Détection de la stabilité de timbre des voyelles : vers une automatisation des tâches

*Christelle Dodane et Christian Guilleminot**

Laboratoire de Phonétique, ELADI, LaSELDI, *membre associé UMR CNRS 5475
30, rue Mégevand – 25000 Besançon, France
Tél. : ++33(0)3 81 66 53 97 – Fax ++33(0)3 81 66 53 00
Mél: christelle.dodane3@freesbee.fr - christian.guilleminot@univ-fcomte.fr

ABSTRACT

Differences of rhythmic patterns in French and English generate differences in the degree of the articulatory tension of vowels and, therefore, stability differences in phonetic vowel quality. In order to study interferences between the two languages during the learning of English by French learners, we have elaborated a method in order to delimit the duration of vowel quality stability, using the tracking of the first three formants frequencies and based on formant ratio theory. This method had been automatized in order to proceed a big quantity of data.

1. INTRODUCTION

Dans le cadre d'un travail sur l'apprentissage de l'anglais par des enfants francophones âgés de 7 ans [Dod02], nous avons été amenée à réaliser une analyse comparée du système prosodique et vocalique de l'anglais et du français, de manière à prévoir les interférences qui pourraient se produire entre les deux langues et en gêner l'apprentissage. L'organisation rythmique de l'anglais engendre un régime de tension musculaire radicalement différent de celui du français [Wen82], ce qui va directement conditionner la prononciation des voyelles : on observe en effet une tension décroissante car la syllabe accentuée se trouve généralement en début de groupe rythmique. Les syllabes atones qui suivent sont de ce fait, affectées par un relâchement articulatoire qui produit une réduction du timbre de leurs voyelles. En revanche, le français est marqué par un régime de tension croissante : à cause de la place de l'accent en finale, l'articulation est très tendue, car l'élément important est toujours à venir et on ne rencontre pas de voyelles relâchées (netteté de timbre, même avec un débit rapide). Or, en raison du phénomène de crible prosodique, les francophones ont tendance à transposer la rythmique du français à l'anglais. La tension articulatoire qui caractérise leur langue maternelle les amène à produire des voyelles trop stables en anglais. Cet excès de stabilité va bouleverser la structure interne du noyau vocalique : la durée de la partie stable sera démesurée par rapport à celle de la transition initiale et de la transition finale, spécialement dans le cas des voyelles longues, normalement marquées en anglais par de longues phases de transition [Dod02]. Le fait de connaître la structure interne des voyelles anglaises produites par les apprenants français nous permettra de déterminer s'ils se rapprochent ou non des voyelles « typiques » de l'anglais, la répartition des événements composant le noyau vocalique servant d'indice sur la progression de leur maîtrise de la prononciation. Dans ce but, nous avons mis au point une méthode qui nous permet de délimiter les différents événements composant le noyau vocalique grâce à la détection de la phase de stabilité de timbre. Cette méthode a été élaborée à partir de l'analyse d'un corpus composé des voyelles typiques de l'anglais et du français. A partir de l'analyse de ce corpus, nous pourrons en outre ébaucher une description du système vocalique de ces deux langues en fonction du critère de stabilité de timbre. Mais l'objectif principal de cet article étant de présenter notre méthode, nous ne donnerons les résultats que pour en permettre l'évaluation [Dod02]. L'automatisation de cette méthode via la création d'un programme informatique nous permet d'envisager l'analyse de grands corpus.

2. FONDEMENTS THÉORIQUES

2.1 De la stabilité articulatoire à la stabilité de timbre

Pour pouvoir comparer la durée relative des événements constituant le noyau vocalique, il est nécessaire de commencer par définir ce qu'est une cible vocalique. Une fois cette notion définie, il nous sera facile de délimiter ces événements au sein du noyau vocalique. Traditionnellement, les chercheurs pensent que l'information essentielle pour déterminer la qualité des voyelles réside dans la localisation des maxima spectraux correspondant aux 2 premiers formants [Lon83] [Sch87] ou aux 3 premiers formants [Syr86] [Fah96]. Cette localisation correspond à la notion de cible vocalique. Pour Lehiste et Peterson [Leh60 : 290], « *the time interval within the syllable nucleus where the formants are parallel to the time axis has been considered as the extent of a vowel target* ». L'« extension de la cible vocalique » dont il est question correspond à la partie stable de la voyelle et elle se mesure habituellement sur une représentation spectrographique en délimitant la zone où les formants sont parallèles à l'axe du temps. Cette stabilité spectrale reflète une position articulatoire stable de la langue et des lèves. Or, la notion de cible n'a de pertinence qu'au travers de la perception [Joh93] [Lad97] : en effet, la cible qu'essaie d'atteindre le locuteur en produisant une voyelle est avant tout définie par des propriétés auditives, plutôt que par la configuration du tractus vocal. La partie stable correspond

donc à un intervalle de temps au cours duquel la stabilité de la structure spectrale engendre une stabilité de timbre pour l'auditeur.

2.2 Les rapports entre les formants

Diverses procédures de mesure permettent de relever les valeurs absolues de F1, F2 et F3 en Hz en un point précis. Elles diffèrent seulement par leur manière de détecter ce point [Bar52], [Lab72], [Len78], [Par85]. Puisque nous voulons mesurer un intervalle de temps au cours duquel le timbre de la voyelle est stable, nous ne pouvons donc pas nous servir de ces méthodes. Par ailleurs, est-il vraiment judicieux d'utiliser les valeurs absolues des trois premiers formants lorsqu'on travaille sur la stabilité de timbre vocalique ? Selon la « formant-ratio theory » formulée par Llyod à la fin du XIXème siècle et citée par Miller [Mil89 : 2115], « *the vowel quality depends on intervals between the resonances, not* ***on*** *their absolute values* ». Si on compare les mesures de fréquences des trois premiers formants pour les hommes, les femmes et les enfants des données de Barney et Peterson [Bar52], on observe une très grande variabilité. Mais, cette variabilité intra et interlocuteurs est en grande partie éliminée si les voyelles sont caractérisées non plus en termes de fréquences absolues, mais en termes de rapports entre leurs formants[Mil89] [Pot50]. Ainsi, les rapports de fréquences entre le premier et le second formant, et entre le second et le troisième formant pourraient servir à éliminer les différences entre les locuteurs [Pet61]. Pour cette raison, de tels rapports jouent un rôle dominant dans l'interprétation auditive et perceptive des voyelles.

2.3 Définition de la stabilité de timbre

En nous inspirant de ces travaux, nous pourrions donc définir la stabilité de timbre comme l'**intervalle de temps où les rapports entre F1, F2 et F3 sont constants**. Mais cet intervalle de stabilité est lui-même soumis à de légères variations. A partir de quel taux d'instabilité l'oreille perçoit-elle une variation de timbre ? En effet, l'oreille ne peut détecter un changement que si la variation de l'excitation dépasse une certaine quantité ou seuil différentiel. Il serait donc plus approprié de définir la stabilité de timbre de la voyelle comme un **intervalle de temps où les rapports entre les trois premiers formants évoluent en dessous d'un certain seuil de perception**. Au-delà de ce seuil, l'auditeur perçoit un changement de timbre. Si nous sommes capables de délimiter la zone de stabilité de timbre, il sera alors facile de délimiter la durée des autres sous-segments, c'est-à-dire les intervalles de temps correspondant à la transition initiale (« tête ») et la transition finale (« queue ») de la voyelle.

3. APPLICATION

3.1 Matériel de parole

Deux locutrices ont été enregistrées en chambre sourde au laboratoire de phonétique de Besançon (magnétophone D.A.T Aiwa HD-S1 et microphone Aiwa) : une locutrice francophone et une locutrice anglophone s'exprimant dans un accent standard. Celles-ci ont lu une liste de mots monosyllabiques de type CVC, chacun commençant par la constrictive sourdre [s] et se terminant par l'occlusive sourde [t]. Ces mots ne diffèrent que par leur noyau vocalique et couvrent la quasi-totalité des monophtongues de l'anglais et du français. La locutrice francophone a ainsi prononcé une liste de 11 mots contenant les 11 voyelles françaises [i], [e], [y], [ɛ], [a], [u], [ɔ], [œ], [ɑ], [o] et [ø] et la locutrice anglophone, une liste de 11 mots contenant les 11 voyelles anglaises [ɪ], [e], [æ], [ʌ], [ʊ] et [ɒ]. Les 4 voyelles oralo-nasales du français on été écartées, car elles n'ont pas d'équivalent en anglais. Leur étude n'est pas pertinente dans une analyse qui vise à décrire l'apprentissage du système vocalique de l'anglais par des apprenants francophones. De même, le [ə] a été exclu de l'analyse, car il ne peut se rencontrer dans les mots monosyllabiques en anglais, ceux-ci étant obligatoirement accentués. Une extraction des trois premiers formants a été réalisée sur la longueur totale de chacune de ces voyelles avec le logiciel Winsnoori[1] (qui permet une extraction des trois premiers formants toutes les 6 ms), générant un total de 22 fichiers. La fiabilité de l'extraction a été expertisée manuellement, de manière à éliminer toute erreur de détection.

3.2 Traitement des données

Winsnoori fournit des tableaux contenant les valeurs des fréquences des trois premiers formants mesurées en hertz. La première action consiste à convertir les valeurs correspondant à F1, F2 et F3 en huitièmes de tons. La formule à appliquer à chaque valeur à convertir est :

$$S = (\log(X) - \log(220)) / (\log(2)/48)$$

X étant la valeur en Hz à convertir et 220 la fréquence de référence.

D'après nos observations, F3 constitue le formant le plus stable. Il nous sert donc de référence de comparaison, pour calculer les rapports R1=F1/F3 et R2=F2/F3. La différence S1=R1-R2 entre ces deux rapports nous permet d'évaluer les variations relatives entre les valeurs de F1, F2 et F3.

3.3. Fixation d'un seuil de variabilité

Il reste maintenant à fixer un seuil en deçà duquel les rapports entre les trois premiers formants, ramenés à la série de nombres S1, évoluent sans qu'il y ait changement de timbre (cf. paragraphe 2.3). Nous avons vu que les voyelles françaises se caractérisaient par une grande

[1] Winsnoori, version 1.2, LORIA, Babel Technologies.

stabilité de timbre, provenant du régime de tension croissant propre à cette langue. Il nous sera donc plus facile de délimiter la zone de stabilité de timbre à partir des voyelles prononcées par la locutrice francophone. Les bornes délimitant la zone de stabilité de chaque voyelle ont été posées en croisant une analyse de leur représentation spectrographique (évolution parallèle des trois premiers formants, correspondant à la définition habituelle de la partie stable de la voyelle) à notre propre jugement auditif (zone sans changement de timbre perceptible). Nous avons ensuite calculé le taux de variabilité affectant la série de nombre S1 sur la zone de stabilité délimitée précédemment, en faisant la soustraction de la valeur maximale à la valeur minimale. Le taux de variabilité fixé correspond au taux de variabilité le plus élevé sur l'ensemble des 11 voyelles françaises, soit 0,09. La figure 1 montre la représentation spectrographique de la voyelle [ɔ] et l'évolution de ses trois premiers formants. Sur la figure 2, sont représentées les valeurs de S1 en fonction du temps. Le rectangle regroupe les valeurs évoluant en dessous du seuil, soit la zone de stabilité de timbre de la voyelle [ɔ].

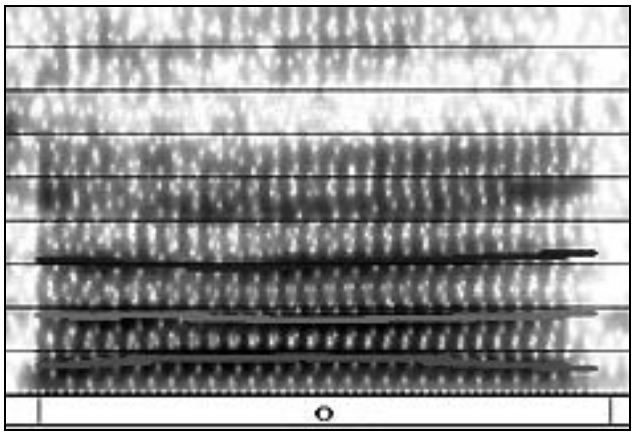

Figure n°1 : Extraction formantique de la voyelle [ɔ] édité avec le logiciel Winsnoori.

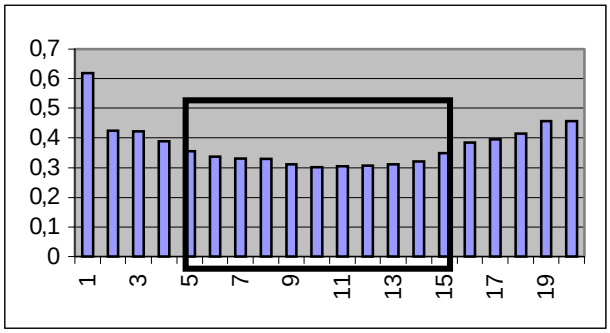

Figure n°2 : Évolution de la série S1 et délimitation de la zone de stabilité de timbre de la voyelle [ɔ].

3.4 Automatisation de la méthode

Un programme informatique a été créé afin d'effectuer les calculs décrits et de donner automatiquement la durée de la phase de stabilité de timbre de la voyelle. Le programme traite chaque fichier comportant l'évolution temporelle des trois premiers formants, édité avec le logiciel Winsnoori, fait la transposition des valeurs fréquentielles en valeur de perception (1/8ème de ton), fait la différence des rapports entre formants (R1, R2) et analyse le résultat S1 pour en extraire la durée de la partie stable en fonction du seuil de variabilité fixé au préalable. Le module qui permet de détecter automatiquement l'intervalle de stabilité de timbre a été programmé de la manière suivante : un rectangle dont la hauteur correspond au seuil de variabilité fixé est centré sur chaque point de la courbe S1. Le programme détecte combien de points consécutifs précédant ou suivant le point considéré sont inclus dans le rectangle. Si la courbe possède n points, il y aura n intervalles. La zone de stabilité de timbre correspond au plus grand intervalle trouvé. Le fichier de résultats fournit la délimitation de la zone de stabilité de timbre, la variabilité affectant cette zone de stabilité, ainsi que la durée absolue (en ms) et relative (en %) de tous les événements composant le noyau vocalique (figure 3).

```
Fichier F33.txt
voyelle o

PS (5 ,15)    var  : 0.053030
DT  = 111 ms
DPS = 58 ms
DTE = 23 ms
DQ  = 30 ms

% DPS/DT = 52.252251 %
% DTE/DT = 20.720720 %
% DQ/DT  = 27.027027 %
```

Figure n°3 : Fichier de résultat fourni par le programme pour la voyelle [ɔ][2]

3.5 Résultats et discussion

La table 1 nous donne les résultats de la détection automatique des événements composant chacune des 22 voyelles analysées. La fiabilité de la détection et l'adéquation du seuil de variabilité ont été testées sur l'ensemble des 22 fichiers. L'optimisation de l'algorithme de détection ne s'est arrêtée que lorsque les zones de stabilité détectées par le programme ont correspondu exactement aux zones de stabilité délimitées manuellement. Le problème central à résoudre pour une automatisation complète de cette détection concerne la qualité de l'extraction formantique. Lorsque celle-ci est bonne, ce qui était le cas pour le corpus présenté, la fiabilité de la détection de la zone de stabilité de timbre est maximale. En revanche, cette méthode a été appliquée à un corpus constitué des productions de 25 enfants

[2] Abréviations utilisées : PS, partie stable ; var, taux de variabilité affectant la partie stable ; DT, durée totale de la voyelle ; DPS, durée de la partie stable ; DTE, durée de la tête de la voyelle ; DQ, durée de la queue de la voyelle.

francophones en apprentissage précoce de l'anglais [Dod02]. Sur un total de 1150 mots prononcés, seuls 295 fichiers comportaient une extraction formantique correcte, soit 25,65 % des mots. Outre le problème de qualité du signal, il s'avère que la détection est beaucoup moins aisée avec des voix d'enfants. Le problème de l'extraction est donc prioritaire si nous voulons réussir une automatisation complète de la détection de la stabilité de timbre de la voyelle, mais il relève du logiciel employé et non de notre méthode de mesure de la stabilité.

Table 1 : durée relative exprimée en % des événements composant les voyelles anglaises (Ang.) par rapport aux voyelles françaises (Fs.).

Ang	DT	DPS	DQ	Fs	DT	DPS	DQ
[ɪ]	12,78	78,1	9,02	[i]	0	100	0
[iː]	33,33	63,7	2,87	[e]	7,91	71,22	20,86
[e]	20	80	0	[ɛ]	8,27	57,14	34,58
[ɜː]	23,96	59,4	16,61	[œ]	6,89	86,20	6,89
[æ]	8,82	91,1	0	[ø]	2,95	91,13	5,91
[ʌ]	22,29	77,7	0	[a]	55,17	35,34	9,48
[ɑː]	10,65	38,2	51,09	[ɑ]	51,82	42,52	5,64
[ɒ]	25,55	58,3	16,11	[ɔ]	20,72	52,25	27,02
[ɔː]	48,71	27,8	23,44	[o]	16,11	67,77	16,11
[uː]	18,06	25,5	56,38	[u]	0	68,18	31,81
[ʊ]	25,15	74,8	0	[y]	6,89	93,10	0

4. CONCLUSION

La mesure de la durée de la zone de stabilité de timbre au sein du noyau vocalique nous semble très pertinente dans une étude comparée de deux langues aussi différentes que le français et l'anglais dans leur « gestion » de la tension articulatoire. Non seulement, elle fournit un indice sur l'état de progression de l'apprenant quant à sa prononciation, mais elle constitue également un excellent indicateur pour travailler sur l'interaction entre les éléments « suprasegmentaux » et « segmentaux », la stabilité de timbre des voyelles dépendant largement de l'organisation rythmique de la langue. La détection automatique permet d'envisager l'analyse de corpus importants.

BIBLIOGRAPHIE

[Bar52] Barney H. L. et Peterson G. E. (1952), "Control Methods used in a study of the vowels", JASA, Vol. 24, pp. 175-184.

[Dod02] Dodane C. (2002), "Influences de la formation musicale sur l'apprentissage précoce d'une langue étrangère", Thèse de Doctorat en cours, Université de Besançon.

[Fah96] Fahey R. P., Diehl R. L., & Traunmüller H (1996), "Perception of back vowels: Effects of varying F1-F0 Bark distance", JASA, Vol. 99 (4), pp. 2350-2357.

[Joh93] Johnson K., Ladefoged P. et Lindau M. (1993), "Individual differences in vowel production", JASA, Vol. 94, pp. 701-14.

[Lab72] Labov W., Yaeger M. et Steiner R. (1972), A Quantitative Study of Sound Change in Progress". Philadelphia: The U.S. Regional Survey.

[Lad97] Ladefoged P. (1997), "Linguistic Phonetic Descriptions", in Hardcastle J. L. et Laver John, The Handbook of Phonetic Sciences, Padstow: Blackwell Publishers, pp. 589-618.

[Leh60] Lehiste I. et Peterson G. E. (1960), "Transitions, glides, and diphthongs", Baken et Daniloff (1991), "Readings in Clinical Spectrography of Speech", Kay Elemetrics. pp. 286-295.

[Len78] Lennig M. (1978), Acoustic Measurement of Linguistic Change : the Modern Paris Vowel System. Philadelphia: Doctorat of Phylosophy, 190 p.

[Lon83] Lonchamp F. (1983), "On Vowel Normalisation", Leeds Experimental Phonetics Symposium, Leeds.

[Mil89] Miller J. D. (1989). "Auditory-perceptual interpretation of the vowel", JASA, Vol. 85, pp. 2114-2134.

[Par85] Paradis C. (1985), An Acoustic Study of Variation and Change in the Vowel System of Chicoutimi and Jonquière. Québec: Doctorat of Philosophy, 326 p.

[Pet61] Peterson G. (1961), "Parameters of vowel quality", Journal of Speech and Hearing Research, Vol. 4, pp. 10-29.

[Pot50] Potter R. K. et Steinberg J. C. (1950), "Towards the specification of speech", JASA, Vol. 22, pp. 807-820.

[Sch87] Schwartz J. L. (1987), "Représentation Auditive des Spectres Vocaliques", Thèse de Doctorat d'Etat, Université de Grenoble.

[Syr86] Syrdal A. K. et Gopal H. S. (1986), "A perceptual model of vowel recognition based on auditory representation of American English vowels", JASA, Vol. *79* (4), pp. 1086-1100.

[Wen82] Wenk et Wioland (1982). "Is French really syllable-timed ?", Journal of Phonetics, Vol. 19.