

Analyses formantiques automatiques en français : périphéralité des voyelles orales en fonction de la position prosodique.

Cédric Gendrot* et Martine Adda-Decker**

* Laboratoire de Phonétique et Phonologie – CNRS/UMR7018

ILPGA, 19, rue des Bernardins – 75005 Paris

** LIMSI-CNRS bât. 508, BP 133, 91403 Orsay cedex

Mél: cgendrot@univ-paris3.fr - madda@limsi.fr

ABSTRACT

The aim of the present study is to highlight peripherality of French vowels in two prosodic positions: (i) word-final syllables (as compared to word-initial syllables), and (ii) in the vicinity (before and after) of pauses. The LIMSI speech alignment system is used and formant values of oral vowels are automatically measured in a total of 25000 segments from two hours of journalistic broadcast speech in French. A tendency to reduction for all vowels (in terms of the shrinking of the vocalic triangle formed by F1 and F2 values) of short duration was clearly observed in a former study (Gendrot & Adda-Decker [6]). We also show that at some extent, the same relationship holds for vowels in both word-final and word-initial syllables.

1. INTRODUCTION ET METHODE

La disponibilité de grands corpus audio et d'outils automatiques d'alignement de plus en plus performants nous a permis de réaliser une étude¹ (Gendrot et Adda-Decker [6]) sur les valeurs formantiques de 25000 voyelles orales en parole continue, et ce de manière automatique. Ces travaux contribuent à l'établissement de valeurs de formants et de leur variabilité en français, peu documentées à l'heure actuelle. Notre intérêt s'est porté principalement sur les variations de F1 et de F2 qui peuvent être interprétées – si on ne tient pas compte de l'effet des lèvres – en termes d'aperture/fermeture (corrélée à F1) et d'antériorité/postériorité (corrélée à F2). Nous avons montré :

- qu'il était possible d'extraire avec des traitements automatiques un pourcentage élevé (94%) des valeurs de formants de manière fiable.

- que l'espace vocalique formé par les formants F1 et F2 diminue progressivement avec la durée des segments analysés ; la durée étant répartie en trois catégories équilibrées (cf. Figure1). Ces variations étaient plus faibles pour les voyelles fermées /i/ et /y/ qui sont, comme

il est reconnu pour le français, mieux caractérisées par la proximité de F3 avec F2 (/y/) ou F4 (/i/).

- que les variations des formants étaient d'une amplitude comparable à celles observées pour des langues à accent lexical telles que l'allemand.

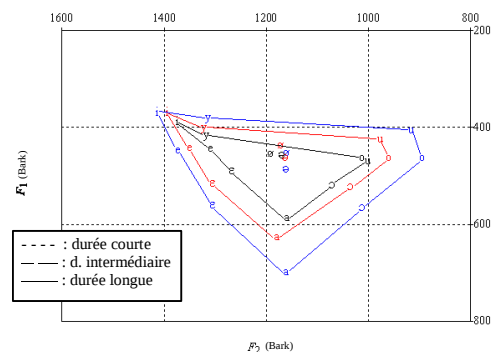


Figure 1 : Valeurs moyennes de F1 et F2 pour les voyelles orales du français en fonction de leur durée (normalisation en Bark).

Le corpus retenu est constitué d'enregistrements radio-et télédiffusés d'émissions journalistiques. Il s'agit de parole publique : l'articulation, sans être soutenue, y reste bonne, afin que la parole puisse être partagée par une large audience. La parole ne peut pas être qualifiée de spontanée, il s'agit plutôt de parole préparée : on observe peu d'hésitations, peu de fragments de mots et les structures syntaxiques restent souvent proches du langage écrit. Les phénomènes de réduction, auxquels nous nous intéressons à travers cette étude, y sont sans doute moindres que dans une vraie parole spontanée. Le corpus utilisé correspond à environ 2 heures de parole utile (15 hommes et 15 femmes) extraites pour la majeure partie d'émissions de France Inter, enregistrées en 1998 et transcrites orthographiquement (20k mots). Il s'agit ici de ressources utilisées dans le cadre du projet MIDL provenant du CTA/DGA, partenaire du projet.

Des analyses statistiques ont été effectuées au moyen d'ANOVAs à 1 facteur (position ou durée) voyelle par voyelle ; les tables de statistiques ne seront pas détaillées ici par manque de place.

2. HYPOTHÈSES

Cette étude vise à explorer plus précisément les variations formantiques des voyelles analysées en fonction de leur position prosodique. En effet, Gendrot et Adda-Decker

¹ Cette étude a été menée dans le cadre du projet MIDL Modélisations pour l'IDentification des Langues du programme interdisciplinaire STIC-SHS Société de l'Information du CNRS, dont les partenaires sont le LIMSI-CNRS, LPP Paris3, CTA/DGA, Télécom Paris, EA1483 Paris3

³ Notons que les intonations conclusives par nature plus basses ont été englobées dans ce résultat par manque d'étiquetage approprié, elles seront analysées dans une étude ultérieure.

Les figures 3a et 3b (page 4) montrent les variations mesurées pour les voyelles en syllabes initiales et finales respectivement, tout en mettant en évidence les différentes catégories de durées utilisées précédemment dans Gendrot et Adda-Decker, à savoir [30–50ms] pour les voyelles courtes, [60–80ms] pour les voyelles à durée intermédiaires et [90–110] pour les voyelles longues. Ces figures indiquent à première vue des tendances similaires, également identiques à celles relevées par la figure 1. Nous pouvons cependant signaler que les voyelles ouvertes /a/, /ɛ/ et /ɔ/ mesurées en syllabe finale de mots ont des valeurs de F1 significativement plus élevées que ces mêmes voyelles mesurées en syllabe initiale de mots. Ces différences observées sur l'axe F1 ont également été notées par Gendrot [7] sur des triangles vocaliques effectués sur la base des catégories de f0 mentionnées ci-dessus et non plus de catégories de durée.

Les résultats diffèrent également concernant la position de la syllabe dans le mot pour les voyelles fermées /i/, /y/, /u/ et /o/. En effet, les variations formantiques pour ces voyelles sur l'axe F1 sont non significatives ou peu importantes lorsque mesurées en syllabe finale de mots, alors que l'on peut remarquer un net détachement du 1^{er} formant des voyelles fermées courtes en syllabe initiale (avec des valeurs plus élevées) par rapport à leurs contreparties plus longues. Nous suggérons ici que les variations observées dans la distribution de ces voyelles, ainsi que les variations formantiques mesurées correspondent à la répartition naturelle de ces voyelles. La position allongée de fin de mot implique donc des voyelles plus périphériques, à l'exception des voyelles antérieures fermées /i/ et /e/. Les variations formantiques de ces voyelles avaient été notées comme plus faibles dans Gendrot & Adda-Decker [6].

4. VARIATIONS FORMANTIQUES AUTOUR DES PAUSES

Nous avons également réalisé des mesures identiques en fonction de la présence ou non de pauses immédiatement avant et après les voyelles analysées. Les pauses ont été déterminées dans la transcription par l'algorithme d'alignement suivant le même principe que les phonèmes (voir Gendrot & Adda [6]). La table 3 résume ainsi les catégories obtenues au moyen d'exemples, ainsi que le nombre d'occurrences recueillies.

Table 3 : Résumé des différentes catégories sélectionnées

	pause précédant V.	pause suivant V.
pause	[pause] <u>a</u> bri	matel <u>a</u> s [pause]
occurrences	avec : 920 sans : 22000	avec : 1600 sans : 21300

Rappelons ici qu'il est particulièrement délicat de se baser sur la syntaxe de la transcription pour déterminer des éventuelles frontières prosodiques comme cela est fait pour l'analyse de phrases lues. En effet, il est fréquent par exemple de constater la présence de pauses entre un

article et le nom qui lui est associé, ce qui va à l'encontre de bon nombre de prédictions, et est souvent considéré comme une (pause de) mise en valeur du mot suivant cette pause. Nous avons ainsi décidé de déduire des frontières prosodiques grâce à la détection de pauses. En effet, la présence d'une pause en français est fréquemment corrélée à une fin de groupe intonatif en français qui est également marquée par un fort allongement vocalique.

Notre principal but était de détecter un allongement vocalique qui pouvait se produire dans l'entourage de frontières syntaxiques et/ou prosodiques. Comme cela est illustré par les figure 4a et 4b (en Annexes), les voyelles précédées ou suivies d'une pause sont caractérisées par des valeurs plus extrêmes, occupant ainsi un espace acoustique plus large, comme nous l'avons noté pour les voyelles les plus longues sur la figure 1. Ces résultats seront plus longuement développés dans la partie suivante.

5. DISCUSSION ET CONCLUSION

Le principal résultat de cette étude est que l'allongement final de mot, ainsi que le contexte immédiat (précédent ou suivant) d'une pause induisent des voyelles plus périphériques selon l'axe F1/F2. Il est intéressant de tenter de relier ces résultats à des travaux récents de prosodie articulatoire. En effet, une question fondamentale de cet axe de recherche est l'analyse supraglottique des voyelles en termes des hypothèses de l'expansion de la sonorité (Straka [9]; Beckman et al. [1]) ou du renforcement des traits distinctifs (de Jong [8]). La seconde hypothèse peut être différenciée de la première sur la base de mesures de formants des voyelles fermées telles que /i/ et /u/. Selon la première hypothèse, un voyelle antérieure fermée /i/ « renforcée » sera plus ouverte (de par la mandibule et/ou les lèvres) avec des valeurs plus élevées du 1^{er} formant, bien que cette ouverture du conduit vocal soit en partie contradictoire avec le trait fermé du /i/. Pour la seconde hypothèse, la voyelle antérieure fermée /i/ « renforcée » sera encore plus fermée (F1 plus bas), plus antérieure et plus étirée (F2 et si plus antérieur, F3 plus élevé, une fréquence plus élevée du 3^{ème} formant semblant être une caractéristique d'un /i/ bien formé en français).

Fougeron [3] a montré que les phonèmes en début de constituant prosodique de haut niveau tel que le groupe intonatif (GI) tendent à être hyperarticulés, ce que l'on pourrait relier ici à l'hypothèse de renforcement des traits distinctifs et que nous observons pour notre corpus (fig. 4a) à l'exception de la voyelle /i/ cependant. Nous avons pu observer cependant qu'une grande proportion (80%) de ces /i/ sont représentés par le pronom « il », qui est plus vraisemblablement réalisé avec une aperture plus grande dans cette position. Quant aux voyelles suivies d'une pause, Tabain [10] a montré que les voyelles finales (principalement /a/) de constituants prosodiques de haut niveau tels que le groupe intonatif tendaient également à être renforcées, ce que nous observons ici sur

la figure 4b. Lorsqu'une voyelle est suivie d'une pause, ses valeurs formantiques sont plus extrêmes sur l'axe F1/F2. A nouveau, ces variations sont plus faibles pour les voyelles antérieures fermées.

De par la faible quantité de voyelles recueillies au contact immédiat d'une pause, nous ne pouvons découper ces voyelles en catégories de durée comme réalisé précédemment. Selon le deuxième résultat de cette étude, les voyelles en syllabe finale et initiale de mot révèlent des variations de durée induisant des variations formantiques similaires à celles observées sur l'ensemble des voyelles, malgré les différences que nous avons signalées. Des mesures effectuées sur un corpus plus conséquent permettront probablement de mesurer des variations formantiques significatives en fonction de la durée dans l'entourage d'une pause.

BIBLIOGRAPHIE

- [1] Beckman M. E., Edwards J. et Fletcher J., 1992, Prosodic Structure and Tempo in a Sonority Model of Articulatory Dynamics, in Papers in lab. phon. II. Cambridge (England), Cambridge University Press, pp.68-89.
- [2] Delattre, P. (1962) Comparing the prosodic features in English, German, Spanish, and French. *Int. Rev. Applied. Linguistics* I, 193-210. volume 4, pages 2379-2382, Philadelphia, 3-6 October 1996.
- [3] Fougeron C., 2001, Articulatory properties of initial segments in several prosodic constituents in French, *Journal of Phonetics*, 29(2), pp. 109-135.
- [4] Gauvain, J.L., Lamel, L. and Adda, G. (2002) The Limsi Broadcast News Transcription System, *Speech Communication*, 37(1-2):89-108.
- [6] 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.. *Eurospeech – Lisbon (Portugal)*, September 2005, pp2453-2456.
- [7] Gendrot, C., (2005). Aspects perceptifs, physiologiques et acoustiques de différentes catégories prosodiques en français, Thèse de doctorat, Université de Paris3-Sorbonne Nouvelle.
- [8] de Jong K. J., 1995, The supraglottal articulation of prominence in English, *JASA*, 97, pp. 491-504.
- [9] Straka G., 1963, La division des sons du langage en voyelles et consonnes peut-elle être justifiée?, *Travaux de Linguistique et de littérature*, Université de Strasbourg, 1, pp. 17-99.
- [1] Tabain M., 2003, Effects of prosodic boundary on /aC/ sequences: acoustic results, *Journal of the Acoustical Society of America*, 113, pp. 516-531.

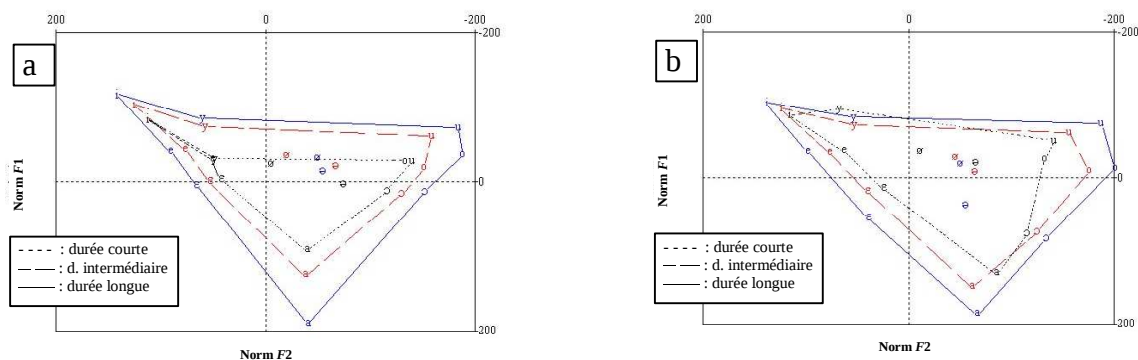


Figure 3ab : Valeurs moyennes de F1 et F2 pour les voyelles en fonction de leur durée (Normalisation lobanov). a(gauche) : voyelles en syllabe initiale de mots b(droite) : voyelles en syllabe finale de mots.

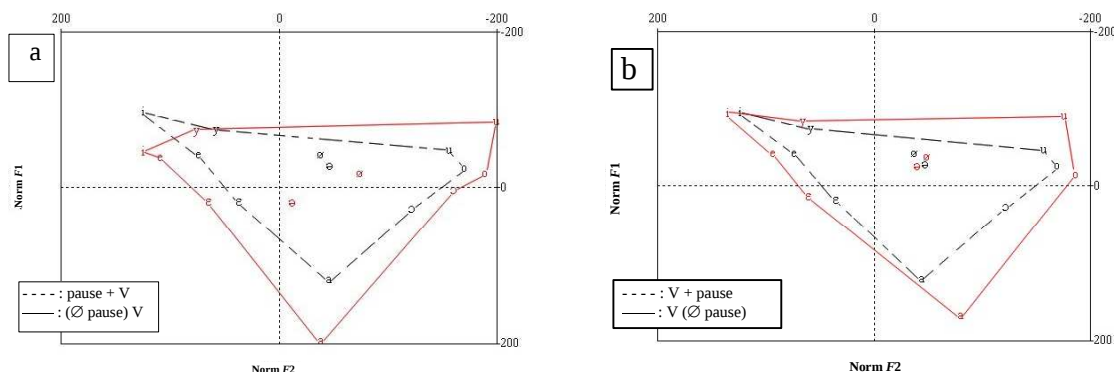


Figure 4ab: Valeurs moyennes de F1 et F2 pour les voyelles en fonction de l'absence/présence de la pause à proximité de la voyelle analysée. a(gauche) : pause précédant la voyelle ; b(droite) : pause suivant la voyelle.