

Le rythme n'est pas élastique : étude préliminaire de l'influence du débit de parole sur la structuration temporelle

Valérie Pasdeloup

Département de Lettres, Université de Rennes 2
6 avenue Gaston Berger CS 24 307, F-35043 Rennes, France
valeriepasde@yahoo.fr

ABSTRACT

We present a preliminary study of temporal structure of reading speech (one speaker) at three different rates (normal, fast and slow). The goal is to constrain rhythmic structure of speech in order to observe how rhythmic patterns depend on speech rate's variations. First results put in light that rhythmic changes linked to speech rate variations do not correspond to a linear processes. At slow rate, temporal structure is not the result of a general decrease of the same material pronounced at fast rate. Contrasts of duration between stressed syllables and unstressed syllables are less marked at fast rate than slow rate. Variation in syllabic duration according to rate changing is more important for stressed syllables ; this last result supports the hypothesis that unstressed syllable is an anchored point in French rhythmic structure.

1. INTRODUCTION

1.1. Problématique du rythme

Mais qu'est-ce que le rythme ?

Notre cadre théorique pour l'étude du rythme dans la parole s'inspire principalement de la théorie des psychologues de la Gestalt ([17], [23], [15]), des travaux de recherche en psychoacoustique sur les rythmes humains ([5], [6], [4], [9], [10], [11], [12], [24]), de la théorie de l'analyse de la scène auditive [2], des théories cognitives en psycholinguistique ainsi que des approches rythmiques de la prosodie (entre autres : [18], [19], [20], [21], [3], [16], [1]) et des travaux sur la structuration temporelle (entre autres : [13], [14], [7]).

Les rythmes humains correspondent à la capacité de l'homme à structurer ses productions motrices et ses perceptions sensorielles. La répétition et la structure caractérisent les rythmes moteurs. Toute production motrice est objectivement rythmée. Toute information perçue est objectivement ou subjectivement rythmée. Ce que nous percevons est un compromis entre l'input sensoriel et le schème spontané, c'est-à-dire une *bonne forme* au sens défini par la théorie de la Gestalt. La prégnance d'une forme doit être importante afin que sa perception puisse correspondre à ses caractéristiques objectives.

Le rythme dans la parole est conditionné à la fois sur le plan de la production et de la perception par la capacité de

l'homme à donner forme à l'information langagière. La production et la perception du rythme dans la parole requièrent des structures rythmiques qui correspondent à des formes prégnantes, c'est-à-dire des formes qui peuvent être facilement produites, reproduites et perçues.

La composante acoustique du rythme dans la parole se réfère, selon nous, non seulement à l'organisation temporelle comme cela est le cas en psychologie, mais aussi à tout élément de nature acoustique susceptible de participer seul ou en interaction avec d'autres éléments à la structuration rythmique, c'est-à-dire permettant de produire un effet de rythme : contraste qualitatif de hauteur mélodique, contraste quantitatif d'intensité et de timbre. L'étude du rythme inclut par conséquent celle de la prosodie (intonation et accentuation).

La majorité des travaux sur l'influence du débit de parole est appliquée à l'étude d'unités segmentales. Très peu de travaux ont été consacrés aux effets du débit de parole sur l'organisation prosodique (un concerne l'intonation (F0, [8] et un autre le niveau macro-prosodique phonétique, sans prise en compte de l'accentuation [25]).

1.2. Objectifs

Ce travail a pour objectif l'étude à la production de la structuration temporelle d'un corpus lu dans trois conditions de débit de parole (normal, très rapide et très lent). Cette analyse se situe dans le cadre d'un projet de recherche plus large qui concerne l'étude du rythme dans la parole pour le français, et plus spécifiquement des gabarits rythmiques (*bonnes formes* au sens défini par la théorie de la Gestalt) et de la pulsation accentuelle.

Notre hypothèse est que la structuration rythmique actualise les formes linguistiques morpho-syntaxiques et énonciatives dans un format qui est conditionné par des lois rythmiques. Des processus perceptifs tels que la segmentation du flux de parole en groupes de petite taille et la détection d'une pulsation accentuelle permettraient de dépasser les limites de la mémoire de travail et ainsi faciliteraient le traitement de l'information continue. Sur le plan de la production, ce formatage par défaut permettrait la mise en forme de la matière verbale.

Notre objectif est de contraindre la structuration rythmique de textes lus en manipulant le débit de parole afin d'observer les contraintes qui opèrent sur le formatage des gabarits rythmiques et la production de la pulsation accentuelle. En effet, on contrôle par ce biais la composante linguistique formelle (morpho-syntaxique),

qui reste identique, tout en manipulant via les variations de débit l'organisation rythmique de la matière phonique.

L'objectif de la présente étude est de déterminer comment la production du flux rythmique, en l'occurrence l'alternance de suites de syllabes inaccentuées et accentuées, est conditionnée par l'organisation temporelle, et ce en relation avec le débit de parole. Nous nous intéresserons principalement aux contrastes de durée syllabique et à l'organisation et à la fréquence des pauses.

2. PROCEDURE EXPERIMENTALE

Corpus :

Le corpus consiste en un texte lu, un conte de fées constitué de 965 mots. Ce texte est lu dans trois conditions de débit de parole (normal, très rapide et très lent), par un locuteur, avec deux répétitions. L'enregistrement a été effectué en chambre sourde.

Analyse expérimentale :

La perception d'une pulsation rythmique présuppose la répétition d'un élément dont la modification qualitative dans le temps, – son accentuation –, produit des accents et des groupements rythmiques. Dans cette approche, la durée de la syllabe inaccentuée constitue en français une référence sur le plan rythmique, une durée standard. Par conséquent, la durée des pauses et des syllabes accentuées est comparée localement à celle des syllabes inaccentuées.

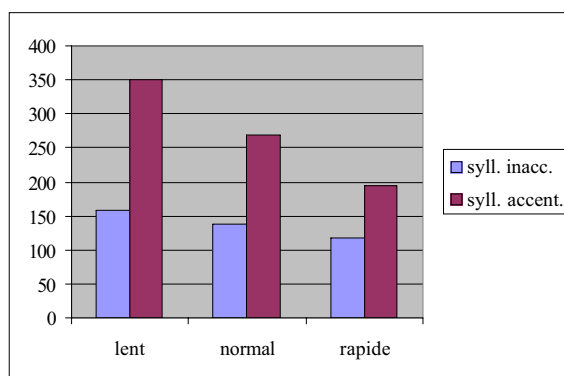
L'étude de la structuration rythmique des énoncés comporte à la fois l'analyse phonétique des paramètres prosodiques (variations de la durée syllabique, de la fréquence fondamentale et de l'intensité) ainsi que leur interprétation phonologique qui permet de déterminer une structure rythmique abstraite dans le cadre d'un modèle théorique donné. Cette représentation phonologique correspond à la structure accentuelle et aux groupements prosodiques. Notre modèle prosodique génère une structure à trois niveaux (groupe accentuel, mot rythmique et séquence rythmique) sur la base de critères linguistiques et rythmiques (pour plus de détails cf. [19], [20], [21] et [22]).

Dans la présente étude, seule l'interprétation des accents est prise en compte. Dans un premier temps, la segmentation phonétique des énoncés en syllabes et pauses et l'étiquetage sont effectués manuellement. Dans un second temps, ces données sont interprétées dans le cadre de notre modèle prosodique, ce qui permet de déterminer le caractère inaccentué ou accentué des syllabes. L'écoute perceptive du rythme d'un énoncé peut aider dans les cas ambigus. Les analyses de variance effectuées sur les durées syllabiques dans les différentes conditions de débit ne prennent pas en compte les pauses.

3. RESULTATS PRELIMINAIRES

Les premiers résultats portent sur l'étude d'une sous-partie du corpus : 94 mots correspondant à environ 120 syllabes selon les réalisations.

La vitesse d'articulation moyenne est de 4.2 syllabes par seconde à débit lent, de 5.2 syll./sec. à débit normal et de 6.8 syll./sec. à débit rapide. Ces données sont proches de celles relevées dans les études précédentes ([8], [25]). Les analyses de variance font apparaître un effet très significatif du débit de parole sur la durée des syllabes (indépendamment de leur caractère accentué ou inaccentué) : les durées syllabiques à débit normal sont significativement différentes des durées à débit lent ($F(114, 113) = 76.54, p = .000$) et à débit rapide ($F(114, 113) = 63.24, p = .000$). Cependant, le rythme n'est pas élastique : la structuration prosodique d'un énoncé produit à débit lent ne correspond pas au ralentissement linéaire du même énoncé produit à débit rapide, comme cela peut être réalisé avec un logiciel d'analyse-synthèse de la parole. Et vive-versa. Quand le débit de parole change, la durée syllabique ne varie pas dans les mêmes proportions selon que la syllabe est accentuée ou inaccentuée (cf. Graph. 1 et Tab. 1). A débit lent, on observe une amplification des contrastes de durée entre syllabes inaccentuées et accentuées. Ce phénomène s'accompagne d'une augmentation marquée de la durée des pauses, réalisées en plus grand nombre. On peut presque parler d'une *sur-exploitation des variations temporelles* à débit lent. Inversement, pour ce locuteur, les variations de F0 semblent sous-exploitées à débit lent, comme cela a déjà été relevé ([8]).



Graphique 1 : Durées moyennes des syllabes inaccentuées et accentuées dans les trois débits (ms)

Tableau 1 : Moyenne des durées syllabiques et des pauses dans les trois conditions de débit

	débit lent	débit normal	débit rapide
syllabes inaccentuées	158ms ±38	137ms ±35	118ms ±32
syllabes accentuées	351ms ±142	270ms ±109	194ms ±57
pauses	714ms ±692	431ms ±469	292ms ±131

Les analyses de variance révèlent un effet très significatif du débit sur la durée des syllabes accentuées : les durées des syllabes accentuées à débit normal sont

significativement différentes des durées des syllabes accentuées à débit lent ($F(47, 46) = 53.96, p = .000$) et à débit rapide ($F(47, 46) = 43.54, p = .000$).

A débit lent, l'allongement moyen des syllabes accentuées est de 121% par rapport à la valeur de référence que constitue la durée moyenne de la syllabe inaccentuée ; la durée des pauses après syllabes accentuées n'est pas prise en compte dans cette moyenne, mais le contraste serait encore plus marqué si on l'intégrait dans ce calcul ; en effet les pauses sont interprétées chez certains auteurs comme la partie silencieuse de l'allongement, ce qui renforce son rôle démarcatif de frontières de groupements prosodiques. Par contre à débit normal, l'allongement moyen des syllabes accentuées est de 96% par rapport à la durée moyenne des syllabes inaccentuées et, à débit rapide, l'allongement moyen des syllabes accentuées tombe à 65% par rapport à la durée moyenne de la syllabe inaccentuée.

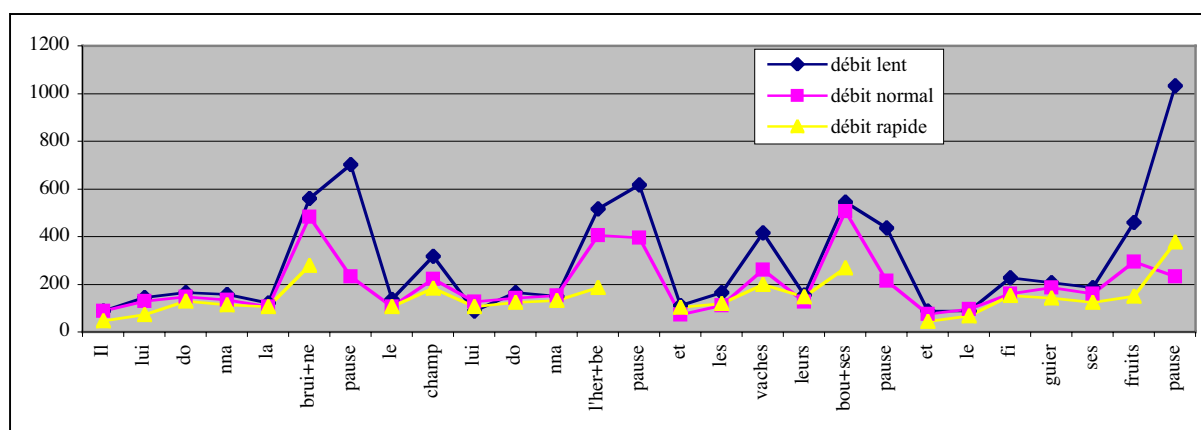
Ainsi, plus on parle lentement, plus on renforce les contrastes de durée, non pas en augmentant d'un même facteur la durée des syllabes inaccentuées et celle des syllabes accentuées, ce qui du fait de la linéarité de la parole poserait des problèmes d'intégration perceptive, mais en maintenant la durée des syllabes inaccentuées à une valeur relativement basse et stable et en augmentant beaucoup celle des syllabes accentuées (cf. graphiques 1 et 2). Par cette stratégie rythmique, le caractère inaccentué d'une syllabe est moins dépendant de la condition de débit de parole, et cette caractéristique est donc facilement identifiable par le système perceptif, puisque la durée moyenne des syllabes inaccentuées à débit lent reste toujours inférieure à celle des syllabes accentuées à débit rapide (158ms vs. 194ms).

Les analyses de variance révèlent néanmoins un effet significatif du débit sur les syllabes inaccentuées : les

durées des syllabes inaccentuées à débit normal sont significativement différentes des durées des syllabes inaccentuées à débit lent ($F(74, 73) = 52.4, p = .000$) et à débit rapide ($F(74, 73) = 63.74, p = .000$). La syllabe inaccentuée subit donc dans des proportions très restreintes mais incontournables les effets des changements de débit.

La durée moyenne des syllabes inaccentuées n'augmente que de 17% puis de 15% lorsqu'on passe respectivement du débit rapide au débit normal, puis du débit normal au débit lent, alors que pour les syllabes accentuées, leur durée augmente de 39% et 30% lorsqu'on passe respectivement du débit rapide au débit normal, puis du débit normal au débit lent. Par conséquent, la syllabe inaccentuée résiste mieux aux changements de débit de parole puisqu'elle varie en moyenne de 35% entre les deux conditions extrêmes de débit, alors que la durée de la syllabe accentuée varie presque du simple au double entre ces deux conditions, 81% (cf. Graph. 1). En raison des contraintes motrices à la production et des contraintes perceptives, la durée de la syllabe inaccentuée ne peut se réduire au-delà d'un certain seuil. A l'inverse, l'allongement d'une syllabe accentuée, renforcée ou non par une pause, peut être considérable.

En ce qui concerne les pauses, leur durée augmente dans de plus fortes proportions quand le débit ralentit (de +66% quand on passe du débit normal au débit lent) que quand le débit s'accélère (de -32% quand on passe du débit normal au débit rapide) (cf. Tab. 1). A débit rapide, chez ce locuteur la stratégie est plutôt de supprimer des pauses, leur nombre diminue de 69% du débit normal au débit rapide, alors qu'il augmente de 31% du débit normal au débit lent. Cette tendance est observée dans d'autres études ([25]). Les pauses les plus longues sont maintenues à débit rapide.



Graphique 2 : Durées syllabiques et pauses de la phrase B36 dans les trois conditions de débit (ms)

4. CONCLUSION

Il est nécessaire tout d'abord de relativiser ses résultats étant donné qu'ils portent sur l'analyse d'un petit nombre de données, sur un locuteur et sur un type bien particulier de lecture, celle du conte.

Cette étude fait apparaître que la structuration temporelle d'un énoncé produit à débit lent ne correspond pas au ralentissement général du même énoncé produit à débit rapide. Les contrastes de durée entre les syllabes inaccentuées et accentuées se renforcent à débit lent et diminuent à débit rapide. A débit rapide, la structuration rythmique est modifiée

par un affaiblissement des contrastes de durée ainsi que par une quasi-suppression des pauses qui apparaissent seulement pour des démarcations majeures de groupes rythmiques. A débit lent, les variations temporelles sont sur-exploitées dans de nombreux domaines : renforcement des contrastes de durée, augmentation et allongement du nombre des pauses, ainsi qu'augmentation du nombre de schwas à l'intérieur et en fin de groupes rythmiques. Cette dernière stratégie a pour effet d'augmenter le nombre de syllabes. A débit lent, le flux rythmique est mieux et plus saucissonné, chaque séquence rythmique étant nettement séparée des autres par une pause importante.

Par ailleurs, la faible variation de durée de la syllabe inaccentuée entre les débits lent et rapide plaide en faveur du rôle de référence de la syllabe inaccentuée sur le plan rythmique en français. La syllabe inaccentuée a très peu d'élasticité relativement à la syllabe accentuée. A un niveau macro-rythmique, les variations ou changements perçus dans le débit de la parole sont liés sur le plan temporel essentiellement à la quantité d'allongements et de pauses produits. Il serait intéressant de vérifier si l'on observe les mêmes phénomènes en musique entre les différents tempos.

L'étude complète du corpus ainsi que celle d'autres types de corpus et d'autres locuteurs devra permettre de déterminer les influences respectives des stratégies mises en place par les locuteurs et des types de documents lus.

BIBLIOGRAPHIE

- [1] C. Astésano. *Rythme et Accentuation en Français. Invariance et Variabilité stylistique*, L'Harmattan Paris, 336 p., 2001.
- [2] A. S. Bregman. *Auditory scene analysis : The perceptual organization of sound*, Cambridge Mass., MIT Press, 1990.
- [3] E. Delais. *Pour une approche parallèle de la structure prosodique*, Doctorat, Université de Toulouse-le-Mirail, 1995.
- [4] D. Deutsch. Grouping mechanism in music, in *The psychology of music*, (ed. D. Deutsch), New York, Academic Press, 99-130, 1982.
- [5] C. Drake. *Processus cognitifs impliqués dans l'organisation du rythme musical*, thèse de doctorat Paris, Université René-Descartes (Paris V), 1990.
- [6] C. Drake. Reproduction of musical rhythms by children, adult musicians and adult nonmusicians, *Perception & Psychophysics*, 53 (1), 25-33, 1993.
- [7] 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.
- [8] C. Fougeron & J. Sun-Ah. Rate effects on French intonation: prosodic organization and phonetic realization, *Journal of Phonetics*, 26, 45-69 Academic Press, 1998.
- [9] P. Fraisse. Les structures rythmiques, *Studia Psychologica*, Publications Universitaires de Louvain, 124p., 1956.
- [10] P. Fraisse. Psychologie des rythmes humains *Colloque sur "Les rythmes"*, *Journal Français d'Oto-Rhino-Laryngologie*, Sup. 7, Lyon, 4 Déc 1967, SIMEP, 23-33 1967.
- [11] P. Fraisse. *Psychologie du rythme*, Presses Universitaires de France, Paris, 360p., 1974.
- [12] P. Fraisse. Multisensory aspects of rhythm, in *Intersensory Perception and Sensory Integration*, R. D. Walk et H. L. Pick eds, Plenum Publishing corporation, New York, 217-248, 1981.
- [13] J.P. Gee & F. Grosjean. Performance Structures: A Psycholinguistic and Linguistic Appraisal *Cognitive Psychology*, 15, 411-458, 1983.
- [14] F. Grosjean & J.-Y. Dommergues. Les structures de performance en psycholinguistique, *L'année Psychologique*, 83, 513-536, 1983.
- [15] P. Guillaume. *La psychologie de la forme* Flammarion, 1979.
- [16] K. Helsloot. *Metrical prosody: a template-and-constraint approach to phonological phrasing in Italian*, Holland Academic Graphics (HIL dissertations ; 16), the Hague, 317, 1995.
- [17] W. Köhler. *Gestalt Psychology*, New York Liveright, 1929.
- [18] Ph. Martin. Structure prosodique et structure rythmique pour la synthèse, *Actes des 15èmes Journées d'Etudes sur la Parole*, Aix-en-Provence 27-30 Mai 1986, GALF, 89-91, 1986.
- [19] V. Padeloup. A prosodic model for French text-to-speech synthesis : A psycholinguistic approach, in *Talking Machines: Theories, Models and Designs* G. Bailly, C. Benoît, and T.R. Sawallis (Eds.) Elsevier Science Publisher, 335-348, 1992.
- [20] V. Padeloup. Exploiting the secondary accent in a prosodic model for French synthesis, *Proc. of International Congress of Phonetic Sciences*, Aix-en-Provence 19-24 Août 1991, 2 : 254-257, 1991.
- [21] V. Padeloup. *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*, thèse de doctorat nouveau régime, Université de Provence Aix-Marseille 1, 1990.
- [22] V. Padeloup. Essai d'analyse du système accentué du français : distribution de l'accent secondaire *Actes des 17èmes Journées d'Etude sur la Parole* Nancy, 20-23 Septembre 1988, 65-70, 1988.
- [23] M. Wertheimer. *Der Gestalttheorie*, Erlangen Weltkreis-Verlag, 1925.
- [24] H. Woodrow. Time perception, in *Handbook of experimental psychology*, S. S. Stevens ed., Wiley New York, 1224-36, 1951.
- [25] B. Zellner. Caractérisation et prédiction du débit de parole en français, Doctorat, Lausanne, 1998.