Physica Macedonica **61**, (2012) p. 79-84

ISSN 1409-7168

FORENSIC SPEAKER IDENTIFICATION THROUGH COMPARATIVE ANALYSIS OF THE FORMANT FREQUENCIES OF THE VOWELS IN THE MACEDONIAN LANGUAGE

V. Pop-Dimitrijoska[1], G. Apostolovska[1], B. Gerazov[2], Z. Ivanovski[2] and J. Jovanovski[3]

[1]*Institute of Physics, Faculty of Natural Sciences and Mathematics, Gazi Baba bb, 1000 Skopje, Macedonia*

[2]*Institute of Electronics, Faculty of Electrical Engineering and Information Technologies, Rugjer Boshkovik bb, 1000 Skopje, Macedonia*

[3]*Faculty of Computer Science and Engineering, Rugjer Boshkovik 16, 1000 Skopje, Macedonia*

Abstract. The main objective of this study is forensic speaker identification from an incriminated recording. The identification was made through a comparative analysis between first three formants F_1, F_2 and F_3 of the voice samples from the questioned and suspects' recordings. The measurements were made with the PRAAT software, for each of the five vowels in the Macedonian language: ***a, e, i, o*** and ***u,*** which were isolated from the recordings. Used methodology of recording examinations employed in this research showed positive identification of the questioned voice. The forensic audio analysis still doesn't have its place in legal and the crime fighting systems in Macedonia. This is a sufficient reason to put a bigger accent on the research of this issue in the future that will contribute in solving many criminal cases which until now, because of the type of generally accepted evidence, were not resolved.

PACS: 89.20.Mn, 43.72.Uv, 43.72.Fx

1. INTRODUCTION

Voice identification is based on the theory that each person's voice is unique like his fingerprint and his DNA. But this analogy is not entirely accurate, given that the "voice print" is not as unique, distinctive and inimitable as a fingerprint. However, voice has individual characteristics that are a result of the organs for sound production, the shape of the vocal tract, the oral cavity, the way of pronunciation, the regional accent, and others, that enable the discrimination of voices, which is not absolute, but is expressed by a probability scale.

A very significant element of speaker identification studies is the analysis of the vowel system used by the speaker. Vowels are voiced segments with a formant structure. Formants are the distinguishing and meaningful frequency components of human speech. They are the resonant frequencies that individualize the speaker, as a result of the anatomical structure of his speech organs. During an analysis of the vowels, the most important is the relative position of the first, second, third, and sometimes fourth formant [5], [3].

1.1. Collecting voice samples and measurements

In this paper we demonstrated a procedure for conducting a voice identification analysis. From an allegedly incriminated recording, taken with a mobile phone, it is necessary to identify the person from the recording. There are 18 suspects whose voice samples have been taken and it is necessary to identify which of those suspects is the one from the incriminated recording, by performing comparative analysis of the formants, as the most important acoustic parameter.

Three voice samples from 18 female individuals, aged 27 to 42 have been taken. They all come from the same language area and speak the Macedonian language, which is also their mother tongue.

The digital recordings of the suspects and the incriminating recording are taken by a mobile phone, iPhone 4s, and they are in a MPEG-4 Audio File Format (.m4a). The recording of the speech samples was done in a quiet environment with short reverberation.

While taking the voice sample from the suspects, they were given a printed version of the speech from the incriminating recording and the text was read in order to capture the speed of the voice, its accent, dialect, etc. In this way the recordings of the suspects will more reliably reflect the dynamics of speech from the incriminating recording. The goal was to minimize the attempts to disguise the voice, by eventually adding false accent, a change in the speed of pronunciation and so on. The best technique for obtaining the prosodic features is the technique of repeating with the speaker [2].

A further analysis is conducted by using the PRAAT software, so vocal segments are carefully and meticulously selected, with duration of 0.02-0.10 seconds and the mean values of the formants are read. Some of the vocals were rejected from the analysis because of:

- the merging of vowels with other speech sounds,
- the difficulties in reading the formants in the vocal segments smaller than 0.02 s,
- the occurrence of noise or whisper when the vocal segments are located at the end of the sentence,
- the occurrence of discontinuities and jumps of the formants in the vocal segments etc.

Only 57 vowels from text were successfully isolated and measurements were made of the mean values of the first three formants F_1, F_2 and F_3 for each of the five vowels in the Macedonian language: ***a***, ***e***, ***i***, ***o*** and ***u***. This was done for all three recordings of the 18 suspects, including the recording of the unknown person, so exactly 513 vowels were measured for every suspect.

2. RESULTS

Because the closest suspect to the incriminating record is required, the distances between the measured values of the formants are calculated and a similarity in the distribution of the formants is examined through the following methods.

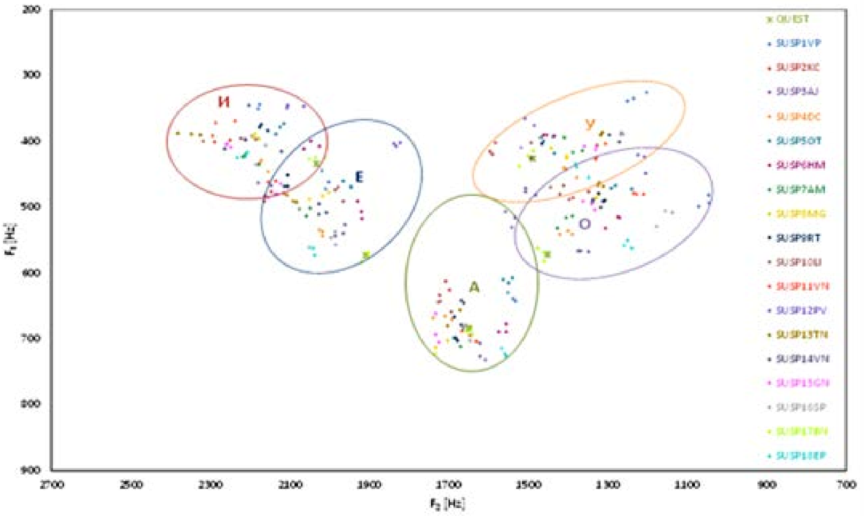

Fig. 1: Distribution of the mean values of formants from all audio samples.

Graphical presentation of formants distribution. Fig. 1 shows the distribution of the mean values of the formants for all the vowels, from all suspects including the unknown person. A most evident overlapping can be seen in QUEST and SUSP17BN. In Fig. 2 are shown distribution of the mean values of those two recordings in order to show their match more clearly.

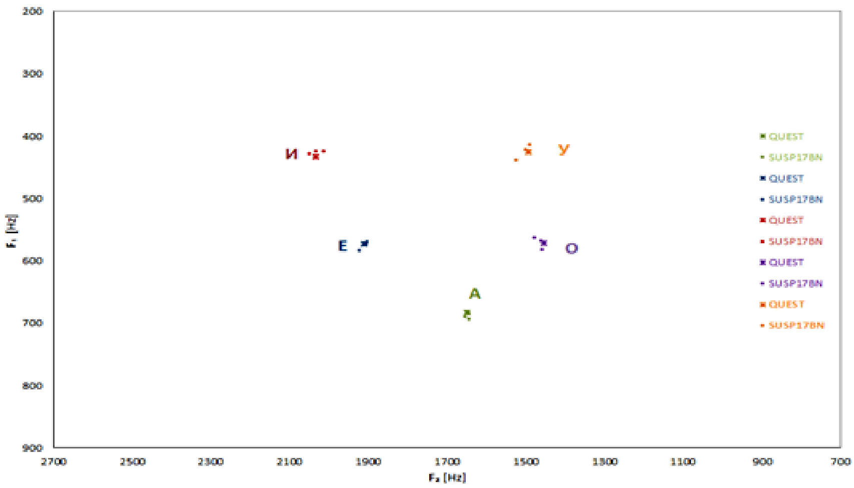

Fig. 2. Matching of the mean values of formants between QUEST and SUSP19BN

Fig. 2: Matching of the mean values of formants between QUEST and SUSP19BN.

Histograms. They show an important visual representation of the distribution of the data. For this purpose, normalized histograms for the formants F_1, F_2 and F_3, for all involved persons were calculated and visualized with Matlab. The widths of the intervals (bin-width) of the histograms are set 50 Hz for F_1, 100 Hz for F_2 and 100 Hz for F_3. The arithmetic means are marked with a red line, the standard deviations are marked with magenta, as follows: ± 1σ, where there are 68.2% of all the data and ± 2σ, where there are 95.4% of all the data. The histograms of the suspects are shown in blue, and the green shows the unknown person. From the histograms, a similarity in the distribution of the bins of the histogram between QUEST and SUSP17B can be observed, and it is the most distinctive in the histograms of the vowels ***e*** and ***o.*** Fig. 3 shows a histogram of the vowel ***e***.

Euclidean distances. From the measures for calculating distance, we used the Euclidean distance, which is selected as a representation for the shortest straight-line distance between the two points, given by the Pythagorean formula. The average value of the Euclidean distances for all vowels will be shown here, first in 2D, between the frequencies (F_1, F_2), and then in 3D, between the frequencies (F_1, F_2, F_3). The summarized results of the computed distances are shown in Table 1, from where it is obvious that the smallest distance is between the formants of QUEST and SUSP17BN.

Percentage of relative deviation*.* The idea is to calculate the relative deviations of the formants i.e. their relative errors, presuming that the two recordings which are compared belong to the same person. The result is expressed as a percentage. A low percentage would mean a slight deviation between the recordings and a possibility that they originate from the same person. The summarized results are shown in Table 1, from where it can be seen that the percentage of deviation is the smallest between QUEST and SUSP17BN, suggesting the possibility that they originate from the same person.

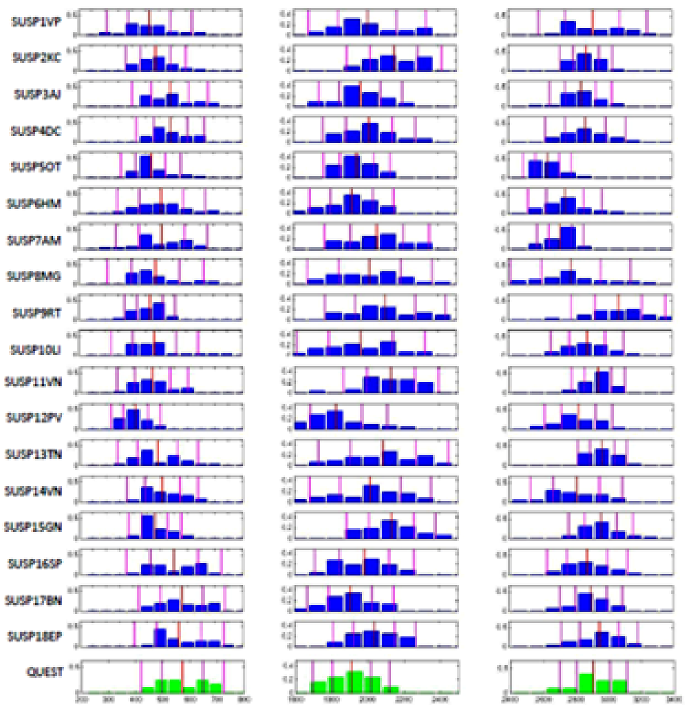

Fig. 3. Histograms of F_1, F_2, F_3, for the vowel *e*. There is evident similarity between formant distribution of QUEST and SUSP17BN

Automatic voice identification with the software - SIS II. In order to receive more reliable results, they are reviewed by using specialized forensic software for audio-analysis and speaker identification - SIS II, from the Russian company Speech Technology Center (STC). The software has a special plugins for speaker identification - Automatic Identification Plugins. The obtained results showed biggest similarity of QUEST with SUSP17BN and they are shown in Table 1.

Using the software for the purposes of this study was temporarily enabled by the company Speech Technology Center (STC) [6].

Table 1. Summarized results of the comparison between formants

Suspects Code	Euclidean distances measured between (F_1, F_2) [Hz]						Euclidean distances measured between (F_1, F_2, F_3) [Hz]						Relat. deviat.	Similarity calculated with SIS II
Vocals	*a*	*e*	*i*	*o*	*u*	Σ/n	*a*	*e*	*i*	*o*	*u*	Σ/n		
SUSP1VP	159	195	238	408	307	261	258	253	316	450	357	327	12.5%	49.3%
SUSP2KC	130	274	264	166	197	206	248	303	305	224	243	264	7.4%	67.2%
SUSP3AJ	88	110	200	136	130	133	189	168	246	159	190	190	4.0%	69.5%
SUSP4DC	122	138	188	111	120	136	315	218	223	156	189	220	4.1%	66.4%
SUSP5OT	131	153	161	221	156	164	409	318	244	311	302	317	10.1%	53.2%
SUSP6HM	107	133	119	190	88	127	288	239	174	296	326	265	6.5%	73.7%
SUSP7AM	89	190	218	100	131	146	334	291	285	204	274	278	6.4%	65.4%
SUSP8MG	105	187	199	189	186	173	259	308	285	255	275	276	6.9%	62.1%
SUSP9RT	94	241	244	177	91	170	211	339	351	340	204	289	7.6%	63.0%
SUSP10LI	94	190	202	131	73	138	154	243	250	171	140	192	5.6%	66.7%
SUSP11VN	147	275	306	255	188	234	218	303	366	265	214	273	8.3%	56.9%
SUSP12PV	218	217	163	286	203	217	265	258	254	320	298	279	13.1%	55.1%
SUSP13TN	115	231	335	188	180	210	183	267	402	221	233	261	8.1%	66.1%
SUSP14VN	90	177	213	188	218	177	206	248	295	254	335	268	7.3%	68.8%
SUSP15GN	106	263	237	147	157	182	207	296	287	233	273	259	7.5%	74.5%
SUSP16SP	65	143	174	320	231	187	188	211	251	444	407	300	7.8%	61.7%
SUSP17BN	32	48	57	43	40	44	63	68	103	72	87	78	0.6%	98.8%
SUSP18EP	132	156	228	215	224	191	198	206	272	228	296	240	4.9%	50.4%

3. CONCLUSIONS

The performed aural and spectrographic comparative analysis, and the results obtained with the different statistical approaches indicate a reliable identification of the unknown person from the incriminating recording, with the person SUSP17BN.

This research demonstrates an attempt of forensic speaker identification, through a comparative analysis of the formants and its aim is to show that under controlled conditions, without a professional and specialized hardware and software equipment, such an identification is possible, based on the analysis and measurement of one of the most important acoustic parameters.

The study identifies the problems of inferring identity from speech in forensics. So when conducting voice identification analysis, we must consider the following problems:

- the complexities of voice sample comparison;
- the probabilistic nature of the technique;
- the difficulties introduced by differential variation within and between voices [1];
- the necessity for both acoustic and auditory comparison, and for the comparison of linguistic and non-linguistic features;
- the expertise required in several areas related to speech science, such as: linguistics, acoustics, statistics and of course forensics.

This research is the first of its kind in the Republic of Macedonia and it includes all recommended norms, technical standards or protocols that are practiced in such expertise and are accepted worldwide [4].

REFERENCES

[1] P. Rose, *Forensic speaker identification*, Taylor & Francis, 2002

[2] M. Heđever, *Uvod u forenzičnu fonetiku i akustiku*, skripta, Sveučilište u Zagrebu, 2009

[3] *Handbook of forensic science* - Jagiellonian University, Institute of Forensic Research, Poland, 2008

[4] S. Cain, *American Board of Recorded Evidence - Voice Comparison Standards,* The Forensic Examiner, April 1999

[5] K. Klus, A. Trawiñska, *Forensic speaker identification by the linguistic-acoustic мethod in Keú and Ies*, Problems of Forensic Sciences, vol. LXXVIII, 160–174, 2009

[6] http://speechpro.com/product/forensic_analysis/ikarlab - SIS II documentation.