Voice Similarities between Brothers*

Do-Heung Ko ** · **SunMee Kang*****

ABSTRACT

This paper aims to provide a guideline for modelling speaker identification and speaker verification by comparing voice similarities between brothers. Five pairs of brothers who are believed to have similar voices participated in this experiment. Before conducted in the experiment, perceptual tests were measured if the voices were similar between brothers. The words were measured in both isolation and context, and the subjects were asked to read five times with about three seconds of interval between readings. Recordings were made at natural speed in a quiet room. The data were analyzed in pitch and formant frequencies using CSL (Computerized Speech Lab), PCQuirer and MDVP (Multi-dimensional Voice Program).

It was found that data of the initial vowels are much more similar and homogeneous than those of vowels in other position. The acoustic data showed that voice similarities are strikingly high in both pitch and formant frequencies. It was also found that the correlation coefficient was not significant between parameters above.

Keywords: Formant Frequencies, Pitch, Voice Similarities, Speaker Identification

1. Introduction

It is assumed that the similarities of voice among family members are high when measured both perceptually and acoustically. If two brothers share common physiological characteristic, and then their voices are very similar acoustically. The voices, as a matter of fact, are often strikingly similar among brothers. It is presumed that these common physiological features come from the same biological mother.

It is a well-known fact that there are many factors affecting voice qualities: (a) the length of vocal tract, (b) the size of the vocal folds, (c) the thickness and elasticity of the vocal folds, (d) the shape and function of articulators, (e) the size of cavities, and (f) speaker's attitude or emotion. It is said that fundamental frequency and possibly

---

* This was supported by a grant from the Interdisciplinary Research Program (Contract No. 1999-2-302-106-5) of the KOSEF.

** Division of Speech Pathology and Audiology, Hallym University

*** Department of Computer Science, Seokyeong University

amplitude are usually affected by (a), (b) and (c), and formant frequencies are affected by (d) and (e). And pitch patterns can be conditioned by all the factors above.

This paper deals with voice similarities among brothers who are supposed to have common physiological characteristics from a biological mother. Statistical data obtained from this experiment can be used as a guideline for modelling speaker identification and speaker verification.

2. Experimental Procedures

2.1 Subjects and Speech materials

Six pairs of brothers who are believed to have similar voices are participated in this study. Unfamiliar listeners were instructed to listen to the recordings and rate their similarity on a numerical scale (1-10). Results showed that most of them indicated very high perceptual scores (8 out of 10) except one pair of brothers. The speech samples obtained from one pair of brothers were eliminated in the analysis. In this experiment, the initial vowels are analyzed because the data of the initial vowels are much more similar and homogeneous than those of vowels in other positions. The words are measured in both isolation and context, and the subjects were asked to read the text five times with about three seconds of intervals. The word list is as below:

Table 1. List of test words

Analysis / Words	Syllable Initial Position	
	in isolation	in context
kaji 'eggplant'	/a/	
munje 'problem'	/u/	
iyagi 'story'	/i/	
moksori 'voice'	/o/	

In context, the carrier sentence is : 'This is called ______.' Sentences are used in analyzing pitch patterns as follows:

(1) Youngho-ga maeu aphayo. 'Youngho is very sick.'
(2) Oje hakkyo-e kassossoyo. 'Yesturday, I went to school.'
(3) sesang-un cham arumdapkunyo. 'The World is so beautiful.'
(4) nori kongwon-enun saramdul-i mani moimnida. 'Children's parks are always crowded.'
(5) onul taehangno-eso yeoguk-ul poassoyo. 'We went to see a drama at Taehangno.'
(6) Chunchon-e yeorum-un maeu teowoyo. 'It's very hot in summer in Chunchon.'

2.2 Acoustic Analysis

Recordings were made with a natural speed in a quiet room. The data were analyzed in two major parameters including pitch in sentences, formant frequencies (F1, F2, F3) in words using CSL (Computerized Speech Lab) and PCQuirer. T-test and correlation coefficient in brothers were measured to see whether or not the statistical results are homogeneous.

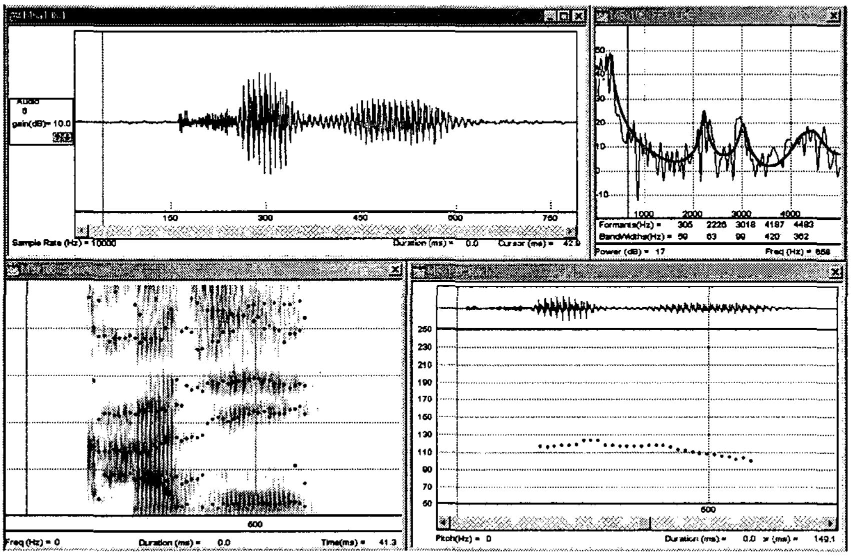


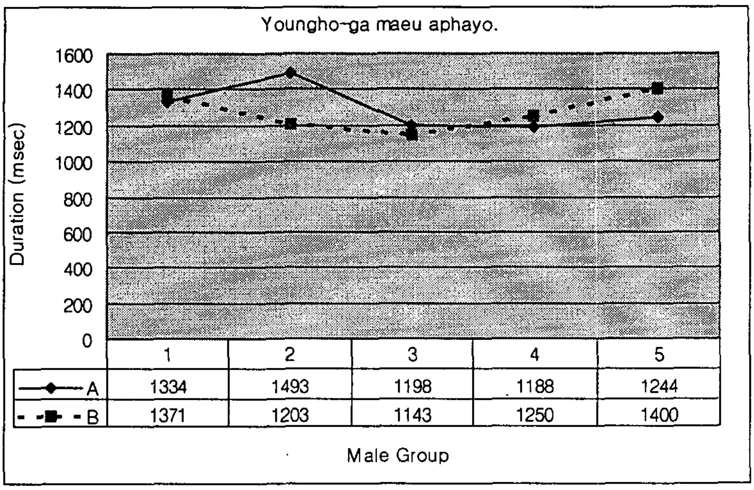
Figure 1. Samples of Acoustic Measurements

3. Results and Discussion

Voice quality plays an essential role in speech communication. However, the nature of speech is complicated because there are many factors affecting voice quality such as frequency (pitch), pressure (loudness), spectrum (timbre), duration (length), timing (rhythm). This experiment is confined some major acoustic parameters such as duration, pitch, and formant frequencies.

3.1 Duration

Even though recordings were made at natural speed, their speaking rates were different for some pairs of brothers as seen in Figure 2. In case pair 2, the difference of utterance durations was 290 msec which is considerably long. However, their pitch patterns were surprisingly similar. The duration of voice between brothers is very similar to that of sisters (Ko (2001)).



	1	2	3	4	5
A	1334	1493	1198	1188	1244
B	1371	1203	1143	1250	1400

Figure 2. A sample utterance duration

3.2 Pitch

As we can see in Figure 3, pitch patterns were very similar. In a sentence which has four peaks, the mean peak values, and the rate of declination were strikingly similar in all pairs of brothers. It is assumed that pitch which is the perceptual correlates of fundamental frequency (Fo) is one of the most important cues for voice similarities because pitch patterns are fairly similar.

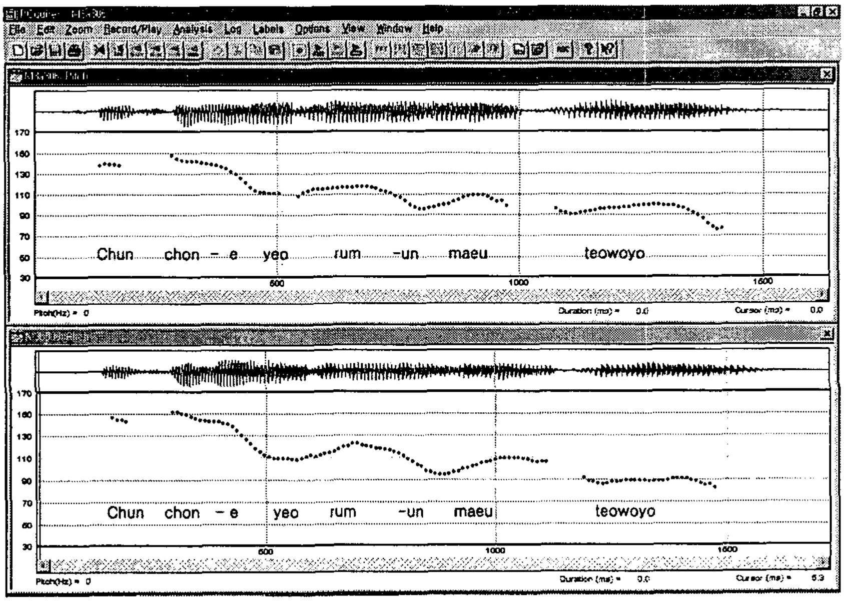
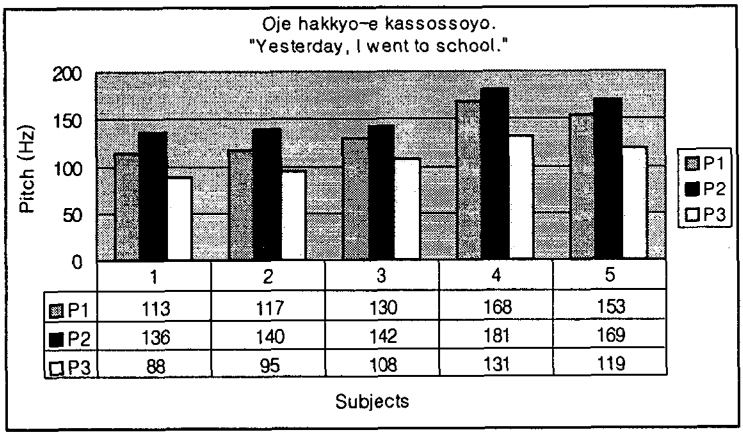


Figure 3. Pitch contour between brothers for Subjects 3A (above) and 3B (below). The test sentence was *Chunchon-e yeorum-un maeu teowoyo.* ('It's very hot in summer in Chunchon.')

As seen in Figure 3, the peak values for subject 3A are 151 Hz (P1), 120 Hz (P2), 110 Hz (P3), and 98 Hz (P4), while those for subject 3B are 157 Hz (P1), 122 Hz (P2), 113 Hz (P3), and 93 Hz (P4). This pair of brothers shows very similar pitch contours. Figure 4 shows the average pitch values for each of five pairs of brothers.



	1	2	3	4	5
P1	113	117	130	168	153
P2	136	140	142	181	169
P3	88	95	108	131	119

Figure 4. Comparison of pitch variation in five pairs of brothers

3.3 Formant frequencies

Figure 5 shows that the mean value of formant frequencies in syllable initial position was fairly similar in pairs of brothers. In addition, as can be seen in Table 2, the mean value of pitch in isolated word was also similar in most cases. The mean values of formant frequencies in the same pair was very consistent. It should be noted, however, that the mean values of F1 were more consistent than those of F2 and F3 throughout the data. This means that F2 and F3 could be more reliable for speaker verification.

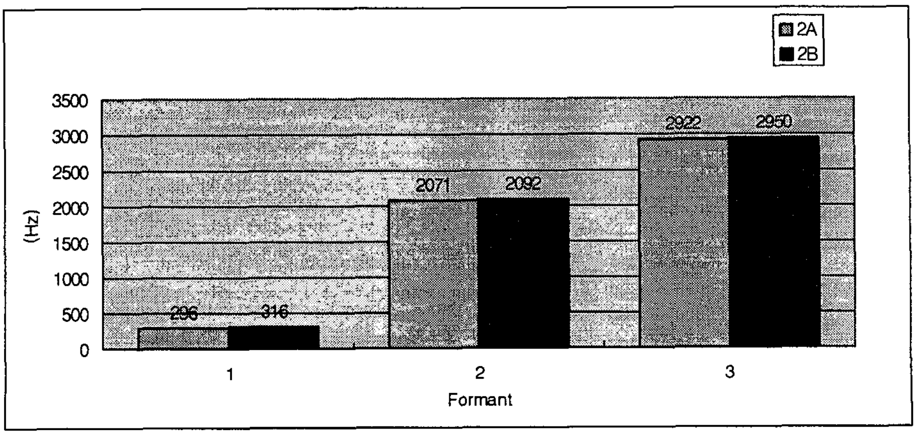


Figure 5. /i/ in iyagi in context

Table 2. Pitch and formant frequencies in five pairs of brothers

	Pitch Mean	SD	F1 Mean	SD	F2 Mean	SD	F3 Mean	SD
Subject 1A	100	1.5	334	5.5	1,091	39.5	2,532	19.2
1B	99	1.5	337	13.6	1,118	31	2,498	32
2A	101	2.1	333	11.4	1,399	11.1	2,319	21.1
2B	102	1.2	343	7.4	1,410	25.6	2,346	33
3A	112	0.6	280	4.4	1,160	22.4	2,291	15.3
3B	111	1.5	281	22.3	1,167	23.6	2,299	23.6
4A	148	1	289	11.9	1,623	20.3	2,369	23
4B	138	1.1	303	18.4	1,585	13	2,384	22.7
5A	128	1.5	300	14.6	1,615	27.8	2,403	13.4
5B	131	1	304	20	1,595	21.3	2,376	12.8

As in Figure 6 and 7, the average values of F1 and F2 in isolation and context were similar in each pair of brothers in most cases. In examining formant frequencies, the initial vowels are analyzed because the data of initial vowels was much more similar and homogeneous than those of vowels in other positions. In many cases vowels in other than initial position was not acoustically similar. The result was found in the study of voice similarities between sisters (Ko, 2001).

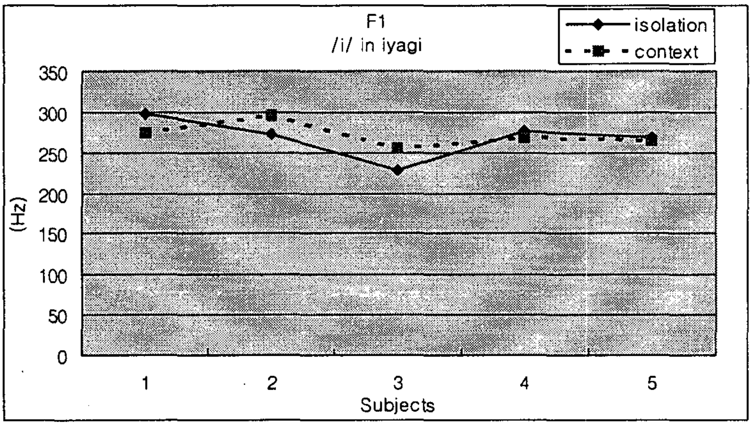


Figure 6. Comparison of F1 average in isolation vs. context

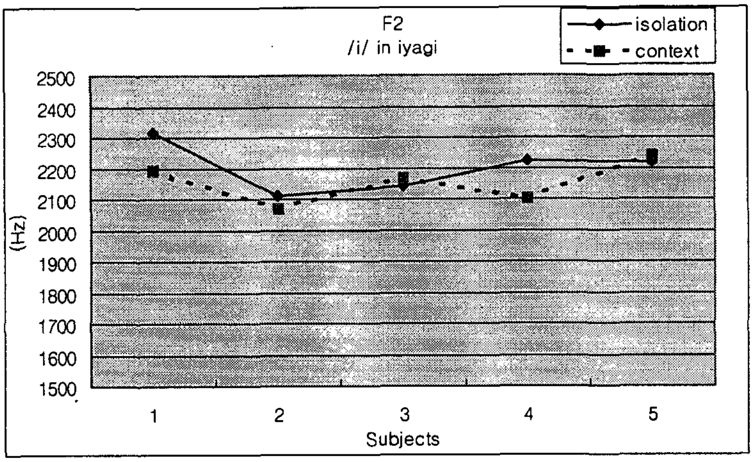


Figure 7. Comparison of F2 average in isolation vs. context /i/ in /iyagi/

3.4 Other parameters

Voice quality is a multi-dimensional component of fluent speech determined both physiologically and acoustically. Thus, there are many other parameters except formant frequencies, pitch and duration. As can be seen in Figure 8-9, there are several parameters providing the information of fundamental frequency, amplitude perturbation, noise, and voice irregularity. For example, F0 (average fundamental frequency), TO (average pitch period), Fhi (highest fundamental frequency), Flo (lowest fundamental frequency), PFR (phonatory F0 range), PER (total pitch period detected), Jitter (variation of frequency), Jita (Absolute Jitter), Shimmer (variation of amplitude). NHR (noise to harmonic ratio), and so on.

Comparing Figure 8 with Figure 9, overall shapes of diagrams are similar in some parameters. For example, The value of mean F0 for 5A was 137 Hz while that for 5B was 139 Hz. However, there are some differences in other parameters. For instance, the value of jitter for subject 5A was 3.1 % while that for 5B was 2.5%.

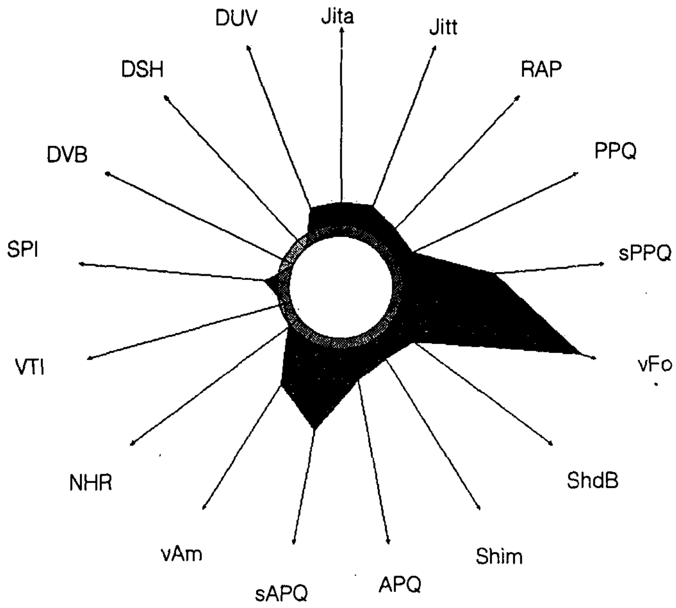


Figure 8. /ə/ in sənsɛŋnim 'teacher' using MDVP (Subject 5A)

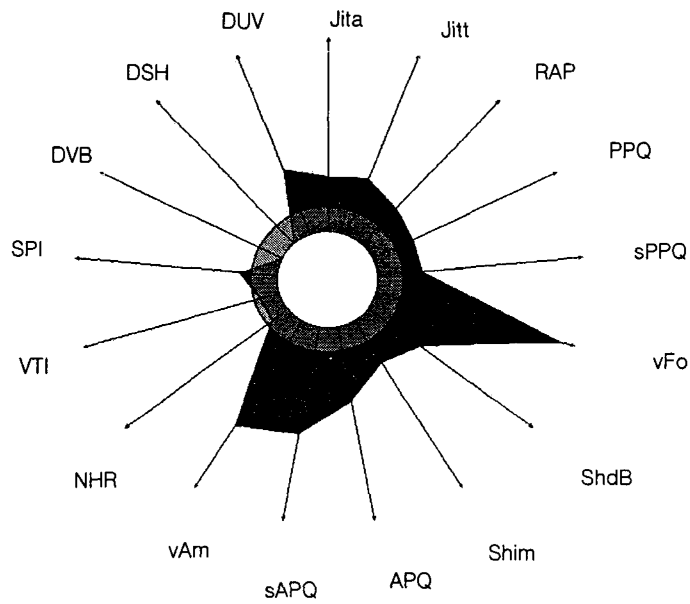


Figure 9. /ə/ in sənsɛŋnim 'teacher' using MDVP (Subject 5B)

As can be seen in appendix 1, MDVP shows 33 both acoustical and physiological parameters. It is suggested that all other parameters be taken into consideration in order to examine the voice similarities between brothers.

4. Concluding Remarks

It is well-known that the shape of an individual's vocal tract is partly genetic, and partly learned. In other words, uniqueness in voice is a product of both physiology and learning. The learned component of the equation could be called vocal habits. These would be items such as rhythm and rate of speech and vowel pronunciation. Thus, it is not a surprising fact that family members sound alike because they share genes and environments.

In this experiment, the authors focused on how pitch and formant frequencies play a role in distinguishing similar voices of five pairs of brothers because the two acoustic parameters are believed to be very important particularly in the fields of speaker verification and speaker identification. According to Kuwabara (2001), voice individuality has been found less sensitive to pitch than for formant manipulation. The current study was consistent with the above finding. Although this study was bound to comparisons of the two major parameters, it was concluded that duration, intonation patterns as a learned component could be other important cues for voice similarities. Finally, in order to examine the voice similarities. other parameters using the sound analyzing tools like MDVP should be taken into consideration.

References

Abberton, E. 2002. "Aspects of voice quality in women." *Proceedings (The 1st International Conference on Speech Sciences (ICSS)*, Korean Association of Speech Sciences, 128-131.

Ko, Do-Heung. 2000. "An Analysis of Phonetic Parameters for Individual Speakers." *Speech Sciences* (Korean Association of Speech Sciences), 7(2), 177~189.

Ko, Do-Heung. 2001. "Voice Similarities between Sisters." *Speech Sciences* (Korean Association of Speech Sciences), 8(3), 43-50.

Kuwabara, Hisao. 2001. "Voice Individuality and related acoustic Parameters." *Proceedings of ICSP 2001 (International Conference on Speech Processing)*, Vol. 2, 829-832.

Kushner, Rachel E. 1995. "Analysis and Perception of Voice Similarities among Family Members." ASA 130th Meeting, St. Louis, MO.

Yang, Byunggon. 2001. "Perceptual Experiment on Number Production for Speaker Identification." *Speech Sciences* (Korean Association of Speech Sciences), 8(1), 7-19.

Received: April 25, 2002
Accepted: May 15, 2002

▲ Do-Heung Ko
Division of Speech Pathology and Audiology
Hallym University
1 Okchon-dong, Chunchon
Kangwon-do, 200-702 KOREA
Tel: +82-33-240-1561 Fax: +82-33-256-3420
E-mail: dhko@hallym.ac.kr
Website: http://www.hallym.ac.kr/~dhko

▲ SunMee Kang
Department of Computer Science
Seokyeong University
#16-1 Chongnung-Dong, Sungbuk-Ku
Seoul, 136-704 KOREA
Tel: +82-2-940-7291 Fax: +82-2-919-0345
E-mail: smkang@skuniv.ac.kr
Website: http://ihci.skuniv.ac.kr

Appendix 1. Other parameters in word Seonsaengnim 'teacher' in isolation using Multi-dimensional Voice Program (MDVP)

Parameters	Name	Unit	Value	
			Subject 5A	Subject 5B
Average Fundamental Frequency	Fo	Hz	136.839	139.371
Mean Fundamental Frequency	MFo	Hz	133.157	138.503
Average Pitch Period	To	ms	7.51	7.22
Highest Fundamental Frequency	Fhi	Hz	163.942	163.296
Lowest Fundamental Frequency	Flo	Hz	96.944	116.635
Standard Deviation of Fo	STD	Hz	24.031	11.112
Phonatory Fo-Range in semi-tones	PFR		10	6
Length of Analyzed Sample	Tsam	s	1.622	1.772
Absolute Jitter	Jita	us	234.525	181.299
Jitter Percent	Jitt	%	3.123	2.511
Relative Average Perturbation	RAP	%	1.592	1.332
Pitch Perturbation Quotient	PPQ	%	1.914	1.431
Smoothed Pitch Perturbation Quotients	sPPQ	%	8.428	1.722
Fundamental Frequency Variation	vFo	%	17.562	7.973
Shimmer in dB	ShdB	dB	1.118	0.874
Shimmer Percent	Shim	%	10.19	7.551
Amplitude Perturbation Quotient	APQ	%	10.645	9.085
Smoothed Ampl. Perturbation Quotients	sAPQ	%	32.594	18.155
Peak-to-Peak Amplitude Variation	vAm	%	41.465	39.692
Noise to Harmonic Ratio	NHR		0.211	0.146
Voice Turbulence Index	VTI		0.037	0.049
Soft Phonation Index	SPI		28.599	20.514
Degree of Voice Breaks	DVB	%	0	0
Degree of Sub-harmonics	DSH	%	0	0
Degree of Voiceless	DUV	%	72.222	66.102
Number of Voice Breaks	NVB		0	0
Number of Sub-harmonic Segments	NSH		0	0
Number of Unvoiced Segments	NUV		39	39
Number of Segments Computed	SEG		54	59
Total Number Detected Pitch Period Periods	PER		69	90