Duration of Turkish Vowels Revisited

Ebru Arısoy[*], *Levent M. Arslan*[*], *Mine N. Demiralp*[**], *Hazım K. Ekenel*[*], *Meltem Kelepir*[**], *H. Mesut Meral*[**], *A. Sumru Özsoy*[**], *Ömer Şayli*[***], *Oytun Türk*[*], *Beste Can Yolcu*[**]

[*]Boğaziçi University, Electrical and Electronics Engineering Dept.
[**]Boğaziçi University, Western Languages and Literatures Dept.
[***]Boğaziçi University, Biomedical Engineering Institute

Abstract

This paper provides an instrumental description of the durational properties of Turkish vowels as part of a larger research project for a fuller acoustic description of the sounds of the language. The basic aim of the paper is to contribute to the very few instrumental phonetic analyses of Turkish. By using audio recording and speech processing tools such as WaveSurfer, HTK, and WavRec, 111 Turkish words uttered by 6 native speakers were recorded and analyzed in terms of mean durations of eight equally distributed Modern Standard Turkish (MST) vowels. It is found that (i) vowels in the initial syllables of multisyllabic words have significantly lower mean durations when compared to vowels in final syllables; (ii) high vowels have lower mean durations than low vowels, confirming the findings of some earlier studies (Şayli 2002), (Şayli & Arslan 2003).

1. Introduction

The results of previous works on the duration of Turkish vowels (Şayli 2002), (Şayli & Arslan 2003) were based on a large database which was not controlled with respect to the number of occurrences of a given vowel in a given environment. Therefore, the aim of this current study is to provide an analysis of the phenomenon based on a controlled set of data and to see whether the results previously reported will be replicated. Specifically, durational properties of the vowels in initial versus final syllables have been investigated.

2. Background

2.1. Definitions

Speech can be described as an acoustic wave that is radiated when air is expelled from the lungs and the resulting flow of air is perturbed by a constriction along the vocal tract (Rabiner et al. 1978). *Speech waveform* (Figure 1a) is the digital representation of this acoustic wave in the computer environment. It shows the change in air-pressure caused by the speech signal as a function of time.

Pitch is a perceptual concept, described in terms of a scale from low to high. To give an example, male adult voice generally has lower pitch than a male child or a woman. Generally, average pitch is around 125 Hz for male speakers and 200 Hz for female speakers. Figure 1b shows the pitch contour of the utterance *ütopik* for a female speaker. It is clear from the figure that the concept of pitch is valid only for the voiced sounds. Pitch contour as in Figure 1b serves as the main ground for the analysis of intonation.

One of the main concepts in acoustic speech analysis is *fundamental frequency*, which is measured in Hertz (Hz). It corresponds to the vibration frequency of the vocal folds in voiced sounds and is highly correlated with the perceived pitch of the speaker, i.e. the lower the fundamental frequency is, the lower the pitch will be perceived. The inverse of fundamental frequency (F0) is the *fundamental period* (T), i.e. T=1/F0. In Figure 2, the fundamental period in a recording of the vowel [a] is marked on the speech waveform for a male and a female speaker. The fundamental period is shorter for the female speaker. Therefore, the fundamental frequency is higher as frequency is inversely proportional to the period.

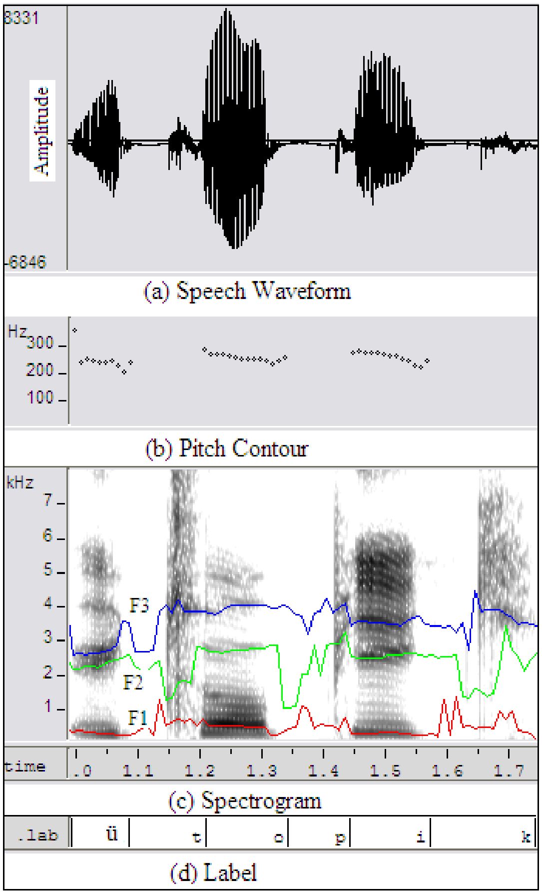

Figure 1: Interface of Wavesurfer 1.5 of *ütopik*

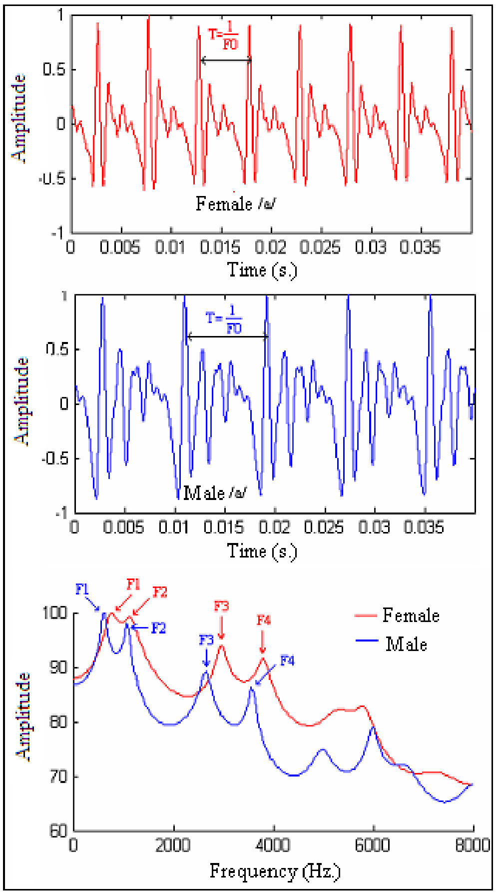

Figure 2: Fundamental frequency and formant frequency

Formant frequencies or simply *formants* are the high-energy regions in the spectrum. Formants depend on the shape and dimension of the vocal tract that changes during the generation of each sound. Moreover, formant frequencies are used to characterize the phoneme being produced as well as the identity of the speaker.[1] Average formant frequencies for female speakers are higher

[1] See Türk et al. (2004), for the claim that formant frequencies are reliable distinctive features especially for vowels.

than male speakers. The third plot in Figure 2 shows the first four formant frequencies (F1, F2, F3, F4) for a female and a male speaker uttering the vowel [a]. All the formant frequencies for the female speaker are higher than the male speaker.

In the process of speech production, the shape of the vocal tract continuously changes to produce different sounds. These changes result in the variation of the frequency characteristics of the speech in time, represented in a *spectrogram*. Figure 1c shows the spectrogram for the word *ütopik*. The vertical axis corresponds to frequency and the horizontal axis corresponds to time. The gray scales in the spectrogram indicate the relative amplitude or the energy at that frequency. Darker colors represent higher energy regions. As a result, formant frequencies are observed as dark-colored horizontal bands in the spectrogram. The straight horizontal first line from the bottom that marks a dark-colored band between 1.2 and 1.3 seconds corresponds to the first formant (F1) in Figure 1. A spectrogram also indicates voicing quite clearly: voiced regions appear as a group of dark lines as in the case of areas corresponding to [ü], [o] and [i] in Figure 1, while the shape of voiceless regions is blurred.

The process of marking the phonetic contents of speech sounds is called *labelling*. In this research, we used Wavesurfer 1.5 as a manual labelling tool. The labels of the word "ütopik" are shown in Figure 1d. Automatic labelling is also available but with varying degrees of reliability in relation to the properties of the sound (See Section 3.2 for a description of our labelling procedure).

Formant frequencies and duration are the main acoustic cues instrumental-computational studies rely on, since they are shown to represent distinctive properties of sounds. For example, there is a clear correspondence between the classification of the vowels with respect to the articulatory parameters such as tongue position as proposed in Selen (1979) and the one with respect to formant frequencies observed in Türk et al. (2004). Another correspondence is observed between the articulatory parameter height and the durational statistics of vowels. Mean durations of non-high vowels (/a/, /e/, /o/, /ö/) are found to be greater than those of high vowels (/ı/, /i/, /u/, /ü/) (Şayli & Arslan 2003).

2.2. Overview

Several studies employed software and/or hardware tools for examining Turkish phonetics and acoustics.

In Türk et al. (2004), formant frequency analysis is performed in different groups of female and male speakers for Turkish vowels. The results show that F1 and F2 characteristics can be used to discriminate among different vowels along the following lines: for male speakers, the vowel space can be classified into three regions using F1 statistics and three regions using F2 statistics as shown in Table 1. For female speakers, F1 divides the vowel space into two, and F2 into three. The formant values for children show significantly higher level of overlap for different vowels. Therefore, discrimination of vowels using F1-F2 statistics is more difficult for children's speech.

Parameter	Vowel(s)	Range of parameter
Average F1 for male speakers	/a/	F1>600 Hz
	/e/, /o/, /ö/	430 Hz < F1< 600 Hz
	/i/, /ü/, /u/	F1 < 430 Hz
Average F2 for male speakers	/a/, /u/, /o/	F2 > 1450 Hz
	/ı/	1450 Hz < F2 < 1650 Hz
	/e/, /i/, /ü/, /ö/	F2 < 1450 Hz
Average F1 for female speakers	/a/, /ı/, /o/, /u/	F1 > 600 Hz
	/e/, /i/, /ö/, /ü/	F1 < 600 Hz
Average F2 for female speakers	/ı/, /i/, /ü/	F2 > 1600 Hz
	/e/, /a/	1400 Hz < F2 < 1600 Hz
	/ö/, /o/, /u/	F2 < 1400 Hz

Table 1. F1-F2 characteristics of Turkish vowels (Türk et al. 2004).

In Şayli (2002) and Şayli et al. (2002) and (2003), mean durations of MST phonemes are investigated as well as the effect of some contextual factors such as the position in word/sentence, the nature of the preceding or the following sounds, number of syllables in the word, the number of words in the sentence. In Şayli & Arslan (2003), additional recordings of an adult male are used and mean durations of vowels and consonants are analyzed. The results are in agreement with the previous study by Şayli (2002) for the classifications with respect to mean durations in one-word and sentence utterances. These studies reveal that (i) mean durations of high vowels (/ı/, /i/, /u/, /ü/) are lower than low vowels (/a/, /e/, /o/, /ö/) , (ii) mean durations of vowels decrease with increasing syllable number in the word and increasing word number in the sentence, and (iii) mean durations of the vowels are the greatest in the word final and the smallest in the word initial positions.

3. Duration Analysis

3.1. Database

A database consisting of 111 words containing each of the eight Turkish vowels in two environments was constructed for our study. The selection of the words for the analysis was based on the following criteria: for each vowel, 5-7 real words where the vowel in question occurs in the initial syllable of a multisyllabic word and 5-7 real words where the vowel occurs in the final syllable. Since [o] and [ö] are restricted to first syllables in Turkish, we used 5-7 made-up words for the final syllable environment for these vowels.[2] To ensure the reliability of automatic labelling of the vowels, the words analyzed are further restricted to those in which the analyzed vowel is preceded or followed by a voiceless stop. The occurrences of the voiceless stops [p], [t] and [k] are more-or-less equally distributed over the database (See Appendix I for the full database).

[2] All Turkish words in our database had word stress on their last syllable. However, the made-up words received different stress patterns for each speaker. Please refer to Figure 4 and note that the word final duration of [o] and [ö] pattern similarly with other sounds. This precludes the possibility that word stress is a variable acting over duration of vowels.

To further control for possible additional variables that could affect duration such as gender, age, and educational background, all 6 speakers participated in this study were female university students or graduates of ages 24± 3.

Recording sessions were carried out using WavRec. The recordings were automatically labelled by HTK (Woodland et al. 1984). Labelled files were also randomly checked manually. Spectrogram analysis and manual labelling were performed using Wavesurfer 1.5. This proved the reliability of the labelling of vowels in the environment of voiceless phonemes done by HTK. Finally, the program for statistical duration analysis was developed in Matlab 6.0.

3.2. Analysis

Figure 3 shows the flowchart of the duration analysis method. In the labelling process, the speech signals are forced-aligned with the corresponding text transcription with Viterbi decoding algorithm (Rabiner 1989). Monophone Hidden Markov Models (HMMs) trained from 200 native Turkish speakers (100 male, 100 female) are used. The mean and the standard deviation of the duration of different vowels in initial versus final syllables are determined using the statistical analysis program developed in Matlab 6.0.

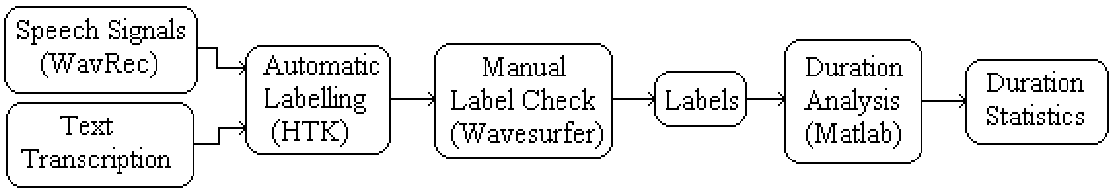

Figure 3. Flowchart of the duration analysis method.

Figure 4 below shows the distribution of vowel durations in the two conditions for all vowels. From Figure 4, a clear distinction between vowel durations in initial versus final syllable is observed. Table 2 shows the distinction in numerical terms. We have also applied a standard t-test with the null hypothesis that the durations of vowels in the two cases are not different. For all vowels, the null hypothesis was rejected in the 99% confidence interval. Therefore, we conclude that initial syllable vowel durations are significantly smaller than those of last syllable with 99% confidence.

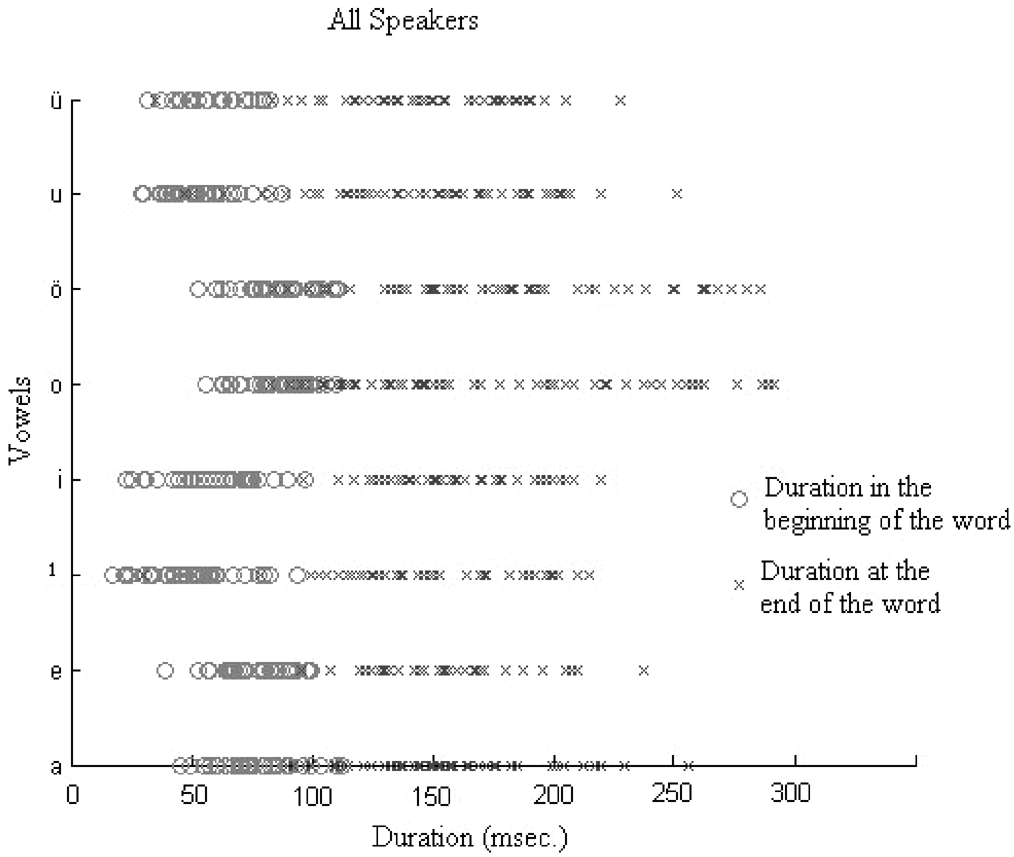

Figure 4. Duration analysis results for all speakers and for all vowels of interest.

	/a/	/e/	/ı/	/i/	/o/	/ö/	/u/	/ü/
Mean (IS)	74.0	76.0	57.4	49.6	84.2	82.4	53.0	58.8
Mean (FS)	154.3	155.4	154.9	148.4	172.8	184.1	148.1	148.0
Std. dev(IS)	15.8	13.2	20.6	17.7	14.0	20.2	14.0	13.4
Std. dev. (FS)	36.5	31.3	28.9	40.1	60.9	51.2	46.0	35.3

Table 2. Means and standard deviations of vowel durations. IS: initial syllable, FS: final syllable.

3.1. Discussion

The analysis incorporating an equal number of occurrences of each vowel in each condition and restricted only to vowels preceding and following voiceless stops replicates the findings of Şayli (2002), Şayli et al. (2002) and Şayli & Arslan (2003) in that (i) high vowels have lower mean durations than low vowels in general and (ii) the position in the word is a determining factor in the duration of vowels. Vowels in the first syllable of a multisyllabic word have significantly lower mean durations when compared to vowels in final syllables across speakers. The current study eliminated a potential bias that could have resulted from the unequal distribution of the vowels in the randomly selected database used in the previous studies.

An additional observation can be that duration range is greater in the final-syllable condition, and relative greatness in the word-final duration range is due to inter-speaker variation (see appendix II for individual speaker statistics). However, no statistical analysis of this observation is carried out in this research.

4. Further Remarks

Instrumental phonetics is a very promising area in Turkish linguistics. Not only because it presents new and interesting ways of reconsidering the data, such as duration, it also provides an important tool for examining and testing linguistic accounts made on articulatory grounds. Following this trail, the previously acknowledged phonological processes like aspiration, word final devoicing, Turkic versus non-Turkic long vowels, the exact nature of "ğ" can be described in much more precise terms. Specifically, durational description of Turkish vowels in other environments and in different syllable settings needs to be made. With accompanying formant frequency analyses of the vowels, it may be possible to reach a computer-sensitive description of Turkish vowels which would pave the way for more advanced studies in the field.

References

Hardcastle, W. J. & Laver, J. (1997). *The Handbook of Phonetic Sciences*. Oxford: Blackwell.

Kopkallı, H. (1993). *A Phonetic and Phonological Analysis of Final Devoicing in Turkish*. University of Michigan. Ph. D. Dissertation.

Rabiner, L. R. & Schafer, R. W. (1978). *Digital Processing of Speech Signals*. New Jersey: Prentice Hall.

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

Şayli, Ö. (2002). *Duration Analysis and Modelling for Turkish Text-to-Speech Synthesis*. Boğaziçi University. M.S. Thesis.

Şayli, Ö., Arslan L. M. & Özsoy A. S. (2002). Duration Properties of the Turkish Phonemes. *11th International Conference on Turkish Linguistics*. Gazimağusa, KKTC.

Şayli, Ö., Özsoy A. S. & Arslan L. M. (2003). Türkçe'de Ses Sürelerine Etki Eden Etmenler ve Etkileri. *XVII. Dilbilim Kurultayı*. Anadolu Üniversitesi.

Şayli, Ö. & Arslan L. M. (2003). *Türkçe'deki seslerin süre özellikleri*. Dilbilim Araştırmaları, 15-26, Boğaziçi University Press.

Türk, O., Şayli, Ö., Özsoy, A. S. & Arslan, L. M. (to appear). Türkçede Ünlülerin Formant Frekans İncelemesi. *Proceedings of the 18th National Conference in Turkish Linguistics*. Ankara University.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, Vol. 77, No. 2, February.

Woodland, P. C., Odell, J. J., Valtchev, V., & Young, S. J. (1994). Large Vocabulary Continuous Speech Recognition using HTK. *Proceedings of the IEEE ICASSP*.

Appendix I: Alphabetic list of words analyzed

akar	a-	ıkın	ı-	köfte	-e	öküzlük	ö-	tepe	-e	utanmaz	u-
akı	a-ı	ıkıntı	ı-	kötö	-ö	öpül	ö-	tepki	-i	ülkü	-ü
apar	a-	ıtır	ı-	kötü	-ü	öpüş	ö-	testi	-i	ütelgi	ü-
apolet	a-	ıtırlı	ı-	kuko	-o	örtü	-ü	tıpa	ı-a	ütopik	ü-
ata	a-a	iki	i-i	kupo	-o	öter	ö-	tilki	-i	ütü	ü-ü
ataç	a-	ikon	i-	kuruntu	-u	ötüş	ö-	tipi	-i	ütücü	ü-
atık	a-	ipek	i-	kutu	-u	pinti	-i	to	-o	ütüsüz	ü-
buto	-o	ipotek	i-	kuytu	-u	po	-o	toka	-a	yaka	-a
coşku	-u	iter	i-	kütük	-ü	pö	-ö	topu	-u	yapı	-ı
çatı	-ı	itiş	i-	leke	-e	rakı	-ı	topuk	-u	yatak	-a
çıta	-a	kapı	-ı	mütö	-ö	sapa	-a	tö	-ö	yeti	-i
ekim	e-	katı	-ı	okan	o-	sıpa	-a	tököç	-ö		
ekip	e-	kete	-e	okoç	-o	sopa	-a	törpü	-ü		
ekol	e-	kıpırtı	ı-	oku	o-u	sopok	-o	tüpü	-ü		
epik	e-	kirpi	-i	okul	o-	söpö	-ö	türkü	-ü		
epitel	e-	ko	-o	opak	o-	şotok	-o	ukala	u-		
eter	e-	kokot	-o	operatör	o-	şükö	-ö	ukap	u-		
etik	e-	koku	-u	otağ	o-	taka	-a	utan	u-		
göpök	-ö	korku	-u	otel	o-	takı	-ı	utanç	u-		
güfte	-e	kö	-ö	öküz	ö-	teke	-e	utangaç	u-		

Note: [letter]- represents sound in first syllable, -[letter] represents sound in last syllable. These symbols following each list item indicates the sound of which duration is taken into account.

Appendix II: Durational statistics for individual speakers

Vowel = /a/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	74.0	58.6	61.2	74.7	83.2	88.3	78.3
Mean (FS)	154.3	163.0	158.4	150.4	127.3	194.7	132.0
Std. dev. (IS)	15.8	4.5	10.0	12.3	14.9	15.3	12.6
Std. dev. (FS)	36.5	50.5	17.1	28.6	22.7	31.1	17.0

Vowel = /e/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	76.0	63.7	75.6	78.1	84.3	78.6	75.7
Mean (FS)	155.4	178.8	161.6	131.8	140.0	189.4	131.0
Std. dev. (IS)	13.2	9.0	10.1	11.9	11.2	10.4	19.1
Std. dev. (FS)	31.3	25.8	9.2	16.3	26.6	30.5	21.4

Vowel = /i/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	49.6	41.7	46.5	40.8	60.8	61.7	46.2
Mean (FS)	148.4	169.5	154.5	131.9	132.9	198.9	102.7
Std. dev. (IS)	17.7	18.4	14.7	6.9	18.0	22.7	15.0
Std. dev. (FS)	40.1	41.0	13.8	22.7	24.9	6.0	37.0

Vowel = /ı/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	57.4	49.6	42.7	51.4	69.2	63.5	65.6
Mean (FS)	154.9	183.6	155.0	136.4	141.8	173.6	137.7
Std. dev. (IS)	20.6	26.1	9.6	25.0	15.4	23.5	9.1
Std. dev. (FS)	28.9	32.2	10.2	37.6	14.3	14.7	16.9

Vowel = /o/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	84.2	84.7	73.9	82.6	87.5	95.5	80.9
Mean (FS)	172.8	191.0	182.0	148.7	152.3	205.1	158.1
Std. dev. (IS)	14.0	9.1	14.7	15.4	11.4	15.5	11.2
Std. dev. (FS)	60.9	76.1	53.1	56.2	51.7	65.9	50.7

Vowel = /ö/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	82.4	82.9	86.4	75.6	90.4	72.9	86.2
Mean (FS)	184.1	205.8	198.7	157.3	168.3	221.9	152.6
Std. dev. (IS)	20.2	13.8	14.8	14.5	20.1	37.6	11.5
Std. dev. (FS)	51.2	57.2	41.9	42.7	26.3	51.1	51.7

Vowel = /u/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	53.0	46.1	45.0	47.2	66.2	55.6	57.8
Mean (FS)	148.1	167.8	152.4	141.3	128.6	192.5	105.9
Std. dev. (IS)	14.0	5.6	11.4	13.2	18.7	13.0	9.7
Std. dev. (FS)	46.0	53.3	29.7	54.7	37.3	14.1	24.0

Vowel = /ü/ (duration statistics in milliseconds)

	All	Spk1	Spk2	Spk3	Spk4	Spk5	Spk6
Mean (IS)	58.8	63.0	50.6	55.1	67.0	57.4	59.4
Mean (FS)	148.0	173.1	142.6	137.3	147.2	171.5	116.2
Std. dev. (IS)	13.4	15.8	11.8	9.8	15.5	12.7	14.6
Std. dev. (FS)	35.3	39.5	22.9	45.2	25.5	24.5	17.1