The Complexity of Speaker Identification based on Voice and Speech and its Application in Forensics

Dr. Mia Šešum & Dr. Ema Petrović,
Department of Special Education and Rehabilitation of the Deaf and Hard of Hearing Persons, Faculty of Special Education and Rehabilitation, University of Belgrade, Belgrade, Serbia & Department of Oriental Studies, Faculty of Philology, University of Belgrade, Belgrade, Serbia

Abstract

The voice and speech of every human being are determined by their anatomical and physiological characteristics, as well as by the speech habits acquired during life. From a forensic point of view, voice and speech are inextricably linked, as voice production is a necessary precondition for speech production. The theory of speaker identification is based on the premise that, given the complexity of the speech system and learned speech behavior, no two people can have an identical voice and speech. Since voice and speech are mutable on conscious and unconscious level, concerning both internal and external factors, and are also traceable through the temporal dimension, speaker identification is considered the most complex discipline within forensic science. The article systematically presents the methods used in the past and those used today to identify speakers, focusing in particular on their advantages and disadvantages. The literature search was carried out in the databases of the Serbian Library Consortium and in relevant internet search engines. The literature review has shown that the rapid development of technology in the 21st century has significantly influenced forensic phonetics and increased the need for its application. However, this has not improved the reliability of the results, as the complexity of speech expression cannot yet be solved by automated speaker identification. Therefore, it is necessary to carefully select and train experts working in this field, in order to ensure reliable results relevant to the needs of court proceedings.

Keywords: Speaker Identification, Auditive-Instrumental Method, Voice and Speech, Spectography, Automatic Method

1. Introduction

Sound is the wave motion that creates the impression of hearing in the ear through the vibration of molecules at a certain frequency. Speech is a signal composed of numerous sounds that are combined into a meaningful whole and serve the purpose of enabling communication between people (Karakoç & Varol, 2017). To produce the human voice, which is the basis of speech, the organs of the digestive and respiratory tracts must be involved (Šešum, 2021). In general, there are two types of sounds produced by the human voice: Speech and non-speech sounds (laughing, crying, screaming, coughing, etc.) (Karakoç & Varol, 2017).

People use language to convey information. Through speech, people express opinions, desires, intentions, and emotions, and it is perceived thanks to the ear and brain's ability to receive sound waves (Alkhatib & Kamal Eddin, 2020). The speaker encodes the message into a continuous, variable waveform that can be preserved, manipulated, and

transmitted during speech production. When this encoded message reaches the listener, it is decoded (Islam et al., 2022).

Forensic phonetics is considered a newer discipline within the forensic sciences. It applies phonetic knowledge to solve criminal cases. The most important aspect of forensic phonetics is speaker identification, i.e., determining the identity of the perpetrator based on their voice and speech (Didla, 2020). Due to their anatomical and physiological nature, voice and speech are biometric evidence (Sharma & Sahu, 2018). Biometrics refers to individual human characteristics, and these characteristics can be used to determine a person's identity (Singh et al., 2018). Like other biometric forensic methods, the goal of speaker identification is to determine the identity of the speaker based on unique personal characteristics (Tsanas et al., 2017). Each person's voice and speech are unique and non-reproducible due to anatomical and physiological characteristics as well as speech habits acquired throughout life (Sharma & Sahu, 2018). Because of the characteristics of voice and speech resulting from these two aspects, it is possible to identify the speaker (Singh & Khan, 2016). Voice and speech are considered one of the most important biometric types of evidence as they play a role in a variety of criminal activities such as threats, extortion, kidnapping, drug trafficking, weapons, etc. Since speech is a verbal behavior, no isolated parameter is sufficient to determine the identity of the speaker; rather, they all combine to create a specific speech profile of the speaker (Braun, 2020). It can be said that forensic phonetics builds a bridge between language and law (Larner, 2015). It is highly interdisciplinary and requires a combination of different language skills and an understanding of voice and language phenomena from different perspectives.

Forensic phonetics is an important segment of forensics (Zhou et al., 2022). The goal of analyzing voice samples depends on the nature of the investigation or court case. A sound recording may contain various information, but its forensic significance depends on the circumstances of a particular case (Meluzzi et al., 2020). Most requests in the field of forensic phonetics are related to speaker identification (Mukattash, 2016). In addition to speaker identification, the field of forensic phonetics also includes the creation of a person's voice profile, the transcription of poorly intelligible recordings and the organization of "voice lines" (determining the identity of the attacker based on the victim's recognition of their voice) (Šešum, 2021). Foulkes & French (2001) found that 70-80% of requests in forensic phonetics concern the identification of speakers.

Since the goal of speaker identification is to determine whether the voice and speech belong to the same person (Leuzzi et al., 2016), the procedure requires comparing the speech of an unknown speaker (perpetrator) with the speech of a known speaker (suspect). More specifically, recordings made in the context of a criminal offence and used as evidence (questioned recordings) are compared with recordings of a suspect, usually made by experts under controlled conditions (suspect's recordings), to determine whether they have enough in common (Mukattash, 2016). Speaker identification is widely used in criminalistics, especially since the use of mobile phones has become an integral part of everyday life (Didla, 2020), which has led to an increase in the number of criminal cases in which recorded speech

appears as the only or one of the pieces of evidence. The ability to identify voice and speech is also relevant to many other security contexts (Jenkins et al., 2021).

The choice of method used to identify the speaker can significantly influence the results and therefore the final opinion of the expert. It is therefore important that experts are aware of the advantages and limitations of the different methods presented in this article, as well as the conditions that must be met for their application. Considering the fact that voice and speech forensics is considered a young forensic discipline, as well as the fact that despite the increasing presence of laboratories of this type, there is still no uniform standard for the application of the speaker identification method, we considered it important to provide an insight into the complexity of the entire procedure of forensic voice and speech processing, as well as the requirements and possibilities of the speaker identification method, based on the review of recent scientific literature.

2. Methodology

When reviewing the literature for the purposes of this article, a basic search was carried out using Google Scholar Advanced Search, Research Gate and the service of the Consortium of Libraries of Serbia for Uniform Cataloguing - KOBSON. The following keywords and phrases were used in the search in Serbian and English: Speaker identification, auditory-instrumental method, voice and speech, spectography, automatic method

3. Results

During the literature search, we found 59 articles that could be systematized according to 3 basic areas: Historical overview of speaker identification, Methods of speaker identification and Factors hindering identification. The articles were mainly research papers published in Serbian or English between 1985 and 2023.

4. Discussion

Due to the multidisciplinary nature of voice and speech forensics, research interest in this area of forensics has been aroused in various scientific disciplines, such as: philology, special education and rehabilitation, physics, electrical engineering... Many scientists tried to contribute to the development of voice and speech forensics within the scope of their professional competences and objective possibilities. Nevertheless, it is striking that most researchers who have dealt with this forensic discipline have limited themselves to theoretical considerations without ever personally trying to apply this method as an expert. For this reason, the theoretical conclusions were often not applicable to real forensic cases. We have tried to focus in this section on the research findings that have been incorporated into the various working standards of forensic phonetics laboratories around the world.

4.1. Historical overview of speaker identification

Historically, it can be said that recognizing voice and speech has always been a way of establishing a person's identity (Šešum, 2021). Before the development of technologies that enabled voice recording, the identification of speakers by their voice and speech was informal and based on personal observations and memories. A key element of this speech

recognition is that it requires prior knowledge of the speaker so that the listener can compare the voice they hear with a voice they recognize. Non-expert speaker recognition is an action that most people perform in everyday life, as it is based on experience with the voice and speech of a known person. Various problems arise in non-expert speaker recognition, including the influence of the passage of time and the reliability of the witness's memory (Lindh, 2017).

One of the oldest examples of lay speaker recognition can be found in the Bible (Genesis, chapter 27, verses 22-25, according to Šešum, 2021), when a blind father tries to recognise which of his sons, he is talking to by saying: "The voice is Jacob's, but the hands are Esau's."

The trial of William Hulett in 1660 is often cited as an argument for the unreliability of lay speech recognition. A witness heard the voice of the murderer of King Charles I, whose face was obscured, and identified him as Hulett, a person he knew well. As a result, Hulett was sentenced to death, but was later acquitted because the real perpetrator confessed to the crime. Such mix-ups were probably not uncommon throughout history, and it is certain that they still occur today, at least in everyday situations.

Probably the most famous case in recent history is the kidnapping of the child of American pilot Charles Lindbergh in 1932 (Lindh, 2017). During the ransom negotiations, Lindbergh briefly heard the language of the kidnapper and was convinced that it matched the language of a man who was later suspected of the crime and who was of German origin. He based his conviction on the influence of the automaticity of the German articulatory base on the pronunciation of the English words. Although the defense managed to disprove his claims due to the long period of time between the kidnapping and the trial in which Lindbergh made his statements (2.5 years), the suspect was sentenced to death and executed soon after. This case has contributed to scholars' interest in the retenstion of speech expression and the definition of lay and expert recognition of speakers (Broeders, 2001).

Until the invention of the sound spectrograph, which enabled the visual representation of voice and speech features, "speech profiles" were part of the criminal archives in police stations in Europe and the United States. Recorded descriptions of criminals' speech were used for various forensic and pseudo-forensic purposes, ranging from physiognomy to speaker identification. One attempt to describe voice and speech is associated with Eduard Sievers, who developed and applied his philological system called "sound analysis" "in 1926 (Li & Mills, 2019). This system was not developed for forensic purposes, but for the needs of literature. Its application was so complex that it was difficult to understand even for language experts, which led to it being abandoned.

The basis for the development of the classical, scientifically based method of speaker identification was the invention of the sound spectrograph, which was developed in 1941 in the laboratory of the renowned scientist Alexander Bell. The spectrograph made it possible to analyze and visually represent speech in three dimensions: Frequency, intensity, and duration. This technology was first used during the Second World War to identify the voices

of enemies. After the war, it was forgotten until the 1960s, when it was used intensively for forensic purposes, first in the United States and then worldwide (Šešum & Kovačević, 2015). Increasing demands for the identification of criminals by voice and speech led to a theoretical review and verification of the method, and technological development contributed to the improvement of the spectrograph's performance. In modern forensics, spectrographic analyses form the basis of the instrumental aspect of voice and speech investigation, and the spectrograph has evolved from a separate, large mechanical device to an easily transportable software-hardware package.

With the advancement of technology, the means of communication have changed rapidly, especially since the beginning of the 21st century. Due to the widespread use of mobile phones, which are cheap, readily available and contain many useful multimedia applications, and the improvement of portable audio recording devices, the need for forensic speaker identification has increased. As a result, security services around the world, as well as private investigators, have increasingly incorporated this type of forensic investigation into their work (Goyal, 2019). Today, almost all major forensic centers worldwide have a forensic phonetics laboratory.

4.2. Methods for identifying speakers

Over time, scientists have developed various methods for analyzing voice and speech. These methods can be categorized into four groups: 1) aural-perceptual methods; 2) spectrographic methods; 3) auditory-instrumental methods and 4) automatic methods for speaker identification (Didla, 2020; Šešum & Kovačević, 2015). The only method that is considered completely reliable for speaker identification is the auditory-instrumental method, which is a combination of two types of analysis: auditory (listening by a phonetician) and instrumental (analyzing voice and speech using software) (Šešum, 2021). This method combines in-depth phonetic-linguistic analysis with computer analysis of acoustic features, calculations, and measurements (Gold & French 2011). In this method, experts prepare an suspect's voice and speech sample for the purpose of analyzing and comparing speech samples, which is compared with a previously submitted questioned recordings. The suspect must first read a universal text, then repeat the sentences dictated in the questioned recording before conducting a free conversation with the expert (Šešum, 2021). After analyzing and comparing the samples, the experts write a report on the similarities of the samples, which is submitted to the court as part of the evidence (Jenkins et al., 2021). The speaker identification result is not an assertion, but an expert opinion relating to the likelihood that the compared speech samples belong to the same person. To give their opinion, the experts rely on an identification scale (Mukattash, 2016; Šešum & Kovačević, 2015), because voice and language experts are not mathematicians and express themselves in degrees of probability rather than percentages (Mukattash, 2016).

Forensic phonetics is a complex scientific field whose application requires specialized knowledge and training of experts (Schilling, 2015, cited in Sheoran & Mahna, 2023). The engagement of an expert is essential throughout the process of forensic voice and speech analysis, from the receipt of material to the presentation of expert reports in court (Cenceschi

et al., 2021). An expert in forensic phonetics/speaker identification can only be a person with the appropriate university education, as he or she must have excellent knowledge of the characteristics of voice and speech and their variability under different circumstances (Schwarz et al., 2011). The requirements for experts in the field of forensic phonetics include, in addition to exceptional knowledge in the field of phonetics and linguistics of a particular language, skills related to the use of techniques and modern technology, as well as knowledge of the relevant legal regulations (Meluzzi et al., 2020). For the results of the expert's opinion to be as reliable as possible, the expert must have as much information as possible about the material presented, which is facilitated by a chain of custody or documentation with data on the recordings. Keeping detailed records of the recordings provides insight into information about the authenticity of the recording and the manipulations made with it, but also ensures transparency and reliability of the process (Maher, 2010).

4.2.1. Auditory-instrumental analysis

The auditory-instrumental method of speaker identification is, as already explained, a combination of two approaches to speech sampling: one based on determining the characteristics of voice and speech by listening to voice recordings, the other based on the technical, visual representation of the acoustic characteristics of voice and speech.

4.2.1.1. Auditory analysis

Auditory analysis was developed first due to its minimal technical requirements and is considered the oldest method of forensic phonetics. Its application does not necessarily require the use of sophisticated technical instruments and software-hardware packages but relies solely on the expertise and experience of the analyst (Šešum, 2022). Auditory analysis of speech samples is an essential part of the speaker identification process because speech contains linguistic, social, pragmatic, behavioral and idiosyncratic features that need to be recognized, identified and compared. Foulkes & French concluded in 2001 that the identification of speakers in most countries is mainly based on the auditory method. The analyses within this method include the examination of voices, smallest speech segments, then syllables, words, sentences, and phonetic phrases as well as suprasegmental speech features. Thanks to auditory analysis, linguistic and dialectal affiliation, speech disorders and many other voice and language features that help to identify the speaker can be determined. Important features for the analysis are pitch, intonation, accent, speed, intensity, the presence of speech disorders, psychological and physiological states, masking, voice quality, etc. (Babić et al, 2017). Rhythm, intonation patterns, use of pauses, duration of segments, etc. are also analyzed to identify speakers (Kulshreshtha, 2012; Lindh, 2017). Breathing patterns are also an important feature in speaker identification, as they are individual and change less over the course of a person's life (Braun, 2020). Kienast & Glitza (2003) found that breathing patterns are characteristic of each speaker. They compared the frequency of breathing cycles, the distribution of breathing types (oral, nasal, combined) and the spectral composition of exhalation.

Although the speaker's gender and age are estimated based on the expert's auditory impression, determining the speaker's gender is considered a more reliable forensic marker than determining age because there are expected frequency ranges within which the fundamental frequencies of the voices range. For example, the expected range for men is 80 to 180 Hz, for women 180-230 Hz and for children 230 to 300 Hz (Kašić, 2003). However, errors can also occur here. For example, women may have a significantly lower voice on the frequency scale than expected, and men a higher voice. To estimate age, there are no frequency ranges on which the expert can rely. Therefore, the assessment is generally limited to broad age categories, with the speaker being categorized as a younger adult male or an older female, for example, with additional age intervals (e.g., 40-60). Changes in lung capacity, vocal cord elasticity or loss of front teeth, which are primarily due to age, will almost certainly affect the change in the quality and/or quantity of the speaker's voice and speech. It is therefore necessary for the expert to obtain answers to a series of prepared questions immediately prior to taking voice and speech samples that relate to the presence of potentially significant factors that may have changed since the questioned recordings recording. If possible, the expert should also be informed of the approximate age of the questioned recordings recording, as the examination of recordings older than ten years is not recommended due to changes that may occur in the speaker's speech production over a long period of time (Šešum, 2022). For example, knowing the age of the recording can be important because in some speakers who have switched to a different type of articulatory base after a certain period of life, remnants of automatisms of the original articulatory base can be detected at both segmental and suprasegmental levels, which can be an important forensic marker (Ivanović and Šešum, 2009). The information whether the woman was in advanced pregnancy during the recorded speech production, on any of the recordings, is crucial. Pregnancy can lead to changes in certain vocal and speech characteristics due to the displacement of the internal organs, especially the elevation of the diaphragm (Šešum, 2021).

Every speaker is characterized by the way they initiate speech and how much noise is in their voice. The onset of the voice is the way in which a speaker begins their speech and is an individual characteristic (Babić et al., 2017). Accent in speech is achieved through the variation of fundamental frequency values. Accent refers to the stress of a syllable in a word. Stressed syllables are usually pronounced louder, longer and at a higher pitch than unstressed syllables. The accent often determines the distinction between words (Harnsberger, 2009, Ahuja, 2018, as cited in Sheoran & Mahna, 2023). Depending on the complexity of the accent system of a particular language, this speech parameter can have a different forensic meaning, i.e., it is more significant in languages with a complex accent system. The auditory impression of fundamental frequency is called pitch (Karakoç & Varol, 2017) and refers to the pitch of the speaker's voice (Hollien, 2012). The variability of pitch can affect the change in the meaning of statements (Sondhi, 2015). Intonation involves changing the fundamental frequency during the production of a sentence or statement. Variations in intonation cause changes in the meaning of statements or in the style of speech (Ahuja, 2018).

Tempo refers to the speed of speech. People speak at a certain speed that corresponds to their habit, so the pace of speech is also a forensic marker (Karakoç & Varol, 2017). Speech rate can be defined as the number of uttered linguistic segments in a unit of time (Braun, 2020). On average, speakers speak around 150 words per minute. However, not everyone's speech rate is the same. Even the same person can speak quickly or slowly for different reasons, as the pace is related to speaking habits, the current emotional state, and the nature of the statement (Ahuja, 2018). When a speaker speaks quickly, this can often result in the omission of some sounds or syllables (Bayram, 2008). The length of speech segments can be observed at different levels, from the length of the pronunciation of individual sounds to the length of an average sentence.

Rhythm refers to the pattern of alternating stressed and unstressed syllables in speech and usually influences the flow and pace of speech (Hollien, 2014). Speech pauses are also regarded as individual markers of a speaker (Bayram, 2008). Speech pauses are a normal occurrence during a conversation, and analysing their frequency and quality is of great forensic importance (Tsanas et al., 2017). Natural conversation contains numerous fluencies and hesitations that manifest themselves in the form of transitional sounds or vocal clusters. These "fillers" usually occur when the speaker is trying to remember something, respond to a question or form a coherent sentence. These sounds are referred to as "pause fillers" and researchers have found that their use is a matter of speech style that can be of great importance in forensic speaker identification (Kuenzel, 1997). Different pause fillers can be used, and their use depends on external factors such as language, region, age of the speaker, etc. (Tsanas et al., 2017).

For the identification of speakers, it is necessary to be familiar with the standard articulation of sounds and their possible deviations, as the non-standard pronunciation of sounds is an important forensic feature. Deviations in pronunciation can result from the dialect, be caused by articulation habits in bilingual speakers or indicate the presence of a speech pathology (Varošanec-Škarić, 2019). Research on the quality and quantity of vowels has shown that the quality of vowel pronunciation differs in each individual dialect compared to the norm. The characteristics of pronunciation that a speaker exhibits can contribute to their identification (Ahuja, 2018, as cited in Sheoran & Mahna, 2023). Sometimes unexpected sounds can be perceived in speech, for example, from ill-fitting dentures. These sounds are perceived as suction because the prosthesis produces them when it detaches from the palate (Braun, 2020). Other sources of strange noises that accompany speech are due to insufficient saliva flow in the speaker's mouth (known as "dry mouth syndrome"), which can be caused by taking certain medications (e.g., beta blockers to regulate blood pressure) or situational stress. The suction-like sound caused by the separation of the tongue from the palate is clearly audible and measurable. On the other hand, excessive salivation can lead to slurping and frequent swallowing, which can also be a significant forensic marker in a particular case (Braun, 2020). All these parameters are observed as part of the auditory analysis (Bayram, 2008).

As has already been emphasized, to identify a speaker, the voice and language used as evidence must be compared with those of the suspect. However, even if there is no suspect, certain information about the perpetrator can be obtained through auditory analysis, i.e., speaker profiling, based solely on the questioned recording (Bayram, 2008). Speaker profiling helps law enforcement agencies to identify the perpetrators and steer the investigation in the right direction to identify the suspect (Morrison, 2019). Speaker profiling is based on extracting information about an unknown speaker based on their voice and language, which may include age, gender, weight, vocalization, dialect, breathing characteristics, articulation, speech style, etc. (Hughes, 2015, Hansen, 2015, Albuquerque, 2020, as cited in Sheoran & Mahna, 2023). It is important to note that speech features related to intonation and lexicon only apply to speakers belonging to the same dialect group, and that changes in dialectal accent are more pronounced in males than in females belonging to the same regional dialect (Kulshreshtha, 2012). One of the best-known examples of successful speaker profiling based on a audio- recording took place in Serbia in 2008, when a serial killer was identified and arrested thanks to the use of this method.

The main challenge in forensic speaker identification is the variability of voice and speech, as they can change under the influence of numerous factors such as age, health, and emotional state, and even the context in which speech production takes place (Albuquerque, 2020, Cerrato, 2000, Hansen, 2015, as cited in Sheoran & Mahna, 2023). In general, the factors that influence the prosody of one's speech are divided into two main groups (Cenceschi, 2019): one refers to fully defined characteristics associated with the dialogic dimension (e.g., rhetorical form, motivational state, emotions) and the other to dimensions that exist independently of the presence of an interaction (e.g., language or dialect, social context, etc.). Since the variability of voice and language is influenced by both internal and external factors and can occur intentionally or unintentionally, it becomes clear how complex this forensic field is and how much knowledge and experience it requires.

Auditory analysis is performed by experts with advanced training in phonetics or related scientific fields based on their knowledge and experience (Hollien, 2012). The auditory method of speaker identification is a fundamental approach to overcoming problems related to insufficient recording quality and mismatched recording channels (Hansen & Hasan, 2015). This is of particular importance as the assessment of degraded acoustic conditions in the recognition of speech features is a major challenge in forensic voice and speech analysis (Sheoran & Mahna, 2023). Auditory analysis is more practical than instrumental analysis as it does not require optimal recording quality or specific technical features that are prerequisites for performing an instrumental analysis. However, unlike instrumental analysis, which is not dependent on a specific duration of speech recording (above a prescribed minimum), the recording duration is crucial in auditory analysis. Experience has shown that the optimal speech duration for auditory analysis is limited to up to 10 minutes, as forensically relevant speech features usually manifest themselves within this time frame. In modern speaker identification, auditory analysis is essential, but must be complemented by data obtained using an instrumental approach.

4.2.1.2. Instrumental analysis

After the auditory analysis, an instrumental analysis of speech samples is carried out. In instrumental analysis, audio recordings are analyzed to extract features of the speech signal such as frequency, amplitude, and duration (Batalla, 2014, cited in Sheoran & Mahna, 2023). These features are used to identify patterns specific to a particular speaker and to compare questioned recordings with suspect's recordings to determine whether they belong to the same speaker. Instrumental analysis includes formant analysis and fundamental frequency analysis performed with a spectrograph (Šešum and Kovačević, 2015).

Spectrography is a technique for graphically decomposing speech as a complex sound into basic acoustic elements available for analysis. A sound spectrograph is a basic technical instrument used to identify voices and speech (Šešum, 2021). Spectrography provides a visual representation of the frequency spectrum of speech, called a spectrogram. The sound spectrogram is sometimes also referred to as a photograph of voice and speech, as it provides a visual representation of their characteristics and contains numerous parameters that can be used in the analysis. Spectrographs have facilitated the automatic recording and quantification of discrete vocal variations and easy conversion between recordings and databases (Li & Mills, 2019).

The most important type of analysis performed by voice and speech experts is vowel formant analysis. Formants represent concentrated acoustic energy generated in resonators (supralaryngeal cavities) depending on their shape. Consequently, the configuration and position of the vowels depend on the gender and age of the speaker (Cenceschi et al., 2021). However, the configuration and arrangement of the formants depends not only on the configuration of the resonators, but also on the anatomy and physiology of the articulators (Lindh, 2017). Formant analysis is based on the determination of the following features: Shape and position of vowel formants, distribution of acoustic energy, duration of speech segments, interrelation of formants and acoustic behaviour of voices in the speech chain (Šešum & Kovačević, 2015). Formant features contribute to the timbre of the voice (Grillo, 2020, cited in Sheoran & Mahna, 2023). (voice color) is an auditory sensation that allows the listener to perceive differences between sounds that are the same in pitch, loudness and length (Crystal, 1985). The timbre of the voice can often indicate the speaker to whom the voice belongs. It is specific to each person due to the complexity of speech production related to tongue position, configuration, and openness of the oral resonator, accompanying sounds, etc. (Karakoç & Varol, 2017).

The three lowest formants most accurately reflect how vowels are articulated and perceived. The position of the first formant (F1) determines the height of the tongue and is inversely proportional. For example, the vowel "A" has the lowest tongue position, and its first formant is the highest on the scale. The second formant (F2) refers to the protrusion of the tongue in the mouth when pronouncing vowels and distinguishes between vowels in the front and back row (e.g., the vowel "I" compared to the vowel "U"). The more the tongue protrudes, the higher the second formant is on the scale. The rounding of the lips during pronunciation is usually associated with fluctuations in the values of the second (F2) and

third (F3) formants (Cenceschi et al., 2021). Since the formant values represent specific frequencies of the sound signal, they are expressed in hertz (Hz). Although they are influenced by various factors, formants that are characteristic of a particular speaker always vary within a certain frequency range, so they cannot be defined in absolute terms (Harrison, 2004). However, among the various measurements available for comparing voice and speech, formant analysis is considered one of the most reliable as it provides a good insight into the characteristics of voice and speech (Cenceschi et al., 2021).

The zeroth formant (F0) refers to the fundamental frequency of the voice (Karakoç & Varol, 2017). The fundamental frequency represents the number of cycles of vibration of the vocal cords in one second; it is expressed in Hertz (Hz) and is a fundamental parameter of instrumental voice analysis (Roach, 2002). The fundamental frequency is considered the most important feature of speech (Titze, 2000) and is one of the most studied parameters of voice and speech. It is accessible because it is easy to calculate even in acoustically sub-optimal environments (Lindh, 2017). The spread and average of the fundamental frequency of the voice indicate its pitch. One problem with relying on average fundamental frequency values is that they are normally distributed in the population (Lindh, 2006), so their forensic value is limited, and they can contribute significantly to forensic analysis mostly when extreme values of this parameter are present. Therefore, in addition to the average baseline frequency values, other characteristics of the parameter are also considered (Lindh, 2017).

4.2.2. Automatic methods for the identification of speakers

Automatic speech recognition is the latest method of speaker identification. It is based on the programming of an automatic system that recognizes human speech independently. Once the speech is identified, the system can use the decoded speech as input for a wide range of different applications (Kydyrbekov et al., 2020). Automatic speech recognition emerged as a solution to overcome the limitations of the classical auditory-instrumental approach to speaker identification. Automatic systems are based on assumed probability patterns rather than individual voice and speech features. In automatic analysis, speech patterns are compared with numerous other patterns stored in the database and the system selects the pattern to which the questioned speech is most similar (e.g., comparing a recorded extortion speech with speech patterns of perpetrators of various crimes stored in the database) (Singh et al., 2018). This approach differs significantly from the classical method of speaker identification, which does not require an existing database and in which an expert recognizes, analyses, and compares the characteristics of each speech sample (Li & Mills, 2019).

Automatic speaker identification methods have several limitations. The most important one is that the speaker identification decision directly depends on the size of the database (Kydyrbekov et al., 2020). Furthermore, unlike other methods of speaker identification, automatic systems are language-independent, meaning that the system does not consider dialects or even the language in which the speech is delivered (Lindh, 2017). Although this feature of automatic systems is often seen as an advantage, it can also be a serious disadvantage in some cases. Ambient noise, such as traffic sounds, other people, wind, etc., can also affect speech quality and significantly reduce the reliability of automatic

speaker identification results (Sheoran & Mahna, 2023). Anil et al. (2005) compared the results of auditory analysis and automatic speaker recognition and found that auditory analysis is more reliable than automatic recognition when the recording conditions are not optimal. They concluded that human hearing is more resistant to environmental noise. In contrast to an automatic system, a person can identify a speaker by voice and speech even in the presence of significant ambient noise by focussing on the speaker and ignoring acoustic distractions. Hautamaeki and colleagues (Hautamaeki et al., 2013) agree with these conclusions and state that experts are more reliable at voice and speech recognition than automatic systems.

Most automatic systems for speaker identification or verification rely solely on vocal tract features, although there have been attempts to include non-anatomical speech features in them. It is important to understand the difference between these aspects of speech, as they are analyzed in different ways and their results are difficult to combine in automatic analysis (Lindh, 2017). Although automatic methods for speaker identification are increasingly used for forensic investigations, their use is usually combined with classical auditory-instrumental methods due to the limitations (Lindh, 2017). Despite the fact that automatic speaker verification has proven useful in security and commercial applications, it is important to emphasize the differences to the forensic application of automatic systems, because in forensic applications the speaker of interest is usually uncooperative, has no interest for the recognition of their voice and speech, the recording conditions and speech styles vary much more, the quality of questioned recordings is often poorer and the outcome of the system's decision is not a binary decision but a quantification of the strength of the evidence (Morrison & Enzinger, 2019).

4.3. Factors that make identification difficult

Forensic speaker identification is a major challenge for experts because the quality of audio recordings in real cases is often very poor and there is usually a significant difference between the speaking style and recording conditions of questioned recordings and suspect's recordings. For example, questioned recordings may contain only a few seconds of speech, be polluted with ambient noise from different sources, vary in intensity, be recorded in a room with reverberation, be recorded through a microphone that is quite far away from the speaker of interest, be recorded via different transmission channels (hand recorder, landline or mobile phone, internet applications), be stored in compressed formats (MP3, amr), which significantly reduces the recording quality (Morrison & Enzinger, 2019). It can be said that most recordings that are the subject of a forensic investigation are made under sub-optimal conditions, both in terms of the acoustic environment and the technical aspects of the recording. Therefore, it sometimes happens that speech and voice, although audible on the recording, are simply not of sufficient quality for analysis and comparison (e.g. low signal-to-noise ratio, speech overlap, etc.). In such cases, the expert is obliged to reject the forensic analysis of recordings that do not fulfil the conditions required for voice and speech examination (Meluzzi et al., 2020).

Over the years, many factors have been identified that complicate the process of speaker identification based on voice and speech. For example, a significant number of cases in this forensic field involve telephone conversations, which are a complicating factor as telephone conversations are usually anonymous, of short duration and of questionable quality. In addition, ambient noise, limitations in the frequency transmission range, interference, signal interruptions, etc. can significantly contaminate the voice sample (Leuzzi et al., 2016). Auditory analysis is most limited by the duration of the speech of interest, which is often very short in questioned recordings (Broeders, 2001), especially in the case of false reports of threats in facilities or transport vehicles. Even surreptitiously recorded audio may be long but not contain enough speech for analysis and comparison, as a few exchanged sentences in a dialogue are generally not sufficient for speech analysis and comparison, both at the auditory and instrumental level (Meluzzi et al., 2020). Interestingly, in auditory analyses, the duration of the recordings is more important than their quality, while in instrumental analyses, the quality of the recordings is more important than their duration. However, since both types of analyses are performed simultaneously for forensic purposes, internationally established common minimum conditions regarding the quality and duration of voice recordings must be met for them to be admissible for examination.

It must be recognized that forensic phonetics is the most complex forensic discipline because, unlike almost all other types of evidence, speech and voice are easily altered and exist over time. One of the main problems in identifying speakers is that speech, unlike fingerprints or DNA samples, is part of human behavior, i.e. it is subject to short- or long-term variability as it is not only due to anatomical and physiological factors. Short-term variability in human voice and speech relates to emotional changes (Braun & Heilmann, 2012), health status (Baken, 1987) and even time of day (Garrett & Healey, 1987); long-term variability relates to ageing (Linville, 2001) or relocation of a speaker, which means a change in the linguistic environment (Kiesewalter, 2019), etc. The use of drugs, certain medications and alcohol can lead to significant changes in a person's voice and speech while under their influence but can also have lasting effects on speech production (Babić et al., 2017). In addition, the human voice and speech can be intentionally masked through the manipulation of vocal and speech characteristics, the use of mechanical barriers and through the use of electronic devices that alter them (Didla, 2020).

Language studies on twins are particularly interesting for forensic purposes because their biometric characteristics are very similar (Jain et al., 2002). Identical twins are characterized by a very similar anatomy, which is also reflected in their voices. On the other hand, same-sex fraternal twins may have similar voices, but the similarity is expected to be at the level of other same-sex siblings who are not twins. Because distinguishing the voices of identical twins is challenging, research in this area is critical to forensic voice and speech science (San Segundo, 2015). Existing evidence shows that even in monozygotic twins where the configuration of the speech tract is identical or nearly identical, linguistic differences that inevitably characterize their speech can relatively easily point to a particular speaker (Šešum, 2013).

Signal variability (and consequently acoustic parameters) is influenced by the quality and context of the recording, which can be crucial in forensics as recordings can range from excellent quality to extremely compressed and from recordings in quiet to those in extremely noisy environments (Cenceschi et al., 2021). In real forensic cases, questioned recordings, and even suspect's recordings contain different types and amounts of ambient noise (Hofstetter et al., 1994). Ubiquitous sounds can mask or obscure speech and cause changes in its auditory, spectral, and other acoustic properties. These changes make it difficult to recognize voice and speech features and affect the reliability of speaker identification (Chakroun & Frika, 2020). Audio recordings are often provided in compressed formats with very unfavorable spectral characteristics and sampling rates. Such recordings are of very poor quality and, in combination with a minimal signal-to-noise ratio, make it considerably more difficult and often impossible to conduct forensic investigations (Meluzzi et al., 2020). The variability of the recording channels can also make it difficult to identify speakers. This is related to changes in the acoustic environment, microphone quality or the quality of the transmission channel, which can affect the quality of the speech signal and reduce the reliability of identification. Signal equalisation techniques such as equalisation and dereverberation can be used to improve the speech signal (Wang, 2021, as cited in Sheoran & Mahna, 2023).

4.3. Development of forensic phonetics in the Republic of Serbia

The development of forensic phonetics in Serbia began in 2007, when the first laboratory for forensic phonetics was opened at the Forensic Centre of the Serbian Ministry of Interior in Belgrade. This laboratory was also the first of its kind in the entire region (Šešum & Kovačević, 2015). The laboratory was designed according to strict sound architecture requirements, including the layout of the rooms and the materials used for construction to avoid reverberation and staff fatigue. Specific software and hardware packages required for voice and speech analysis were procured from the Russian manufacturer "Speech technology center", as they were rated by the lab's experts as the highest quality on the market at the time. The accompanying equipment, including noiseless computers, recording devices, microphones, amplifiers, sound reproducers, speakers, headphones, etc., was purchased through the usual channels from commercial suppliers on the market.

The staff structure of the laboratory consisted of a speech and hearing specialist and an electrical engineer. The laboratory experts successfully completed the required training under the guidance of world-leading experts in the field of forensic phonetics. The training took place between 2008 and 2011 at the National Scientific Police in Madrid, the Federal Security Service in Moscow and by experts from the "STC" center in St. Petersburg (Šešum & Kovačević, 2015). In addition to voice and speech analysis, the employees were also actively involved in scientific research and thus contributed to the development of forensic voice and speech analysis in Serbia and the region. This is of crucial importance, as analyses of this type depend on the linguistic characteristics of the region. Today, forensic voice and speech analysis is also used in private practice and in laboratories in the region that have been modelled on the Serbian laboratory (e.g., the laboratory in Sarajevo, Bosnia). Forensic

voice and speech analysis has found its place in academic education and is studied at the Faculty of Special Education and Rehabilitation in Belgrade and at the Faculty of Philology.

5. Conclusion

The widespread availability of mobile phones has significantly increased the frequency of long-distance calls and enabled the easy recording of communications, which has affected the increase in police investigations and court proceedings where voice and speech are used as evidence. As a result, the identification of speakers by voice and speech has only recently shown its true potential. To establish the identity of a speaker, experts must be able to recognize, isolate and compare all the characteristics of voice and speech that occur in the speech production of both unknown and known speakers. They must also be alert to possible attempts to manipulate voice and speech with the aim of deceiving the experts. It is therefore important that only experts with an appropriate university education in speech, voice, and language, as well as appropriate training and sufficient experience in forensic analysis, are commissioned to identify speakers. Future experts must be trained in the use of the classical, auditory-instrumental method, as it is considered the most reliable method and the only one whose results are recognized in court, despite the constant advances in automatic identification methods, whose weaknesses have not yet been successfully overcome.

Bibliography

1. Alkhatib, B., & Kamal Eddin, M. M. W. (2020). Voice Identification Using MFCC and Vector Quantization. *Baghdad Science Journal*, *17*(3), 1019. https://doi.org/10.21123/bsj.2020.17.3(Suppl.).1019
2. Anil, A., Damien, D., Filippo, B., & Andrzej, D. (2005). Aural and automatic forensic speaker recognition in mismatched conditions. *International Journal of Speech, Language and the Law*, *12*(2), 214–234. DOI:10.1558/sll.2005.12.2.214
3. Babić, I., Otuzbir, S., & Hodžić, I. (2017). The Significance of the Forensic Phonetic in the Voice Identification as an Effective Protectuon and Safety Measure. *Nauka i Tehnologija*, *5*(9), 157–165.
4. Baken, R. J. (1987). *Clinical Measurement of Speech and Voice*. Taylor & Francis.
5. Bayram, L. (2008). *Voice and Speech Analyzes in Forensic Sciences*. Seçkin Publishing.
6. Braun, A. (2020, October, 5). *Nonverbal Vocalisations – A Forensic Phonetic Perspective In Laughter and Other Non-Verbal Vocalisations* Workshop, Bielefeld, Germany, 19–23. DOI: https://doi.org/10.4119/lw2020-918
7. Braun, A., & Heilmann, C. M. (2012). *SynchronEmotion*. Peter Lang.
8. Broeders, A. (2001, October 16-19). *Forensic Speech and Audio Analysis Forensic Linguistics.* 13th INTERPOL Forensic Science Symposium, Lyon, 54–84. https://deliverypdf.ssrn.com/delivery.php?ID=644113004017115116119008095110092097029078004022011049090080084098125113027002096126119126029119053124101020126089115101006070122047012042084066073127065089126002092007079067091089086095095096013018112082112085070012027118075019080007121111113097087096&EXT=pdf&INDEX=TRUE

9. Cenceschi, S. (2019). *Speech analysis for automatic prosody recognition*. [Ph.D. thesis, Department of Computer Engineering at the Politecnico di Milano]. https://hdl.handle.net/10589/144841
10. Cenceschi, S., Meluzzi, C., & Trivilini, A. (2021). The Variability of Vowels' Formants in Forensic Speech. *IEEE Instrumentation & Measurement Magazine, 24*(1), 38–41. Doi: 10.1109/MIM.2021.9345600.
11. Chakroun, R., & Frikha, M. (2020). Robust features for text-independent speaker recognition with short utterances. *Neural Computing and Applications, 32*, 13863–13883. https://doi.org/10.1007/s00521-020-04793-y
12. Crystal, D. (1985). *A Dictionary of Linguistics and Phonetics.* Basil Blackwell Ltd.
13. Didla, G. S. (2020). A Review of Voice Disguise in a Forensic Phonetic Context. *International Journal of English Literature and Social Sciences, 5*(3), 721–725. DOI: 10.22161/ijels.53.25
14. Foulkes, P., & French, P. (2001). Forensic phonetics and sociolinguistics. In R. Mesthrie (Ed.), *Concise Encyclopedia of Sociolinguistics* (pp. 329-332). Elsevier.
15. Garrett, K. L., & Healey, E. C. (1987). An acoustical analysis of fluctuations in the voices of normal adult speakers across three times of day. *Journal of the Acoustical Society of America, 82*, 58-62. DOI: 10.1121/1.395437
16. Gold, E., & French. P. (2011). International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law 26*(1): 293–307. DOI:10.1558/ijsll.v18i2.293
17. Goyal, A. S. (2019). Identification of source mobile hand sets using audio latency feature. *Forensic Science International*, *298*(1), 332–335. https://doi.org/10.1016/j.forsciint.2019.02. 031
18. Hansen, J.H.L., & Hasan, T. (2015). Speaker Recognition by Machines and Humans: A tutorial review. *IEEE Signal Processing Magazine, 32*, 74–99. doi: 10.1109/MSP.2015.2462851.
19. Harrison, P. (2004). *Variability of formant measurements*. [M.A. thesis, Department of Language and Linguistic Science, University of York]. https://www.jpfrench.com/wp-content/uploads/harrison-formant-dissertation.pdf
20. Hautamäki, R. G., Hautamäki, V., Rajan, P., & Kinnunen, T. (2013). Merging human and automatic system decisions to improve speaker recognition performance. *Proceedings of the Annual Conference of the International Speech Communication Association, 2519–2523.* DOI:10.21437/Interspeech.2013-422
21. Hofstetter, E.M., & Rose, R.C., & Reynolds, D. A. (1994). Integrated models of signal and background with application to speaker identification in noise. *In IEEE Transactions on Speech and Audio Processing*, 2, (2), (pp. 245-257). doi: 10.1109/89.279273.
22. Hollien, H. (2012) About forensic phonetics. *Linguistica, 52*(1), 27–53. DOI:10.4312/linguistica.52.1.27-53
23. Islam, R., Abdel-Raheem, E., Tarique, M. (2022). A Novel Pathological Voice Identification Technique through Simulated Cochlear Implant Processing Systems. *Applied Sciences, 12*(5), 2398. https://doi.org/10.3390/app12052398

24. Ivanović, M., & Šešum, M. (2009, jun 15–18). *Jedan tip regionalne redukcije neakcentovanih slogova kao forenzički marker*. Zbornik radova, 53. Konferencija ETRAN, Vrnjačka Banja.
25. Jain, A., Phrbhakar, & S. Pankanti, S. (2002). On the similarity of identical twin fingerprints. *Pattern Recognition, 35*(11), 2653–2663. https://doi.org/10.1016/S0031-3203(01)00218-7
26. Jenkins, R. E., Tsermentseli, S., Monks, C. P, Robertson, D. J., Stevenage, S. V., Symons, A. E., & Davis, J. P. (2021). Are super-face-recognisers also super-voicerecognisers? Evidence from cross-modal identification tasks. *Applied Cognitive Psychology, 35*(3), 590–605. https://doi.org/10.1002/acp.3813
27. Karakoç, M. & Varol, A. (2017). Visual and Auditory Analysis Methods for Speaker Recognition in Digital Forensic. *International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey,* 1113-1116. doi: 10.1109/UBMK.2017.8093505.
28. Kašić, Z. (2003). *Fonetika*. FASPER, Beograd.
29. Kienast, M., & Glitza, F. (2003). Respiratory Sounds as an Idiosyncratic Feature in Speaker Recognition. *Proceedings of XVth International Congress of Phonetic Sciences, Barcelona.* 1607- 1610.
30. Kiesewalter, C. (2019). *Zur subjektiven Dialektalität regiolektaler Aussprachemerkmale des Deutschen.* Steiner.
31. Kulshreshtha, M., Singh, C., Sharma, R. (2012). Speaker Profiling: The Study of Acoustic Characteristics Based on Phonetic Features of Hindi Dialects for Forensic Speaker Identification. In Neustein, A., Patil, H. (eds) *Forensic Speaker Recognition (pp.71-100).* Springer. https://doi.org/10.1007/978-1-4614-0263-3_4
32. Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *International Journal of Speech, Language and the Law, 4*(1), 48–83.
33. Kydyrbekov, A. & Othman, M., Mamyrbayev, O., Akhmediyarova, A. & Zhumazhanov, B. (2020). Identification and authentication of user voice using DNN features and i-vector. *Cogent Engineering, 7*, 1–21. Doi: 10.1080/23311916.2020.1751557.
34. Larner, S. (2015) From intellectual challenges to established corpus techniques: introduction to the special issue on forensic linguistics. *Corpora, 10*(2). 131–143. DOI: 10.3366/cor.2015.0071
35. Leuzzi, F., Tessitore, G., Delfino, S., Fusco, C., Gneo, M., Zambonini, G., & Ferilli, S. (2016). *A statistical approach to speaker identification in forensic phonetics field.* 5th International Workshop on New Frontiers in Mining Complex Patterns, held in ECML-PKDD. DOI:10.1007/978-3-319-61461-8_5
36. Li, X., & Mills, M. (2019). Vocal Features: From Voice Identification to Speech Recognition by Machine. *Technology and Culture 60*(2), 129-160. doi:10.1353/tech.2019.0066.
37. Lindh, J. (2017). *Forensic Comparison of Voices, Speech and Speakers Tools and Methods in Forensic Phonetics.* [PhD thesis, Department of Philosophy, Linguistics and Theory of Science University of Gothenburg].

https://gupea.ub.gu.se/bitstream/handle/2077/52188/gupea_2077_52188_4.pdf?sequence=4&isAllowed=y

38. Lindh, J. (2006, January 1). *Preliminary f0 statistics and forensic phonetics.* Annual conference of IAFPA, Department of Linguistics, Geteborg University. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=68466944694e11951683b23dc4c71d44558057b8
39. Linville, S. E. (2001). *Vocal Aging*. Singular.
40. Maher, R. C. (2010). Overview of audio forensics. In H.T. Sencar et al. (Eds.), *Intelligent Multimedia Analysis for Security Applications* (pp.127–144). Springer.
41. Meluzzi, C., Cenceschi, S., & Trivillini, A. (2020). Data in Forensic Phonetics from theory to practice. *TEANGA the Journal of the Irish Association for Applied Linguistics, 27,* 65–78. Doi: 10.35903/teanga. v27i.223.
42. Morrison, G., & Enzinger, E. (2019). Multi-laboratory evaluation of forensic voice comparison systems under conditions reflecting those of a real forensic case (forensic_eval_01) – Introduction. *Speech Communication, 112*(c), 37–39. https://doi.org/10.1016/j.specom.2019.06.007
43. Mukattash, B. (2016). The Role of Forensic Phonetics in Legal Investigation: A Case Study of Two Speaker-Identified/Unidentified Recorded Samples. *Journal of Literature, Languages and Linguistics, 29,* 31–37.
44. Roach, P. (2002). *A Little Encyclopedia of phonetics*. University of Reading.
45. San Segundo, E. (2014). Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics. *International Journal of Speech Language and the Law, 22* (2), 249–253. Doi: 10.13140/RG.2.2.34536.01284.
46. Schwartz, R., Campbell, J. P., Shen, W., Sturim, D. E., Campbell, W. M., Richardson, F. S., Dunn, R.B. & Granville, R. (2011). *USSS-MITLL 2010 human assisted speaker recognition.* In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference, Prague, Czech Republic, 5904–5907. https://doi.org/10.1109/ICASSP.2011.5947705
47. Šešum, M (2013). Komparativna analiza formantnih struktura glasova sestara i glasova monozigotnih bliznakinja. *Beogradska defektološka škola, vol. 19*(3), 515–527.
48. Šešum, M. (2021). Forenzička fonetika-identifikacija govornika. U: Knežević, S. (ur.), *Forenzičko računovodstvo, istražne radnje, ljudski faktor i primenjeni alati* (pp. 828–859). Beograd: Fakultet organizacionih nauka
49. Šešum, M. (2022, april 9–10). *Uloga surdologa i logopeda u forenzičkoj analizi glasa i govora.* V simpozijum logopeda Srbije: Timski rad u logopediji i defektologiji., Beograd.
50. Šešum, M., Kovačević, J. (2015). Forenzička fonetika. U *Vodič za primenu Zakonika o krivičnom postupku Republike Srbije (pp.90–106).* Grafolik, Beograd.
51. Sharma, P. & Sahu, N. (2018). A Review and Analysis of Voice Identification System. *International Journal of Innovative Knowledge Concepts*, *6*(5), 2454–2415. DOI 11.25835/IJIK-54

52. Sheoran, S., & Mahna, D. (2023). Voice Identification And Speech Recognition: An Arena Of Voice Acoustics. *European Chemical Bulletin, 12*(5), 50–60. DOI: 10.31838/ecb/2023.12.si5.008
53. Singh, N. & Khan, R. A. (2016, March 16–18). *Underlying of text independent speaker recognition*. 3rd International Conference on Computing for Sustainable Global Development, at BVICAM, New Delhi, 11–15. https://ieeexplore.ieee.org/document/7724216
54. Singh, N., Agrawal, A., & Khan, R. A. (2018). Voice Biometric: A Technology for Voice Based Authentication. *Advanced Science, Engineering and Medicine, 10*(7), 1–6. doi:10.1166/asem.2018.2219
55. Sondhi, S. K. (2015).). Acoustic analysis of speech under stress. *International Journal of Bioinformatics Research and Applications,11*(5). 417–432. DOI:10.1504/IJBRA.2015.071942
56. Titze, I.R. (2000*). Principles of Voice Production*. National Center for Voice and Speech.
57. Tsanas, A., Gomez, P., & San Segundo, E. (2017). Exploring Pause Fillers in Conversational Speech for Forensic Phonetics: Findings in a Spanish Cohort Including Twins. *Proceedings of ICPRS 2017 : 8th International Conference on Pattern Recognition Systems At: Madrid, Spain*, 1–6. Doi: 10.1049/cp.2017.0161.
58. Varošanec-Škarić, G. (2019). *Forenzična fonetika*. Ibis grafika.
59. Zhou, Y., Liu, Y., & Niu, H. (2022). Perceptual Characteristics of Voice Identification in Noisy Environments. *Applied Sciences, 12*(23), 12129. https:// doi.org/10.3390/app122312129