Identification perceptive d'accents étrangers en français

Bianca Vieru-Dimulescu & Philippe Boula de Mareüil

LIMSI-CNRS
BP 133 – 91403 Orsay CEDEX, France
Tél.: ++33 (0)1 69 85 80 70 – Fax : ++33 (0)1 69 85 80 88
Courriel : bianca@limsi.fr ; mareuil@limsi.fr

ABSTRACT

A perceptual experiment was designed to determine to what extent naïve French listeners are able to identify foreign accents in French: Arabic, English, German, Italian, Portuguese and Spanish. They succeed in recognising the speaker's mother tongue in more than 50% of cases (while rating their degree of accentedness as average). They perform best with Arabic speakers and worst with Portuguese speakers. The Spanish and Italian accents on the one hand and the English and German accents on the other hand are the most mistaken ones. Phonetic analyses were conducted; clustering and scaling techniques were applied to the results, and were related to the listeners' reactions that were recorded during the test. Emphasis was laid on differences in the vowel realisation (especially concerning the phoneme /y/).

1. INTRODUCTION

Peut-on reconnaître l'origine d'une personne à partir d'un échantillon de parole ? Dans certains cas cela est possible : des expériences récentes l'ont confirmé sur des accents régionaux en anglais [1] et en français [2]. On sait moins dans quelle mesure des auditeurs naïfs se montrent capables d'identifier des accents étrangers en français. Le but de cet article est de combler ce manque. Des études préliminaires [3] [4] ont suggéré que le timbre des voyelles permet d'identifier l'origine d'un accent en français mieux que la prosodie (du moins le rythme). De même, la plupart des études en la matière se concentrent sur le segmental (*i.e*. la chaîne de phonèmes, en particulier vocaliques) [5] [6]. C'est cet aspect que nous développons ici.

Dans la perspective de mener des tests perceptifs, nous avons porté notre intérêt sur les accents avec lesquels des Français sont le plus susceptibles d'être familiers : allemand, anglais, arabe, espagnol, italien, portugais (plus néerlandais, uniquement conservé pour illustrer quelle réaction on peut avoir face à un accent étranger). Le choix de ces langues a été établi en recoupant des statistiques sur l'immigration et le tourisme en France. Nous avons enregistré une quarantaine de locuteurs de ces différentes langues maternelles, à la fois en lecture et en parole spontanée.

La section suivante présente le protocole et les résultats d'une expérience d'identification perceptive des 6 accents étudiés. Et la section 3, avant de conclure, examine certains traits phonétiques relevés par les auditeurs, en matière de timbre vocalique. L'analyse acoustique s'appuie sur de la lecture, des éléments de comparaison avec la parole spontanée étant fournis.

2. EXPÉRIENCE PERCEPTIVE

2.1. Locuteurs et corpus

Parmi les locuteurs d'origine allemande, anglaise, arabe, espagnole, italienne ou portugaise dont les enregistrements ont été collectés, 36 ont été utilisés pour l'expérience proprement dite (6 par langue), 6 autres (1 de chaque langue) et une locutrice néerlandaise pour une phase de familiarisation. Dans le test, on avait autant d'hommes que de femmes. Les locuteurs étaient européens ou venaient, pour les Arabes, de différents pays du Maghreb — mais une étude antérieure a montré la difficulté à discriminer les différentes origines possibles, algérienne, marocaine ou tunisienne, de locuteurs parlant français [3]. Aucun hispanophone n'était catalan ni latino-américain. L'âge moyen des locuteurs (tous étudiants) était de 24 ans, leur durée de résidence en France (dans la région parisienne) était en moyenne de 15 mois et l'âge auquel ils avaient commencé à apprendre le français était en moyenne de 17 ans.

Les locuteurs ont lu, entre autres, le texte du projet « Phonologie du Français Contemporain » (PFC) [7], et parlé librement pendant quelque 5 minutes en situation de face à face avec l'expérimentateur. Des éléments de comparaison nous sont ainsi disponibles, à travers les données du projet PFC dont nous avons repris et étendu le protocole — les locuteurs étaient également enregistrés dans leur langue maternelle.

Pour le test perceptif, un échantillon de parole spontanée, d'une dizaine de secondes, a été retenu pour tous nos locuteurs, selon les critères suivants : absence de référence culturelle et d'erreur morpho-syntaxique typique d'une origine donnée, pas trop d'hésitations et cohérence thématique de l'énoncé. Les stimuli sélectionnés ont ainsi fait l'objet d'une expérience perceptive, décrite dans ce qui suit.

2.2. Protocole et tâche des sujets

Le test se déroulait dans une chambre isolée, les auditeurs écoutaient les stimuli à travers des enceintes avec un niveau d'écoute confortable — préalablement égalisé à l'aide du logiciel Goldwave (http://www.goldwave.com). Les stimuli, au format Wave, étaient échantillonnés à 22,05 kHz, 16 bits, mono. Le test était réalisé à travers une interface conviviale [4] [8], qui permet entre autres choses d'entrer des informations sur la familiarité avec les différents accents et de saisir les réponses.

Les sujets étaient avertis qu'ils seraient amenés à porter des jugements sur des échantillons de parole non native

en français. Ils écoutaient pour commencer un stimulus et les commentaires qu'une collègue avait faits sur la prononciation qui se trouvait être celle d'un accent néerlandais. Ensuite, en guise de familiarisation avec les accents étudiés, les sujets écoutaient un extrait de parole spontanée illustrant chacun des accents allemand, anglais, arabe, espagnol, italien et portugais. À chaque fois étaient indiqués l'origine et le degré d'accent évalué par les auteurs : de très faible à très fort. Une bonne image de la diversité des accents était ainsi fournie par ces locuteurs qui, bien sûr, n'étaient pas utilisés par la suite.

Lors de la phase suivante, le test proprement dit, les auditeurs écoutaient 36 stimuli présentés dans un ordre aléatoire différent pour chaque auditeur. Comme lors de la phase de familiarisation, chaque stimulus pouvait être réécouté, arrêté au milieu ou repris à partir d'un certain point ; mais il était impossible de revenir en arrière une fois passé au stimulus suivant. Les auditeurs devaient attribuer un degré d'accent au stimulus en question, sur une échelle continue graduée de 0 à 5. Les degrés proposés étaient paraphrasés de la façon suivante : (0) pas d'accent, (1) petit accent, (2) accent modéré, (3) assez fort accent, (4) fort accent, (5) très fort accent.

Les auditeurs, munis d'un microphone, étaient invités à réagir verbalement à l'écoute du stimulus (en l'imitant voire en le caricaturant) ou à écrire leurs commentaires dans une fenêtre de texte. Ces données étaient enregistrées stimulus par stimulus, et les consignes suggéraient simplement de préciser quels traits non natifs dans la prononciation et l'intonation du locuteur leur semblaient marquants. Les sujets devaient déterminer la langue maternelle du locuteur ayant produit le stimulus, avant de traiter un autre extrait. Le choix était forcé (sans distracteur ni classe rejet) parmi allemand, anglais, arabe, espagnol, italien et portugais. En cas d'hésitation entre deux réponses, les sujets avaient droit à une seconde chance : une option leur était facultativement laissée à cet effet. Mais elle a été très peu utilisée (seulement 30 fois pour l'ensemble des auditeurs), nous ne l'avons donc pas analysée.

2.3. Auditeurs

Le test perceptif a été soumis à 25 auditeurs de région parisienne, de langue maternelle française, sans problèmes d'audition connus et membres d'un laboratoire d'informatique (le LIMSI). Ils n'étaient pas rémunérés pour cette tâche. La table 1 reporte, entre parenthèses, le nombre d'auditeurs qui se disaient avant le test capables de reconnaître tel ou tel accent en français.

Table 1 : nombre d'auditeurs s'estimant capables de reconnaître tel ou tel accent en français (entre parenthèses) et rapportant un niveau faible, moyen ou bon dans les langues correspondantes.

Niveau	allemand (18)	anglais (20)	arabe (20)	espagnol (12)	italien (3)	portugais (6)
faible	13	0	23	14	24	24
moyen	9	5	2	10	1	0
bon	3	20	0	1	0	1

Ces chiffres ne vont pas de pair avec le niveau des auditeurs dans les langues correspondantes, également consigné : par exemple, 23 sujets sur 25 déclaraient qu'ils n'avaient pas ou que peu de connaissances en arabe, alors que 20 d'entre eux se sentaient capables de reconnaître un accent arabe en français.

2.3. Résultats

Des jugements des auditeurs, il ressort que le degré d'accent de nos locuteurs est moyen (2,66 sur 5) : 2,18 pour les Allemands ; 3,00 pour les Anglais ; 2,37 pour les Arabes ; 2,94 pour les Espagnols ; 3,07 pour les Italiens ; 2,39 pour les Portugais. Les résultats du test d'identification montrent que nos auditeurs arrivent bien à distinguer les accents étrangers présentés (*cf.* table 2). Le taux global d'identification correcte (52,2 %) est très supérieur au hasard (16,7 %). Des tests de khi-deux révèlent que pour chaque langue on est significativement au-dessus du seuil de hasard [$ddl = 3$; $p < 0{,}01$]. Et pour chaque origine linguistique la réponse majoritaire est la bonne — en gras, dans la diagonale de la table 2. Il en va de même pour 27 locuteurs sur les 36 présentés, qui sont bien identifiés.

Table 2 : matrice de confusion pour 36 stimuli et 25 auditeurs (% par rapport à 6×25 = 150 réponses).

origine \ réponse	allemand	anglais	arabe	espagnol	italien	portugais
allemand	**63,3**	14,7	6,0	3,3	4,7	8,0
anglais	28,0	**48,7**	8,7	8,7	2,7	3,3
arabe	6,0	1,3	**77,3**	2,0	5,3	8,0
espagnol	2,7	2,7	5,3	**58,7**	19,3	11,3
italien	5,3	3,3	7,3	34,0	**40,0**	10,0
portugais	17,3	8,0	16,7	20,7	12,0	**25,3**

Les locuteurs dont l'origine a le mieux été reconnue sont les Arabes, principale communauté immigrée vivant en France. Les moins bien reconnus sont les Portugais, dont les phonèmes sont « proches » du français, ce qui peut expliquer leur accent peu marqué. En outre, le stéréotype chuintant qui est souvent associé à l'accent portugais est loin de la réalité.

Il est intéressant de souligner que le degré d'accent de nos locuteurs portugais (2,39) est supérieur à celui des arabes (2,37), même si cette différence n'est pas significative d'après un test de Student. Entre ces cas extrêmes, les confusions les plus fréquentes sont dans l'ordre italien/espagnol — ce qui est corroboré par des études antérieures sur l'accent espagnol en italien et l'accent italien en espagnol [8] — et anglais/allemand. Des techniques d'échelonnement multidimensionnel (*scaling*) et de *clustering* permettent de visualiser ce fait : elles rassemblent Italiens et Espagnols dans un même groupe, Anglais et Allemands dans un autre groupe, et fait apparaître Arabes et Portugais dans deux groupes à part (*cf.* figure 1, qui montre une sorte de distance perceptive entre les différents accents).

Des analyses de variance (ANOVA) ont également été conduites sur les réponses comptées comme correctes (1) ou fausses (0) avec le facteur aléatoire Sujet et les deux facteurs intra-sujet Familiarité (avec l'accent) et Degré

d'accent. Selon que les auditeurs se sont majoritairement déclarés capables de reconnaître l'accent en français (comme dans le cas des Allemands, Anglais et Arabes) ou non (comme dans le cas des Espagnols, Italiens, Portugais), deux groupes de Familiarité ont été distingués. En ce qui concerne le Degré d'accent, les locuteurs ont été séparés en trois groupes équilibrés, moyennant les évaluations des auditeurs. Les ANOVA montrent un effet majeur de la Familiarité [$F(1,24) = 56,5$; $p < 0,01$] et du Degré d'accent des locuteurs [$F(2,48) = 21,4$; $p < 0,01$]. On a également une interaction entre les deux [$F(2,48) = 4,1$; $p = 0,02$].

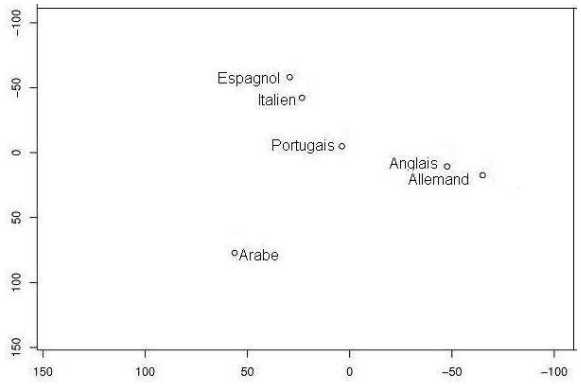

Figure 1 : résultat de l'échelonnement multidimensionnel — algorithme classique prenant en entrée une matrice de dissimilitude qui peut être calculée à partir des distances (ici euclidiennes) entre lignes de la matrice de confusion.

Les auditeurs ont déclaré s'être appuyés sur des indices dont on peut vérifier la pertinence par des mesures acoustiques. Parmi les indices segmentaux, nos auditeurs ont relevé : le *r* « roulé » qui leur évoquait des pays du Sud ; [i] à la place de /e/ dans le cas des locuteurs de langue maternelle arabe ; *yé* à la place de *je*, [v] à la place de /b/ et [s] à la place de /z/ pour les Espagnols ; [z] à la place de /s/ pour les Allemands ; [u] à la place de /y/ ou l'inverse, ainsi qu'une mauvaise réalisation des nasales, sans rapprochement avec une origine particulière, mais plutôt signe d'un accent étranger en général. D'où les analyses acoustiques suivantes, avec le tracé de triangles vocaliques des locuteurs — les études sur les consonnes et la prosodie sont en cours.

3. ANALYSES PHONÉTIQUES

Les mesures présentées dans cette section ont été faites sur le texte PFC (400 mots), lu par les 36 locuteurs utilisés dans le test de perception. Ce matériau commun, moins restreint que nos échantillons de parole spontanée, facilite l'analyse et se prête mieux aux comparaisons entre locuteurs. Même si en contrepartie, il reflète moins bien le vernaculaire (la façon naturelle de parler), ce texte permet la comparaison directe, sur le même corpus, avec des natifs français.

Grâce à l'alignement automatique dérivé du système de reconnaissance de la parole du LIMSI [9], le texte PFC a été segmenté en phonèmes en utilisant des modèles acoustiques indépendants du contexte pour le français standard. De la même façon ont été segmentés les textes lus par 6 locuteurs normands ou vendéens (3 hommes, 3 femmes) parmi les plus jeunes de ceux qui ont été étudiés dans [2]. Ces locuteurs n'avaient « pas d'accent », ou plus précisément leur degré d'accent avait été estimé à moins de 1 sur 5 par 25 auditeurs de région parisienne suivant un protocole très proche de celui de la présente étude.

Sur cette base ainsi segmentée, des études acoustiques ont été menées, facilitées par le traitement automatique de la parole. Un script a été écrit pour le logiciel PRAAT (http://www.fon.hum.uva.nl/praat/) afin d'extraire les fréquences des formants en différents points des voyelles — le texte PFC en comprend plus de 500 par locuteur. Comme la méthode de segmentation est automatique, des filtres ont été prévus (adaptés à chaque voyelle, hommes et femmes distingués) pour écarter les valeurs aberrantes à un horizon tolérant de ±500 Hz en moyenne par rapport à des valeurs de référence [10]. Seulement 4 % des phonèmes ont ainsi été rejetés. Les valeurs des premiers formants ont ensuite été normalisées à l'aide de diverses procédures décrites par [11]. Utilisant la normalisation de Nearey, les triangles vocaliques correspondant aux différents accents (ou origines linguistiques) sont donnés figure 2, où les valeurs des formants au tiers, à la moitié et aux deux tiers des voyelles sont moyennées. On peut noter que les triangles des locuteurs anglais et allemands sont plus réduits que les autres — ce qu'on peut relier à la réduction vocalique que connaissent leurs langues d'origine. L'antériorisation du /u/ anglais est également remarquable, de même que le fait que parmi les /e/, le plus proche des [I] est celui des Arabes.

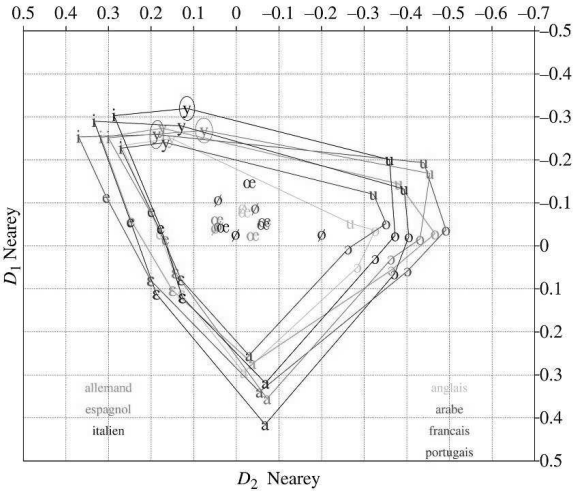

Figure 2 : triangles vocaliques (après filtrage des probables erreurs de mesure et normalisation de Nearey) correspondant au français natif ou parlé avec différents accents, pour le texte PFC. De gauche à droite sont entourés d'ellipses les /y/ des Arabes, des Italiens et des Espagnols.

La réalisation du /y/ français est particulièrement différente entre les locuteurs espagnols ou italiens notamment (chez qui elle est plus proche du [u]) et les locuteurs arabes (chez qui elle tend vers le [i]). Les uns privilégient le trait [+antérieur], les autres le trait [+arrondi]. Ce phénomène souvent caricaturé est connu [12], même si on n'en a pas d'explication . On peut le retrouver dans des transcriptions ludiques telles que *tou*

m'as toué ou bien *Itats-Inis*. Il est bien mis en évidence par scaling ou clustering à partir d'une caractérisation de chaque origine par les coordonnées moyennes de son /y/ dans le plan F1/F2. Une représentation graphique sous forme de dendrogramme en est fournie figure 3. En introduisant le troisième formant (F3), en utilisant un autre algorithme (agglomératif plutôt que divisif) et quelle que soit la distance utilisée (euclidienne ou Manhattan), on obtient des résultats très semblables. On retrouve la même tendance sur les phrases spontanées présentées aux auditeurs, que nous avons transcrites, alignées et analysées comme le texte lu.

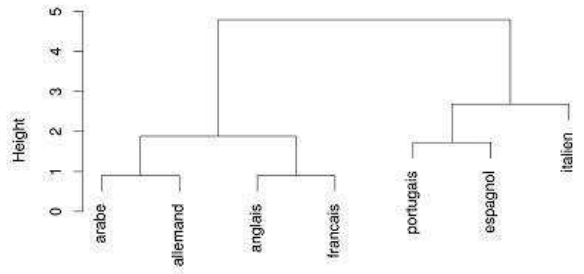

Figure 3 : dendrogramme issu du clustering divisif utilisant une distance euclidienne, à partir d'une caractérisation dans l'espace acoustique de la prononciation du /y/ en français standard et dans chaque accent.

En ne représentant plus les moyennes par langue d'origine mais par locuteur, on obtient par le même algorithme divisif utilisant une distance euclidienne un dendrogramme où les locuteurs arabes d'une part et les locuteurs italiens et espagnols, d'autre part, sont assez bien regroupés : 5 Arabes sur 6 du côté gauche, 5 Italiens et 4 Espagnols sur 12 du côté droit de l'arbre. Ce /y/ pourrait donc être un bon indice discriminant.

4. CONCLUSION

Ainsi, il a été possible de mener une étude à base de perception, d'analyse de données et de traitement automatique de la parole, sur les accents allemand, anglais, arabe, espagnol, italien et portugais en français. Ces accents étrangers sont ceux avec lesquels nous avons le plus de chance d'être exposés. Quelque quarante locuteurs ont été enregistrés, parlant français avec un accent jugé moyen. Leur origine a été bien identifiée par des auditeurs français natifs à partir d'échantillons de parole — à 52 %, soit 10 % de plus que 6 accents régionaux en français, étudiés dans une tâche similaire [2]. Ces auditeurs ont indiqué de façon opportune les indices acoustiques qui leur paraissaient saillants. Pour hiérarchiser ces derniers, nous poursuivons nos mesures de VOT (*Voice Onset Time*) pour les consonnes occlusives, et devons lier les paramètres rythmiques [4] avec le ratio de durée des syllabes accentués / non accentuées. Il restera à comparer les résultats avec les productions en langue maternelle de nos locuteurs.

Plusieurs perspectives s'ouvrent pour le traitement automatique : ajout de variantes liées aux accents étrangers (ex. /y/ → [u], avec les modèles acoustiques français) ; alignements avec les modèles acoustiques des langues d'origine (ex. [ɹ] anglais), en vue d'une identification automatique des accents et d'une amélioration des scores de reconnaissance de la parole non native. La synthèse de la parole mériterait également d'être mise à profit : mieux que des imitateurs humains qui ont trop tendance à renforcer certains traits, cet outil est un bon instrument de simulation. Enfin, nous espérons que cette étude pourra être utile pour l'apprentissage et l'enseignement du français langue étrangère.

5. REMERCIEMENTS

Nous sommes reconnaissants à Martine Adda-Decker, Cécile Woehrling et Cédric Gendrot pour la mise à disposition, le lancement ou l'adaptation de scripts notamment exploitant l'alignement automatique.

6. BIBLIOGRAPHIE

[1] C.G. Clopper & D.B. Pisoni. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, 32: 111-140, 2004.

[2] C. Woehrling & P. Boula de Mareüil. Identification d'accents régionaux en français : perception et catégorisation. *Bulletin PFC*, 6 : 89-103, 2005.

[3] P. Boula de Mareüil, B. Brahimi & C. Gendrot, Role of segmental and suprasegmental cues in the perception of Magrebian-accented French. *Interspeech-ICSLP*, Jeju, 2004.

[4] B. Vieru-Dimulescu, P. Boula de Mareüil & M. Adda-Decker. Identification perceptive d'accents étrangers en français : premiers résultats, *10es Journée PFC*, 2006.

[5] B. Lauret. *Aspects de Phonétique Expérimentale Contrastive : « l'accent» anglo-américain en français.* Thèse de doctorat, Université Paris III, 1998.

[6] J.E. Flege, C. Schirru & I.R.A. MacKay, Interaction between the native and second language phonetic subsystems, *Speech Communication*, 40(4) : 467-491, 2003.

[7] E. Delais-Roussarie & J. Durand (éd.), *Corpus et variation en phonologie du français. Méthodes et analyses.* Presses Universitaires de Mirail, Toulouse, 2003.

[8] B. Vieru-Dimulescu & P. Boula de Mareuil, Contribution of prosody to the perception of a foreign accent: A study based on Spanish/Italian modified speech. *ISCA Workshop on Plasticity in Speech Perception*, London, pages 66-69, 2005.

[9] M. Adda-Decker, P. Boula de Mareüil, G. Adda & L. Lamel. Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, 46 (2): 119-139, 2005.

[10] C. Gendrot & M. Adda-Decker, Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Interspeech*, Lisboa, pages 2453-2456, 2005.

[11] P.M. Adank. *Vowel normalization : a perceptual-acoustic study of Dutch vowels.* Thèse de doctorat, Radboud University Nijmegen, 2003.

[12] B.L. Rochet, *Perception and Production of Second-Language Speech Sounds by Adults, in W. Strange* (ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues, York Press, York, pages 379-410, 1995.