Caractériser la distinctivité du système vocalique des locuteurs

Christine Meunier [1] & Alain Ghio[1]
(1) Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
christine.meunier@lpl-aix.fr, alain.ghio@lpl-aix.fr

RESUME

L'objectif de notre étude est de caractériser les locuteurs du français grâce à un indice de distinctivité lors de la production de voyelles en parole spontanée. Cette distinctivité est le plus souvent établie selon la dispersion de l'espace vocalique. Des travaux précédents (Huet & Harmegnies, 2000) ont proposé un indice plus dynamique prenant en compte le rapport entre la dispersion de l'ensemble des voyelles du système et la dispersion moyenne de chaque voyelle dans sa catégorie. Nous nous inspirons de ces travaux pour proposer un indice de distinctivité (ID) en vue d'établir des profils de locuteurs. Nos premiers résultats confirment des différences interlocuteurs. L'indice lui-même n'est pas toujours en lien avec la dispersion globale du système et permet de mettre en évidence une interaction plus fine entre voyelle et système. Suite à cette première étape nous envisageons d'évaluer cet ID selon différents facteurs (langue, type de parole, populations pathologiques).

ABSTRACT

The characterization of the distinctivity in speakers' vowel production.
The objective of our study is to characterize the French speakers thanks to a cue of distinctiveness in the production of vowel in spontaneous speech. Distinctiveness is most often derived from the dispersion of vowel space. Previous work (Huet & Harmegnies, 2000) has proposed a more dynamic cue taking into account the relationship between the dispersion of the whole vowels of the system and the average dispersion of each vowel in its category. To go on with this view we propose a cue of distinctiveness (ID) in order to provide speakers' profiles. Our first results confirm differences between speakers. The cue itself is not always related to the overall dispersion of the system but highlights a more precise interaction between the vowel and the system. Following this first step, we plan to evaluate this ID according to different factors (language, type of speech, pathological populations).

MOTS-CLES : distinctivité ; acoustique; voyelles; parole spontanée
KEYWORDS: distinctiveness ; acoustics ; vowels ; spontaneous speech

1 Introduction

L'objectif de notre étude est de caractériser les locuteurs du français au travers d'un indice de distinctivité lors de la production de voyelle en parole spontanée. L'organisation des réalisations vocaliques a depuis longtemps suscité l'intérêt des chercheurs au travers de plusieurs objectifs : la comparaison des langues, les variations dues à la situation de parole ou encore la comparaison de populations (saines/pathologiques, L1/L2). D'un point de vue méthodologique, le système vocalique présente l'avantage de faire apparaître des variations graduelles au sein d'un même mode de

10.21437/JEP.2018-54

production, ce qui n'est pas le cas pour les consonnes. A cet égard, la *centralisation* du système vocalique est apparue comme une conséquence des facteurs énumérés ci-dessus. Notamment (Smiljanić & Bradlow, 2009) précise que l'hyper-articulation des voyelles, associée à une parole « claire », augmente la distance F1-F2 et occasionne moins de chevauchement entre les réalisations de chaque catégorie de voyelle, ce qui les rend plus distinctes. Toutefois, l'hyper-articulation des voyelles peut être reliée à deux dimensions : une augmentation de l'espace vocalique global (du système) et/ou une réduction de l'espace de production de chaque catégorie de voyelle. Huet & Harmegnies (2000) propose un indice *Phi* permettant de de calculer le rapport entre la variation à l'intérieur du système (CM_{inter}) et la variation moyenne de chaque catégorie de voyelle (CM_{intra}). Cet indice a permis aux auteurs de montrer que la parole spontanée induisait un indice bien plus faible que des situations de lecture chez un même locuteur. C'est donc sur la base de cet indice que nous souhaiterions apporter une contribution concernant la comparaison des locuteurs en parole conversationnelle en français.

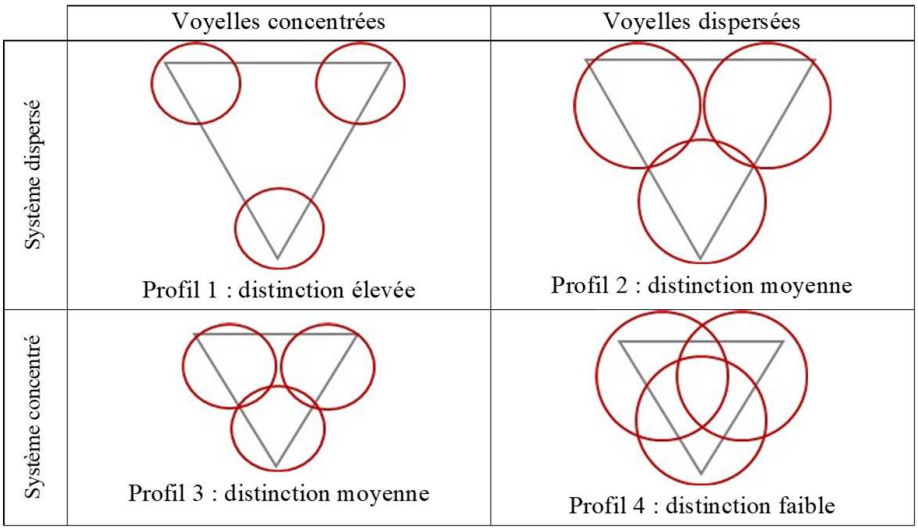

Figure 1 : Représentation schématique de la production des voyelles selon deux paramètres de dispersion: celui du système vocalique et celui de chaque catégorie de voyelle. Quatre profils peuvent être proposés suggérant des confusions plus ou moins importantes entre les voyelles.

L'interaction entre la dispersion/concentration du système vocalique et celles des catégories de voyelle peut être représentée schématiquement (Figure 1). Nous observons ici que le degré de distinctivité (donc la possibilité ou non de chevauchement des réalisations vocaliques) est fonction de ces deux facteurs. Ces profils représentent des hypothèses très caricaturales et nous envisageons évidemment que ces deux paramètres évoluent de façon corrélée c'est-à-dire qu'une concentration du système pourrait entraîner une concentration des voyelles. Toutefois, ça n'est pas ce qui a été observé dans un changement de situation de parole : la lecture de mots isolés serait plutôt similaire au profil 1 alors que le passage à une parole plus enchaînée engendrerait un profil 4 (Meunier, Espesser, & Frenck-Mestre, 2006). Ce que nous cherchons, dans cette première étape, est de faire apparaître des différences interlocuteurs dans un type de parole très relâché.

Pour ce faire, et à partir des travaux de (Huet & Harmegnies, 2000), nous avons calculé un Indice de Distinctivité (ID) permettant d'exprimer le rapport entre la dispersion du système et celle des voyelles chez des locuteurs de façon à obtenir des profils différents. A plus long terme, nous envisageons plusieurs types de comparaison : locuteurs de différentes langues (et donc systèmes vocaliques variés), locuteurs dans des situations de parole différentes (Huet & Harmegnies, 2000)

et locuteurs affectés par des pathologies de la parole (comme l'ont présenté Audibert & Fougeron, 2012).

2 Méthodologie

2.1 Corpus

Les analyses ont porté sur un corpus de parole conversationnelle, le *Corpus of Interactional Data* CID, (Bertrand et al., 2008). Ce corpus, enregistré en 2003, comprend des enregistrements audio et vidéo de dialogues spontanés entre des locuteurs français natifs (8 conversations d'une heure chacune entre deux locuteurs, soit 16 locuteurs, 10 femmes et 6 hommes). Dans chaque dialogue, les locuteurs entretenaient une conversation familière. Pour notre étude, nous avons sélectionné 10 locuteurs[1] : 5 femmes (AB, AC, BX, LL, ML) et 5 hommes (AG, AP, EB, LJ, SR). L'ensemble du corpus a bénéficié d'une transcription orthographique enrichie. Cette transcription a ensuite été phonétisée, puis alignée automatiquement de façon à obtenir une annotation phonétique (Bertrand et al., 2008).

Notre analyse porte exclusivement sur les voyelles orales. Les voyelles dont la durée était inférieure à 30ms ou supérieure à 300ms ont été exclues de nos analyses. En effet, les voyelles très courtes sont souvent (même si ça n'est pas systématiquement) le produit d'erreurs d'alignement dus à des omissions ou réductions non perçues pas les transcripteurs. De même, les voyelles extra-longues incluent très souvent des pauses remplies que nous ne souhaitions pas inclure dans cette étude. Enfin, un filtre établi par (Gendrot & Adda-Decker, 2005) a été appliqué de façon à exclure les valeurs de formant aberrantes dues à des erreurs de détection. En conséquence, un total de 37452 voyelles a été analysé dans notre étude (Table 1).

Voyelles	AB	AC	AG	AP	BX	EB	LJ	LL	ML	SR	Total
@	510	453	554	607	399	570	513	309	480	609	5004
A	925	959	1057	852	880	654	1087	473	1069	873	8829
e	1530	1278	1398	1303	1303	1106	1427	606	1289	1306	12546
i	716	449	437	525	482	450	697	318	528	484	5086
o	298	207	235	236	273	183	326	120	288	263	2429
u	132	118	105	119	116	41	115	60	129	87	1022
y	293	205	338	310	158	249	323	117	262	281	2536
Total	4404	3669	4124	3952	3611	3253	4488	2003	4045	3903	**37452**

Table 1: nombre de voyelles étudiées par locuteur et par voyelle

On constate que le nombre total de voyelles produites est très variable selon le locuteur (2003 pour LL et 4488 pour LJ) et est fonction du temps de prise de parole dans le dialogue pour chaque locuteur. De même, on observe que l'effectif de chaque catégorie de voyelle est très hétérogène. La voyelle /e/ est 12 fois plus représentée que la voyelle /u/. Cela tient à deux facteurs : le premier, bien évidemment, est la fréquence des voyelles dans le lexique, ainsi que la fréquence du lexique dans le

[1] Ces 10 locuteurs ont été sélectionnés pour deux raisons : 1/ pour une partie d'entre eux, nous disposons de corpus de lecture (mots et textes) avec lequel des comparaisons sont envisagées ; 2/ pour une autre partie, nous disposons d'annotations phonétiques fines (alignement corrigé par un expert humain) qui nous permettent des analyses plus poussées.

discours ; le deuxième réside dans le fait que le phonétiseur ne fait pas de distinction pour les voyelles moyennes qui sont regroupées en archiphonèmes /e/ (/e/, /ɛ/), /o/ (/o/, /ɔ/) et @ (/ə/, /ø/, /œ/). Quoiqu'il en soit, on observe que le système du français montre une majorité de voyelles antérieures (ou centrale). Cette tendance est accentuée par une sur-représentation de ces voyelles dans le discours. Elle n'est pas anodine pour le calcul du centre de gravité du système vocalique des locuteurs qui pourra ainsi tendre vers l'avant. Nous aurons à prendre en compte ce phénomène dans l'analyse des données (voir section ci-dessous)

2.2 Mesures

Les trois premiers formants ont été estimés automatiquement à l'aide d'une méthode de prédiction linéaire (autocorrélation) avec un algorithme Viterbi de façon à sélectionner les meilleurs candidats en imposant une contrainte de continuité fréquentielle (ESPS package, Entropic, 1997). Par la suite les mesures en Hertz des trois formants ont été converties en Bark selon la formule de Traunmüller (1990). En effet, dans la mesure où notre étude porte sur une analyse des distances euclidiennes entre chaque voyelle et son barycentre pour les trois formants, il nous a semblé important d'homogénéiser au mieux chacune des dimensions de l'espace. En effet, si on considère les variations de F1 de 348 à 685 Hz pour les femmes (Gendrot & Adda-Decker, 2005), cela représente une dynamique de 337Hz mais seulement 2.9 Barks d'écart. Si on considère les variations de F2 de 1140 à 2365 Hz, cela représente 1225 Hz d'écart (4 fois plus que pour F1) mais seulement 4.8 Barks. Enfin, pour F3, les auteurs mesurent des variations de 2687 à 3130 Hz (ΔF=443 Hz) ce qui représente environ 1 Bark. L'homogénéité n'est pas parfaite mais bien meilleure en Bark qui, nous le rappelons, est une échelle psychoacoustique imitant les mécanismes de la perception humaine, notamment de distinctivité, ce qui est parfaitement cohérent avec notre travail.

Pour chaque locuteur et pour F1, F2 et F3, nous avons calculé: 1/ la moyenne en Bark des valeurs de chaque catégorie de voyelle (**MOY_VOY**) et 2/ la moyenne en Bark pour l'ensemble des voyelles produites (**MOY_SYS**)**.** Ces moyennes nous permettent de calculer, pour chaque locuteur, les distances euclidiennes en trois dimensions (F1, F2 et F3) entre la valeur de chaque voyelle et la moyenne 1/ de sa catégorie et 2/ du système.

2.2.1 Calcul des Distances Euclidiennes par catégorie de voyelle (DE_3D_VOY)[2]

DE_3D_VOY représente la dispersion des voyelles par rapport au centre de gravité de leur catégorie (dispersion des voyelles). Elle est obtenu en appliquant la formule suivante :

$$\sqrt{(F1v - F1V)^2 + (F2v - F2V)^2 + (F3v - F3V)^2}$$

Où $F1v$ représente la valeur de F1 en Bark d'une voyelle et $F1V$ la valeur moyenne en Bark de F1 pour la catégorie de cette voyelle (MOY_VOY). Par exemple, on soustrait à un exemplaire d'une voyelle /a/ la moyenne des valeurs de toutes les voyelles /a/ chez un même locuteur, ce qui donne la distance entre cet exemplaire et la moyenne de sa catégorie. La même formule a été appliquée pour F2 et F3. Cette mesure nous donne donc la distance euclidienne de chaque voyelle par rapport à son

[2] Par rapport à l'étude de (Huet & Harmegnies, 2000), nos DE_3D_VOY correspondent au CM_{intra}, tandis que nos DE_3D_SYS correspondent au CM_{inter}

centre de gravité sur trois dimensions. Nous pouvons ensuite calculer, pour chaque catégorie de voyelle la moyenne de ces DE. On obtient donc une valeur de dispersion pour chaque catégorie de voyelle de chaque locuteur.

2.2.2 Calcul des Distances Euclidiennes pour le système (DE_3D_SYS)

DE_3D_SYS représente la dispersion des voyelles par rapport au centre de gravité du système (dispersion du système). Elle est obtenu en appliquant la formule suivante :

$$\sqrt{(F1v - F1S)^2 + (F2v - F2S)^2 + (F3v - F3S)^2}$$

Où $F1v$ représente la valeur de F1 en Bark d'une voyelle et $F1S$ la valeur moyenne en Bark de F1 pour l'ensemble des voyelles produites (MOY_SYS). La même formule a été appliquée pour F2 et F3. Cette mesure nous donne donc la distance euclidienne de chaque voyelle par rapport au centre de gravité de l'ensemble des voyelles produites (pour chaque locuteur) sur trois dimensions. Nous pouvons ensuite calculer, pour chaque locuteur la moyenne de ces DE. Toutefois, le calcul de cette moyenne est obtenu en regroupant les valeurs par catégorie de voyelle de façon à ne pas faire entrer dans le calcul le poids des effectifs de voyelles (Table 1). On obtient donc une valeur de dispersion du système vocalique pour chaque locuteur.

2.2.3 Calcul de l'Indice de Distinctivité (ID)

L'Indice de Distinctivité (ID) a pour avantage de fournir une information dynamique sur l'espace de production des voyelles et de mettre en lien la dispersion du système avec celle des catégories de voyelle. Il est donc obtenu en calculant le rapport entre la dispersion du système et la dispersion de chaque catégorie de voyelle :

ID = moyenne DE_3D_SYS / moyenne DE_3D_VOY

3 Résultats

3.1 Centre de gravité (CG)

Nous avons, dans un premier temps, estimé le centre de gravité des productions vocaliques des locuteurs sur un plan F1/F2 (Figure 2). Ces CG sont calculés à partir des moyennes des catégories de voyelle et ne tiennent donc pas compte des effectifs de chaque catégorie de voyelle (Table 1).

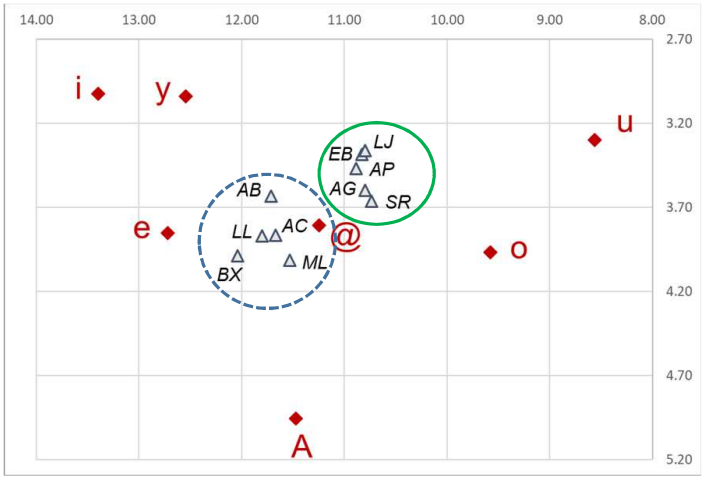

Figure 2 : Centre de gravité des productions des 10 locuteurs sur un plan F1/F2 (en Bark). Chaque locuteur est représenté par un triangle gris (hommes: cercle vert continu ; femmes: cercle bleu pointillé). En rouge sont représentées les valeurs moyennes de chaque catégorie de voyelle produites par l'ensemble des locuteurs.

Il en ressort une nette distinction homme/femme, les locutrices montrant un centre de gravité plus bas et plus antérieur tandis que celui des locuteurs est plus central et plus haut. Ce résultat est conforme aux observations habituelles et s'explique par la taille plus réduite du conduit vocal des locutrices. Globalement, le CG moyen de l'ensemble des locuteurs tend plutôt vers l'avant, ce qui n'est pas surprenant étant donné qu'une majorité des voyelles du français est plutôt antérieure (ou centrale) tandis qu'il n'y a que 2 voyelles d'arrière.

3.2 Distances euclidiennes 3D (F1-F2-F3)

Les DE_3D pour les voyelles et le système ont été calculées selon la méthode expliquée en 2.2.1 et 2.2.2. On note ainsi que les DE_3D_SYS sont toujours supérieures aux DE_3D_VOY (Table 2), ce qui est conforme à nos attentes puisque la dispersion autour de chaque voyelle est logiquement moins grande que la dispersion de toutes les voyelles par rapport au centre du système.

	femmes					hommes				
	AB	AC	BX	LL	ML	AG	AP	EB	LJ	SR
DE_3D_VOY	1.39	1.78	1.53	1.58	1.51	1.63	1.73	1.63	1.40	1.50
DE_3D_SYS	1.94	2.22	2.07	2.11	2.01	2.20	2.23	1.97	1.94	2.12
ID	1.40	1.24	1.35	1.33	1.33	1.35	1.29	1.21	1.39	1.41

Table 2 : distance euclidienne 3D (F1, F2, F3) pour chacun des locuteurs et chacun des espaces (espace système et moyenne des espaces voyelles). L'ID correspond au rapport entre les deux DE_3D

On notera également un certain équilibre dans la production des locuteurs dans la mesure où, le plus souvent, lorsque la dispersion du système est faible, la dispersion de chaque voyelle tend à être également minimisée (AB, LJ) et inversement (AC, AP). Il semble donc que les locuteurs tendent vers un équilibre dans la production de leurs voyelles de façon à maintenir un minimum de distinctivité. Il est malgré tout possible de distinguer des profils différents en comparant les Indices de Distinctivité (ID).

3.3 Indices de distinctivité

Comme expliqué plus haut (2.2.3), l'ID se présente comme un rapport entre la dispersion du système et la dispersion des catégories de voyelles. La Figure 3 nous montre une représentation hiérarchisée des locuteurs les plus distinctifs (à gauche) vers ceux qui le sont le moins (à droite). Trois locuteurs (SR, AB et LJ) sont clairement au-dessus de la moyenne des locuteurs tandis que trois autres (AP, AC et EB) sont clairement en-dessous. Les quatre autres locuteurs se situent autour de la moyenne.

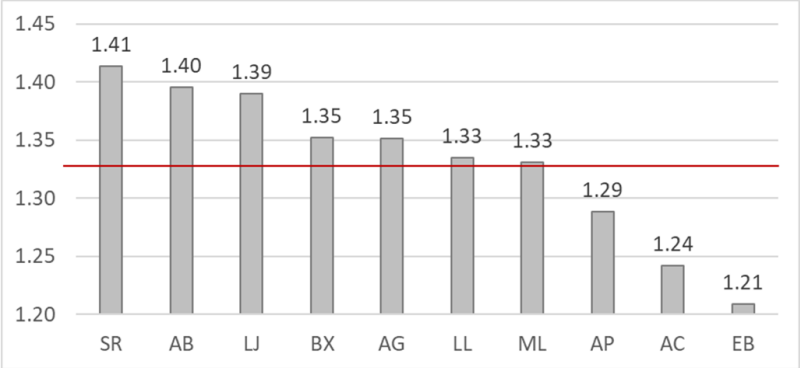

Figure 3: Indice de distinctivité (ID=rapport entre dispersion du système et dispersion des voyelles) calculé pour les 10 locuteurs. En rouge, la moyenne des ID des locuteurs (ID=1.33).

L'échelle laisse supposer que les différences sont importantes alors qu'en réalité, nous ne pouvons, en l'état, rien dire sur la magnitude de ces différences. Toutefois, nous pourrions poser l'hypothèse qu'une valeur 1 est une limite basse car en dessous de 1, cela signifie que la dispersion de chaque voyelle est plus forte que celle de tout le système, ce qui intuitivement est une limite. Cela aurait donc du sens de soustraire 1 à notre indice actuel de distinctivité, la valeur 0 indiquant alors une distinctivité nulle (dispersion de chaque voyelle étant égale à celle de tout le système). Dans ce cadre, une valeur à 1.41 (locuteur SR, Figure 3), qui deviendrait 0.41 après soustraction, indiquerait alors une nette différence avec 1.21 (locuteur EB) qui deviendrait 0.21. Dans tous les cas, pour pouvoir donner un sens à cette magnitude, il sera nécessaire de mettre en lien ces différences avec une mesure de l'intelligibilité des locuteurs.

Figure 4: DE_3D_SYS (gris clair) et DE_3D_VOY (gris foncé) en Bark classés par ordre décroissant de DE_3D_SYS pour chaque locuteur (axe de gauche). En points rouges (axe de droite), l'ID pour chaque locuteur

Nous avons par ailleurs mis en lien l'ID avec les résultats concernant la dispersion du système, d'une part, et la dispersion des voyelles, d'autre part. Cette comparaison, représentée sur la Figure 4, met en évidence le fait qu'un ID élevé n'est pas forcément lié à un système dispersé. En effet, les locuteurs LJ et AB présentent des ID élevés mais les systèmes les moins dispersés. De même le locuteur SR a l'ID le plus élevé des locuteurs mais ne présente pas le système le plus dispersé. Ce qui explique l'ID élevé de ces trois locuteurs, c'est la faible dispersion des catégories de voyelle. Inversement, on notera que la locutrice AC, présentant un système fortement dispersé, a un ID faible en raison d'une forte dispersion des catégories de voyelle. Notre mesure de l'ID apparaît donc bien comme une combinaison des deux facteurs de dispersion et n'est pas systématiquement liée à la taille du système vocalique.

4 Conclusions et perspectives

Notre étude avait pour objectif de déterminer un indice de distinctivité basé sur l'indice *Phi* de Huet & Harmegnies (2000) et permettant de rendre compte de la relation dynamique entre la dispersion des catégories de voyelles et la dispersion de l'ensemble du système. Nous avons vu que cet indice permet de différencier les locuteurs. Par ailleurs, bien que les données sur les CG des systèmes des locuteurs montre une nette différence Homme/Femme, nous n'avons retrouvé cette différence ni dans les dispersions (système ou voyelles), ni dans la mesure de l'ID. Nous avons constaté également qu'un ID élevé n'est pas lié à un système plus large mais plutôt à la combinaison des deux facteurs.

Toutefois, nos conclusions doivent s'arrêter là et nous ne pouvons rien dire sur la magnitude de ces différences d'ID. En effet, rien ne nous permet de dire, à ce stade, que ces différences sont importantes ou qu'elles aient un impact sur l'intelligibilité des locuteurs. Pour cela, il nous faudra mettre en place une évaluation perceptive des productions de ces locuteurs. Notre objectif à plus long terme sera d'utiliser l'ID pour caractériser des contextes différenciés. En premier lieu, des travaux précédents ont pu mettre en évidence des espaces de réalisation distincts selon la langue parlée (Al-Tamimi & Ferragne, 2005; Meunier et al., 2006 ; Gendrot & Adda-Decker, 2007). Nous faisons l'hypothèse que, selon l'organisation du système vocalique mais également des propriétés lexicales présents dans une langue donnée, des profils de distinctivité différents pourraient émerger selon les langues. De même, on peut supposer que la magnitude des différences entre les ID des locuteurs sera plus importante si on observe leur production en lecture ou en parole spontanée (comme l'ont mis en évidence Huet & Harmegnies, 2000). On pourra alors observer si la hiérarchie du classement des locuteurs selon l'ID est conservée quelle que soit la situation de parole, ce qui supposerait des profils spécifiques aux locuteurs. Des expériences de type bite-block pourraient également nous permettre de mettre en évidence les phénomènes de compensation et de réajustement dynamique des dispersions. Enfin, les pathologies de la parole caractérisées par un déficit moteur (dysarthrie) montrent une désorganisation et/ou une centralisation du système vocalique pour lesquelles des systèmes de mesure très fins ont été proposés (Audibert & Fougeron, 2012). Nous pensons que la mesure de l'Indice de Distinctivité pourrait apporter des réponses complémentaires sur cette désorganisation qui n'est pas systématiquement une réduction du système vocalique. Cet indice de distinctivité pourrait finalement fournir une mesure de prédiction d'intelligibilité sachant qu'une faible distinctivité pourrait entrainer des problèmes de décodage et de compréhension.

Remerciements

Les données audiovisuelles ont été enregistrées et mises à disposition dans le cadre du Centre d'Expérimentation sur la Parole du Laboratoire Parole et Langage à Aix-en-Provence.

Références

AL-TAMIMI J. E., FERRAGNE E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. *Proceedings of Interspeech*, 2464–1468.

AUDIBERT N., FOUGERON C. (2012). Distorsions de l'espace vocalique : quelles mesures? Application à la dysarthrie. *Proceedings of JEP-TALN-RECITAL,* 217–224.

BERTRAND R., BLACHE P., ESPESSER R., FERRE G., MEUNIER C., PRIEGO-VALVERDE B., & RAUZY S. (2008). Le CID — Corpus of Interactional Data — Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique Des Langues*, *49* (3), 105–134.

GENDROT C., ADDA-DECKER M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Proceedings of Interspeech*, 2453-2456.

GENDROT C., & ADDA-DECKER M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1417–1420.

HUET K., & HARMEGNIES B. (2000). Contribution à la quantitication du degré d'organisation des systèmes vocaliques. *Journées d'Etudes sur la Parole,* 225–228.

MEUNIER C., ESPESSER R., & FRENCK-MESTRE C. (2006). Aspects phonologique et dynamique de la distinctivité au sein des systèmes vocaliques: une étude inter-langue. In *Journées d'Etude sur la Parole*, 333-336.

SMILJANIC R., BRADLOW A. R. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, *3*(1), 236–264.

TRAUNMÜLLER, H. (1990). "Analytical expressions for the tonotopic sensory scale". *The Journal of the Acoustical Society of America*. 88: 97.