The Effects of Voice Disguise, Made by Raising or Lowering Fundamental Frequency, on Other Formant Frequencies

The Digital Forensics Research Conference (DFRWS EU 2024)
19-22 March 2024
Zaragoza / SPAIN

Muharrem DAVULCU[1], Ayşe Zeynep AYDIN [2]

[1]Turkish Gendarmerie Forensic Department, M.Sc.
[2]Hacettepe University, Department of English Linguistics, Asst.Prof.

Türkiye Gendarmerie Forensic Department

Abstract

Voices, which are deliberately tried to be changed during the speech in order to hide their identities by the speakers, are increasingly being the subject of forensic investigations. Although voice disguise can be done in different ways, one of the simplest method is to change the fundamental frequency, that is, to make the sound low pitch or high pitch. In this study, in which F0, F1, F2, F3 and F4 parameters of normal and disguised recordings of 15 female and 15 male speakers were evaluated, it was aimed to find out which acoustic properties changed and which remained consistent in order to help detect disguised voices. For features affected by voice disguise, the goal is to determine whether these changes are systematic, whether the same change always occurs for a given change condition.

Introduction

With the development of telecommunication technology, the services and areas of activity provided by voice biometrics are also developing. The ever-increasing number of customer service calls, the increasing number of conversations parallel to mobile phone technology, and voice messages on social media are rapidly increasing. For this reason, voice recordings that are the subject of a crime or directly contain criminal elements, such as threats and insults, are increasing, and these voice recordings are becoming the subject of forensic investigations in increasing numbers every day. In voice recordings that are the subject of a crime or directly contain criminal elements, it is often seen that the speaker tries to disguise his voice in order to hide his identity. Voice disguising can be done in many ways, but one of the simplest methods is to change the fundamental frequency, that is, to make the sound low pitch or high pitch. Although various studies have been conducted on voices changed by this method, no study has been found that examines the interaction of all formant frequencies.

Research Aim and Objectives

In forensic voice investigations based on auditory and visual examination, the aim is to reveal how much the formant frequencies (F1, F2, F3 and F4) are affected by the change in the fundamental frequency (F0) in voice disguises made by making the sound low pitch or high pitch.

Our research questions that we will examine for this purpose are;
For Turkish phonemes /a/, /e/, /ɪ/, /i/, /o/, /ö/, /u/ and /ü/;

1. To determine whether other formant frequencies (F1, F2, F3 and F4) are affected by the change in F0,
2. If the change occurring in F0 also affects other formant frequencies (F1, F2, F3 and F4), to reveal whether this interaction is systematic or not,
3. If the change occurring in F0 also causes changes in other formant frequencies, the question is whether it can be determined which formant frequency of which phoneme is least affected by this change.

Methodology

The sample of the study consists of 15 female and 15 male participants who speak standard Turkish and have no articulation disorders. Before the research, a text with semantic integrity was prepared in which Turkish vowel phonemes were voiced at the beginning, middle and end of the word, and the words belonging to each phoneme were voiced three times in the text. Following the examination of the records, the process of obtaining the fundamental frequencies and formant frequencies of the phonemes was done with the Praat 6.1.53 software. By listening to the recordings with the mentioned software, frequency values were taken by hearing the phoneme, including 5 points for each phoneme, and by determining the place where the energy density of the phoneme was highest. Since each vowel phoneme was voiced 9 times in the text, a total of 153 formant frequency values, including 9 F0, 36 F1, 36 F2, 36 F3 and 36 F4, were determined for each phoneme and speakers' charts were created. In the charts created, there are a total of 2448 frequency values of individuals, 1224 of which are normal pronunciations and 1224 are modified pronunciations. A total of 73,440 frequency values were examined for 30 speakers. All frequency values were transferred to Minitab 19 software. These values were compared with paired sample t test.

Results

Table 1. F0 Means of Female Speakers and Disguising Methods

Nickname	F0 Mean (Normal Pronunciation)	F0 Mean (Disguised Pronunciation)	Disguising Method
FM1	295	312	High pitch
FM2	258	341	High pitch
FM3	254	328	High pitch
FM4	197	201	High pitch
FM5	228	381	High pitch
FM6	211	216	High pitch
FM7	273	386	High pitch
FM8	214	217	High pitch
FM9	270	267	Low pitch
FM10	247	243	Low pitch
FM11	250	222	Low pitch
FM12	213	184	Low pitch
FM13	182	179	Low pitch
FM14	250	344	High pitch
FM15	259	280	High pitch
Mean	240	273	10 High pitch 5 Low pitch

Table 2. F0 Means of Male Speakers and Disguising Methods

Nickname	F0 Mean (Normal Pronunciation)	F0 Mean (Disguised Pronunciation)	Disguising Method
M1	139	117	Low pitch
M2	133	136	High pitch
M3	169	143	Low pitch
M4	169	168	Low pitch
M5	172	415	High pitch
M6	153	170	High pitch
M7	190	344	High pitch
M8	158	338	High pitch
M9	168	149	Low pitch
M10	160	171	High pitch
M11	180	184	High pitch
M12	168	234	High pitch
M13	119	133	High pitch
M14	172	206	High pitch
M15	107	150	High pitch
Mean	157	203	11 High pitch 4 Low pitch

Of the 15 female speakers in total, 10 chose the high pitch method and 5 chose the low pitch method. The mean normal pronunciation fundamental frequency (F0) of 15 female speakers is 240. The mean disguised pronunciation fundamental frequency (F0) is 273. Of the 15 male speakers in total, 11 chose the low pitch and 4 chose the high pitch method. The mean normal pronunciation fundamental frequency (F0) of 15 male speakers is 157. The mean modified pronunciation fundamental frequency (F0) is 203 (Table 1-2).

Table 3. T Test P Values for All Female Speakers

T TEST P VALUES FOR ALL FEMALE SPEAKERS					
Phoneme	**FO NOR-FO DIS**	**F1 NOR-F1 DIS**	**F2 NOR-F2 DIS**	**F3 NOR-F3 DIS**	**F4 NOR-F4 DIS**
/a/	0	0,471	0,003	0,959	0,065
/e/	0	0	0,874	0,727	0,004
/ɪ/	0	0,799	0,002	0	0
/i/	0,002	0,078	0	0,001	0
/o/	0	0,005	0,376	0,291	0,203
/ö/	0	0,291	0	0,006	0,001
/u/	0	0,01	0,127	0,094	0,714
/ü/	0	0,002	0,579	0,154	0,732

When the T test p values of all female speakers are examined; It is seen that there is a significant difference in the F0NOR-F0DIS comparison. In the speaker-based evaluation, it was seen that not all speakers could change their fundamental frequencies at the same rate. When the values of all female speakers are compared, it is seen that there is a significant difference between the F0 values of normal and disguised sounds. The values obtained for other formant frequencies are presented in Table 3.

Table 4. T Test P Values for All Male Speakers

T TEST P VALUES FOR ALL MALE SPEAKERS					
Phoneme	**FO NOR-FO DIS**	**F1 NOR-F1 DIS**	**F2 NOR-F2 DIS**	**F3 NOR-F3 DIS**	**F4 NOR-F4 DIS**
/a/	0	0,02	0,045	0,609	0,012
/e/	0	0	0,374	0,413	0,013
/ɪ/	0	0	0,444	0,001	0,005
/i/	0,006	0	0	0,007	0,034
/o/	0	0	0,54	0	0
/ö/	0	0	0,188	0,09	0,016
/u/	0	0	0,014	0	0,021
/ü/	0	0	0,392	0	0,003

When the T test P values of all male speakers are examined; It is seen that there is a significant difference in the F0NOR-F0DIS comparison. In the speaker-based evaluation, it was seen that not all speakers could change their fundamental frequencies at the same rate. When the values of all male speakers are compared, it is seen that there is a significant difference between the F0 values of normal and disguised sounds. The values obtained for other formant frequencies are presented in Table 4.

Conclusion

It has been determined that the change in F0 affects all formant frequencies (F1, F2, F3 and F4) at different rates, this interaction is not systematic and the rate of influence varies based on person and phoneme. It is considered that the value least affected by the fundamental frequency change in male and female speakers is /e/ and /a/ phonemes at the F3 formant frequency.

Reference List

-Künzel, H.J., González-Rodríguez, J., & Ortega-Garcia, J. (2004). Effect of voice disguise on the performance of a forensic automatic speaker recognition system. *The Speaker and Language Recognition Workshop*.

-Masthoff, H. (2013). A report on a voice disguise experiment. International Journal of Speech Language and the Law. 3. 160-167.

-Perrot, P., Aversano, G., Chollet, G. (2007). Voice Disguise and Automatic Detection: Review and Perspectives. In: Stylianou, Y., Faundez-Zanuy, M., Esposito, A. (eds) Progress in Nonlinear Speech Processing. Lecture Notes in Computer Science, vol 4391. Springer, Berlin, Heidelberg.

-Zetterholm, E. Voice Imitation. A phonetic study of perceptual illusions and acoustic success. Dissertation. Department of Linguistics and Phonetics, Lund University. 2003.