Modeling Shortening and Lengthening in Connected Speech

Antônio R.M. Simões

The University of Kansas at Lawrence

1. INTRODUCTION

Because duration is not linguistically significant in Brazilian Portuguese (hereinafter BP), studies of BP duration do not have immediate application to the teaching of BP. Studies of BP duration will primarily benefit applied areas such as artificial speech, automatic recognition of the speaker, and other applied areas of this type. The teaching of BP will eventually benefit, after other parameters besides duration, have been studied. Scholars such as D.H. Klatt (1974), Lindblom (1979), have supported this type of study for the purpose of building speech models. The understanding of temporal patterns of sound segments can be considered as a first step toward constructing speech models. The major purpose of the present analysis is to provide BP with a model as effective as D.H. Klatt's (1976) model for American English.

A number of scholars have attempted to develop models to describe the duration patterns of speech sounds in several languages, e.g. Lindblom and Rapp (1973), in their study of Swedish, and Klatt (1976), in his study of American English. Their work provided them with a starting point for the analysis and modeling of other parameters such as the fundamental frequency and loudness at both the perceptual and acoustical level. Similar expectations and procedures apply to the present work. The experimental protocol of the present investigation is fully described in my doctoral work related to the present analysis (Simões:1987).

The model being developed here is intended (1) to uncover the factors or language components that affect sound-segment duration, (2) to determine if these factors interact in series or simultaneously, and (3) to describe vowel-duration change by means of rules for the specific case of BP. We have to look at segment duration as two linguistic phenomena: shortening and lengthening in terms of an unmarked duration. I would like to mention that the use of "unmarked" here is synonymous with Klatt's (1976) use of the term "inherent." I have proposed in my former investigation the following definition of an unmarked duration: the *unmarked duration* of a given sound-segment μ is the *median* duration of all occurrences of that sound-segment μ in a text of over one-thousand words.

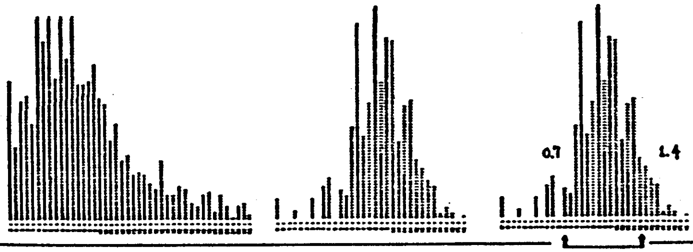
2. DISCUSSION

From the experimental analysis of the temporal organization of the three extreme vowels [i], [a], and [u], the relationship between lower level linguistic components (phonetics/ phonology) and higher level linguistic components (syntax, discourse) is studied. Data for this investigation were obtained from one informant, PM, from Rio de Janeiro who read a text for children containing over one thousand words.

Klatt's (1976) model, based on American English, is examined in order to know how it could be adapted to BP. From the standpoint of modeling BP vowel duration and based on the acoustic and statistical analysis results, rules of vowel duration change are proposed for the specific case of the informant in this study, These rules apply from inner domains (phonemes, syllables) to outer domains (utterances).

Analyses of all data from these vowels in the corpus consistently show functions positively skewed, namely lower values concentrate on the left side of the abcissa, in such a way that higher values will spread rightward. A parameter such as duration is expected to be positively skewed. For this reason, the median was chosen as the measurement type for this parameter, instead of the mean. In figure 1A a common pattern of the data is shown. Besides the use of the median in the establishement of an unmarked duration, I have found that the common logarithm of the skewed function can be used as well. Figures 1A-1C illustrate this additional technique. Common logarithms of positively skewed distributions were taken so that a normal distribution such as the one seen in 1B would be obtained. Figure 1C shows how significant lengthening can be determined by considering values greater than 1.4 times the median duration of the segment studied. Shortening processes appear on the left side of the abcissa and the cutoff point for significant shortening processes is 0.7.

Figure 1: The transformation of a positively skewed distribution into a normal distribution for the establishment of unmarked (inherent) and marked (significant reduction and expansion) sound segment duration.



The unmarked (inherent) values are obtained from the grand median, namely all stressed and unstressed positions. The grand medians (unmarked duration) are given in Table 1 and the measurements in stressed and unstressed positions are given in Table 2. Measurements from tables 1 and 2 are used in the application of the model (equation 1).

Table 1: PM's unmarked durations.

		items
i,a,u	60ms	1275
i	56ms	379
a	64ms	636
u	48ms	260

Table 2: Vowel duration in stressed and unstressed position in PM's speech.

	+stress	items	-stress	items
i,a,u	120ms	239	52ms	1036
i	108ms	78	48ms	301
a	120ms	120	60ms	510
u	112ms	35	44ms	225

In the present study the speaker's speech is affected by phonological, word, syntactic and semantic factors in both lengthening and shortening processes. These remarks do not necessarily reflect how production processes take place. These conclusions might be related only to the model chosen here and to the purpose of synthesizing speech. These factors will be used in an adaptation of Klatts' (1976) model to predict and consequently to create rules for sound segment synthesis of duration in BP. The accuracy of Klatt's model is tested mainly around the linear equation $D_o = K \times (D_i - D_{min}) + D_{min}$ (eq 1) where D_o for duration output, is the duration sought at any given point in the text; K is a constant value for each phonological environment, each position in a word, each position in a sentence, and each type of semantic factor; D_i is the unmarked duration for each sound-segment; and D_{min} is the minimum reduction the inherent sound-segment duration can have (see Klatt:1976:1215).

The value of K is established by operating on the original equation (1) making equation (2) $K = D_o - D_{min} / D_i - D_{min}$. Due to possible complete reductions of sound segments in PM's speech at the end of a word and especially at the end of a word at the end of a sentence, the rules proposed in this study attempt to reflect this phonological characteristic of PM, common in BP.

3. THE APPLICATION OF THE MODEL

The present work is based on results and analysis done exclusively at the acoustic domain. According to Klatt (1976:1208) "duration often serves as a primary perceptual cue in the distinctions between (1) inherently long versus short vowels, (2) voiced versus voiceless fricatives, (3) phrase-final versus non-final syllables, (4) voiced versus voiceless postvocalic consonants, as indicated by changes to the duration of the preceding vowel in phrase-final positions, (5) stressed versus unstressed or reduced vowels, and (6) the presence or the absence of emphasis." Conclusions (3), (5) and (6) are relevant to the speech of PM at the phonetic level. Conclusions (1), (2) and (4) do not apply to PM's speech because of the phonology and phonotactics of BP, namely BP vowels are not distinctive in terms of duration and most syllables in BP are open. Klatt suggests rules to reduce or expand sound-segments in English that apply recurrently from local to outer units according to the equation (1). While Klatt's work assesses the importance of the linguistic information by means of perceptual tests, the present work filters out significant linguistic information by means of statistical analysis. Klatt (1973b) already had preliminary attempts to apply a percent change model which applies several rules simultaneously, which demonstrated that this model failed. Lindblom and Rapp (1973:47) did propose a model for Swedish with recurrent rules going from outer units (sentences) into inner units (clauses and words). Lindblom et al (1981) then, moved to Klatt's approach going from inner to outer units. There are many more details in the Lindblom et at (1981) model that make it quite attractive to test for other languages as well. Because of the simplicity and effectiveness of Klatt's model, which uses only four rules and nine parameters to predict average vowel duration occurring in 56 situations (Klatt:1976:1217) the present study tested this model first.

The rules of the present study are listed below. They were established for the prediction of shortening and lengthening of the three vowels [i, a, u] according to equation (1). Although I proposed at first to have equation (1) operating in series from domain 1 to domain 4, I do not have an argument against models, in which all involved factors operate through one functional form, viz. simultaneously. I have not yet tested this possibility myself, but I hope to have an opportunity to run such rules in series and simultaneously. The value of K here is obtained by means of equation (2) and the unmarked duration of each vowel initializes each process in any position within a word in the case of applying the rules in series. The subsequent D_i values are the outputs (D_o) of the application of the rule just applied. **Level 1 (Sound segment domain)** - Rule 1: Initialization. From a given sound duration inventory the unmarked sound segment duration is set. D_i of PM's vowels are: [i] 56ms, [a] 64ms, [u] 48ms; Rule 2: If the phonetic voiceless fricative [s] follow the vowel within the same syllable, shorten the vowel by $K = -1.17$ (65% decrease). In case the consonant is [x], shorten the vowel by 25%, viz. $K = .17$. Rule 3: If a phonetic voiced consonant follows the vowel within the same syllable, no change, $K = 1.0$. **Level 2 (Word domain)** - Rule 4: If the vowel is in postonic position, decrease the vowel by 25%, viz. $K = .17$; Rule 5: If the vowel is in pretonic position, not preceded by a consonant, decrease it by 10%, i.e. $K = .67$. In case the vowel is preceded by a consonant, increase it by 42%, viz. $K = 2.4$; Rule 6: If the vowel is in inmediate pretonic position, increase it by 13%, viz $K = 1.43$; Rule 7: If the vowel is in stressed position, increase the vowel by 90%, viz. $K = 4$. **Level 3 (Sentence domain)** - Rule 8:

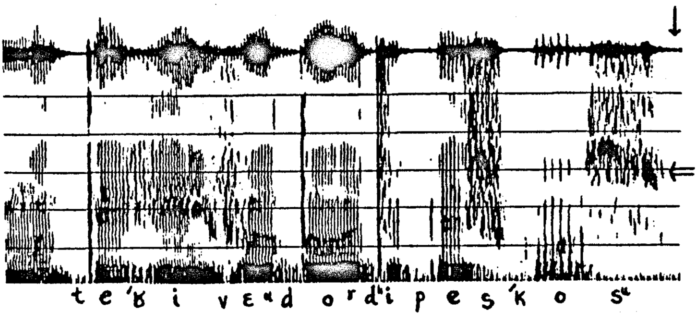
If the vowel is at the beginning of a sentence or a pause, no change, *K* = 1.; Rule 9: If the vowel is in sentence medial position, decrease by 13%, viz *K* = .57. Rule 10: If the vowel is in sentence final position without physical pause, increase the vowel by 20%, *K* = 1.67, but if the vowel is in a major final position, and a physical pause follows, increase the vowel by 32%, *K* = 2.07. **Level 4 (Semantic domain)** - Rule 11: If vowel is within a focused word, increase the vowel by 60%, making *K* = 3; Rule 12: If the vowel is in an exclamatory word, increase the vowel by 80% by making *K* = 3.7

Rules 10-12 do not apply to vowels in postonic position and stressed vowels are the only ones with a minimum duration. When comparing both sets of rules in Klatt (1976) and in the present work, one has to bear in mind that all but one rule in Klatt's model will *shorten* vowels for the inherent vowel duration in Klatt's model is the *longest* vowel (1976:1217). In the present model the phonological operations can either shorten to a complete reduction or lengthen the vowel segments.

4. CONCLUSION

This model and its duration changes rules are not intended to reflect mental processing. The preceding discussion gives an idea of the difficulties found in the first incursions in this type of analysis subsequent proposition of a model. I have already mentioned the need to use the median, instead of the longest vowel, because the BP syllables are normally CV sequences, especially in connected speech. Analyses of the signal from connected speech, as shown in Figure 2 will show that in connected speech very often the whole formant structure of the nucleus disappears.

Figure 2: The lack of formant structure where a vowel is supposed to be (↓). The transition in the preceding sound (⇐) seems to contain that vowel.



Thanks to a reanalysis of the spectrograms produced for this study and to an article by Parker and Diehl (1981), the interpretation of the temporal organization of BP vowels at the discourse level may acquire a new view. As depicted in Figure 2, vowels in connected speech commonly *enter* the preceding consonant in the speech of my informant. It may be necessary for BP, then, to eliminate the element D_{min} from the equation (1) making it a much simpler equation, viz.

equation (3) $D_o = K * D_i$

or, if equation (1) is to be maintained, the duration of a vowel is not its nucleus, but instead the preceding consonant and whatever is left of the vowel (Parker and Diehl:1981). In the latter option, the D_{min} is kept. A model such as Klatt's (1976), can only account to the cases of complete reductions of the formant structures if the duration of a vowel is considered as being the duration of both the vowel and the preceding sound. If the duration of a vowel is considered to be what is the formant structure, then equation (1) used in Klatt's work needs to be simplified by cancelling the element D_{min}. In case this assumption is true, a simpler linear equation that does not include D_{min} has to be used: $D_o = K * D_i$ (eq 3).

REFERENCES

Câmara, J.M., Jr. (1977), *Para o estudo da fonêmica portuguesa*. Rio de Janeiro: Padrão.

Câmara, J.M., Jr. (1979), *História e estrutura da língua portuguesa*. Rio de Janeiro: Padrão.

Chafcouloff, M. et al (1976), Effets de la coarticulation sur les caracteristiques acoustiques des contoïdes fricatives du français. In *Travaux de l'Institut de phonétique d'Aix*, 3, 61-113.

Fant, G. (1970), *Acoustic theory of speech production*, 2nd. ed. The Hague: Mouton.

Klatt, D. H. (1974), The duration of [s] in English words. In *Journal of speech hearing research*, 17, 1, 51-63.

Klatt, D. H. (1975b), Perception of segment duration in sentence contexts. In *Structure and process in speech perception*, A. Cohen and S. Noteboom, eds. Heidelberg: Springer Verlag.

Klatt, D. H. (1976), Segmental duration in English. In *Journal of the acoustical society of America*, 59, 1208-21.

Klatt, D. H. (1975a), Vowel lengthening is syntactically determined in a connected discourse. In *Journal of phonetics*, 3, 129-40.

Lindblom, B., B. Lyberg, and K. Holmgren (1981), *Durational patterns of Swedish phonology: do they reflect short-term memory processes?* Indiana, Bloomington: Indiana University Linguistic Club.

Lindblom B. and K. Rapp (1973), Some temporal regularities of spoken Swedish. Publication no. 21. Stockholm: Institute of Linguistics of the University of Stockholm (unpub.).

Parker, E. M. and R. L. Diehl (1984), Identifying vowels in CVC syllables: effects of inserting silence and noise. In *Perception & psychophysics*, 36, (4), 369-80.

Simões, A. R. M. (1980), *Étude acoustique des consonnes [s] et [z] du Brésilien*. Mémoire de D.E.A., ms. Aix-en-Provence, France: Institut de phonétique.

Simões, A. R. M. (1990), Predicting sound segment duration in connected speech: an acoustical study of Brazilian Portuguese. In the First international conference on speech synthesis, ETRW-GALF90, September 23-28. Autrans, France: Institut de la communication parlée.

Simões, A. R. M. (1987), *Temporal organization of Brazilian Portuguese vowels in continuous speech: an acoustical study*, Ph. D. diss. (unpub.). Austin, Texas: The University of Texas.