

UTILIZATION OF VOICE ANALYSIS METHOD IN CRIMINAL INVESTIGATIONS

Jadranka Otašević, PhD

Faculty of Special Education and Rehabilitation, University of Belgrade, Serbia

Saša Atanasov, PhD

Faculty of Law, University of Priština, Kosovska Mitrovica, Serbia

Gordana Čolić

Abstract: Contemporary technological and educational achievements enable the emergence and development of completely new identification methods based on individual physical characteristics of persons. Today, forensic voice identification is a powerful tool in the fight against crime, in situations where it is necessary to identify a suspect or to acquit an innocent person. Just as there are no two identical fingerprints, two identical retinas, or two identical handwritings, there are no two identical voices. Based on 20-30 seconds of effective speech, forensic experts can perform a voice expertise and identify with high reliability whether a particular voice belongs to a particular person. Comprehensive analysis results in a large amount of data, so by using the statistical analysis method, the necessary forensic markers for forming the final expert opinion are obtained. In this type of expertise, interdisciplinary approach is applied in which several scientific disciplines are included - linguistics, phonetics, acoustics, speech-language therapy, psychology, mathematical statistics, law and criminalistics. It is not only the voice of the speaker that undergoes a detailed analysis, but also the speech on the whole. This means that not only the voice quality (loudness, pitch, harshness, nasality), but also the characteristics of language (Morphology, Semantics, Syntax and Lexis) are examined.

Keywords: crime, voice analysis, forensic expertise, identification.

INTRODUCTION

By listening and analyzing of voice recordings, based on series of specific speech characteristics, forensic experts are able to make a suspect's profile and establish their identity. Based on detailed analysis, it is possible to determine the

part of the country in which the suspect was born, lived or went to school, create their socio-linguistical profile (verbal capacity, education background, colloquial speech), as well as psycho-linguistic profile (temperament, expressiveness, composure). These are the actual facts that provide full scientific meaning to the Socrates' quote "Speak, so that I may see you".

Uniqueness and singularity of each human voice happens to be the result of a long path of evolution lasting for millions of years and standing as one of survival factors of human kind. Voice of a particular person is determined by a person's anatomy and physiology (Chermak, Bellis & Musiek, 2007). Through voice, one can express deepest emotions and complex intellectual thoughts. It is the voice we use to choose our friends or partners and randomly create an image of the speaker. Voice is actually a designation of our identity, just as there are no two identical fingerprints, handwritings or eye retinas, there are no two identical voices. Identifying a particular person based on their voice is called forensic speaker identification. Today, this form of identification is a powerful tool against various forms of crimes. It is used for identification of voices recorded in kidnapping, drugs manufacturing and dealing, threats, extortions, human trafficking, war crimes being recorded on a video tape and other criminal actions (Gibbons, 2014). The main task is to listen to the recorded voice and perform detailed analysis of the speech signal and consequently identify array of speaker-specific information serving as strong forensic "markers" to confirm the identity of a perpetrator with high reliability. There are several things which may serve as forensic markers: some form of speech pathology, use of filler words or characteristics and specificities of each particular voice. There have been frequent demands to identify whether someone disguised their voice, closed their nose, put a handkerchief over mouth or was impersonating someone. Voice record analysis depends on the recording quality, conversation duration, whether it was masked by any background noise coming from other sources (traffic noise or coffee shop hum). Distortions and background noise not only mask certain particularities of the speech signal, but influence the alterations in speech characteristics. Psychoactive substances like alcohol and drugs can also cause significant changes in speech characteristics resulting in misleading conclusions when such voice is being compared with a disputable sample.

Development of technical systems for voice recordings analysis and invention of special software for voice comparison have significantly improved the identification based solely on voice identification. However, although advanced software programs can be applied, human factor which comprises knowledge and experience of experts and professionals still remain irreplaceable (Cooper, 2005). According to the international standards, its specific expertise can be executed solely by an expert holding academic degree in the field of speech along with some expertise in the field of electro acoustics. In Serbia there are no legislative regulations, technical standards or protocols (or at least available to general public), resulting in executing of the expertise in accordance with the Code of the The International Association for Forensic Phonetics (IAFP).

PROCESS OF EMERGENCE OF VOICE AND SPEECH

Speech is a continuous process in time and its elements should be regarded as segments being parts of a wider whole. Thus, through analysis we will obtain fuller characteristics of isolated segments. In a detailed analysis, the expertise begins with the evaluation of a voice, followed by a syllable, a word or a sentence. Basic vocal sound is produced by air passing through the larynx, causing the vibration of vocal cords. The sound is generally perceived by the ear (but also by tactile and kinesthetic perception), which brings us to three basic characteristics of sound: pitch, loudness and quality. As a sound phenomenon, there are four processes of speech production: respiration, phonation, resonance and articulation. All these processes provide individual characteristics to each particular speaker. Speech originates from simultaneous and synchronized action of these processes. It is a group of modifications, blending and segmenting of sound and movement in time (Hedeber, 2012). All speech processes are time-related. However, some of these processes “happen in time and others change over time” (Škarić, 1991; 213). If we study speech, in particular its specificities, it is necessary to convey a research of spectral, intensity-related and time-related components of speech. We may assume that the majority of speech specificities are being manifested *inter alia* through specific changes of certain time segments. Sometimes, these changes are visible only in “macro-segments” (prolonged time sequences, in duration of several minutes) and sometimes in “micro-segments” of speech, lasting only for milliseconds. These changes are not always relevant and we sometimes fail to perceive it, but its identification and analysis may contribute to finding the perspective of specificities of the speech inquired. Each speech deposition consists of components related to content (linguistic component) and form (manner of respiration, phonation, resonance and articulation). Both components are present in final realization of speech, while careful and precise acoustic analysis may determine which processes were involved in forming particular segments of speech signal (Jessen, 2008).

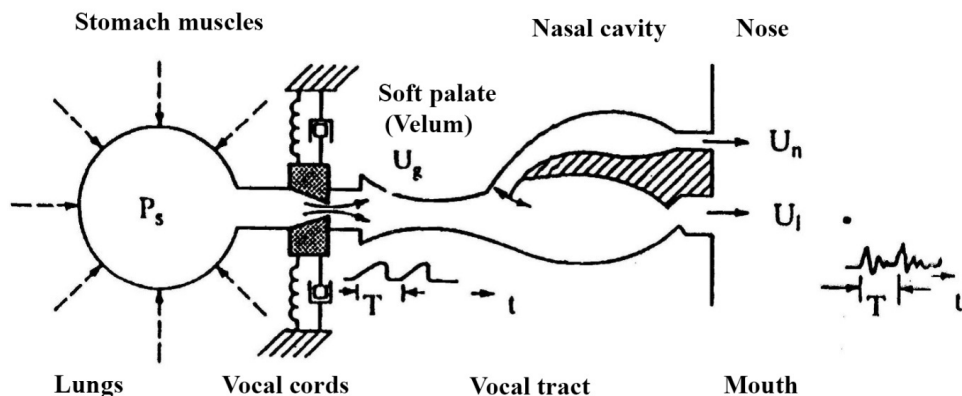


Fig.1. Speech Mechanism Scheme

Theoretical foundation of vocal identification is based on presumption that each individual voice is distinct and different from any other, which can be confirmed by voice analysis. There are two basic factors which make each voice unique during the speech process. First factor is the size, shape and connection of vocal cavities (throat, nose, and mouth) and the length and tension of vocal cords. Cavities of the vocal tract act as resonators (similar to pipe organ's tubes) amplifying certain tones produced by the vocal cords. Thus, typical sounds are created which is being registered as a print in a voice analysis. The other factor which gives unique characteristics of speech process of each individual speaker are the acquired motor patterns which directly involve articulation organs (lips, teeth, tongue, soft palate, jaws) in speech production. Humans acquire these motor patterns through process of learning and impersonating other people from their immediate surrounding and they are singular features of each particular individual (Bell-Berti & Hariss, 1982).

FORENSIC SPEAKER RECOGNITION

Speaker identification based on vocal analysis may be defined as a combination of auditive (hearing) and spectrographic (instrumental) comparison between one or more familiar voices with the voice unknown, in order to either identify or eliminate the investigated individual. The onset of voice analysis method was first encountered in America in the late 1940s when it was used for military intelligence purposes. The intense use of voice analysis technique for forensic purposes began in the late 1960s at the time when it was officially introduced by the Police forces in America as one of the identification methods. Forensic speaker recognition comprises comparison of familiar voice of a suspect with the unfamiliar voice with aim to identify or eliminate the suspect as a perpetrator. The specificity of forensic recognition lies in mandatory precondition of conditional cooperativeness of undisputed person and the uncontrolled conditions of acquired recording of a disputed voice. In such recognition procedure, elements of recognition in the form of verification and identification stand out clearly (Jovičić, 2009). From the forensic point of view, the process of individualization of a person swings from general population towards the individual, while the speed of reduction of its characteristics depends on disputable and undisputable features standing at a disposal of a forensic investigator.

Biometric technology is based on analysis of range of human features, such as physiological, biological and behavioral. Voice is characteristic in biometrics as it comprises physiological and behavioral features. Biometric methods of vocal analysis are based on non-linguistic information, i.e. acoustic features of speech being characteristic for each particular speaker and are interpreted and used accordingly. What is characteristic for forensic methods is that they use all available markers which can be registered in a speech signal, as well as linguistic infor-

mation comprised in each voice message (Kašić&Đorđević,2009). Besides, it is utterly important to emphasize that the quality of voice recordings in a forensic recognition depends on many unpredicted and uncontrolled conditions, while the recording cannot be repeated. Text context which is being investigated is also out of control of a forensic investigator (they receive the content to be analyzed and processed). Duration of the voice recording is also an important element for quality identification. One should bear in mind that speech expression is determined by a combination of variables each having its own degree of freedom (Jovičić, 2001). This triggers a problem that speech expression spoken out consecutively by the same speaker is not the same. In general, we can say that variations in speech expression can be recognized as voice variations and speech variations (Coulthard, Johnson & Wright, 2017). So, we can speak of intra-speaker variations based on inherent characteristics of each speaker, being the main marker of individuality, but are subject to changes (hoarseness as a result of age, respiratory infections, and vocal cords' surgery). Intra-speaker variations at the context level (phonetic and linguistic level) also include variations of emotional nature (Jovičić & Kašić, 2009). The essential thing is to discriminate speech characteristics of two speakers and to determine criteria for finding similarities between these two speakers with more or less probability. The essence of forensic speaker recognition is to actually determine intra-speakers variations and intra-speakers distances and reach a decision on similarities between these two voices based on matching zone of the two statistical divisions of the features. This assessment can be given partially for each particular feature. The final estimation is given after considering the complete milieu of observations. At this point, knowledge and experience of forensic investigators stand out, as they are able to integrate all the results of analysis and circumstances involved (Rose, 2002). Reaching final evaluation is performed upon Voice Comparison Standards of the International Association for Identification passed in 1991 and extended in 2000 by the Forensic Science Service, based on needs and experience from practice. Formulated scale of verbal evaluation is applied in most countries in Europe, including Serbia at its Forensic Acoustics and Phonetics Laboratory of the Life Activities Advancement Center as shown in Table 1 (Association of Forensic Sciences Providers, 2009).

Table 1. Vocal Identification Verbal Scale Grades

IAI standards (1991)	Forensic Science Service (2000)	Acoustic Phonetics and Forensics Laboratory Belgrade
Complete identification	Very strong likelihood	Very high degree of likelihood
-	Strong likelihood	High degree of likelihood
Probable identification	Moderately strong likelihood	Significant degree of likelihood
-	Moderate likelihood	Certain degree of likelihood
Possible identification	Limited likelihood	Likelihood exists and person cannot be excluded
Inconclusive case	Inconclusive case	Inconclusive case

Possible elimination	Limited unlikelihood	Unlikelihood exists and person cannot be excluded
-	Moderate unlikelihood	Certain degree of unlikelihood
Probable elimination	Moderately strong unlikelihood	Significant degree of unlikelihood
-	Strong unlikelihood	High degree of unlikelihood
Complete elimination	Very strong unlikelihood	Very high degree of unlikelihood

Types of Forensic Voice Investigation

Nowadays, several voice analysis methods for speaker identification are used. Purpose of each applied method is to maximize the accuracy of identification. Most common methods are: spectrogram analysis, auditive-linguistic analysis and auditive-instrumental analysis (hybrid analysis) which includes use of computer programs.

Spectrography

This method provides a visual image – voiceprint based on acoustic sound analyzers. Visual image is called spectrogram. Spectrography is based on Fourier transform of sound recording, representing the transformation of a sound from time domain into frequency domain. Periodical sound signal is decomposed into its constituent frequencies – accordions (Author, 2011). Thus, sound may be analyzed based on distribution of constituent frequencies determined by its amplitudes (loudness). This method requires adequate program support (Heđever, 2009).

This method enables detection of inter-speaker differentiations, while inter-speaker variables are not taken into consideration, which seriously compromises the validity of identity identification. These are the prevailing reasons for the standpoint that spectrography may be useful in a wider auditive-instrumental method.

Auditive-linguistic analysis

This method is used for analyzing features in speech expression which indicate speaker's identity in acoustic, linguistic and phonetic domain. The analysis is performed by prolonged listening and hearing speech expressions, resulting in identification of speaker's characteristics. Forensic investigator finds the duration of analyzed voice recording extremely important, as more material provides more diverse feature characteristics of the speaker. This analysis is being conducted depending on disposable voice material which needs to be analyzed. We have a

situation with two voice recordings, one of a perpetrator, the other of a suspect. In this case, forensic investigator can independently perform analysis and reach elements necessary for identification. The other situation is when there is just one voice recording, recording of a suspect. In this identification method, role of a victim or a witness who tries to remember the voice they have heard remains crucial. The identification is being done by voice line-up. Voices of witnesses may be lined-up in two tactical methods: standard, simultaneous method and sequential method, done in sequences. In the first case, witness will hear all the voices and then they will be asked to declare if they have recognized the perpetrator's voice among them. In the second case (sequential voice line-up) the witness will be asked to immediately declare whether the voice belongs to the perpetrator or not, right after hearing the particular voice. In a sequential voice line-up, voices are not being presented simultaneously, but sequentially (Simonović, 1999). The identification of a person by their voice should be selective. Voices line-ups (both suspects' and fillers') should be similar in its characteristics (phonetic, linguistic and acoustic) to the extent that the voice of a suspect does not stand out individually. The sole act of identification is performed in such a manner that the witness, having described the voice of a perpetrator, enters into separate room from which the identification is conducted. There they receive instructions and explanations on the identification procedure (for example, they are explained that the individuals they are about to identify cannot see or hear the witness, instructed not to hurry and, in case they make an identification, they should point at the person and explain the features they recognized the suspect by). This preliminary instructions should be carefully executed, all witness' questions must be answered, due to their high stress level before and during the identification process. After the witness answers affirmatively that they understood the identification procedure, the administrator opens a door to the next room with line-up and asks the witness the following question: "Does among the voices you are about to hear happen to be the voice which you have described earlier or is it not the case?", which is a factual commencement of the identification procedure. When it comes to practical experience, in Serbia and some countries in the region, the witness is usually being presented with a voice of a suspect along with three or four more voices of innocent citizens (Author, 2016).

Although there are opinions that, due to its sensitivity, auditive line-up should rather be considered as indicial evidence, the court practice stands on a point that the result of identification represents the evidence at the trial and serves as a ground for court verdict (Ilić, P.G. et.all., 2013).

Auditive-instrumental analysis and voice analysis programs

This contemporary method is used to increase the power of objective voice identification and suppress the impact of subjective factor as much as possible. Man and computer are two inseparable components, as it is the human factor

that performs selection of parameters and speech segments which are subject to comparison, as well as numeric interpretation and explanation.

PRAAT program enables monitoring of voice features in time and frequency domain, process of synthesis and analysis of voices, as well as speech signal manipulating. The program is able to process downloaded audio signal by filtrating it or distinguishing some specific frequency segments, marking audio recording segments which are parts of sentences, words or syllables. Program is equipped to analyze wave sound by frequency, amplitude, duration, loudness, jitter, etc. It also calculates and graphically shows pitch of voice, intensity of sound, spectrogram, jitter and shimmer.

MDVP Program (Multi Dimensional Voice Program) by KayPENTAX is highly suitable for voice forensic investigation as it has the option to follow-up and calculate significantly larger amount of parameters than the abovementioned PRAAT. Among other things, MDVP can give us the data on respiratory and aerodynamic parameters of a voice through numerical, graphic and image displays. Using Kay Elemetrics Corp. Model 4337, while singing vowel A, out of 34 total voice parameters which can be monitored by this model, 19 were monitored, as shown in Fig. 2 and 3.

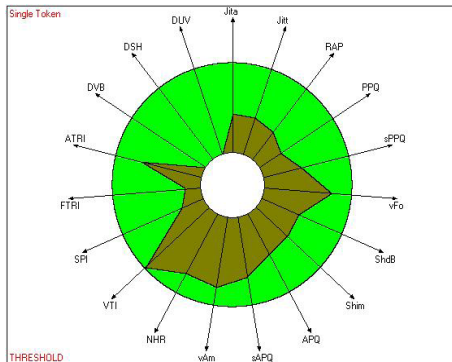


Fig. 2 Male voice singing vowel A

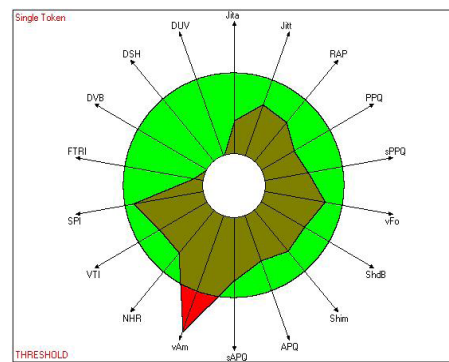


Fig.3 Female voice singing vowel A

CONCLUSION

Voice forensic investigation in identification purposes is important, because in the past there have been many cases in criminalistic and court practice that the voice or speech recording of one or more individuals happened to be the only clue and evidence of the criminal act. Regardless of the court, this forensic discipline became a powerful tool in the hands of police investigations. Recent research studies support the idea of fully automated speech analysis, aiming at future elimination of subjective impression of an investigator performing the analysis and reaches conclusions. Practical experience has shown that, for the time being, this option remains just a theory. The issues of perspective and problems of voice

forensic analysis open new fronts and challenges to voice analysis forensic investigators which needs to be solved in the future, in a manner which will determine the prosperity of their profession.

REFERENCES

1. Atanasov, S. (2016). Uloga svedoka u otkrivanju i dokazivanju krivičnog dela. Doktorska disertacija. Univerzitet u Prištini - Kosovska Mitrovica: Pravni fakultet.
2. Association of Forensic Sciences Providers.(2009). Standards for the formulation of evaluative forensic science expert opinion. *Science and Justice*, 49, 161-164.
3. Bell-Berti, F. & Hariss K. S. (1982). Temporal patterns of coarticulation: Lip rounding. *Journal of the Acoustical Society of America*, 71, 449-454.
4. Chermak, G. D., Bellis, T. J. & Musiek, F. E. (2007). Neurobiology, Cognitive Science, and Intervention. In: *Handbook of (Central) Auditory Processing Disorder: Volume II - Comprehensive Intervention*, San Diego, USA: Plural Publishing Inc.
5. Cooper, A.J. (2005). The Significance of the Serial Copying Management in the Forensic Analysis of Digital Audio Recordings. *The International Journal of Speech Language and the Law*, 12(1), 49-62.
6. Coulthard, M., Johnson, A. & Wright, D. (2017). *An introduction to Forensic Linguistics: Language in Evidence*. Routledge.
7. Gibbons, J.P. (2014). *Language and the Law*. Routledge.
8. Heđever, M. (2009). *Uvod u forenzičnu fonetiku i akustiku*, Skripta iz kolegija forenzična fonetika i akustika. Zagreb: Edukacijsko rehabilitacijski fakultet.
9. Heđever, M. (2012). *Osnove fiziološke i govorne akustike*. Zagreb: Edukacijsko rehabilitacijski fakultet, Sveučilište u Zagrebu.
10. Ilić, P.G., Majić, M., Beljanski, S. & Trešnjev, A. (2013). *Komentar Zakonika o krivičnom postupku*, četvrto izdanje. Beograd: Službeni glasnik.
11. Jessen, M. (2008). Forensic Phonetics. *Language and Linguistic Compass* 2/4 (pp.671-711) Wiesbaden, Deutschland: *Bundeskriminalamt*.
12. Jovičić, S.T. (2001). Forenzički aspekti prepoznavanja govornika, *Nauka Tehnika Bezbednost*, 1, 41-60.
13. Jovičić, S.T. (2009). Forenzičko prepoznavanje govornika: između teorije i prakse.
14. I. naučni skup sa međunarodnim učešćem: *Pravo i forenzika u kriminalistici* (pp.324-332) Kragujevac, Srbija: Kriminalističko policijska akademija Beograd.

15. Jovičić, S.T. & Kašić, Z., (2009). Intra-spikerske varijacije u govoru: forenzičke implikacije, *International Conference on Applied linguistics*, Novi Sad, Srbija.
16. Kašić, Z. & Đorđević, J. (2009). About linguistic expertise in crime solving. Proceeding of First conference with international participation: *Law and forensics in criminal law*, (pp. 333-340) Belgrade: Academy of Criminalistic and Police Studies.
17. Nešić, L., Kovačević, J. & Stevović-Otašević, J. (2011). Forenzička analiza govornog signala, *Pravni život*, 60 (5-6), 87 – 102.
18. Rose, P. (2002) *Forensic speaker recognition*, London and NY: Taylor & Francis.
19. Simonović, B. (1999). Predočavanje radi prepoznavanja (sudsko psihološki i kriminalističko taktički aspekti), *Bezbednost*, 4, 501-512.
20. Škarić, I. (1991). Fonetika hrvatskoga književnog jezika. In R. Katičić (Ed.), *Povijesni pregled, glasovi i oblici hrvatskoga književnog jezika* (pp. 61-372), Zagreb, Jugoslavija: Hrvatska akademija znanosti i umjetnosti i Globus nakladni zavod, Zagreb.
21. Voice Identification and Acoustic Analysis Subcommittee of the International Association for Identification. (1991). Voice Comparison Standards. *Journal of Forensic Identification*, 41(5), 373-392.