

# APPLICATION OF VOICE BIOMETRICS IN PROTECTION SYSTEMS AND CRIME FIGHTING

Saša Paunović<sup>1</sup>, Lazar Nešić<sup>2</sup>, Jovan Kovačević<sup>3</sup>

*Ministry of Interior, Belgrade, Republic of Serbia*

*<sup>1</sup>sasa.paunovic@mup.gov.rs , <sup>2</sup>lazar.nesic@mup.gov.rs , <sup>3</sup>jovan.kovacevic@mup.gov.rs*

General survey

DOI: 10.7251/JIT1202059P

UDC: 351.755.62:57.081.1

**Abstract:** Modern communication relies increasingly more on the verbal communication between a machine and a human, aiming to govern certain resources and robots, increase the security of certain means, initiate certain processing protocols, faster financial transactions... This paper illustrates the possibility of using the voice biometrics in modern living, from simple examples, such as starting the motor of a vehicle, through opening security gates, to proving fraud and embezzlement. Special emphasis has been put on the systems of automatic speaker identification and forensic speaker recognition.

**Key words:** Biometry, Biometry systems, Speech, Speaker, Voice recognition, Identification, Security.

## INTRODUCTION

The best method of human communication is speech as the process of articulating sounds with certain meaning [5]. In order for people to be able to execute "distant" communication (over great physical distances), it is necessary to develop devices which will enable the transfer of sound from one location to another. That is how telephone and radio devices emerged. However, after some time it did not suffice anymore. Right after the Second World War the need was felt to analyze particular speech features and depict speech in a visual form. It has set the foundation for the development of new machines and devices for voice processing.

The analysis of specific characteristics of voice has begun before the computer invention. Melville Bell, father of Alexander Graham Belle has started the research on translation of the speech into a visual form in 1867. The idea was to monitor the movement of enemy troops over the radio. However, the technology was not developed enough to support this idea. Later on, the police continued to develop Melville

Bell's idea until it reached its peak through today's software use.

Voice biometrics represents the biometric model of physical characteristics of the human voice, numerical sound carrier and the way of speech of each individual [7].

When we are talking about the speech identification through voice recognition, it is necessary to distinguish the difference between speech identification and the identification of the speaker. Speech identification represents the recognition of the spoken words while the biometrical method of identification represents the speaker identification, or the person who is speaking.

The purpose of voice recognition is the authentication of a certain person based on the unique voice characteristics, given that the voice represents the unique and constant identifier [1]. The voice uniqueness is represented on one hand in its pitch, and tone, and on the other in the difference in verbaliza-

tion, which is the difference in use of the muscles of mouth, jaw and tongue. Nobody can change one's voice to the point of unrecognizing. It is interesting to say that the voice-print of twins is not identical due to automatised speech patterns. Moreover, speech uniqueness is represented in a way people articulate sounds, make pauses while speaking, their beginning of the speech - its speed, rhythm and intonation - all these characteristics can contribute to speech identification.

Voice biometrics is applied more often in private and public sector, given that it is simple to use, acceptable by the public, and is cost effective compared to other biometric technologies. However, biometrics is not that reliable due to the fact that it is used for verification and not for identification purposes. For that reason, it is used as an additional tool in biometrics; in combination with other biometric methods.

## APPLICATIONS OF VOICE BIOMETRICS IN PROTECTION SYSTEMS AND CRIME FIGHTING

### How does this biometric function?

The technology of voice biometrics is the least invasive, but at the same time the least reliable one. This technology is based on voice recognition, by making the speaker say a phrase into a microphone, which is followed by software analysis that compares the voice with the speech database [2].

The process of identification is divided into two phases (steps). The first phase is the encryption and the second one is verification or identification.

First step in acquiring the biometric sample is the voice-print. Sample can be identified as a phrase, text or a serial numbers that are implemented into a database which serve to identify a certain person. Step one consists of a speaker speaking into a microphone. Microphone is the one that transforms an acoustic wave into an electric signal, which is more modified with the use of sound cards in computers. Moreover, this electric signal is adapted into various applications and infiltrated in a certain amplification range (value

interval -1 to 1) quantified by the corresponding levels etc. Finally, the signal is represented in such a way that it is possible to apply a vast number of sophisticated processes which will acquire all the significant information from the signal.

The separation is performed when an algorithm in given sample is searched for unique voice characteristics such as tone, pitch, time frame etc, which is then followed by the formation of biometrical sample which is stored into the database. This process lasts between 2 and 8 seconds. So, in this phase, the spoken phrase is transformed into a digital format (from an analogue) and is stored for future comparison reference.

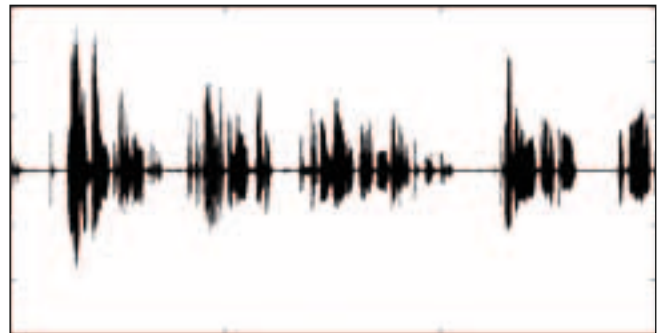


FIGURE 1. RECORDING OF 15 SECOND SPEECH WITH THE SAMPLING RATE OF 8 KHz

The next phase is the verification/identification which is performed in a way that an individual produces speech. An individual reads a text or a series of numbers into a microphone. Voice samples are then converted into digital characteristics which are then compared to algorithm characteristics in the database. This process lasts for 0.5 seconds.

Systems that are applied in speech and individual recognition can be divided into dependent, independent and integrated systems.

Speaker articulates a certain phrase that represents name, place of birth, favorite color or various numbers. This method is very simple to use because identification is verified by comparing a phrase with a database.

Compared to the dependent system, in the **independent system** quality analyses of the voice and the spoken words do not affect identification pro-

cess. In other words, this does not require the pronunciation of a certain word; the identification can be performed with the pronunciation of any word. More specifically, the authentication is done when a person speaks out a specific text, for example a series of numbers from 0-9. It is worth mentioning that specifics in pronunciation of these series of numbers should be followed for the voice identification to be safer and faster.

Speaker verification in the integrated system consists of two steps:

1. System identifies the speech by recognizing a certain text (password, personal information or a number).
2. Spoken words are compared to a database for identification. This type of system is used in bank transactions.

For example Via Voice software (developed by IBM) enables user identification via phone by combining two sources of information: voice print and user's password and personal information. The process looks like this: during a conversation, Via Voice poses a series of random questions to a user; this is followed by verification of given responses via voice print. When a sample is matched with the one in the database, Via Voice concludes that the user in question is the right one. Also, voice authentication depends on the user's right or wrong answer and the voice-print assessment. If it is the right user, this process will be short since it is composed of one question. In case of the false identification, the authentication process lasts longer since Via Voice poses more questions until the person gives the wrong answer or until it concludes that the voice-prints do not match.

## TECHNOLOGY DEVELOPMENT

Technology of voice-prints is getting more advanced in the past couple of years. Before its modernization, every word had to be spoken out separately and slowly so the system could recognize it. With today's technology, this is not necessary since the system is capable of recognizing fluent speech and the computer can recognize 160 per minute. It is also im-

portant to emphasize that the system integrates neural nets, in other words it memorizes the way how every person sounds. Still, errors are unavoidable since the biometry represents a mathematical model of every person's physical trait in everyday life. However, this happens in a very small number of cases.

## APPLICATIONS

Given the biometry characteristics, it is not surprising that its use is widespread. It is used to access locations and buildings, personal belongings, money transactions, etc. It is also used in security forces, usually by police in executing court orders.

### Physical access to certain locations and objects

The application of biometry in this area of use is represented by a sample of voice-print with a combination of other authentication models such as codes, fingerprints and iris prints. All of these methods contribute to a maximum security. Institutions that use these kinds of methods are banks, security buildings, apartment buildings, etc.

### Safe and faster access to personal devices

The application of this biometry is very significant concerning the use of cell phones, laptops and other personal devices since it offers a maximum protection from the unauthorized use. This can be overcome by denying access to anyone whose voice print does not match the owners.

Smart phones are more and more present in today's society. They usually recognize speech commands and allow hand-free use. For example, a person speaks out the name of a person that he/she wishes to call and the phone will automatically recognize that person from the memory and dial the number automatically.

### Money transactions

Application of this type of biometry can be noted in ATM machines which lead to less unauthorized use.

In other words, ATM machines (cash dispensers) require the authentications with a PIN number. How-

ever, what if someone is standing behind us when we are typing the PIN number, or what if the camera that is set above the ATM machine records our PIN number? All of this can lead to illegal use of a PIN number. This can be avoided by the combination of biometry and a PIN number. Moreover, system would have to identify the person that is using an ATM machine (either with a spoken word or an answer to a certain question) and a PIN number. Once all the necessary information is verified, he/she is able to withdraw the money. This type of ATM machine use does not require a lot of time and is safe at the same time.

This type of biometry is also used in call centers within a bank when customers use telephone banking. The majority of banks have call centers that base their authentication systems on voice-print biometry. To sum up, clients are using call centers to do their money transactions and they are usually speaking with a bank official which then starts up the operations on a computer of a local or central bank.

### Security forces

This type of technology is very much applied in security forces either public or secret. One of the applications is seen in the surveillance of telephone lines. This is very significant today since the classic ways of monitoring are not applied as much. How does this work? For example, in an effort to make operator's job easier, the computer would monitor large number of lines simultaneously until the key word appears on any line. When those words come up, the recorded conversation is forwarded to a human operator so he/she would verify the content. It is important to say that if the right person is expected, then those systems would be combined with an automatic recognition of the speaker so that the security of the search for the right person would be maximized.

### Suspect identification, investigation and the execution of court orders

The use of this type of biometry is important in police investigations, since it provides the possibility of identifying the suspect through a voice recording. For example, when a kidnapper calls and asks for a money ransom, a voice-print can identify his/

her voice through a specialized investigation. Apart from its use in identifying the kidnappers, it is used in telephone frauds, voice recordings and transcripts.

The dependent system is mostly applied here, and it is important to develop mechanism for filtering the information due to existence of a large number of data.

In order for a voice print to be used as evidence in court, it is necessary to spot and represent the voice characteristics of the recording in question and compare them to voice characteristics of another recording. This can be done by executing appropriate analysis such as audio-linguistic phonetic or instrumental (computer).

Audio-linguistic phonetic analysis can be performed by a phonetic (phonetically). It consists of defining speech characteristics with a goal of pointing out their regularities such as: pathological patterns, speech irregularities, the use of buzzwords and maxims, various accents, etc. The result of this analysis depends on the knowledge and experience of the expert and is not objective.

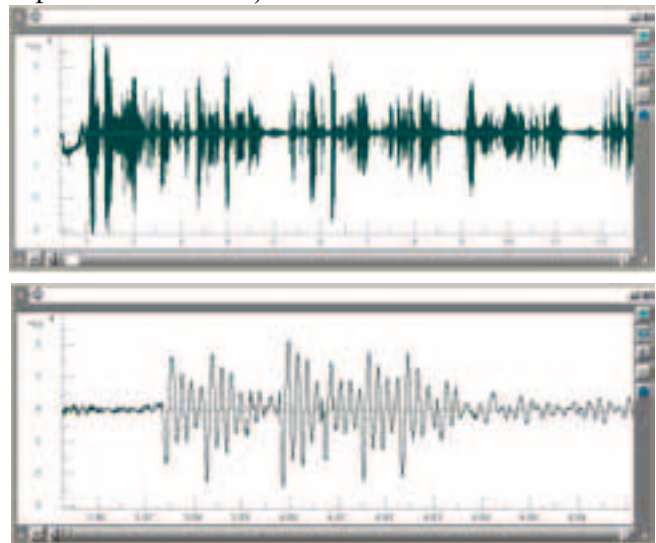


FIGURE 2. PRESENTATION OF THE SPEECH SIGNAL IN A 12 SECOND PERIOD AND ITS ONE SECOND FIGURE

Instrumental analysis is done by an engineer using modern computer technologies. Voice recording that is to be processed is usually already digitalized and depending on the way it was digitalized it is necessary for both recordings (the controversial and non-controversial one) to be compared and adjusted to similar characteristics, (Figure2.). That, as well as



other operations that are performed during instrumental analysis, are done by the use of special software adapted for forensic speech identification. After the recording adjustment, frequency of the speaker's voice is specified and compared to each other. Furthermore, signal spectrogram [7] (or digital voice print) is observed and compared to format and shape values of similar or identical voices on the controversial and non voice recording (Figure 4).

The whole instrumental analysis is based on the application of Furriers transformations which form the base of signal transformation from a time frame into a spectral field. Similar voice identification is performed in automatic systems for the speaker identification, with a difference that forensic identification requires an expert, compared to an automated identification which is done by a computer.

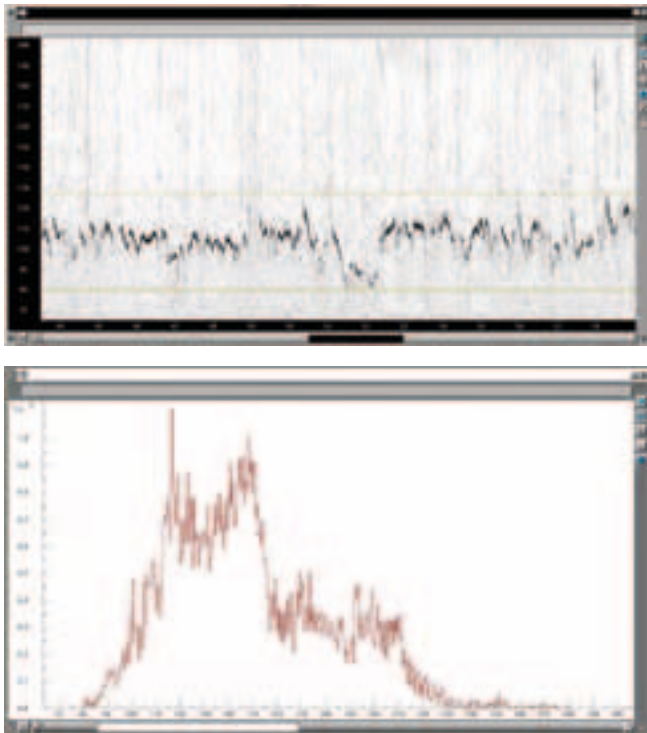


FIGURE 3. CEPSTROGRAM AND HISTOGRAM OVERVIEW OF THE BASIS FREQUENCY OF SPEECH SIGNAL

It is important to say that it is necessary to get consent from the suspect in order for this analysis to be performed.

In order for a suspect's voice sample to be adequate for the analysis, it is necessary for a sample to

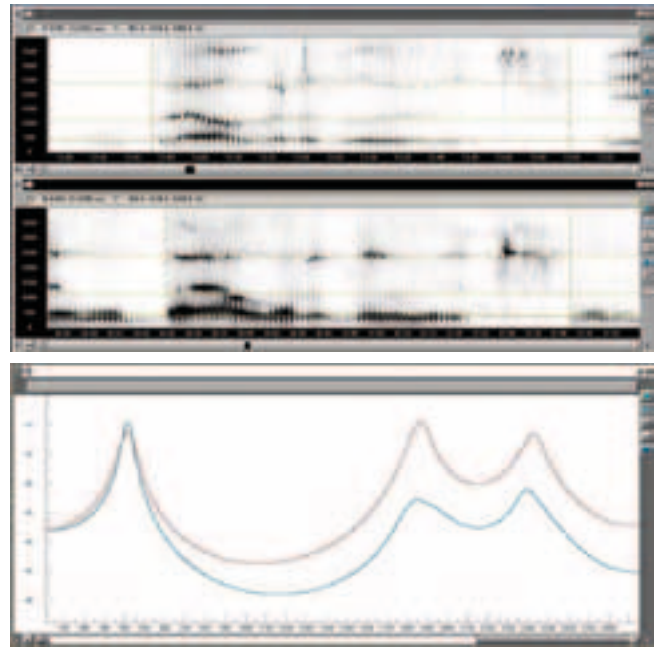


FIGURE 4. SPECTROGRAM OF THE SPOKEN SENTENCE OF CONTROVERSIAL AND NON-CONTROVERSIAL RECORDING OF THE FIRST THREE VOICE FORMATS

contain all of speech characteristics such as accent and different speech patterns under various emotional states. This is why investigations require three different ways of recordings. Suspect is first required to read a certain text, which is followed by a sentence repetition which is dictated to him. These sentences are specifically suitable for the analysis since they instigate suspect's voice to be tested in natural circumstances.

During the analysis different markers are used, but the voice pitch cannot be used as a basis for a conclusion that a person in question is the same one since the voice changes depend on the circumstances. Here is an example: Video surveillance camera recorded a thief during a night time robbery by which, the high tone of his voice was recorded. After his arrest, his voice was low since he was scared and tired. This leads us to a conclusion that the tone of the voice is not an element by which we can claim that it is the same person that we are trying to identify. Even if the tone of the voice is confirmed, it cannot automatically serve as evidence, given those two completely different circumstances.

It is also important to determine the content of spoken words especially if the sound recording is of poor

quality or if the person speaking has a foreign accent. During the identification it is important to remove the presence of other sounds such as crying, breathing, barking, etc. Moreover, it is interesting to say that during the analysis forensics compare and process voices by describing voices in the context of phonetic units-vowels and consonants. For example, forensic phonetic expert can conclude that a vowel "I" is different in two samples or that a consonant "T" can be pronounced in a specific way in both samples. Forensics analyzes taken material from the language point of view as well as non language characteristics such as voice quality, tone, pitch, intensity frequency, strength and quality of speech. The quality of speech determination is based on pronunciation, accent, consonant and vowel pronunciation, as well as sounds such as p,t,k,b,d,g which require pronounced intonation etc.

It is also important to say that this biometry is applied in the field of executing court decisions on parole prisoners. Parole prisoners are required to "call-in" to a machine at a certain time to confirm their presence in a specific place. State of Louisiana requires parole prisoners to call an official and speak a certain phrase in order to confirm that they are at a specific place [8].

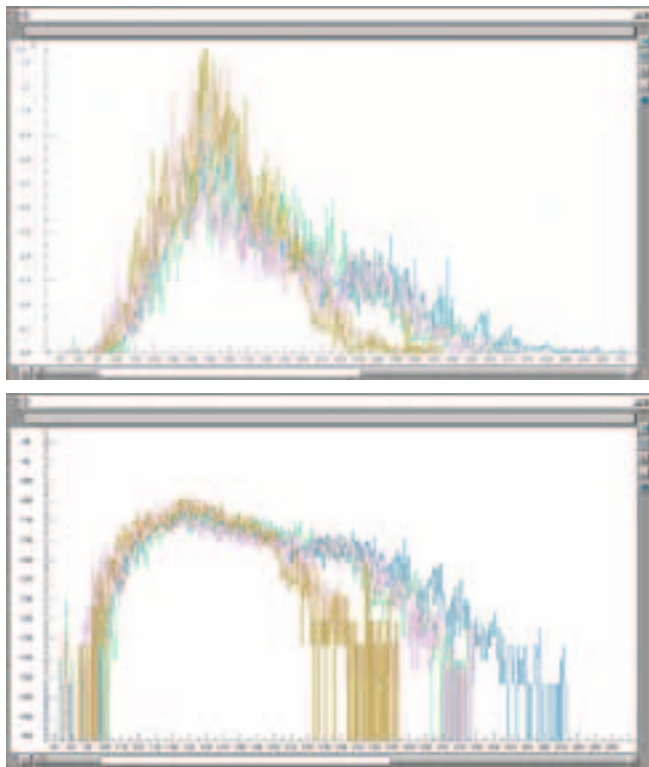


FIGURE 5. HISTOGRAMS FOR THE VOICE FREQUENCY OF THE SAME PERSON IN LINEAR AND DECIBEL SCALE

## E-transactions

E-transactions are becoming more frequent ways of payment in a contemporary world in comparison to credit cards. This enables payments to be processed quicker; however, it also increases a chance of fraud, which can be decreased by applying this type of biometry. This topic will be covered later on.

## On-line shopping

This type of biometry enables additional safety measure during on-line shopping. In other words, in today's world, on line shopping is becoming more frequent since it does not require a lot of time and it can be done "from the chair". However, this type of shopping is followed by various risks such as unauthorized credit card use, which means that somebody other than us can get a hold of our credit card and purchase items instead of us. This problem is solved by a Visa credit card company by applying voice biometrics prior to online purchases to identify and confirm a credit card user. Besides voice biometrics, the company is offering to its customers changing of their passwords via telephone which is time and cost efficient [8].

## E-bay shopping

During online e-bay shopping, customer has to follow certain rules so that his/her site entry is not forbidden in the future [8]. However, this problem is not solved, since the same person can open another account and have access to the same web site. By applying this biometry, problem of this kind can be easily solved by making a user to say a certain word into a microphone during a registration process. This means that the access will be granted to a user, only if the system recognizes and confirms user's identity.

## Other applications "Smart" cars

Besides smart cell phones, other type of mass use is a smart car where communication is enabled through build-in electronic equipment via voice commands. This equipment includes FM radio, CD, air conditioning, ventilation as well as power windows, seats, lights, etc. Modern technology is enabling built in satellite navigation which gives a driver the opportunity

to be navigated to a specific place, telephone and internet lines as well as other entertainment systems such as TV, DVD, video games. However, it is important to say that these contemporary technologies should not be used while driving since they can distract a driver.

This type of biometry can be applied instead of card while entering or exiting the building or a garage. In other words, saying a certain word is safer and faster way than a card swiping which can be damaged or lost.

### **Applications in school system**

Voice biometry is applied in schools in the United States since 2007. It was used in distributing school lunches or snacks to students. They would speak out their name and ID number which would be matched to their snack or lunch.

### **Voice biometrics in a contemporary business world**

Voice biometrics is contributing to the employee productivity as well, which leads to a higher company profit. When we are writing or processing a certain report we do it by typing in the text into a computer. That is the way we send an e-mail which requires a certain amount of time. By applying this biometric method, much greater efficiency is accomplished in the particular case, because the time needed for some of the above mentioned actions is decreased, considering that it is much faster and easier to dictate the text than type it in. Research has shown that by using this method we can decrease the amount of time needed to finish the text of 900 words by 6, 5 minutes [2].

Some companies, such as Bell Canada, are enabling access to certain data to their employees by authenticating their identity through a phone call. This means that, by applying this biometry employees are not required to use a lap top so that they can access the desired data. Moreover, this represents more secure and faster way of accessing information.

One of the advantages can be illustrated in the following example: employees in various companies forget passwords and the procedure that requires their change is expensive. Per Say Company has pointed out that other companies spend millions of dollars on

the password maintenance that their employees use to access company's database. This kind of biometry enables employees to change their passwords by themselves without calling technical support. Password change can be executed in the following way: password change is processed when a system poses a question to a user who wants to change his/her password by authenticating it. So, if the voice is recognized by the system, the employee will be allowed to change his/her password.

The Gartner Group research has shown that 20-50% of the calls to a technical support are about password change. From the economic point of view, the average cost of password change amounts up to 32 USD per call, and the application of automated voice recording would decrease this average price to only 0.50 per call [3].

The application of voice biometry also enables auditors to access company's data by comparing their voices to a specific sample in the database [6].

In the field of social insurance, in the US, employers pay benefits for their employees online with the use of PIN number and voice sample which increases safety and security.

### **ADVANTAGES**

Each method has its advantages and disadvantages. As far as voice biometry is concerned, research has shown that the voice is more reliable than codes. Also, it is less likely that the voice sample will fall into the wrong hands and be used for fraud.

Furthermore, voice biometry is simple and inexpensive to use since it does not require special hardware; all you need is a computer and a microphone, and most importantly it is accepted by society.

It is important to emphasize that this is the only biometrical method which gives a user an option to use the system from far away. Furthermore, the inscription process into a system is short and the authentication is quick. Moreover, biometric sample requires very small space in a computer database, as well as SIM cards in cell phones.

Today's software packages enable users to convert their voice into a text, which is directly incorporated into a word processing software or e-mail. Also, all menu functions can be voice accessed.

Physically challenged persons who are not able to use their hands to type or have vision problems can benefit from Automatic Voice Recognition.

## DISADVANTAGES

Besides the above mentioned advantages there are some disadvantages to this method. The one such is a possibility that somebody records someone's voice and commits a fraud. However, the percentage of unauthorized use is very small which is supported by IBM research center. Their research shows that users, who use this method while registering online, namely hackers, would be falsely identified as real users in 0.00001% of cases [11]. In order to prevent this type of an "attack", system can ask the user to repeat a random choice of words or phrases in a certain order.

Furthermore, the disadvantage of the voice biometry is that the voice is susceptible to changes which can lead to an error. Basic factors that can lead to voice modification are [4]:

- Age (voice changes over the years);
- Disease (voice changes due to a flu, sore throat, etc)
- Acoustics (voice determined by an environment where authentication is performed)
- Person's emotional state (voice changes due to stress)
- Incorrect pronunciation of previously determined words or phrases
- Distance from the microphone or use of various types of microphones

## CONCLUSION

Over time, the technology has been developing, so along with the need for transfer of spoken information, wide scope of possibilities of voice utilization has been developed, such as the "means" of personal identification, issuing orders at robotic

machines, security measures, parameters in poly-graphic research, etc.

Nowadays, the voice as the sound that carries certain information is not the only thing that matters, but voice biometrics is equally important – the numeric model of sound and pronunciation characteristic of each individual. It enables us to use voice as a unique and unchanging identifier in the world of modern machines and computers.

Like any other biometric model, this one is not perfect either, it has certain advantages and disadvantages. The advantages are its reliability, easy to use, price – inexpensive for implementation, social acceptance, etc. The shortcomings are very few and found in the fact that voice is susceptible to changes, while software and algorithms based on the voice biometrics are still not robust enough.

As there is striving for the improvement of both hardware and software, we can expect that besides the methods relating to the voice verification, we will also see the use of programs for personal identification on the basis of voice, i.e. along with the identification based on papillary lines there will be equally important identification by voice available. This possibility will significantly contribute to the curbing of crimes related to threats, ransom and extortion demands.

With the advent of technologies and computer capacities, voice biometrics is gaining in importance. New possibilities of use are appearing, advantages are being discovered and shortcomings removed. The popularity of this biometric model is growing, so in the future, voice biometrics will hold a significant position in all areas of social functioning.

### *Authorship statement*

*Author(s) confirms that the above named article is an original work, did not previously published or is currently under consideration for any other publication.*

### *Conflicts of interest*

*We declare that we have no conflicts of interest.*



**REFERENCES:**

- [1] CCERT-PUBDOC-2006-09-167, article from the web site: <http://www.cert.hr/sites/default/files/CCERT-PUBDOC-2006-09-167.pdf>, accessed on 10.09.2012.
- [2] Dragon Naturally Speaking Review, article from the web site: <http://www.dragonnaturallyspeakingreview.com/> accessed on 23.11.2011.
- [3] Gonzalez-Rodriguez, J., Toledano, D.T., Ortega-Garcia, J. (2008). Voice biometrics, Handbook of Biometrics, Springer, pp. 151-170.
- [4] Gravnes, H. (2005), User's trust in biometric Authentication Systems, Master Thesis, Gjøvik University College, Norway
- [5] Jovicic, S. (1999), *Govorna komunikacija*, Nauka, Beograd (*in serbian*)
- [6] Markowitz, J. A. (2000), Voice biometrics, *Communication of the ACM* 43, 9, 66-73.
- [7] Myers, L. (2004), An Exploration of Voice Biometrics, article from the web site: [http://www.sans.org/reading\\_room/white-papers/authentication/exploration-voice-biometrics\\_1436](http://www.sans.org/reading_room/white-papers/authentication/exploration-voice-biometrics_1436), accessed on 15.10.2012.
- [8] Spectrogram, article from the web site: <http://en.wikipedia.org/wiki/Spectrogram>, accessed on 14.04.2010
- [9] Vasilijevic, A., Petrinovic, D. (2011), Perceptual Significance of Cepstral Distortion Measures in Digital Speech Processing, *Automatika*, 52(2), 132–146.
- [10] Vaskovic, V., Todorovic, M. (2008), Biometric payment technology, *Zbornik YUInfo*, (*in serbian*)
- [11] Voice Biometric Authentication Systems, article from the web site: [http://www.authenticate.com/solutions/voice\\_biometrics.html/](http://www.authenticate.com/solutions/voice_biometrics.html/), accessed on 18.10.2011.

Submitted: November 11, 2012.

Accepted: December 4, 2012.