinctions parmi les différents types de comportements imitatifs décrits dans la littérature. Le mimétisme semble relever des processus d'adaptation des espèces à leurs milieux. Cette première manifestation des comportements imitatifs peut se définir comme une adaptation non contrôlée (parfois biologique) de l'individu en réponse à son environnement. A l'autre extrémité du spectre définissant l'imitation en relation avec l'intention, nous avons proposé de considérer la mimésis, telle que définie par Donald (1993). Ici, l'intention du sujet est intrinsèquement liée à la définition de ce comportement, puisqu'il implique que le sujet produisant de la mimésis représente (pour lui ou pour les autres) un événement.

1.3 Buts, imitations, émulations et temporalité

La notion de but constitue un autre facteur habituellement décrit dans la littérature sur l'imitation. Quand elle est associée à l'imitation, cette notion permettrait de distinguer l'achèvement de l'action, du processus fait pour l'accomplir. En d'autres termes, il est possible de copier, littéralement, l'action d'un autre (ce que fait l'élève qui n'a pas appris quand il louche sur la copie de son voisin) mais il est également possible d'essayer de reproduire le résultat d'une action sans utiliser le procédé mis en œuvre par le modèle pour accomplir son action (ce que peut faire un copiste reproduisant un tableau ou un orang-outan utilisant ses doigts pour faire faire un clin d'œil à sa paupière). De cette nouvelle distinction, émerge la définition des comportements d'émulation qui différent de l'imitation « classique ».

De plus, comme le terme « imitation », le terme « émulation » peut avoir différentes acceptions, ce que notent Call & Carpenter (2002, p. 212) :

> One mechanism that, in our opinion, has begun to suffer some of the same problems that have plagued the term imitation is emulation

Ainsi, il faudra que nous accordions quelques lignes à la définition de ce terme. Celui-ci pourrait en effet être utilisé opportunément pour décrire ce qui se passe chez un apprenant de L2 qui commence son apprentissage.

Enfin, il sera nécessaire de considérer la notion de but(s) durant la réalisation d'une action imitative : ceux-ci modèlent différents aspects de l'action des individus et d'un point de vue définitoire, servent à estimer quel type de comportement nous observons.

1.3.1 Définitions de l'émulation

Comme le notent Call et Carpenter (2002, p. 212), le mal dont pâtit le terme « émulation » est équivalent à celui affectant le terme « imitation ». Selon eux –et nous les rejoignons dans cette considération–, certains termes ont, au fil du temps, reçu tellement d'acceptions qu'il est maintenant difficile de les utiliser ou de les lire sans avoir de doute quant à leur signification précise :

> Some theoretical terms have become too broad or too narrow. For instance, the term *imitation* is used with different meanings by different researchers. While some researchers use it in a general way to denote copying behavior, others prefer to reserve it for those case in which the organism not only copies behavior but also acquires a novel behavior. (Call & Carpenter, 2002, p. 211)

Dans le langage courant, le terme « émulation » renvoie à un sentiment de rivalité, pouvant pousser des personnes à égaler ou surpasser leur entourage dans certains domaines. On y trouve donc une notion de similarité, puisqu'en regardant ou en étant en présence de l'autre, un sujet peut être amené à adopter une démarche similaire à l'autre. Cependant, le terme renvoie plus à la motivation apportée par l'autre qu'à une imitation.

Autre acception, issue cette fois du domaine informatique, l'émulation dans une définition assez générique du terme désigne le fait de simuler le fonctionnement d'une machine X sur une machine Y structurellement différente, afin de pouvoir utiliser sur la machine Y, des logiciels initialement prévus pour la machine X. Dans ce cadre, émuler un programme ou une machine a une forte connotation imitative, cependant, les moyens mis en œuvre pour faire comme l'autre diffèrent. Toutefois, nous ne retrouvons pas dans cette acception la notion de but de l'action (soit, son résultat) qui est prégnante dans la description proposée par Call & Carpenter (2002).

Faisant référence à la situation de « l'apprentissage social [23]» au sens de Whiten & Ham (1992, p. 248), c'est-à-dire la situation où « *B learns some aspect of the behavioral*

[23] Social learning

similarity from A», Call & Carpenter considèrent que l'observation d'un démonstrateur peut amener à obtenir des informations de trois ordres différents :

- Les **Buts**, qui correspondent à l'intentionnalité du modèle, ce qu'il cherche à produire
- Les **Actions**, qui équivalent aux *patterns* moteurs utilisés pour aboutir au but
- Les **Résultats**, qui sont les changements dans l'environnement effectivement produits par le modèle.

<table>
<tr><td colspan="3">Un sujet B…</td><td>Le sujet B produit une …</td></tr>
<tr><td rowspan="4">Comprend (et adopte) le But de A</td><td rowspan="2">Copie l'Action</td><td>Reproduit le Résultat</td><td>Imitation</td></tr>
<tr><td>Ne reproduit pas le Résultat</td><td>Imitation ratée</td></tr>
<tr><td rowspan="2">Ne copie pas l'Action</td><td>Reproduit le Résultat</td><td>Emulation du but</td></tr>
<tr><td>Ne reproduit pas le Résultat</td><td>Emulation du but</td></tr>
<tr><td rowspan="4">Ne comprend pas (ou n'adopte pas) le But de A</td><td rowspan="2">Copie l'Action</td><td>Reproduit le Résultat</td><td>Mimique</td></tr>
<tr><td>Ne reproduit pas le Résultat</td><td>Mimique</td></tr>
<tr><td rowspan="2">Ne Copie pas l'Action</td><td>Reproduit le Résultat</td><td>Emulation</td></tr>
<tr><td>Ne reproduit pas le Résultat</td><td>Emulation ratée (ou pas d'apprentissage social)</td></tr>
</table>

Tableau 1 : Trois sources d'informations en imitation d'après Call & Carpenter (2002, p. 217) : La table se lit de gauche à droite. Les trois sources d'information sont représentées dans chaque colonne. En fonction de ce que l'observateur perçoit de la situation, ce dernier peut qualifier le type de comportement imitatif observé.

Call & Carpenter se servent de cette triade **But / Action / Résultat** pour distinguer différents types de comportements. Nous reproduisons dans le Tableau 1 ces différents comportements imitatifs.

Cette proposition de Call & Carpenter permet donc –dans un cadre où les comportements imitatifs doivent être vecteurs d'apprentissage– de distinguer différents types de comportements, en analysant à la fois le comportement du modèle et celui de l'imitateur.

Les trois aspects de cette proposition renvoient à des notions évoquées précédemment. L'intention de l'imitateur a été décrite comme un élément définitoire des comportements imitatifs. Nous retrouvons cette notion d'intentionnalité cette fois associée au démonstrateur par la notion de But. Si l'imitateur comprend l'intention du démonstrateur, il peut chercher à l'adopter, mais cela n'implique pas obligatoirement qu'il reproduise les Actions et/ou le Résultat ; on parlera alors d'émulation du but, *i.e.* reproduire un résultat par un autre moyen, plutôt que d'imitation. De même, un sujet peut ne pas comprendre l'intention du démonstrateur, mais toutefois copier les Actions et aboutir au Résultat. Nous retrouvons dans cette dernière distinction la notion de « mimique » évoquée par Donald (1993, pp. 168–169) dans le cas où l'Action est copiée, et la notion d'émulation si l'Action n'est pas copiée…

Si ces distinctions permettent de classer les comportements imitatifs en fonction des sources d'informations fournies par le démonstrateur et ce qu'elles provoquent dans le comportement du sujet imitant, il convient toutefois de rappeler que cette analyse se limite à des cas d'apprentissage social. C'est pourquoi Call & Carpenter (2002, p. 215) proposent de limiter l'usage de ces termes à des situations précises, où la notion d'apprentissage social s'applique ; par exemple, l'étude du développement de l'enfant. Cependant, leur proposition permet de mieux appréhender une assertion comme celle de Kuhl & Meltzoff (1995) disant que « *Imitation can result in non-identical, but fonctionnaly equivalent of the modeled behavior.* ». En effet, en fonction de son appréhension du But (compris et intégré ou non), un sujet qui ne copierait pas l'Action, produirait alors de l'émulation (ou de l'émulation du but), puisqu'il en reproduirait le Résultat !

La notion de but chez le modèle peut donc rejoindre la notion de l'intentionnalité du sujet. Adjoint aux Actions et au Résultat, le cadre de travail proposé par Call & Carpenter permet de décrire différentes variantes des comportements imitatifs en apprentissage social. Bien que les comportements imitatifs chez l'être humain adulte n'entrent pas systématiquement dans ce cadre, la proposition de Call & Carpenter est utile pour qui veut définir et décrire les imitations.

1.3.2 Buts et sous buts dans l'accomplissement d'une action

Quelques lignes plus haut, nous avons repris une citation de Kuhl et Meltzoff (1995) indiquant qu'une imitation (ou émulation) pouvait être différente du modèle tout en conservant la même fonction, soit amener au même résultat. Ainsi, il est possible d'aboutir au même point en adoptant des procédures différentes. Cela indique que dans une série d'actions, certaines peuvent être oubliées car elles ne sont pas pertinentes pour l'atteinte du résultat final. En d'autres termes, le but ultime d'une série d'action (atteindre un objet dans une boîte munie d'un verrou) peut être décomposé en sous-buts (saisir la boîte, faire jouer le verrou, ouvrir le couvercle, et saisir l'objet).

A partir de cette situation très simple, nous pouvons proposer deux scénarii :

- Le verrou de la boîte est enclenché (1)
- Le verrou de la boîte n'est pas enclenché (2)

Dans le scénario (1), pour atteindre l'objet dans la boîte, il est nécessaire de faire jouer le verrou, car ce dernier est fermé. Cependant, dans le scénario (2), manipuler le verrou est une action superflue. Ce deuxième scénario peut être propice à faire émerger un comportement *–a priori–* spécifique à l'être humain décrit comme une *over-imitation*. C'est-à-dire :

> […] in what have been known as *overimitation*, young children copy the explicit actions of an adult demonstrator even when a more efficient method of achieving the demonstrated outcome is available and when copying the adult's actions results in failure to bring about the demonstrated outcome. (Nielsen & Tomaselli, 2010, p.1)

Si un démonstrateur montre à un enfant la boîte du scénario (2), et l'ouvre devant lui après en avoir touché le verrou, il serait alors fort probable que l'enfant produise toutes les actions perçues, y compris celles superflues pour l'atteinte du but.

L'exemple que nous avons choisi est encore trop simple pour montrer la portée du phénomène. Lyons *et al.* (2007) ont entraîné des enfants âgés de 3 à 5 ans à identifier les parties pertinentes ou non-pertinentes de nouvelles séquences d'actions accomplies par un adulte manipulant des objets du quotidien ; par exemple : récupérer un jouet dans une jarre en plastique après en avoir tapoté le côté trois fois avec une plume. Les enfants ont ensuite observé l'adulte montrant des séquences d'actions équivalentes (dont la pertinence des différentes actions pouvait être évaluée) sur de nouveaux objets. Malgré leur entraînement,

ces enfants ont par la suite reproduit les actions superflues, y compris quand il leur a été demandé de ne copier que les actions essentielles à l'atteinte du but.

Pour Arbib (2012, p. 195), les jeunes enfants tendent ainsi à focaliser plus particulièrement leur attention sur les actions spécifiques plutôt que sur le but d'une séquence, lorsqu'il s'agit d'imiter immédiatement. Arbib base son propos sur les travaux de Horner et Whiten (2004) qui comparent la manière de faire des chimpanzés et des jeunes enfants.

Dans leur étude, les chimpanzés avaient tendance à ignorer les actions non pertinentes à l'accomplissement du but, tandis que les enfants reproduisaient des actions non nécessaires, voire complètement arbitraires. Pour Horner et Whiten, la différence entre les membres des deux espèces réside dans le focus attentionnel porté sur les intentions, les actions et/ou les résultats. Les chimpanzés auraient tendance à se focaliser sur le résultat uniquement, favorisant alors des stratégies d'émulation ou d'émulation du but (en nous plaçant dans la perspective de Call et Carpenter, 2002), tandis que les jeunes enfants seraient plus focalisés sur les actions, étant alors soit dans un comportement de mimique, si l'aboutissement ne les intéresse pas, soit dans un comportement d'imitation, s'ils ont saisi l'intention du démonstrateur. Sans entrer plus dans le détail, cette tendance à l'*over-imitation* serait typique de l'espèce humaine, et aurait également une dimension assez universelle, d'après Nielsen et Tomaselli (2010), qui ont testé des populations très variées : des enfants du Bush Sud-africains comme des enfants de grandes métropoles australiennes.

Il convient toutefois de noter que ce comportement n'est pas systématique, puisque les enfants ne copient pas aveuglément tout ce qu'ils observent. Cela dit, il semble que l'adulte comme démonstrateur ait un effet sur la production des enfants, comme s'ils étaient poussés à copier leurs actions (Arbib, 2012, p. 196). Bien que ce comportement puisse être vu comme inadapté, il est plutôt considéré que l'*over-imitation* a un rôle non-négligeable dans la transmission d'actions et de patterns complexes, dont le développement et la transmission de la culture humaine (Arbib 2012., p. 197).

Afin de clore ces quelques considérations sur l'*over-imitation*, il semble opportun de souligner les travaux de McGuigan, Makinson, & Whiten (2011) sur la comparaison entre deux groupes d'enfants (âgés de 3 et 5 ans) et des adultes. Ils trouvent que les enfants de 5 ans reproduisent plus d'actions non-pertinentes que ceux de 3 ans. Par ailleurs, les adultes, malgré leur capacité de jugement plus élevé, dans le contexte expérimental de leur étude, reproduisent également les actions non-pertinentes avec une fidélité plus grande que celle des

enfants[24] (*Ibid.*, p11). Que l'adulte soit capable de copier des séquences de gestes complexes de manière machinale nous paraît naturel, qualifier sa conduite d'*over-imitation* nous semble plus périlleux. En effet, il nous semble possible que l'effet obtenu par ces chercheurs soit dû à la tâche et à la consigne données aux sujets. La tâche consistait à récupérer un objet dans une boîte, au moyen d'un outil. La boîte avait deux ouvertures, une sur le dessus qui ne permettait pas d'accéder à l'objet et une autre sur le côté qui, elle, permettait de l'atteindre. Les actions du démonstrateur étaient toujours présentées dans le même ordre :

- Non pertinentes qui se concentraient sur le dessus de la boîte (toucher le dessus, l'ouvrir, insérer l'outil par le dessus et taper trois fois dans la boîte)
- Puis pertinentes (ouvrir la porte latérale, insérer l'outil et en ressortir l'objet).

Il était ensuite dit au sujet : « *c'est votre tour maintenant* »[25]. En regard des discussions que nous venons d'avoir sur l'importance du but dans le type d'actions imitatives mise en œuvre par des sujets, il nous semble que la consigne est biaisée : s'il était explicitement dit de récupérer l'objet, les adultes auraient-ils reproduit toute la séquence –actions non-pertinentes comprises– avec autant de fidélité ? Comme très peu d'informations était données aux sujet-imitateurs, nous pensons que les intentions du démonstrateur ont été mal perçues. *De facto,* nous pensons qu'il faut prendre avec plus de prudence la propension de l'adulte à produire de l'*over-imitation* : ce pourrait n'être qu'une stratégie pour faire face à une situation dont le but est incertain.

En revanche, cette étude apporte une indication intéressante car Mc Guigan et collègues (2011) ont recherché l'effet des démonstrateurs adultes ou enfants sur les sujets-imitateurs : tous leurs groupes de sujet ont exprimé une préférence en copiant plus les actions des démonstrateurs adultes. Leur analyse les conduit à penser que les imitateurs, auraient tendance à reproduire plus volontiers les actions du démonstrateur si son niveau d'expertise est reconnu. Allant plus loin, ils ajoutent :

> A more specific version of the above hypothesis is that participants saw the irrelevant actions performed by an adult model as more 'intentionnal' than those performed by a child: perhaps, a seemingly irrelevant action performed by an adult should be worthwhile to copy as it is likely that they intended to perform it, whereas a child may more likely have produced the action accidentally. (McGuigan *et al.*, 2011, p. 12)

[24] We found that rather than growing out of such a tendency, adults continued to copy in an unselective fashion, adopting irrelevant action with an even higher fidelity than the children.
[25] Now it's your turn (McGuigan et al., 2011)

En substituant le terme « enseignant » au terme « adulte », il semble que cette remarque puisse s'appliquer à la situation de correction phonétique : si l'enseignant arrive à être perçu comme un expert de son domaine, les apprenants pourraient s'adonner plus volontiers à cette pratique !

1.3.3 Emulation et L2, construction hypothétique

Précédemment, la notion d'émulation a été envisagée dans le domaine des apprentissages sociaux, comme relative au but du démonstrateur, ses actions et les résultats de ces actions. Toutefois, nous avions également évoqué (au moment de définir l'émulation, p. 45) le sens informatique du terme « émulation », en indiquant que l'enjeu de l'émulation informatique était de faire fonctionner des logiciels (*softwares*) sur un matériel (*hardware*) qui n'est à l'origine pas prévu à cette fin. Cette idée nous offre l'occasion de développer une métaphore sur l'apprenant de langue étrangère. Celui-ci, en tentant de parler une L2, émulerait de manière imparfaite le *software* « L2 » sur la base des ressources dont il dispose pour faire fonctionner son *software* « natif » (sa L1) : des processus cognitifs dédiés au traitement de sa L1 et le *hardware* qui les sous-tendent, *i.e.* sa structuration cérébrale et ses organes phonatoires.

La situation évoquée par une telle métaphore n'est pas sans rappeler le concept de surdité phonologique en L2, énoncé par Polivanov (1931, pp. 79–80), et illustré par Troubetzkoy (1939, p. 54)[26] au moyen de la métaphore du crible phonologique

> à travers [du]quel passe tout ce qui est dit. Seules restent dans le crible les marques phoniques pertinentes pour individualiser les phonèmes. […] L'homme s'approprie le système de sa langue maternelle. Mais s'il entend parler une autre langue, il emploie involontairement pour l'analyse de ce qu'il entend le « crible phonologique ».

Cette métaphore du crible phonologique –explication potentielle de la source des erreurs phonétiques en L2– se trouve développée par ailleurs dans des publications plus actuelles qui tentent d'établir un modèle de la perception de la parole.

Par exemple, Kuhl *et al.* (2008, p. 982-985) décrivent un modèle de perception en partant de données développementales. Leur modèle, le *Native Language Magnet theory – expanded* (*NLM-e*), postule un glissement d'un état initial où le nourrisson est potentiellement

[26] Edition originale : 1939

capable de discriminer toutes les unités phonologiques de différentes langues à un état final chez l'adulte où les représentations phonologiques sont suffisamment stabilisées pour ne pas être affectées par une courte période d'exposition à une L2 (Kuhl *et al.*, 2008, p. 989-991). Ainsi, au fil de son développement, l'individu organiserait son espace cognitif perceptuel pour faciliter le traitement des sons de sa ou de ses langues, au détriment des autres langues, qu'il ne parle pas ou bien auxquelles il n'est pas exposé. Au cours de sa maturation, l'individu mettrait alors en place un système optimisé pour son *software* originel.

En termes de *hardware*, ces chercheurs documentent le fait que l'exposition à des sons nouveaux conduirait au développement de structures neuronales, de manière assez rapide chez l'enfant, moins rapide chez l'adulte dont la plasticité cérébrale serait moindre. Pour ce dernier, une exposition prolongée serait nécessaire. A défaut de nouvelles structures leur permettant de catégoriser efficacement les sons de la L2, les apprenants subiraient alors l'effet des « aimants perceptifs », tels que décrits par Kuhl et collègues. Cet effet lisserait la perception des sons : il équivaudrait, comme le note Nguyen (2005, p. 432) à « *la réduction des différences au sein des catégories* ».

En admettant que l'apprenant subisse les effets d'un tel système de perception de la parole (ce que, sans prendre position en faveur d'un modèle particulier de la perception des langues, nous sommes tentés de croire), il semble acceptable de penser que le traitement d'une langue inconnue sera problématique : puisque le système perceptif serait calibré pour une langue X, le traitement de la langue Y sera mal aisé.

Nous avons ici pris l'exemple de la catégorisation des sons de parole, mais la recherche sur le bilinguisme pose de nombreuses questions sur le traitement cognitif des L2, allant dans le sens de notre métaphore initiale autour de l'apprenant comme émulateur « informatique ». Par exemple, le traitement mémoriel de certains aspects linguistiques varierait en fonction de leur statut de L1 (mémoire procédurale) et L2 (mémoire déclarative) (Köpke, 2007, pp. 17–18), faisant pencher la balance en direction de notre métaphore initiale : pour parvenir à utiliser une L2, l'individu doit mettre en place des stratégie cognitives hors de son système habituel.

Considérer ainsi qu'un apprenant de L2 émule –au sens informatique– la langue nouvelle sur la base du système de sa L1 revient simplement à tenter de comprendre ses difficultés et leur origine. Ces difficultés se manifestent par des erreurs à tous les niveaux linguistiques, très fréquentes au début de son apprentissage et qui parfois subsistent –voire

persistent– alors que l'apprenant a atteint un niveau de locuteur indépendant dans la L2, *i.e.* quand celui-ci-émule efficacement la L2.

Du point de vue de l'enseignant, se pose alors la question d'aider au mieux l'apprenant à émuler la L2. Pour ce faire, l'enseignant doit être conscient des limites et des contraintes du système de l'apprenant. Au niveau phonétique et en termes d'imitation, cette perspective offre un regard neuf sur les *outputs* erronés et systématiques des apprenants en L2, et rappelle à l'enseignant que son action doit entraîner l'apprenant à élargir son mode de fonctionnement. Il revient alors à l'enseignant de faire les choix pertinents pour tenir compte de la situation de l'apprenant.

1.3.4 Imitations : temporalité et contexte

Lorsqu'on les lie à l'imitation, les notions de temporalité et de contexte sont connexes. En effet, pour distinguer imitation « décalée », « immédiate » et « différée », la présence ou l'absence du sujet produisant le comportement modèle peut servir de critère pour ranger les imitations « immédiate » et « décalée » dans une première catégorie (en présence du modèle), et l'imitation « différée » dans une seconde (en son absence) (Baudonnière, 1997, pp. 110–111).

Souhaitant garder notre ligne de conduite pour cette toute première partie –soit, tenter de définir les comportements imitatifs indépendamment de leurs fonctions–, nous devrons nous limiter à indiquer que :

- l'imitation immédiate est décrite comme se produisant simultanément à l'observation du comportement,
- l'imitation décalée se produit en présence du modèle, mais avec un court temps de latence entre l'observation du comportement et sa reproduction
- l'imitation différée (ou « vraie », selon Guillaume, 1925) a lieu en l'absence du modèle, ultérieurement à l'exposition au modèle.

Considérer le critère de temporalité, ou de moment de la performance (Messum, 2007a, p. 99) dans l'observation des imitations peut nous aider à décrire et à essayer de comprendre ce que font les imitateurs. Le moment de production d'un comportement peut par exemple nous informer sur la capacité d'un sujet à le reproduire à loisir, ou non. Dans le premier cas, peut-on considérer que l'imitateur a suffisamment intégré le comportement, qu'il en a des

représentations mentales (qu'elles soient des patterns d'activation moteurs ou des représentations phonologiques) ?

Par exemple, dans le cadre d'une situation de correction phonétique, entendre un apprenant parvenir à prononcer la cible une fois ou deux lors d'une séance, ne garantit pas que l'apprenant la produise toujours correctement dans les jours suivants. En effet, la production réussie d'un son, une fois et dans un contexte précis, ne conduit pas automatiquement à la réussite de la prononciation de ce même son dans d'autres contextes phonotactiques, voire même d'autres contextes de communication.

Ceci étant dit, notre ligne de conduite initiale trouve ici sa limite : il devient impossible de gloser plus avant à propos de l'imitation en évitant d'évoquer ses fonctions. Il devient donc nécessaire que nous décrivions des comportements imitatifs en contexte, en tentant de leur attribuer une fonction.

1.4 Résumé des distinctions composant le spectre des comportements imitatifs

Afin de clore ce premier ensemble, nous nous proposons de rappeler les critères définitoires des comportements imitatifs et les grandes distinctions que nous avons pu faire dans la description de leur variété au moyen de quelques figures de synthèse.

Les partis impliqués dans les comportements imitatifs sont à définir comme une triade dont l'interaction a pour produit l'imitation. En effet, l'imitation produite par un sujet résulte de son contact avec les deux autres membres de cette triade, comme nous l'indiquons dans la Figure 1.

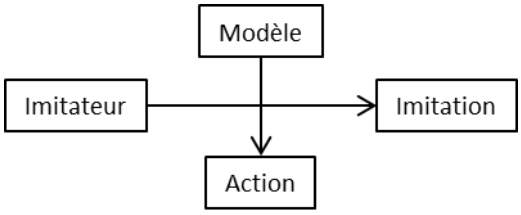


Figure 1 Situation d'imitation réduite à l'essentiel de ses composantes

Nous avons ensuite souligné que la description du type de comportement imitatif pouvait être nuancée par l'implication du sujet dans son acte, c.a.d. son intentionnalité. Nous avons alors décrit les comportements relevant du mimétisme, qui ont une dimension automatique et non intentionnelle, puis les comportements relevant de la mimésis, qui sont produits de manière intentionnelle et transmettent la représentation d'un événement. Nous illustrons cette distinction en Figure 2.

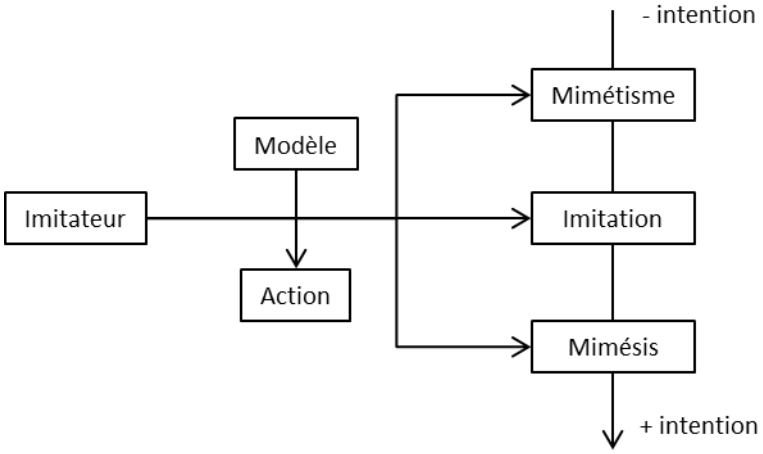


Figure 2 : Distinction des comportements imitatifs en fonction de l'intentionnalité prêtée au sujet lors de sa production.

Dans la suite de notre propos, nous avons considéré les notions de but, d'action et de résultat pour affiner nos descriptions en fonction des actes accomplis par le modèle, de leur compréhension par l'imitateur, et de la réussite de l'imitateur dans sa reproduction. Nous renvoyons le lecteur au Tableau 1 où cette distinction émulation/imitation proposée par Call & Carpenter (2002, p. 217) est illustrée.

Enfin, nous avons constaté que le moment de la reproduction peut avoir lieu en même temps que la démonstration (immédiate), juste après et en présence du modèle (décalée) ou bien plus tard et en l'absence du modèle (différée). En fonction de ce moment, les définitions du comportement imitatif peuvent changer, de même que la ou les fonctions qui peuvent lui être attribuées.

Ces fonctions sont l'objet des pages suivantes.

2 Imitation(s) : formes et fonctions

Afin d'éviter les désagréments d'une multitude de termes aux significations souvent voisines, il nous a semblé plus prudent de décortiquer les composantes de la situation d'imitation pour en saisir certaines nuances. Il convient à présent de contextualiser l'imitation produite par des individus, pour en observer la manifestation (le type d'imitation) et les fonctions (pourquoi est-elle produite). Outre sa fonction d'apprentissage, abondamment débattue dans la communauté des éthologues et des psychologues développementalistes, les comportements imitatifs pourraient aussi servir des fonctions d'adaptation à l'environnement, de représentation et de communication. Ceci dit, nous verrons également que les comportements imitatifs sont susceptibles d'être produits par l'être humain à tous les âges de sa vie… Eminemment naturels, les comportements imitatifs pourraient être vus comme un ensemble de stratégies, permettant à l'être humain de faire face aux situations qu'il rencontre dans son environnement.

2.1 Imitations & fait sociaux

Du fait même de sa définition, l'imitation est un comportement lié à la vie en société : en effet, la production d'un comportement imitatif résulte de l'observation par un sujet Y, d'un comportement produit par X. Dans un premier temps, nous nous intéresserons plus particulièrement à des actions de l'individu résultant du contact avec le groupe dans lequel il évolue. Par rapport aux descriptions du début de ce chapitre, nous nous situerons plutôt dans la partie mimétisme du spectre des comportements imitatifs. Nous rapporterons en premier lieu deux expériences de psychologie sociale, traitant de la pression exercée par le groupe sur l'individu, puis nous nous intéresserons au langage et à la notion de sociolecte.

2.1.1 Pressions des groupes & Conformation sociale

Précédemment, nous avons associé mimétisme et mouvements de foule en rapportant des exemples donnés par Baudonnière (1997, pp. 37–41) à propos des hochements de têtes des fans de *heavy metal*, ou encore de la propagation des mouvements de panique. Quand le groupe agit de conserve, l'individu isolé, sans avoir de raison rationnelle, aurait tendance à

intégrer son action à celle du groupe pour se synchroniser avec lui. Ces phénomènes massifs sont souvent le résultat d'un déclencheur, qui va entraîner à sa suite le reste du groupe :

- Dans le règne animal, la migration des gnous constitue un exemple intéressant à observer. Lorsqu'ils se trouvent confrontés à une rivière, les animaux hésitent à la traverser jusqu'à ce que le premier animal s'engage dans la traversée. A sa suite, le troupeau entier (de quelques dizaines à milliers d'individus…) se met à suivre de manière irrépressible.

Figure 3 La migration des gnous : traversée de la rivière Mara, photo d'Andréi Gudkov

- Pour prendre un exemple à propos de l'humain, il suffit de s'imaginer dans une salle de spectacle. Le bref moment de silence précédent les applaudissements (et le suspense qui accompagne cet instant : y aura-t-il applaudissements ?) est souvent brisé du fait d'individus isolés avant que les autres ne suivent.
- A l'inverse, une anecdote datant de l'ère stalinienne du communisme nous a été rapportée par Soljenitsyne[27] : lors d'un congrès du Parti Unique, la terreur instaurée par Staline était telle que personne n'osait s'arrêter d'applaudir afin de montrer sa ferveur (et d'éviter de se retrouver sur la liste des membres du parti à expurger). Qui oserait donner le signal en arrêtant d'applaudir ?

[27] L'Archipel du Goulag

Ces comportements sont parfois métaphorisés comme résultant de l'effet de « mouton de Panurge », en référence à un épisode du *Quart Livre* de Rabelais : Panurge jeta un de ses moutons à la mer, ce qui eut pour effet d'entraîner à sa suite (et à sa perte) tout le reste de son troupeau. Cet épisode illustre l'instinct grégaire (et le décrie). Cependant, s'associer au groupe peut être une stratégie efficiente pour améliorer sa survie : dans le cas des gnous, traverser la rivière en groupe rend plus compliquée la prédation aux crocodiles…

Par ailleurs, ce type d'effet illustre la propension de l'être humain à agir en synchronie. Les applaudissements, par exemple, ont au début un caractère assez chaotique, mais il n'est pas rare que le groupe finisse par taper des mains à l'unisson. Il devient d'ailleurs difficile de se désynchroniser du groupe et de faire exprès d'applaudir à contre temps de la majorité (nous relatons notre expérience personnelle…).

L'effet du groupe sur le comportement de l'individu a intéressé les psychologues. Une expérience intéressante au sujet des effets du groupe sur le comportement de l'individu a été produite par Solomon Asch dans les années 50. Durant cette expérience, un sujet naïf participait à un test en compagnie d'autres personnes, des complices de l'expérimentateur. La tâche consistait en une prise de décision : à chaque *trial*, il était présenté aux sujets deux images. Sur la première, un segment unique était dessiné et sur la seconde, trois autres segments, dont un avait la même longueur que le segment de la première image. Les sujets devaient dire quel segment parmi les trois de la seconde image correspondait au segment de la première image (voir Figure 4). Le sujet naïf, répondant en dernier, subissait une pression sociale de la part des complices, qui répondaient parfois en choisissant une mauvaise réponse. Malgré la flagrance de l'erreur, le sujet naïf tendait à se conformer à l'avis du groupe dans 37% des cas (Asch, 1955)[28]. Certes, il ne s'agit plus ici d'imitation au sens où nous l'entendions précédemment, *i.e.* une reproduction de gestes corporels, cependant un tel comportement laisse supposer que, dans certains cas, l'être humain adapte son comportement pour se conformer au groupe.

[28] A noter que l'étude originale est parue en 1951, nous indiquons cette référence que nous avons pu consulter. Une vidéo de l'expérience est disponible ici : goo.gl/US7Cc8

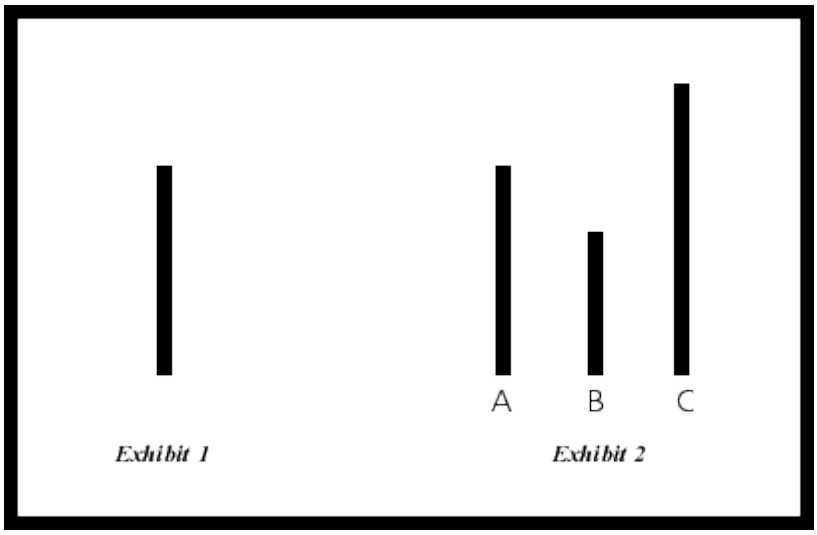


Figure 4 : Exemple de stimulus des études d'Asch sur la décision et l'opinion

Cependant, cet effet du groupe sur l'individu a donné lieu à des *spin-off* assez amusants : l'expérience de l'ascenseur en constitue un bon exemple. Une personne –le sujet de l'expérience– entre dans un ascenseur occupé par plusieurs autres personnes (des complices de l'expérimentateur). A mesure que les complices changent de manière synchrone de position, se tournent vers le même côté ou lèvent la tête, le sujet de l'expérience tend à se conformer au comportement dominant et suit le mouvement des complices. Ce comportement a été le ressort comique de caméras cachées des années 50[29], mais des étudiants en psychologie de l'université de l'Arkansas se sont essayés à le reproduire assez récemment.

De tels résultats peuvent faire froid dans le dos, si l'on pense aux implications sociales que pourraient avoir cette propension de l'humain au conformisme. Cependant, ces exemples nous semblent à propos pour faire ressortir que :

- La production de comportements imitatifs peut avoir un côté irrépressible, voire irrationnel et/ou automatique.
- Ce type de comportement peut constituer une réponse d'adaptation à son contexte de production.

En effet, dans le cas de l'ascenseur, on peut penser que le sujet naïf respecte un usage implicite de la situation, qui indiquerait que tous les sujets présent dans ce lieu doivent regarder la même direction. Dans le cas de l'étude de Solomon Asch, il peut sembler plus

[29] Caméra cachée originale : goo.gl/hrvMR1 ; et une redite : goo.gl/mqI3k0

facile (ou confortable) au sujet de rejoindre l'avis dominant que de contester le point sur lequel tous les autres sujets semblent être tombés d'accord.

En nous référant aux définitions des comportements imitatifs précédentes, nous pourrions alors estimer que les situations dont nous venons de parler réfèrent à un phénomène de mimétisme social, dans la mesure où ces sujets se conforment à leur milieu, peut-être dans un but d'intégration au groupe.

2.1.2 Imprégnations mutuelles et usages du langage

Dans la première partie de notre exposé, nous avons brièvement rapporté des phénomènes d'imprégnations linguistiques dans les partis politiques, en faisant référence au Parti Communiste de l'époque de Georges Marchais. Nous pouvons difficilement en dire plus sur ce temps que nous n'avons pas connu… Cependant, les communicants politiques actuels, dans leur recours systématique aux éléments de langage, produisent (de manière certes très consciente) une parole mimétique (dans son contenu lexical et pragmatique).

Les éléments de langage sont ces formules toutes faites que l'on retrouve dans la bouche de plusieurs membres d'un même parti politique, à différents moments d'une même journée. D'une part, cela permet aux communicants politiques d'avoir des réponses et des arguments à apporter sur des sujets qu'ils ne maîtrisent pas. D'autre part, cela donne une impression de corps au public qui les écoute : en utilisant les mêmes éléments de langage, les politiques incarnent le même discours et ce faisant sont identifiés comme membre, qui de la majorité gouvernementale, de l'opposition, de tel ou tel parti. Ainsi, faire comme l'autre, s'imiter mutuellement, remplirait dans ce cadre une fonction identitaire.

De même, l'usage du langage qui est fait par certains groupes sociaux renvoie à cette fonction identitaire : par exemple l'usage d'argot au sein d'un même groupe. Gadet (2007) définit l'argot comme « *un lexique parasite non standard, accompagnant généralement un usage phonique et grammatical populaire* ». Originellement, l'argot remplit deux fonctions :

- Cryptique : dans le but de ne pas se faire comprendre par les locuteurs qui ne parlent pas l'argot
- Identitaire : afin que les locuteurs utilisant un argot particulier puissent se reconnaître entre eux.

Actuellement, il semblerait que la dimension cryptique de l'argot se réduise au bénéfice de la fonction identitaire , ce que notent Goudaillier (2002) ou Gadet (2007) en faisant référence au parler des jeunes de cités :

> L'usage de la langue qu'ont [les jeunes des cités] est adapté à des pratiques communicatives de solidarité entre pairs […] dont ils apprécient la valeur identitaire et cohésive (reconnaissance entre membre du groupe – *nous* / exclusion des autres – *eux*). (Gadet, 2007, p125)

Ainsi, partager un même lexique, accompagné de spécificités phonétiques (/r/ glottalisé, par exemple), prosodiques (déplacement de l'accentuation), syntaxiques ou morphologiques remplit cette fonction d'identification au groupe à un niveau linguistique, au même titre que ce que notait Baudonnière (1997, pp. 67-68) à propos des tenues vestimentaires au sein des groupes de jeunes.

Les exemples que nous venons de citer dans les points précédents relèvent de comportement imitatifs assez spontanés, et ainsi, pourraient être classés sous l'étiquette du mimétisme. La variété de ces exemples laisse supposer que l'être humain (et de nombreux autres êtres vivants) tendent à se conformer aux actions de leur groupe. Cette conformation aurait une fonction identitaire (pour le langage et l'habillement), mais aussi une dimension survivaliste (dans les applaudissements des membres du parti, ou la propagation des mouvements de foule), ou simplement une dimension d'adaptation immédiate à ce qu'il convient de faire sur le moment (les applaudissements à la fin du spectacle).

2.2 Imitations & communication

Une large part de la communication humaine se fait via le langage, que celui-ci soit oral, signé ou écrit. Toutefois, communiquer au moyen du langage suppose que les interlocuteurs partagent de manière suffisante un même code linguistique : il faut pouvoir émettre des messages mais aussi pouvoir les décoder. C'est le schéma classique de la communication proposé par Jakobson (1963).

Ceci dit, il est important de remarquer que la compétence linguistique de l'être humain n'est ni acquise dès la naissance, ni figée une fois acquise :

- L'enfant ne naît pas sachant parler
- L'individu mature peut voir sa compétence linguistique décliner progressivement (à cause de maladies dégénératives), ou de manière abrupte (suite à un AVC entraînant une aphasie, par exemple).

Par ailleurs, le nombre important de langues différentes laisse également ouverte l'émergence de situations où deux individus sains et matures ne partagent pas le même code linguistique et se retrouvent de fait confrontés à des difficultés de communication.

Ici, nous décrirons donc quelques situations où les sujets ne disposent plus ou pas encore de leur accès au langage et dans lesquelles les comportements imitatifs revêtent alors des fonctions communicatives : les imitations simultanées chez les jeunes enfants, l'imitation chez l'enfant autiste et le recours à l'imitation dans les crises d'un sujet apraxique.

2.2.1 Imitations et communications entre jeunes enfants

Jusqu'au milieu du 20^{e} siècle, le nouveau-né et les bébés jouissaient de peu de considérations de la part des psychologues, quant à leur faculté d'interaction avec leur environnement. Nadel (2011, p. 8) rappelle en effet que le bébé a longtemps été assimilé à « *un tube digestif ouvert aux deux bouts* » toutefois capable d'émettre des informations sporadiques supposées informer son environnement de son état (faim, soif). Depuis, les psychologues développementalistes ont largement dépassé cette vision archaïque des capacités du très jeune enfant, pour montrer que celui-ci percevait son environnement et était capable de le comprendre et d'agir dessus.

Certains travaux sur l'imitation et le développement des jeunes enfants sains et présentant un trouble du spectre autistique (Nadel, 1986, 2005, 2011; Nadel et al., 1999; Nadel & Potier, 2002a, 2002b) semblent particulièrement appropriés pour comprendre le développement de l'imitation ainsi que son rôle. Nous ne débattrons pas de thématiques liées à la présence ou l'absence d'imitation néonatale (Meltzoff & Moore, 2005; Zazzo, 1957) qui reste parfois un sujet controversé, ni au développement proprement dit des comportements imitatifs de l'enfant. Nous nous intéresserons aux aspects de l'imitation traitant de la communication entre enfants de 2 à 4 ans.

Nous pouvons en effet nous demander ce qu'il se passe lors d'interactions entre des enfants dont la capacité linguistique est insuffisante pour communiquer en tant que sujets

parlants. Les enfants restent-ils isolés, ou bien parviennent-ils à avoir des interactions fructueuses les uns avec les autres ? Afin de comprendre la nature de ces interactions, Nadel (2011, pp. 58–67) propose deux *set up* expérimentaux différents, où des dyades d'enfants de 2 à 4 ans ont l'opportunité d'agir sans intervention directrice d'un adulte.

Le premier *set up* expérimental est constitué d'une pièce où plusieurs objets en double exemplaire sont disposés à l'intention des enfants (par exemple : deux chapeaux de cowboy, deux poupées, deux parapluies, *etc.*). Un adulte est présent dans la salle, mais son rôle est de ne pas s'intéresser à ce que font les enfants pour ne pas les distraire de leur interaction. Avant d'entrer dans la salle, l'expérimentateur indique simplement aux enfants qu'on leur a préparé une surprise et qu'ils peuvent prendre les objets. Ainsi, les enfants ont l'alternative de faire des jeux solitaires avec les objets ou bien d'entrer en contact avec leurs semblables.

Dans ce contexte, Nadel a observé que les enfants en groupes ne collectionnaient pas les objets identiques, mais tendaient au contraire à porter chacun le même objet que leur semblable : ses résultats indiquent une fréquence de port d'objets identiques bien supérieures à celle des ports solitaires (d'un facteur 4) et un temps de port des objets identiques occupant 70% du temps de la situation. Par ailleurs, Nadel note que lorsqu'un objet est abandonné par un enfant qui se saisit alors d'un nouvel objet, les autres enfants faisaient de même dans un délai très réduit. Enfin, concernant ce qui est accompli par les enfants portant des objets semblables, il s'agissait d'imitation immédiate : un comportement synchronique.

Le second set up expérimental reprend le même principe que le premier (les enfants sont dans une pièce et font ce qu'ils souhaitent) mais les objets disposés dans la salle sont des exemplaires uniques. Dans cette configuration, les interactions entre les enfants étaient plus courtes, sans imitations et parfois conflictuelle. Pourtant, les mêmes dyades testées dans un contexte avec des objets doublés se comportaient différemment. Ainsi, il semble que l'imitation synchrone d'enfants ayant environ 30 mois ait comme médiateur des objets identiques en tous points : une simple ressemblance du média ne suffirait pas (par exemple, une canne et un parapluie).

Pour observer l'effet de l'âge des enfants sur la production d'imitation synchrone, Nadel a par ailleurs testé des dyades d'enfants de 12 à 48 mois. Ses résultats indiquent que ce type d'imitation est transitoire : ce comportement devient de plus en plus fréquent de 12 à 30 mois, où il y a un pic d'utilisation de l'imitation gestuelle synchronique, puis sa fréquence d'utilisation décline jusqu'à 48 mois. A ce moment, du développement, les enfants peuvent

interagir par des moyens verbaux et n'ont donc plus besoin de l'imitation synchrone, qui est remplacée par des jeux de faire semblant (Nadel, 2011, pp. 66–67).

L'analyse proposée par Nadel pour certains aspects de ces imitations synchrones propose de les contextualiser comme un apprentissage des communications verbales à venir. En effet, la manière dont se dérouleraient ses interactions synchroniques entre jeunes enfants présenterait des aspects similaires à ceux de la conversation. Nadel (2011, pp.69-81) décrit que ces interactions constituent une action conjointe (comme la conversation), présentant une alternance entre les enfants (comme des tours de parole) et le partage de thèmes communs, évoluant au fil de l'interaction.

Malgré l'aspect synchronique de ces comportements entre jeunes enfant, ces interactions ne peuvent se résumer à un dialogue de sourd. Non seulement les enfants font la même chose en même temps, mais ils surveillent également le fait que leur semblable suive les actions : ils s'attendent. Par ailleurs, Nadel (2011, p. 73) remarque qu'il y a une alternance dans l'origine des thèmes de l'interaction : tour à tour, l'enfant imite et est imité par son semblable. D'autres codes rappellent la conversation : le mécanisme d'invite à l'interaction synchrone décrit par Nadel (*Ibid.*, p.74-75) en constitue un exemple frappant. Lorsqu'un enfant veut inviter l'autre à une interaction synchronique, il tend à son semblable un objet identique à celui qu'il a en main. Si l'autre enfant accepte le don, la conversation d'imitation synchronique peut démarrer. En cas de refus, celle-ci est momentanément délayée.

Dans le cas des jeunes enfants, la production d'imitations synchrones revêt donc une fonction de communication. Par ailleurs, l'organisation de cette communication est proche de celle des interactions verbales. Enfin, ce type de comportement imitatif disparaît du répertoire de l'enfant quand celui-ci acquiert un moyen plus efficace pour interagir avec ses semblables. Par conséquent, cette fonction communicative de l'imitation est transitoire chez l'enfant.

Cependant, l'enfant avant 4 ans n'est pas le seul type de sujet humain présentant des difficultés d'accès au langage. Il semble également possible qu'une telle fonction de l'imitation se manifeste chez des sujets plus âgés, privés durablement ou ponctuellement, de leur capacité à communiquer verbalement.

2.2.2 Imitation et communication chez le sujet mature : « Brother John », jeux de mimes et langues étrangères

Le sujet adulte, également, peut subir l'effet d'une perte de sa capacité à communiquer. Nous en envisagerons brièvement deux cas :

- Une perte ponctuelle ou durable liée à une pathologie (par exemple, les aphasies)
- Une difficulté de communication en lien avec la barrière de la langue (soit, une L2).

Dans ces situations, le recours au pouvoir évocateur des gestes (ceux qui peuvent avoir une dimension symbolique) pourrait alors servir de substitut au langage.

Donald (1993, pp. 82–89) rapporte par exemple le cas de « *Brother John* », initialement étudié par Lecours et Joanette (1980). *Brother John* est décrit par ces derniers comme un homme de 50 ans, membre d'un ordre religieux et souffrant depuis plus de 25 ans de crises épileptiques passagères. Ces crises temporaires étaient plus ou moins durables et fréquentes :

- Courtes (1 à 5 minutes, plusieurs fois par jour)
- Longues (1 à 11 heures, environ une fois par mois).

Durant ces crises, les facultés linguistiques de *Brother John* étaient partiellement diminuées, sans que ces facultés mémorielles ne soient atteintes. Ainsi, le religieux pouvait relater *a posteriori* l'expérience vécue pendant ses crises. Lecours et Joanette ont décrit que les longues crises épileptiques de Brother John manifestaient les symptômes d'une aphasie globale, avant que le religieux ne recouvre graduellement ses facultés linguistiques au fil du spectre des symptômes aphasiques (par exemple, en passant par une phase de jargon néologique).

Malgré la sévérité de ces longues crises, Brother John restait tout de même conscient de son environnement, et pouvait accomplir, dans une certaine mesure, certains actes sociaux. Par exemple, il est rapporté dans cette étude de cas un épisode helvétique, où, arrivant dans une ville inconnue, Brother John sent se manifester les symptômes annonciateurs d'une longue crise. Voici ce que Donald rapporte :

> He found himself at the peak of one if his seizure as he arrived at his destination, a town he had never seen before. He took his baggage and managed to disembark. Although he could not read or speak, he managed to find a hotel and show his medic-alert bracelet to the concierge, only to be sent away. He then found another

> hotel, received a more sympathetic reception, communicated by mime[30], and was given a room. (Donald, 1993, p. 84)

Lorsque nous définissions les comportements imitatifs en fonction de degré d'intentionnalité, nous avions souligné que la pratique du mime requérait un accès aux représentations et que cette pratique avait un pouvoir évocateur particulier, dû à sa force symbolique. Dans le cas de Brother John, le mime basé sur la reproduction des formes et des mouvements de l'environnement ainsi que la situation dans l'espace sert de système palliatif de communication sans dimension linguistique. Donald (Ibid. p85) poursuit d'ailleurs :

> […] both gestural ability and practical knowledge were intact. He could imitate or reproduce on demand a wide variety of gestures. […] All of this was achieved in the absence of visual or oral language, and in the absence of internal speech as well.

Pour Donald, le cas de Brother John est primordial pour montrer l'indépendance entre les représentations sémantiques du monde et le langage symbolique. Dans notre travail, ce cas nous intéresse pour la force évocatrice du mime et des imitations, susceptibles de supporter des fonctions de communication en l'absence du langage symbolique.

Nous avons d'ailleurs –du moins, nous pouvons le penser– tous ressenti le potentiel communicatif des gestes et des imitations dans certaines situations simples de la vie courante :

- Lors d'un voyage dans un pays dont la langue nous est inconnue, il peut arriver que nous cherchions notre chemin ou que nous souhaitions savoir les éléments qui composent le plat que nous nous apprêtons à commander au restaurant. De ces situations, il peut s'ensuivre un ballet gestuel pour arriver à ce que les interlocuteurs se comprennent (de notre expérience, il y a souvent des surprises malgré tout).
- Les jeux reposant sur une contrainte de nos facultés habituelles sont assez courants. Tout le monde connaît colin-maillard, où la vue est occultée, les exercices littéraires de l'Oulipo comme La Disparition de Georges Pérec, où l'usage de la lettre « e » est interdit ou bien encore les jeux de mimes. Dans ces derniers, les joueurs choisissent de se priver de leur moyen habituel de communication afin de faire deviner les uns aux autres des concepts, des mots, des personnages ou des expressions.

[30] Nous soulignons

Les comportements imitatifs, produits par des enfants comme par des adultes, peuvent avoir une fonction communicative. Celle-ci peut être une étape normale dans le développement de la communication (les enfants) ou bien une stratégie palliative à une perte (pathologique ou non pathologique) de l'accès au langage symbolique.

2.3 L'imitation comme stratégie d'apprentissage au long de la vie

La fonction d'apprentissage de l'imitation est probablement la plus abondamment observée dans la littérature. Nous avons en effet souligné au moment des définitions que cette fonction d'apprentissage était indissociable, pour certains chercheurs, de la notion même d'imitation.

En ce qui nous concerne, nous avons nuancé notre point de vue de sorte à isoler les différents comportements imitatifs et leur fonction. Ceci dit, nous pensons également que les comportements imitatifs, dans certains contextes, peuvent être les médiateurs de l'apprentissage. Dans cette partie, nous nous arrêterons un court instant sur l'apprentissage gestuel (dans les sports ou l'artisanat, par exemple) avant de nous intéresser aux situations d'apprentissage du langage chez l'enfant et d'une langue étrangère chez le sujet mature.

2.3.1 Imitations gestuelles et apprentissages

Le langage est un outil formidable pour transmettre du sens, et de nombreux apprentissages sont prodigués au moyen d'instructions explicites. Les recettes de cuisine constituent un exemple frappant d'enseignements délivrés par ce médium. Par exemple, la recette d'œuf mollet indique les étapes suivantes :

a. Mettre de l'eau à bouillir et la saler.
b. Quand l'eau bout, y plonger doucement les œufs.
c. Laisser cuire cinq minutes à découvert.
d. Passer les œufs sous l'eau froide et écaler.

Une personne qui ne connaîtrait pas cette manière de faire les œufs (ô combien délicieuse) serait en mesure de réussir cette recette du premier coup. S'il la refait souvent, il est probable qu'elle se souvienne des différentes étapes de la recette et que l'apprentissage de la cuisson des œufs mollets ait été fructueux.

Ceci dit, il aurait également été possible de réaliser cet apprentissage au moyen d'imitations synchrones ou décalées : si le sujet suit un cours de cuisine et qu'on y réalise la recette, l'enseignant va lui montrer les étapes qui peuvent être reproduite, pendant la démonstration ou juste après. La recette de l'œuf mollet étant très simple, la recette peut suffire au sujet pour atteindre une compétence suffisante dans cette réalisation. Cependant, d'autres domaines des activités humaines nécessitent souvent notre aptitude à imiter comme médiateur de nos apprentissages.

Cette dernière remarque renvoie à une observation formulée par Brass & Heyes (2005, p. 489) dans l'introduction d'un de leurs travaux :

> Could you learn to tango by telephone? Maybe, but it would be much easier to learn by watching the steps than by listening to instructions.

Pour apprendre, il est parfois plus simple de s'en remettre à notre capacité de reproduire les gestes que nous observons que de suivre des instructions verbales. La facilité déconcertante avec laquelle nous sommes capables de reproduire les gestes pourrait être à l'origine de comportements humains très répandus (peut être planétaires).

Par exemple, dans la danse et la culture populaire occidentale des années 90, la chanson « *La Macarena* » est restée célèbre pour sa chorégraphie extrêmement accessible et qui a fait danser des millions de personnes sur les mêmes pas de danse. Personne ne leur a donné la recette : ils ont simplement imité ce qu'ils ont vu dans un clip vidéo, ce qu'a fait leur voisin au barbecue dominical ou le G.O. de leur club de vacances. Encore aujourd'hui, mettre ce morceau de musique dans une soirée de trentenaires peut potentiellement déclencher un phénomène d'imitation synchrone et immédiat…

Nous ne pourrions faire le tour de toutes les situations d'apprentissage où les comportements imitatifs sont impliqués, car elles sont trop nombreuses. Notons simplement que les activités humaines où le corps et le geste ont une grande importance sont souvent appropriées pour être transmises au moyen d'imitations, comme les danses et les sports, l'artisanat, certains jeux, les rites culturels…

2.3.2 Apprentissages linguistiques chez l'enfant et le sujet mature

Acquérir sa langue maternelle ou apprendre une langue étrangère constitue un enjeu d'adaptation : il s'agit pour celui qui apprend de pouvoir s'intégrer à son environnement et interagir avec les autres membres de l'espèce. On relève dans la littérature de rares cas d'être humain ayant appris une langue maternelle trop tard ou ayant refusé d'apprendre une langue étrangère.

Dans le premier cas, nous faisons référence aux enfants sauvages (par exemple, Victor de l'Aveyron) qui ont intéressé les chercheurs travaillant autour de la question d'une période critique pour l'apprentissage de la langue maternelle (ou des langues étrangères). Outre le rapport du Dr. Itard qui a tenté de socialiser l'enfant[31], son histoire est relatée par le film de François Truffaut, « *L'enfant sauvage* ».

Le second cas auquel nous faisons référence est relaté par l'écrivain et réalisateur Emmanuel Carrère : dans un reportage pour Envoyé Spécial « *Le soldat perdu* », le film « *Retour à Kotelnitch* » et le livre autobiographique Un roman russe, Carrère raconte son voyage sur les traces d'un prisonnier de guerre hongrois, interné dans l'hôpital psychiatrique de Kotelnitch en Russie. Il y est resté enfermé 55 ans, dans un isolement quasi absolu, sans vouloir parler russe. Il n'était, au début de son internement, nullement atteint de pathologies psychiatriques : son cas serait dû à son refus d'apprendre le russe et à « l'omission » du personnel soignant qui n'a pas cherché à savoir quelle langue il parlait.

Pourtant, si l'on se prête au jeu, il est possible de reproduire les sons d'une langue inconnue (même si cela peut rester imprécis). L'apprentissage de l'oral en classe de langue étrangère repose d'ailleurs grandement sur ce ressort : sans forcément comprendre le contenu des énoncés auxquels ils sont confrontés, les apprenants débutants essayent tout de même de reproduire des mots qui ne font pas encore partie de leur système. Ils produisent alors une imitation en parole.

Puren (1988, pp. 151–159) mentionne d'ailleurs dans son histoire des méthodologies l'existence d'une « méthode imitative » au début du vingtième siècle. Parfois aussi appelée « méthode d'imitations », elle est décrite comme faisant partie de la méthode directe. Le terme de méthode imitative faisait alors référence à l'acquisition de la langue par les enfants qui, dans les termes des didacticiens de l'époque rapportés par Puren (1988, p. 152),

[31] Mémoire et rapport sur Victor de l'Aveyron (1801-1806), BNF-Gallica

« *produisent une imitation acoustique pure* » des sons entendus dans leur entourage bien qu'ils ne les comprennent pas. Cette méthode essayait donc d'exploiter la capacité d'imitation humaine comme un médiateur de l'apprentissage.

D'après Puren, considérer la notion d'imitation dans l'apprentissage de la langue était à l'époque un moyen d'envisager les questions liées à l'enseignement de la prononciation. Puren, dans son Histoire des méthodologies cite un propos qui devrait intéresser le didacticien s'intéressant à l'enseignement de la prononciation comme le chercheur en sciences du langage :

> La méthode d'imitation a tort de croire l'oreille infaillible et qu'il suffise de bien parler la langue pour bien l'enseigner à des étrangers. F.M. Vipan, d'après Camerlynck (1907, p. 177)[32], cité par Puren (1988, p. 155).

Ce propos préfigure avec quelques décennies d'avance l'hypothèse de travail de la MVT selon laquelle l'apprenant serait trompé par sa perception : on y retrouve en filigrane des notions liées à la perception des L2, notamment la difficulté de catégorisation des sons d'une L2.

En termes d'imitation et d'apprentissage/acquisition des langues, il peut à présent être opportun de faire référence à la notion de crible phonique proposée par Billières (1988) afin de cheminer du parcours de l'apprenant de L2 à celui de l'enfant (Figure 5).

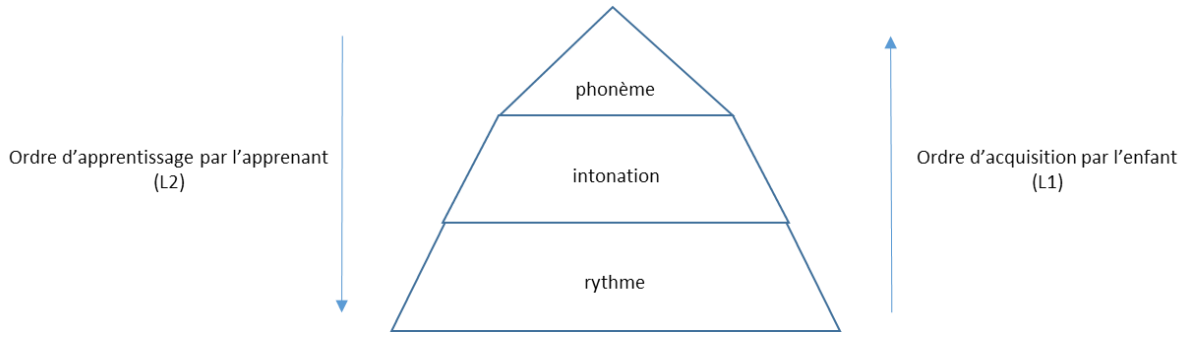


Figure 5 : Le crible phoniqueet les paliers représentants différentes couches de l'oral. D'après Billières (1988).

Ce schéma nous semble important pour résumer l'expérience de l'apprenant débutant une L2 en termes d'imitations phonétiques. Lors des premiers cours, exposé à un système linguistique nouveau, l'apprenant va chercher à comprendre ce qu'il perçoit. Très rapidement,

[32] Camerlynck G., (1907). « La phonétique et ses applications possibles à l'enseignement de la prononciation. Discussions. Extraits d'une communication faite au Congrès de l'Association belge des Professeurs de langues vivantes, tenu à Gand du 18 au 22 septembre 1906 ». *Les Langues Modernes* (5), pp. 172-178

son objectif consistera donc à segmenter les mots et à les reproduire, ce qui, dans une certaine mesure, est possible dès les premières minutes d'un cours d'une langue jusqu'alors inconnue. Ceci dit, il est probable que l'apprenant se raccroche aux phonèmes, car ils véhiculent le sens porté par le lexique. En reproduction, il est tout à fait probable que les structures profondes de l'oral (rythme et intonation) soient celles de la L1 de l'apprenant, sur lesquelles sont collés les mots de la L2… Métaphoriquement, l'apprenant produirait alors une imitation parolière de surface : il ne produirait que la partie émergée de l'iceberg, le premier palier (Figure 5).

A contrario, l'enfant ne prononce pas les mots d'emblée : les vocalisations sont au début limitées aux sons articulés aux extrêmes de l'appareil phonateur, comme /p/ /t/ /m/, et à quelques voyelles ouvertes (Renard, 1979, p. 16). Les combinaisons consonnantiques du français sont d'ailleurs produites avec précision assez tardivement par les enfants français dont les taux d'erreurs phonétiques restent parfois très élevés jusqu'à trois ans (Nocaudie, Köpke, Giraudo, & Calderone, 2015). Cependant, malgré ces erreurs de surface, les aspects rythmiques et prosodiques seraient acquis assez tôt par l'enfant. D'après de Boysson-Bardies (1996), les babillages des enfants de dix-huit mois seraient complètement imbibés par la langue de l'entourage de l'enfant. Nous retrouvons ici une trace des paliers évoqués dans la Figure 5 : l'enfant acquérant sa L1 fixerait en premier le niveau suprasegmental de sa langue (rythme et intonation) avant d'en acquérir le niveau segmental véhiculant le lexique et les concepts.

Disposant de plus de temps que l'apprenant pour acquérir sa L1 ainsi que la communication, l'enfant n'est par ailleurs pas court-circuité par un système linguistique préexistant. Ainsi, nous pourrions maladroitement dire que l'enfant peut être assimilé à un support vierge de toute inscription (l'enfant percevrait tout de même des sons avant sa naissance : voir Granier-Deferre & Schaal, 2005). Cette image suggèrerait que l'enfant a des possibilités initiales très étendues en ce qui concerne sa perception et son registre de vocalisations.

A notre sens, l'enjeu de l'acquisition de la L1 par l'enfant consiste à sélectionner des éléments pertinents pour pouvoir agir dans son environnement. En fonction de ce qu'il capte, celui-ci modèlerait son système perceptif de façon à faciliter le traitement des unités qu'il rencontre fréquemment. C'est par exemple ce que décrivent Kuhl *et al*. (2008) dans la théorie du *Native Language Magnet expanded*, quand ils proposent que les enfants forgent des liens entre perception et production des phonèmes par le biais de ses interactions sociales (soit ses

care givers). Pour ces chercheurs, l'exposition de l'enfant aux sons de sa langue et le fait qu'il les imite vocalement lui permettraient de faire un *mapping* audio articulatoire et donc de renforcer ses représentations phonologiques. Pour Kuhl *et al.* (2008) l'acquisition de la L1 aurait pour médiateur les comportements imitatifs produits par l'enfant.

Messum (2007a) adopte un point de vue différent : dans sa thèse, celui-ci développe l'idée que l'enfant ne ferait pas ses premières vocalises par imitation. Ce seraient les imitations du *care giver* qui permettraient à l'enfant de comprendre peu à peu comment les choses sont dites. En effet, dans sa perspective, le très jeune enfant n'aurait pas encore appris à imiter et ses premières vocalises seraient dues au hasard. En revanche, le *care giver* induirait le sens des vocalises de l'enfant et proposerait en retour une imitation orientée de ce que l'enfant aurait dit. Ainsi, Messum estime que ce ne sont pas les imitations de l'enfant qui jouent un rôle prépondérant dans les premiers développements langagiers de l'enfant, mais bien celles de l'adulte.

Quel que soit le point de vue adopté sur la question, nous pouvons tout de même mettre en exergue le rôle des comportements imitatifs (de l'enfant ou de l'adulte) dans l'acquisition et l'apprentissage de l'oral. Imiter et être imité, une dichotomie fondamentale pour Nadel (2005), a aussi une fonction dans nos apprentissages linguistiques.

Durant ce chapitre, nous avons décrit les différents types de comportements imitatifs, envisagés en fonction de l'intentionnalité qu'a le sujet de les produire, puis en fonction de sa compréhension des actions du modèle, sa reproduction du comportement observé et l'atteinte du but de l'action.

Par la suite, nous avons illustré au moyen d'exemples trois grandes fonctions des comportements imitatifs : l'adaptation au milieu ; la communication et l'apprentissage. Concernant la fonction d'apprentissage des comportements imitatifs, nous avons finalement considéré le cas de l'acquisition de la langue maternelle par l'enfant et de l'apprentissage d'une langue étrangère par le sujet mature.

Chapitre 2 : Des structures neuronales et cognitives de l'imitation à l'intégration des processus d'imitation en perception et production de la parole.

Nous avons précédemment décrit l'imitation comme un phénomène complexe, faisant intervenir trois composantes : un modèle, un comportement produit par le modèle et un sujet le reproduisant. Nous avons par ailleurs souligné diverses fonctions des comportements imitatifs. Ceux-ci servent, à différents niveaux, l'adaptation au groupe et à l'environnement, la communication ainsi que les apprentissages. Ainsi, l'imitation est un phénomène comportemental ayant une dimension sociale. Celle-ci est doublée d'aspects cognitifs : pour Petit & Pascalis (2009), les comportements imitatifs résultent de compétences cognitives inscrites dans nos contextes d'interactions avec autrui.

Jusqu'à maintenant, nous n'avions pas abordé (ou très brièvement) l'imitation au niveau de notre fonctionnement cognitif. Nous avions simplement résumé le problème de correspondance, qui pose la question de savoir comment un sujet parvient à déterminer quels muscles activer lorsqu'il observe un comportement qu'il souhaite reproduire (Brass & Heyes, 2005). En posant la question de la cognition et de l'imitation, se pose également la question des structures cérébrales de l'imitation et de leur fonctionnement. Si notre travail ne se focalise nullement sur les aspects neurologiques, il nous semble nécessaire de donner quelques éléments sur des sujets comme les neurones miroirs, ou les éventuelles zones cérébrales actives lors de l'imitation.

Nous souhaitons ensuite décrire une hypothèse relative à la formation de la capacité d'imitation à un niveau cognitif. Cette hypothèse, dite « généraliste », propose que les processus d'imitation soient mêlés à des mécanismes d'apprentissage par association et de contrôle des actions (Brass & Heyes, 2005, p. 489).

Enfin, ce chapitre est le lieu où nous souhaitons définir les comportements imitatifs en parole. Ce faisant, nous lierons ces comportements à des aspects de la cognition langagière qui nous semblent nécessaire à l'émergence de ces comportements. Ainsi, nous évoquerons tour à tour le stockage mémoriel des unités lexicales, la perception/production de la parole

comme un système dynamique et évolutif et l'action simultanée des processus de perception et de production de la parole.

1. Structures neuronales et dynamique de l'imitation

Se poser la question des structures neuronales de l'imitation revient à se demander quelles sont les structures qui permettent à un sujet de faire face au problème de correspondance. Un imitateur doit pouvoir percevoir les actions du modèle, puis activer les parties de son corps qui lui permettront d'accomplir l'action observée. Il semble évident que les zones du cerveau associées à la perception (visuelle et auditives) et la production (zones motrices) sont nécessaires à l'imitation. Par conséquent, le débat qui nous occupe doit porter sur les éventuelles structures cérébrales qui permettraient de faire le lien entre ces modalités de perception et de production. Le système de neurones miroirs semble être un candidat approprié pour résoudre ce problème de correspondance (Nehaniv & Dautenhahn, 2002, p. 56).

1.1 La question des neurones miroirs

Depuis leur découverte dans l'aire prémotrice F5 du cerveau de certains primates (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996), la question des neurones miroirs et d'un système équivalent chez l'être humain déchaîne passions et spéculations. Ces cellules cérébrales déchargent à la fois quand l'animal produit une action spécifique, mais également lorsqu'il observe l'action en question. Il a alors rapidement été proposé que ces neurones puissent être un médiateur de l'imitation, mais aussi de la compréhension des actions, de l'empathie, de l'émergence et du développement du langage (Arbib, 2013), entre autres (Brass & Heyes, 2005, p. 489).

Les premiers neurones miroirs découverts ont été détectés lors d'actions de saisies d'objets par les singes, mais d'autres types de neurones miroirs ont été révélés par la suite (toujours chez les singes). Par exemple :

- des neurones s'activant lors d'observations de la bouche d'autres singes lors d'ingestion de nourriture ou d'émission de signaux et qualifiés ainsi de neurones miroirs communicatifs (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003)

- d'autres répondraient plus spécifiquement à l'utilisation d'outils (Ferrari, Rozzi, & Fogassi, 2005).

Heyes (2010a, p. 789), devant l'enthousiasme de la communauté scientifique au sujet de ces cellules, pose cependant une question intéressante : « *What is so mesmerising about mirror neurons* ? ».

Dans un sujet comme l'imitation en parole, ces neurones revêtent en effet un caractère fascinant : si l'humain dispose effectivement d'un système miroir lui permettant d'activer des représentations pendant qu'il perçoit une action, et ainsi d'intégrer la chaîne motrice nouvelle pour être ensuite capable de la reproduire, alors le problème de correspondance serait résolu… Ceci dit, il semble que l'explication ne soit pas aussi simple : comme le notent Brass & Heyes (2005), il y a encore peu de preuves expérimentales qui pointent vers l'implication des neurones miroirs comme explication du développement du langage chez l'être humain. De plus, les difficultés d'apprentissage des sons d'une langue nouvelle abondent également en ce sens. En revanche il est assez probable que les neurones miroirs jouent un rôle dans les comportements imitatifs, sans leur être exclusivement dédiés.

Afin de faire ressortir le caractère fascinant des neurones miroirs, il peut être intéressant de développer succinctement l'hypothèse développée par Arbib (2012) concernant le rôle du système miroir dans l'émergence et l'évolution du langage dans l'espèce humaine.

Sur la base de la découverte des neurones miroirs de saisie chez le singe, puis la détection de zones a priori équivalentes chez l'humain à proximité de l'aire de Broca suite aux travaux de Grafton *et al.* (1996), Arbib (2012, pp. 28-29) postule que l'aire de Broca est issue de l'évolution des zones de neurones miroirs pour la saisie des singes. Ainsi, ces zones miroirs auraient servi de « marche pied » pour faire émerger la propriété de parité des premiers protolangages et rendre le cerveau humain apte au traitement du langage (ce que Arbib désigne comme « *language ready brain* »). Pour Arbib, l'hypothèse des systèmes miroirs pourrait ainsi fournir le chaînon manquant des théories sur l'évolution du langage arguant que la communication non verbale y aurait joué un rôle crucial.

Le potentiel pouvoir explicatif des neurones miroirs dans le domaine de l'évolution de l'espèce constitue d'ailleurs pour Heyes (2010a) la première raison leur fournissant un caractère fascinant. Les deux autres raisons que voit Heyes à la fascination qu'exercent les neurones miroirs sont exprimées par les notions d'atomisme et de télépathie :

- L'atomisme renvoie à la tradition de pensée de la Grèce antique où une unité petite et indivisible, permettait d'expliquer la composition de la matière. Or, Heyes (2010a.) remarque que les neurones miroirs semblent avoir ce caractère atomique. Pour cette raison, il peut devenir tentant d'expliquer un pan de plus en plus large des comportements sociaux de l'être humain par l'effet de ces particules miroirs, comme les comportements de violence dans la société.
- La notion de télépathie intervient dans la mesure où les neurones miroirs permettraient la compréhension de l'état d'autrui, sans qu'il y ait eu communication. En effet, si les neurones miroirs activent les représentations du sujet observant, ils devraient aussi provoquer chez l'observateur les mêmes états émotionnels que ceux du modèle.

Un certain nombre de ces traits fascinants relèvent de l'*a priori* considérant que les neurones miroirs sont des structures neuronales immuables, présentes chez tous les individus, dans des proportions et à des fins équivalentes. Un tel point de vue défendrait l'idée que les neurones miroirs sont une adaptation de l'organisme humain dans le sens suivant :

> A characteristic of an organism, C, is an adaptation for a particular function, F, if C evolved because it helped organisms to do F. (Heyes, 2010b, p. 576)

Or, la manière dont se développent ces neurones est encore mal connue. L'hypothèse que défend Heyes (2010b, p 579-580) est que les neurones miroirs sont le produit corollaire (*a by-product*) d'un apprentissage par association :

- La présence d'un système miroir chez le nouveau-né est controversée. Bien que ces derniers fassent preuve d'imitation néonatale, celle-ci concernerait uniquement un répertoire restreint d'action (la protrusion de la langue) de manière transitoire.
- L'expérience et l'entraînement de l'individu dans l'accomplissement d'actions spécifiques modulerait l'activation de leur système de neurones miroirs. Ainsi, les danseurs de ballet verraient leurs neurones miroirs décharger plus quand ils observent d'autres danseurs plutôt que des pratiquants de capoeira (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005).
- Les neurones miroirs seraient également sujets à la plasticité cérébrale. Heyes (2010b, p. 580) rapporte en effet que l'expérience sensorimotrice d'un individu peut améliorer ou au contraire faire décliner leur activité…

Envisager ainsi que les neurones miroirs sont un produit neuronal dérivé de notre expérience avec le monde contribue à faire diminuer leur caractère extraordinaire. Ils représenteraient le produit d'un apprentissage spécifique, destiné à améliorer le traitement en perception et

production de ces comportements appris. S'ils facilitaient le traitement bimodal des comportements observés, ce type de neurones s'érigerait en candidats privilégiés supportant au moins un des mécanismes cérébraux des imitations.

Enfin, il nous semble plus sage de parler de « mécanisme miroir » plutôt que de « neurones miroirs » ou de « système miroir » comme l'indique Rizzolatti dans sa correspondance avec Arbib (rapportée par Arbib 2012, p. 124) :

> I also used the term "mirror system" for many years. I am afraid that this term is misleading. There is not such a thing as a mirror system. There are many centers[33] in the brain of birds, monkeys and humans that are endowed with a mechanism – the mirror mechanism– that transform representation in its motor counterpart. […] The term "mechanism" avoids the notion that mirror neurons have a specific behavioral function.

Cette dernière assertion laisse entendre que les neurones miroirs sont dispersés dans le cortex cérébral et qu'ils supportent un type particulier de mécanisme cognitif dédié à l'établissement de correspondances entre l'expérience perçue et sa potentielle reproduction. De plus, plutôt que d'attribuer l'explication des comportements imitatifs aux seuls neurones miroirs, Rizzolatti leur attribue un rôle de rouage, ils ne seraient alors plus qu'une part de l'explication des mécanismes cognitifs conduisant à la production de comportements imitatifs.

1.2 Court inventaire des zones cérébrales liées à l'imitation gestuelle

Maintenant que nous avons dit quelques mots des neurones miroirs, il nous semble opportun de nous intéresser aux zones du cerveau qui seraient actives lors d'imitations. Brass & Heyes (2005, p. 492) ont passé en revue la littérature en neurosciences de l'imitation. D'après leur revue, les principales zones du cerveau impliquées dans les comportements imitatifs (celles qui seraient actives à la fois durant la production d'une action et durant son observation) seraient relativement peu nombreuses :

- Le gyrus frontal inférieur, siège entre autres de la zone de Broca
- Le cortex prémoteur dorsal et ventral, liés à la planification des mouvements
- Le lobule pariétal supérieur, impliqué dans la perception spatiale de l'individu
- Le cortex pariétal inférieur, dont l'activité servirait l'analyse des objets

[33] Nous soulignons

- Le sillon temporal postérieur supérieur, en lien avec les aires auditives et visuelles primaires

Il est intéressant de noter que ces aires cérébrales impliquées dans l'imitation sont aussi bien des aires liées à la perception de l'environnement du sujet que des aires liées à sa motricité, soit à sa production. Celles-ci ne semblent de plus pas spécifiques à l'imitation.

Ceci dit, certaines zones précédemment citées, dont le gyrus frontal inférieur, sont source de débats dans la communauté, quant à leur implication dans l'imitation, notamment celle de l'aire de Broca.

Depuis l'étude de cas de Broca, qui estimait que cette zone du cerveau est le siège de « *la faculté de coordonner les mouvements propres au langage articulé* » (Broca, 1861), les fonctions qui lui sont associées sont plus variées et nuancées (Tzourio-Mazoyer, 2003, pp. 78-84). On attribue à cette aire un rôle de contrôle exécutif du discours (la sélection et la manipulation des éléments sémantiques) ainsi qu'une dissociation en deux pôles :

- Un pôle qui participe aux tâches phonologiques (partie operculaire du gyrus frontal inférieur)
- Un pôle impliqué dans le traitement sémantique (parties triangulaire et orbitaire du même gyrus).

Tzourio-Mazoyer (2003, p. 82) souligne par ailleurs dans sa revue que la zone de Broca serait aussi significativement plus active durant la lecture labiale, que pendant l'observation de mouvement de bouche sans fonction linguistique.

Pour Heiser, Iacoboni, Maeda, Marcus, & Mazziotta (2003), il ne semble faire aucun doute que la partie operculaire du gyrus frontal inférieur (la partie postérieure de l'aire de Broca) et son homologue dans l'hémisphère droit sont des structures neurales essentielles de l'imitation des mouvements de doigts. Afin de parvenir à cette conclusion, ces chercheurs ont utilisé la technique rTMS (*repetitive transcranial magnetic stimulation*) pour inhiber les parties operculaires (et une zone contrôle) durant une tâche d'imitation de mouvements de doigts. Leurs résultats indiquent un taux d'erreur significativement plus élevé à la tâche durant l'imitation quand les parties operculaires étaient soumises à la rTMS. Ils suggèrent alors que les parties operculaires droites et gauches participent à un mécanisme miroir commun d'association directe (« *direct matching* »). Ils soutiennent dans leur conclusion qu'il y aurait une continuité évolutionniste entre la reconnaissance des actions, l'imitation et le langage.

A contrario, l'étude de Makuuchi (2004) présente des résultats plus mesurés : pour ce dernier, l'activité dans la partie operculaire serait affectée par le timing d'exécution du mouvement à reproduire, suggérant alors que l'aire de Broca servirait de « mémoire tampon » pour la reproduction d'actions différées. Ainsi, Makuuchi estime que l'aire de Broca n'a pas de rôle fondamental dans l'imitation des actions.

En ce qui nous concerne, le point de vue de Makuuchi semble parcellaire : il n'inclut à aucun moment la possibilité que l'imitation soit souvent une action exécutée en différé et semble considérer que l'imitation ne s'applique qu'à des comportements nouveaux. Ainsi, Makuuchi fonde son argumentaire contre l'implication de Broca sur la méthodologie employée par les autres chercheurs : ces derniers n'auraient pas proposé de tâche d'imitation car ils ne définiraient pas l'imitation de la même manière. Nous nous retrouvons encore dans une controverse liée au débat terminologique sur le terme « imitation »…

D'une part, cette dernière remarque souligne l'importance de définir avec rigueur les termes utilisés pour décrire le type de comportement que le chercheur observe. D'autre part, elle met en exergue l'impact des termes choisis sur la méthodologie et l'interprétation de leurs résultats.

L'imitation gestuelle semble donc mobiliser de nombreuses zones cérébrales, qui seraient actives durant la production d'actions, mais également durant leur observation. Ces zones sont par ailleurs impliquées dans des processus liés à la perception et à la production des actions, ainsi qu'à leur planification. Certaines zones citées font toutefois débat, mais nous ne prendrons pas part à ce dernier pour plutôt nous intéresser à deux études s'intéressant à l'activité cérébrale durant des tâches d'imitation en parole.

1.3 Exploration des zones cérébrales en imitation de la parole

Face à ces premiers constats sur le substrat neurologique des imitations gestuelles, impliquant des zones cérébrales afférentes à la perception et la production d'activités motrices, il convient d'observer si l'imitation de la parole –une activité mobilisant perception et production– suit un pattern similaire d'activations cérébrales. Nous pourrions en effet attendre que l'imitation de la parole mobilise des zones liées à sa perception, à sa production et à sa planification.

Durant une série de cinq tâches testant différents degrés d'imitation en parole (perception, imitation implicite, lecture de phonèmes, imitation délibérée et inhibition de l'imitation), Garnier *et al.* (2013) ont observé l'activité cérébrale de 15 sujets francophones au moyen d'imagerie par résonnance magnétique fonctionnelle (IRMf).

En termes de production parolières, cette étude rapporte que les locuteurs ont produits les effets d'imitation attendus : globalement la production des locuteurs changeait au fil des tâches pour, en fonction des consignes, se rapprocher ou s'éloigner de la référence entendue. Ci-après, nous nous focaliserons plutôt sur leurs résultats abordant les aspects neurologiques de l'imitation de la parole (*Ibid.*, p. 7-8).

- En tâches de perception (écoute) et production (lecture) : leurs résultats rapportent une activation des réseaux neurologiques classiques de la parole. Entre autres :
    - En perception :
        - Activation bilatérale du gyrus temporal supérieur, plus particulièrement du cortex auditif
        - Activation bilatérale du gyrus frontal inférieur (parties operculaire et triangulaire) et latéralisée (gauche) de la partie orbitaire.
        - Activations dans le lobe frontal
            - Cortex préfrontal dorsolatéral
            - Cortex prémoteur
            - Aire motrice supplémentaire
    - En production :
        - Activation bilatérale des cortex
            - Prémoteur
            - Moteur primaire
            - Sensorimoteur
        - Activation de l'aire motrice supplémentaire
        - Activation bilatérale du gyrus frontal inférieur (parties operculaire et triangulaire)

En ce qui concerne les différentes tâches d'imitation, Garnier et collègues (2013, p. 8) ne notent pas de différence fondamentale dans les régions cérébrales impliquées lors des tâches d'imitation, qu'il soit demandé aux sujet une simple production, une imitation délibérée ou une inhibition de l'imitation.

Garnier *et al.* (*Ibid.*, p 8-9) indiquent simplement que leurs analyses supplémentaires soulignent une modulation de l'activité de l'aire de Wernicke et du cortex auditif en perception sans toutefois pouvoir observer de tendance nette (+/- activation) en fonction du degré d'imitation. Il en va de même en production pour l'Insula gauche, le gyrus supramarginal et le cortex auditif droit.

Comme ils le soulignent dans leurs résultats, l'imitation de la parole, quelle que soit l'intentionnalité du sujet, semble mobiliser des zones cérébrales identiques, reflétant la mise en œuvre d'un mécanisme unique de l'imitation. Il semble par ailleurs, que ce mécanisme cérébral d'imitation de la parole ait de grandes similarités avec les mécanismes cérébraux de l'imitation gestuelle (nous nous y attendions). Enfin, leur discussion met en lumière l'idée intéressante que l'activité cérébrale de certaines zones diffère en fonction de l'étape de la boucle imitative (de la perception à la production) : il semblerait que l'activité cérébrale diffère plus en fonction de la tâche durant les processus perceptifs. Le degré d'imitation serait donc modulé par l'activité de perception (*Ibid.*, p. 12), supportant alors l'idée que la perception actualiserait nos représentations motrices de manière automatique.

Reiterer, Hu, Sumathi, & Singh (2013) se sont quant à eux intéressé à l'activation cérébrale des imitateurs en fonction de leur degré de compétence imitative de sons inconnus. Cette étude est un prolongement de Reiterer et al., 2011) sur le phénomène d'effort/efficacité corticale, selon lequel une bonne compétence imitative se traduirait par une activité corticale réduite, par comparaison avec une compétence d'imitation plus mauvaise.

Il semble que leur étude de 2013 (p. 11) confirme ce principe d'efficacité/effort cortical puisque les imitateurs faisant preuve d'une compétence réduite dans ce domaine consomment plus d'espace neuronal dans les tâches d'imitation et y ont une activité plus intense. Cela traduirait leur effort cognitif pour (mal) accomplir la tâche d'imitation. En termes de différence structurelle, le lobe pariétal inférieur, *a priori* impliqué dans la mémoire de travail auditive, serait plus actif chez les sujets les moins talentueux.

Ces résultats tendent à rejoindre la proposition de Garnier *et al.* (2013) selon laquelle les activités cérébrales de traitement perceptif jouent un rôle particulièrement important durant l'imitation de la parole.

1.4 Une inhibition de l'imitation ?

Si notre cerveau possède des zones actives pendant l'observation et la production d'un comportement dont la fonction serait de lier perception et production, nous pouvons nous demander pourquoi nous arrivons à faire preuve de singularités dans nos comportements, *i.e.* pourquoi la production de comportements imitatifs n'est pas automatique et systématique.

Le temps accordé à la définitions des termes utilisés pour évoquer l'imitation nous a permis de souligner le spectre très étendu des types de comportements imitatifs : de l'automatisme des réponses mimétiques à la préméditation de certains types d'imitations. La revue de littérature de Brass & Heyes (2005, p. 493) rapporte plusieurs études sur des patients ayant une lésion située dans le lobe préfrontal et qui présenteraient un comportement échopraxique. Ils ne pourraient s'empêcher d'imiter les mouvements d'un modèle malgré des consignes émises par ailleurs.

Brass, Derrfuss, Matthes-von Cramon, & von Cramon (2003) ont montré au cours de tâches d'inhibition de l'imitation de mouvements de doigts que des sujets présentant une lésion dans le lobe frontale avaient des difficultés à réprimer les réponses imitatives, par comparaison avec des sujets ayant une lésion cérébrale autre et des sujets non lésés. Il est donc probable que les sujets ne présentant pas de lésion du lobe frontal disposent d'un mécanisme neurologique permettant d'inhiber les imitations.

Afin de porter un éclairage sur cette question, Brass *et al.* (2005; 2001) ont produit des études de neuro-imagerie sur les mécanismes d'inhibition des réponses imitatives. Leurs résultats tendent à indiquer que l'inhibition de l'imitation provient des structures cérébrales dont la fonction est liée à la distinction entre soi et les autres : ils relèvent une activation du cortex médian préfrontal et de la jonction temporo-pariétale.

En parole, l'étude de Garnier et collègues (2013, p. 12) indiquent que l'inhibition de l'imitation mobilise les mêmes aires cérébrales que celles évoquées par Brass *et al.* (2005, 2001).

1.5 Synthèse : l'imitation comme système dynamique d'ajustements permanents

Ces quelques éléments de littérature tendent à montrer que les structures neurologiques de l'imitation se complèteraient de structures dédiées à l'inhibition des représentations motrices automatiquement activées par l'observation d'une action.

Notre vue (relativement parcellaire) des structures neuronales des comportements imitatifs tend à nous diriger vers un modèle cognitif (assez simplifié) de l'activation de l'imitation reposant sur une binarité de repoussement/attraction entre l'individu et son contexte (dans un sens très large).

Par certains côtés, cette vue est assez modelée par la théorie du seuil d'activation des langues du sujet bilingue proposée par Paradis (1993) selon laquelle le seuil d'activation d'une langue X d'un individu diminuerait en fonction de sa fréquence d'utilisation, tandis que le seuil des autres langues qu'il parle augmenterait de manière dynamique.

Nous pensons que l'individu sera plus ou moins enclin à produire de l'imitation en fonction de la perception qu'il a du contexte dans lequel il se trouve (Figure 6).

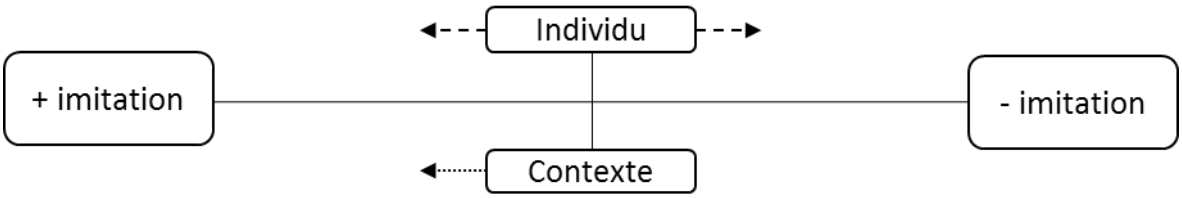


Figure 6 : Continuum dynamique d'activation de réponse imitative.

Le curseur sur le continuum représente le lien entre l'expression du soi de l'individu (les structures inhibitoires) et l'effet que produit le contexte sur lui (les représentations motrices activées par sa perception). Le contexte peut tirer l'individu vers l'imitation (flèche en points vers la gauche) et alors faire dériver le curseur dans cette direction :

- La propagation des mouvements de foule (panique, ola) constituent un bon exemple d'effet brutal du contexte sur le comportement de l'individu. La manière dont nous respectons certaines conventions sociales illustrerait aussi ce type d'effet, dans une

dynamique d'ajustement; un type de mimétisme social. Dans ces cas-ci l'imitation ne serait pas inhibée car l'individu serait « attiré » (*urged*) par l'effet du contexte.

L'individu peut au contraire tirer son type d'action vers la droite du continuum si l'effet du contexte sur lui n'est pas assez fort (flèche en tirets vers la droite), l'imitation serait inhibée, l'effet du contexte repoussé (*repel*). Enfin, l'individu peut déclencher (*trigger*) de manière délibérée l'imitation (flèche en tirets vers la gauche) :

- Dans des situations où l'individu aurait l'envie ou la nécessité de reproduire des actions de manière délibérée (jouer, se servir d'un nouvel automate, évoquer quelque chose ou quelqu'un)

Par ailleurs, cette représentation tirée d'une analyse partielle des aspects neurologiques de l'imitation est conceptuellement très flexible. En effet, elle donne un squelette de base pour décrire une dynamique de l'imitation à plusieurs niveaux :

- Au niveau terminologique, elle peut intégrer trois types d'imitation évoqués au chapitre précédent : le mimétisme, l'imitation, l'émulation.

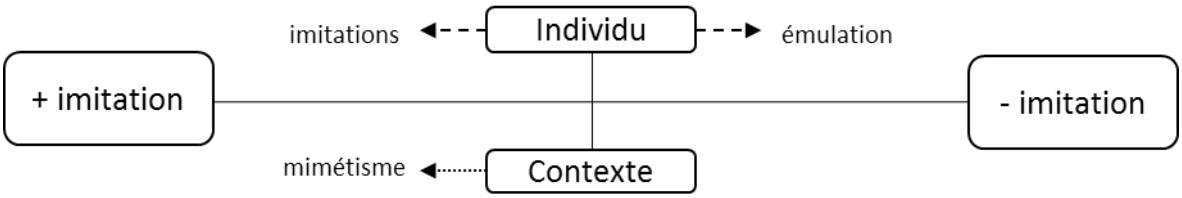


Figure 7 : Insertion des types d'imitation à un continuum dynamique des imitations

- Appliquée en parole, on y retrouverait la distinction entre différents comportements comme la convergence, la divergence phonétique et l'impersonnation (*cf.* Figure 10).
- Quel que soit le niveau d'analyse, il est possible d'envisager que le système s'actualise et que la part imitée des réponses soit ajustée en permanence en produisant à chaque fois des variations plus ou moins franches dans le comportement de l'individu.

Une telle représentation laisse supposer que les individus sont capables de réguler volontairement la part imitative de leur comportement en en changeant « brutalement » (plus ou moins d'imitation), ou bien de laisser plus libre cours à un équilibrage automatique entre attirance et repoussement, entre imitation et inhibition de l'imitation.

2. Expérience du sujet et apprentissage sensorimoteur dans le développement de la capacité d'imitation

Ayant abordé les structures cérébrales de l'imitation et ce qu'elles nous suggèrent quant à la dynamique comportementale que ces structures semblent sous-tendre, il nous faut à présent aborder le mécanisme cognitif qui permettrait de prendre en charge le problème de correspondance.

Deux types de modèles cognitifs se confrontent à ce sujet : des modèles dits « spécialistes » et d'autres dits « généralistes » (Brass & Heyes, 2005, pp. 489–490). Les modèles spécialistes postulent que l'imitation possède des structures dédiées à son fonctionnement tandis que les modèles de type généraliste considèrent que le traitement cognitif de l'imitation serait géré de manière partagée entre plusieurs structures non exclusives à l'imitation.

Nous avons observé que certaines zones cérébrales actives durant l'imitation, le sont aussi durant l'observation d'une action, sa planification ou sa production. Par conséquent, il nous semble plus pertinent de porter notre attention sur un modèle de type généraliste, pour décrire la manière dont se développerait la capacité d'imitation (Heyes, 2001, 2010b; Heyes, Bird, Johnson, & Haggard, 2005).

Plutôt que de supposer qu'une structure interne établisse à la demande la correspondance entre stimuli visuels et production motrice (ce que proposent les modèles spécialistes), l'*Associative Sequence Learning Hypothesis* (ASL) propose que le mécanisme d'imitation repose sur des ensembles de liens bidirectionnels entre les représentations motrices et sensorielles (Heyes, 2001, p. 258).

Ainsi, l'ASL présume une liaison forte entre nos sensations et notre répertoire d'actions, et elle décrit ce lien comme des associations verticales de deux types (Figure 8) :

- *Associations verticales directes* : formées par la co-activation des représentations sensorielles et motrices survenant lors d'une expérience personnelle d'observation/action (Brass & Heyes, 2005, p. 491; Heyes, 2001, p. 258)

- *Associations verticales indirectes* : dont la création passe par la médiation d'une autre modalité que la simple observation, par exemple, quand une instruction verbale est donnée pendant l'observation et/ou l'observation de l'action (Heyes, 2001, p. 258).

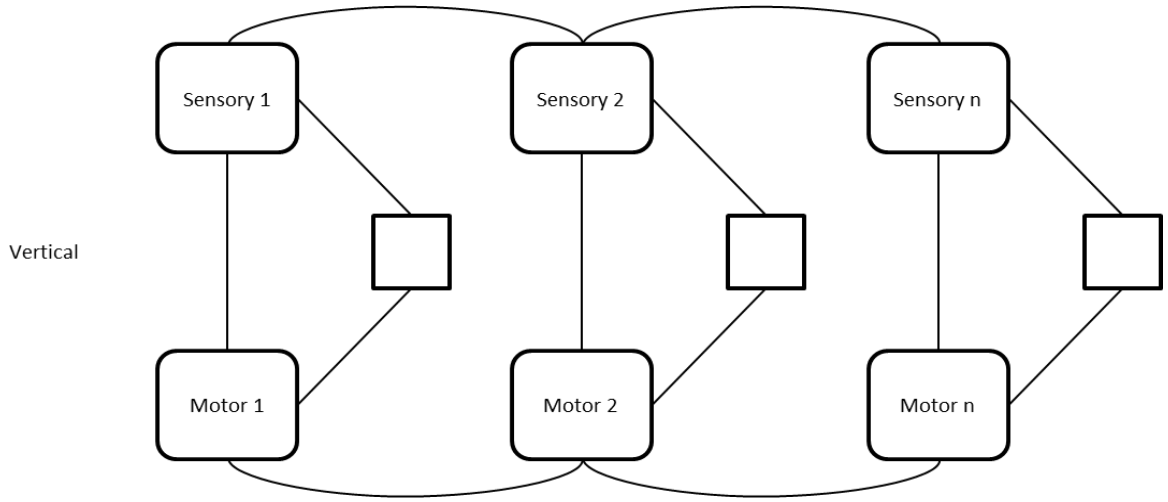


Figure 8 : Associations verticales directes et indirectes postulées par le modèle ASL (Brass & Heyes, 2005; Heyes, 2001; Heyes et al., 2005). Celles-ci se traduiraient au niveau neurologique, par un renforcement du lien entre les neurones sensitifs et moteurs spécifiques aux actions imitées, d'après une règle hebbienne (Heyes, 2010b, p. 577).

Ainsi, pour les tenants de l'ASL, pour que des représentations motrices soient actives lors de l'observation d'une action, il faut que l'observateur ait déjà observé et/ou accompli une action similaire. L'expérience du sujet serait alors primordiale pour initier les mécanismes imitatifs.

Par ailleurs, Heyes *et al.* (2005) ont testé l'effet de l'expérience du sujet sur les imitations gestuelles dans une série de deux expérimentations :

- Durant la première, les sujets devaient fermer ou ouvrir la main dès qu'ils commençaient à détecter du mouvement dans le stimulus vidéo qui présentait alternativement une main s'ouvrant ou se fermant.
    - Les résultats de ce premier test indiquaient une réponse plus rapide lorsque le mouvement demandé au sujet et le stimulus observé étaient congruents (ouverture/ouverture ou fermeture/fermeture *vs.* ouverture/fermeture ou fermeture/ouverture).
- Le second test reprenait la procédure du premier test, hormis le fait que les groupes de sujets recevaient chacun un entraînement différent :

    - Le premier groupe s'entraînait à répondre de manière congruente (fermeture de la main si fermeture sur le stimulus, et inversement).
    - Le second groupe s'entraînait à répondre à l'inverse (fermeture de la main si ouverture en vidéo et vice-versa).
- Par ailleurs, les sujets des deux groupes devaient effectuer des tâches congruentes (faire le même mouvement que la vidéo) et non congruentes (faire le mouvement opposé à la vidéo).
    - Le résultat principal de ce test indiquait qu'un effet d'imitation automatique (facilitation des réponses congruentes) peut être diminué, voire aboli, par un entraînement approprié.

Cette étude indique donc que les réflexes imitatifs peuvent être modulés par une expérience appropriée ; en d'autres termes, elle pointe que les liens entre nos représentations motrices et sensorielles sont susceptibles de subir une évolution, soit par une exposition fréquente à des répertoires d'actions, soit par un apprentissage.

Cette dernière remarque n'est pas anodine dans le contexte de l'apprentissage de la correction phonétique au moyen de la MVT dont la proposition est d'agir sur les représentations phonologiques de l'apprenant pour en améliorer l'*output* moteur : la prononciation des sons d'une L2…

D'ailleurs, nous devrions peut être parler d'*output* perceptuo-moteur lorsqu'il s'agit de geste parolier, en nous référant à la théorie PACT (*Perception for Action Control Theory*, Schwartz, Basirat, Ménard, & Sato, 2012) :

> [PACT] considers that speech perception is the set of mechanisms that enable not only to understand, but to control speech, considered as a communicative process. This leads to two consequences. Firstly, perception and action are costructured in the course of speech development, which involves both producing and perceiving speech items. In consequence, the perceptual system is intrinsically organized in reference to speech gestures […]. Secondly, perception provides action with auditory (and possibly visual) templates, which contributes to define the gesture […]. In PACT, the communication unit through which parity may be achieved, is neither a sound, nor a gesture, but a perceptually shaped gesture.[34]

En nous plaçant dans la perspective de PACT, ces considérations introduisent dans notre propos l'idée que le geste parolier aurait une valeur fonctionnelle intrinsèquement liée à sa perception. Il semble alors plus périlleux de considérer les notions de renforcement vertical de

[34] Nous soulignons

l'ASL au domaine de la parole puisque la dichotomie entre représentation motrice et perception serait ici fusionnée en une seule unité.

Ceci dit, un compromis (probablement insatisfaisant) pourrait être de postuler que la liaison verticale entre représentation motrice et sensorielle des sons de parole est forgée très tôt dans le développement du langage (pour PACT, le développement langagier serait une structuration commune dans les modalités perceptives et productives). Communiquer étant une activité quotidienne, il pourrait alors être acceptable de penser que ces liaisons perception/action sont particulièrement fortes, au point d'être insécable et de former les unités perceptuo-motrices proposées dans PACT. Un tel point de vue pourrait alors expliquer la difficulté des apprenants de langue étrangère à imiter les sons de la L2 (nous y revenons au chapitre 3).

3. Intégration d'aspects imitatifs à la perception et à la production de la parole

Il convient à présent de clarifier le spectre des différents comportements d'imitation en parole. Pour ce faire, nous proposons de souligner quelques aspects de traitement de la parole qui nous semblent nécessaires à une capacité d'imitation. Dans un premier temps, nous nous intéresserons au stockage des unités linguistiques au travers duquel nous évoquerons les comportements dits « d'impersonnation » (ou imitation vocale). Par la suite, nous nous prononcerons en faveur d'une dynamique de la perception/production de la parole, dans laquelle nous contextualiserons les comportements d'ajustement phonétique décrits par la *Communication Accomodation Theory* (Giles *et al.,* 1991). Enfin, nous envisagerons les activités langagières comme l'action conjointe de processus de perception et de production, un point de vue séduisant dans notre contexte.

3.1 La question du stockage lexical en regard des imitations délibérées de la parole

Les locuteurs sont capables d'imiter (plus ou moins fidèlement) la voix d'autres locuteurs, au point qu'ils pourraient parvenir à se faire passer pour autrui. Quand un locuteur essaye « d'usurper » la voix d'un autre en changeant son comportement vocal (pour une raison frauduleuse, humoristique ou autre) nous parlerons soit d'imitation vocale (Révis et al., 2013), soit d'*impersonnation* (Zetterholm, 2009a). En effet, le but du sujet produisant de l'impersonnation est de parvenir à leurrer son auditoire, pour lui suggérer qu'il est une autre personne. La vignette de la Figure 9 représente un cas typique d'impersonnation à visée humoristique.

Figure 9 : Les rois du rire Tronchet (1996). Situation d'impersonnation : Monsieur Paintex imite délibérément la voix de la femme de l'adjoint au maire Lemortier.

Nous avons choisi d'illustrer l'impersonnation par cet exemple à cause de la force évocatrice de la bande dessinée, où la voix des personnages est contenue dans les phylactères. Le changement vocal de Paintex est d'ailleurs illustré dans la seconde vignette présentée. L'auteur y marque la modulation phonétique de l'imitateur, au niveau segmental (« chûûûcres » pour « sucres »), et prosodique (allongements très marqués : « cafêê », « lêê », « prendrêê », « chûûûcres »). Cette capacité de modulation de la voix est accessible à tout locuteur, la différence principale entre les individus dans ce domaine résidant plutôt dans la notion de talent. Les impersonnateurs reconnus vivent de cette activité et constituent une population expérimentale experte dans certains domaines de recherche (voir chapitre 4).

Par ailleurs, cette situation souligne le fait que l'imitation vocale est souvent une imitation différée (*i.e.* produite en l'absence du locuteur modèle, parfois longtemps après l'exposition de l'imitateur au modèle). Cette idée suggère que les locuteurs stockent des représentations mentales de la manière de parler de leurs interlocuteurs, auxquelles ils peuvent accéder à loisir. Ainsi, il nous est nécessaire de considérer les alternatives théoriques du stockage des unités linguistiques.

Dans ce domaine, deux types de modèles théoriques s'opposent : les premiers sont dits « abstractionnistes » et les seconds « exemplaristes »

> Abstractionnist models, on the one hand, are based on the assumption that an abstract and speaker-independent phonological representation is associated with

> word in the mental lexicon. In exemplar models, on the other hand, words and frequently-used grammatical constructions are represented in memory as large sets of exemplars containing fine phonetic information. (Nguyen, Wauquier, & Tuller, 2009, p. 193).

Les positions abstractionnistes sont une réponse à la question de savoir comment les auditeurs font pour traiter la variabilité intrinsèque à la parole. Dans cette optique, les modèles abstractionnistes postulent que les items lexicaux sont stockés en mémoire après une procédure de normalisation, consistant, dans le traitement de la parole, à évacuer les différentes manières de prononcer les mots (la variabilité). Selon cette vue, il ne resterait en mémoire de l'individu qu'une seule unité dans le lexique mental : une représentation phonologique permanente.

A contrario, les modèles exemplaristes supposent que les représentations stockées en mémoire sont constituées de morceaux[35] (*chunks*) comportant plusieurs exemplaires de la même unité. Chaque exemplaire serait stocké avec un certain nombre de détails sur son contexte de perception/production, détails d'ordre phonétique, sémantique, *etc.* (Nguyen et al., 2009, pp. 194–195).

Le comportement que nous observons, l'imitation, tend à faire pencher notre opinion en faveur des modèles de type exemplariste. Pour pouvoir imiter la voix d'un autre individu ou faire l'accent marseillais quand on est toulousain, il nous semble nécessaire que l'information stockée en mémoire à long terme contienne des détails phonétiques fins, ou des représentations de certaines unités liées au souvenir d'un locuteur.

Par ailleurs, ce type de stockage permettrait d'améliorer notre perception de la parole produite par nos interlocuteurs habituels. D'après Goldinger (1998), après une exposition à la voix d'une personne, le traitement ultérieur de mots prononcé par cette même personne se verrait facilité.

Bien que nous soyons ainsi tentés de nous en remettre aux seules sirènes des théories exemplaristes, il nous faut ici souligner que la revue de littérature de Nguyen *et al*. (2009) pointe que les deux approches accumulent des arguments et des résultats expérimentaux en leur faveur.

[35] Nous aurions été tentés de traduire « chunk » par grumeaux. Ceux-ci suggèrent en effet l'agglomération de particules très fines.

> [This review] suggests that dichotomy that is sometimes established between the exemplar-based and abstractionnist approaches to speech perception is to a large extent artificial. (p. 209)

C'est pourquoi ces derniers proposent de dépasser la dichotomie exemplariste/abstractionniste en considérant la perception (et la production de la parole) comme un système dynamique dans lequel des aspects des deux approches sont considérés.

3.2 Pour une dynamique de la perception/production de la parole

Lorsque deux amis discutent, il est possible d'avoir l'impression que ces personnes partagent leur manière de parler : ils utilisent les mêmes mots, peuvent avoir le même accent, les mêmes attitudes, les mêmes intonations. Si nous nous intéressons au niveau phonétique, ce type de cas illustre ce qui est désigné dans la littérature sous l'appellation de convergence phonétique, comme une tendance à partager le même comportement parolier.

La convergence a été définie dans le cadre de la Communication Accomodation Theory (CAT) par Giles et al. (1991, p. 7) :

> « Convergence » has been defined as a strategy whereby individuals adapt to each other's communicative behavior in terms of a wide range of linguistic-prosodic features including speech rate, pausal phenomena and utterance length, phonological variants, smiling, gaze, and so on.

Plusieurs aspects de cette définition sont particulièrement importants dans notre réflexion. En premier lieu, nous devons souligner la notion d'adaptation introduite ici. Si nous nous remémorons les définitions du chapitre précédent, il peut être intéressant de considérer la convergence comme une stratégie mimétique du locuteur pour faciliter son interaction avec son interlocuteur en s'adaptant aux codes de la situation de communication.

Par ailleurs, notre exemple introductif présente la faiblesse de ne pas prendre en compte la potentielle dynamique qui pourrait être en jeu dans ce comportement. En effet, il est tout à fait concevable que la mise en place de cette stratégie se fasse au fil même de la conversation entre deux inconnus. De nombreux travaux récents se sont en effet intéressés à l'émergence de cette stratégie :

- En conversation (Nathalie Lewandowski, 2012; Pardo, 2000, 2006)
- En situation de laboratoire (Cole & Shattuck-Hufnagel, 2011; Dufour & Nguyen, 2013; Namy, Nygaard, & Sauerteig, 2002)

- En étude longitudinale (Pardo, Gibbons, Suppes, & Krauss, 2012)
- Sur des voyelles isolées (Sato et al., 2013)

Les résultats de ces études tendent à montrer que les locuteurs parviennent à percevoir les détails phonétiques fins puisque leur production change à mesure des tâches qui leur sont proposées, pour devenir plus similaire à la production du modèle entendu ou de leur interlocuteur.

De plus, il semble que l'effet de convergence soit robuste en conversation, comme le souligne l'étude de Pardo (2006), cet effet perdurant après la fin de l'interaction test. De tels effets sont en ligne avec la proposition de Sato et collègues (2013) pour qui le système perceptuomoteur des locuteurs s'ajusterait et se calibrerait en permanence sous l'effet de leur perception ; illustrant alors une dynamique du système langagier.

Afin de revenir à nos précédentes considérations sur la dynamique de l'imitation, la convergence phonétique représenterait le mécanisme attracteur lié au contexte de la communication dans un continuum dynamique du spectre des imitations.

Si nous nous replaçons dans un tel contexte, il faut également envisager l'autre extrémité d'un tel continuum. La CAT propose une stratégie opposée à la convergence phonétique, sous le terme de divergence :

> « Divergence » was the term used to refer to the way in which speakers accentuate speech and nonverbal differences between themselves and others. (Giles *et al*., 1991, p. 8)

La divergence propose donc que les locuteurs puissent, au lieu de s'adapter interactivement avec leur interlocuteur, s'en éloigner pour marquer des différences. Bourhis, Giles, Leyens, & Tajfel (1979) et Bourhis & Giles (1977) ont provoqué expérimentalement cet effet. Un locuteur A d'une communauté donnée se trouvaient en interaction avec un locuteur B ayant dénigré la communauté A. Le locuteur A, durant la tâche expérimentale tendait alors à diverger, pour manifester sa désapprobation. D'après les études que nous venons de citer, les stratégies de divergence sont variées et probablement très fortement liées au contexte communicatif. Hormis pour considérer une dynamique de l'imitation en parole, elles sortent du cadre de notre propos.

Enfin, Giles *et al*. (1991, p. 10) évoquent dans la CAT un troisième type de stratégie équivalent à la mise en place d'un statu quo comportemental, une tentative de préservation de l'authenticité de l'identité du locuteur :

> « Speech maintenance » is a value (and possibly conscious and even effortful) act of maintaining one's group identity.

Les stratégies communicatives envisagées par la CAT évoquent toutes la possibilité offerte au locuteur de moduler sa production en fonction de la perception qu'il peut avoir du contexte. Ceci dit, les stratégies de divergences et de maintenance semblent être plus fortement ancrées dans des enjeux culturels et sociaux que la stratégie de convergence. Si l'on se réfère aux mécanismes supposés de l'inhibition de la communication, nous pouvons en effet penser cela, dans la mesure où ces mécanismes de divergence et maintenance reposeraient sur des structures liées à la distinction entre le moi du sujet (son identité ?) et les autres.

Par ailleurs, il semble que l'effet de convergence puisse également affecter le traitement d'unité et de processus de bas niveaux comme la perception et la production des voyelles (Sato et al., 2013). Enfin, des effets similaires ont été repérés dans des contextes expérimentaux non-interactifs, sans enjeu social lié à la perception/production de la parole (Dufour & Nguyen, 2013; Goldinger, 1998; Namy et al., 2002).

Nous avons tracé les contours de trois stratégies de communication impliquant une dynamique de la production parolière. Celle-ci évoluerait à mesure de l'interaction à divers niveaux (phonétiques, prosodiques, lexicaux…).

Ayant défendu plus tôt qu'il y avait un lien particulièrement fort entre action et perception, il nous semble qu'il faille envisager les deux processus comme joints dans un système dynamique : c'est-à-dire, un système dont l'état n'est pas figé et dont le fonctionnement est largement dépendant du contexte (Tuller, Nguyen, Lancia, & Vallabha, 2010). En admettant ce type de point de vue, il devient concevable d'envisager que les traces phonétiques suscitées par les comportements décrits par la CAT constituent une fenêtre sur l'état du système linguistique du locuteur à un instant donné.

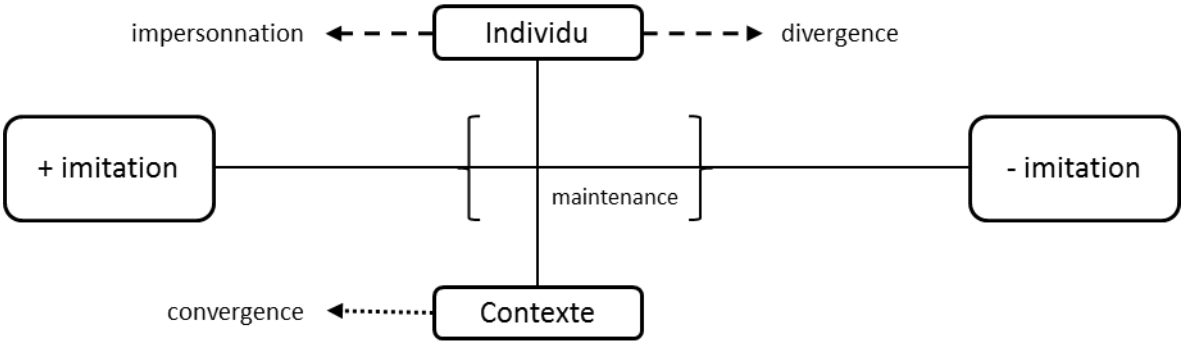


Figure 10 : Continuum dynamique des comportements imitatifs en parole, reprenant les notions d'attraction (*urge*) repoussement (*repel*) et déclenchement (*trigger*) et « zone » de maintenance.

La Figure 10 illustre une dynamique des comportements imitatifs paroliers. L'influence du contexte se traduirait dans le comportement parolier de l'individu par une propension à la convergence, une attraction vers un comportement imitatif diffus [+ imitation]. L'individu aurait par ailleurs la possibilité de déclencher l'impersonnation, et donc de moduler sa voix de manière à reproduire sciemment la voix d'un autre. Vers l'extrêmité [- imitation] du continuum, en fonction du contexte, l'individu peut se positionner dans une stratégie de divergence. L'intervalle central délimite une zone dans laquelle l'individu pourrait tenter de maintenir stable sa manière de parler : ne produisant ni de convergence, ni d'amplification de ses différences avec l'interlocuteur.

3.3 Aspects de synchronie en perception/production de la parole

Jusqu'à présent, nous avons défini l'impersonnation comme la tentative d'un locuteur de copier le comportement vocal d'un autre ainsi que la convergence phonétique, qui consiste en une adaptation mutuelle des locuteurs au comportement parolier de leurs interlocuteur (nous laissons ici de côté la divergence et la maintenance).

Alors que l'impersonnation est ponctuelle, déclenchée par le locuteur et aisément remarquable par les auditeurs, la convergence serait au contraire plus diffuse dans le temps et moins aisément détectable par les auditeurs (nous verrons cependant dans le chapitre 4 de notre travail, que les tests perceptifs sont privilégiés à cette fin, (Pardo, 2006, 2013a).

Levitan et Hirschberg (2011) distinguent différents types de similarité phonétique dans la conversation :

- *Proximité* : elle dénote une proximité ponctuelle des locuteurs sur certains aspects de leur production parolière
- *Convergence* : elle équivaut à une réduction de la distance d'un même aspect de la production de deux locuteurs (soit, comme la convergence de Giles *et al*, 1991)
- *Synchronie* : il s'agit ici d'une coordination relative des locuteurs au moment de l'alternance des tours de parole entre locuteurs.

Ce dernier phénomène nous intéresse particulièrement. D'une part, il pourrait s'agir en parole d'un comportement approchant l'imitation synchrone, telle que Nadel l'a décrit pour les enfants en bas âge (Nadel, 2011). D'autre part, la distinction convergence/synchronie peut indiquer une frontière entre tâche de fond imitative (une subtile modification dynamique du système) et pic ponctuel d'activité imitative (déclencher immédiatement un comportement semblable). Révis *et al.* (2013) distinguent d'ailleurs stratégie de convergence (tentative du locuteur de s'adapter aux caractéristiques globales de la voix cible) et stratégies de synchronie (tentative du locuteur de reproduire des variations immédiates de la parole du modèle) durant l'impersonnation. Enfin, le comportement de synchronie tendrait à indiquer que la parole en conversation constitue une activité jointe des interlocuteurs, dont les processus de perceptions seraient activés en production et vice versa.

Ce dernier point de vue est celui exprimé par Pickering & Garrod (2013) dans leur modèle intégratif de la production et de la compréhension du langage (aussi dénommé comme *Simulation Theory, ST*, Gambi & Pickering, 2013).

Avant de proposer un résumé des principes de la *ST*, il peut être souhaitable de faire référence à une situation assez commune en conversation : lorsqu'un locuteur fini l'énoncé de son interlocuteur :

- Loc A: i'm afraid I burnt the ceiling
- Loc B: But have you
- Loc A: burned myself? Fortunately not.

(Exemple rapporté par : Pickering & Garrod, 2013, p. 330)

Dans cette situation, le locuteur A a été capable de prédire la fin de l'énoncé de B. En s'aidant du contexte précédent, il est parvenu à inférer le contenu de l'énoncé à venir. Ce type d'exemple indique une activation simultanée des processus de perception et production de la parole.

La ST considère en premier lieu que la production du langage est une forme d'action et sa compréhension une forme de perception de l'action. Les deux modalités seraient en effet imbriquées l'une dans l'autre, dans la conversation, mais aussi au niveau des réseaux cérébraux engagés dans la production et la compréhension (ce que nous avons d'ailleurs tenté de défendre dans les premières parties de ce chapitre). Suite à ce constat, la ST postule que l'imbrication des modalités perceptives et productives du langage sert à faciliter la prédiction de ce qui va être dit, prédiction qui doit à son tour faciliter production et compréhension du langage. A propos des prédictions, elles seraient faites sur la base d'un système de représentation structurées du langage, à différents niveaux (sémantique, syntaxique, phonologique, phonétique, etc.) (Pickering & Garrod, 2013, p. 337).

En production, la ST postule qu'une commande de production initie l'acte de parole par deux canaux en simultané :

- un canal productif implémentant l'action et initiant le mouvement pour prononcer ce qui doit être dit.
- un canal de simulation créant une copie efférente initiant un modèle prédictif (*forward model*) de ce qui va être dit.

Les productions des deux canaux passeraient alors dans deux modules de compréhension puis comparés dans un moniteur. En fonction de l'adéquation entre énoncé prédit et effectivement prononcé, le locuteur saurait s'il a bel et bien produit ce qu'il avait planifié (Pickering & Garrod, 2013, p. 338).

En perception, le décodage d'un énoncé du locuteur B par un auditeur A provoquerait une « imitation secrète » (*covert imitation*) internalisée par A en même temps que B produit son énoncé. De même qu'en production, la ST postule deux canaux en perception :

- le canal perceptif classique où A traite effectivement ce que B produit
- un canal de simulation, avec copie efférente, et imitation synchrone et secrète

Comme en production, la simulation et le percept effectif seraient comparés dans un module de compréhension.

La route duelle en production permettrait donc à l'énonciateur de monitorer sa production tandis que la route duelle en perception permettrait à l'auditeur d'induire ce qui va être dit, et de l'exprimer ouvertement si le contexte communicatif le permet (ce que nous donnions en exemple précédemment.

Ce modèle nous intéresse particulièrement car il joint perception et production de la parole, ce qui dans un contexte d'étude de l'imitation a une certaine pertinence. Il est également propice à l'émergence de nouveaux points de vue sur des problématiques en lien avec les L2 et la difficulté que pourraient avoir les locuteurs de L2 à agir de manière synchronique.

Par exemple, si le système langagier du locuteur est incomplet, et que ses représentations ne sont pas suffisamment structurées, les routes de simulation en production et en perception ne pourraient pas être empruntées efficacement. De même, si certains niveaux (sémantique, syntaxique, phonologique, prosodique) sont insuffisamment complets, nous pourrions imaginer que cela soit source d'interférence, en compréhension comme en production.

Synthèse

Après avoir discuté des structures neuronales de l'imitation, réparties dans des zones allouées à la production et à la perception des actions (gestes ou parole), nous avons évoqué l'émergence cognitive de l'imitation selon l'hypothèse ASL. Cette dernière postule que le mécanisme d'imitation résulte de l'expérience et des apprentissages des individus. Nous avons ensuite défini différents comportements paroliers imitatifs en relation avec des apports théoriques de traitement du langage. Que l'imitation soit gestuelle ou parolière, nous proposons que la production d'imitation s'inscrive dans une dynamique d'attirance, repoussement ou déclenchement, nous permettant de décrire au niveau de l'individu, l'expression du spectre des différents comportements imitatifs

Chapitre 3 : Remédiation phonétique et imitation

Après avoir défini le spectre des comportements imitatifs (gestuels et paroliers) ainsi que leur dynamique d'utilisation chez les sujets humains, nous visons ici à mettre en contexte ces observations dans le domaine de l'enseignement/apprentissage de la prononciation en L2, plus particulièrement au moyen de la méthode d'intégration phonétique verbo-tonale (MVT). Avant de nous intéresser aux caractéristiques de la situation de correction phonétique, nous formulerons quelques remarques sur le travail phonétique dans l'enseignement des L2. Suite à ces propos liminaires, nous défendrons l'idée que l'interaction de correction phonétique et ses enjeux reposent sur les mécaniques de l'imitation parolière. Nous envisagerons alors les pratiques proposées par la MVT à l'enseignant pour remédier à la prononciation défectueuse de l'apprenant.

1. Propos liminaires : De la phonétique dans l'enseignement de la L2

S'il n'est pas proposé de travail phonétique à l'apprenant, il est implicitement considéré que ce dernier doit par lui-même parvenir à acquérir une prononciation naturelle de la langue cible. Il est alors espéré de l'apprenant qu'il s'imprègne des habitudes langagières des natifs à mesure de ses contacts avec eux. Comme le notent Delvaux, Demolin, & Soquet (2004) les locuteurs tendraient à conformer leurs productions phonétiques à celles auxquelles ils sont confrontés. En d'autres termes, une telle situation suppose que l'apprentissage phonétique de l'apprenant se déroule selon des mécanismes mimétiques.

Ceci n'est pas sans rappeler la méthode imitative du début du vingtième siècle (Puren, 1988, pp. 151–159) considérant que l'apprenant de L2 pourrait, comme l'enfant acquérant sa L1, reproduire les sons entendus dans son entourage. Cependant, les enseignants de L2, nous y compris, remarquent qu'en l'absence de travail phonétique spécifique, les apprenants peinent à atteindre la « *prononciation claire et naturelle* » attendue d'eux aux niveaux avancés (B2 et plus du CECR). La correction phonétique a pour but d'aider l'apprenant à développer la maîtrise de sa prononciation dans le contexte de l'interaction de classe.

1.1 Un effacement de la pratique phonétique ?

Parfois décrite comme parent pauvre de la didactique des langues étrangères (Rivenc, 2002, p. 26), la phonétique reste souvent absente des méthodes d'enseignement à disposition des enseignants, qui, eux-mêmes sont rarement formés à ces pratiques durant leur formation initiale (Renard, 2002a, p. 6).

L'absence dans les méthodes de matériel destiné au travail phonétique ne serait pas intrinsèquement problématique si les enseignants étaient formés dans ce domaine. Or, malgré les progrès récents des offres de formation en phonétique corrective (introduction de la MVT dans les curricula FLE de plusieurs universités, cours en ligne dédié, Billières, Alazard, Astésano, & Nocaudie, 2013), le travail phonétique, à un niveau institutionnel, pâtit d'une forte déconsidération.

En effet, le travail phonétique se trouve parfois relégué en dehors de la classe. Par exemple, certaines institutions[36] investissent des sommes considérables dans des solutions logicielles qui se proposent d'accomplir en lieu et place de l'enseignant, une correction de la prononciation de l'apprenant.

Pourtant, ce type d'approche présente, outre son coût financier astronomique, des défauts manifestes :

- L'absence de contrôle de l'enseignant peut amener à une fossilisation des erreurs de l'apprenant.
- Les logiciels ne tiennent pas compte de la nécessité de produire ou non une correction.
- Les logiciels qui proposent de filtrer la voix de l'apprenant pour la lui faire réécouter ne semblent pas corriger les aspects prosodiques (rythme et intonation) pourtant essentiels à la structuration de la parole.
- Ces logiciels sont opaques dans leur fonctionnement (protégé par le ©).
- Leur fondement scientifique est controversé.

Ceci étant dit, ces logiciels occupent une niche didactique car les enseignants sont désemparés face au travail phonétique. Or, la situation des enseignants face à cette problématique ne risque pas de s'améliorer s'ils se trouvent constamment absouts d'y faire face par les institutions pédagogiques. Nous cédons ici la parole à Renard (2002a, p. 6) qui exprime au mieux notre pensée :

[36] Dont l'Université de Toulouse 2 Jean Jaurès il y a quelques années, avant d'y mettre un terme.

> Et pourtant, la phonétique dont ont besoin les enseignants n'est pas la mer à boire, une fois débarrassée de la terminologie linguistique qui opacifie son langage, une fois démystifiée[37] et *reliée à la pratique*[38].

En réponse à ces constats sur l'évanouissement du travail phonétique et cette assertion de Raymond Renard, nous évoquerons quelques pistes, à notre sens, explicatives de l'effacement du rôle de l'enseignant dans le travail phonétique. Pour ce faire, nous considérerons quelques aspects de la MVT et de la méthode articulatoire (MA).

1.2 Un point de vue partisan à l'encontre de la méthode articulatoire

Les défenseurs de la MVT (dont nous faisons partie) ont l'habitude dans leurs écrits de comparer les préceptes des deux méthodes afin de souligner la supériorité de la MVT sur la MA. Nous ne nous prêterons au jeu que très brièvement.

Historiquement, la MVT est ancrée dans le courant méthodologique structuro global audiovisuel (SGAV) dont la spécificité est d'envisager que l'apprentissage langagier est une activité structurante, requérant une assimilation des principes de la langue nouvelle par approximations successives. Le point de vue SGAV sur l'apprentissage des langues a conduit à un rejet fort des méthodes traditionnelles dont la MA, les exercices à base d'enregistrement en laboratoire de langue ou reposant sur la binarité des oppositions phonologiques (Renard, 2002b, p. 13-14).

Les griefs portés à l'encontre de la MA par les verbo-tonalistes concernent :

- La négligence de l'aspect perceptif en lien avec le geste articulatoire.
- L'absence de prise en compte des phénomènes de coarticulation des phonèmes.
- Le présupposé selon lequel connaître la position des articulateurs permet une bonne production des phonèmes, en ignorant alors la possibilité de compensation articulatoire.
- Le désintérêt des aspects prosodiques et de la prise en compte globale du corps dans la phonation.

[37] Nous soulignons
[38] Mis en valeur par l'auteur

Nous renvoyons le lecteur intéressé par ces aspects vers Alazard (2013), Billières (2002), Billières *et al.* (2013) et Renard (2002b) pour une analyse plus poussée des faiblesses de la MA.

A contrario, la grande force de la MA réside :

- Dans son postulat simple et intuitif : les apprenants n'arriveraient pas à prononcer les sons correctement car ils ne parviennent pas à les articuler.
- Dans les nombreux matériaux théoriques et pratiques disponibles pour mettre en œuvre les activités de MA, largement diffusés du fait de l'idée initiale de la MA.

Ainsi, la simplicité du propos de la MA lui assure une bonne diffusion. En effet, les connaissances phonétiques requises de l'enseignant pour mettre en place des activités de MA se résument à l'articulation des phonèmes. Celle-ci est par ailleurs bien documentée et disponible dans de nombreux ouvrages appliqués à l'enseignement de la L2.

1.3 Quelques forces et faiblesses de la MVT découlant de son postulat

Malgré ses aspects novateurs, pratiques et sa souplesse d'emploi, la MVT ne bénéficie pas (ou peu) des avantages de la MA.

Théoriquement, le verbo-tonaliste averti (nous entendons : formé et opérationnel) n'a pas besoin de matériel autre que son corps (procédés posturo-mimo-gestuels) et sa voix (procédés vocaux) pour produire une remédiation efficace.

La MVT fonde son propos sur le principe de crible phonologique (Polivanov, 1931; Troubetzkoy, 1939) selon lequel les sons des langues étrangères reçoivent une interprétation phonologiquement inexacte, puisqu'ils seraient filtrés par un système calibré pour la langue maternelle. Cette observation se trouve métaphorisée en MVT par le concept de surdité phonologique qui considère que les apprenants de L2 adultes se comportent comme des « *durs d'oreilles* » (Guberina, 1978, 1991).

La MVT, objet didactique et donc forcément opportuniste, se contente de cette métaphore simplifiant les nuances que l'on peut trouver dans les théories de la perception de la parole en lien avec la catégorisation des sons de parole dont celles développées par Best (1994) ou Kuhl *et al.* (2008) constituent de bons exemples.

Ainsi, la MVT postule que l'erreur phonétique en L2 est due au biais perceptif de l'apprenant, lié au crible phonologique de sa L1. En conséquence, la MVT se retrouve dans l'impossibilité de produire des supports didactiques standardisés puisque l'erreur phonétique qui dépendrait de l'apprenant nécessiterait alors une remédiation personnalisée.

Par ailleurs, la MA propose des exercices mécaniques, décontextualisés de l'erreur et anticipe des erreurs alors qu'il peut ne pas y en avoir. La MA applique donc un adage de type « *prévenir, c'est guérir* ». En opposition, l'usage de la MVT se veut « *diffus dans l'apprentissage de la langue* » et la mise en place d'un procédé de remédiation n'est qu'une réponse à une erreur effectivement produite dans la communication (Renard, 2002b, p. 15).

Ainsi, la remédiation phonétique MVT repose entièrement sur l'expertise du praticien verbo-tonaliste, qui doit diagnostiquer au cas par cas les erreurs des apprenants. A notre sens, cette dernière remarque illustre la force de la méthode en tant qu'objet didactique pouvant être mis en œuvre à tout moment du cours pour répondre à tout type d'erreur phonétique.

Elle met aussi en lumière un problème fondamental quant à sa diffusion et sa vitalité dans les sphères didactiques actuelles : si les enseignants n'y sont pas formés, cette pratique ne peut ni se développer ni être reconnue à juste titre ; la MVT reste trop souvent un secret d'initiés.

Enfin, la MVT en tant que pratique phonétique doit être « démystifiée », comme nous le soulignions plus haut en citant Raymond Renard. C'est également un enjeu pour sa diffusion et sa compréhension par le public des enseignants : la table ronde sur la MVT[39] organisée le 4 juin 2013 (Billières *et al*., 2013) propose une réflexion collective autour du terme « magie » accolé à la MVT.

Bernard Harmegnies y propose une réflexion fort intéressante, que nous synthétisons ici :

- Un caractère « *magique* » de la MVT est de produire des effets manifestes, et cela rapidement.
- Cela pourrait relever du « *miracle* » pour l'apprenant, qui n'est pas au fait des rouages à l'œuvre en MVT.
- Or, il n'y a en MVT rien de magique, simplement la mise en œuvre de techniques, issues d'une méthodologie fondée sur une réflexion rationnelle.
- Ainsi, la MVT relèverait de la « *prestidigitation* » plutôt que de la magie des contes de fées, puisque le praticien peut fournir une explication technique des effets produits.

[39] Disponible ici : goo.gl/LLbQos

- Contrairement au prestidigitateur, le verbo-tonaliste veut et doit dévoiler ses techniques à son public.
- Il doit également constituer une alliance avec le monde scientifique pour légitimer ses pratiques en les exposant à la communauté académique.

Ce dernier point nous semble fondamental : si, en didactique, l'efficacité pratique précède le théorique (et son dogmatisme) (Renard, 2002b, p. 21), il importe en revanche que la MVT en tant que construction théorique soit éprouvée et criblée par la recherche scientifique.

Ainsi, la MVT serait dévoilée, démystifiée et légitime aux yeux de l'enseignant chez qui le préjugé de « sorcellerie » phonétique a encore la vie dure.

De nos jours, les méthodologies rigoureuses d'enseignement ont laissé la place à un éclectisme opportuniste, autorisant l'enseignant de L2 à piocher à loisir dans les méthodes en fonction de ses objectifs (Alazard, 2013, p. 34).

Dans ce contexte, il devient alors difficile de ne pas considérer la formation des enseignants à la MVT comme un enjeu majeur dont la résolution permettrait de ne plus laisser de côté la problématique du travail phonétique. En dehors de son ancrage SGAV, la MVT présente la flexibilité nécessaire pour s'adapter à l'ambiance didactique actuelle, ainsi qu'aux contraintes des acteurs du milieu.

Nous nous proposons en suivant de situer les pratiques de phonétique corrective au plus près de leur réalité en décrivant les partis impliqués dans ce type d'interaction didactique, les enjeux pour chacun des partis, et les moyens à dispositions des uns et des autres pour atteindre leur but. Nous envisagerons plus spécifiquement la mise en œuvre de la MVT mais considérerons sporadiquement le point de vue de la MA.

2. Apprentissages et corrections phonétiques au travers du prisme de l'imitation : nature de l'interaction, dimension cognitive et enjeux pratiques

Le but de la correction phonétique est de parvenir à corriger une erreur de prononciation d'un apprenant, soit, de lui faire produire un geste phonatoire dont le résultat sera perçu par l'enseignant comme phonologiquement juste.

La correction phonétique est donc une pratique orale où l'enseignant produit un modèle à destination de l'apprenant qui, à son tour, doit le répéter. Si l'enseignant estime que l'output de l'apprenant est erroné, un nouveau modèle est produit, capté par l'apprenant, redit, rejugé, *etc*. En d'autres termes, la correction phonétique est une suite d'interactions orales à visée imitative que l'on peut modéliser à l'aide d'un schéma issu de domaines non-linguistiques (Figure 11).

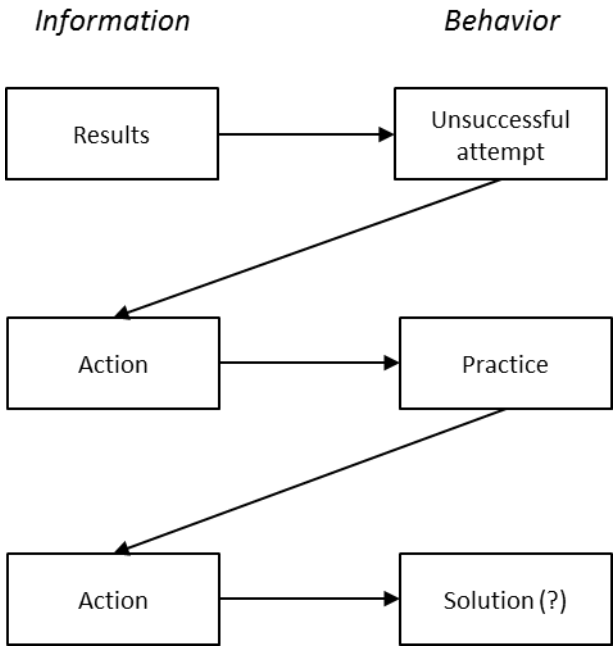


Figure 11 : "Shifting between source of information in a social learning task", d'après Call & Carpenter, (2002, p. 220). Les cadres de la colonne de gauche représentent la source d'information de l'imitateur. A droite, le comportement de l'imitateur (essai infructueux, entraînement, éventuellement solution). L'imitateur peut d'abord observer le résultat de l'action, puis à mesure de ses essais, porter son attention sur différents aspects des actions qui conduisent à ce résultat.

Indépendamment de toute méthodologie corrective, la situation de remédiation phonétique en classe présente donc les caractéristiques d'une situation d'imitation telle que nous les avons décrites précédemment : modèle (sujet et action produite) et imitateur.

En appliquant cette modélisation à la correction phonétique, le cadre « résultat » équivaut au son que l'apprenant doit imiter. Après un essai infructueux, l'enseignant produit une action (éventuellement en lien avec les préceptes d'une méthodologie de correction, MA ou MVT), afin de fournir un supplément d'information à l'apprenant. Il peut s'ensuivre plusieurs échanges de ce type pour atteindre (éventuellement) une solution, *i.e.* un essai fructueux.

Envisagée comme une succession d'imitations, la situation de correction phonétique fait émerger à nouveaux les questionnements à l'origine de la correction phonétique. En les posant en termes imitatifs, nous pouvons nous demander :

1. Pourquoi les essais de l'apprenant sont-ils infructueux, soit pourquoi les imitations du geste phonatoire sont-elles rarement concluantes ?
2. Quelles actions mettre en œuvre pour améliorer l'*output* imitatif ?

Répondre à la première question revient à prendre position sur la deuxième question : c'est en estimant la source de l'erreur phonétique que l'on peut chercher à proposer des solutions pour l'enseignant, et que l'on se place alors dans une perspective méthodologique de correction.

En termes d'imitation, la question 1 peut être traitée de deux manières différentes (et corollairement, la question 2) :

1. Nous pourrions considérer que l'erreur phonétique est produite par l'apprenant car le geste lui est inconnu. Il lui est en effet difficile de l'observer, car la bouche de l'enseignant n'est pas transparente : le geste articulatoire présente une opacité qui rendrait difficile à l'apprenant de traiter le problème de correspondance.
    - En considérant que l'erreur phonétique est liée au geste articulatoire, le praticien se positionne dans la tradition de la MA et envisage des actions agissant sur le système moteur de l'apprenant en éliminant explicitement l'opacité du geste à reproduire.
2. L'autre possibilité que nous envisageons est de supposer que le geste importe autant ou moins que le percept auquel il est rattaché. Même en ayant une connaissance du

geste à effectuer, l'apprenant produirait une erreur car ce geste ne signifierait perceptivement rien pour lui.

- Postuler que l'erreur phonétique en L2 est liée à un problème perceptif place l'enseignant dans la perspective de la MVT. Les actions mises en œuvre en MVT ont donc pour but premier d'élargir la capacité de traitement phonologique de l'apprenant par un entraînement sensorimoteur.

En fonction du point de vue considéré, celui de l'apprenant ou de l'enseignant, les enjeux de la situation de remédiation phonétique ainsi que ses aspects imitatifs diffèrent.

2.1 Correction phonétique, expérience sensorimotrice et crible phonologique : enjeux pour l'apprenant

Si nous relions ces remarques aux propositions de l'*Associative Sequence Learning Hypothesis* (Brass & Heyes, 2005; Heyes, 2001; Heyes et al., 2005), nous pourrions considérer que les procédés de la MA visent à former des associations verticales indirectes tandis que ceux de la MVT tentent de former des associations verticales directes (Figure 12).

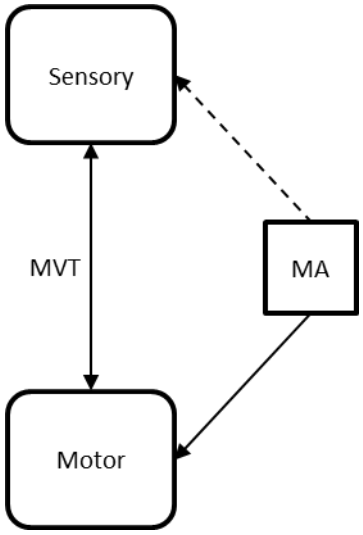


Figure 12 : Associations verticales entre représentations motrices et sensorielles et nature des liens en fonction des méthodes de correction : MVT vs. MA

En effet, la MA privilégie une approche explicite de la phonation. Avant d'être produits par les apprenants, les sons leur sont présentés au moyen de schémas anatomiques et fonctionnels détaillant la configuration des articulateurs, si les cordes vocales produisent une vibration, *etc*. Après ces explications et un peu de gymnastique articulatoire, les sons ciblés sont produits dans des exercices allant du plus simple (sons isolés) au plus complexe (sons en contexte dans des mots, puis des phrases). Ainsi, la MA propose en premier lieu une approche centrée sur le mouvement articulatoire, au détriment du canal auditif, pourtant intimement lié aux gestes de la phonation (Alazard, 2013; Billières, 2000). En d'autres termes, pour la MA, la représentation phonologique vient après la représentation motrice.

En ce qui concerne la MVT, l'approche prône un apprentissage implicite de la prononciation en contexte de communication. Puisque l'on considère que la source de l'erreur phonétique est perceptive, le but des procédés de la MVT est de faire percevoir une différence entre ce que l'apprenant a produit et ce qu'il devrait produire dans le contexte d'apparition de l'erreur. A mesure des essais de l'apprenant, le verbo-tonaliste produit donc des modulations vocales (suprasegmentales et/ou segmentales) qui ont pour but de guider perceptivement l'apprenant vers le son (et le geste) juste.

En conséquence, la MVT propose une intégration sensorielle des schémas moteurs, par approximations successives en liant systématiquement expérience sensorimotrice de l'apprenant et *feedback* adapté produit par l'enseignant. Ce faisant, la MVT peut être positionnée dans une perspective similaire à celle de la théorie PACT (Schwartz et al., 2012) qui considère que les gestes paroliers sont des unités communicatives modelées par la perception (*perceptually-shaped units*).

Dans PACT, la dimension communicative de l'acte phonatoire revêt en effet une importance particulière, dans la mesure où la parole est produite à cette fin. Ainsi, les unités perceptuo-motrices envisagée par PACT dans la communication sont caractérisées par une valeur double :

- Cohérence articulatoire, qui est due à sa nature gestuelle.
- Valeur perceptive, qui est nécessaire pour avoir une valeur fonctionnelle.

Or, l'objectif des procédés de la MVT est de faire remarquer à l'apprenant la valeur fonctionnelle de l'unité ciblée par la correction, cela, dans le contexte original de la communication dont est issue l'erreur phonétique.

Dans le chapitre précédent (pp.86-88), nous postulions que la dichotomie entre représentations motrices et sensorielles proposée par l'*Associative Sequence Learning* pourrait ne pas avoir cours pour les unités parolières si nous adoptions la perspective de PACT, sauf à supposer que ces liaisons se forment très tôt dans le développement. Si tel était le cas, les liaisons entre représentations motrices et perceptives seraient particulièrement fortes, au point de fusionner en unités perceptuo-motrices.

En ce qui nous concerne, envisager ainsi les représentations des unités de parole constitue un renforcement de la notion de crible phonologique chère à la MVT. En enracinant ensemble représentations motrices et perceptives de la parole, une activité quotidienne, il semble alors particulièrement mal aisé d'affranchir l'activité parolière en L2 des réseaux d'activation de la L1, de ces habitudes sensorimotrices.

Une telle vue fait écho à la métaphore d'émulation cognitive que nous développions à propos de l'apprenant de L2 durant notre chapitre premier. En percevant les sons d'une L2, l'apprenant active à la fois des représentations phonologiques erronées, mais également des patterns moteurs non-pertinents pour la communication en L2. Ceci mettrait l'apprenant débutant dans une situation délicate, équivalent peu ou prou à une situation de handicap, puisque le système sensorimoteur de sa L1 se trouve à la fois inadapté pour faire face efficacement au traitement de la L2 mais aussi le seul sur lequel il puisse se reposer pour tenter de la traiter. Cet état conduirait à une imitation défectueuse des sons de la L2.

2.2 Du diagnostic de l'erreur à la production d'une remédiation ciblée : focus sur l'enseignant

Les pratiques de remédiation phonétique doivent donc permettre à l'apprenant d'élargir ses possibilités de traitement de la L2 en intégrant à son système des représentations phonologiques et des patterns moteurs nouveaux. Ici, nous développerons les procédures de remédiation proposées par la MVT à cette fin. Ce faisant, nous soulignerons d'une part une grande force de la MVT, sa flexibilité d'emploi ainsi qu'une de ses faiblesses d'autre part, sa difficulté pour l'enseignant débutant

2.2.1 Instantané de la situation de remédiation en MVT

In vivo, la correction phonétique par la MVT, en tant que pratique diffuse dans l'enseignement, doit donner lieu à des échanges d'une durée réduite entre l'enseignant et l'apprenant. Cet échange démarre lorsque l'enseignant remarque une prononciation défectueuse (segmentale ou suprasegmentale) sur laquelle il aimerait agir.

Après avoir fait un diagnostic de l'erreur, l'enseignant propose un stimulus correctif à l'apprenant, en prenant pour principe de placer le son ciblé dans un contexte optimal pour que l'apprenant parvienne à le percevoir. L'enseignant « crée » ce contexte optimal en produisant lui-même des modulations vocales de l'énoncé original au moyen de plusieurs types de procédés :

- Modulations prosodiques (procédé privilégié)
- Prononciation nuancée de la cible (allophones)
- Manipulation segmentale de l'entourage phonétique de la cible
- Indications posturo-mimo-gestuelle.

En fonction des reproductions de l'apprenant, l'enseignant module ses propres productions pour finir la séquence en proposant l'énoncé original à l'apprenant. En d'autres termes, les échanges de remédiation en MVT voient à chaque nouvelle proposition de l'enseignant, l'ajout, la suppression ou la modulation d'un procédé déjà utilisé. Ainsi, chaque nouvelle proposition de l'enseignant diffère de la proposition précédente. Ceci peut être modélisé comme suit :

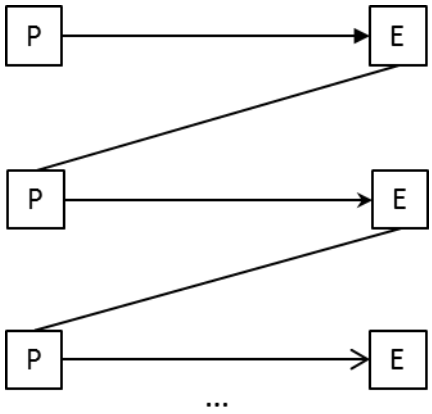


Figure 13 : Modélisation d'une situation de correction phonétique.P représente le praticien et E l'étudiant. Les différents types de flèches issus de P suggèrent les modulations successives produites par P afin de guider E dans sa reproduction.

Cette représentation est très proche de celle que nous avons donnée au début de cette partie, cependant, nous y explicitons le choix rationnel de l'enseignant (selon les principes de la MVT) qui le conduisent à ajouter une nouvelle source d'information en direction de l'apprenant.

Ici réside un caractère paradoxal de la MVT (un aspect « magique ») pour qui n'y est pas familier : la modulation produite par l'enseignant de manière à guider l'apprenant vers le son cible n'a pas vocation à être imitée telle quelle, même si cela peut se produire. Un observateur extérieur partageant le même système phonologique que l'enseignant jugerait en effet que la modulation produite est volontairement fausse. C'est la raison pour laquelle le verbo-tonaliste doit expliquer ces techniques qui semblent contre-intuitives, ce que nous proposons de faire en suivant.

2.2.2 Techniques de remédiation verbo-tonales

Dans l'explication des techniques de remédiation de la MVT, il est couramment fait références à deux notions opératoires qui permettent de poser le diagnostic de l'erreur segmentale et de construire, en relation avec ce diagnostic, un énoncé correctif. Ces deux notions, la tension et le timbre, servent à classifier consonnes et voyelles :

- *La tension* peut se définir comme « *l'énergie neuro-musculaire dépensée pour produire la parole* » (Renard, 1979, p. 127). La tension est reliée au domaine physiologique, puisque classer les sons en fonction de leur tension revient simplement à les classer en fonction de leur articulation.
    - La MVT considère que les erreurs de prononciation des consonnes sont essentiellement dues à une Tension trop forte (erreur T+) ou trop faible (erreur T-).
    - L'erreur de tension peut aussi être considérée dans la prononciation des voyelles
- *Le timbre* d'un son fait référence à sa composition spectrale (soit, la répartition des fréquences dans le spectre). C'est une notion liée à l'impression subjective qu'il produit chez un individu (Billières *et al.*, 2013). La MVT classe les sons vocaliques en fonction de leur timbre en opposant les sons :
    - Plus clairs, *i.e.* favorables aux fréquences hautes

- Plus sombres, *i.e.* favorables aux fréquences basses
- Les erreurs de timbre sont notées C+ (pour son trop Clair) ou C- (son trop sombre)

En MVT, il est donc considéré que l'erreur segmentale est le résultat d'une tension ou d'un timbre inadapté. Le premier pas de la remédiation phonétique est donc d'arriver à porter un diagnostic précis de l'erreur produite par l'apprenant, et ceci, en temps réel.

Une fois le diagnostic posé, le praticien verbo-tonaliste peut mettre en place un énoncé correctif au moyen des procédés suivants, que nous présentons de manière isolée mais dont le cumul est possible. Ces procédés sont présentés de manière plus détaillée, accompagnés d'exemple dans la ressource en ligne de l'UOH (goo.gl/18l4wW) à laquelle nous avons participé (Billières *et al.*, 2013) :

- Modulation prosodique :
  - Ce procédé est privilégié en MVT, car il dénature moins la chaîne parolière originale.
  - Il s'agit d'utiliser l'influence du mouvement rythmico-intonatif pour influencer le timbre ou la tension du son cible.
    - Par exemple, un ralentissement du débit diminue la tension, de même qu'une désaccentuation ; inversement, accélération et accentuation augmentent la tension (et éclaircit le timbre dans le cas de l'accentuation.) (notion de *spectral tilt*, Campbell & Beckman, 1997)
    - Le mouvement intonatif montant tend à augmenter la tension et à éclaircir le timbre des voyelles, et inversement pour un mouvement descendant.
- Prononciation nuancée :
  - Pour la production de ce procédé, le son cible est produit avec un timbre intermédiaire à l'opposé de l'erreur produite par l'apprenant.
    - Par exemple, si l'apprenant produit une voyelle C- par rapport à la cible, l'enseignant proposera une voyelle C+, avant de revenir pas à pas vers la cible.
  - Il est également possible de substituer momentanément le son cible par un autre son (toujours à l'opposé de l'erreur), puis de revenir vers le son cible petit à petit.
- Entourages facilitant

    - Ce procédé utilise les effets de la coarticulation et de la combinaison des sons au sein de la syllabe ($C_1V\ C_2$) où :
        - C_1 influence le timbre de V
        - V a un effet sur la tension de C_1
        - C_2 tend à raccourcir ou allonger C_2
    - En manipulant l'entourage du son cible, l'enseignant peut donc mettre en valeur les propriétés du son cible.
- Gestualité accompagnatrice
    - Elle accompagne la production des autres procédés et peut donner une indication visuelle sur la proprioception.
        - Par exemple, on peut indiquer le relâchement du corps pour indiquer à l'apprenant qu'il a produit un son T+
    - Elle est particulièrement appropriée pour décrire le mouvement mélodique et le découpage en groupes rythmiques.

Il importe particulièrement de rappeler que la production de ces procédés peut (et doit souvent) se faire de manière cumulée et à la volée. En pratique de classe, l'enseignant n'a pas le temps de jeter un œil sur son *vade-mecum* du verbo-tonaliste. Il doit pouvoir mobiliser immédiatement les connaissances requises pour diagnostiquer puis remédier à l'erreur de l'apprenant. Ces connaissances comprennent :

- Le classement des consonnes sur l'axe de la tension
    - Diagnostic et prononciation nuancée/déformée
- Le classement des voyelles sur l'axe de la tension et l'axe clair/sombre
    - Diagnostic, prononciation nuancée/déformée et entourage facilitant (effet de V sur C_1)
- Le classement des consonnes sur l'axe clair/sombre
    - Entourage facilitant (effet de C_1 sur V)
- Le degré de résistance des consonnes en position finale de syllabe
    - Entourage facilitant (effet de C_2 sur V)

Pour être maîtrisée, la pratique verbo-tonale demande donc à l'enseignant de se former pour intégrer ces principes de correction, qu'ils lui viennent naturellement et qu'il parvienne à les produire sans temps mort, comme les katas du karatéka. En tant qu'enseignant débutant, la tâche n'est pas aisée d'autant plus que l'enseignant met en jeu sa crédibilité dans cette

pratique… Ceci étant dit, une fois l'enseignant expert dans la pratique verbo-tonale, il devient capable de faire face avec beaucoup de souplesse à tout type d'erreur phonétique, en ayant l'espoir de produire un effet bénéfique sur la prononciation de l'apprenant.

Nous proposons en Figure 14 un schéma récapitulatif de la correction des phonèmes.

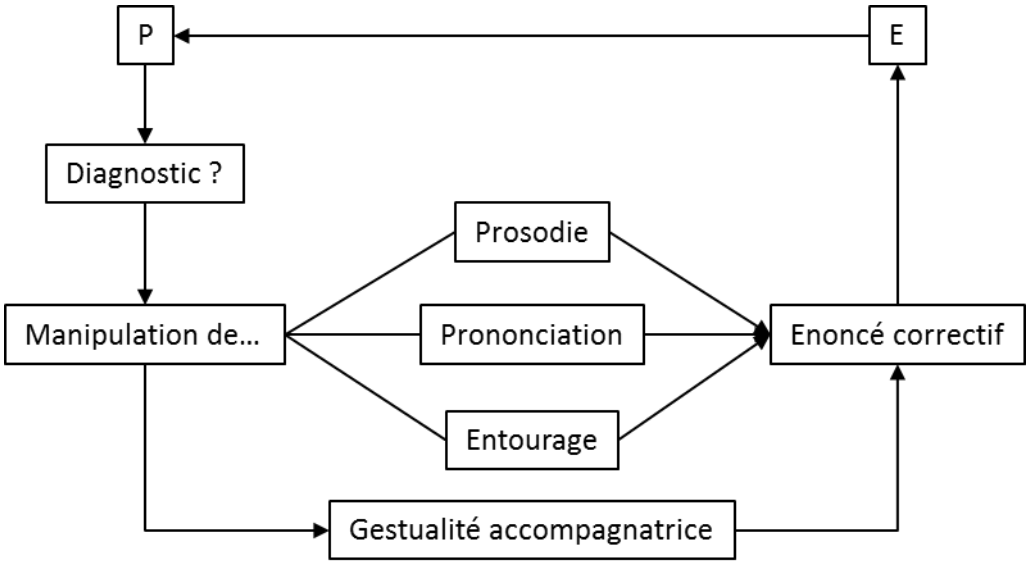


Figure 14 : Cycle de l'interaction MVT pour la correction des phonèmes, celui-ci se répète à chaque nouvelle production de l'étudiant (E). Le professeur (P) module sa production en temps réel, en fonction de la production de E.

En plus de la réflexion liée au diagnostic, les manipulations prosodiques et de prononciations nuancées demandent à l'enseignant une maîtrise particulière de sa phonation pour être réalisées de manière efficace. En plus de douter sur son analyse de la situation, l'enseignant débutant la MVT peut être amené à douter de ses capacités vocales, s'il n'a pas eu l'occasion de suivre un des rares stages pratiques de formation existant.

Bien que nous ayons reçu une formation théorique et pratique à la MVT dans le cadre de nos études universitaires, nos premiers pas en tant que praticien verbo-tonaliste se sont d'abord limités aux manipulations prosodiques les plus basiques et aux prononciations nuancées les plus intuitives.

Par ailleurs, doutant de notre compétence nous n'avons pas osé aborder le travail sur l'intonation et le rythme que propose la MVT au moyen des logatomes. Le logatome, une suite syllabique sans signification, est destiné à présenter spécifiquement le rythme et l'intonation à l'apprenant (Billières, 2002, p. 53).

Pour désactiver la charge sémantique des énoncés, l'enseignant a plusieurs procédés à disposition (Billières, 2002, p. 54) :

- Le « dadada » qui consiste à substituer à chaque syllabe un « da »
- Le « *mumming sound* » en utilisant le son [m]
- La voix logatomique qui, métaphoriquement, équivaudrait à parler avec une grosse boule de coton dans la bouche.

La production du logatome peut précéder un énoncé sémantiquement chargé pour présenter le contour intonatif à venir aux apprenants. Lors du logatome, ils peuvent ainsi s'imprégner de la prosodie de l'énoncé et déjà en dégager la structuration.

En ce qui concerne l'enseignant, la production de logatomes n'est pas un exercice évident. En effet, les logatomes doivent respecter au mieux les caractéristiques prosodiques de la phrase originale, soit :

- Comporter le même nombre de syllabes
- Avoir un patron intonatif le plus similaire possible

Dans le cas d'énoncés courts, le nombre de syllabe n'est pas réellement problématique, mais dès que celui-ci excède cinq syllabes, des omissions peuvent survenir. Enfin, la question de la reproduction du patron intonatif mérite d'être soulevée : dans la mesure où la production du logatome s'accompagne, pour l'enseignant aussi d'une perte de sens, on peut se poser la question de savoir si le patron intonatif logatomique est fidèle à l'original.

En résumé, la MVT propose à l'enseignant de nombreuses solutions pour avoir une action sur la prononciation de l'apprenant, tant au niveau segmental qu'au niveau prosodique. Ces procédés visent à contrecarrer les effets du biais perceptif de l'apprenant au moyen de modulations produites par l'enseignant. Ceci étant dit, produire ces modulations demande à l'enseignant une maîtrise élevée de sa parole :

- Au niveau segmental, pour produire des prononciations nuancées et/ou déformée
- Au niveau prosodique, pour manipuler le contexte des segments visé et pour produire des logatomes.

Ce dernier point soulève une question intéressante en termes d'imitation parolière, puisque le logatome se veut l'imitation prosodique d'un énoncé. Savoir évaluer la similarité prosodique d'un modèle et de son imitation parolière au niveau prosodique constitue une piste intéressante pour proposer par la suite des outils d'entraînement au verbo-tonaliste.

Chapitre 4 : Problèmes méthodologiques de l'étude de l'imitation en parole

La question de la méthodologie dans les études sur l'imitation en parole est particulièrement épineuse. L'imitation est un phénomène complexe, tandis que la parole est un objet d'étude aux dimensions multiples. L'imitation résulte d'une interaction : elle consiste, pour un sujet, à « *reproduire tout ou partie d'un comportement perçu chez un autre sujet, de manière à ce qu'un tiers puisse percevoir la production de l'imitateur comme ressemblant à celle du modèle* ». Cette définition de l'imitation, que nous estimions être une définition *a minima*, puisqu'elle évacue toute notion de fonction qui pourrait court-circuiter la simple définition physique de l'imitation, laisse déjà entrevoir de potentielles complexités méthodologiques… Il faut en premier lieu tenir compte du modèle et de ses caractéristiques qui seront l'objet –pour tout ou partie– de l'imitation. Il faut également pouvoir estimer la réussite de l'imitation : une imitation en est une si et seulement si un tiers perçoit la production de l'imitateur comme semblable à celle du modèle. Il faut de plus, dans notre cadre d'étude de la parole, pouvoir comparer différentes productions aux niveaux qui nous intéressent, *i.e.* phonologiques et phonétiques. Ce chapitre sera donc consacré aux aspects méthodologiques de l'étude de l'imitation en parole.

Les problèmes posés par ce questionnement des méthodes sont nombreux :

- L'étape du recueil des données, qui, en fonction du phénomène d'imitation que l'on souhaite étudier (convergence phonétique, *impersonnation*[40], imitation phonétique), nécessitera la mise en place de tâches expérimentales spécifiquement calibrées en vue de faire émerger le phénomène à observer : souhaite-t-on observer une parole mimétique (un changement incontrôlé de la manière de prononcer du sujet suite au contact avec l'autre) ou bien une imitation de la parole de l'autre (la tentative d'un sujet de se faire passer pour ou de copier l'autre).
- Des données de parole recueillies, il faut pouvoir *estimer, évaluer ou mesurer* la quantité (s'il y a) d'imitation produite par les sujets expérimentaux.

[40] Nous utilisons ce terme pour le sens suivant : quand un locuteur vise à se faire passer pour un autre. Les imitateurs professionnels relèvent, pour nous, de l'impersonnation.

- Sur quel(s) critères(s) peut-on se baser pour dire qu'il y a « imitation » dans les données que nous obtenons ?
- Quelle(s) mesure(s) acoustique(s) effectuer sur les itérations produites par les sujets qui imitent et comment les interpréter ?

Ces questions, récurrentes dans la littérature sur l'imitation en parole, ont occupé une part importante de notre réflexion car elles constituent un défi permanent dans l'étude des imitations, ce que rapporte Pardo (2013b) :

> Results from multiple studies examining phonetic convergence offer an array of often confusing and disparate findings. Reconciling such diverse findings is difficult without a clear rationale for engaging in one acoustic measure over another.

Cette assertion sur la convergence phonétique peut s'appliquer sans réelle restriction à toute étude sur les phénomènes imitatifs en parole. Par conséquent, nous tenterons d'expliciter les considérations théoriques qui ont abouti à nos choix méthodologiques

Ainsi, l'objectif de ce chapitre sera double : dans un premier temps, nous proposerons une revue méthodologique des tâches expérimentales choisies par la communauté scientifique pour faire émerger les phénomènes d'imitation en parole. Par la suite, nous rapporterons les approches utilisées dans la littérature pour tenter d'évaluer et de mesurer les effets se rapportant à l'imitation en parole. Celles-ci sont diverses et relèvent des tests psycholinguistiques, des mesures acoustiques et des techniques de traitement du signal. Dans un troisième temps, nous discuterons les apports de ces approches multiples et expliciterons celle(s) que nous retiendrons dans nos pratiques expérimentales. Il semblerait en effet que la combinaison de plusieurs approches soit le meilleur moyen de rendre compte des effets d'imitation phonétique en parole.

1 Faire émerger l'imitation(s) en parole : designs expérimentaux pour le recueil de données

Les productions imitatives résultent de l'interaction de plusieurs paramètres : il faut un modèle qui produit un comportement ainsi qu'un sujet qui perçoit l'action du modèle afin que cette dernière soit reproduite. Ainsi, les designs expérimentaux visant à étudier l'imitation en parole vont chercher à mettre les sujets dans un contexte favorable à la production de

comportements d'imitation. Cependant, les tâches choisies pour faire émerger l'imitation en parole vont dépendre du phénomène que le chercheur souhaite observer. Il sera ainsi possible de distinguer différents paradigmes expérimentaux en fonction de la nature du comportement imitatif cible de l'expérimentation. D'un côté, nous trouverons les expérimentations qui s'intéressent à la convergence phonétique (Giles et al., 1991). De l'autre côté, certains paradigmes expérimentaux se concentrent sur une imitation, dite « de laboratoire », qui propose d'observer les comportements d'adaptation à sens unique, puisque leurs occurrences se produisent chez un humain face à une « machine » parlante (entendre : l'équivalent moderne d'un magnétophone). Le comportement ainsi observé serait fondamentalement différent, la dimension interactive propre à la convergence en étant absente.

Cette première partie vise à faire le point sur les différentes tâches proposées en contexte expérimental pour que des sujets reproduisent tout ou partie de ce qu'ils entendent (convergence phonétique ou imitation de laboratoire).

1.1 Tâches conversationnelles et convergence phonétique

A mesure de nos interactions avec notre environnement, il se peut que notre comportement langagier change au point que nos partenaires habituels de communication trouvent notre manière de parler différente (Goldinger, 1998). Ce phénomène, peut affecter différents niveaux de notre production linguistique, comme le niveau lexical (Garrod & Doherty, 1994) ou le niveau phonétique (Namy et al., 2002; Pardo, 2000, 2006). Quand celui-ci est particulièrement affecté, nous parlerons alors de convergence phonétique. Ce comportement serait assez incontrôlé de la part de ceux qui le produisent, mais il pourrait dépendre des objectifs sous-jacents du locuteur qui prend part à la communication. En ce sens, il est possible qu'un sujet converge ou ne converge pas.

Les pratiques expérimentales qui ont pour but d'observer ce phénomène essaient donc de réunir des conditions favorables pour son apparition en proposant des protocoles qui permettent :

- De connaître le comportement langagier normal des différents locuteurs engageant la conversation
- D'obtenir des données exploitables pour évaluer et mesurer le degré de convergence de chaque locuteur

Ainsi, les paradigmes expérimentaux explorant la convergence phonétique proposent des tâches de nature conversationnelle (la convergence est vue dans la *Communication Accomodation Theory* comme une stratégie de communication) auxquelles sont souvent associées des pré- et des post-tests (Natalie Lewandowski, Jilka, Rota, Reiterer, & Dogil, 2007; MacLeod, 2014; Pardo, 2000, 2006; Schweitzer & Lewandowski, 2014).

Les pré-tests permettent d'obtenir des échantillons « naturels » de la manière de parler des locuteurs qui seront ensuite comparés aux items recueillis durant la tâche de conversation et le post-test. Le post-test, souvent de même nature que le pré-test, est fait pour observer si le comportement phonétique du locuteur a changé durant la tâche de conversation et si ce nouveau comportement persiste. La nature de la tâche conversationnelle sera détaillée en suivant. Les lignes qui suivent présentent la nature des pré- et post-test.

Comme nous le verrons par la suite, les tâches conversationnelles pour l'étude de l'imitation ont pour but de contraindre les locuteurs à l'utilisation de *tokens* cibles (des items lexicaux, de longueurs variables). Les pré- et post-tests sont alors construits de manière à obtenir des itérations de ces *tokens*. Paradoxalement, bien que la convergence phonétique arrive normalement en contexte de conversation, certains chercheurs, comme Macleod (2014), utilisent une tâche de lecture de phrases comme tests préliminaires. D'autres, comme Pardo (2006) font prononcer les *tokens* cibles dans diverses phrases préconstruites (« Dites, XX, encore une fois » ; « Le numéro X est le XX »), afin d'éviter d'avoir un type de parole trop différent de la parole spontanée. Au niveau de la temporalité, ces pré- et post-tests sont parfois administrés juste avant ou juste après les tâches conversationnelles, mais les délais de passation de ces phases expérimentales varient de quelques instants à plusieurs semaines (Pardo, 2006).

Les paradigmes expérimentaux en parole s'appuient donc sur la différence entre les items de différentes phases expérimentales pour attester –ou non– de la convergence phonétique : le chercheur espère ainsi qu'un *token* produit au cours d'un pré-test aura des caractéristiques différentes d'un *token* produit plus tardivement dans l'expérimentation, en d'autres termes, qu'il se produira un glissement dans la manière de parler du locuteur : un *shift*[41]. Quel que soit le type d'imitation en parole étudié, cette notion de *shift* sera centrale : s'il va en direction du modèle, on pourrait alors dire que le paramètre mesuré converge, en cas

[41] Quand nous parlerons de « shift », nous ferons référence à des paramètres mesurés sur le plan acoustique, par exemple temporel ou en terme de hauteur.

d'absence de *shift*, le locuteur maintiendrait son comportement normal, si le *shift* se produit à l'opposé du comportement du modèle, il y aurait alors divergence…

1.1.1 Map Task

La Map Task constitue une tâche particulièrement privilégiée dans les études sur la convergence phonétique, même si, originellement, cette tâche a simplement été conçue pour étudier la parole spontanée (Anderson et al., 1991; Brown, Anderson, Yule, & Schillcock, 1983) en « cadrant » cette dernière : durant cette tâche, le topic et l'objectif de la conversation sont connus de l'expérimentateur. Initialement développée en langue anglaise, la Map Task a été adaptée au français ces dernières années et un corpus multimodal a été recueilli pour le français (Bard et al., 2013 ; Gorisch, Astésano, Bard, Bigi, & Prévot, 2014). Cependant, nous n'avons pas connaissance d'étude de la convergence phonétique en français utilisant ce protocole conversationnel.

Concrètement, la Map Task est une tâche conversationnelle dyadique ayant pour but le partage d'informations au sein de la dyade afin de résoudre un problème commun (tracer un chemin sur une carte). Brown *et al.* (1983) décrivent quatre caractéristiques de la Map Task qui nous semblent d'importance pour le choix de cette tâche dans l'étude de l'imitation.

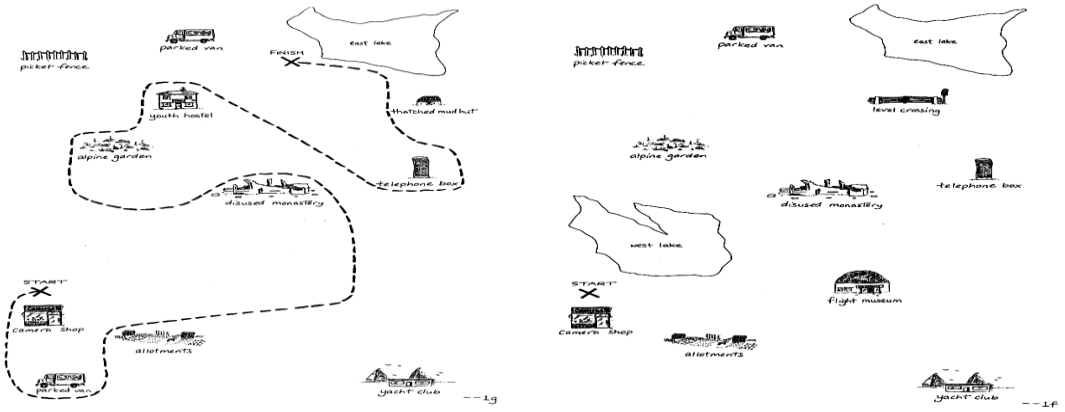


Figure 15 : Exemple de cartes pour la Map Task

1. La Map Task a pour but d'inciter les locuteurs à communiquer dans le cadre d'une tâche coopérative.

Cet aspect coopératif nous semble important pour espérer voir émerger de la convergence : le fait que les locuteurs partagent un but commun devrait favoriser l'apparition du comportement ciblé. Par ailleurs, cet objectif place un des interlocuteurs dans une position de « donneur d'instruction » (celui qui dispose du chemin sur sa carte)

et l'autre interlocuteur dans le rôle du « receveur » (celui qui doit tracer le chemin). Cette distribution peut être alternée si on souhaite éviter un effet social du rôle sur la direction du *shift*, ou au contraire privilégiée (soit, chaque locuteur tient toujours le même rôle) si le chercheur souhaite observer l'effet de la position sociale sur la direction du *shift* (est-ce le donneur ou le receveur qui converge avec l'autre ?).

2. L'accomplissement de la Map Task ne peut se faire qu'au moyen d'une communication fructueuse
3. Le succès de la communication peut être mesuré en comparant le tracé des sujets à la solution attendue de la Map Task, ainsi qu'en observant le temps mis par les sujets pour l'atteinte du but.

Les points (2) & (3) permettent donc d'observer si la performance communicative des sujets est efficace.

4. Les repères topographiques et le chemin de la carte sont fixés à l'avance, assurant ainsi à l'observateur qu'il soit fait référence à des cibles précises durant la conversation.

Le point (4) est primordial dans le cadre de l'étude de l'imitation car ce sont les repères topographiques qui vont constituer la liste de *tokens* cible de l'étude du *shift* : ceux-ci vont être dits (ou lus) dans les pré- et post-tests, et réitérés plusieurs fois par chaque interlocuteur lors de la phase de conversation, dans la mesure où le chemin à tracer est établi par rapport à ces repères (*cf* Figure 15, page précédente).

Ainsi, selon ces caractéristiques, nous pourrions considérer que la Map Task est une tâche conversationnelle taillée pour l'étude de la convergence phonétique. Comme le note en plus Anderson et collègues (1991, pp. 352–353), les éléments de la Map Task peuvent être adaptés aux besoins de l'expérimentateur :

- la seule restriction concernant les repères topographique est l'inventivité de l'illustrateur
- les média de communication (canal audio ou audiovisuel) sont contrôlables
- l'appariement des interlocuteurs est aussi contrôlé, ce qu'a fait Pardo (2006) pour éviter un effet du genre, rapporté par précédemment par Namy et al. (2002)

Malgré une apparence très favorable pour l'étude de l'imitation, cette tâche pourrait présenter certains écueils limitant en fait son utilité. D'après Lewandowski (2012, p. 118), les rôles de « Donneur d'instruction » et de « Récepteur des instruction », qui sont assignés au début de la

tâche, conduiraient à des disparités dans les productions des sujets. Après avoir évalué un corpus de Map Task, Lewandowski rapporte en effet que le temps de parole en fonction du rôle conversationnel est déséquilibré. De plus, les structures syntaxiques utilisées par le « Receveur » seraient moins complexes, le vocabulaire moins riche et les tours de parole plus courts.

Par conséquent, la Map Task présente des aspects intéressants pour l'émergence de comportement imitatifs en situation de conversation quasi spontané, cependant, il est possible qu'utiliser cette tâche présente quelques inconvénients.

Une autre critique en défaveur de la Map Task relève aussi des rôles de « Donneur » et « Receveur », mais en ayant cette fois-ci trait aux types de phrases qui peuvent être recueillies. Durant la Map Task :

- Le « Donneur » aurait tendance à émettre en majorité des instructions, des énoncés présentant donc une modalité assertive ou impérative
- Le « Receveur » émettrait majoritairement des demandes de précisions ou des confirmations, des énoncés qui présenteraient plutôt une modalité interrogative ou assertive

De fait, la répartition des modalités utilisées par les participants à la Map Task semble manquer d'équilibre et de diversité, biais a priori favorisé par les rôles de « Donneur » et de « Receveur » (Bradlow, Baker, Choi, Kim, & Van Engen, 2007).

1.1.2 Diapix : le principe des images différenciées

L'épreuve du Diapix constitue une alternative intéressante pour l'obtention de données de parole spontanées dont le topic peut –à l'avance– être maîtrisé par l'expérimentateur. La tâche du Diapix a été développée initialement par une équipe du département de linguistique de Northwestern University (Bradlow et al., 2007; Van Engen et al., 2010). Cette épreuve vise à dépasser les limitations de la Map Task, tout en en gardant les principes essentiels :

- Coopération requise entre les participants
- Dialogue cadré par la tâche, mais non scripté
- Assurance de recueillir plusieurs occurrences des cibles

Pour ce faire, Bradlow et collègues (*Ibid.*) s'inspirent du principe de la Map Task (deux images présentant des similarités et des différences sont données aux sujets) mais ils en changent :

- Le type d'image donnée
- La tâche à accomplir avec ces images

En effet, le Diapix est une tâche qui consiste à trouver les différences entre deux images (jeu fréquent dans les journaux sous l'appellation « jeu des 7 erreurs »). Les paires d'images du Diapix présentent dix différences que les participants doivent réussir à trouver en communiquant ce qu'ils voient sur l'image qui leur a été donnée. A propos des différences, elles sont de deux natures puisque il est possible qu'un élément soit changé (la couleur des chaussures d'un personnage, rouge ou verte selon l'image) ou qu'il soit simplement absent du décor (cf Figure 16). Ainsi, comme la Map Task, le Diapix favorise la coopération des interlocuteurs pour l'accomplissement d'un but commun. Cependant, la nature différente de la tâche offre aux locuteurs la possibilité d'utiliser des ressources langagières plus variées. Le Diapix semble donc plus profitable que la Map Task en ce qui concerne le recueil de données de parole tout en gardant un équilibre entre les deux interlocuteurs.

Figure 16 : Images d'un Diapix

Toutefois, cette tâche présente également quelques défauts qu'il nous faut noter. Lewandowski (2012, pp. 194–195) pointe ainsi les avantages et inconvénients de cette technique d'obtention de données :

> Eliciting data with the Diapix task, on the one hand, allows for the collection of rich speech in a communicative setting, which is an undeniable prerequisite for

> investigating naturally-occurring convergence. On the other hand, though, deciding for a quasi-spontaneous setting, takes away the possibility of controlling the amount of data at hand, which is possible in more controlled settings. It is obvious that every speaker uttered our lexical target items a varying number of times […].This led to a variable amount and type of data for each pair of speakers.

Ainsi, les deux tâches peuvent pâtir de biais spécifiques :

- Les données obtenues par une Map Task risquent d'être biaisées à cause du rôle du « Donneur », qui peut être en position d'hyper énonciateur
- Le type de données obtenues par un Diapix peut être biaisé ; car il n'y a pas de chemin obligé. Lewandowski (2012, p. 195) rapporte qu'un locuteur peut utiliser des termes que son partenaire n'utilisera pas ou peu, empêchant ainsi l'obtention de *tokens* à comparer.
- Enfin, il n'est pas à exclure que les locuteurs utilisent des stratégies d'évitement des répétitions d'items lexicaux au moyen de l'utilisation, par exemple de pronoms personnel ou de déictiques : il serait peu naturel que les interlocuteurs n'utilisent pas ces moyens d'expression, par économie.

En résumé, le Diapix –comme la Map Task– semblent des tâches fructueuses pour qui souhaite étudier la convergence phonétique, bien que toutes deux présentent certains inconvénients. Leur design commun (celui du Diapix est inspiré de la Map Task, [Van Engen *et al.*, 2010, p. 515]) offre un terrain favorable à l'émergence de la convergence phonétique par la coopération requise entre les locuteurs pour l'atteinte efficace d'un but commun. Les deux tâches partagent également un inconvénient commun : bien que l'expérimentateur connaisse à l'avance les *tokens* cibles, il ne peut garantir le nombre d'occurrences spontanées de ces *tokens*.

1.1.3 Rôles des participants et maîtrise du sens de convergence

La définition de la convergence phonétique donnée par Giles et al. (1991, p. 7) laisse supposer que la convergence est un phénomène mutuellement partagé entre les interlocuteurs,

puisqu'il s'agit de « *s'adapter l'un à l'autre*[42] ». Il se pose alors la question, dans le contexte des tâches conversationnelles précédemment décrites d'arriver à déterminer si :

- Un locuteur converge et pour quelle(s) raison(s) ?
- Chaque locuteur converge avec son partenaire.

Du point de vue de l'expérimentateur, la question peut avoir une importance particulière et faire partie intégrante de sa problématique, ou bien, l'expérimentateur peut tenter de forcer la convergence d'un des locuteurs en direction de l'autre afin de pouvoir concentrer son analyse sur la production du locuteur en question. En fonction des cas de figure, les questions posées sont différentes.

Si l'expérimentateur laisse libre cours à ses sujets, la problématique –outre l'attestation de la convergence et de ses effets– va avoir une teinte sociolinguistique. Il peut s'agir de déterminer, parmi les facteurs affectant l'individu ou la situation de communication, celui qui va expliquer la stratégie d'adaptation des sujets. Il faut alors contrôler l'appariement des sujets pour provoquer certains effets recherchés. C'est ce que font Namy et al. (2002) qui proposent des paires de locuteurs avec des sexes opposés et qui trouvent un effet du genre du locuteur sur la convergence et ce que cherche à éviter Pardo (2006) en proposant de n'apparier que des locuteurs de même sexe. Dans ce genre de tâche, la question du sens de convergence en dépendance avec les caractéristiques intrinsèque du sujet peut, en elle-même, être l'objet de recherche. Ainsi, Kim (2011, 2012) s'intéresse-t-elle à la question de la convergence des locuteurs en fonction de leur L1 et de celle de leur partenaire.

Enfin, l'expérimentateur peut essayer de maîtriser le sens de convergence dans les tâches conversationnelles en donnant des consignes explicites à un des locuteurs. En ce sens, l'expérimentateur ne laisse plus libre cours à l'émergence du phénomène, ce que Lewandowski (2012) jugeait essentiel pour l'étude de ce phénomène de manière authentique. Cette approche a pourtant été retenue par Pardo, Jay & Krauss (2010, p. 2256), mais pour une question de recherche différente :

> To assess the impact of an explicit attempt to match phonetic attributes, one member of each pair of talkers was instructed to try to imitate the speech of the other talker during the course of the conversation.

L'émission d'une telle consigne résulte, dans l'approche de Pardo *et al.* (*Ibid.*), du fait que l'effet de la convergence est de magnitude assez restreinte d'une part, et contrainte par les

[42] Individuals adapts to each other's behavior

spécificités de la situation de communication d'autre part. De plus, par cette consigne, cette équipe vise à créer une nouvelle condition expérimentale qui renvoie à l'aspect intentionnel – ou non– du comportement imitatif, dont nous avons souligné auparavant l'importance comme élément définitoire primordial des différents types de comportements imitatifs. Ce faisant, Pardo & collègues sortent peut être du cadre du comportement même de la convergence phonétique et proposent une approche de l'imitation qui sera retenue dans d'autres types de designs expérimentaux.

1.2. Les tâches d'impersonnation

Avant même de voir une personne, nous sommes capables de l'identifier au moyen de sa voix (Laver, 1994; Laver & Trudgill, 1979; Révis, 2013) en nous fondant sur ses spécificités physiologiques et son origine géographique, comme si chaque personne disposait d'une empreinte vocale . Bien sûr, nous ne sous-entendons nullement qu'il existe une empreinte vocale qui équivaudrait aux empreintes digitales (voir à ce propos : Boë & Bonastre, 2012), idée par ailleurs qualifiée de « mythe » par Révis (2013, p. 101‑102).

Cependant, il n'est pas rare que des individus d'une même société (re)connaissent les voix de personnalités politiques au point de pouvoir, dans les dîners entre amis, imiter les traits caractéristiques de leur manière de parler. Un phrasé incisif, un peu heurté (et quelques mouvements d'épaule) peuvent suffire à faire penser à un ancien président de la V^{ème} République. *A priori*, tout locuteur peut essayer d'imiter l'autre afin de se faire passer pour lui ; en d'autres termes, de déguiser sa voix[43], de faire une imitation *vocale*.

Sur le plan scientifique, certaines études ont cherché à comprendre les stratégies mises en place par les imitateurs pour copier la voix des personnalités, comme celle de Bessler (1991) qui a étudié la manière de parler d'un imitateur du général De Gaulle, avec sa voix naturelle et lors de ses impersonnations. Cette étude a montré que Tisot, l'imitateur de De Gaulle, parvenait à modifier son niveau moyen de fréquence fondamentale, ainsi que son étendue. Dans l'absolu, la manière de procéder de ce genre d'étude diffère peu des études sur la convergence phonétique et des études d'imitation de laboratoire : en ce qui concerne les

[43] Imitation vocale (chez Révis), impersonnation (chez Zetterholm), « voice disguise » (chez Farrùs) désignent la même chose : imiter la voix d'un autre afin de se faire passer pour lui. Nous préférons le terme d'impersonnation pour désigner la tentative d'un locuteur d'imiter la voix d'un autre afin d'être pris pour l'autre. La notion de « voice disguise » (déguisement de la voix) recouvre un ensemble de procédés, dont l'impersonnation (cf. infra, 4.1.2.2)

indices acoustiques, il s'agit d'observer un *shift* entre les différentes productions des sujets expérimentaux. Les différences les plus notables entre les tâches d'impersonnation et les autres tâches imitatives résident dans le type de consignes données aux sujets, ainsi qu'au choix des sujets eux-mêmes : les groupes expérimentaux incluent souvent un professionnel réputé du monde du spectacle. (Attal-Fiocchi & Jarzé, 2014; Brzostek & Deschanvres, 2014; Farrús, Wagner, Anguita, & Hernando, 2008b; Mary et al., 2012; Mary, Anish Babu, Joseph, & George, 2013; Mejvaldova, 2002; Révis et al., 2013; Zetterholm, 2002, 2002, 2006, 2009a). Ces sujets particuliers jouent sur leur capacité à imiter la voix des célébrités et l'érigent en marque de fabrique, par exemple : Laurent Gerra, réputé pour ses imitations de Jacques Chirac (et impliqué dans les travaux de J. Révis), ou, Yves Lecoq, une des voix des Guignols de l'Info. Pour les chercheurs, les imitateurs professionnels constituent donc des sujets de choix, dans la mesure où ils représentent des sujets à la performance hors norme :

> The performance of a professional impersonation artist probably represents state-of-the-art performance in this area and material of that kind should, therefore, represent interesting test cases. (Eriksson & Wretling, 1997)

Dans ce type de tâche, il est proposé au professionnel et/ou à l'individu lambda d'imiter –de se faire passer pour– la cible à des fins et dans des conditions variables. En cela, les tâches d'impersonnation diffèrent d'autres tâches imitatives, car le but de la tâche proposée aux sujets expérimentaux consiste explicitement à usurper la voix d'un autre. Nous présenterons ici les spécificités de deux approches des tâches d'impersonnation qui diffèrent par leur finalité, mais se rejoignent par les dispositifs expérimentaux qu'elles mettent en œuvre. Ces dispositifs s'articulent autour de trois éléments :

- La cible de l'imitation, qui, généralement, jouit d'une certaine célébrité dans la communauté linguistique se prêtant à l'expérience ;
- Le professionnel, qui, en termes de capacités imitatives, représenterait un sujet expert dans son domaine ;
- L'individu lambda, qui serait a priori capable de produire de l'imitation vocale, mais supposément de manière moins convaincante que le professionnel.

Nous discuterons ici le rôle attribué à chacun de ces protagonistes dans les paradigmes expérimentaux de l'impersonnation, en portant une attention toute particulière à la place qui y est réservée à l'expert.

1.2.1 Approche centrée sur l'imitateur expert - objet d'étude

Les travaux de Eriksson & Wretling (1997) constituent un point de départ intéressant pour décrire les méthodologies associées à l'étude de l'impersonnation. En effet, il nous semble que les tâches proposées aux imitateurs experts ont peu évolué depuis leur étude intitulée « *How flexible is the human voice – A case study of mimicry* ». Y faire référence, revient –en ce qui nous concerne– à souligner les deux éléments essentiels de leur titre que sont la notion « d'étude de cas » et le questionnement sur la flexibilité de la voix humaine.

L'étude de cas semble inhérente au sujet car il ne semble pas chose aisée d'avoir un accès à l'imitateur professionnel : Eriksson & Wretling (1997), Zetterholm (Zetterholm, 2002a, 2002b, 2006) ou l'équipe de J. Révis[44] recueillent des imitations produites par des professionnels de premier plan, respectivement « artiste professionnel de l'impersonnation [suédois] », « deux professionnels depuis plus de dix ans [suédois, eux aussi] » et « Laurent Gerra[45] ». En ce sens, ces études mettent en scène l'exceptionnel, puisque la réussite de l'imitation (en terme de sa réception par un auditeur naïf) serait garantie par le statut de l'expert.

Le questionnement sur la flexibilité de la voix humaine conduit par ailleurs à proposer différents types de situations expérimentales avec comme présupposé que l'imitateur aura une performance excellente lors de ses reproductions de différentes voix. En ce sens, Eriksson & Wretling (1997) proposent à leur sujet expert d'imiter trois voix différentes (et célèbres, en Suède), prélevées d'enregistrements télévisuels, afin d'observer la capacité d'adaptation de l'expert. La parole à imiter est continue et dure environ trente secondes pour chaque voix différente. A partir des enregistrements faits consécutivement par l'imitateur à partir de l'extrait, les chercheurs suédois procèdent à différents relevés d'indices phonétiques afin d'observer le *shift* attendu. Ce type de tâche est très répandu dans les rares études existantes sur l'impersonnation : Zetterholm (2001) propose à l'expert de travailler à partir d'enregistrements TV, Mejvaldova (2002) à partir d'archives de la Radio tchèque, Révis *et al.* (2013) à partir d'un discours politique enregistrés et Farrus *et al.* (2008b) adoptent une approche identique. Cela pourrait être lié au fait que les voix cibles de l'imitation sont exclusivement des voix célèbres : ce sont ces voix que les imitateurs professionnels

[44] C'est-à-dire, Révis & collègues (2013), ainsi que les étudiantes d'orthophonie de l'AMU, dirigées par J. Révis : Attal-Fiocchi & Jarzé et Brzostek & Deschanvres, entre autres.
[45] Paraît-il, un très célèbre imitateur français, depuis plus de vingt ans.

impersonnent et qui ont l'avantage d'être enregistrées avec une bonne qualité, dans des studios TV ou radio…

En ce qui concerne les consignes données aux imitateurs dans ces études, certaines se rejoignent dans l'idée selon laquelle ces experts, contrairement à leurs habitudes, doivent imiter dans le but de se faire passer pour la voix cible, et non, pour divertir :

> These imitations were not intended to be entertaining, but explicitly meant to mimic the target voices and speech style as closely as possible. (Eriksson & Wretling, 1997)

Il nous semble important de faire cette précision car la limite entre imitation et caricature peut être ténue : comme le fait remarquer Laver (1994, p.28), imiter une voix est un procédé stéréotypant, et non une exacte copie du comportement langagier de la cible. Ainsi, si le chercheur souhaite observer dans quelle mesure le sujet expert est capable de reproduire son modèle, il faut que ce dernier s'affranchisse de son habitude. Cela dit, le chercheur doit garder en tête qu'une imitation se fait toujours sur la base de la voix de l'imitateur et de sa perception de la cible. Ainsi, comme le note Zetterholm (2002), « *certains traits importants peuvent être exagérés, d'autres moins importants* [pour l'imitateur] *peuvent être négligés ; l'audience aura quand même l'impression d'une imitation réussie*[46] ».

Le couple voix célèbre/expert imitateur représente donc la base des tâches d'impersonnation. En fonction du questionnement du chercheur, cette dernière peut se trouver sujette à des variations. Les premières situations que nous avons décrites s'intéressaient aux cas où le chercheur tente de découvrir le degré de flexibilité de la voix de l'expert : ce dernier est –souvent– seul point de référence. Il nous semble dommage qu'un certains nombre d'études (notamment celles des équipes suédoises) ne proposent pas de contraster leurs résultats sur les experts en utilisant des groupes contrôle composés de sujets naïfs. Comme nous le notions en introduction de ce point méthodologique (4.1.2), nous sommes tous, peu ou prou, capables de déguiser notre voix. Cela peut être dû au présupposé dont nous parlions ci-avant : la notoriété de l'expert, son entraînement et même son « talent » (Révis, 2013, p. 78 (note 80)) garantissent sa performance, bien que des sujets naïfs puissent imiter de manière convaincante (certes, avec beaucoup de variabilité interindividuelle) (Nocaudie & Astésano, 2012) ou dans une mesure plus limitée qu'un professionnel (Révis et al., 2013). Ainsi, la seule réputation de l'expert peut conduire à un biais d'omission dans le design des études, soit en ce

[46] Some important features may be exaggerated some less important may be neglected in the voice imitation; the audience will still get the impression of a successful impersonnation

qui concerne la flexibilité vocale (l'étude seule de l'expert, sans contrepoint, interdit les comparaisons), soit en ce qui concerne le jugement sur la réussite du procédé imitatif, ce que relèvent Attal-Fiocchi & Jarzé (2014) :

> Du fait de son excellente réputation, nous avons pris pour acquise la qualité des imitations réalisées par Laurent Gerra. […] Pour autant les performances ne sont pas identiques [d'une imitation à l'autre[47]]. […] Il aurait pu être intéressant de mettre en place une évaluation perceptive des enregistrements par un jury afin de comparer leurs impressions avec nos mesures prosodiques. (pp. 80-81)

Cela étant dit, notre critique principale à propos de l'absence des sujets naïfs porte essentiellement sur le recueil du matériel phonétique : issu de sources différentes, il permettrait d'effectuer des comparaisons des stratégies mises en œuvre par les différents groupes de l'étude, ce que proposent Révis *et al.* (2013).

A propos du recueil de données, d'autres procédures que celle proposée par Eriksson et Wretling (1997) méritent d'être commentées. Les études de Zetterholm (2002, 2002, 2002) sont en l'occurrence assez particulières car le matériau phonétique étudié n'a pas été construit à des fins scientifiques mais à des fins publicitaires. Zetterholm s'est basée sur une série de douze enregistrements d'un imitateur professionnel suédois commandée par un opérateur de téléphonie. Le matériau phonétique de ces études a donc été recueilli *a posteriori*. En considérant l'imitation phonétique seule, certains aspects de ces études demeurent problématiques, ce que relèvent également Révis et al. (2013)

- Chacun des douze enregistrements parodie une personnalité suédoise différente (homme politique, *speaker*…). A cette fin, l'imitateur a lui-même établi le contenu des sketchs en se basant sur des procédés autres que la seule imitation phonétique de la voix cible. Les douze textes ont donc un contenu lexical différent, ainsi que des tournures syntaxiques spécifiques aux voix cibles afin de faciliter aux auditeurs l'identification des voix parodiées.
- En conséquence, le contenu de chacun des douze sketchs n'a pas d'homogénéité et seul un *token* leur est commun [le terme *mobilsvar* (répondeur téléphonique)]
- Par ailleurs, Zetterholm s'intéressant à la seule flexibilité de la voix de l'expert, ne propose pas de comparer les productions de l'imitateur aux voix de référence.

[47] Pour ce mémoire d'orthophonie, Attal-Fiocchi & Jarzé comparent des productions de L. Gerra au fil d'une tournée de spectacle. La formulation originale était « d'un soir à l'autre ».

Ces trois études de Zetterholm révèlent une absence de méthodologie au niveau du recueil des données qui limitent ainsi fortement les analyses possibles sur l'imitation phonétique à proprement parler. Du fait de la rareté des études dans ce domaine, l'apport de Zetterholm (et de ses études suivantes) est cependant intéressant en ce qui concerne l'appréhension de la flexibilité du comportement vocal d'un même locuteur.

Pour prolonger cette petite revue méthodologique des tâches centrées sur l'imitateur professionnel, il nous semble opportun de souligner l'originalité de l'approche de Révis *et al*. (2013). D'une part, le professionnel –Laurent Gerra– n'est pas l'unique parti de l'étude puisque ses productions sont confrontées à celle d'un groupe de sujets naïfs. Ce faisant, Révis, De Looze et Giovanni assoient le statut de l'expert, et prenne le « risque » de le remettre en question, tout en ouvrant une perspective comparatiste. Cela dit, l'originalité de leur approche réside aussi dans la manière dont le recueil des productions imitatives est effectué. Cette équipe propose à leurs sujets experts et naïfs d'imiter la voix de Jacques Chirac, le cinquième président de la V^{ème} République, à partir d'un discours politique prononcé en 2002. Dans un premier temps, le discours est donné aux sujets sans qu'ils connaissent sa provenance. Les sujets doivent alors produire une première lecture du texte avec leur voix naturelle. Par la suite, il leur est annoncé que ce texte a été originellement prononcé par J. Chirac et qu'ils doivent essayer de lire le discours une nouvelle fois en imitant spontanément l'ancien président, et sans avoir écouté sa voix auparavant.

L'expert a effectué uniquement cette tâche, car il était habitué à imiter cette voix qui fait partie de son registre habituel ; les sujets naïfs ont quant à eux effectué une troisième lecture après écoute du discours original. A la suite de ces enregistrements, des mesures d'indices prosodiques ont été effectuées, afin de comparer les trois types de lecture : avec leur voix naturelle, en imitation spontanée et en imitation après écoute (pour les seuls sujets naïfs). Ce protocole expérimental nous semble particulièrement intéressant, car il se soucie de comparer les productions de l'expert et des sujets naïfs. Par ailleurs, ne pas faire écouter le discours original pour la tâche d'imitation spontanée exploite l'idée selon laquelle nous disposons de représentations mentales du comportement parolier d'autres locuteurs : les résultats de Révis et collègues montrent en effet que dès cette tâche, les sujets naïfs également changent leur comportement langagier en adaptant, par exemple, leur niveau moyen de fréquence fondamentale. Enfin, ce protocole présente une rigueur proche des travaux de phonétique en laboratoire.

De fait, la limite principale que nous voyons aux designs expérimentaux des tâches d'impersonnation réside dans la relative absence des sujets naïfs dans la composition des groupes expérimentaux. Les études que nous venons d'évoquer prennent pour sujet un (Mejvaldova, 2002; Révis et al., 2013; Zetterholm, 2002, 2006) ou deux (Zetterholm, 2002) imitateurs professionnel, et seul le travail de Révis et collègues inclut un groupe contrôle pour contraster les résultats du sujet expert. Dans ce dernier cas, le groupe de sujets naïfs comptait seulement quatre locuteurs, sans talent réputé pour l'imitation. Ce design, en se focalisant sur l'expert, fait en effet briller l'exceptionnelle performance du professionnel. Cependant, il pourrait être bénéfique de proposer des tâches d'impersonnation à de plus larges groupes de locuteurs naïfs afin de détecter plus finement les stratégies d'impersonnation des locuteurs naïfs. De manière idoine, il serait intéressant d'arriver à constituer des groupes d'imitateurs experts, afin de s'éloigner du design de l'étude de cas et de pouvoir déterminer les stratégies des locuteurs en fonction des caractéristiques des voix perçues.

Dans les études de cette revue, les voix imitées sont d'ailleurs exclusivement des voix célèbres. Le fait qu'elles soient célèbres facilite l'impersonnation des sujets –experts ou naïfs– car du fait de leur statut, les sujets ont une certaine familiarité avec ces voix. Ils connaissent leur image et ils les ont vus parler (et donc se mouvoir en parlant, ce qui nous pensons, est un point important). Les sujets ont donc à disposition des représentations de la voix à imiter, du corps qui la produit, et de certaines de ses habitudes comportementales. L'imitation étant d'après Laver (1994) un processus stéréotypant, nous pouvons supposer que les sujets –experts ou naïfs– poussés à imiter ces voix cherchent à exploiter l'ensemble de ce qu'ils connaissent de la voix et de l'habitude corporel de l'autre pour produire leur impersonnation.

Finalement, en proposant des voix célèbres à l'imitation, l'expérimentateur favorise l'expert qui a l'habitude de reproduire ces voix, au point que les résultats produits à partir de ces stimuli puissent être critiqués. Certes, comme le note Révis (2013, p. 78, note 80), l'objectif poursuivi dans ces approches est « *d'étudier un modèle de fonctionnement à son plus haut niveau de performance afin d'en observer le comportement* ». Cependant, il faudrait sortir du paradigme des tâches incluant une voix célèbre pour mettre l'expert sur la même ligne de départ que le sujet naïf. C'est ce qui est proposé par Révis *et al.* (2012 a & b) lorsqu'ils ont cherché à montrer les capacités perceptives supérieures de leur expert en le soumettant au paradigme expérimental proposé par Dupoux (1997) concernant la surdité accentuelle en langue étrangère des locuteurs natifs. Faire sortir l'expert de son habitude professionnelle (soit, sa zone de confort), –c'est-à-dire : lui proposer des voix inconnues à

imiter– permettrait peut-être de dissocier des processus qui relèvent de l'impersonnation d'autres processus relevant de l'imitation purement phonétique.

En ce sens, nous pensons qu'il faudrait proposer aux sujets d'imiter des voix inconnues et de contrôler l'image qu'ils perçoivent du sujet à imiter (uniquement la voix, une vidéo du visage de la cible, ou de l'ensemble de son corps). L'expérimentateur pourrait ainsi contrôler certains aspects du processus d'imitation, notamment sa partie perceptive.

1.2.2 Impersonnation et reconnaissance automatique du locuteur

Les auditeurs semblent capables d'identifier les autres au moyen de leur voix. Cependant, il semble également qu'un imitateur compétent puisse être en mesure de leurrer parfaitement son auditoire. Révis (2013, p. 67) relate à ce propos une anecdote tout à fait intéressante :

> Entre « deux manips », je me suis amusée […] à faire écouter mes enregistrements à des personnes de mon entourage qui n'étaient pas au courant des travaux entrepris sur l'imitation. J'avais tout mélangé en ordre aléatoire : les différentes productions des locuteurs non-imitateurs, Laurent Gerra dans ses deux voix naturelles, Jacques Chirac lui-même et son imitation par Laurent Gerra. Les participants ont massivement identifié l'imitation réalisée par Laurent Gerra comme étant Chirac, plus que l'enregistrement de Chirac lui-même !

Ce récit souligne d'abord que l'imitateur professionnel a une compétence d'imitation supérieure au locuteur dit « naïf », puisqu'il est choisi en priorité. Ensuite, l'histoire rapportée par Révis indique qu'un tiers peut parvenir à usurper la voix d'un autre, au point que le locuteur original puisse être considéré comme le « faussaire » par un auditeur qui en connaît par ailleurs la voix… Dans le cadre du divertissement, l'usurpation de la voix d'un autre est permise par la finalité de l'imitation, qui est d'amuser un auditoire. En revanche, dans un cadre plus légal, l'imitation peut s'envisager comme une pratique délictueuse. En imaginant que la reconnaissance vocale devienne une pratique courante dans le domaine bancaire, un malandrin doué pour l'impersonnation pourrait-il tromper le système de reconnaissance vocale de la banque et assécher les comptes de ses victimes ? Ce dernier point intéresse plus particulièrement le second type d'approche de l'impersonnation que nous avons relevé dans la littérature. L'impersonnation y est envisagée comme un moyen parmi d'autres (voir Tableau 2 ci-dessous) pour « déguiser sa voix » :

> Voice disguise is the purposeful change of perceived age, gender, identity, personality of a person. It can be realized mechanically by using some particular means to disturb the speech production system, or electronically by changing the sound before it gets to the listener. (Perrot, Aversano, & Chollet, 2007)

De manière idoine, certains chercheurs comme Lau *et al.* (2004, p. 148) considèrent d'ailleurs l'imitation comme une forme technologiquement simple d'attaque. Ainsi, l'imitation de la voix d'un autre –soit, l'impersonnation– apparaît comme un objet central de la recherche autour de la problématique de l'identification des locuteurs.

	Electronique	Non électronique
Conversion	GMM (Gaussian Mixture Model) …	Imitation (impersonnation)
Transformation	Dispositif électronique Logiciel de modification de la voix	Altération mécanique Altération prosodique/phonétique

Tableau 2 : typologie du déguisement de la voix, d'après Perrot & collègues (2007). Nous rajoutons entre parenthèse la précision « impersonnation » car Perrot *et al*. donnent la signification suivante au terme « imitation » : l'imitation produite par un professionnel. Les notions d'altérations (prosodique ou phonétique) n'indiquent pas spécifiquement de valeur imitative dans cette typologie.

En ce sens, ces chercheurs aussi ont besoin de recueillir des données d'imitations parolières de qualité, afin de pouvoir, par exemple, procéder au test de la robustesse des technologies de reconnaissance vocale. Certaines études, comme celles de Farrús, Wagner, Anguita & Hernando (2008a; 2008b) proposent un design de recueil de données extrêmement proche des études citées précédemment, soit : un ou plusieurs imitateurs professionnels, reproduisant des discours de voix célèbres, extraits de documents radiophoniques ou télévisuels. A quelques différences près, la méthodologie de recueil de l'étude de Lau & collègues (2004) est très proche de la méthodologie que nous avons décrite ci-avant. Cependant :

- ils envisagent qu'un locuteur lambda puisse aussi procéder à une tentative d'usurpation de la voix d'un autre et choisissent donc des locuteurs sans expertise particulière dans le domaine de l'imitation ;

- en raison de la spécificité de leur recherche, ils proposent des modèles issus du corpus YOHO[48] spécifique au domaine de la vérification de la voix (J. Campbell, 1995) ;
- afin de simuler le choix d'une cible comme le ferait une personne mal intentionnée, ils laissent le choix des voix à imiter à leurs sujets d'expérience, qui doivent en sélectionner trois (parmi des voix leur semblant proches à éloignées).

Par certains côtés, cette approche de l'impersonnation semble plus contrainte que la précédente en ce qui concerne la limitation du lexique employé par les sujets. L'exemple de l'utilisation du corpus YOHO semble assez révélatrice : si l'on considère que la séquence à prononcer pour s'identifier est fixe –une sorte de mot de passe– il paraît alors important de tester le système sur des séquences précises. Si cette approche de l'imitation ne se distingue pas par la manière dont sont recueillies les données, elles permettent de considérer les problématiques liées au traitement de l'imitation sous un jour rarement abordé par les études en convergence ou celles centrées sur l'imitateur expert. Par exemple, dans les études de l'équipe de Farrús (2008 a & b), le recueil du corpus donne lieu à l'analyse subséquente de certains facteurs du timbre (notamment le *jitter* et le *shimmer*, respectivement la variation cyclique de la *f0* et de l'intensité) entre le modèle ciblé et son imitation. Comme le remarque Révis (2013, p. 57), certains de ces facteurs ne sont qu'effleurés par tout un pan de recherche sur l'imitation car « *il est difficile de rapprocher ces résultats d'une réalité physiologique […]. Il est en effet difficile d'imaginer qu'un imitateur puisse avoir une prise consciente sur le réglage de son* shimmer. ».

Afin d'asseoir l'importance de ces études dans notre revue de littérature, il semble important de faire référence au travail des équipes indiennes de Leena Mary (2012, 2013). Bien que leurs travaux s'inscrivent dans la problématique de la reconnaissance vocale, ces chercheurs s'intéressent également à « *l'art d'imiter par les artistes professionnels de l'imitation*[49] ». De plus, ces études présentent une particularité méthodologique dans le recueil de données, dans la mesure où leur corpus d'imitation présente deux modalités : dépendante ou indépendante du texte. La première correspond aux types de recueil décrits précédemment puisqu'il est important de pouvoir comparer la parole imitée à son modèle sur des cibles précises. Cet aspect vaut aussi bien pour les études en convergence qui font prononcer des

[48] YOHO est un corpus particulier et très contraint au niveau lexical : il s'agit exclusivement de suites de nombres à deux chiffres, compris entre 21 et 97 (dit « *twenty one* » ou « *ninety seven* »). Les 138 locuteurs du corpus prononcent des séries de 3 nombres (par exemple : « 22, 38, 92 »).

[49] The art of mimicking by professional artists. Mary et al (2012)

cibles lexicales lors de pré- et post-tests que pour les études de la flexibilité vocale des imitateurs où le texte est contraint par le discours des personnalités qu'ils impersonnent (au point que les mots supplémentaires sont supprimés des corpus[50]). La seconde partie du corpus, indépendante de tout texte, est recueillie dans le but d'illustrer les problématiques de la phonétique légale : à partir d'un outil de reconnaissance, estimer l'authenticité de la parole d'un locuteur dont l'identité est sujette à caution. *De facto*, bien qu'elles ne soient pas liées directement à notre préoccupation principale, ces études envisageant l'impersonnation comme un outil pour tester des systèmes de reconnaissance vocale nous fournissent des indications précieuses, aussi bien théoriques que techniques et peuvent nourrir notre réflexion méthodologique concernant le recueil de données et leur traitement.

Comme le montre cette courte revue, les deux types d'approches de l'impersonnation (centrée sur l'imitateur expert ou bien sur le système de reconnaissance) proposent des designs de recueil des données particulièrement semblables, cela, bien que leurs finalités diffèrent. Un imitateur (professionnel et/ou naïf) est confronté à un texte lexicalement limité (audio et/ou écrit), qu'il doit reproduire au mieux après écoute de la voix originale ou bien à partir des représentations qu'il a de la voix de la personne. Une fois le recueil établi, le domaine de recherche envisagé par le chercheur ainsi que sa problématique dictent le traitement des données.

1.3 De l'imitation de laboratoire

Ce que nous appelons « *imitation de laboratoire* » s'inscrit en complémentarité des approches de la convergence, et dans une certaine mesure, des études sur l'impersonnation. En effet, cette appellation doit nécessairement faire référence à des études qui se distinguent de celles menées à partir de tâches conversationnelles. Nous voyons en cela deux raisons principales. D'une part, le contexte d'émergence de la convergence phonétique est la conversation. En ce sens, le comportement observé dans ce contexte ne peut être le même qu'un comportement émergeant d'une simple répétition de mot, cela, même si les deux comportements partagent de grandes similarités. D'autre part, le contrôle incomplet que le chercheur a sur les *tokens* cible des études sur la convergence conduit tout de même aux limitations que nous avons décrites ci-avant. Celles-ci indiquent l'impossibilité du chercheur à maîtriser le contenu des données, leur ajoutant alors une part d'inconnu ou de potentiels

[50] Par exemple, Zetterholm (2009b) n'analyse qu'un seul mot cible de son corpus…

déséquilibres en ce qui concerne le nombre d'occurrence des cibles lexicales. Il convient alors de distinguer la convergence phonétique d'autres phénomènes d'imitation.

A propos de l'impersonnation, les études qui ont été décrites un peu plus haut rejoignent dans une certaine mesure la conception que nous avons de l'imitation de laboratoire, puisque le chercheur sélectionne les stimuli et contrôle les tâches en donnant des consignes spécifiques. Cependant, nous estimons que le comportement visé par ces études constitue une exception dans le spectre des imitations parolières, du fait de la rareté des sujets et de leur expertise. Par ailleurs, nombre de ces études sélectionnent des modèles issus de documents authentiques (radiophoniques et/ou télévisuels). Ce faisant, il apparaît que cette manière de procéder tend à s'éloigner des principes fondamentaux de la « *phonologie de laboratoire* » (D'Imperio, 2005; Pierrehumbert, Beckman, & Ladd, 2000) –une dimension que nous souhaitons intégrer à l'étiquette « *imitation de laboratoire* »–. Ainsi, la spécificité du comportement ciblé par ces études et l'approche méthodologique qu'elles retiennent limitent dans une certaine mesure leur domaine d'investigation à l'exploration de l'exceptionnel.

Après un bref aperçu des préceptes de la phonologie de laboratoire et une revue méthodologique de quelques études relevant de l'imitation au laboratoire, nous expliciterons cette dernière approche qui pourrait viser à extraire des situations d'imitations parolières de leur contexte naturel d'occurrence pour les transposer au milieu contrôlé du laboratoire. De cette manière, l'expérimentation produirait un contexte favorable à l'observation d'imitation(s) ; soit, « *la création d'une situation dans laquelle des observations cruciales peuvent être faites*[51] » (Ohala & Jaeger, 1986, p. 2).

1.3.1 La parole au laboratoire

Malgré les critiques que nous avons pu faire à l'encontre des études sur l'impersonnation et sur la convergence phonétique, nous ne cherchons pas à les exclure du cadre que nous allons décrire à présent. Au contraire, cette parenthèse sur les méthodes de la communauté de chercheurs en phonologie de laboratoire devrait faire ressortir avec une certaine clarté, l'appartenance de cette littérature à cette tradition de recherche. En effet, de

[51] L'intégralité de la phrase d'Ohala & Jaeger : « An experiment, then is simply the creation –contrivance, if one prefers- of a situation in which crucial observations, those relevant to a given question, may be made in such a way that they will be free from as many anticipated distorting influences as possible »

nombreux indices laissent penser que les différentes études passées en revue jusqu'à présent se sont conformées à ce paradigme méthodologique. Parfois, les paramètres définitoires des processus mis en expériences –convergence ou impersonnation– imposent des choix pouvant impliquer une perte de contrôle sur le contenu des données. Cependant, les approches retenues sont claires et les designs expérimentaux sont discutés, et donc, de plus en plus raffinés. Par exemple, la discussion concernant le choix de la tâche optimale pour l'étude de la convergence –*Map Task* ou *Diapix*– témoigne de la vitalité des pratiques expérimentales autour de notre sujet. Pour reprendre les termes de Ohala et Jaeger (1986, p. 1) « *Gradually the discipline's practitioners discover how to create and control the opportunities they need in order to make those crucial observations […].* ». Ce point précis motive grandement la tenue de cette discussion méthodologique.

L'aspect le plus attrayant de la phonologie de laboratoire réside dans son ouverture théorique. Comme le notent plusieurs articles fondateurs (Ohala & Jaeger, 1986; Pierrehumbert et al., 2000) ou de synthèse (D'Imperio, 2005), la phonologie de laboratoire ne se veut ni une théorie pouvant orienter les explications du chercheur, ni un cadre de travail au sens donné par Pierrehumbert et collègues (2000., p. 4).

> Frameworks are packages of assumptions about the fundamental nature of language, and the research strategy for empirical investigation is driven by top-down reasoning about the consequences of the framework. […] One framework can replace another via a paradigm shift, if incorporating responses to successive empirical findings makes the prior framework so elaborate and arcane that a competitor becomes more widely attractive.

Le dernier point noté par Pierrehumbert *et al.* semble particulièrement important pour définir l'approche de la phonologie de laboratoire de manière contrastive avec un cadre de travail. En effet, si le *framework* peut être remplacé par un autre (comme le modèle issu de *Sound Patterns of English* a pu l'être par la théorie de l'optimalité, D'Imperio, 2005; Lyche, 2005), l'approche de la phonologie de laboratoire se veut une méthode générique, opérationnelle et rigoureuse ; soit, une approche méthodologique irremplaçable dans l'étude phonologique ou de la parole.

La phonologie de laboratoire regroupe une communauté de chercheurs, d'horizons variés (puisqu'ils peuvent venir, entre autres, des sciences du langage, de la psychologie, d'ingénierie), dont les préoccupations principales sont l'explication scientifique de la structuration mentale du langage et de ses corrélats physiques (D'Imperio, 2005; Pierrehumbert et al., 2000). En termes de philosophie de recherche, cela revient, selon

Pierrehumbert, Beckmann & Ladd (*Ibid.*, p. 2) à adopter la posture de recherche des sciences dites matures, telles la biologie ou la physique. Ce faisant, le langage (dimension sociale comprise) est vu comme un phénomène naturel, certes, complexe, comme peut l'être le système climatique de notre planète. En conséquence, l'étude expérimentale du langage est souhaitable (nécessaire, même), malgré sa variabilité. Comme nous le soulignions quelques lignes plus haut, les situations expérimentales –prônées par cette démarche– sont justement construites pour pallier cette variabilité en la maîtrisant autant que possible.

Par ailleurs, le caractère naturel des phénomènes linguistiques implique que la communauté des chercheurs s'engage dans des opérations de formalisation et de modélisation de ces phénomènes (D'Imperio, 2005, p. 242). En ce qui concerne l'étude de la parole, formalisation et modélisation sont intimement liées. En effet, le signal que nous étudions est dérivé de programmes informatiques qui traitent et permettent de visualiser le signal. PRAAT (Boersma, 2001) par exemple, permet un lien entre le signal modélisé et certains formalismes ou méthodes d'annotation. Cependant, la modélisation proposée par ce logiciel est, pour certains, imparfaite, menant alors les membres de la communauté à proposer des outils supplémentaires afin de corriger les imperfections de ces modélisations. C'est par exemple le cas de l'algorithme MOMEL (Hirst & Espesser, 1993), qui vise à obtenir une image de la *f0* plus lisse (car l'algorithme « élimine » l'influence des événements articulatoires sur la *f0*) ou les propositions de De Looze (2010) qui corrigent certaines instabilités caractéristiques des représentations de la *f0*. L'existence de tels outils découle, à notre sens, de la mise en expérimentation du langage et non du seul progrès technologique. Comme l'indiquent Ohala & Jaeger (1986, p. 3) :

> A great many of the experiments in the mature disciplines do involve intruments and complex procedures. But it is important to emphasize that the complexities are simply a reflection of the advances made in those fields, that is, the accumulated wisdom about what steps are necessary to make observations of the phenomena in a way that is free from distortion. […] Naturally […] the discipline gains experience in recognizing previously overlooked sources of error and in finding ways to compensate for them.

De plus, la modélisation mathématique ne doit pas se borner à la représentation du signal pur. Un titre de conférence tel que « *Mathématiques et phonologie : quels outils mathématiques pour la modélisation en phonologie* » qui a eu lieu en 2006 à Orléans indique la volonté commune de chercheurs d'horizons différents pour s'inscrire dans une démarche conjointe de

recherche, ce que remarquent Bergougnioux et collaborateurs (2007) dans l'article introductif d'un numéro spécial de revue, séquelle de cette conférence :

> Il y a une convergence dans les recherches des mathématiciens et des phonologues qui s'est opérée à partir des outils élaborés en statistique, en modélisation, […] à partir de l'usage que peuvent en avoir des linguistes qui se revendiquent de différentes écoles.

Ce type d'interaction entre disciplines *a priori* éloignées, est particulièrement bienvenu puisque la discussion interdisciplinaire pousse fréquemment la communauté scientifique à changer ses pratiques.

Enfin, il semble judicieux de souligner encore l'importance d'une approche méthodologique partagée. Les avancées dans ce domaine, comme l'indiquent Pierrehumbert et al. (2000, p. 6) sont parfois aussi importantes que les avancées théoriques. Pour ces derniers, une querelle théorique n'est pas synonyme de querelle méthodologique : des chercheurs dont les avis divergent au niveau théorique peuvent tout à fait partager les mêmes méthodes expérimentales et cela « *contribue à la cohésion des communautés de recherches qui sont suffisamment diverses pour « survivre » à long terme* »[52]. A notre sens, les études évoquées précédemment (4.1.1 & 4.1.2), s'inscrivent alors pleinement dans cette démarche de phonologie de laboratoire : elles partagent des méthodes (pour le moment, nous nous sommes limités au recueil de données), les discutent et adoptent la manière de faire des sciences matures. La suite de ce propos présentera quelques études représentatives de ce que nous appelons « imitation de laboratoire », en référence à la démarche que nous venons de décrire.

1.3.2 Une revue d'imitations au laboratoire

L'approche de la phonologie de laboratoire permet de tester toute théorie, pourvu « *qu'elle soit testable expérimentalement* » (D'Imperio, 2005). Les études sur la convergence phonétique testaient principalement le cadre théorique décrit par la Communication Accomodation Theory (Giles et al., 1991) et les études sur l'impersonnation s'intéressaient alternativement aux limites de la flexibilité vocale des locuteurs et à des problèmes d'ingénierie en reconnaissance vocale. Toutes, du fait de leur méthode rigoureuse appartiennent à cette approche expérimentaliste. Bien que l'approche théorique et les

[52] These shared methods are one reason why research paradigms in the established sciences are not as incommensurate as Kuhn claims, and they contribute to the cohesion of research communities which are diverse enough for long-term vitality.

problématiques qu'elles traitent soient diverses, les études que nous allons présenter en suivant sont –à notre sens– représentatives de ce que nous appelons « *imitation de laboratoire* ». En effet, l'objectif de cette revue sera de souligner la convergence des méthodes et des tâches employées dans des domaines aux problématiques variées : les tâches imitatives en parole permettraient de mettre en évidence des phénomènes divers comme la nature du stockage lexical (Goldinger, 1998), la détection de l'aptitude de *phonetic compliance* (Delvaux, Huet, Piccaluga, & Harmegnies, 2014) ou encore le biais perceptif bilatéral entre les auditeurs dits analytiques ou synthétiques (Postma-Nilsenová & Postma, 2013). Les phénomènes d'imitations étant de manière définitoire une interface entre perception et production langagière, les tâches imitatives transposées au laboratoire présentent un intérêt particulier pour l'étude des dynamiques de la parole.

L'étude de Goldinger (1998), intitulée « *Echoes of echoes ? An episodic Theory of Lexical Access* » semble un point de départ naturel pour entamer cette revue. En effet, cette étude est assez révélatrice des problématiques qui peuvent être traitées au moyens de tâches imitatives. Par ailleurs, il s'agit également d'une étude fréquemment citée dans la littérature sur l'imitation en parole. Cette étude traite du débat concernant la nature du stockage lexical dans la mémoire, thématique que nous avons survolée dans le second chapitre de ce texte.

Dans son étude de 1998, Goldinger vise à défendre la vision exemplariste du stockage lexical en proposant une procédure expérimentale basée sur la tâche de *shadowing*. La tâche de *shadowing* consiste en une répétition rapide d'énoncés entendus, généralement des mots isolés (Dufour & Nguyen, 2013). Au moment où Goldinger publie son étude, celui-ci remarque que les chercheurs utilisant cette tâche de *shadowing* s'intéressaient peu à la parole produite par le sujet, mais plutôt aux temps de réaction entre l'écoute et la production :

> The typical dependent measure in shadowing is the latency betwenn stimulus and response onsets. A seldom-used secondary mesasure is the speech output itself. (Goldinger, 1998, p. 255)

Sur la base de sa revue de littérature, Goldinger propose alors d'utiliser la tâche de *shadowing* pour sa valeur imitative afin de tester les prédictions d'un modèle exemplariste du stockage lexical (MINERVA 2, Hintzman, 1986). Goldinger considère comme principe de base que le *shadowing* est construit sur des processus perceptifs et cognitifs :

> Shadowing is not a shallow activity –words do not "travel directly" from the ears to the vocal tract in a reflex arc. […] If shadowing is a truly cognitive process, models like MINERVA 2 may predict performance. (Goldinger, 1998, p. 256)

Cette étude trouve sa place dans notre revue d'imitations au laboratoire car elle permet de faire le lien entre ce que nous avons dit de la phonologie de laboratoire et la méthodologie en imitation. En effet, le questionnement de Goldinger n'est pas directement lié aux performances imitatives des locuteurs puisque ce dernier s'intéresse plutôt aux processus cognitifs qui permettraient cette performance. Ce faisant, Goldinger opère bien dans le cadre de la phonologie de laboratoire, puisqu'il propose, par l'expérimentation de mettre à l'épreuve un modèle théorique. Par ailleurs, cette étude propose d'utiliser un paradigme expérimental que nous retrouvons souvent dans la littérature, sous diverses formes : les répétitions rapides de mots isolés ou *shadowing*. Cette tâche a souvent été utilisée pour tester la rapidité de traitement lexical des locuteurs, et dans le cas de Goldinger, pour trouver des traces d'échos dans la mémoire lexicale. En termes d'imitation, il est maintenant assez communément admis que le *shadowing* est une tâche favorable à l'émergence d'imitation parolière.

Dans le cadre de cette revue méthodologique, il convient alors de nous intéresser à l'étude de Dufour et Nguyen (2013) qui proposent justement de quantifier l'imitation contenue dans les réponses de tests type *shadowing*. D'un point de vue méthodologique, cette étude en particulier semble cruciale dans la mesure où elle nous renseigne sur l'effet de convergence que l'on peut attendre d'une telle tâche.

Dans cette étude, Dufour *et al.* proposent de tester l'effet d'une consigne supplémentaire au paradigme du *shadowing*, c'est-à-dire : tester les degrés d'intentionnalité en imitation. A cette fin, ils font répéter à deux groupes de locuteurs méridionaux des mots contenant les voyelles cibles /e/ et /ɛ/. Or cette distinction phonologique n'est pas faite par les locuteurs de cette région. Dufour et Nguyen espèrent alors que l'exposition à ces mots prononcés par un locuteur de français standard aura une influence sur la production des locuteurs méridionaux, notamment sur la hauteur du premier formant. S'ils ne donnaient pas de consigne supplémentaire, nous nous trouverions dans une situation de *shadowing* classique et les auteurs ne pourraient déduire de leurs données des éléments de réponse permettant de quantifier l'imitation contenue dans le *shadowing*. C'est pour cela que Dufour *et al.* ont divisé leurs sujets en deux sous-groupes auxquels des consignes différentes sont données :

> Half of the participants were instructed that, upon hearing the word, they were to **repeat it as naturally and clearly as possible**. The other half were instructed that, upon hearing, they were to **repeat it by imitating the speaker's specific pronunciation.** (Dufour & Nguyen, 2013, p. 3)

Cette étude trouve les résultats suivants :

- Lors du prétest, il n'y a pas de différence de Formant 1 sur les voyelles cibles prononcées par les deux sous-groupes
- Pendant le test, une différence entre /e/ et /ɛ/ apparaît dans les productions des deux groupes, indiquant que l'exposition à la parole d'un autre a induit un changement dans la production des sujets d'expérience.
- Cette différence est significativement plus grande dans le sous-groupe ayant reçu pour consigne d'imiter explicitement la locutrice modèle.
- Pendant le post-test, la différence de Formant 1 persiste pour les deux sous-groupes et la distance entre eux s'est résorbée.

Ainsi, cette étude montre, en accord avec la littérature sur le *shadowing*, que la production des locuteurs est influencée par ce qu'ils entendent. De plus, leurs résultats montrent que l'effet de la consigne d'imitation explicite permet d'obtenir un plus grand *shift* comportemental pendant le test, mais qu'à long terme (soit : durant le post-test) un effet de même magnitude persiste chez les deux groupes.

Ces résultats semblent montrer que le *shadowing* est une tâche positivement biaisée pour la production de comportements imitatifs en parole, notamment au niveau segmental. Par ailleurs, dans le cas d'une recherche de convergence phonétique sur le niveau des segments, cette étude indique qu'il est judicieux de former des groupes homogènes au niveau dialectal, ce que proposent également Delvaux & Soquet (2007). Enfin, ce paradigme expérimental de *shadowing* peut également être utilisé dans le cas d'expérimentation sur le niveau suprasegmental. A ce propos, l'étude de Michelas & Nguyen (2011) qui utilisent un paradigme identique adapté pour le niveau prosodique, conclut également que l'exposition aux stimuli dans des tâches de répétition a une influence sur la production des sujets. Cependant, les différences entre imitation implicite et explicite obtenues par Dufour & Nguyen (2013) au niveau segmental, n'ont pas été répliquées dans le travail de Michelas & Nguyen (2011) : dans cette dernière étude, les auteurs n'ont pas trouvé de différence significative entre les tâches de *shadowing* et d'imitation explicite.

Le *shadowing* représente le canon méthodologique de l'imitation implicite au laboratoire, auquel il peut être adjoint des consignes engageant les sujets vers des comportements plus explicites d'imitation. Comme pour toute tâche imitative, le but de l'expérimentateur consiste à générer un *shift*, un changement phonétique dans la production du sujet (Delvaux et al., 2004). Dans certaines études citées, les expérimentateurs constituaient des groupes de sujets homogènes au niveau dialectal, puis les exposaient à des

stimuli issus de variantes dialectales différentes. Ce cas se présente dans l'étude de Delvaux *et al.* (*Ibid.*) où des locuteurs liégeois sont exposés à des stimuli de locuteurs bruxellois (et vice versa), ainsi que dans l'étude de Dufour et Nguyen (2013) où des locuteurs méridionaux sont exposés à des stimuli de français standard. Dans cet ordre d'idée, il peut être opportun de souligner l'approche retenue dans l'étude de Adank, Hagoort & Bekkering (2010). En effet, leur étude présente une particularité fort intéressante puisque ces chercheurs néerlandais, s'intéressant à la compréhension du langage, proposent de « créer » une variante dialectale :

> Our study evaluated the effect of several forms of training on the comprehension of accented speech. Listeners heard sentences spoken in an unfamiliar accent of Dutch. This unfamiliar accent was obtained by systematically altering the pronunciation of vowels in stressed lexical positions, with the aim of creating a nonexistent accent of Dutch. (Adank, Hagoort & Bekkering, p. 2)

L'effet recherché par cette création dialectale est de mettre sur un pied d'égalité les sujets expérimentaux vis-à-vis des tâches à accomplir en compréhension : comme le soulignent les auteurs, « r*elative familiarity with an accent affects the processing of accented speech.* » (Adank *et al.* 2010, p. 1904). Il est intéressant de noter que cette approche s'accorde avec la critique que nous avions formulée à l'encontre des tâches d'impersonnation qui proposaient systématiquement aux imitateurs professionnels d'imiter des voix qu'ils avaient l'habitude de traiter.

En ce qui concerne l'étude d'Adank *et al.* (2010), les sujets expérimentaux entendaient cette variété artificielle de hollandais et subissaient un entraînement spécifique à leur groupe expérimental (au nombre de six), afin d'observer l'effet de l'entraînement sur la compréhension de cette variété artificielle. Le groupe témoin n'avait pas d'entraînement, et les cinq autres groupes avaient chacun un entrainement différent :

- Le groupe « auditeur » devait écouter et imaginer ce que le locuteur disait en hollandais standard.
- Le groupe « répétiteur » devait écouter et répéter la phrase entendue en conservant leur manière de parler habituelle (en imitant le moins possible…)
- Le groupe « transcripteur » devait écouter puis transcrire ce qu'ils ont entendu
- Le groupe « imitateur » devait écouter puis répéter la phrase entendu avec sa prononciation précise
- Enfin, le groupe « imitateur plus bruit » devait écouter la phrase, puis imiter la phrase entendue alors qu'un brouhaha était joué dans leur casque.

La méthodologie ainsi que les résultats de cette étude nous intéressent particulièrement. En premier lieu, les consignes d'imitation n'ont pas ici pour but d'étudier la production des sujets, mais de souligner l'effet bénéfique de l'imitation d'une variété dialectale inconnue sur la compréhension de la dite variété. Nous sommes alors tentés de dresser un parallèle avec l'apprentissage des langues étrangère, plus particulièrement sur l'importance des pratiques de production orale, dont (osons-le dire) les pratiques de correction phonétique. En effet, selon les résultats de cette étude, imiter la prononciation spécifique de phrases perçues favoriserait la compréhension des phrases produites dans la variété entendue (Adank *et al.* 2010, p. 1906) ! En second lieu, la création d'une variété dialectale artificielle permet de gommer l'expérience linguistique des sujets d'expérience. Ce choix peut être crucial pour l'expérimentateur qui s'intéresse à certaines capacités du sujet humain, comme le talent phonétique (Jilka, Baumotte, Lewandowski, Reiterer, & Rota, 2007) ou le concept mal connu de *phonetic compliance* (Delvaux et al., 2014).

Hormis le travail de Adank et collègues, les études présentées jusqu'à maintenant proposaient toutes comme cible de l'imitation des voix ou des variétés dialectales dont les sujets d'expériences avaient une connaissance. C'était particulièrement le cas dans les études sur l'impersonnation, où les professionnels imitaient les voix de personnalités publiques de premier plan. A propos des études sur la convergence, comportement qui n'affleure qu'en contexte communicatif, il semble simplement requis que la communication soit possible entre les locuteurs. En effet, la valeur adaptative du phénomène, ainsi que son caractère automatique, non-contrôlé, suppose que ce dernier puisse se produire indépendamment du niveau de compétence des locuteurs. En revanche, des codes linguistiques inhabituels, voire inconnus des sujets d'expérience peuvent également être mis à profit dans le cadre de tâches imitatives.

Par exemple, dans le cas des études sur le talent phonétique menés par les équipes allemandes de l'université de Stuttgart, l'hindi a été utilisé comme langue éloignée pour apporter un indice de talent phonétique :

> Besides German and English we also added Hindi as a language that the test participants were not familiar with. The Hindi imitation of words and short phrases focus exclusively on the segmental aspects of speech, especially the perception of those sounds and phonotactic structures not present in the native language and the ability to reproduce the perceived acoustic patterns. (Jilka *et al.*, 2008, p. 231)

Un locuteur qui, sans expérience préalable, parviendrait à reproduire les patrons phonotactiques d'une langue inconnue aurait un talent particulier dans ce que Delvaux et collaborateurs (2014, p.1) appellent la « *compliance phonétique*[53] ». Pour les scientifiques belges, ce concept peut se définir comme « *the intrinsic individual ability to produce speech sounds that are unusual in the native language, and constitutes a part of the ability to acquire L2 phonetics and phonology.* ». Faire référence ici à ce phénomène semble pertinent, car sa détection et son évaluation pourraient résulter d'un paradigme de tâches d'imitation de sons inconnus.

La question posée par Delvaux et collaborateurs a une résonnance assez fondamentale dans le champ de la recherche sur l'apprentissage et le traitement d'une L2. Par leur démarche, ceux-ci visent à expliquer une part de la variabilité individuelle de performance observée pendant les expérimentations dans ce domaine. Comme ils le font remarquer, la variabilité inter-individuelle obtenue dans ces expérimentations est telle qu'il est difficile de déterminer l'effet même des variables indépendantes choisies par les scientifiques.

En se référant à la théorie du score vrai (*true score theory, TST*), la variabilité inter-individuelle résulterait d'effets aléatoires durant le recueil de données, comme l'expliquent Delvaux et al. (2014) : cela peut être considéré comme un « *bruit* » dans la mesure. Ainsi, d'après la TST[54], toute observation X de comportement humain est composé de l'effet aléatoire E et du score vrai T, de telle sorte que :

$$X = E + T$$

En d'autres termes, si la mesure X était absoute de tout effet E, nous aurions alors une mesure du *true score* T :

$$E \rightarrow 0 \Rightarrow X \rightarrow T$$

Comme le font remarquer Delvaux et collaborateur, il est alors crucial pour l'expérimentateur d'adopter une méthodologie rigoureuse afin de minimiser l'effet de E (*i.e.* contrôler la variabilité inter-individuelle) pour pouvoir expliquer la variabilité attendue des facteurs contrôlés de l'expérience. Or, la proposition de ces chercheurs est de considérer qu'il existe

[53] Compliance : initialement, terme de biologie ou médecine servant à désignant l'élasticité de certains tissus comme les poumons.

[54] Dans ce paragraphe, nous reprenons l'explication de Delvaux et collègues (2014, pp. 1-2) et la simplifions pour n'en garder que la ligne directrice, notamment la dernière équation qui est un peu abusive. Nous renvoyons le lecteur à l'article original pour une explication plus rigoureuse.

une composante non aléatoire et dépendante du sujet de l'expérience habituellement agglomérée à *E*. Cette composante, qu'ils appellent *C*, serait, comme *E*, une source de variabilité incontrôlable par l'expérimentateur, mais, à la différence de *E*, ne serait pas aléatoire :

> The *C* component has initially been misinterpreted as part of the random error *E*, whereas it is in fact a systematic, non-random factor, independent from *E*. (Delvaux *et al*., 2014, p. 2)

Dans le cadre d'expérimentations sur le traitement des sons d'une L2, la composante *C* représenterait alors l'effet de la compliance phonétique, c'est-à-dire, l'effet fixe de l'habileté des sujets à reproduire des sons inhabituels dans leur langue maternelle. Ainsi, pour estimer *T* au plus juste, l'expérimentateur devrait minimiser *E*, et connaître *C*, de telle sorte que[55] :

$$si\ E\ \rightarrow 0 \Rightarrow X = C + T \Rightarrow\ X - C = T$$

Afin de détecter cette composante *C*, Delvaux et collègues proposent d'utiliser une tâche d'imitation de sons inconnus construits à partir du synthétiseur Klatt, ce qui leur permet de contrôler pleinement le contenu acoustique des stimuli perçus par les sujets.

Les études que nous venons de passer en revue portent sur des thématiques variées (traitement cognitif des items lexicaux, questionnement méthodologique sur une tâche expérimentale, compréhension langagière et comportement d'imitation, nature même des données recueillies durant l'expérimentation). Ce faisant, leur rigueur théorique ainsi que leur approche méthodologique constamment mise en question témoignent de leur appartenance aux études relevant de la communauté de la phonologie de laboratoire. Par ailleurs, ces études ont été choisies pour montrer une convergence de paradigmes expérimentaux : celui des tâches imitatives au laboratoire. Ce paradigme de tâches ne se limite pas à l'étude même de l'imitation, mais peut (et même, doit) nous renseigner sur des questions fondamentales traitant des dynamiques de la perception et de la production du langage chez le sujet humain.

1.3.3 De l'imitation *au* laboratoire à l'imitation *de* laboratoire

Outre cette volonté d'appartenance à une communauté de recherche, l'étiquette « *imitation de laboratoire* » permet de différencier un ensemble de paradigmes expérimentaux

[55] L'équation suivante n'illustre pas au plus juste la relation des composantes, *X*, *E*, *C* et *T*. En effet, la composante C s'exprime mieux en termes de variance pour un ensemble d'observations X.

d'autres paradigmes sur les tâches imitatives comme ceux ayant trait à la convergence phonétique ou à l'impersonnation. Afin de désamorcer d'éventuelles critiques, il convient de noter que nous n'excluons pas ces deux phénomènes de l'étude de l'imitation au laboratoire de manière générale. Ils représentent simplement des cas suffisamment singuliers pour que nous voyions des frontières bien dessinées entre les approches :

- La convergence phonétique est un comportement lié à l'interaction langagière entre au moins deux interlocuteurs. De fait, les tâches visant à faire émerger la convergence phonétique exigent un contexte conversationnel. En cela, il s'agit d'imitation **au** laboratoire et non d'imitation **de** laboratoire.
- L'impersonnation constitue une forme particulière d'imitation en parole, et plus particulièrement d'imitation vocale, puisque ce comportement résulte de l'intention d'un locuteur de se faire passer pour autrui au moyen de sa voix. Les paradigmes de l'impersonnation représentent un cas particulier d'imitation **de** laboratoire bien que leur design expérimental relève souvent trop de l'étude de cas.

Ainsi, l'imitation **au** laboratoire désigne pour nous la mise en contexte de toute situation d'imitation dans un cadre expérimental alors que l'imitation **de** laboratoire fera référence à une situation plus particulière. Ici, nous souhaitons clarifier quelques critères qui permettent de passer de l'imitation **au** laboratoire à l'imitation **de** laboratoire, que nous pouvons définir comme : une situation expérimentale d'étude rigoureuse de l'imitation en parole dans un contexte non communicatif.

En d'autres termes, l'imitation de laboratoire met l'imitateur dans une situation biaisée d'imitation, car le modèle en est souvent absent. Il n'est présent que sa voix dématérialisée, et l'imitateur doit alors se reposer uniquement sur sa perception auditive (et/ou visuelle) pour parvenir à accomplir la tâche qui lui est demandée.

a- Déconstruction de l'imitation

Malgré son apparente simplicité, la situation d'imitation, complexe par nature, est difficile à transposer au milieu contrôlé du laboratoire. Cette transposition requiert de l'expérimentateur une analyse détaillée du phénomène. En cela, une définition rigoureuse de l'imitation (ou du comportement que l'on cherche à observer) permet de circonscrire les éléments de la situation expérimentale. A partir de la définition générale que nous avons donnée précédemment *« un sujet produit de l'imitation quand il reproduit tout ou partie d'un comportement (et/ou de ses caractéristiques) perçu chez un autre sujet, de manière à ce qu'un*

tiers puisse percevoir la production de l'imitateur comme ressemblant à celle du modèle » nous pouvons définir les éléments majeurs des paradigmes d'imitation au laboratoire.

- La constitution d'un matériel modèle

Ce sont les stimuli auxquels le sujet d'expérience sera exposé, et, supposément qui doivent provoquer un *shift* comportemental chez le sujet, *i.e.* un changement phonétique.

- La présence d'un imitateur

Il s'agit du sujet d'expérience. De nombreux paradigmes expérimentaux utilisent des sujets naïfs, bien que les approches autour de l'impersonnation fassent appel à des sujets experts.

- Le recueil de la production de l'imitateur

Cette étape constitue le recueil principal de données de parole imitée. L'analyse de ces données doit permettre de mettre à l'épreuve les hypothèses posées par l'expérimentateur. Les types de tâches utilisés pour effectuer les recueils de données ont été abondamment décrits dans les pages qui précèdent. Souvent, plusieurs étapes de recueils sont nécessaires (pré-tests et post-tests), car il faut pouvoir témoigner du changement comportemental des sujets.

- L'évaluation de la valeur imitative de la production de l'imitateur

Une fois les données recueillies, l'expérimentateur n'est pas encore certain qu'il y ait eu imitation. Cette étape cruciale mérite que nous y consacrions toute notre attention, car elle est source de nombreux débats quant à la méthode à appliquer pour ce faire !

b- Contrôle de la matière phonique perçue et produite

L'étude de la convergence phonétique se déroule dans un cadre conversationnel. Comme le phénomène est interactif, on peut supposer que les deux locuteurs prenant part à la conversation vont subir une influence mutuelle, *i.e.* converger. Or, ce que nous décrivons comme imitation de laboratoire se déroule dans un contexte non-interactif, soit, à partir de stimuli pré-enregistrés.

Disposer d'ensembles de stimuli (qu'ils soient issus d'un synthétiseur [Delvaux et al., 2014] ou de voix naturelles) permet un contrôle de la matière phonique perçue par le sujet. Ce faisant, l'expérimentateur peut estimer quelles sont les cibles susceptibles de subir un changement phonétique ou d'être spécifiquement reproduites (Dufour & Nguyen, 2013) et contrôler le temps d'exposition des sujets à chaque type de stimuli (ce qui, par certains côtés

était une critique que nous portions à l'encontre du recueil de données en convergence phonétique). *De facto*, en contrôlant précisément le contenu phonologique et prosodique de la parole perçue par les sujets-imitateurs, l'expérimentateur peut alors émettre des hypothèses sur le traitement de la matière phonique au fil de l'expérience par l'ensemble des groupes expérimentaux. En effet, l'avantage de l'imitation de laboratoire, par rapport à l'approche conversationnelle de l'imitation, est de proposer le même modèle à tous les sujets.

En parallèle, le contexte de la situation expérimentale contrôlée permet à l'expérimentateur d'obtenir de chaque sujet une quantité équivalente de productions imitatives et ainsi pallier un problème récurrent des expérimentations en convergence phonétique. Nous avions en effet souligné le risque de ne pas obtenir un nombre suffisant d'occurrences des *tokens* ciblés durant les tâches conversationnelles, en raison des stratégies discursives, notamment l'utilisation de pronoms et de déictiques.

c- Contrôle de l'intentionnalité

Le critère d'intentionnalité a été précédemment utilisé pour circonscrire des comportements imitatifs différents, du mimétisme à la mimésis. En parole, l'intention du sujet, qu'elle soit explicite ou sous-jacente, doit impliquer la production de comportements différents. Durant la revue méthodologique précédente, nous avons par exemple décrit des tâches illustrant deux extrêmes de ce paradigme d'intentionnalité en imitation phonétique : les tâches de conversation phonétique et les tâches d'impersonnation.

En convergence, il est attendu que le phénomène se produise de manière incontrôlée de la part des interlocuteurs, car il s'agit là d'un critère définitoire du phénomène de convergence. Seule l'étude de Pardo, Jay & Krauss (2010) proposait à un sujet de chacune de leur dyade de s'engager explicitement dans un comportement d'imitation. Habituellement, les tâches expérimentales pour la convergence phonétique ne comportent donc pas de consigne à ce sujet.

A contrario, les consignes des tâches d'impersonnation comportaient systématiquement des directives précises quant à l'engagement intentionnel des sujets : ceux-ci devaient imiter la voix de leurs cibles durant le recueil de données, soit, tenter d'usurper leur identité vocale.

En imitation de laboratoire, le critère d'intentionnalité doit également être contrôlé, afin de calibrer au mieux les tâches expérimentales. Par exemple, dans l'étude d'Adank et al.

(2010), le « groupe répétiteur » devait répéter les phrases entendues. Si un des sujets tendait à imiter les caractéristiques vocales du sujet entendu, l'expérimentateur lui rappelait de ne pas s'engager dans un comportement imitatif. En parallèle, si un des sujets du groupe « imitateur » semblait ne pas tenter de reproduire les caractéristiques vocales des modèles entendus, l'expérimentateur le rappelait également à l'ordre. D'autre part, il arrive également que la référence à l'imitation soit volontairement exclue des consignes : Goldinger (1998) émet une hypothèse sur la trace que laisse un stimulus entendu dans la mémoire lexicale des sujets ; il se sert de l'énoncé reproduit pour détecter un changement phonétique et travaille alors sur un effet de convergence phonétique *et non* sur une imitation volontaire. Enfin, l'étude de Dufour & Nguyen (2013) montrait clairement un effet de l'intentionnalité sur la production des sujets. Dans cette dernière étude, la consigne était différente pour chacun des deux groupes (un imitait, l'autre faisait un simple *shadowing*) mais il pourrait aussi être envisageable de proposer une gradation de l'intentionnalité à un seul et même groupe, au cours de différents blocs expérimentaux.

1.4 Synthèse

Faire émerger l'imitation en parole requiert une approche rigoureuse. Que ce soit en situation conversationnelle, ou bien en situation d'imitation de laboratoire, l'expérimentateur se doit de contrôler un grand nombre de paramètres. Dans les deux cas, l'objectif sous-jacent de l'expérimentateur est le même : il s'agit de provoquer un changement phonétique chez les sujets d'expérience, et recueillir des données qui lui permettront de le documenter.

En conversation, l'expérimentateur doit pouvoir diriger les débats sans imposer de script prédéfini aux sujets, afin de recueillir des *tokens* cibles qui lui serviront lors des analyses à venir. Cette contrainte préside au choix des tâches expérimentales, et devrait également avoir une incidence sur le contenu des pré- et post-tests. La Figure 17, qui montre le *set up* expérimental en convergence phonétique, synthétise les différentes étapes de recueil de données : on peut y distinguer trois blocs, du pré-test au post-test, au sein desquels la dimension temporelle est primordiale. Pour être analysés, les items recueillis sont généralement regroupés par phase de test, voire même, en fonction du moment de la phase de test (par exemple : début, milieu ou fin du test).

En imitation de laboratoire, la question des cibles de changement phonétique est résolue par le contrôle absolu de l'expérimentateur sur le corpus modèle. En effet, un interlocuteur étant remplacé au profit d'un corps désincarné, la dimension communicative disparaît de la situation expérimentale d'imitation de laboratoire (voir Figure 18) et le même matériel audio (voire audiovisuel) est ainsi présenté à chaque sujet un même nombre de fois. De fait, il importe particulièrement que le matériel audiovisuel ait des caractéristiques phonétiques/phonologiques contrôlées afin de pouvoir induire la direction du shift comportemental que ce matériel doit provoquer chez les sujets. De même, les consignes données aux sujets doivent maîtriser l'intentionnalité procurée au sujet pour la tâche imitative.

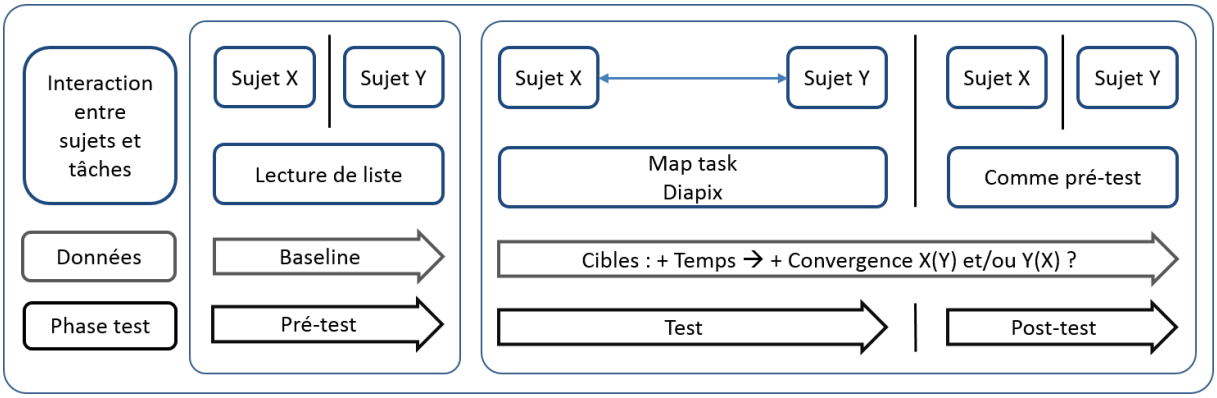


Figure 17 : Forme canonique du recueil de données de convergence phonétique**.** Beaucoup de designs proposent trois phases expérimentales. Le pré-test constitue la ligne de base de chaque sujet, *i.e.* sa parole naturelle avant exposition à un autre sujet, où il est recueilli des *tokens* qui doivent émerger durant la tâche conversationnelle. Les deux autres phases permettent le recueil des *tokens* pour comparaison avec la ligne de base. Il est attendu que l'effet de convergence phonétique se manifeste au cours du test et qu'il persiste après la fin de l'interaction entre X et Y, d'où le recueil de post-test.

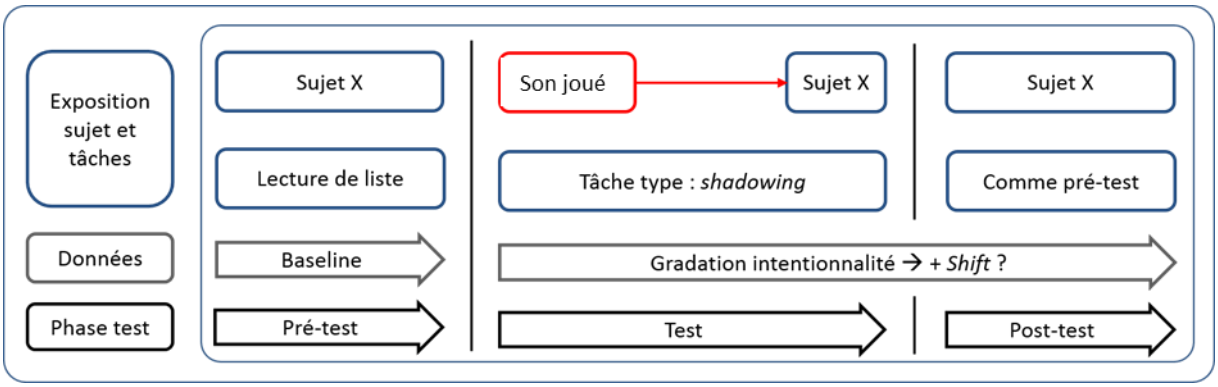


Figure 18 : Paradigme expérimental de l'imitation de laboratoire. Certains designs proposent le recours au pré-test pour le recueil d'une ligne de base. Cependant, cette étape est parfois mise de côté au profit d'une gradation de l'intentionnalité imitative durant la tâche. Par rapport à la figure XXX, on peut noter qu'un seul sujet intervient dans la situation de recueil.

La situation d'imitation de laboratoire semble laisser l'expérimentateur libre de tester des hypothèses variées en perception et en production de la parole. Sur une note plus générale, il semble que l'imitation de la parole, par nature à l'interface des versants de perception et de production puisse être un candidat privilégié pour traiter expérimentalement des problématiques intéressant de vastes domaines des sciences du langage : traitement mémoriel, changements linguistiques, nature des systèmes linguistiques : ces domaines en lien avec les dynamiques humaines en perception et production de la parole sont autant de thèmes que nous pouvons voir traités par une approche rigoureuse de l'imitation en parole.

2 Evaluer ou mesurer l'imitation en parole ?

Une fois le matériel sonore recueilli se pose la question de l'analyse des données. En effet, une problématique récurrente dans l'imitation de la parole concerne la méthode utilisée pour évaluer ou mesurer le degré d'imitation contenu dans une réponse. Dans une certaine mesure, le problème de correspondance (*i.e.* déterminer comment un sujet fait le lien entre ce qu'il voit et ce qu'il produit par la suite), initialement appliqué au traitement cognitif, se pose à nouveau face aux items de parole recueillis, cette fois, en des termes nouveaux. A savoir : les indices acoustiques sur lesquels se basent les auditeurs pour juger de la réussite d'une imitation parolière sont-ils les mêmes que ceux utilisés par l'imitateur pour adapter sa production aux caractéristiques de l'énoncé entendu ? En d'autres termes, nous pourrions nous demander si détecter un changement phonétique dans la production d'un locuteur peut suffire à établir certaines correspondances entre perception et (re)production de la parole ? Ces questions demeurent épineuses, et elles devraient être traitées avec la plus grande attention et certaines précautions. C'est pourquoi aucune position tranchée ne sera prise sur ces questions. Cela dit, attester du fait qu'une production d'un locuteur est une imitation parolière (ou comporte des aspects d'imitation) n'est pas non plus une tâche triviale.

Afin d'amorcer cette discussion, nous nous réfèrerons à Pardo (2013a, 2013b), pour qui le paradigme d'analyse de l'imitation en parole –plus particulièrement, de la convergence phonétique– devrait s'articuler autour de deux approches complémentaires :

> In order to derive a valid, and comprehensive assessment of acoustic-phonetic convergence, future studies should employ perceptual assessment of phonetic convergence in concert with measures of multiple acoustic dimensions. (Pardo, 2013b)

Cette proposition s'accorde particulièrement avec un aspect définitoire de l'imitation souligné précédemment : pour qu'une production gagne le statut d'imitation, il faut qu'un tiers estime qu'il y a une ressemblance entre le modèle et son imitation supposée. Ce constat pose alors un certain nombre de questions quant à la nature du tiers (auditeur, système d'évaluation automatique) et des ressemblances entre modèle et imitation (physique, détaillées, générales) d'une part ; tout en soulignant d'autre part que l'étude de l'imitation en parole est, par essence, une approche comparatiste. En se plaçant dans le paradigme de Pardo, l'évaluation de similarités parolières aurait donc une dimension subjective (résultat des choix d'un ensemble d'individus) à laquelle il faudrait corréler des facteurs objectifs (indices

acoustiques ayant une réalité physiquement mesurable) ; ces deux volets de l'analyse seraient indissociables. Au sens de Pardo, une approche systématique de l'évaluation de l'imitation consisterait donc à comparer les résultats de tests perceptifs avec les tendances du *shift* du sujet données par les mesures acoustiques.

Cela dit, la considération de Pardo (2013) sur la double approche de la convergence phonétique est assez récente et spécifique à l'étude de ce phénomène en particulier (voir Tableau 3).

Authors	Year	Behavior	Language	Acoustic measures	Imitation rating	Statistics
Eriksson & Wretling	1997	Impersonnation	Swedish	Total & word durations Segment deviation F0 mean Vowel space (F1 & F2)	-	Correlation (word durations) Descriptive
Goldinger	1998	Lab imitation	English	Duration F0	Perceptual (Relative) AXB	
Zetterholm	2000	Impersonnation	Swedish	Duration (utterance/word) F0 mean Formatn frequencies (F1-F4)	Perceptual (absolute)	Descriptive
Mejvaldova	2002	Impersonnation	Czech	F0 : contours and mean Duration : flow, syllable/word duration	-	Descriptive
Namy, Nygaard & Sauerteig	2002	Lab imitation	English	-	Perceptual (Relative) AXB	ANOVA
Zetterholm	2002	Impersonnation	Swedish	F0 mean Peaks deviation Duration Word deviation (beginning)	-	Descriptive
Pardo	2006	Convergence	English	-	Perceptual (Relative) AXB	ANOVA
Zetterholm	2006	Impersonnation	Swedish	F0 mean Duration (word) Articulation rate Formant frquencies (F1 & F2)	Auditory analysis ((3 trained listeners)	-
Farrus, Wagner, Anguita & Hernando	2008	Impersonnation	Catalan	Segment duration Jitter/Shimmer F0 various measures	System response	Identification Error rate Imitation Rate
Pardo, Jay & Krauss	2010	Convergence	English	Articulation rate Vowel spectra	Perceptual (Relative) AXB	ANOVA
Cole & Shattuck-Hufnagel	2011	Lab imitation	English	Tone & accent labelling	Kappa between different imitations' phonologic cues	-
Michelas & Nguyen	2011	Lab imitation	French	F0 : tonal targets	-	Mixed Logit Model
Mixdorff, Cole & Shattuck-Hufnagel	2012	Lab imitation	English	Accent and boundary location Syllable duration	Kappa between different imitations' phonologic cues	-
Pardo, Gibbons, Suppes & Krauss	2012	Convergence	English	Duration Vowel spectra	Perceptual (Relative) AXB	ANOVA
Pardo	2013	Lab imitation	English	Duration F0 Vowel formants	Perceptual (Relative) AXB	Mixed effect regression model

Tableau 3 : Aspects méthodologiques d'une sélection d'études en imitation de la parole. Classement par date puis auteur. On y observe le glissement du paradigme des pratiques expérimentales dans les études de Pardo, et des approches diverses sur les plans acoustiques et perceptifs en fonction des chercheurs.

Les approches d'évaluations de l'imitation sont donc plurielles, mais on peut toutefois repérer les deux tendances (perceptives et acoustiques) décrites par Pardo (2013), parfois isolée, d'autres conjuguées.

Afin d'éclaircir en quoi ces deux approches semblent nécessaires, nous ferons une revue méthodologique des procédés utilisés pour évaluer l'imitation. Travaillant en phonétique, il semble naturel de nous pencher en premier lieu sur les mesures acoustiques : un *a priori* naturel serait que les mesures acoustiques d'une production imitatives tendraient à se rapprocher de celles de son modèle. Variées, les mesures acoustiques utilisées dans la littérature pour évaluer l'imitation conduiraient cependant à s'inscrire dans un cercle vicieux. En effet, la focale parfois trop grossissante de l'analyse acoustique et la grande variabilité parolière intra- et inter-individuelle limiterait grandement le potentiel explicatif des indices acoustiques seuls (4.2.1). Par la suite, nous nous intéresserons à l'évaluation perceptive de la parole imitée, qui serait utilisée pour sa valeur holistique. Bien qu'elle permette de classer les items en fonction de leur similarité perçue avec les modèles, cette approche se révèlerait être insuffisante pour amorcer un semblant d'explication des processus liant perception et production phonétique (4.2.2). Enfin, nous envisagerons d'autres approches d'évaluation/mesure de la similarité visant à appréhender plus particulièrement la courbe de fréquence fondamentale (4.2.3 & 4.2.4).

2.1 Les mesures acoustiques : de bons indicateurs de l'imitation ?

Lorsqu'il s'agit de déterminer si les productions d'un locuteur sont plus ou moins similaires à celles du modèle auquel il a été exposé, une approche simple et naturelle consiste à s'intéresser au contenu du signal de parole. Comme le recueil de données d'imitation en parole vise à provoquer un *shift* comportemental du sujet de l'expérimentation, il paraît alors évident pour l'expérimentateur d'en trouver des traces dans le signal. En poussant cette idée à son paroxysme (de manière très naïve, certes), l'idée sous-jacente à cette démarche serait de dire que toutes les valeurs des indices acoustiques d'imitations successives d'un locuteur doivent converger avec toutes les valeurs des indices acoustiques du modèle de l'imitation, soit, que le *shift* comportemental se produit en direction du modèle.

En fonction du niveau d'étude envisagé –segmental ou prosodique–, l'expérimentateur peut choisir de mesurer de nombreux paramètres acoustiques, comme le montrent les revues de littérature de Pardo (2013a, pp. 2–3) et de Lewandowski (2012, pp. 114–116). Pour les seules études relevant de l'adaptation dyadique en conversation ou de la convergence, ces chercheuses relèvent une grande variété d'indices mesurés comme par exemple des aspects de durée (Goldinger, 1998; Pardo, 2010), le débit de parole (Pardo, 2013;

Pardo et al., 2010), des aspects relatifs à la fréquence fondamentale (Babel & Bulatov, 2012; Goldinger, 1998; Levitan & Hirschberg, 2011), au spectre des voyelles (Babel, 2012; Pardo et al., 2012, 2010) ou encore à l'usage de variantes phonémiques (Delvaux et al., 2014; Delvaux & Soquet, 2007).

D'autres études citées précédemment, ayant cette fois trait au phénomène d'impersonnation, ont proposé d'utiliser des relevés des valeurs de *f0* –moyenne (Farrús et al., 2008a; Mejvaldova, 2002; Révis et al., 2013; Zetterholm, 2002, 2002, 2006), étendue (Attal-Fiocchi & Jarzé, 2014; Brzostek & Deschanvres, 2014; Farrús et al., 2008a, 2008b; Révis et al., 2013) et/ou contours (Mary et al., 2012; Mejvaldova, 2002)– ainsi que des relevés d'aspects temporels –durée des segments (Eriksson & Wretling, 1997; Farrús et al., 2008a, 2008b, Mejvaldova, 2002, 2002, Zetterholm, 2000, 2002, 2006), présence et durée des pauses (Brzostek & Deschanvres, 2014; Révis et al., 2013)– afin de décrire la flexibilité vocale des imitateurs professionnels. Ces dernières études ont par ailleurs scruté d'autres aspects comme le débit de parole (Mejvaldova, 2002; Révis et al., 2013), les valeurs des formants vocaliques (Eriksson & Wretling, 1997; Zetterholm, 2000, 2006) ou encore la déviation temporelle (entre modèle et imitation) de certains événements phonétiques –début d'item lexical (Eriksson & Wretling, 1997; Zetterholm, 2002) ou pics de *f0* (Zetterholm, 2002)–.

Devant une telle variété de mesures, il paraît relativement peu évident de déterminer la -ou lesquelles- pourraient constituer un indicateur satisfaisant de l'imitation. Bien que les mesures des études en impersonnation illustrent la flexibilité vocale des locuteurs, les mesures produites dans les études en tâche de convergence ont une dimension plus ambiguë. En effet, Pardo (2013) remarque :

> Although multiple acoustic attributes are represented across these studies, this practice does not provide a comprehensive assessment of convergence. […] First, talkers likely converge on multiple attributes simultaneously. Second, talkers might converge on some attributes at the same time that they diverge on other attributes. Third, talkers might converge on one set of attributes for one set of items or talkers and another set of attributes for other items or talkers.

Ainsi, il semble que les mesures produites durant les tâches d'impersonnation soient interprétables plus facilement que les mesures issues de tâches de convergence. Les premières devraient en effet refléter la flexibilité vocale du locuteur, la tentative intentionnelle d'usurper la voix de l'autre résultant des consignes émises durant le recueil de données. Malgré tout, ces premières mesures ne nous renseignent pas réellement sur le ou les paramètres acoustiques que l'impersonnateur contrôle réellement. Par exemple : si le *jitter* ou le *shimmer* d'un

locuteur converge (Farrús *et al.*, 2008a), peut-on dire que le locuteur opère un contrôle sur ce dernier, ou bien est-ce là le résultat d'un processus plus global, comme le laisse penser Révis (2013, p.57) ? A propos des mesures acoustiques issues de tâches conversationnelles, ces dernières semblent particulièrement bruitées comme l'illustre le propos de Pardo (2013c), que nous citions page précédente. D'une part, les différences entre les lignes de base des locuteurs et les items des tâches expérimentales semblent quantitativement réduites puisque, d'après Lewandowski (2012, p. 120), la magnitude de l'effet de convergence serait assez réduite. D'autre part, il apparaît que l'utilisation de mesures acoustiques seules soit vouée à une impasse, dans la mesure où identifier un paramètre acoustique particulier responsable de l'effet d'imitation semble être une tâche impossible (Lewandowski, *Ibid.*, p. 123).

Malgré le fait que les mesures acoustiques revêtent toujours un intérêt pour le phonéticien, ces dernières se révèlent donc particulièrement mal aisées à interpréter et à lier les unes aux autres dans le cadre des processus d'imitations parolières. Cependant, quelques cas spécifiques (comme l'impersonnation, ou la reproduction de sons inconnus) nous semblent pouvoir être investigués au moyen des seules mesures acoustiques tout en donnant un indice suffisant des processus mis en jeu, car les consignes données durant l'expérience et les points de mesures ciblés sont suffisamment explicites pour lever certains doutes du chercheur en imitation quant aux variations observées : celles-ci sont-elles dues au phénomène investigué ou à la variation intrinsèque à la production de la parole ?

Finalement, toute mesure acoustique semble pouvoir rendre compte du fait qu'il y a eu imitation tant que la comparaison du modèle et de sa reproduction va dans le bon sens. Cependant, aucune mesure acoustique particulière ne semble pouvoir certifier plus qu'une autre la réussite d'une imitation ou le degré de similarité entre modèle et reproduction à un niveau global. Le dilemme du chercheur procédant à des mesures acoustiques mérite alors d'être rappelé : doit-il être tenté de cumuler toutes les mesures acoustiques imaginables afin d'être certain de ne pas manquer l'effet qu'il recherche ou bien se concentrer sur quelques mesures de son choix et risquer de ne rien trouver ? Par ailleurs, il convient également de mettre en exergue l'importance de l'auditoire comme juge des productions imitatives : l'oreille humaine pourrait finalement être un recours non négligeable pour obtenir un indice du degré d'imitation contenu dans la parole par le biais d'évaluations perceptives.

2.2 Approches perceptives de l'évaluation de l'imitation

L'évaluation perceptive de l'imitation n'est pas un procédé auquel les chercheurs ont automatiquement recours. Par exemple, la majorité des études que nous citions précédemment dans notre revue sur l'impersonnation (Attal-Fiocchi & Jarzé, 2014; Brzostek & Deschanvres, 2014; Eriksson & Wretling, 1997; Mejvaldova, 2002; Révis et al., 2013; Zetterholm, 2002, 2002, 2002, 2006, 2009a) ne proposaient pas d'évaluation perceptive de la réussite d'une imitation. Prenant pour acquis que leur sujet était un expert, son imitation était forcément réussie, sa cible, reconnaissable. Cependant, le cas de l'impersonnation demeure particulier car l'expertise du locuteur et/ou son intention de changer sa voix rendent caduque la nécessité du test perceptif : des mesures acoustiques comparatives peuvent suffire à repérer le *shift* comportemental de l'impersonnateur, que l'imitation soit réussie ou non.

Dans le cas d'autres phénomènes imitatifs en parole, comme la convergence, la nécessité de mener des évaluations perceptives peut émerger. En effet, le phénomène de convergence phonétique implique qu'il y a un changement des caractéristiques acoustiques de la production des locuteurs. Cependant, le manque de cohérence dans les patterns et la faible magnitude des effets conduit à une impasse : il semble risqué de ne repérer la convergence phonétique qu'au moyen des mesures acoustiques.

Pour Pardo (2000, 2006), les tests perceptifs représentent donc un format privilégié de l'évaluation de la convergence phonétique :

> Perceptual measures provide a more global appraisal of convergence that is not limited to the particular acoustic-phonetic attribute a researcher might choose. First, it is not possible to characterize the phenomenon without including multiple dimensions. Talkers can converge on one dimension at the same time that they diverge or produce random variation in other dimensions. […] this flexibility extends across different items. (Pardo, 2013, p. 2)

Ainsi, la « mesure[56] » perceptive permettrait d'envisager les productions imitatives d'un point de vue plus englobant, holistique. En effet, nous nous rapprocherions alors d'une situation commune où un auditeur évalue la production d'un tiers en lui trouvant un air de déjà vu, *i.e.* où un juge attribue à une production parolière son statut d'imitation, ou de non imitation. Ici seront décrits les types de tests perceptifs utilisés dans la littérature sur l'imitation en parole afin d'en évaluer l'utilité et les limites pour le chercheur. Le paradigme

[56] Le terme de mesure perceptive ne nous convient pas. Nous lui substituerons celui « d'évaluation » un peu plus loin.

AXB (Goldinger, 1998; Namy et al., 2002; Pardo, 2006) constituera ce que nous appellerons une évaluation relative, et le paradigme AX (Mary et al., 2012, 2013), une évaluation absolue.

2.2.1 Format AXB : une évaluation *relative* d'énoncés imitatifs

Relevant de ce que nous avons décrit comme une triade entre un comportement modèle, un imitateur et un comportement produit, les comportement imitatifs doivent être évalués *a minima* par paires « imitation *vs* modèle ». En effet, l'imitation se définissant par sa relation au modèle, un test visant à évaluer des aspects d'imitation doit permettre à l'auditeur d'établir la présence ou l'absence de cette correspondance modèle/imitation. La solution la plus naturelle pourrait être de présenter à l'auditeur des modèles X accompagnés d'une imitation A. Pourtant, ce paradigme AX n'est pas très fréquent dans la littérature, notamment dans celle qui traite de la convergence phonétique, où un paradigme AXB est préféré. Nous nous arrêterons quelques instants sur ce dernier paradigme.

Les tests d'évaluation perceptive AXB (Goldinger, 1998; Namy et al., 2002; Pardo, 2000) ou ABX (Dupoux *et al.*, 1997) proposent à des auditeurs d'entendre des sons par triplets, où X est le modèle, A et B sont deux reproductions (ou simples itérations) de celui-ci, opposables selon leur conditions de recueil (par exemple : ligne de base *vs.* conversation) ou leurs caractéristiques phonologiques. Par exemple, Pardo (2006) apparie des repères géographiques prononcés par deux sujets durant une *Map Task* de la manière suivante :

- X est prononcé par un premier sujet (S1) dans la conversation
- A & B apparient alternativement des items correspondants à X, prononcés par un second sujet (S2)
    - Pendant le pré-test
    - Pendant la conversation
    - Pendant le post-test
- Les paires possibles se décrivent ainsi :
    - $A_{\text{pré-test}}$ *vs.* $B_{\text{conversation}}$
    - $A_{\text{post-test}}$ *vs.* $B_{\text{conversation}}$
    - $A_{\text{pré-test}}$ *vs.* $B_{\text{post-test}}$
- Afin d'éviter un biais lié à l'ordre de présentation des items, chaque triplet doit être présenté dans les deux ordres possibles

- Enfin, on obtient des ensembles supplémentaires en échangeant les places de S1 et S2 dans le paradigme, en position X ou A & B.

Pour l'auditeur, le but de la tâche est alors de déterminer qui de A ou de B ressemble le plus à X en termes phonétiques, de manière très rapide. En fait, l'auditeur doit faire un choix forcé entre les deux items A et B, que ceux-ci lui semblent très différents ou bien très similaires. En ce sens, nous souhaitons souligner que l'évaluation perceptive proposée par le paradigme AXB est relative : A (ou B) est choisi au détriment de B (ou A), comme ressemblant phonétiquement plus que B (ou A) à X.

En fait, depuis Goldinger (1998) et Namy *et al.* (2002) les résultats des tests de type AXB constituent un moyen de quantifier l'imitation phonétique :

> [NDR : Pardo évoque l'étude de Goldinger] Imitation was quantified as the percentage of shadowed items that sounded more similar to a model's talker than baseline item. (Pardo, 2013a, p.2).

Goldinger (1998, p259, note 4) confirme la capacité des auditeurs à détecter l'imitation dans ce paradigme, bien qu'il soit insuffisant à nous renseigner sur les bases perceptuelles sur laquelle repose l'identification de l'imitation. En ce qui concerne Pardo, les résultats de ces tests constituent une mesure à privilégier de la convergence phonétique : elle systématise son utilisation dans ses travaux (Pardo, 2000, 2006, 2013, Pardo et al., 2012, 2010).

Le taux de convergence phonétique, issu des tests AXB, augmente si les auditeurs choisissent le plus souvent des items produits durant la tâche de conversation que des items produits lors du recueil de la ligne de base. Soit x un item recueilli, S_x le score de x au test AXB, et n le nombre total de triplets contenant x, le rapport estimant le taux de convergence $Conv(x)$ doit être le suivant :

$$Conv(x) = \frac{S_x}{n} \times 100$$

avec :

$$S_x \in R$$

et

$$0 \leq S_x \leq n$$

Le résultat d'une telle expression prend du sens, si et seulement si, un seuil critique permet de l'interpréter. La proposition de Pardo est de considérer qu'il y a convergence phonétique si on obtient un score dépassant les 50% de chance que les items issus du *shadowing* soient choisis au détriment des autres. Cependant, un seuil si proche des probabilités associées à un jet de pièce (50/50) demande un grand nombre d'observation pour être fiable. C'est d'ailleurs une des limites que Pardo voit à l'utilisation des mesures perceptives :

> Just as talkers might converge on different attributes across different items, listeners are not likely to be uniform in their appraisal of similarity. Therefore, it is necessary to collect data from multiple listeners per talker pair, and to incorporate all levels of variability in the analyses. (Pardo, 2013a)

Outre cette première limite liée au nombre d'auditeurs nécessaires pour obtenir un index fiable de la convergence phonétique, Pardo (2013a) rejoint l'avis de Goldinger (1998) sur l'information réelle fournie par ces tests : une simple évaluation de la convergence sans indices sur les paramètres acoustiques saillants qui permettent au locuteur de fournir son jugement. En ce sens, nous privilégierons le terme « évaluation perceptive » au terme « mesure perceptive » utilisé par Pardo.

> The exclusive use of perceptual measures does not provide information about which acoustic-phonetic attributes were employed by the talkers. (Pardo, 2013a)

Bien qu'il soit possible d'orienter l'attention de l'auditeur vers certains aspects de la parole (par exemple : la prononciation des locuteurs ou la musicalité des énoncés), les tests perceptifs, qu'ils adoptent le paradigme AXB ou AX, laissent toujours une inconnue quant à la cause de leur résultats.

2.2.2 A propos du paradigme AX et d'une évaluation *absolue* de l'imitation

Ce que nous dénommons paradigme AX, consiste en fait en une approche comparative des plus simples : il s'agit d'apparier les items –une imitation et son modèle– puis de les soumettre au jugement d'un tiers. Cela conduit à une approche très différente de la matière phonique que celle proposée par le paradigme AXB tout en posant également un certain nombre de problèmes liés à la nature même de la tâche AX. Dans un premier temps, il convient de constater que l'évaluation fournie ici, se base donc sur les qualités intrinsèques de l'énoncé évalué ; contrairement à l'évaluation faite dans le paradigme AXB, qui propose de choisir le plus proche de X, au détriment du plus lointain. Ainsi, ayant qualifié le paradigme

AXB d'évaluation *relative* (car les items sont classés en fonction des autres), il conviendrait alors de qualifier le paradigme AX d'évaluation *absolue*, puisque chaque élément X de chaque paire AX est évalué isolément.

Dans la tâche AX, un auditeur doit donc comparer un modèle A à une reproduction X. Pour ce faire, l'expérimentateur met à disposition de l'auditeur une échelle graduée (de 1 à 5 pour Hermes, 1998a ; de 1 à 6 dans l'étude de Mary *et al*., 2013) afin que la distance (ou l'absence de distance) entre A et X soit évaluée perceptivement. Jusqu'à présent, les exemples présentés concernaient toujours des fragments de paroles. Cependant, Hermes (1998a) propose deux tests AX différents pour évaluer la dissimilarité prosodique d'énoncés et de leurs formes resynthétisées : le premier test utilise leur forme sonore et le second leur représentation de la *f0*. Le troisième volet de cette étude (Hermes, 1998b) étend la forme de test AX à une approche computationnelle où la similarité de deux contours prosodiques est mesurée au moyen de plusieurs calculs de distance entre courbes. Ces études soulignent plusieurs points conceptuellement intéressants à propos de l'évaluation et/ou de la mesure[57] de l'imitation.

Notons en premier que les notions de similarité et de dissimilarité sont concomitantes : évaluer l'une revient à évaluer l'autre. Comme le note Hermes (1998b, p. 75), il peut être possible de différencier les deux en fonction de la manière dont une valeur numérique obtenue au moyen de l'évaluation est interprétable. Si un résultat élevé, signifie que la similarité est haute, alors on peut dire qu'il s'agit d'une évaluation de la similarité, et inversement si un résultat élevé signifie une grande dissimilarité. Dans le cas de mesures objectives, l'interprétation d'un test de dissimilarité comme indicateur de similarité doit pouvoir se faire sans *a priori* en raison de la nature stable de la mesure. A propos des évaluations perceptives, la même démarche nous semble sujette à caution : il faudrait d'abord pouvoir s'assurer qu'un changement de consigne ou d'échelle d'évaluation puisse se faire sans biais particulier.

Par ailleurs, en proposant d'évaluer/mesurer les mêmes items sous différentes modalités, ces études impulsent une dynamique de validation des différents types de notation de l'imitation. Idéalement, un ensemble de productions imitatives devrait obtenir une même répartition des résultats quelle que soit la méthode retenue d'évaluation/mesure. Ainsi, en constatant la corrélation entre une mesure automatique et une évaluation perceptive, Hermes

[57] Dans les paragraphes suivants, nous utiliserons systématiquement le terme « évaluation » plutôt que « mesure ». Il nous semble important de distinguer autant que possible les deux termes. Cependant, les deux termes seraient ici interchangeables, sans pâtir d'amalgame conceptuel.

(1998b) ouvre une piste prometteuse d'un point de vue expérimental, si l'objectivation et l'automatisation des évaluations perceptives était rendu possible par de telles approches.

Ce point-ci semble crucial, car les évaluations effectuées dans le paradigme AX sont soumises aux mêmes limites que les évaluations issues de tests AXB. Nous le mentionnions plus tôt, les auditeurs auraient potentiellement une variabilité interindividuelle dans leur manière de percevoir la dis/similarité, comme le souligne Pardo (2013). De plus, face à une échelle de notation il semble que certains correcteurs rechignent à utiliser les extrêmes ou, au contraire, les privilégient. Enfin, par rapport aux tests AXB, les tests AX ne donnent pas d'informations supplémentaires en ce qui concerne les raisons physiques de la note, *i.e.* quels éléments acoustiques ont subi des modifications. En ce sens, les tests AX peuvent être intégrés à une approche du type de celle de Pardo (2013a). Ils bénéficient alors des mêmes avantages et pâtissent des mêmes limites que les tests AXB.

2.2.3 Pour une double approche de l'imitation en parole

Les études sur l'imitation en parole cherchent principalement à évaluer :

- l'effet de l'exposition à la parole de l'autre
- l'effet de différentes consignes sur la production d'un imitateur
- le niveau d'expertise d'un imitateur (par rapport à un autre).

Ces objectifs soulignent la nécessité de comparer les items entre eux, ce qui a été argué précédemment. Pour ce faire, nous avons jusqu'à présent envisagé deux possibilités –les mesures acoustiques et les tests perceptifs–, qui, isolément, semble insuffisants pour à la fois détecter les aspects imitatifs ou fournir des arguments explicatifs sur la manière dont les locuteurs produisent de l'imitation.

La granularité très fine des mesures acoustiques, qui permet habituellement d'obtenir des informations précieuses sur le signal de parole, pose un problème relativement insoluble dans le cadre de l'étude de l'imitation. En effet, en fonction des locuteurs et/ou des items, les paramètres acoustiques mesurés adoptent des comportements erratiques. Par exemple, les *patterns* de convergence concernant les voyelles dans les études de Pardo (2010) et de Babel (2012), sont quasiment opposés : les locuteurs de Babel convergent plus sur /æ/ et /a/ et de manière moins significative sur /i/ et /u/ tandis que les locuteurs de Pardo suivent des tendances inverses, convergeant sur /i/ et /u/, divergeant même sur /æ/ et /a/. Ce simple

exemple suffit à souligner les avantages et inconvénients liés aux mesures acoustiques. Du fait de leur focale grossissante, les mesures acoustiques s'intéressent à des composantes précises du signal de parole, cherchant alors –dans le cadre des imitations parolières– à déterminer les paramètres acoustiques sur lesquels l'imitation repose. Cependant, cette même focale conduit la mesure acoustique à être « inconstante », *i.e.* à manquer de fiabilité, puisque la variabilité des productions parolières d'un même locuteur peut se traduire par une variété dans la composition du cocktail de paramètres acoustiques en jeu dans ses différentes productions imitatives. A ceci, il convient d'ajouter la variabilité interindividuelle, exacerbant encore le travers que nous venons de souligner.

D'un point de vue métaphorique, l'utilisation de mesures acoustiques seules pour évaluer la similarité de segments de paroles pourrait donc équivaloir à estimer la similarité de deux maisons différentes en observant une à une les briques dont elles sont construites. Ceci étant dit, les mesures acoustiques jouissent tout de même d'une stabilité certaine liée à leur statut de « mesure » : cette dernière est due à l'outil dont la mesure est issue, outil dans lequel le chercheur peut avoir confiance s'il le connaît suffisamment pour en connaître les limites, comme le notait déjà Duhem dans le domaine de la physique théorique (Duhem, 2007, p. 23 & 192; éd. originale 1906).

A l'opposé des mesures acoustiques, les tests perceptifs proposent d'évaluer la similarité des items de manière holistique. Si nous devions reprendre notre métaphore précédente, l'évaluation perceptive de la similarité en parole équivaudrait à évaluer la similarité de deux maisons différentes en un coup d'œil, sans s'intéresser aux différences structurelles de celles-ci, soit : sans observer les matériaux de construction sous le crépi de la maison. Cette approche présente des avantages dont ne bénéficie pas l'approche de mesurage acoustique : ici, les items de parole sont évalués dans leur ensemble par des groupes d'auditeurs, palliant ainsi le besoin d'élargir le champ de la mesure acoustique.

> When listeners perform a perceptual similarity task, they are effectively collapsing across these multi-dimensionnal acoustic patterns, providing a more reliable measure of phonetic convergence than any single acoustic-phonetic attribute. (Pardo, 2013, p. 2)

Ainsi, ces tests fournissent une information relativement fiable –en tenant compte de la variabilité des auditeurs– et leurs résultats permettent :

- d'évaluer (contrairement à Pardo qui estime « mesurer ») le taux de convergence

- de classer les items en fonction de leur similarité avec le modèle (*i.e.* la qualité de l'imitation), à mesure qu'un ensemble d'items recueille un nombre grandissant de jugements.

Cela dit, l'évaluation perceptive évoquée souffre de deux dérives qu'il nous semble important de souligner :

- son holisme, qui constitue un avantage, est aussi paradoxalement, un inconvénient puisque l'évaluation ne nous renseigne pas sur le détail du contenu
- en tant qu'évaluation, les résultats obtenus ont une certaine instabilité, car les évaluateurs ne fournissent pas automatiquement le même résultat pour un même item (ou bien deux évaluateurs)[58]

S'ils sont employés séparément, les évaluations perceptives, holistiques mais instables, ainsi que le mesurage acoustique, fragmenté mais stable, semblent inaptes à investiguer les comportements imitatifs en parole. En revanche, la proposition de Pardo (2013) consistant à employer les deux approches de manière parallèle nous semble plus satisfaisante. En d'autres termes, une utilisation conjointe de ces deux approches amène le chercheur à s'inscrire dans une dynamique vertueuse reflétant la complexité du phénomène imitatif.

En effet, il est nécessaire pour l'expérimentateur d'obtenir un diagnostic de la similarité perceptuelle entre deux items, afin d'avoir une base solide dans le but de comparer les items les mieux imités avec les items les moins bien imités. Cette double approche comparatiste a été raffinée par Pardo (2000, 2006, 2010, 2013) au fil de ses recherches (*cf.* Tableau 3). En fonction des résultats de ses tests AXB et des données acoustiques recueillies, J. S. Pardo propose dernièrement de jauger l'importance des paramètres acoustiques dans le jugement des auditeurs au moyen de modèles statistiques mixtes (Baayen, 2008).

Les travaux de Pardo, qui constituent une pierre angulaire de notre réflexion méthodologique, présentent pour nous les limites de se focaliser uniquement sur la prononciation des phonèmes et de ne proposer dans les tests perceptifs que l'écoute de mots isolés. Hormis les tests perceptifs, qui sont une approche globale de la matière phonique, Pardo évalue l'importance des indices acoustiques dans la convergence sur des bases statistiques. Ainsi, son approche repose à la fois sur une évaluation perceptive –instable– dont les indices de mesures acoustiques –stables, mais très locaux– sont fusionnés dans une approche probabiliste pour en estimer l'impact dans les résultats de l'évaluation perceptive.

[58] Contrairement à la mesure que fourniraient deux double-décimètres pris au hasard dans le commerce.

Ces travaux soulignent la complexité intrinsèque de l'étude de l'imitation en parole, qui se traduit par une nécessaire approche plurielle et comparatiste pour qui veut estimer ou mesurer la similarité entre un modèle et son imitation.

3 Approches comparatives des contours intonatifs

Nous venons de défendre une approche double pour l'évaluation/mesure de l'imitation en parole. A propos du niveau prosodique, et plus particulièrement du contour intonatif, il semble approprié de nous pencher sur les approches à disposition afin de pouvoir comparer des courbes de *f0* dans l'optique de proposer par la suite une sélection de plusieurs méthodes.

Certains travaux cités précédemment, comme ceux de Révis *et al.* (2013), proposaient de mesurer la hauteur de la *f0* en certains points, d'évaluer sa moyenne, ou son étendue. Si elle est intéressante pour illustrer la flexibilité vocale d'un locuteur par rapport à une cible de référence, cette seule approche de mesurage acoustique nous semble insuffisante dès lors que l'on cherche à déterminer si la forme même du contour (sa configuration tonale) a été reproduite. En effet, la mesure locale ne nous renseigne pas sur la succession des tons et les indices comme la moyenne ou l'étendue donnent de simples informations sur la distribution des valeurs de fréquence fondamentale dans la production d'un imitateur. Cependant, indépendamment de la forme du contour, ces indices documentent les changements comportementaux des imitateurs, notamment les stratégies mises en œuvre par ces derniers.

A ce propos, Révis *et al.* (*Ibid.* p.8) indiquent dans la discussion de leur article que les imitateurs utiliseraient deux grands types de stratégies : de convergence ou de synchronie. La conjugaison de ces deux types de stratégies serait le signe d'une imitation potentiellement réussie. Les stratégies de convergence réfèrent à des stratégies d'ajustement global de la voix de l'imitateur à la voix de la cible, par exemple : quand un imitateur élève le niveau moyen de sa fréquence fondamentale, car sa voix est plus grave que celle de la cible. Cependant, l'emploi de cette seule stratégie ne serait pas suffisant pour arriver à imiter de manière convaincante un locuteur.

En effet, il faudrait en plus parvenir à mettre en place des stratégies de synchronie, c'est-à-dire : à imiter des variations instantanées de la parole de l'autre. Par variation instantanée, il est sous-entendu des événements comme la fréquence, la position et la durée des pauses, ou encore, le contour intonatif. A ce sujet, il peut être intentionnel de la part de l'imitateur de ne reproduire que les variations instantanées perçues tout en gardant les caractéristiques naturelles de sa voix : c'est le cas de l'enseignant produisant un logatome afin de faire percevoir à l'apprenant la syllabation et l'intonation spécifique d'un énoncé. Ce dernier cas occupe une place centrale dans notre travail :

- En terme de contrôle de prosodique, on peut se demander si des locuteurs naïfs ou experts parviennent à produire sciemment les mouvements prosodiques préconisés par la MVT dans le cas de la correction des erreurs segmentales, ou suggérés par le modèle entendu dans le cas de reproduction logatomique.
- Aucun outil n'existe pour mesurer la validité des logatomes produits par l'enseignant.

L'objet de cette sous-partie est donc de considérer les méthodes et approches pour comparer des formes ; en l'occurrence, les forme décrites par des courbes de *f0*. Nous proposerons dans un premier temps quelques considérations concernant la problématique de la comparaison des formes. Nous soulignerons ensuite une approche globale des courbes de *f0* sous une forme « brute », qui essaierait de saisir la distance physique entre deux courbes. Consécutivement, nous considèrerons le recours à l'annotation phonologique de la *f0* afin de catégoriser les événements tonals qui doivent servir à faire émerger les points d'ancrage d'une forme du patron intonatif. Nous présenterons finalement une méthode pour comparer les formes, issue de domaines hors cadre des sciences du langage (la géométrie) et que nous nous proposons d'appliquer à des patrons intonatifs stylisés.

3.1 Formes et comparaisons : le cas de la courbe intonative

Définir la notion de forme n'est pas aussi évident qu'il y paraît. Il n'y aurait en effet pas de définition universelle de ce qu'est une forme (Veltkamp, 2001, p 1). Nous pouvons alors nous poser la question de savoir si décrire un contour intonatif comme une « forme » est légitime. Ce point théorique semble important, car la caractérisation d'une courbe de *f0* en tant que « forme » pourrait résoudre une part de nos problèmes méthodologiques par le recours à des techniques et approches du domaine de la « correspondance des formes » (ci-après : *shape matching*).

A propos de la notion de forme, Veltkamp (*Ibid.*) indique :

> Impressions of shape can be conveyed by color or intensity patterns (texture), from which a geometrical representation can be derived. […] Here, we […] consider shape as something geometrical. We […] use the term shape for a geometrical pattern, consisting of a set of points, curves, surfaces, solids, etc. (p. 1-2)

Lorsqu'il étudie l'intonation, le chercheur s'intéresse à la courbe de fréquence fondamentale (*f0*). La courbe de *f0*, qui représente la fréquence de vibration des cordes vocales, est la

modélisation du corrélat physique de l'attribut perceptuel du son appelé mélodie (Rossi, Di Cristo, Hirst, Martin, & Nishinuma, 1981). En d'autres termes, l'intonation (la variation de *f0* en fonction du temps) est décrite par une forme que l'on peut visualiser. Un enjeu majeur, en ce qui nous concerne, est alors de déterminer si comparer des formes sonores (la mélodie perçue) équivaut à comparer les formes géométriques en résultant (la représentation graphique de leur *f0*). La question de la comparaison des formes sonores sera effectuée au moyen de tests perceptifs, tels que nous les avons décrits précédemment. En ce qui concerne le second volet de cette question, nous développons en suivant des méthodes relevant du *shape matching*.

Veltkamp (2001, p.1-2) décrit l'approche de *shape matching* comme suit :

> Shape matching deals with transforming a shape and measuring the ressemblance with another one, using some similarity measure. So, shape similarity measures are an essential ingredient in shape matching.

Le *shape matching* vise donc à estimer la distance physique (dissimilarité) ou la proximité (similarité) entre deux formes. Cette notion concomitante de dis/similarité ayant été abordée précédemment, nous focaliserons notre attention sur les types de problèmes liées au *shape matching*, telles que Veltkamp (2001, pp. 3-4) les décrit et qui font écho à nos préoccupations. En effet, d'après ce dernier, lorsque l'on compare deux patterns au moyen d'une mesure de dis/similarité, on a affaire alternativement à un problème de :

- computation : si on cherche à calculer la dis/similarité entre deux formes
- décision : si, pour un seuil donné, on doit décider si la similarité entre deux formes est plus petite que le seuil
- optimisation : si on souhaite trouver la transformation minimisant la dis/similarité entre le pattern transformé et l'autre pattern.

Or, ces séries de problèmes ont des résonnances dans différents types d'applications nous intéressant (recherche, reconnaissance, classification, alignement, approximation des formes).

Par exemple, les prosodistes travaillant sur de grands corpus de parole pourraient trouver utile de disposer d'outils leur permettant de localiser automatiquement le type de contours intonatifs qu'ils recherchent (problèmes de décision et de computation). En ce qui concerne la mesure de la dis/similarité entre deux représentations de la *f0*, nous nous situons d'une part dans les applications de reconnaissance et de classification, et dans celles d'alignement et d'approximation d'autre part.

En effet, cherchant à savoir si un imitateur est parvenu à reproduire un contour intonatif, et à quel degré il y est parvenu, nous nous appliquons à reconnaître et à classifier des formes, soit, à « *déterminer si une forme donnée correspond suffisamment à un modèle*[59] » (problème de décision) (Veltkamp, 2001). Pour rendre des formes comparables, il peut être nécessaire de les transformer au préalable, nous appliquons soit un alignement (comme une interpolation linéaire ou le *Dynamic Time Warping*, (Kim, 2012; Rilliard et al., 2011), soit nous cherchons à approximer la forme pour nous absoudre d'une partie de sa complexité avant comparaison, c'est-à-dire : « *construire une forme à partir de moins d'éléments (points, segments, triangles, etc.), qui est toujours similaire à la forme originale*[60] » (Veltkamp, *Ibid.*). Pour parvenir à résoudre ces questions, nous devons répondre au problème de computation, soit : calculer la dis/similarité entre les contours.

3.2 Mesures de la distance prosodique et perception de la similarité

La courbe de *f0* décrit des pics et des creux montrant la variation de la fréquence de vibration des cordes vocales, soit, la réalité physique de l'intonation. Elle est à la fois porteuse d'informations linguistiques (systémiques : rythme et accentuation, pragmatiques : accents emphatiques, par exemple) et paralinguistiques (émotions du locuteur, registre, état de santé). Cela dit, la *f0* n'est ni une courbe continue ni une courbe lisse. En effet, lors de la production de consonnes dévoisées, les cordes vocales ne vibrent pas et la *f0* perd alors sa continuité. Par ailleurs, les événements micro-prosodiques, conséquences de l'articulation des phonèmes, ont une influence sur la forme de la courbe (Di Cristo, 1978). Par exemple, lors de la production d'une occlusive dévoisée, la courbe de *f0* diminue légèrement lors de la tenue (contact des articulateurs supérieurs et inférieurs), puis elle augmente brusquement au moment de l'explosion. Ainsi, la courbe de *f0* au plus proche de sa réalité physique (ce que nous désignerons comme « *f0* brute ») est une forme complexe, discontinue et bruitée. Ces formes brutes de *f0* peuvent être l'objet d'un calcul de distance.

[59] Determine wether a given shape matches a model sufficiently close
[60] Construct a shape of fewer elements (points, segments, triangles, etc.), that is still similar to the original

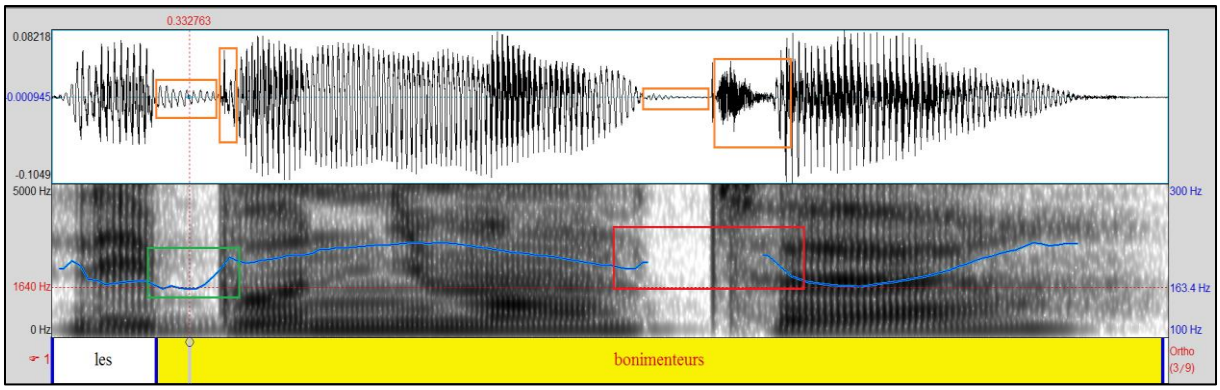


Figure 19 : Illustration d'événements micro-prosodiques. Oscillogramme et spectrogramme d'une production de « Les bonimenteurs ». La *f0* est représentée par la courbe bleue. Le cadre vert indique la variation de *f0* lors de la production de /b/. On observe une légère descente de la courbe avant une montée brusque. Sur l'oscillogramme, on remarque une perte d'amplitude de l'onde puis une explosion (cadres orange). Le cadre rouge montre la perte de *f0* lors du dévoisement du /t/. Sur l'oscillogramme, on observe un vide acoustique, suivi d'une explosion (cette production de /t/ est presque affriquée…)

Plus haut, nous avons évoqué les études de Hermes (1998a, 1998b) s'intéressant à l'évaluation de contours de *f0* sous une modalité perceptive (1998a) et aussi au moyen de mesures physiques (1998b). L'approche de Hermes a levé un verrou méthodologique dans notre travail, puisque :

- Il propose une double approche (perceptive et computationnelle) de la comparaison des *f0*, allant dans le sens de ce que nous défendons pour l'étude de l'imitation en parole
- Les mesures proposées semblent adaptées à notre objet de recherche

Par ailleurs, ces études laissent entrevoir la possibilité d'automatiser l'évaluation de la similarité prosodique, en sélectionnant la ou les mesures qui reflètent le mieux les résultats de tests perceptifs :

> The basic idea underlying the testing of these measures is that, if the perceptual mechanism that processes the similarity between two pitch contours predominantly gives weight to the similarity represented by such measure, the correspondence between the ratings obtained with this measure and the perceptual ratings obtained from listeners will be high. If this is the case, it is concluded that such a physical measure well represents the physical basis on which these perceptual ratings are based. (Hermes, 1998b, p. 74)

La mesure de similarité qui obtiendrait la meilleure corrélation avec les résultats de tests perceptifs pourrait, selon cette perspective, représenter un candidat de choix en vue d'une

implémentation dans un outil d'entraînement à la production de contours mélodiques spécifiques.

En ce qui concerne les mesures de similarité entre deux courbes, Hermes (1998b) en passe quatre en revue :

- *Mesure de tunnel* : ce procédé consiste à repérer si le contour évalué dépasse ou non les valeurs plafond et plancher établies par rapport au contour de référence ; si le contour évalué « sort » du tunnel, il est évalué comme non similaire (*cf.* Figure 20).

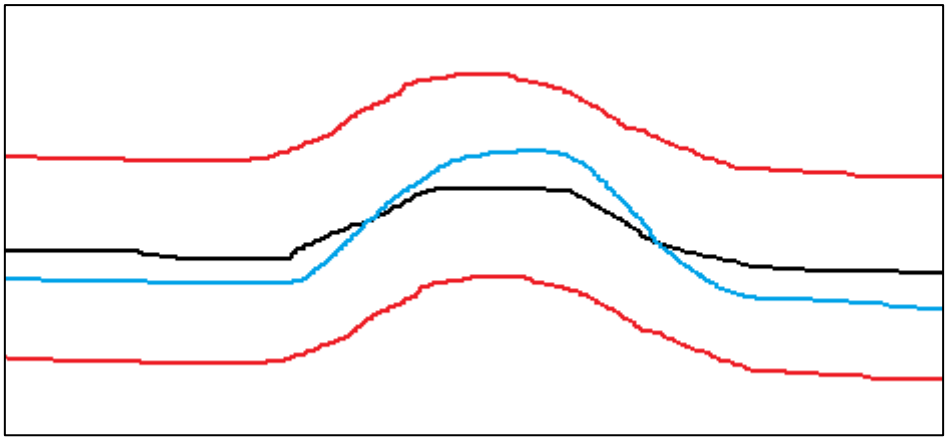


Figure 20 : Concept de la mesure de tunnel.La courbe noire représente le contour de référence et les courbes rouges, le tunnel. La courbe bleue représente le contour évalué. Ce dernier est jugé similaire (l'imitation est réussie…), s'il ne dépasse pas des limites du tunnel.

Cependant, cette mesure est, d'après Hermes (1998b, p. 75) un candidat non idéal pour évaluer la similarité de deux courbes mélodiques, car les événements micro-prosodiques seraient susceptibles de fausser les résultats de la mesure en cas de valeurs de tunnel trop étroites. De plus, si les valeurs du tunnel sont trop larges, la mesure est susceptible de produire un biais que nous illustrons dans la Figure 21.

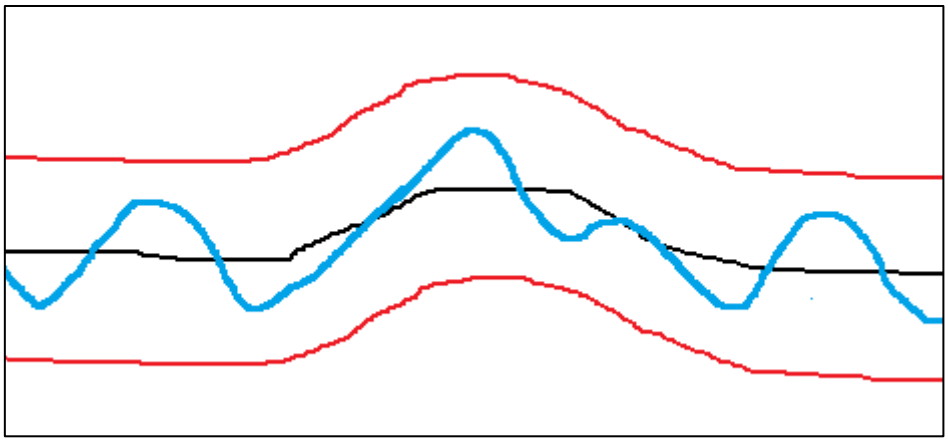


Figure 21 : Biais potentiel de la mesure de tunnel. Bien qu'elle soit intégrée dans les limites du tunnel, la courbe bleue a un contour très différent de celui de la courbe de référence. Cependant, comme leur distance maximum n'est pas très grande, elle pourrait être évaluée comme similaire à la référence…

Hormis la mesure de tunnel dont nous venons d'évoquer les biais, Hermes décrit encore trois mesures de similarité :

- *Différence moyenne absolue* : cette mesure cherche à définir la distance moyenne entre les contours évalués, cela au fil du temps
- *Différence de la moyenne des moindres carrés* : cette mesure établit la distance entre les courbes, en attribuant un poids quadratiquement plus important aux distances les plus grandes
- *Coefficient de corrélation* : cette mesure exprime la distance entre les contours en calculant le coût de la transformation d'un contour en un autre au moyen d'une transformation linéaire. Plus ce coût élevé, plus les contours sont originellement différents.

De ses investigations, Hermes (1998b) estime que les deux meilleurs candidats pour refléter les résultats des évaluations perceptives sont la différence de la moyenne des moindres carrés et le coefficient de corrélation. Cette dernière mesure a également fait l'objet d'une évaluation positive par Rilliard *et al.* (2011), de manière concomitante avec l'utilisation d'une opération d'alignement et transformation des contours, le *Dynamic Time Warping* (ci-après, DTW).

En effet, Hermes avait pu se passer d'une étape de transformation des contours, dans la mesure où les formes qu'il comparait avaient la même longueur : à partir d'un enregistrement de référence dont était extraite la *f0*, Hermes (1998a) produisait les items à comparer par resynthèse. Il disposait donc d'une *f0* brute, et de formes stylisées de cette courbe brute. Ainsi, bien que la configuration du contour soit de plus en plus lisse à mesure

des resynthèses, la longueur de la forme était toujours identique. Comme les mesures proposées par Hermes calculent des différences point par point, elles n'étaient pas applicables en l'état par Rilliard et collègues (2011).

A propos du DTW, il peut être utilisé comme Rilliard et al. (*Ibid.*) en tant que méthode d'interpolation non-linéaire, mais il peut également donner, de la même manière que le coefficient de corrélation calculé par Hermes, une indication de la distance entre les contours. En effet, le DTW sert à aligner et normaliser des formes. Par exemple, si deux contours doivent être comparés, le DTW alignera dans un premier temps les éléments qui se ressemblent (les pics avec les pics, les creux avec les creux, les blancs avec les blancs). Ainsi, si les deux formes à comparer ont une configuration identique, tous les événements remarquables par l'algorithme devraient se trouver alignés paire à paire ; suite à quoi des points sont ajoutés pour normaliser la longueur des formes. Dans le cas où les formes comparées ont des configurations différentes, l'alignement devrait être plus coûteux. Ce coût d'alignement des formes, peut donner un indice de la similarité entre elles. C'est la méthode que choisit Kim (2011, 2012) (en l'appliquant aux coefficient MFCC, et non à des contours de *f0*) comme mesure automatique de la convergence phonétique. D'après ses travaux, les résultats obtenus au moyen de cette méthode étaient fidèles aux résultats des tests perceptifs menés par ailleurs.

Finalement, il nous faut noter que ces mesures sont holistiques et qu'elles présentent donc une limitation idoine aux évaluations perceptives : le score obtenu étant global, elles constituent alors essentiellement un moyen supplémentaire de classifier la réussite de l'imitation (au niveau prosodique) en reflétant théoriquement la perception des sujets. En d'autres termes, comme les évaluations perceptives, ces mesures automatiques de la similarité prosodique ne répondent qu'à la question « *Ce contour est-il une imitation réussie* ? » et non à la question « *En quoi ce contour est-il une bonne imitation* ? ». Par ailleurs, il nous faut noter que les résultats obtenus sur des courbes brutes de *f0* tendent à gommer l'aspect temporel de la comparaison en raison de la déformation subie par les formes au moment de l'alignement/interpolation. C'est pourquoi nous aimerions envisager une mesure permettant d'obtenir une strate supplémentaire de comparaison, *i.e.* un niveau phonologique, tout en conservant les aspects temporels des formes comparées.

3.3 Propriétés souhaitées d'une fonction de distance et *Turning Function (T-Function)*

Dans les domaines mathématiques, de nombreux problèmes d'une branche de la discipline ont été résolus par le biais des techniques et méthodes d'autres domaines mathématiques (Singh, 2010). En ce qui concerne l'analyse phonologique (et linguistique), l'interaction avec les mathématiques est souhaitée, sinon souhaitable (Bergounioux et al., 2007). Une telle remarque laisse s'ouvrir des perspectives originales si nous en venons à considérer les contours intonatifs comme des formes géométriques : quoi de plus naturel qu'avoir recours aux techniques d'un domaine spécialisé dans l'étude des formes pour approcher la résolution de notre problème sur la comparaison des formes ?

Veltkamp (2001) note qu'un ensemble de points peut être considéré comme une forme ; or, annoter les tons d'un contour intonatif revient à déterminer un ensemble de points dans un repère orthogonal, paramétrant la forme du contour en fonction du temps (en abscisses) et de la hauteur (en ordonnées). Nous renvoyons à la sous-partie suivante la manière dont nous pouvons annoter les contours pour nous concentrer sur une méthode permettant de comparer deux ensembles de points.

Arkin *et al.* (1991, p. 209) listent un ensemble de propriétés que devrait posséder une fonction exprimant la dissimilarité entre deux formes A et B, soit : $d(A, B)$. Nous reprenons leur liste et la commentons :

- $d(A, B)$ doit être une métrique, c'est-à-dire :
    - $d(A, B) \geq 0$ pour tout A et B.
        - Une absence de distance équivaut à un score de 0 donné par la mesure, tout autre score doit être positif
    - $d(A, B) = 0$ si et seulement si $A = B$
        - Comparer deux formes et obtenir une distance nulle signifie que ce sont les deux mêmes formes
    - $d(A, B) = d(B, A)$ pour tout A et B
        - L'ordre de comparaison ne doit pas avoir d'importance
    - $d(A, B) + d(B, C) \geq d(A, C)$
        - On retrouve ici l'inégalité triangulaire. Arkin et collègues estiment cette propriété importante dans la reconnaissance de forme, pour

éviter des cas où $d(A,B)$ et $d(B,C)$ seraient très petits tandis que $d(A,C)$ serait très grande.

- $d(A,B)$ doit être invariant :
    - à la translation
        - Une telle propriété est intéressante pour l'étude des contours intonatifs, si les patrons à comparer ne sont pas centrés sur le même point dans l'espace de représentation
    - à la rotation
        - Etant donné que la *f0* est paramétrée par le temps et la hauteur, l'orientation de sa forme étant toujours la même, il est peu probable que cette propriété nous soit utile
    - aux changements d'échelle
        - Cette propriété nous semble fondamentale dans la mesure où à formes équivalentes mais à échelles différentes, nous obtiendrons un bon score de similarité
- $d(A,B)$ doit correspondre à notre intuition concernant la ressemblance des formes A et B

Comme le remarquent Arkin *et al.*(1991), cette dernière propriété de la fonction de distance est fondamentale puisque le résultat donné par la fonction doit correspondre à la dissimilarité qu'un œil (ou une oreille…) humain donnerait lorsqu'il estime la dissimilarité. Il s'agit pour nous d'un enjeu majeur car nous nous posons la question de savoir si la ou les mesures que nous allons proposer pour estimer la dissimilarité prosodique ont une réalité perceptive ! Cette dernière considération est d'ailleurs équivalente à la remarque d'Hermes (1998b, p74) que nous rappelions quelques pages plus haut.

De plus, le respect de ces propriétés (notamment, l'invariance aux translations, rotations et changements d'échelles) lorsque nous déterminons la distance entre deux patrons intonatifs stylisés permettrait de dépasser la critique émise à l'égard des transformations de *f0* brute qui oblitèrent les aspects temporels de la *f0* afin de rendre les contours comparables.

Dans leur article, Arkin *et al.* (1991) définissent par ailleurs une fonction de distance appelée Turning Function (ci-après, *T-Function*) et démontrent que cette dernière respecte bien les propriétés que nous venons de rappeler. La *T-Function* s'applique à des représentations de formes polygonales dans un plan ; elle semble donc appropriée pour

comparer des patrons intonatifs stylisés à partir d'un ensemble de points entre lesquels des segments de droite sont tirés.

La *T-Function* est ainsi décrite (Veltkamp, 2001, p. 10) :

> The cumulative angle function, or turning function, $\Theta A(s)$ of a polygon or polyline A gives the angle between the counterclockwise tangent and the x-axis as a function of the arc length s. $\Theta A(s)$ keeps track of the turning that takes place, increasing with left hand turns, and decreasing with right hand turns.

En d'autres termes, la *T-Function* opère une transformation des formes qu'elle parcourt. Afin de représenter concrètement ces transformations, nous renvoyons à la Figure 22.

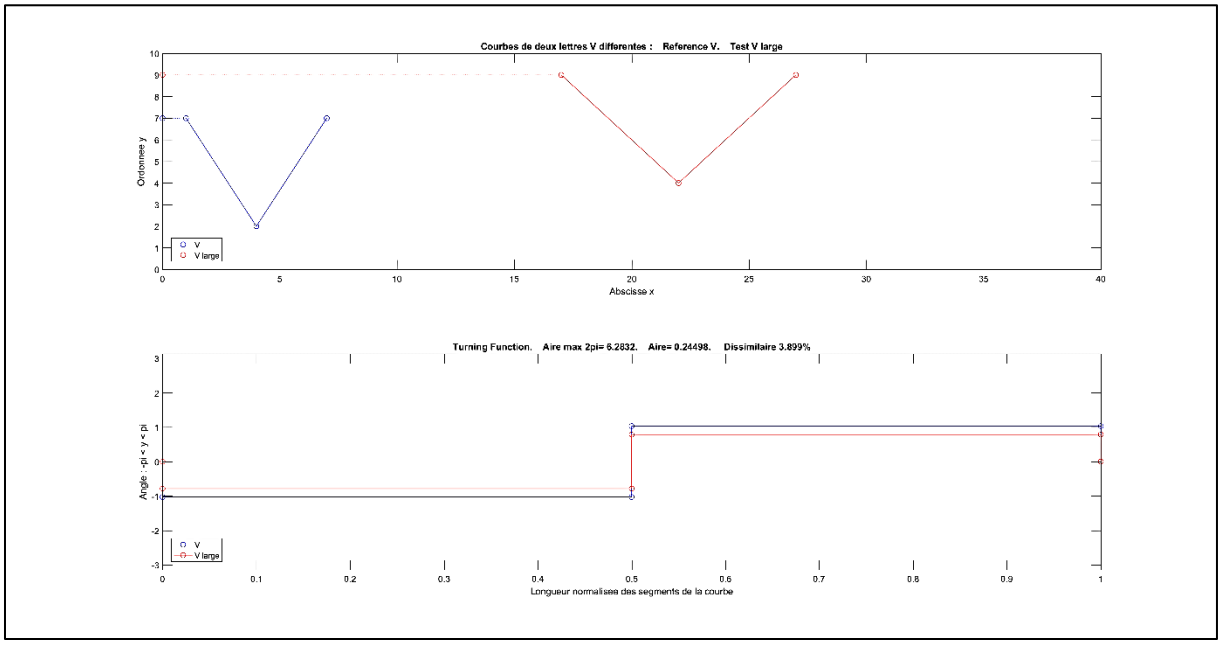


Figure 22 : T-Function de deux formes simples (V_{bleu} et V_{rouge}).

Dans le premier espace de la Figure 22 sont représentées les formes sur lesquelles la transformation proposée par la *T-Function* va être opérée ; celles-ci sont dessinées par des traits pleins. Les lignes pointillées constituent des segments imaginaires permettant de prendre en compte le premier angle, ici descendant, de la figure. L'espace en dessous correspond à l'espace de transformation où est représenté l'output de la *T-Function*. Dans ce dernier espace, les segments verticaux et leur longueur représentent la force des angles décrits par la figure transformée, et les segments horizontaux représentent alors les segments de la polyligne en tenant également compte de leur longueur. Quelques commentaires supplémentaires sur ce type d'image que nous allons retrouver plusieurs fois en suivant :

- L'espace de transformation est gradué de 0 à 1 en abscisses. La *T-Function* normalise la longueur totale des segments d'une forme. Ce faisant, la proportion des figures n'est pas altérée lors de la transformation
- En ordonnées, ce même espace est gradué de $[-\pi; +\pi]$.
- Etant donné les informations précédentes, l'aire de l'espace de transformation de la *T-Function* correspond à 2π (sa hauteur multipliée par sa longueur).
- La dissimilarité entre les figures (ici, entre V_{bleu} et V_{rouge}), calculée dans l'espace de transformation, est donnée par l'aire comprise entre la transformation de la forme bleue et la transformation de la forme rouge.
- Nous retrouvons ces données résumées au-dessus de l'espace de transformation :
    - Aire maximum : $2\pi \approx 6{,}382$
    - Aire (comprise entre les transformations) $= 0{,}24498$
    - Dissimilaire à $3{,}899\% \approx \frac{0{,}24498}{6{,}382} \times 100$; ainsi, ce pourcentage est un rapport entre l'aire maximale de l'espace et l'aire calculée entre les courbes.

Par conséquent, plus l'aire est proche de 0, plus la similarité entre les figures évaluée est grande, ici, V_{bleu} est très similaire à V_{rouge}.

Afin de poursuivre l'illustration de cette fonction, observons la Figure 23.

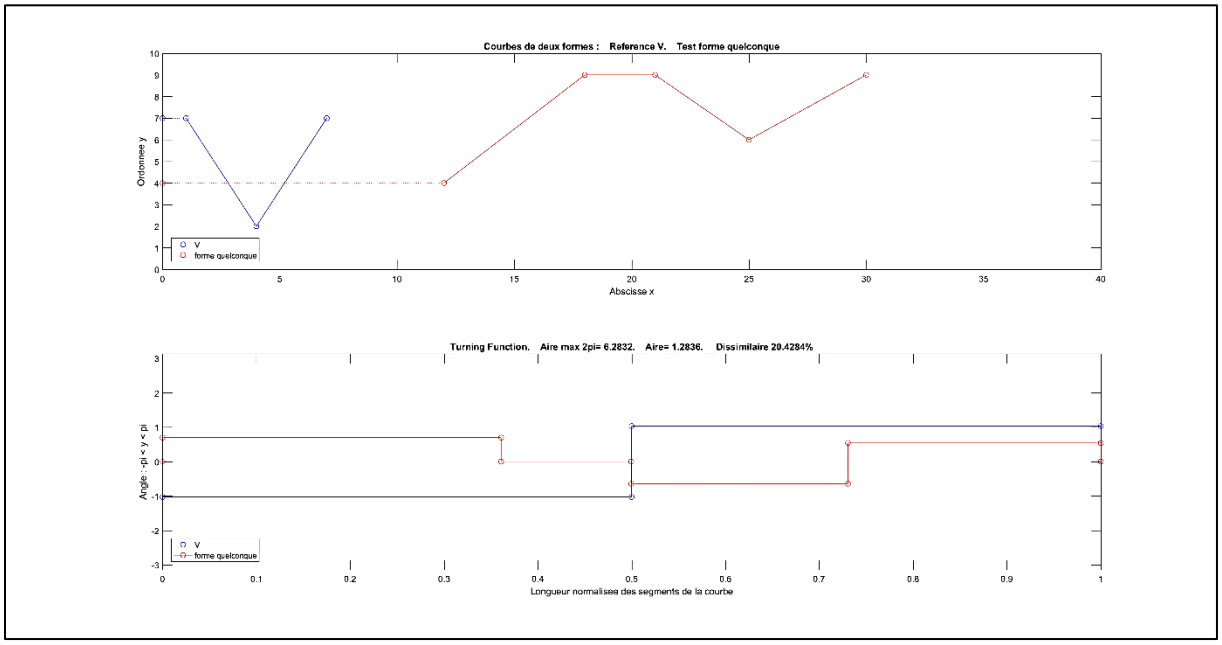


Figure 23 : *T-Function* de deux formes très différentes (V_{bleu} et X_{rouge},)

Quelques remarques sur la Figure 23 :

- Nous pouvons observer, malgré la longueur très différente de V_{bleu} et X_{rouge} que les distances sont bien normalisées dans l'espace de transformation.
- Les résultats donnés pour l'aire et le pourcentage de dissimilarité sont bien plus élevés que précédemment. A l'œil, les deux formes non transformées sont en effet très différentes !

En raison de ses propriétés, cette transformation semble donc appropriée pour pouvoir ensuite calculer la dissimilarité entre deux formes polygonales. Nous expliciterons en suivant différentes manières d'annoter la *f0* dans le but d'obtenir les points nécessaire à l'émergence de formes comparables à l'aide de la *T-Function*.

3.4 Stylisations et annotations prosodiques : émergence de formes

La fonction que nous avons choisie pour opérer nos comparaisons nécessite une stylisation de la *f0* à base de lignes droites. Cependant, il nous semble important d'aborder brièvement le problème évoqué par t'Hart (1991, p. 3368) entre stylisation en lignes droites et en courbes paraboliques. En effet, les questions que nous nous posons ne sont pas uniquement mathématiques ou géométriques : il faut tout de même que les stylisations choisies aient une validité perceptive. Or, la *f0* sous sa forme brute est essentiellement courbe plutôt que droite. Ainsi, une critique qui pourrait être portée à notre encontre serait de dire que la stylisation droite est une simplification abusive de la forme originale.

Des algorithmes comme MOMEL (Hirst & Espesser, 1993) permettent ainsi de modéliser la courbe de *f0* au moyen de fonctions *spline*, donnant alors une transformation proche de la *f0* originale, tout en gommant les variations micro-prosodiques dues à l'articulation de phonèmes. Une telle modélisation serait ainsi tellement proche de la *f0* originale, que distinguer perceptivement une *f0* brute d'une *f0* momelisée serait quasi impossible. Cela dit, d'après t'Hart (1991, p. 3370), distinguer une *f0* à stylisation parabolique (de type MOMEL, donc) d'une *f0* à stylisation droite ne serait pas non plus une tâche aisée. Cependant, il convient d'apporter une légère nuance à notre propos : t'Hart indique que les stylisations rectilignes devraient comporter un plateau de 30 à 40 ms, les versions avec des pics ayant été plus facilement distinguées. Cette dernière remarque légitime notre choix de travailler à partir d'une stylisation droite, malgré les réserves admises par t'Hart : la *T-*

function étant assez sensible au bruit, nous préférerons garder des stylisations droites et sans plateaux qui ajoutent du bruit dans la transformation.

Afin d'obtenir ce type de stylisation, il faut donc déterminer des points clefs du contour intonatif. Pour ce faire, plusieurs alternatives s'offrent à nous en fonction des règles d'annotation que nous choisissons.

- A) Une annotation tirée de l'analyse phonologique (inspiré de la théorie métrique auto-segmentale, Jun & Fougeron, 2000; et plus particulièrement Welby, 2006).

Pour ce type d'annotation, nous repérerons les événements les plus saillants du contour intonatif, et procéderons à une annotation des tons hauts et tons bas du contour. Nous retiendrons par ailleurs l'idée défendue par Di Cristo (1999) d'une bipolarisation accentuelle des constituants prosodique (soit : l'idée qu'un constituant est marqué par un accent initial et un accent final). Voici les règles que nous retiendrons :

  - Les cibles tonales annotées sont les points d'inflexion de la courbe indépendamment des perturbations microprosodiques dues à l'articulation de la parole
  - Les cibles tonales sont donc placées sur des voyelles, où la *f0* est plus stable
  - En accord avec la méthode d'analyse auto segmentale, le texte sert de référence pour l'alignement des cibles tonales
  - Considérant les tons en tant que catégories phonologiques, il convient de proposer un inventaire de ces dernières
    - L1 : première cible basse
    - H1 : première cible haute, proéminence marquant l'accent initial
    - L2 : cible basse avant l'accent final haut (H2)
    - H2 : dernière cible haute du constituant, accent primaire du français
    - X2 : cible basse, en fin de constituant, au fort poids métrique (Astésano & Bertrand, 2016; Welby, Bertrand, Portes, & Astésano, 2016)
  - Nous notons par ailleurs des points d'inflexion qui ne sont pas des cibles tonales, mais qui servent à rendre compte des phénomènes de plateaux. Ils sont notés avec la lettre I (pour inflexion), et un diacritique (+, - ou =) indiquant la configuration de la courbe après le point I. Par exemple :
    - I+ : succède à un plateau bas
    - I- : succède à un plateau haut

- I = : précède un plateau

Ainsi, ce type d'annotation vise à faire ressortir le pattern du contour, et les contrastes de *f0* entre deux cibles tonales (si certaines syllabes ne contiennent pas de cible) seront linéarisés. Nous espérons également que ce type d'annotation, outre son caractère phonologique, nous permettra d'obtenir des informations sur les différences phonétiques de patterns identiques. Une question que cette approche vise à traiter serait en effet de savoir si deux patterns ayant la même succession de ton que leur modèle, peuvent être systématiquement distingués.

- B) Une annotation syllabique en mettant un point au milieu de chaque voyelle considérant alors chaque voyelle comme le lieu d'une pulsation rythmique

Plutôt que de prendre pour base une analyse phonologique, il nous semble également intéressant de mettre en place un système d'annotation plus proche des segments syllabiques. En effet, bon nombre des productions que nous souhaitons comparer sont supposées contenir le même nombre de syllabes :

- Notre corpus d'imitations (chapitre 5) consiste en des réitérations d'énoncés entendus.
- Notre corpus de logatomes (chapitre 6) consiste en des réitérations délexicalisées d'énoncés entendus.

D'une part, nous pensons qu'une annotation syllabique permettra d'avoir une vue plus précise des imitations en ce qui concerne le rythme. D'autre part, nous estimons que le comptage des syllabes a une importance dans la production des logatomes : il est probable qu'une même forme puisse émerger d'énoncés avec un nombre différent de syllabe. Par ce type d'annotation, nous souhaitons donc ajouter une dimension supplémentaire à l'évaluation des logatomes. Ces annotations ont été produites de manière semi-automatique, au moyen d'un script (cf. Annexes) pour le logiciel PRAAT (Boersma, 2001).

- C) Une annotation automatique, en nous servant des points déterminés par un algorithme comme MOMEL (Hirst & Espesser, 1993), à partir desquels est ensuite estimée la *spline* reliant chaque point.

Obtenir automatiquement les points permettant de calculer par la suite des *T-Functions* semble une approche séduisante. Cependant, le manque de contrôle sur le résultat de l'algorithme pourrait également conduire à un risque d'erreur accru dans la détermination de

la forme du contour intonatif. Comme l'indiquent Campione & Veronis (2000, p.42), MOMEL produit certains types d'erreur de manière récurrente en plaçant certains points de manière erronée (environ 4%), voire en omettant des points à proximité des silences (environ 6%). Ces taux d'erreur semblent acceptables si des corrections manuelles sont systématiquement appliquées. Cela dit, dans notre travail, le risque d'erreur lié à l'utilisation de MOMEL pourrait être accru :

- Les énoncés évalués sont généralement courts, et par conséquent, très souvent encadrés par du silence, cas problématique pour l'algorithme
- Pour obtenir une bonne *T-Function*, il faudrait n'avoir d'erreur ni sur l'énoncé testé, ni sur l'énoncé de référence : si une des deux formes est erronée, la *T-Function* en résultant sera biaisée.

Malgré ces inconvénients, l'utilisation de MOMEL pour la détection de la forme des patrons intonatifs nous semble une piste prometteuse, vue la potentialité qu'offre l'automatisation d'une telle tâche en termes d'applications logicielles ou de recherche.

Dans ce travail, nous exploiterons chacune de ces approches et nous chercherons alors à estimer leurs avantages et inconvénients vis-à-vis des résultats donnés par les *T-Functions*.

4 Synthèse : du recueil à l'analyse de l'imitation parolière

Cette revue de littérature sur la méthodologie de l'étude de l'imitation parolière nous a dans un premier temps permis de rendre compte de l'importance d'adapter le recueil de données au type de comportement observé. Dans les préliminaires de ce travail, nous avons accordé une importance particulière à la définition des imitations ; ces premiers pas se trouvent justifiés dans ce contexte. Afin de proposer un protocole expérimental cohérent, une définition rigoureuse du comportement observé est nécessaire. Ainsi, nous avons en premier lieu décrit les tâches utilisées dans la littérature pour provoquer chez des sujets (naïfs ou experts) la production de parole à caractère imitatif. La Table 4 rappelle ces tâches en les classant en fonction du comportement visé par l'expérimentateur.

Conversationnel	**Imitation de laboratoire**	**Impersonnation**
Map Task	Shadowing	Imitation de la voix (Texte fixe)
Diapix	Imitation de sons inconnus	Imitation de la voix (pas de texte fixe)
Rorshach matching task	Gradation d'imitations	
Columbia Game Corpus		

Tableau 4 : Synthèse des types de tâches proposées dans les études ayant trait à l'imitation, classée en fonction du type de tâche.

Dans un second temps, nous nous sommes intéressés à la problématique de l'évaluation/mesure de l'imitation en parole. Nous avons alors exprimé les difficultés relevées dans la littérature à ce sujet. D'une part, les mesures acoustiques classiques semblent inconstantes car elles ne révèlent pas de *pattern* systématique permettant de jauger le degré d'imitation contenu dans une production. D'autre part, les évaluations perceptives semblent avoir une relative fiabilité en présentant l'inconvénient d'avoir un pouvoir explicatif quasi nul en ce qui concerne les raisons phonétiques d'une imitation parolière réussie. Ceci dit, nous avons estimé que les deux pans (acoustique et perceptif) de l'évaluation/mesure traditionnelle des imitations parolières sont complémentaires. La partie gauche de la Table 5 (description

acoustique et évaluation perceptive) présente des approches *a priori* adaptées à l'étude de l'imitation prosodique.

<table>
<tr><th colspan="2" rowspan="2">Description acoustique</th><th rowspan="2">Evaluation perceptive</th><th colspan="2">Mesures de distance</th></tr>
<tr><th>Mesure Holistique</th><th>Mesure non Holistique</th></tr>
<tr><td>Etendue de f0</td><td>Débit de Parole</td><td>Test AX</td><td>Facteur d (DTW)</td><td>T-Function (annotations manuelles)</td></tr>
<tr><td>Moyenne de f0</td><td>Temps de Phonation</td><td rowspan="2">Test AXB</td><td>Différence de la moyenne des moindres carrés</td><td rowspan="2">T-Function (annotation Momel)</td></tr>
<tr><td>Points de mesures locales de la f0</td><td>Présence et durée des pauses</td><td>Coefficient de corrélation</td></tr>
</table>

Tableau 5 : Synthèse des approches possible pour l'évaluation/mesurage de l'imitation prosodique

Par la suite, nous nous sommes concentrés sur l'étude de la fréquence fondamentale et nous avons décrit des approches permettant de mesurer la distance entre des contours intonatifs (Table 5, partie à droite). Or, la littérature présente peu d'études adoptant cette approche pour l'étude de l'imitation parolière : celles-ci se limitent souvent aux tests perceptifs et/ou à la description acoustique de paramètres relativement isolés. Par ailleurs, contrairement à d'autres paramètres acoustiques très locaux, le contour intonatif englobe le contenu parolier et en révèle la structure. Ainsi, nous pensons qu'envisager l'étude de l'imitation prosodique en proposant de considérer la forme de la fréquence fondamentale pourrait permettre de mesurer efficacement la part prosodique du degré d'imitation d'un énoncé. En effet, nous pensons que le contour intonatif est un élément assez aisément perceptible de la production parolière et qu'il pourrait peut-être servir de base à l'auditeur humain dans son jugement de similarité ; dans une perspective *gestaltiste*, ce dernier serait d'abord le fruit d'une perception globale des formes.

Enfin, rappelons que la MVT préconise l'usage systématique de modulations prosodiques pour produire des énoncés correctifs. Pouvoir évaluer la similarité entre deux formes prosodiques pourrait permettre de proposer des outils d'entraînement à la production de ces modulations.

Chapitre 5 : Exploration des liens entre jugements perceptifs et mesures de la similarité prosodique

Dans ce chapitre, nous nous proposons de passer un même corpus de parole imitative au crible de jugements de la similarité prosodique issus de sources différentes. L'objectif premier de cette démarche expérimentale sera de pouvoir à terme contourner les difficultés de mesure de l'imitation à un niveau prosodique rencontrées jusqu'à présent. Pour cela, il nous est nécessaire de mesurer par de multiples méthodes la similarité prosodique d'énoncés et de leur modèle, puis de comparer les *outputs* obtenus par ces méthodes diverses aux résultats de tests perceptifs de jugement de la similarité.

En effet, nous intéressant à l'évaluation de l'imitation, il importe particulièrement qu'une mesure objective de la similarité prosodique reflète de manière satisfaisante la perception d'auditeurs humains. En ce sens, les mesures objectives de la similarité prosodiques doivent répondre à la même exigence que la propriété d'intuition de l'humain émise par Arkin *et al.* (1991, p. 209) concernant les fonctions de distance des formes géométriques :

- Si l'œil humain estime qu'il y a une haute similarité entre deux formes, le résultat de la mesure de distance doit indiquer la même tendance.

Ce faisant, si un type de mesure objective est corrélé positivement et de manière systématique aux résultats perceptifs, cela pourrait alors vouloir dire que les attributs acoustiques paramétrant ce type de mesure ont une importance particulière dans le mécanisme perceptif procédant au jugement de la similarité de stimuli acoustiques (Hermes, 1998b, p. 74). Ce type de mesure pourrait alors être utilisé pour estimer la similarité prosodique d'un énoncé et de son modèle avec une fiabilité suffisante pour, à terme, se passer des jugements perceptifs. Dans cet ordre d'idée, disposer d'une mesure répondant à ces critères laisse entrevoir la possibilité d'applications utiles à la fois dans le champ de la correction phonétique et dans la recherche en phonétique.

Cette partie de notre travail résume donc l'approche expérimentale (*cf.* Figure 24) mise en œuvre pour l'évaluation systématique de la similarité des formes sonores. Nous relaterons

dans un premier temps la constitution de notre premier corpus d'imitations parolières ainsi que les premiers résultats que nous en avions tiré quant à la production de parole imitative.

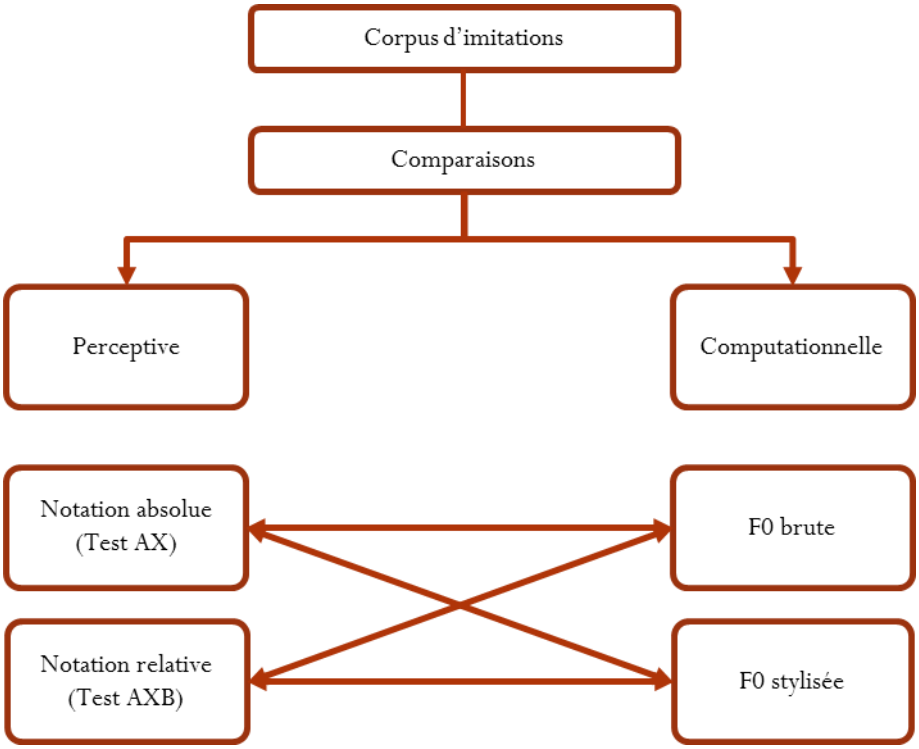


Figure 24 : Vue globale du protocole expérimental de comparaison des formes sonores et de leur représentation graphique.Après recueil d'un corpus d'imitation parolière, les imitations et leur modèle sont systématiquement comparés par le biais de plusieurs méthodes illustrées par l'arborescence. La relation (indiquée par les connecteurs) entre ces différents types de résultats est par la suite évaluée.

Après avoir rappelé les raisons pour lesquelles nous n'avons pas adopté une approche classique de mesurage acoustique, nous présenterons les différentes méthodes d'évaluation de la similarité prosodique, conjuguant de manière parallèle perception et mesure objective, qui ont été utilisées dans ce protocole. Les résultats obtenus par ces méthodes seront finalement mis en relation les uns avec les autres afin qu'ils puissent être évalués, puis discutés.

1. Constitution du matériel expérimental : du corpus d'Edimbourg au Corpus Imitation

Afin de pouvoir évaluer la similarité prosodique entre des imitations et leur modèle, il faut disposer d'un corpus de parole imitée. Lors d'un précédent travail (Nocaudie, 2012; Nocaudie & Astésano, 2012), nous avions justement recueilli en laboratoire un corpus de parole contrôlée. Originellement, ce corpus était destiné à observer les stratégies d'imitation au niveau prosodique de locuteurs francophones natifs. Pour ce faire, nous nous étions appuyés sur un corpus source, dit *Corpus d'Edimbourg* (ci-après *CE*) (Astésano, Bard, & Turk, 2007; Astésano, Bertrand, Espesser, & Nguyen, 2012), dont les enregistrements ont servi de stimuli pour le recueil du Corpus Imitation (ci-après *CI*). Nous présenterons brièvement le *CE*, avant de récapituler les conditions de recueil et les tâches expérimentales du *CI*. Conséquemment, nous émettrons quelques remarques au sujet de nos premiers traitements du *CI*, avant d'expliciter le raisonnement qui nous a conduits à considérer la notion de forme.

1.1 Le Corpus d'Edimbourg

Le CE est un corpus de parole contrôlée dont les énoncés présentent une ambiguïté syntaxique. L'ambiguïté réside dans la portée de l'adjectif dans des syntagmes nominaux coordonnés du type :

- **Les termitières et les fourmilières napolitaines**, voyez-vous, sont assez peu répandues.

A la lecture d'une telle phrase, il est permis de se demander si l'adjectif « napolitaines » qualifie seulement « les fourmilières » ou bien s'il qualifie « les termitières et les fourmilières ». A l'oral, cette ambiguïté peut être levée à l'aide des indices prosodiques (accent final, accent initial, frontières). Ainsi, en fonction de la manière de prononcer ce groupe nominal, il est possible de comprendre deux sens différents :

- *Cas 1* : [Les tomates] [**et les oignons rouges**]
- *Cas 2* : [**Les tomates et les oignons**] [**rouges**]

Nous avons sélectionné des syntagmes nominaux du *CE* lus par une locutrice. Celle-ci réalise en *Cas 1* une frontière après le premier nom (*N1*), marquée par une pause silencieuse et un

allongement sur la dernière syllabe du *N1*, ainsi qu'une reprise sur le *N2* avec un débit plus rapide. En *Cas 2*, aucune pause n'est réalisée après *N1* et l'allongement est moins marqué à la frontière (au niveau de *N2*), le débit reste plus constant.

Dans le CE, la taille des items lexicaux (noms et adjectifs) varie par ailleurs de une à quatre syllabes, ce qui implique également une variété dans la structure prosodique des énoncés permettant alors d'observer les indices prosodiques nécessaires à la linéarisation des énoncés. Pour le recueil du *CI*, nous avons choisi uniquement des noms et adjectifs de trois à quatre syllabes car ils favorisent l'émergence d'événements prosodiques plus variés.

1.2 Recueil du Corpus Imitation : reproductions de la structure prosodique perçue

Dans la mesure où le *CE* est constitué uniquement d'énoncés ambigus, les locuteurs du *CI* seraient alors contraints de se reposer sur des indices d'ordre phonétique ou prosodique pour désambiguïser ce qu'il entendent. Afin de nous indiquer ce qu'ils avaient compris, les locuteurs devaient par la suite reproduire ces structures perçues. Nous détaillons ci-après le protocole de recueil du *CI*.

1.2.1 Matériel linguistique

Un sous ensemble de 16 énoncés du *CE*[61] (uniquement les syntagmes nominaux) ont été sélectionnés comme stimuli pour les tâches de recueil du *CI*. Il s'agissait uniquement de noms de trois à quatre syllabes combinés avec des adjectifs de une à quatre syllabes, cela dans les deux cas syntaxiques évoqués précédemment.

Nous avons de plus manipulé la présence/absence de pause dans les énoncés pour créer des conditions syntaxique supplémentaires pour obtenir les cas suivants :

- *Cas 1* : [Les tomates] [+ Pause] [**et les oignons rouges**]
- *Cas 1 bis* : [Les tomates] [-Pause] [**et les oignons rouges**]
- *Cas 2* : [**Les tomates et les oignons**] [-Pause] [**rouges**]
- *Cas 2 bis* : [**Les tomates et les oignons**] [+Pause[62]] [**rouges**]

[61] Voir Annexe XX pour la liste complète des énoncés sélectionnés
[62] Pause ajoutée de 300 ms, moyenne des pauses observées en Cas 1.

L'ajout ou la suppression des pauses dans les *Cas 1 & 2 bis* avait pour but de créer un conflit dans les indices prosodiques perçus par les locuteurs. Nous pensions notamment que la pause supprimée serait réintroduite par les imitateurs de manière irrépressible, tandis que la pause ajoutée serait ignorée.

Durant chaque bloc expérimental, chaque locuteur du *CI* a entendu un ensemble de 8 phrases différentes (noms de trois à quatre syllabes associés aux adjectifs de une à quatre syllabes) déclinées dans ces 4 conditions syntaxiques (soit : 8 * 4 stimuli différents). Tous les stimuli ont été entendus et répétés 3 fois dans chaque bloc expérimental pour aboutir à un total de 96 *trials* (32 stimuli * 3 répétitions) par bloc.

Par ailleurs, les stimuli étaient présentés à l'écoute de manière pseudo aléatoire avec les restrictions suivante :

- Une même condition syntaxique ne pouvait être entendue plus de 3 fois d'affilée
- Le même contenu ne pouvait être entendu plus de de fois de suite.

1.2.2 Population expérimentale et conditions d'enregistrement

8 locuteurs francophones natifs (âge 20-26 ; 5 femmes et 3 hommes) ont participé à l'enregistrement du *CI*. Au moment de l'enregistrement, ils ont rapporté n'avoir pas ou plus de pratique musicale et théâtrale, ni de troubles de l'audition. Ces locuteurs ont été recrutés parmi les étudiants de l'université Toulouse 2 Jean Jaurès et ils ne devaient faire d'étude ni en spécialité langues étrangères, ni en spécialité des sciences du langage.

La passation de l'expérience durait 45 minutes en moyenne, selon la rapidité du sujet. Après avoir accueilli le sujet, nous l'installions dans le studio d'enregistrement PETRA (goo.gl/D9d2ZC). Il était assis devant un écran d'ordinateur, un micro équipé d'un filtre anti-pop à hauteur de bouche. Les stimuli sonores étaient compilés dans un diaporama, et le sujet devait simplement appuyer sur un bouton pour écouter un nouveau stimulus avant d'effectuer la tâche qui lui était demandée.

Suite au recueil du *CI*, nous n'avons conservé les enregistrements que de 6 locuteurs pour des raisons liées à la qualité vocale (1 locutrice, enrouée le jour du recueil) ou à un stress fort induit par la situation expérimentale (1 locuteur).

1.2.3 Tâches expérimentales du *CI* : une gradation imitative

Le but initial du *CI* était de pouvoir observer les stratégies d'imitation prosodique de locuteurs dans un milieu contrôlé. Dans les protocoles pour la convergence phonétique, l'expérimentateur propose habituellement un pré-test au sujet, afin de recueillir une *baseline* exempte de l'influence phonétique d'un autre locuteur (Nathalie Lewandowski, 2012; Pardo, 2006). Ici, les comportements imitatifs visés différant de la convergence phonétique, un tel pré-test n'a pas été mené.

En effet, nous nous sommes placés dans un paradigme de gradation d'imitations en fonction de l'intention du sujet en proposant trois blocs consécutifs : répétition, imitation et exagération. Il était donc attendu que la consigne des blocs ait un effet sur la production des sujets. Les consignes des trois tâches étaient les suivantes :

- *Tâche 1* (répétition : *REP*) : Vous allez entendre des phrases : vous devrez dire ces phrases en respectant la structure entendue.
- *Tâche 2* (imitation : *IMI*) : Vous allez entendre des phrases : imitez la structure entendue.
- *Tâche 3* (exagération : *EXA*) : Vous allez entendre des phrases : vous imiterez jusqu'à exagération la structure entendue.

Notons également que ces trois tâches ont toujours été passées dans l'ordre Répétition, Imitation, Exagération, dans une gradation du comportement imitatif.

Les stimuli issus du *CE* se prêtaient particulièrement bien à cette approche, car l'ambiguïté syntaxique à résoudre servait dans la première tâche à camoufler la notion d'imitation derrière la tâche de décision syntaxique. Dans un premier temps, il était ainsi attendu des sujets qu'ils produisent une imitation prosodique minimale. Par la suite, les consignes étant explicites, la fidélité prosodique attendue était plus forte.

L'autre originalité de ce protocole de recueil réside dans la troisième tâche proposée : l'exagération. Les deux premières tâches de ce protocole sont assez classiques puisqu'on y oppose imitation implicite (*REP*) et explicite (*IMI*). Ce type de tâches peut être retrouvé dans d'autres protocoles (voir par exemple : Michelas & Nguyen, 2011). En proposant cette troisième tâche (*EXA*), nous espérions observer des stratégies nouvelles chez les locuteurs, qui en essayant de forcer le trait devaient tendre vers des comportements propres aux imitateurs professionnels en essayant de changer leurs habitudes vocales (Mejvaldova, 2002; Révis et al., 2013).

1.3 Quelques notes sur le Corpus Imitation : des essais de mesurage acoustiques à la comparaison des formes sonores

Avant d'avoir divergé vers la recherche d'une mesure de la similarité prosodique, nos travaux précédents se concentraient sur quelques aspects acoustiques des productions des locuteurs du *CI*. Bien que cette approche de mesurage acoustique ait fini par nous sembler particulièrement insatisfaisante pour les raisons que nous avons évoquées au chapitre , il paraît tout de même opportun d'émettre quelques remarques empiriques et statistiques quant au contenu du *CI*.

1.3.1 Evolution des productions au fil des tâches

Afin de nous rendre compte de la manière dont les consignes des blocs ont été perçues par les sujets, nous avons procédé à une première écoute du corpus. En premier lieu, il nous semble important de noter que les sujets ont paru réussir à distinguer les structures du *Cas 1* et du *Cas 2* dès la tâche de simple répétition, les indices prosodiques majeurs des différents cas ayant été apparemment bien reproduits :

- Montée intonative et pause du *Cas 1*
- Frontière de mot ou d'Accentual Phrase entre *N1* et *N2* du *Cas 2*. (Garnier, Baqué, Dagnac, & Astésano, 2016)

En ce qui concerne les *Cas bis* (dont la présence/absence de pause a été manipulée), il nous faut indiquer que :

- Une pause a été majoritairement produite en *Cas 1 bis* (alors que le stimulus n'en comportait plus)
- Les pauses du *Cas 2 bis* ont peu été reproduites (sauf en *IMI* et en *EXA*)

Il apparaît enfin assez clairement que les sujets ont changé leur production au fil des tâches, en passant d'une transmission *a minima* de la structure perçue à des tentatives d'imitations plus élaborées de la structure entendue, jusqu'à altérer leurs habitudes vocales dans certains cas. Un locuteur a notamment pris un registre plus haut, pour reproduire plus fidèlement le registre de la locutrice modèle, lors de la tâche d'exagération.

Bien qu'elle révèle quelques pistes quant aux stratégies d'imitation des locuteurs, l'écoute du *CI* permet surtout de se rendre compte de la variété des stratégies des locuteurs et de la variabilité de leurs productions au cours du recueil.

1.3.2 Production/omission des pauses (P)

Nous reprenons ici l'analyse de l'occurrence et de la durée des pauses, réalisée initialement à l'occasion de nos travaux précédents (Nocaudie, 2012; Nocaudie & Astésano, 2012). Ces analyses portaient sur un sous-ensemble du *CI*, constitué de 144 productions des sujets (1 longueur de phrase * 4 conditions syntaxiques * 3 répétitions * 3 tâches * 4 sujets). Nous avions alors réalisé des ANOVAs, avec comme variables indépendantes :

- les locuteurs (le modèle : *CRI vs.* le groupe d'imitateurs : *SUJ*)
- les blocs expérimentaux (*REP*, *IMI*, *EXA*)
- les conditions syntaxiques
    - Cas 1 = *C1o*[63]
    - Cas 1 bis = *C1m*
    - Cas 2 = *C2o*
    - Cas 2 bis = *C2m*

et comme variables dépendantes :

- l'occurrence des pauses (notée 0 ou 1)
- la durée des pauses (en secondes).

Condition	**Répétition (REP)**	**Imitation (IMI)**	**Exagération (EXA)**	**Modèle (CRI)**
C1o	70 %	90 %	100 %	100 %
C1m	80 %	100 %	80 %	0 %
C2o	0 %	0 %	30 %	0 %
C2m	35 %	90 %	100 %	100 %

Tableau 6 : Taux d'occurrence des pauses en fonction de la tâche expérimentale et de la condition syntaxique. Ces taux portent sur 144 réponses des locuteurs du *CI*. La dernière colonne indique les pauses produites par le modèle.

Dans la condition *C1o* qui contient P dans le modèle, les locuteurs ont produit significativement moins de P durant la tache de *REP* (p=.0019), avant de parvenir à approcher la performance de *CRI* durant *IMI* (p>.05) et atteindre sa performance en *EXA* (p=1). Ainsi, les locuteurs omettaient P dans presque 1 cas sur 4 en répétition.

[63] o = original ; m = modifié

Le taux de production de P dans la condition *C1m* est particulièrement intéressant si on le rapproche de celui observé en *C1o*. Rappelons que le modèle entendu en *C1m* a vu sa pause effacée. Malgré cela, les locuteurs ont réinsérer de manière irrépressible une pause silencieuse entre les deux noms du syntagme nominal. Ainsi, l'analyse statistique révèle une absence de différence significative entre le taux de pause de *CRI* en *C1o* et les taux de pause des *SUJ* en *C1m*. De plus, en *IMI*, les locuteurs parviennent à produire 100 % de pauses en *C1m* ! Etonnamment, le taux de pause des *SUJ* en *C1m* retombe à 80 % pour la condition EXA. Ainsi, il semblerait que :

- la frontière prosodique de *C1m* aurait été perçue, et reproduite, malgré l'effacement de P, que les *SUJ* ne peuvent s'empêcher de réinsérer afin de marquer la structure, saillante par son AF
- En ce qui concerne *EXA*, les approches stratégiques des sujets ont pu diverger, certains ayant cherché à reproduire la structure prosodique originale (pause comprise), les autres ayant calqué leur production pour reproduire le plus fidèlement possible ce qu'ils ont entendu en ne produisant pas la pause.

A propos des taux de pauses observés en *C2o* (absentes dans le modèle entendu), aucune différence significative n'est à signaler pour les tâches *REP* et *IMI*. Cependant, nous pouvons observer une différence significative en EXA ($p < .05$) où certains sujets ont choisi de réintégrer une pause après le second nom, peut-être pour marquer la frontière prosodique et la structure syntaxique sous-jacente de manière plus prononcée.

En *C2m*, la pause incongrue intégrée au modèle original est peu reproduite en *REP* ($p < .001$), tandis que les sujets ont fidèlement imité cette pause en *IMI* et en *EXA*, accentuant le marquage de la frontière prosodique entre le second nom et l'adjectif de manière quasi-systématique.

Vus les taux de pauses produites, il semble que cet indice prosodique soit une cible aisément perceptible (et reproductible) pour les imitateurs. Par ailleurs, les résultats obtenus concernant la condition *C1m* semblent indiquer que les locuteurs natifs ont un biais en faveur de la production d'une pause en cas d'allongement en fin de mot, et de montée intonative. Ce résultat en production imitative pourrait donner une piste pour expliquer la manière dont les groupements prosodiques sont perçus : si la pause est acoustiquement absente, mais qu'elle

est reproduite, cela pourrait indiquer que AF en frontière majeure (IP) est représenté avec une pause.

1.3.4 Durée des pauses et des AF aux frontières

Pour ces analyses, nous considérions les durées des pauses, ainsi que les durées des AF à l'entourage de la position des frontières les plus fortement marquées par la locutrice de référence ; soit :

- L'AF en fin de premier nom pour les *Cas 1* et *Cas 1 bis*
    - [Les baga**telles**][et les balivernes sottes]
- L'AF en fin de second nom pour les *Cas 2* et *Cas 2 bis*
    - [Les bagatelles et les bali**vernes**] [sottes]

A propos de la durée de P dans le *Cas 1* (*C1o*), les imitateurs ont atteint une performance approchant celle du locuteur de référence dès la tâche d'imitation, et ont persisté durant la tâche d'exagération. En condition manipulée (*C1m*), il faut attendre la tâche d'exagération pour que les imitateurs produisent des pauses dont la durée approche celle de référence. En parallèle, les aspects de durée des AF sur le premier nom produit par les imitateurs ne sont pas significativement différents de la durée des AF de CRI. Ainsi, les imitateurs parviendraient à ajuster la durée de la pause, alors qu'il n'est pas nécessaire de modifier la production des AF, dont les valeurs approchent déjà celles de la référence.

Concernant les *Cas 2* (*C2o*) et *Cas 2 bis* (*C2m*), notons en premier lieu que très peu de pauses étaient produites en *C2o*, pour toutes les tâches. Il n'y avait donc aucune différence à attendre pour ce cas. En *C2m*, en revanche, les pauses produites restent significativement plus courtes en REP et en IMI. En EXA, la durée de P est donc finement ajustée. Quant aux AF, si leur durée est légèrement plus courte en REP de *C2o*, les imitateurs atteignent une performance similaire à celle de CRI dès la seconde tâche. Aucune différence de durée des AF n'avait été relevée pour la condition *C2m*.

1.3.5 Une discussion lapidaire

Ces résultats, que nous venons de survoler, montrent l'approche que nous envisagions d'adopter au début de notre travail. L'aspect essentiel que nous en retenons est que les imitateurs semblent capables de faire des ajustements assez fins de leur production lorsqu'il

leur est explicitement demandé d'imiter une structure. Pour ce faire, ces derniers se basent sur les indices prosodiques les plus perceptibles (ici, la présence de la pause, puis sa durée). Par ailleurs, lorsqu'il n'est pas nécessaire d'ajuster un paramètre (en raison d'une similarité déjà prononcée avec le modèle, comme la durée des AF), il ne se passe rien…

Ces tendances sont plutôt en accord avec ce qui est décrit dans les conclusions de Révis *et al.* (2013) concernant le comportement vocal des imitateurs. Lorsqu'ils devaient imiter la voix d'un autre, ceux dont le registre vocal était différent de la voix cible tentaient de changer le leur, alors que ceux ayant un registre naturellement approchant n'avaient pas d'ajustement à faire. Ceci indique que les locuteurs percevraient les distances entre leur voix et celle de leurs modèles et qu'ils parviendraient à moduler leur propre production en fonction des distances ressenties.

En guise de conclusion sur cette première approche du CI, soulignons que la poursuite d'une approche de mesurage acoustique des productions imitatives permet d'obtenir des détails très fins sur la production des imitateurs, afin d'en étudier le comportement vocal et probablement, les représentations phonologiques.

Cependant, une telle approche ne constitue plus la coloration principale de notre travail, car cette approche ne nous permet pas de savoir si une production imitative est une imitation fidèle ou non de son modèle, pour le niveau prosodique qui nous intéresse.

C'est pourquoi la suite de ce travail se focalise sur la similarité entre prosodie perçue et représentation graphique, en envisageant systématiquement une taille de constituant allant du syntagme nominal (chapitre 5) à la phrase complète (chapitre 6).

1.4 Une parenthèse sur la comparaison des objets

Au début de notre cheminement, nous pensions qu'évaluer le degré d'imitation contenu dans un énoncé serait une tâche relativement aisée. Nous postulions qu'il suffirait de produire assez de mesures acoustiques différentes pour retrouver les *patterns* de l'imitation. C'est pourquoi nos travaux précédents s'intéressaient en premier lieu aux pauses, à leur présence, leur absence et leur durée ainsi qu'à la hauteur et à la durée des Accents initiaux et des Accents Finaux, et que nous envisagions mesurer le débit, la durée de chaque syllabe pour faire de multiples comparaison.

Pourtant, une revue attentive de la littérature traitant de la convergence phonétique et de l'imitation en général (chapitre 4) montre que cette démarche est assez vaine si l'on cherche à savoir en quoi une reproduction d'un modèle sera perçu comme une bonne ou une mauvaise imitation. Ainsi, la mesure de multiples paramètres acoustiques isolés a été une option que nous avons longtemps considérée avant de conclure que le pouvoir prédictif de ces paramètres isolés est quasi nul si l'on veut pouvoir décider de la réussite d'une imitation. En conséquence, nous pensons à présent que les paramètres acoustiques que nous avions l'habitude de considérer représentent un niveau d'analyse à la granularité trop fine pour estimer la similarité prosodique entre deux énoncés.

En procédant par analogie, considérons que nous voulions estimer la similarité des volumes représentés dans la Figure 25.

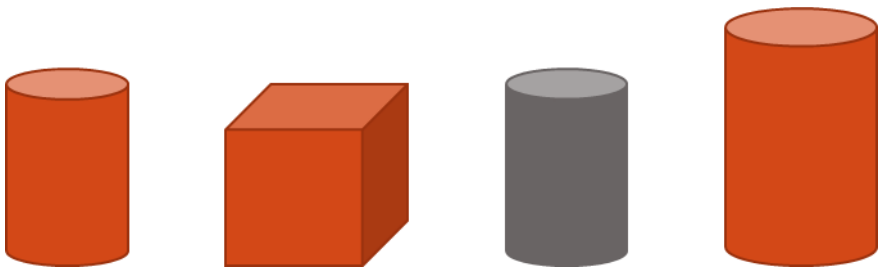


Figure GGG : Figure 25 : Trois cylindres et un cube.

En nous limitant aux trois paramètres que sont la forme, la hauteur et la couleur des volumes, leur comparaison deux à deux aboutit à considérer qu'il existe entre chaque volume au moins un critère de similarité commune (excepté le cube rouge et le cylindre gris). Le cube est clairement différent de tous les autres volumes, mais le critère de couleur reste suffisamment flagrant pour qu'on puisse considérer qu'il existe une similarité partielle entre le cube et deux des cylindres. Si nous supposons que ces volumes ont une masse, peut être que la masse du

cube et celle du cylindre gris seraient suffisamment proches pour que nous arrivions à les trouver similaires à leur tour… Il serait également possible d'affiner encore et encore ces critères.

Ce qui nous semble important dans ces considérations sont les points suivants :

- Comparer des objets se fait sur la base de critères catégorisables
- Le choix de ces critères est arbitraire
- Le nombre de critères de comparaison est théoriquement infini[64]

Pour en revenir à l'étude de la similarité prosodique en parole, en étudiant des paramètres acoustiques isolés, nous nous mettions dans une position délicate puisque nous multiplions les critères de comparaison à bas niveau, ce qui ne nous permettait pas d'appréhender l'ensemble d'une production. Il semble maintenant plus logique de commencer par considérer un niveau plus global, puis de descendre par la suite –et si nécessaire– dans des niveaux plus fins, dans une analyse par strates.

La mise en place du premier niveau de cette démarche, l'étude de la forme, est l'objet du reste de ce chapitre. Rappelons qu'il est nécessaire que la mesure de similarité ait une bonne adéquation avec les résultats des tests perceptifs, afin d'être interprétable, *i.e.* valide.

[64] En tout cas, il peut vite devenir très grand.

2. Jugements perceptifs et mesures de la similarité prosodique

Afin de produire une évaluation systématique de la similarité prosodique d'une imitation et de son modèle, nous avons sélectionné un sous-ensemble du *CI* que nous décrivons ci-après. Ce matériel linguistique a ensuite été passé au crible de deux tests perceptifs de jugement de la similarité (un test AX et un test AXB) et de différentes mesures objectives de la similarité des formes prosodiques. Nous décrirons en suivant les conditions de passation des tests de jugement perceptif et ces différentes mesures.

2.1 Description du matériel linguistique

Pour cette étude, nous avons extrait 4 phrases du *CI* en fonction de plusieurs critères :

1. Les phrases sont issues du *Cas 1*, dont la désambiguïsation syntaxique est marquée par une pause silencieuse entre les deux noms du syntagme nominal. La présence de la pause y présente un intérêt particulier.
    a. D'une part, nous l'avons décrite précédemment comme un indice prosodique privilégié par les imitateurs pour reproduire la structure prosodique, elle devrait par conséquent constituer un repère important pour les auditeurs jugeant la similarité entre les modèles (contenant toujours la pause) et leurs imitations (contenant très souvent une pause).
    b. D'autre part, la pause présente un intérêt particulier pour évaluer la robustesse des méthodes objectives de mesure de la similarité prosodique. Celles-ci doivent pouvoir prendre en compte la pause lorsqu'elles évaluent des constituants prosodiques de plus grande taille.
2. Les phrases choisies ont 2 longueurs de constituants et sont prononcées par 4 sujets qui étaient appariés lors du recueil du CI :
    a. Sp1 (féminin) et Sp5 (féminin) ont imité
        i. Les bagatelles et les balivernes sottes
        ii. Les bonimenteurs et les baratineurs fades
    b. Sp3 (masculin) et Sp7 (féminin) ont prononcé
        i. Les bagatelles et les balivernes saugrenues
        ii. Les bonimenteurs et les baratineurs fabuleux

Chaque sujet a dit 9 fois chacune des phrases (3 répétitions * 3 tâches expérimentales lors du recueil du CI), aboutissant à 18 phrases par sujet. Les tests d'évaluation des formes sonores et de leur représentation graphique ont donc été menés sur un ensemble de 72 phrases du CI (4 sujets * 18 imitations).

2.2 Tests perceptifs de jugement de la similarité prosodique

Comme l'indique Pardo (2013), la réponse des auditeurs aux tests perceptifs de jugement de la similarité permet d'obtenir une information globale et efficace sur le degré d'imitation des énoncés évalués, puisque la perception des auditeurs agrège les dimensions multiples des patterns acoustiques. Ceci dit, nous décrivons en suivant deux tests perceptifs dont les paradigmes ont été décrit au chapitre précédent. Ceux-ci diffèrent par le système de notation de la similarité que les auditeurs utilisent :

- Le test AX consiste en une notation absolue entre le modèle et son imitation.
- Le test AXB propose une évaluation relative en forçant le choix entre deux items imités et leur modèle.

2.2.1 Test AX : évaluation absolue de la similarité prosodique

2.2.1.1 Population

15 auditeurs naïfs (8 femmes et 7 hommes) ont participé au test AX de jugement de la similarité. Aucun de ces auditeurs, francophones natifs âgés de 25 à 32 ans, n'a reporté de trouble de l'audition ou de la parole.

Le recrutement des auditeurs s'est fait par bouche à oreille. Avant le test, ils ont signifié leur accord de participation à l'étude en signant un formulaire de consentement éclairé, leur expliquant leurs droits comme sujet de l'expérience et les devoirs de l'expérimentateur.

La participation au test n'était pas rémunérée.

2.2.1.2 Tâche expérimentale

Lors du test AX de jugement de la similarité, les auditeurs doivent estimer la ressemblance entre un item A (le modèle) et un item X (l'imitation). A cette fin, les sujets disposent d'une échelle de type *likert*, graduée de 1 à 5 (Mary et al., 2013).

Les consignes données aux auditeurs étaient les suivantes :

- Vous allez écouter des phrases.
- A est la phrase de référence, et X est la phrase que vous devez juger.
- Les phrases que vous allez entendre ont le même contenu sémantique, on vous demande de juger leur ressemblance en terme de musicalité.
- Au-delà des mots, on vous demande donc de juger ce qui relève de la mélodie et du rythme.
- En vous servant de l'échelle de 1 (moins ressemblant) à 5 (ressemble fidèlement) notez le degré de ressemblance de X avec A.

Il était également signifié aux sujets qu'ils pouvaient jouer chaque phrase A et X jusqu'à 5 fois avant d'indiquer leur note. L'expérimentateur était aussi disponible pour leur apporter des précisions quant aux consignes.

2.2.1.3 Passation

Le test AX était scripté pour être présenté dans Lancelot, l'environnement HTML du logiciel Perceval (André, Ghio, Cavé, & Teston, 2003). Après accueil du sujet, l'expérimentateur lançait la session expérimentale, lisait la consigne avec le sujet (et répondait à d'éventuelles question) et restait pendant la session d'entraînement à la tâche (4 *trials* non présentés durant le test). Le sujet accomplissait ensuite le test sans que l'expérimentateur intervienne.

Il était prévu que la durée effective de passation soit d'environ 20 minutes. Les sujets les plus rapides ont effectué la tâche en quinze minutes, tandis que les sujets les plus lents l'ont accomplie en un peu plus de trente minutes.

Durant ce temps de passation, les sujets ont effectué 72 *trials* afin d'évaluer, conformément aux consignes, les 72 imitations et leur modèle issus du *CI*.

Les *trials* étaient randomisés par le logiciel et présentés sur un ordinateur portable à écran tactile. Afin d'écouter les sons, les auditeurs disposaient d'un casque haute-fidélité[65].

2.2.1.4 Exploitation en vue des résultats

A l'issue des passations, les 72 imitations ont donc reçu 15 notes comprises entre 1 (très peu ressemblant) et 5 (très ressemblant) sensées évaluer leur similarité prosodique avec le modèle.

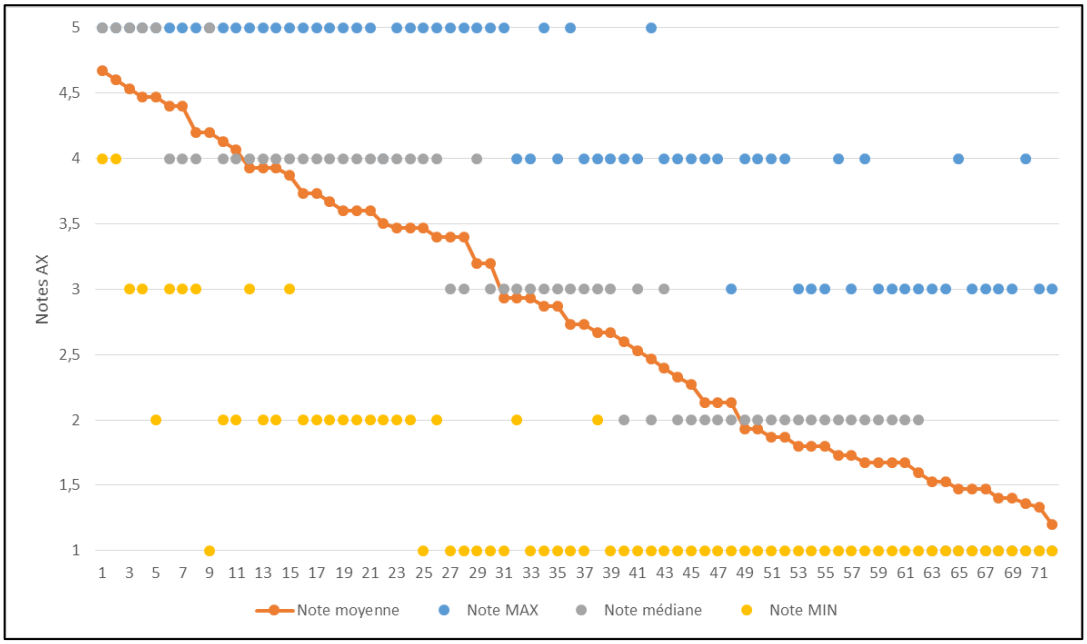
Dans le cas où un sujet n'aurait pas donné de note lors d'un *trial*, la note médiane de l'échelle de notation lui était attribuée automatiquement (soit, la note de 3). Ce cas s'est présenté 2 fois sur 1080 *trials* pour l'ensemble de l'étude (15 sujets * 72 *trials*), soit dans 0,2% des cas.

Suite à ce recueil de scores, la moyenne des notes obtenues pour chaque phrase X a été calculée. A partir de ces moyennes, des rangs ont été attribués aux phrases. Celle obtenant la moyenne la plus élevée obtient le rang 1 du classement AX (la mieux imitée), la suivante le rang 2, et ainsi de suite jusqu'au rang 72. En cas d'égalité de score entre deux phrases X ou plus, un rang égal à la moyenne des rangs qu'elles devraient occuper dans le classement a été attribué à chacune d'entre elles (par exemple : 7 ; 8 ; 9 → 8 ; 8 ; 8). Avant d'aborder les résultats concernant les rangs obtenus par les différents sujets, nous montrerons en premier ce que l'on peut apprendre de la manière dont les notes ont été réparties par les auditeurs.

Les Figures 26a & 26b montre la répartition des scores obtenus par chaque phrase lors de ce test (les imitations sont triées de gauche à droite, de celle qui a reçu la meilleure note moyenne à celle qui a eu la moins bonne). La Figure 26a représente les données descriptives de la dispersion des notes obtenues par les 72 imitations évaluées (note Maximum, Médiane, Moyenne et Minimum). Le diagramme empilé (Figure 26b) montre quant à lui les effectifs obtenus par phrase et par score. Ces répartitions sont assez remarquables pour leurs extrêmes qui sont bien marqués. Ainsi, les auditeurs semblent n'avoir eu aucun mal à déterminer quelles productions des imitateurs étaient soit de très bonnes ou soit de très mauvaises reproductions du modèle.

[65] Beyer Dynamic DT 770 pro : goo.gl/t0r3Io

a)



b)

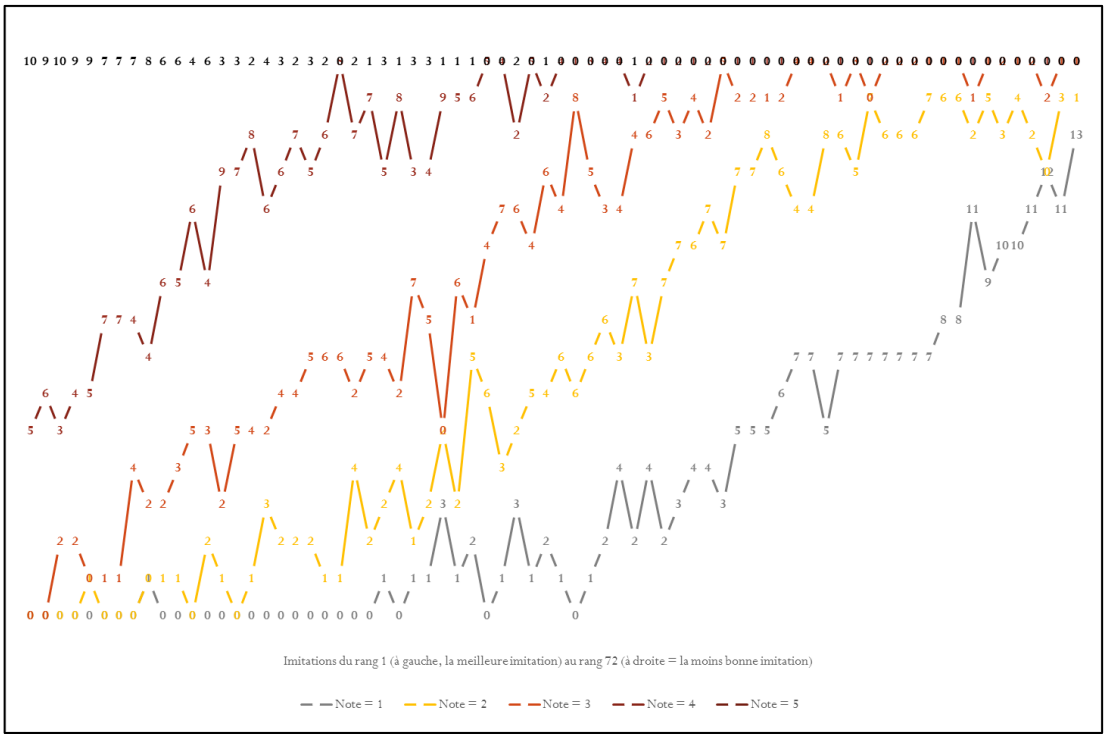


Figure 26a & 26b : Dispersion des scores du test AX pour 72 imitations

a. . Notes moyenne, médiane, maximum et minimum obtenues par chaque paire AX.

b. Effectif de chaque score pour chaque paire AX. Le total de chaque colonne est égal à 15, les scores sont empilés (soit, le nombre total de notes obtenues par chaque phrase).

Dans les deux figures, les phrases sont classées du rang 1 (le meilleur) au rang 72 (le moins bon), de gauche à droite.

En ce qui concerne les accords inter-juges, nous commencerons par souligner les écarts de notes que nous avons relevés dans la Table 7.

Ecart de notes pour une même phrase X

Minimum	Maximum	Moyen	Médian
1	4	2,7	3

Tableau 7 : Ecarts de notes pour une même phrase X

Ces valeurs, ainsi que les Figures précédentes indiquent que les auditeurs tombent rarement d'accord lorsqu'ils utilisent une échelle graduée. C'est un problème récurrent lié à l'utilisation des échelles psychométriques[66], qui varie en fonction des juges :

- Certains auditeurs attribuent très rarement une des deux notes extrêmes et sont donc biaisés vers le centre (ou de manière asymétrique dans le cas du refus d'une seule note)
- D'autres attribuent difficilement une note moyenne et sont alors biaisés sur les côtés

Nous pouvons observer ces phénomènes sur la Figure 27. Sur cet histogramme, on peut remarquer que les juges J2, J14 et J15 utilisent très peu la note la plus basse (1). *A contrario*, J10 et J12 utilisent très peu les notes maximum (4 & 5). J14 par ailleurs suit une tendance plus centrale. La Table 8 indique les valeurs moyennes et médianes du nombre d'attribution de chaque note.

Note	**1**	**2**	**3**	**4**	**5**
Nombre moyen d'attribution de la note	15,93	16,8	14,4	15,53	9,33
Nombre médian d'attribution de la note	18	16	14	14	10

Tableau 8 : Données descriptives sur l'attribution des notes par les juges

[66] Vautier 2016 propose un cours vidéo au sujet du problème des mesures/évaluation en psychologie : goo.gl/8Iz6h3

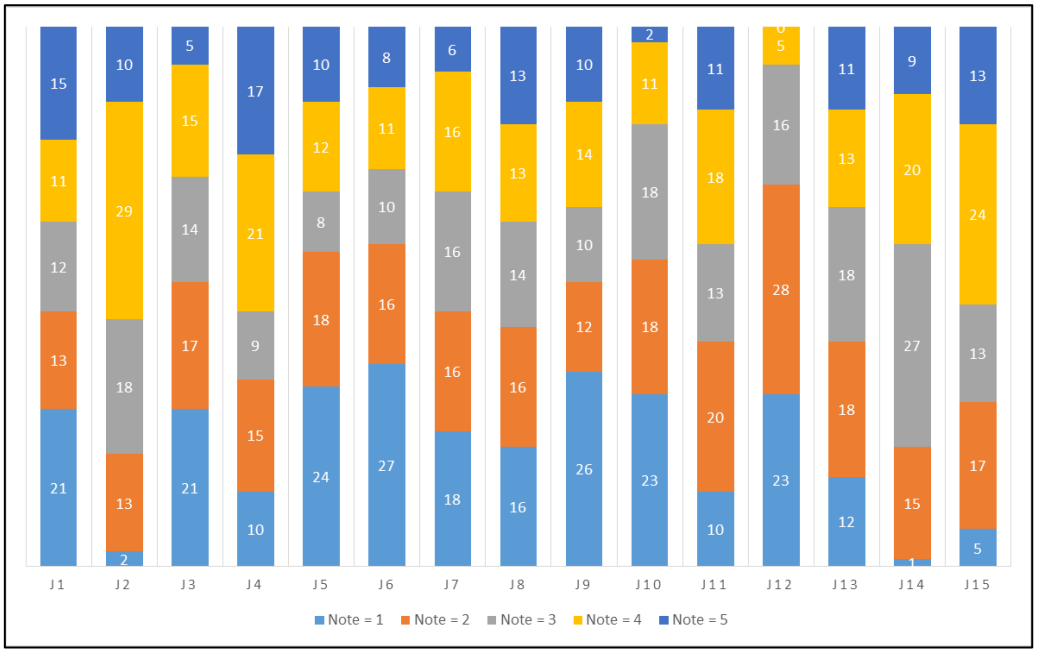


Figure 27 : Répartition des notes attribuées par chaque juge (J1 à J15).Le nombre d'attribution de chaque note est indiqué dans les secteurs de l'histogramme (total de chaque colonne = 72, soit, le nombre de *trial* effectué par chaque juge).

Ces données descriptives laissent présager des accords inter-juges assez faibles (Figure 27 & Table 9), tout en nous assurant cependant que les extrêmes du classement des 72 phrases, pris de manière isolée, obtiennent des notes homogènes (Figures 26a & 26b). La Table 9 présente les κ de Cohen effectués pour estimer le taux d'accord entre nos juges.

	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12	J13	J14	J15
J1	-	.267	.392	.257	.439	.369	.448	.406	.380	.222	.251	.146	.293	.273	.269
J2	.267	-	.265	.400	.281	.194	.211	.401	.260	.143	.187	.062	.365	.402	.484
J3	.392	.265	-	.308	.365	.359	.358	.415	.480	.418	.220	.319	.439	.288	.315
J4	.257	.400	.308	-	.278	.275	.297	.470	.316	.242	.334	.124	.323	.381	.415
J5	.439	.281	.365	.278	-	.452	.460	.471	.426	.298	.312	.318	.331	.219	.281
J6	.369	.194	.359	.275	.452	-	.382	.396	.396	.413	.209	.333	.338	.245	.257
J7	.448	.211	.358	.297	.460	.382	-	.359	.359	.243	.417	.281	.292	.341	.277
J8	.406	.401	.415	.470	.471	.396	.359	-	.382	.356	.303	.276	.432	.303	.473
J9	.380	.260	.480	.316	.426	.396	.359	.382	-	.370	.181	.376	.337	.215	.247
J10	.222	.143	.418	.242	.298	.413	.243	.356	.370	-	.199	.401	.236	.178	.192
J11	.251	.187	.220	.334	.312	.209	.417	.303	.181	.199	-	.174	.285	.347	.378
J12	.146	.062	.319	.124	.318	.333	.281	.276	.376	.401	.174	-	.245	.018	.098
J13	.293	.365	.439	.323	.331	.338	.292	.432	.337	.236	.285	.245	-	.306	.396
J14	.273	.402	.288	.381	.219	.245	.341	.303	.215	.178	.347	.018	.306	-	.301
J15	.269	.484	.315	.415	.281	.257	.277	.473	.247	.192	.378	.098	.396	.301	

Tableau 9 : Kappa de Cohen pondérés par un facteur quadratique entre les juges du test AX. (Facteur quadratique : 0 ; .437 ; .75 ; .937 ; 1)

Pour calculer ces valeurs de κ, nous avons choisi de les pondérer par un facteur quadratique[67]. Ainsi, les différences d'un degré de notation sur un *trial* se verraient sanctionner moins sévèrement que les différences plus larges. L'interprétation du κ ne fait pas consensus, c'est pourquoi les quelques statistiques descriptives que nous avons données précédemment nous rassurent face à ces valeurs de κ qui ne sont pas outrageusement élevées. Ainsi, le meilleur taux d'accord est observé entre J2 et J15 ($\kappa = .484$), mais plusieurs juges sont en complet désaccord :

- J12 et J14 ($\kappa = .018$),
- J12 et J2 ($\kappa = .062$),
- J12 et J15 ($\kappa = .098$).

Nous avions précédemment remarqué que J12 avait une distribution de score assez atypique, puisqu'il n'a jamais attribué le score maximum lors du test. Il semble alors assez normal qu'il tombe rarement en accord avec J2, J14 et J15 qui attribuent très peu la note minimum et sont donc parmi les juges les plus cléments de l'échantillon.

[67] Ce lien renvoi vers un document sur le Kappa que nous n'avons pas intégré aux éléments bibliographiques : goo.gl/jaOEXs

La conjonction de ces κ pondérés ainsi que les dispersions des données que nous avons illustrées précédemment nous indiquent que les juges du test AX n'ont pas fait d'accord parfaits. Ceci n'est pas étonnant car le nombre de scores possibles pour l'évaluation impliquait qu'il y aurait une variabilité inter-juge liée à l'usage de l'échelle de *likert*. Bien qu'effectivement présente, cette dernière n'empêche cependant pas de voir se dessiner des tendances nettes dans ces résultats, que nous allons continuer de décrire en nous intéressant, cette fois, à la perception qu'ont eue les juges de la production des locuteurs du *CI*.

2.2.1.5 Evaluation de la performance des locuteurs du CI

Nos premières remarques avaient trait à la distribution des données recueillies lors du test AX. Les Figures 26a & 26b montraient le classement ordonné des 72 productions évaluées. Cependant, nous n'avons pas abordé la distinction entre les productions des différents locuteurs, dont les performances pourraient être uniformément réparties dans ce classement. La Figure 28 présente cette distinction.

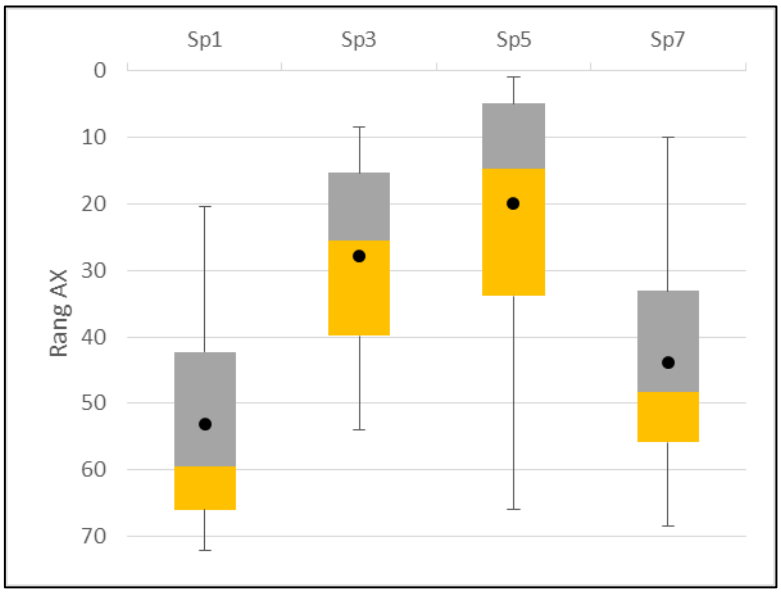


Figure 28 : Répartition interquartile des rangs (de 1, le meilleur, à 72, le moins bon) obtenus par les sujets au test AX. Les points noirs représentent la moyenne des rangs. La médiane peut être visualisée par la ligne de démarcation entre jaune et gris.

Au vue de cette figure, il semble assez évident de souligner que les différents imitateurs ont été perçus comme plus ou moins doués en ce qui concerne leur performance imitative.

Bien que l'étendue des rangs obtenus par *Sp5* soit la plus large, la plus grande part des productions de ce sujet se concentre dans le haut du classement : la moitié de ses productions occupe des rangs compris entre 1 et 15.

L'autre sujet dont la performance se démarque est *Sp3*. Hormis le fait que la moyenne des rangs obtenus par les productions de ce sujet permet de le classer comme le second le plus performant, il est remarquable de constater que l'étendue de ses rangs obtenus est la plus réduite. Nous pourrions alors considérer *Sp3* comme l'imitateur dont la production a été la plus constante au fil des tâches du recueil du *CI*.

Les productions de Sp7 obtiennent quelques bonnes places dans le classement AX. Cependant les productions de ce sujet sont majoritairement évaluées comme peu fidèles au modèle, 75% de ses productions étant classées en deçà du rang 35.

Enfin, Sp1 est évalué comme le sujet le moins performant dans la tâche de reproduction de la structure prosodique perçue.

Afin d'illustrer encore ces tendances, nous donnons dans la Table 10, la moyenne des scores obtenus par les différents sujets.

Sujet	*Moyenne des scores/5*
Sp1	1,992
Sp3	3,181
Sp5	3,575
Sp7	2,374
Tous sujets	2,78

Tableau 10 : Moyenne des scores obtenus sur l'ensemble des productions évaluées au test AX. Chaque moyenne est calculée à partir de 270 scores (18 phrases évaluées par sujet * 15 juges au test AX) ; la moyenne de tous les sujets, quant à elle est calculée à partir de 1080 scores.

2.2.1.6 Discussion : apports du test AX à l'évaluation de la similarité prosodique

Il était attendu de ce test AX, qu'il nous permette d'obtenir une évaluation de la similarité prosodique entre les modèles et leurs imitations. Il semble en effet que les auditeurs soient parvenus à établir une hiérarchie parmi les imitations évaluées. Ceci dit, il nous semble difficile de pouvoir expliquer les fondements sur lesquels les auditeurs ont effectué leur choix.

C'est d'ailleurs un des principaux inconvénients que nous évoquions quand nous discutions les méthodologies de l'étude de l'imitation en parole : les tests perceptifs sont décrits dans la littérature comme une évaluation holistique, et nous ne pouvons qu'en convenir.

Etant donnée la nature de nos stimuli, nous pensons pouvoir évoquer quelques pistes d'explications partielles concernant l'évaluation des productions du sujet *Sp7*. Lors du recueil du *CI*, le sujet *Sp7* était un de ceux qui omettait le plus grand nombre de pause. Il en va de même, dans l'échantillon sélectionné pour cette étude de la similarité prosodique perçue, comme l'indique la Table 11.

Pause ou Non Pause	*Nombre de réalisations*	*Proportion*	*Note moyenne (sur 5)*
Pause	6/18	33,3%	2,97
Non Pause	12/18	66,6%	2,07

Tableau 11 : Evaluation AX de Sp7 et pauses produites.

Si l'on se fie aux données décrites par cette table, la présence ou l'absence de la pause peut être considérée comme un indice ayant une influence forte dans le jugement des auditeurs de ce test AX. Ce résultat serait à lier avec ce que nous disions de la production des locuteurs du *CI* : la pause était un paramètre prosodique majeur dans le repérage perceptif de la structure du *Cas 1* et avait été presque systématiquement reproduite. Nous pourrions appuyer ce propos en indiquant que la moyenne des scores reçus par les énoncés sans pause (17 / 72 productions pour l'ensemble des sujets) est de 1,93 ; soit, dans la tranche basse des notes obtenues par les 72 phrases. Cependant, nous ne pouvons avancer ces propos de manière sûre si l'on se penche sur les productions de *Sp1* (Table 12).

Pause ou Non Pause	*Nombre de réalisations*	*Proportion*	*Note moyenne (sur 5)*
Pause	14/18	77,7%	2,2
Non Pause	4 /18	22,2%	1,46

Tableau 12 : Evaluation AX de Sp1 et pauses produites.

Nous remarquons, certes, que la même tendance se dessine entre *Sp1* et *Sp7* : les énoncés sans pause reçoivent une moins bonne évaluation de la part des auditeurs. Pourtant, malgré une plus grande proportion de pauses produites, *Sp1* reçoit une évaluation globale moins bonne que celle de *Sp7*. Ainsi, la pause pourrait avoir une importance notable dans le jugement perceptif de la similarité prosodique de deux énoncés sans que ce critère soit absolument décisif. Les auditeurs portent probablement leur attention sur d'autres paramètres prosodiques lors de leur jugement. Enfin, notons tout de même que notre tentative d'explication demeure au niveau de pistes : la quantité de données analysée ici ne permet pas de conclure de manière définitive.

En guise de perspective sur la question du poids de la pause dans le jugement perceptif de la similarité prosodique, il pourrait être pertinent de manipuler les énoncés *Cas 1* du *CI* contenant une pause, de la même manière que nous avions créé les modèles *Cas 1 bis* ; puis de les soumettre à un nouveau test AX. Si les énoncés dont on a effacé la pause obtenaient des scores systématiquement moins élevés que les imitations originales, nous pourrions alors avoir une meilleure indication de l'importance de la pause dans ce processus de jugement perceptif de la similarité prosodique.

Afin de poursuivre nos investigations sur l'évaluation de la similarité prosodique, nous avons par la suite soumis les mêmes données à un test AXB. Nous le décrivons en suivant.

2.2.2 Test AXB : évaluation relative de la similarité prosodique

2.2.2.1 Population

24 auditeurs naïfs (15 femmes et 9 hommes) ont participé au test AXB de jugement de la similarité. Aucun de ces auditeurs, francophones natifs âgés de 21 à 32 ans, n'a reporté de trouble de l'audition ou de la parole.

Le recrutement des auditeurs s'est fait par bouche à oreille et affichage sur le campus de l'université Toulouse 2 Jean Jaurès. Avant le test, les auditeurs ont exprimé leur accord de participation à l'étude en signant un formulaire de consentement éclairé leur expliquant leurs droits comme sujet de l'expérience ainsi que les devoirs de l'expérimentateur à leur égard.

A l'issue du test, un dédommagement d'un montant de 10 € a été remis aux sujets. Ce financement a été obtenu par le biais de la cellule de valorisation de l'UT2J[68].

2.2.2.2 Tâche expérimentale

Lors du test AXB de jugement de la similarité, les auditeurs effectuent un choix forcé entre les imitations A et B. Par leur choix, ils indiquent laquelle des deux phrases périphériques ressemble le plus à la phrase X.

Les consignes données aux auditeurs étaient les suivantes :

- Vous allez écouter des phrases.
- Vous devez dire quelle phrase de A ou de B, ressemble le plus à la phrase X.
- Les phrases que vous allez entendre ont le même contenu sémantique, on vous demande de juger leur ressemblance en terme de musicalité.
- Au-delà des mots, on vous demande donc de juger ce qui relève de la mélodie et du rythme.

L'expérimentateur était aussi disponible pour leur apporter des précisions quant aux consignes.

Il était également indiqué aux sujets qu'ils pouvaient jouer chaque phrase A, X et B jusqu'à 5 fois avant d'indiquer leur choix. Habituellement, le choix forcé dans le protocole AXB doit se faire très rapidement. Cependant, le test AXB est souvent effectué sur des mots-cibles isolés (Goldinger, 1998; Pardo, 2006) ; or nos stimuli avaient un empan bien plus large. Il nous a paru plus simple pour la mémoire de travail des sujets de les autoriser à écouter plusieurs fois les syntagmes nominaux en triplets.

2.2.2.3 Appariements entre phrases A et B et constitution des groupes

Comme le test AXB propose une évaluation à choix forcé entre les items A et B, il est nécessaire que chaque phrase évaluée :

[68] Appel à projet Emergence, janvier/février 2016, remporté pour un financement total de 4000 €, pour le projet Verbo Tonal Method – Trainer.

- Soit opposée à toutes les autres phrases équivalentes de l'échantillon (locuteur et contenu segmental)
- Se retrouve, pour chaque opposition, dans la position A et dans la position B.

Ces contraintes impliquent que pour un nombre n de phrases à évaluer il faut constituer un nombre N_{AXB} de triplets équivalent à :

$$N_{AXB} = n(n-1)$$

Ici, il a été décidé d'obtenir des classements par locuteur *et* par phrase du CI. En effet, le test AX nous a permis d'obtenir une information globale sur la performance imitative des locuteurs en classant l'ensemble de leurs phrases d'un point de vue perceptif. Pour ce test AXB, nous voulions obtenir une granularité intra-individuelle plus fine tout en évitant d'éventuels biais liés à la perception de la voix des locuteurs du CI.

Chacune des 4 phrases évaluées a été imitée 9 fois par 2 locuteurs différents, ce qui donne un total de *trials* pour $n = 9$, équivalent à :

$$N_{AXB} = 9(9-1) \times 4 \times 2 = 576$$

Etant donné ce nombre important, et la grande monotonie des contenus à évaluer, il a été décidé de créer 4 groupes de sujets, pour un total de 144 *trials* à accomplir par groupe. La Table 13 montre la répartition des trials par groupe expérimental.

Phrase	**Les bagatelles et les balivernes sottes**		**Les bagatelles et les balivernes saugrenues**		**Les bonimenteurs et les baratineurs fades**		**Les bonimenteurs et les baratineurs fabuleux**	
Dite par	**Sp1**	**Sp5**	**Sp3**	**Sp7**	**Sp1**	**Sp5**	**Sp3**	**Sp7**
Ensemble 1	G1	G4	G2	G3	G4	G1	G3	G2
Ensemble 2	G2	G3	G1	G4	G3	G2	G4	G1

Tableau 13 : Répartition des trials AXB par groupes expérimentaux.Un triplet AXB présent dans l'Ensemble 1, a son équivalent BXA dans l'Ensemble 2, et inversement. Chaque case où un groupe est noté correspond à 36 *trials*.

Cette répartition des *trials* assure :

- que chaque groupe expérimental entendra les 4 locuteurs du CI et les 4 phrases différentes sélectionnées pour l'étude,
- que la production d'un même locuteur du CI sera évaluée de manière partielle par chaque groupe expérimental,
- que l'ensemble des phrases sera évalué en position AXB et BXA.

2.2.2.4 Passation du test AXB

Le test se déroulait dans la plateforme expérimentale CLOE de l'UT2J (goo.gl/L6Y7DT). Cette salle peut accueillir jusqu'à 8 participants sur des postes informatiques. Cependant, lors des passations de la tâche AXB, nous ne disposions que de 4 casques haute-fidélité identiques limitant alors notre capacité d'accueil à 4 participants par session.

Le test AXB a lui aussi été scripté pour être présenté dans Lancelot, l'environnement HTML du logiciel Perceval (André et al., 2003). Après accueil du ou des sujets, l'expérimentateur attribuait un groupe au(x) sujet(s), de manière à équilibrer au mieux la répartition des auditeurs masculins et féminins dans les groupes.

L'expérimentateur lançait ensuite la session expérimentale, lisait les consigne avec le(s) sujet(s) (et répondait à d'éventuelles question). Il restait pendant la session d'entraînement à la tâche (4 *trials*). Le test était ensuite accompli sans que l'expérimentateur n'intervienne.

Dans le cas de passation en groupe, l'expérimentateur expliquait les consignes collectivement après que chacun des participants a été installé. Cependant, il s'assurait individuellement, avec chaque sujet, que la tâche et la manipulation de l'interface du test étaient comprises avant de laisser les sujets démarrer le test.

Lors des tests préalables de la procédure expérimentale, le temps de passation de l'expérience avait été évalué à environ 45 minutes (60 minutes de présence prévue pour les sujets, explication des consignes comprise). Les auditeurs ont effectivement mis de 30 à 60 minutes pour accomplir la tâche qui leur était demandée.

Durant ce temps de passation, les sujets ont effectué 144 *trials*. Les *trials* étaient randomisés par le logiciel et présentés sur un ordinateur de bureau. Afin d'écouter les sons, les auditeurs disposaient d'un casque haute-fidélité[69].

2.2.2.5 Le système de votes AXB

L'*output* du test AXB consiste en une série de 0 et de 1 : à chaque trial, le sujet attribue 1 à A s'il le choisit (et donc, 0 à B), et inversement s'il choisit B. En d'autres termes, le test AXB constitue un système d'élection, dont on peut faire ressortir des tendances. Nous avons par exemple exprimé au chapitre 4, une utilisation possible des votes AXB, en présentant la mesure du taux de convergence phonétique proposée par Pardo (2013, p. 2). L'alternative que nous avons choisie consiste à effectuer un comptage des votes pour comparer la distribution des votes obtenus par les phrases à une distribution théoriquement idéale de l'*output* d'un test AXB.

En émettant l'hypothèse que les objets du test AXB sont effectivement catégorisables du mieux au moins bien imité, il est assez simple de déterminer cette distribution idéale. Imaginons que les formes de la Figure 29 sont les objets d'un test AXB sur la similarité perçue de formes géométriques. On cherche à estimer le degré de similarité perçue entre les formes périphériques (des polygones) et la forme centrale (un cercle). Le cercle sera en position X des triplets, les autres formes en position A et B (ou B et A).

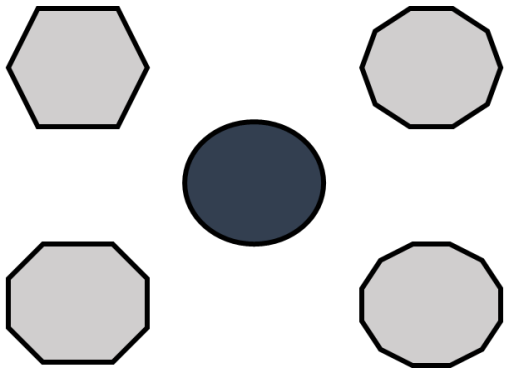


Figure 29 : Un cercle, au centre, encadré par un hexagone, un octogone, un dodécagone et un décagone(sens de rotation antihoraire, en partant du nord-ouest).

[69] Beyer Dynamic DT 770 pro

On peut calculer en premier lieu le nombre de triplets AXB et BXA possibles pour les formes, N_{forme} :

$$N_{forme} = 4 \times (4 - 1) = 12$$

Pour un nombre x de sujets évaluant n formes, le nombre total V_{total} de votes distribués sera égal à :

$$V_{total} = xN_{AXB}$$

En supposant que nos formes sont évaluées par 2 sujets, nous aurons dans notre cas :

$$V_{forme} = xN_{forme} = 2 \times 12 = 24$$

En ce qui concerne la répartition de ces votes entre les formes, considérons en premier lieu le nombre maximal d'apparition d'une forme dans les triplets : il est égal à $2(n - 1)$.

Pour x sujets faisant le test AXB sur n formes, nous pouvons calculer le score maximum V_{max} qu'une forme peut obtenir au test :

$$V_{max} = 2x(n - 1)$$

La distribution théorique idéale d'un test AXB est donnée par les termes d'une suite arithmétique de premier terme

$$U_0 = V_{max}$$

et de raison

$$r = -2x$$

où x est égal au nombre de sujets. Cette suite a n termes (le nombre d'objets évalués) et la somme de ses termes est égale à V_{total}.[70] En d'autres termes, la distribution théorique du test AXB correspond à une fonction affine ayant un intercept égal V_{max} et un coefficient directeur égal à $-2x$.

Reprenons notre exemple issu de la Figure 29, où 2 sujets évaluent 4 formes. D'après les indications que nous venons de donner, la distribution idéale du test AXB de ces 4 formes est donnée dans la Table 14. Etant donné que les polygones sont assez facilement

[70] Nous aurions dû considérer une suite de premier terme 0 et de raison 2x. Malheureusement, notre raisonnement a démarré par le haut, en considérant le nombre maximum de vote…

catégorisables, nous pourrions alors supposer que les résultats d'un tel test suivront la distribution que nous venons de décrire.

Forme	**Nombre théorique de votes**
Dodécagone	12
Décagone	8
Octogone	4
Hexagone	0

Tableau 14 : distribution idéale des votes d'un test AXB évaluant 4 formes par 2 sujets.

La question que nous pouvons alors nous poser pour le test AXB que nous avons effectuée sur les stimuli du CI est la suivante : les auditeurs ont-ils réussi à classer les imitations entendues de manière nette (leurs votes suivent-ils une distribution théorique) ou bien, dans le cas contraire, ont-ils eu des difficultés à distinguer les imitations les unes des autres ?

2.2.2.6 Distributions des votes du test AXB

Lors de ce test AXB, nous avons comparé entre elles les reproductions d'un même modèle par un locuteur. Nous allons donc montrer les distributions de l'évaluation des imitations de 8 modèles (2 phrases * 4 locuteurs). Chaque modèle du CI avait été reproduit 9 fois par les locuteurs au cours du CI. Les distributions dont nous allons parler concernent alors 72 triplets AXB-BXA évalués 6 fois. En d'autres termes, chaque triplet a été entendu 6 fois, et chaque imitation a été entendue 96 fois (48 fois par position A ou B).

Rang (du meilleur au moins bon)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Nombre de votes attendus	96	84	72	60	48	36	24	12	0

Tableau 15 : Distribution théorique des votes attendus pour l'évaluation AXBd'un bloc de 9 imitations par 6 sujets, en faisant l'hypothèse que chaque objet évalué est suffisamment mieux ou moins bien imité que les autres.

Pour chaque modèle reproduit 9 fois par un locuteur du CI, le nombre total de votes à attribuer est donc de 432 (72 triplets * 6 évaluations). La distribution théorique AXB pour ces valeurs est indiquée en Table 15. Ayant fait évaluer 8 modèles, nous avons recueilli un total de 3456 votes durant ce test.

Afin de comparer les résultats observés des votes AXB, aux votes attendus représentés par la distribution théorique, nous utiliserons un test multinomial de conformité (*multinomial test of Goodness-of-fit*) au moyen d'une simulation de Monte Carlo. En effet, le test de χ^2 de conformité n'est pas applicable ici, car le nombre de votes attendus pour le rang 9 est égal à 0. Or, un des critères d'applicabilité du χ^2 est d'avoir des effectifs théoriques au moins supérieurs ou égaux à 5 dans toutes les catégories (de plus, dans le test du χ^2 les rapports entre effectifs observés et effectifs théoriques sont divisés par les effectifs théoriques… Enfin, le 0 attendu dans le rang 9 pose également un problème pour l'usage de la méthode de Monte Carlo, c'est pourquoi nous avons choisi d'égaliser les ratios de scores attendus pour prendre en compte l'hypothèse selon laquelle les imitations extrêmes des distributions seront plus dures à distinguer les unes des autres. Nous présentons ces ratios dans la Table 16.

Ces tests ont été conduits dans le logiciel de statistique `R` (*ver* 3.2.1) (CRAN, 2016). Nous avons utilisé la fonction `xmonte` du package `XNomial`, qui permet d'approximer le résultat d'un test exact multinomial de conformité au moyen d'une simulation de Monte-Carlo. Le package propose également la fonction `xmulti`, qui n'était pas utilisable ici, en raison du trop grand nombre de comparaison à effectuer pour obtenir le résultat du test exact (de l'ordre de 1^{e}16 comparaisons).

Rang	*Rang 1*	*Rang 2*	*Rang 3*	*Rang 4*	*Rang 5*	*Rang 6*	*Rang 7*	*Rang 8*	*Rang 9*	*Total*
Vote théorique	96	84	72	60	48	36	24	12	0	432
Ratio théorique	8	7	6	5	4	3	2	1	0	36
Correction vote	84	84	84	60	48	36	12	12	12	432
Ratio corrigé	7	7	7	5	4	3	1	1	1	36

Tableau 16 : Votes et Ratio théoriques et corrigés pour le test AXB

Après description des résultats, nous testerons l'hypothèse nulle selon laquelle toutes les distributions des votes AXB sont conformes à la distribution théorique que nous venons de définir.

Afin de faciliter la lecture des figures suivante, nous destinons la Table 17 au lecteur qui y trouvera la manière dont sont codées les productions des imitateurs (par exemple : S01_E-Bag-C1-2).

Sxx	**R ou I ou E**	**Bag ou Boni**	**C1**	**1 ou 2 ou 3**
Le numéro des sujets (1, 3, 5 ou 7)	La tâche expérimentale lors du recueil du *CI* Répétition (R), Imitation (I) ou Exagération (E)	Bag : Bagatelles et balivernes Boni : Bonimenteurs et baratineurs	Toutes les phrases sont issues du *Cas 1*.	La première, seconde ou troisième production lors de la tâche de recueil du *CI*

Tableau 17 : Codage des phrases issues du CI

2.2.2.7 Résultats du test AXB

- *Sujet : Sp1*

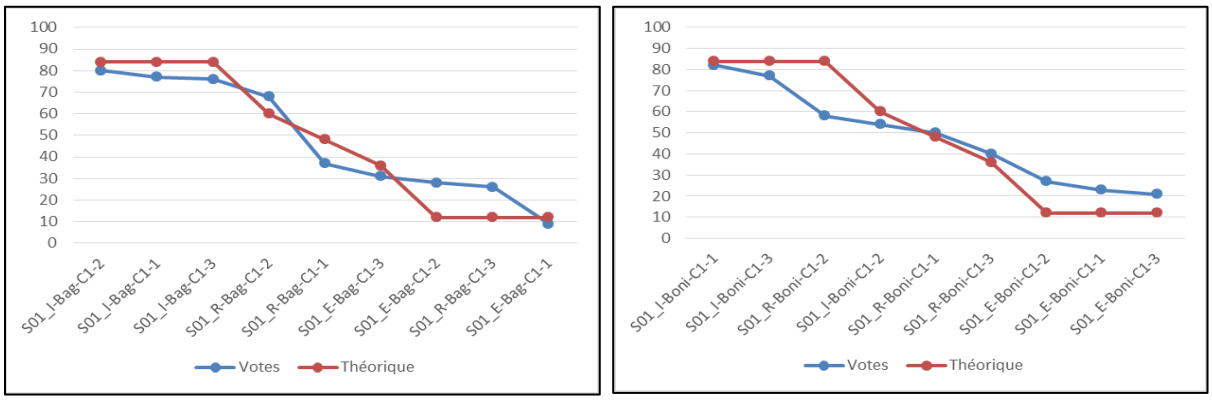


Figure 30a & 30b : Votes AXB obtenus par le sujet Sp1. A gauche, les productions émanant du modèle « *Les bagatelles et les balivernes sottes* » (30a), à droite, celles issues de « *Les bonimenteurs et les baratineurs fades* » (30b). La courbe rouge représente la distribution théorique.

La distribution de la Figure 30a présente la particularité d'avoir deux plateaux assez distincts démarquant les imitations de Sp1 perçues comme mieux et moins bien imitées. Il semble donc que les auditeurs sont parvenus à distinguer ces deux groupes de phrases, sans pouvoir hiérarchiser clairement les productions au sein de ces groupes. La distribution observée des votes obtenus n'est pas conforme aux ratios de la distribution théorique :

```
P value (LLR) = 2e-05 ± 1.414e-05
1e+05  random trials
Observed:  80 77 76 68 37 31 28 26 9
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

Sur la Figure 30b, nous remarquons un palier regroupant les trois moins bonnes imitations de ce bloc, conformément à ce qu'attend la distribution que nous étudions. Cependant, il semble y avoir une déviation importante entre les votes observés et les votes théoriques, ce qu'indique le résultat du test exact :

```
P value (LLR) = 0 ± 0
1e+05  random trials
Observed:  82 77 58 54 50 40 27 23 21
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

- *Sujet : Sp3*

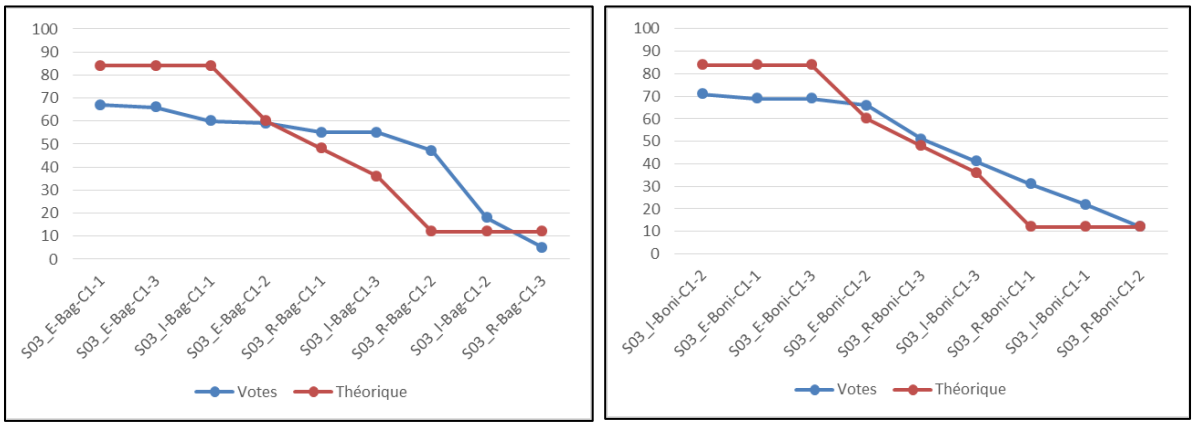


Figure 31a & 31b : Votes AXB obtenus par le sujet Sp3.A gauche, les productions émanant du modèle « *Les bagatelles et les balivernes saugrenues* » (31a), à droite, celles issues de « *Les bonimenteurs et les baratineurs fabuleux* » (31b). La courbe rouge représente la distribution théorique.

Les deux distributions des votes obtenus par *Sp3* présentent de très nets effets de plateaux :

- Le plateau de la Figure 31a s'étend sur les 7 productions de Sp3 perçues comme les meilleures. Cette distribution indique donc que les sujets ne sont pas parvenus à distinguer ces imitations les unes des autres. Elle est assez caractéristique d'un test AXB dont le résultat est indécis. Malgré l'éventualité de regroupement des votes que nous avions prise en compte dans notre distribution théorique, le test de conformité n'est pas concluant.

```
P value (LLR) = 0 ± 0
1e+05  random trials
Observed:  67 66 60 59 55 55 47 18 5
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

- Sur la Figure 31b, deux tendances se dégagent. Un plateau sur les 4 imitations les mieux notées, puis une répartition des votes suivant une distribution quasi linéaire. Dans ce cas, le sujet aurait produit de meilleures imitations, difficile à distinguer entre elles, puis des imitations de moins en moins bonnes mais différenciables les unes des autres. Malgré ce plateau initial, la distribution observée n'est pas conforme à la distribution attendue :

```
P value (LLR) = 0 ± 0
1e+05  random trials
Observed:  71 69 69 66 51 41 31 22 12
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

- *Sujet : Sp 5*

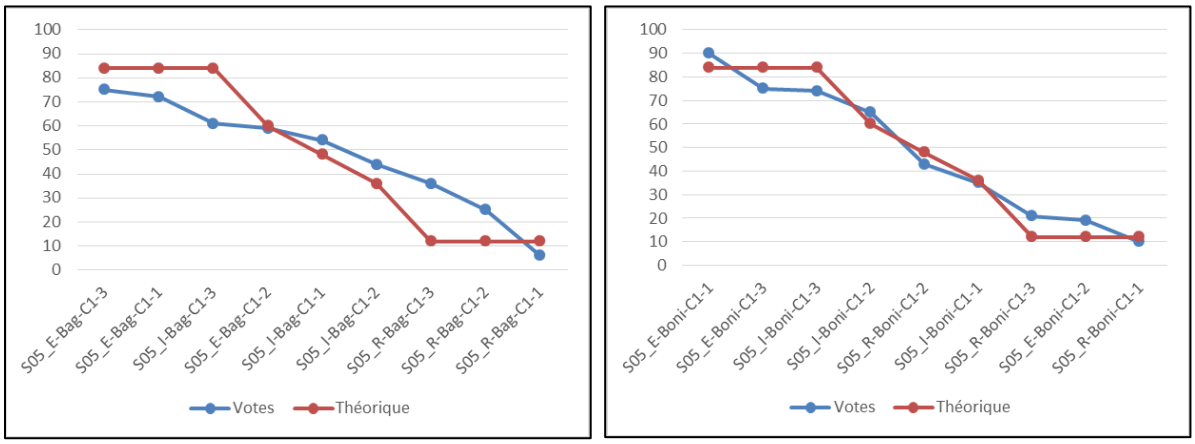


Figure 32a & 32b : Votes AXB obtenus par le sujet Sp5. A gauche, les productions émanant du modèle « *Les bagatelles et les balivernes sottes* » (XXXa), à droite, celles issues de « *Les bonimenteurs et les baratineurs fades* » (XXXb). La courbe rouge représente la distribution théorique.

La distribution des votes des auditeurs pour les phrases « *Bonimenteurs* » du sujet *Sp5* (Figure 32b) est remarquablement proche de la distribution théorique : les juges semblent être parvenus à catégoriser très clairement les productions de ce sujet. Le test multinomial de conformité indique d'ailleurs que les deux distributions se ressemblent :

```
P value (LLR) = 0.11768 ± 0.001019
1e+05  random trials
Observed:  90 75 74 65 43 35 21 19 10
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

Concernant la distribution observée en Figure 32a, la tendance semble moins bien dessinée puisqu'il semble y avoir un palier dans le milieu de la distribution, puis une tendance dégressive insuffisamment marquée pour que le test de conformité soit concluant :

```
P value (LLR) = 0 ± 0
1e+05  random trials
Observed:  75 72 61 59 54 44 36 25 6
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

- *Sujet : Sp7*

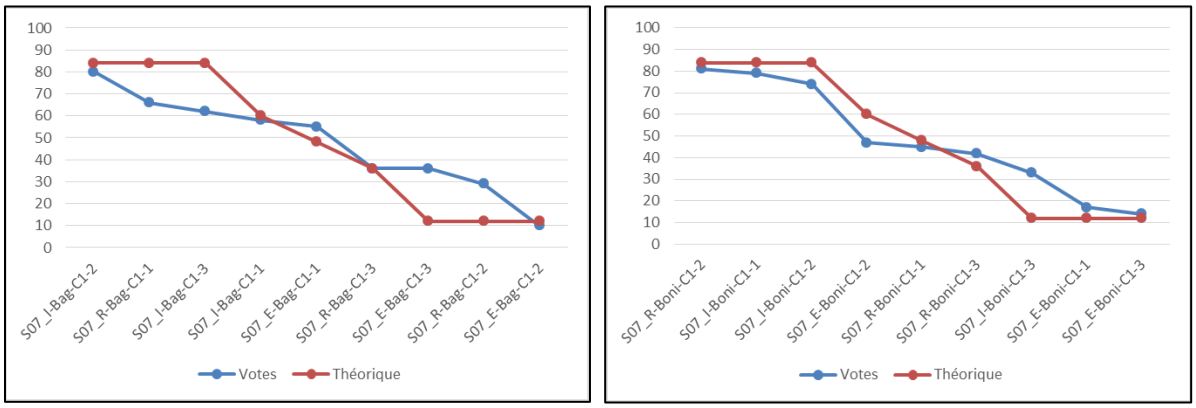


Figure 33a & 33b : Votes AXB obtenus par le sujet Sp7. A gauche, les productions émanant du modèle « *Les bagatelles et les balivernes saugrenues* » (33a), à droite, celles issues de « *Les bonimenteurs et les baratineurs fabuleux* » (33b). La courbe rouge représente la distribution théorique.

Les productions de *Sp7* récoltent des votes par paliers :

- En Figure 33a, hormis l'imitation classée 1 qui se détache du lot, les imitations classées 2 à 5 se trouvent sur un même plateau. Les trois imitations suivantes sont aussi sur un palier, mais plus haut que celui prévu par la distribution théorique. Le test de conformité conclut d'ailleurs, que les deux distributions sont différentes

```
P value (LLR) = 0 ± 0
1e+05  random trials
Observed:  80 66 62 58 55 36 36 29 10
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

- Le début de la distribution des votes observée dans la Figure 33b semble coïncider avec la distribution théorique. Cependant, le second plateau de la distribution est positionné en milieu de distribution, plutôt qu'en fin.

```
P value (LLR) = 0.00011 ± 3.316e-05
1e+05  random trials
Observed:  81 79 74 47 45 42 33 17 14
Expected Ratio:  7 7 7 5 4 3 1 1 1
```

2.2.2.8 Discussion AXB

En faisant passer ce test AXB, nous nous sommes demandés si les différentes imitations des locuteurs du CI étaient suffisamment différentes les unes des autres pour pouvoir être catégorisées, *i.e.* savoir si les votes attribués à chaque ensemble de phrases évaluées suivraient la distribution théorique que nous avions définie. Mis à part les productions du paradigme « *les bonimenteurs et les baratineurs* » de *Sp5* dont les votes obtenus sont fidèles à nos attentes pour ce test, l'évaluation AXB des autres blocs d'imitations ont abouti à des résultats plus incertains.

Les distributions observées se situaient entre deux distributions théoriques possibles pour un test AXB, que nous illustrons en Figure 34.

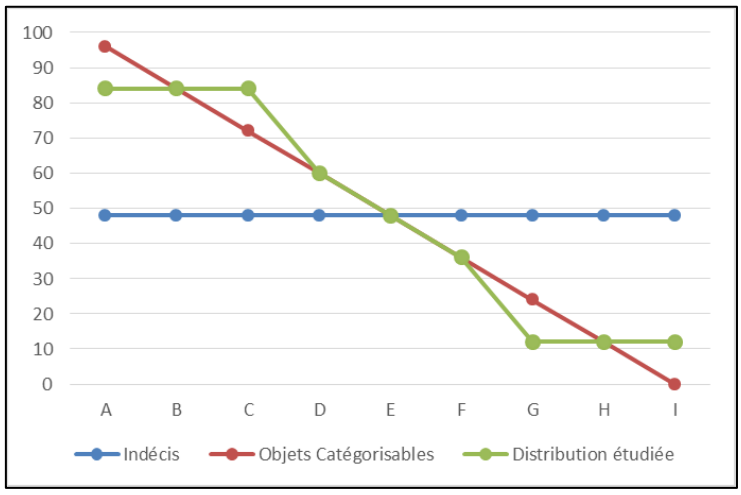


Figure 34 Trois distributions théoriques pour différentes issues possibles des tests AXB. La distribution verte correspond à celle que nous avons testé précédemment, nous discutons plus particulièrement les autres distributions.

La courbe rouge de la Figure 34 montre la distribution que nous avons définie en premier lieu : on y passe d'un rang à l'autre en ajoutant ou en soustrayant deux fois le nombre de sujets au nombre de votes obtenu. Cela représente l'issue attendue du test AXB si les objets de A à H sont clairement différentiables les uns des autres. Quant à la courbe bleue, elle représente une issue que nous qualifions d'indécise[71], puisque chaque objet évalué y obtient un vote dans la moitié des triplets où il est présent. Enfin, la distribution que nous avons comparée à nos données postule que les trois items les mieux et les moins bien évalués

[71] Qui peut potentiellement être une issue souhaitée si l'hypothèse est que les sujets du test ne parviendront pas à faire la différence entre les objets comparés

recevront un nombre équivalent de votes. Les résultats des tests de conformité à la courbe verte n'étaient pas concluants (sauf pour *Sp5*, soit, 1 seul cas sur 8) car les groupes d'imitations perçues à un même niveau étaient rarement centrés. Ceci illustre la difficulté que nous pouvons rencontrer quant aux hypothèses à émettre sur l'issue d'un test AXB d'évaluation de la similarité prosodique, sans avoir d'information préalable sur cet aspect. Une métrique évaluant cette similarité pourrait permettre de pallier ce problème, si tant est que nous connaissons suffisamment le lien entre une telle métrique et le jugement perceptif de la similarité[72].

Les distributions théoriques rouge et bleue de la Figure 34 peuvent révéler des faits intéressant :

- La courbe rouge est intéressante pour connaître la hiérarchie perçue entre les imitations
- Des distributions du type de la courbe bleue apportent un autre genre d'information : quand, lors d'un test AXB, des parties de la distribution sont en palier, cela indique que les juges ne sont pas parvenus à catégoriser les objets individuellement.

Cela n'est pas problématique en soi, puisque ces paliers indiquent simplement la proximité perceptive entre plusieurs productions. Cependant, cela peut devenir plus épineux si l'on cherche à savoir si des productions sont de bonnes ou de mauvaises imitations :

- Dans le cas d'une distribution complètement plate des votes (courbe bleue, Figure 34), on ne sait pas si les imitations sont toutes très réussie ou très mauvaises ! (voir les productions de Sp3, Figure 31a).
- De même, s'il y avait deux paliers (voir les productions de *Sp1* Figure 30a), on ne pourrait pas connaître la distance entre les objets des deux paliers, nous serions justes en mesure de dire qu'il y a une différence d'au moins un degré entre les deux groupes d'objets représentés par ces paliers.

Ainsi, le test AXB seul donne, par son système de choix forcé, une indication réduite du degré de similarité contenu dans les imitations, nous devrions plutôt dire qu'il exprime les différences entre les énoncés A et B. Ce test permet d'établir une hiérarchie entre les items

[72] Et le serpent se mord un peu la queue dans notre cas, puisque ces tests ont pour but de trouver notre méthode de calcul de la similarité prosodique..

(ou entre des groupes d'items) mais il ne fournit pas, tout comme le test AX, d'explications quant aux raisons qui permettent de faire cette hiérarchie.

Enfin, le caractère relatif de l'évaluation AXB nous a contraints à produire uniquement des comparaisons intra-locuteur. Certes, la trop grande variabilité du *CI* en termes de phrases à imiter conduit aussi à cet état de fait, car chaque phrase n'a été dite que par une paire de locuteurs. Ceci dit, vue l'écart de performance perçue entre les locuteurs appariés (*Sp1 vs. Sp5* et *Sp3 vs. Sp7*) durant le test AX, il aurait été probable que les auditeurs fassent subir aux locuteurs les moins performants un délit de « tractus vocal[73] », en choisissant systématiquement le locuteur qu'ils ont ressenti être le meilleur.

Le travail de Hirst (2016), qui propose une technique de clonage de la *f0*, permettant d'extraire et de coller la *f0* d'un sujet sur la production d'un autre, nous ouvre une porte méthodologique pour de futurs tests ne risquant pas de pâtir de biais de perception des locuteurs. En utilisant une telle technique, nous pouvons envisager de proposer des comparaisons inter-individuelles de la similarité prosodique au moyen d'un test AXB.

2.2.3 Discussion : quelle congruence entre Tests AX et AXB ?

Ayant fait passer deux tests perceptifs de natures différentes sur un même set de données, la question que nous pouvons à présent nous poser est celle de la congruence de leurs résultats. Il serait en effet inquiétant que ces tests indiquent des tendances inverses, car les résultats des deux tests seraient alors invalides. Nous nous proposons donc d'observer dans un premier temps s'il y a une corrélation positive ou négative entre les scores du Test AX et le nombre de votes obtenus au Test AXB.

Bien que les imitations aient été évaluées par bloc lors du Test AXB, nous calculons tout de même la corrélation r de Pearson pour l'ensemble des résultats ; la Figure 35 montre le nuage de dispersion des données : r = .68 ($p.$ < 001, df = 70, t = 7.8159). Il y aurait donc une corrélation positive relativement forte entre les résultats obtenus par les imitations lors de ces deux tests (lorsque le score AX augmente, le nombre de votes reçus augmenterait également).

Si nous observons le nuage de dispersion, on remarque en effet que les points sont répartis dans une ellipse. Cependant, celle-ci est assez évasée : un bon nombre de points ayant

[73] Equivalent en parole du délit de faciès… Humour…

reçu un score bas au test AX ont tout de même reçu un grand nombre de votes au test AXB. Ceci est lié à la nature du test AXB et à ce que nous en disions dans la discussion précédente. Du fait de l'évaluation relative conduite dans le paradigme AXB et de notre design expérimental nous contraignant uniquement à des comparaisons intra-individuelles, certaines piètres imitations ont mécaniquement reçu un grand nombre de votes car les reproductions auxquelles elles étaient opposées dans les triplets AXB étaient encore moins convaincantes pour les auditeurs.

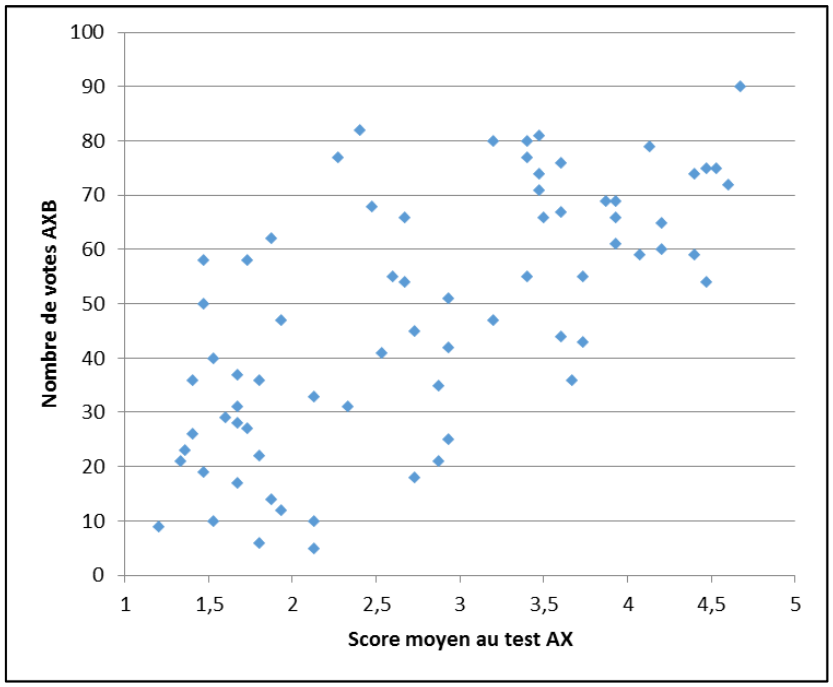


Figure 35 : Nuage de dispersion des 72 imitations évaluées perceptivement, en fonction du score moyen obtenu au Test AX en abscisses et du nombre de votes obtenus au test AXB en ordonnées

Ainsi, cette corrélation de .68 pour l'ensemble de ces données est biaisée car les imitations ont été évaluées toutes ensemble lors du test AX, et par blocs de phrases originales * locuteurs lors du test AXB. C'est pourquoi nous discutons en suivant les résultats par phrase et par locuteur du *CI*.

Phrases	**Corrélation**	**t value**	**df**	**p value**
72 imitations	0.68265	7.8159	70	3.988e-11

Tableau 18 : Corrélation de Pearson : résultats des tests AX & AXB pour l'ensemble des sujets/imitations

a. *Sujet = Sp1*

Sp1 était le locuteur dont les évaluations AX étaient les moins réussies : 61% (11/18) de ses productions ont obtenu un score inférieur à 2, alors que la moyenne de tous les scores au test AX est de 2,78. Si l'on regarde séparément les phrases *Bag* et les phrases *Boni* (Figure 36), on peut observer deux patterns similaires : dans les deux cas, un groupe de trois imitations se détache des autres au test AX. Il semble qu'elles reçoivent plus de vote au Test AXB. Cette tendance est plus flagrante pour les phrases *Bag* que pour les phrases *Boni*.

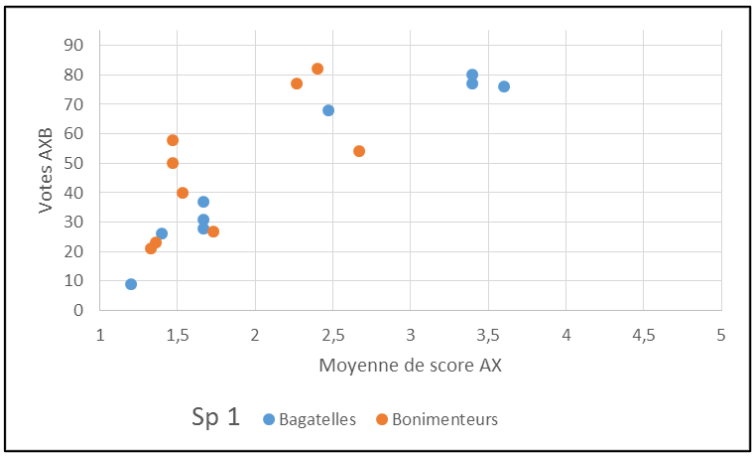


Figure 36 : Dispersion des scores de Sp1 aux tests AX et AXB

Les imitations de *Sp1* dont le score AX est compris entre 1 et 2 forment un cluster sur la gauche de la figure : ce sont les évaluations des phrases Boni qui ont un comportement plus chaotique en ce qui concerne les évaluations reçues.

Phrases	**Corrélation**	**t value**	**df**	**p value**
Ensemble	0.81975	5.7255	16	3.128e-05
Bag	0.96297	9.4499	7	3.103e-05
Boni	0.69213	2.5371	7	0.03883

Tableau 19 : Corrélation de Pearson : résultats de Sp1 aux tests AX & AXB.

Cette tendance se retrouve dans le taux de corrélation assez faible observé pour cet ensemble de phrase, par comparaison aux phrases Bag, où le taux de corrélation entre les résultats donnés par les deux tests est très élevé.

b. Sujet = Sp3

Sp3 avait été évalué au test AX comme un sujet plutôt performant : la moyenne de ses scores était supérieure à 3. D'après ce test AX, c'était également le sujet qui faisait preuve du plus de constance dans sa production imitative, étant donné que l'étendue de ses notes obtenues était la plus réduite.

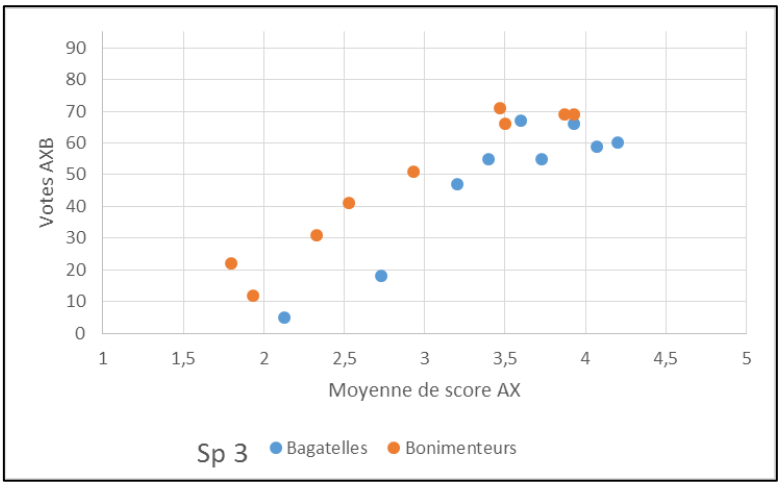


Figure 37 : Dispersion des scores de Sp3 aux tests AX et AXB

Par ailleurs, les distributions des votes obtenus par *Sp3* au test AXB présentaient des plateaux très marqués :

- Les phrases *Bag* ayant les mieux évaluées dénombraient respectivement 67, 66, 60, 59, 55 et 55 votes ;
- Les phrases *Boni* idoines ont quant à elles eu 71, 69, 69 et 66 votes.

Ce sont les phrases que nous retrouvons ici dans le cluster de points en haut à droite du nuage de la Figure 37, dont les scores AX sont compris dans une fourchette allant de 3,5 à 4,2. Ce resserement aurait rendu difficile pour les auditeurs du Test AXB, la tâche de différencier ces bonnes imitations.

Phrases	**Corrélation**	**t value**	**df**	**p value**
Ensemble	0.88418	7.571	16	1.125e-06
Bag	0.91914	6.1734	7	0.000457
Boni	0.967717	10.159	7	1.928e-05

Tableau 20 : Corrélation de Pearson : résultats de Sp3 aux tests AX & AXB

Ainsi, pour ce qui concerne les productions de *Sp3*, le test AX semble apporter un éclairage au résultat du test AXB. Par ailleurs, les corrélations observées entre les résultats des deux tests perceptifs (Table 20) indiquent une bonne adéquation entre les deux types de décisions –relative et absolue– fournies par les auditeurs de nos tests.

c. *Sujet = Sp5*

Sp5 a été le locuteur en moyenne le mieux évalué par les auditeurs du test AX (3,75 pour ses 18 productions). Ceci dit, c'est aussi le locuteur avec l'étendue de score la plus large, notamment dans le cas des phrases *Boni*. On retrouve dans le nuage de points de la Figure 38, quelques agglomérats de points qui illustrent l'idée selon laquelle un ensemble de phrases obtenant des scores proches au test AX recueilleraient des votes AXB à part équivalente (tendance une première fois repérée chez *Sp3*).

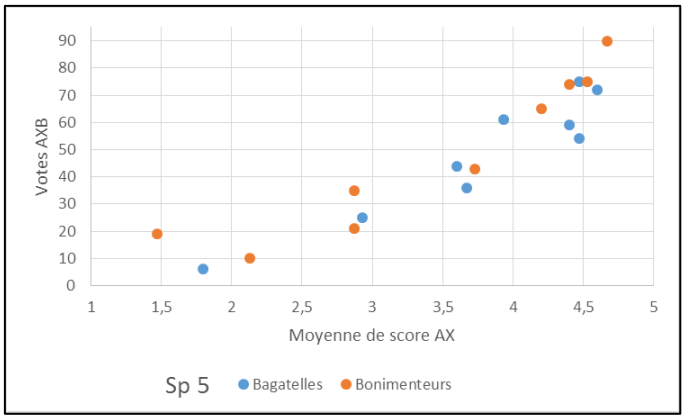


Figure 38 : Dispersion des scores de Sp5 aux tests AX et AXB

Parallèlement, les phrases *Boni* de *Sp5* étaient les seules à respecter la distribution théorique AXB que nous avions définie : l'étendue remarquable des scores AX de ces phrases pourrait impliquer que chacune constitue un degré différent d'imitation perçus par les juges.

Phrases	**Corrélation**	**t value**	**df**	**p value**
Ensemble	0.92312	9.6031	16	4.811e-08
Bag	0.94174	7.4078	7	0.00014
Boni	0.93315	6.8676	7	0.0002382

Tableau 21 : Corrélation de Pearson : résultats de Sp5 aux tests AX & AXB

Encore une fois, les taux de corrélation observés entre les résultats des deux tests laissent supposer qu'une imitation ayant eu un score élevé au test AX obtiendra un nombre de vote élevé au test AXB.

d. *Sujet = Sp7*

Les productions de *Sp7* n'avaient pas obtenu de très bonnes évaluations lors du test AX, ce locuteur avait donc été classé troisième sur quatre, avec une moyenne de 2,374 pour ses 18 productions. *Sp7* était toutefois parvenu à produire quelques bonnes imitations dans les phrase *Boni*, dont 3 productions avaient un score moyen supérieur à 3,5. Ceci dit, les évaluation de la majeure partie de ses productions des phrases *Bag* (6/9) ont reçu un score compris entre 1,5 et 2. Nous aurions dû attendre un nombre de votes AXB très homogènes pour ces productions.

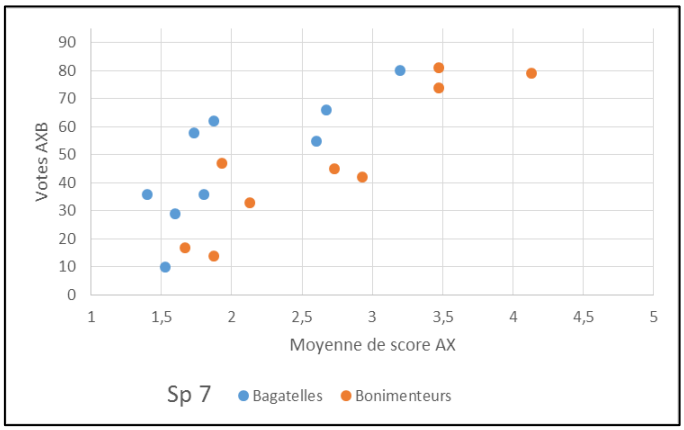


Figure 39 : Dispersion des scores de Sp7 aux tests AX et AXB

Il semble que ce ne soit pas le cas, comme l'illustre le nuage de points en Figure 39 : les auditeurs du test AXB ont clairement attribué plus de votes à ces productions, à score AX pourtant approchant.

Phrases	**Corrélation**	**t value**	**df**	**p value**
Ensemble	0.77624	4.9251	16	0.00015
Bag	0.78621	3.3662	7	0.01198
Boni	0.90424	5.6026	7	0.00081

Tableau 22 : Corrélation de Pearson : résultats de Sp7 aux tests AX & AXB

Cette dispersion des votes pour les phrases *Bag* est relevée par le taux de corrélation de Pearson de « seulement » r = 79.

En revanche, les phrases Boni présentant une plus grande étendue de scores AX, ont également des votes AXB sensiblement mieux répartis. Le taux de corrélation entre les deux tests est d'ailleurs plus satisfaisant, comme le montre la Table 22.

2.2.4 Synthèse

Nous avons décortiqué les résultats de deux tests de jugement perceptif de la similarité prosodique sur un même ensemble de données. Entre les résultats de ces deux tests, dont les systèmes de notation de la similarité étaient différents, nous retrouvons de manière quasi systématique une bonne adéquation, illustrée par des taux positifs de corrélation assez forts.

A la vue de ces résultats, nous pourrions avancer que:

- il est très probable qu'une imitation ayant obtenu un score moyen élevé à un test AX, obtiendra un grand nombre de votes à un test AXB, comparativement à une production ayant eu un score moyen moindre au même test AX.
- Il est également très probable qu'un groupe d'imitations ayant obtenu un grand nombre de votes proches à un test AXB, auront de même des scores moyens assez proches à un test AX.

Ainsi, il semble y avoir un lien logique entre les résultats de ces deux tests d'évaluation de la similarité perceptive.

Par conséquent, la nature de l'information apportée par ces tests quant aux phrases évaluées diffère en raison de la nature du système de notation employé dans les deux paradigmes :

a. Le test AX propose une évaluation absolue, où chaque phrase est évaluée indépendamment des autres phrases de l'échantillon. Ainsi, il offre la possibilité d'établir une hiérarchie globale entre des imitations dont les contenus segmentaux et les locuteurs imitateurs peuvent différer.
b. Le test AXB propose une évaluation relative, où chaque phrase est évaluée par comparaison avec une autre phrase. Etant donné cette nature du test, les phrases

opposées les unes aux autres doivent donc avoir le même locuteur et le même contenu segmental. En conséquence, le test AXB ne permet d'établir qu'une hiérarchie par degrés au sein d'un même groupe.

Une différence fondamentale entre test AX et AXB de jugement perceptif de la similarité réside donc dans le type de question auxquelles ces tests sont susceptibles d'apporter une réponse :

- Si l'on cherche à classer des phrases sur un continuum des imitations du moins bien imité au mieux imité, il vaut mieux choisir le test AX.
- Si l'on cherche seulement à savoir s'il y a des différences individuelles perçues dans les énoncés d'un groupe d'imitation, le test AXB donnera cette indication, sans évaluer la réussite des imitations.

En ce qui concerne ce protocole expérimental d'exploration des liens entre similarité prosodique perçue et mesurée, nous retiendrons essentiellement les résultats du test AX comme point de comparaison pour évaluer la pertinence des mesures objectives de la similarité. En effet, l'information apportée par le test AXB, par son caractère relatif, ne nous permettra pas de jauger la congruence entre les mesures objectives et les évaluations perceptives.

2.3 Mesures de la similarité prosodique : comparaison des contours intonatifs

Les tests perceptifs nous ont permis d'obtenir une première information sur la similarité prosodique perçue entre les 72 imitations du *CI* et les modèles correspondants du *CE*. Il s'agit à présent d'observer s'il y a une congruence entre les résultats donnés par les différentes mesures de la similarité que nous allons décrire ci-après et les résultats obtenus aux tests de jugement par les auditeurs décrits précédemment.

Calculer objectivement le degré d'imitation entre deux contours de *f0* revient à déterminer la distance physique entre ces contours, *i.e*. à savoir le degré de correspondance des formes des deux contours (*shape matching*). En d'autres termes, nous allons ici appliquer les principes de base du *shape matching* en faisant une transformation des formes à comparer, puis en calculant leur similarité au moyen d'une mesure (Veltkamp, 2001). L'étape de transformation du *shape matching* consiste simplement à rendre les formes comparables, pour pouvoir par la suite leur appliquer une ou des mesures de similarité.

En ce qui concerne les contours de *f0*, il importe que les transformations servent à la fois de normalisation des hauteurs (pour pouvoir comparer des voix aux registres différents) et des aspects temporels (afin que les formes aient la même longueur). En fonction du type de forme que nous considérons, la transformation et les mesures appliquées vont différer. Avant de les décrire une à une, nous les indiquons dans la Table 23.

	Similarité prosodique de deux contours prosodiques		
Type de *f0*	**Brute**		**Stylisation rectiligne**
Transformation	Dynamic Time Warping		Turning Function
Mesure	Norme L_2 pondérée	Coefficient Z_r	Norme L_2

Tableau 23 : Transformations et mesures de similarité utilisée en fonction de la stylisation de la *f0*.

Dans un premier temps, nous présenterons les mesures de similarité prosodique issues de formes de *f0* brutes. Cette étape de notre protocole constitue une réplication des approches de Hermes (1998b) et Rilliard *et al.* (2011) puisque nous reprenons pour notre étude :

- Les mesures proposées par Hermes
- La transformation utilisée par Rilliard *et al.* afin de rendre des contours de *f0* brutes comparables en normalisant leurs longueurs.

La seconde étape de notre protocole considère la *f0* stylisée de manière rectilinéaire. Elle provient de nos considérations liées à l'annotation phonologique des contours prosodiques. Nous remarquions lors des séances d'annotation qu'à patrons intonatifs similaires, il semble y avoir une différence de réalisation phonétique (localisation des événements et étendue du patron) entre deux énoncés. Nous voulions donc pouvoir « mesurer » la différence phonétique entre des patrons phonologiques similaires pour savoir estimer l'importance du détail phonétique dans la perception de la similarité prosodique.

2.3.1 Procédure de mesure de la similarité prosodique de *f0* brute

Les courbes de fréquence fondamentale sont définies par deux paramètres : le temps (en ms ou s) et la hauteur (en Hz). Pour chacune des 72 imitations étudiées, nous avons donc extrait le décours temporel de la hauteur de *f0* avec un taux d'échantillonnage de 1000 valeurs par secondes (un taux similaire à celui utilisé par Rilliard *et al.*, 2011) au moyen d'un simple script Praat. Ce grand nombre de valeurs obtenues permet de dessiner des courbes de *f0* qualifiables de « brutes », puisque l'on peut y observer les variations microprosodiques (qui reflètent l'influence de l'articulation des segments sur la courbe de *f0*). Nous donnons un exemple de ces courbes Figure 40.

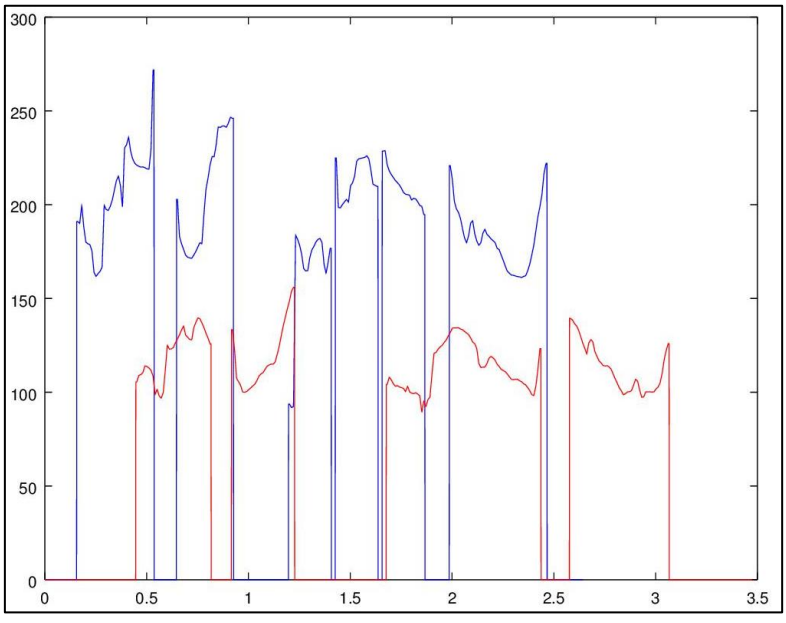


Figure 40 : Extraction de la *f0* de la phrase « les bagatelles et les balivernes saugrenues » dite par la locutrice de référence en bleu (féminin) et Sp3 en rouge (masculin). Temps en secondes en abscisses et hauteur en Hertz en ordonnées.

En observant cette figure, il apparaît clairement que la distance entre les deux courbes originales est particulièrement grande : non seulement la longueur des courbes diffère, mais le registre des deux locuteurs éloigne encore les deux contours. Il est donc requis de normaliser les aspects temporels des courbes ainsi que leurs hauteurs relatives avant de mesurer leur similarité.

L'ensemble de ces opérations a été réalisé dans la même routine écrite pour le logiciel Octave[74]. Nous détaillons en premier lieu la procédure de normalisation.

2.3.1.1 Normalisation et transformation des courbes

Pour contourner les différences de hauteur relative entre les locuteurs, nous avons divisé toutes les valeurs de *f0* de chaque contour par la valeur maximale de ces contours. Ainsi, chaque contour de *f0* a vu ses valeurs reportées dans une fourchette comprise entre 0 et 1, une échelle relative à la moyenne des locuteurs. Cette première transformation fait perdre l'échelle de la hauteur mais garde les proportions des courbes originales.

Même s'il existe d'autres méthodes pour rapporter les valeurs de *f0* de deux locuteurs dans échelles comparables (conversion en demi-tons, ou échelle ERB, par exemple), cette méthode nous a paru la plus simple et la moins coûteuse afin de pouvoir comparer les *f0* de deux locuteurs aux registres très différents.

Après avoir normalisé la hauteur des courbes, il faut également en normaliser les longueurs afin de pouvoir par la suite mesurer la distance entre chaque point de chaque courbe. Une possibilité pour normaliser la longueur des courbes pourrait consister à compter la différence de nombre de points entre les deux courbes puis à intercaler des points de manière linéaire (par exemple, un point sur deux). Cependant, en procédant à une interpolation linéaire, nous risquerions de décaler des événements comparables et donc d'obtenir en suivant une mesure de similarité faussée.

Pour contrebalancer cet effet de l'interpolation linéaire, Rilliard *et al.* (2011) proposent d'utiliser le *Dynamic Time Warping* (DTW). Le DTW, que nous avons évoqué au chapitre méthodologique de notre travail, est une technique d'interpolation non linéaire qui consiste à calculer le chemin optimal entre deux matrices. En d'autres termes, le DTW tente de retrouver

[74] Nous remercions chaudement Benjamin Boulbène et Julien Dupouy pour l'écriture de cette routine, palliant ainsi les lacunes de l'auteur.

dans les matrices des ensembles de données qui se ressemblent, les fait correspondre puis procède ensuite à une interpolation aux endroits où il manque des données.

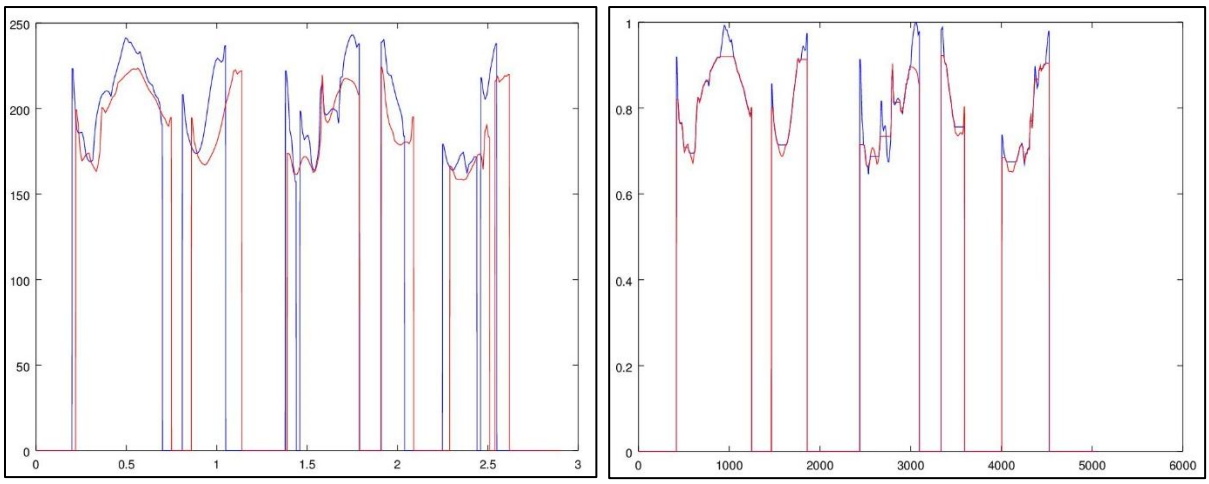


Figure 41 : Un exemple de DTW sur deux courbes de *f0*, extraites de la phrase « *Les bonimenteurs et la baratineurs fades* » dite par la locutrice modèle (en bleu) et par *Sp5* (en rouge). A gauche les *f0* originales, à droite les mêmes après DTW.

En l'appliquant sur des courbes de *f0*, le DTW va d'abord rechercher les pics, les creux et les blancs des deux formes à normaliser, puis aligner ces événements identiques. Ainsi, faire un DTW sur deux contours intonatifs consiste à faire un alignement tonal ainsi qu'une normalisation de la longueur de la courbe. La Figure 41 montre un exemple de courbes de *f0* avant DTW et après DTW. On peut y observer que les deux courbes de la partie gauche de la figure sont décalées ; dans la partie droite, tous les événements sont alignés. On peut aussi remarquer à certains endroits de la partie droite, l'interpolation effectuée par le DTW, au moyen des plateaux qui ne sont pas dans la figure originale.

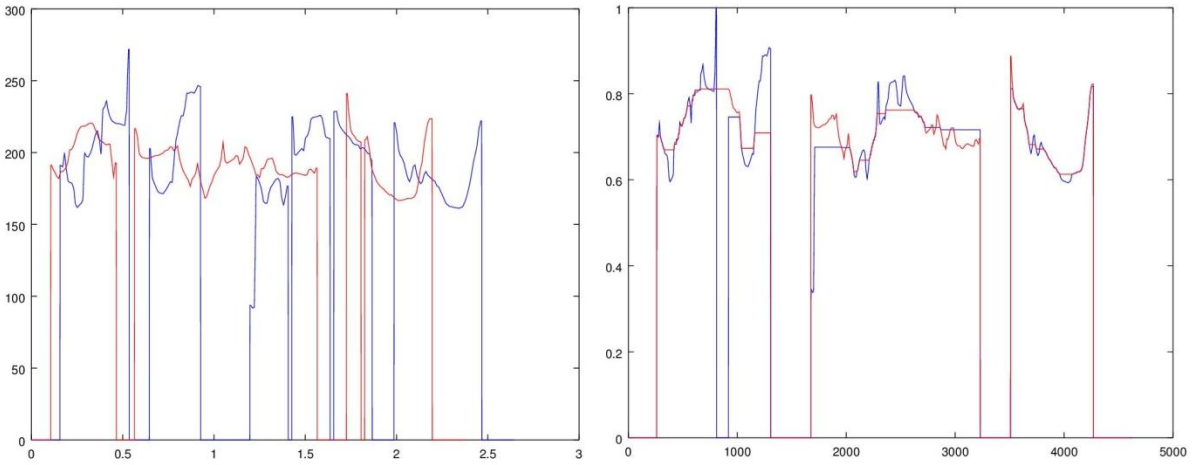


Figure 42 : DTW des courbes de *f0* issues de la phrase « *les bagatelles et les balivernes saugrenues* » dite par la locutrice modèle (en bleu) et par *Sp7* (en rouge)

Le DTW est particulièrement efficace si les formes originales se ressemblent, comme les courbes de la Figure 41. Cependant, s'il y a des différences notables entre les deux formes à aligner, le résultat donné par le DTW semble moins optimal, comme l'illustre la Figure 42. Sur la partie gauche de cette figure, nous pouvons remarquer que la courbe rouge ressemble peu à la courbe bleue. Il y manque notamment la pause et les proéminences sont moins marquées. La partie droite de la figure montre le résultat d'un DTW non optimal : les difficultés d'alignement se repèrent au moyen des plateaux dans les deux premiers segments. Les derniers morceaux de courbes ont en revanche été alignés de manière satisfaisante.

2.3.1.2 Mesures appliquées aux courbes normalisées

Suite à la transformation des contours à comparer, nous leur avons appliqué deux mesures de similarité initialement proposées par Hermes (Hermes, 1998b).

Nous avons ici repris sa méthode où $w(t)$ représente le décours temporel du facteur de pondération (soit, la somme du *subharmonic sumspectrum* du signal modèle), W l'intégrale de $w(t)$ de 0 jusqu'à T (T étant la durée totale de l'énoncé de référence), $f1$ et $f2$ les deux contours intonatifs à comparer.

A partir de ces facteurs, Hermes propose le calcul de :

- La norme L_2 (ou différence de la moyenne des moindres carrés) :

$$L_2 = \left\{ \frac{1}{W} \int_0^T w(t) \left| f_1(t) - f_2(t) \right|^2 dt \right\}$$

- Du coefficient de corrélation r :

$$r = \frac{\frac{1}{W} \int_0^T w(t) f_1(t) f_2(t) dt}{\sqrt{\left\{ \frac{1}{W} \int_0^T w(t) \left| f_1(t) \right|^2 dt \, \frac{1}{W} \int_0^T w(t) \left| f_2(t) \right|^2 dt \right\}}}$$

qu'il est nécessaire de transformer en Z de Fischer (ci-après Z_r) afin de pouvoir comparer les différents r :

$$Z_{f_1 f_2} = \frac{1}{2} ln \frac{1 + r_{f_1 f_2}}{1 - r_{f_1 f_2}}$$

Des courbes ainsi traitées, nous retirons deux estimations différentes de la similarité prosodique : L_2 qui mesure les changements rapides dans les contours de *f0* et Z_r qui est une mesure plus holistique de la similarité des formes.

- L_2 est une mesure de **dissimilarité,** (plus l'indice L_2 a une valeur élevée, plus la dissimilarité entre les formes est grande). D'après Hermes (1998b), L_2 estime la distance perceptive entre deux contours, où un poids quadratiquement plus élevé est donnée aux endroits où les différences entre les formes sont plus grandes.
- Z_r est une mesure de similarité (plus l'indice Z_r est élevé, plus la similarité entre les formes évaluées est grande). Z_r exprime dans quelle mesure un contour intonatif peut être obtenu à partir d'un autre au moyen d'une transformation linéaire. En d'autres termes, Z_r exprime la distance entre la forme des contours.

2.3.1.3 Résultats des mesures L_2 et Z_r

Nous comparons ci-après les résultats obtenus pour la comparaison des 72 imitations avec leurs modèles, au moyen des deux mesures que nous venons de décrire. Nous attendons de cette comparaison, une bonne corrélation entre les mesures. En effet, il serait particulièrement incohérent que deux mesures différentes de dis/similarité appliquées aux mêmes objets donnent des résultats chaotiques. La Figure 43 montre les scores obtenus par les imitations en fonction de Z_r en abscisses et L_2 en ordonnées.

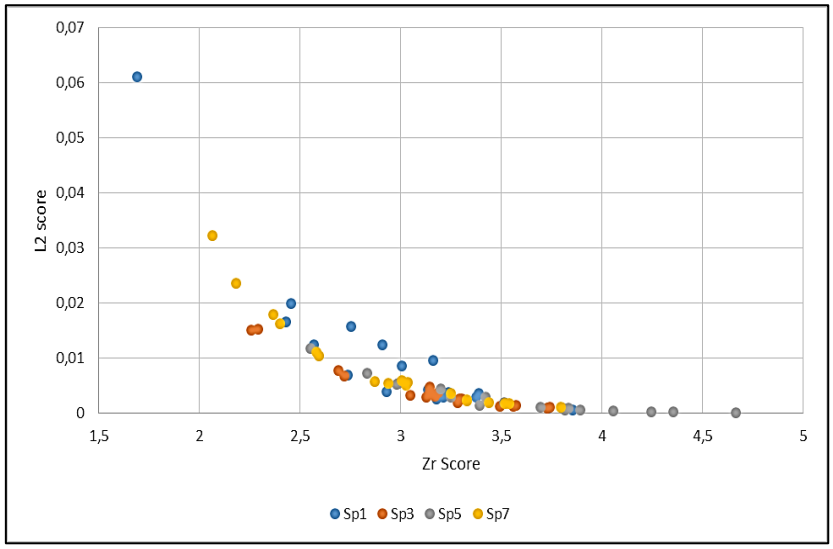


Figure 43 : Nuage de point des scores des mesures issues de la *f0* brute. Z_r (abscisses) et L_2 (ordonnées). Rappelons que L_2 est une mesure de dissimilarité et Z_r une mesure de similarité. Les différents imitateurs peuvent être repérés par différentes couleurs de points.

Malgré quelques *outliers*, la dispersion de ces points semble indiquer une relation assez forte entre les deux mesures. Un Z_r élevé semblent impliquer un L_2 très bas et inversement. Par ailleurs, nous pouvons noter qu'en deçà d'un certain seuil sur l'indice Z_r, les scores L_2 augmentent drastiquement. Ceci pourrait être dû à la différence de nature des deux mesures : L_2 attribue un poids quadratiquement plus important aux écarts les plus grands entre modèle et imitation. Cette donnée pourrait expliquer que la relation entre Z_r et L_2 ne soient pas simplement linéaire.

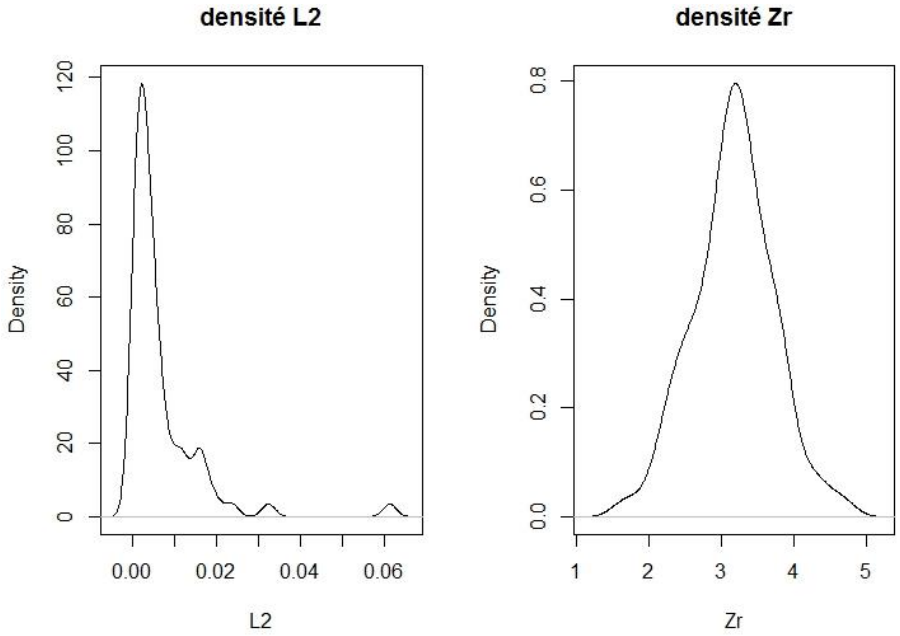


Figure 44 Densité des scores de mesures de la similarité L_2 et Z_r pour 72 imitations.

Avant de calculer notre corrélation entre L_2 et Z_r, nous proposons d'observer la répartition de leurs données au moyen d'une courbe de densité (Figure 44). Malgré une symétrie apparente de la courbe pour Z_r, les données de L_2 présentent une asymétrie remarquable. Il semble que la répartition des données ne suive pas une courbe gaussienne.

Bien que le non-respect du critère de normalité ne soit pas une entorse trop importante pour la passation d'un test paramétrique de corrélation (McDonald, 2014), nous avons préféré choisir le test de corrélation de Spearman. Ce dernier évalue la corrélation sur la base des rangs des événements, plutôt que sur leurs valeurs réelles. Nous donnons cependant, à titre indicatif, le résultat donné par le test de Pearson. Tous les tests de corrélation décrits dans ce chapitre ont été obtenus au moyen du logiciel `R`, avec la commande `cor.test` en spécifiant, en fonction du test, la méthode à appliquer (`spearman` ou `pearson`).

`Spearman's rank correlation rho`
`S = 122390`
`p-value < 2.2e-16`
 	
`sample estimates: rho = -0.9678436`
	`Pearson's product-moment correlation`
	`t = -10.326, df = 70`
	`p-value = 1.037e-15`
	`95 percent confidence interval:`
	`-0.8547699 -0.6649830`
	`sample estimates: cor = -0.776957`

Tableau 24 : Résultats des tests de corrélation des mesures issues de la *f0* brute (L_2 et Z_r)

Les résultats donnés par ces tests indiquent une corrélation négative assez forte entre les deux mesures. Le caractère négatif de la corrélation était attendu, dans la mesure où ces deux indices s'interprètent à l'inverse, L_2 étant une mesure de dissimilarité et Z_r une mesure de similarité. De même, le taux élevé de corrélation est satisfaisant : il montre qu'il y a une bonne cohérence entre les deux mesures, malgré quelques écarts qui sont probablement dues au fait que L_2 et Z_r ne traitent pas les grandes différences de la même manière.

2.3.1.4 Test AX vs. mesure Z_r

Après nous être assurés que L_2 et Z_r proposent des mesures concordantes, il nous faut à présent comparer leurs mesures aux évaluations perceptives fournies par les tests AX. Nous représentons dans la Figure 45 les scores obtenus par les 72 imitations au moyen de l'indice Z_r et des évaluations du test AX.

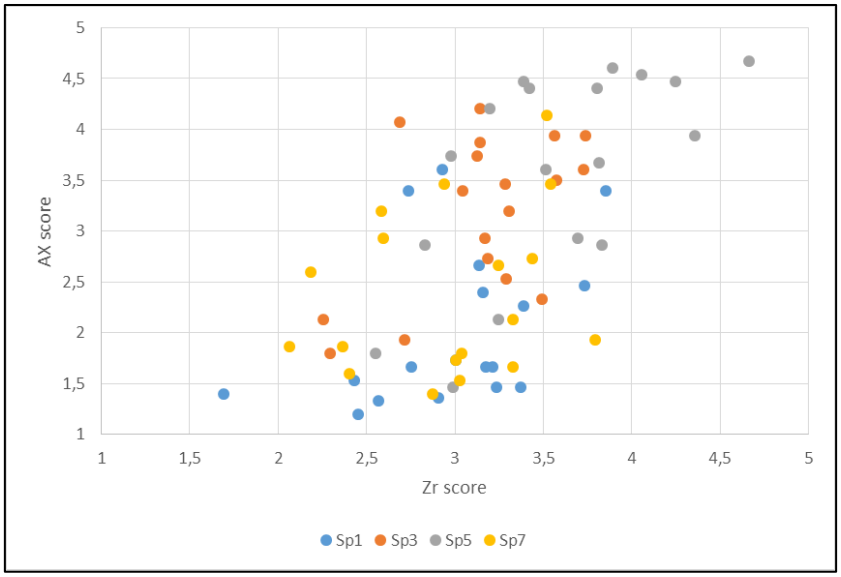


Figure 45 : Nuage de points des scores *Zr* et AXpour 72 imitations du CI. *Zr* et AX évaluant la similarité, les scores élevés indiquent les meilleures imitations.

L'inspection visuelle de ce graphique montre une certaine dispersion des points de ce nuage. Sa forme elliptique suggère cependant une corrélation positive modérée entre l'évaluation perceptive AX et la mesure Z_r. Notons toutefois que l'intervalle de score Z_r [3 :3,5] semble le plus problématique en ce qui concerne sa concordance avec AX. De même, certaines productions de Sp1 (en bleu) et Sp7 reçoivent de piètres évaluations AX, malgré un score Z_r plutôt élevé.

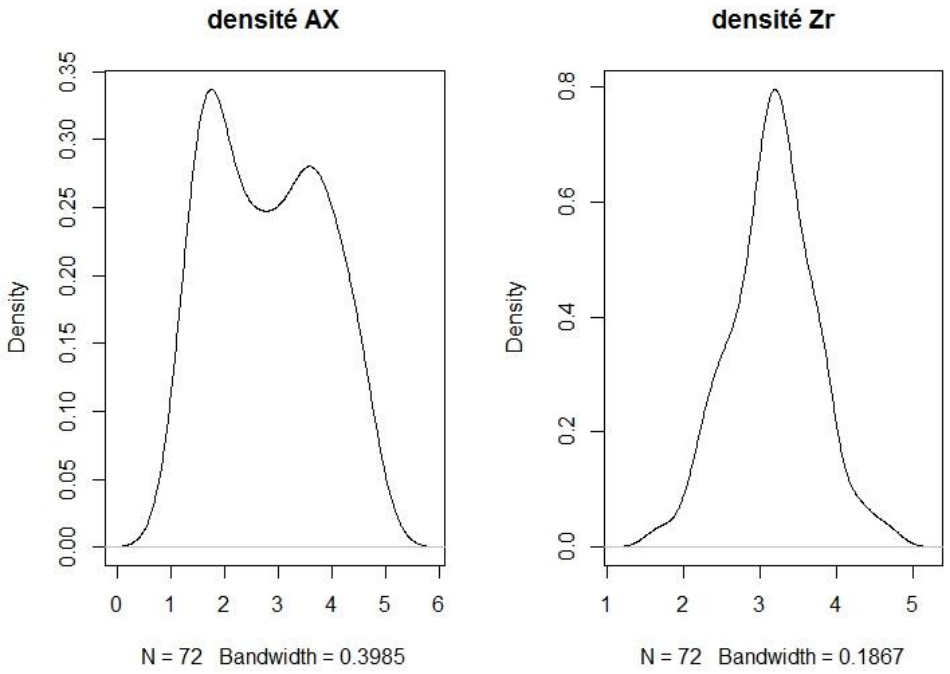


Figure 46 : Courbes de densité des scores AX et *Zr*.

La répartition des données pour les scores AX et Z_r est non normale, comme nous pouvons l'observer sur la Figure 46. La courbe des scores AX présente deux modes et celles des scores Z_r est particulièrement resserrée.

De même que précédemment, nous préférerons utiliser un test non paramétrique de corrélation (Spearman), mais présenterons également, à titre indicatif, le résultat donné par le test de Pearson.

Spearman's rank correlation rho S = 27862 p-value = 7.966e-07 sample estimates: rho = 0.5520291	Pearson's product-moment correlation t = 5.879, df = 70 p-value = 1.279e-07 95 percent confidence interval: 0.3960083 0.7117934 sample estimates: cor = 0.5749311

Tableau 25 : Résultats des tests de corrélation entre AX et *Zr*

Les taux de corrélation observés indiquent une corrélation positive modérée entre les scores perceptifs AX et la mesure $Z_{r.}$ Ainsi, la mesure Z_r semble être un candidat à considérer pour évaluer automatiquement la similarité prosodique entre une imitation et son modèle.

2.3.1.5 Test AX vs. mesure L_2

Ayant testé le lien entre les tests perceptifs et notre première mesure automatique (Z_r), il convient à présent d'adopter la même approche pour la mesure L_2. Les Figures 47a & 47b montrent les scores obtenus par les 72 imitations au test AX (ordonnées) et par la mesure L_2 (abscisses).

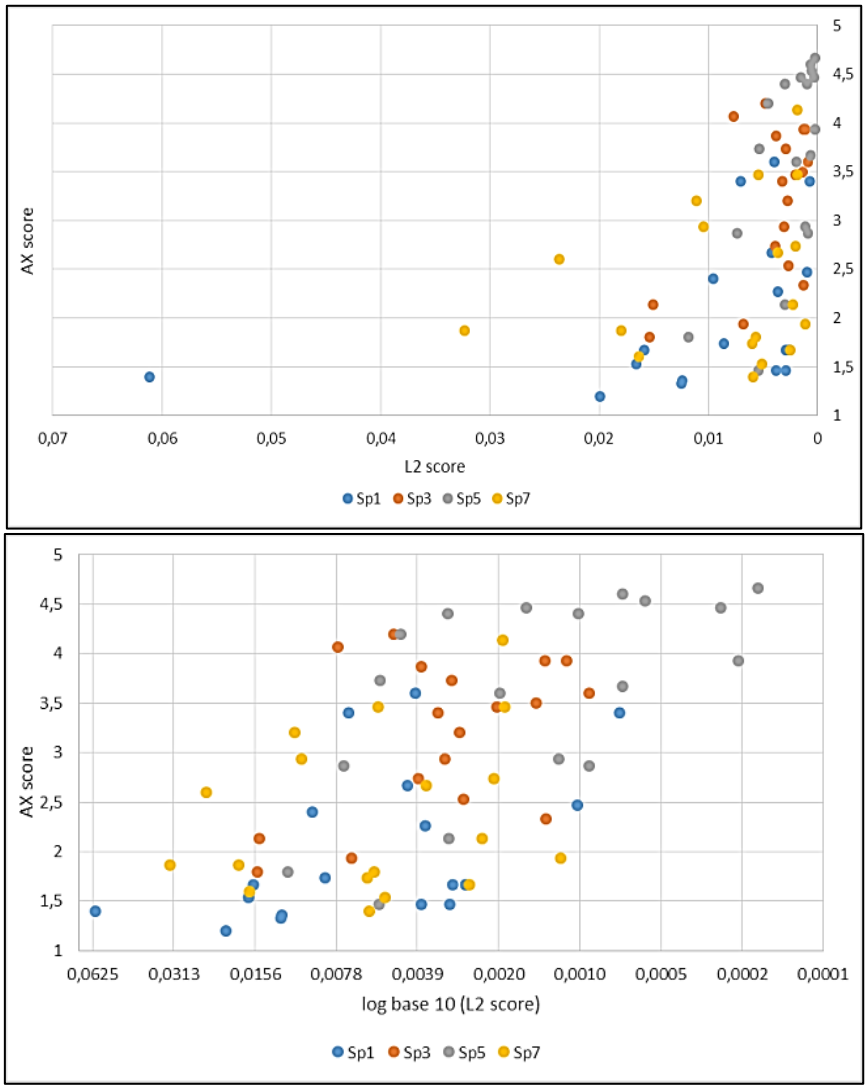


Figure 47a & 47b : Nuages de points des scores *L2* et AX.Notons que les valeurs de l'axe des abscisses sont présentées en ordre inverse. En haut (47a), les valeurs originales de L_2, en bas (47b) les valeurs de L_2 ont été converties en log base 10.

La répartition des valeurs de L_2 (Figure 47a) est particulièrement asymétrique. Par ailleurs, la majeure partie des points est concentrée dans des valeurs comprises entre 0,01 et 0. Afin de mieux saisir visuellement une éventuelle tendance pouvant se dégager de ces données, nous avons converti les valeurs de L_2 en log base 10 (Figure 47b). Nous retrouvons

alors une dispersion elliptique des données en fonction de L_2 et AX laissant suggérer une corrélation moyennement forte entre les deux indices.

Contrairement à ce que nous avons pu constater précédemment (Figure 45) dans la relation entre Z_r et AX, les points des productions de Sp1 (en bleu) semblent mieux répartis.

En ce qui concerne les taux de corrélation en mesure L_2 et score AX, nous avons encore une fois préféré choisir un test non paramétrique (Spearman), dont le résultat est le même aussi bien avec les valeurs originales de L_2 qu'avec ses valeurs converties en log base 10. Ceci dit, nous donnerons également les résultats du test de Pearson qui subit l'influence de cette conversion.

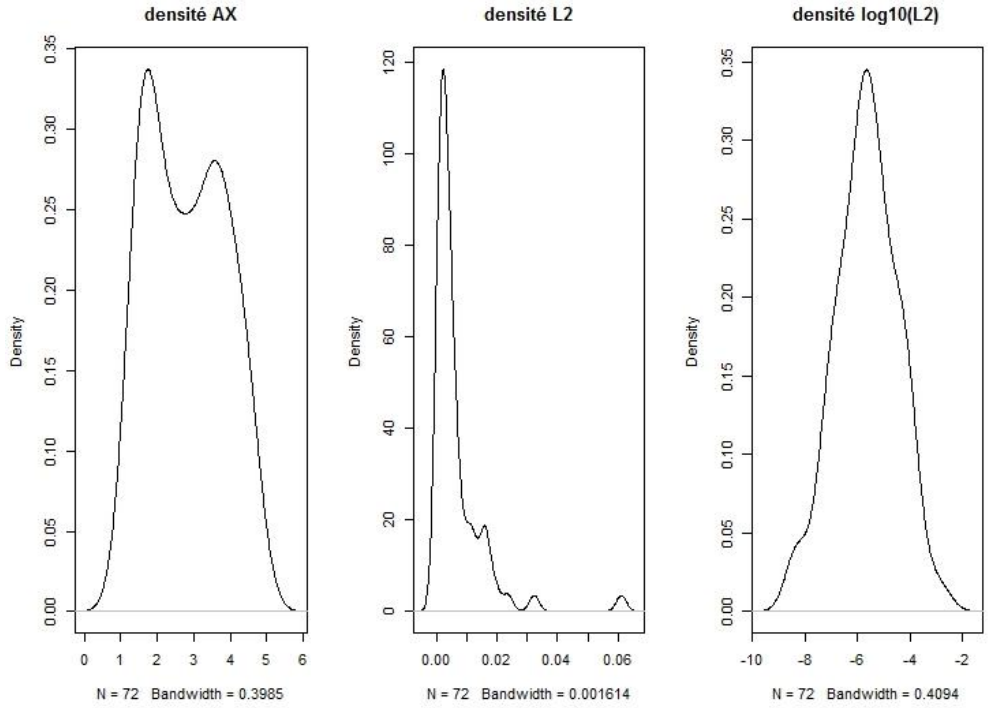


Figure 48 : Courbe de densité des scores AX, L_2 et log base 10 (L_2)

Encore une fois, la répartition de ces données ne suit pas la courbe gaussienne. Ceci nous conforte dans notre choix du test non paramétrique de corrélation.

Spearman	Pearson (L_2)	Pearson (Log10 de L_2)
S = 98854 p-value = 9.831e-08 sample estimates: rho = -0.5893948	t = -4.2371, df = 70 p-value = 6.785e-05 95 percent confidence interval: -0.6187056 -0.2458597 sample estimates: cor = -0.4517949	t = -6.5083, df = 70 p-value = 9.716e-09 95 percent confidence interval: -0.7403542 -0.4457300 sample estimates: cor = -0.6139963

Tableau 26 : Résultats des tests de corrélation entre *L2* et AX

Les taux de corrélation indiquent un lien négatif entre les scores AX (évaluation de la similarité) et les mesures L_2 (mesure de la dissimilarité). La force du lien entre AX et L_2 semble légèrement plus élevée que le lien entre AX et Z_r. Cependant, la différence entre les deux est très faible.

2.3.1.6 Résultats par Sujets

Lors du test AX, il était ressorti une hiérarchie entre les sujets. *Sp5* avait obtenu les meilleures évaluations, suivi d'assez près par *Sp3*. *Sp7*, puis *Sp1* fermaient la marche. Les quatre cadrans de la Figure 49 montrent de manière croisée les résultats des mesures L_2 et Z_r pour chacun de ces sujets. La taille des bulles indique le score moyen obtenu au test AX : plus une bulle est grosse, meilleur est le score.

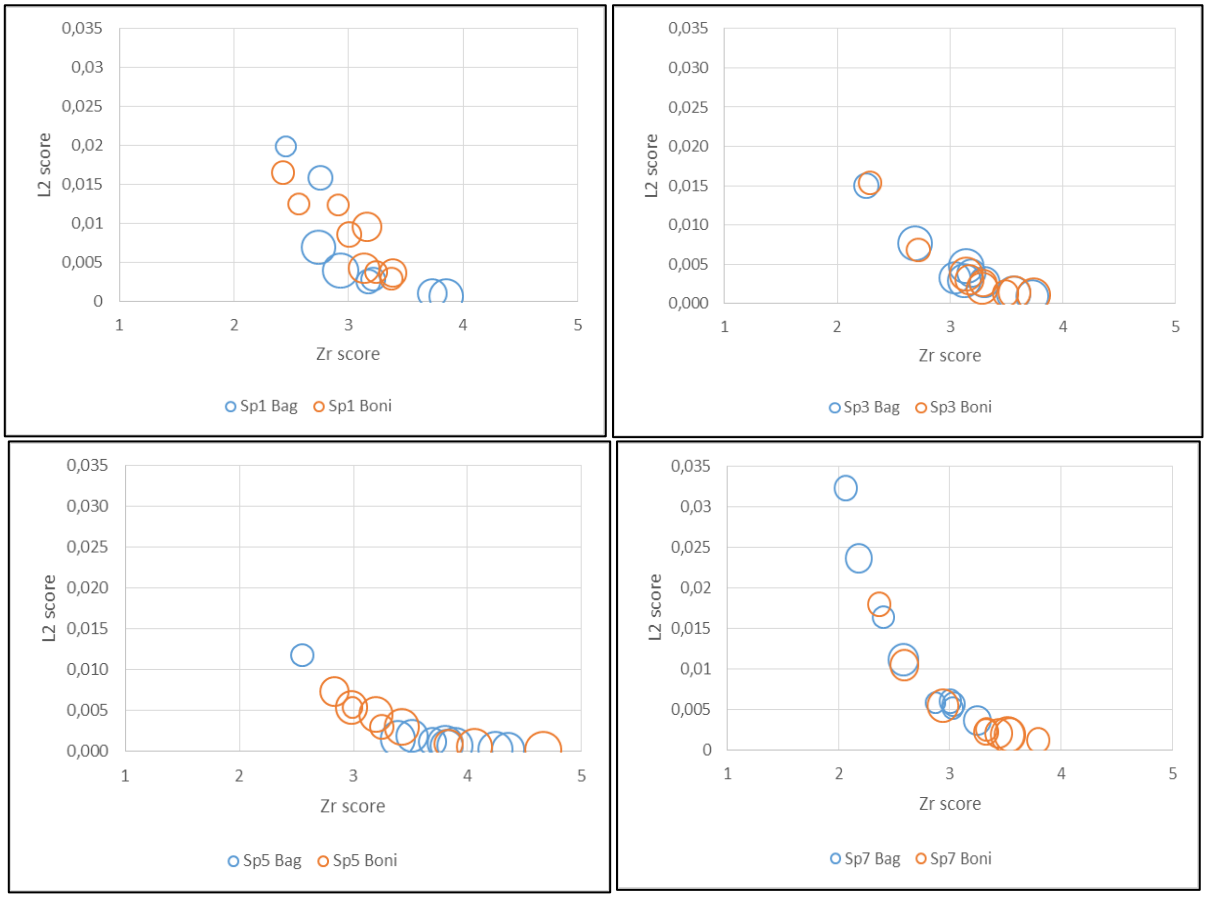


Figure 49 : Nuage de point par sujet. En abscisses, le score $Z_{r,}$; en ordonnées le score L_2. La taille des bulles est proportionnelle au score du test AX.

Notons qu'une production de *Sp1* a été enlevée de cette représentation afin de pouvoir conserver la même étendue en ordonnées pour tous les sujets.

Il semble que nous retrouvions ici la même hiérarchie dans les résultats obtenus au moyen des mesures L_2 et Z_r : *Sp5* obtient les meilleurs scores aux deux indices, les productions de *Sp3* sont également évaluée comme bonnes par les deux indices. *Sp1* & *Sp7* présentent des résultats plus étendus et évalués comme moins bon.

Il semble qu'en dessous de 3 sur l'indice Z_r, les scores L_2 tendent à augmenter rapidement et dépassent alors 0,005.

La Figure 50 représente les classements obtenus par les productions des 4 sujets en fonction des évaluations AX et des mesures L_2 et Z_r.

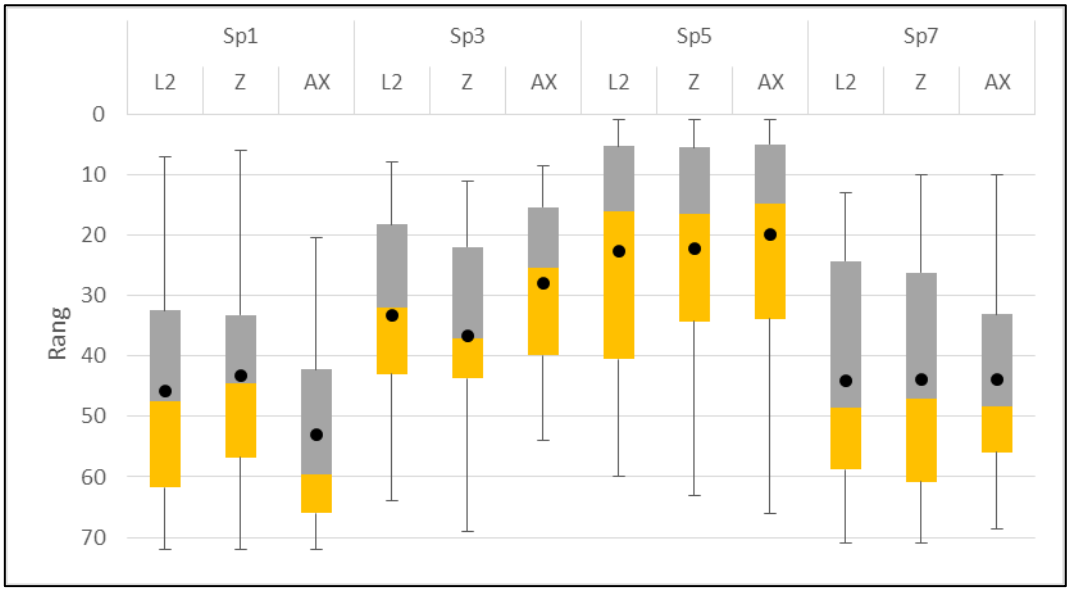


Figure 50 : Répartition interquartile des rangs obtenus par les sujets au moyen de 3 méthodes d'évaluations. Les points noirs représentent le rang moyen pour chacun des classements.

Cette figure souligne qu'il y a une bonne homogénéité entre les différentes méthodes de classement des sujets. Les auditeurs humains comme les mesures L_2 et Z_r ont établi une hiérarchie similaire en ce qui concerne la réussite des sujets à la tâche d'imitation.

Ceci étant dit, il semble que les auditeurs humains aient eu un avis un peu plus tranché que l'évaluation automatique :

- *Sp1* obtient de moins bonnes évaluations perceptives
- *Sp3* obtient quant à lui des scores un peu meilleurs

2.3.1.7 Discussion

Nous avons comparé des mesures de la similarité prosodique issues de *f0* brutes avec les résultats de tests perceptifs au moyen de tests de corrélation. Ceux-ci suggèrent une congruence modérément forte entre les mesures automatiques (L_2 et Z_r) et les évaluations perceptives (AX) de la similarité prosodique des 72 imitations de ce corpus.

De manière générale, l'inspection des rangs obtenus par les productions des sujets souligne la cohérence entre ces mesures et ces évaluations, puisque une hiérarchie similaire entre les sujets se dégage de ces différentes méthodes d'évaluation/mesure.

La corrélation est légèrement plus forte entre L_2 et AX qu'entre Z_r et AX :

- Z_r évalue le coût de transformation d'une forme en une autre, soit une simple mesure de distance.
- L_2 évalue les distances entre les deux formes et attribue un poids plus élevé aux grandes distances. En d'autres termes, L_2 tente de saisir les différences que l'on pourrait repérer perceptivement

Conceptuellement, L_2 semble donc plus approprié que Z_r pour évaluer la similarité prosodique tel que le feraient des auditeurs. Ceci étant dit, la différence entre les deux taux de corrélation est assez minime.

Par ailleurs, ces taux de corrélation relativement moyens pourraient refléter une différence fondamentale entre les procédures de mesure que nous venons de tester et ce qui a été demandé aux auditeurs lors de l'évaluation perceptive.

En effet, lors des évaluations AX, il était demandé aux auditeurs d'évaluer la similarité prosodique en se fondant sur la mélodie et le rythme. Or, pour effectuer les mesure L_2 et Z_r ce dernier aspect est probablement annihilé par la transformation préalable des courbes de *f0*.

De fait, les objets à comparer n'ayant pas la même durée (contrairement à ceux d'Hermes, (1998b) nous étions contraints de procéder à leur transformation simplement pour les rendre comparables. La procédure de *Dynamic Time Warping* déformant les contours originaux pour les aligner avant interpolation, nous pouvons alors considérer que les aspects rythmiques (en tant que décours temporel des tons) contenus dans les courbes originales ont été oblitérés. Ce faisant, les mesures L_2 et Z_r proposent une mesure partielle de la similarité prosodique, puisque le rythme n'y est pas pris en compte.

Ainsi, il semble y avoir un décalage entre la consigne donnée pour l'évaluation perceptive et les mesures utilisées par ailleurs. Cette différence pourrait expliquer les taux de corrélation modérément forts relevés précédemment puisque l'information sur la similarité prosodique donnée par L_2 et Z_r se limite à la configuration globale des formes.

Par conséquent, L_2 et Z_r sont des candidats viables pour obtenir une information sur la similarité prosodique entre une imitation et son modèle si l'expérimentateur accorde peu d'importance aux aspects rythmiques pour privilégier la configuration tonale des contours évalués.

En ce qui nous concerne, il nous faut pallier les limitations que nous venons de relever. En effet, les aspects rythmiques ont une importance particulière dans l'enseignement

de l'intonation : les logatomes produits par l'enseignant doivent évoquer au mieux l'énoncé original et plus particulièrement le rythme qui est premier en prosodie (Astésano, 2001; Billières, 1988, 2002; Di Cristo & Hirst, 1993).

2.3.2 Procédure de mesure de la similarité prosodique issue de *f0* stylisées

Pour contourner les problèmes liés à la déformation temporelle de la courbe de *f0* brute lors de la mesure de la similarité prosodique, il nous a paru approprié de tester une méthode issue de la géométrie. Nous avons donné au chapitre précédent les grandes lignes de cette méthode lorsque nous évoquions le *shape matching* et les *T-Function* au moyen d'exemples de formes simples. Nous n'avions par contre pas donné d'exemple de la méthode lorsqu'elle est appliquée à des courbes de *f0*.

En suivant, nous présenterons donc les traitements effectués sur les stimuli originaux afin de pouvoir leur appliquer la fonction de courbure[75].

2.3.2.1 Un exemple de stylisation puis transformation d'une courbe de f0 par la T-Function

Afin de pouvoir appliquer une *T-Function*, il est nécessaire d'avoir deux lignes polygonales. Celles-ci consistent en une série de points reliés par des segments. Nous montrons en Figure 51 une phrase originale dont une annotation tonale a été réalisée.

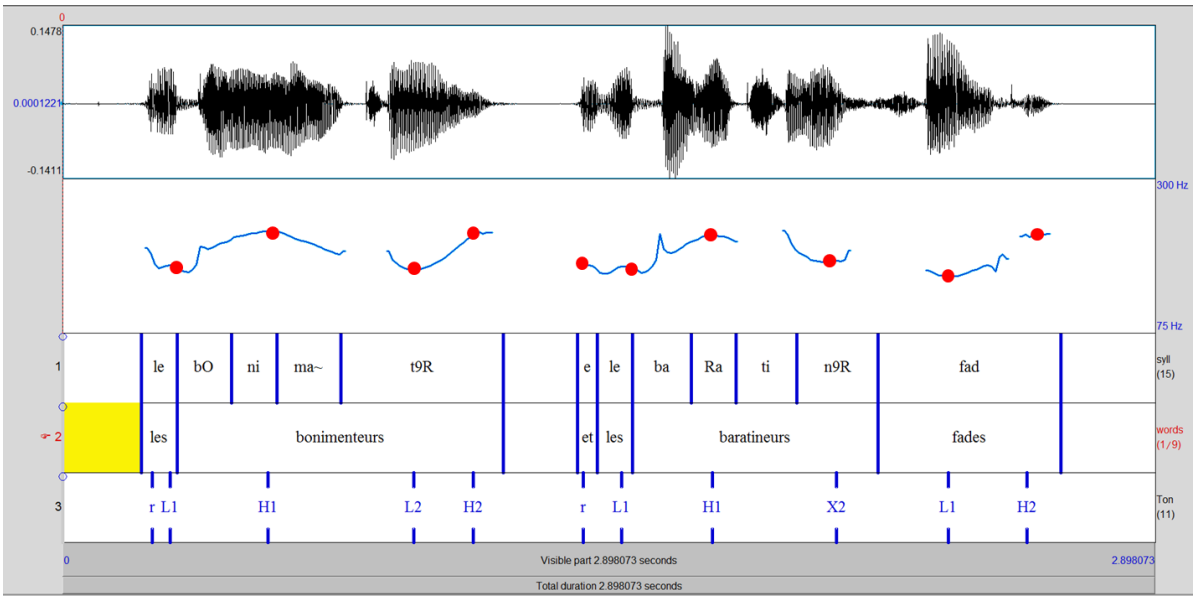


Figure 51 : Phrase « Les bonimenteurs et les baratineurs fades » dite par Sp5. Les points relevés pour la *T-Function* sont notés par les marqueurs rouges. Tier 1 : annotation syllabique,, tier 2 annotation lexicale, tier 3 annotation tonale.

Suite à l'annotation des phrases, les valeurs associées (temps et hauteur) aux différents points ont été relevées au moyen d'un script PRAAT et stockées dans des fichiers individuels. Ces

[75] Nous remercions Kevin Lepan, qui a effectué son stage de Master 2 sous notre direction, durant lequel il a adapté et implémenté les fonctions de courbures.

fichiers indiquent donc les coordonnées des points de chaque phrase, entre lesquels sont tirés les segments pour la stylisation.

La Figure 52 montre dans sa partie haute la stylisation de la *f0* de la phrase « *Les bonimenteurs et les baratineurs fades* » dite par *Sp5* (Figure 51) au moyen de la polyligne rouge. La polyligne bleue représente la stylisation de la *f0* du modèle.

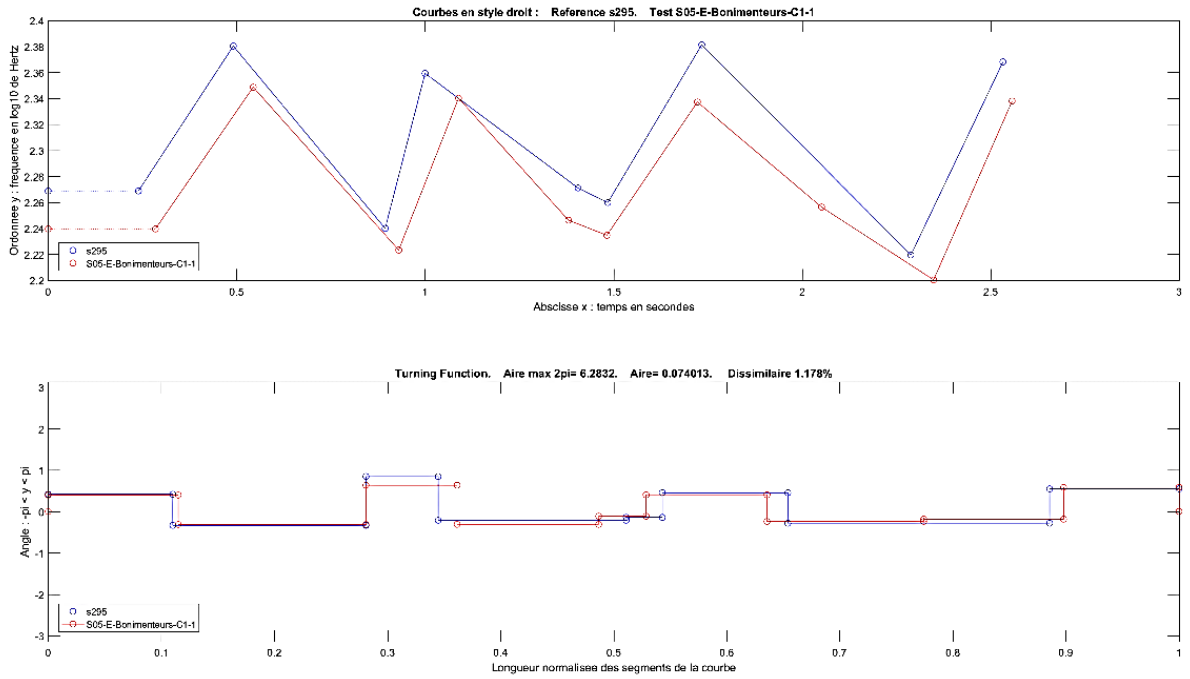


Figure 52 : Stylisation et *T-Function* de la fondamentale de la phrase « Les bonimenteurs et les baratineurs fades », dite par Sp5 et le modèle.

Les courbes stylisées sont transformées par la *T-Function* dans la partie basse de la figure. Rappelons que les segments verticaux représentent l'écartement des angles des polylignes originales ; les segments horizontaux rapportent la longueur de chaque segment des polylignes, sur une longueur totale ramenée à 1.

Ainsi, au lieu de forcer l'alignement entre les polylignes en les déformant, la *T-Function* conserve leur proportion en les transformant dans un même espace.

Finalement, le calcul de l'aire entre les courbes transformées indique la distance entre celles-ci. Ce dernier calcul est effectué au moyen de la norme L_2.

2.3.2.2 Le problème de l'échelle

Dans la mesure où la *T-Function* est une fonction qui se base sur une représentation des points dans l'espace, le choix de l'échelle d'unité des abscisses et ordonnées a un impact majeur sur l'output de la fonction.

En effet, s'il y a une différence d'ordre de grandeur entre les deux axes, les angles (et leur transformation) n'apporteront aucune information sur la configuration de la forme du contour, car les différences seront peu perceptibles. Nous présentons ce problème dans la Figure 53.

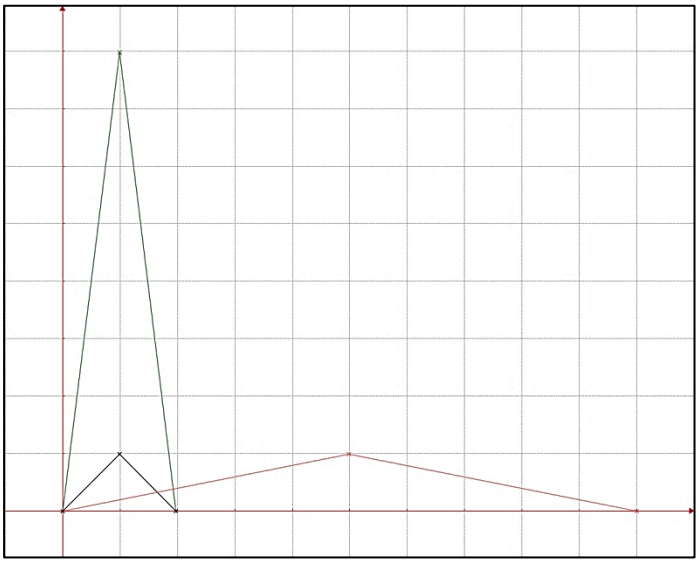


Figure 53 : Problèmes de proportion et représentation des formes. Le V_{noir} a une proportion de 1x:1y, le V_{vert} une proportion de 1x:8y, le V_{rose} une proportion de 5x:1y

On peut observer trois formes sur cette figure. Nous n'y indiquons pas d'échelle, mais nous avons changé le rapport de proportion des abscisses ou des ordonnées de la de V_{vert} et V_{rose}. On peut alors remarquer que le changement de proportion (soit, celui de l'unité) sur l'axe des *x* ou sur l'axe des *y* a une incidence particulière sur la forme représentée. L'angle de V_{vert} est particulièrement resserré, tandis que celui de V_{rose} est très large..

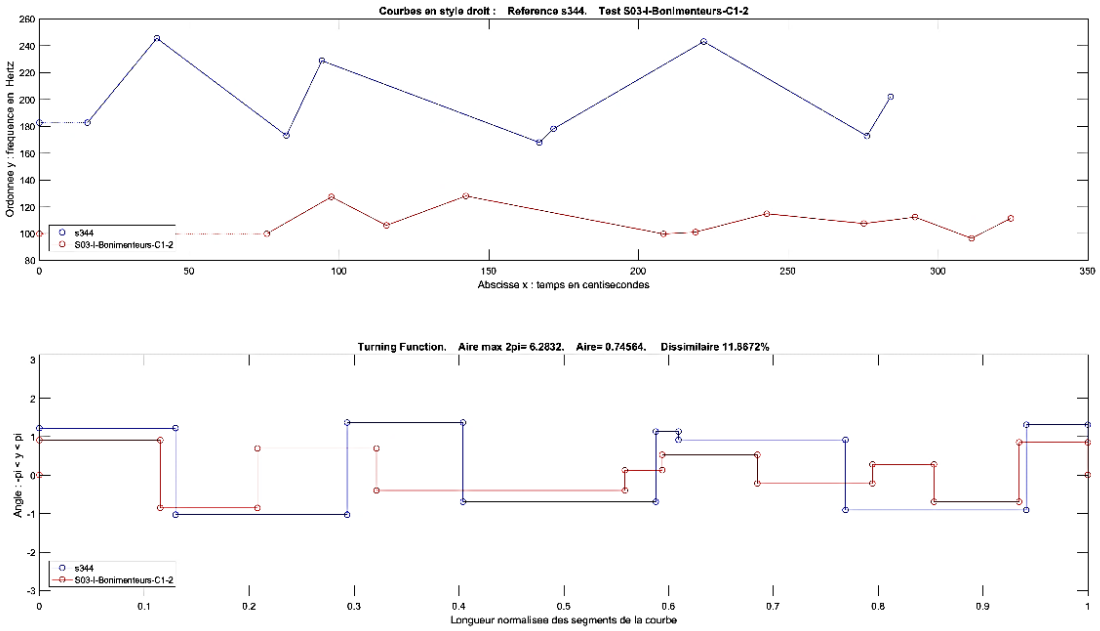
Ainsi, pour éviter des déséquilibres dans la représentation des *f0* stylisées, il sera nécessaire de convertir les valeurs du temps et de la hauteur dans des échelles ayant un rapport de proportion assez équilibré.

En ce qui concerne cette étude, nous avons étudié deux alternatives qui ont un rapport présentant ce rapport de proportion harmonieux :

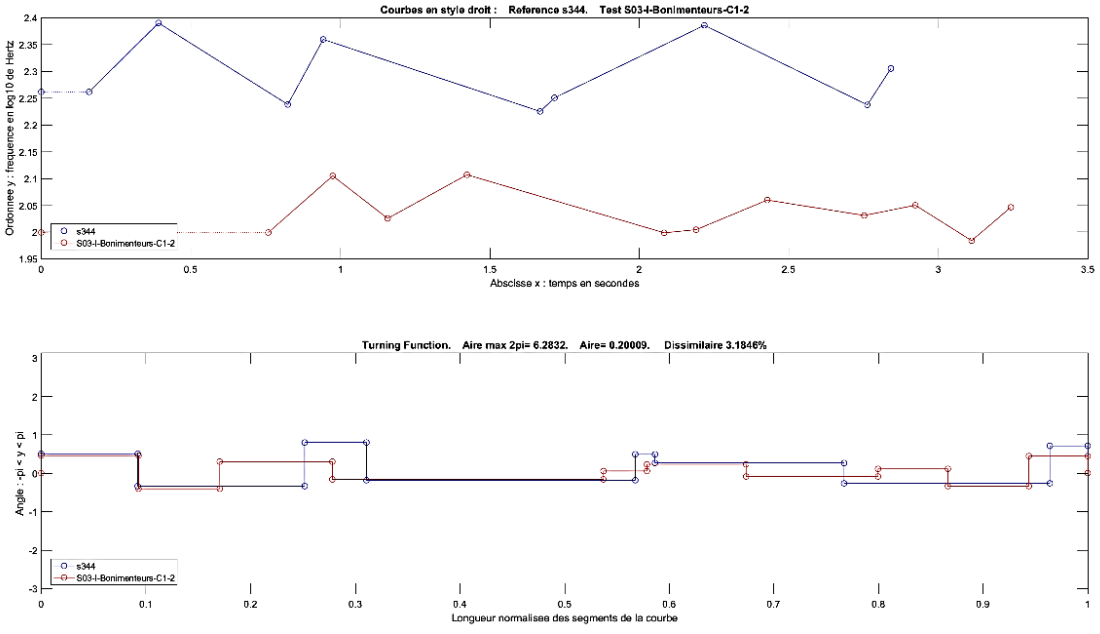
- Temps en centisecondes et hauteur en Hertz
- Temps en secondes, hauteur en log10(Hertz)

Nous illustrons l'influence de ces ratios dans les Figures 54a & 54b.

a)



b)



Figures 54a & 54b : Différence de résultat de la *T-Function* en fonction des unités choisies pour tracer la stylisation. Les segments bleus montrent la stylisation du contour intonatif de la locutrice modèle, en rouge la stylisation d'un contour de *Sp3* (masculin)

Bien que le couple Hertz/centiseconde propose des valeurs dans des ordres de grandeur comparable, ce choix d'unité ne semble pas optimal. Les Figures 54a & 54b soulignent que les différences de registre et d'étendue de *f0* sont amenuisées par la conversion de la hauteur

en log10. La courbe de référence (bleue) a une étendue moindre en log10, tandis que la courbe test (rouge) a une étendue un peu plus grande. Ainsi, nous avons choisi le couple log10(Hertz)/secondes car il « lisse » certaines différences interindividuelles.

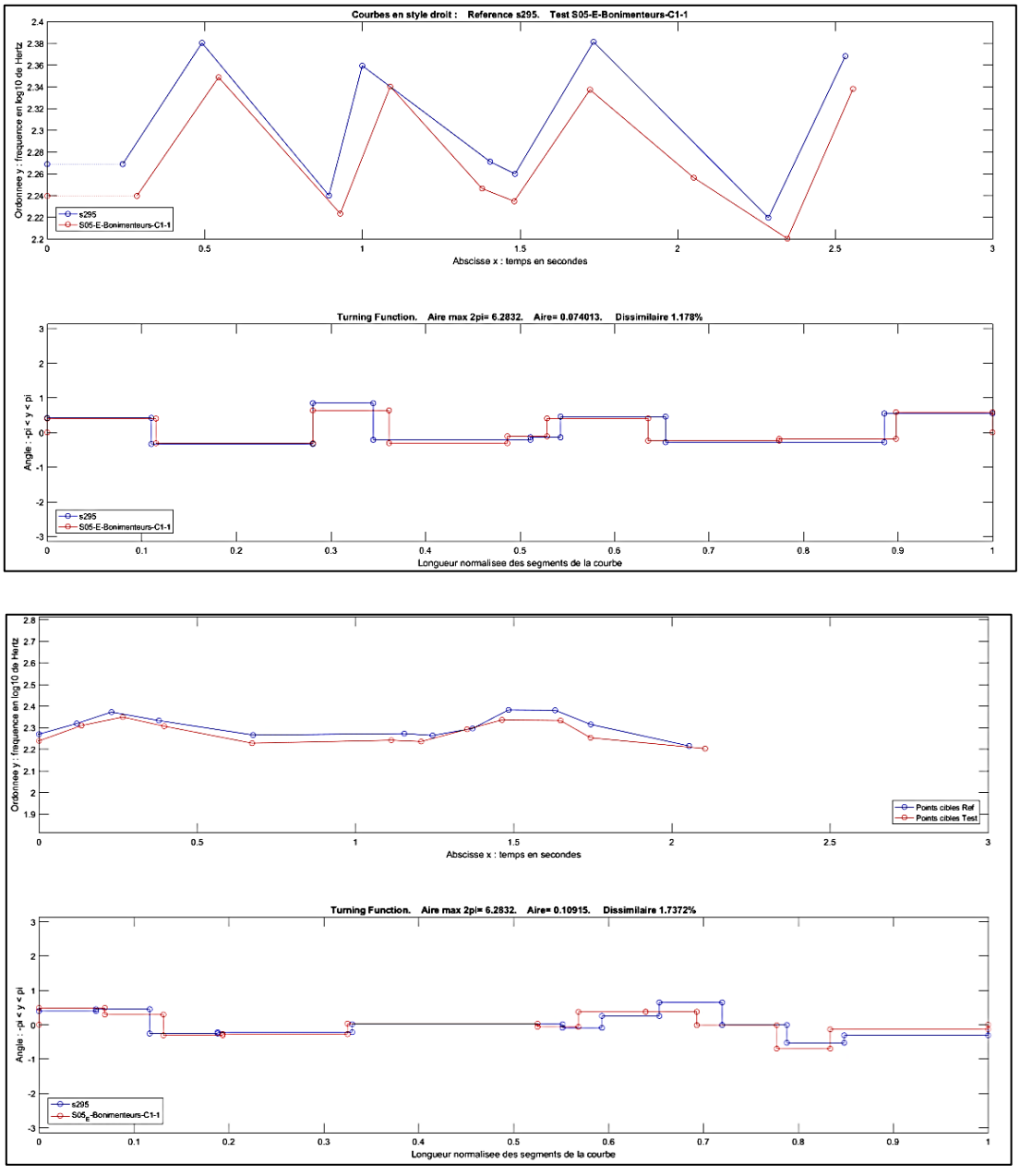


Figure 55a & 55b : Stylisations rectilignes de la *f0* de la phrase « *Les bonimenteurs et les baratineurs fades* » dite par la locutrice modèle et Sp5. Stylisation phonologique (55a) et syllabique (55a).

2.3.2.3 Systèmes d'annotations

Une fois le problème d'unité résolu se pose la question du choix d'annotation pour faire émerger la forme. Pour ces tests, nous avons retenu deux approches :

- Une annotation phonologique des énoncés, où l'annotation des tons permet de faire émerger les points d'ancrage de la stylisation rectiligne.
- Une annotation syllabique, où le milieu de chaque voyelle devient un point d'ancrage.

Nous illustrons Figure 55a & 55b (page précédente) ces deux stylisations issues d'une même phrase.

La stylisation phonologique (ci-après TSP, pour *T-Function* Stylisation Phonologique) privilégie le patron tonal du contour puisque les points d'ancrage visent à le faire ressortir. La pulsion rythmique y est sous-entendue car le patron intonatif conserve des aspects de durée.

La stylisation syllabique considère en premier lieu la pulsion rythmique en s'ancrant sur chaque syllabe (ci-après TSR, pour *T-Function* Stylisation Rythmique). Dans cette dernière, le patron tonal est implicitement pris en compte.

2.3.2.4 Résultats des T-Functions : TSP vs. TSR

Les deux systèmes de stylisation étant différents, nous pouvons dans un premier temps nous demander si TSP et TSR proposent des mesures concordantes de la similarité prosodique. Nous illustrons sur la Figure 56 la dispersion de ces mesures en fonction de TSP et TSR.

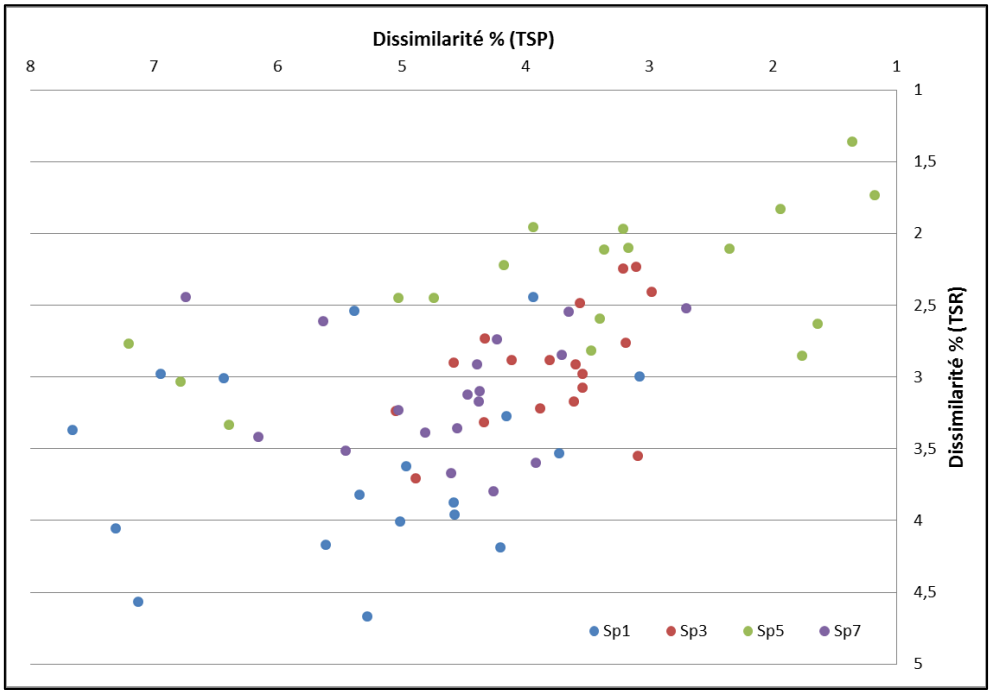


Figure 56 : Taux de dissimilarité des 72 imitations du CI en fonction de des mesures issues de TSP (abscisses) et TSR (ordonnées). L'ordre des valeurs des axes a été inversé : les meilleures imitations dans le cadran nord-est du graphique, inversement, les moins bonnes dans le cadran sud-ouest.

Bien que ce nuage de point semble être assez diffus, nous retrouvons certains caractères communs à toutes les mesures et évaluations présentées jusqu'à présent :

- Les productions de Sp5 sont évaluées comme les meilleures, et celle des Sp1 comme les moins bonnes.
- Les scores des productions de Sp3 sont assez concentrés.
- La tendance est moins nette pour les imitations de Sp7.

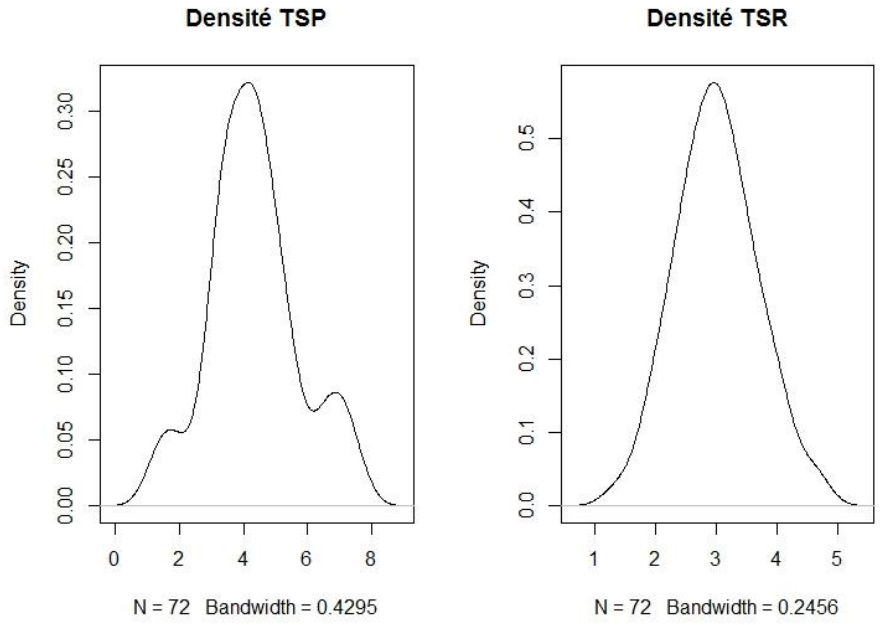


Figure 57 : Courbes de densité des scores TSP (gauche) et TSR (droite)

L'inspection visuelle de la distribution des mesures TSP montre une courbe présentant trois modes. Nous pouvons donc vraisemblablement douter de sa normalité. C'est pourquoi nous choisirons un test de corrélation de Spearman, non paramétrique. Nous présenterons cependant, à titre indicatif, les résultats du test de Pearson.

Spearman's rank correlation rho	Pearson's product-moment correlation
S = 28398 p-value = 1.25e-06 sample estimates: rho = 0.5434112	t = 5.2433, df = 70 p-value = 1.596e-06 95 percent confidence interval: 0.3413558 0.6791476 sample estimates: cor = 0.5310298

Tableau 27 : Résultats des tests de corrélation entre TSP et TSR

La corrélation entre mesures TSP et TSR est positive mais modérément forte. Il semble en effet qu'il y a certains désaccords entre les deux mesures, notamment dans l'évaluation des moins bonnes imitations qui sont très dispersées.

2.3.2.4 TSP vs. AX

Ayant comparé TSP et TSR, il convient à présent d'observer s'il y a une congruence entre ces mesures issues de stylisation rectiligne de la *f0* et leur évaluation perceptive. Nous présentons dans un premier temps la corrélation entre mesure TSP et test AX.

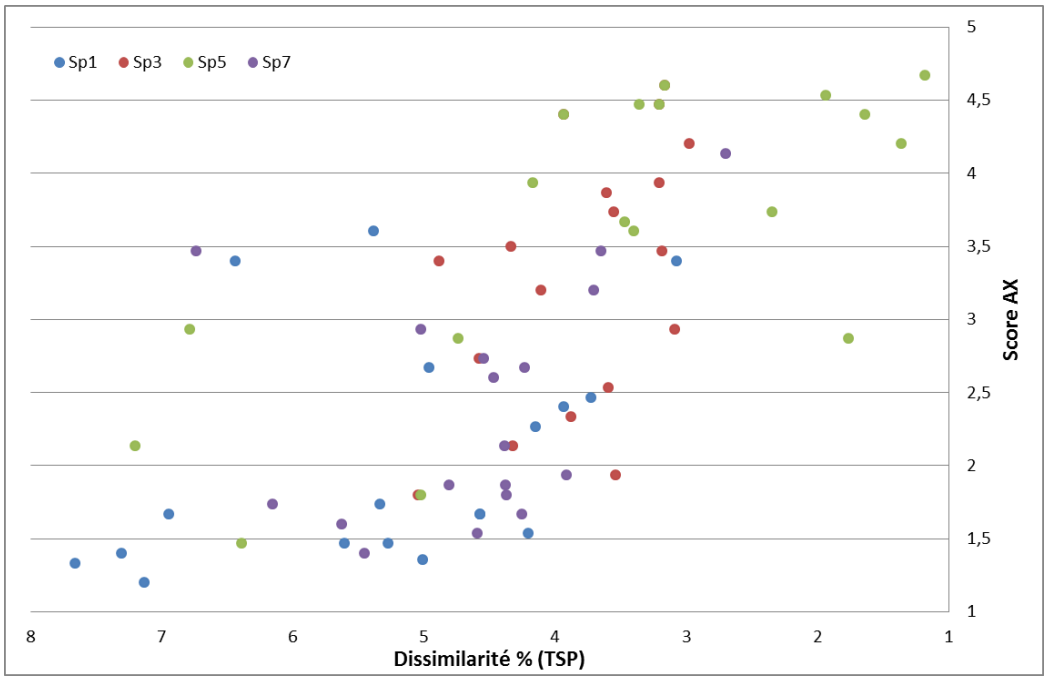


Figure 58 : Dispersion des scores évaluant les 72 imitations en fonction de la mesure TSP (abscisses, valeur inversées) et du test AX (ordonnées). Les imitations évaluées comme les meilleures se trouvent en haut à droite du graphique.

La distribution des données montre encore les mêmes tendances entre les sujets (*Sp5* > *Sp3* > *Sp7* > *Sp1*). Hormis quelques outliers ayant un score AX supérieur à 3 et un TSP inférieur à 6, les données sont plutôt regroupées.

L'inspection des courbes de densité (Figure 59) révèle des distributions non-normales des scores perceptifs d'une part (deux modes) et des scores TSP (3 modes). Il semble donc plus sage d'observer en priorité le résultat d'un test de Spearman, non-paramétrique.

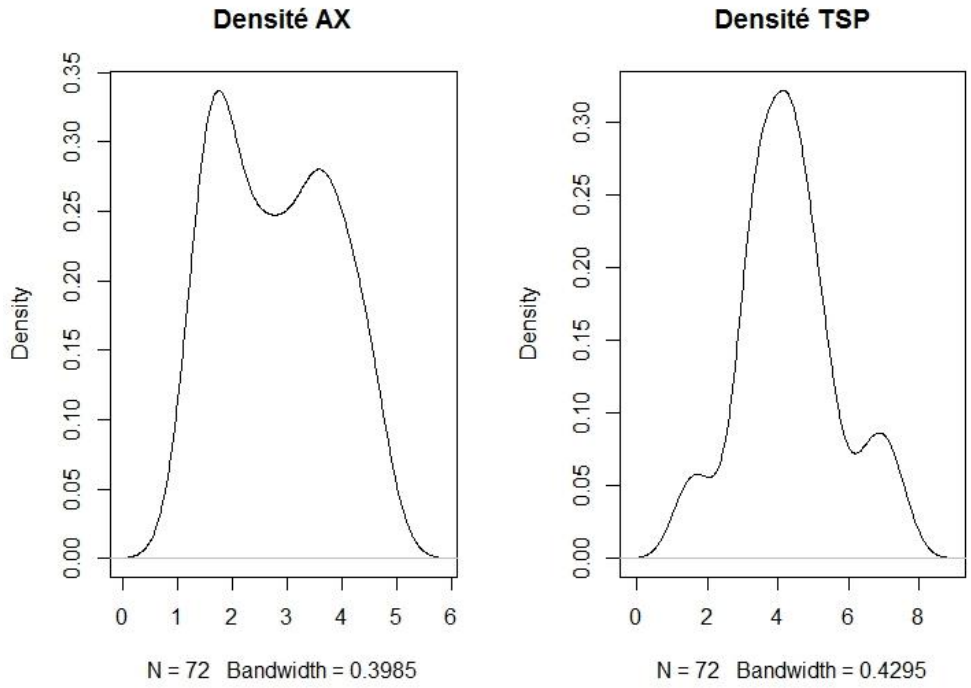


Figure 59 : Courbes de densité des scores AX (gauche) et TSP (droite)

Dans la Table 28, nous pouvons observer un taux de corrélation négatif entre TSP et AX. Le caractère négatif de ce taux était attendu : TSP est une mesure de dissimilarité, tandis que l'évaluation AX note la similarité.

Par ailleurs, ce sont les meilleurs taux de corrélation entre mesure automatique et évaluation perceptive constatés jusqu'à présent. TSP qui évalue la distance entre les patrons tonals de deux énoncés semble refléter la perception des auditeurs de manière satisfaisante.

Spearman's rank correlation rho S = 106670 p-value < 2.2e-16 sample estimates: rho = -0.7150942	Pearson's product-moment correlation t = -7.1653, df = 70 p-value = 6.235e-10 95 percent confidence interval: -0.7666177 -0.4931152 sample estimates: cor = -0.6504741

Tableau 28 : Résultat des tests de corrélation entre TSP et AX

2.3.2.5 TSR vs. AX

TSR est la dernière mesure que nous avons testée en relation avec l'évaluation perceptive AX. Nous montrons Figure 60, le nuage de dispersion des mesures TSR croisées avec les scores perceptifs AX des 72 imitations du *CI*.

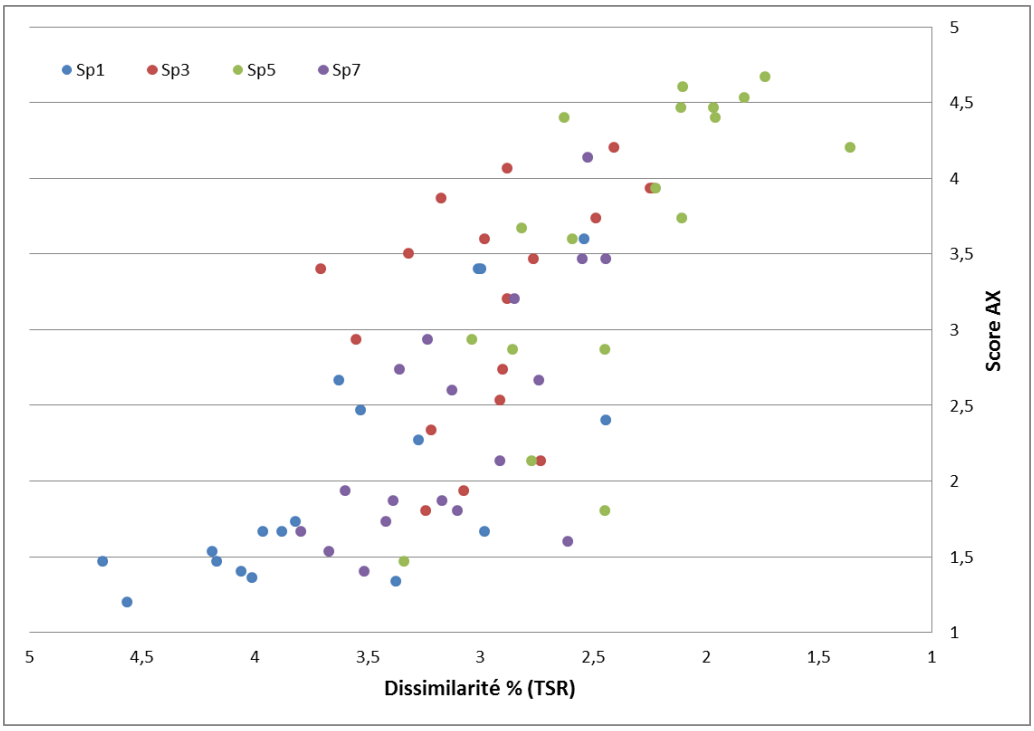


Figure 60 : Dispersion des scores de dissimilarité TSR (abscisses, valeurs inversées) et des scores de similarité perceptive AX (ordonnées). Les imitations évaluées comme les meilleures se trouvent en haut à droite du graphique.

Ce nuage de dispersion est particulièrement intéressant. Il semble y avoir une limite en deçà et au-delà de laquelle ni la mesure TSR ni les auditeurs humains ne se contredisent. En effet, les points aux extrémités du nuage sont concentrés autour du maximum et du minimum relevés pour les deux indices. Le centre de la distribution reste plus incertain.

L'inspection des courbes de densité (Figure 61) dont l'aspect sort de la normalité, comme précédemment, nous pousse à considérer en priorité le résultat d'un test de corrélation non-paramétrique (Spearman). Nous donnons toutefois le résultat du test de Pearson dans la Table 29, à titre indicatif.

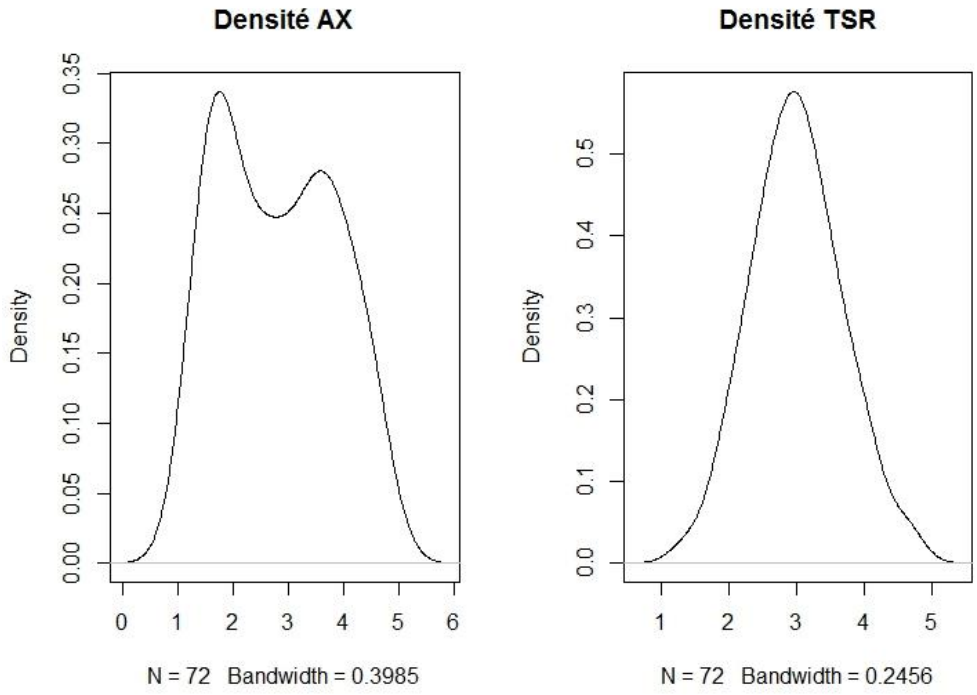


Figure 61 : Courbes de densité des scores d'évaluation perceptive AX (gauche) et des scores de dissimilarité TSR (droite)

Comme pour l'indice TSP, l'indice TSR entretient une corrélation négative avec le score perceptif AX. Enfin, les taux de corrélation observés ici sont meilleurs que tous les taux de corrélations observés précédemment, laissant suggérer que l'indice TSR pourrait être le plus approprié pour évaluer la similarité prosodique entre une imitation et son modèle.

Spearman's rank correlation rho S = 110100 p-value < 2.2e-16 sample estimates: rho = -0.7701781	Pearson's product-moment correlation t = -9.9666, df = 70 p-value = 4.592e-15 95 percent confidence interval: -0.8472438 -0.6494804 sample estimates: cor = -0.7659074

Tableau 29 : Résultats des tests de corrélation entre les indices TSR et AX

2.3.2.6 Discussion

Afin de compenser le fait que les mesures L_2 et Z_r oblitèrent les aspects temporels des contours dont nous souhaitions comparer la similarité, nous avons proposé d'utiliser une méthode considérant les représentations de la *f0* comme des objets géométriques.

Pour ce faire, les contours intonatifs ont été stylisés au moyen de points d'ancrage (patrons intonatifs, TSP, ou centre des syllabes, TSR) puis il leur a été appliqué une *T-Function*. Cette transformation a l'avantage de conserver les proportions des contours, et, par conséquent, de maintenir des aspects temporels dans le calcul de similarité.

Les taux de corrélation relevés pour les méthodes TSP et TSR étaient supérieurs à ceux relevés pour les mesures L_2 et Z_r. Il semble donc que TSP et TSR sont des candidats plus appropriés que les mesures issues de contours bruts pour mesurer la similarité prosodique de deux énoncés.

Des deux mesures issues des *T-Function*, TSR s'est révélée être meilleure que TSP. Il est probable que les auditeurs se sont basés également sur le rythme (et en filigrane le patron intonatif) pour effectuer leur jugement perceptif de similarité. TSR, qui vise à capturer prioritairement cette dimension, se révèle ici comme un candidat intéressant pour évaluer la similarité entre une phrase originale et sa reproduction délexicalisée. Ceci constitue le prochain pas de notre travail (Chapitre 6).

Dans la mesure où la corrélation entre les *T-Function* et les évaluations perceptives AX est plutôt fort, nous pensons que les *T-Function* pourraient se substituer aux tests perceptifs, en gardant une certaine prudence, toutefois. En effet, d'autres tests sont nécessaires sur des corpus plus étendus pour affiner notre compréhension de ces résultats. Il pourrait notamment être bénéfique de faire diminuer la variabilité interindividuelle dans la perception de la similarité prosodique afin de pouvoir par la suite établir des seuils critiques permettant d'interpréter avec fiabilité le résultat des mesures automatiques.

Ceci étant dit, les *T-Functions* se révèlent être des outils pertinents dans un cadre d'études linguistiques, puisqu'il est assez simple de les appliquer une fois l'annotation tonale d'un contour intonatif réalisée.

Enfin, les deux méthodes que nous avons utilisées pour la stylisation des contours présentent l'avantage d'être potentiellement automatisables :

- La TSP est reproductible en utilisant un algorithme comme MOMEL (Hirst & Espesser, 1993)
- La TSR peut être réalisée de manière semi-automatique en utilisant un aligneur segmental comme SPPAS (Bigi, 2015) puis un script PRAAT pour placer les points d'ancrage.

Cela dit, il nous faut noter que les *T-Function* sont particulièrement sensibles au bruit ou aux événements exceptionnels : un point déviant ou manquant peut avoir une influence considérable sur la transformation du patron stylisé et donc, fausser proportionnellement le résultat de la mesure de similarité. Il peut par exemple être nécessaire de corriger les valeurs de *f0* manuellement quand la voix du locuteur présente des craquements.

Dans ce chapitre, nous avons estimé la similarité prosodique de 72 imitations avec leur modèle par des tests perceptifs (AX et AXB) et des mesures automatiques issues de *f0* brutes (L_2 & Z_r) et de *f0* stylisées (TSP & TSR). Globalement, toutes les mesures automatiques reflètent les jugements de similarité obtenus durant le test perceptif AX, le test AXB ayant été laissé de côté à cause de son fonctionnement. Cependant, TSP, et plus particulièrement TSR, qui conservent des aspects rythmiques dans le calcul de la similarité prosodique semblent donner les résultats les plus fidèles à la perception des auditeurs.

Chapitre 6 : Evaluation de la similarité prosodique d'imitations délexicalisées : une application de la TSR

Le chapitre précédent présentait notre exploration d'un même corpus d'imitations parolières au moyen de diverses mesures de la similarité prosodique que nous liions systématiquement aux résultats de tests perceptifs. L'objectif premier de cette étude était de sélectionner une mesure en laquelle nous aurions suffisamment confiance pour évaluer automatiquement la similarité prosodique d'une imitation et de son modèle.

Au terme de cette étape de notre travail, nous avons déterminé que les résultats obtenus par la *T-Function* Stylisation Rythmique semblent correspondre à ceux de notre tâche d'évaluation : cette mesure obtenait une corrélation satisfaisante avec les jugements perceptifs des auditeurs. En effet, la TSR dont les points d'ancrages sont les voyelles de chaque syllabe tient particulièrement compte du décours temporel dans les formes comparées tout en conservant également des traces du patron intonatif, soulignant alors l'importance du rythme dans la perception de la prosodie.

Dans le chapitre à venir, nous nous proposons de poursuivre nos investigations en nous focalisant sur une situation que peuvent rencontrer les enseignants utilisant la MVT : la production de logatomes. Lors de ce moment du cours, l'enseignant est amené à reproduire, en les délexicalisant, intonation et rythme d'un énoncé entendu précédemment. Cette tâche demande à l'enseignant un contrôle prosodique particulier pour reproduire de manière systématique et avec une précision suffisante le pattern prosodique perçu.

Pour évaluer la réussite de locuteurs naïfs et experts dans cette tâche, nous appliquerons des mesures issues de *T-Functions* sur un corpus d'imitations lexicalisées et délexicalisées. Ce faisant, nous nous intéresserons d'une part au contrôle prosodique des locuteurs et à l'apport de la TSR dans notre recherche d'autre part. Enfin, nous présenterons un projet de logiciel d'entraînement au logatome à destination de l'enseignant.

Nous décrirons donc les étapes de ce protocole expérimental avant de nous intéresser aux résultats obtenus au moyen de la TSR sur ce corpus de logatomes.

1. Corpus Logatome (CL) : acquisition et traitement des données

Pour cette étude, il a été décidé de recueillir un nouveau corpus. L'idée sous-jacente à ce recueil était de replacer les sujets d'expérience dans la situation d'un enseignant de langue étrangère préparant le travail d'un dialogue avec ses apprenants. Pour cela, nous devions disposer au préalable d'un matériel sonore adapté.

Nous avons fait le choix du dialogue car leur utilisation est assez fréquente dans les niveaux de langue les moins avancés : de nombreuses méthodes de FLE proposent des dialogues comme amorce des leçons jusqu'au niveau B1. De plus, le travail sur le rythme et l'intonation est crucial dans les premiers temps de l'apprentissage. Il semble donc concevable qu'un enseignant s'entraîne à produire les logatomes correspondants à un dialogue préenregistré pour les niveaux A1 à B1 du Cadre Européen Commun de Référence (Conseil de l'Europe, 2001).

1.1 Création du dialogue modèle

Le dialogue a été écrit par l'auteur de ce travail. Le thème sélectionné est typique des méthodes pour les niveaux débutants : il s'agit d'une interaction de la vie courante, où deux personnes partageant le même logement planifient leur déjeuner, font une liste, puis vont faire les courses (Figure 62). Dans le dialogue « *Au marché* », A et B discutent du déjeuner à venir puis vont au marché où ils rencontrent le marchand C.

Le dialogue « *Au marché* » a été enregistré par trois locuteurs francophones natifs (âge 28-31), étudiants en doctorat de sciences du langage. Les rôles A et B étaient tenus par des femmes et le rôle C par un homme.

L'enregistrement a eu lieu dans le studio de la plateforme expérimentale PETRA (goo.gl/D9d2ZC). Les trois comédiens étaient debout, placés chacun face à un micro muni d'un filtre anti-pop.

Après avoir pris connaissance du dialogue, celui-ci a été répété trois fois dans son intégralité avant de l'enregistrer. Il était demandé aux comédiens d'éviter d'avoir un débit trop lent, afin que la conversation paraisse naturelle. Il leur était également demandé de ne pas parler trop vite car les phrases devaient pouvoir être logatomisées par la suite.

Cinq enregistrements de « *Au marché* » ont été faits. Par la suite, l'expérimentateur en a monté une version finale dans le logiciel PRAAT.

Durant cette étape, les cinq versions ont été segmentées phrase par phrase. Puis, l'expérimentateur a écouté les différentes versions de chaque segment et en a inspecté visuellement la courbe de *f0*. Les segments où la voix était craquée ont été éliminés d'office, car les valeurs de *f0* associées à ce type de voix sortent du registre normal des locuteurs. Les segments paraissant les plus « authentiques » à l'écoute et dont la courbe de *f0* présentait des contrastes intéressants ont finalement été sélectionnés pour le montage final.

- **A** : Tu vas au marché ? |
- **B** : Oui, | il nous manque des choses pour le déjeuner. |
- **A** : Tu voulais faire quoi ? |
- **B** : Alors, | j'avais prévu de faire | une bonne salade, | avec des lentilles, | des tomates et des oignons rouges. | Et un magret de canard. |
- **A** : Encore de la viande ? | On pourrait manger du poisson, non ? |
- **B** : Si tu veux. | Je vais passer chez le poissonnier, | voir s'il a une belle dorade, ou quelques rougets. |
- **A** : Et pour le dessert ? |
- **B** : Je n'ai rien prévu, | tu sais moi, les douceurs, c'est pas mon truc. |
- **A** : Je t'accompagne, | je t'aiderai à choisir le dessert. |
- […]
- **C** : A qui le tour ? |
- **A** : A nous ! |
- **C** : Qu'est-ce qu'il vous faudra ? |
- **B** : Alors, | on va prendre des tomates et des oignons rouges. | Environ un kilo de chaque. |
- **C** : Avec ceci ? |
- **A** : Vous nous mettrez aussi quatre endives, s'il vous plaît. |
- **C :** Ce sera tout ? |
- **A** : Oui, merci. |
- **C** : Ça fera quatre cinquante. |
- […]
- **B** : On a le poisson et les légumes, | il nous manque juste le dessert. | Tu voulais prendre quoi ? |
- **A** : Un bon gâteau, | c'est dimanche ! | On va chez le pâtissier ? |
- **B** : Ça marche, | dépêchons-nous, | je commence à avoir faim ! |

Figure 62 : Dialogue « Au marché » utilisé pour l'enregistrement du CL. La lettre en début de ligne indique les rôles de chaque personnage. Les barres verticales marquent la segmentation du dialogue pour créer les stimuli destinés à l'enregistrement du CL.

Les phrases du dialogue ont par la suite été segmentées une nouvelle fois pour créer les stimuli modèles du Corpus Logatome (CL) (Figure 62) dont l'intensité a été normalisée à 70 dB au moyen d'un script PRAAT pour égaliser le niveau sonore des stimuli. Suite à cette nouvelle segmentation, nous avons donc obtenu 40 stimuli présentés en Table 30.

Numéro Stim	Syllabes	Segment	Numéro Stim	Syllabes	Segment
p01	5	tu vas au marché	**p21**	3	à qui le tour
p02	1	oui	**p22**	2	à nous
p03	10	il nous manque des choses pour le déjeuner	**p23**	5	qu'est-ce qu'il vous faudra
p04	5	tu voulais faire quoi	**p24**	2	alors
p05	2	alors	**p25**	11	on va prendre des tomates et des oignons rouges
p06	5	j'avais prévu de faire	**p26**	7	environ un kilo de chaque
p07	4	une bonne salade	**p27**	4	avec ceci
p08	5	avec des lentilles	**p28**	12	vous nous mettrez aussi quatre endives s'il vous plaît
p09	8	des tomates et des oignons rouges	**p29**	3	ce sera tout
p10	6	et un magret de canard	**p30**	3	oui merci
p11	5	encore de la viande	**p31**	6	ça fera quatre cinquante
p12	9	on pourrait manger du poisson non	**p32**	9	on a le poisson et les légumes
p13	3	si tu veux	**p33**	7	il nous manque juste le dessert
p14	9	je vais passer chez le poissonnier	**p34**	5	tu voulais prendre quoi
p15	12	voir s'il a une belle dorade ou quelques rougets	**p35**	4	un bon gâteau
p16	5	et pour le dessert	**p36**	3	c'est dimanche
p17	5	je n'ai rien prévu	**p37**	7	on va chez le pâtissier
p18	10	tu sais moi les douceurs c'est pas mon truc	**p38**	2	ça marche
p19	4	je t'accompagne	**p39**	4	dépêchons nous
p20	9	je t'aiderai à choisir le dessert	**p40**	7	je commence à avoir faim

Tableau 30 : 40 stimuli originaux du Corpus Logatome et comptage des syllabes. Le nombre de syllabe correspond à ce qui a effectivement été prononcé par les locuteurs.

La longueur des stimuli du CL varie de 1 à 12 syllabes. Or, d'après Billières (2002), la production de logatomes devient difficile pour l'enseignant au-delà de 4 à 5 syllabes en raison des capacités de traitement de la mémoire de travail.

A mesure que nous soulèverons des problème dans la description de ce protocole, nous formulerons des hypothèses quant à la similarité prosodique et sa mesure. Ces hypothèses seront récapitulées avant les résultats.

H1 : Nous pourrons donc nous demander si le score de similarité prosodique diminue à mesure que le nombre de syllabes de l'énoncé à reproduire augmente, *i.e.* s'il y a une limite au-delà de laquelle la tâche devient trop difficile pour les locuteurs.

1.2 Recueil du Corpus Logatome

Lors du recueil du CL, 4 locuteurs (2 naïfs, 1 homme et 1 femme et 2 experts, 1 homme et 1 femme également) ont produit des imitations parolière lexicalisées et délexicalisée, *i.e.* : ils ont écouté chaque stimulus que nous venons de décrire et en ont alternativement reproduit l'intégralité du contenu ou simplement son logatome.

1.2.1 Population expérimentale

Les 4 locuteurs du CL étaient des locuteurs francophones natifs et ont été recrutés sur le campus de l'UT2J. Leur participation à l'enregistrement n'était pas rémunérée. Dans ce recueil, nous avons souhaité opposer une population experte à une population naïve :

- Les 2 locuteurs experts ont une formation de Français Langue Etrangère, sont formés à la MVT et en ont une pratique fréquente durant leur carrière.
    - Expert 1 (E1) : locuteur masculin
    - Expert 2 (E2) : locuteur féminin
- Parmi les locuteurs naïfs :
    - 1 locuteur est en cours de formation à l'UT2J pour devenir professeur de FLE mais n'a pas encore de pratique courante de la MVT.
        - Non Expert 1 (NE1) : locuteur féminin

- 1 locuteur n'a aucune formation en didactique, mais connait les principes de la MVT du fait de son implication dans un projet y ayant trait.
    - Non Expert 2 (NE2) : locuteur masculin

Il était intéressant pour nous d'avoir ces deux types de locuteurs car nous attendions une meilleure performance des locuteurs experts.

De prime abord, la production de logatome n'est pas une tâche intuitive : il faut parvenir à court-circuiter le niveau sémantique pour produire uniquement le rythme et l'intonation de l'énoncé entendu. En quelques sortes, il s'agit à la fois d'un procédé d'imitation prosodique (dont nous avons vu que la réussite dépendait du locuteur) et d'un procédé métalinguistique qui demande d'avoir une conscience phonologique du système intonatif du français.

H2 : Les locuteurs experts devraient obtenir de meilleurs scores de similarité prosodique que les locuteurs naïfs.

H3 : Un effet d'entraînement à la tâche n'est pas exclure pour les locuteurs naïfs, *i.e.* leurs scores de similarité prosodique pourraient augmenter durant le test.

1.2.2 Tâches expérimentales

Pour cet enregistrement de corpus, les locuteurs ont accompli 2 tâches expérimentales présentées dans 2 blocs consécutifs :

- **Bloc 1 : Phrase-Logatome-Phrase (ci-après : PLP)**

Durant ce bloc, les locuteurs entendaient un stimulus avant de reproduire la phrase entendue, son logatome, puis à nouveau la phrase entendue.

- **Bloc 2 : Logatome-Phrase-Logatome (ci-après : LPL)**

Après avoir entendu le stimulus, les locuteurs en reproduisaient le logatome, la phrase originale puis une dernière fois le logatome.

Hormis l'ordre de production des phrases ou des logatomes, les consignes pour les deux blocs étaient similaires. Voici ce qui était indiqué aux locuteurs :

- Vous allez entendre les phrases d'un dialogue.
- Vous devez dire ces phrases en essayant de reproduire leur mélodie.

- Durant ce bloc [*i.e.* PLP ou LPL], après écoute, dites [P ou L], [L ou P], puis [P ou L]
- Faites les logatomes sur la syllabe « da »
- Vous allez pouvoir écouter l'ensemble du dialogue plusieurs fois avant de démarrer

Etant donné que les logatomes sont une imitation partielle d'une phrase dite ou entendue, nous nous attendions à ce qu'il soit plus simple pour les sujets de produire le logatome du bloc PLP (produit après une phrase dite) que le premier logatome du bloc LPL (produit après écoute d'une phrase), en raison de l'expérience sensorimotrice liée à la production de la première phrase du bloc PLP.

H4 : Le logatome du bloc PLP obtiendrait de meilleurs scores de similarité prosodique que le logatome 1 du bloc LPL.

H5 : Les P obtiendront des meilleurs scores que les L

Nous avons choisi de faire produire plusieurs phrases et logatomes d'affilée pour plusieurs raisons. D'une part, il n'est pas rare que l'enseignant pratiquant la MVT présente une nouvelle phrase en la faisant précéder de son logatome : ces tâches visaient à simuler cette situation. D'autre part, nous nous demandions si la production successive de phrases et logatomes améliorerait ou dégraderait la similarité prosodique entre les imitations et le modèle original. Nous pourrions en effet penser que les locuteurs subissent l'influence de la dernière trace acoustique (ou sensorimotrice) stockée en mémoire de travail pour produire chaque énoncé.

H6 : La similarité prosodique avec le modèle original sera dégradée à mesure des productions successives d'un même *trial*.

Chaque bloc expérimental comptait 120 stimuli :

- 40 stimuli de « *Au marché* » randomisés, soit l'intégralité du dialogue
- 3 répétitions du dialogue en bloc

Ainsi, les locuteurs entendaient les 40 phrases du dialogue dans un ordre aléatoire, avant d'être exposés une seconde fois à ces 40 phrases, puis une troisième fois.

L'intégralité du CL représente donc un total de 40 segments modèles * 3 répétitions par bloc * 3 L ou P par trial * 2 blocs * 4 locuteurs, soit 2880 imitations (1440 P et 1440 L) auxquelles il faut associer les 40 stimuli originaux.

1.2.3 Passation

Le protocole de recueil du CL était scripté dans Lancelot, l'environnement HTML du logiciel Perceval (André et al., 2003) et l'enregistrement se déroulait dans le studio de la plateforme PETRA.

Un micro sur pied muni d'un filtre anti-pop était installé au milieu du studio, et réglé à hauteur de bouche du sujet, de façon à ce qu'il reste debout durant l'expérience.

Après accueil du sujet, l'expérimentateur lançait la session expérimentale, lisait la consigne avec le sujet et lui expliquait, si nécessaire ce qu'était un logatome.

Il lui faisait ensuite écouter l'intégralité du dialogue pour que le sujet se familiarise avec le matériel sonore, de la même manière que pourrait le faire un enseignant préparant un cours. Le dialogue était écouté au moins 3 fois, mais le sujet pouvait demander 2 nouvelles écoutes pour un total maximum de 5 écoutes intégrales.

Il était ensuite proposé au sujet de se familiariser à la tâche expérimentale en reproduisant 4 énoncés de différentes longueurs issus du dialogue. A l'issue de l'entraînement, le test était lancé par l'expérimentateur.

Les sujets ne manipulaient pas l'interface de test : l'expérimentateur lançait manuellement chaque stimulus quand le sujet était prêt pour la production suivante. Ainsi, les locuteurs testés n'étaient pas distraits de leur tâche par le dispositif expérimental et l'expérimentateur pouvait s'assurer que chaque *trial* était effectué correctement avant de passer au suivant.

La passation totale du protocole d'enregistrement a duré environ 1h15 par sujet. L'enregistrement des blocs en eux même durait en moyenne 25 minutes (soit environ 50 minutes pour les deux blocs). Le temps restant était réparti dans l'accueil du sujet, le réglage du matériel (hauteur des micros, sensibilité de la capture), l'explication des consignes, l'écoute du dialogue et la pause entre les blocs si le sujet souhaitait en faire une.

1.2.4 Traitement des données

Les enregistrements de chaque bloc expérimental (environ 25 minutes chacun) ont d'abord été segmentés en morceaux plus petits de manière semi-automatique. Nous avons en premier lieu appliqué la commande « *to Textgrid silence* » de PRAAT (Boersma, 2001), pour repérer les temps de parole. Après correction du Textgrid, un code permettant d'identifier chaque segment a été intégré dans la Tier d'annotation initialement prévue pour distinguer le silence de la parole (Table 31). La transcription orthographique de chaque segment a également été insérée sur une autre Tier d'annotation afin de préparer l'alignement automatique. Ces deux opérations ont été réalisées au moyen d'un script pour PRAAT.

Exemple : NE1_LPL_p13_1_X2

Sujet	**Bloc**	**pxx**	**1 ou 2 ou 3**	**X1 ou X2 ou X3**
Experts (E1 ou E2) Non Experts (NE1 ou NE2)	PLP ou LPL	Enoncé issu du dialogue, modèle numéroté de 01 à 40	Première, seconde ou troisième écoute du dialogue dans le bloc	Position de l'imitation dans le bloc PLP ou LPL (soit, X1 = L ou P en fonction du bloc)

Tableau 31 : Codage des stimuli du CL pour leur identification. L'exemple donné en première ligne indique : Production du locuteur Non Expert 1, durant le bloc Logatome-Phrase-Logatome, modèle phrase 13, première écoute, production 2 (P) du *trial*.

Suite à ces premières opérations, nous avons donc obtenus 2880 fichiers .wav et leurs Textgrid. Nous avons alors utilisé SPPAS (Bigi, 2015) pour procéder à l'alignement automatique des phonèmes de chaque imitation. Chaque Textgrid a ensuite été inspecté manuellement pour corriger les erreurs d'alignement les plus flagrantes, ainsi que les aberrations liées à l'écart entre la transcription orthographique et ce qui a effectivement été dit par les locuteurs du CL (notamment le nombre de « *da* » dans les productions de logatomes).

Les mêmes traitements ont été appliqués aux 40 stimuli modèles.

Finalement, un script PRAAT a annoté chaque milieu de voyelle d'un point, sur lequel le temps et la *f0* ont été relevés dans le but d'obtenir les données nécessaires pour appliquer la TSR entre les modèles (X0) et les imitations (X1, X2 et X3).

Dans la mesure où le contenu segmental des productions P et des productions L diffère fondamentalement, la méthode d'obtention des points d'ancrage risque de provoquer un biais dans le résultat de la TSR. En effet, les syllabes des L étant toujours ouvertes (« da »), le

noyau vocalique devrait être plus long que leur équivalent dans les P. Comme le modèle est de type P, il faut s'attendre à ce que la mesure TSR donnée pour les L soit sensiblement moins bonne, dans la mesure où le milieu de la voyelle peut systématiquement être décalé.

H5 bis : Le système d'annotation des points en milieu de noyau vocalique provoquera un biais dans la TSR, se traduisant par des scores de similarité meilleurs pour P, par comparaison avec L.

1.2.5 Rappel des hypothèses

Avant de relater nos résultats, nous récapitulons ici les hypothèses formulées au fil de la description de notre protocole expérimental de recueil d'imitations lexicalisées et délexicalisées.

H1 : Le score de similarité prosodique diminuera à mesure que le nombre de syllabes de l'énoncé à reproduire augmente, *i.e.* il y a une limite au-delà de laquelle la tâche devient trop difficile pour les locuteurs.

H2 : Les locuteurs experts devraient obtenir de meilleurs scores de similarité prosodique que les locuteurs naïfs.

H3 : Un effet d'entraînement à la tâche n'est pas exclure pour les locuteurs naïfs, *i.e.* leurs scores de similarité prosodique pourraient augmenter durant le test.

H4 : Le logatome du bloc PLP obtiendrait de meilleurs scores de similarité prosodique que le logatome 1 du bloc LPL.

H5 : Les P obtiendront des meilleurs scores que les L

H5 bis : Le système d'annotation des points en milieu de noyau vocalique provoquera un biais dans la TSR, se traduisant par des scores de similarité meilleurs pour P, par comparaison avec L.

H6 : La similarité prosodique avec le modèle original sera dégradée à mesure des productions successives d'un même *trial*.

Notons que ces hypothèses ne seront pas testées dans l'ordre où elles sont présentées ici.

2. Analyse du Corpus Logatome

Nos analyses du CL ont été réalisées dans le logiciel R (CRAN, 2016). Nous avons été particulièrement guidés par l'ouvrage de Baayen (2008).

2.1 Notes liminaires

Tous les résultats que nous allons décrire à présent concernent des scores relatifs à la similarité prosodique entre un modèle (lexicalisé) et son imitation, que celle-ci soit lexicalisée (P) ou délexicalisée (L).

Ces scores de similarité ont été obtenus au moyen de calcul de la similarité suite à une transformation de type *T-Function*. Il est nécessaire de rappeler que ces mesures évaluent la dissimilarité : plus le score est élevé et plus l'écart entre les objets comparés est grand. Elles donnent deux outputs :

- Le calcul de l'aire comprise entre les courbes de *f0* transformées par la *T-Function*, dont les valeurs sont contenue dans l'intervalle [0 ; 2π]
- Un taux de dissimilarité, qui exprime la proportion entre l'aire calculée et l'aire maximum théorique, dont les valeurs sont bornées dans l'intervalle [0 ; 100]

Etant donné qu'il est plus facile de se représenter la dissimilarité en termes de pourcentage, nous avions initialement choisi de retenir les valeurs exprimées par les taux de dissimilarité.

Cependant nos premières inspections visuelles nous ont indiqué que la distribution des données était systématiquement asymétrique. Afin de rendre ces distributions plus symétriques, il a donc été décidé de transformer l'ensemble des données en les convertissant en logarithmes naturels puis en leur ajoutant un (afin d'éviter les scores négatifs ou nuls). Ainsi, les scores de dissimilarité sont ici exprimés log(x% +1).

Enfin, dans notre explication du protocole expérimental du recueil du CL, nous avons uniquement évoqué la mesure TSR. Or, dans notre cheminement nous avons produit trois autres mesures approchantes :

- Une mesure issue d'une stylisation semi-automatique :

    - La TSR multipoint (TSRm), similaire à la TSR, à ceci près que les voyelles pouvaient recevoir plusieurs points d'ancrage en fonction de leur durée (jusqu'à trois points).
- Deux mesures intégralement automatisées, plutôt similaires à la TSP (annotation tonale) :
    - Détection et annotation tonale de la *f0* avec l'algorithme MOMEL (Hirst & Espesser, 1993)
    - Détection de la *f0* avec l'algorithme YIN (de Cheveigné & Kawahara, 2002), puis annotation tonale avec MOMEL.

Pour le test des hypothèses rappelées précédemment, nous nous limiterons à la TSR normale. L'ajout de points de la TSRm semblait initialement une bonne idée pour tenter de rendre compte au mieux de la topographie de la *f0*. Cependant, il nous semble que cela « fausse » partiellement la mesure : dans certains cas, le nombre de points d'ancrage entre modèle et imitation devient inégal et l'information sur la pulsion syllabique est alors moins précise. Ainsi, en voulant capturer une finesse supplémentaire, la TSRm devient moins efficace que la TSR pour comparer les logatomes aux productions originales puisqu'un nombre de points inégal entre modèle et imitation peut alternativement être dû à une syllabe omise durant la reproduction ou à des points ajoutés à cause des durées vocaliques.

Les versions complètement automatisées de la mesure seront évoquées après ces résultats. Une observation des scores de dissimilarité obtenus par ces méthodes complètement automatiques laisse supposer l'émergence d'erreurs dans le processus, probablement liées à la détection même de la *f0* et/ou bien au choix des points d'ancrage pour la *T-Function*. Après le test de nos hypothèses, nous décrirons plus en détail le contexte d'automatisation de ces mesures et leurs limites actuelles.

Enfin, nous avons éliminé toutes les phrases :

- Dont les scores de dissimilarité (avant transformation) étaient supérieurs à 15%
    - Ces phrases ont été jugées comme des *outliers*. Ces scores très hauts en terme de dissimilarité étaient selon nous dus à des erreurs de détection de la *f0* (ou son absence de détection) ou à des voix craquées, ayant faussé la mesure.
- Qui n'avaient pas de score de dissimilarité dans au moins une des quatre méthodes
    - Il s'agissait essentiellement de segments très courts comme « oui », « alors » ou bien contenant beaucoup de consonnes sourdes et/ou fricatives comme

« c'est dimanche », « avec ceci ». Dans ces cas-là, l'algorithme n'avait pas assez de points pour effectuer la transformation de la *f0* puis le calcul de la similarité.

Le corpus retenu pour analyse représente finalement 2055 phrases et logatomes sur un total de 2880.

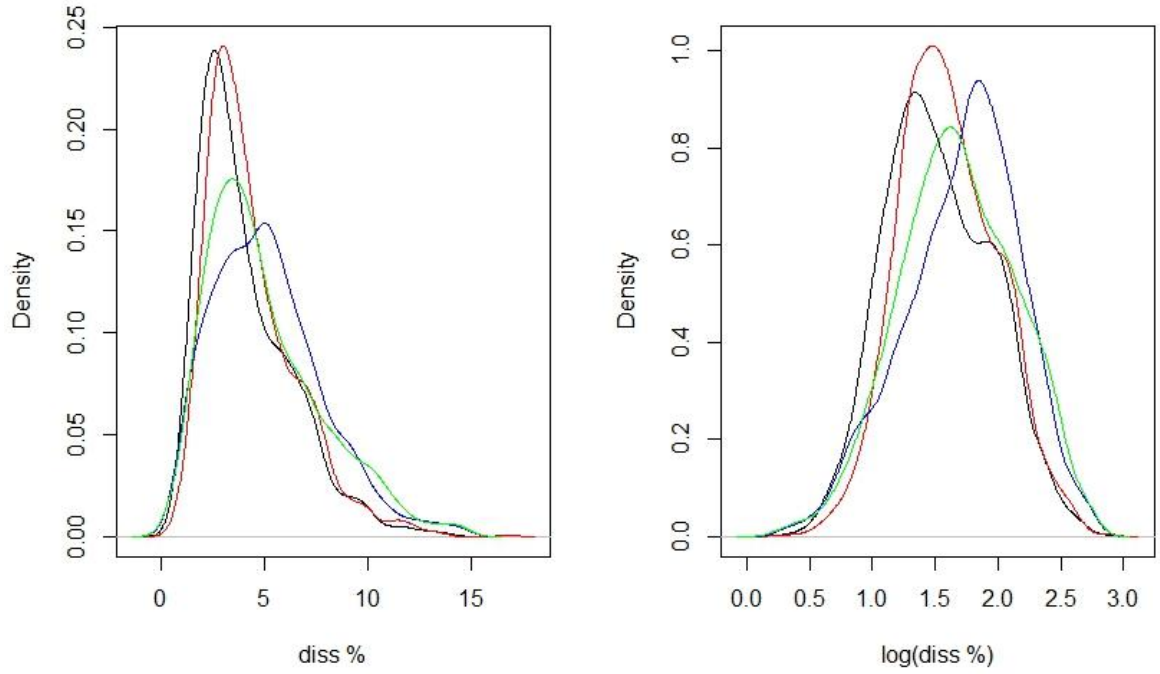


Figure 63 : Courbes de densité des scores de dissimilarité de 2055 imitations en % (gauche) et en log(score+1) (droite). TSR classique (noir), TSR multipoints (rouge), Automatique Momel (bleu), YIN + Momel (vert)

La Figure 63 illustre la densité de distribution des données obtenues par les différentes mesures que nous venons d'évoquer. Dans la partie gauche, nous pouvons observer que les scores supérieurs à 10% de dissimilarité sont très peu fréquents et que les distributions sont asymétriques. La partie droite montre l'allure des distributions après conversion des scores. La queue de la distribution est moins étendue et les distributions sont plus symétriques.

2.2 Comparaison des scores de dissimilarité des Phrases vs. Logatomes

Nous testons ici les hypothèses **H5** et **H5bis** qui indiquent que les Phrases (P) auraient de meilleurs scores de similarité prosodique que les Logatomes (L).

- **H5** considère que P serait mieux reproduite que L par les locuteurs car la prosodie de L n'est pas ancrée sur le contexte sémantique habituel
- **H5bis** considère que P obtiendra des scores meilleurs que L en raison du système d'annotation, *i.e.* l'adéquation entre le contenu segmental de P *vs*. P et P *vs*. L.

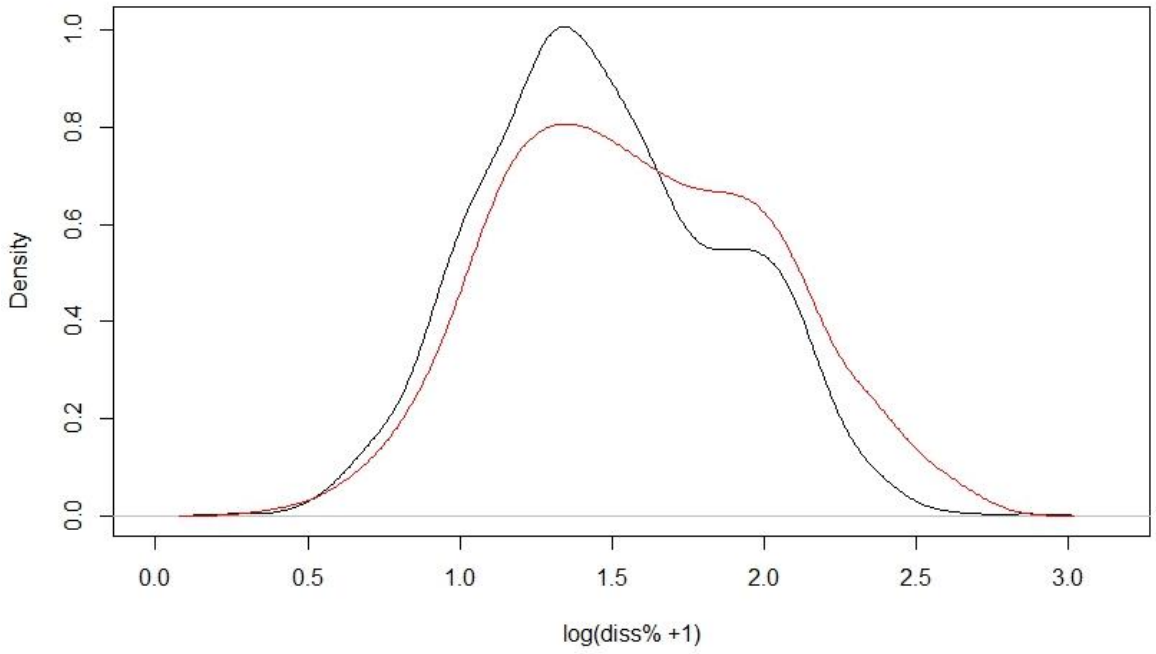


Figure 64 : Courbes de densité des scores de dissimilarité en log(score+1) pour les imitations lexicalisées (Phrases = P, courbe noire) et délexicalisées (Logatomes = L, courbe rouge)

L'inspection visuelle des distributions des scores des phrases et des logatomes révèle deux modes. Nous pouvons donc douter de leur normalité. Le test de normalité de Shapiro-Wilk indique en effet que les deux ensembles de données ne suivent pas une distribution normale.

Shapiro-Wilk normality test	
Phrases	Logatomes
W = 0.99083	W = 0.99169
p-value = 4.531e-06	p-value = 1.714e-05

Tableau 32 : Test de normalité des scores de dissimilarité de P et L

Malgré ces distributions non-normales, nous considérons que l'équilibre de population des deux échantillons (1017 L vs. 1039 P) ne met pas en danger la robustesse d'une analyse de variance (Howell, 2006, pp. 325–326).

L'inspection des moyennes des scores de dissimilarité indique que les imitations de type P ont une dissimilarité moins grande avec le modèle que les imitations de type L.

	Diss% (e.s.)	Log(Diss% + 1) (e.s.)
Type P	3.75724 (±.09951)	1.4783 (±.01861)
Type L	4.37621 (±.07074)	1.58307 (±.01323)

Tableau 33 : moyennes des imitations de type P et de type L. Les scores sont exprimés en taux de dissimilarité. Plus il est élevé, plus grande est cette dernière.

Une analyse de la variance à un facteur (Type P *vs*. Type L) révèle une différence statistiquement significative entre ces moyennes de scores de similarité des imitations de type P et de type L ($F(1, 2054) = 31.707$, $p. < .001$).

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
L_ou_P	1	5.64	5.6417	31.707	2.039e-08
Residuals	2054	365.47	0.1779		

Tableau 34 : Résumé de l'Anova à un facteur pour les imitations de types P et L

De manière corollaire, la moyenne des scores de l'ensemble des imitations par bloc expérimental indique une similarité prosodique meilleure durant la production des imitations du bloc PLP que durant les imitations de LPL, comme le suggère la Table 35.

	Diss% (e.s.)	Log(Diss% + 1) (e.s.)
Bloc PLP	3.93772 (±.10029)	1.5102 (±.01873)
Bloc LPL	4.18936 (±.07095)	1.54974 (±.01325)

Tableau 35 : moyennes de scores de dissimilarité des imitations produites durant les blocs PLP et LPL

Une analyse de la variance à un facteur (Bloc PLP *vs.* Bloc LPL) indique une différence statistiquement significative entre la moyenne des scores des imitations en fonction du bloc (F (1, 2054) = 4,38, *p.* < .05).

```
Analysis of Variance Table

            Df Sum Sq  Mean Sq  F value    Pr(>F)
Bloc         1   0.79  0.78969     4.38    0.03649 *
Residuals 2054 370.32  0.18029
```

Tableau 36 : Résumé de l'Anova à un facteur pour les productions des deux blocs

Ainsi, nos premières analyses tendent à montrer que les productions de type P reçoivent des scores de dissimilarité moindres que les imitations de type L. De manière mécanique, les imitations du bloc PLP (en mélangeant les types P et les types L) reçoivent en moyenne des scores de dissimilarité moins élevés que les imitations du bloc LPL en raison de la surreprésentation des imitations de type P dans le bloc PLP (2P pour 1 L *vs.* 2L pour 1P).

Aucune de nos hypothèses ne peut être infirmée ou validée ici, malgré des différences statistiquement significatives. En effet, **H5bis**, afférente au comportement de la mesure même, rend compliquée la tâche de conclure sur ce point.

Plusieurs solutions nous sont alors offertes :

- Considérer que les imitations prosodiques de type L sont systématiquement de moins bonne qualité que les imitations prosodiques de type P quel que soit le sujet qui les produit car les locuteurs ne parviennent pas à les produire.
- Accepter **H5bis** et considérer que la mesure TSR est légèrement biaisée lorsque qu'on compare une phrase et son logatome, en raison de la différence de contenu segmental. Il faudrait dans ce cas tenir compte de ce biais pour interpréter le résultat de la mesure.

Afin de pallier ce problème nous proposons d'étudier l'impact du facteur Sujet sur ces différences de score entre production de type P et production de type L au moyen d'une analyse de variance à plusieurs facteurs (Type L ou P * Sujet).

Analysis of Variance Table					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
L_ou_P	1	5.64	5.6417	34.5402	4.862e-09
Sujet	3	28.94	9.6466	59.0593	< 2.2e-16
L_ou_P:Sujet	3	2.02	0.6725	4.1175	0.006369
Residuals	2048	334.52	0.1633		

Tableau 37 : Résumé de l'Anova à plusieurs facteurs (Type * Sujet)

Ces résultats de l'analyse de variance indiquent :

- Une différence significative entre la moyenne des scores de dissimilarité des productions de type L et de type P
- Une différence significative entre la moyenne des scores de dissimilarité en fonction du sujet qui les produit (nous y reviendrons plus loin)
- Une différence significative des moyennes d'interaction Type * Sujet.

Afin de saisir l'impact du sujet sur la moyenne des scores, nous proposons l'utilisation du test HSD de Tukey (Honestly Significant Difference) dans la mesure où nos échantillons sont assez équilibrés (N_{E1} = 503 ; N_{E2} = 525 ; N_{NE1} = 519 ; N_{NE2} = 509)[76]. De plus, la fonction `R` que nous avons utilisée (`TukeyHSD`) intègre une correction pour d'éventuels déséquilibres dans la taille des échantillons.

	diff	lwr	upr	p adj
P:E1-L:E1	-0.10145050	-0.2107999781	0.007898977	0.0918367
P:E2-L:E2	-0.03541522	-0.1424628733	0.071632434	0.9739770
P:NE1-L:NE1	-0.07819921	-0.1858530372	0.02945460	0.3496615
P:NE2-L:NE2	-0.20559793	-0.3143043093	-0.096891544	0.0000003

Tableau 38 : Rapport du test HSD de Tukey pour l'interaction Type * Sujet

Les différences rapportées dans la Table 38 montrent que les moyennes de scores de dissimilarité des imitations de type L de tous les sujets sont plus élevés que leurs moyennes des imitations de type P (la dissimilarité est donc plus grande dans les productions de type L). Ceci étant dit, cette différence n'est significative que pour le sujet NE2.

Ce résultat porte un éclairage sur les hypothèses **H5** et **H5bis**. En effet, nous avons pu observer que les productions de type L ont une dissimilarité prosodique systématiquement plus élevée que les productions de type P. Ceci tendrait à nous faire accepter les deux hypothèses. Cependant, nous avons finalement constaté l'impact particulier du sujet Non

[76] Nous indiquons ici les effectifs d'imitation retenues pour chaque sujet (E1 & E2, locuteurs experts ; NE1 & NE2, locuteurs non experts)

Expert NE2 dans nos résultats précédents. Ainsi, il nous semble falloir relativiser l'hypothèse H5 en l'acceptant partiellement :

Les P sont systématiquement mieux notées que les L, mais cette différence n'est significative que dans le cas d'un sujet naïf, sans aucune expertise dans les domaines linguistiques.

Ce faisant, nous disposons également d'une piste nouvelle pour interpréter la mesure de TSR (et donc **H5bis**) : d'après nos résultats, il conviendrait d'appliquer une légère correction à la baisse de la mesure quand elle compare une phrase et son logatome.

Enfin, ce premier aperçu des différences entre nos sujets tend à confirmer partiellement **H2**, selon laquelle les imitateurs experts auraient une meilleure performance que les locuteurs non-experts.

2.3 Test de H3 : un effet d'entraînement au fil de la tâche ?

Etant donné que les imitateurs devaient faire de nombreuses imitations des mêmes modèles, nous avons postulé qu'il pourrait y avoir un effet d'entraînement lors du recueil du CL.

H3 postulait en effet que la performance des locuteurs progresserait à mesure du recueil, *i.e,* le taux de dissimilarité prosodique de leurs imitations diminuerait.

Afin de tester cette hypothèse, nous avons effectué une Anova à deux facteurs (Bloc et Répétition) pour déterminer s'il y avait une augmentation significative de la performance moyenne des sujets au cours de la tâche.

Nous en rapportons les résultats Table 39.

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Bloc	1	0.8	0.7897	4.377	0.0365
as.character(Repetition)	2	0.2	0.0771	0.427	0.6524
Bloc:as.character(Repetition)	2	0.3	0.1595	0.884	0.4133
Residuals	2050	369.9	0.1804		

Tableau 39 : Résumé de l'analyse de variance à deux facteurs (Bloc * Répétition)

Hormis la différence significative entre les moyennes des scores en fonction du bloc expérimental (déjà rapportée précédemment) nous ne relevons aucun effet significatif du moment de la répétition dans le bloc. En d'autres termes, les imitations prosodiques produites

au début de la tâche ne sont pas significativement plus réussies que les imitations produites à la fin de la tâche.

Nos investigations, en intégrant les différents Sujets comme facteur, n'ont apporté aucune information supplémentaire.

A la vue de ces résultats, il semble sage de repousser **H3**. Il n'y aurait pas eu d'effet d'entraînement durant cette tâche. Dans le cas des sujets experts et/ou performants, nous pourrions admettre que leur compétence d'imitation prosodique est plafonnée. Dans le cas de sujets naïfs et/ou non performants, nous pourrions supposer qu'un entraînement plus long et durable est nécessaire pour améliorer cette capacité d'imitation/contrôle prosodique.

2.4 Test de la dégradation du pattern prosodique au cours d'un même trial

Nous testons ici **H6**, selon laquelle la similarité prosodique entre l'imitation et le modèle serait dégradée au fur et à mesure des trois productions de chaque *trial*.

De manière parallèle, nous pourrons tester **H4**, qui postulait que les logatomes produits durant les *trials* PLP seraient plus similaires au modèle que les logatomes en première position des *trials* LPL.

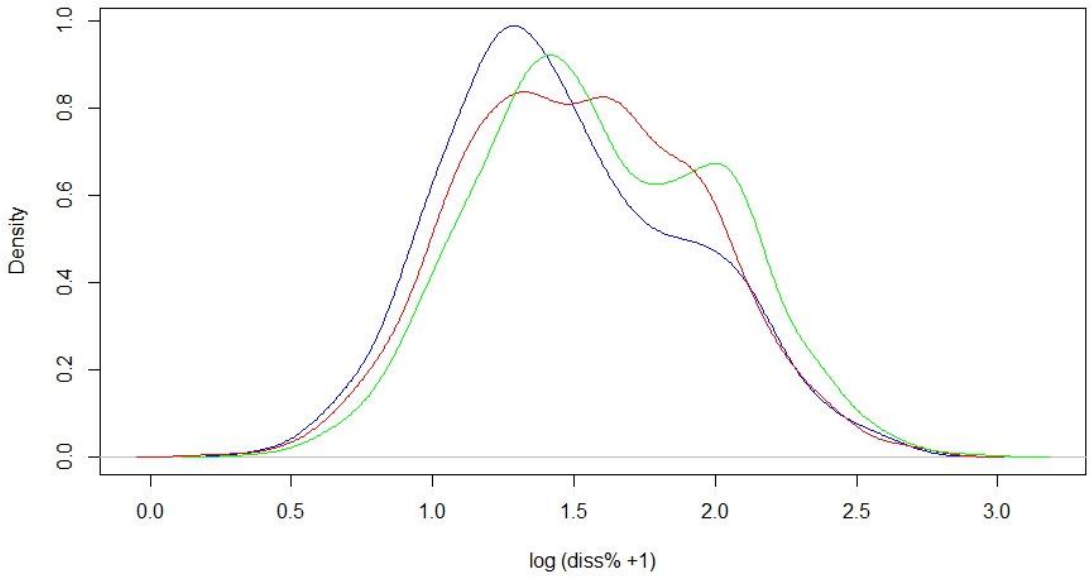


Figure 65 : Courbes de densité des imitations en fonction de leur position dans le trial.Première production (X1 = bleu), seconde (X2 = rouge), troisième (X3 = vert).

Les courbes de densité de X1 (en bleu) et de X3 (en vert) présentent une légère asymétrie, et la courbe de X3 a deux modes apparents. La courbe de X2 (en rouge) semble un peu plus normale.

Shapiro-Wilk normality test		
X1	X2	X3
W = 0.98411	W = 0.99574	W = 0.99062
p-value = 8.631e-07	p-value = 0.05751	p-value = 0.0002403

Tableau 40 : Test de normalité des distributions des scores en fonction de leur position dans les trials

Encore une fois, ces distributions (à l'exception de celle de X2) ne sont pas normales. Cependant, l'équilibre de population entre les échantillons (N_{X1} = 686, N_{X2} = 686, N_{X3} 684) rend moins périlleux la violation des conditions d'application des Anovas.

L'inspection des moyennes de score de dissimilarité en fonction de la position de l'imitation dans le trial montre que les imitations produites en première position (X1) présentent une dissimilarité avec le modèle moins grande que les imitations produites en seconde (X2) ou en troisième position (X3).

	Diss % (s.e.)	Log(Diss% + 1) (s.e.)
Position X1	3.78090 (±.08644)	1.47054 (±.01612)
Position X2	4.02277 (±.12224)	1.52683 (±.02280)
Position X3	4.38750 (±.12233)	1.59319 (±.02281)

Tableau 41 : Moyenne des scores de dissimilarité en fonction de la position de l'imitation dans les trials (X1, X2 ou X3).

Une analyse de la variance à un facteur (Position) des scores de dissimilarité indique que les différences de moyennes que nous avons observées précédemment sont statistiquement significatives.

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Position	2	5.16	2.58177	14.484	5.672e-07
Residuals	2053	365.95	0.17825		

Tableau 42 : Analyse de variance des scores de dissimilarité en fonction de leur position

L'analyse supplémentaire fournie par le test HSD de Tukey confirme qu'il y aurait une dégradation de la similarité prosodique des imitations avec leur modèle, significative sur X2, et plus fortement sur X3.

	diff	lwr	upr	p adj
X2-X1	0.05629062	0.002823374	0.1097579	0.0362842
X3-X1	0.12265153	0.069145220	0.1761578	0.0000003

Tableau 43 : Rapport du test HSD de Tukey pour le facteur position

Pour aller plus loin dans nos analyses, nous avons conduit une autre Anova en prenant pour facteurs la position (X1, X2 ou X3) et le type (P ou L) de l'imitation afin de pouvoir observer leur interaction.

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Position	2	5.16	2.5818	14.6909	4.625e-07
L_ou_P	1	5.61	5.6148	31.9498	1.803e-08
Position:L_ou_P	2	0.07	0.0351	0.1995	0.8192
Residuals	2050	360.27	0.1757		

Tableau 44 : Résumé de l'Anova à deux facteurs (Position * Type) des scores de dissimilarité

Ce test n'apporte pas d'information supplémentaire quant à la significativité de l'interaction Position * Type dans la dégradation des patterns prosodiques. Ceci étant dit, l'observation des différences individuelles au moyen du test HSD de Tukey pourrait apporter une granularité supplémentaire à notre analyse. Nous rapportons dans la Table 45 les différences relevées au sein des blocs ainsi qu'une différence donnant une première piste pour **H4**.

Bloc	diff	lwr	upr	p adj	Résumé
PLP					
X1:P-X2:L	-0.16583534	-0.25721901	-0.074451663	0.0000037	X1P<X2L**
X3:P-X2:L	-0.03040362	-0.12198371	0.061176462	0.9341438	Ns
X3:P-X1:P	0.13543172	0.04445265	0.226410785	0.0003270	X1P<X3P**
LPL					
X2:P-X1:L	-0.05323279	-0.14448330	0.038017723	0.5559154	Ns
X2:P-X3:L	-0.16224680	-0.25342931	-0.071064288	0.0000063	X2P<X3L**
X3:L-X1:L	0.10901401	0.01723529	0.200792731	0.0093486	X1L<X3L**
H4					
X2:L-X1:L	0.04494311	-0.04697107	0.136857303	0.7302953	Ns

Tableau 45 : Rapport du test HSD de Tukey pour l'interaction Position* Type.La colonne Résumé a été ajoutée par nos soins afin de faciliter la lecture des différences. Les astérisques relèvent les différences significatives d'après le test.

Dans ces différences relevées par le test HSD, nous retrouvons une tendance illustrée précédemment puisque les items de type P, quelle que soit leur position, ont en moyenne des scores de dissimilarité moindre que les items de type L. Cependant, cette différence n'est pas statistiquement significative quand l'item P succède à l'item L.

Par ailleurs, nous remarquons une dégradation systématique de X1 à X3 pour des items de même type. Ce fait est intéressant à souligner, car les items de même type pâtissent des mêmes biais de mesure (**H5bis**). Nous pouvons alors estimer avec une confiance mesurée que la similarité prosodique entre un modèle initial et des imitations successives se dégrade à au fur et à mesure de la production des imitations. Ceci tendrait donc à valider notre hypothèse **H6**.

Enfin, il semble que notre hypothèse **H4** doive être rejetée car nous n'avons pas trouvé de différence significative entre les moyennes de score des logatomes X1 du bloc LPL et les logatomes X2 des blocs PLP.

2.5 Performance individuelle des sujets en imitation prosodique

Nous testons ici **H2**, selon laquelle les sujets Experts (E1 et E2) auraient de meilleures performances en imitation prosodique que les sujets Non Experts (NE1 et NE2). Les experts sont en effet des enseignants de langue en exercice, praticiens de MVT. Nous attendons donc de leur part une expertise dans la production des logatomes, un procédé courant de leur pratique.

L'inspection des courbes de densité des scores des 4 sujets laisse supposer que le critère de normalité ne sera pas rempli. En ce qui concerne la performance qu'on peut attendre de chacun, nous sommes surpris de noter que le sujet NE1 (en bleu) semble avoir surpassé les deux experts (E1, en noir, E2, en rouge). Enfin, NE2, paraît avoir eu la performance la plus basse au cours de la tâche.

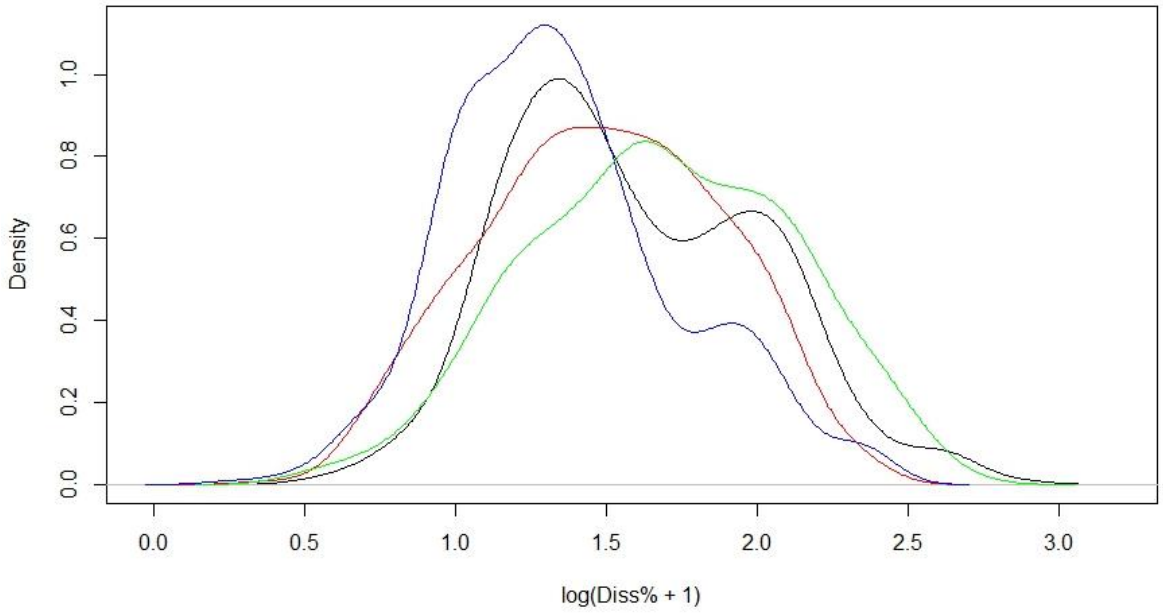


Figure 66 : Courbes de densité des scores d'imitation prosodique des 4 locuteurs du CL : E1 (noir), E2 (rouge), NE1 (bleu), NE2 (vert).

Shapiro-Wilk normality test			
E1	E2	NE1	NE2
W = 0.97978	W = 0.9907	W = 0.9783	W = 0.99261
p-value<.001	p-value = 0.00217	p-value <.001	p-value = 0.01294

Tableau 46 : test de normalité des scores d'imitation des différents sujets.

Le test de normalité confirme en effet que ce critère ne sera pas rempli. Cependant, encore une fois, nos effectifs dans les échantillons sont équilibrés (N_{E1} = 503, N_{E2} = 525, N_{NE1} = 519, N_{NE2} = 509). Ceci étant dit, l'inspection des moyennes de scores des 4 sujets semble confirmer l'impression que nous avions en regardant les courbes de densité : NE1, malgré son statut de sujet naïf aurait eu la meilleure performance en termes d'imitation prosodique. Les deux experts suivraient sa performance et NE2 fermerait la marche.

	Diss % (s.e.)	Log(Diss% + 1) (s.e.)
Sujet : E1	4.40330 (±.09767)	1.59616 (±.01821)
Sujet : E2	3.77590 (±.13667)	1.48738 (±.02548)
Sujet : NE1	3.22499 (±.13705)	1.36306 (±.02555)
Sujet : NE2	4.87897 (±.13772)	1.67932 (±.02567)

Tableau 47 : Moyenne des scores de dissimilarité par sujet

Nous avons donc conduit une Anova à 1 facteur (Sujet) pour estimer si les différences de moyennes de scores entre les sujets étaient statistiquement significatives. (Table 48).

```
Analysis of Variance Table

            Df Sum Sq Mean Sq F value    Pr(>F)
Sujet        3  28.97  9.6562  57.913 < 2.2e-16 ***
Residuals 2052 342.15  0.1667
```

Tableau 48 : résumé de l'anova à un facteur (Sujet)

Ce test relève un effet significatif du facteur Sujet sur la moyenne de scores de dissimilarité prosodique. Pour saisir les tendances internes à cet effet, nous avons conduit un test HSD de Tukey.

	diff	lwr	upr	p adj	Résumé
E2-E1	-0.10878282	-0.17428795	-0.04327770	0.0001202	E2<E1 **
NE1-E1	-0.23310326	-0.29879340	-0.16741313	0.0000000	NE1<E1**
NE2-E1	0.08315819	0.01715123	0.14916515	0.0066769	NE2>E1*
NE1-E2	-0.12432044	-0.18930786	-0.05933302	0.0000056	NE1>E2**
NE2-E2	0.19194101	0.12663335	0.25724866	0.0000000	NE2>E2**
NE2-NE1	0.31626145	0.25076823	0.38175467	0.0000000	NE2>NE1**

Tableau 49 : Rapport du test HSD de Tukey pour l'anova à un facteur (Sujet)

L'observation de ces différences au moyen du HSD de Tukey semble confirmer la hiérarchie que nous constations précédemment à propos de la performance d'imitation prosodique des sujets experts et naïfs.

Le sujet NE2 a une performance significativement inférieure à celle des autres sujets, ce qui se traduit par une moyenne de scores de dissimilarité plus élevée que les autres. Ceci étant dit, le sujet NE1 a déjoué nos pronostics puisqu'il surclasse les deux experts. A l'issue de ce test, nous constatons donc la hiérarchie de performance suivante entre nos sujets :

➔ NE1 > E2 > E1 > NE2

Ce premier test, tous blocs et types d'imitations confondus permet de dégager une première tendance quant à la performance de ces sujets. A présent, nous poursuivrons nos analyses en conduisant une Anova pour étudier l'interaction entre les blocs expérimentaux (PLP vs. LPL) et les sujets.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sujet	3	28.97	9.6562	58.1249	< 2e-16
Bloc	1	0.82	0.8183	4.9257	0.02657
Sujet:Bloc	3	1.10	0.3655	2.2001	0.08613
Residuals	2048	340.23	0.1661		

Tableau 50 : Résumé de l'anova à deux facteurs (Bloc * Sujet)

Ce test relève, comme précédemment, un effet statistiquement significatif du facteur Sujet sur les moyennes de scores, de même pour le facteur Bloc. Ces deux résultats ne nous surprennent pas : nous venons de débattre des sujets, et le test de **H5** et **H5bis** a illustré la différence intrinsèque entre les scores des deux blocs, liée à la surreprésentation des phrases de type P (mieux notées) dans le bloc PLP et de type L (moins bien notées) dans le bloc LPL. En revanche, nous sommes surpris de noter que l'interaction de ces deux facteurs ne semble pas statistiquement significative.

Afin d'observer plus finement les interactions entre ces deux facteurs, nous proposons le recours au test HSD de Tukey (Table 51)

	diff	lwr	upr	p adj	Résumé
Intra individuel					
(PLP vs. LPL)					
E1:PLP-E1:LPL	-0.077932343	-0.18821033	0.03234564	0.3867563	Ns
E2:PLP-E2:LPL	0.037083311	-0.07085934	0.14502597	0.9678908	Ns
NE1:PLP-NE1:LPL	-0.050542660	-0.15910747	0.05802215	0.8517433	Ns
NE2:PLP-NE2:LPL	-0.070870626	-0.18049840	0.03875715	0.5085206	Ns
Inter individuel					
(PLP)					
E2:PLP-E1:PLP	-0.051317145	-0.16042917	0.05779488	0.8448898	Ns
NE1:PLP-E1:PLP	-0.219534583	-0.32895598	-0.11011319	0.0000000	NE1<E1**
NE2:PLP-E1:PLP	0.086820430	-0.02291669	0.19655755	0.2415211	Ns
NE1:PLP-E2:PLP	-0.168217438	-0.27657527	-0.05985961	0.0000719	NE1<E2**
NE2:PLP-E2:PLP	0.138137575	0.02946093	0.24681422	0.0029839	NE2>E2**
NE2:PLP-NE1:PLP	0.306355013	0.19736776	0.41534227	0.0000000	NE2>NE1**
(LPL)					
E2:LPL-E1:LPL	-0.166332799	-0.27545391	-0.05721169	0.0001079	E2<E1**
NE1:LPL-E1:LPL	-0.246924266	-0.35635237	-0.13749616	0.0000000	NE1<E1**
NE2:LPL-E1:LPL	0.079758714	-0.03041046	0.18992789	0.3541409	Ns
NE1:LPL-E2:LPL	-0.080591467	-0.18874190	0.02755896	0.3159929	Ns
NE2:LPL-E2:LPL	0.246091512	0.13719131	0.35499171	0.0000000	NE2>E2**
NE2:LPL-NE1:LPL	0.326682980	0.21747517	0.43589079	0.0000000	NE2>NE1**

Tableau 51 : Rapport du test HSD de Tukey pour l'anova à deux facteurs (Sujet * Bloc)

En premier lieu, il est intéressant de noter que les différences intra individuelles entre les blocs ne sont pas significatives, malgré le léger biais de la TSR dans la mesure des

imitations de type L et P. Cela souligne également qu'il n'y a pas eu d'entraînement des sujets à la tâche durant la session, malgré de nombreuses itérations des mêmes imitations (**H3**).

En ce qui concerne les différences interindividuelles, nous pouvons noter les différences suivantes :

- PLP
    - NE1 a surclassé tous les autres sujets de manière significative, malgré son statut de naïf
    - La différence entre les deux experts n'est pas significative
    - Bien qu'E1 ne soit pas significativement meilleur que NE2, E2 est significativement meilleur que NE2
    - Ainsi, dans ce bloc, nous pouvons estimer la hiérarchie comme suit
        - → NE1 > E2 ≥ E1 ≥ NE2
- LPL
    - NE1 surclasse encore significativement les autres sujets, sauf E2, qui atteint une performance approchante
    - La performance d'E2 dépasse celle d'E1 et de NE2
        - → NE1 ≥ E2 > E1 ≥ NE2

Par ces tests, nous avons pu affiner notre compréhension de la performance des sujets en fonction des blocs. Nous proposons à présent de nous intéresser à ces différences de performance en fonction des sujets et du type d'imitation (L ou P). Nous avons donc conduit une Anova avec ces facteurs (Sujet * Type d'imitation).

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sujet	3	28.97	9.6562	59.1181	< 2.2e-16
L_ou_P	1	5.61	5.6129	34.3639	5.315e-09
Sujet:L_ou_P	3	2.02	0.6725	4.1175	0.006369
Residuals	2048	334.52	0.1633		

Tableau 52 : Résumé de l'anova pour l'interaction Sujet * Type d'imitation

Nous pouvons noter ici que l'interaction entre Sujet et Type d'imitation semble avoir un effet statistiquement significatif sur les moyennes de scores de dissimilarité. Pour investiguer les différences individuelles, nous recourons au test HSD de Tukey (Table 53).

L'observation des différences intra individuelles révèle que le sujet NE2 uniquement voit sa performance d'imitation prosodique significativement dégradée lors de l'imitation délexicalisée, ce que nous relevions précédemment lors du test de **H5** & **H5bis**.

En ce qui concerne les différences interindividuelles dans la production du même type d'imitation (P ou L), nous remarquons que :

- Imitation de type P
    - NE1 a une performance significativement meilleure que tous les autres sujets
    - La différence entre les experts n'est pas significative
    - Comme précédemment, malgré cette absence significative de différence entre experts, seul l'expert E2 obtient des scores significativement meilleurs que NE2.
        - → NE1 > E2 ≥ E1 ≥ NE2

	diff	lwr	upr	p adj	Résumé
Intra individuel					
(L *vs.* P)					
E1:P-E1:L	-0.10145050	-0.2107999781	0.007898977	0.0918367	Ns
E2:P-E2:L	-0.03541522	-0.1424628733	0.071632434	0.9739770	Ns
NE1:P-NE1:L	-0.07819921	-0.1858530372	0.029454609	0.3496615	Ns
NE2:P-NE2:L	-0.20559793	-0.3143043093	-0.096891544	0.0000003	P<L**
Inter individuel					
(Type L)					
E2:L-E1:L	-0.14179944	-0.2506210089	-0.032977870	0.0020321	E2<E1**
NE1:L-E1:L	-0.24465476	-0.3535804721	-0.135729052	0.0000000	NE1<E1**
NE2:L-E1:L	0.13593918	0.0264819175	0.245396436	0.0041976	NE2>E1**
NE1:L-E2:L	-0.10285532	-0.2109215983	0.005210953	0.0755678	Ns
NE2:L-E2:L	0.27773862	0.1691365846	0.386340648	0.0000000	NE2>E2**
NE2:L-NE1:L	0.38059394	0.2718875561	0.489300321	0.0000000	NE2>NE1**
(Type P)					
E2:P-E1:P	-0.07576416	-0.1833484258	0.031820108	0.3915015	Ns
NE1:P-E1:P	-0.22140348	-0.3294860537	-0.113320898	0.0000000	NE1<E1**
NE2:P-E1:P	0.03179175	-0.0768061055	0.140389606	0.9871512	Ns
NE1:P-E2:P	-0.14563932	-0.2522705783	-0.039008056	0.0009273	NE1<E2**
NE2:P-E2:P	0.10755591	0.0004023906	0.214709428	0.0483600	NE2>E2*
NE2:P-NE1:P	0.25319523	0.1455414035	0.360849049	0.0000000	NE2>NE1**

Tableau 53 : Rapport du test HSD de Tukey pour les interactions Sujet * Type d'imitation

- Imitation de type L
    - La différence de performance entre NE1 et E2 n'est pas significative
    - NE1 et E2 ont une performance moyenne significativement meilleure que celle de E1

- E1 lui-même a une performance moyenne significativement meilleure que celle de NE2
- NE2 a une performance moyenne significativement moins bonne que celle de tous les autres locuteurs
  - → NE1 ≈ E2 ≥ E1 > NE2

Etant donnés ces résultats, notre hypothèse **H2** postulant que les sujets Experts auraient de meilleures performances que les sujets Non Experts se trouve partiellement confirmée. En effet, un de nos deux sujets naïfs a surpassé tous les autres sujets.

Nous pouvons voir plusieurs explications à cela. En premier lieu, nous devrions nous interroger sur la naïveté réelle de NE1 en termes de comportement vocal. Ce sujet pouvait être considéré comme débutant la MVT. Celle-ci lui a seulement été enseignée pendant un à deux semestres dans sa formation universitaire. Pourtant sa performance a été très élevée. Peut-être pourrions-nous considérer que ce temps de formation a été suffisant à ce sujet pour maîtriser le procédé du logatome et/ou que NE1 présente un talent phonétique particulier.

Malgré cet écart par rapport à notre hypothèse **H2**, les experts E1 et E2 sont parvenus à avoir un niveau de production plus élevé que le sujet naïf NE2. Ce dernier, contrairement à NE1, peut être considérer comme un vrai sujet naïf : bien qu'il ait été familier des principes de la MVT, il n'en a pas eu de formation pratique et elle ne lui a pas été enseignée en présentiel. Son seul contact avec la méthode avait été le cours en ligne de l'UOH (Billières et al., 2013).

Ainsi, cette étude des performances individuelles nous permet de considérer que **H2** est probablement vraie, en pondérant notre propos de la manière suivante : il convient de définir avec plus de rigueur la notion de naïf dans notre cadre.

Par ailleurs, ces analyses nous rassurent une nouvelle fois quant à **H5bis** et le supposé biais de mesure de la TSR : bien que les imitations de type P reçoivent en moyenne des scores de dissimilarité inférieurs aux imitations de type L, cette différence n'était significative que dans le cas d'un sujet réellement naïf, pour qui la tâche était plus difficile et donc, moins réussie.

2.6 Effet de la longueur du modèle sur la production du logatome

Afin de clore ces analyses sur la similarité prosodique dans le CL, nous nous intéresserons à l'effet de la longueur du modèle entendu sur le logatome. **H1** postulait en effet que le score de similarité prosodique diminuerait à mesure que le nombre de syllabes de l'énoncé à reproduire augmente.

Les locuteurs du CL ont entendu des modèles faisant de 1 à 12 syllabes. En deçà de 4 points, la TSR ne peut pas produire de mesure. Nous avions donc conservé, pour les analyses précédentes, les énoncés faisant au moins 4 syllabes. Par ailleurs, **H1** traite explicitement des logatomes, c'est pourquoi nous avons uniquement conservé pour ces analyses les imitations de type L, soit : 1017 observations.

Ceci résulte en un ensemble de 8 échantillons aux populations très inégales (Figure 67).

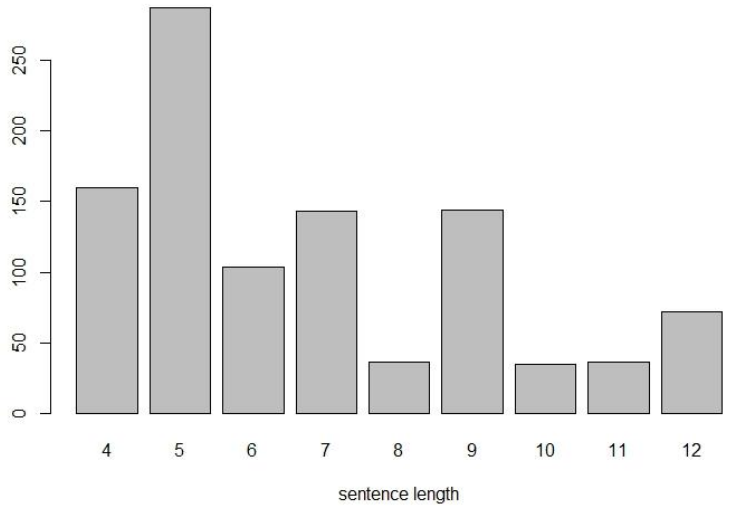


Figure 67 : Taille des échantillons « Type L » en fonction de la longueur du modèle

En raison de ces échantillons de tailles variées, nous ne soumettrons pas ces données à une analyse de variance : en effet, nous ne nous attendons pas à ce que leurs distributions soit normales. Nous nous limiterons donc à une description des données et aux pistes qu'elles nous suggèrent.

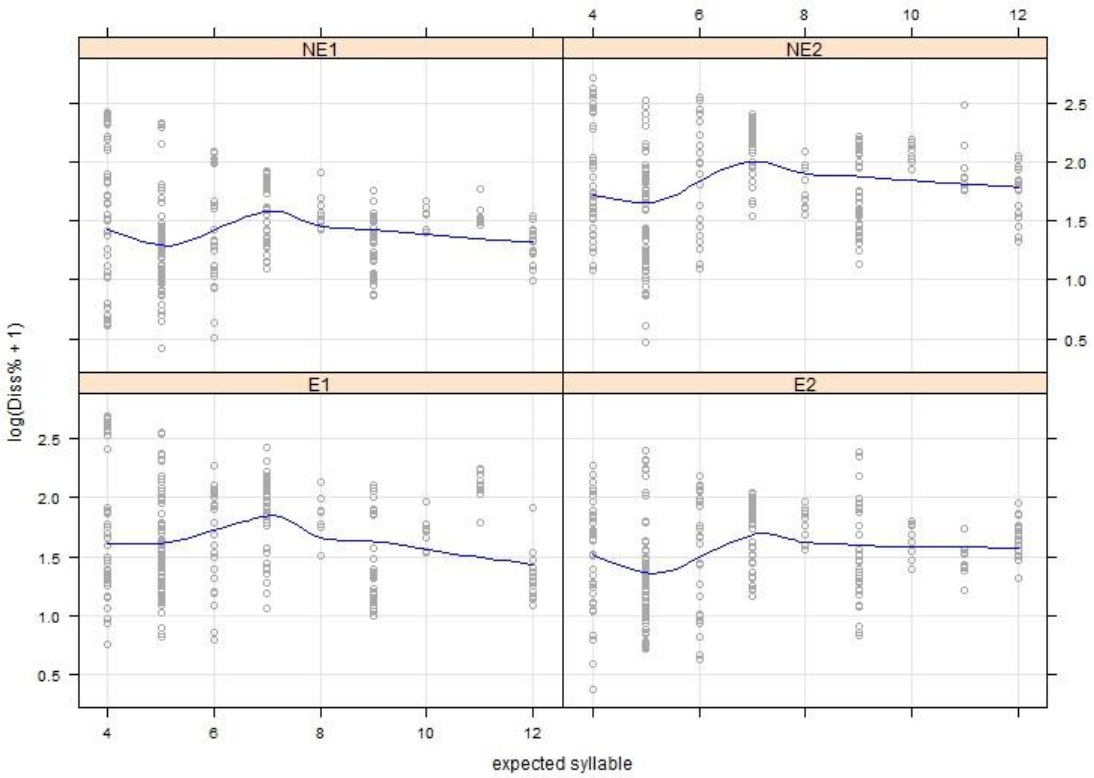


Figure 68 : Score de dissimilarité (ordonnées) en fonction du nombre de syllabes du modèle (abscisses)

La Figure 68 présente les scores de dissimilarité de chacun des sujets en fonction du nombre de syllabes contenues dans le modèle. Nous y repérons plusieurs mouvements intéressants :

- Etonnament, les items de 4 syllabes semblent avoir une dissimilarité plus grande que les items de 5 syllabes, ceci, dans les productions de tous les sujets
- Les items de 5 syllabes sont les items les plus similaires au modèle chez tous les sujets
- La dissimilarité augmente sur les items de 6 et 7 syllabes, puis diminue sur les items de 8 syllabes.
- La dissimilarité stagne ou diminue sur les items 8 et plus syllabes.

Il nous faut formuler plusieurs remarques consécutives à ces observations.

En premier lieu, le fait que les logatomes issus d'un modèle faisant 4 syllabes aient des scores moins bons que les imitations issues de modèles faisant 5 syllabes est contre intuitif. Pour ce qui concerne la mesure, la tendance des scores des imitations résultant de modèles de 8 syllabes nous surprend également. Nous pensons que ces résultats sont liés à l'annotation ou à la transformation et mesure TSR :

- En deçà de 4 syllabes, le moindre écart par rapport au modèle pourrait résulter en une mesure disproportionnée de la similarité, notamment s'il y a un point manquant ou en trop
- Au-delà de 8 syllabes, l'écart requis pour faire dévier la mesure de dissimilarité doit probablement être plus grand.

Cela illustre une limite des mesures issues de *T-Function*, qui est leur sensibilité au bruit : l'étape du choix des points d'ancrage pour faire émerger les formes à comparer est donc cruciale.

Ceci étant dit, la TSR semble avoir été un choix approprié pour évaluer la similarité prosodique de logatome et de modèles lexicalisés : le fait que la mesure TSR s'ancre sur les syllabes des énoncés implique que leur omission ou leur ajout conduit à une mesure de dissimilarité plus élevée. Nous pouvons le constater en observant la Figure 69.

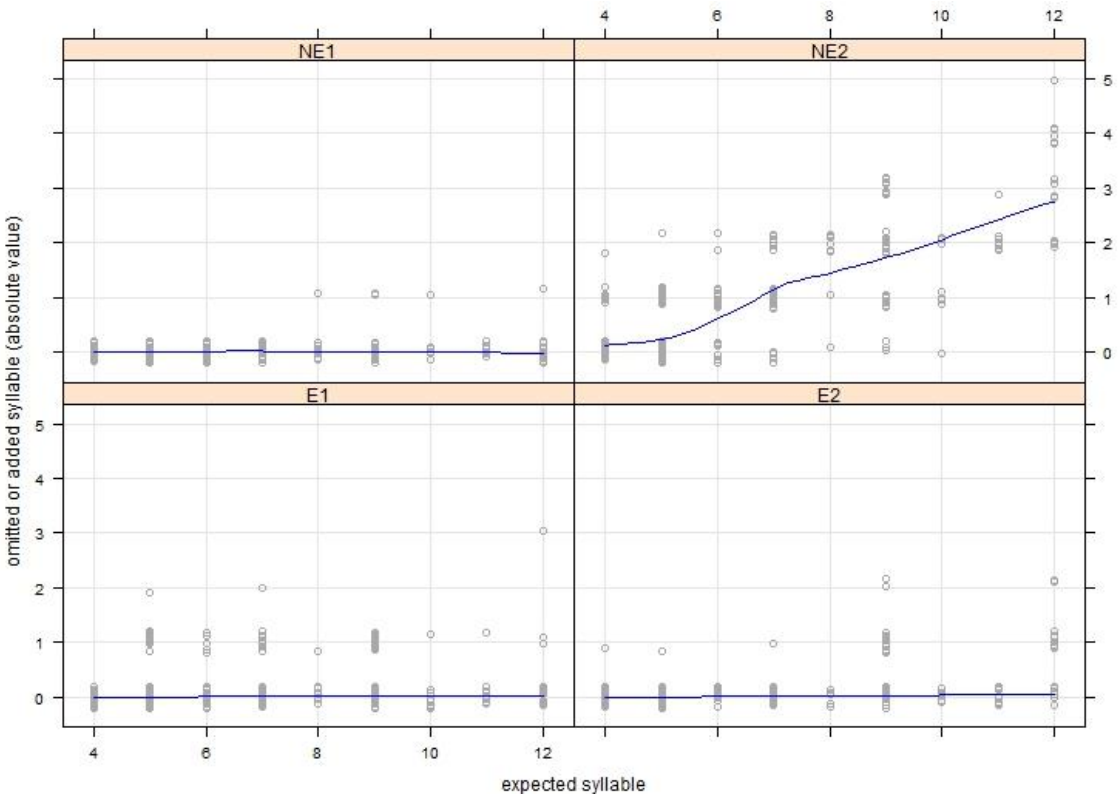


Figure 69 : Nombre de syllabes du modèle (abscisses) vs. Syllabes omises ou ajoutées (ordonnées, en valeur absolue). Nous avons introduit du *jitter* dans les valeurs des ordonnées afin de pouvoir observer la densité d'effectifs pour chaque longueur de phrase.

Sur cette figure, nous pouvons observer que les sujet NE1, E1 et E2 tendent à produire exactement le nombre de syllabes attendues. *A contrario*, NE2 omet ou ajoute presque

systématiquement des syllabes lorsqu'il produit des logatomes excédent 5 syllabes. Ces données descriptives pourraient expliquer pourquoi les imitations de type L de NE2 sont systématiquement moins bien évaluées que celles des autres sujets.

Enfin, cette description des données tend à montrer que les locuteurs Experts (+ NE1) ont peu de difficulté à traiter le nombre de syllabes des énoncés les plus longs. Ils font preuve d'une bonne capacité de transposition du modèle lexicalisé au logatome, délexicalisé.

Finalement, nous ne pouvons réellement décider si **H1** est vraie ; tout au plus pouvons-nous donner quelques pistes. Il semble qu'un trop petit nombre de syllabe fausse la TSR. De 5 à 7 syllabes, les résultats sont conformes aux prédictions de **H1**. Au-delà, les résultats donnés par la TSR nous laissent songeurs. De plus amples tests seraient nécessaires pour saisir le comportement de la TSR sur les énoncés les plus longs.

2.7 Synthèse et discussion

Pour cette étude, 2 locuteurs naïfs et 2 locuteurs experts dans le domaine de la MVT ont participé à l'enregistrement d'un corpus d'imitation lexicalisées (Phrases) et délexicalisées (Logatomes) à partir d'un dialogue préenregistré. Après écoute d'un modèle, les locuteurs devaient produire 3 énoncés :

- Dans le bloc PLP, une phrase, un logatome puis une phrase.
- Dans le bloc LPL, un logatome, une phrase puis un logatome.

Pour tenter d'infirmer ou de confirmer nos six hypothèses émises lors de la description du protocole d'enregistrement, nous avons appliqué systématiquement la mesure TSR entre chaque imitation et le modèle entendu avant sa production.

Au terme de nos analyses :

- Nous avons partiellement admis **H1** : il semble que la longueur du modèle ait une influence sur la réussite de la production du logatome, surtout si le locuteur est un vrai débutant en MVT.
- **H2** a également été partiellement acceptée : les locuteurs experts tendent à avoir de meilleures performances en termes d'imitation prosodique que notre locuteur vrai débutant (NE2). Notre autre locuteur Non Expert (NE1) a semblé faire preuve d'un réel talent dans ce domaine et/ou sa formation à la MVT a suffi à la préparer à l'exercice.
- Nous n'avons pas relevé d'effet d'entraînement au cours de la tâche, et nous avons donc rejeté **H3**.
- Nous avons également rejeté **H4** : contrairement à ce que nous pensions, prononcer une phrase avant un logatome n'améliore pas sa similarité avec le modèle ; *i.e.* entendre le modèle peut suffire à atteindre une performance satisfaisante.
- En ce qui concerne **H5** et **H5bis** :
    - **H5** : Tous les sujets ont obtenu de meilleurs scores de similarité prosodique durant les productions de type P. Nous avons donc accepté **H5** stipulant que les imitations lexicalisées tendent à être mieux réussies que les imitations délexicalisées.
    - **H5bis** : Nous avons également accepté H5bis : la TSR semble être légèrement biaisée en faveur des productions de type P. Cependant, ce biais n'est pas

suffisamment fort pour nous empêcher de distinguer un locuteur qui n'accomplit pas la tâche de manière satisfaisante.

- Enfin, nous avons remarqué une dégradation de la similarité prosodique entre le modèle entendu et les imitations successives : la seconde, puis la troisième imitation de chaque *trial* étaient en moyenne moins similaire au modèle que la première. Nous avons donc accepté **H6**.

En ce qui concerne la TSR, son usage pour l'exploration de ce corpus semble s'être révélé pertinent. La mesure a été à même de saisir la performance satisfaisante des locuteurs Experts, l'excellente performance du locuteur NE1, pourtant plus naïf que les Experts en MVT, et la performance moindre du locuteur NE2, un vrai débutant en MVT.

Cependant, nous avons également soulevé de nouvelles questions quant à cette mesure qui semble :

- Particulièrement sensible dès lors que le nombre de point est très réduit (4 ou moins)
- Moins sensible quand le nombre de points dépasse un certain seuil (8 ou plus)

Ainsi, il nous semble nécessaire de mener de nouveaux tests en contrôlant mieux la longueur des stimuli, pour mieux comprendre le comportement de la méthode TSR.

Par ailleurs, en utilisant la méthode TSR, nous avons privilégié le niveau rythmique, du fait de l'ancrage de l'annotation sur la voyelle de chaque syllabe. Il pourrait également être intéressant d'observer si la mesure TSP, laissée de côté pour ces analyses, est plus robuste à la taille des énoncés comparés.

A propos de la production de logatomes, notre étude de ce corpus a montré que des locuteurs formés aux pratiques de la MVT et ayant une bonne connaissance du système phonologique du français parvenaient à accomplir cette tâche de manière satisfaisante et avec une certaine constance.

Nos dernières données descriptives suggèrent d'ailleurs que ces locuteurs parvenaient à reproduire des énoncés comprenant jusqu'à 12 syllabes. En revanche, l'étude de ce corpus souligne également qu'une formation théorique et un entraînement à la pratique sont nécessaires pour produire ce procédé de manière convaincante.

Ceci rejoint l'idée que nous développions à la fin de notre chapitre 3 : la MVT est une méthode séduisante, flexible et efficiente ; cependant, sa maîtrise n'est pas aisée. La théorie ne suffit pas.

Enfin, notre étude a également souligné la dégradation du pattern prosodique original à mesure des imitations. Cependant, les raisons de cette dégradation demeurent peu claires. Il serait nécessaire que nous produisions des comparaisons croisées systématiques entre chaque énoncé de chaque trial en fonction de leurs position (X0 = modèle, X1, X2 et X3 = 3 imitations successives) :

- X0 vs. X1 ou X2 ou X3
- X1 vs. X2 ou X3
- X2 vs. X3

Ce design de comparaison pourrait tester l'hypothèse selon laquelle la similarité prosodique serait toujours plus grande entre deux imitations successives. En d'autres termes, les locuteurs se baseraient sur l'écho de la forme prosodique en mémoire de travail pour produire la forme prosodique suivante.

Si une telle hypothèse était confirmée, il faudrait alors considérer que les locuteurs font preuve d'une adaptation permanente à ce qu'ils entendent (ils imitent), tout en réinventant à chaque instant leur pratique (ils font preuve d'individualité dans leur comportement).

Finalement, la mesure TSR offre des perspectives expérimentales intéressantes, cependant il est primordial d'approfondir notre compréhension de ce type de mesures (TSR et TSP) pour pouvoir par la suite systématiser leur usage avec une confiance suffisante.

3. Perspectives logicielles pour l'entraînement de l'enseignant aux procédés et techniques de la MVT

Si, initialement, nous voyions l'imitation prosodique comme un enjeu pour l'apprentissage phonétique de l'apprenant, notre travail doctoral nous a conduit à considérer le point de vue de l'enseignant, et plus particulièrement celui de l'enseignant en devenir. Dans nos premiers travaux (Nocaudie, 2012; Nocaudie & Astésano, 2012), l'étude de l'imitation prosodique en L1 française était un prérequis nécessaire à l'étude de l'imitation prosodique en L2 française. Cependant, notre recherche méthodologique nous a conduit à reconsidérer notre direction.

Durant nos années de thèse, nous avons en effet constaté l'importance de participer à la formation des enseignants dans le domaine de la remédiation phonétique. Si l'apprenant de L2 doit acquérir de nouveaux comportements paroliers, l'enseignant souhaitant aider son apprenant à mieux prononcer doit comprendre les causes de l'erreur de prononciation et connaître les moyens d'y remédier. Plus important : il doit également être capable de les mettre en œuvre. En d'autres termes, l'enseignant doit maîtriser des techniques vocales.

En ce sens, nous avons souligné dans notre chapitre 3 le développement des formations universitaires à la remédiation phonétique. Celles-ci se développent à la fois en présentiel, mais aussi sur Internet (Billières et al., 2013). Dans ce dernier cas, l'audience est potentiellement plus large, mais le public n'a pas accès au praticien pour l'aider à forger ses premières armes pratiques. C'est pourquoi nous pensons qu'il faut élargir l'offre en direction de l'enseignant pour l'aider à faire ses premiers pas et à oser la pratique de la MVT.

Dans cet ordre d'idée, nous avons répondu à un appel à projet de la cellule de valorisation de l'université de Toulouse Jean Jaurès en janvier 2016, afin d'obtenir un budget pour le développement d'une application sur la MVT pour un public enseignant.

Le projet a été retenu et financé pour un montant total de 4000€. Cette somme nous a permis :

- De rémunérer les participants au test AXB (voir chapitre 5)
- De proposer un stage d'une durée de 5 mois à un étudiant de Master 2 professionnel en informatique, qui aurait pour charge de développer le logiciel d'entraînement.

En avril 2016, nous avons donc recruté Kévin Lepan, étudiant en Master 2 professionnel Imagerie et Multimédia à l'université Paul Sabatier de Toulouse[77].

Nous avons mentionné au chapitre 5 que Kévin avait écrit les algorithmes pour l'adaptation de la *T-Function* aux contours prosodiques. Cependant, son projet de stage était de produire effectivement un outil opérationnel pour l'entraînement au logatome. La compétence informatique de Kévin a ainsi constitué un accélérateur pour notre recherche, mais elle permettra surtout d'en rendre les résultats concrets disponibles pour le public dans un temps relativement proche.

3.1 Verbo Tonal Method Trainer (VTM-T)

VTM-T est le résultat concret de notre recherche et du stage de K. Lepan. Il s'agit d'une application logicielle dont l'objectif est de permettre à l'enseignant de s'entraîner à certaines pratiques vocales.

Lors du dépôt du projet[78] à la cellule de valorisation, nous avions prévu d'intégrer deux modules :

- Un module logatome exploitant les mesures de similarité prosodique, pour rendre concret le contour intonatif et lui permettre d'évaluer sa performance
- Un module « triangle vocalique » pour permettre à l'enseignant de visualiser et de comprendre le continuum phonétique, la variabilité de la parole, et l'entraîner à produire des sons hybrides.

Une version alpha de l'application VTM-T est maintenant opérationnelle, et nous souhaitons la faire tester par différents types de publics enseignants afin d'obtenir un *feedback* quant à son utilisation.

[77] La tutelle technique de ce stage a été assurée par Jérôme Farinas, UT3-UPS, IRIT, équipe SAMOVA
[78] Cf. Annexes

3.2 Module Turning Function

Il s'agit du module d'entraînement à l'imitation prosodique. Les locuteurs peuvent enregistrer des sons (ou les charger depuis leur disque dur) et les comparer dans la foulée. Ils obtiennent très rapidement un résultat quand à la similarité entre les deux sons. La Figure 70 présente l'interface du module *T-Function* de VTM-T.

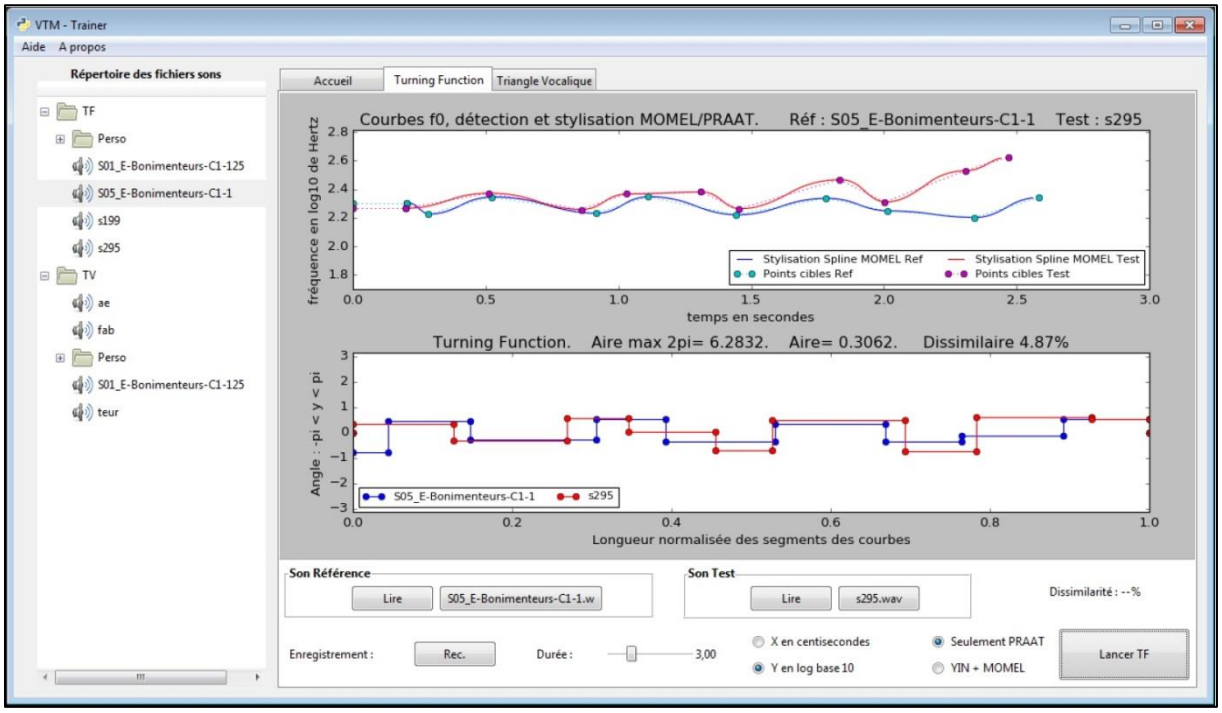


Figure 70 : Interface VTM-T, module Turning Function.

A l'heure actuelle, le module *T-Function* fonctionne sur une mesure de similarité de type TSP, *i.e.* les formes à partir desquelles la mesure de similarité est calculée émergent d'une annotation tonale du contour de *f0*. L'utilisateur a encore le choix d'appliquer deux mesures différentes :

- Une TSP issue de YIN + Momel (de Cheveigné & Kawahara, 2002; Hirst & Espesser, 1993)
- Une TSP issue de l'algorithme de Momel sous Praat

Par ailleurs, l'utilisateur a également le choix des unités pour le temps et la hauteur de *f0*.

La limite actuelle de ce module T-Function réside dans les erreurs qui peuvent émerger lors de la détection de la *f0* et, consécutivement, lors du choix des points cibles par l'algorithme Momel, parfois problématique (Campione & Véronis, 2000). En effet, nous avons vu que les mesures TSR & TSP peuvent être grandement faussées par des points

aberrants, manquants ou ajoutés. C'est pourquoi la figure est adjointe au score : l'utilisateur doit pouvoir savoir si le procédé de mesure de la similarité a marché comme prévu.

Enfin, il a été développé une fonctionnalité pour compter le nombre de « *da* », au cas où l'utilisateur souhaite s'entraîner au logatome. La TSP ne comptant pas les syllabes, cette information demeure essentielle pour que l'utilisateur puisse estimer sa performance (fonctionnalité qui est prête, mais non implémentée dans la version actuelle de VTM-T).

3.3 Module triangle vocalique

Ce module a été développé dans l'idée de permettre à l'enseignant de comprendre par l'expérience le classement des sons vocaliques utilisés par la MVT pour effectuer le diagnostic des erreurs vocaliques ainsi que pour la production des procédés de prononciation nuancée/déformée et de phonétique combinatoire. Par ailleurs, ce module doit lui permettre de prendre conscience de la variabilité intrinsèque des sons de parole.

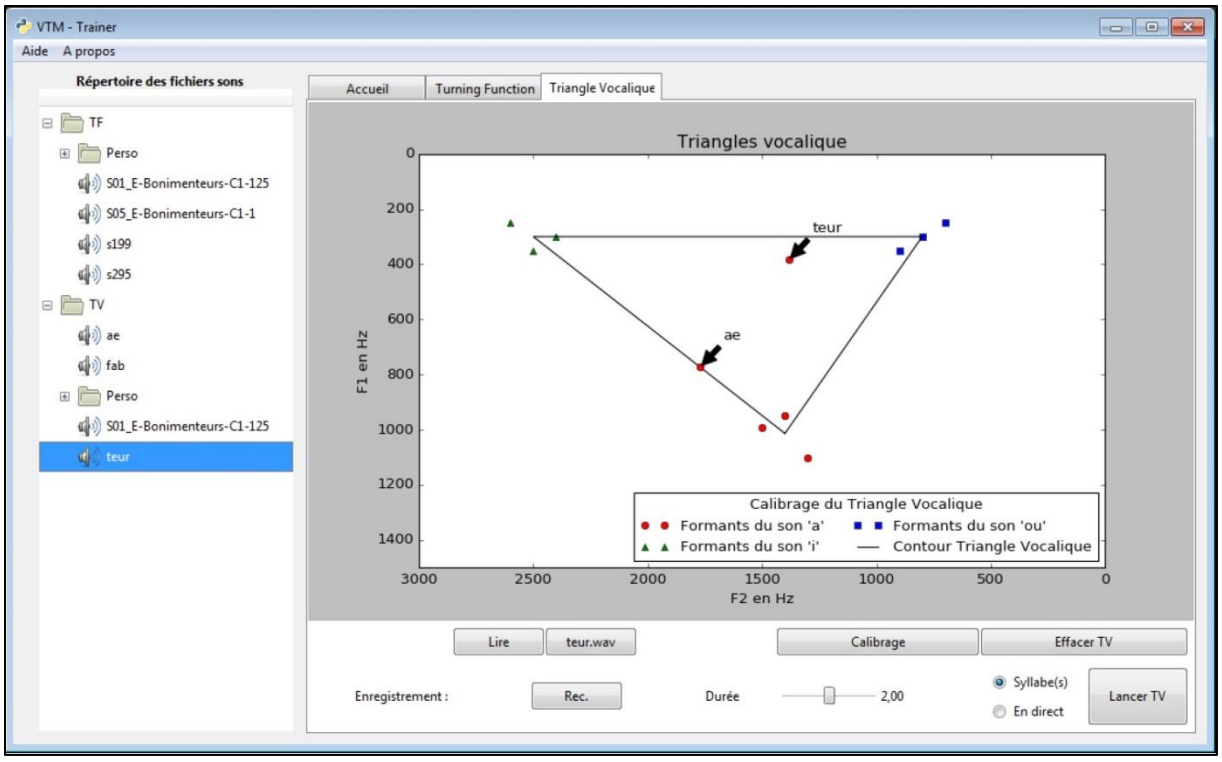


Figure 71 : Capture d'écran de VTM-T, module triangle vocalique.

Ce module est très simple dans son fonctionnement. L'utilisateur commence par calibrer les extrémités de son triangle vocalique en faisant plusieurs productions des sons /i/, /u/ et /a/. Il peut ensuite produire des voyelles isolées ou dans une syllabe et voir chaque production représentée dans le triangle. (Figure SSS).

En l'état, le module Triangle Vocalique ne prend en compte que les deux premiers formants de chaque voyelle se conformant aux représentations du triangle vocalique présentées dans les cours de MVT.

Par ailleurs, l'inscription des points se déroule en léger différé. A terme, il pourrait être tout à fait intéressant de pouvoir observer en direct le déplacement du repère dans le triangle à mesure de la modulation d'une voyelle.

3.4 Quelques pistes pour améliorer VTM-T

Ces deux premiers modules constituent une base intéressante pour l'enseignant qui veut comprendre par l'expérience la théorie de la MVT et s'entraîner à sa pratique. Evidemment, VTM-T demeure en l'état une version alpha qui peut et doit être améliorée en peaufinant d'une part les modules existants et en lui ajoutant d'autres.

- En ce qui concerne le module T-Function, une attention particulière devrait être portée à l'amélioration de la détection de la *f0* et du choix des points cibles.
    - Il pourrait être intéressant par exemple de pouvoir ajuster manuellement la position des points cibles.
    - La question de la détection de la *f0* est plus épineuse : en l'absence de meilleurs procédés, créer une passerelle avec un logiciel d'analyse acoustique pourrait permettre de s'assurer que la détection a bien fonctionné.
    - De plus, le travail sur les méthodes de mesures de la similarité prosodique et leur relation avec les évaluations perceptives doit continuer afin de pouvoir fixer des seuils de réussite et ainsi éclaircir la signification des valeurs chiffrées à l'utilisateur.
    - Enfin, ce travail sur les méthodes de mesure devrait aboutir à la sélection d'une méthode unique (la plus fiable pour coût de calcul modéré) et ainsi simplifier l'expérience de l'utilisateur.

- A propos du module Triangle Vocalique, pouvoir observer en direct les modulations de la voix des productions des voyelles permettrait de mieux rendre compte de la notion de continuum des sons de parole.
    - Il pourrait également être intéressant de calibrer chaque voyelle du triangle (au lieu de seulement ses extrémités) et de pouvoir appliquer des filtres sur la figure pour visualiser les aires de dispersion de chaque voyelle.
    - De même, rendre la figure interactive en permettant de rejouer les sons en cliquant sur leur repère dans la figure pourrait améliorer l'expérience de l'utilisateur de ce module.

En termes de perspectives, nous pensons qu'envisager la création d'autres modules permettrait d'élargir la portée de VTM-T à d'autres aspects de la MVT que la seule production de procédés vocaux.

En l'état, VTM-T est encore assez brut et tous les sons sont à enregistrer par l'utilisateur. Il a cependant été prévu de fournir un corpus d'exemples sonores illustrant les fonctionnalités de VTM-T.

Dans cet ordre d'idée, fournir des corpus d'erreurs produites par des apprenants de tous horizons pourrait mener à proposer un module d'entraînement au diagnostic, au choix des procédures, et à leur correction.

Par exemple, l'utilisateur pourrait jouer un exemple sonore auquel serait associé un quizz reprenant les étapes du diagnostic, de choisir les modulations à effectuer puis de s'y entraîner…

En guise de conclusion, nous souhaitons souligner que VTM-T ouvre des perspectives pratiques intéressantes dans le domaine de la formation des enseignants à la remédiation phonétique par la MVT. L'idée de développement de cet outil résulte de nos premières recherches, et, nous l'espérons, sera nourri par notre recherche à venir.

Discussion générale et perspectives

Rappel du cadre théorique

Pour ce travail, les comportements d'imitation en parole, convergence et divergence phonétique (Giles et al., 1991; Pardo, 2006), impersonnation (Eriksson & Wretling, 1997; Révis et al., 2013; Zetterholm, 2009a) et imitation de laboratoire (Dufour & Nguyen, 2013; Namy et al., 2002) ont été replacés dans le cadre traditionnel de l'étude générale des comportements imitatifs, afin de souligner :

- La variété de ces comportements au travers de différents spectres des imitations (mimétisme et mimésis, émulation, mimique et imitation).
- Les fonctions de ces différents types d'imitation, indiquant l'importance de l'environnement du sujet dans sa construction individuelle et dans son mode d'action.

Bien que confinant à l'obsession terminologique, cette première étape de notre travail était essentielle. En effet, la description des comportements imitatifs est difficile à cause d'un nombre réduit de termes aux acceptions multiples, se recouvrant partiellement (Call & Carpenter, 2002). De plus, la navigation entre les littératures francophones et anglophones rend plus compliquée encore l'établissement d'un vocabulaire descriptif clair. Enfin, il nous semblait primordial de lier imitation gestuelle et imitation parolière pour illustrer leur dynamique commune.

L'inspection des facteurs neurologiques et cognitifs de l'imitation (Brass & Heyes, 2005; M. Garnier et al., 2013) nous ont en effet conduit à formuler l'idée que le spectre des comportements imitatifs constitue un mécanisme d'ajustement permanent de l'action de l'individu avec son environnement. Nous pensons que ces mécanismes d'imitations agissent en tâche de fond sur un mode automatique, et qu'ils pousseraient (*urge*) l'individu à se conformer aux usages du groupe social avec lequel il interagit. Cependant, l'individu pourrait également, en fonction de son expérience et/ou des enjeux de la communication, s'engager dans un comportement de non conformation (*repel*), ou déclencher un comportement d'imitation (*trigger*).

Appliqué au langage et à la parole, ce cadre d'analyse suppose que les comportements imitatifs en parole (convergence, divergence, maintenance, impersonnation) sont prégnants

dans la communication des individus et qu'ils servent les fonctions que nous avons décrites durant notre chapitre premier : adaptation, communication, apprentissage.

Ceci étant dit, il convient de rappeler l'importance de l'expérience du sujet pour la mise en œuvre des comportements imitatifs. En effet, nous avons indiqué que les liaisons perceptuomotrices, selon l'*Associative Learning Hypothesis* (Brass & Heyes, 2005; Heyes, 2001; Heyes et al., 2005), seraient forgées par les approximations successives produites par le sujet en interaction avec son environnement. Nous avons également souligné que les unités communicatives, les sons de langage, ont une double dimension, perceptive et motrice (Schwartz et al., 2012).

Si nous pensons à l'apprenant de L2, dont le système perceptuomoteur est conditionné par des années d'expériences d'utilisation de sa L1 (Best, 1994; Kuhl et al., 2008), ces considérations permettent de mettre en exergue plusieurs idées fortes :

- L'importance de fournir à l'apprenant de nombreuses interactions orales, afin de favoriser le développement de ses capacités perceptuomotrices.
- La nécessité de multiplier les origines linguistiques (variétés dialectales) des locuteurs dont on lui présente la parole pour élargir son expérience avec la langue cible.
- Et, non des moindres, la prise en compte du système de la L1 de l'apprenant dans son apprentissage.

Le recours à la MVT vise particulièrement à apporter des solutions quant à ce dernier problème en proposant des pratiques de remédiation personnalisées, au moyen de modulations prosodiques et segmentales produites par l'enseignant. D'un point de vue imitatif, la MVT faciliterait la construction des liens entre perception de la parole et expérience motrice en faisant ressortir, par ses procédés, les spécificités du comportement ciblé. Cependant, il a été souligné que la mise en œuvre des techniques vocales de la MVT pouvait être mal aisée et que cette mise en œuvre représentait aussi, pour l'enseignant, un enjeu d'imitation et de contrôle prosodique.

Cette dernière assertion a constitué la raison pour laquelle notre partie expérimentale s'est focalisée sur l'évaluation et la mesure de la similarité prosodique dans des imitations produites par des locuteurs de L1 française.

Apports étude 1 : exploration des liens entre évaluations subjectives et mesures objectives de la similarité prosodique

Notre première étude s'est concentrée sur les aspects méthodologiques de l'évaluation/mesure de la similarité prosodique. En effet, nous avions relevé durant notre cadre théorique que les mesures acoustiques décrites dans la littérature à ce sujet (par exemple, (Pardo, 2013, 2013) avaient un caractère insatisfaisant dans notre cadre puisqu'elles n'étaient pas adaptées à l'étude de la similarité des mouvements prosodiques, mais des paramètres acoustiques locaux. Par un design d'évaluations/mesures répétées sur un même corpus de 72 imitations, nous avons exploré les liens entre :

- Tests de jugement perceptif de la similarité prosodique
    - Test AX
    - Test AXB
- Mesures de la similarité prosodique
    - Calculées après la transformation de *f0* brutes au moyen *du Dynamic Time Warping* (Hermes, 1998b; Rilliard et al., 2011).
        - Norme L_2
        - Coefficient Z_r
    - Calculées (norme L_2) après la transformation, au moyen de la *T-Function* (Arkin et al., 1991; Veltkamp, 2001), de *f0* stylisées en lignes droite selon :
        - Un système d'annotation tonal (TSP)
        - Un système d'annotation rythmique (TSR)

En ce qui concerne les jugements perceptifs, nos résultats ont souligné la différence de philosophie entre les tests AX, qui permettent d'obtenir une information absolu de la similarité prosodique entre un modèle et son imitation, et les tests AXB, qui se limitent à établir des degrés de hiérarchie dans des groupes d'imitation. Les deux tests avaient une bonne corrélation dans leur output, mais il a été décidé que le test AX convenait mieux à l'interprétation de nos résultats ultérieurs. A propos du test AXB, nous avons remarqué que son usage était limité dans notre cadre, puisqu'il ne permet pas d'obtenir d'information sur la valeur absolue de la réussite d'une imitation sur le plan perceptif. Il demeure cependant approprié dans l'étude de la convergence phonétique, tel que le propose Pardo (2006).

A propos des mesures de la similarité, les deux types d'approches étudiées ont révélé avoir une certaine validité, puisque leur corrélation avec les résultats des tests AX était

satisfaisante. Il semble toutefois que les mesures TSP et TSR soient de meilleures candidats à l'évaluation de la similarité prosodique. Ceci serait lié aux différences de propriétés des transformations des formes utilisées :

- La transformation des courbes de *f0* brute requiert l'usage du DTW. Ce faisant, il se produit une déformation des aspects temporels, *i.e.* une altération des aspects rythmiques en jeu dans le contour intonatif.
- La transformation des courbes de *f0* stylisées par la *T-Function*, transpose la configuration des formes dans un nouvel espace de représentation, en conservant leurs proportions dans les deux dimensions.

Ainsi, le DTW fait perdre une dimension dans le calcul final de la similarité prosodique, tandis que la *T-Function* garde un paramètre supplémentaire. En conséquence, TSP et TSR ont obtenu de meilleures corrélations avec les tests perceptifs qui mentionnaient dans leurs consignes mélodie et rythme. A ce sujet, la TSR a obtenu la meilleure corrélation avec les évaluations perceptives, suggérant une prééminence du rythme en prosodie (Astésano, 2001; Billières, 1988; Di Cristo & Hirst, 1993).

Par ailleurs, notre discussion des résultats a fait ressortir un problème inhérent aux évaluations perceptives, qui est celui de la perception par les auditeurs du degré de compétence vocale supposée des imitateurs. Nous avons remarqué que les auditeurs étaient parfois plus sévères que les mesures automatiques dans leur notation de certains imitateurs. Nous pensons qu'au cours du test perceptif, les auditeurs ont fini par avoir un préjugé sur certains imitateurs, qui les auraient conduits à sous-évaluer la performance de ces imitateurs lors de leur notation.

Faire évoluer la méthodologie de ce type de test permettrait peut-être de contourner ce biais de jugement. Hirst (2016) propose une technique de clonage des contours prosodiques qui permettrait d'éliminer purement et simplement ce biais d'identification des locuteurs : il pourrait être pertinent de dupliquer la *f0* des imitateurs pour la coller sur la voix du modèle. Ce procédé pourrait nous permettre d'obtenir une évaluation perceptive plus précise de la similarité prosodique.

Apports de l'étude 2 : Evaluation de la similarité prosodique d'imitations délexicalisées : une application de la *T-Function* sur Stylisation Rythmique

Notre seconde étude a été consacrée à l'étude de la performance d'imitation prosodique de 4 locuteurs (2 naïfs et 2 experts) lors du recueil d'un corpus d'imitations lexicalisées et délexicalisée (logatomes). Ces sujets ont été instruits de reproduire trois fois d'affilée la prosodie d'énoncés entendus, en alternant phrases et logatomes.

Durant cette étude, nous avions plusieurs objectifs :

- Observer la différence de performance d'imitation prosodique entre des locuteurs naïfs et experts
- Observer si la qualité du logatome diminuait en relation avec la longueur des énoncés
- Déterminer s'il y avait
    - Un effet d'entraînement à la tâche d'imitation prosodique (effet tout au long de la tâche)
    - Un effet de dégradation du patron prosodique original entre les imitations successives d'un même trial expérimental.
- Connaître le comportement de la mesure TSR lorsqu'on l'applique pour comparer
    - Phrase vs. Phrase,
    - Phrase vs. Logatome

Dans ce cadre expérimental, la mesure TSR a bien réagi car les résultats qui nous ont été fournis par la mesure étaient cohérents avec nos hypothèses.

A propos de la mesure en elle-même, nous avons remarqué que la comparaison Phrase vs. Phrase donnait en moyenne de meilleurs résultats de similarité que la comparaison Phrase vs. Logatome. Cependant, nos résultats ont aussi souligné que cette différence n'était statistiquement significative que pour les productions d'un locuteur naïf, dont la performance était moindre.

L'explication que nous voyons à ces différences de scores est liée à l'annotation des points d'ancrage précédant la transformation *T-Function*. Dans la stylisation TSR, les points d'ancrage sont placés au milieu des voyelles : ainsi lorsqu'on compare deux phrases au même contenu segmental, les décalages temporels liés à la durée des voyelles sont probablement

moins grands que lors de la comparaison d'une phrase et d'un logatome pour lequel toute les syllabes sont ouvertes (dadada), *i.e.* où toutes les voyelles sont potentiellement plus longues que dans la phrase.

Il se pourrait de plus que la différence significative observée entre les phrases et les logatomes du locuteur naïf soit due au grand nombre de syllabes omises par ce locuteur lors de ses productions, comme l'ont montré nos résultats descriptifs sur le nombre de syllabes omises par les sujets.

Au terme de cette première étude utilisant la TSR, nous pouvons estimer que le résultat de la mesure TSR doit être légèrement ajusté lorsque sont comparés une phrase et son équivalent logatome. Cependant, il sera nécessaire d'effectuer une étude perceptive pour évaluer la taille de l'ajustement. Enfin, rappelons que la mesure TSR semble très sensible à certaines conditions :

- Lorsque le nombre de points est de 4 ou moins, il est probable que les décalages soient amplifiés.
- Au-delà de 8 points, ceux-ci seraient minimisés.

Des tests sur l'application de la TSR aux formes prosodiques seront nécessaires pour évaluer le poids du nombre de points dans le résultat de la mesure.

Outre ces considérations sur la mesure même, les résultats donnés par la TSR indiquent que les locuteurs experts méritent leur statut. Cependant, ils ont été dépassés par la performance étonnante d'un des locuteurs non experts. Cela illustre d'une part la grande variabilité de compétence imitative que nous avions une première fois illustrée dans les résultats des tests perceptifs sur le Corpus Imitation. NE1 (Corpus Logatome) et Sp5 (Corpus Imitation) semblaient particulièrement talentueux dans leur reproduction des contours prosodiques. Ainsi, ce type de mesure semble être un bon candidat pour détecter un aspect du talent phonétique : le talent prosodique.

Par ailleurs, un résultat sur lequel nous aimerions revenir concerne la dégradation du pattern prosodique au cours des trois imitations successives de chaque *trial*. Ce dernier résultat suggérait que les locuteurs gardaient en mémoire la dernière trace acoustique produite/entendue et que cette trace avait une influence sur la production suivante. Pour l'heure, nous souhaiterions indiquer que ce résultat est à prendre avec précaution, dans la mesure où nous n'avions pas fait de comparaisons entre les imitations successives, mais

uniquement entre le modèle original et ces différentes imitations. Cependant, cela constitue une piste de recherche intéressante.

Enfin, nous proposons l'application des évaluations de la similarité prosodique dans une solution logicielle qui permettrait à l'enseignant d'avoir une information sur la qualité de ses imitations prosodiques. Le logiciel (*Verbo Tonal Method-Trainer*) est déjà utilisable et il pourra être testé dans les mois à venir.

Perspectives de recherche

Les premières perspectives découlant naturellement de notre recherche concernent l'amélioration de la méthodologie des tests perceptifs d'une part, et l'affinage de notre compréhension des mesures de la similarité prosodique d'autre part.

Lors de notre travail, les aspects méthodologiques et la réflexion associée à ces aspects ont eu une importance cruciale que nous réaffirmons ici. Par exemple, nous pensons qu'il sera judicieux de construire de nouveaux tests perceptifs avec le corpus de 72 imitations utilisé pour notre première étude. Cela nous permettra de faire fructifier les différentes mesures que nous avons produites sur ces phrases, et ainsi d'en affiner notre compréhension. Il sera également intéressant de faire grandir le nombre de jugements perceptifs obtenus sur ces phrases, afin d'obtenir une masse d'information suffisamment importante pour fiabiliser les liens entre mesures et évaluations de la similarité prosodique.

A propos des mesures, nous pensons continuer à développer TSP et TSR. Après la première phase de validation de ces mesures, l'usage que nous en avons fait durant notre expérimentation sur le corpus de logatome s'est révélé être pertinent. Nous pensons que ce type de mesure pourra être utilisé dans nos recherches futures, mais aussi dans d'autres cadres expérimentaux que le nôtre. Il nous semble cependant nécessaire d'approfondir notre compréhension de ces mesures, en ce qui concerne leur biais potentiels ou leurs limites.

A terme, l'utilisation des mesures de similarité prosodique pourrait être intégrée dans les recherches sur le talent phonétique (Jilka et al., 2007). Les protocoles pour évaluer le talent des locuteurs sont particulièrement lourds et les mesures de similarité prosodiques pourraient, au moins partiellement, détecter les locuteurs les plus doués au niveau prosodique.

Par ailleurs, ces mesures pourraient trouver leur place dans des recherches sur l'apprentissage des langues secondes ou sur l'acquisition de la L1 : il serait par exemple intéressant de soumettre les babillages des enfants à des mesures de similarité avec des patrons prosodiques de leur langue maternelle pour étudier l'émergence du système prosodique dans le processus d'acquisition.

En ce sens, une autre perspective directe de notre recherche sera de rendre disponible nos outils à la communauté scientifique. Cette recherche a été conduite essentiellement de manière individuelle. Or, nous sommes très curieux et excités de découvrir les usages qui pourraient être faits de ces mesures par d'autres chercheurs ayant des problématiques de recherche différentes que les nôtres.

A propos de *Verbo Tonal Method-Trainer* (VTM-T), son développement nous offre plusieurs perspectives :

- Il convient de poursuivre le développement de l'outil, tant sur le plan technique qu'au niveau du contenu théorique intégré au logiciel (la version actuelle est encore assez épurée).
- Il sera nécessaire de soumettre chaque module de l'outil à une phase de tests approfondis.
- Il pourrait être intéressant de produire une étude longitudinale pour étudier l'effet de l'outil VTM-T dans le développement de la compétence d'imitation prosodique.

En termes de recherche, notre thèse dessine des perspectives à court et à long terme. Il y a en effet un travail à poursuivre sur nos données acquises durant cette expérience et les outils utilisés ici pourraient être raffinés. De plus, la situation de remédiation proposée par la MVT peut encore fournir matière à de nombreuses études, tant sur l'enseignant que sur l'apprenant et leur interaction.

Par ailleurs, l'étude de la similarité prosodique pourrait être appliquée à d'autres domaines. Par exemple, l'évaluation de la similarité prosodique pourrait servir dans l'évaluation de la qualité de parole des patients ayant subi une opération de la cavité buccale, et plus particulièrement du larynx.

Pistes dans le domaine didactique

Au niveau didactique, les premières implications de notre recherche concernent les enseignants. En effet, VTM-T s'adresse à eux comme le premier outil pratique d'entraînement aux procédés de la MVT.

Ceci étant dit, cette recherche vise à réaffirmer l'importance du travail phonétique dans les pratiques de classe. Cependant, comme nous le signifions dans notre chapitre 3, la MVT, qui à notre avis est le candidat idéal pour avoir une pratique adaptée au contexte didactique actuel, demeure encore trop peu diffusée. Nous estimons donc qu'il faudra produire un effort de diffusion de la MVT pour la présenter sous un jour accessible aux futurs enseignants, mais aussi aux enseignants déjà en poste (et la tâche est plus compliquée).

Nous avions souligné que la MVT était le couteau suisse de l'enseignant, capable de remédier à toute situation. Or, les enseignants de FLE sédentaires ne rencontrent pas toutes les situations. Ainsi, nous pensons qu'il serait intéressant d'envisager la rédaction de petits livres pratiques adaptant le travail phonétique en MVT au contexte d'enseignement spécifique des enseignants (par exemple : correction phonétique des locuteurs germanophones, italophones, etc.). En effet, quand elle est présentée comme une méthode aux possibilités presque infinies (nous exagérons, certes), il est possible que la MVT fasse peur aux enseignants et qu'ils doutent d'être capables de la mettre en pratique. C'était aussi pour cette raison, pour lever les doutes des enseignants, que nous avions souhaité le développement de VTM-T.

En ce qui concerne les comportements imitatifs, nous souhaiterions simplement rappeler qu'ils constituent un vecteur particulièrement puissant d'apprentissage par expérience chez l'être humain. En tant que tel, il pourrait être pertinent pour l'enseignant d'essayer de les exploiter dans ses pratiques de classe, en invitant explicitement l'apprenant, sur le mode du jeu, à imiter la manière de parler des locuteurs natifs entendus.

En effet, si l'on considère à nouveau le système des comportements imitatifs en parole, nous pourrions émettre l'hypothèse que l'apprenant se trouve dans un état de maintenance. En l'invitant à produire de l'impersonnation, l'enseignant pourrait ainsi pousser l'apprenant à abandonner ses réticences et à adopter un rôle et des attitudes qui lui sont encore étrangères.

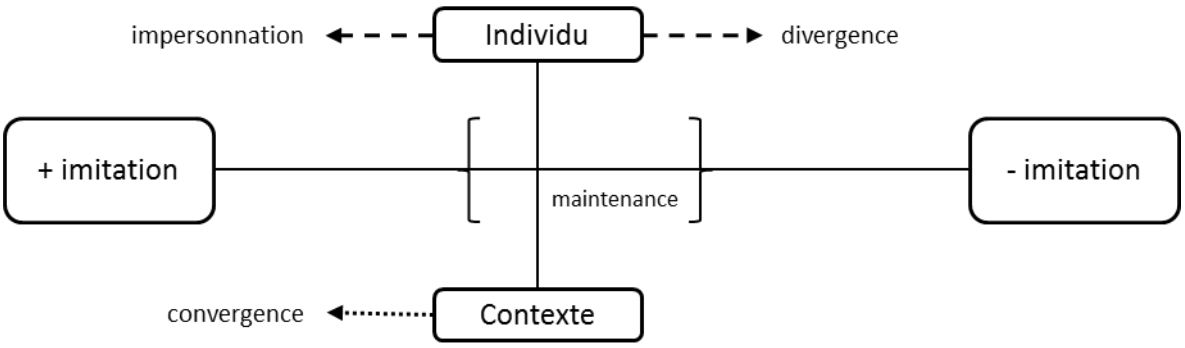


Figure 72 : Système dynamique des comportements imitatifs en parole (rappel)

Ainsi, notre travail tend à montrer que si les sujets humains sont communiquants, ils sont également dans ce cadre, des sujets imitants. Cependant, ne pouvant imiter que ce dont nous avons fait l'expérience, il convient pour l'enseignant de fournir à l'apprenant de quoi constituer un nouveau répertoire de comportements adaptés à son environnement.

Dans cet ordre d'idée, la MVT est un outil qui semble particulièrement approprié aux mécanismes imitatifs, puisque son mode de fonctionnement a pour but de fournir une expérience perceptive à l'apprenant.

Pour conclure, nous souhaitons souligner que notre recherche est duelle : en la nourrissant des contextes didactiques de la MVT, nous avons fait émerger des questions de recherche fondamentale en parole. Ceci étant dit, les résultats de ces recherches doivent aussi être reversés sur le terreau fertile que représente pour nous la MVT. C'est pourquoi nous avons proposé Verbo Tonal Method Trainer et que nous envisageons de continuer à travailler dessus. Nous réaffirmons ici l'importance du lien entre recherche et pratiques didactiques.

Bibliographie

Abramson, J. Z., Hernández-Lloreda, V., Call, J., & Colmenares, F. (2013). Experimental evidence for action imitation in killer whales (Orcinus orca). *Animal Cognition*, *16*(1), 11–22. https://doi.org/10.1007/s10071-012-0546-2

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation Improves Language Comprehension. *Psychological Science*, *21*(12), 1903–1909. https://doi.org/10.1177/0956797610389192

Alazard, C. (2013). *Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Étrangère*. Université Toulouse le Mirail-Toulouse II. https://tel.archives-ouvertes.fr/tel-00944968/

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., … others. (1991). The HCRC map task corpus. *Language and Speech*, *34*(4), 351–366.

André, C., Ghio, A., Cavé, C., & Teston, B. (2003). PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. In *Proceedings of XVth ICPhS* (pp. 1421–1424). Barcelone, Espagne.

Anisfeld, M. (1996). Only tongue protrusion modeling is matched by neonates. *Developmental Review*, *16*(2), 149–161.

Arbib, M. A. (2012). *How the Brain Got Language* (Oxford University Press). Oxford. Retrieved from

Arbib, M. A. (2013). Précis of How the brain got language: The Mirror System Hypothesis. *Language and Cognition*, *5*(2–3), 107–131. doi.org/10.1515/langcog-2013-0007

Aristote. (1997). *Poétique*. (B. Gernez, Ed.). Paris, France: Les Belles Lettres.

Arkin, E. M., Chew, P., Huttenlocher, D., Kedem, K., & Mitchel, J. S. B. (1991). An Efficiently Computable Metric for Comparing Polygonal Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(3), 209–216.

Asch, S. E. (1955). Opinions and Social Pressure. *Scientific American*, *193*(5), 31–35.

Astésano, C. (2001). *Rythme et accentuation en français: invariance et variabilité stylistique*. L'Harmattan.

Astésano, C., Bard, E. G., & Turk, A. (2007). Structural influences on Initial Accent placement in French. *Language and Speech*, *50*(3), 423–446.

Astésano, C., & Bertrand, R. (2016). Accentuation et niveaux de constituance en français: enjeux phonologiques et psycholinguistiques. *Langue Française*, (3), 11–30.

Astésano, C., Bertrand, R., Espesser, R., & Nguyen, N. (2012). Perception des frontières et des proéminences en français. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 1: JEP* (pp. 353–360). Grenoble.

Attal-Fiocchi, M., & Jarzé, A. (2014). *Variabilité prosodique au cours de représentations successives d'un même sketch : stéréotype ou spontanéité ?* (Mémoire d'orthophonie). Aix-Marseille, Marseille.

Baayen, R. . (2008). *Analyzing Linguistic Data: A practical introduction to Statistics using R* (Cambridge University Press). New York.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231–248.

Bard, E. G., Astésano, C., D'Imperio, M., Turk, A., Nguyen, N., Prévot, L., & Bigi, B. (Eds.). (2013). Aix Map Task: A new French resource for prosodic and discourse studies. In *TRASP 2013* (pp. 15–19). Aix-en-Provence.

Baudonnière, P.-M. (1997). *Le mimétisme et l'imitation: un exposé pour comprendre, un essai pour réfléchir*. Flammarion.

Bergounioux, G., Bergounioux, M., Nguyen, N., & Wauquier, S. (2007). Introduction au n° spécial «Mathématiques et phonologie». *Mathématiques et Sciences Humaines. Mathematics and Social Sciences*, (180), 1–7.

Bessler, P. (1991). La caricature de De Gaulle par Tisot : Etude phonostylistique, *12*, 19–32.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, *167*, 224.

Bigi, B. (2015). SPPAS - Multi-Lingual Approaches to the Automatic Annotation of Speech. *"The Phonetician"-International Society of Phonetic Sciences*, (111–112), 54–69.

Billières, M. (1988). Crible phonique, crible psychologique et intégration phonétique en langue seconde. *Travaux de Didactique Du Français Langue Étrangère*, *19*, 5–29.

Billières, M. (2000). *Didactique de l'enseignement de la prononciation. Aspects prosodiques, phonétiques, psycholinguistiques et méthodologiques.* (HDR). Toulouse 2 Jean Jaurès, Toulouse.

Billières, M. (2002). Le corps en phonétique corrective. In *Apprentissage d'une langue étrangère/seconde 2. La phonétique verbo-tonale* (De Boeck Université, pp. 37–70). Bruxelles: Renard R.

Billières, M., Alazard, C., Astésano, C., & Nocaudie, O. (2013). Phonétique corrective en FLE : Méthode Verbo-Tonale.

Boë, L.-J., & Bonastre, J.-F. (2012). L'identification du locuteur: 20 ans de témoignage dans les cours de Justice. In *Actes de la conférence jointe JEP-TALN-RECITAL 2012* (pp. 417–424). Grenoble. Retrieved from http://anthology.aclweb.org/F/F12/F12-1053.pdf

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5:9/10*, 341–345.

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In *Language, Ethnicity and Intergroup Relation* (London Academic Press, pp. 119–135). Giles, H.

Bourhis, R. Y., Giles, H., Leyens, J. P., & Tajfel, H. (1979). Psycholinguistic distinctiveness: Language divergence in Belgium. In *Language and Social Psychology* (Blackwell, pp. 158–185). Oxford: Giles, H. and St. Clair R.

Bradlow, A. R., Baker, R. E., Choi, A., Kim, M., & Van Engen, K. J. (2007). The Wildcat Corpus of Native- and Foreign-Accented English. *Journal of the Acoustical Society of America*, *121*(5), 3072.

Brass, M., Derrfuss, J., Matthes-von Cramon, G., & von Cramon, D. Y. (2003). Imitative response tendencies in patients with frontal brain lesions. *Neuropsychology*, *17*(2), 265–271. https://doi.org/10.1037/0894-4105.17.2.265

Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: a functional double dissociation. *Neuropsychologia*, *43*(1), 89–98. https://doi.org/10.1016/j.neuropsychologia.2004.06.018

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, *9*(10), 489–495. https://doi.org/10.1016/j.tics.2005.08.007

Brass, M., Zysset, S., & von Cramon, D. Y. (2001). The Inhibition of Imitative Response Tendencies. *NeuroImage*, *14*(6), 1416–1423. https://doi.org/10.1006/nimg.2001.0944

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bull. Soc. Anat. Paris*, *36*(6), 330–357.

Brown, G., Anderson, A., Yule, G., & Schillcock, R. (1983). *Teaching talk* (Cambridge University Press). Cambridge, UK.

Brzostek, M., & Deschanvres, M. (2014). *Flexibilité prosodique chez l'imitateur professionnel : Comment passer d'une cible à l'autre ?* (Mémoire d'orthophonie). Université Aix-Marseille, Marseille.

Call, J., & Carpenter, M. (2002). Three Sources of Information in Social Learning. In *Imitation in Animals and Artifacts* (The MIT Press, pp. 211–228). Cambridge: Dautenhahn, K. & Nehaniv, C.L.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cerebral Cortex*, *15*, 1243–1249.

Campbell, J. (1995). Testing with the YOHO CD ROM voice verification corpus. In *ICASSP 95*.

Campbell, N., & Beckman, M. E. (1997). Stress, prominence and spectral tilt. In *Proceedings of an ESCA Workshop* (pp. 67–70). Athens, Greece: Botinis, Kouroupetroglou & Carayiannis.

Campione, E., & Véronis, J. (2000). Une évaluation de l'algorithme de stylisation mélodique MOMEL. *Travaux Interdisciplinaires Du Laboratoire Parole et Langage d'Aix-En-Provence (TIPA)*, *19*, 27–44.

Carpenter, M., & Call, J. (2009). Comparing the imitative skills of children and nonhuman apes. *Revue de primatologie*, (1). https://doi.org/10.4000/primatologie.263

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893.

Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? In *INTERSPEECH*.

Conseil de l'europe. (2001). Cadre européen commun de Référence pour les Langues : apprendre, enseigner, évaluer. Didier.

Coulon, M., Hemimou, C., & Streri, A. (2012). Effects of seeing and hearing vowels on neonatal facial imitation. In *Proceedings of ISICS*. Aix-en-Provence.

CRAN. (2016). R (Version 3.2.1) [Windows].

de Boysson-Bardies, B. (1996). *Comment la parole vient aux enfants : de la naissance jusqu'à deux ans* (Odile Jacob). Paris.

de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*(4), 1917. https://doi.org/10.1121/1.1458024

De Looze, C. (2010). *Analyse et interprétation de l'empan temporel des variations prosodiques en français et en Anglais*. Université de Provence-Aix-Marseille I. Retrieved from https://tel.archives-ouvertes.fr/tel-00470641/

Delvaux, V., Demolin, D., & Soquet, A. (2004). Interactions mimétiques entre locuteurs: une étude expérimentale. *Actes Des XXVèmes Journées D'étude Sur La Parole*, 153–156.

Delvaux, V., Harmegnies, B., & Soquet, A. (2005). Mimésis et contrôle phonétique: Implications pour l'apprentissage d'une langue seconde. In *BILLIÈRES, M., GAILLARD, P. & SPANGHERO-GAILLARD, N., Actes du premier colloque international de didactique cognitive, Université de Toulouse II, Toulouse* (pp. 126–131).

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2014). Phonetic compliance: a proof-of-concept study. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01375

Delvaux, V., & Soquet, A. (2007). The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. *Phonetica*, *64*(2–3), 145–173. https://doi.org/10.1159/000107914

Derrida, J. (2006). *L'animal que donc je suis*. (M.-L. Mallet, Ed.). Paris: Editions Galilée.

Di Cristo, A. (1978). *De la microprosodie à l'intonosyntaxe*. Thèse d'état

Di Cristo, A. (1999). Vers une modélisation de l'accentuation du français : première partie. *Journal of French Language Studies*, *9*(2), 143–179.

Di Cristo, A., & Hirst, D. (1993). Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix*, *15*, 9–24.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180.

D'Imperio, M. (2005). La Phonologie de Laboratoire : notions de base et applications. In *Phonologie et phonétique -forme et substance-* (Hermes Lavoisier, pp. 241–264). Paris: Nguyen N., Wauquier-Graveline S. & Durand, J.

Donald, M. (1993). *Origins of the Modern Mind - Three Stages in the Evolution of Culture & Cognition* (Reprint). Cambridge, Mass.: Harvard University Press.

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00346

Duhem, P. (2007). *La théorie physique, son objet, sa stucture* (Vrin (éd. originale 1906)). Paris, France.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing" deafness" in French? *Journal of Memory and Language*, *36*, 406–421.

Eriksson, A., & Wretling, P. (1997). How flexible is the human voice ? - A case study of mimicry. In *Proceedings of Eurospeech '97* (Vol. 2, pp. 1043–1046). Rhodes.

Farrús, M., Wagner, M., Anguita, J., & Hernando, J. (2008a). How vulnerable are prosodic features to professional imitators? In *Odyssey* (p. 2).

Farrús, M., Wagner, M., Anguita, J., & Hernando, J. (2008b). Robustness of prosodic features to voice imitation. In *Proceedings of Interspeech 2008* (pp. 613–616). Brisbane, Australia.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex: Mirror neurons for mouth actions in F5. *European Journal of Neuroscience*, *17*(8), 1703–1714. https://doi.org/10.1046/j.1460-9568.2003.02601.x

Ferrari, P. F., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience*, *17*(2), 212–226.

Gadet, F. (2007). *La variation sociale en français* (Ophrys). Paris.

Galef, B. G., & Benett, G. (1988). Evolution and learning before Thorndike: A forgotten epoch in the history of behavioral research. *Evolution and Learning*, 39–58.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609.

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00340

Garnier, L., Baqué, L., Dagnac, A., & Astésano, C. (2016). Perceptual investigation of prosodic phrasing in French. In *Proceedings of Speech Prosody 2016*. Boston.

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00600

Garrod, S., & Doherty, G. (1994). Covnersation, co-ordination and convention: an empirical investigation of how groups establish linguistic convention. *Cognition*, *53*, 181–215.

Giles, H., Coupland, N., & Coupland, J. (1991). Accomodation theory: Communication, context and consequence. In *Contexts of Accomodation* (Cambridge University Press & Editions de la Maison des Sciences de l'Homme, pp. 1–68). New York: Giles, H., Coupland, J. & Coupland N.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Gorisch, J., Astésano, C., Bard, E. G., Bigi, B., & Prévot, L. (2014). Aix Map Task corpus: The French multimodal corpus of task-oriented dialogue. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavic: European Language Resources Association (ELRA).

Goudailler, J.-P. (2002). De l'argot traditionnel au français contemporain des cités. *La linguistique*, *38*(1), 5. https://doi.org/10.3917/ling.381.0005

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representation in humans by positron emission tomography. *Experimental Brain Research*, *112*(1), 103–111.

Granier-Deferre, C., & Schaal, B. (2005). Aux sources foetales des réponses sensorielles et émotionnelles du nouveau-né. *Spirale*, *33*(1), 21. https://doi.org/10.3917/spi.033.0021

Guberina, P. (1978). Etude préliminaire sur les moyens de rendre l'écoute en langues étrangères et suggestions en vue d'une approche méthodologique de l'enseignement des langues étrangères aux enfants. In *Rétrospection* (ArTresor Naklada, pp. 280–317). Zagreb, 2003: Roberge C.

Guberina, P. (1991). Rôle de la perception auditive dans l'apprentissage précoce des langues. In *Rétrospection* (ArTresor Naklada, pp. 517–524). Zagreb, 2003: Roberge C.

Guillaume, P. (1925). *L'Imitation chez l'enfant: Étude psychologique*. Librairie Félix Alcan.

Harmegnies, B., Delvaux, V., Huet, K., & Piccaluga, M. (2005). Oralité et cognition: pour une approche raisonnée de la pédagogie du traitement de la matière phonique. *Revue Parole*, *34*, 265.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J. C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience*, *17*(5), 1123–1128. https://doi.org/10.1046/j.1460-9568.2003.02530.x

Hermes, D. J. (1998a). Auditory and Visual similarity of Pitch Contours. *Journal of Speech, Language and Hearing Research*, *41*, 63–72.

Hermes, D. J. (1998b). Measuring the Perceptual Similarity of Pitch Contours. *Journal of Speech, Language and Hearing Research*, *41*, 73–82.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261.

Heyes, C. (2010a). Mesmerising mirror neurons. *NeuroImage*, *51*(2), 789–791. https://doi.org/10.1016/j.neuroimage.2010.02.034

Heyes, C. (2010b). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, *34*(4), 575–583. https://doi.org/10.1016/j.neubiorev.2009.11.007

Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research*, *22*(2), 233–240. https://doi.org/10.1016/j.cogbrainres.2004.09.009

Hintzman, D. L. (1986). "Schema abstraction" in a multiple trace memory model. *Psychological Review*, *93*, 411–428.

Hirst, D. (2016). On the automatic comparison and clonage of native and non-native speech prosody. In *Proceedings of Speech Prosody 2016*. Boston.

Hirst, D., & Espesser, R. (1993). Automatic Modelling of Fundamental Frequency Using a Quadratic Spline Function. *Travaux de l'Institut de Phonétique d'Aix*, *15*, 75–85.

Horner, V., & Whiten, A. (2004). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition*, *8*(3), 164–181. https://doi.org/10.1007/s10071-004-0239-6

Howell, D. C. (2006). *Méthodes statistiques en sciences humaines (6e ed.)* (De Boeck). Louvain-La-Neuve.

Jacobson, S. W. (1979). Matching Behavior in the Young Infant. *Child Development*, *50*(2), 425. https://doi.org/10.2307/1129418

Jakobson, R. (1963). *Essais de linguistique générale* (Editions de Minuit). Paris, France.

Jilka, M., Anufryk, V., Baumotte, H., Lewandowska, N., Rota, G., & Reiterer, S. (2008). Assessing individual talent in second language production and perception. In *New Sounds 2007:*

Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech. (pp. 224–239). Florianópolis, Federal University of Santa Catarina:

Jilka, M., Baumotte, H., Lewandowski, N., Reiterer, S., & Rota, G. (2007). Introducing a comprehensive approach to assessing pronunciation talent. *Proceedings of the 16th ICPhS, Saarbrücken*, 1737–1740.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In *Intonation* (pp. 209–242). Springer.

Kim, M. (2011). Phonetic convergence after perceptual exposure to native and nonnative speech: Preliminary findings based on fine-grained acoustic-phonetic measurement. In *Proceedings of The 17th International Congress of Phonetic Sciences*.

Kim, M. (2012). *Phonetic accommodation after auditory exposure to native and nonnative speech*. NORTHWESTERN UNIVERSITY.

Köpke, B. (2007). Language attrition at the crossroads of brain, mind, and society. In *Language Attrition, theoretical perspectives* (John Benjamins, Vol. 33). Amsterdam: Köpke, B., Schmid, M. S., Keijzer, M. & Dostert, S.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLEM-e). *Phil. Trans. Soc. B*, *363*, 979–1000.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425. https://doi.org/10.1121/1.417951

Lau, Y. W., Wagner, M., & Tran, D. (2004). Vulnerability of speaker verification to voice mimicking. In *Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on* (pp. 145–148). IEEE.

Laver, J. (1994). *Principles of Phonetics* (Cambridge University Press). Cambridge.

Laver, J., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In *Social Markers in Speech* (Cambridge University Press). Cambridge: Klaus Rainer Scherer & Howard Giles.

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.

Lewandowski, N. (2012). *Talent in nonnative phonetic convergence*. Universität Stuttgart.

Lewandowski, N., Jilka, M., Rota, G., Reiterer, S. M., & Dogil, G. (2007). Phonetic convergence as a paradigm of showing phonetic talent in foreign language acquisition. In *Proceedings of the 8th Annual Conference of the Cognitive Science Society of Germany*. Saarbrücken.

Lyche, C. (2005). Des règles aux contraintes : quelques aspects de la théorie de l'optimalité. In *Phonologie et phonétique -forme et substance-* (Hermes, pp. 209–240). Paris: Nguyen N., Wauquier-Graveline S. & Durand, J.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, *104*(50), 19751–19756.

MacLeod, B. (2014). Investigating the effects of salience and regional dialect on phonetic convergence in Spanish. In *Variation within and across Romance Languages: Selected papers from the 41st Linguistic Symposium on Romance Languages (LSRL), Ottawa, 5-7 May 2011* (John Benjamins, pp. 351–378). M. Côté & E. Mathieu.

Makuuchi, M. (2004). Is Broca's Area Crucial for Imitation? *Cerebral Cortex*, *15*(5), 563–570. https://doi.org/10.1093/cercor/bhh157

Mary, L., Anish Babu, K. K., & Joseph, A. (2012). Analysis and detection of mimicked speech based on prosodic features. *International Journal of Speech Technology*.

Mary, L., Anish Babu, K. K., Joseph, A., & George, G. M. (2013). Evaluation of mimicked speech using prosodic features. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 7189–7193). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6639058

McDonald, J. . (2014). *Handbook of Biological Statistics (3rd ed.)* (Sparky House Publishing). Baltimore, Maryland.

McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children: From over-imitation to super-copying. *British Journal of Psychology*, *102*(1), 1–18.

Mejvaldova, J. (2002). Caractéristiques temporelles de la parole imitée. In *Speech Prosody 2002 proceedings*. Aix-en-Provence: Laboratoire Parole et Langage.

Meltzoff, A. N., & Moore, M. K. (2005). Imitation et développement humain : les premiers temps de la vie*. *Terrain*, (44), 71–90. https://doi.org/10.4000/terrain.2455

Messum, P. R. (2007c). How children learn to pronounce: not by imitation but by their mothers' vocal mirroring. *Unpublished MS Submitted to 15th ICPhS, Saarbrucken*.

Messum, P. R. (2007b). Mirroring, not imitation, for the early learning of L1 pronunciation. *The Journal of the Acoustical Society of America*, *122*(5), 2997. https://doi.org/10.1121/1.2942701

Messum, P. R. (2007a). *The role of imitation in learning to pronounce.* University of London, London. Retrieved from http://discovery.ucl.ac.uk/1444832/

Michelas, A., & Nguyen, N. (2011). Uncovering the Effect of Imitation on Tonal Patterns of French Accentual Phrases. In *INTERSPEECH* (pp. 973–976).

Mitchell, R. W. (2002). Imitation as a Perceptual Process. In *Imitation in Animals and Artifacts* (MA: MIT Press, pp. 441–469). Cambridge: Nehaniv, C. L. & Dautenhahn, K.

Morau, H. (1893). Note sur le mimétisme, à propos de quelques insectes tropicaux. *Bulletins de la Société d'anthropologie de Paris*, *4*(1), 707–712. https://doi.org/10.3406/bmsap.1893.5486

Nadel, J. (1986). *Imitation et communication entre jeunes enfants* (PUF). Paris, France.

Nadel, J. (2005). Imitation et autisme. *Autisme, Cerveau et Développement*, 341–356.

Nadel, J. (2011). *Imiter pour grandir. Développement du bébé et de l'enfant avec autisme* (Dunod). Paris, France.

Nadel, J., Guérini, C., Pezé, A., & Rivet, C. (1999). The evolving nature of imitation as a transitory means of communication. In *Imitation in Infancy* (Ma : Cambridge University Press, pp. 209–234). Cambridge: Nadel J. & Butterworth G.

Nadel, J., & Potier, C. (2002a). Imiter et être imité dans le développement de l'intentionnalité. In *Imiter pour découvrir l'humain* (Presses Universitaire de France, pp. 83–104). Paris.

Nadel, J., & Potier, C. (2002b). Imitez, imitez, il en restera toujours quelque chose : le statut développemental de l'imitation dans le cas d'autisme. *Enfance*, *54*(1), 76–85. https://doi.org/10.3917/enf.541.0076.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender Differences in Vocal Accommodation:: The Role of Perception. *Journal of Language and Social Psychology*, *21*(4), 422–432. https://doi.org/10.1177/026192702237958

Nehaniv, C. L., & Dautenhahn, K. (2002). The Correspondence Problem. In *Imitation in Animals and Artifacts* (MIT Press, pp. 41–62). Cambridge, Mass.: Nehaniv, C. L. & Dautenhahn, K.

Nguyen, N. (2005). Perception de la parole. In *Phonologie et phonétique, forme et substance* (Hermès Science, Lavoisier, pp. 425–448). Paris: Noël Nguyen, Sophie Wauquier-Gravelines & Jacques Durand.

Nguyen, N., Wauquier, S., & Tuller, B. (2009). The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories. In *Approaches to Phonological Complexity* (Walter de Gruyter, pp. 193–217). Berlin: Pellegrino F., Marsico E., Chitoran I. & Coupé C.

Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman Children and the Origins of Human Cultural Cognition. *Psychological Science*, *21*(5), 729–736. https://doi.org/10.1177/0956797610368808

Nocaudie, O. (2012). *Imitation, convergence phonétique et apprentissage des langues étrangères au travers du prisme de la méthode verbo-tonale d'intégration phonétique*. Toulouse 2 Le Mirail, Toulouse.

Nocaudie, O., & Astésano, C. (2012). Prosodic structuring imitation in French L1 context-A first step towards correcting phonetic-prosodic features in L2 French. In *Proceedings of ISICS*. Aix-en-Provence.f

Nocaudie, O., Köpke, B., Giraudo, H., & Calderone, B. (2015). Exploring phonotactic and morphonotactic constraints in the acquisition of consonant clusters in L1 French. Presented at the 3rd International Conference on Phonotactics and Phonotactic Modeling (PPM 2015), Vienne.

Ohala, J. ., & Jaeger, J. (1986). Introduction. In *Expimental phonology* (Academic Press). Orlando: Ohala J.J. & Jaeger J.

Paradis, M. (1993). Linguistic, psycholinguistic and neurolinguistic aspects of "interference" in bilingual speakers: The activation threshold hypothesis. *International Journal of Psycholinguistic*, *9*(2), 133–145.

Pardo, J. S. (2000). *Imitation and coordination in spoken communication*. Yale, New Haven.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–93. https://doi.org/10.1121/1.2178720

Pardo, J. S. (2010). Expressing Oneself in Conversationnal Interaction. In *Expressing Oneself/Expressing One's Self: Communication, Cognition, and Identity* (Psychology Press/Taylor & Francis., pp. 183–196). Hove, England: E. Morsella.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00559

Pardo, J. S. (2013). Phonetic convergence in shadowed speech: A comparison of perceptual and acoustic measures. In *INTERSPEECH* (pp. 530–534). Retrieved from http://lpp.ilpga.fr/PDF/IS130085/IS130085.PDF

Pardo, J. S. (2013). Reconciling diverse findings in studies of phonetic convergence. In *Proceedings of Meetings on Acoustics* (Vol. 19, p. 60140). Acoustical Society of America. Retrieved from http://scitation.aip.org/content/asa/journal/poma/19/1/10.1121/1.4798479

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190–197.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception & Psychophysics*, *72*(8), 2254–2264.

Perrot, P., Aversano, G., & Chollet, G. (2007). Voice disguise and automatic detection -review and perspective-. In *Progress in Nonlinear Speech Processing* (Springer, pp. 101–117). Berlin: Yannis Stylianou, Marcos Faundez-Zanuy & Anna Esposito.

Petit, O., & Pascalis, O. (2009). Dossier Imitation-Introduction générale. *Revue de Primatologie*, (1). Retrieved from http://primatologie.revues.org/279

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pierrehumbert, J. B., Beckman, M. E., & Ladd, D. R. (2000). Conceptual Foundations of Phonology as a Laboratory Science. In *Phonological Knowledge: Its Nature and Status* (Cambridge University Press, pp. 273–303). Cambridge: Burton-Roberts N., Carr P. & Docherty G.

Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux Du Cercle de Linguistique de Prague*, *4*, 76–96.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Animal behaviour: elephants are capable of vocal learning. *Nature*, *434*(7032), 455–456.

Postma-Nilsenová, M., & Postma, E. (2013). Auditory perception bias in speech imitation. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00826

Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research*, *25*(3), 632–640. https://doi.org/10.1016/j.cogbrainres.2005.08.020

Puren, C. (1988). *HIstoire des méthodologies* (Nathan-Clé International). Paris.

Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., … Ackermann, H. (2011). Individual Differences in Audio-Vocal Speech Imitation Aptitude in Late Bilinguals: Functional Neuro-Imaging and Brain Morphology. *Frontiers in Psychology*, *2*.

Reiterer, S. M., Hu, X., Sumathi, T. A., & Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology*, *4*.

Renard, R. (1979). *Introduction à la méthode verbo-tonale de correction phonétique* (Didier). Bruxelles.

Renard, R. (2002a). *Apprentissage d'une langue étrangère/seconde. 2. La phonétique verbo-tonale* (De Boeck Université). Bruxelles.

Renard, R. (2002b). Une phonétique immergée. In *Apprentissage d'une langue étrangère/seconde 2. La phonétique verbo-tonale* (De Boeck Université). Bruxelles: Renard R.

Révis, J. (2013). *La voix et soi. Ce que notre voix dit de nous* (De Boeck Solal). Louvain La Neuve.

Révis, J., De Looze, C., & Giovanni, A. (2013). Vocal Flexibility and Prosodic Strategies in a Professional Impersonator. *Journal of Voice*, *27*(4), 524.e23-524.e31.

Rilliard, A., Allauzen, A., & de Mareüil, P. B. (2011). Using Dynamic Time Warping to Compute Prosodic Similarity Measures. In *INTERSPEECH* (pp. 2021–2024). Retrieved from http://perso.limsi.fr/mareuil/publi/IS110831.pdf

Rivenc, P. (2002). Place et rôle de la phonétique dans la méthodologie SGAV. In *Apprentissage d'une langue étrangère/seconde 2. La phonétique verbo-tonale* (De Boeck Université, pp. 25–34). Bruxelles: Renard R.

Roch Lecours, A., & Joanette, Y. (1980). Linguistic and Other Psychological Aspects of Paroxysmal Aphasia. *Brain and Language*, (10), 1–23.

Rossi, M., Di Cristo, A., Hirst, D., Martin, P., & Nishinuma, Y. (1981). *L'intonation, de l'acoustique à la sémantique* (Klincksieck). Paris, France.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436.

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J.-L., & Nguyen, N. (2013). Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00422

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336–354.

Schweitzer, A., & Lewandowski, N. (2014). Social Factors in Convergence of F1 and F2 in Spontaneous Speech. In *Proceedings of the 10th International Seminar on Speech Production*. Cologne.

Singh, S. (2010). *Le dernier théorème de Fermat* (Librairie Fayard).

Speed, M. P. (1993). Muellerian mimicry and the psychology of predation. *Animal Behavior*, *45*, 571–580.

t'Hart, J. (1991). F0 stylization in speech: Straight lines versus parabolas. *Journal of the Acoustical Society of America*, pp. 3368–3370.

Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science*, *291*(5513), 2564–2569.

Tchernichovski, O., & Nottebohm, F. (1998). Social inhibition of song imitation among sibling male zebra finches. *Proceedings of the National Academy of Sciences*, *95*(15), 8951–8956.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*(4), i. https://doi.org/10.1037/h0092987

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993). Imitative Learning of Actions on Objects by Children, Chimpanzees, and Enculturated Chimpanzees. *Child Development*, *64*(6), 1688. https://doi.org/10.2307/1131463

Tronchet, D. (1996). *Les rois du rire* (Albin Michel). Paris.

Troubetzkoy, N. S. (1939). *Grundzüge der Phonologie* (Edition original 1939). Pragues.

Tuller, B., Nguyen, N., Lancia, L., & Vallabha, G. K. (2010). Nonlinear Dynamics in Speech Perception. In *Nonlinear Dynamics in Human Behavior* (Springer, pp. 135–150). Berlin: Huys R. & Jirsa V. K.

Tzourio-Mazoyer, N. (2003). Les réseaux neuraux dédiés au langage chez l'adulte. In *Cerveau et langage* (Hermes Sciences, Lavoisier, pp. 67–102). Paris: Etard, O. & Tzourio-Mazoyer N.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of Native-and Foreign-accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and Speech*, *53*(4), 510–540. https://doi.org/10.1177/0023830910372495

Veltkamp, R. C. (2001). Shape matching: similarity measures and algorithms (pp. 188–197). IEEE Computer Society. https://doi.org/10.1109/SMA.2001.923389

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, *34*(3), 343–371. https://doi.org/10.1016/j.wocn.2005.09.001

Welby, P., Bertrand, R., Portes, C., & Astésano, C. (2016). Realization of the French initial accent: stability and individual differences. In *TIE Conference 2016–7th conference on Tone and Intonation in Europe (TIE), Canterbury* (pp. 1–3).

Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. In *Advances in the Study of Behavior* (Slater P.J.B., Rosenblatt J.S., Beer C. & Milinsky M., pp. 239–283). New York: Academic Press.

Wozniak, R. H. (1999). Classics in the History of Psychology -- Introduction to Thorndike (1911). Retrieved July 29, 2015, f

Zazzo, R. (1957). Le problème de l'imitation chez le nouveau-né. *Enfance*, *10*(2), 135–142. https://doi.org/10.3406/enfan.1957.1350

Zentall, T. R. (2006). Imitation: definitions, evidence, and mechanisms. *Animal Cognition*, *9*(4), 335–353. https://doi.org/10.1007/s10071-006-0039-2

Zetterholm, E. (2000). The significance of phonetics in voice imitation. In *8th Aust. Int. Conf. Speech Sci. & Tech* (pp. 342–347). Canberra.

Zetterholm, E. (2002). A case study of successful voice imitation. *Log Phon Vocol*, *27*, 80–83.

Zetterholm, E. (2002). A comparative survey of phonetic features of two impersonators. In *Fonetik* (Vol. 44, pp. 129–132).f

Zetterholm, E. (2002). Intonation pattern and duration differences in imitated speech. In *Speech Prosody 2002, International Conference*. Retrieved from http://www.isca-speech.org/archive_open/sp2002/sp02_731.html

Zetterholm, E. (2006). Same speaker – different voices A study of one impersonator and some of his different imitations. In *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (pp. 70–75). Auckland: Australasian Speech Science & Technology Association.

Zetterholm, E. (2009a). Impersonation–reproduction of speech. *Working Papers in Linguistics*, *49*, 176–179.

Zetterholm, E. (2009b). Voice imitation-different ways of saying mobilsvar. *Working Papers in Linguistics*, *48*, 193–207.

Annexes

Annexe A : Formulaire de consentement éclairé

Annexe B : Divers scripts praat utilisés durant notre travail

Annexe C : Article paru dans les actes des JEP 2016

Annexe D : Article paru dans les actes de Speech Prosody 2016

D’autres annexes sont fournies sur un CD-Rom (Sons et Figures)

Unité de Recherche Interdisciplinaire Octogone-Lordat (EA 4156)
Université de Toulouse 2 Jean Jaurès
5 allées Antonio Machado
31058 Toulouse CEDEX 9

Formulaire de consentement libre, éclairé et exprès

Expériences comportementales en psycholinguistique

Je certifie avoir donné mon accord pour participer à une étude comportementale de psycholinguistique. J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi.

Au cours de cette expérience, j'accepte que soient recueillies des données de perception de la parole. Je comprends que les informations recueillies sont strictement confidentielles et à usage exclusif des investigateurs concernés.

J'ai été informé que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'accepte que les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par l'Unité de Recherche Interdisciplinaire Octogone-Lordat (EA 4156).

_____

Expérience :

Date : ……/……/……..

Nom du volontaire :

Signature, précédée de la mention « lu et approuvé »

Nom de l'expérimentateur :

Signature :

Annexe : Script PRAAT (pour Windows, PRAAT v 5.3.50)

Script pour la modification manuelle d'un grand nombre de TextGrids

```
#################################################################
### Read in all WAV and TextGrids files with the same name
### in specified directories, open Editor, allow you to change boundaries before
saving
###
### Olivier Nocaudie
### nocaudie@univ-tlse2.fr
### 27/07/2016
### Praat 5.3.50
###
### Adapted from Read-in-WAV-and-TextGrid
### Rafèu Sichel-Bazin
### rsichelb@uos.de
### 13/06/2013
### Praat 5.2.22
###
### Itself adapted from readin-files-simple.praat
### Pauline Welby
### welbyp@tcd.ie
### September 2008
### Praat 5.0.32
#################################################################
form Select directories
      comment Where are the WAV files kept?
      sentence wav_dir
      comment Where are the TextGrid files kept?
      sentence txt_dir
      comment Where should the new TextGrid files saved ?
      sentence output_dir
      comment Should new TextGrid have a prefix in its name ? (If not, comment
following line with hash)
      sentence output_prefix
endform

# lists all .wav files
Create Strings as file list... list 'wav_dir$'\*.wav
# loop that goes through all files

numberOfFiles = Get number of strings
for ifile to numberOfFiles
   select Strings list
   fileName$ = Get string... ifile
   baseFile$ = fileName$ - ".wav"
   # Read in the Sound files with that base name
   Read from file... 'wav_dir$'\'baseFile$'.wav
   Read from file... 'txt_dir$'\'baseFile$'.TextGrid
```

```
   #Open Editor and pause so you can work, save when you proceed
   select Sound 'baseFile$'
   plus TextGrid 'baseFile$'
   View & Edit
      beginPause ("Modify TextGrid if necessary and/or proceed")
      endPause ("Continue", 1)
   select Sound 'baseFile$'
   Remove
   select TextGrid 'baseFile$'
   Write to text file... 'output_dir$'\'output_prefix$'_'baseFile$'.TextGrid
   #Clear object list
   Remove
endfor
select all
Remove
```

```
Script pour la segmentation des fichiers longs

form
sentence directory
sentence output output
sentence name dialogue
endform


do ("Open long sound file...", "'directory$'\'name$'.wav")
Read from file... 'directory$'\'name$'.TextGrid

select TextGrid 'name$'
numberofintervals = Get number of intervals... 5
for i from 1 to numberofintervals
      label5$ = Get label of interval... 5 i
      if label5$ <> ""

            start = Get start point... 5 i
            end = Get end point... 5 i
            select LongSound 'name$'
            Extract part... start end rectangular 1 no
            Save as WAV file... 'directory$'\'output$'\'label5$'.wav

            select TextGrid 'name$'
            do ("Extract part...", start, end, "no")
            Save as short text file... 'directory$'\'output$'\'label5$'.TextGrid


      select TextGrid 'name$'
      endif
endfor
plus Sound 'name$'
select all
Remove
```

```
Script de préparation pour la TSR/TSRm

# Un ensemble de TextGrid et .wav en entrée
# Tier 1 doit être les phonèmes
# 1) Le script crée une point tier et ses points en fonction de la durée des
voyelles (et semi consonnes) annotées dans la tier phonèmes
# 2) Le script relève les valeurs du pitch et les écrit comme étiquettes des
points
# 3) Le script rapporte les valeurs temporelles des points en les centrant à
zéro, ainsi que les valeurs de pitch dans un fichier tabulé, sauvé dans le
dossier output
# 4) Le script duplique la point tier et la nettoie de ses labels pour la
préparer à l'annotation tonale, puis sauve le textgrid dans le dossier output

# INPUT (& OUVERTURE)
form
   sentence directory
   sentence output
   sentence output_prefix
endform


#Crée la liste de fichiers . wav et les compte

Create Strings as file list... fileList 'directory$'\*.wav
nombredefichiers = Get number of strings

#Début loop global
for i from 1 to nombredefichiers

#Variable sans extension
 file$ = Get string... i
   basefile$ = file$ - ".wav"

#Ouvre TextGrid & Insère la point tier Ton
   Read from file... 'directory$'\'basefile$'.TextGrid
   do ("Insert point tier...", 1, "Ton")

 #Variable voyelle
 voyelles = Get number of intervals... 2

            #Loop for 1 : crée les points de la point tier en fonction du
contenu des intervals de la tier phon
            for k from 1 to voyelles
            label_voy$ = Get label of interval... 2 'k'

                  #Loop if label = voyelle, et loop if durée, met un certain
nombre de points en fonction de cette valeur
                  if label_voy$ = "a" or label_voy$ = "e"  or label_voy$ = "A"
or label_voy$ = "@" or label_voy$ = "i" or label_voy$ = "a~" or label_voy$ = "o"
```

```
or label_voy$ = "9" or label_voy$ = "y" or label_voy$ = "o~" or label_voy$ =
"U~" or label_voy$ = "u" or label_voy$ = "eu" or label_voy$ = "E" or label_voy$
= "j" or label_voy$ = "w"
                  start = Get start point... 2 'k'
                  end = Get end point... 2 'k'
                  dur = end - start
                        if dur <= 0.1
                        time = start + dur / 2
                        Insert point... 1 'time'
                        elif dur > 0.1 and dur < 0.2
                        time1 = start + dur / 3
                        Insert point... 1 'time1'
                        time2 = start + (2 * dur / 3)
                        Insert point... 1 'time2'
                        elif dur >= 0.2
                        time1 = start + dur / 4
                        time2 = start + dur / 2
                        time3 = start + (3 * dur / 4)
                        Insert point... 1 'time1'
                        Insert point... 1 'time2'
                        Insert point... 1 'time3'
                        else
                        #clôture loop if durée
                        endif
                  else
                  #clôture loop if label
                  endif
            #clôture loop for 1
            endfor

      #Ouvre .wav correspondant au Textgrid ouvert précédemment

      Read from file... 'directory$'\'file$'

      #Création Pitch
      To Pitch (ac)... 0 75 15 no 0.03 0.45 0.01 0.35 0.14 600
      Down to PitchTier

      select TextGrid 'basefile$'
      point = Get number of points... 1
      corr = Get time of point... 1 1

            #Récupère les valeurs de pitch et les mets en label des points
correspondants
            for u from 1 to 'point'
                  select TextGrid 'basefile$'
                  point_time = Get time of point... 1 'u'
                  select PitchTier 'basefile$'
                  pitch = Get value at time... 'point_time'
                  select TextGrid 'basefile$'
```

```
                  #NB : la variable "fixed$(pitch,3)" permet de considérer des
valeurs numériques comme des suites de caractères en fixant le nombre de
décimales (ici, 3),
                  #nécessaire pour les champs demandant du texte dans praat
(comme string$, mais qui ne limite pas les décimales)
                  do ("Set point text...", 1, u, fixed$(pitch,0))

                  #Centrage des données temporelles
                  new_time = point_time - corr

                  #Ecriture des valeurs temps (corrigées) et pitch dans un
fichier txt
                  fileappend
"'directory$'\'output$'\'output_prefix$'_'basefile$'.txt"
'new_time''tab$''pitch''newline$'
            endfor

      #Duplique la point tier créée précédemment et la prépare pour l'étape
d'annotation tonale
      do ("Duplicate tier...", 1, 1, "Ton_label")

      point_label = do ("Get number of points...", 1)
            for g from 1 to 'point_label'
                  select TextGrid 'basefile$'
                  label_label$ = Get label of point... 1 'g'
                  if label_label$ <> ""
                        do ("Set point text...", 1, g, "")
                  else
                  endif
            endfor

      Save as short text file...
'directory$'\'output$'\'output_prefix$'_'basefile$'.TextGrid
select Strings fileList
#Fin loop global
endfor

#Nettoyage fenêtre objet
   select all
Remove
```

Acquisition d'informations acoustiques

```
# INPUT (& OUVERTURE)
form
   sentence directory
   sentence output_dir
   sentence data_name
   positive pitch_floor
   positive pitch_ceiling

endform


# Crée la liste de fichiers . wav et les compte

Create Strings as file list... fileList 'directory$'\*.wav
nombredefichiers = Get number of strings

# Début loop global
for i from 1 to nombredefichiers

# Variable sans extension
 file$ = Get string... i
   basefile$ = file$ - ".wav"

   # Ouvre TextGrid & wav.
   Read from file... 'directory$'\'basefile$'.TextGrid
   Read from file... 'directory$'\'file$'

   # Création Pitch
      To Pitch (ac)... 0 pitch_floor 15 no 0.03 0.45 0.01 0.35 0.14
pitch_ceiling
      Down to PitchTier


      # Infos globales f0
      select TextGrid 'basefile$'
      sentence = Get number of intervals... 4
            if sentence > 3

                  last_bound = sentence - 1
                        repeat
                              select TextGrid 'basefile$'
                              do ("Remove left boundary...", 4, last_bound)
                              remove = Get number of intervals... 4
                              last_bound = remove - 1
                        until remove = 3
            else
            endif
```

```
      sentence2 = Get number of intervals... 4

            for z to sentence2
                        select TextGrid 'basefile$'
                        sentence$ = Get label of interval... 4 'z'
                        if sentence$ <> ""
                              start_sent = Get start point... 4 'z'
                              end_sent = Get end point... 4 'z'
                              sentence_duration = end_sent - start_sent
                              select Pitch 'basefile$'
                              min_pitch = do ("Get minimum...", start_sent,
end_sent, "Hertz", "Parabolic")
                              max_pitch = do ("Get maximum...", start_sent,
end_sent, "Hertz", "Parabolic")
                              mean_pitch = do ("Get mean...", start_sent,
end_sent, "Hertz")
                              range_pitch = 12*log2(max_pitch/min_pitch)
                              fileappend "'output_dir$'\'data_name$'.txt"
'basefile$''tab$''min_pitch''tab$''max_pitch''tab$''mean_pitch''tab$''range_pitc
h''tab$''sentence_duration''tab$'
                        else
                        endif
            endfor



      # Durées syllabiques et débit

      select TextGrid 'basefile$'
      syllables = Get number of intervals... 3

      syll_count = 0
      syll_add = 1
      repeat
            syll_label$ = Get label of interval... 3 'syll_add'
                  if syll_label$ <> ""
                  syll_count = syll_count + 1
                  syll_add = syll_add + 1
                  else
                  syll_add = syll_add + 1
                  endif
      until syll_add = syllables
      fileappend "'output_dir$'\'data_name$'.txt" 'syll_count''tab$'
      flow = syll_count / sentence_duration
      fileappend "'output_dir$'\'data_name$'.txt" 'flow''tab$'
```

```
            for k from 1 to syllables
                  syll_label2$ = Get label of interval... 3 'k'
                        if syll_label2$ <> ""
                              syll_start = Get start point... 3 'k'
                              syll_end = Get end point... 3 'k'
                              syll_dur = syll_end - syll_start
                              fileappend "'output_dir$'\'data_name$'.txt"
'syll_dur''tab$'
                        else
                        endif
            endfor

      fileappend "'output_dir$'\'data_name$'.txt" 'newline$'
      select Strings fileList
endfor

select all
Remove
```

Quelle(s) mesure(s) de similarité prosodique comme évaluation de l'imitation ?

Olivier Nocaudie, Corine Astésano

U.R.I. Octogone-Lordat (E.A. 4156), Université de Toulouse, UTM

nocaudie@univ-tlse2.fr, corine.astesano@univ-tlse2.fr

RESUME

La performance imitative des locuteurs varie de celle du professionnel, expert, à celle du naïf, plus ou moins talentueux. L'étude de l'imitation souligne la difficulté pour trouver des indices mesurables de la réussite d'une imitation. Dans cette étude exploratoire, des contours de *f0* recueillis au fil de tâches d'imitation sont testés au moyen d'une double approche : mesure objective par le biais de deux mesures de la similarité prosodique reportées dans la littérature et évaluation perceptive par un panel de 15 d'auditeur naïfs. Nos premiers résultats indiquent une bonne corrélation entre les deux approches et soulèvent la question du choix de l'indice mesurable qui rendrait le mieux compte d'une imitation au niveau tonal. Ils soulignent également la variabilité interindividuelle des comportements imitatifs en parole tout en ouvrant des perspectives intéressantes dans le domaine de la formation à la phonétique corrective par la Méthode Verbo-tonale.

ABSTRACT

Which measure(s) of prosodic similarity as an evaluation of imitation?

Imitative proficiency across speakers is highly variable. Studies on imitation underlines how difficult it is to find measurable cues to assess a successful imitation. In this exploratory study, *f0* contours stem from imitations tasks are tested in a double approach: objective measurements of prosodic similarity using two measures reported in the literature and perceptive evaluation by a panel of 15 naïve listeners. Our first results indicate a good correlation between the two approaches and they raise the question concerning the selection of the measurable factor assessing a successful imitation at a tonal level. Meantime, these results underline an imitative proficiency's variability across speakers while opening perspectives in the domain of phonetic correction using the Verbo Tonal Method.

MOTS-CLES : parole, prosodie, imitation, mesures objectives, évaluations perceptives.
KEYWORDS: speech, prosody, imitation, objective measurements, perceptive evaluation.

Introduction

Les études sur l'imitation de la parole rapportent différents types de comportements imitatifs, comme la convergence (une adaptation mutuelle des interlocuteurs au fil de la conversation, Pardo, 2006), l'impersonnation (la tentative d'usurper la voix de l'autre, Révis, De Looze, & Giovanni, 2013), la simple imitation (Mixdorff, Cole, & Shattuck-Hufnagel, 2012) ou le *shadowing* (Dufour & Nguyen, 2013; Goldinger, 1998). Définir ces comportements en se basant sur des facteurs comme le contexte de production (Lewandowski, 2012) ou l'intention de l'imitateur (Donald, 1993) aboutit à les différencier. Pourtant, ils partagent une similitude définitoire majeure : pour être qualifiés de comportements imitatifs, la production du locuteur doit être perçue par un tiers comme similaire au modèle de l'imitation. Par conséquent, l'étude de l'imitation en parole vise à faire produire, puis à observer des changements comportementaux dans la manière de parler de locuteurs (naïfs ou experts), au niveau lexical ou phonétique. Pour ce qui est du niveau phonétique, deux questions ne cessent de représenter un défi méthodologique : quels sont les paramètres du signal sonore perçus puis imités en priorité par les locuteurs ; de quelle manière évaluer et comparer les paramètres choisis, entre leur modèle et leur(s) imitation(s) ? Choisir quel(s) paramètre(s) acoustique(s) mesurer et lier aux résultats d'évaluations perceptives de l'imitation demeure un choix crucial (Pardo, 2013).

Les imitateurs professionnels parviennent à ajuster globalement leur voix aux spécificités de leurs voix cibles, mais ils sont aussi capables d'imiter en reproduisant des variations instantanées de contours intonatifs ou de fréquence et durée des pauses (stratégies de synchronie) (Révis et al., 2013). En revanche, d'après ces derniers travaux, les locuteurs naïfs semblent limités à des stratégies d'ajustement global. Cette dernière remarque soulève une série de questions connexes, en lien avec les stratégies de synchronie : **(1)** jusqu'à quel point un locuteur naïf est-il capable de reproduire un patron

prosodique perçu (variation instantanée) ; **(2)** est-il possible d'entraîner un locuteur à reproduire fidèlement des contours intonatifs ; **(3)** comment peut-on évaluer leur réussite –et leur échec– dans l'accomplissement de cette tâche

Les deux premières questions ont une pertinence certaine dans le domaine de l'enseignement de la prononciation à des locuteurs de langue seconde (L2), plus particulièrement en se plaçant dans la cadre théorique proposé par la Méthode Verbo-Tonale d'intégration phonétique (MVT). La MVT postule que les erreurs de prononciation en L2 seraient dues à un biais de perception de la L2. Afin de neutraliser les effets de ce biais, la MVT propose une rééducation perceptive, au moyen d'un ensemble de procédés correctifs où l'influence de la prosodie sur les segments phonétiques joue un rôle majeur. Un enseignant recourant à la MVT doit donc avoir une conscience et un contrôle prosodique efficace, notamment lorsqu'il doit produire des énoncés délexicalisés afin de faciliter la perception par l'apprenant des caractéristiques rythmiques et intonationnelles de la langue cible (Billières, Alazard, Astésano, & Nocaudie, 2013).

Intrinsèquement, une séquence de correction phonétique représente un cas typique d'interaction imitative. En effet, durant une interaction MVT, l'enseignant comme l'apprenant doivent imiter ou réitérer des sons de son interlocuteur. L'apprenant doit répéter le modèle proposé par l'enseignant, ce qui peut conduire à questionner le lien entre perception et (re)production de la parole chez le sujet devenant bilingue. De son côté, l'enseignant doit produire systématiquement des patrons prosodiques phonologiquement cohérents, soulevant ainsi la question du contrôle de sa production, plus particulièrement au niveau mélodique.

Si les questions **(1)** et **(2)** peuvent être liées à la fois à l'enseignant et à l'apprenant, la présente étude se focalise sur la capacité de l'enseignant à reproduire systématiquement des patrons prosodiques. En effet, avant même de pouvoir tester la capacité de l'apprenant à (re)produire les paramètres phonétiques d'un énoncé, il faut s'assurer que l'enseignant même est capable d'imiter les éléments prosodiques saillants de la parole. Or, la pratique de la MVT implique la reproduction de paramètres prosodiques de manière maîtrisée et la mise en valeur certains événements pour faciliter à l'apprenant la perception des sons cibles. Ainsi, cette étude propose de tester en premier lieu la capacité des locuteurs de L1 à produire des imitations de contours prosodiques. Ce faisant, nous nous intéresserons plus particulièrement à notre question **(3)**, à savoir : évaluer le degré de réussite d'une imitation. La méthode d'évaluation de la (dis)similarité prosodique pourrait conduire à la création d'un outil d'évaluation du contrôle prosodique de l'enseignant de L2, et ainsi être utile à leur formation dans le domaine de la correction phonétique.

La littérature sur l'imitation parolière en langue française est réduite, et peu d'études ont abordé plus spécifiquement la reproduction des indices prosodiques (voir cependant Michelas & Nguyen, 2011 à propos de l'accent initial). Cette communication est la poursuite d'une autre étude préliminaire qui décrivait la capacité de locuteurs à imiter les indices prosodiques de phrases contrôlées, au cours de trois tâches d'imitation, d'une simple répétition à une exagération (Nocaudie & Astésano, 2012).

Matériel linguistique : un corpus d'imitation(s)

Notre corpus d'imitation est issu d'un corpus de phrases (*Corpus d'Edimbourg : CE*) présentant une ambigüité syntaxique qui peut être résolue par la production des indices prosodiques pertinents. L'ambigüité syntaxique dérive de la manipulation de la portée de l'adjectif, comme dans « les gants et les bas lisses », où l'adjectif (A) « lisses » qualifie alternativement le second nom « bas » uniquement ([les gants][et les bas lisses]) (*Condition 1 : Cond-1*) ou bien les deux noms ([les gants et les bas][lisses]) (*Condition 2 : Cond-2*). La longueur des noms et des adjectifs varie de une à quatre syllabes. Le contrôle de l'ambigüité syntaxique et de la longueur des constituants nous permet d'observer les indices prosodiques (proéminences, frontières, tons, pauses…) utilisés dans la linéarisation syntaxique des énoncés oraux (voir Astésano, Bard, & Turk, (2007) pour le détail de la constitution du *CE)*.

Du *CE*, 16 énoncés prononcés par une locutrice ont été sélectionnés comme stimuli auditifs pour la constitution du corpus *Imitation* (*CI*). Ces phrases comportaient uniquement deux longueurs de noms (noms tri- et quadrisyllabiques) combinées à des adjectifs de une à quatre syllabes, prononcées dans les deux conditions syntaxique (*Cond-1* & *Cond-2*). 8 locuteurs naïfs ont dit ces phrases au fil de trois tâches différentes : a) une simple répétition (*REP*), b) une imitation (*IMI*), c) une exagération des phrases de la locutrice (*EXA*). Les données de 2 locuteurs ont été exclues du *CI* en raisons de facteurs physiologiques ayant eu une incidence sur leur voix (stress induit par la situation expérimentale, timbre éraillé). Durant chaque tâche, les imitateurs répétaient chaque phrase 3 fois, dans un ordre aléatoire, nous permettant d'obtenir un total de 864 phrases (16 énoncés * 2 conditions syntaxiques * 3 occurrences * 3 tâches * 6 imitateurs). Afin d'évaluer la capacité implicite des locuteurs à imiter la parole, l'attention des locuteurs n'était pas focalisée sur le

fait d'imiter durant la tâche a). Il leur était simplement demandé de « *dire la phrase entendue en préservant sa structure* ». Durant les tâches b) & c), il leur était demandé explicitement de produire une imitation ou une exagération des énoncés entendus. Cependant, l'expérimentateur n'a pas dirigé l'attention des imitateurs sur les indices prosodiques.

La présente étude vise à comparer des données perceptives (Test AX) et objectives (issues d'un algorithme) obtenues à partir d'un extrait du *CI*. Nous avons sélectionné un sous-corpus de 4 énoncés de 2 longueurs différentes pour tester la robustesse de notre algorithme de mesure de la similarité prosodique. En effet, toutes les phrases sont prises dans la condition *Cond-1* car la désambigüisation syntaxique est marquée par une pause silencieuse entre le premier et le second nom. De fait, la présence de cette pause présente un intérêt particulier pour tester la robustesse de l'algorithme dans la mesure où celui-ci a une nette tendance à aligner les silences lorsqu'il évalue la similarité prosodique.

Les tests ont été menés sur la production de 4 imitateurs (Sp1, Sp3, Sp5 & Sp7), qui étaient appariés pour prononcer certains énoncés. Sp1 (femme) & Sp5 (femme) ont dit « *Les bagatelles et les balivernes sottes* » et « *Les bonimenteurs et les baratineurs fades* » ; Sp3 (homme) & Sp7 (femme) ont prononcé « *Les bagatelles et les balivernes saugrenues* » et « *Les bonimenteurs et les baratineurs fabuleux* ». Nos résultats ont été calculés sur 18 énoncés par sujet, produisant alors un nombre total de 72 couples de phrases (modèle *vs.* imitation) : [(2 énoncés * 3 répétitions * 3 tâches) * 4 sujets].

Méthode : mesures objectives & évaluations perceptives d'imitations prosodiques

Un problème constant dans l'évaluation de l'imitation en parole réside dans la relative absence de congruence entre la multitude de paramètres acoustiques pouvant être mesurés, alors que ces derniers divergent ou convergent avec le modèle. La mélodie, et son corrélat physique, la *f0* ont été décrits comme une cible de choix pour l'imitateur (Révis et al., 2013). Par ailleurs, les procédés correctifs de la MVT reposent largement sur la manipulation des contours mélodiques. Ainsi, notre méthode se concentre principalement sur la mesure de la distance physique entre deux paires de contours *f0* d'une part et sur l'évaluation perceptive de leur ressemblance d'autre part.

Dynnamic Time Warping (DTW) & (dis)similarité prosodique

Mesurer objectivement l'imitation d'un contour mélodique équivaut à trouver s'il y a une distance physique entre le contour modèle et sa reproduction, *i.e.* connaître le degré de correspondance entre les deux formes de contours intonatifs.

Ceci étant dit, comparer des formes impose une normalisation tonale et un alignement temporel des pics et des creux décrits par les contours de *f0* (DTW). Cette méthode d'interpolation non linéaire, qui force l'alignement entre le modèle et sa reproduction, améliorerait les scores de corrélation entre les formes, notamment si les contours intonatifs sont fonctionnellement similaires, *i.e.* s'ils partagent le même patron accentuel (Rilliard *et al.*, 2011). La distance entre les paires de contours intonatifs a été calculée au moyen de deux mesures similaires à celles proposées par Hermes (1998) où : *w(t)* correspond au décours temporel du facteur de poids (la somme des *subharmonic sumspectrum* du contour imité), *W* est l'intégrale du contour de 0 à *T* (avec *T*, la durée totale du contour) et *f1* & *f2*, la paire de contours testée par l'algorithme.

Pour ces travaux, la procédure de normalisation de la *f0* choisie diffère de celle d'Hermes. En effet, nous avons divisé chaque valeur de *f0* par la valeur maximale de *f0* de l'énoncé (*f1 = 1/p1max*). Cette procédure de normalisation permet de comparer plus facilement les imitateurs masculins et féminins en ramenant la variation des courbes de *f0* dans des valeurs comprises entre 0 et 1. Par ailleurs, cette procédure de normalisation facilite le travail de la partie DTW de l'algorithme. Les valeurs de *f0* ont été extraites à l'aide du logiciel PRAAT (Boersma, 2001) et le taux d'échantillonnage était d'une valeur par milliseconde (Rilliard, Allauzen, & de Mareüil, 2011).

Après normalisation et DTW, nous avons calculé la différence de la moyenne des moindres carrés (ci-après, L_2) de chaque paire de contours comme suit :

$$L_2 = \left\{\frac{1}{W}\int_0^T w(t)\left|f_1(t) - f_2(t)\right|^2 dt\right\} \qquad (1)$$

Le coefficient de corrélation *r* entre deux contours *f1* et *f2* ont ensuite été calculés de la manière suivante :

$$r = \frac{\frac{1}{W}\int_0^T w(t) f_1(t) f_2(t) dt}{\sqrt{\left\{\frac{1}{W}\int_0^T w(t)\left|f_1(t)\right|^2 dt \frac{1}{W}\int_0^T w(t)|f_2(t)|^2 dt\right\}}} \qquad (2)$$

Afin de pouvoir comparer les coefficients de corrélation, Hermes propose de transformer *r* en *Z* de Fisher (ci-après, *Zr*), que nous calculons ainsi :

$$Z_{f_1 f_2} = \frac{1}{2} ln \frac{1+r_{f_1 f_2}}{1-r_{f_1 f_2}} \qquad (3)$$

L_2 mesure les changements rapides de *f0*, tandis que *Zr* est une mesure holistique de la proximité de la forme de deux contours. Il est intéressant de tester leur complémentarité vis-à-vis des jugements perceptifs que nous recueillons par ailleurs.

Finalement, chaque phrase reproduite a reçu un rang en fonction de ses scores *L2* et *Zr*. L_2 est une mesure de dissimilarité (soit, plus L_2 a une valeur élevée, plus grande est la dissimilarité entre le modèle et sa reproduction). La phrase avec le L_2 le plus bas a donc été classée comme rang 1, celle avec le L_2 le plus haut, rang 72. *Zr* est une mesure de similarité (plus *Zr* est haut, plus la similarité entre les deux contours est grande). La phrase avec le plus haut *Zr* a donc reçu le rang 1, celle avec le plus bas, le rang 72.

Ainsi, nous avons obtenu 2 classements différents (en fonction de L_2 ou de *Zr*), qui seront comparés au classement dérivé des résultats des évaluations perceptives.

Test AX de jugement de la similarité

Comme le note Pardo (2013), l'imitation en parole devrait être évaluée objectivement et subjectivement, *i.e.* physiquement et perceptivement. A cette fin, nous avons complété les mesures objectives décrites ci-dessus avec des données issues d'une tâche AX de jugement de la similarité. Cette tâche permet d'obtenir une évaluation holistique de chaque phrase imitée (X) comparée à son modèle (A). 15 auditeurs naïfs ont pris part à la tâche. Tous étaient de L1 française (âge : 25-32) et ne présentaient pas de trouble de la parole ou de l'audition.

Il était demandé aux locuteurs de noter la ressemblance de X avec A en termes de musicalité de la parole (rythme, variation tonale). Le test AX a été passé au moyen du logiciel Lancelot (environnement HTML de PERCEVAL (André, Ghio, Cavé, & Teston, 2003)). Les paires de phrases étaient randomisées par le logiciel, et présentées en modalité auditive dans des écouteurs de qualité professionnelle. Les auditeurs pouvaient écouter chaque couple de phrase jusqu'à cinq fois avant de leur attribuer une note sur une échelle Likert allant de 1 (moins similaire) à 5 (très similaire) en touchant l'écran tactile de l'ordinateur de passation.

Pour chaque phrase X, la moyenne des 15 scores obtenus a été calculée. La phrase obtenant la moyenne la plus élevée a obtenu le rang 1 du classement AX, etc. En cas d'égalité de score entre deux phrases X ou plus, un rang égal à la moyenne des rangs qu'elles devraient occuper dans le classement a été attribué à chacune d'entre elles (par exemple : 7 ; 8 ; 9 → 8 ; 8 ; 8).

Résultats

Nous décrirons tout d'abord les résultats comparant les classements obtenus à partir des scores objectifs (L_2 & *Zr*) et du score perceptif (AX) pour tous les locuteurs. La Figure 1 montre la distribution des rangs pour les trois types de scores, qui donneront par la suite lieu à un calcul de corrélation. Sur ce diagramme, le point indique le score moyen obtenu par le locuteur et les boîtes montrent la distribution interquartile des rangs de chacune des 18 phrases prononcées par un sujet, parmi l'ensemble de 72 phrases évaluées. Les barres de confiance montrent les valeurs minimales maximales de rangs obtenus par chaque sujet, en fonction du classement. Nous rappelons que le rang 1 montre la production qui a obtenu le meilleur jugement de similarité.

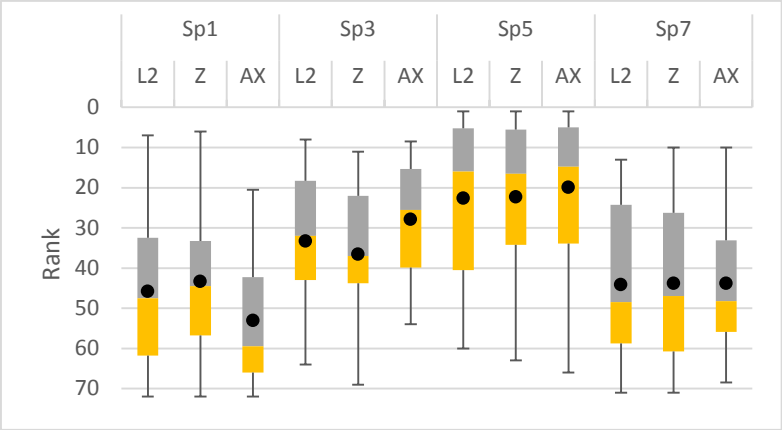


Figure 1 : Distribution des rangs (de 1 à 72, en ordonnées) obtenus par chaque sujet (Sp1, Sp3 ; Sp5 & Sp7) en fonction des types de scores (L_2, *Zr* & AX). Les points représentent le rang moyen de chaque sujet pour les 18 énoncés imités.

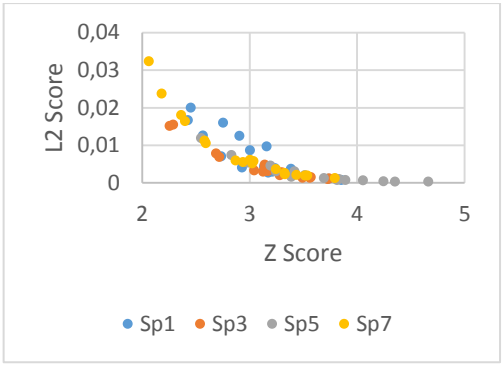
A la vue de cette distribution des rangs obtenus par chaque sujet, il semble raisonnable de classer Sp5 comme le sujet le plus performant dans les tâches d'imitation (Rang moyens : L_2 = 22,76 ; *Zr* = 22,28 ; AX = 19,88), suivi clairement par Sp3 (L_2 = 33,33 ; *Zr* = 36,56 ; AX = 27,92). Les rangs des classements objectifs de Sp7 (L_2 = 44,17 ; *Zr* = 43,83) et de Sp1 (L_2 = 45,83 ; *Zr* = 43,33) sont très proches, mais leur rangs du classement AX (Respectivement : AX(Sp7) = 43,82 ; AX(Sp1 = 53,06) peut refléter la plus grande dispersion de leurs rangs dans le quartile inférieur : le meilleur rang obtenu par Sp1 est meilleur que celui de Sp7, mais cette valeur isolée peut fausser le calcul de la moyenne de leurs rangs. De fait, Sp7 a obtenu un plus grand nombre de bons rangs que Sp1 au fil des procédures d'évaluation, comme le résume le diagramme.

Le calcul de la corrélation a été fait au moyen de Real Statistics pour Excel (Zaiontz, 2015). Afin de comparer les données objectives et subjectives, nous proposons d'utiliser le r_s de Spearman, qui est un indice de corrélation entre des données *ordonnées* (rangs des classements). Les tests bilatéraux de Spearman menés sur nos données montrent des corrélations positives :

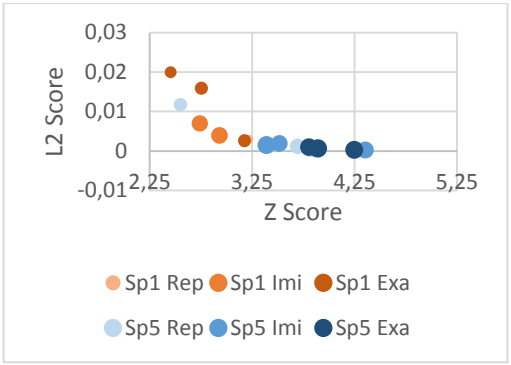
- *Zr* vs. AX → r_s = .554, p < .0001, t (71) = 5.562
- L_2 vs. AX → r_s = .589, p < .0001, t (71) = 6,092

Les deux indices de corrélation dépassent les valeurs critiques de r_s admises pour N = 72 (*rs-crit* = .382, *t-crit* = 3.43). Ainsi, la relation linéaire entre les classements objectifs et perceptifs semble robuste.

La Figure 2A montre la relation entre L_2 et *Zr* pour les 72 phrases testées. Les points dans le coin en bas à droite sont les phrases qui ont été jugées très similaires à leur modèle. Les résultats donnés par l'algorithme soulignent la différence de performance en imitation entre les différents sujets. La Figure 2B illustre la différence de performance entre les deux locuteurs qui ont été respectivement classés le moins (Sp1) et le plus (Sp5) performants au fil des tâches, d'après les résultats de l'algorithme et du panel d'auditeurs.



2A



2B

Figure 2 (A & B) : La Figure 2A montre les scores L_2 (axe *y*) & Zr (axe *x*) de 71 phrases. La Figure 2B illustre en trois dimensions l'évaluation de la performance de 2 locuteurs imitant la phrase « Les bagatelles et les balivernes sottes » au cours de REP, IMI & EXA. Les scores Zr et L_2 sont respectivement sur les axes des *x* et des *y*, la taille des points représente le score moyen au test AX. Un point d'exception a été retiré de ces représentations (Sp1 REP L2=xx, Zr=xx, AX =xx) afin d'améliorer la lisibilité de ces représentations.

Nous prévoyions que les imitations les plus conscientes (réalisées durant les tâches *IMI* & *EXA*) présenteraient une reproduction plus précise des indices prosodiques. Si cette prédiction se retrouve dans les résultats de Sp5 dont la performance, de mieux en mieux notée au fil des tâches, semble montrer un bon contrôle prosodique ; notre prévision se trouve mise à mal par la performance de Sp1 dont certaines phrases produites en *REP* obtiennent de meilleures notes qu'en *IMI* ou en *EXA*. Ainsi, nos résultats mettent en relief une grande variabilité interlocuteur en termes de contrôle prosodique au cours des tâches.

Discussion & Conclusion

Cette étude visait à tester méthodologiquement une approche double d'évaluation/mesure de l'imitation des formes prosodiques. A terme, nous souhaitons développer un algorithme suffisamment robuste pour être implémenté dans un outil d'entraînement des enseignants qui leur permettrait d'évaluer la précision de leur performance prosodique.

Précédemment, le DTW, dont le terme *d* indique le coût d'alignement entre deux contours, a été utilisé par Kim (2012) en tant que mesure de la convergence phonétique. Dans notre cadre, le DTW est principalement utilisé comme méthode d'interpolation (Rilliard et al., 2011) préliminaire au calcul de deux mesures de la similarité prosodique (initialement rapportées par Hermes, 1998). D'après ce dernier, L_2 mesure la distance perceptive entre deux contours, où une distance élevée a un facteur de poids quadratiquement plus élevé. *Zr* exprime une distance globale entre deux formes de contours, *i.e.* sa valeur indique le coût de la transformation d'un contour en un autre. Etant données leurs natures différentes, il était intéressant de les corréler tous deux aux résultats des évaluations perceptives.

Ces deux méthodes de mesure de la similarité prosodique ont obtenu une bonne corrélation avec les résultats du test AX conduits auprès de 15 auditeurs naïfs : les bonnes et les mauvaises imitations ont été repérées par l'algorithme comme par les auditeurs. La différence de corrélation entre AX et *L2/Zr* pourrait refléter la nature différente des mesures objectives. Cela étant dit, ces résultats ouvrent des perspectives intéressantes dans l'évaluation automatique de l'imitation au niveau prosodique. La question des mesures automatiques reste cependant ouverte : une étude sur plus de sujets et plus de stimuli pourrait nous renseigner sur la nécessité de conserver L_2 et/ou *Zr*. De même, de plus amples investigations seront nécessaires à la détermination d'une valeur seuil pour L_2/*Zr*, au-delà de laquelle le résultat d'un test perceptif de jugement de la similarité prosodique pourrait être estimé avec une précision suffisante.

Ces premiers résultats, encourageants, nous poussent à étendre cette approche dans de multiples directions : (1) plus d'énoncés du *CI* seront notés, objectivement et subjectivement ; (2) un corpus de phrases et de leur reproduction délexicalisée par des locuteurs sera constitué afin de tester la robustesse des mesures.

Par ailleurs, il pourrait être pertinent d'envisager l'utilisation d'autres méthodes de transformation que le DTW. Ce dernier requiert un grand nombre de valeurs de *f0*, conduisant à un temps de calcul relativement long. Parmi les méthodes de comparaison de formes passées en revue par (Veltkamp, 2001), la fonction de cumulation des angles semble pouvoir être appliquée à des contours de *f0* stylisés à partir de quelques points remarquables. Cette approche permettrait de prendre en compte explicitement certains détails des patrons prosodiques, comme les durées et les pentes et ainsi améliorer la description de ce qu'est une imitation réussie. Enfin, le raffinement de ces mesures, devrait à terme nous conduire à limiter drastiquement le recours aux tests perceptifs, en sélectionnant les mesures objectives corrélant au mieux avec des résultats perceptifs extensifs.

Les tâches accomplies pour recueillir le *CI* visaient à souligner la capacité d'imitation de locuteurs naïfs, soit, un comportement approchant celui d'un enseignant de langue étrangère sans contrôle prosodique particulier, lors de tentatives de correction phonétique. Pour certains locuteurs (par exemple, Sp1), les scores objectifs fournissent une aide au diagnostic du niveau de contrôle et de conscience prosodique, et dans une certaine mesure, du talent phonétique. Lewandowski (2012) rapporte en effet la complexité à déterminer le talent phonétique d'une personne. Ce type de mesures automatique pourrait donc être utilisé comme un indice pour détecter une composante du talent phonétique, plus précisément au niveau prosodique.

Finalement, nos perspectives de recherches se concentreront sur certains aspects spécifiques de l'entraînement à la MVT, plus particulièrement, sur la correction des indices prosodiques reproduits. A court terme, nous espérons construire une interface permettant à l'enseignant d'améliorer leur compétence d'utilisation des procédés de correction de la MVT, dans ce cas, la production de phrases délexicalisées servant à porter l'attention de l'apprenant sur la syllabification et le rythme de la langue en cours d'apprentissage.

Références

ANDRE, C., GHIO, A., CAVE, C., & TESTON, B. (2003). PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. In *Proceedings of XVth ICPhS* (p. 1421−1424). Barcelone, Espagne.

ASTÉSANO, C., BARD, E. G., & TURK, A. (2007). Structural influences on Initial Accent placement in French. *Language and speech*, *50*(3), 423−446.

BILLIERES, M., ALAZARD, C., ASTESANO, C., & NOCAUDIE, O. (2013). Phonétique corrective en FLE : Méthode Verbo-Tonale. http://w3.uohprod.univ-tlse2.fr/UOH-PHONETIQUE-FLE/

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5:9/10*, 341−345.

DONALD, M. (1993). *Origins of the Modern Mind - Three Stages in the Evolution of Culture & Cognition* (Reprint). Cambridge, Mass.: Harvard University Press.

DUFOUR, S., & NGUYEN, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology*, *4*.

GOLDINGER, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251−279.

HERMES, D. J. (1998). Measuring the Perceptual Similarity of Pitch Contours. *Journal of Speech, Language and Hearing Research*, *41*, 73−82.

KIM, M. (2012). *Phonetic accommodation after auditory exposure to native and nonnative speech*. NORTHWESTERN UNIVERSITY.

LEWANDOWSKI, N. (2012). *Talent in non-native phonetic convergence*. Universität Stuttgart, Stuttgart.

MICHELAS, A., & NGUYEN, N. (2011). Uncovering the Effect of Imitation on Tonal Patterns of French Accentual Phrases. In *INTERSPEECH* (p. 973–976).

MIXDORFF, H., COLE, J., & SHATTUCK-HUFNAGEL, S. (2012). Prosodic Similarity–Evidence from an Imitation Study. In *Speech Prosody 2012*.

NOCAUDIE, O., & ASTÉSANO, C. (2012). Prosodic structuring imitation in French L1 context-A first step towards correcting phonetic-prosodic features in L2 French. In *Proceedings of ISICS*. Aix-en-Provence.

PARDO, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382−93.

PARDO, J. S. (2013). Reconciling diverse findings in studies of phonetic convergence. In *Proceedings of Meetings on Acoustics* (Vol. 19, p. 060140). Acoustical Society of America.

REVIS, J., DE LOOZE, C., & GIOVANNI, A. (2013). Vocal Flexibility and Prosodic Strategies in a Professional Impersonator. *Journal of Voice*, *27*(4), 524.e23−524.e31.

RILLIARD, A., ALLAUZEN, A., & DE MAREÜIL, P. B. (2011). Using Dynamic Time Warping to Compute Prosodic Similarity Measures. In *INTERSPEECH* (p. 2021–2024).

VELTKAMP, R. C. (2001). Shape matching: similarity measures and algorithms (p. 188−197). IEEE Computer Society.

ZAIONTZ, C. (2015). *Real Statistics Using Excel*. Consulté à l'adresse www.real-statistics.com

Evaluating prosodic similarity as a means towards L2 teacher's prosodic control training

Olivier Nocaudie, Corine Astésano

U.R.I Octogone-Lordat (E.A. 4156), Université de Toulouse, UTM, Toulouse, France

nocaudie@univ-tlse2.fr, corine.astesano@univ-tlse2.fr

Abstract

Studies on professional impersonators and naïve speakers has underlined that speech imitation proficiency varies across speakers. Imitation in speech supposes that a speaker succeeds in reproducing specific features of the perceived speech. Because of the inherent variability of human speech behaviors, the question lies open whether different speakers can accurately imitate phonetic features, and more specifically prosodic patterns. This exploratory study proposes to test *f*0 contours' imitation of 4 sentences originally pronounced by a female speaker, by 4 naïve listeners undertaking 3 different tasks: mere repetition, imitation and exaggeration of the perceived sentences., Two tests were performed: imitated sentences and models were time-warped and objective comparisons were performed using two (dis)similarity measures reported in the literature; a panel of 15 listeners evaluated perceptually the same set of sentences during an AX similarity judgment task. Similarity scores were used to build multiple rankings in order to observe the correlation between the two tests' rankings and to evaluate prosodic imitation proficiency across speakers/listeners. This research has implication for L2 phonetic correction using the Verbo-Tonal Method, which requires excellent prosodic awareness and control by the teacher in the production of lexicalized and delexicalized sentences.

Index Terms: speech imitation, prosodic patterns, time-warping methods, perception of prosodic similarity.

Introduction

Studies on speech imitation report different types of human imitation behavior such as convergence (mutual adaptation during the course of the interaction) [1], voice disguise (attempt of impersonating someone else's voice) [2], [3], or mere imitation (simple mimicry [4], shadowing [5], [6]). On the one hand, these behaviors may be defined as different in so far as they depend on factors like contexts of production [7] or imitator's intention [8]. On the other hand, they share a major common trait, namely their qualification as imitative speech behavior: the speaker's production must sound similar to its model, whatever imitation characteristics are used. Thus, speech imitation's studies aim to elicit and observe behavioral shifts in the way speakers talk, may it be at a lexical or a phonetical level. For the latter, uncovering what feature in the signal is being imitated and how it is being assessed often remains a methodological puzzle. Indeed, it is delicate to choose what acoustical features to measure and to link to the results of perceptual tests [9].

Professional impersonators tend to use global adjustment to the voice's target specificities and are also able to imitate instant variations (synchrony strategies) like intonation contours or duration of pauses. Naïve speakers, however, seem to be limited to convergence strategies (global adjustment to the voice) [2] [3]. This observation raises some interconnected questions related to synchrony strategies: (1) To what extent can a naïve speaker reproduce a perceived prosodic pattern (instant variations); (2) How can we assess their success or failure in doing so; (3) Is it possible to train a speaker to reproduce intonation, and more generally any prosodic feature, more accurately?

Questions (1) & (3) have a specific relevance in the domain of teaching pronunciation to L2 speakers, more particularly in the framework of the Verbo-Tonal Method (hereafter VTM). VTM postulates that errors of pronunciation in L2 are due to a L1 bias in the perception of the L2. To neutralize the effect of this bias, VTM proposes to exercise the speaker's ear using a wide array of correction processes where prosody has a crucial role. A teacher using VTM must have specific prosodic awareness and control, more particularly when (s)he is required to delexicalize (or logatomize) a sentence in order to facilitate the perception of the rhythmic and intonational features of the target language, by drawing the learners' attention on these prosodic features.

Per se, live phonetic correction performance represents a typical imitative interaction. Indeed, during VTM interaction, both the teacher and the trainee have to imitate or reiterate some speech features. The trainee (L2 learner) has to repeat the teacher's linguistic model, which leads one to question the link between speech perception & (re)production in bilingual learners. The teacher has to produce coherent phonological and prosodic patterns consistently, which raises the question of production control, more specifically at a prosodic level.

If questions (1) and (3) may apply to both the teacher and the trainee, the present study focuses on the teacher's aptitude to consistently reproduce prosodic patterns. Indeed, before addressing learners' ability to imitate/(re)produce linguistic features, one has to make sure that the *imitee* (L1 teacher) is actually able to consistently reproduce (hence, imitate) his/her own speech. As mentioned before, phonetic correction in the VTM framework implies repetition of prosodic features in a consistent way; it also implies that the teacher is able to emphasize some prosodic realizations to facilitate learners' perception of the target features. We therefore propose to first test L1 speakers' ability to control their imitation of prosodic features. In doing so, we address question (2), *i.e.* assessing for speakers' success or failure at imitating prosodic features. Ultimately, the methods used to assess prosodic (dis)similarity is intended to evaluate teachers' prosodic control and be used as a tool for their training.

Few studies have tackled the issue of speech imitation in French, and more specifically on prosodic cues' imitation (see however [10] for Initial Accent reproduction). The present

study is following up on our previous preliminary study describing speakers' ability to imitate prosodic features of controlled sentences on an 'imitation scale' going from a simple repetition to an exaggerated mimicry [11].

Linguistic material: An imitation corpus

The corpus originally consists of syntactically ambiguous sentences that can be disambiguated via prosodic cues. Syntactic ambiguity derives from the manipulation of the adjective scope on two coordinated nouns, as in "les gants et les bas lisses" (*the smooth gloves and stockings*), where the adjective (A) "lisses" either qualifies the second noun "bas" only ([les gants][et les bas lisses]; Low Adjective attachment hereafter *Low*), or either the two nouns "gants et bas" ([les gants et les bas][lisses]; High Adjective attachment, hereafter *High*). Sentences vary in terms of Noun and Adjective lengths, from one to four syllables. Manipulating syntactic ambiguity and constituents' lengths allows us to uncover the prosodic cues (prominences, boundary tones, pauses …) used for syntactic linearization of spoken utterances. For more details on this corpus, see [12].

A subset of 16 sentences spoken by a female speaker was selected for our imitation tasks. These sentences consisted of two Noun lengths (tri- and quadri-syllable nouns) combined with one- to four-syllable lengths of Adjective, in the two syntactic readings conditions (*Low* and *High*). 8 native listeners/imitators of French were instructed to speak out sentences in three different tasks performed in separate blocks: a) a mere repetition *(Rep*); b) an imitation (*Imi*); and c) an exaggerated imitation (*Exa*) of the speaker's sentences. 2 speakers were discarded for voice quality problems or experiment-induced stress. In each block, listeners/imitators repeated each sentence 3 times, in a random order, giving rise to a total of 864 sentences (16 sentences * 2 syntactic conditions * 3 repetitions * 3 tasks * 6 speakers). In order to evaluate the implicit ability to imitate speech, the attention of the speakers was not drawn to imitation in task a); they were instructed to just "*say the sentence while preserving the intended structure*". In tasks b) and c), they were explicitly asked to imitate and to exaggerate the sentences. Their attention was however not drawn to prosodic features.

In the present exploratory study comparing objective and subjective data, we chose to select a subset of 4 sentences from this corpus according to two criteria chosen to evaluate the robustness of the algorithm used to test prosodic similarity: 1) the sentences were all taken from the *Low* attachment syntactic condition because syntactic disambiguation is marked by a silent pause between the first and the second noun. The presence of acoustic silence is of particular interest to test for robustness insofar as the algorithm is overly biased towards silence alignment when evaluating prosodic similarity; 2) the sentences were chosen to illustrate two different phrase lengths. We also chose to run the present tests on 4 listeners/imitators only (Sp1, Sp3, Sp5 and Sp7), who were paired to imitate the following sentences:

- Sp1 (female) & Sp5 (female)
    - Les bagatelles et les balivernes sottes
    - Les bonimenteurs et les baratineurs fades
- Sp3 (male) & Sp7 (female)
    - Les bagatelles et les balivernes saugrenues
    - Les bonimenteurs et les baratineurs fabuleux

Altogether, our results will be computed on 18 sentences by subject, yielding a total number of 72 sentences ([2 sentences * 3 repetitions * 3 tasks] * 4 subjects).

Method: Objective measurements & perceptual evaluations of prosodic imitation

Section 3 describes our methodology for evaluating imitated *f0* contours (dis)similarity with our speaker's model. It also presents the perceptual evaluation task that was undertaken for comparison with the objective measurements.

One problem raised by the assessment of imitation in speech lies in the absence of congruence between perceptual judgments of imitation and the multitude of acoustic features either converging with or diverging from the model [7], [9].

Pitch, and its physical correlate *f0* is reported to be the main feature targeted by imitators [3]. It is also the primary cue used for corrective feedback during VTM correction. Our method will thus focus on the measurement of the physical distance between pairs of *f0* contours on the one hand, and on the perceptive evaluation of their resemblance on the other hand.

Dynamic Time-Warping (DTW) & (dis)similarity measures

Assessing imitation of *f0* contours objectively amounts to find if there is a physical distance between these contours, *i.e.* to answer the question of the shapes' matching of the contours.

Shape matching however supposes tonal normalization and temporal alignment of *f0* peaks and valleys (DTW). The distance between two *f0* contours was computed through two measures similar to the method proposed by Hermes [13] where $w(t)$ is the temporal course of the weighting factor (*i.e.* the sum of the reference signal's subharmonic sumspectrum), W its time integral from 0 to T (T being the duration of the utterance), f_1 and f_2 the tested pitch contours of sentence pairs.

We however chose to use a different normalization procedure than that of Hermes, and divided each *f0* values by the maximum *f0* of the utterance ($f_1=p_1/p_1max$). This normalization procedure allows for comparing male and female listeners/imitators by bringing *f0* variations on a comparable scale from zero to one, relative to speakers' mean *f0*. This will later help peaks' and valleys' comparisons using the DTW algorithm's comparison. The sampling rate was one *f0* values per millisecond [14] extracted with Praat [15].

After normalization, the root mean square difference (L_2) between two contours was computed as follows:

$$L_2 = \left\{\frac{1}{W}\int_0^T w(t)\left|f_1(t) - f_2(t)\right|^2 dt\right\}^{1/2} \quad (1)$$

A correlation coefficient (r) between the two contours f_1 and f_2 was then computed as follows:

$$r = \frac{\frac{1}{W}\int_0^T w(t) f_1(t) f_2(t) dt}{\sqrt{\left\{\frac{1}{W}\int_0^T w(t)\left|f_1(t)\right|^2 dt \frac{1}{W}\int_0^T w(t)|f_2(t)|^2 dt\right\}}} \quad (2)$$

Hermes [13] however reports that r needs to be transformed in Fischer's Z (hereafter Zr) to allow for correlation's comparison:

$$Z_{f_1 f_2} = \frac{1}{2} ln \frac{1+r_{f_1 f_2}}{1-r_{f_1 f_2}} \quad (3)$$

L_2 measures rapid changes in the f0 contour while Zr is a holistic measure of contour shapes.

Before computing L_2 and Zr for each pair of *f0* contours, Dynamic Time Warping was performed on the tested contours to force the alignment between the model and its reproduction (non-linear *f0* interpolation). It has been reported that such an alignment would overall improve the correlation, especially

when the contours are functionally similar, *i.e.* when they share the same accentual pattern [14].

Finally, each sentence was ranked relatively to the others, depending on their L_2 and *Zr* scores:

- L_2 is a **dissimilarity measure** (the higher the L_2, the higher the dissimilarity). The sentence with the lowest L_2, was ranked 1 while the one with the highest L_2 was ranked 72.
- *Zr* is a **similarity measure** (the higher the *Z*, the higher the similarity). The sentence with the highest *Z* was ranked 1, the second highest was ranked as 2, and so on.

Two objective rankings were thus obtained, which will be compared to the ranking derived from the perceptual evaluation's results (see 4.3.2).

AX similarity judgment test

As argued by [9], imitation in speech should be assessed both objectively and subjectively, *i.e.* physically and perceptively. To this end, we complemented the objective measurements described above with an AX similarity judgment task, which allows for an absolute rating of each reiterated sentence (X) compared to the model (A). 15 naive listeners participated to the AX judgment task. All were French native speakers (age 25-32) and did not report any hearing or speech disorder.

Listeners were instructed to rate the resemblance of X with A in terms of the 'musical' features of speech (rhythm, tonal variations). The task was run on a computer using the Lancelot software (HTML environment of PERCEVAL [16]). Sentences were randomized by the software and auditorily presented using high quality headphones. Listeners could hear each pair of sentences up to five times before giving their rating on a scale from 1 (less similar) to 5 (perfect match) by clicking on the corresponding button with the computer mouse.

Results were computed by calculating the mean score of each X sentence. In case of ties, we attributed to groups of tied sentences a rank equal to the mean of their consecutive original ranking.

Results

Distribution of objective and perceptual rankings

We first describe the results comparing the rankings obtained from objective scores (L_2 and *Zr*) and the perceptual scores (AX) across speakers. Figure 1 shows the distribution of the 3 different scores, which will give rise to the calculation of correlation coefficients. Box plots show the global mean ranking (dots) and the interquartile distribution of the ranks by subject on 18 sentences (note here that the 3 different imitation tasks are merged for now). Whiskers indicate minimum and maximum ranking values. Lowest mean rankings indicate better judgment in *f0* contour comparisons.

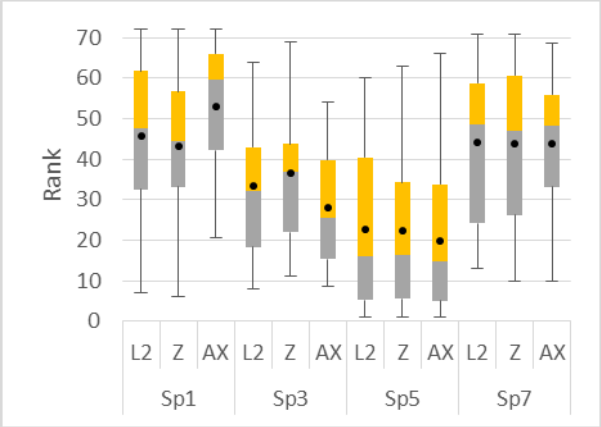


Figure 1: *Distribution of* L_2, Zr *and AX ranks per subject (Sp1, Sp3, SP5, Sp7), classified from 1 to 72 sentences (y axis). Dots represent the mean ranks over the 18 sentences produced by each speaker.*

Given this distribution of ranks across the subjects, it seems that Sp5 can be classified as the most proficient subject (Mean rank L_2 = 22,76, *Zr* = 22,28 and AX = 19,88) followed clearly by Sp3 (L_2 = 33,33, *Zr* = 36,56, AX = 27,92). Objective rankings of Sp7 (L_2 = 44,17, *Zr* = 43,83) and Sp1 (L_2 = 45,83, *Zr* = 43,33) are close to each other, but their perceptual rankings (respectively (AX(Sp7) = 43,82 ; AX(Sp1) = 53,06) may reflect the dispersion of their ranks in the inferior quartile: Sp1's best rank is greater than Sp7's, but it may act as an outlier for the computation of their mean score. Overall, Sp7 obtained a greater amount of good ranks than Sp1 during every evaluation process task, as shown in the box plot.

According to Hermès [13], L_2 measures the perceptual distance between two contours, where quadratically more weight is given to larger distance. *Zr* expresses the distance between the contours' shape, *i.e.* to what extent can a pitch contour be obtained from another by performing a linear transformation. Given their different nature, it is of interest to correlate them both with perceptual evaluation results in order to later determine a threshold on L_2 and/or *Zr* beyond which we could estimate fairly accurately the results of perceptual judgement of prosodic similarity. Figure 2 show the correlation between L_2 and *Zr* scores for the 72 sentences. The points in the lower right corner represent sentences estimated as highly similar with the model.

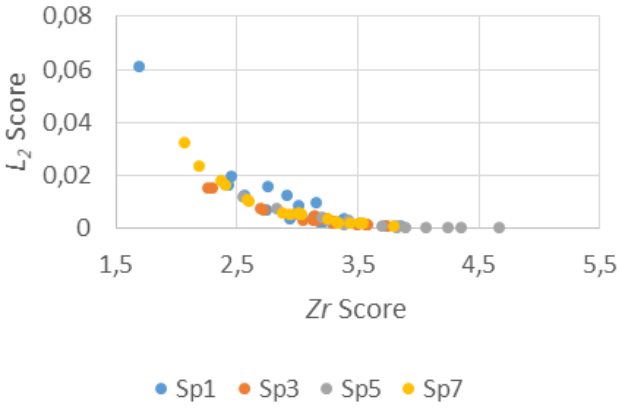


Figure 2: L_2 *(RMS difference) and* Zr *scores (Fisher transform of r) of the 72 sentences.*

Correlation between Zr, L_2 and AX rankings

The correlation was computed using *Real Statistics* for Excel [17]. A test of correlation between objective and subjective measures based on *ordered* data (ranks) is possible when using the r_s of Spearman, which allows for the comparison between different rankings. Pairwise two-tailed tests show fairly good correlation of *Zr* ranking with AX ranking (r_s = .554, p < .0001, t (71) = 5.562), while the correlation between L_2 and AX rankings were slightly stronger (r_s = .589, p < .0001, t (71) = 6.092). Both correlation values are indeed above r_s's critical value for N = 72 (r_s-*crit* = .382; *t-crit* = 3.43). The linear relationship between objective and perceptive rankings thus seems pretty robust.

Imitation tasks and performance

Results given by the algorithm underline the difference of imitation proficiency across speakers/listeners. Figure 3 illustrates proficiency differences between the two paired speakers which respectively are the less (Sp1) and the most (Sp5) proficient in the tasks, as rated both by the algorithm and the panel of listeners. We predicted that the more conscious imitations (tasks *IMI* and *EXA*) would be produced as most prosodically accurate. However, both objective and perceptual results indicate great imitation performance variation across speakers. Whereas Sp5 seemingly shows a better control with increasing performance throughout the three tasks, some of Sp1's *REP* sentences exhibit better rating than other sentences produced during the *EXA* task. Note that bigger dots in the lower right corner indicate perceptively better rated imitations

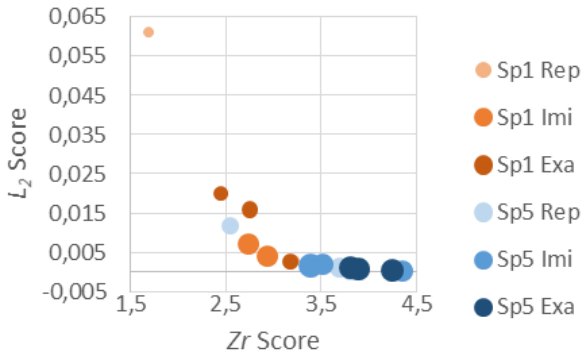


Figure 3: *Illustration of speakers' proficiency in performing the 3 imitation tasks for 9 repetitions of sentence 'Les bagatelles et les balivernes sottes'.* L_2 *and* Zr *scores are on the y and x axes; AX judgments' mean scores are represented by bullets' size.*

Discussion

This preliminary study was intended to methodologically test the validity of comparing objective and perceptual evaluations of prosodic similarity in imitation. Our ultimate goal is to find a sufficiently robust algorithm method, which could be implemented as an automatic tool evaluating teachers' proficiency at imitating prosodic contours.

Originally, the DTW was used by [18] as a measure of convergence in speech, expressed as *d*, the output of DTW giving the cost of alignment between two contours. With our aim in mind, we chose here to use the DTW as an interpolation method (as proposed by [14]) to compute two measures of prosodic similarity (initially reported by [13]): the first measure (L_2) models rapid perceptual processes, while the second measure (*Zr*) models holistic perceptual processes related to contours' shapes.

Both measures correlated well with the perceptual test performed on 15 listeners for this exploratory study: bad and good imitations were consistently spotted by the algorithm too. The difference of correlation between L_2 and *Zr* may reflect the nature of both measures, as discussed earlier. That being said, these results encourage us to elaborate on the automatic investigation of imitation but the question lies open whether both L_2 and *Zr* measures are to be kept in a near future to continue our experiments. In other words, studies on a larger database are necessary to set a threshold on L_2 and/or *Zr* beyond which the results of perceptual judgement are accurately enough estimated and to evaluate potential discrepancies between these two factors of similarity.

As our first results were encouraging, it is planned to expand this approach in multiple directions:

- More sentences from the imitation corpus will be rated, both objectively and subjectively.
- A new corpus consisting of sentences and their delexicalized reproductions by human speakers will be constituted, in order to further test these types of measures.

Besides, it may be of interest to test a method of shape matching involving a different transformation than DTW, which requires thousands of *f0* values, and quite a long computing cost. Among the methods of shape matching reviewed by [19], the cumulative angle function could be applied to *f0* contours, stylized with the help of much fewer sampling points. It could lead to refine prosodic patterns analysis (slopes and timing), which might be a satisfactory substitution to the *Zr* measure. Ultimately, it is intended to limit the use of perceptual tests in the assessment of prosodic imitation in speech, by selecting the factors of objective similarity best correlating with extensive perceptual results.

The tasks performed to gather the corpus intended to underline the capacity of naïve speakers to imitate, in the same way a prosodically unaware teacher could do when trying to correct phonetics of L2 learners. For some speaker (as Sp1) the algorithm may help diagnose if they exhibit or not prosodic awareness and control, and to some extent, talent. As underlined by [20], talent, as an individual factor is complex to assess. This type of objective approach could be used to detect that part of the talent of individuals resorting to prosodic ability.

Finally, our perspective will be to focus on specific training of VTM, more precisely, on the correctness of prosodic cues reproduction. Ideally, our research should lead to build a user-interface allowing teachers to train specific VTM processes, in this case, delexicalization used to help focus on syllabification and rhythm.

Acknowledgements

This study is supported by the Agence Nationale de la Recherche grant ANR-12-BSH2-0001 (PI: Corine Astésano)

We would like to thank Albert Rilliard, LIMSI, CNRS, France, for his advices on the topic of prosodic similarity comparison; and Benjamin Boulbène & Julien Dupouy, France for their involvement with implementing the algorithm.

References

[1] J. S. Pardo, « On phonetic convergence during conversational interaction. », *J. Acoust. Soc. Am.*, vol. 119, n° 4, p. 2382−93, 2006.

[2] E. Zetterholm, « A comparative survey of phonetic features of two impersonators », in *Fonetik*, 2002, vol. 44, p. 129–132.

[3] J. Revis, C. De Looze, et A. Giovanni, « Vocal Flexibility and Prosodic Strategies in a Professional Impersonator », *J. Voice*, vol. 27, n° 4, p. 524.e23−524.e31, juill. 2013.

[4] H. Mixdorff, J. Cole, et S. Shattuck-Hufnagel, « Prosodic Similarity–Evidence from an Imitation Study », in *Speech Prosody 2012*, 2012.

[5] S. D. Goldinger, « Echoes of echoes? An episodic theory of lexical access. », *Psychol. Rev.*, vol. 105, n° 2, p. 251−279, 1998.

[6] S. Dufour et N. Nguyen, « How much imitation is there in a shadowing task? », *Front. Psychol.*, vol. 4, 2013.

[7] N. Lewandowski, « Talent in non-native phonetic convergence », Universität Stuttgart, Stuttgart, 2012.

[8] M. Donald, *Origins of the Modern Mind - Three Stages in the Evolution of Culture & Cognition*, Reprint. Cambridge, Mass.: Harvard University Press, 1993.

[9] J. Pardo, « Reconciling diverse findings in studies of phonetic convergence », in *Proceedings of Meetings on Acoustics*, 2013, vol. 19, p. 060140.

[10] A. Michelas et N. Nguyen, « Uncovering the Effect of Imitation on Tonal Patterns of French Accentual Phrases. », in *INTERSPEECH*, 2011, p. 973–976.

[11] O. Nocaudie et C. Astésano, « Prosodic structuring imitation in French L1 context-A first step towards correcting phonetic-prosodic features in L2 French », in *Proceedings of ISICS*, Aix-en-Provence, 2012.

[12] C. Astésano, E. G. Bard, et A. Turk, « Structural influences on Initial Accent placement in French. », *Lang. Speech*, vol. 50, n° 3, p. 423−446, 2007.

[13] D. J. Hermes, « Measuring the Perceptual Similarity of Pitch Contours », *J. Speech Lang. Hear. Res.*, vol. 41, p. 73−82, 1998.

[14] A. Rilliard, A. Allauzen, et P. B. de Mareüil, « Using Dynamic Time Warping to Compute Prosodic Similarity Measures. », in *INTERSPEECH*, 2011, p. 2021–2024.

[15] P. Boersma, « Praat, a system for doing phonetics by computer. », *Glot Int.*, vol. 5:9/10, p. 341−345, 2001.

[16] C. André, A. Ghio, C. Cavé, et B. Teston, « PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception », in *Proceedings of XVth ICPhS*, Barcelone, Espagne, 2003, p. 1421−1424.

[17] C. Zaiontz, *Real Statistics Using Excel*. 2015.

[18] M. Kim, « Phonetic accommodation after auditory exposure to native and nonnative speech », NORTHWESTERN UNIVERSITY, 2012.

[19] R. C. Veltkamp, « Shape matching: similarity measures and algorithms », 2001, p. 188−197.

[20] M. Jilka, H. Baumotte, N. Lewandowski, S. Reiterer, et G. Rota, « Introducing a comprehensive approach to assessing pronunciation talent », *Proc. 16th ICPhS Saarbr.*, p. 1737–1740, 2007.