Keller, E., & Zellner, B. (1996). A timing model for fast French. *York Papers in Linguistics*, 17, University of York. 53-75

A Timing Model for Fast French

Eric Keller and Brigitte Zellner

Laboratoire d'analyse informatique de la parole (LAIP)

Informatique — Lettres

Université de Lausanne

CH-1015 LAUSANNE, Switzerland

Abstract

Models of speech timing are of both fundamental and applied interest. At the fundamental level, the prediction of time periods occupied by syllables and segments is required for general models of speech prosody and segmental structure. At the applied level, complete models of timing are an essential component of any speech synthesis system.

Previous research has established that a large number of factors influence various levels of speech timing. Statistical analysis and modelling can identify order of importance and mutual influences between such factors. In the present study, a three-tiered model was created by a modified step-wise statistical procedure. It predicts the temporal structure of French, as produced by a single, highly fluent speaker at a fast speech rate (100 phonologically balanced sentences, hand-scored in the acoustic signal). The first tier models segmental influences due to phoneme type and contextual interactions between phoneme types. The second tier models syllable-level influences of lexical vs. grammatical status of the containing word, presence of schwa and the position within the word. The third tier models utterance-final lengthening.

The complete segmental-syllabic model correlated with the original corpus of 1204 syllables at an overall $r = 0.846$. Residuals were normally distributed. An examination of subsets of the data set revealed some variation in the closeness of fit of the model.

The results are considered to be useful for an initial timing model, particularly in a speech synthesis context. However, further research is required to extend the model to other speech rates and to examine inter-speaker variability in greater detail.

1. Introduction

Previous research on the prediction of speech timing have documented influences at three major levels: the phoneme or segmental, the syllabic and the phrase level.

1.1. Models Based on the Prediction of Segmental Durations

The most influential statistical model for spoken French text has probably been the model proposed by O'Shaughnessy (1981, 1984). On the basis of numerous readings of a short text containing all phonemes of French, a model of durations of acoustic segments suitable for synthesis by rule was proposed. In this model, 33 rules for the modification of segment duration according to segment type, segment position and phoneme context served to specify basic phoneme durations.

For sound classes that did not involve prepausal lengthening, the model was able to predict the durations for 281 segments of a text with a standard deviation of 9 ms. But it was less accurate for the prediction of prepausal vowel durations, because of the greater variability of segments in such positions. Moreover, this model was not able to predict silent inter-lexical pauses.

O'Shaughnessy's statistical model is constructed around the hypothesis that speech timing phenomena can be captured by the segment, as if this unit "possesses an inherent target value in terms of articulation or acoustic manifestation" (Fujimura, 1981). However, recent measures have indicated that syllable-sized durations are generally less variable than subsyllabic durations, and thus may represent more reliable anchor points for the calculation of a

general timing structure than segmental durations (Barbosa and Bailly, 1993; Keller, 1993; Zellner, 1994). The taking into account of explicit syllable-level information is further supported by the observation that stress variations and variations of speech rate tend to modify at least syllable-sized units.

Barkova's model (1985, 1991) attempts to solve these deficiencies by adding calculated coefficients to the formula for predicting segment durations:

$$Dur\ Seg = DurI + k_{Syll} + k_{Ac}$$

where *DurI* is the intrinsic duration of the segment, k_{Syll} is a syllabic coefficient, and k_{Ac} an accentuation coefficient. The exact manner in which these coefficients are obtained is not described; it is only noticed that they can vary from a minimum to a maximum interval, according to the position of the segment in the speech chain, and according to the acoustic properties of the speech sound.

The syllabic coefficient depends on the nature of the word (lexical / grammatical), and on the position in the word (initial, medial, final syllable). The coefficient of accentuation depends on the next consonant, on the presence / absence of a syntactic boundary in the case of a final vowel, or on the presence / absence of clusters in the case of a final consonant, as well as on the syllabic structure near a pause.

According to Bartkova, a comparison of predicted and measured durations in 10 sentences gives rather good predictions, since the mean difference on segmental duration is about ±15 ms.

However, it would seem that beyond the opacity of the coefficients, a divergence between predicted and measured durations of the order of 15 to 30 ms can be a major handicap for short segments. In our corpus, for example, the mean duration for /d/ was 50 ms. In the case of such a short phoneme, a 15-30 ms divergence would correspond to an error of 30-60% with respect to its measured duration.

1.2. Required Macro-Timing Information

Since the segmental unit cannot capture the overall temporal structure of speech, the next level which can be expected to encapsulate temporal phenomena is the syllable. This appears to be a good candidate. According to some psycholinguists, it is considered to be the minimal perception unit, and according to numbers of phoneticians and phonologists, it is the minimal unit of rhythm (see Delais, in press).

It has been shown that quite a number of parameters are involved in variations of syllabic duration. The most important are: the position in the prosodic group, the position in the word, degree of stress, the length of the prosodic group, the position according to the stressed syllable, the position according to the local speech rate (as measured by cycles of speeding up and slowing down), semantic focus, proximity of syntactic boundaries, the status of the word (lexical or grammatical), and emotional factors (Bartkova, 1985, 1992; Campbell,1992; Delais, 1994; Duez, 1985, 1987; Fant and al, 1991; Fònagy, 1992; Grégoire, 1899; Grosjean et al, 1975, 1983; Guaïtella, 1992; Konopczynski, 1986; Martin, 1987; Mertens, 1987; Monnin et al, 1993; Pasdeloup, 1988, 1990, 1992; Wenk et al, 1982; Wunderli, 1987). Some of these factors may be redundant; for instance, in many cases of read text, lexeme-final position may be redundant with phrase-final position.

In view of existing information, it thus seems best to issue from segmental predictions, and to consider syllabic information as additional information, which is not captured at the segmental level. One of the important points to consider in the present study will be the selection of non-redundant and relevant information.

Beyond the syllabic level, it is likely that a good predictive model will eventually need to incorporate further information at the word or the phrase

level. For example, the prediction of pauses for slow speech requires phrasal knowledge, which is not captured at the segmental or at the syllabic level. In the area of word group boundaries in French speech, a great deal of work has been accomplished to determine the nature of these groups — syntactic groups, prosodic groups, rythmic groups, intonational groups, the congruence between these labels — and to calculate the automatic generation of such groups and potential inter-group pauses (Delais, 1994; Grosjean et al, 1975; Keller et al., 1993; Martin, 1987; Monnin et al, 1993; Pasdeloup, 1988; Saint-Bonnet et al, 1977). These effects will have to be integrated into a general timing model for a given language, but were not taken into account in the present study.

In the current study, the objective was to account for a single speaker's syllable durations with the smallest number of segmental and syllabic factors. At each succeeding level, relevant parameters were chosen so as to explain the greatest proportion of the variance in the residue of the previous analysis. In this manner, a three-tier model, based successively on segmental, syllabic and phrasal information, was constructed.

2. Method

2.1. Procedure

2.1.1. The corpus

A highly fluent speaker of French (a professor of French literature) was recorded with 277 sentences, the first 100 of which were analysed for the present study. The speaker was instructed to speak quite rapidly, with a normal, unexaggerated intonation. The resulting readings have generally been judged by listeners as highly intelligible and well-pronounced. No dialectal particularities were noted.

Recording occurred in studio conditions on DAT-tape. The digitized data was transferred to Macintosh computer and was downsampled to 16 kHz.

2.1.2. Time labelling

The time occupied by each phoneme was labelled with the Signalyze™ program according to detailed instructions on how to handle phoneme-to-phoneme transitions (Thévoz and Enkerli, 1994). Specifically, transitions in the acoustic corpus was analyzed according to three articulatory levels: labial, lingual and laryngeal. For example, the coarticulatory overlap at the /e/-/s/ transition was marked by symbols representing the following events: “onset of friction, associated with the lingual level”, followed at a given time interval by an “offset of fundamental frequency, associated with a cessation of vocal cord activity”. The following possible states were distinguished:

1. Labial system: aperture, occlusion, friction, burst, error
2. Lingual system: aperture, occlusion, friction, burst, palatal, transient movement, error
3. Laryngeal system: aperture, occlusion, transient movement, diminution, error
4. “Error” refers to any state that occurs inadvertently, such as during a speech error.

To examine the reliability of transcriptions, two judges compared judgements concerning how and where points of transition between inferred articulatory states were to be marked. Two measures of interjudgemental agreement were used:

Robustness (agreement in the application of criteria to state transition), scored 1 = low agreement, 2 = agreement in general, but some further discussion required, and 3 = excellent agreement.

Precision, scored 1 = more than two Fo periods difference, 2 = 1-2 Fo periods difference and 3 = less than 1 Fo period difference in measurement.

Both measures showed good to excellent interjudgemental agreement. Over the 50 types of state transitions examined, there were no cases of low robustness or low precision. The average robustness was 2.53 and the average precision was 2.68.

A total of 4544 phonemes and 1203 syllables were analyzed in this manner.

2.2. Analysis and Results

A modified step-wise statistical regression technique was used to develop a well-fitting model of this speaker's timing behaviour. In accordance with previous observations on factors that influence speech timing, it was decided to model three major levels: the segmental, the syllabic and the phrase level. In step-wise fashion, each succeeding level was made to model the residue left by the previous level. Three different models were thus established, the Segmental, the Syllabic and the Phrase Model (Figure 1).

(Figure 1.)

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones. An initial issue concerned the calculation of segmental duration in a corpus where coarticulatory transition zones are marked explicitly. Does phoneme duration correspond to the zone of the signal which is unambiguously marked for a given phoneme (zone B in figure 2), or does it include one or both zones of coarticulatory overlap with adjoining phonemes (zones A and C in figure 2)?

The issue was resolved with reference to durational variation. The combination of zones A, B and C (with an average coefficient of variation of 0.375) turned out to be systematically less variable than the unambiguous zone B (with an average coefficient of variation of 0.412) (see Table 1). Also, combinations of zones A and B, or of B and C, were less variable than zone B alone. The transition zones can thus be considered to be “buffer zones” whose function, in part, may well be to “regularise” phoneme duration. For the purpose of the present research it was thus decided to consider the combined duration of A, B and C as “phoneme duration”. Syllable durations were constructed from phoneme durations by taking into account transitional overlaps. As a net effect, the segmental duration entering the statistical modelling procedure is slightly more regular than more commonly measured phoneme durations. Nevertheless, it is not believed that the modelling results of the present study seriously depend on this manner of proceeding; the size and resilience of the measured effects suggest that as long as transitions are handled in systematic fashion, the predictive pattern should remain largely identical.

(Figure 2.)

(TABLE I.)

(Figure 3.)

(TABLE II.)

Segmental transformation and grouping. Raw segment durations were non-normal in their distribution. Among the common transformations, the log10 transformation produced the closest approximation to a normal distribution (Figure 3). All calculations of the segmental portion of the model were thus performed on log10-transformed durations.

Subsequent to transformation, phonemes were grouped according to their mean durations and their articulatory definitions. Eight classes could be identified (Table 2). Groups showed roughly comparable coefficients of variation, and an inspection of histograms and normal probability plots showed roughly normal distributions for all classes whose *N* was greater than 100.

To test Model 1 in the syllabic context, square root-transformed syllable durations were calculated on the basis of coefficients produced by the linear model for segmental durations, and by taking into account mean durations of phoneme-to-phoneme transitions. These calculated syllable durations were compared to the square root-transformed measured syllable durations. The correlation coefficient was $r = .647$ ($N = 1203$, $p<.0001$) (Figure 5). The residue from the model (= observed - predicted) was termed “Delta 1” and served as the basis for further factorial modelling at the syllabic level.

A Linear Model for Segmental Durations. Using the Data Desk® statistical package on the Macintosh, a general linear model for discontinuous data (based on an ANOVA) was calculated with partial (non-sequential, Type 3) sums of squares. The following main and interaction factors (up to two-way[1]) were postulated:

Duration (log10(ms)) = constant + previous type + current type + next type + previous type * current type + current type * next type + previous type * next type

(TABLE III)

(**Figure 4)**

(Figure 5.)

In the partial sums of squares solution, all factors were significant at $p<.05$, with the exception of "previous type" and "next type", taken alone, and the interaction term "previous type * next type" (Table 3). The residual error was 101.137 / 196.070 = 0.516, that is, the model explained about 48.4% of the variance. Expressed in terms of a Pearson product-moment correlation, the model's predicted segmental durations correlated with empirical phoneme durations at $r = 0.696$.

Syllable Durations and Delta 1. Another means of testing the model is a comparison with measured syllable durations. In contrast to phoneme durations, where a log transformation served to provide roughly normal distributions, square roots had to be applied to measured syllable durations in order to approximate normal distributions (Figure 4).

2.2.2. Model 2: The Syllabic Model

Syllabic Factors Predicting Delta 1. After considerable experimentation with a variety of factors described in the literature, a three-factor model, including two-way interactions, was retained for analysis:

delta 1 = constant + function + position + schwa + function * position + function * schwa + position * schwa,

where *"function"* distinguishes whether the syllable is found in a lexical or a function word, *"position"* identifies three types of position in the word which are

(1) “monosyllabic and polysyllabic-initial”, (2) “polysyllabic pre-schwa” and (3) “other”, and “schwa” indicates whether or not a schwa is present in the syllable. Again, a general linear model for discontinuous data was calculated with partial (Type 3) sums of squares. The results of the ANOVA showed that all main and interaction factors were significant at $p<.05$ (Table 4). The residual error of 3277.29 / 5432.93 = .6 indicated that the model explained 40% of the variance in Delta 1.

(TABLE IV.)

2.2.3. Model 2 and Delta 2. Syllable durations obtained from the segmental model were combined with those from the present linear model for Delta 1 to produce the Syllabic Model (Model 2). The predictions correlated with observed square root-transformed syllable durations at $r = .723$ (N=1203) (Figure 6). The residual data was termed Delta 2.

2.2.4. Model 3: The Phrase Model

Inspection of the predictions of Models 1 and 2 (Figures 5 and 6) showed a noticeable deviation from the regression line in the higher values. Specifically, these models underestimated most syllable durations in the > 280 ms range. Furthermore, an examination of Delta 2 revealed that the residual error was most pronounced for utterance-final syllables ending in a consonant. Consequently, a correction term was calculated, which was applied to such syllables in Model 3.

The predictions of Model 3, which incorporates segmental and syllabic modelling as well as the phrase-final correction term, correlated with the observed square root-transformed syllable durations at $r = .846$ (Figure 7). The residual values from Model 3 vary quasi-randomly around 0. At the present time,

it appears that only more sophisticated rules for the generation of the schwa vowel may still be able to improve this model's predictive capacity to some degree.

(Figure 6)

(Figure 7.)

2.3. Stability

The Phrase Model was examined for its predictive stability by performing Pearson product-moment correlations between various subsamples of the data and the model's prediction. The resulting data is presented in Table 5.

(TABLE V)

It can be seen that the model's predictive capacity varies considerably from one subset to the next. For example, the correlation was only .726 for the fourth slice of 100 syllables in the set, while it had been .884 for the first slice. Even when slices of 300 syllables are compared, considerable variability prevails. The reasons for these instabilities are presently being investigated.

3. Discussion

By a modified step-wise procedure, a general model for the prediction of the fast-speech performance of a highly fluent speaker of French was constructed. The initial model incorporates segmental information concerning type of phoneme and proximal phonemic context. The subsequent model adds information on whether the syllable occurs in a function or a lexical word, on whether the syllable contains a schwa and on where in the word the syllable is located. The final model adds information on phrase-final lengthening. The effects of these three levels are demonstrated on a single sentence in Figure 8. In view of current discussions surrounding segmental and syllabic contributions to timing models, it is interesting to note that segmental information accounts for a major portion of the variance explained by the model. As Figure 8 shows, segmental information alone successfully predicts several cases of major syllable lengthening.

The overall correlation of 0.846 between predictions of Model 3 and the data set from which the model is derived is encouraging. This correlation level corresponds roughly to the average inter-speaker correlation of $r = 0.833$ for phrase-final syllable durations, as measured between the readings of a short text by 12 speakers in the Caelen-Haumont corpus (Caelen-Haumont, 1991; see Keller, 1994). This means that the model behaves as differently from its target data as one natural speaker would behave with respect to another speaker. Although this may be an acceptable initial predictive level for synthesis purposes, further improvements in the modelling would be welcome. Preliminary indications suggest that such improvements may come about through predictions of the presence vs. the absence of schwa, through explicit predictions of the effects of speech rate manipulation, and in longer texts, through a better modelling of pauses. Further information on possible improvements may also be gained through an examination of cases of high delta 3 values in subsets of the present data set. These effects are currently being studied.

It is worth noting that in the present fast-speech corpus, no phrase-level effects were identified, other than phrase-final lengthening. This is in contrast to our findings on the production of French at a normal speech rate, where a fairly systematic increase of lexeme-final syllable durations was observed over the extent of the prosodic phrase (Keller *et al.*, 1993). It seems likely that in conditions of considerably accelerated speech rate, our speaker sacrificed some of the "niceties" of phrase-internal timing modulation, and limited himself to a single, phrase-final durational marker.

Considerably more work also needs to be done before the generalisability of the present model can be tested. The examination of the model's stability has shown that predictions begin to show comparable strength at about 300 syllables or more. Consequently, systematic testing of these predictions for another

speaker would involve a completely new research study. Nevertheless, a few quick examinations of predictions for another speaker's sentences suggest that the model may indeed be generalisable to more than one speaker of French (Figure 9).

(Figure 8.)

(Figure 9.)

Acknowledgements

The authors are grateful to the following members of the LAIP team for their invaluable assistance in scoring and creating the present corpus: Nicolas Thévoz, Alexandre Enkerli, Hervé Mesot, Cédric Bourquart, Nicole Blanchoud, and Thomas Styger. Particular thanks go to Prof. J. Local (York University, UK) for his many ideas and his encouragement. Prof. A. Wyss of the University of Lausanne is cordially thanked for his participation as a subject for this study. This research is supported by the Fonds National de Recherches Suisses (Projet Prioritaire en informatique and ESPRIT Speech Maps) and by the Office Fédéral pour l'Education et la Science (COST-233).

References

Barbosa, P., & Bailly, G. (1993). Generation and evaluation of rhythmic patterns for text-to-speech synthesis. Proceedings of an ESCA Workshop on Prosody (pp. 66-69). Lund, Sweden.

Bartkova, K. (1985). Nouvelle approche dans le modèle de prédiction de la durée segmentale. 14ème JEP (pp188-191). Paris.

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

Caelen-Haumont, G. (1991). Stratégies des locuteurs et consignes de lecture d'un texte: Analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodiques, Thèse d'Etat, Aix-en-Provence.

Campbell, W.N. (1992). Syllable-based segmental duration. Talking Machines. Theories, Models, and Designs (pp211-224). Elsevier Science Publishers.

Delais, E. (1994). Prédiction de la variabilité dans la distribution des accents et les découpages prosodiques en français. XXèmes Journées d'Etude sur la Parole (pp379-384). Trégastel.

Delais, E. (sous presse). Rythme et structure prosodique en Français. Proceedings of Congrès Annuel de l'Association pour l'Etude de la Langue française, Aix-Marseille

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.

Duez, D., Nishinuma, Y. (1985). Le rythme en français. Travaux de l'Institut de Phonétique d'Aix, 10, 151-169

Fant, G., Kruckenberg, A., Nord, L. (1991). Durational correlates of stress in Swedish, French and English. Journal of Phonetics. 19, 351-365.

Fònagy, I. (1992). Fonctions de la durée vocalique. In P. Martin (Ed.), Mélanges Léon. (pp. 141-164). Editions Mélodie-Toronto.

Fujimura, O. (1981). Temporal organisation of articulatory movements as a multidimensional phrasal structure. Phonetica, 38, 66-83.

Grégoire, A. (1899). Variation de la durée de la syllabe en français. La Parole, 1, 161-176.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. Linguistics, 21. 501-529.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français. Phonetica, 31, 144-184.

Keller, E. (1993). Prosodic Processing for TTS Systems: Durational Prediction in English Suprasegmentals. Final Report, Fellowship, British Telecom.

Keller, E., Zellner, B., Werner, S., & Blanchoud, N. (1993). The Prediction of Prosodic Timing: Rules for Final Syllable Lengthening in French. Proceedings, ESCA Workshop on Prosody (pp. 212-215). Lund, Sweden.

Keller, E. (1994). Fundamentals of phonetic science. In E. Keller (ed.), Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art and Future Challenges (pp. 5-21). Chichester, UK: John Wiley.

Konopczynski, G. (1986). Vers un modèle développemental du rythme français: Problèmes d'isochronie reconsidérés à la lumière des données de l'acquisition du langage. Bulletin de l'Institut de Phonétique de Grenoble, 15. 157-190.

Martin, Ph. (1987). Structure rythmique de la phrase française. Statut théorique et données expérimentales. Proceedings des 16e JEP (pp255-257). Hammamet.

Mertens, Piet. (1987). L'intonation du français. De la description linguistique à la reconnaissance automatique. Thèse doctorale, Katholieke Universiteit Leuven.

Monnin, P & Grosjean, F. (1993). Les structures de performance en français: caractérisation et prédiction. L'Année Psychologique, 93, 9-30.

O'Shaughnessy, D. (1981). A study of French vowel and consonant durations. Journal of Phonetics, 9, 385-406.

O'Shaughnessy, D. (1984). A multispeaker analysis of durations in read French paragraphs. Journal of the Acoustical Society of America. 76, 1664-1672.

Pasdeloup, V. (1988). Analyse temporelle et perceptive de la structuration rythmique d'un énoncé oral. Travaux de l'Institut de Phonétique d'Aix, 11, 203-240.

Pasdeloup, V. (1990). Organisation de l'énoncé en phases temporelles: Analyse d'un corpus de phrases réitérées, (pp 254 - 258). 18émes Journées d'Etudes sur la Parole. Montréal, 28 - 31 Mai.

Pasdeloup, V. (1992). Durée intersyllabique dans le groupe accentuel en Français. Actes des 19émes Journées d'Etudes sur la Parole. (pp531-536). Bruxelles.

Saint-Bonnet, M., Boe, J. (1977). Les pauses et les groupes rythmiques: leur durée et disribution en fonction de la vitesse d'élocution. VIIèmes Journées d'Etude sur la Parole, (pp337- 343). Aix en Provence.

Thévoz, N., & Enkerli, A. (1994). Critères de segmentation: Rapport intermédiaire. LAIP-Lausanne.

Wenk, B. J. & Wiolland, F. (1982). Is French really syllable-timed? Journal of Phonetics, 10, 177-193.

Wiolland, F. (1984). Organisation temporelle des structures rythmiques du français parlé. Etude d'un cas. Rencontres régionales de Linguistique, BLLL (pp293 - 322).

Wunderli, P. (1987). L'intonation des séquences extraposées en français. Tübingen: Narr, 1987.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State-of-the-Art and Future Challenges (pp. 41-62). Chichester, UK: John Wiley.

Tables and figures

2.2. Analysis and Results

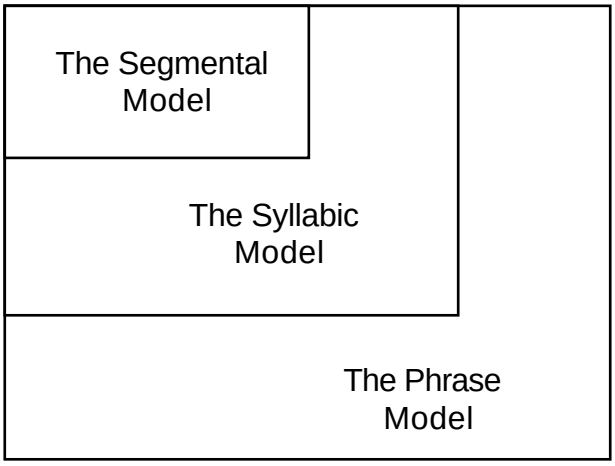

Figure 1. The Segmental, Syllabic and Phrase Models. Each subserquent model incorporates the modelling effects of the previous level.

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones.

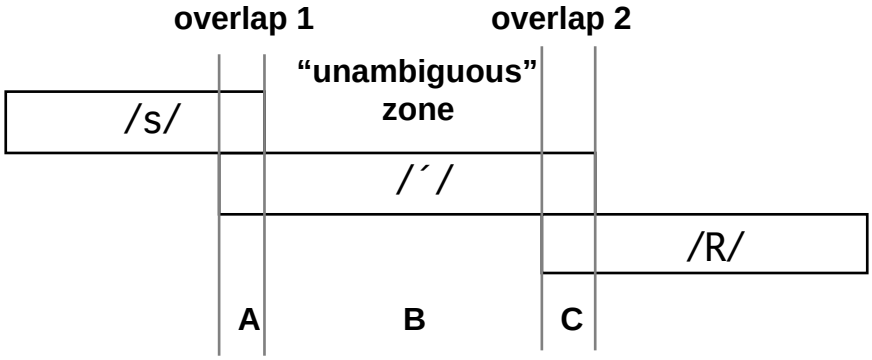

Figure 2. What constitutes a phoneme? B is a portion of the signal that is unambiguously marked for the phoneme /´/, while A and C are transitory zones with adjoining phonemes.

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones.

TABLE I. Coefficients of variation for zones A, B and C as well as various combinations of these zones

	A	**B**	**C**
Average coefficient of variation (s.d./ mean) for 34 phonemes	1.6379	0.4123	1.7472
	A + B	**B + C**	**A + B + C**
Average coefficient of variation for 34 phonemes	0.3916	0.3933	**0.3751**

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones.

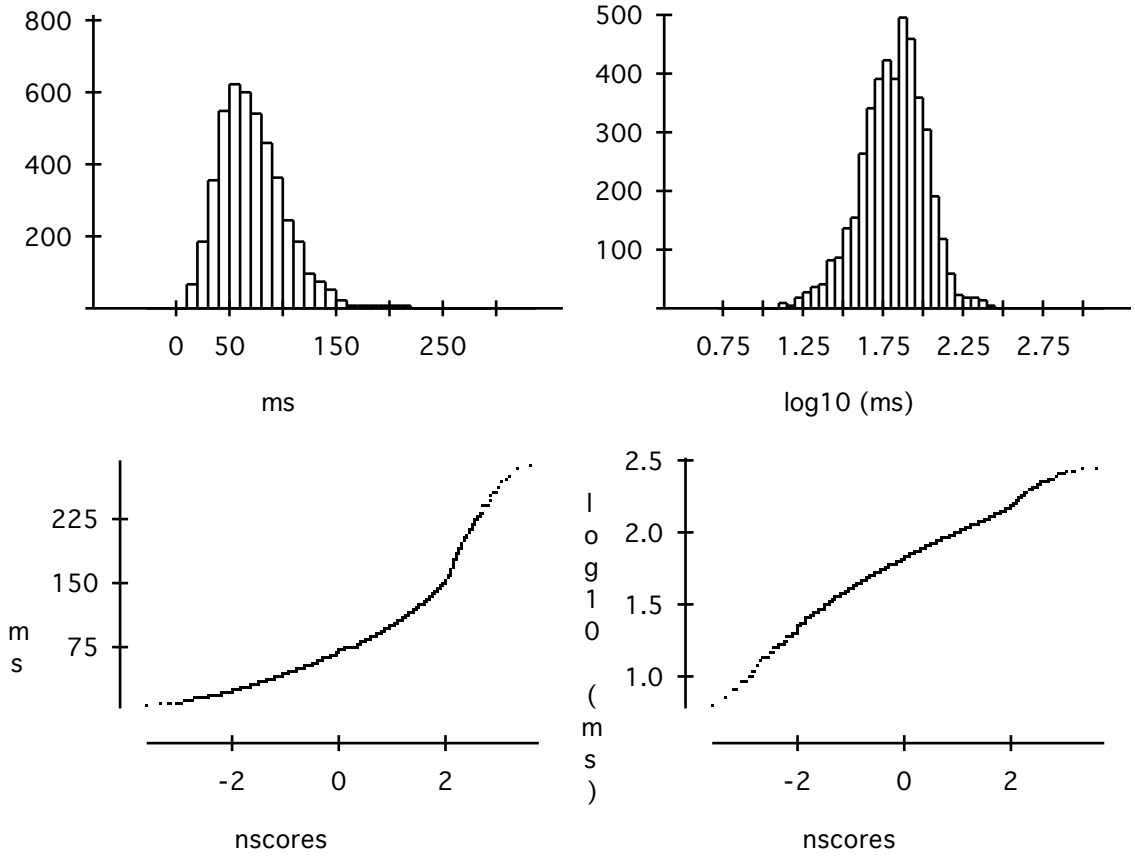

Figure 3. The distribution of segment durations before and after the log 10 transformation. Above: histograms, below: normal probability plots.

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones.

TABLE II. Mean durations for phoneme classes (N = 4544)

Phoneme type	Name	Mean duration (ms)	Coefficient of variation (s.d. / mean)	Frequency (*N*)
œ, Ø	AntRound	109.45	0.4881	71
ßsf	Fric	105.17	0.2708	357
œ~, ´~, a~, o~	Nas	97.78	0.3585	334
o	PostMidRnd	94.92	0.3130	60
p, t, k	UnvPlos	92.94	0.3475	504
a, e, ´, ø, u, i, y	OthVow	69.62	0.4089	1557
b, z, m, ˜, g, v, Ω, n, d, ÷	VcdCons	61.72	0.3669	892
R, j, w, l, ¥	SemiVLiquids	43.63	0.4908	769
Mean		90.23	0.3648	539

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones.

Segmental transformation and grouping.

TABLE III. The Segmental Model: Analysis of Variance for Segmental Data (N = 4544) Using Partial Sums of Squares

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	14903.8	14903.8	642500	≤ 0.0001
previous	8	0.123239	0.015405	0.66410	0.7236
current	7	3.13402	0.447717	19.301	≤ 0.0001
next	8	0.267002	0.033375	1.4388	0.1748
previous * current	50	3.24144	0.064829	2.7948	≤ 0.0001
current * next	50	5.04499	0.100900	4.3498	≤ 0.0001
previous * next	60	1.79531	0.029922	1.2899	0.0665
Error	4360	101.137	0.023197		
Total	4543	196.070			

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones.

Segmental transformation and grouping.

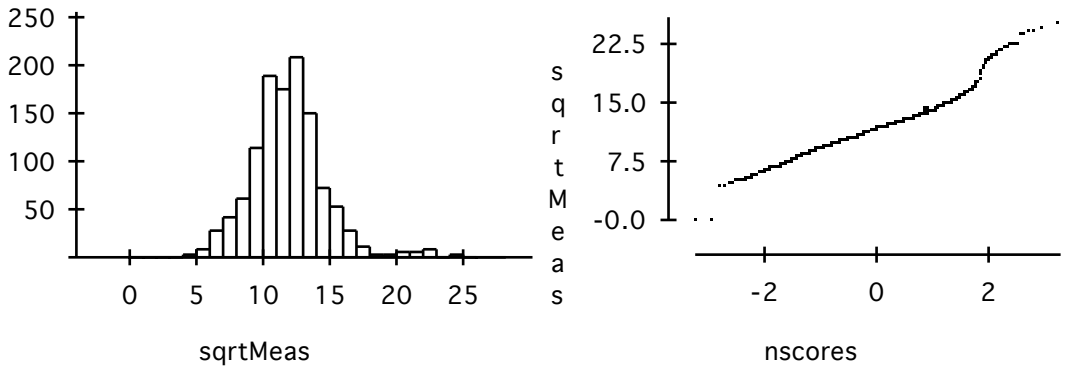

Figure 4. Syllable durations in ms were square-root transformed in order to approximate a normal distribution.

2.2.1. Model 1: The Segmental Model

Segmental Durations and Overlap Zones.

Segmental transformation and grouping.

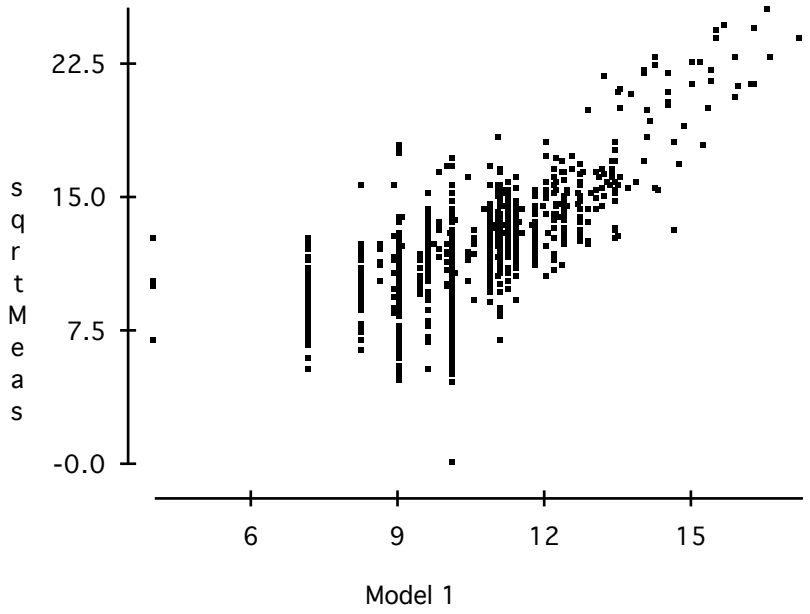

Figure 5. Prediction of the Segmental Model (Model 1): Syllable durations predicted exclusively on the basis of segmental durations (r = .647). Values are in sqrt(ms).

2.2.2. Model 2: The Syllabic Model

Syllabic Factors Predicting Delta 1.

TABLE IV. Analysis of Variance for Delta 1 (N = 1203) Using Partial Sums of Squares

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	2663.53	2663.53	969.58	≤ 0.0001
function	1	176.508	176.508	64.252	≤ 0.0001
position	2	98.5753	49.2877	17.942	≤ 0.0001
schwa	1	149.296	149.296	54.347	≤ 0.0001
function * position	2	97.3872	48.6936	17.725	≤ 0.0001
function * schwa	1	27.5860	27.5860	10.042	0.0016
position * schwa	2	63.0467	31.5234	11.475	≤ 0.0001
Error	1193	3277.29	2.74710		
Total	1202	5432.93			

2.2.4. Model 3: The Phrase Model

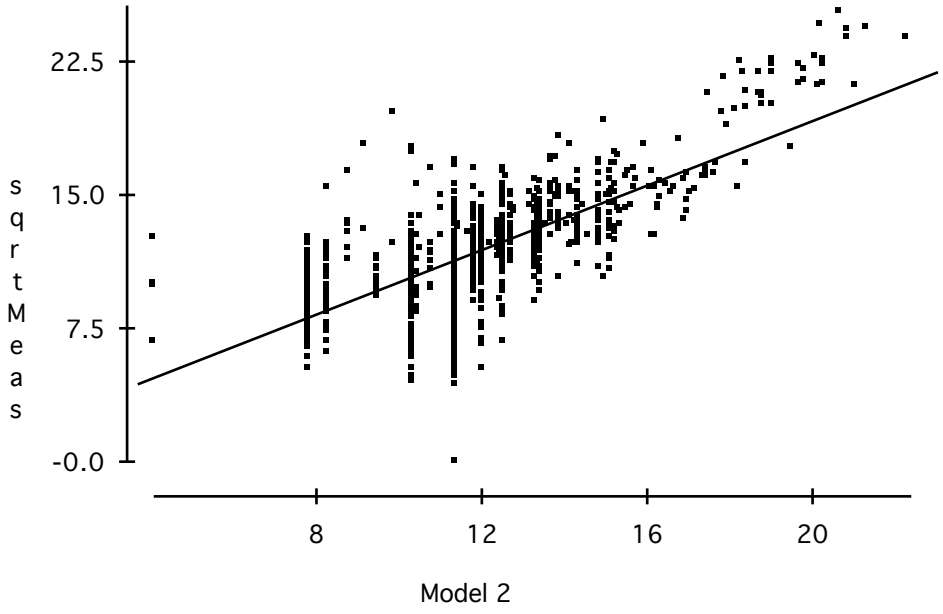

Figure 6. Prediction of the Syllabic Model (Model 2): Syllable durations predicted on the basis of segmental durations and syllable-level factors (r = .723). Values are in sqrt(ms).

2.2.4. Model 3: The Phrase Model

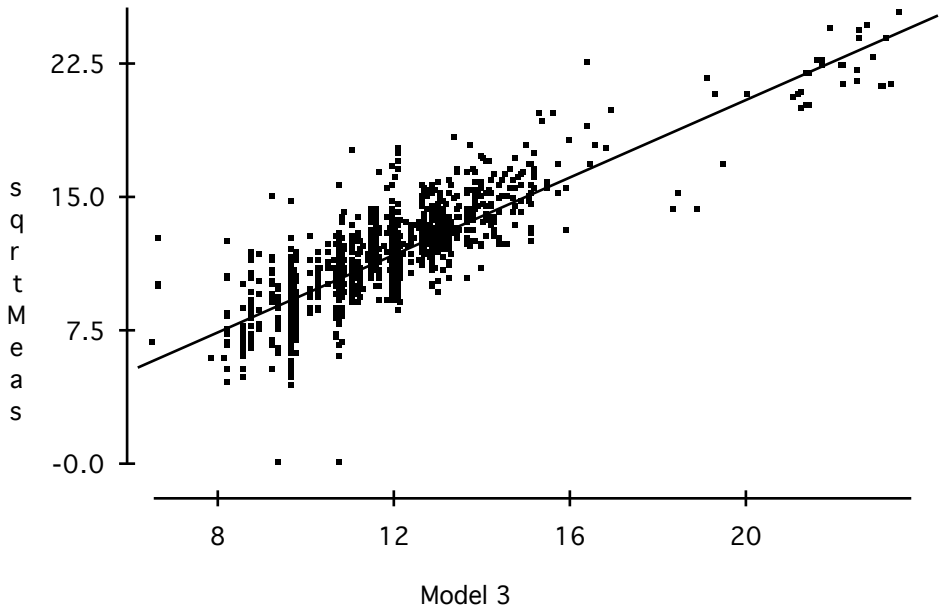

Figure 7. Prediction of the Phrase Model (Model 3): Syllable durations predicted on the basis of segmental durations, syllable-level factors and phrase-final lengthening (r = .846). Values are in sqrt(ms).

2.3. Stability

TABLE 5. Pearson Product-Moment Correlations between Various Subsets of the Dataset and the Phrase Model's Prediction

	slices of 50 syllables	slices of 100 syllables	slices of 200 syllables	slices of 300 syllables
1st slice	0.9	0.884	0.878	0.869
2nd slice	0.87	0.872	0.789	0.805
3rd slice	0.853	0.852	0.838	0.874
4th slice	0.89	0.726	0.885	0.838
5th slice	0.866	0.823	0.841	
6th slice	0.852	0.868	0.838	

3. Discussion

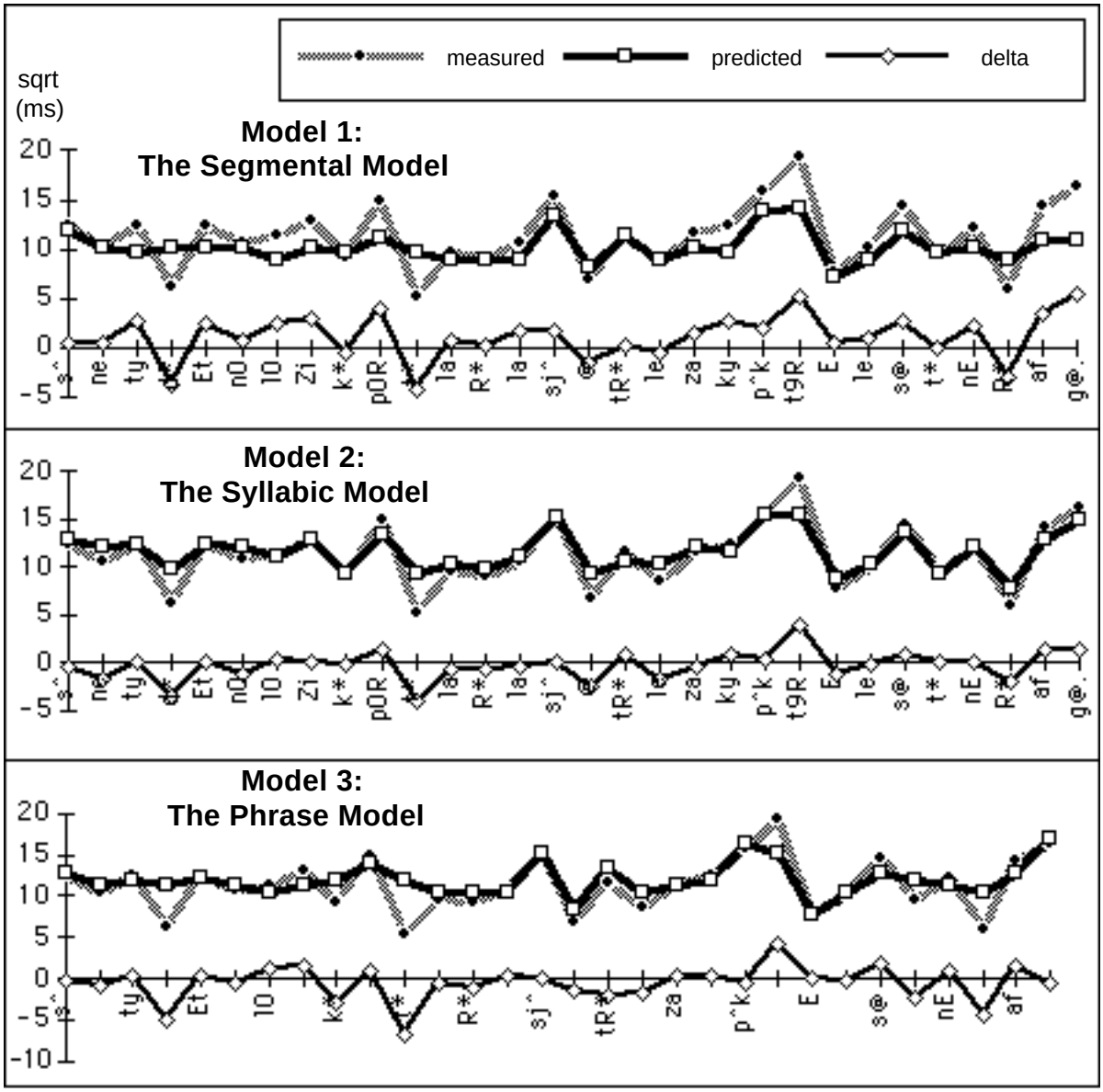

Figure 8. A comparison of predictions of the three models and measured syllable durations for the sentence "Son étude ethnologique porte sur la relation entre les acupuncteurs et les centenaires afghans".

3. Discussion

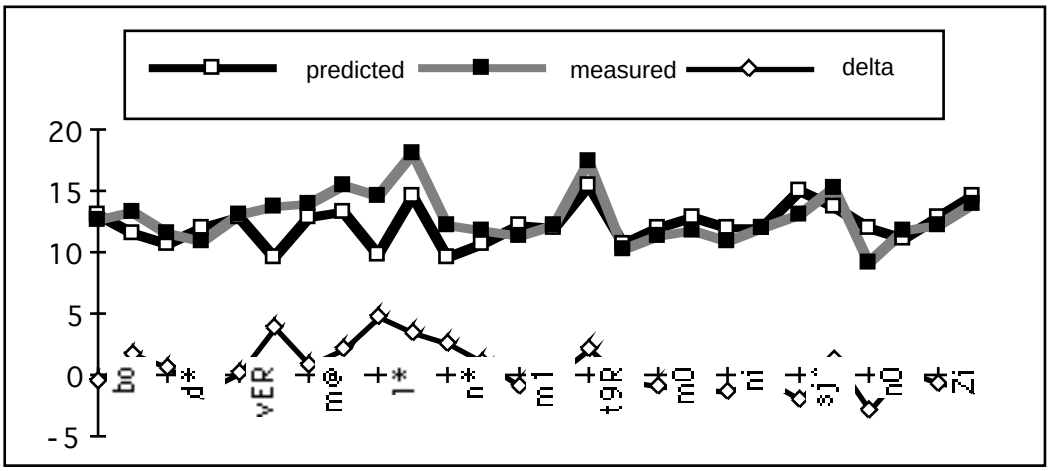

Figure 9. A comparison of predictions of Model 3 and the measured syllable durations of another speaker of French for the fast reading of the sentence “Beaucoup de gouvernements voient le CERN comme un moteur de modernisation technologique”.

Footnotes

[1] For reasons of insufficiency in per-cell observations, calculation complexity and theoretical difficulty of interpretation, three-way interactions were not calculated.