Sensibilité au débit et marquage accentuel des phonèmes en français

Valérie Pasdeloup°, Robert Espesser° & Malika Faraj*

*Université de Rennes 2
°Laboratoire Parole et Langage, UMR 6057 CNRS, Université de Provence, France
valeriepasde@yahoo.fr, robert.espesser@lpl.univ-aix.fr, malika.faraj@free.fr

ABSTRACT

The aim of this work is to determine the way the prosodic scene reorganises itself according to speech rate variations in French. We present the temporal structure study of a one thousand word speech corpus. The corpus was produced at three different rates (normal, fast and slow) by one speaker with two repetitions. The goal is to study the relationship between speech rate sensitivity and accentual markedness of phoneme. Results put in light that phoneme does not behave the same way if stressed or not and if consonantic or vocalic. Unstressed phonemes are less rate sensitive than stressed ones. Vowels are more rate sensitive than consonants, especially when stressed. Nevertheless, consonants are rate sensitive in such proportions when stressed that it is not possible to say, as usually said, that it is the vowel which carries stress.

1. INTRODUCTION

1.1. L'organisation de la scène prosodique

Le décalage entre le percept d'une forme visuelle et le stimulus sur lequel il se base est un phénomène bien connu dans les illusions visuelles étudiées par la théorie de la Gestalt. Ce décalage illustre un des aspects génératifs de la perception. On peut expérimenter un phénomène similaire dans la perception du flux rythmique d'un texte lu à différents débits de parole. Par exemple, à débit lent on a l'impression subjective que l'ensemble du flux rythmique (c'est-à-dire toutes les syllabes) est produit plus lentement qu'à débit normal, alors qu'en fait toutes les syllabes ne sont pas également sensibles aux variations de débit : les syllabes inaccentuées sont moins sensibles au débit que les syllabes accentuées [14, 15].

Cette étude de la scène prosodique se situe dans le cadre de la théorie de la *Gestalt*. Cette théorie qui a beaucoup été appliquée à la perception des sons [8, 9] et à celle de la musique [3, 4, 5] a eu peu d'écho dans l'étude de la prosodie [12, 2, 1, 15]. Une des lois fondamentales de la théorie de la forme est que nous percevons des *figures*, des *formes*, qui se dégagent d'un *fond* en vision et en audition. Ces formes correspondent à des « objets sensibles » qui possèdent des caractéristiques spécifiques afin de pouvoir ainsi émerger d'un fond [10]. Les figures n'ont aucune existence autonome puisqu'elles n'existent qu'en relation avec un fond. La figure a comme caractéristique fonctionnelle de posséder une forme et une organisation alors que le fond est une continuité amorphe, indéfinie qui n'a pas de contours propres. En parole, dans l'optique de la théorie de la Gestalt, les suites de syllabes non-accentuées constituent le fond de la scène prosodique et les syllabes accentuées constituent les figures qui *émergent* de ce fond. Ainsi, une suite de syllabes constitue en tant que suite amorphe et indéfinie le fond de la scène prosodique, tant que n'émerge pas un accent qui donne forme à une syllabe particulière. Il n'y a donc pas à proprement parler de syllabes inaccentuées, mais des syllabes qui reçoivent ou non un accent.

1.2. Objectifs

La plupart des recherches sur l'influence du débit de parole est consacrée à l'étude des unités segmentales. Peu de travaux ont été consacrés en français aux effets du débit de parole sur l'organisation prosodique [6, 7, 18].

Cette recherche prend place dans un projet plus large sur les gabarits rythmiques et la pulsation accentuelle en français. Le but est de contraindre la structuration rythmique de textes lus en manipulant le débit de parole afin d'observer les contraintes qui opèrent sur la production de la pulsation accentuelle et le formatage des gabarits rythmiques. Dans une étude précédente [14, 15], nous avons mis en évidence que les variations de débit n'ont pas la même influence dans la scène prosodique sur les syllabes selon qu'elles jouent le rôle de forme ou de fond : le fond de syllabes inaccentuées est moins élastique que les formes (les syllabes accentuées) qui en émergent. L'objectif du présent travail est d'étudier comment ce phénomène opère à l'intérieur de la syllabe sur la voyelle et la consonne. Les voyelles sont-elles plus sensibles au débit que les consonnes ? Nous présentons ici les résultats relatifs à l'étude de la structuration temporelle des phonèmes d'un corpus lu de mille mots dans trois conditions de débit de parole par une locutrice.

2. MÉTHODOLOGIE EXPÉRIMENTALE

2.1. Corpus

Le corpus est un conte d'environ 1000 mots, lu en chambre sourde, dans trois conditions de débit (normal, rapide et lent), par une locutrice (la 1ère auteure), avec deux répétitions. La meilleure répétition a été ensuite retenue à chaque débit. Les corpus lus correspondent à environ 1200 syllabes pour chaque condition de débit. Pour les trois débits, on totalise 8081 phonèmes : 2660 à débit rapide, 2698 à débit normal et 2723 à débit lent. Parmi ces 8081 phonèmes, on décompte : 4085 consonnes, 3527 voyelles (dont 423 schwas constituant

un noyau vocalique), 108 schwas extra-métriques (ne constituant pas le noyau vocalique d'une syllabe, en général devant pause) et 361 semi-voyelles.

2.2. Analyse expérimentale

En parole, la composante acoustique du rythme correspond selon nous à tout stimulus acoustique qui seul ou en interaction avec d'autres permet de produire un percept rythmique : contrastes mélodiques, de durée, d'intensité et de timbre. L'étude de la structuration rythmique des énoncés du corpus inclut par conséquent celle de la prosodie (intonation et accentuation) : l'analyse phonétique des paramètres prosodiques (F_0, durée syllabique, phonémique et pauses principalement) et leur interprétation phonologique afin de déterminer une structure rythmique abstraite dans le cadre d'un modèle théorique donné. La représentation phonologique correspond à l'accentuation et aux groupements rythmiques. Notre modèle rythmique distingue quatre niveaux prosodiques [12, 13] :

- la syllabe qui constitue l'unité rythmique minimale et qui peut être accentuée ou non-accentuée (les syllabes accentuées sont indiquées en gras et les limites entre les syllabes par des tirets) ;
- le groupe accentuel qui est le groupement rythmique minimal (indiqué par les symboles < >) ; il est constitué d'une syllabe accentuée précédée généralement d'une ou de quelques syllabes non-accentuées et est soumis à des contraintes de taille ;
- le mot rythmique (indiqué par les symboles [ ]) qui est la plus petite structure prosodique qui organise un groupe de sens (petit groupe syntactico-sémantique) [17] ; il est constitué généralement d'un ou de deux groupes accentuels et est soumis à des contraintes de taille ;
- la séquence rythmique (indiquée par les symboles / /) qui est une structure prosodique de niveau hiérarchique supérieur au mot rythmique qui organise une unité discursive ; elle est constituée en général de plusieurs mots rythmiques mais ne semble pas soumise à des contraintes de taille.

(1) Le rythme$_{2syll}$ d'la parole$_{3syll}$ n'est pas$_{2syll}$ élastique$_{3syll}$ =>

/ [<lə - **ritm**> <dla - pa - **rɔl**>] [<ne - **pa**> <e - las - **stik**>] /

(2) Le rhinocéros$_{5syll}$ de Constantinople$_{5syll}$ n'est pas$_{2syll}$ élastique$_{3syll}$ =>

/ [<lə - **ri**> <no - se - **rɔs**>] [<də -**kɔ̃**> <stɑ - ti - **nɔpl**>] [<ne - **pa**> <e - las - **stik**>] /

Dans notre modèle, les règles phonologiques d'accentuation et d'intonation sont basées sur des contraintes linguistiques au sens strict (morpho-syntaxiques et lexicales) et rythmiques (nombre de syllabes des unités lexicales, des constituants morpho-syntaxiques, des groupes accentuels et des mots rythmiques). Ainsi les énoncés (1) et (2) ci-dessus qui ont la même structure syntaxique mais dont les constituants syntaxiques sont composés d'un nombre différent de syllabes n'auront pas nécessairement la même structure prosodique (pour plus de détails cf. [13]) :

Dans un premier temps, l'étiquetage phonétique des énoncés et leur segmentation phonémique sont effectués avec le logiciel développé au LORIA (D. Foher & Y. Laprie : http://www.loria.fr/equipes/parole/) puis corrigés manuellement. Le logiciel code identiquement les voyelles orales et nasales à double timbre correspondant aux archiphonèmes : /E Œ O A E~/. On obtient ainsi 11 types de consonnes et 17 types de voyelles, en excluant les schwas extra-métriques et les semi-voyelles qui n'ont pas été pris en compte dans l'analyse des phonèmes. La syllabation est effectuée avec un script sous Praat et corrigée manuellement. Dans un second temps, l'analyse phonétique des paramètres prosodiques est réalisée. Enfin, ces données sont interprétées dans le cadre de notre modèle prosodique ce qui permet de déterminer le caractère accentué ou non accentué des syllabes. Dans la présente étude, seule l'interprétation de l'accent, c'est-à-dire le statut accentué ou inaccentué des phonèmes, a été pris en compte. Pour une étude de l'influence du débit sur les pauses et les syllabes dans ce corpus cf. [15].

3. RÉSULTATS

3.1. Taux d'articulation et durée des phonèmes

Le taux d'articulation est de 15.31 phonèmes/s à débit rapide (R), 12.33 phon/s à débit normal (N) et 9.88 phon/s à débit lent (L) (hors pauses ; schwas extra-métriques et semi-voyelles inclus). Pour les syllabes, le taux d'articulation est de 6.8 syll/s à débit R, 5.4 syllabes à débit N et 4.4 syll/s à débit L.

Table 1 : Moyenne des durées syllabiques des Consonnes et Voyelles Inaccentuées et Acc. dans les trois débits

	débit Rapide	débit Normal	débit Lent	moyenne
C I	62.06ms ±25	68.41ms ±28	79.57ms ±33	70.01
C A	79.06ms ±34	95.35ms ±41	119.06ms ±51	97.82
V I	57.83ms ±18	68.08ms ±21	80.55ms ±26	68.82
V A	68.99ms ±24	104.21ms ±39	144.62ms ±75	105.94

La Table 1 fait apparaître que les phonèmes accentuées (A) sont plus sensibles au débit que les phonèmes inaccentuées (I). Ce phénomène semble plus manifeste chez les voyelles (V) accentuées qui sont plus sensibles au débit que les consonnes (C) accentuées : comparée au débit N, la durée des VA varie en moyenne de –34% à débit R et de +39% à débit L, alors que celle des CA varie en moyenne de –17% à débit R et de + 25% à débit L. Les CI et VI manifestent une insensibilité au débit très proche : 18ms de différence moyenne entre les deux conditions extrêmes de débit R et L pour les CI contre 23ms pour les VI. Les contrastes accentuels de durée (A-I/I%) se renforcent quand le débit ralentit de façon plus marquée chez les V (de 19% à débit R à 80% à débit L) que chez

les C (de 27% à débit R à 50% à débit L). Les contrastes sont plus forts chez les V que chez les C à débit N et L, mais pas à débit R (27% pour les C contre 19% chez les V). De plus, dans chaque débit, la durée moyenne des V est très proche de celle des C, et ce surtout chez les inaccentuées (à débit R : CI ≥ VI ; à débit N et L : CI ≅ VI).

3.2. Traitement statistique

La durée des phonèmes a été évaluée en fonction de trois facteurs : *Débit*, facteur ordonné à 3 niveaux (Rapide, Normal, Lent) ; *Accent*, facteur à 2 niveaux (Inaccentué, Accentué) ; *CLasse*, facteur à 2 niveaux (Consonne, Voyelle). Un modèle linéaire mixte, où le phonème est le facteur de groupement, a permis de traiter la répétition des 28 groupes non équilibrés de phonèmes (17 consonnes et 11 voyelles) ([16], http://www.R-project.org/). Ainsi, les variations inter-phonémiques de durée ont été neutralisées. De plus, l'utilisation du logarithme de la durée a stabilisé la variance. Ce premier modèle ayant montré que seules les composantes linéaires du débit sont significatives, le facteur débit a été considéré ensuite comme une variable numérique classique, ce qui simplifie le modèle. A chaque débit a été associée la durée totale correspondante du corpus lu (hors pauses) : 175s pour le débit R, 225s pour le débit N et 275s pour le débit L. La variable *Débit* a été centrée sur le débit R afin de tester certaines hypothèses spécifiques à ce débit.

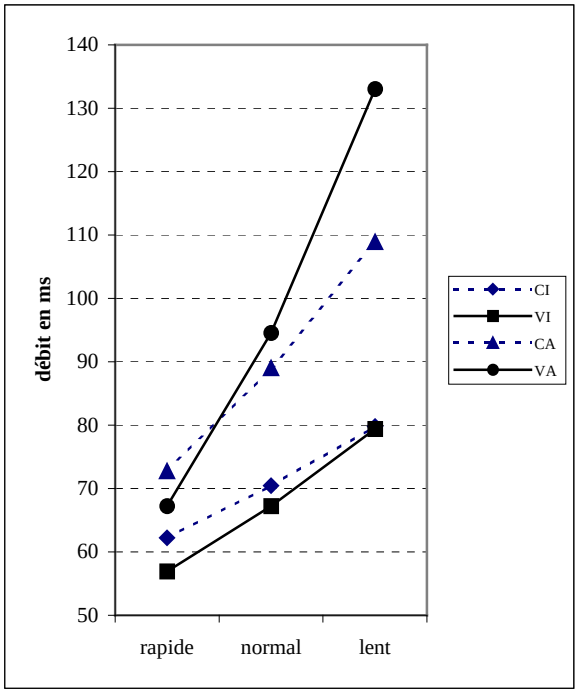

Figure 1 : Durées estimées des consonnes et des voyelles par le modèle mixte dans les trois conditions de débit

La table 2 des régresseurs du modèle montre que tous les coefficients d'interaction avec le débit sont significatifs (S) et positifs : *DébitC:AccentA*, *DébitC:CLasseV*, *DébitC:AccentA:ClasseV*. On a donc quatre droites de régression distinctes CI, VI, CA et VA (Fig. 1). La significativité des autres coefficients *AccentA* (S), *ClasseV* (Non S) et *AccentA:ClasseV* (NS) n'a de valeur que pour le débit R.

Table 2 : Régresseurs du modèle mixte centré

	Value	Std.Error	DF	t-value	p-value
(Intercept)	4.1308	0.0612	7577	67.47	0.0000
Débit Centré	0.0025	0.0002	7577	14.54	0.0000
AccentA	0.1556	0.0164	7577	9.48	0.0000
ClasseV	-0.0894	0.0973	26	-0.92	0.3667
DC:AccentA	0.0016	0.0002	7577	6.27	0.0000
DC:CLasseV	0.0008	0.0002	7577	3.47	0.0005
AccentA:CLV	0.0108	0.0244	7577	0.44	0.6583
DC:AccA:CLV	0.0019	0.0004	7577	5.30	0.0000

3.3. Interprétation

L'interaction significative *DébitC:AccentA* montre que les phonèmes inaccentués sont moins sensibles au débit que les phonèmes accentués (cf. Table 1 et Fig. 1). Ces résultats sont comparables à ceux trouvés par Duez [6]. En français, les variations de débit n'ont donc pas le même effet sur les phonèmes selon qu'ils jouent le rôle de fond ou de forme dans la scène prosodique. L'interaction *DébitC:CLasseV* montre que l'effet du débit est plus marqué pour les V que pour les C. La moins grande élasticité des consonnes a été observée dans d'autres travaux [6, 11]. La double interaction *DébitC:AccentA:CLasseV* précise que l'accentuation renforce la sensibilité au débit des V. Par conséquent, les V sont plus sensibles au débit que les C, et ce surtout chez les accentuées et de façon très peu marquée mais significative chez les inaccentuées.

Selon notre hypothèse, cette différence de sensibilité au débit des C et des V serait liée à des *contraintes de matière* (contraintes motrices de contrôle articulatoire et contraintes sensori-motrices proprioceptives et auditives) et ne serait donc pas au sens strict phonologique, c'est-à-dire liée à des *contraintes formelles*. En revanche, la différence de sensibilité au débit des phonèmes accentués et inaccentués serait phonologique, puisqu'elle résulterait de la structuration formelle de la scène prosodique.

Par ailleurs, la plus faible sensibilité des C au débit a pour conséquence qu'à débit R les V ne se distinguent plus des C à la fois chez les inaccentuées et chez les accentuées (*AccentA:ClasseV* : NS). En extrapolant à débit encore plus rapide, on peut émettre l'hypothèse que la distinction entre les C et les V deviendrait significative, les C devenant alors plus longues que les V.

Enfin, du point de vue des contrastes accentuels de durée entre les phonèmes inaccentués et accentués (écart A-I), les contrastes se renforcent quand le débit ralentit et sont significativement plus marqués chez les V que chez les C à débit N et L. A débit R, le contraste accentuel subsiste (*AccentA* : S) mais ne diffère plus significativement entre les V et les C (*AccentA:ClasseV* : NS). En extrapolant à débit encore plus rapide, on peut supposer que ce contraste deviendrait significatif, le contraste accentuel de durée chez les C devenant plus marqué que chez les V.

4. CONCLUSION

Dans cette étude, nous avons montré qu'en français la consonne comme la voyelle est plus sensible au débit quand elle est accentuée qu'inaccentuée. Ce phénomène de plus grande sensibilité au débit, que nous avions déjà observé pour la syllabe accentuée [14, 15], opère donc sur les deux constituants syllabiques. La consonne est cependant moins sensible au débit que la voyelle, et ce surtout chez les accentuées. En ce qui concerne la programmation motrice, nous émettons l'hypothèse que seule la durée des phonèmes accentués serait planifiée (plus précisément le contraste temporel). La durée des phonèmes non-accentués ne serait pas planifiée. Par conséquent, la grande sensibilité des phonèmes accentués au débit correspondrait à des variations de haut niveau dans le système (commandes motrices), alors que la très faible sensibilité des phonèmes inaccentués correspondrait à des variations de bas niveau.

Dans la scène prosodique, les objets consonnes et voyelles - bien que différemment contraints dans leur matérialité - seraient phonologiquement identiques pour constituer soit des formes en émergeant, soit le fond en restant amorphes et indéfinis. Les consonnes et les voyelles inaccentuées, du fait de leur relative insensibilité au débit, participeraient conjointement à l'illusion perceptive du débit : l'auditeur a l'impression subjective que le fond de la scène prosodique accélère ou décélère selon les différents débits, alors qu'en fait c'est principalement la durée de présentation des formes qui se modifie sur un fond relativement stable. Quand le débit s'accélère, la durée de présentation des phonèmes accentués diminue. Quand le débit ralentit, la durée de présentation des phonèmes accentués s'allonge.

En conclusion, sur le plan de la substance, ce n'est uniquement la voyelle qui "porte l'accent" dans la syllabe. La consonne a une contribution non négligeable dans la réalisation du contraste temporel entre les syllabes inaccentuées et accentuées. Du fait de la moins grande sensibilité de la consonne au débit, cette contribution se renforce quand le débit s'accélère. Enfin, en ce qui concerne la relation entre le marquage phonologique suprasegmental et la variation phonétique, on observe que les phonèmes marqués par l'accent sont plus sensibles à la variation que les phonèmes non-marqués.

REMERCIEMENT : Nous souhaitons remercier Daniel Hirst pour ses conseils ainsi que pour la réalisation de nombreux scripts sur Praat.

BIBLIOGRAPHIE

[1] C. Astésano. *Rythme et Accentuation en Français : Invariance et Variabilité stylistique*, Paris, L'Harmattan, 2001.

[2] E. Couper-Kuhlen. English Speech Rhythm : *Form and function in everyday verbal interaction, Amsterdam*, John Benjamins Publishing Company, 1993.

[3] D. Deutsch. Grouping mechanism in music, in Deutsch D. (ed.), *The psychology of music*, New York, Academic Press, 99-130, 1982.

[4] C. Drake. *Processus cognitifs impliqués dans l'organisation du rythme musical*, Thèse doctorale, Université René-Descartes, Paris , 1990.

[5] C. Drake. Reproduction of musical rhythms by children, adult musicians and adult nonmusicians, *Perception & Psychophysics*, 53 (1), 25-33, 1993.

[6] D. Duez. *Contribution à l'étude de la structuration temporelle de la parole en français*, Thèse de Doctorat d'Etat, Université de Provence, Aix-Marseille 1, 1987.

[7] C. Fougeron & S.-A. Jun. Rate effects on French intonation: prosodic organization and phonetic realization, *Journal of Phonetics*, 26, 45-69, Academic Press, 1998.

[8] P. Fraisse. Les structures rythmiques, *Studia Psychologica*, Publications Universitaires de Louvain, 124p.,1956.

[9] P. Fraisse. *Psychologie du rythme*, Paris, Presses Universitaires de France, 360p., 1974.

[10] P. Guillaume *La psychologie de la forme*, (1937), Paris, Flammarion, 1979.

[11] A. Kozhevnikov & L. A. Chistovich C. Speech articulation and perception, in *Joint Publications Research Service*, 543p., 1965.

[12] V. Pasdeloup. *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*, Thèse de doctorat, Université de Provence Aix-Marseille 1, 1990.

[13] V. Pasdeloup. A prosodic model for French text-to-speech synthesis: A psycholinguistic approach, in Bailly G., Benoît C. & Sawallis T.R. (eds.), *Talking Machines: Theories, Models and Designs*, Elsevier Science Publisher, 335-348, 1992.

[14] V. Pasdeloup. Figures et fond dans la scène prosodique : leur résistance face aux variations du débit de parole, *1er Symposium International « Interface Discours-Prosodie »*, 8-9 sept. 2005.

[15] V. Pasdeloup, R. Espesser & M. Faraj. Rate sensitivity of syllable in French: a perceptual illusion? *3rd International Conference on « Speech Prosody »* , Dresde, 2-5 mai 2006

[16] J. C. Pinhero & D. M. Bates. *Mixed-Effects Models in S and S-Plus*, Springer, 2001.

[17] J. Vaissière. "La structuration acoustique de la phrase française", *Annali della Scuola Normale Superiore di Pisa*, 3(10), 529-560, 1980.

[18] B. Zellner. *Caractérisation et prédiction du débit de parole en français*, Doctorat, Lausanne, 1998.