Zellner, B. (1998). Fast and Slow Speech Rate: a Characterisation for French. *ICSLP, 5th International Conference on Spoken Language Processing.* (Volume 7, pp. 3159 - 3163), December 1998, Sydney (Australia).

Fast and Slow Speech Rate: a Characterisation for French

Brigitte Zellner

LAIP, IMM, Faculté des Lettres
University of Lausanne, 1015 Lausanne, Switzerland
Brigitte.Zellner.@imm.unil.ch

ABSTRACT

This paper is concerned with the evaluation of speech rate in French. Usually, this dynamic parameter is described as a unidimensional quantitative dimension. It is shown that the slowing down of speech has also major qualitative effects that must be taken into account. The theory on slowing down speech is thus revised.

1. INTRODUCTION

Usually, speech rate is described as a single quantitative dimension such as the number of syllables produced per second (for example, Grosjean, 1975; Bertrand & Casolari, 1996). However, many phonetic studies have shown that changes in speech rate have numerous effects at various levels of the temporal structure. For example for French, such effects have been observed at the prosodic level (for example, Fougeron, 1995), at the syllabic level (for example, Duez, 1987; Lacheret-Dujour, 1990) and at the segmental and articulatory levels (for example, Lacheret-Dujour, 1990; Vaxelaire, 1995). Hence, a characterisation of speech rate with a single parameter is not satisfactory. This observation is reinforced by a verification with speech synthesis. Increasing or decreasing the number of syllables per second is not a satisfactory manner of creating natural sounding fast or slow synthetic speech.

This paper reports on part of a doctoral study (Zellner, 1998) and it shows that the characterisation of speech rate has to take account qualitative features. When compared with other studies, our results suggest that some speakers may use different "strategies" in slowing down their speech production.

2. METHODOLOGY

In our study, 50 sentences were read by a highly fluent French speaker at two speech rates (9 sentences containing less than 12 syllables, 20 sentences containing 13 to 20 syllables, 21 sentences containing 21 to 30 syllables). As judged by ten native speakers of French, the readings were considered to be highly intelligible with no dialectal accent. The signals were carefully segmented and checked by two experts. Then, phonetic syllabic boundaries were controlled. In particular, the syllables containing a branching coda or a branching onset and the cases of "liaisons" and "enchainements" effects were examined.

The raw syllabic durations were then normalised with a logarithmic transformation. The minor and major prosodic boundaries were identified on the basis of a proximal textual analysis, looking for succeeding grammatical and lexical words (Zellner, 1996, to appear). This algorithm is similar to the one developed by Sorin (1987).

A database was then drawn up, providing for each syllable the syllabic and segmental contexts, the length (number of syllables) of the sentence, the length of the prosodic group, the type of word in which the syllable is located (lexical or grammatical), the length of the word, the position of the syllable in the word, the syllabic structure, the duration of each syllabic component (onset, nucleus, rhyme), the length (number of segments) of the syllable, the duration of each segment, the potential presence of a schwa. Then, the 968 syllables produced at the fast speech rate were compared to the 1001 syllables produced at the slow speech rate. The statistical analyses were performed with the DataDesk 5.0 package, on a Macintosh.

3. RESULTS

As expected, the comparison between the two speech rates shows an increased speech time (phonatory time + pauses) at the slow speech rate. This increase can be explained with respect to various parameters ranked according to their contribution to this "time differential " — i.e., the total slowing down time for the 50 sentences.

3.1. Pauses

It has largely been reported in the literature that the main mechanism to slow down speech consists in the production of silent pauses. (For example, Saint-Bonnet & Boe, 1977; Grosjean, 1975; 1979; Barbosa, 1994). However, it appears in our analysis that the silent pauses produced at the slow speech rate represent only 7.9% of the total time differential.

The pausal analysis shows two major features. First, there is a lengthening of the pauses already produced at the fast speech rate. This lengthening represents 1.07% of total slowing.

Then, additional pauses are produced at the slow speech rate (6.83% of total time differential). These pauses emerge in longer prosodic groups. The average length of the prosodic groups slowed down by a new pause is 7.11 syllables, a significant difference with the other groups at the slow speech rate (F-ratio = 16.91, $p \leq 0.0001$).

The number of syllables produced between two pauses is variable. Pauses tend to be produced around either a previous minor boundary or a previous major boundary, either at a new interlexical boundary — i.e., often after a period of slowing down. From one speech rate to the other, the location of pauses is different, since 57% of pauses are placed after a minor prosodic boundary at the fast speech rate versus 71% at the slow speech rate. At the fast speech rate, pauses are produced not far from the first major prosodic boundary which is located around the middle of the sentence. At the slow speech rate, pauses are not so narrowly conscentrated. No correlation appears between the average speech rate per sentence and the number of pauses. In other words, for this speaker, the "local" speech rate is not correlated to the frequency of pauses.

In conclusion, two different pause patterns for the two speech rates were found. The importance of the pausal mechanism in slowing down speech appeared to be weaker with this speaker than suggested by the literature (Barbosa, 1991, 1994).

Barbosa (1994) has suggested that the presence or the absence of a silent pause participates in the same syllabic lengthening event with a fork into the distribution of the durations within the rhythmic unit. Indeed, our data confirm that pauses and syllabic lengthening are the major indicators of the temporal structure of an utterance where a slowing down indicates a less proximal interlexical relation. Having said that, it does not appear that pausing and lengthening should be considered "equivalent" in terms of phonatory strategies, since one introduces a stop and the other one introduces a continuation of the gesture. In that sense, our results do not confirm Barbosa's claim.

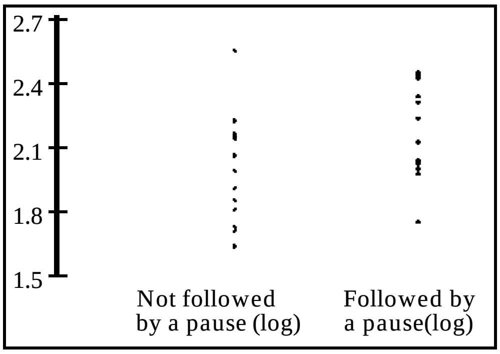


Figure 1 Duration of rhymes at fast speech rate

In our study, the durations of the rhymes preceeding a pause are not significantly longer than the durations of the other rhymes, suggesting that there is no significative difference between the two groups of syllables. Thus, the dependency between pausing and lengthening does not appear reliably (figure 1).

3.2. New Verbal Material

The production of new verbal material is another mechanism to slow down the speech rate. This strategy represents 8.8% of total time differential.

80% of these syllables are located inside the prosodic group. In 88% of the cases, the new syllable contains a schwa. The other cases concern diaeresis events (ex: in/te/llec/tuel -> in/te/llec/tu/el), and 2% are cases of hyperarticulation. (ex: a été (ai)guillé -> a ét**é ai**guillé).

The major effect of this mechanism is to change the phonological structure of the utterance.

3.3. Segmental Lengthening

Finally, the most poweful manner of slowing down speech is the lengthening of the durations of the segments. The segmental lengthening represents 69.96% of total time differential. Each segment is lengthened by 34.5% *on the average*. However the comparison of the durational distributions of segments reveals a non-linear relation between the two speech rates.

Groups Fast	Duration (log)	
1	1.572	j,w, l, ?
2	1.659	˙, Â, d
3	1.750	n, y, ë, v, i, u, J, O, g, E, z
4	1.842	e, a, b, m, t
5	1.939	E$, O$, ˆ, Ø, A$, o, k, π, p, π$, s, f
6	2.038	S

Groups Slow	Duration (log)	
1	—	—
2	1.668	˙, j, ?
3	1.753	l, ë, w, Â, d, y
4	1.831	n, v, i, g, z, O
5	1.935	J, b, u, E, m, p, t, a, k
6	2.055	e, A$, o, f
7	2.125	O$, s, π, E$, S, ˆ, Ø

Table 1: Duration groups of segments (Fast, n=2175, slow, n=2295)

The grouping of the segments in durational classes based on the normalised averaged duration of the segment (cf. table 1) shows that some segments are not in the same group, when ranked according to the speech rate. For example, the segment [˙] is in the group {˙, R, d} at the fast speech rate, but belongs to the group {˙, j} at the slow speech rate.

This is probably due to the fact that segments do not share the same temporal elasticity — it is easier to lengthen a [R] than a [˙]. Moreover, some segments may have a different phonological status — for example, the French schwa seems to have a fairly stable duration, independently of the speech rate.

In conclusion, the most poweful mechanism to slow down the speech production is to lengthen the duration of speech units.

3.4. The Interlexical Cohesion

Slowing down speech changes the structure of the interlexical cohesion thanks to the insertion of new syllables and pauses. Since the pausal mechanism is more active at prosodic boundaries, the "enchainements" (ex: "il par**le à** sa fille" — he spokes to his daughter—) are more directly concerned than the "liaisons" events (ex: "nou**s a**vons" — we have —) .

As a matter of fact, the analysis of occurrences of liaisons and enchainements in this study shows that for the two speech rates, the two events are differently distributed. The Chi 2 tests reject the hypothesis of a uniform distribution of liaisons and enchainements (Fast: $\chi 2$ = 21.778, p= 0.0001, df=1; Slow: $\chi 2$ = 19.036, p= 0.0001, df = 1). Liaisons are relatively more frequent inside the prosodic group while enchainements are less sensitive to the presence of a prosodic boundary, i.e., they are equally distributed inside a group and at its boundaries.

Hence, the effect of adding new syllables and new pauses is more sensitive to the enchainement process. Our speaker produced 18% more of enchainements at the fast speech rate while the number of liaisons remained stable despite of the speed variation. At the slow speech rate, 63% of the suppressed enchainements are replaced with the insertion of a new syllable containing a schwa followed by a pause.

In this study, the enchainement process is impeded by the phonological structure of the speech chain at the slow speech rate. The largest number of schwas at the end of lexemes plus the insertion of pauses restrain the interlexical binding process during the speech production.

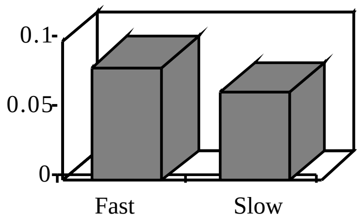
4. DISCUSSION

The relevant literature suggests that speech production is generally slowed down in two ways: by the pausal mechanism and by the lengthening of speech units (for example, Barbosa, 1994). Usually, these two mechanisms were not scaled. This was the source of defective algorithms for speech synthesis and this explains why most of the synthetic rhythms sound unnatural.

In this study, a systematic comparison of 50 sentences read by the same speaker at two speech rates shows that slowing down is obtained first by lengthening the duration of segments, second by producing additional syllables, and third by producing pauses. From the qualitative point of view, slowing down has a major effect on the interlexical binding process. It is claimed that taking account of this effect would allow a better simulation of one's speech rate organisation.

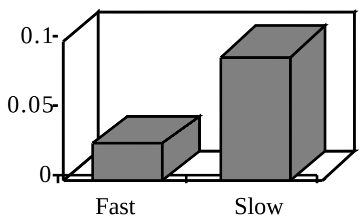
In this respect, it is proposed to calculate a first indicator of *interlexical proximity* by taking the ratio between the number of enchainements and the number of words (this remains the same whatever the speech rate). Interlexical proximity is high when the indicator is high.

Nb enchainements / Nb words



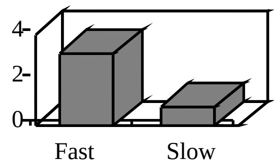
To improve the estimation of the interlexical cohesion, one has to take account of pauses since they impede interlexical binding. An indicator of *interlexical distance* is thus computed with the ratio of the number of pauses and the number of words. The interlexical distance is high when the indicator is high.

Nb pauses / Nb words



The *interlexical cohesion degree* then corresponds to the ratio between the indicator of word proximity and the indicator of word distance. It indicates for a specific speech rate the proximity between words according to both the enchainement and the pausal process. When the cohesion is high, the temporal integration degree is also high.

Interlexical cohesion degree



In that case, at the fast speech rate, interlexical cohesion degree is high because both of the smaller number of pauses and of the largest number of enchainements.

As it can be seen, this indicator is interesting since it captures a two dimensional phenomenon that caracterises speech rate change in French.

CONCLUSION

In this study, it was shown that the evaluation of speech rate has to take into account various quantitative and qualitative parameters.

In French, except for Lacheret-Dujour (1991) who concentrated more on the phonological dimension, the slowing down mechanism of inserting new syllables was not explicitly reported and scaled. In our speaker's productions, this mechanism is as frequent as the pausal strategy. It also appeared that, contrary to Barbosa's data (1994), the occurrences of pauses were not clearly related to the lengthening of the preceeding speech units. It is proposed that these differences be considered among the possible strategies of slowing down speech in French.

REFERENCES

Barbosa, P. A. (1994). *Caractérisation et génération automatique de la structuration rythmique du français.* Thèse de Doctorat. U.R.A. CNRS n°368 - INPG/ENSERG, Université Stendhal, Grenoble.

Bartkova, K. (1991). Speaking rate in French application to speech synthesis. *XIIème Congrès International des Sciences Phonétiques,* (pp. 482-485). Aix en Provence. Actes.

Bertrand, R., et Casolari, F. (1996). Approche prosodique et pragmatique des modulations. *XXIes Journées d'Etudes sur la Parole,* (pp. 179 - 182.). Avignon.

Duez, D. , & Nishinuma, Y. (1987). Vitesse d'élocution et durée des syllabes et de leurs constituants en français parlé. *Travaux de l'Institut de Phonétique d'Aix, 11.* 157-180.

Fougeron, C. & Jun, S-A. (1995). Properties of french intonation at fast speech rate. *XIIIème Congrès International des Sciences Phonétiques, 3* (pp. 488-491). Stockholm.

Grosjean, F. , & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica, 31.* 144-184.

Grosjean, F. , & Collins, M. (1979). Breathing, pausing, and reading. *Phonetica, 36.* 98-114.

Lacheret–Dujour, A. (1990). *Contribution à l'analyse de la variabilité phonologique pour le traitement automatique de la parole continue multilocuteur.* Thèse de doctorat de l'Université de Paris 7.

Lacheret–Dujour, A. (1991). Le débit de la parole: un filtre utilisé pour la génération des variantes de prononciation en français parisien. *XIIème Congrès International des Sciences Phonétiques* (pp. 194-197). Aix en Provence. Actes.

Saint-Bonnet, M., Boe, J. (1977). Les pauses et les groupes rythmiques: leur durée et disribution en fonction de la vitesse d'élocution. *8èmes Journées d'Etude sur la Parole,* (pp. 337- 343). Aix en Provence.

Sorin, C., Larreur, D. & Llorca, R. (1987). A rythm-based prosodic parser for text-to-speech systems in French. *XIème Congrès International des Sciences Phonétiques* (pp. 125-128). Talinn, Estonie. Proceedings.

Vaxelaire, B. (1994). Variation de geste et débit. Contribution à une base de données sur la production de la parole, mesures cinéradiographiques, groupes consonantiques en français. *Travaux de l'Institut de Phonétique de Strasbourg, 24.* 109-146.

Zellner, B. (1996). Structures temporelles et structures prosodiques en français lu. *Revue Française de Linguistique Appliquée: La communication parlée.* 1. (pp.7-23).Paris.

Zellner, B. (1998). *Caractérisation et prédiction du débit de parole en français. Une étude de cas.* Thèse de Doctorat. Faculté des Lettres, Université de Lausanne.

ACKNOWLEDGEMENTS

With many thanks to Prof. Eric Keller (LAIP) for his precious remarks, and to my European colleagues of COST-258 for their interest in this study.