

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 of (before and after) pauses. The LIMSI speech alignment system is used [1] 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 [2]). We show that at some extent, a similar 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 étude (Gendrot et Adda-Decker [2]) 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 est 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. Figure 1). Ces variations sont 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.

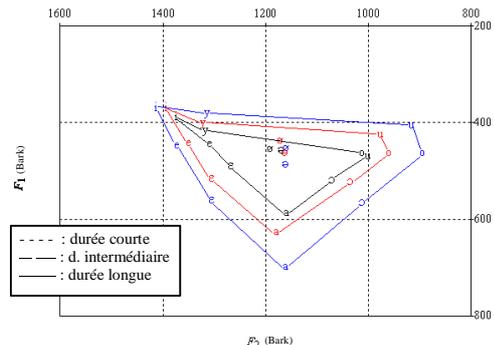


Figure 1 : Valeurs moyennes de F1 et F2 pour les voyelles orales du français en fonction de leur durée (normalisation en Bark).

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 [2] avaient conclu en comparant le français et l'allemand d'après un protocole expérimental identique à celui-ci, que des réductions vocaliques importantes pouvaient être mesurées en fonction de la durée à la fois en français et en allemand (cf. Figure 1). Le français n'étant pas une

langue à accent lexical, et typiquement considéré comme une langue « syllable-timed » avec un allongement final en fin de groupes de sens (ou groupes rythmiques), des phénomènes de réduction vocalique plus faibles pouvaient être attendus, tel que cela était suggéré par Delattre [3]. Nous analyserons séparément dans cette étude des positions prosodiques initiales et finales en français afin de vérifier si des variations de durée (réparties sur les trois catégories équilibrées de durée établies dans Gendrot & Adda-Decker [2]) impliquent également des variations formantiques, et de même ordre.

Deux facteurs pourraient expliquer un allongement vocalique en français : (i) La présence d'accents de focalisation/emphase qui, en français, se positionnent plus vraisemblablement sur les syllabes initiales de mots et qui impliquent – en plus d'une intensité accrue et d'une f0 augmentée – un allongement vocalique. (ii) La proximité de frontières est une cause fréquente d'allongements vocaliques et peut également être corrélée à la présence de pauses.

Dans ce présent travail, nous nous intéresserons au deuxième facteur de variation : dans un premier temps, les voyelles de syllabes finales de mots seront analysées puisque potentiellement plus longues, et comparées aux voyelles en syllabes initiales de mots (généralement non allongées, sauf en cas d'insistance). Nous séparerons également les valeurs en fonction des trois catégories de durée précédemment établies. Dans un deuxième temps, nous effectuerons les mêmes analyses formantiques du point de vue de la présence/absence de pauses dans l'entourage vocalique. Nous discuterons enfin de ces résultats en termes de prosodie articulatoire.

3. VARIATIONS FORMANTIQUES EN SYLLABE INITIALE/FINALE DE MOT

Nous avons ainsi séparé les mesures de formants en fonction de la position de la voyelle orale dans le mot en considérant les syllabes initiales (5200 occurrences) et finales (5300) de mots polysyllabiques. Pour ce faire, nous avons eu recours à la transcription à la fois orthographique et phonémique, en considérant comme règle de base que chaque voyelle alignée (les semi-voyelles étant donc exclues) correspond à une syllabe.

La figure 2 révèle que les voyelles en syllabes finales de mot ont des valeurs plus extrêmes que les syllabes initiales de mot sur les axes F1/F2, à l'exception des voyelles antérieures fermées /i/ et /e/. Les tables 1 et 2 détaillent la répartition des voyelles en fonction de leur position dans le mot, et en fonction de leur catégorie de durée.

Dans Gendrot [4], il a été suggéré que les procédures de normalisation telles que Lobanov, Nearey et Bark ne modifient sensiblement pas les résultats présentés ici de par une quantité importante et équilibrée de données. Une normalisation (Lobanov) sera toutefois utilisée pour

faciliter la présentation des résultats cumulés hommes/femmes ici.

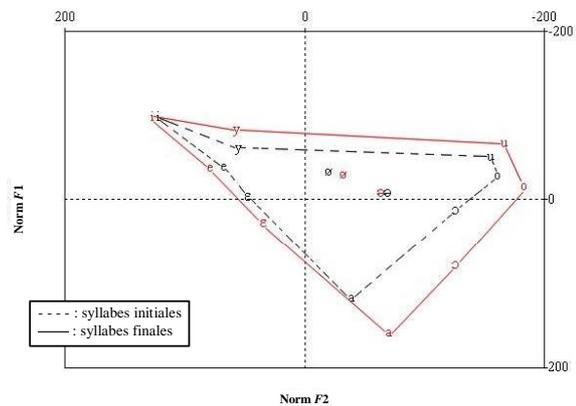


Figure 2 : Valeurs moyennes de F1 et F2 pour les voyelles en syllabe initiale et finale (Normalisation Lobanov, toutes durées confondues).

Afin de mieux interpréter ces résultats, nous avons également illustré une mise en regard de la distribution des voyelles en fonction de leurs valeurs de f0. Des catégories de f0 équilibrées (uniformes) ont ainsi été déterminées sur le même principe que les catégories de durée :

	f0 basse	f0 moyenne	f0 haute
hommes	≤ 110 Hz	$110 < f0 \leq 140$	> 140 Hz
femmes	≤ 160 Hz	$160 < f0 \leq 210$	> 210 Hz

Nous pouvons observer sur ces tables que les voyelles longues sont également celles qui ont des valeurs de f0 plus élevées, plus particulièrement en syllabe finale de mots¹. Les voyelles longues sont également plus fréquentes lorsque positionnées en syllabe finale de mots.

Table 1 : répartition (en %) des voyelles en syllabe initiale de mots en fonction de leurs catégories de durée et de f0.

f0 \ durée	bas	moyen	haut	total (durée %)
court	14	14	9	37
moyen	11	14	14	39
long	5	7	12	24
total(f0 %)	30	35	34	100

Table 2 : répartition (en %) des voyelles en syllabe finale de mots en fonction de leurs catégories de durée et de f0.

f0 \ durée	bas	moyen	haut	total (durée %)
court	10	11,5	8,5	30
moyen	8	11	14	33
long	7,5	10	19	37
total(f0 %)	25	33	42	100

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

¹ 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.

différentes catégories de durées utilisées précédemment dans Gendrot et Adda-Decker [2], 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 [4] 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 en syllabe finale de mot, les variations formantiques pour ces voyelles sur l'axe F1 sont non significatives ou peu importantes. A l'inverse en syllabe initiale, un net détachement du 1^{er} formant des voyelles fermées courtes (avec des valeurs plus élevées) par rapport à leurs contreparties plus longues peut être observé. 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 allongeante 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 [2].

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-Decker [2]). 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 (dur moy = 100ms) sans:22000 (dur moy=70ms)	avec :1600 (dur moy=120ms) sans:21300 (dur moy=68ms)

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 déterminer 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 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

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 supra-glottique des voyelles en termes des hypothèses de l'expansion de la sonorité (Straka [5]; Beckman et al. [6]) ou du renforcement des traits distinctifs (de Jong [7]). 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, une 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 [8] 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 ouverture plus grande dans cette position. Quant aux voyelles suivies d'une pause, Tabain [9] 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 les axes F1/F2. A nouveau, ces variations sont plus faibles pour les voyelles antérieures fermées.

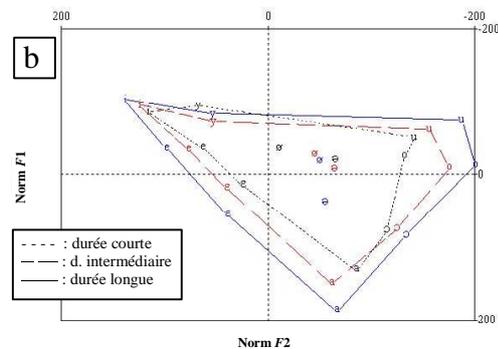
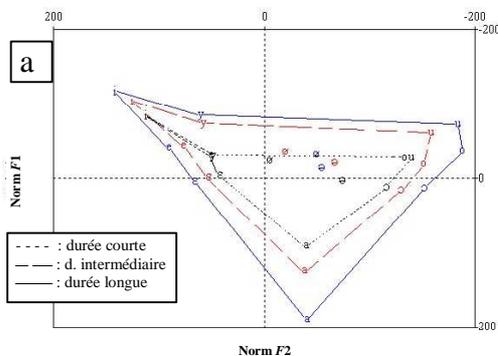
De par la faible quantité de voyelles recueillies au contact immédiat d'une pause, nous ne pouvons découper ces voyelles en catégories de durée comme réalisé

précédemment. Malgré tout, des mesures effectuées sur le corpus ESTER ont permis de mesurer des variations formantiques significatives en fonction de la durée dans l'entourage d'une pause, du même ordre que celles observées pour l'ensemble des voyelles.

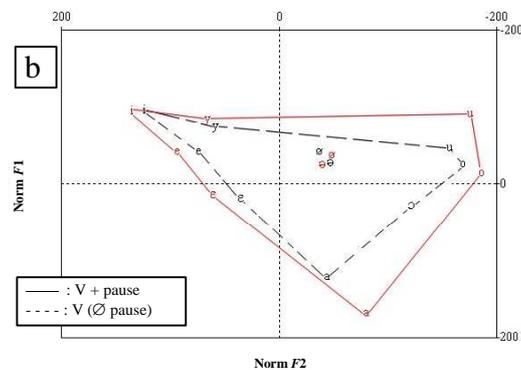
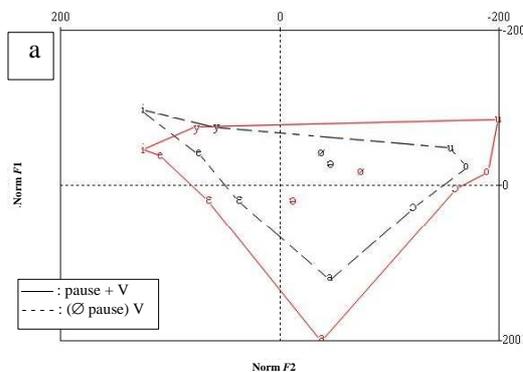
Le résultat principal de cette étude est la mise en évidence de la périphéralité des voyelles françaises dans deux positions prosodiques: (i) en syllabe finale de mot (comparées aux voyelles en syllabe initiale de mot (ii) dans l'entourage immédiat (avant et après) de la pause. Nous avons également montré que 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.

BIBLIOGRAPHIE

- [1] Gauvain, J.L., Lamel, L. & Adda, G. The Limsi Broadcast News Transcription System, *Speech Communication*, 37(1-2):89-108, 2002.
- [2] Gendrot, C. & Adda-Decker, M. 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, 2005.
- [3] Delattre, P. Comparing the prosodic features in English, German, Spanish, and French. *Int. Rev. Applied. Linguistics I*, 193-210. volume 4, pages 2379-2382, 1965.
- [4] Gendrot, C. Aspects perceptifs, physiologiques et acoustiques de différentes catégories prosodiques en français, Thèse de doctorat, 2005. Université de Paris3-Sorbonne Nouvelle.
- [5] Straka G. 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, 1963.
- [6] Beckman M. E., Edwards J. & Fletcher J. 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, 1992.
- [7] de Jong K. J. The supraglottal articulation of prominence in English, *JASA*, 97, pp.491-504, 1995.
- [8] Fougeron C. Articulatory properties of initial segments in several prosodic constituents in French, *Journal of Phonetics*, 29(2), pp. 109-135, 2001.
- [9] Tabain M. Effects of prosodic boundary on /aC/ sequences: acoustic results, *JASA*, 113, pp. 516-531, 2003.



Figures 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.



Figures 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.