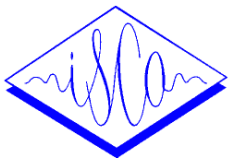


The ESCA Workshop
on Speech Synthesis
Autrans, France
September 25-28, 1990

PREDICTING SOUND SEGMENT DURATION IN CONNECTED SPEECH: AN ACOUSTICAL STUDY OF BRAZILIAN PORTUGUESE

Antônio R. M. Simões

Department of Spanish and Portuguese
The University of Kansas at Lawrence
Lawrence KS 66045 USA

ABSTRACT

The relationship between lower level linguistic components (phonetics/phonology) and higher level linguistic components (syntax, discourse) is studied from the experimental analysis of the temporal organization of the three extreme vowels [i], [a], and [u]. D. Klatt's (1976) model, based on American English, is examined in order to know how it could be adapted to Brazilian Portuguese. Following this comparison, rules are proposed to predict sound-segment duration of Brazilian Portuguese in continuous speech. Data for this investigation were obtained from one speaker from Rio de Janeiro who read a text for children containing over one thousand words.

1. INTRODUCTION

Although duration is not linguistically significant in Brazilian Portuguese (hereinafter BP), there are reasons for concentrating so much effort in analyzing duration as scholars have already noted (Klatt, 1974). The present paper attempts to bring a contribution to speech synthesis models by investigating the temporal patterns of sound segments in connected speech in BP.

Accounts on the general trends as well as on the general principles of BP have already been given elsewhere (Head, 1964; Câmara, 1968). The goals of the present work are: (1) to uncover the factors or language components that affect sound-segment durations by both shortening and lengthening vowel duration, (2) to determine if these factors interact in series or simultaneously, and (3) to determine vowel-duration change rules for the specific case of this BP speaker fitting Klatt's (1976) model to predict duration for the three vowels studied.

Segment duration is reflected as two linguistic phenomena: reduction and expansion in terms of an unmarked or inherent duration. The term "inherent" is borrowed from Klatt (1975, 1976), but the present study uses the term unmarked, instead. Thus, the unmarked duration of a given sound-segment μ is defined here as the median duration of all occurrences of that sound-segment μ in a text of over one-thousand words.

2. THE EXPERIMENTAL PROTOCOL

The experimental protocol is organized according to three main procedures: the production of the corpus (recordings), the production of spectrograms for sound-segment segmentation, and data analysis. In the production of the corpus, a single speaker of BP, PM, 32 years old, from Rio de Janeiro, was recorded. PM read a 1286-word text for children in two recording sessions. The recording sessions took place one week apart. PM was asked to read the same text three times in each session, totalling six readins of the same story. The third reading of the second session, i.e. the sixth recording, is the one used in this analysis. The recordings were made at the language laboratory at the University of Texas, at Austin. An acoustically isolated recording booth was used and the recording

was done using a AKG dynamic, unidirectional cardiod microphone situated at 40 cm from the subject's mouth. Before each recording the system was calibrated. The tape used is an AMPEX tape, 1/4", 1.5 mil, mylar.

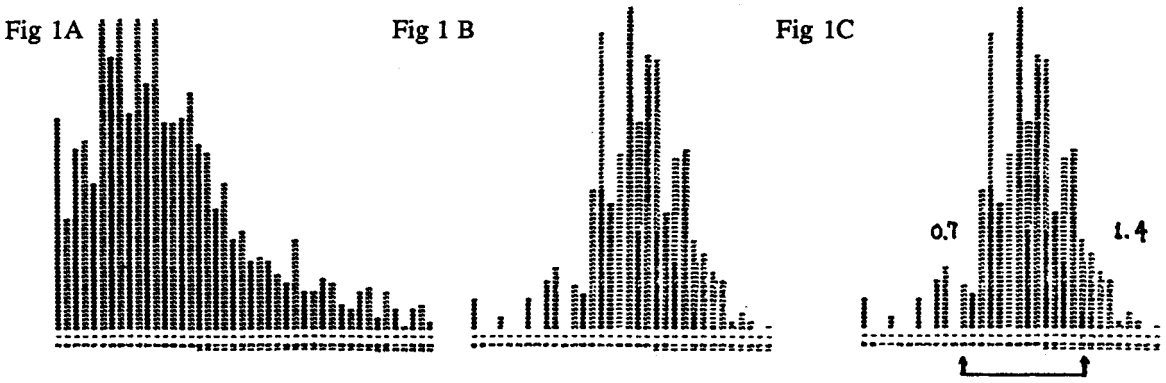
Some 700 spectrograms were produced to observe and measure the 1286-word text and its sound-segments in this study. The spectrograms were obtained through a stereo tape deck TEAC A-2300SX coupled to a Digital Kay Sonagraph 7800. The technique for spectrograms analysis in connected speech especially developed for this study combines two spectrograms for each sample analyzed. One spectrogram has the regular broad band and a second one is made using the amplitude contour. Superimposed on the top of each spectrogram there is an oscillographic image. A full description of this spectrographic techniques as well as techniques for sound segmentation on spectrograms can be seen in Simões (1987).

The statistical analysis was made using the S Statistical Package, implemented in the UNIX operating system (Unix is a trademark of Bell Labs), and run on a Digital Equipment Corporation VAX 11/780 at the UT Austin campus. The analysis of variance used here is a local program (ANOVA8) which is based on the PMDP series. It was developed by Thain Marston while a graduate student in Experimental Psychology at UT Austin.

3. DATA ANALYSIS

Analysis of all data from these vowels in the corpus consistently shows 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. Data information in figure 1A is only valid for segment expansion, i.e. lengthening. In order to consider both expansion and reduction of segments in terms of an unmarked duration (cf. definition above), common logarithms of positively skewed distributions were taken so that a normal distributions such as the one seen in 1B would be obtained. In Figure 1C, Klatts' (1975) is adapted to determine which segments were greater than 1.4 times the median duration of the segment studied. Shortening processes appear on the left side of the abcissa and the cuttof 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 or intrinsic) 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 $D_O = K \quad (D_i - D_{min}) + D_{min}$ (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,m 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.

4. 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 (1), (3), (5) and (6) are relevant to the speech of PM at the phonetic level. Conclusions (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 assess 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 equation (1) operates in series from domain 1 to domain 4 the present study does not have an argument against models, in which all involved factors operate through one functional form, simultaneously. I have not yet tested this possibility myself, but Klatt seems to have had problems with its implementation as it is noted in the preceding paragraph. The value of K is obtained by means of equation (2) and the unmarked

duration of each vowel initializes each process in any position within a word. 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.

When comparing both sets of rules in Klatt (1976) and in the present work, one has to bear in mind that all rules in Klatt's model will normally *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.

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.
Head, B. F. (1964), *A comparison of the segmental phonology of Lisbon and Rio de Janeiro*, Ph. D. diss. (unpub.). Austin, Texas: The University of Texas.
Klatt, D. H. (1974), The duration of [s] in English words. In *Journal of speech hearing research*, 17, 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.).
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. (1987), *Temporal organization of Brazilian Portuguese vowels in continuous speech: an acoustical study*, Ph. D. diss. (unpub.). Austin, Texas: The Univ. of Texas.