Rate sensitivity of syllable in French: a perceptual illusion?

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, malikafaraj@free.fr

Abstract

This study takes place within the theoretical framework of Gestalt theory. The aim of this work is to determine the way the prosodic scene reorganises itself according to the variation of speech rate. In other words: how do the forms constituted by stressed syllables interact with the ground of unstressed syllables?
We present a study of the temporal structure 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 constrain the rhythmical structure of speech in order to observe how rhythmic patterns depend on the variation of speech rate.
Results show that rhythm is not elastic: temporal structuring produced at a slow rate is not the consequence of a linear decrease. When speech rate changes, syllabic duration does not vary in the same way for stressed and for unstressed syllables. Unstressed syllables have very little elasticity compared with stressed syllables. This last result supports the hypothesis that the unstressed syllable is an anchor point in the rhythmic structure of French.

1. Introduction

1.1. Figures and ground in the prosodic scene

This study of the prosodic scene is situated within the theoretical framework of Gestalt theory ([12], [11]) and of numerous psychoacoustic studies on human rhythms based on this framework. As a matter of fact, this theory had many applications to sound and music perception ([9], [10], [20], [4], [5], [6]), although it has less frequently been applied to the field of prosody ([13], [15], [3], [2]). Indeed, prosody is still often represented as pure abstract structure. The generative structural principles from which phonological representations are built are usually not rooted in motor, perceptual and cognitive constraints which govern other human productions and their percepts.

One of the basic laws of Gestalt is that we perceive *figures, forms* in vision and audition which emerge from a *ground*. These forms correspond to "sensitive objects" which have specific caracteristics in order to be able to emerge from a ground ([11]). The figures do not have any autonomous existence since they exist only in relation to the ground. To have a form and an organization are the functional caracteristics of the figure, while the ground is an amorphous and undefined continuity which does not possess its own contours.

In the application of the Gestalt theory to speech, the succession of unstressed syllables constitutes the ground of the prosodic scene and the stressed syllables constitute the figures which emerge from this ground. Thus, a succession of syllables constitutes, in so far as an amorphous and undefined suite, the ground as long as an accent does not emerge which will give form to a given syllable. Therefore, there is strictly no such thing as unstressed syllable, but only syllables which receive or do not receive an accent.

In our opinion, the acoustic component of speech rhythm corresponds to all the acoustic elements which, either alone or in interaction with each other, can generate a perceptual effect of rhythm: contrasts of pitch, duration, intensity, vocalic quality etc. Consequently, the study of rhythm includes that of prosody (intonation and accentuation).

1.2. Aim of the study

Most works on the effects of speech rate are concerned with the segmental level rather than the suprasegmental one. Concerning those which are concerned with the suprasegmental level, studies focus on variation in duration or pitch ([8], [21]).

This study is part of a project on rhythmic templates and stress pulsation in French. According to our rhythmic law-based prosodic model, the rhythmic structuring actualizes linguistic forms (lexical, morpho-syntactic and enunciative sub-components) in a form conditioned by rhythmic laws ([13], [15]). The goal is to constrain the rhythmic structuring of read texts by manipulating speech rate. In this way it is possible to observe how the rhythmic flow reorganizes its matter and form while the same formal (lexical and morpho-syntactic) linguistic structure is retained.

The aim of the present study is to determine the way the prosodic scene reorganises itself as a result of variation of speech rate. In other words: how do the forms constituted by stressed syllables interact with the ground of unstressed syllables? We present a study of the temporal structure 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 (for some preliminary results on part of the corpus cf: [16], [17]).

2. Method

2.1. Corpus

The corpus is a one thousand word fairy-tale, which corresponds to about 1200 syllables depending on the different productions. The corpus was produced at three different rates (normal, fast and slow) by one speaker (the first author) with two repetitions. The recording was made in a sound treated recording booth.

2.2. Experimental analysis

The rhythmic structuring study of the corpus includes the phonetic analysis as well as its phonological interpretation. By phonetic analysis, we mean the realization of prosodic

parameters: mainly the syllable duration and the pitch contour. By phonological interpretation, we mean the determination of an abstract rhythmic structure in the framework of a given theoretical model. Phonological representation corresponds to accentuation (the number and type of accents) and rhythmic phrasing (the number and types of rhythmic groups). Our rhythmic model distinguishes four prosodic levels ([13], [14], [15]) :

- The syllable is the minimal rhythmic unit which can be stressed or unstressed. Stressed syllables will be shown in bold characters as in example (1).

(1) « Le débit d'parole » : lə - **de** - bi - dpa - **rəl**

- The stress group is the smallest rhythmic group: in French, this consists of one stressed syllable preceded by a few unstressed ones (on average 3). Stress groups will be shown by < > as in example (2).

(2) <lə - **de**> <bi - dpa - **rəl**>

- The rhythmic word is the smallest prosodic structure which organizes a meaning group ([19]). This small syntactico-semantic group is usually formed out of one or two stress groups and is limited in size to a maximum of around 7 syllables. The rhythmic words will be shown by [ ] as in example (3).

(3) [<lə **de**> <bi dpa **rəl**>]

- The rhythmic sequence is the major prosodic structure, which refers to the discourse level. This large syntactico-semantic group is not limited in size. Rhythmic sequences will be shown by / / as in example (4).

(4) / [<lə **de**> <bi dpa **rəl**>] [<ã frã **se**>] [<e **lã**>] /

In our model, accentuation and phrasing rules are based both on linguistic criteria (lexical and morpho-syntactic) and rhythmic criteria (such as size: syllable number of lexical units and rhythmic words). Therefore the sentences (5) and (6) which have the same syntactic structure but a different number of syllables will not necessarily be given the same prosodic structure ([15]) :

(5) Le débit$_{3syll}$ d'parole$_{2syll}$ est lent$_{2syll}$ =>
/ [lə **de** bi dpa **rəl** e **lã**]$_{7syll}$ / or
/ [lə **de** bi dpa **rəl**]$_{5syll}$ [e **lã**]$_{2syll}$ /

(6) Le rhinocéros$_{5syll}$ de Constantinople$_{5syll}$ est lent$_{2syll}$ =>
/ [lə **ri** no se **rɔs**]$_5$ [də **kɔ̃** stã ti **nɔpl**]$_5$ [e **lã**]$_2$ /$_{12syll}$

First of all, phonetic alignment was carried out using the aligner developped by LORIA (D. Foher & Y. Laprie: http://www.loria.fr/equipes/parole/). This automatic phonetic alignment was then corrected manually. We filtered out the extreme durations of pause due to some artefacts during the reading: values above 2s at fast rate, 3s at normal rate and 4s at slow rate. As the next step, a phonetic analysis of the prosodic parameters was made. These data were then interpreted in the framework of our rhythmic model. In the present study, only the accent interpretation was taken into account, that is to say the stressed or unstressed status of the syllables.

3. Results

3.1. Articulation rate and syllable duration

The articulation rate (pause excluded) is 6.8 syll/s at fast rate, 5.4 syll/s at normal rate and 4.4 syll/s at slow rate. These values are comparable to those found in previous studies in French ([8], [21]). Rate variations modify the syllable duration as expected. Nevertheless, rhythm is not elastic: temporal structuring produced at a slow rate is not the consequence of a linear decrease of the same material pronounced at a fast rate. When rate changes, syllabic duration does not vary the same way if the syllable is stressed or not (cf. Table 1, figure 1).

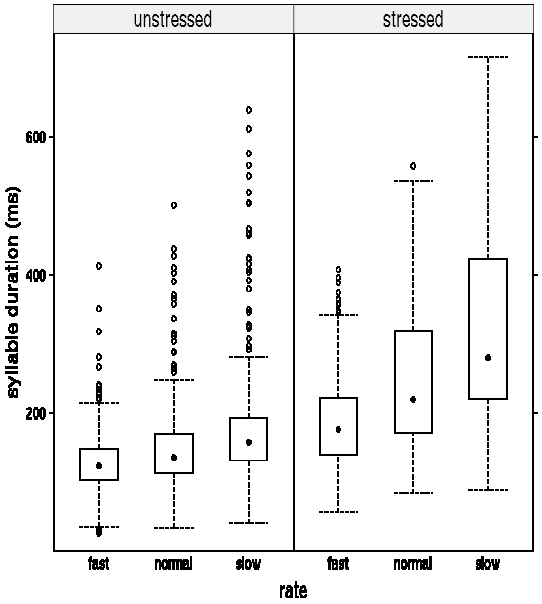

Figure 1 : *Boxplots[1] of the unstressed and stressed syllable duration in the three rate conditions*

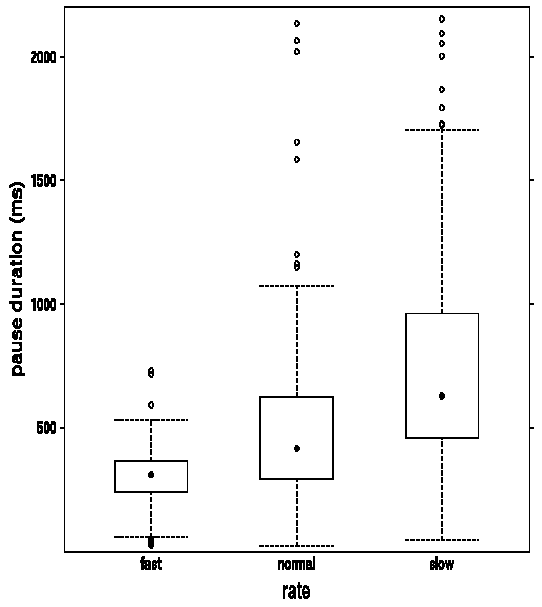

Figure 2 : *Boxplots of pause duration in the three rate conditions*

[1] The boxes figure the interquartile ranges: 50% of the data, the black circles figure the median and the empty circles figure the outliers.

Table 1 *: Mean syllable and pause duration in the three rate conditions*

	Fast rate	Normal rate	Slow rate
Unstressed syllables	126ms ±42	144ms ±54	170ms ±71
Stressed syllables	187ms ±63	252ms ±106	320ms ±134
Pauses	302ms±145	513ms±377	785ms±556

3.2. Statistic analysis

To evaluate the influence of rate on syllable duration, we run a mixed linear model ([18]) with the logarithm of the syllable duration as the dependent variable, the rate as a factor with 3 ordered levels (fast, normal, slow), stress as a 2 level factor (unstressed, stressed). The grouping factor was the segmental label of the syllable (for ex.: "pa", "ma"). The 3529 syllable durations were thus grouped into 374 syllable blocks (i.e. different segmental labels). Consequently, the effect of intrinsec duration of the 374 syllable blocks was taken into account in the model.

Results show that the two factors rate and stress as well as their interaction are highly significant. Table 2 of the model regressors shows that the rate (the linear component "rate.L") is significant. The positive value "stressed", as expected, shows that the stress increases all the syllable duration. The significant positive interaction term ("rate.Linear: stressed") shows that the incresase of rate has a stronger effect on the stressed syllables. Table 3 shows the values of the syllable duration (with stress and rate factors) which were computed on a linear scale from the estimated coefficiants presented in Table 2.

To evaluate the effect of rate on the pause duration, we ran a linear model with logarithm of the pause duration as dependent variable and rate as a factor with 3 ordered levels. Table 4 shows a significant effect of the factor rate on pause duration which decreases as rate increases.

Table 2 *: Regressors of the model "syllable"*

	Value	Std.Error	t-value	p-value
(Intercept)	5.244	0.0198	264.71	0.000
Rate.Linear	0.211	0.0076	27.72	0.000
Rate.Quadratic	0.002	0.0077	0.33	0.739
Stressed	0.224	0.0107	20.89	0.000
Rate.Linear: stressed	0.141	0.0126	11.16	0.000
Rate.Quadratic: stressed	0.002	0.0125	0.19	0.846

Table 3 *: Estimated syllable duration in the three rate conditions by the model "syllable"*

	Fast rate	Normal rate	Slow rate
Unstressed syllables	163ms	190ms	220ms
Stressed syllables	174ms	237ms	322ms

Table 4 *: Regressors of the model "pause"*

	Value	Std.Error	t-value	Pr (>/ t /)
(Intercept)	5.98	0.038	156.349	<2e-16
Rate.Linear	0.63	0.070	8.990	<2e-16
Rate.Quadratic	-0.02	0.061	-0.393	0.695

3.3. Rate and stress

Speech rate does not have the same effect on the forms and on the ground of the prosodic scene (cf. Fig. 1). Unstressed syllables have very little elasticity compared with stressed syllables. Our results thus quantify what Duez [7] mentioned for French. The ground of unstressed syllables is more stable and resistant to rate variation than the forms constituted by stressed syllables. Compared to normal rate, unstressed syllables are on average 14% shorter at fast rate and 16% longer at slow rate, while stressed syllables are on average 26% shorter at fast rate and 36% longer at slow rate (these percentages are calculated from Table 3). This result supports the hypothesis that the unstressed syllable is an anchor point in rhythmic structure of French.

Table 5 **:** *Number of unstressed syllables, stressed syllables and pauses in the three rate conditions*

	Fast rate	Normal rate	Slow rate
Syllables	1174	1174	1193
Unstressed syllables	746	709	728
Stressed syllables	428	465	465
Pauses	78	172	231

The number of syllables produced is stable in the three rate conditions (cf. Table 5). The percentage of stressed syllables is slightly smaller at fast rate: 36% against 39% at normal and slow rates. Similar result was found in Dutch [2]. This means that the rate variations have almost no effect on the size of the stress group: on average between 2.52 and 2.74 syll., i.e. one stress every 2-3 syllables. Consequently, we can hypothesize that the skeleton of stress pulsation is not rate sensitive.

3.4. Rate and pause

The duration and the number of pauses decreases as the rate increases (cf. Tables 1 and 5, Figure 2). Compared to normal rate, pause duration is on average 41% shorter at fast rate and 53% longer at slow rate. The number of pauses decreases by an average of 55% between the normal and fast rate conditions and increases by an average of 34% between the normal and slow rate conditions. This tendency was also observed in other studies ([21]).

3.5. Accentual contrast and prosodic strategies

Prosodic strategies are different according to speech rate since the accentual contrasts have phonetically different manifestations. In French, a stressed syllable contrasts phonologically with the preceding unstressed syllables. The syllabic duration contrasts strengthen when the speech rate decreases.

At fast rate, the duration contrasts between the stressed syllables and the unstressed ones are weakened compared to normal rate. There is a functional synergy between this phenomenon and the strong reduction of the number and duration of pauses. In some cases, major rhythmic boundaries (corresponding to rhythmic sequences in our model) can be produced without any pause, just with syllabic lengthening and F_0 resetting.

At slow rate, the duration contrasts between the stressed syllables and the unstressed ones are strengthened compared

to normal rate. There is a functional synergy between this phenomenon and the increase of the number and duration of pauses. Furthermore, for this speaker at slow rate, this hyper-usage of duration contrasts seems to be combined with a hypo-usage of pitch contrasts. In some cases, accentual contrast is realized just by temporal contrast.

4. Conclusion

In this study, we have shown that, in French, the variation of speech rate does not have the same effect in the prosodic scene on syllables which are the form (i.e. stressed) as on those which are the ground (i.e. unstressed). Unstressed syllables are less rate sensitive than stressed ones. As expressed by Gestalt theory, the forms constituted by stressed syllables have distinct caracteristics from those of the ground of unstressed syllables. The ground of the prosodic scene is more stable and resistant to rate variation than the forms which emerge from the ground: the unstressed syllable duration increases by an average of 35% between the fast and slow rate conditions, while the stressed syllable duration increases by an average of 85% between those two conditions. Still the fact remains that rate variations have slight but unavoidable effects on unstressed syllable duration. The lesser elasticity of unstressed syllables supports the hypothesis that this syllable is an anchor point in the rhythmic structure of French. These results of rate effects on syllabic duration need to be validated with recordings from other speakers. Concerning the motor programming of syllable duration, we can hypothesize that only the duration of stressed syllables will be planned (that is to say the duration contrast). When speech rate changes, the variation of the unstressed syllable duration would correspond to an intrinsic low level variation of the motor system, while that of the stressed syllable would correspond to a high level variation of the system (i.e. motor commands).

To conclude, we can notice that it is not a linear rhythmic elasticity (i.e. a linear compression or lengthening of all the syllables) which is responsible for the perception of speech acceleration and deceleration. For example, at fast rate the impression can be given that the whole rhythmic flow (that is to say all the syllables) is pronounced faster than at normal rate, whereas in fact unstressed syllables are on average 14% shorter and stressed syllables on average 26% shorter. From the accentual contrast point of view, the contrast in syllabic duration weakens when rate increases and strengthens when rate decreases. There seems, consequently, to be a perceptual illusion: we have the impression that ground of the prosodic scene accelerates or decelerates according to the different rates, whereas in fact it is mainly the duration of the presentation of the forms which is shorter or longer. When the rate increases, the duration of the presentation of the stressed syllables shortens. When the rate decreases, the duration of the presentation of the stressed syllables lengthens. The difference between the experienced percept of a visual form and the sensory stimulus on which it is based is a well-known phenomenon observed in visual illusions by Gestalt theory.

We wish to thank Daniel Hirst for his helpful comments and useful scripts running on Praat.

5. References

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

[2] CASPERS J. (1994) Pitch movements under time pressure : effects of speech rate on the melodic marking of accents and boundaries in Dutch, Doctoral dissertation, University of Leiden.

[3] COUPER-KUHLEN E. (1993) English Speech Rhythm : Form and function in everyday verbal interaction, Amsterdam, John Benjamins Publishing Company.

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

[5] DRAKE C. (1990) Processus cognitifs impliqués dans l'organisation du rythme musical, Thèse doctorale, Paris, Université René-Descartes (Paris V).

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

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

[8] FOUGERON C. & JUN S.-A. (1998) Rate effects on French intonation: prosodic organization and phonetic realization, Journal of Phonetics, 26, 45-69, Academic Press.

[9] FRAISSE P. (1956) Les structures rythmiques, Studia Psychologica, Publications Universitaires de Louvain, 124p.

[10] FRAISSE P. (1974) Psychologie du rythme, Paris, Presses Universitaires de France, 360p.

[11] GUILLAUME P. (1937) La psychologie de la forme, Paris, Flammarion (1979).

[12] KÖHLER W. (1929) Gestalt Psychology, New York, Liveright.

[13] PASDELOUP V. (1990) Modèle de règles rythmiques du français appliqué à la synthèse de la parole, thèse de doctorat, Université de Provence Aix-Marseille 1.

[14] PASDELOUP V. (1991) 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.

[15] PASDELOUP V. (1992) 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.

[16] PASDELOUP V. (2004) Le rythme n'est pas élastique : étude préliminaire de l'influence du débit de parole sur la structuration temporelle, Actes des 25èmes Journées d'Etude sur la Parole, Fès (Maroc), 19-22 Avril 2004, 397-400.

[17] PASDELOUP V. (2005) Figures et fond dans la scène prosodique : leur résistance face aux variations du débit de parole, *IDP 2005*, 8-9 septembre 2005.

[18] PINHERO J. C. & BATES D. M. (2001) Mixed-Effects Models in S and S-Plus, Springer.

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

[20] WOODROW H. (1951) Time perception, in Stevens S. S. (ed.), Handbook of experimental psychology, New York, Wiley, 1224-36.

[21] ZELLNER B. (1998) Caractérisation et prédiction du débit de parole en français, Doctorat, Lausanne.