DURATION ANALYSIS AND MODELLING FOR TURKISH TEXT-TO-SPEECH SYNTHESIS

by

Ömer Şayli

B.S. in E. E., Boğaziçi University, 1999

Bogazici University Library

39001101685249

14

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science
in
Electrical and Electronics Engineering

Boğaziçi University

2002

ACKNOWLEDGEMENTS

I am thankful to my thesis supervisor Assoc. Prof. Levent M. Arslan for introducing me to this challenging and beneficial subject and to Prof. A. Sumru Özsoy for her motivation and criticisms during the preparation of this thesis. I would also like to thank Professor Bülent Sankur for serving on my thesis committee and for his revisions on the thesis.

I am grateful to my family and my friends for their encouragement and support.

ABSTRACT

DURATION ANALYSIS AND MODELLING FOR TURKISH TEXT-TO-SPEECH SYNTHESIS

Naturalness in TTS systems plays a big role in the acceptability of the TTS synthesis outputs. Rhythm, intonation, stress pattern, pitch and duration (timing) are the most important parameters which effect naturalness of the TTS system output. The task of the timing component in a TTS system is to compute duration information for sub-elements which are to be used in synthesis output. Duration modelling is a very challenging part of a TTS system since very little is known about the underlying process responsible for speech timing of humans.

To analyze and model duration for Turkish TTS systems, spoken utterances of 1-words and sentences of an adult male are used which are recorded at high digital quality. Firstly, coverage of the Turkish by this spoken text corpus is investigated, which is found to be well enough. Afterwards, analysis of the durations of Turkish phonemes is done. Effects of factors that can be computed from text on the durations are found to determine which of them should be included in the duration models.

To model duration, four models have been implemented. First two models use mean durations of the phonemes and mean durations of the triphones. Third model uses mean durations of the nodes of trees for triphones for duration prediction. The last model is an additive model where the effects of factors are found by regression analysis.

ÖZET

TÜRKÇE YAZIDAN SESE ÇEVİRİ SİSTEMLERİ İÇİN SÜRE ANALİZİ ve SÜRE MODELLEME

Doğallık, Yazıdan-Sese-Çeviri (YSÇ) sistemlerinin kabul edilirliğini belirlemede önemli bir göreve sahiptir. YSÇ sistemince üretilen ses çıktısının doğallığını etkileyen en önemli parametreler ritim, entonasyon, vurgu örüntüsü, temel sıklık ve süre bilgisidir. YSÇ sistemlerindeki süre biriminin görevi sentezde kullanılan parçacıkların süre uzunluk bilgisini hesaplamaktır. YSÇ sistemleri için süre modelleme, insanlardaki ses üretim mekanizması tam olarak anlaşılamadığı için oldukça zordur.

Türkçe YSÇ sistemlerinde kullanılmak üzere süre analizi ve modelleme yapmak için yetişkin bir erkek tarafından söylenen ve yüksek kalitede kaydedilen tek-kelimeler ve cümleler kullanıldı. Öncelikle söylenip kaydedilen metnin Türkçe'nin ne kadarlık bir kısmını kapsadığı araştırıldı ve yeterince iyi olduğu bulundu. Bundan sonra Türkçe'deki seslerin süre analizi yapıldı. Yazıdan bulunabilen etmenlerin süre üzerindeki etkileri, süre modellemede kullanılmak üzere araştırıldı.

Süre modellemesi için dört model geliştirildi ve uygulandı. Denenen ilk iki model Türkçe'deki seslerin ve üçlü öbeklerin ortalama sürelerini kullanmaktadır. Üçüncü model, üçlü öbekler için ağaç yapısındaki düğüm noktalarının ortalama sürelerini süre tahmininde kullanmaktadır. Son model, süreyi etkilediği bulunan etmenler için toplam modelini kullanmakta ve etki değerleri doğrusal bağlanım ile bulunmaktadır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

c	For phonetic symbol ǰ
C	For phonetic symbol č
G	For phonetic symbol :
H_0	Null hypothesis
H_A	Alternative hypothesis
j	For phonetic symbol ž
n	Sample size
S	For phonetic symbol š
$\mathbb{R}$	Real numbers
S^2	Sample variance
$\overline{y}$	Sample mean
α	Significance level
μ	Mean
σ^2	Variance
ANOVA	Analysis of Variance
$d.f$	degree of freedom
Hz	Hertz
i.i.d	independent and identically distributed
ms	milli seconds
p.d.f	probability distribution function
Q-Q	Quantile-quantile
T	Consonant
TTS	Text-to-Speech
V	Vowel

1. INTRODUCTION

The importance of man-machine interface has increased over the years and the area is promising a great potential for widespread use in the coming years. Sound (speech) is one of the crucial elements of man-machine interfaces. In sound technology there are three main areas [1]:

- Voice Response Systems (Text-to-Speech (TTS) Systems)
- Speaker Recognition Systems
- Speech Recognition Systems

In voice response systems, objective is to produce speech utterances that correspond to any text. Differentiating a particular speaker's voice from others is the goal of speaker recognition systems. Ability to understand the message contained in the utterances of humans is aimed by speech recognition systems. Like in other areas, advances in digital signal processing and computer technology have led to a great advancements in these.

Our interest in this study is on the TTS systems. Main application of TTS systems in man-machine interfaces is to respond to a request for information by spoken messages. Most important parameter in the quality of TTS systems' outputs is intelligibility. Other aesthetic factors such as quality and naturalness of the utterance produced by a TTS system have effect on the usefulness and acceptability by we, humans. There are two main approaches to the implementation of a TTS system. First one is speech-synthesis-by-rule systems where goal is to model human speech production mechanism as closely as possible. This approach is really challenging because of the difficulty in discovery of parameters for controlling the synthesizer, i.e. pitch, intensity, and vocal track response parameters [1]. A second approach is to concatenate isolated speech elements which are *sub-segments* of high quality recorded spoken words, phrases or sentences. These speech sub-segments can be *monophones*, *diphones*, *triphones* or even the complete word sets. The latter approach is used in

current Text-to-Speech (TTS) systems because it provides a good balance between intelligibility, quality and naturalness while being technologically easier to implement. Perfect human like sounding mechanism could be realizable using the former approach.

1.1. Problem Statement

The type of TTS systems we consider is concatenation based ones. In such a system the vocabulary storage is *'sub-elements'* extracted from uttered words, phrases or sentences. These sub-elements can be monophones, diphones or even the complete word sets. These systems are parametric because they do not store the waveforms of speech sub-elements but their LPC (linear predictive coding) coefficients. In this way, low storage requirement is achieved. In synthesis, such a TTS system first finds appropriate elements from the sub-element storage database for the given input text. It does a Viterbi search in selecting the units using how much they resemble the input text as the criterion. For example, let the input text be /kalem/. If the sub-segments are monophones, the system will first search the database for /k/. If it exists, it searches the database further to see if there exists any /k/ which has space on the left and /a/ on the right. If it can not find /Space,k,a/ it will search the database to find the most resembling one. The system will continue doing this search for the other units in the text input. Using the selected sub-segments from the database, waveforms are produced from LPC coefficients and finally they are concatenated to produce the synthesizer output.

It can be thought that this concatenation of sub-units to form the desired utterance in a TTS system could be successful enough but it fails when a spoken sentence is very different from a sequence of small sub-units uttered in isolation or in other 'context' than the desired sentence. For example, in a sentence, words could be as short as half their duration when spoken in isolation [2]. Other problems which lead to unnaturalness in the synthesized output are stress pattern, rhythm, pitch, intonation, prosody and duration.

The problem we will work on will be the duration, or the so called timing. By

duration, we mean the duration of speech sub-elements. The task of the timing component in a TTS system is to compute duration information for sub-elements to be used in concatenation from symbolic input such as phoneme symbols, stress and accent markings. There is very little known about the underlying process responsible for speech timing. It has been found that the duration component is unusual characteristic in that it is neither purely rule based nor purely statistically based. Instead, it borrows from both of these opposing approaches [3].

Interestingly, it has been analyzed that speech timing can be predicted from text only up to a point. For example, in a study done for some other languages, an analysis of durations of the same vowel spoken in identical context by the same speaker corrected for per-utterance speaking rate showed at least 8 per cent of the total variance is text-independent [3, 4]. Apparently, speech accelerates and decelerates in the course of even a brief utterance in a way that can not be predicted from text. This leads to inherent variability of speech.

1.2. Objectives

Our first objective in this study is to analyze the durational properties of Turkish phonemes and to find out the effect of factors (that could be computed from text) on these durations. The next goal is to derive and implement models that could compute duration information of sub-elements to be used in concatenation for a given text to a TTS system. There are several reasons why this is a hard task: one is that there are many factors that affect timing and joint effect of these are quite complex. Also the number of factorial constellations (cases) that can occur in a language like Turkish is vast [3].

1.3. Organization of the Thesis

For the used spoken corpus in this thesis, statistical coverage properties are analyzed in the second chapter. Results of the study on the durational properties of Turkish phonemes and the effects of factors (that could be computed from text) on

Table 1.1. Used symbols for the Turkish phonemes

Phoneme symbol in Turkish Alphabet	Phoneme symbol in this thesis
a	a
b	b
c	c
ç	C
d	d
e	e
f	f
g	g
ĝ	G
h	h
ı	I
i	i
j	j
k	k
l	l
m	m
n	n
o	o
ö	O
p	p
r	r
s	s
ş	S
t	t
u	u
ü	U
v	v
y	y
z	z

those are presented in Chapter 3. In this chapter and afterwards, used symbols for the phonemes of Turkish in this study and their equivalent in Turkish Alphabet are given in Table 1.1 (also, V is used to denote vowels and T is used to denote consonants throughout the thesis). Review of some models used in the literature along with the derived and implemented models for duration modelling are discussed in Chapter 4. Also performance comparisons of the implemented models are given in this chapter. Finally, results of this study are discussed.

2. DATA COVERAGE

In an ideal world, a speech synthesizer should be able to synthesize any arbitrary word sequence with complete intelligibility and naturalness. However, there is a trade-off between flexibility of vocabulary and sentences at the expense of naturalness, as shown in Figure 2.1. For example, arbitrary words and sentences can be synthesized, which do not sound very natural. Conversely a system can produce vary natural sounding utterances for a very constrained set of word sequence. This applies to articulatory, rule-based, and concatenative methods of speech synthesis [5]. To produce

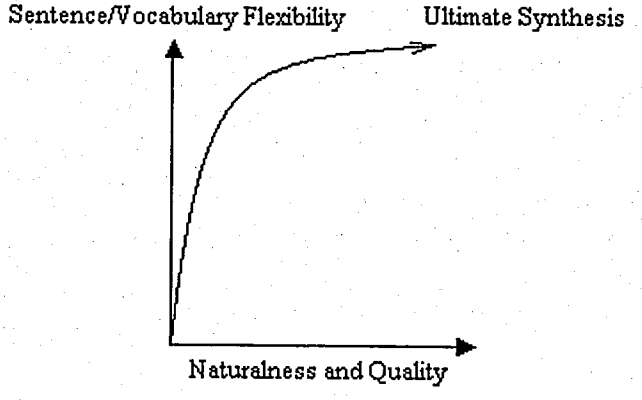

Figure 2.1. Synthesis development trade-off schematic

highly natural sounding utterance, it is desired to have a very large database of words, phrases or even complete sentences, from which the sub-units will be extracted to be used in concatenation. However it is indeed difficult to have a database which covers all possible words, phrases and it is literally impossible to have a spoken database of sentences that can be uttered. Since we have limited time, energy and resources, we should decide what to cover and how much to cover to get a good compromise between *quality* and *sentence and/or vocabulary flexibility* for the synthesizer.

2.1. Reasoning for Selection of Triphones for Coverage

In concatenative text-to-speech synthesis there is an important assumption that speech is produced as a sequence of distinct sounds. However, this assumption does not

hold most of the time because of our inability to move our vocal tract system abruptly to produce distinct sounds in consequence. Rather, there is generally a transition region from one phoneme to the subsequent one. Therefore, context dependent modeling seems to be a more natural way of representing fluent speech. Thus triphones model these transitions thereby achieving well acoustic modeling of phones. Of course use of higher order models would be better but then a much larger corpus would be needed. To illustrate it, consider our language. In Turkish, there are 29 distinct phonemes in the alphabet, including the background silence 30. A corpus of size $30^1 = 30$ would be enough to cover all the monophones. The number of left or right diphones is $30^2 = 900$. The number increases to $30^3 = 27000$ when we consider all the possible triphones. This number goes up to $30^4 = 810000$ for covering all the possible four-phones. In fact, number of the most used triphones are much smaller than this intimidating numbers. But these numbers show the trend how the database corpus should increase when multiple-phonemes are chosen as the units instead of monophones. So a good compromise would be to use a corpus which covers a fairly high percentage of the most frequent triphones [6].

A more detailed study should have a spoken text corpus that covers the Turkish language not just based on the text coverage, but also for every possible combination of the factors such as [7]

I. Identity of the current segment
II. Identity of the stress
III. Identity of the previous segment
IV. Identity of the following segment
V. Identity of intonation
VI. Speaking rate(speed)
VII. Number of the preceding syllables in the word
VIII. Number of the following syllables in the word
IX. Number of the preceding syllables in the phrase
X. Number of the following syllables in the phrase
XI. Number of the preceding syllables in the utterance

XII. Number of the following syllables in the utterance
XIII. Syllable type (i.e. TV, VT, TVT, V, TVTT)

Factor I is simply the number of phonemes in the alphabet (8 vowels and 21 consonants). Factors III and IV can also be the number of phonemes, but one should find groups (i.e. voiced fricatives, unvoiced stops) to reduce the unnecessary complexity. Factor II is the type of stress of the segment (stressed vs. unstressed). Factor V is the sample of intonation patterns. Factor VI may have the values slow, normal and fast. Factors VII to X may have the values segment is at the boundary, segment is 1-syllable away and 2 or more syllable away from the boundary. Factors XI to XII may have the values segment is at the boundary and segment is 1 or more away from the boundary. Factor XIII is the syllable type the segment is in. There are 10 possible syllable types; TTV, TTVT, TTVTT, TV, TVTT, TVT, V, VT, VTT and T. Some of these does not occur in Turkish words but in the foreign words passed to Turkish from other languages.

Moreover, ideally the corpus should be balanced, i.e frequency of every combination is more or less the same. Bias can result if a segment occurs more frequently in some environments than the others. It is unfortunate that we don't have a spoken and labelled (by linguists) corpus, such as TIMIT for English, which covers our language in a good percentage in terms of the factors stated above.

2.1.1. Statistical Properties of Triphones

It is a promising fact that the number of all possible triphones, 27000, is much higher than number of triphones encountered in the language due to grammar and syntax constraints. In a study done by Yapanel [6], the number of the most commonly used triphones is investigated. Yapanel [6] conducted a statistical study on a text containing 2.2 million words. As a result of this study, it was found that on the order of 11.000 distinct triphones are found, which is nearly half of the 27.000. From Table 2.1, we can see that just the most frequent 1000 triphones covers 80 per cent of the Turkish text corpus used. This is a significant decrement compared to 27000 possible

Table 2.1. The most frequent triphones versus coverage of the corpus

Number of the most frequent triphones	Covered per cent of the Turkish (text corpus)
100	31.80
500	64.44
1000	80.08
1500	87.80
2000	92.15
2500	94.91
3000	96.66
4000	98.51
5000	99.53

triphones. From the Figure 2.2 we see that, 92 per cent of the corpus is covered using the most frequent 2000 triphones, and we have 98 per cent coverage using the most frequent 4000 triphones. Practically, the most commonly occurring 5000 triphones cover the Turkish language completely. For practical purposes, we can deduce that modelling the statistics of 2000 most frequent triphones would be sufficient.

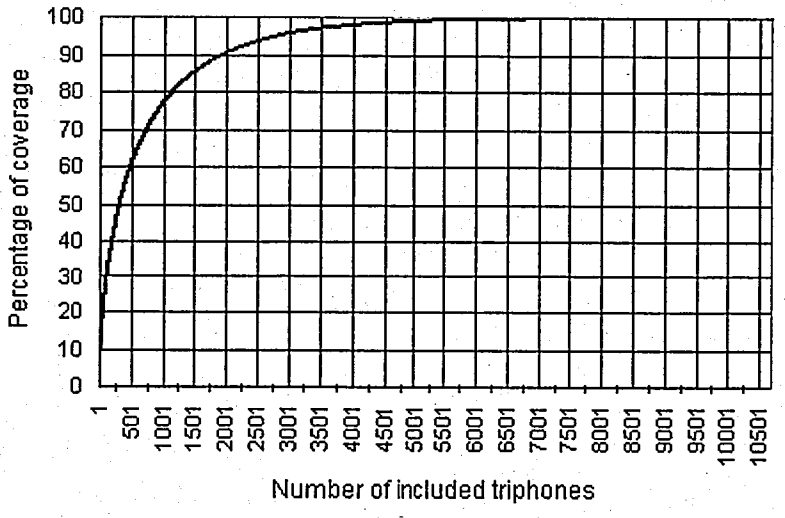

Figure 2.2. Coverage of the text corpus versus number of the most frequent triphones

2.2. Coverage Properties of the Database

We have a database of one adult male speaker. There are mainly two types of utterances. Spoken words and sentences. Utterances are recorded at 16 kHz with 16-bit representation. After recording, the labelling of the data is done, which is exhausting and time consuming task. In labelling, one decides and labels boundaries of phonemes of spoken utterances with the aid of spectograms and waveforms. In labelling using spectogram, boundaries are decided to be placed when the dominant formants begin to change. Using the waveform of an utterance, labels are put when the amplitude envelope characteristics begin to differ. In fact, there are no *true boundaries* between phonemes. There are always transitions between segments. What makes labelling difficult is that these transition characteristics differ in every utterance. Hence it is very important to adopt a labelling convention, although a convention which covers all possible cases is difficult to derive. For some languages, there are big databases, for which labelling is done by linguistics, such as TIMIT database for the English language. Unfortunately for our language currently such a database does not exist, a very big drawback for speech studies. To infer statistical behaviors from a spoken corpus, it has to be shown that this corpus covers enough amount of the most frequent triphones. The database we use is sufficient for our purposes as shown in the following sections.

2.2.1. Coverage in the Word Corpus

Word database contain approximately 7898 spoken words. Analysis of this data base from the viewpoint of triphone coverage is illustrated in Figure 2.3. In the Figure, *Coverage of the Triphones* refers to the coverage of the Turkish -text corpus used in the study conducted by Yapanel [6]- by the corresponding number of the most used triphones. *Uniform Coverage Curve* is calculated by the following formula;

$$\text{Uniform Coverage Guess} = \text{Covered Percent of Triphones} \times \text{Covered Percent of Turkish by corresponding Triphones}$$

Intuition behind Uniform Coverage Guess is simple. How much percent of the Turkish

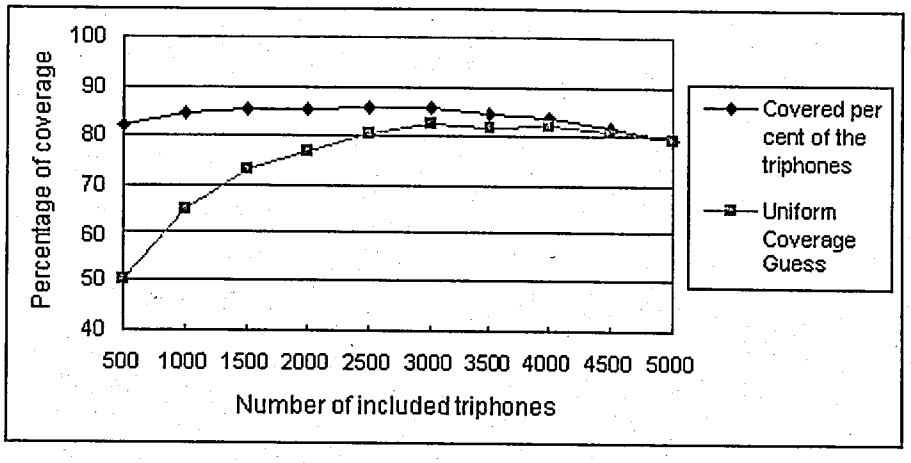

Figure 2.3. Coverage of Turkish by the spoken word corpus

is covered by the used corpus can be found by multiplication of the percentage coverage of the most used triphones and the coverage percentage of the Turkish by this to be covered triphones. Here the assumption is that occurrence frequencies of the covered triphones are equal. This assumption fails to hold when the percentage of the coverage decreases. We expect the Uniform Coverage Guess plot to increase first and then decrease. The low coverage prediction for the low number of triphones is because of the low coverage of the Turkish by small number of triphones. Covered percent of the triphones decreases monotonically which is the factor of the decrease for the Uniform Coverage Curve in the high number triphone coverage region. In the mid-triphone coverage region, Uniform Coverage Curve makes a peak which can be used as a somewhat optimistic prediction of coverage of the Turkish by the spoken database. It can be easily seen from the Figure 2.3 that triphone coverage of the spoken word database is fairly good. Percentage coverage of the triphones does not fall below 79 per cent up to the most frequent 5000 triphones. It covers 85.8 per cent of the most frequent 3000 triphones and 81 per cent of the most frequent 4500 triphones. Here we can be comfortable to rely on the Uniform Coverage Guess because of the high coverage of the triphones. As expected, Uniform Coverage Guess plot first increases then starts to decrease. Uniform Coverage Guess takes peak 82.4 per cent at the most frequent 3000 triphones, which is very promising. So we can be very confident that this corpus covers the Turkish Language with a very high probability of 82.4 per cent.

However redundancy in this word corpus can be deduced from Figure 2.4. We

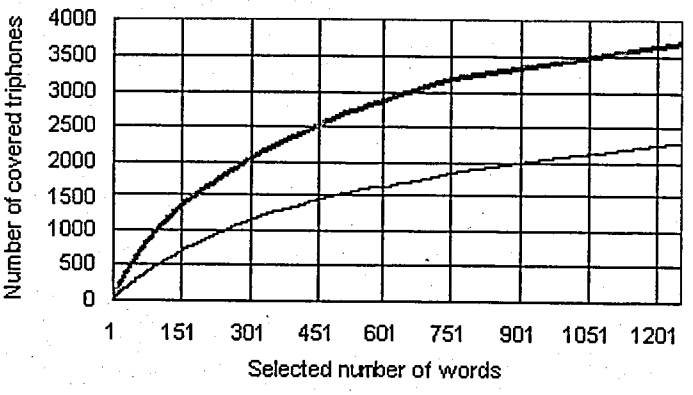

Figure 2.4. Number of covered triphones (of the 4500 most frequent triphones) with greedy sentence selection versus random selection in the 1-word corpus, solid line greedy selection, thin line random selection

have implemented a *Greedy Selection* algorithm which selects the minimal number of words from the corpus which cover maximum number of the triphones. From the figure, we see that 1200 words selected using this Greedy Algorithm covers 3600 of the 4500 the most used triphones, which is 15 per cent of the whole words. Although this means that more triphones could be covered using 7898 words, those would be mostly rarely used ones. Also crudely this means most of the covered triphones occur more than 6 in the spoken word corpus.

2.2.2. Coverage in the Sentence Corpus

Sentence database contain approximately 205 spoken sentences. We have also analyzed this corpus from the viewpoint of its triphone coverage. Results are indicated in Figure 2.5. As can be seen from the figure, the percentage coverage of the Turkish by spoken word database is much lower than the spoken word corpus. This is natural if we look at the ratio of the number of words in the corpuses, as calculated below

$$\frac{\text{Number of words in the spoken word corpus}}{\text{Number of words in the spoken sentence corpus}} = \frac{7898}{1167} \cong 6.77$$

This explains the reason why the coverage of recorded sentence corpus is worse. However collecting sentence corpus is much more difficult because of labelling. We have

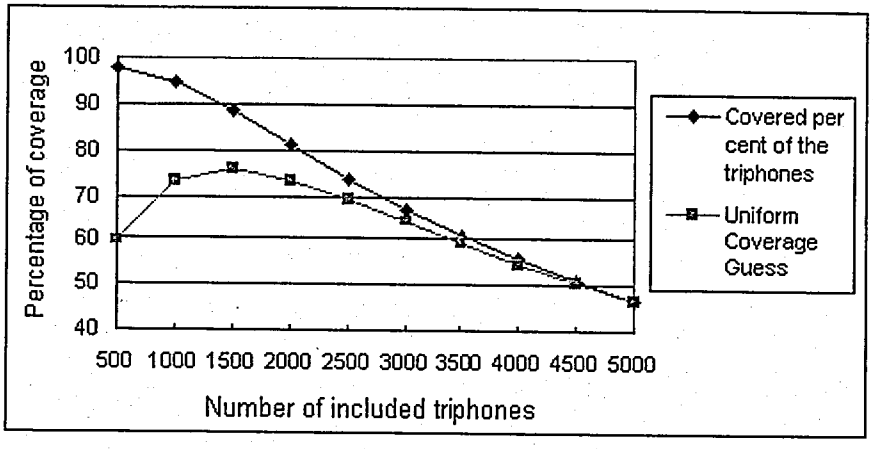

Figure 2.5. Coverage of Turkish by the spoken sentence corpus

mentioned the difficulty of labelling at the start of Section 2.2.2. In spoken sentences, this difficulty increases as the speaking rate (speed) increases, since the durations of segments shrink compared to utterance of one word only and become indistinguishable from transitions. Also transitions are much wider than in the 1-word environment. For example, for a spoken text in which /a/ occurs, finding a region in the spectrogram where only formant frequencies of /a/ exist is difficult, but a region where formants of /a/ as well as neighboring phonemes' formants exist can be found. Hence it becomes important how to divide these transitions. This dividing is done to minimize the problems that can occur when concatenating the segments in the synthesis. Coverage

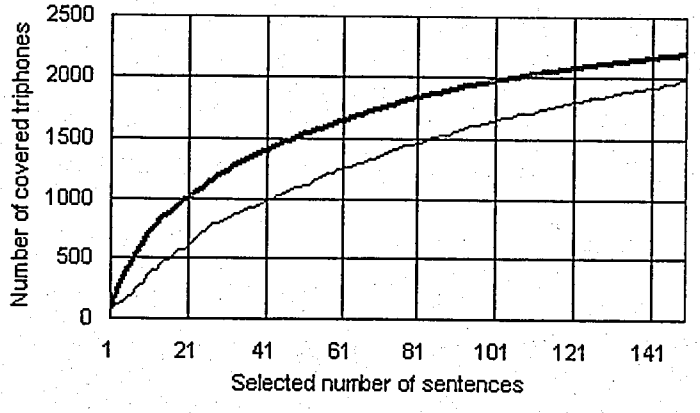

Figure 2.6. Number of covered triphones (of the 4500 most frequent triphones) with greedy sentence selection versus random selection in the sentence corpus, solid line greedy selection, thin line random selection

analysis results are indicated in Figure 2.5. It is seen that the coverage of the triphones

declines sharply. After the most frequent 2000 triphones, coverage of the triphones by the sentence corpus decreases below 80 per cent. Uniform Coverage Guess Curve gets its peak at the most frequent 1500 triphones with 76.3 per cent coverage. So we can conclude that sentence corpus covers approximately 76 per cent percent of the Turkish and covers 80 per cent of the most frequent 2000 triphones.

Although there are much smaller number of words than the word corpus, redundancy exist also in the sentence corpus. 105 words selected using the Greedy Algorithm covers 2001 of the 4500 the most used triphones whereas 150 randomly selected words cover 1984 of them.

2.2.3. Conclusion for Coverage of Turkish with the Spoken Corpus

The word corpus covers 82 per cent of the Turkish while the sentence corpus covers 76 per cent of the Turkish. We refer to the text corpus used in the study of Ü. Yapanel [6] as Turkish, which contains 2 million words -how many of them are distinct are not indicated in that study-. We are confident that coverage of the word corpus is high enough for duration analysis. Coverage of the sentence corpus is also reasonable. Covered triphones in the word corpus occur mostly more than once, which is better for deducing statistical models. Having a small sentence corpus and large word corpus is not bad because after finding duration models for both word domain and sentence domain one can use the segments from the word corpus using the duration model for sentence domain to use in 'sentence utterance synthesis'. This is the ultimate goal and real world scenario since gathering data for spoken words are easier. If we can find a good duration model for sentence domain, with a very small set of sentence corpus and large word corpus, a satisfactory synthesizer can be developed with natural timing.

3. DURATION ANALYSIS

To derive useful models for the duration, it is mandatory to analyze duration properties of the Turkish Phonemes and to find factors which effect these durations. In the literature, there are a number of good studies on the phonetics of Turkish [8, 9, 10]. However we were able to find only one study investigating durational properties of Turkish phonemes by Prof. Dr. Nevin Selen [10] which is a detailed study on the phonetics and acoustics of the Turkish.

In this chapter, we present our findings about the durational properties of the Turkish Phonemes and the general factors which effect these durations in 1-word and sentence environments.

3.1. Analysis Tools

As indicated in Sections 2.2.1 and 2.2.2, our database consist of two environments. Spoken sentences (205 units) and 1-words (7898 units). To investigate the durations of the phonemes, one had to convert these data into a form which can be analyzed easier and processed properly. Because of its high execution speed, C++ programming language was chosen to 'read' the database and to convert the information into matrix form which can be analyzed with MATLAB. MATLAB is chosen for mathematical analysis because of the statistical tools available in this environment and our familiarity with it (written programs and used data are in the accompanying CD with this thesis).

3.1.1. Database Construction

For the spoken utterance, label information were written to the files which contain timing information for letters. Database construction from this data is done as shown in the Figure 3.1. Algorithm is as follows: For each label file, corresponding 'sentence' is extracted. Timing information for the letters of this sentence is known. For each letter in the found sentence, the following feature factors are found and coded:

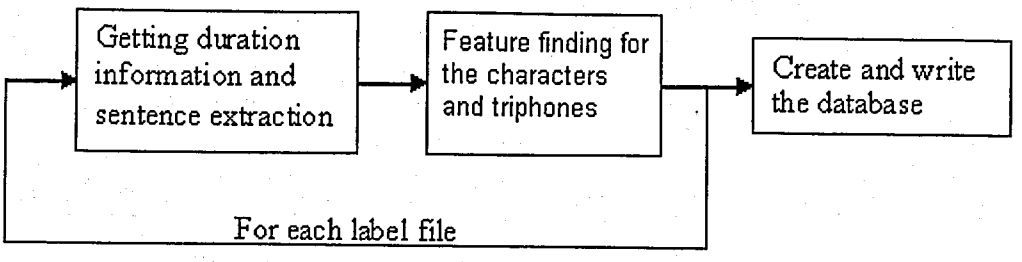

Figure 3.1. Database construction flowchart

I. Identity of the current segment (29 values)

II. Preceding identity type (3 levels: consonant, vowel, punctuation)

III. Following identity type (3 levels: consonant, vowel, punctuation)

IV. Identity of the preceding segment

- If vowel, preceding vowel identity (8 levels)
- If consonant, preceding consonant identity (21 levels)

V. Identity of the following segment

- If vowel, following vowel identity (8 levels)
- If consonant, following consonant identity (21 levels)

VI. Number of syllables in the word (7 levels)

VII. Number of words in the sentence (7 levels)

VIII. Word position (3 levels: initial, middle, final)

IX. Sentence position (3 levels: initial, middle, final)

X. Syllable pattern (10 levels: V, VT, TV, T, TVT, VTT, TTV, TTVT, TVTT, TTVTT)

Some of the factors stated in Section 2.1, stress, speaking rate (speed) and intonation, are not coded as we don't have tools currently that can automatically and reliably generate these information from the waveform of spoken utterance. After coding is complete for all the label files, this information is written to text files in matrix form for all the letters of the Turkish Alphabet. For the triphones, factors VI, VII, VIII and IX are coded and written to a database. Factors II, III, IV, V and X are not coded for the triphones because for the *center* of letter of this triphone, these information is known (levels TTVT, TVTT and TTVTT are missed for factor X since length of triphone is 3 letters). Database written in matrix form is then used easily in MATLAB

for analysis and modelling purposes.

3.1.2. Statistical Tools

Mathematical tools used in investigating properties of the Turkish phonemes are statistical methods. Namely, they are box plot, distribution fitting, Q-Q plots, confidence intervals, ANOVA analysis and multiple comparison procedure.

3.1.2.1. Box Plot. A quick comprehension of the general durational properties of the phonemes can be gained by looking at the box plots. Box plot is a graphical way of looking at the distribution of the data in different groups. Box plot produces a box and whisker plot for each group. The box extends from the lower quartile to the upper

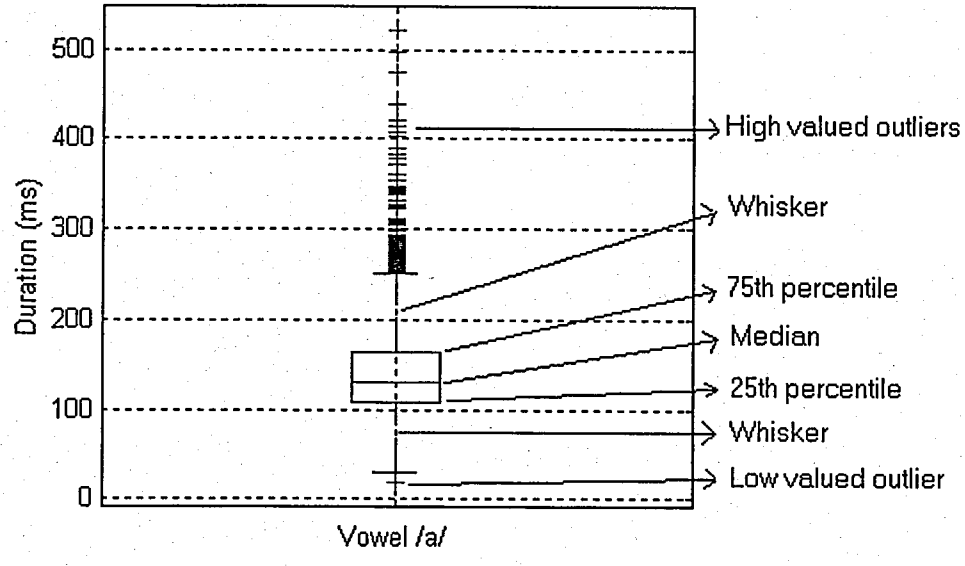

Figure 3.2. Box plot example, duration of vowel /a/ in 1-word environment

quartile and has lines at the lower quartile (the 25th percentile), median (the 50th percentile), and upper quartile (the 75th percentile) values. Probability for the data to fall in the range below 25th percentile is 25 per cent, to fall in between 75th percentile and 25th percentile is 50 per cent and to fall above 75th percentile is 25 per cent. Hence, probability of falling in the 'box' is the biggest with 50 per cent. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Each whisker extends to the most extreme data value within 1.5 interquartile range

of the box. Interquartile range is the difference between the 75th percentile and 25th percentile and it is a measure for the spread. Outliers are data with values beyond the ends of the whiskers. An example of box plot is given for duration of vowel /a/ in the 1-word environment in Figure 3.2.

3.1.2.2. Distribution Fitting. Several distribution functions (beta, chi-square, exponential, half-normal, laplace, logistic, student's t, weibull) are tried to find out which one best fits the histograms of the durations of the phonemes. With the aid of Q-Q plots (explained in Section 3.1.2.3), distribution of durations of the phonemes are found to be well approximated by the gamma distribution and log-normal distribution. Normal distribution is also good but data histograms deviate more from it. Normal, log-normal and gamma distributions are given in equations 3.1, 3.2 and 3.3 ($\Gamma(a)$, used in gamma distribution, is the gamma function defined by equation 3.4).

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \sigma > 0 \tag{3.1}$$

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(m-\ln x)^2}{2\sigma^2}}, \quad x,\sigma > 0 \tag{3.2}$$

$$f(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}, \quad x \geq 0 \tag{3.3}$$

$$\Gamma(a) = \int_0^\infty y^{a-1} e^{-y} dy, \quad a > 0 \tag{3.4}$$

3.1.2.3. Q-Q and Deviation Plots. To compare expected probability distribution to the actual data histogram, a quantile-quantile (Q-Q) plot is used usually. A Q-Q plot shows the relationship between the quantiles of the expected distribution and the actual data. An agreement between the two is illustrated by a straight line. Straight line shows

the quantiles of the expected distribution type. Quantiles of the data is shown on the same plot. If data has the same underlying distribution as the expected distribution, quantiles of the data will be quite close to the straight line. Otherwise, if the plotted points deviate significantly from a straight line, the hypothesized distribution model is not appropriate. In fact, the determination of whether or not the data plot as a straight line is subjective [11]. Deviation from the straight line in the Q-Q plot is more formally shown in the deviation plots, which is constructed as follows: All data points are ranked from smallest to largest and each is paired with an expected distribution value for a sample of that size from an expected probability distribution. The deviation in the detrended normal plot is the difference between the standardized value for a case and its expected distribution value.

3.1.2.4. Confidence Interval. Confidence interval provides an interval within which the value of the parameter (i.e. mean duration) is expected to lie with a certain probability (i.e. 95 per cent) [11]. There are a number of formulas used in calculating confidence intervals for various assumptions [11, 12, 13]. The one we used is determination of a confidence interval for the mean μ of a normal distribution (assuming that the durations of phonemes are distributed normally) with unknown variance σ^2, given in equation 3.5 [14]

$$\overline{y} - t_{\alpha/2}\sqrt{\frac{S^2}{n}} \leq \mu \leq \overline{y} + t_{\alpha/2}\sqrt{\frac{S^2}{n}} \tag{3.5}$$

for $100(1-\alpha)$ per cent confidence interval, where $t_{\alpha/2}$ is the value at which cumulative t distribution with $(n-1)$ d.f is equal to $(1-\frac{\alpha}{2})$. The parameters α and $(1-\alpha)$ are called significance level and confidence level, respectively. Statistically *true mean value* lies $100(1-\alpha)$ per cent of the time in the estimated interval. Equation 3.5 is valid practically for distributions other than normal for large sample size (i.e. larger than 30). This is true for the word database but may fail for some phonemes (i.e. /j/ which is uttered only two times) in the sentence database.

3.1.2.5. ANOVA. To analyze the differences along with similarities of the phonemes and the factors affecting on the phonemes' durations, One-Way Analysis of Variance (ANOVA) and multiple comparison techniques are used. The purpose of one-way ANOVA is to find out whether data from several groups have a common mean. That is, to determine whether the groups are actually different in the measured characteristic. One-way ANOVA is a simple special case of the linear model. The one-way ANOVA form of the model is [11],

$$d_{ij_i} = \mu_i + \epsilon_{ij_i} \tag{3.6}$$

where i represents factor levels (i.e. vowels, consonants or sentence position), j_i is the number of observations for each factor level, μ_i is the mean value for each factor level, ϵ_{ij_i} is a random error component (which is assumed normally distributed and i.i.d) and d_{ij_i} is the observation value. If we are interested in the equality of the treatment means for k factor levels, the appropriate hypotheses are

$$\begin{aligned} H_0 &: \ \mu_1 = \mu_2 = ... = \mu_k \\ H_A &: \ \mu_i \neq \mu_j \text{ for at least one pair } (i,j) \end{aligned} \tag{3.7}$$

The test statistic is an F test with k-1 and N-k degrees of freedom, where N is the total number of observations and k is the number of factor levels. Test statistic is ratio of the sum of squares of the differences *between* the treatment averages and the grand average divided by the degree of freedom $(k-1)$ to sum of squares of the differences of observations *within* treatments from the treatment average divided by the degree of freedom $(N-k)$, called MS_E. ANOVA returns also a P-value for the null hypothesis that the means of the groups are equal. Probability of taking a value greater than the one obtained from data for the test statistic is the P-value. A low P-value (high F value) for this test indicates evidence to reject the null hypothesis in favor of the alternative. In other words, there is evidence that at least one pair of means are not equal. For significance level of 0.05, any test resulting in a P-value under 0.05 would be significant, and therefore, one would reject the null hypothesis in favor of the alternative hypothesis.

3.1.2.6. Multiple Comparison Procedure. In a one-way analysis of variance, you compare the means of several groups to test the hypothesis that they are all the same, against the general alternative that they are not all the same. Sometimes this alternative may be too general. You may need information about which pairs of means are significantly different, and which are not. A test that can provide such information is called a "multiple comparison procedure". When there are many group means, there are also many pairs to compare. Ordinary t-tests are not appropriate in this situation, since the alpha value would apply to each comparison, so the chance of incorrectly finding a significant difference would increase with the number of comparisons. Multiple comparison procedures are designed to provide an upper bound on the probability that any comparison will be incorrectly found significant. We have used Tukey-Kramer's multiple comparison procedure since it does control the overall error rate. Overall significance level is exactly α when the sample sizes are equal and at most α when the sample sizes are unequal. It makes use of the distribution of the studentized range statistic [11]

$$q = \frac{\overline{y}_{max} - \overline{y}_{min}}{\sqrt{MS_E/n}} \tag{3.8}$$

where n is sample size, $\overline{y}_{max}$ and $\overline{y}_{min}$ are the largest and the smallest sample means, respectively, out of a group of p sample means. For unequal sample sizes, Tukey's test declares two means significantly different if the absolute value of their sample differences exceeds

$$T_\alpha = q_\alpha(a, f)\sqrt{MS_E\left(\frac{1}{2n_i} + \frac{1}{2n_j}\right)} \tag{3.9}$$

where n_i and n_j are sample sizes of groups i and j respectively, $q_\alpha(a, f)$ is the upper α percentage point of q, f is the number of degrees of freedom associated with the MS_E (estimate of variance within treatments). $100(1 - \alpha)$ per cent confidence interval for

all pairs of means is as follows (for groups i and j):

$$\begin{aligned} \overline{y}_i - \overline{y}_j - q_\alpha(a,f)\sqrt{MS_E\left(\tfrac{1}{2n_i}+\tfrac{1}{2n_j}\right)} &\leq \mu_i - \mu_j \\ &\leq \overline{y}_i - \overline{y}_j + q_\alpha(a,f)\sqrt{MS_E\left(\tfrac{1}{2n_i}+\tfrac{1}{2n_j}\right)}, \quad i \neq j \end{aligned} \tag{3.10}$$

where $\overline{y}_i$ and $\overline{y}_j$ are group means of groups i and j respectively.

3.2. A Look at the General Durations of the Phonemes

There are some alternative analysis to phoneme inventory in Turkish (i.e. where number of vowels and consonants differ). Vowels have primary importance, since the nucleus of a syllable is a vowel. Vowels and consonants are phonetically classified as shown in the Tables 3.1 and 3.2.

Table 3.1. Vowel classification table

	Unrounded		Round	
	Wide	Narrow	Wide	Narrow
Back	a	I	o	u
Front	e	i	O	U

Table 3.2. Consonant classification table

	Nasals	Fricatives	Affricate	Stops	Semi-vowels	Whisper
Voiced	m, n	z, v	c	b, d, g	r, y, l, j	
Unvoiced		f, s, ş	C	p, t, k		h

The analysis presented in this study is based on the labelling convention developed by a non-linguist. Some deviation from the results here is predicted with a more linguistic approach to labelling. The symbol /G/ which is analyzed as representing vowel length in linguistic studies has been left out of the scope of the duration analysis because the labelling convention adopted in this work needs further refinement for this symbol.

3.2.1. Vowels

For the vowels, a very significant property was observed from box plots (Figures 3.3 and 3.4) and confidence intervals of the vowel means (Figures 3.5 and 3.6). Mean durations of the wide-vowels (/a/, /e/, /o/, /O/) are higher than the narrow-vowels (/I/, /i/, /u/, /U/). So from the duration viewpoint, vowels could be classified into two categories, wide vowels which have high mean durations and narrow vowels which have low mean durations.

Another interesting observation is that for all the vowels, outliers exist mostly for extreme high values. Confidence intervals of the mean durations of wide vowels lie in the range between 127 ms and 152 ms in the 1-word environment, between 105 ms and 118 ms in the sentence environment. Confidence intervals of the mean durations of narrow vowels lie in the range between 105 ms and 118 ms in the 1-word environment, between 75 ms and 90 ms in the sentence environment. Mean durations of the vowels in the 1-word and sentence environments are shown in the Figure 3.7.

Table 3.3. Mean durations of vowels (in ms)

Vowels	a	e	I	i	o	O	u	U
1-word environment	139	135	115	116	133	133	113	109
Sentence environment	112	105	81	82	109	110	81	84

Table 3.4. Ratio of mean durations in 1-word environment to mean durations in sentence environment for the vowels

Vowels	a	e	I	i	o	O	u	U
Compression values	1.24	1.29	1.42	1.42	1.22	1.22	1.40	1.30

From the Figures 3.3 and 3.4, it may be suspected that some vowels have the same probability distribution. To measure the similarity, ANOVA analysis is done, firstly to test the hypothesis that the means of the vowels are all the same in 1-word and sentence environments. It resulted in the conclusion that vowels have different means (in both environments) since ANOVA analysis give P-value of 0.00 in both environments for

this hypothesis. Tukey's test has been conducted to find the confidence intervals of mean differences of vowel pairs with overall significance level 0.05. Two vowels could be said to be similar from duration viewpoint if the confidence interval for mean difference includes zero. Vowels found to be similar with this analysis are given in Tables 3.5 and 3.6. From these Tables, it is seen that similar vowels /a/ and /o/, /a/ and /O/, /e/ and /o/, /e/ and /O/, /o/ and /O/ are all wide vowels. Also /I/ and /i/, /I/ and /u/, /I/ and /U/, /i/ and /u/, /i/ and /U/, /u/ and /U/ are similar which are all narrow vowels. Close relationship in sentence environment between /a/ and /o/, /i/ and /U/ and between /I/ and /U/ disappears in the 1-word environment.

Mean durations of the vowels in the sentence environment are all shorter than in the 1-word environment. Mean of the vowels are within the range of 109 ms and 139 ms in the 1-word environment and between 81 ms and 112 ms in the sentence environment. Sentence environment seems to affect mean durations of the vowels in a linear type compression by a factor between 1.20 and 1.42 (Table 3.4). An interesting observation is that vowel classification into two is also seen in the compression of the mean durations in the sentence environment compared to 1-word environment. Duration compression value can be defined as ratio of mean duration in 1-word environment over mean duration in sentence environment. Mean duration compression values are in the range between 1.21 and 1.29 for wide-vowels, in the range between 1.40 and 1.42 for narrow vowels except /U/, for which contraction value is 1.30. Duration compressions are lower for wide vowels which have larger mean durations than narrow vowels.

Maximum likelihood parameter estimates of the vowels for normal, lognormal and gamma distributions are given in the Tables 3.7 and 3.8. To illustrate the comparison between the relative frequencies of the vowels to the theoretical values of expected distributions, Q-Q plots (Figures 3.8, 3.14, 3.12, 3.18, 3.10, 3.16) and deviations from the theoretical distributions (Figures 3.9, 3.15, 3.11, 3.17, 3.13, 3.19) of the vowel /a/ are given. It is seen that gamma and log-normal distributions model the duration of the vowel /a/ quite well and better than normal distribution. For the low and mid-range of duration values, lognormal and gamma distributions fit with the relative histogram of the data quite well, but discrepancies occur for the high duration range (starting from

300 ms in the 1-word environment and 200 ms in the sentence environment). Normal distribution is not appropriate to model duration of the vowel /a/ as it deviates from it both in the low and high duration range. For other vowels, modelling performances of these distributions are same.

Table 3.5. Similar vowels in 1-word environment, '•' denotes similar vowels, 'S' denotes same vowel pair

Vowels	a	e	I	i	o	O	u	U
a	S					•		
e		S			•	•		
I			S	•			•	
i			•	S			•	
o		•			S	•		
O	•	•			•	S		
u			•	•			S	•
U							•	S

Table 3.6. Similar vowels in sentence environment, '•' denotes similar vowels, 'S' denotes same vowel pair

Vowels	a	e	I	i	o	O	u	U
a	S				•	•		
e		S			•	•		
I			S	•			•	•
i			•	S			•	•
o	•	•			S	•		
O	•	•			•	S		
u			•	•			S	•
U			•	•			•	S

Boğaziçi Üniversitesi Kütüphanesi

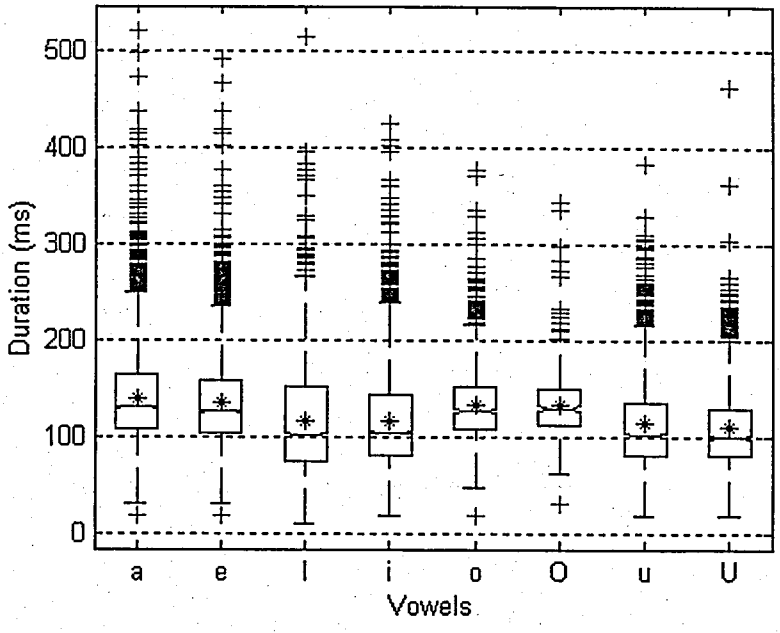

Figure 3.3. Boxplot of the durations of vowels in 1-word environment, stars are the means

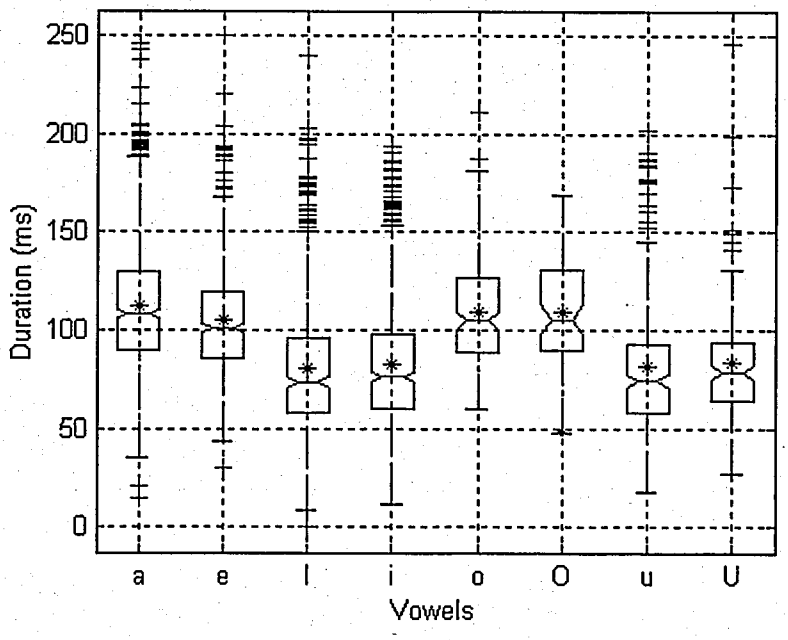

Figure 3.4. Boxplot of the durations of vowels in sentence environment, stars are the means

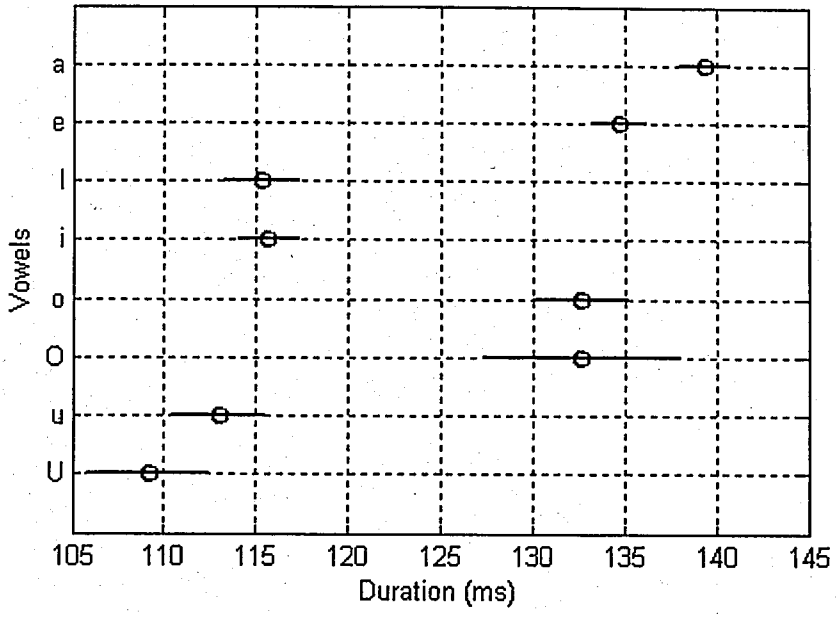

Figure 3.5. 95 per cent confidence intervals of the vowels' means in 1-word environment

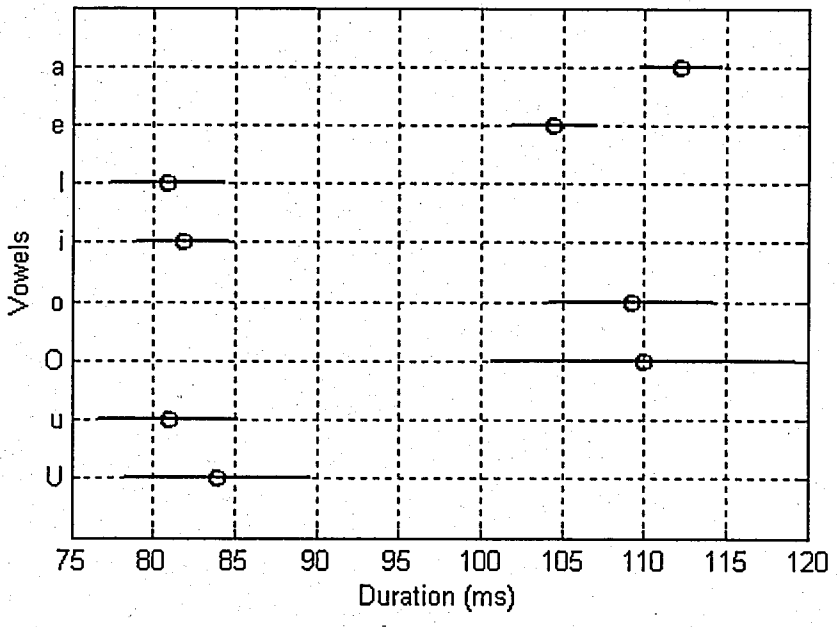

Figure 3.6. 95 per cent confidence intervals of the vowels' means in sentence environment

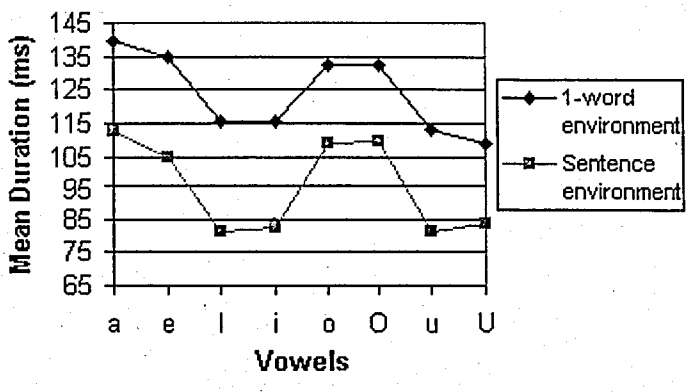

Figure 3.7. Mean durations of vowels in 1-word and sentence environments

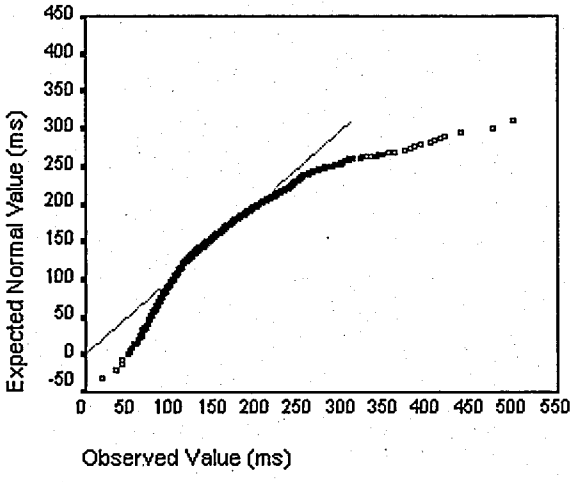

Figure 3.8. Normal Q-Q plot, vowel /ȧ/ in the 1-word environment

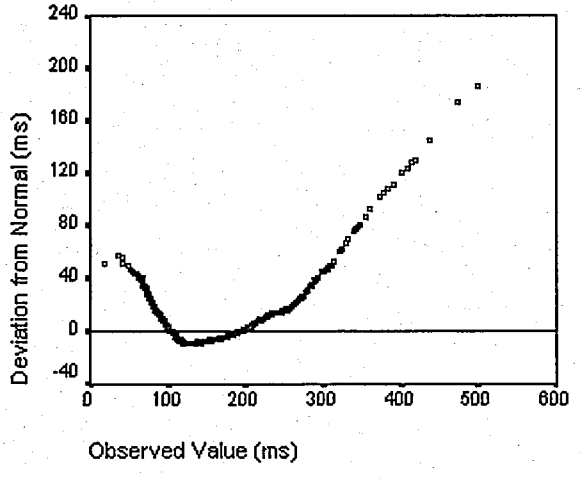

Figure 3.9. Deviation from normal distribution, vowel /a/ in the 1-word environment

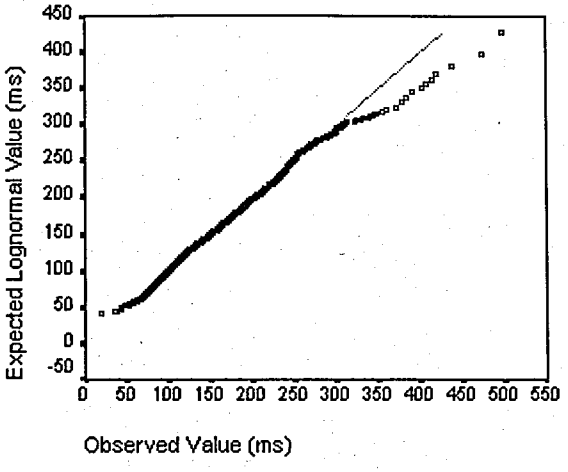

Figure 3.10. Log-Normal Q-Q plot, vowel /a/ in the 1-word environment

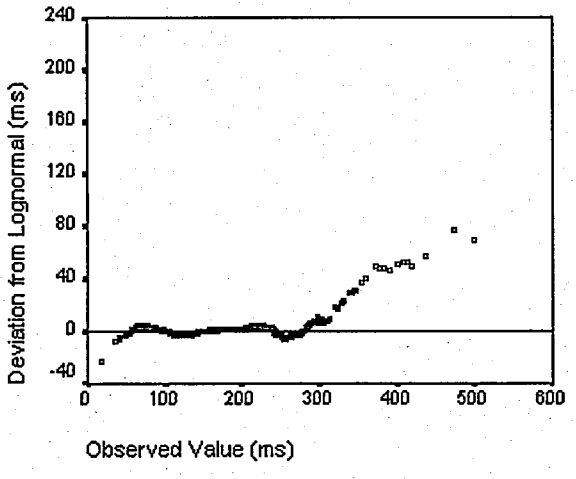

Figure 3.11. Deviation from log-normal distribution, vowel /a/ in the 1-word environment

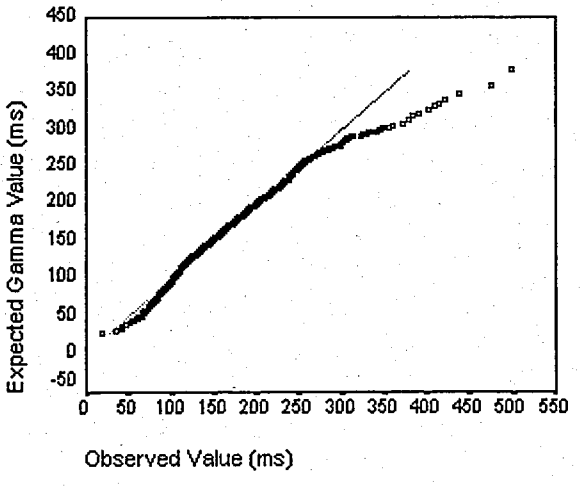

Figure 3.12. Gamma Q-Q plot, vowel /a/ in the 1-word environment

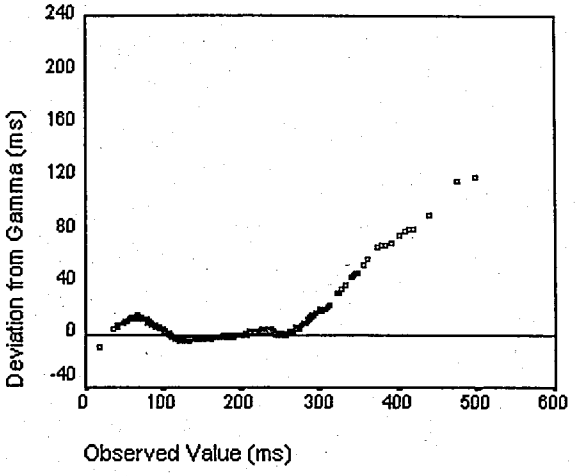

Figure 3.13. Deviation from gamma distribution, vowel /a/ in the 1-word environment

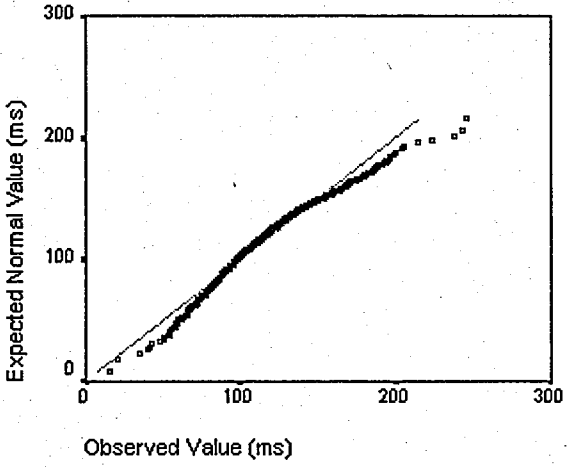

Figure 3.14. Normal Q-Q plot, vowel /a/ in the sentence environment

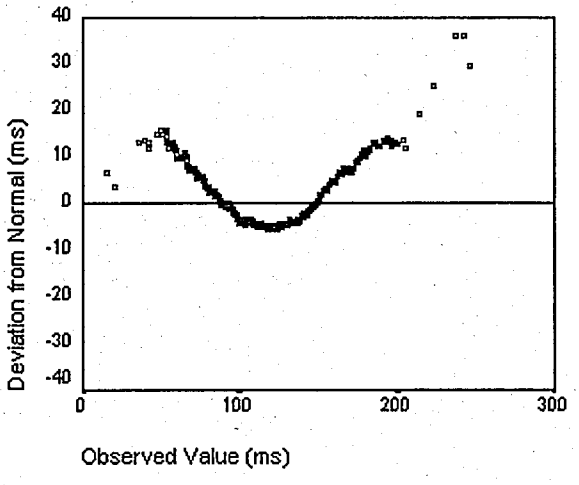

Figure 3.15. Deviation from normal distribution, vowel /a/ in the sentence environment

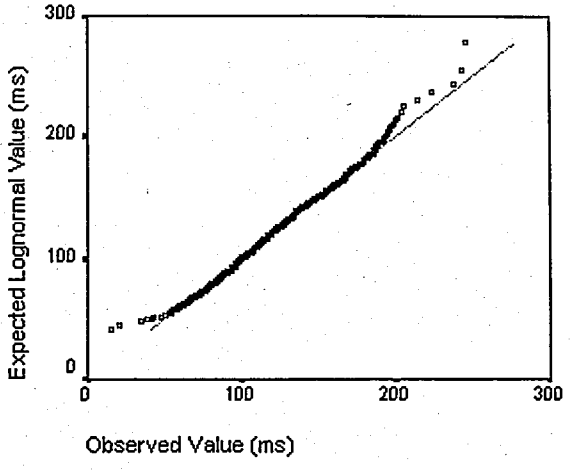

Figure 3.16. Log-Normal Q-Q plot, vowel /a/ in the sentence environment

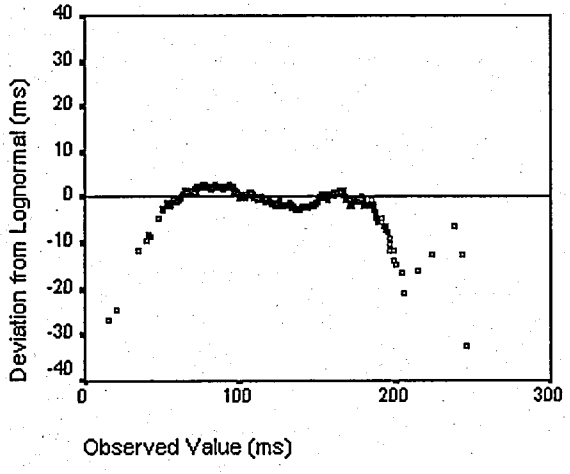

Figure 3.17. Deviation from log-normal distribution, vowel /a/ in the sentence environment

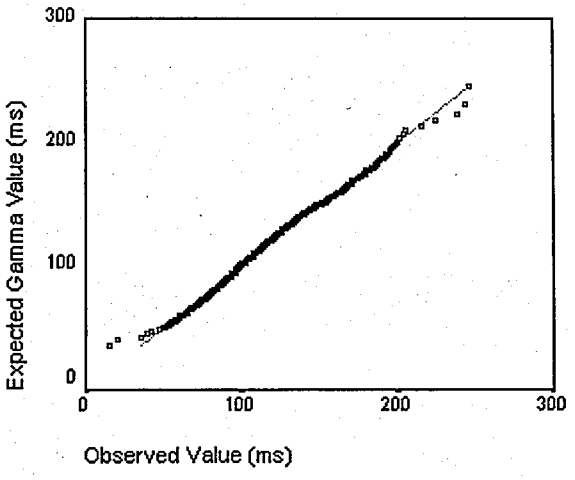

Figure 3.18. Gamma Q-Q plot, vowel /a/ in the sentence environment

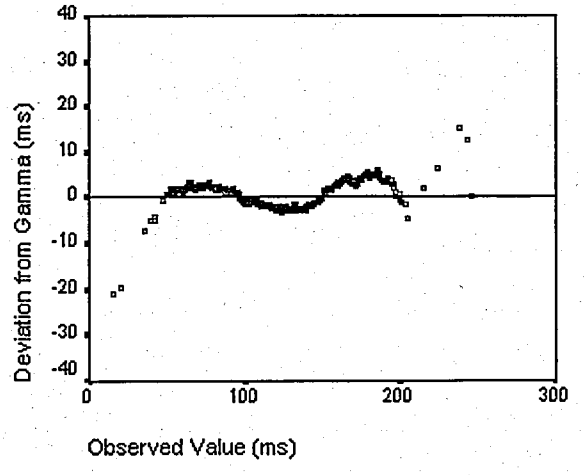

Figure 3.19. Deviation from gamma distribution, vowel /a/ in the sentence environment

Table 3.7. Estimated distribution parameters of vowels in 1-word environment

Vowel	Frequency	Normal pdf		Lognormal pdf		Gamma pdf	
		μ	σ	m	σ	a	b
a	7201	139.39	45.80	4.8878	0.3121	10.2663	13.5777
e	4762	134.70	43.68	4.8556	0.3051	10.6965	12.5929
I	2351	115.31	56.94	4.6316	0.4874	4.4692	25.8015
i	3830	115.67	50.77	4.6597	0.4297	5.651	20.4695
o	1468	132.61	37.78	4.8504	0.2706	13.6781	9.6948
O	427	132.66	34.56	4.858	0.2422	16.9317	7.8349
u	1563	112.99	47.66	4.6449	0.4079	6.2325	18.1286
U	954	109.17	43.96	4.6216	0.3768	7.1655	15.2361

Table 3.8. Estimated distribution parameters of vowels in sentence environment

Vowel	Frequency	Normal pdf		Lognormal pdf		Gamma pdf	
		μ	σ	m	σ	a	b
a	907	112.25	32.48	4.6785	0.2969	12.0138	9.3432
e	766	104.91	28.09	4.6175	0.2706	14.2089	7.3836
I	371	81.16	38.23	4.2879	0.4798	4.7683	17.021
i	601	82.23	32.74	4.3314	0.405	6.5566	12.5422
o	171	109.19	28.11	4.6605	0.2569	15.5186	7.036
O	65	109.79	26.57	4.668	0.2545	16.5064	6.6516
u	239	81.46	34.90	4.3185	0.4024	6.2889	12.9524
U	148	83.80	31.73	4.3674	0.3451	8.3641	10.0186

3.2.2. Consonants

Durational properties of consonants are observed to be much more complex than those of vowels as can be seen from box plots (Figures 3.20 and 3.21) and confidence intervals for the means (Figures 3.22 and 3.23) of the consonants' durations.

Like vowels, consonants have different means from each other (at least one of them has different mean than the others) in both environments since ANOVA analysis give P-value of 0.00 in both environments for the hypothesis that means of the consonants are equal. Mean durations of the consonants in the sentence environment are all shorter than in the 1-word environment. Consonants have mean durations between 62 ms and 144 ms in the 1-word environment and between 41 ms and 123 ms in the sentence environment (Table 3.9). Mean duration contraction values are in the range of 1.11 and 1.49, except phoneme /j/ which has compression value 1.62 (Table 3.9).

In the 1-word environment, consonants could be classified into four classes, taking only mean duration into consideration. Durations of the unvoiced fricatives /s/ and /S/ are the highest while phoneme /d/ has the lowest mean. Second class is comprised of the phonemes /k/, /C/ and /j/. Given the labelling convention adopted on the data in this study, the third class members are identified as /p/, /t/, /f/, /z/ and /n/. The hybrid nature of this class indicates that the labelling convention needs refinement. Rest of the phonemes (/b/, /d/, /g/, /c/, /v/, /h/, /m/, /l/, /r/ and /y/) make up the fourth class. In sentence environment, duration difference between some of these classes diminishes and they seem to merge, resulting in three classes. Phonemes /s/ and /S/ have the highest mean durations as in the 1-word environment. Also mean duration of /C/ is high. Second class in the sentence environment is composed of the phonemes /p/, /t/, /k/, /f/, /z/, /c/, /j/, /m/ and /n/. Finally third class is /b/, /d/, /g/, /v/, /h/, /l/, /r/ and /y/.

Looking at the durations of the consonants again from the classical classification viewpoint (Table 3.2), it is observed that mean durations of the unvoiced stops (/p/, /t/, /k/), unvoiced fricatives (/f/, /s/, /S/), and phonemes /C/, /j/ and /z/ are

high in the 1-word environment. This holds in the sentence environment, but mean durations of these phonemes come close to the other phonemes which have lower means. However unvoiced fricatives /s/ and /S/ still continue having much higher means than others. Durations of the voiced stops (/b/, /d/ and /g/), semivowels (/r/, /y/ and /l/), whisper (/h/) and voiced fricative phoneme /v/ are low in both environments.

Tukey's test has been conducted also for the consonants to find the confidence intervals of mean differences of consonant pairs with overall significance level 0.05. Two consonants are said to be similar from duration viewpoint if the confidence interval for mean difference includes zero. Similar consonants found by this analysis are given in Tables 3.10 and 3.11.

Maximum likelihood parameter estimates of the consonants for normal, lognormal and gamma distributions are given in the Tables 3.12 and 3.13. But for the consonant /j/, number of occurrence is two in the sentence environment, hence not enough data exists. As a representative of modelling performance of the consonants' durations, deviations of the duration data of consonant /b/ from normal, lognormal and gamma distributions are given in the Figures 3.25, 3.28, 3.26, 3.29, 3.27 and 3.30. Like for the vowel /a/, gamma and lognormal distributions are good at modelling distribution of the duration of consonant /b/, except for extreme high values (starting from about 150 ms) in the 1-word environment and extreme low (below near 30 ms) and high values (higher than 80 ms) in the sentence environment. Normal distribution deviates from data distribution both in the low and high duration range. For the other consonants, modelling performances of these distributions are alike.

Table 3.9. Mean durations (in ms) and mean duration compression values of consonants

Consonants	1-word environment	Sentence environment	Compression value
b	68	55	1.24
c	75	67	1.11
C	118	105	1.13
d	58	47	1.23
f	99	71	1.40
g	64	48	1.33
h	67	52	1.29
j	119	73	1.62
k	124	83	1.49
l	69	56	1.23
m	84	72	1.17
n	101	72	1.40
p	105	76	1.38
r	75	60	1.24
s	134	112	1.20
S	144	123	1.17
t	104	79	1.32
v	66	52	1.27
y	64	45	1.42
z	109	80	1.36

Table 3.10. Similar consonants in 1-word environment, '•' denotes similar consonants, 'S' denotes same consonant pair

	b	c	C	d	f	g	h	j	k	l	m	n	p	r	s	S	t	v	y	z
b	S	•				•	•			•								•	•	
c	•	S					•			•				•				•		
C			S					•	•											
d				S		•														
f					S							•	•				•			
g	•			•		S	•			•								•	•	
h	•	•				•	S			•								•	•	
j			•					S	•				•		•		•			•
k			•					•	S											
l	•	•				•	•			S								•		
m											S									
n					•							S	•				•			
p					•			•				•	S				•			•
r		•												S						
s								•							S					
S																S				
t					•			•				•	•				S			•
v	•	•				•	•			•								S	•	
y	•					•	•											•	S	
z								•					•				•			S

Table 3.11. Similar consonants in sentence environment, '•' denotes similar consonants, 'S' denotes same consonant pair

	b	c	C	d	f	g	h	j	k	l	m	n	p	r	s	S	t	v	y	z
b	S			•	•	•	•	•		•				•				•		
c		S			•			•			•	•	•	•			•			•
C			S					•							•					
d	•			S		•	•	•										•	•	
f	•	•			S		•	•	•	•	•	•	•	•			•	•		•
g	•			•		S	•	•		•								•	•	
h	•			•	•	•	S	•		•				•				•	•	
j	•	•	•	•	•	•	•	S	•	•	•	•	•	•	•	•	•	•	•	•
k					•			•	S				•				•			•
l	•				•	•	•	•		S				•				•		
m		•			•			•			S	•	•							•
n		•			•			•			•	S	•							•
p		•			•			•	•		•	•	S				•			•
r	•	•			•		•	•		•				S				•		
s			•					•							S					
S								•								S				
t		•			•			•	•				•				S			•
v	•			•	•	•	•	•		•				•				S	•	
y				•		•	•	•										•	S	
z		•			•			•	•		•	•	•				•			S

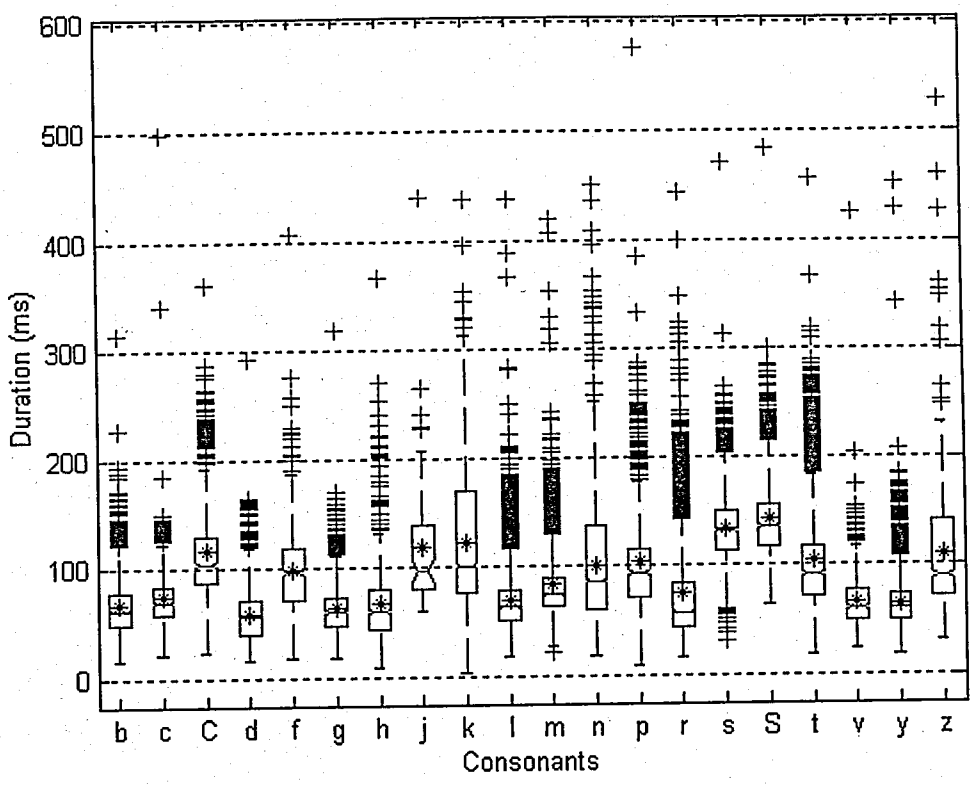

Figure 3.20. Boxplot of the durations of consonants in 1-word environment, stars are the means

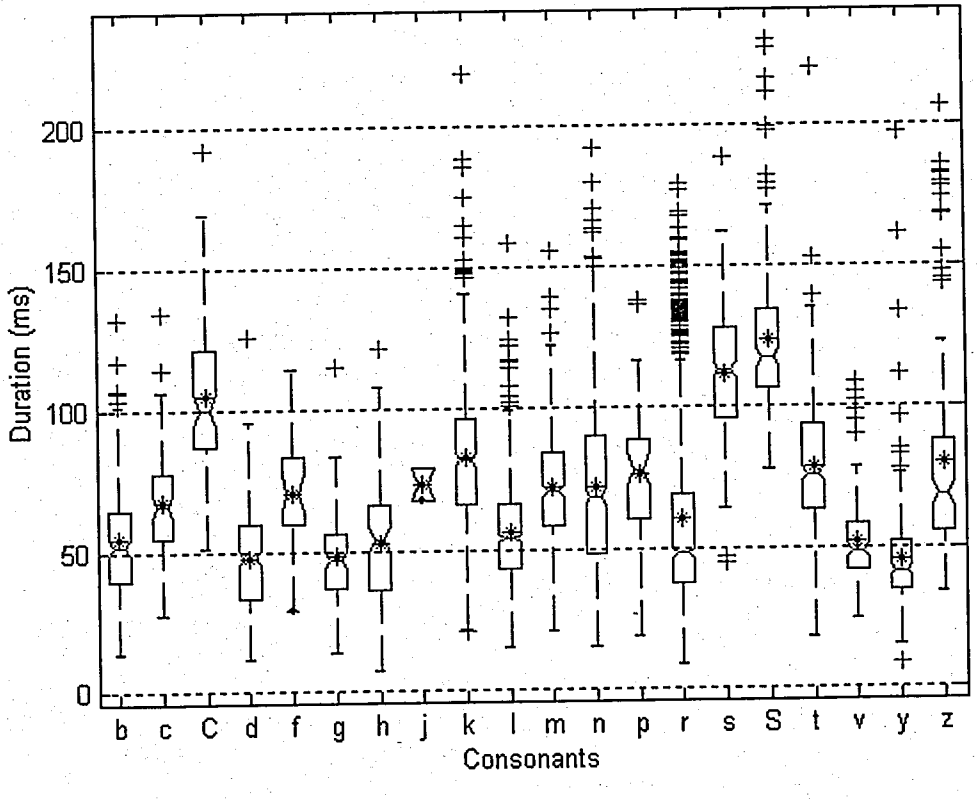

Figure 3.21. Boxplot of the duration of consonants in sentence environment, stars are the means

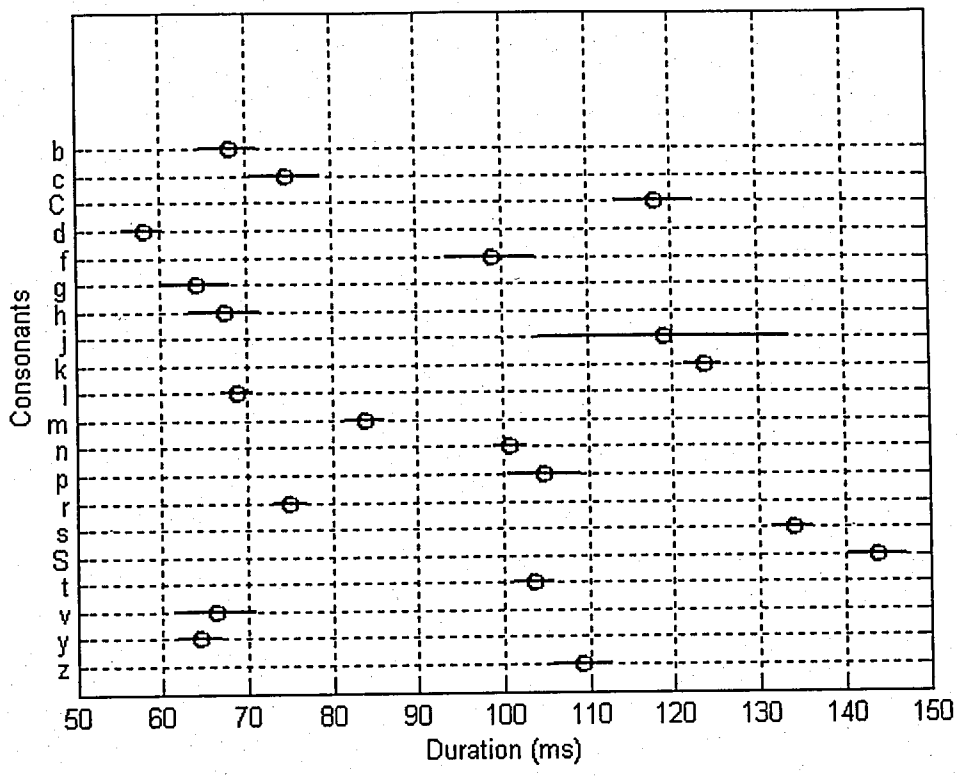

Figure 3.22. 95 per cent confidence intervals of the consonants' means in 1-word environment

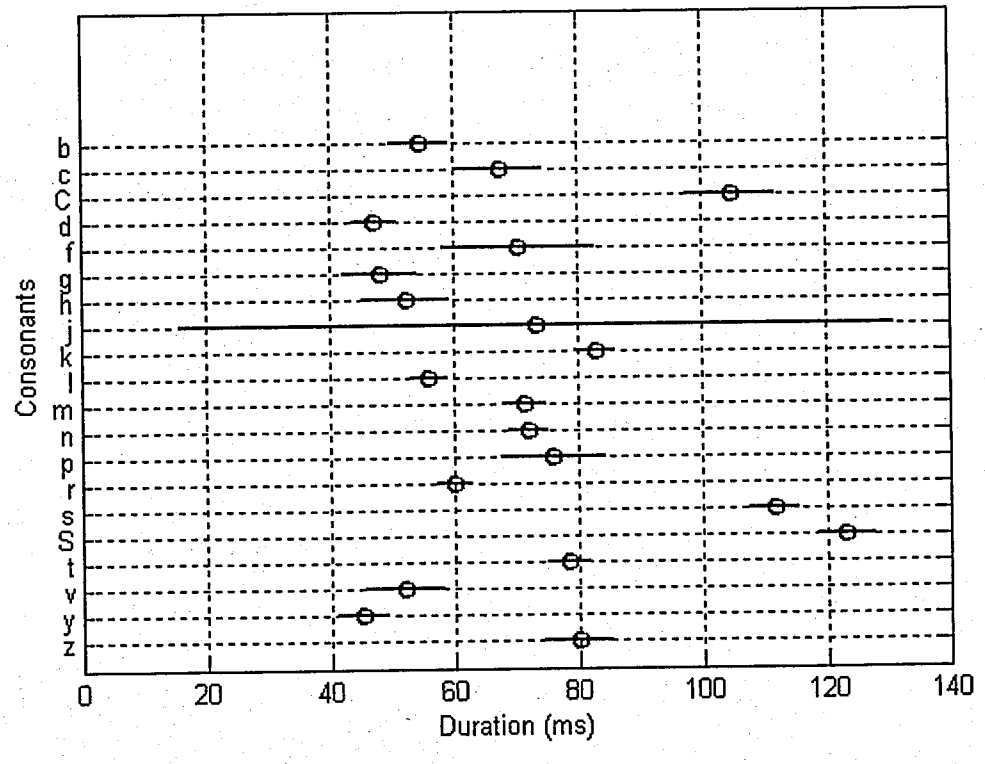

Figure 3.23. 95 per cent confidence intervals of the consonants' means in sentence environment

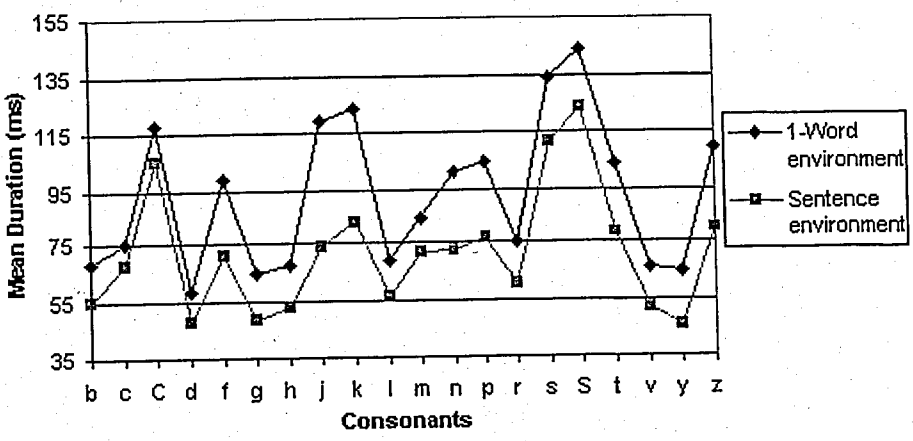

Figure 3.24. Mean durations of consonants in 1-word and sentence environments

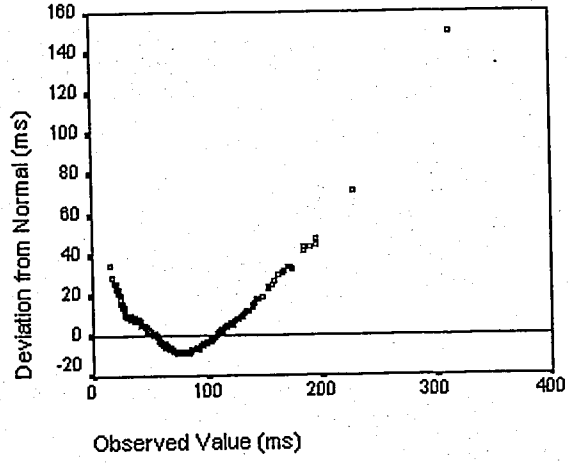

Figure 3.25. Deviation from normal distribution, consonant /b/ in the 1-word environment

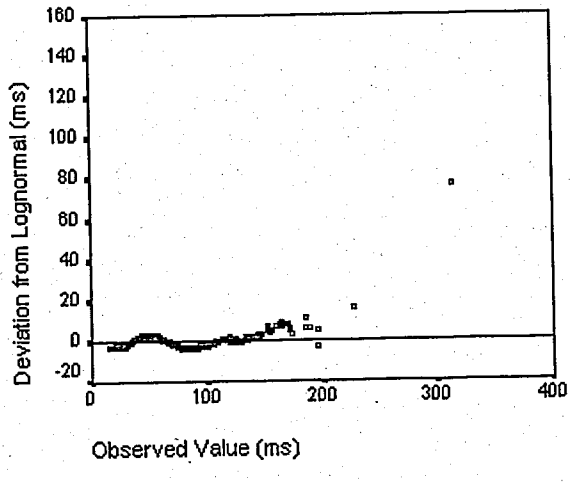

Figure 3.26. Deviation from lognormal distribution, consonant /b/ in the 1-word environment

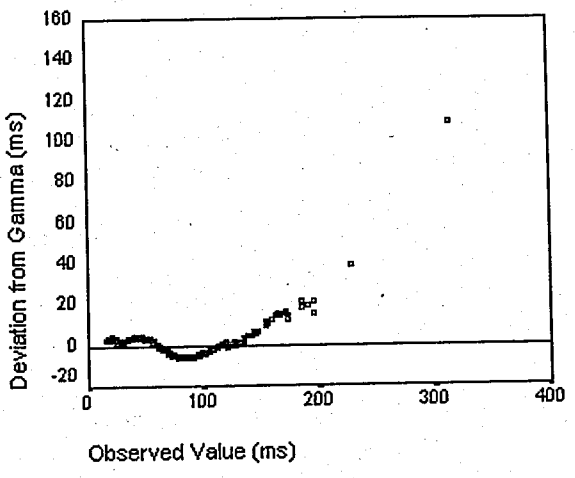

Figure 3.27. Deviation from gamma distribution, consonant /b/ in the 1-word environment

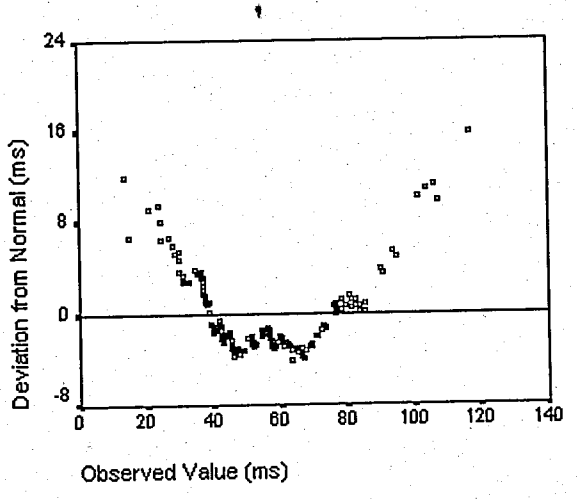

Figure 3.28. Deviation from normal distribution, consonant /b/ in the sentence environment

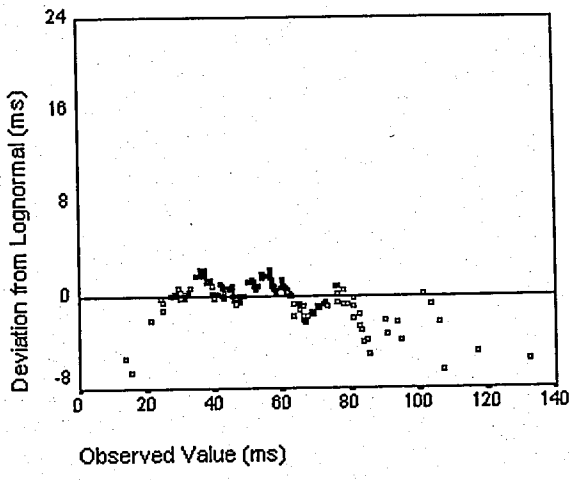

Figure 3.29. Deviation from lognormal distribution, consonant /b/ in the sentence environment

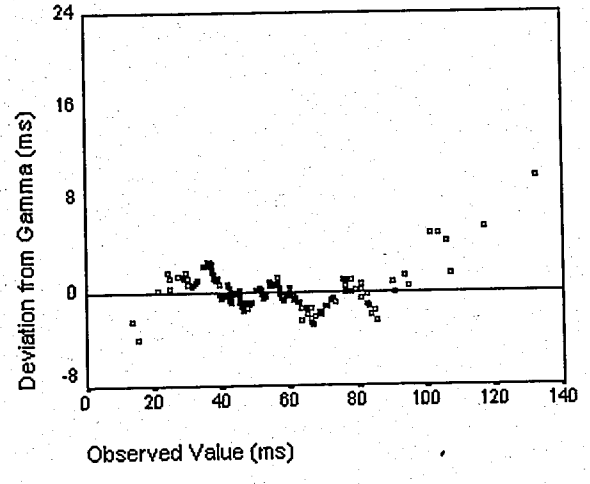

Figure 3.30. Deviation from gamma distribution, consonant /b/ in the sentence environment

Table 3.12. Estimated distribution parameters of consonants in 1-word environment

Consonant	Frequency	Normal pdf		Lognormal pdf		Gamma pdf	
		μ	σ	m	σ	a	b
b	998	68.034	29.6589	4.1351	0.4142	6.0498	11.2456
c	640	74.6989	30.0025	4.2603	0.312	9.5631	7.8111
C	539	118.1097	47.0461	4.7066	0.349	7.8486	15.0485
d	1931	58	23.8699	3.9742	0.4272	5.9603	9.731
f	442	98.8145	41.3378	4.5088	0.4247	6.0815	16.2485
g	738	64.2095	25.8262	4.0951	0.3594	7.6148	8.4322
h	636	67.4282	39.0126	4.0798	0.5067	3.9674	16.9957
j	76	118.9566	61.4958	4.6826	0.4162	5.3624	22.1836
k	2835	123.5423	63.482	4.6932	0.4964	4.2113	29.3356
l	3769	68.8485	28.6597	4.1656	0.3516	7.6983	8.9433
m	2285	83.7968	35.3726	4.3616	0.3479	7.6473	10.9578
n	3707	100.7088	51.3809	4.4872	0.5044	4.1585	24.2173
p	622	104.7587	55.2459	4.544	0.4517	4.8068	21.794
r	3636	74.999	48.0463	4.1634	0.5245	3.4038	22.034
s	1970	133.8422	33.4304	4.8658	0.2527	16.3675	8.1773
S	964	143.7897	40.3771	4.9346	0.2524	14.9832	9.5968
t	2320	103.5956	48.8103	4.5487	0.4172	5.6058	18.4802
v	531	66.2232	28.6288	4.1282	0.3442	7.8772	8.4069
y	1805	64.3327	28.0235	4.0996	0.3414	7.9207	8.1221
z	914	109.1682	53.7663	4.5927	0.4364	5.1509	21.1939

Table 3.13. Estimated distribution parameters of consonants in sentence environment

Consonant	Frequency	Normal pdf		Lognormal pdf		Gamma pdf	
		μ	σ	m	σ	a	b
b	173	54.6621	19.7527	3.9363	0.3686	7.8706	6.9451
c	75	67.4302	19.9546	4.1671	0.3044	11.5258	5.8504
C	77	104.9068	26.8117	4.6206	0.259	15.5693	6.7381
d	320	47.281	18.0716	3.7744	0.4242	6.2804	7.5283
f	30	70.5573	20.4516	4.2106	0.3216	11.0662	6.3759
g	106	48.0987	14.7838	3.8262	0.3148	10.7925	4.4567
h	81	52.4803	22.0373	3.8541	0.5046	4.8617	10.7946
j	2	73.3201	7.7245	4.2921	0.1055	179.8576	0.4077
k	384	82.9443	25.6375	4.3722	0.3074	11.0396	7.5133
l	574	55.9799	17.7954	3.9757	0.3176	10.3174	5.4258
m	357	71.6022	20.435	4.2292	0.2955	12.0996	5.9177
n	519	72.0962	30.9104	4.1831	0.4483	5.428	13.2823
p	56	76.1213	21.9921	4.2863	0.3232	11.0367	6.8971
r	596	60.1555	34.1165	3.9662	0.4932	3.9845	15.0974
s	249	111.6479	22.0774	4.6944	0.2102	24.0785	4.6368
S	171	123.2197	27.9402	4.792	0.204	22.8903	5.383
t	345	78.6369	22.7401	4.3231	0.2949	12.133	6.4812
v	99	52.1154	16.9401	3.9071	0.3003	10.9507	4.7591
y	231	45.2063	19.9163	3.7414	0.3627	7.3245	6.1719
z	127	80.06	36.3557	4.3023	0.3833	6.3752	12.558

3.3. Factors Affecting Durations of Turkish Phonemes

The usage of the term "factor" refers to categorization such as syllable number in the word which has *levels* one syllable, two syllables, three syllables, or more. These levels are computed from text in text-to-speech synthesis. Afterwards they are joined into feature vectors and then given to timing module. Following factors are found to affect duration in several studies for other languages [3].

I. Phonetic segment identity (number of phonemes)
II. Identities of surrounding segments
III. Syllabic stress
IV. Word importance (can be predicted from features such as of the word and its neighbors, its lexical identity, word frequency, and whether the word occurred in the sentences preceding the current sentence (given vs. new information))
V. Location of the syllable in the word, in the sentence, in the phrase
VI. Speaking rate (speed)
VII. Intonation

Additional factors have been claimed effecting durations of the phonemes, but evidence from empirical studies were found to be less certain than for the above factors ([15] contains results of a study which is investigation of factors which are speculated to affect durations of American English Phonemes). Although these additional factors could have importance, their effect is expected to be low compared to the above factors. In a study done for the American English [4], 94.5 per cent of the vowel duration variance that can be predicted from text is found to be predicted by the above first five factors. In this study, factors that could be computed from the text, which are stated in Section 3.1.1, are taken into consideration. They cover the above factors I, II and V along with syllable pattern, number of words in the sentence and number of syllables in the word. Basically these are the contextual factors.

The effect of these factors on the durations of phonemes is investigated via ANOVA analysis. In this analysis, whether a factor effects general (overall) mean

durations of vowels and consonants or not is investigated. Results of this ANOVA analysis for the vowels and consonants are given in the Tables 3.14 and 3.15. In these tables, factor column contains the factors whose effect on the mean durations of vowels and consonants are investigated. Level number is the number of levels particular factor can have. The d.f. ratio refers to the degree of freedom of the numerator and denominator of the F-ratio statistic.

3.3.1. Mean Length in Initial and Middle of Word

Mean length of sounds in initial and middle position in a word is affected by the nature of the preceding element, i.e. with respect to whether the preceding element is a vowel, consonant or pause. The effect of these three levels are inspected on the mean durations of phonemes in word initial and middle position. From the ANOVA Tables 3.14 and 3.15, it is determined that preceding phoneme type has effect in both environments, with P-values zero. Effect of preceding phoneme type on the durations of vowels and consonants can be seen more clearly in the Table 3.16 and Figures 3.31, 3.32, 3.33 and 3.34. For the vowels and consonants in the 1-word environment, preceding vowel has more lengthening effect on the duration of the sound in contrast to preceding consonant and pause. In the sentence environment, this also holds for the consonants. For the vowels in the sentence environment, preceding pause has the longest lengthening effect on the durations. Mean durations of vowels and consonants in both environments are lowest when preceding phoneme type is consonant.

3.3.2. Mean Length in Middle and Final of Word

Mean length of sounds in middle and final position in a word is affected by the nature of the preceding element, i.e. with respect to whether the following element is a vowel, consonant or pause. The effect of these three levels are inspected on the mean durations of phonemes in word middle and final position. From the ANOVA Tables 3.14 and 3.15, it is determined that following phoneme type has effect on the preceding phoneme in both environments, with P-values zero. Effect of following phoneme type on the durations of vowels and consonants can be seen more clearly in the Table 3.17

and Figures 3.35 3.36 3.37 and 3.38. Durations of vowels and consonants are longest when followed by pause in both environments. For the vowels, duration is longer when the vowel is followed by vowel than when followed by consonant, in both environments. Durations of consonants are longer when followed by consonant than when followed by vowel.

3.3.3. Effect of Preceding Vowel

This factor has eight levels corresponding to the eight vowels in the Turkish. Effect of preceding vowels on the vowels and consonants is depicted in Figures 3.39, 3.40, 3.41, 3.42 and Table 3.18. The P-value is zero for the consonants in both environment. So preceding vowel has effect on the following consonant duration. The P-value is 0.0143 and 0.1598 for the vowels, in 1-word and sentence environments respectively. So the preceding vowel has effect in the following vowel duration in the sentence environment but not in the 1-word environment, for significance level 0.05. However, data points in the sentence environment for the vowels which are preceded by vowels is quite low (nine observations). So this P-value for the vowels in the sentence environment is not so reliable which are preceded by vowels.

The ranking of preceding vowels making mean durations of consonants from longest to shortest is /I/, /i/, /e/, /a/, /u/, /U/, /o/ and /O/. This holds in both environments. It is interesting that consonant durations are longer in general when preceded by narrow vowels.

The ranking of preceding vowels making durations of following vowels from longest to shortest is /e/, /i/, /u/, /o/, /a/, /I/ and /U/ in the 1-word environment and /u/, /I/, /a/ and /o/ in the sentence environment.

3.3.4. Effect of Preceding Consonant

This factor has twenty-one levels corresponding to the twenty-one consonants in the Turkish. Effect of preceding consonant on the vowels and consonants is depicted

in Figures 3.43, 3.44, 3.45, 3.46 and Table 3.26. The P-values are zero in both environments, so preceding consonants effect following phonemes' durations.

General vowel mean duration is decreasing when vowels are preceded by the consonant order /c/, /z/, /d/, /n/, /j/, /m/, /v/, /f/, /y/, /l/, /s/, /b/, /h/, /g/, /r/, /t/, /k/, /p/, /S/ and /C/ in the 1-word environment. The order of consonants is /z/, /d/, /f/, /h/, /l/, /n/, /p/, /b/, /g/, /v/, /m/, /c/, /k/, /y/, /s/, /t/, /r/, /S/ and /C/ in the sentence environment. In general, vowel mean is high when preceded by /z/, /d/ and low when preceded by /S/, /C/, /t/ and /r/.

General consonant duration is decreasing when consonants are preceded by the consonant order /c/, /r/, /h/, /j/, /k/, /s/, /l/, /z/, /y/, /v/, /f/, /b/, /S/, /p/, /d/, /C/, /m/, /g/ and /n/ in the 1-word environment. The order of consonants is /r/, /h/, /k/, /C/, /S/, /l/, /y/, /t/, /v/, /s/, /p/, /m/, /c/, /z/, /g/, /d/, /b/, /n/ and /f/ in the sentence environment. In general, consonant mean is low when preceded by /n/ and high when preceded by /r/, /h/ and /k/.

3.3.5. Effect of Following Vowel

This factor has eight levels corresponding to the eight vowels in the Turkish. Effect of following vowels on the vowels and consonants is depicted in Figures 3.47, 3.48, 3.49, 3.50 and Table 3.19. Like the preceding vowel factor, P-values are below significance level 0.05 except for the vowels in the sentence environment, for which the observation number is nine. So this P-value for the vowels in the sentence environment is not so reliable which are followed by vowels.

The vowels have longest durations when followed by the vowel /i/. Vowels have less durations when followed by /u/, /e/, /a/ and /o/, in the order of decreasing mean durations of preceding vowels. Mean durations of vowels are 181 ms, 163 ms and 132 ms when followed by vowels /i/, /u/ and /e/ respectively. Again we see that following narrow vowels have much more lengthening effect of the durations of preceding vowels than the others. It is not appropriate to make comments about the effect of following

vowels' effect on the preceding vowels because of the so few data points.

For the consonants, mean durations are in decreasing order when followed by the vowels /i/, /I/, /o/, /U/, /u/, /O/, /a/ and /e/ in the sentence environment and /I/, /i/, /O/, /U/, /u/,/a/, /e/ and /o/ in the 1-word environment. Mean durations of the consonants are highest when followed by the narrow vowels /i/ and /I/ in both environments. Mean durations of the consonants are lowest when followed by the wide vowels /a/, /e/ and /o/ in the one-word environment and /a/, /e/ and /O/ in the sentence environment.

3.3.6. Effect of Following Consonant

This factor has twenty-one levels corresponding to the twenty-one consonants in the Turkish. Effect of following consonant on the vowels and consonants is depicted in Figures 3.51, 3.52, 3.53, 3.54 and Table 3.27.

General vowel mean duration is decreasing when vowels are followed by the consonant order /j/, /z/, /v/, /C/, /h/, /r/, /g/, /c/, /k/, /n/, /t/, /f/, /d/, /l/, /s/, /y/, /S/, /b/, /p/ and /m/ in the 1-word environment. The order of consonants is /z/, /g/, /v/, /r/, /C/, /h/, /s/, /l/, /y/, /f/, /b/, /c/, /k/, /n/, /t/, /d/, /p/, /S/ and /m/ in the sentence environment. In general, vowel mean is high when followed by the phonemes /z/, /v/, /C/, /h/, /r/ and low when followed by the phonemes /S/, /p/ and /m/.

General consonant duration is decreasing when consonants are followed by the consonant order /h/, /v/, /C/, /y/, /p/, /c/, /t/, /b/, /k/, /r/, /m/, /d/, /l/, /g/, /f/, /n/, /z/, /s/, /j/ and /S/ in the 1-word environment. The order of consonants is /y/, /t/, /p/, /b/, /c/, /k/, /C/, /l/, /h/, /d/, /r/, /m/, /v/, /n/, /g/, /z/, /f/, /s/ and /S/ in the sentence environment. In general, consonant mean is high when followed by the phonemes /y/, /p/, /t/, /b/, /h/, /c/, and low when followed by the phonemes /s/, /S/, /z/, /n/ and /g/.

3.3.7. Effect of Number of Syllables

This factor has seven levels: one syllable, two syllables, three syllables, four syllables, five syllables, six syllables or more. Effect of number of syllables in the word the vowels and consonants are in is depicted in Figures 3.55, 3.56, 3.57, 3.58 and Table 3.20.

For the syllable number factor, P-value is much lower than 0.05 significance level in both environments for the vowels and consonants, so syllable number is an effecting factor of the durations. From the figures and table, it is seen that mean durations of consonants and vowels decrease with increasing syllable number. In 1-word environment, mean of the vowels decreases from 189 ms in one syllable words to 109 ms in six syllable words. In sentence environment, mean of the vowels decreases from 108 ms in one syllable words to 88 ms in six syllable words.

Mean of consonants decreases from 148 ms in one syllable words to 74 ms in six syllable words, in 1-word environment. In sentence environment, general durations of consonants decrease from 73 ms in two syllable words to 61 ms in six syllable words. It is interesting that in this case, general durations of consonants in one syllable words is lower than consonants in four syllable words.

It can be said that the decrease in general durations of the phonemes with increase in the syllable number can be attributed to the speaking rate increase as syllable number increases.

3.3.8. Effect of Word Position

This factor has three levels: a phoneme's position in the word can be word-initial, word-middle or word-final. Effect of word position on the vowels and consonants is depicted in Figures 3.59, 3.60, 3.61, 3.62 and Table 3.21.

For the word position factor, P-value is much lower than 0.05 significance level

for the vowels and consonants, so word position is an effecting factor of the durations.

General means of vowels and consonants are highest in word-final position in both environments. General means of vowels and consonants are higher in word-middle position than in word initial position in both environments. Hence the the order of word positions general mean durations of vowels and consonants are high to low is word-final, word-initial and word-middle.

3.3.9. Effect of Syllable Pattern

This factor has ten levels corresponding to the ten syllable patterns a phoneme can be in: V, VT, TV, T, TVT, VTT, TTV, TTVT, TVTT and TTVTT, 'V' representing vowel and 'T' representing consonant. Although a consonant alone (T) can not be a syllable in Turkish, it is included to overcome the syllabification problem encountered in some words whose origin is not Turkish. Effect of word position on the vowels and consonants is depicted in Figures 3.63, 3.64, 3.65, 3.66 and Tables 3.22 and 3.23.

General vowel mean is decreasing when the syllable pattern is in the following order; VTT, TTVTT, VT, TVTT, V, TV, TTVT, TVT and TTV in the 1-word environment. The order is VT, VTT, TTVT, TVTT, V, TVT, TV and TTV in the sentence environment.

General consonant mean is decreasing when the syllable pattern is in the following order; VTT, TVT, TTVTT, TVTT, VT, TTVT, TTV and TV in the 1-word environment. The order is VT, VTT, TVT, TTVT, TVTT, TV and TTV in the sentence environment.

The syllable pattern orders are similar for vowels and consonants. For the vowels and consonants, duration means are higher in general for the syllable patterns in which vowel is followed by two consonants (i.e. VTT). Also duration means are lower for the syllable patterns ending with vowels (i.e. TTV, TV).

3.3.10. Effect of Sentence Position

This factor has three levels: a word's position -the phoneme occurs in- in the sentence can be sentence-initial, sentence-middle or sentence-final. Effect of sentence position on the vowels and consonants is depicted in Figures 3.67, 3.68 and Table 3.24.

For the sentence position factor, P-value is much lower than 0.05 significance level for the vowels and consonants, so sentence position is an effecting factor of the durations.

For the vowels, mean vowel duration is same for the sentence-final (102 ms) and sentence-initial positions (102 ms) and lower in sentence-middle position (95 ms). Mean consonant duration is highest in the sentence-final position (79 ms), lowest in the sentence-middle position (66 ms) and in the middle in the sentence-initial position (71 ms). So the position order is same for the consonants as in the word position factor making mean duration high to low, sentence-final, sentence-initial and sentence-middle.

3.3.11. Effect of Number of Words

This factor has seven levels according to the number of words in the sentence the phoneme is in: one word, two words, three words, four words, five words, six words or more words. Effect of number of words in the sentence the vowels and consonants is in is depicted in Figures 3.69, 3.70 and Table 3.25.

For the word number factor, P-value is much lower than 0.05 significance level for the vowels and consonants, so word number is an effecting factor of the durations.

Like syllable number factor, general durations of vowels and consonants decrease as word number in the sentences increase. General mean durations of vowels decrease from 112 ms in two word sentences to 96 ms in six word sentences. General mean durations of consonants decrease from 85 ms in two word sentences to 69 ms in six word sentences.

Table 3.14. ANOVA analysis of the factors on general duration means of vowels and consonants in 1-word environment

Factor	Level number	Vowels			Consonants		
		d.f. ratio	F ratio	P value	d.f. ratio	F ratio	P value
Preceding phoneme type	3	2/22553	65	0	2/31355	705	0
Following phoneme type	3	2/22553	7471	0	2/31355	16479	0
Preceding vowel	8	6/178	2.42	0.0286	7/19039	44	0
Preceding consonant	20	19/20474	51.6	0	19/6344	23.5	0
Following vowel	8	4/191	26.07	0	7/20796	65.88	0
Following consonant	20	19/18797	53.47	0	19/6344	22.58	0
Number of syllables in the word	7	6/22549	369	0	6/31351	407	0
Word position	3	2/22553	6075	0	2/31355	1807	0
Syllable pattern	9	8/22547	21.2	0	7/31322	339.95	0

Table 3.15. ANOVA analysis of the factors on general duration means of vowels and consonants in sentence environment

Factor	Level number	Vowels			Consonants		
		d.f. ratio	F ratio	P value	d.f. ratio	F ratio	P value
Preceding phoneme type	3	2/3265	48.9	0	2/4569	55	0
Following phoneme type	3	2/3265	99.36	0	2/4569	338	0
Preceding vowel	8	3/3	2.93	0.2004	7/2757	13.79	0
Preceding Consonant	20	19/2926	8.6	0	18/940	4.49	2.10^{-9}
Following vowel	8	3/4	3.75	0.1172	7/2986	12.52	7.10^{-16}
Following consonant	20	19/2707	9.03	0	18/940	6.26	10^{-14}
Number of syllables in the word	7	6/3261	11.5	9.10^{-13}	6/4565	7.11	10^{-7}
Word position	3	2/3265	136	0	2/4569	51	0
Syllable pattern	9	7/3260	21.48	0	6/4565	25.6	0
Sentence position	3	2/3265	14.81	4.10^{-7}	2/4569	65.22	0
Number of words	7	6/3261	5.76	6.10^{-6}	6/4565	11.26	10^{-11}

Table 3.16. Mean durations of vowels and consonants with respect to preceding phoneme type (ms)

	1-Word environment			Sentence environment		
	Vowel	Consonant	Punctuation	Vowel	Consonant	Punctuation
Vowels	162	127	134	98	96	116
Consonants	99	76	81	74	64	63

Table 3.17. Mean durations of vowels and consonants with respect to following phoneme type (ms)

	1-Word environment			Sentence environment		
	Vowel	Consonant	Punctuation	Vowel	Consonant	Punctuation
Vowels	131	115	199	98	94	117
Consonants	73	88	179	63	75	95

Table 3.18. Mean durations of vowels and consonants with respect to preceding vowel

(ms)

		/a/	/e/	/I/	/i/	/o/	/O/	/u/	/U/
1-Word environment	Vowels	150	182	146	176	151		170	112
	Consonants	99	100	111	103	85	77	96	91
Sentence environment	Vowels	83		121		65		129	
	Consonants	71	72	84	81	66	51	71	70

Table 3.19. Mean durations of vowels and consonants with respect to following vowel

(ms)

		/a/	/e/	/I/	/i/	/o/	/O/	/u/	/U/
1-Word environment	Vowels	110	132		181	108		163	
	Consonants	71	70	81	80	70	74	72	72
Sentence environment	Vowels	27	19		143	138			
	Consonants	59	58	68	69	64	60	63	63

Table 3.20. Mean durations of vowels and consonants with respect to syllable number

(ms)

		One	Two	Three	Four	Five	Six	More
1-Word environment	Vowels	189	145	131	118	111	109	103
	Consonants	148	105	89	82	78	74	71
Sentence environment	Vowels	108	104	99	95	92	88	89
	Consonants	66	73	71	68	67	61	68

Table 3.21. Mean durations of vowels and consonants with respect to word position

(ms)

	1-Word environment			Sentence environment		
	Word initial	Word middle	Word final	Word initial	Word middle	Word final
Vowels	119	100	168	101	85	108
Consonants	86	73	111	67	66	76

Table 3.22. Mean durations of vowels and consonants with respect to syllable pattern in 1-word environment (ms)

	V	VT	TV	TVT	VTT	TTV	TTVT	TVTT	TTVTT
Vowels	131	142	129	125	153	106	126	132	150
Consonants		98	74	101	103	79	87	100	103

Table 3.23. Mean durations of vowels and consonants with respect to syllable pattern in sentence environment (ms)

	V	VT	TV	TVT	VTT	TTV	TTVT	TVTT	TTVTT
Vowels	103	128	94	97	123	93	110	109	
Consonants		76	62	74	76	60	73	72	

Table 3.24. Mean durations of vowels and consonants with respect to sentence position (ms)

	Sentence initial	Sentence middle	Sentence final
Vowels	102	95	102
Consonants	71	66	79

Table 3.25. Mean durations of vowels and consonants with respect to word number (ms)

	One	Two	Three	Four	Five	Six	More
Vowels	96	112	96	101	101	96	96
Consonants	64	85	75	71	71	69	67

Table 3.26. Mean durations of vowels and consonants with respect to preceding consonant (ms)

	1-word		Sentence	
	Vowel	Consonants	Vowels	Consonants
b	125	76	99	55
c	148	90	94	59
C	107	69	78	66
d	145	71	109	56
f	129	76	105	47
g	122	64	98	56
h	123	82	101	73
j	136	79	140	
k	114	79	92	67
l	128	79	101	65
m	132	68	94	59
n	142	61	101	54
p	114	71	100	59
r	119	88	86	76
s	126	79	90	61
S	112	73	80	66
t	115	72	88	63
v	131	77	98	62
y	128	77	91	64
z	145	78	112	56

Table 3.27. Mean durations of vowels and consonants with respect to following consonant (ms)

	1-word		Sentence	
	Vowel	Consonants	Vowels	Consonants
b	103	94	91	80
c	117	97	91	77
C	124	100	98	76
d	112	88	90	73
f	112	78	92	58
g	117	82	112	67
h	123	101	97	73
j	142	64	126	
k	116	93	91	77
l	112	84	95	75
m	100	89	78	70
n	115	78	91	69
p	102	97	89	81
r	123	90	102	72
s	109	67	97	55
S	106	56	82	51
t	113	96	91	89
v	128	101	104	69
y	107	98	92	108
z	135	75	112	63

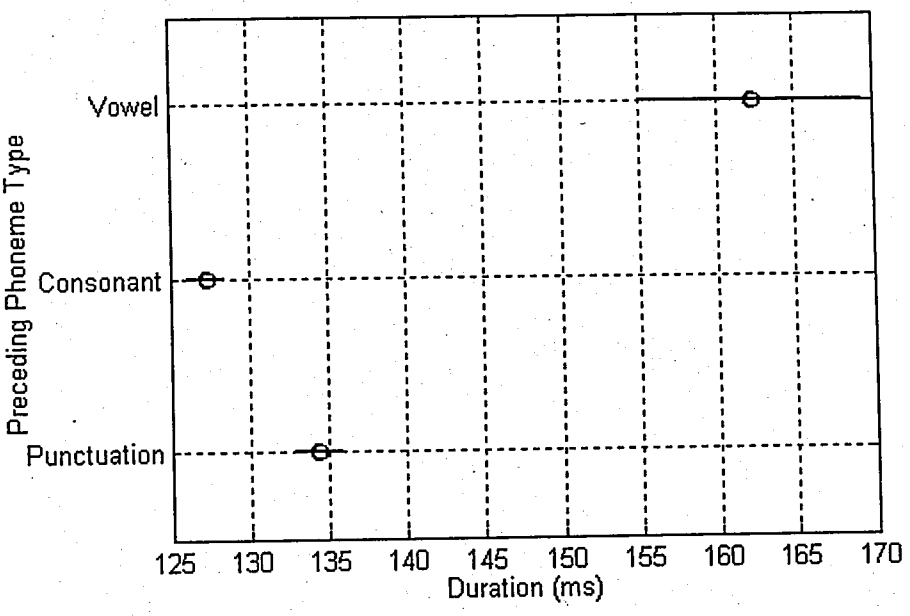

Figure 3.31. 95 per cent confidence intervals of the vowels' means with respect to preceding phoneme type, 1-word environment

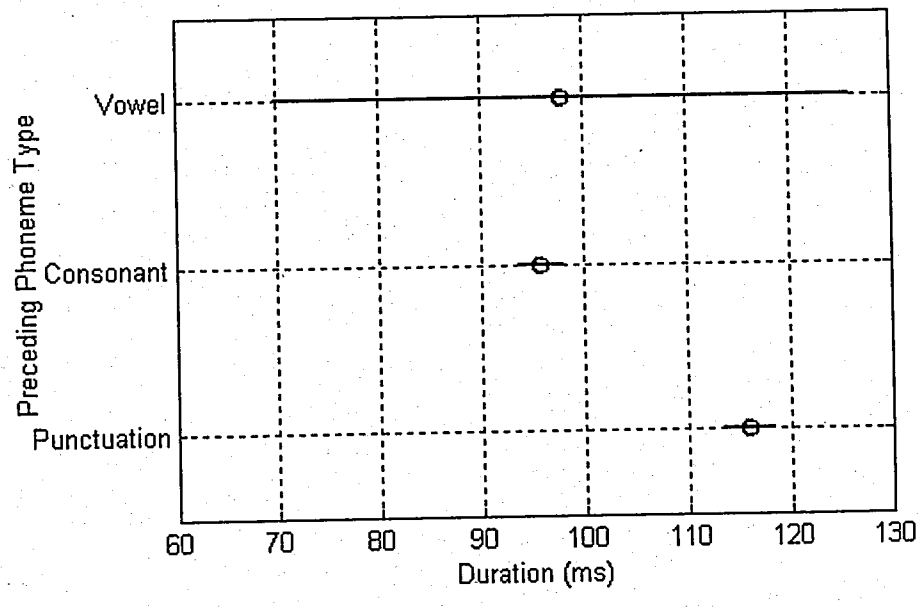

Figure 3.32. 95 per cent confidence intervals of the vowels' means with respect to preceding phoneme type, sentence environment

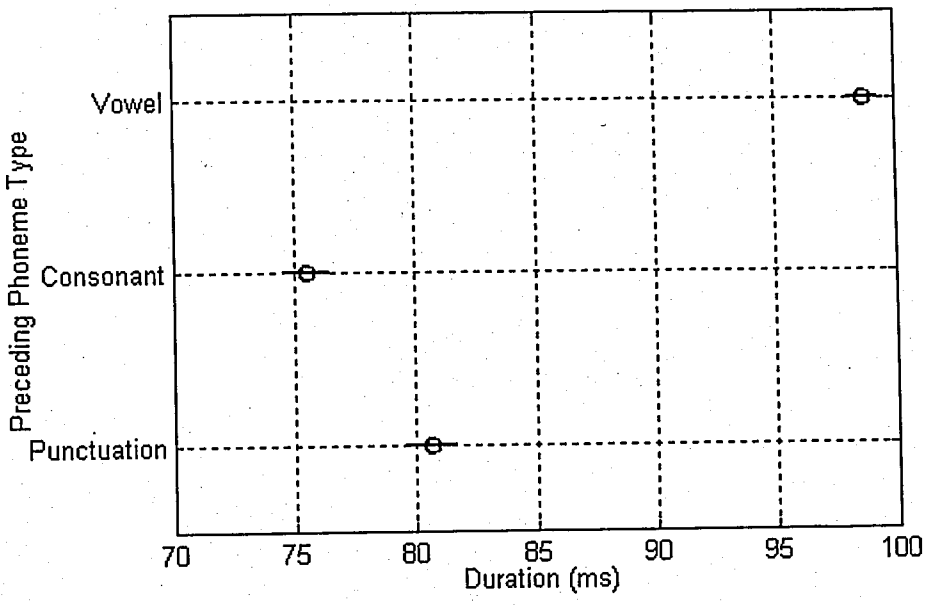

Figure 3.33. 95 per cent confidence intervals of the consonants' means with respect to preceding phoneme type, 1-word environment

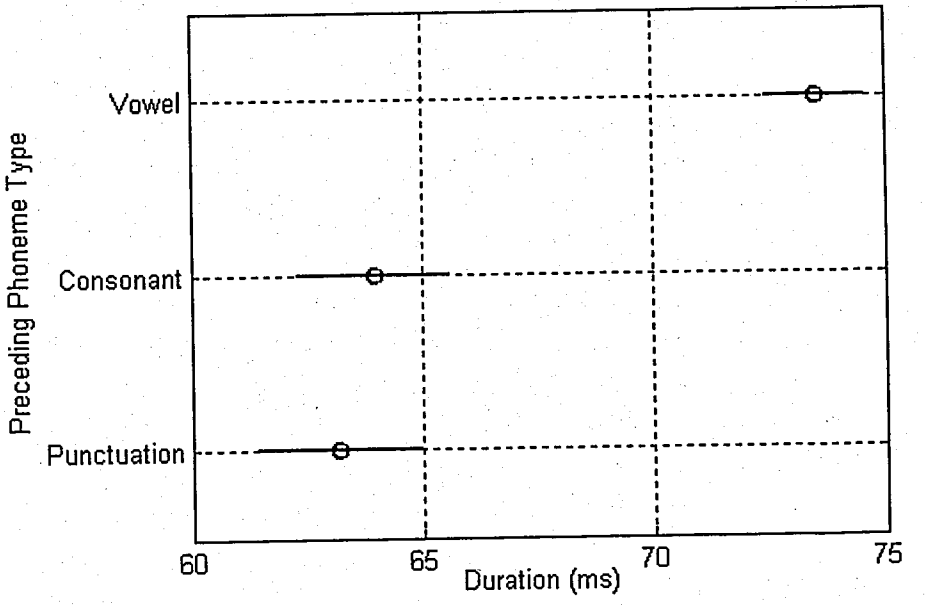

Figure 3.34. 95 per cent confidence intervals of the consonants' means with respect to preceding phoneme type, sentence environment

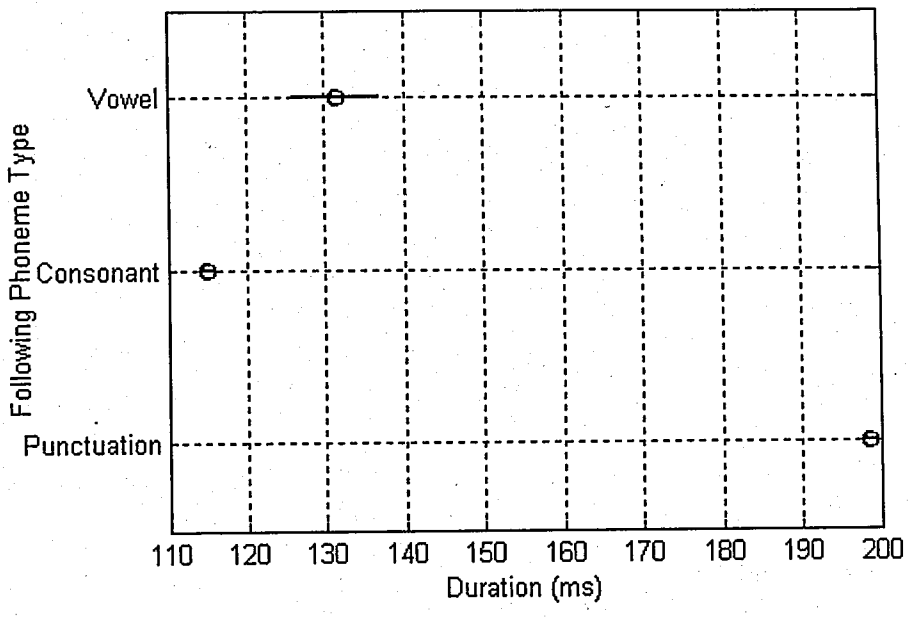

Figure 3.35. 95 per cent confidence intervals of the vowels' means with respect to following phoneme type, 1-word environment

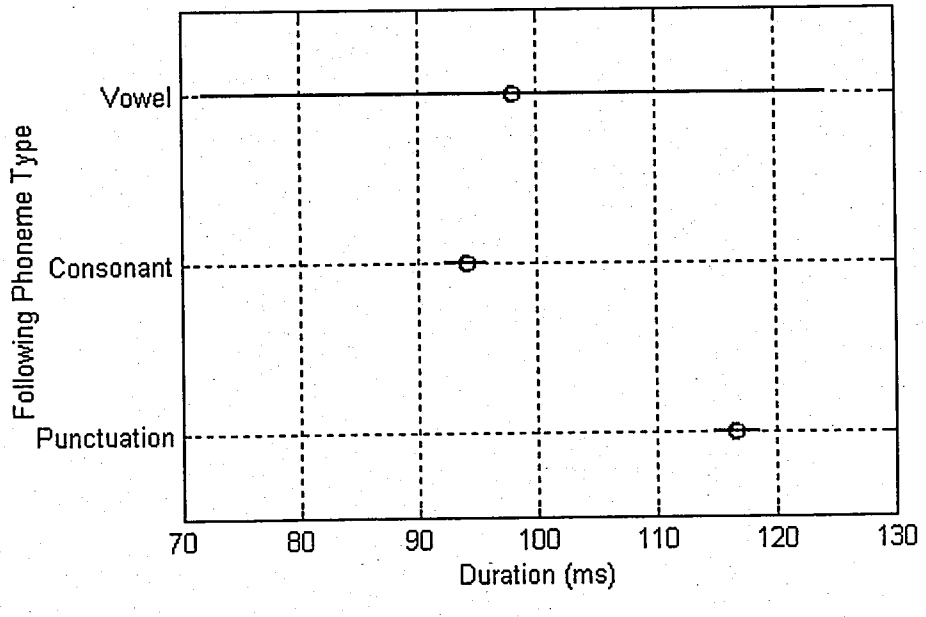

Figure 3.36. 95 per cent confidence intervals of the vowels' means with respect to following phoneme type, sentence environment

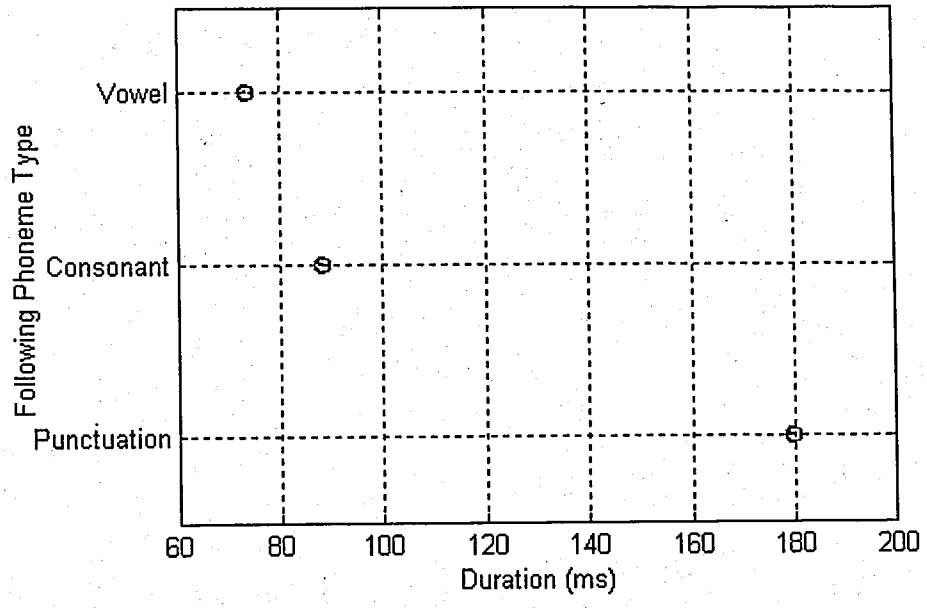

Figure 3.37. 95 per cent confidence intervals of the consonants' means with respect to following phoneme type, 1-word environment

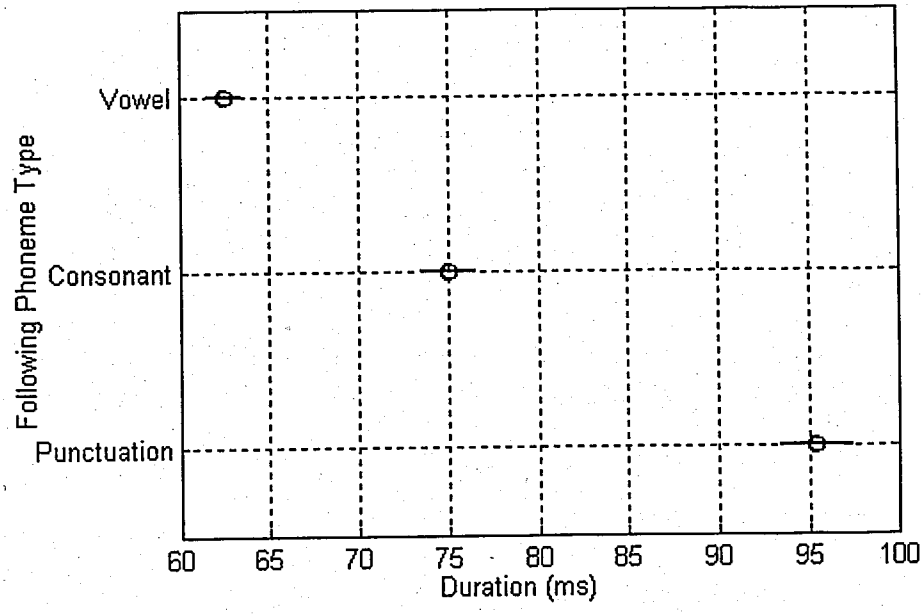

Figure 3.38. 95 per cent confidence intervals of the consonants' means with respect to following phoneme type, sentence environment

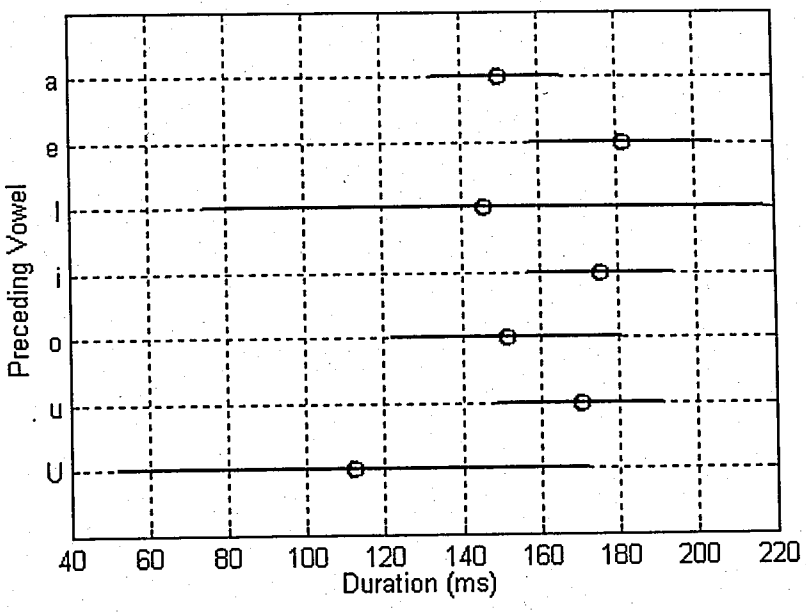

Figure 3.39. 95 per cent confidence intervals of the vowels' means with respect to preceding vowel, 1-word environment

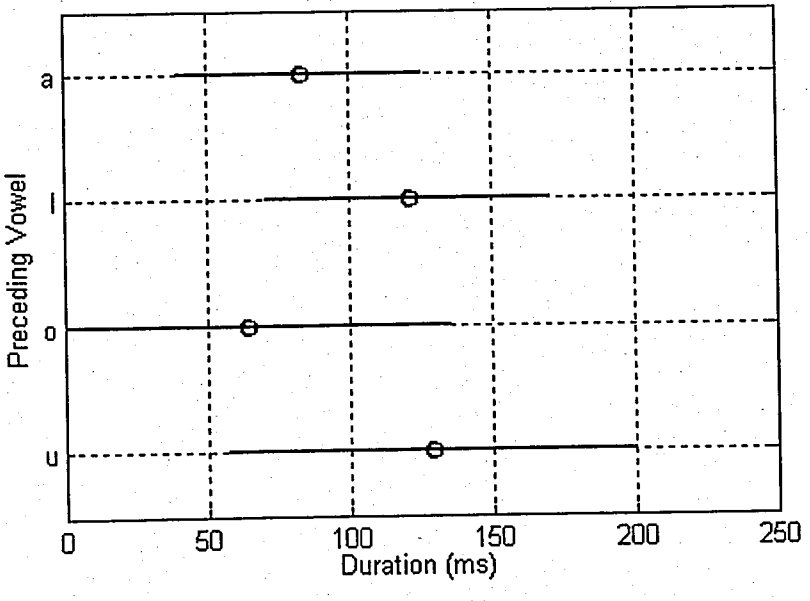

Figure 3.40. 95 per cent confidence intervals of the vowels' means with respect to preceding vowel, sentence environment

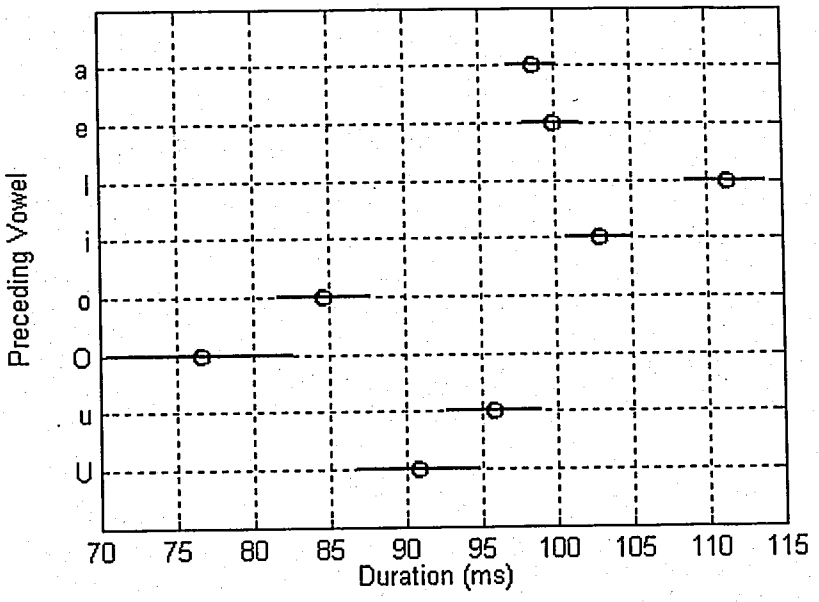

Figure 3.41. 95 per cent confidence intervals of the consonants' means with respect to preceding vowel, 1-word environment

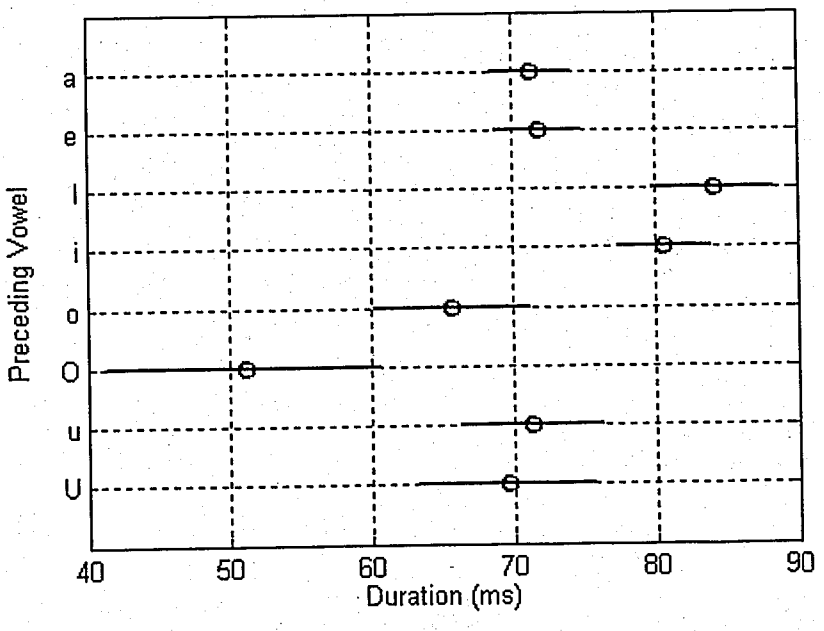

Figure 3.42. 95 per cent confidence intervals of the consonants' means with respect to preceding vowel, sentence environment

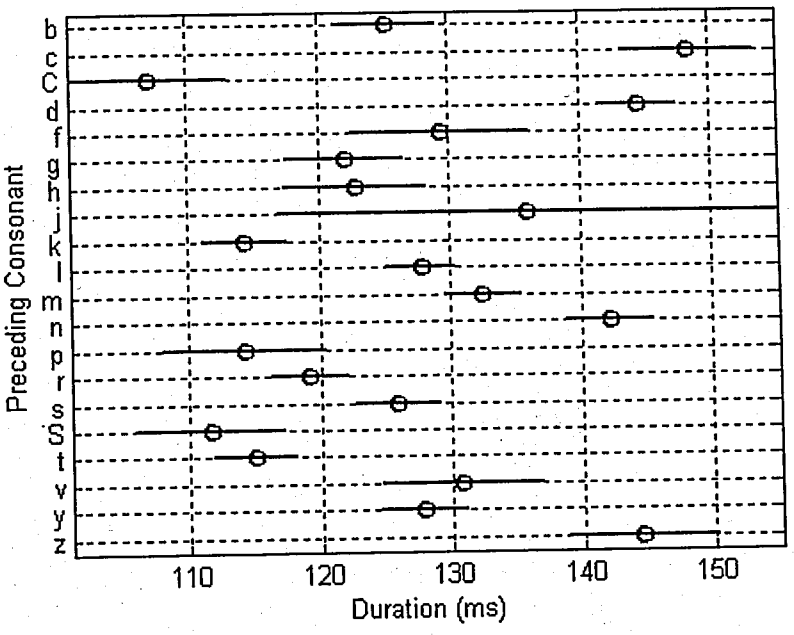

Figure 3.43. 95 per cent confidence intervals of the vowels' means with respect to preceding consonant, 1-word environment

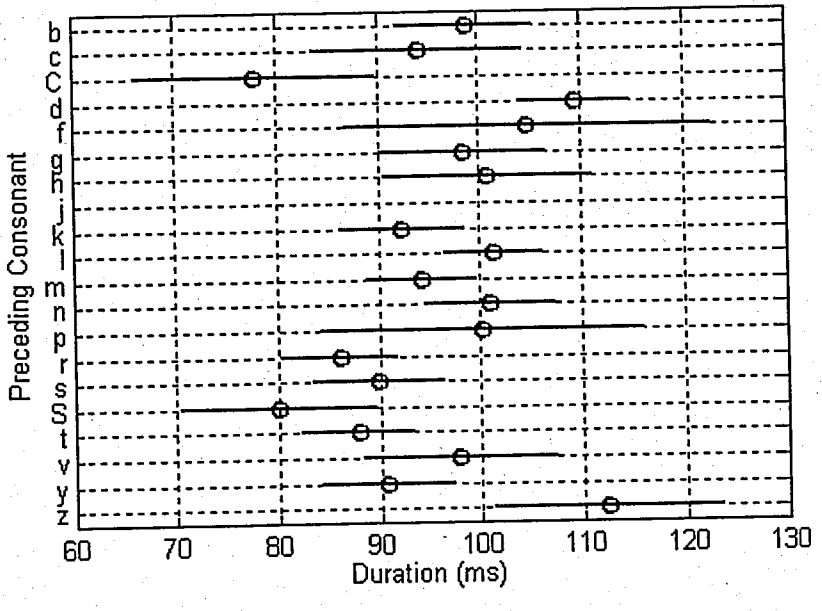

Figure 3.44. 95 per cent confidence intervals of the vowels' means with respect to preceding consonant, sentence environment

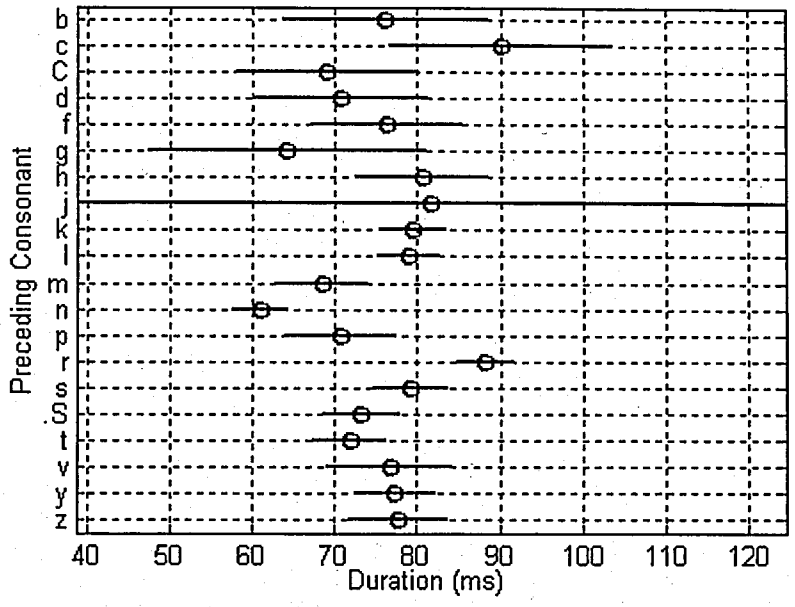

Figure 3.45. 95 per cent confidence intervals of the consonants' means with respect to preceding consonant, 1-word environment

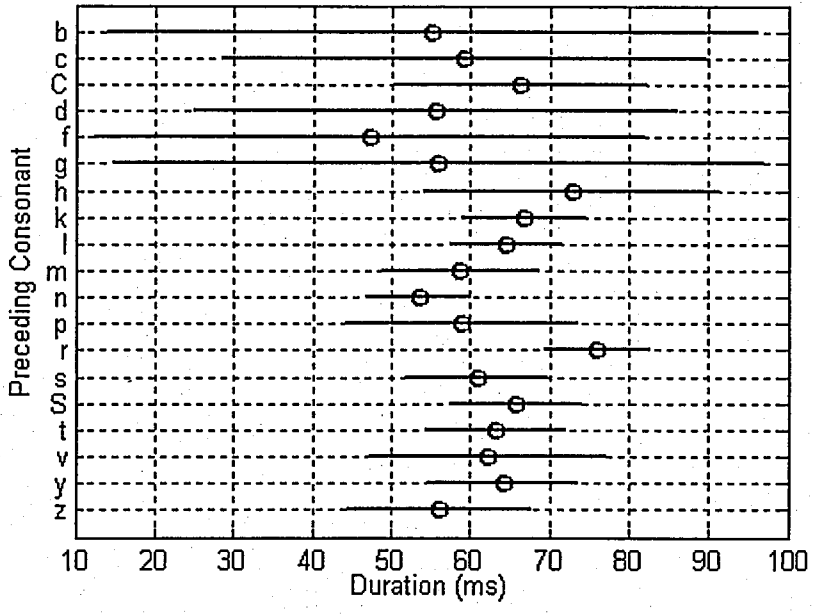

Figure 3.46. 95 per cent confidence intervals of the consonants' means with respect to preceding consonant, sentence environment

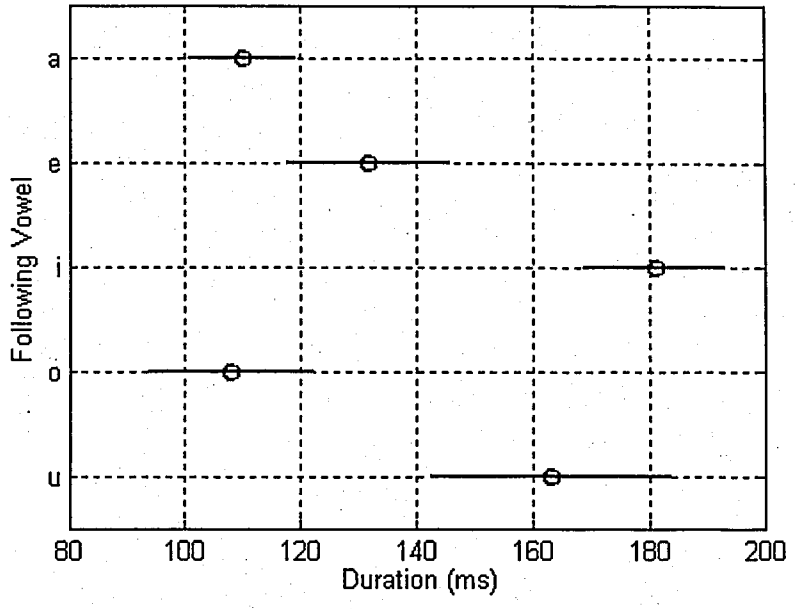

Figure 3.47. 95 per cent confidence intervals of the vowels' means with respect to following vowel, 1-word environment

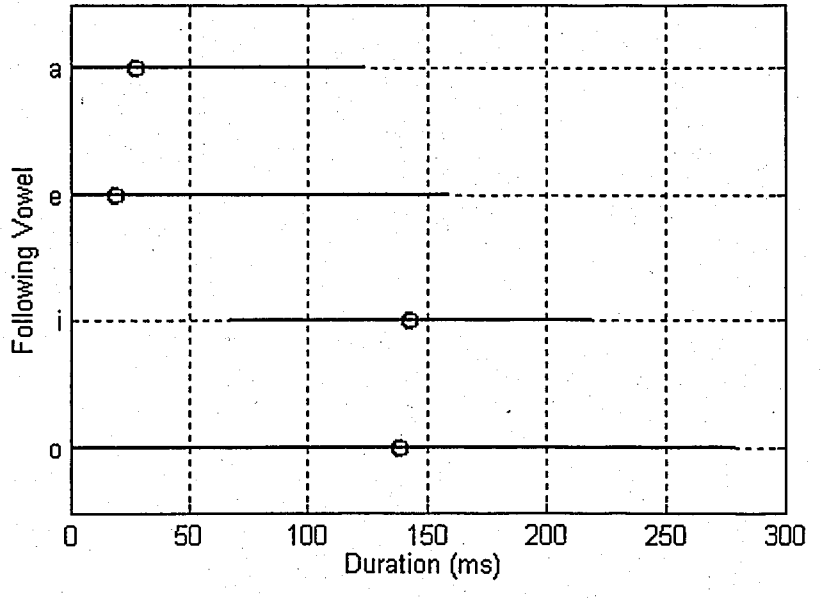

Figure 3.48. 95 per cent confidence intervals of the vowels' means with respect to following vowel, sentence environment

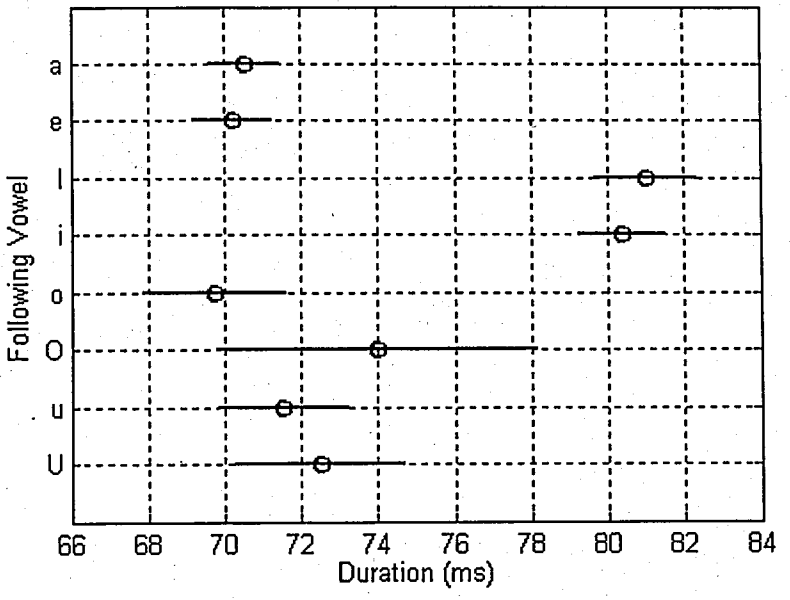

Figure 3.49. 95 per cent confidence intervals of the consonants' means with respect to following vowel, 1-word environment

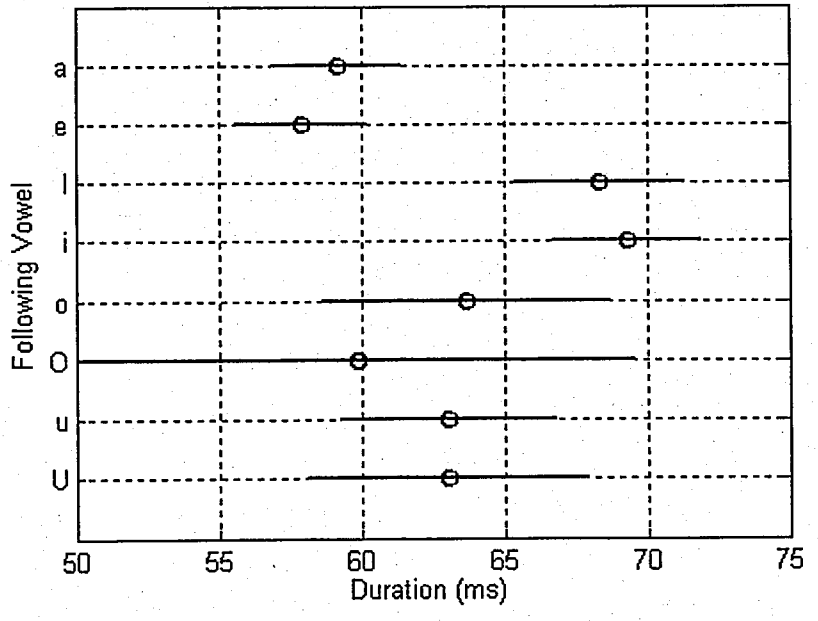

Figure 3.50. 95 per cent confidence intervals of the consonants' means with respect to following vowel, sentence environment

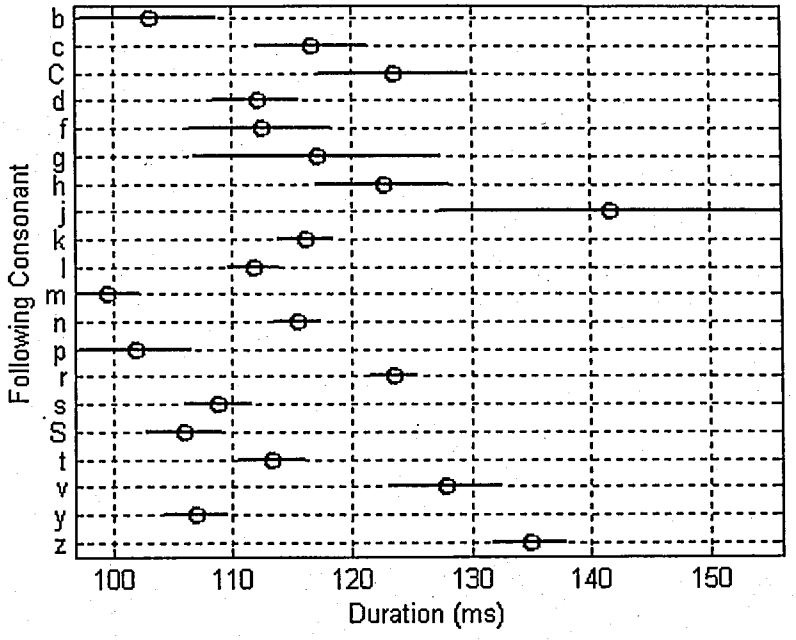

Figure 3.51. 95 per cent confidence intervals of the vowels' means with respect to following consonant, 1-word environment

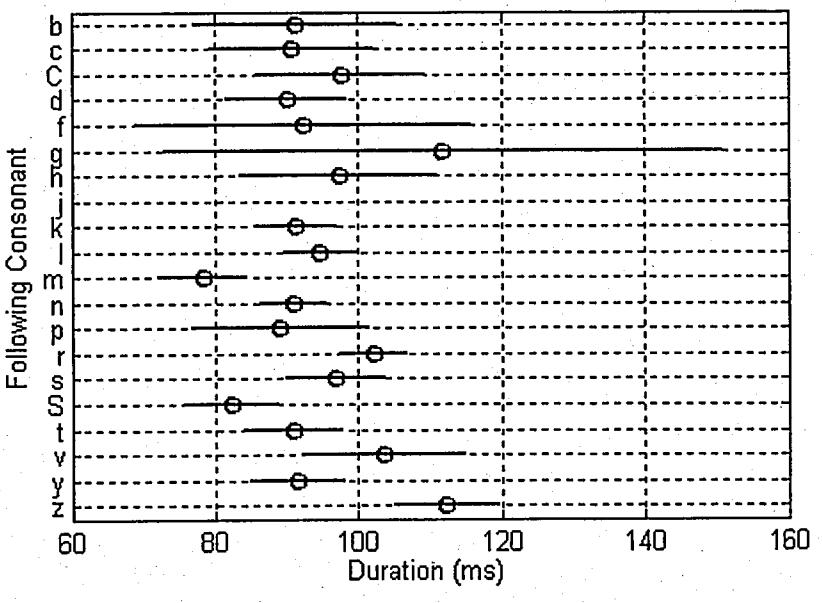

Figure 3.52. 95 per cent confidence intervals of the vowels' means with respect to following consonant, sentence environment

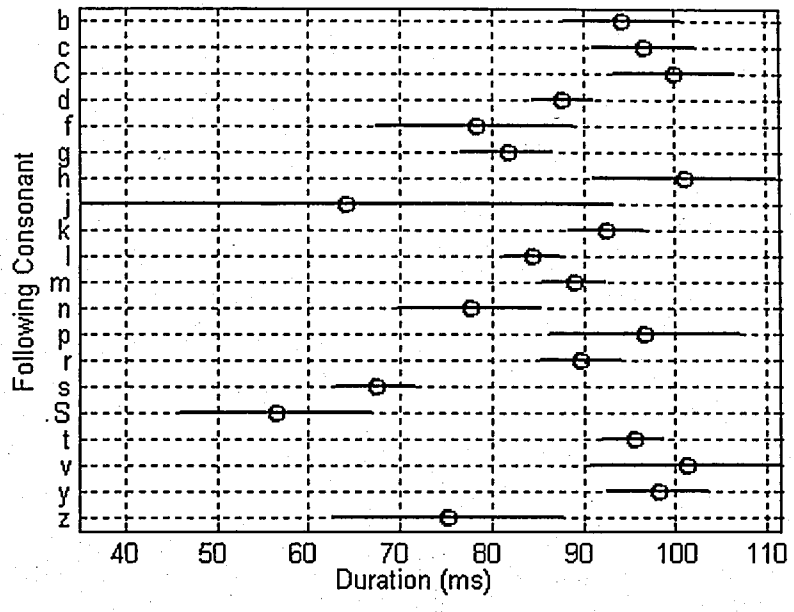

Figure 3.53. 95 per cent confidence intervals of the consonants' means with respect to following consonant, 1-word environment

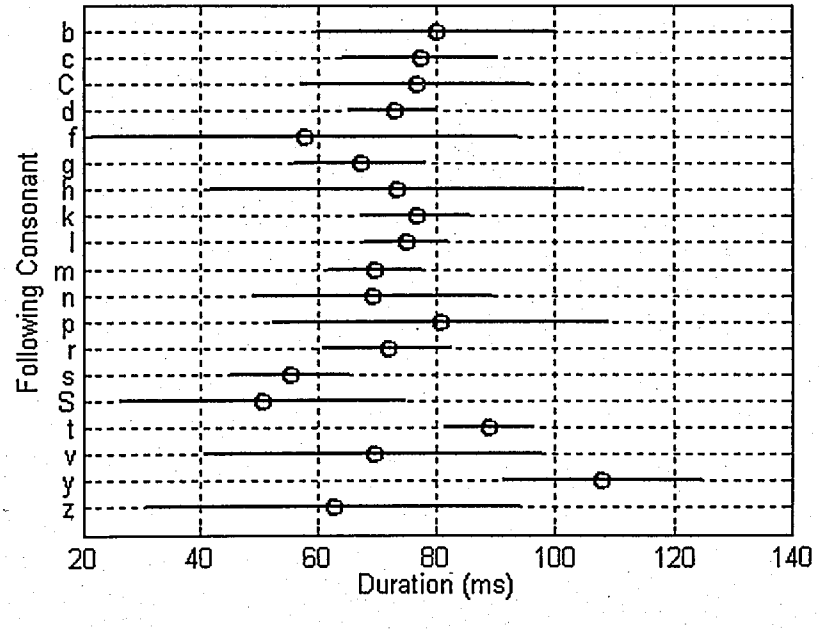

Figure 3.54. 95 per cent confidence intervals of the consonants' means with respect to following consonant, sentence environment

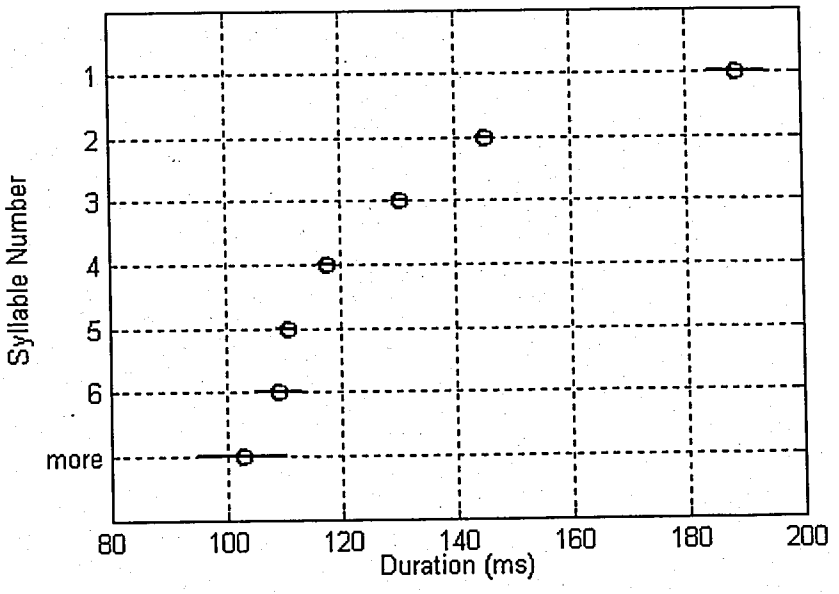

Figure 3.55. 95 per cent confidence intervals of the vowels' means with respect to syllable numbers, 1-word environment

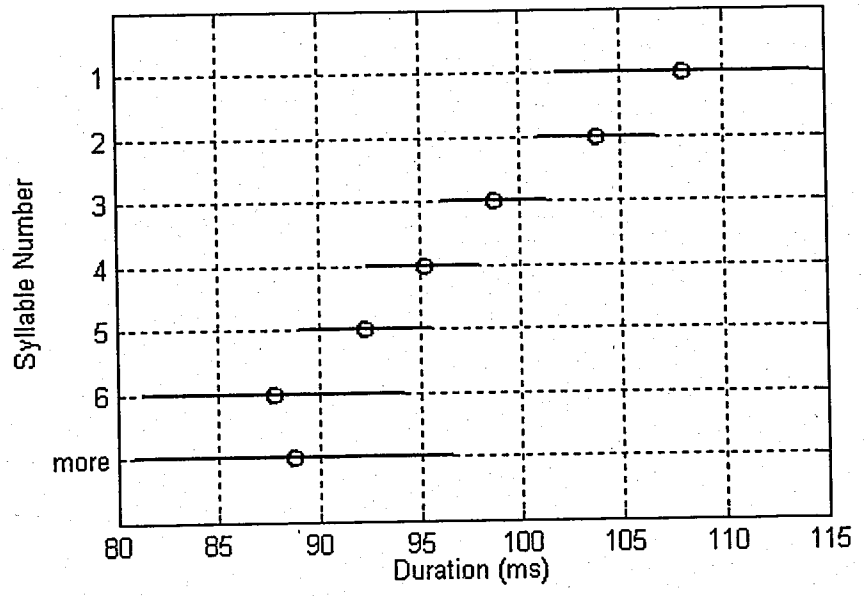

Figure 3.56. 95 per cent confidence intervals of the vowels' means with respect to syllable numbers, sentence environment

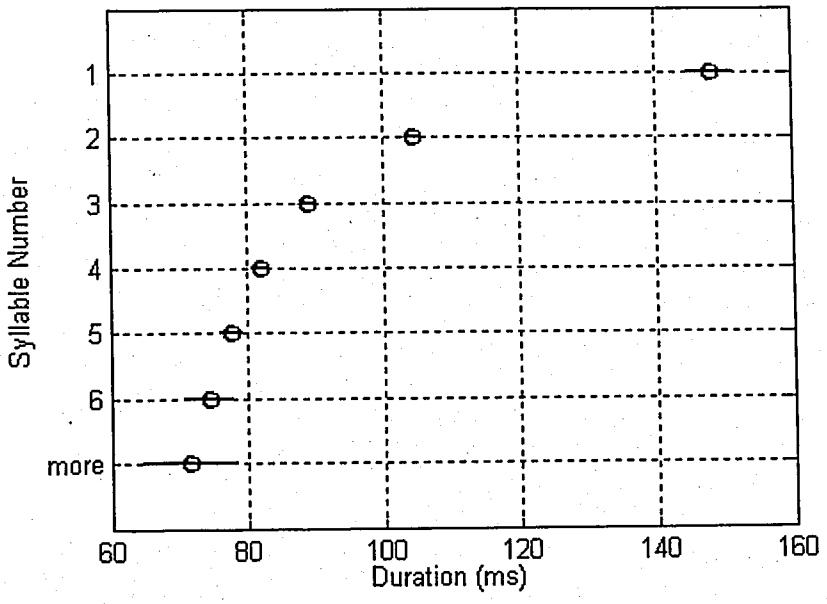

Figure 3.57. 95 per cent confidence intervals of the consonants' means with respect to syllable numbers, 1-word environment

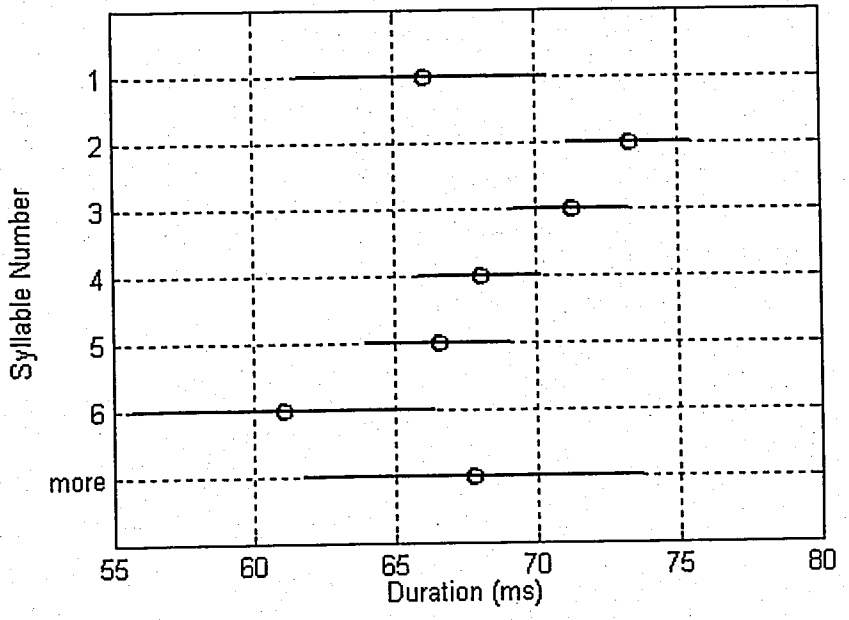

Figure 3.58. 95 per cent confidence intervals of the consonants' means with respect to syllable numbers, sentence environment

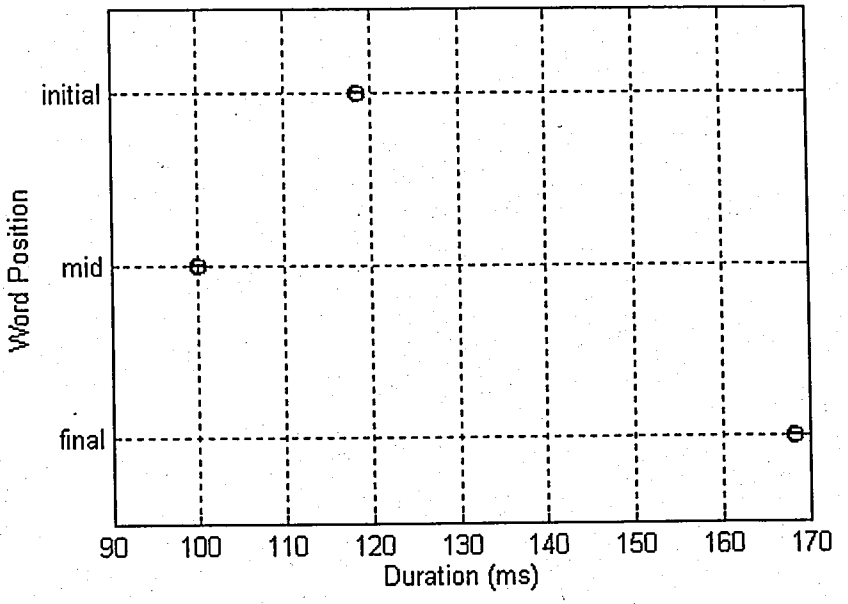

Figure 3.59. 95 per cent confidence intervals of the vowels' means with respect to word positions, 1-word environment

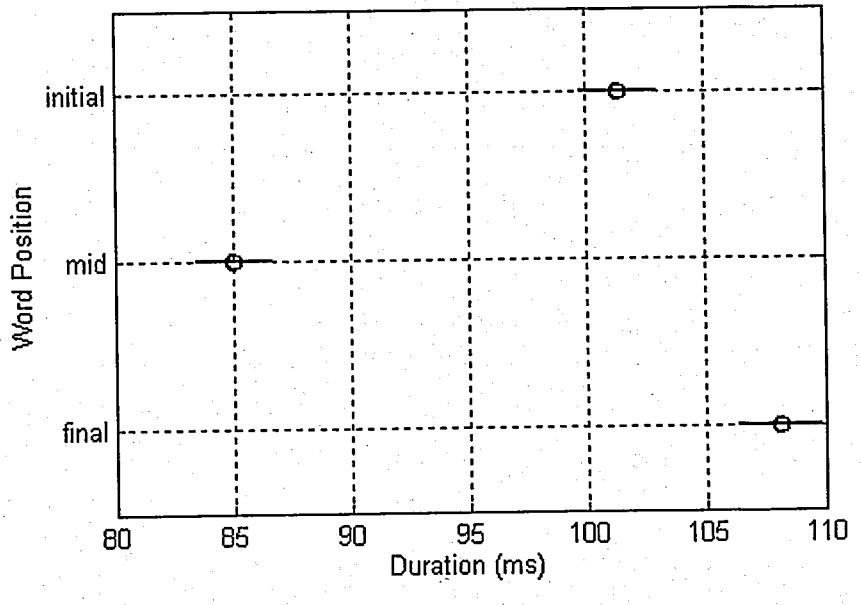

Figure 3.60. 95 per cent confidence intervals of the vowels' means with respect to word positions, sentence environment

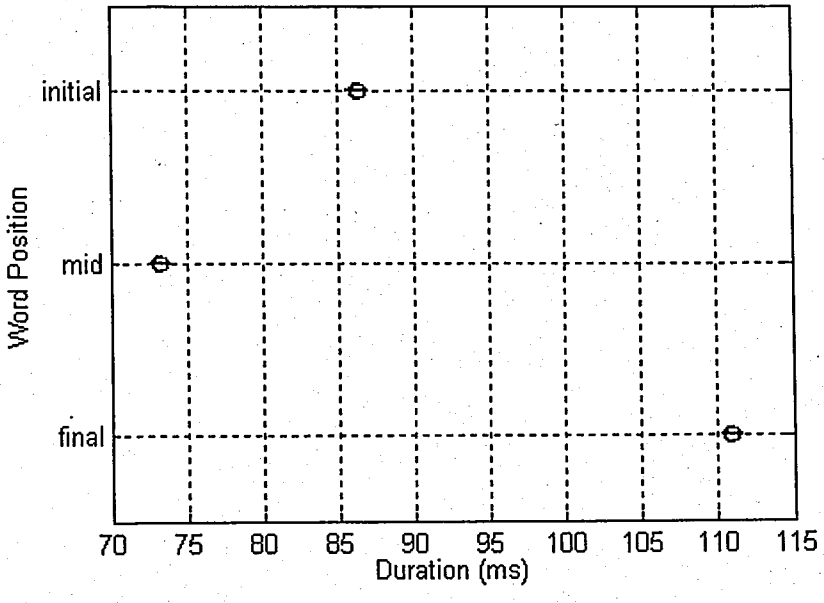

Figure 3.61. 95 per cent confidence intervals of the consonants' means with respect to word positions, 1-word environment

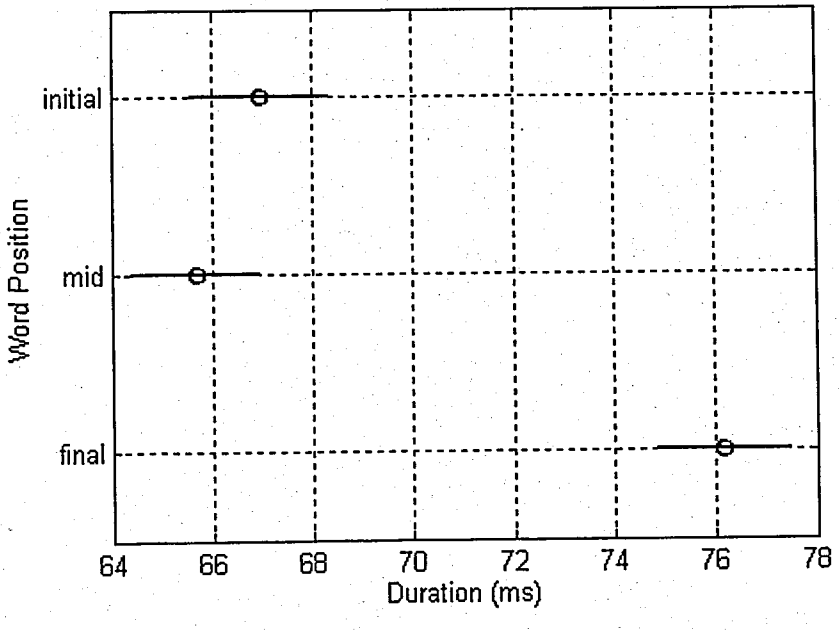

Figure 3.62. 95 per cent confidence intervals of the consonants' means with respect to word positions, sentence environment

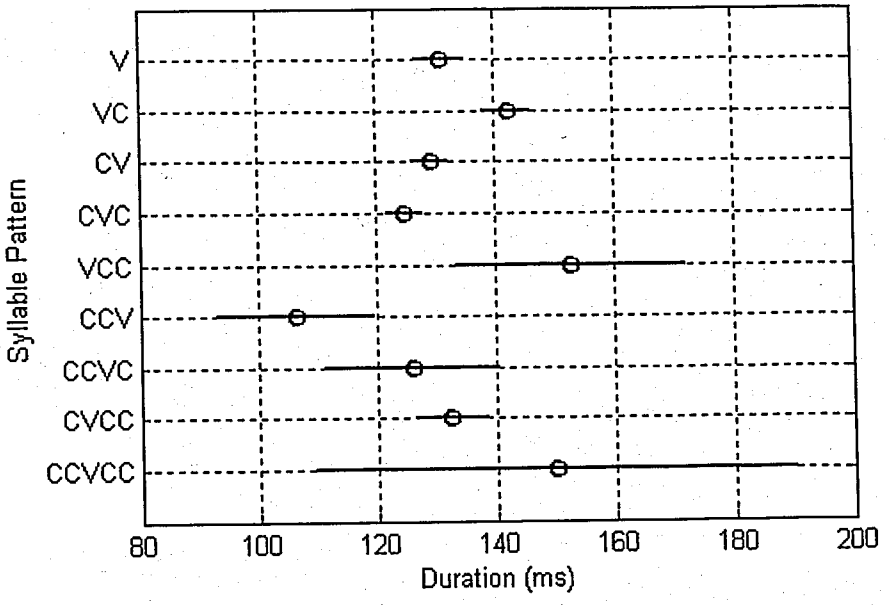

Figure 3.63. 95 per cent confidence intervals of the vowels' means with respect to syllable patterns, 1-word environment

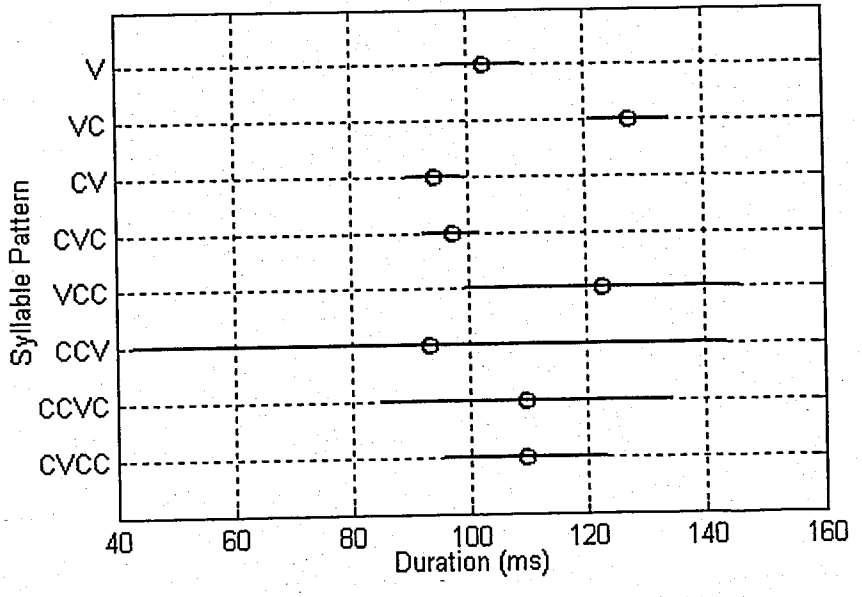

Figure 3.64. 95 per cent confidence intervals of the vowels' means with respect to syllable patterns, sentence environment

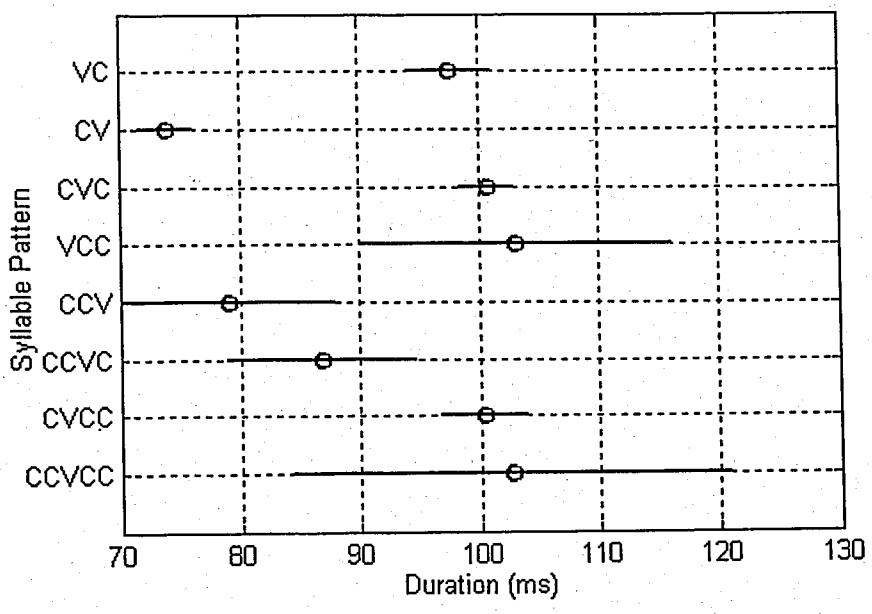

Figure 3.65. 95 per cent confidence intervals of the consonants' means with respect to syllable patterns, 1-word environment

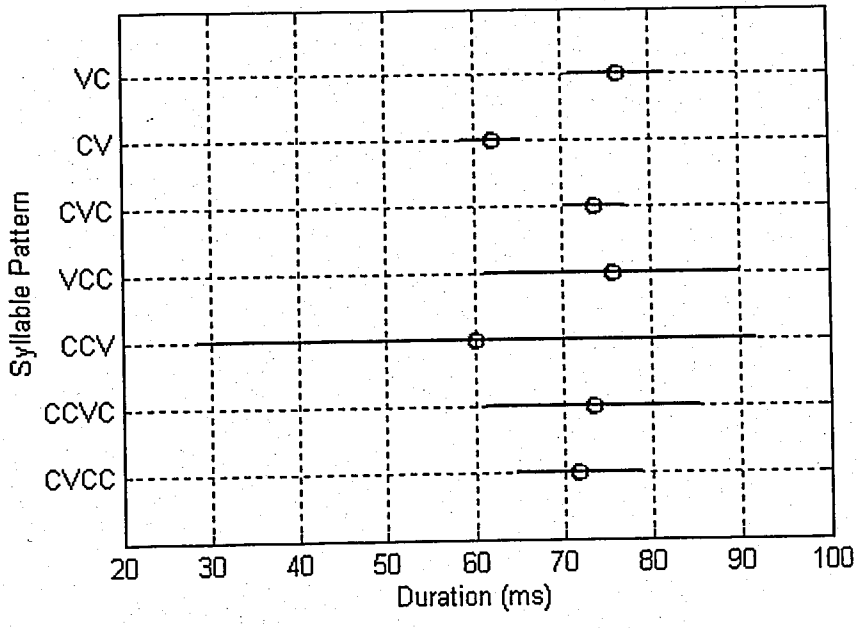

Figure 3.66. 95 per cent confidence intervals of the consonants' means with respect to syllable patterns, sentence environment

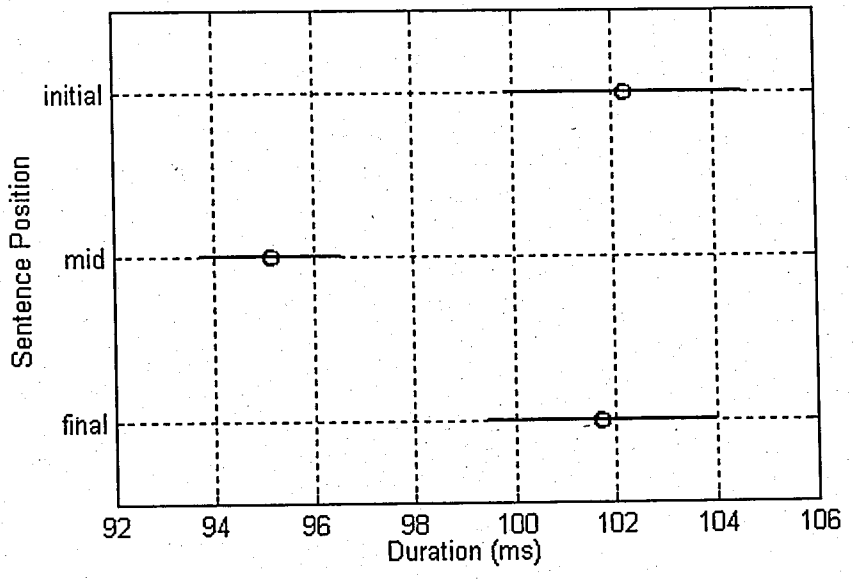

Figure 3.67. 95 per cent confidence intervals of the vowels' means with respect to sentence positions

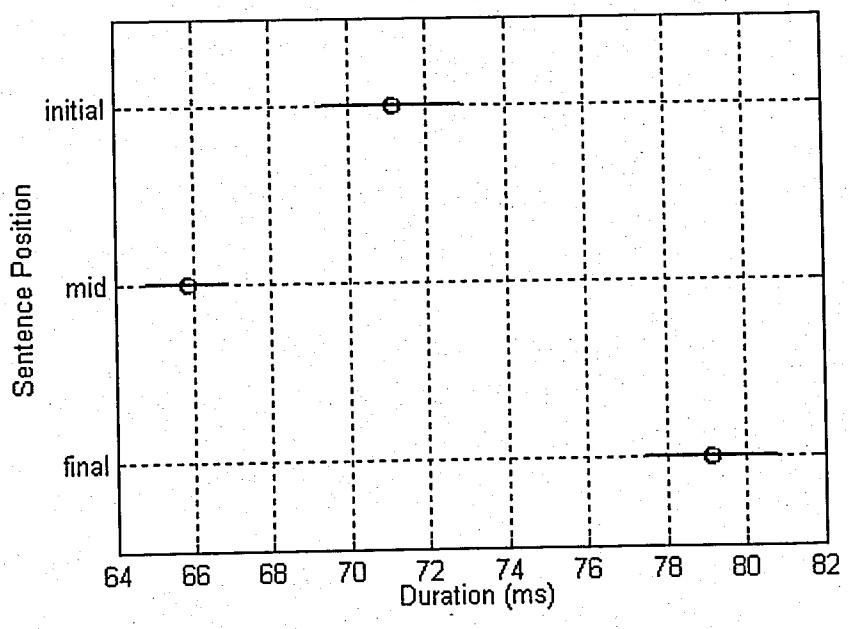

Figure 3.68. 95 per cent confidence intervals of the consonants' means with respect to sentence positions

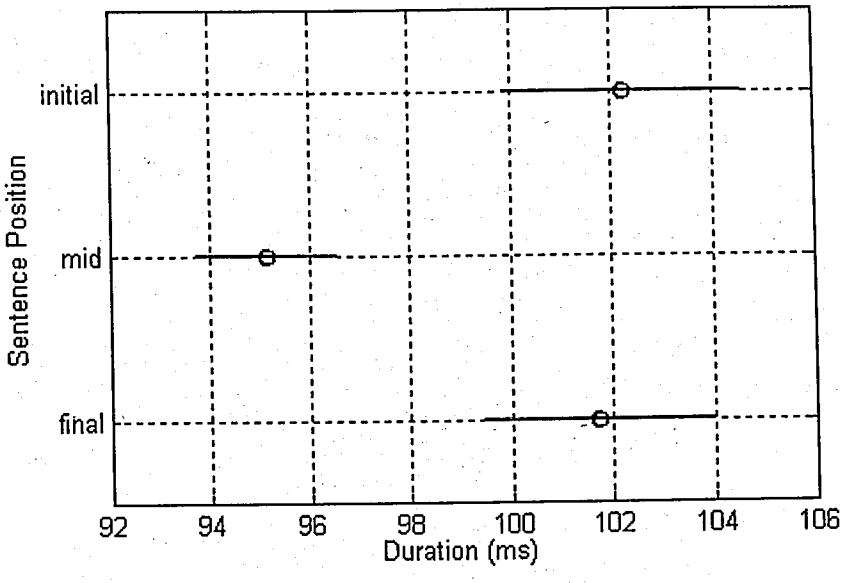

Figure 3.67. 95 per cent confidence intervals of the vowels' means with respect to sentence positions

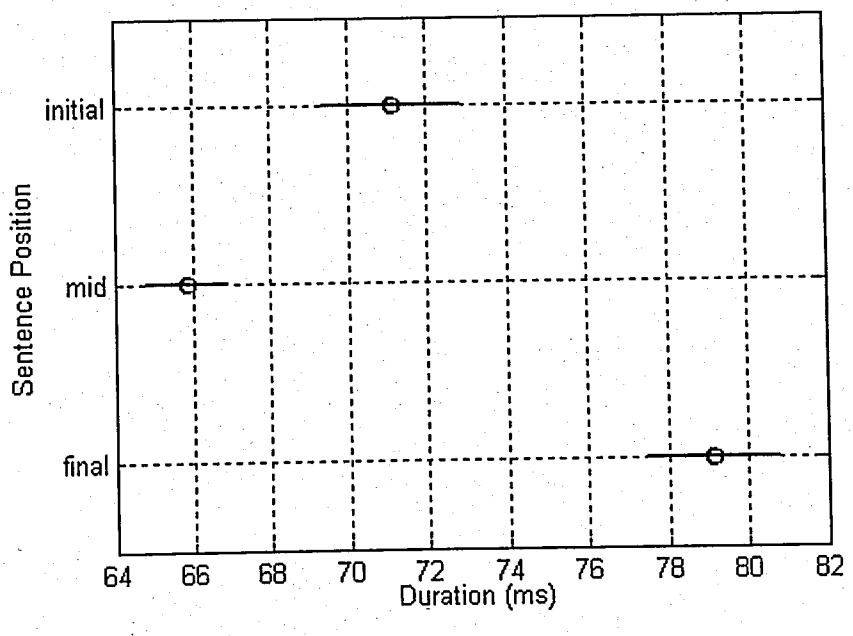

Figure 3.68. 95 per cent confidence intervals of the consonants' means with respect to sentence positions

4. DURATION MODELLING

As stated in the introduction, there is very little known about the underlying process responsible for speech timing. Moreover, speech timing can be predicted from text only up to a point [3]. Hence duration modelling for TTS systems remains a challenging research area.

In this chapter, firstly used notation will be introduced, which is used by Van Santen [3]. Then some models used for duration prediction in the literature are discussed. Finally, derived and implemented models for duration prediction are studied with their duration prediction performances on the spoken corpus.

4.1. Duration Component in TTS

TTS system is composed of several modules and duration component is one of those. Thinking TTS duration component as a black box, inputs to this black box can be described as discrete vectors. To give an example, input vector to the TTS duration component can be a vector like the one below,

$$\vec{f} = < /o/,\ \text{stressed, accented,}\ldots,\ \text{word-final} > \tag{4.1}$$

which is a typical vector representing the properties of phoneme /o/ with properties stressed, accented and in word-final position. Here, elements of such a vector is a level on a factor. A factor can be represented by a set. An example for word position factor is given in the below.

$$\text{Word position} = \{\text{word-initial, word-middle, word-final}\} \tag{4.2}$$

The set of all vectors $\vec{f}$ forms the factorial space

$$S = F_1 \times F_2 \times \ldots \times F_N \tag{4.3}$$

where $F_1, \ldots, F_N$ represents the factors. How much of this factorial space should be covered in the training data base depends on the model used.

$$\mathrm{DUR} : S \rightarrow \mathbb{R} \tag{4.4}$$

Task of duration component is to give a duration value for each input vector. In segmental concatenation based system, it gives segmental durations to inputs like the one in equation 4.1. Stating in another way, duration component maps discrete vectors onto the real numbers, $\mathbb{R}$, as shown in the above equation [3].

4.2. Statistical Models in the Literature

In this section, some commonly used models in the TTS systems will be reviewed. In historical order, they are Lookup table model, Additive and Multiplicative Models, Klatt's Model, Classification and Regression Tree Model and finally Sum-of-Products Models. The material in this section is taken from the study of J. P. H. Van Santen [3, 16] and D. H. Klatt [2].

4.2.1. Lookup Table

In the lookup table model, using the training data base average duration for each feature vector is found to be used in duration prediction. This model is quite simple -in fact the simplest- statistical model. However, difficulty with this model is that training data base should cover the feature space completely in order to find average durations for each feature vector [3].

4.2.2. Additive and Multiplicative Models

Duration prediction with additive model is done according to the formula below [3];

$$\mathrm{DUR}(\vec{f}) = A_1(f_1) + \ldots + A_N(f_N) \tag{4.5}$$

for a feature vector $\vec{f} = \langle f_1, \ldots, f_N \rangle$. Here f_i represents a value on the i-th factor. For example, if the i-th factor is word-position, then f_i can be 'word-initial' and f_i' 'word-final'. The effect of factor i on the duration is given by the parameter $A_i(f_i)$ when it has level f_i. To give an example, if feature vector is $\vec{f} = \langle f_1, f_2, f_3 \rangle$ corresponding to the word position factor levels, $A_1(word-initial), A_2(word-middle), A_3(word-final)$ represent the effects of the word-position factor.

When the effects of one factor are changed by another factor, these two factors are called to *interact in the additive sense.* An example is given by the below duration prediction formula [3];

$$\begin{aligned} \mathrm{DUR}(\vec{f}) \;=\; & [A_1(stressed) + B_1(stressed) \times C_3(/a/)] \\ & -[A_1(unstressed) + B_1(unstressed) \times C_3(/a/)] \end{aligned} \tag{4.6}$$

where A, B and C are three per-factor mapping of stress factor levels to duration values. Here stress factor and vowel identity factor interact *additively.* If the "+" and "-" in the equation 4.6 are replaced by "×" and "÷", interaction becomes multiplicative. The effects are measured as fractions in the multiplicative interactions instead of raw durations.

A big advantage of additive systems is that the needed coverage of feature vectors in the training data base is much lower than others, i.e. lookup table models where complete coverage is required. Additive and multiplicative models are used frequently in the TTS systems because of relatively simple parameter estimation.

4.2.3. Klatt's Model

The Klatt model [17] captures the interaction between postvocalic voicing and phrasal position. This model assumes that;

- Each phonetic segment type has an inherent duration that is specified as one of its distinctive properties

- Each rule tries to effect a percentage increase or decrease in the duration of the segment.
- Segments cannot be compressed shorter than a certain minimum duration.

The model is summarized by the formula:

$$\mathrm{DUR} = \mathrm{MINDUR} + \frac{(\mathrm{INHDUR} - \mathrm{MINDUR}) \times \mathrm{PRCNT}}{100} \tag{4.7}$$

where INHDUR is the inherent duration of a segment, MINDUR is the minimum duration of a segment if stressed, and PRCNT is the percentage shortening determined by applying rules determined from experiments (i.e. phrase-final lengthening, polysyllabic shortening). This equation can be rewritten as:

$$\mathrm{DUR}(V, C, P) = S_{1,1}(V)S_{1,2}(C)S_{1,3}(P) + S_{2,1}(V) \tag{4.8}$$

where V denotes vowel identity factor, C the class of the postvocalic consonant (voiced vs. voiceless), P the phrasal position factor, $S_{2,1}(V)$ is the minimum duration of vowel V, $S_{1,1}(V)$ is the net duration defined as the difference between the inherent duration and the minimum duration and finally $S_{1,2}(C)$ and $S_{1,3}(P)$ are constants tied to the postvocalic consonant and to phrasal position. To clarify, each $S_{i,j}$ is a parameter vector, each parameter corresponding to a level on the j-th factor. The subscript i (having values 1 and 2) refers to the fact that there are two product terms in equation 4.8.

The problem with the Klatt's model is that it is not an accurate description of some of the interactions that have been observed.

4.2.4. Sum-of-Products Models

The sums-of-products model derives from analysis of variance. The ANOVA customarily is used not for modelling purposes but for hypotheses testing, in particular testing for the existence of main effects and (additive) interactions. However, underly-

ing this statistical technique is a model in which some observed variable (Obs) is the sum of a set of interaction terms:

$$Obs\left(\vec{f}\right) = \sum_{I \epsilon K} D_I\left(\vec{f}_{[I]}\right) \tag{4.9}$$

Here, K is some collection of subsets of the set of factors, $\{1, \ldots, N\}$, I is one of these subsets, and $\vec{f}_{[I]}$ is the sub-vector of $\vec{f}$ corresponding to subset I. To illustrate, for $I = \{1, 2, 4\}$, $\vec{f}_{[I]}$ is the vector $< \vec{f}_1, \vec{f}_2, \vec{f}_4 >$. The interaction terms $D_I\left(\vec{f}_{[I]}\right)$ are constrained to have zero sums. The additive model corresponds to the special case where K consists of the singleton sets $\{1\}, \ldots, \{N\}$. The reason that the ANOVA is rarely used for predictive modelling purposes is that the interaction terms are, except for the zero assumption, completely unconstrained, and hence can 'model' any interaction pattern [3].

The sums-of-products model [18], attempts to make interaction terms more meaningful by dropping the zero sum assumption (which is made only for reasons of mathematical convenience), and replacing it with the assumption that each term is a product of single-factor parameters. Thus the interaction terms have the form:

$$D_I\left(\vec{f}_{[I]}\right) = \prod_{i \epsilon I} s_{Ii}\left(\vec{f}_i\right) \tag{4.10}$$

Again when K consists of the singleton sets $\{1\}, \ldots, \{N\}$, the additive model emerges as a special case. Equation 4.10 also generalizes the multiplicative model, which can be obtained by letting $I = \{1, \ldots, N\}$ and $K = \{I\}$.

In the use of this sums-of-products model for duration modelling, input domain of the duration module is described as a factorial space. Duration is modelled in two phases. First, space is divided along some standard distinctions such as vowels vs. consonants, ultimately producing a tree. Afterwards, the cases subsumed under each terminal node of the tree is modelled by a sum-of-products model.

According to sum-of-products models, the duration for a phoneme/context combination described by the feature vector $\vec{f}$ is given by:

$$\mathrm{DUR}\left(\vec{f}\right) = \sum_{i \epsilon K} \prod_{j \epsilon I_i} S_{i,j}\left(f_j\right) \tag{4.11}$$

Here, K is a set of indices, each corresponding to a product term, I_i is the set of indices of factors occurring in the i-th product term. For example, in the following model of Klatt [17]:

$$\mathrm{DUR}(V, C, P) = exp\left(S_{1,2}(C)S_{1,3}(P) + S_{2,3}(P) + S_{3,1}(V)\right) \tag{4.12}$$

there are three product terms with index sets {2,3}, {3}, and {1}. The factors are indexed as 1 (V), 2 (C), and 3 (P). The concept *product* refers to "product of one or more".

As an another example, for the additive model, $K = \{1, \ldots, N\}$ and $I_i = \{i\}$ and for the multiplicative model $K = \{1\}$, and $I_1 = \{1, \ldots, N\}$. We see that other models are merely instances of sum-of-products models.

It is shown that the structure of sum-of-products models (i.e. the index sets I_i) can be inferred from data by subtracting certain marginal means [18]. This is important since the number of distinct sum-of-product models grows extremely rapidly with the number of factors (roughly given by $2^{2^{N-1}-1}$), so that it is in practice not possible to fit each model to a given data set.

4.2.5. Classification and Regression Tree Model (CART)

In CART, in the training phase, a tree is formed by successively dichotomizing the factors (e.g., the stress factor is split into 1-stressed, 2-stressed vs. unstressed) to minimize the variance of the durations under the two newly formed subsets of the speech corpus. For each node of the tree, the observed average duration of the

associated subset of the speech corpus is listed. In other words, CART is a general purpose statistical method that imposes little structure on the data. In a way, it is a condensed lookup table [3, 19].

4.3. Derived and Implemented Models for Duration Modelling

Implemented models in this thesis for duration prediction are;

- Duration Prediction Using Mean Durations of the Phonemes
- Duration Prediction Using Mean Durations of the Triphones
- Tree-Based Modelling of Triphone Durations
- Linear Additive Model

First three models are implemented using C++ programming and the last one in the MATLAB programming environment. For the first three models, parameters for duration prediction models are found in the first analysis of the training data. For the Linear Additive Model, model parameters are found after analyzing and converting data into suitable matrix format which can then be handled easily in MATLAB.

4.3.1. Duration Prediction Using Mean Durations of the Phonemes

This model is the simplest of all. Duration prediction is done using mean durations of the phonemes of Turkish in the training data base. For a given text input to the duration module, for each phoneme in the sentence, mean durations of the phonemes in the training data base is given as duration prediction. Needed feature space coverage in the training data base is just the number of the phonemes in Turkish, twenty-nine.

4.3.2. Duration Prediction Using Mean Durations of the Triphones

In this model, duration prediction is done using the mean durations of triphones in the training database. Needed feature space coverage in the training data base is the number of the most frequent triphones in the Turkish, for desired coverage.

Duration prediction using triphones is more complex than duration prediction using mean durations of phonemes. Main complexity comes from the memory requirement to hold the the most frequent triphones and their mean durations.

4.3.3. Tree-Based Modelling of Triphone Durations

This model is in a way combination of CART model and duration prediction using mean durations of the triphones. The root of the tree is a triphone. From the root, the first level of leaves represent the sentence position (sentence-initial, sentence-middle, sentence-final) of the triphone in the sentence. Second level of the tree represents word number (one to six or more) in the sentence, forth level is the word position (word-initial, word-middle, word-final) of the triphone in the the word. Final level of the tree is formed according to the number of syllables (one to six or more) in the word the triphone is in. The tree is shown in Figure 4.1. For each triphone in the data base,

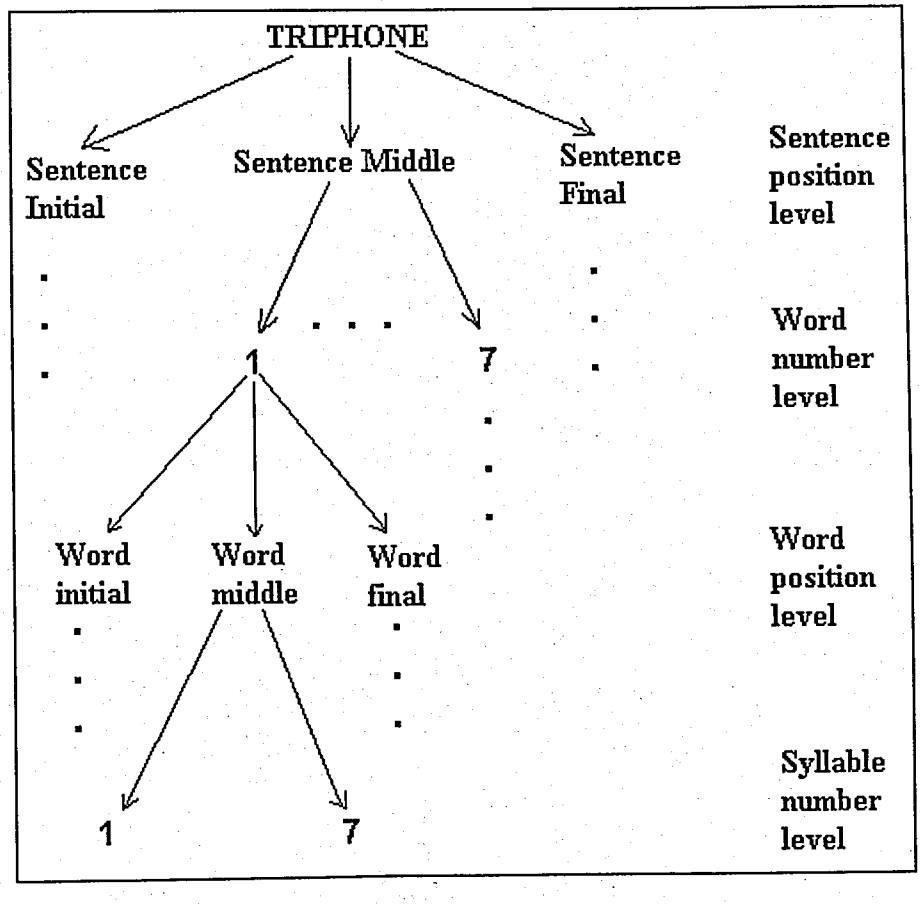

Figure 4.1. Tree used in tree-based modelling of triphones

this tree is formed and for each node of the tree, mean durations are found using the

training data base. For the duration prediction, this nodes of the tree for each triphone is used.

Needed feature space coverage in the training data base enormously big for this tree-based modelling of triphone durations. For a triphone, required occurrence number is the multiplication of the level numbers of the factors sentence position (3), word number (7), word position (3) and syllable number (7), which is 441. For 90 per cent coverage of Turkish (Section 2.1.1) nearly 2000 triphones should be included. Hence 2000×441 makes 882000. This requires that a very big corpus should be used for training.

4.3.4. Linear Additive Model

In this model, every factor level is assumed to effect duration of a segment in an additive manner (Section 4.2.2). The factors considered to have effect are the ones that could be computed from text, as mentioned in Section 3.1.1. These factors are;

- Identity of the current segment (29 values)
- Preceding identity type (3 levels: consonant, vowel, punctuation)
- Following identity type (3 levels: consonant, vowel, punctuation)
- Identity of the preceding segment
    - o If vowel, preceding vowel identity (8 levels)
    - o If consonant, preceding consonant identity (21 levels)
- Identity of the following segment
    - o If vowel, following vowel identity (8 levels)
    - o If consonant, following consonant identity (21 levels)
- Number of syllables in the word (7 levels)
- Number of words in the sentence (7 levels)
- Word position (3 levels: initial, middle, final)
- Sentence position (3 levels: initial, middle, final)
- Syllable pattern (10 levels: V, VC, CV, C, CVC, VCC, CCV, CCVC, CVCC, CCVCC)

for every phoneme in the word and/or sentence. Duration model for a particular occurrence of a phoneme (say /a/) is given by the formula for feature vector $\vec{f_i}$;

$$DUR_{/a/}(\vec{f_i}) = \beta_0 + \beta_1 \times f_{1,i} + \ldots + \beta_{N,i} \times f_{N,i} + \epsilon_i \tag{4.13}$$

where $f_{j,i}$'s are factors' level values, N is the total number of factor levels and ϵ_i is random error component assumed normally distributed with mean zero and variance σ_i^2. β's weight the effect of levels to the phoneme duration.

In the training phase, the database is converted into matrix form. The model given in equation 4.13 can be written in matrix notation as

$$\mathbf{DUR} = F\beta + \epsilon \tag{4.14}$$

where **DUR** is an vector of the duration observations, $\boldsymbol{F}$ is matrix of the levels of the independent variables, $\boldsymbol{\beta}$ is a vector of the regression coefficients and $\boldsymbol{\epsilon}$ is an vector of random errors. The least squares estimator of $\boldsymbol{\beta}$, which minimizes $\epsilon^T\epsilon$, is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T\,\mathbf{DUR} \tag{4.15}$$

These prediction coefficients are obtained for every phoneme in the training phase. Duration prediction is done according to the fitted regression model,

$$\mathbf{DUR} = \boldsymbol{F}\hat{\boldsymbol{\beta}} \tag{4.16}$$

Needed feature space coverage in the training database is the number of independent parameters. Number of independent parameters is simply the sum of level numbers minus one for each factor. In this case, it is (represented by NOIP)

$$\begin{aligned} \text{NOIP} &= \textstyle\sum_{i=1}^{12}(LN_i - 1) \\ &= (28 + 2 + 2 + 7 + 20 + 7 + 20 + 6 + 6 + 2 + 2 + 9) \\ &= 111 \end{aligned} \tag{4.17}$$

where LN_i represents level number for factor i, i from one to twelve represents the factors; identity of the current segment, preceding identity type, following identity type, preceding vowel identity, preceding consonant identity, following vowel identity, following consonant identity, number of syllables in the word, number of words in the sentence, word position, sentence position and syllable pattern respectively. In this computation, some factors can not happen simultaneously (i.e. previous phoneme can not be a vowel and consonant at the same type). Although required feature coverage seems quite low (111 compared to the feature space with size $29 \times 3 \times 3 \times 8 \times 21 \times 8 \times 21 \times 7 \times 7 \times 3 \times 3 \times 10 = 3.24 \times 10^{10}$), more data is needed for good estimation of regression parameters. Interpolation becomes more accurate when more data is used.

4.4. Experiment Setup

The database used for analyzing duration properties of the Turkish phonemes are also used for testing the performances of the implemented duration models. Restating here, database (Sections 2.2.2 and 2.2.1) consists of the 7895 spoken 1-words and 205 sentences (consisting of 1167 words). For each implemented model, model parameters are found separately for 1-word environment and sentence environment. In each environment, 90 per cent of the data are used for training phase and 10 per cent of the data are used for testing phase. Selection of training data and test data are done randomly. The randomly selection of training and test data is done twenty times to get reliable results. For each set of these data, each model is trained and duration prediction performance is measured on the test data. For all the models, the symbol /G/ is modelled as a consonant in the modelling, which has to be modified in the future research.

4.5. Comparison of the Performances of the Models

To evaluate the performances of implemented models, five metrics are used. They are mean error, mean error percentage, standard deviation of error, standard deviation percentage of error and percentage error mean. The term 'error' represents the difference between actual duration and predicted duration of a segment. First metric, mean error is simply the mean of the absolute value of error. Mean error percentage

is defined by the equation below.

$$\text{Mean error percentage} = 100 \times \frac{\text{Mean}(|\text{ True duration - Predicted duration }|)}{\text{Mean duration of segments}} \quad (4.18)$$

The numerator, *True duration - Predicted duration*, is calculated for every segment for which duration is predicted. This metric calculates percentage of mean absolute error with respect to mean duration of segments. Similarly, fourth metric, standard deviation percentage gives standard deviation of error as a percentage of mean duration of segments.

$$\text{Standard deviation percentage} = 100 \times \frac{\text{Standard deviation of error}}{\text{Mean duration of segments}} \quad (4.19)$$

Finally, percentage error mean is mean of the *percentage error*, given in the equation below.

$$\text{Percentage error mean} = 100 \times \text{Mean}\left(\frac{|\text{ True duration - Predicted duration }|}{\text{True duration}}\right) \quad (4.20)$$

In addition to these, R^2 values are computed for linear additive model. R^2 value indicates how well the additive model performs. It is defined by the equation below,

$$R^2 = 1 - \frac{SS_E}{SS_T} = 1 - \frac{\mathbf{DUR}^T\mathbf{DUR} - \hat{\beta}^{\mathbf{T}} \times \mathbf{F}^{\mathbf{T}} \times \mathbf{DUR}}{\hat{\beta}^{\mathbf{T}} \times \mathbf{F}^{\mathbf{T}} \times \mathbf{DUR} - \frac{\left(\sum_{i=1}^{n} \text{DUR}_i\right)^2}{n}} \quad (4.21)$$

where n is number of observations and DUR_i's are individual observed durations. R^2 value can be 1 at maximum. It is a measure of reduction in the variability of observed values ($\mathbf{DUR}$) obtained by regressor variables (variables in vector $\mathbf{F}$) in the model. High R^2 value is desirable but it is possible for poor predictor models to have large R^2 values.

Calculated performance metrics are given in the Tables 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10, 4.11 and 4.12. In these tables, the symbol /G/ is in the consonants class since it has been modelled as a consonant in the modelling. General performance

results are plotted in the Figures 4.2 and 4.3.

It can seen from these that in 1-word environment, the best model is Linear Additive Model with 16.9 mean error percentage and 24.3 σ percentage. The decreasing performance order of other models is Triphone Mean, Triphone Tree and Phoneme Mean models with 18.5, 18.6 and 31.5 mean error percentages respectively. In sentence environment, the best models are Linear Additive and Triphone Mean models with 22.8 and 23.4 mean error percentages. Performance of Triphone Tree model is quite close with 24.6 mean error percentage. Phoneme Mean model has 26.8 mean error percentage.

Although being quite simple and crude, Phoneme Mean Model has 31.5 and 26.8 mean error percentages in 1-word and sentence environments respectively. This model could be used if quite few data is available for training. Also its simplicity is a plus for implementation.

Performances of Triphone Mean and Triphone Tree Models are quite close. It is because of the fact that Triphone Tree Model requires a lot of data for decent training. Morover Triphone Tree Model is more complex. So Triphone Mean Model is preferable to Triphone Tree Model.

Although mean error percentages of Linear Additive Model are low, it has relatively higher standard deviation percentages compared to the other models. It has 24.3 and 30.4 standard deviation percentages compared to 19.1 and 21.2 of Triphone Mean Model, in 1-word and sentence environments respectively.

Overall, the best models are Linear Additive and Triphone Tree Models.

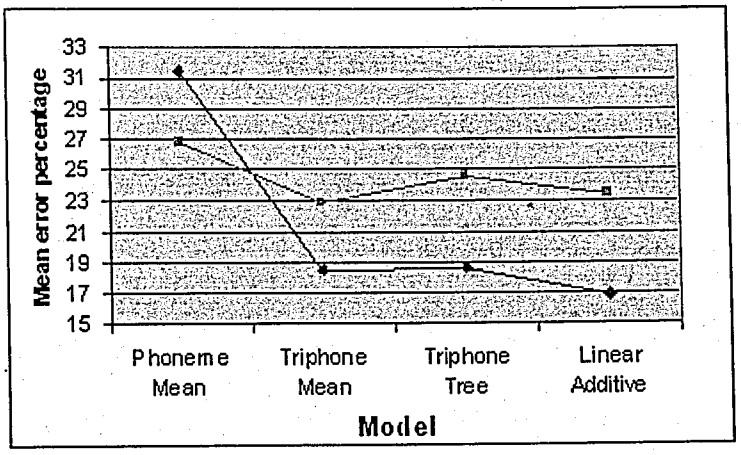

Figure 4.2. Mean error percentages, diamonds and squares represent results of the four models in 1-word and sentence environments, respectively

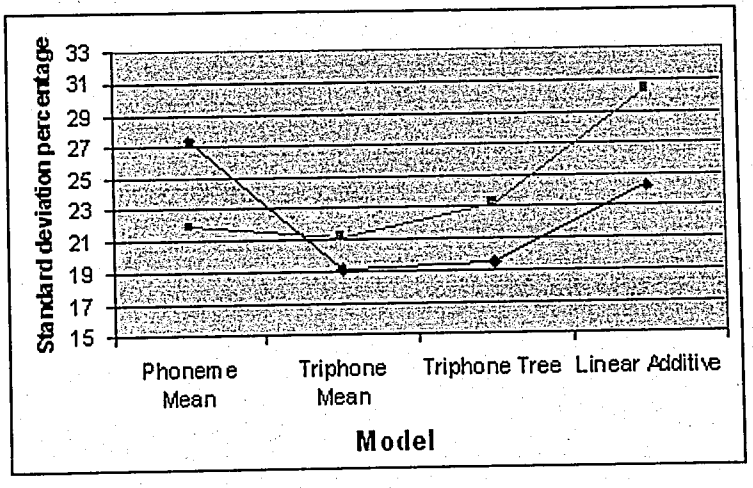

Figure 4.3. Standard deviation percentages, diamonds and squares represent results of the four models in 1-word and sentence environments, respectively

Table 4.1. General error results of the models, 1-word environment

Model	Error type	Error computation for		
		General	Vowels	Consonants
Phoneme Mean	Mean error (ms)	32.6	36.0	30.2
	Mean error percentage	31.5	28.5	33.7
	σ (ms)	28.0	29.7	26.7
	σ percentage	27.4	23.4	30.4
	Percentage error mean	36.3	32.5	39.0
Triphone Mean	Mean error (ms)	19.6	23.9	16.4
	Mean error percentage	18.5	18.7	18.3
	σ (ms)	20.2	21.4	18.7
	σ percentage	19.1	16.7	20.8
	Percentage error mean	21.4	20.9	21.7
Triphone Tree	Mean error (ms)	19.8	23.6	16.9
	Mean error percentage	18.6	18.5	18.8
	σ (ms)	20.7	22.2	19.0
	σ percentage	19.5	17.3	21.1
	Percentage error mean	21.4	20.4	22.2
Additive	Mean error (ms)	17.2	19.2	15.8
	Mean error percentage	16.9	15.0	18.3
	σ (ms)	24.8	27.0	23.1
	σ percentage	24.3	21.1	26.6
	Percentage error mean	19.0	16.4	21.1
	R^2 (over 1)	0.67	0.67	0.67

Table 4.2. General error results of the models, sentence environment

Model	Error type	Error computation for		
		General	Vowels	Consonants
Phoneme Mean	Mean error	20.9	24.3	18.4
	Mean error percentage	26.8	25.3	27.8
	σ	17.1	20.3	14.9
	σ percentage	21.9	21.2	22.3
	Percentage error mean	31.6	29.7	33.0
Triphone Mean	Mean error	18.5	21.8	16.1
	Mean error percentage	22.8	22.3	23.2
	σ	17.1	19.8	15.1
	σ percentage	21.2	20.3	21.9
	Percentage error mean	27.1	25.8	28.0
Triphone Tree	Mean error	19.8	22.7	17.7
	Mean error percentage	24.6	23.3	25.6
	σ	18.6	21.6	16.4
	σ percentage	23.2	22.1	24.0
	Percentage error mean	28.2	26.1	29.8
Additive	Mean error	18.2	20.1	16.7
	Mean error percentage	23.4	21.0	25.2
	σ	23.6	26.4	21.6
	σ percentage	30.4	27.5	32.5
	Percentage error mean	27.1	24.3	29.1
	R^2 (over 1)	0.47	0.45	0.48

Table 4.3. Error results of the phoneme mean and additive models for the vowels, 1-word environment

Phoneme	Phoneme mean model			Additive model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
a	35.5 (25.5)	29.2 (20.9)	27.4	21.6 (15.5)	30.3 (21.7)	16.0
e	33.5 (24.8)	29.2 (21.6)	26.5	16.9 (12.5)	25.1 (18.5)	12.9
I	45.1 (39.2)	33.2 (28.8)	50.4	18.5 (16.1)	26.3 (22.8)	19.2
i	39.6 (34.2)	31.7 (27.3)	41.3	19.7 (17.0)	27.2 (23.5)	19.5
o	27.7 (20.8)	25.5 (19.2)	22.0	17.2 (12.9)	23.6 (17.8)	13.4
O	23.5 (17.6)	22.8 (17.0)	18.6	13.8 (10.4)	17.9 (13.4)	11.1
u	35.4 (31.4)	28.2 (25.0)	37.0	18.9 (16.8)	25.6 (22.8)	19.6
U	33.0 (29.6)	30.6 (27.3)	33.0	18.0 (16.2)	24.3 (21.7)	18.3

Table 4.4. Error results of the phoneme mean and additive models for the vowels, sentence environment

Phoneme	Phoneme mean model			Additive model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
a	24.7 (21.9)	20.6 (18.2)	24.8	20.4 (18.1)	27.3 (24.2)	20.4
e	22.5 (21.2)	17.7 (16.7)	23.5	17.6 (16.7)	23.8 (22.5)	18.2
I	28.0 (34.6)	25.3 (31.2)	46.6	22.3 (27.7)	29.9 (37.1)	36.7
i	24.6 (29.9)	20.5 (24.8)	36.0	20.0 (24.3)	25.8 (31.3)	28.7
o	25.2 (23.0)	17.4 (15.7)	24.9	22.5 (20.5)	27.8 (25.4)	21.9
O	20.6 (18.3)	11.3 (10.0)	19.3	19.6 (17.4)	24.8 (22.0)	18.6
u	23.0 (28.2)	21.8 (26.4)	32.3	20.5 (25.1)	25.6 (31.3)	28.2
U	22.3 (26.0)	22.5 (25.9)	28.5	22.4 (26.3)	27.7 (32.4)	27.9

Table 4.5. R^2 values of additive model for the vowels (over 1)

Vowels	1-word environment	Sentence environment
a	0.57	0.33
e	0.70	0.35
I	0.76	0.55
i	.072	0.45
o	0.68	0.62
O	0.76	0.86
u	0.73	0.60
U	0.74	0.66

Table 4.6. R^2 values of additive model for the consonants (over 1)

Consonants	1-word environment	Sentence environment
b	0.27	0.24
c	0.66	0.73
C	0.84	0.71
d	0.51	0.50
f	0.58	0.81
g	0.40	0.56
G	0.59	0.76
h	0.70	0.49
j	0.91	
k	0.79	0.27
l	0.59	0.26
m	0.73	0.47
n	0.79	0.51
p	0.74	0.84
r	0.79	0.71
s	0.48	0.39
S	0.74	0.58
t	0.75	0.45
v	0.66	0.35
y	0.62	0.53
z	0.66	0.66

Table 4.7. Error results of the triphone mean and triphone tree models for the vowels, 1-word environment

Phoneme	Triphone mean model			Triphone tree model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
a	25.8 (18.5)	22.8 (16.4)	19.7	25.7 (18.5)	24.0 (17.2)	19.3
e	23.3 (17.3)	20.3 (15.1)	18.5	22.0 (16.3)	20.4 (15.2)	17.4
I	22.4 (19.4)	19.9 (17.2)	23.6	22.8 (19.7)	21.7 (18.8)	23.8
i	22.6 (19.5)	20.0 (17.2)	23.0	22.7 (19.6)	20.7 (17.9)	22.1
o	24.9 (18.8)	21.6 (16.4)	20.6	23.8 (18.0)	22.6 (17.1)	19.2
O	19.2 (14.5)	19.1 (14.4)	14.1	19.9 (15.0)	18.3 (13.8)	15.4
u	24.1 (21.3)	21.1 (18.6)	25.3	23.7 (20.9)	21.2 (18.7)	24.1
U	23.3 (21.4)	22.6 (20.7)	24.3	23.6 (21.6)	22.7 (20.8)	25.0

Table 4.8. Error results of the triphone mean and triphone tree models for the vowels, sentence environment

Phoneme	Triphone model			Triphone tree model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
a	22.9 (20.4)	20.2 (18.0)	22.4	23.8 (21.2)	19.5 (17.4)	22.6
e	20.8 (19.9)	18.6 (17.8)	22.1	21.4 (20.4)	19.9 (19.1)	20.9
I	21.2 (26.2)	21.9 (27.0)	33.5	26.9 (33.5)	29.6 (36.8)	39.3
i	21.2 (26.0)	19.0 (23.2)	31.2	17.7 (21.5)	16.2 (19.7)	28.5
o	23.9 (22.0)	18.5 (16.9)	22.7	25.4 (23.3)	17.5 (16.1)	22.7
O	19.6 (17.8)	12.1 (11.0)	19.2	15.5 (14.1)	5.8 (5.3)	14.8
u	20.4 (25.3)	18.4 (22.8)	29.3	25.2 (31.3)	23.3 (28.9)	32.9
U	22.5 (27.0)	20.0 (24.0)	25.4	28.0 (33.5)	32.8 (39.3)	33.5

Table 4.9. Error results of the phoneme mean and additive models for the consonants, 1-word environment

Phoneme	Phoneme mean model			Additive model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
b	21.0 (30.7)	19.5 (28.6)	37.6	19.4 (28.5)	26.6 (39.0)	34.1
c	18.3 (24.5)	20.3 (26.5)	25.7	14.4 (19.2)	21.2 (28.1)	19.5
C	32.5 (27.6)	30.2 (25.6)	29.4	16.3 (13.9)	21.9 (18.7)	15.6
d	18.3 (31.7)	15.1 (26.1)	41.1	12.4 (21.5)	17.5 (30.4)	25.2
f	28.5 (28.6)	27.5 (27.3)	35.9	22.4 (22.5)	32.5 (32.6)	27.4
g	18.3 (28.6)	17.1 (26.7)	32.0	15.4 (24.1)	21.3 (33.4)	26.2
G	16.9 (27.7)	15.4 (25.2)	33.6	13.4 (22.0)	18.0 (29.4)	26.3
h	28.2 (40.5)	32.8 (46.5)	51.0	19.9 (28.7)	26.7 (38.4)	36.5
j	40.7 (36.4)	27.0 (23.3)	39.1	28.2 (25.9)	38.8 (34.8)	27.0
k	52.3 (42.5)	35.9 (29.1)	52.7	19.3 (15.6)	29.3 (23.8)	22.7
l	18.8 (27.4)	19.4 (28.4)	29.6	13.0 (19.0)	18.2 (26.6)	20.9
m	24.7 (29.5)	26.9 (32.1)	30.3	13.3 (15.9)	18.9 (22.5)	17.4
n	42.2 (42.2)	29.9 (29.9)	53.6	15.4 (15.4)	24.0 (24.0)	17.2
p	39.5 (37.0)	44.3 (41.2)	43.2	20.2 (19.0)	33.1 (30.9)	22.4
r	35.4 (47.2)	31.6 (42.2)	53.6	14.9 (19.9)	22.2 (29.6)	21.9
s	23.7 (17.7)	22.2 (16.6)	19.9	18.0 (13.4)	24.9 (18.6)	15.4
S	29.9 (20.7)	27.2 (18.8)	20.7	15.5 (10.7)	21.8 (15.1)	10.9
t	34.9 (33.4)	34.4 (32.9)	37.2	17.4 (16.6)	25.7 (24.6)	20.1
v	18.9 (28.6)	15.7 (23.7)	30.2	15.0 (22.7)	20.6 (31.3)	24.3
y	17.8 (27.6)	21.5 (33.2)	28.6	12.8 (19.9)	18.7 (28.9)	21.2
z	41.8 (37.8)	36.0 (32.4)	42.1	21.0 (19.0)	36.1 (32.5)	20.3

Table 4.10. Error results of the phoneme mean and additive models for the consonants, sentence environment

Phoneme	Phoneme mean model			Additive model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
b	15.6 (29.7)	12.3 (23.5)	39.3	17.7 (33.7)	22.4 (42.6)	44.0
c	14.9 (22.9)	11.8 (18.0)	27.8	21.1 (32.5)	24.9 (38.1)	35.4
C	20.0 (18.3)	15.1 (13.7)	18.8	25.9 (23.7)	34.0 (31.1)	24.2
d	15.4 (32.6)	10.5 (22.2)	43.4	12.6 (26.5)	15.8 (33.3)	31.7
f	19.5 (23.1)	9.6 (11.5)	22.9	29.8 (37.0)	38.3 (47.7)	41.3
g	12.6 (26.2)	11.2 (22.8)	31.6	14.0 (28.7)	19.0 (38.9)	32.2
G	11.8 (30.0)	7.7 (19.2)	34.6	11.3 (28.2)	14.0 (34.9)	30.8
h	16.9 (33.0)	14.0 (27.0)	47.3	29.0 (56.6)	38.6 (75.0)	76.7
j						
k	17.7 (21.5)	15.3 (18.5)	24.2	17.9 (21.7)	24.0 (29.1)	24.2
l	13.5 (24.2)	11.3 (20.3)	28.1	13.5 (24.1)	17.3 (30.9)	27.4
m	15.7 (22.6)	12.3 (17.6)	26.8	14.8 (21.2)	18.0 (25.8)	24.8
n	25.4 (34.7)	19.3 (26.3)	44.5	18.0 (24.6)	23.7 (32.4)	29.5
p	15.7 (20.5)	11.8 (15.1)	22.0	26.5 (34.5)	32.0 (41.4)	35.2
r	25.4 (42.6)	22.0 (36.8)	48.6	15.5 (26.1)	20.0 (33.6)	30.8
s	17.9 (16.3)	14.1 (12.8)	18.3	19.9 (17.3)	24.5 (22.2)	19.0
S	20.6 (16.7)	17.8 (14.3)	16.6	21.0 (17.1)	28.4 (23.2)	17.8
t	18.5 (23.5)	12.6 (16.0)	26.3	16.3 (20.8)	21.4 (27.2)	22.7
v	11.6 (21.9)	9.7 (18.1)	22.9	13.5 (25.4)	17.0 (31.9)	27.4
y	11.7 (26.8)	10.4 (23.3)	31.9	12.7 (29.1)	17.8 (40.6)	32.6
z	27.2 (32.0)	27.2 (32.0)	31.5	26.2 (31.3)	32.2 (38.5)	34.4

Table 4.11. Error results of the triphone mean and triphone tree models for the consonants, 1-word environment

Phoneme	Triphone model			Triphone tree model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
b	19.7 (29.0)	19.1 (28.0)	35.5	19.1 (28.0)	18.9 (27.7)	35.0
c	13.7 (18.4)	13.5 (18.2)	18.8	17.1 (23.0)	19.8 (26.8)	21.1
C	19.2 (16.3)	20.7 (17.6)	17.8	19.3 (16.3)	19.9 (16.8)	17.8
d	13.1 (22.6)	13.1 (22.6)	26.1	13.9 (24.0)	12.9 (22.3)	27.9
f	24.1 (24.3)	24.4 (24.6)	30.0	24.8 (25.0)	24.5 (24.8)	32.4
g	16.6 (25.9)	18.4 (28.7)	27.5	16.6 (25.9)	14.4 (22.4)	29.3
G	16.6 (26.6)	16.7 (26.7)	33.7	16.5 (26.4)	16.6 (26.7)	30.8
h	17.9 (26.7)	16.5 (24.6)	32.8	19.7 (29.5)	19.3 (28.8)	34.2
j	25.7 (21.3)	21.7 (18.0)	21.9	24.4 (20.0)	21.1 (17.3)	24.3
k	19.0 (15.4)	21.4 (17.3)	19.7	20.1 (16.3)	23.0 (18.6)	20.5
l	13.8 (20.1)	13.9 (20.2)	22.0	14.2 (20.6)	13.6 (19.7)	23.1
m	13.4 (16.0)	13.6 (16.3)	17.4	13.5 (16.1)	12.3 (14.7)	17.5
n	15.8 (15.7)	17.3 (17.2)	17.8	16.5 (16.4)	19.9 (19.8)	18.1
p	21.7 (20.7)	27.3 (26.0)	23.0	23.3 (22.3)	30.1 (28.8)	24.3
r	15.1 (20.2)	16.7 (22.3)	22.8	15.5 (20.8)	17.1 (22.9)	22.6
s	19.8 (14.8)	19.5 (14.5)	16.9	19.7 (14.7)	17.5 (13.1)	17.3
S	18.4 (12.8)	22.4 (15.6)	12.9	17.2 (12.0)	14.9 (10.3)	12.2
t	18.4 (17.8)	18.6 (17.9)	21.2	19.2 (18.5)	20.5 (19.8)	22.3
v	17.3 (26.3)	21.5 (32.7)	26.2	16.1 (24.4)	17.6 (26.6)	25.0
y	13.3 (20.7)	13.1 (20.5)	22.3	13.3 (20.8)	14.3 (22.3)	22.4
z	21.0 (19.2)	27.7 (25.4)	21.3	21.8 (20.0)	25.5 (23.3)	22.2

Table 4.12. Error results of the triphone mean and triphone tree models for the consonants, sentence environment

Phoneme	Triphone model			Triphone tree model		
	Mean error	σ of error	Mean percentage	Mean error	σ of error	Mean percentage
b	14.8 (27.1)	11.5 (21.1)	30.8	22.6 (41.0)	12.9 (23.4)	53.4
c	12.8 (19.0)	10.6 (15.8)	18.2	21.4 (32.3)	21.6 (32.5)	27.4
C	21.5 (20.7)	20.3 (19.5)	22.2	24.3 (23.2)	27.0 (25.8)	32.2
d	11.8 (25.1)	10.0 (21.1)	32.4	12.8 (27.0)	10.1 (21.2)	31.7
f	14.7 (20.7)	7.6 (10.7)	24.9	11.4 (16.2)	7.2 (10.2)	18.5
g	12.7 (26.5)	9.5 (19.8)	31.2	14.1 (29.7)	13.9 (29.3)	28.5
G	11.5 (27.5)	5.8 (13.9)	32.0	12.1 (29.2)	6.9 (16.7)	25.0
h	19.4 (37.3)	15.3 (29.4)	62.6	12.4 (23.6)	12.3 (23.4)	24.9
j						
k	18.0 (21.6)	15.7 (18.9)	25.5	20.2 (24.4)	20.3 (24.5)	26.2
l	12.4 (22.1)	10.6 (18.8)	26.1	14.5 (26.0)	12.4 (22.1)	29.5
m	16.2 (22.6)	12.1 (17.0)	24.0	18.7 (26.1)	13.1 (18.3)	29.8
n	17.7 (24.6)	15.2 (21.1)	30.6	19.9 (27.5)	16.6 (23.0)	32.4
p	14.6 (19.1)	11.7 (15.3)	20.8	21.0 (27.9)	20.1 (26.8)	29.5
r	16.9 (28.2)	17.3 (28.8)	31.4	16.9 (28.1)	13.7 (22.7)	35.5
s	20.3 (18.2)	17.5 (15.7)	19.7	19.5 (17.5)	13.2 (11.8)	19.0
S	19.1 (15.5)	18.5 (15.0)	15.6	21.7 (17.6)	24.3 (19.7)	20.0
t	16.3 (20.7)	15.6 (19.8)	23.6	19.6 (25.0)	21.1 (26.9)	25.7
v	12.8 (24.5)	9.8 (18.8)	28.2	15.6 (30.2)	15.1 (29.3)	29.4
y	13.0 (28.9)	13.8 (30.5)	29.0	13.3 (29.1)	14.0 (30.7)	32.4
z	26.4 (33.0)	21.8 (27.4)	36.7	24.0 (29.9)	19.5 (24.2)	30.6

5. CONCLUSION

In this thesis, as far as we know, a *first* attempt has been done to analyze and model durations of Turkish phonemes. To do this, a software system has been developed in C++ programming language. Some parts of the system are in MATLAB.

The analysis and modelling have been done using a corpus of spoken 7898 1-words and 205 (1167 words) sentences. It should be emphasized that the analysis and results presented in the duration analysis (Chapter 3) is based on the labelling convention developed by a non-linguist. So some deviation from the results in the duration analysis is predicted with a more linguistic approach to labelling. In these analysis, durational properties of Turkish phonemes and the effect of contextual factors on the phonemes are investigated.

For duration modelling, four models are implemented. They are Phoneme Mean Model, Triphone Mean Model, Triphone Tree Model and Linear Additive Model. Linear Additive and Triphone Mean Models are found to be better than others.

5.1. Further Research

It is very crucial that future research on duration analysis and modelling for Turkish needs a 'labelling convention'. A study should be done to develop a consistent labelling convention which can be applied easily for different data. Also use of a larger (and probably more accurate) symbol set for the sounds of Turkish could be useful.

Larger text corpus should be used for in future research that also includes every possible combination of the factors of interest (Section 2.1). This corpus should be spoken by as much persons as possible (with different dialects, ages etc.).

More models could be developed to model durations of the phonemes. For example, neural networks could be used.

APPENDIX A: USED CORPUS

In this section, spoken words and sentences are written according to the symbol convention introduced in Section 1.1.

A.1. Sentences

A.1.1. Sentences Containing Two Consecutive Vowels

1. ayla uncuoGlu ve oya baSer
2. blok apartmanlarda her daire kendine Ozel balkon yapmaya kalkarsa damlarda yer bulmak zorlaSIr
3. Cok fazla gUvenilir fakat daha pahalI mikroiSlemcilerde yUrUtUlUr
4. fedailerine kIzmIStIk
5. konaklardaki sadece harem dairesi halkevlerinden bUyUktU
6. tUrkuaz mavisi beyaz kadar yeSildir

A.1.2. Other Sentences

1. aCIkca sOylemekte tereddUt ediyordum
2. ahmet vardar ve uGur mumcu devamlI televizyondalar
3. akustik dalgalar fiziGin temel konularIndan biridir
4. akustik debimetre uygulamasInIn piyasaya ilk giren modelinde kullanIlan bir prensiptir
5. alparslan tUrkeS
6. ancak yansImalarIn Siddeti yansItIcI yUzeyin Islak veya kuru olmasIna gOre deGiSir
7. aristoteles bu konuda gOzlemlerini anlatIyordu
8. artIk bahar geldi
9. asuman akbaS
10. avrupa bize dUSman kesildi diyebilirsiniz
11. aynI ivmelerle yavaSlayarak son noktaya ulaSIrlar
12. aySe arslan
13. aySin ertUzUn ve ahmet denker oturuma katIldI

14. balkanlar ve ortadoGudaki topluluklar imparatorluktan kopmuSlardIr
15. bankamIza hoSgeldiniz
16. baSka bir iSlem yapmak istiyorsanIz hatta kalIn lUtfen
17. baSlangICta osmanlI sultanlarInIn tanIdIGI ayrIcalIklar sanayiyi tUmUyle COkertmiSti
18. batIdaki bu geliSme osmanlI imparatorluGunun gerilemesinde etken olmuStur
19. berkay tamer
20. berna laCin ve duygu asena
21. bizim yetiStireceGimiz bebekler tedavi edecek
22. bizler evlatlarImIza ninni sOylerken oGlum paSa olsun gibi sOzleri sOylemeyelim
23. bodruma derslik aCIldI
24. boGaziCinde bir vapur gezisi
25. bohCacI kadInlar arabuluculuk gOrevini UstlenmiSlerdi
26. borcunuz beS milyon lira
27. bu dalgalar uzayda ISIk ISInlarIna benzer Sekilde yayIlIrlar ve CeSitli cisimlerden yansIrlar ve etrafa saCIlIrlar
28. bu eklemin dinamiGi dikkate alInmamIStIr
29. bu eser milattan Once beSinci yUzyIlda boyanmIS
30. bu halde anten en kuvvetli yansImayI alacak Sekilde ayarlanIr
31. bu hiyerarSinin yUrUtUlmesi her kullanIcInIn ayrI ayrI deGerlendireceGi faktOrlere baGlIdIr
32. bu iki iSaretin yol farkIna baGlI olarak deGiSen uzaklIklarda gOlgeler meydana gelir
33. bu mUnasebetsiz bir SakaydI
34. bu tebliGde minimum dUzeyde tutulmuStur
35. bu yapIlanma seyahat biCimine gOre deGiSiyordu
36. bu gOlgede yaSamak heponlarIn lehine
37. buharlaStIrIcI boru aksamI emniyeti saGlar
38. bUlent ecevit
39. buna gOlge veya hayalet gOrUntU adI verilir
40. buradaki kavramlar bir pratik Ornek ile gOsterilmiStir
41. butUr kristallerden dOrdUncU harmonikte calISacak Sekilde kesilmiS olanlar tercih sebebidir
42. bUyUk ihtimalle bunu gOz OnUne almamIStIk
43. CabuklaStIrabilirsek iyi olacak

44. CeSitli rezaletlere yol aCtIklarI sOylenirdi
45. CIkIStaki bir basInC dUSUrUcU vana otomatik operasyon saGlar
46. CiCekler aCtI kIrlar Senlendi
47. Cocuklar hocalarIn sopalarI altInda esner ve titrerdi
48. CoGu iSlevler daGIlIr
49. daG baSInI duman almIS
50. deGiSik merhale safhalarI incelendi
51. denebilir ki tUrk CaGdaSlaSma eylemi otorite boSluGu yaratmIStIr
52. deniz suyuyla soGutma sistemlerinde klor solUsyonu eklenir
53. derbeder bir kaldIrIm gOrUntUsU
54. derginin idari bOlUmU ilkin sultanahmete taSIndI
55. devletin olaGan alISIlmIS gelirleri bu daralmadan dolayI azalmIStIr
56. doGal gaz Once karbon ve metanIn ayrIStIrIlmasI ile kararlI hale getirilir
57. dOnUSUmden aCIk seCik olarak gOrUlmektedir ki eksen yUzeyseldir
58. dUGUnde Celenk yollamak yerine para yatIr
59. dUzlemde karakteristik denklemin kutuplarI deGiSik yerlerdedir
60. eGer alIcI antene sadece bir dalga ulaSIrsa tek ve net bir gOrUntU elde edilir
61. eklemlere uygulanacak kuvvet ve moment elde edilir
62. ekonominin gUCsUzleSmesi ticaretin yabancIlarIn eline geCmesi Onemlidir
63. ekrem pakdemirli
64. ekspres servis Su anda mUmkUn mU
65. elkoyucu batIlI devletler bu oluSumu hazIrlamIStIr
66. elli gUnlUk dOnemde amasya genelgesi yayInlanmIStI
67. emin adImlarla ilerliyordu
68. emrah gUrSahbaz
69. erdal demirtaS
70. erkeGin dunyasI kamusal kadInInki ise Ozel ve mahremdi
71. esas iSaretle birlikte yakIndaki binalardan ve tepelerden yansIyan iSaretler de gelir
72. estetik aCIdan Cok muhteSem bir gOrUntUydU
73. evlenme tOrenlerine dayalI geleneksel kUltUrUn Ozellikleri tam anlamIyla yansIyordu
74. fabrikalarda ince sepet yumurta istenir
75. fantazi kurmak egzersiz yapanlara
76. frekans bOlgesi davranISInI izlemekte yarar vardIr
77. giderleri karSIlayamamasI dIS borClanmayI zorunlu kIlmIStIr

78. gidilecek ve gezilecek yerler devlet tarafIndan kIsItlanmIStI
79. gOkSin Ilgaz
80. gOzler kanlanIr mezara varIlIrken
81. gUmUSdere durmaz akar
82. gUneS ufuktan Simdi doGar
83. gUray ateS
84. haftada iki kere pirzola ve salata yapmasInIn yanInda iCki iCti
85. hakikatI manalI bulmak zorundaydIm
86. hakkImIzI eninde sonunda alIrIz
87. haluk bingOl
88. hava Cok soGuk ve kasvetli
89. hayvana binmeleri yasaklandIGIndan Oteden beri arabaya binerlerdi
90. heder olmak gUzel midir acaba
91. hepimiz bunun bilincindeyiz
92. herhangi bir denetim iSlevinin taramasI temelde dUSUk seviyede olmalIdIr
93. hesabInIzda sekiz yUz lira var
94. hiC istemediGim bir kavgaya Sahit oldum
95. hiCbir koSula vergiye gUmrUk uygulamasIna baGlI deGildi
96. imkansIz gibi bir olay
97. insan haklarI bildirgesinin otuz numaralI maddesi SOyle der
98. iptal etmek iCin tekrar giriniz
99. iSaretler verici antenden alIcI antene elektromagnetik dalgalar halinde gelir
100. ismet eroGlu
101. ismet inOnU ve mustafa kemal atatUrk
102. istediGiniz hesaba Su anda ulaSIlamIyor
103. istiklal mahkemeleri Cok gOrev yaptI
104. jokeri tutturmak iCin aSIk olmak lazIm
105. kabiliyetli kiSinin harcI ancak bu
106. kafalarImIz Cok karIStI
107. kahverengi en hoSuma giden renk
108. kahyanIn tavIrlarI etkileyiciymiS
109. kalite hava durumuna ve mevsimlere gOre deGiSecektir
110. kamuran akkor
111. kapalICarSIda dolaSmaktan da men edilmiSlerdi

112. karISmanIn en aza indirilmesi iCin anten telsizlerden mUmkUn olduGu kadar uzaGa konmalIdIr
113. karSIt cinsler arasIndaki sosyal iliSki kontrol altInda tutulurdu
114. kin ve intikam duygularI doruGa CIkmIStI
115. klor beslemesi durduGunda alarm vermelidir
116. kontrol paneli izole edilmiS UCUncU bir odaya monte edilmelidir
117. kutlay karaman
118. levent arslan
119. leyla ile mecnun tarihi bir aSk OykUsUymUS
120. lUtfen dOrt nolu tuSa basInIz
121. maCtan ilginC bir enstantane tUrUyor
122. mahCup etmeseydiler pek Onemi olmazdI
123. mani olmak mUmkUn mU ki
124. meclis araStIrmasI olumlu sonuC verdi mi
125. memleketin meseleleri fazla
126. meral mansuroGlu ve barIS manCo
127. merkezsel gUC Cevre Uzerindeki denetimini yitirmiStir
128. mesut yIlmaz
129. mUfettiSlerin gOrev alanI dISIndaki yerlerle haberleSme kopmuStu
130. muhabere teknikleri konulu konferansa Uye misiniz
131. muhafaza etmek isterim
132. muharrem karakaS
133. mUmkUn olamayan doGrusallIk deGerlerine eriSmek artIk kolaydIr
134. mUmtaz soysal
135. mUslUman araplar baGImsIzlIGa kavuSmuSlardIr
136. mutlaka neden ister misiniz
137. naCizane kulunuza bir SarkI baGIslayIn
138. necmettin erbakan
139. nefretin ve Siddetin hIzla yayIldIGI bir zamandI
140. o dOnemde yaSanan balkan savaSI yenilgisi kin ve intikam duygularInIn artmasIna yol aCmIStI
141. okumak herkesCe desteklenmeli
142. OlCmenin avantajI gecikme sUresinin az olmasInI saGlamasIdIr
143. Olmesine ramak kalmIStI sanki

144. on yIlda on beS milyon genC
145. optimizasyon iCin bUyUk hesaplamalar ve karmaSIk programlarIn yUrUtUlmesinde bir bilgisayara gerek vardIr
146. ordu birliklerinin dISInda herkes kazIm karabekirin komutlarInI yerine getireceklerdi
147. OrneklenmiS iSaretten bu iSaretin UretildiGi sUrekli iSaretin nasIl elde edileceGini gOrmek lazIm
148. orta seviyeler iCin bir kural yoktur
149. osmanlI toplumunda kadIn ve erkeGe iki ayrI dunya sunulmuStu
150. oturduklarI yerler birbirinden ayrIlmIStI
151. Oyle ki senelerce koskoca osmanlI hUkUmeti bunlarla baSa CIkamaz oldu
152. Ozellikle Cukur semtler iCin bu durum sOzkonusudur
153. parmaksIz olmak Cirkinlik demek anlamIna gelmiyor
154. sarI vadinin tohumlarI Cok mor
155. sayaClama sistemi iCin elemanlarIn sIvI fazda tutulmalarI Onem kazanmaktadIr
156. sebepsiz yere sUrgUne gOnderildi
157. Sekilde verilen algoritmanIn sonuClarI yOrUnge hIzIna baGlIdIr
158. seyirciler naklen yayIn seyretmekteler
159. sIcaklIk kompanzasyonu iCin standart bir bilgisayar kullanIlabilir
160. simgesel bazI konuSma vasItalarI doGurmuStu
161. sipahiler avrupanIn derebeylerine benzeyen gUC odaklarI haline gelmiSlerdir
162. sivil konutlarIn baSlIca OzelliGi pencere kafesleriydi
163. sUleyman demirel
164. takip etmek ayIp deGil mi
165. tarkan demirbaS
166. tazminat Odemesi bu ayIn sonuna kadar gerCekleSmeli
167. tek sebep buysa bence haksIzsIn
168. tekke Uzerinde yoGun bir baskI vardI
169. teklif edilmesi doGru deGil
170. temel dUzenleyici denetim yUksek tarama hIzlI yedekli denetleClerde yapIlmalIdIr
171. temsilciler kurulunun yasal uzantIsIdIr
172. terk etmek daha zor
173. tIkInmaya takatim kalmamIStI
174. togay bayatlI eski spor yazarlarImIzdandIr
175. trakya bOlgemiz Ulkemizin deGiSik iklimi olan yOrelerimizdendir

176. tUrbin kullanIldIGI durumda dUzgUn hattIn kullanIlmasI SarttIr
177. turgut aytekin ve selma gUneri
178. tUrkan Soray ve zeki mUren
179. tUrki cumhuriyetler baGImsIz olmalI bence
180. tUrkiye bUyUk millet meclisi anayasasI CIkarIlmIStI
181. tUrkiyeden CeSitli sivil OrgUtlerden onbeS kadIn vardI
182. UCU tuSlayIn lUtfen
183. ulaSIm araClarInda birlikte oturulmasI yasaktI
184. vatanIn sinesindeki o eski mUzmin yaralar artIk sarIlacaktI
185. ver allahIm Su kuluna bir akIl
186. vergiler arttIkCa hoSnutsuzluk baSlamIStI
187. vicdanI titremeyen bir fert yok mu
188. yaGmur Ozkan kombinasyonu iSe yaradI
189. yalInayak sokrates fena bir tiyatroymuS
190. yansIma ekranda birden fazla gOrUntUnUn UstUste CIkmasIna sebep olur
191. yarattIk her yaStan
192. yedek olarak depolamak tavsiye edilmektedir
193. yirminci yUzyIlIn bUyUklerinden sayIlIrdI
194. yUksek OGrenim hakkI talep edilirken baSka bir eGitim alInmIS olacaktI
195. yunanistandaki kadInlarla buluSmasI atinada gerCekleSti
196. yUrUyelim arkadaSlar
197. zavallI rumeli ateSler iCinde yandI
198. zeytin dalI barISIn simgesi
199. zira sIvI klorun artIGI karIStIrIcIyI tIkayabilir
200. zuhal olcay ve cihan Unal

A.2. Words

A.2.1. Words Containing Two Consecutive Vowels

Table A.1. Words containing two consecutive vowels

aaa	diana	iddialarI	koordine	oneal	tabii
aidatlarInI	diananIn	iddialarInI	koreografisini	oo	teaching
aile	die	iddialI	kuafOrde	philadelphia	teaS
aileleri	duasI	iddiasI	laik	puan	teessUf
ailenin	duayeni	iddiasInI	laikliGe	puanla	teoman
ailesinin	eee	iddiasIyla	laiklik	rauf	teorik
ait	email	ideallerin	laura	raund	terfian
aittir	enbiey	ideolojik	liizing	renoir	tiryakioGlu
alaaddin	engineering	iguanagiller	maalesef	rio	tuana
alnIaCIk	entellektUel	ilie	maaS	saat	tuena
anteplioGlu	erbain	industrial	maaSalara	saatlerde	tUsiadIn
antikacIoGlu	euro	inSaat	mafiaboy	saatleri	uefa
antonio	faaliyet	irticai	medea	saatlerinde	uncuoGlu
atrium	faaliyete	ismail	mesai	saatlik	uu
azraili	faaliyeti	israil	mevduat	saatte	venezUella
bahCelievler	facia	israilliler	michael	Sairin	video
bakIrcIoGlu	faiz	itfaiye	milguet	sanayii	videodan
bauhaus	faizi	jeomorfolog	milvauke	SanlIurfa	yazIcIoGlu
boa	fiilen	jeostratejik	moshoeu	sarIaGIz	zaaf
camiaya	fuarI	jiujitsu	muamele	sarIer	ziraat
carrefeur	fuarlar	joella	muayenesinin	seans	
cezaevi	fuat	kamuoyu	mUebbet	selUloit	
cezaevinde	galleria	kamuoyuna	mUesses	Seriat	
cezaevine	gaziantep	kamuoyunun	mukataa	SeriatCI	

Table A.2. Words containing two consecutive vowels, continued

civaoGlu	gaziosmanpaSa	karaaGar	mUracaat	sezai
commercial	hayIrdua	katliamlar	mUtearife	Siir
coolio	heloise	kItaat	nail	sofuoGlu
daima	hercai	koalisyon	nazlIoGlu	suadiye
dair	iade	koalisyondan	nihai	suat
daniel	iddia	kocaeli	noel	suavi
daughter	iddiadIr	kocaelindeki	nUkleer	Suurluluk
detroit	iddialar	kocaelispor	nuriosmaniye	taahhUtte

A.2.2. Some of the Other Words

Table A.3. Word list

a	aC	aCIkladI	baGlanmak	bak
abacIlIk	acaba	aCIkladIGI	baGlI	bakan
abartIcIlIk	aCacak	aCIklama	baGlIdIr	bakan
abbas	acaGIz	aCIklamada	baGnazlIk	bakana
abdal	aCan	baba	baha	bakanI
abdullah	acar	babasI	bahane	bakanlar
abdurrahman	acentalarI	babasI	bahar	bakanlIGI
abede	acI	bacayI	baharatCIlIk	bakanlIGIna
abelard	acIbadem	baGdaSIk	baharda	bakanlIGInIn
abelya	aCIdan	baGdaStIrma	bahattin	bakarak
aberasyon	aCIGa	baGfaS	bahCe	cuma
abi	aCIGIz	baGImlaSma	bahCeli	cumalIkIzIk
abone	aCIk	baGImsIz	bahCIvan	cumartesi
aboneden	aCIkCa	baGImsIz	bahname	cumhurbaSkanI

Table A.4. Word list, continued

cadde	Cakmak	davranmaya	egemen	filarizleme
caddesi	Cala	davulcu	eGer	filarmoni
cirit	Calan	dayalI	eGilimli	filelerle
citibank	CaldI	dayanIklI	eGinti	film
civarI	CalI	dayanmIyor	eGiS	filmi
civarIndadIr	CalISan	dayI	eGitim	filmleri
civelek	CalISanlar	de	eGitime	filolarInI
ceza	CalISIyor	debimetre	eGitimi	filozofluk
cezalarI	CalISma	december	eGitimin	final
cezanIn	CalISmada	dede	fettanca	finalde
cezasI	dar	dedi	fevkalade	finali
cezasI	daraldI	edilen	fevzi	finans
cezayirli	darbe	edilerek	feyyaz	finansal
check	dardanel	ediliyor	fezleke	garanti
CadIrdaysanIz	darIcada	edilmemesi	fIkIh	gardiyanlarIn
CadIruSaGI	darlaSma	edilmesi	fIkrasI	gargar
CaGcIllaStIrma	dava	edilmiS	fIndIkzade	garplIlaSma
CaGdaS	davacI	edilmiStir	fIrIldakCIlIk	gastrit
CaGIrdI	davalaSmak	edimli	fIrInlama	gavur
CaGIrmadIGImIz	davasI	edince	fIrkata	gaye
CaGIyla	davaya	edip	fIrsat	gayeli
CaGlar	davayI	ediyor	fIrsatCI	gayemiz
CaGlayan	davet	ediyorlar	fIrsatI	gayrImenkul
CaGrIsI	davetli	ediyorum	fIrtInasIna	gayrimenkul
cahilce	david	ediyoruz	fideci	gayrimUsavi
Cakandemir	davos	efes	figen	gaz
CakIcI	davranabilmesi	efim	fihristleme	gazete
CakIr	davrandI	eflatun	fikret	gazeteci
CakIroGlu	davranISlarI	efsane	fikrisabit	gazetecilerin
CakISIk	davranmalIsInIz	ege	filan	gazetelerde

Table A.5. Word list, continued

gazeteleri	hadise	Isfahan	iCmek	kabinesinin
gazetesinin	hafakan	ISIGInda	icra	kabloda
gazolin	hafIzalarImIzdan	ISIk	iCten	kablonun
ge	hafi	ISIkCI	iCtenlik	kabul
gebe	hafif	ISIklIIIk	iCtihatlara	kaburga
gebrelenmek	hafta	ISIldamak	idam	kabusuna
gebze	haftaki	IsInma	idare	kaC
geC	haftalarda	IsIrIk	idarenin	kaCakCIIIGI
gece	haftalIk	IsItIlmak	idaS	kaCar
geCecek	hak	Islami	idi	kaCIncI
geceleyin	hakan	Islavist	idrak	kaCIrtma
geCen	hakanIn	IstampacIIIk	ifade	kaCIyoruz
geCen	hakaret	Istanbul	ifadesini	kaCtI
geCer	hakem	iC	j	kadar
geCerli	hakemi	icadiye	jakuzi	kadarIyla
gecesi	hakemler	icaz	jale	kadarki
geCici	hakim	iCe	jandarma	kademeler
geciktirdiGi	hakimler	iCeren	japon	kaderi
geCirdi	hakkari	iCeriGi	japonya	kaderini
haber	hakkI	iCerisinde	japonyada	kadIkOy
haberdar	hakkInda	iCi	je	kadIn
haberi	hakkIndaki	iCin	jelatin	kadInIn
haberine	hadi	iCinde	jennifer	kadInlar
haberler	I	iCindeki	judith	kadInlara
haberleri	Ih	iCinden	jurnal	kadInlarIn
hacmine	IhtIrmak	iCine	ka	kadInlarla
haczetmek	IklIm	iCirtme	kabadayIca	kadife
haddini	Ilgaz	iCiSleri	kabak	kadir
hadIm	Irak	iCiSleri	kabalak	kadri
hadi	IrkCI	iCki	kabiliyetli	kadro

Table A.6. Word list, continued

kadrosu	leyla	magazin	neler	okurumuz
kafa	leylekgagasI	maGdur	nerede	okurumuzun
laboratuvarlardan	lezzetli	maGlup	nereden	okuyan
lacivert	liberal	mahalle	neredeyse	okuyanlarIn
laCkalaSmak	lider	mahallesi	nereye	okuyor
laf	lideri	mahallileSmek	nergis	okuyucularIm
lafarj	mablak	mahcuz	neriman	okyanus
lagos	maC	mahkeme	neSesizlik	okyay
laGvetmek	maCa	mahkemesi	neSet	ol
lahana	macaristan	mahkemesindeki	nesim	olabileceGini
laktoz	maCI	ne	neSriyat	olabilir
lale	maCIn	necdet	net	olabilirdi
lambiri	maCInda	necmettin	netaS	olabiliyor
langa	maCIndan	necmiye	ocak	olacaGI
lanoz	maClarda	neden	ocaka	olacaGIna
laponca	maClarI	nedeni	ocakta	olacaGInI
lar	maCta	nedeniyle	oda	olacaGIz
laso	madalya	nedenle	odaklaStIrmak	olacak
latife	madde	nedenlerle	odasI	olacaktIr
latife	maddenin	nedenli	odasInda	olaGan
lavrence	maddesi	nedense	ohal	olaGanUstU
layIk	maddesinde	nedim	okan	olamayacaGInI
lazanya	maddi	nedir	okanIn	Oz
lazIm	madem	nefes	oktay	Ozal
le	madra	nefret	oktayIn	Ozalp
lekesiz	madrid	nefyedilmek	okul	OzaydInlI
lento	mafya	negam	okullarda	Ozbay
ler	mafyasI	negatif	okullarInIn	OzcanIn
leva	maGazalar	nehirde	okulu	Ozdemir
levent	maGazasI	nejat	okumuS	Ozden

Table A.7. Word list, continued

Ozel	pamuklu	raGmen	sabin	Sadan
OzelleStirme	panayIr	rahat	sabit	SaSal
Ozellikle	panda	rahatlayacak	saCIStIrmak	SaSIrtma
Ozellikler	pandanIn	rahatsIz	sadakatlerinden	SaSma
Ozemek	pandora	rakam	saddam	Sehvetli
Ozenme	panel	rakamIn	sadece	SehzadebaSI
Ozer	pangalos	rakamlara	sadeleSmek	Seref
Ozerklik	panik	rakamlarI	sadi	Seker
Ozetin	pankart	rakamsal	saf	Sekerbank
Ozetlenme	panorama	rakibi	safCa	SereflikoChisar
OzgUn	papalIk	rakibine	saffet	Sekerpare
OzgUnleStirme	papazla	rakibini	saffete	Sekilde
Ozkan	papazlar	rakip	safhalarI	Sekillendirmek
Ozlem	papsu	raks	safinaz	Sekillenir
Oztezcan	para	ralli	saG	Serbetli
OzUmsenme	para	ramazan	saGa	Simdiden
Ozyineleme	parabellum	ramp	saGcI	Simdiye
pabetland	paradigma	randImanlI	saGda	SirpenCe
pabuCcu	parafe	rant	saGduyulu	SiSe
padiSahlar	paragrafIn	rantabl	saGIn	SiSecam
pahalI	paraketeci	rantCIsInIz	saGIr	SiShane
pahasIna	paralar	rapor	saGladI	SiSkin
pak	rabItasIzlIk	raporda	saGlam	Sok
paketi	radikal	raporu	saGlamak	SOyle
paketlemek	radon	raporun	saGlamaya	taban
paketten	radyo	raporunu	saGlanacak	tabanca
pakistan	rafet	rapten	saGlanan	tabelasInda
pakistanda	rafIyla	sabah	saGlandIGI	tabi
palamut	raftan	sabaha	saGlanIr	tabla
palazlamak	raGbet	sabancI	Sad	tablo

Table A.8. Word list, continued

tacik	uCak	umudu	Unsal	var
taCsIzlar	uCakla	umursama	Uretim	varamadIGImIz
taffarel	uCurumdan	umut	UrettiGi	vardI
tahIl	ucuz	umutlu	UrkUtUcU	vardIr
tahir	ufalmak	UC	UrUndUl	varGUCleriyle
tahkim	uGra	UCgeninden	UrUnlerdeki	varlIGInI
tahmin	uGradI	Ucret	UrUnleri	varlIk
tahran	uGradIGI	Ucreti	UrUnlerini	varsa
tahrilli	uGraSan	Ucretlilere	UskUdar	varyete
tahsildar	uGraSIlma	UCUncU	vadede	vashington
tahsin	uGraSmaya	Ulke	vadeli	ya
tahsis	uGrayan	Ulkede	vadesiz	yabancI
tahta	uGur	Ulkeden	vagon	yaG
tahtIrevan	ulagay	Ulkeler	vahdetivUcut	yaGdIrdI
takabilmelerini	ulaSan	Ulkelerde	vahSilik	yaGhane
takdir	ulaSIncaya	Ulkelerdeki	vakfe	yaGISOlCer
takdirde	ulaStI	Ulkeleri	vakfI	yaGlanma
takIldIGI	ulaStIGInI	Ulkelerin	vakfInIn	yaGmur
takIm	ulaStIrIlacak	Ulkemize	vakIf	yaGmurdereli
takImI	ulaStIrma	Ulkenin	vakko	yahu
takImIn	uludaG	Ulkeye	vali	yahudi
taki	uluG	Ulkeyi	valide	yahut
takibe	uluGbay	Ulkeyle	valisi	yakaladIlar
takip	ulus	UlkUleStirme	vallahi	yakalanan
takmak	ulusal	Umit	van	yakalattIGI
takrir	uluslararasI	Umitlendirmek	vana	yakalayabilecekti
taksim	uluyol	Unal	vandaki	yakIn
taksime	uluyolun	Universite	vanet	yakIndan
taksirli	umarIz	Universitesi	vanspor	yakInlarIndaki
uCaGI	umman	UnlU	vapur	yakInsaklIk

Table A.9. Word list, continued

yakISIksIz	zamanlarda
yakinen	zamklamak
yaklaSIk	zammI
yaklaSImI	zarar
yaklaSmak	zararlI
yaktI	zaten
yalan	zaten
yalanladI	zaten
yalanlayan	zaten
yalCIn	zavallI
yalCInsu	zayIf
yalI	ze
yalnIz	zebra
yalnIzca	zehirledi
yalpa	zekasI
zafer	zekayla
zaGcI	zekeriya
zagor	zeki
zahiri	zemin
zahmet	zemine
zam	zengin
zamaldinov	
zaman	
zamanda	
zamandIr	
zamanInda	
zamanIymIS	
zamanIymIS	
zamanki	
zamanlar	

REFERENCES

1. Rabiner, L. R. and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall Inc., Englewood Cliffs, N. J., 1978.

2. Klatt, D. H., "Review of Text-to-Speech Conversion for English", *J. Acoust. Soc. Am.*, Vol. 82, No. 3, pp. 737–793, September 1987.

3. Van Santen, J. P. H., "Chapter 5: Timing", in Richard Sproat (editor), *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, pp. 115-139, Kluwer Academic Publishers, 1998.

4. Van Santen, J. P. H., "Contextual Effects on Vowel Duration", *Speech Communication*, Vol. 11, pp. 513–546, 1992.

5. Yi, J. R. W. and J. R. Glass, "Natural-Sounding Speech Synthesis Using Variable-Length Units", *Proceedings of ICSLP*, 1998.

6. Yapanel, Ü., *Garbage Modeling Techniques for a Turkish Keyword Spotting System*, M.S. Thesis, Boğaziçi University, 2000.

7. Shih, C. and B. Ao, "Duration Study for the Bell Laboratories Mandarin Text-to-Speech System", in J. P. H. Van Santen, R. W. Sproat, J. P. Olive and J. Hirschberg (editors), *Progress in Speech Synthesis*, pp. 383-399, Springer-Verlag, New York, 1997.

8. Demircan, Ö., *Türkiye Türkçesinin Ses Düzeni, Türkiye Türkçesinde Sesler*, Türk Dil Kurumu Yayınları, Ankara, 1979.

9. Banguoğlu, T., *Türk Grameri*, Türk Tarih Kurumu Basımevi, Ankara, 1959.

10. Selen, N., *Söyleyiş Sesbilimi, Akustik Sesbilim ve Türkiye Türkçesi*, Türk Dil Kurumu Yayınları, Ankara, 1979.

11. Montgomery, D. C., *Design and Analysis of Experiments*, John Wiley & Sons, 2001.

12. Devore, J. and N. Farnum, *Applied Statistics for Engineers and Scientists*, Duxbury Press, Pacific Grove, CA, 1999.

13. Kanji, G. K., *100 Statistical Tests*, Sage Publications, 1993.

14. Kreyszig, E., *Advanced Engineering Mathematics*, John Wiley & Sons, 1993.

15. Crystal, T. H. and A. S. House, "Segmental Durations in Connected-Speech Signals: Current Results", *J. Acoust. Soc. Am.*, Vol. 83, No. 4, pp. 1553–1573, April 1988.

16. Van Santen, J. P. H. and R. Sproat, "Chapter 2: Methods and Tools", in Richard Sproat (editor), *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, pp. 7-30, Kluwer Academic Publishers, 1998.

17. Klatt, D. H., "Interaction Between Two Factors That Influence Vowel Duration", *J. Acoust. Soc. Am.*, Vol. 54, No. 4, pp. 1102–1104, 1973.

18. Van Santen, J. P. H., "Analyzing N-way Tables with Sums-of-Products Models", *Journal of Mathematical Psychology*, Vol. 37, pp. 327-371, 1993.

19. Riley, M., "Tree-Based Modelling for Speech Synthesis", in G. Bailly and C. Benoit (editors), *Talking Machines: Theories, Models, and Designs*, pp. 265-273, Elsevier Science Publishers B. V., Amsterdam, 1992.

REFERENCES NOT CITED

Allen, J., M. S. Hunnicutt, and D. Klatt, *From Text to Speech: The MITalk System*, Cambridge University Press, Cambridge, 1987.

Bellegarda, J. R., K. E. A. Silverman, K. Lenzo and V. Anderson, "Statistical Prosodic Modelling: From Corpus Design to Parameter Estimation", *IEEE Trans. Speech Audio Processing*, Vol. 9, No. 1, pp. 52–66, January 2001.

Butler, C., *Statistics in Linguistics*, Basil Blackwell, Oxford, UK, 1985.

Campbell, W. N., "Syllable-Based Segmental Duration", *Talking Machines: Theories, Models, and Designs*, Elsevier Science Publishers B. V., pp. 211–224, 1992.

Chung, G., *Hierarchical Duration Modelling for a Speech Recognition System*, M.S. Thesis, M.I.T., May 1997.

Crystal, T. H. and A. S. House, "Segmental Durations in Connected-Speech Signals: Syllabic Stress", *J. Acoust. Soc. Am.*, Vol.83, No. 4, pp. 1574–1585, April 1988.

Emerard, F., L. Mortamet and A. Cozannet, "Prosodic Processing in a Text-to-Speech Synthesis System Using a Database and Learning Procedures", *Talking Machines: Theories, Models, and Designs*, Elsevier Science Publishers B. V., pp. 225–253, 1992.

Harris, M. S. and N. Umeda, "Effect of Speaking Mode on Temporal Factors in Speech: Vowel Duration", *J. Acoust. Soc. Am.*, Vol. 56, No. 3, pp.1016–1018, September 1974.

House, A. S., "On Vowel Duration in English", *J. Acoust. Soc. Am.*, Vol. 33, pp. 1174–1178, September 1961.

Klatt, D. H., "Linguistic Uses of Segmental Duration in English: Acoustic and Per-

ceptual Evidence", *J. Acoust. Soc. Am.*, Vol. 59, No. 5, pp. 1208–1221, May 1976.

Kaiki, N., K. Takeda and Y. Sagisaka, "Linguistic Properties in the Control of Segmental Duration for Speech Synthesis", *Talking Machines: Theories, Models, and Designs*, Elsevier Science Publishers B. V., pp. 255–263, 1992.

O'Shaughnessy, D., "A Multispeaker Analysis of Durations in Read French Paragraphs", *J. Acoust. Soc. Am.*, Vol. 76, No. 6, pp. 1664–1672, December 1984.

Port, R., "Linguistic Timing Factors in Combination", *J. Acoust. Soc. Am.*, Vol. 69, pp. 262–273, January 1981.

Sproat, R. (editor), *Multilingual Text-to-Speech Synthesis:The Bell Labs Approach*, Kluwer Academic Publishers, Massachusetts, 1998.

Umeda, N., "Vowel Duration in American English", *J. Acoust. Soc. Am.*, Vol. 58, No. 2, pp. 434–445, August 1975.

Umeda, N., "Consonant Duration in American English", *J. Acoust. Soc. Am.*, Vol. 61, No. 3, pp. 846–858, March 1977.

Van Santen, J. P. H., "Deriving Text-to-Speech Durations from Natural Speech", in G. Bailly and C. Benoit (editors), *Talking Machines: Theories, Models, and Designs*, pp. 275-285, Elsevier Science Publishers B. V., Amsterdam, 1992.