

Modelage perceptuel du contrôle de la matière phonique en L2 : Variabilité interindividuelle

Véronique Delvaux* #, Kathy Huet*, Myriam Piccaluga*, Bernard Harmegnies*

*Laboratoire des Sciences de la Parole, Académie Wallonie-Bruxelles, Place du Parc 20, 7000 MONS, Belgique.

Fonds National de la Recherche Scientifique.

Parole-academie@umh.ac.be

ABSTRACT

This paper reports on an exploratory study of the processes involved in the acquisition of new phonetic control regimes in L2 learning. The focus is on the acquisition of long VOT initial stops by French-L1 English-L2 beginners. Francophones were asked to repeat /ta/ stimuli varying in VOT duration and burst intensity. Three experiments were carried out in order to compare the objective properties of these speech productions, as well as their subjective properties as indexed by typicality and similarity judgements from French and American English listeners.

Keywords: L2 learning, VOT, burst, interindividual variability, production, typicality, similarity.

1. INTRODUCTION

L'idée que, dans le cadre de l'apprentissage de la L2, le sujet recourt mal à propos aux stratégies perceptuelles et productives qui lui sont habituelles en L1 a depuis longtemps inspiré linguistes et pédagogues [4, 6]. Plus récemment, les scientifiques intéressés à la cognition humaine [e.g. 5], ont fait l'hypothèse que les difficultés de l'apprenant en L2 sont liées à la mise en œuvre inappropriée de stratégies routinisées par le biais de l'usage de la L1. Les pédagogues de l'oral en langue étrangère, cherchant à prendre en charge cette propension de l'apprenant, ont proposé divers moyens d'intervention dans le cadre de démarches qu'il est de coutume de rassembler sous l'appellation *correction phonétique*. De cela résulte un corps *technique* de *procédés didactiques* appuyés essentiellement sur l'expertise des praticiens [4]. Les enseignants de langue s'accordent en général sur l'intérêt de ces techniques. Néanmoins, l'étude objective non seulement de leur efficacité mais aussi, plus profondément, de leur mode de fonctionnement, n'a fait l'objet que de fort peu de développements [1, 3].

2. OBJECTIFS EXPÉRIMENTAUX

Le projet de recherche auquel ressortit la présente contribution concerne la manipulation expérimentale de l'input perceptuel auquel est exposé l'apprenant en L2. Nous étudions les effets provoqués par cette seule manipulation sur les compétences de traitement de la matière phonique de l'apprenant (tant en réception qu'en émission). La question traitée est donc celle des

mécanismes d'un possible *shaping* du comportement phonique du sujet via la présentation de stimuli auditifs manipulés expérimentalement compte tenu des liens entre L1 et L2. Ce type d'intervention expérimentale se rapproche en fait du procédé didactique dit de la *prononciation nuancée*, par laquelle le formateur est réputé modifier le modèle dans une 'direction opposée' à celle de l'erreur de l'apprenant [4, 6].

Afin de procéder à une simulation en laboratoire des mécanismes de base impliqués dans ce procédé, nous nous centrons ici sur un aspect ponctuel de la maîtrise de l'oral, à savoir le contrôle du voisement dans la production des occlusives pour la paire de langues *français-anglais*. Nous cherchons plus particulièrement à faire apparaître, chez des sujets francophones à faible maîtrise de l'anglais, des perceptions et des productions de séquences *occlusive-voyelle* comparables à celles des anglophones. Nous nous intéressons ainsi aux déterminants de l'acquisition du contrôle d'un contraste qui, du point de vue phonologique, est présent dans les deux langues, mais du point de vue phonétique, s'y réalise différemment. Il n'y a donc pas ici, pour le sujet, à construire de nouvelles catégories phonologiques, mais bien à mettre en œuvre de nouvelles stratégies de contrôle susceptibles d'incarner autrement des catégories existantes. Dans cette étude, nous construisons notre corpus sur la base du logatome formé de l'occlusive dentale et de la voyelle /a/, réalisé en isolation.

3. DISPOSITIFS EXPÉRIMENTAUX

Trois expériences sont présentées ici. Pour l'expérience 1, on a soumis des sujets francophones à une tâche de production lors de laquelle ils devaient répéter des sons de parole. L'expérience 2 a été menée auprès de sujets anglophones à qui il a été demandé de juger de la typicalité des échantillons sonores recueillis à la faveur de l'expérience 1. L'expérience 3 a été menée sur les sujets des deux groupes, qui devaient juger, cette fois, de la similarité entre les stimuli-modèles et les productions-réponses de l'expérience 1. Les expériences 2 et 3 permettent donc de compléter l'étude des 'propriétés objectives' des productions des francophones (expérience 1) par une analyse des 'propriétés subjectives' de ces productions, tant du point de vue des auditeurs anglophones que francophones.

Deux groupes de sujets volontaires (14 francophones de Belgique natifs dont la connaissance et la pratique de l'anglais est minimale et 17 anglophones américains natifs, débutants en langue française), tous sélectionnés sur base d'un questionnaire de biographie linguistique, ont participé aux expériences. Vu le caractère exploratoire de la présente communication, nous nous centrerons, pour l'analyse des résultats, sur les productions de 4 des sujets francophones S4, S6, S8, S10 (choisis parce qu'ils présentent des comportements très contrastés), le corpus de résultats en perception impliquant, lui, tous les sujets à titre d'auditeurs.

3.1. Dispositif pour l'expérience 1

Les stimuli utilisés ici trouvent leur origine dans un segment de signal issu de la réalisation de /ta/ par une anglophone américaine native. La production se compose d'un *burst* de 20 ms, suivi d'une phase d'aspiration de 40 ms (soit un VOT de 60 ms), et enfin d'une voyelle de 210 ms (débutant par une attaque *breathy* de 30 ms). De cette production naturelle, nous avons conservé la durée du *burst*, l'intensité de l'aspiration et la voyelle dans son ensemble. Seules ont été manipulées la durée de la phase d'aspiration (et donc le VOT) et l'intensité globale du *burst*. L'ensemble de ces manipulations nous ont permis de construire 36 stimuli se différenciant par 4 intensités de *burst* (*burst* écologique multiplié en amplitude instantanée par un facteur 4, 2, 1 et 0,5) et par 9 VOT (de 20 ms - caractéristique d'une sourde non-aspirée de L1 - à 100 ms, par pas de 10 ms). Chaque fichier son contenant un stimulus a été construit de manière à présenter une durée de 1500 ms. Seuls les sujets francophones étaient ici sollicités. Leur tâche consistait, à partir de l'audition d'un modèle (l'un des 36 stimuli) à le répéter « le plus fidèlement possible, comme si il s'agissait d'un mot d'une langue étrangère à prononcer le mieux possible ». Pour chaque amplitude de *burst*, trois présentations des stimuli se succédaient : les stimuli étaient présentés soit dans l'ordre croissant sur le continuum de durée du VOT, soit dans l'ordre décroissant, soit enfin dans un ordre aléatoire. Afin d'éviter les effets de contamination, chacune de ces phases de présentation était séparée de la suivante par la présentation de six stimuli « neutres » (le premier stimulus de la série en question). De manière à obtenir plusieurs productions d'un même modèle dans les mêmes conditions de présentation, cette tâche a été répétée trois fois pour chaque sujet. Au total, chacun a produit 540 [ta]. Le corpus a été segmenté de façon semi-automatique, et le VOT ainsi que la durée de la voyelle /a/ ont été mesurés dans chaque production.

3.2. Dispositif pour l'expérience 2

Pour cette expérience, qui a été menée uniquement auprès des 17 auditeurs anglophones, les productions de l'expérience 1 ont été utilisées à titre de stimuli. Parmi celles-ci, nous avons retenu un sous-ensemble

correspondant au stimulus à VOT de 60 ms pour chacune des 4 intensités du *burst* et dans chacun des 3 types de présentation (montante, descendante et aléatoire), soit 12 productions par locuteur francophone, séparées par deux stimuli « neutres » à chaque changement de locuteur. Chaque auditeur anglophone a ainsi été confronté à 196 stimuli. La tâche consistait en un jugement de typicalité, exprimé par chaque auditeur, pour chaque stimulus, en situant un curseur sur une échelle continue, l'auditeur ayant à se prononcer sur le caractère plus ou moins « native-like » de la production entendue (de 0 : « pas du tout » à 100 : « tout à fait » par pas de 1). Un indice (pourcentage moyen) de typicalité a ainsi été obtenu pour chaque sujet francophone dans divers sous-ensembles de stimuli en fonction de l'intensité du *burst*.

3.3. Dispositif pour l'expérience 3

L'expérience 3 a été menée auprès des 17 sujets anglophones et – du fait de la mortalité expérimentale – d'un sous-groupe de l'échantillon francophone (5 sujets). Les sujets étaient confrontés à des paires de stimuli, qu'ils devaient évaluer en termes de leur similarité. Le premier était le modèle présenté à un sujet francophone lors de l'expérience 1 et le second, la production en résultant. Les auditeurs se prononçaient en positionnant un curseur sur une échelle continue (de 0 à 100 par pas de 1) variant de « - » (pas du tout) à « + » (tout à fait) similaire. Un indice (pourcentage moyen) de similarité, pour chaque série de stimuli, a été calculé pour chaque catégorie d'auditeurs (francophones ou anglophones) sur chaque sujet francophone.

4. ANALYSE DES RESULTATS

4.1. Propriétés objectives des productions

Les résultats de l'expérience 1 montrent que les locuteurs ont effectué la tâche de (re)production avec un certain succès: le VOT produit augmente à mesure que le VOT entendu augmente. Cependant, l'éventail des durées des productions est réduit par rapport à celui des stimuli. De plus, on constate une importante variabilité inter-individuelle dans les productions. En particulier, S4 s'est révélé incapable de modifier le VOT de ses productions. Pour les trois autres sujets, le VOT produit augmente globalement avec le VOT entendu, mais seules les valeurs mesurées pour S8 correspondent aux valeurs du stimulus d'entrée; S10 tend à produire des VOT systématiquement plus longs que ceux entendus (en moyenne 27 ms), tandis que les VOT produits par S6 sont d'un empan réduit (entre 45 et 85 ms) par rapport aux stimuli (entre 20 et 100 ms).

On observe également une importante variabilité inter-individuelle dans la façon dont les VOT produits varient en fonction de l'intensité du *burst* du modèle (Fig.1 gauche: exemples pour S8 et S10). Pour S10, le

VOT produit n'évolue pas du tout en fonction du VOT du modèle lorsque celui-ci présente un burst très intense. Plus l'intensité du burst diminue, meilleure est la relation entre VOT produit et VOT entendu. Dans une large mesure, l'intensité du burst a l'effet opposé pour S8. Ces résultats sont à mettre en relation avec les autres propriétés articulatoires et acoustiques des /ta/ produits par ces deux sujets, telles que la durée de la voyelle /a/ (Fig.1 droite). Ainsi, pour S8 la voyelle produite est d'autant plus longue que le burst du modèle est peu intense. A l'inverse, pour S10 la durée de la voyelle varie peu en fonction du VOT entendu lorsque le burst est relativement peu intense, alors qu'elle s'accroît avec la durée du VOT entendu lorsque le burst est (très) intense. S8 privilégie donc un accroissement de la durée du /a/ au détriment d'un allongement du VOT en réponse aux stimuli avec un burst peu intense, tandis que S10 adopte le comportement inverse. De même, S8 fait peu varier la durée de la voyelle et beaucoup plus celle de la consonne (en termes de VOT) en reproduisant des stimuli à burst intense, alors que S10 fait exactement le contraire. Dans le cas de S6, le VOT produit et la durée de la voyelle covarient positivement; la durée totale du /ta/ produit est sensiblement plus longue lorsque le modèle proposé présente un burst très peu intense.

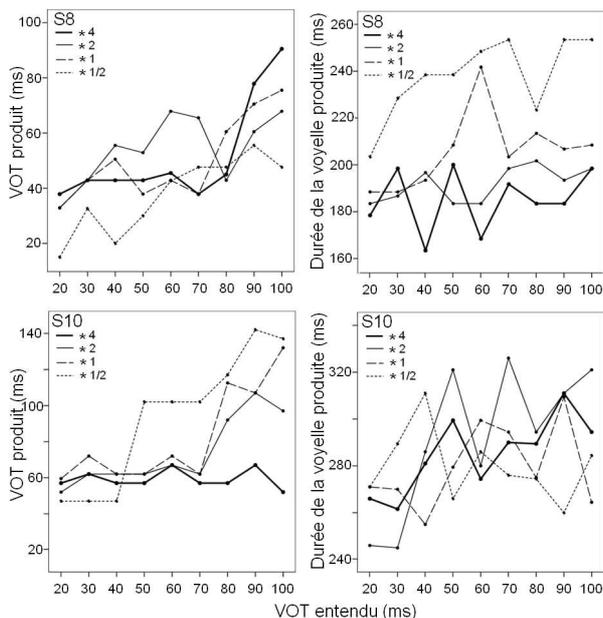


Figure 1. Expérience 1. VOT produit et durée de la voyelle produite pour S8 et S10 en fonction du VOT entendu et de l'intensité du burst.

4.2. Propriétés subjectives des productions

Les scores de typicalité et de similarité obtenus lors des expériences 2 et 3 varient en fonction des auditeurs concernés. Le score moyen attribué aux productions jugées (sur l'échelle de 0 à 100) varie selon les auditeurs entre 34 et 76 pour l'expérience 2 et 25 et 67 pour l'expérience 3. Globalement, un tiers des

auditeurs (francophones et anglophones confondus) tend à répondre de façon bipolaire, c'est-à-dire à attribuer à une majorité des productions un score de 0 ou 100. Une moitié des auditeurs distribue ses réponses le long de l'échelle de façon globalement équivalente, tandis qu'une minorité d'auditeurs privilégie les scores moyens. On notera également que dans l'ensemble, les anglophones tendent à répondre de façon plus 'contrastée' (vis-à-vis des différents locuteurs, mais pas vis-à-vis des différentes intensités de bursts, cf. fig.2) que les francophones, ce qui peut être attribué au fait que l'évaluation s'opère en référence à des modèles phoniques explicites (expérience 3) ou implicites (expérience 2) qui ressortissent à des catégories phonologiques en anglais mais pas en français.

L'analyse des résultats montre également la variabilité des réponses des auditeurs en fonction du locuteur (Fig.2). Le locuteur S4 est jugé peu performant (score moyen < 45), que ce soit en termes de typicalité par rapport à l'anglais ou en termes de similarité par rapport au modèle, ce qui s'explique aisément au vu des mesures de production. Les locuteurs S8 et S10 obtiennent des scores de typicalité et de similarité proches de 50 (sauf pour des bursts de très faible intensité), alors que leur VOT produit est proche de 50 ms. Quant à S6, il est particulièrement bien évalué par les auditeurs anglophones (≈ 75), mais pas par les francophones (≈ 50), et ce spécialement lorsque l'intensité du burst est égale ou supérieure à l'intensité écologique (< 20).

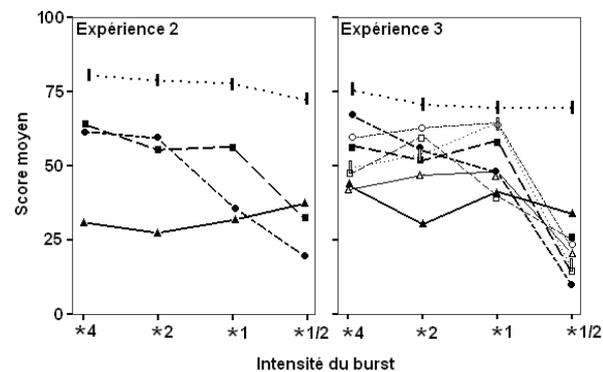


Figure 2. Scores moyens en fonction de l'intensité du burst (triangle: S4, rectangle: S6, cercle: S8, carré: S10; symboles pleins/vides: auditeurs angl./franc.).

5. DISCUSSION

Comme l'a montré l'expérience 1, la manipulation du modèle s'avère un mode d'action efficace pour le modelage du comportement d'émission du locuteur de L1 à qui l'on tente de faire émettre des sons de parole d'une manière proche de celle qui caractérise les productions des natifs de L2. Un tel résultat tend à conforter le procédé didactique de prononciation nuancée, qui consiste notamment à présenter aux apprenants des modèles à reproduire déviant par rapport aux productions ciblées. On observe cependant

une importante variabilité des réactions individuelles à ce genre de procédé. Le cas de S4 est, de ce point de vue, particulièrement intéressant. Ce sujet, qui s'avère incapable de modifier le VOT de ses productions, est en effet pourtant celui dont une autre expérience [2], non relatée ici, avait montré qu'il était le plus performant dans une tâche de discrimination des VOT. En supposant que ce locuteur s'est véritablement impliqué dans la tâche de production (ce que nous ne pouvons qu'escompter), on peut attribuer son manque de succès à un déficit de contrôle de la production de la matière phonique, mais également éventuellement à un déficit perceptuel lié au fait que dans l'expérience 1 l'attention de l'auditeur n'est pas explicitement attirée sur les propriétés acoustiques différentielles d'une paire de stimuli, mais bien sur une similarité globale recherchée entre un modèle entendu et une production à réaliser. Au total, une façon d'interpréter les résultats concernant les aspects temporels des productions est de poser que les sujets sont sensibles à l'accroissement (par rapport à leur langue maternelle) du VOT des stimuli auxquels ils sont exposés, et qu'ils y réagissent en effectuant des modifications temporelles dans leurs propres productions, soit en augmentant la durée de la voyelle, soit en produisant un VOT effectivement plus long (ce qui requiert un contrôle différent de la synchronisation entre articulateurs laryngaux et supra-laryngaux), soit encore en combinant les deux stratégies. L'augmentation de la durée de la voyelle peut, d'un certain point de vue, être considérée comme un pis aller au regard du modèle proposé (puisque la voyelle y est de durée constante), même si certaines études ont montré que la production et la perception du trait de voisement pour les occlusives dépend en fait de multiples indices (p.ex. la fréquence et la pente de F1 à l'attaque de la voyelle, le débit de parole, l'amplitude du burst et du bruit d'aspiration, la durée de la voyelle, voir [e.g. 7]). Ces observations sont néanmoins à combiner avec celles concernant l'effet sur la production du VOT de l'intensité du burst présent dans le stimulus. À l'évidence, certains sujets y sont sensibles alors que d'autres ne semblent pas l'être, et les observations rapportées suggèrent de profondes interactions entre sujet, durée du VOT et intensité du burst qui, dans certains cas, semble accroître la saillance perceptuelle de la durée allongée du VOT.

La confrontation des productions des sujets avec les évaluations des auditeurs montre par ailleurs que, hormis les comportements objectifs dont rendent compte les mesures acoustiques, les manipulations des stimuli ont également des répercussions sur les propriétés subjectives des productions. Ainsi, les scores de typicalité et similarité très faibles obtenus par S8 et S10 pour les productions liées à un modèle à burst fortement diminué témoignent du fait que les stratégies de production consistant à faire varier le rapport entre durée de la voyelle et durée de VOT ne sont pas bien accueillies par les auditeurs. D'ailleurs, S6 n'y recourt pas et est particulièrement bien évalué, du moins par les

auditeurs anglophones. La différence de jugement de similarité entre anglophones et francophones dans le cas de S6 (bursts d'intensité égale ou supérieure à l'écologique) s'explique soit par le fait que les anglophones réagissent plus positivement que les francophones à une production très proche de la valeur canonique de 60 ms pour le VOT, soit par le fait que d'autres aspects de la production (ici non mesurés) sont récompensés par les anglophones (attaque breathy, finale creaky de la voyelle, etc.). Dans le cas des modèles à burst peu intense, S6 a réalisé des /ta/ avec un burst relativement intense, ce qui a été jugé inadéquat par les francophones, mais pas par les anglophones. En admettant que les anglophones ont bien fait le départ entre les tâches de similarité et de typicalité, ce résultat suggère que l'effet de l'intensité du burst sur le VOT perçu diffère en anglais et en français, et appelle à de nouvelles expériences perceptuelles sur la relation entre burst et délai d'attaque du voisement en fonction de la langue maternelle de l'auditeur.

En conclusion, nos observations soulignent non seulement la sensibilité des locuteurs au modelage perceptuel, mais surtout la complexité et la variabilité des processus imbriqués dans ce type de tâche. À l'évidence, cette recherche exploratoire confirme l'intérêt d'investigations systématiques à tous les niveaux (contrôle du stimulus, perception de celui-ci par un sujet, production par ce dernier, perception de la production par des tiers natifs ou non) de la chaîne production-perception.

BIBLIOGRAPHIE

- [1] Billières, M., Didactique des langues et phonétique ; la place du verbo-tonal, *RPA* 114, 1995, 43-64.
- [2] Delvaux, V., Huet, K., Piccaluga, M. & Harmegnies, B., Pour une approche expérimentale de l'acquisition du contrôle de la matière phonique en FLE. *Actes Didcog 2007*, 1-6.
- [3] Harmegnies, B., Delvaux, V., Huet, K. & Piccaluga, M., Oralité et cognition : pour une approche raisonnée de la pédagogie du traitement de la matière phonique, *RPA* 37-39, 2005, 277-348.
- [4] Intravaia, P., *Formation des professeurs de langue en phonétique corrective ; le système verbo-tonal*, Didier érudition, CIPA, Paris, Mons, 2000.
- [5] Kuhl, P.K., Linguistic Experience and the "Perceptual Magnet Effect" in Strange, W. (éd.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Y Press, Baltimore, 1995, 121-154.
- [6] Renard, R. (éd.), *Apprentissage d'une langue étrangère/seconde, 2. La phonétique verbo-tonale*, De Boeck Université, 2002.
- [7] Repp, B., Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants, *Lang. Speech* 22, 1979, 173-189.