Technical Note

Impact of the Passage of Time on the Correct Identification of the Speaker Using the Auditory Method

Stefan BRACHMAŃSKI* , Bartosz HUS, Piotr STARONIEWICZ

Faculty of Electronics, Photonics and Microsystems, Department of Acoustics, Multimedia and Signal Processing
Wrocław University of Science and Technology
Wrocław, Poland

*Corresponding Author e-mail: stefan.brachmanski@pwr.edu.pl

Courts in Poland, as well as in most countries in the world, allow for the identification of a person on the basis of his/her voice using the so-called voice presentation method, i.e., the auditory method. This method is used in situations where there is no sound recording and the perpetrator of the criminal act was masked and the victim heard only his or her voice. However, psychologists, forensic acousticians, as well as researchers in the field of auditory perception and forensic science more broadly describe many cases in which such testimony resulted in misjudgement. This paper presents the results of an experiment designed to investigate, in a Polish language setting, the extent to which the passage of time impairs the correct identification of a person. The study showed that 31 days after the speaker's voice was first heard, the correct identification for a female voice was 30% and for a male voice 40%.

Keywords: speaker recognition; crime acoustics; aural identification.

1. Introduction

Most courts around the world allow auditory identification of a person, i.e., testimony in which the witness is able to identify the speaker or other auditory impression. Auditory identification is one of the oldest methods of identifying a speaker from the voice. It was the first method accepted by courts of various countries, e.g., in the USA, in the state of Florida it has been used since 1907 (HOLLIEN, 1990; 2002). Auditory identification is a complex technique due, among other things, to the temporal fluctuation of speech features and parameters caused by the psychophysical state of people involved in an event, as well as external acoustic conditions (HOLLIEN, 1990; 2002; HOLLIEN *et al.*, 2016). In Poland, this method is used when there is no sound recording and the identification of a person can only be made on the basis of an auditory assessment by the injured person or by witnesses to the event (The Code of Criminal Procedure [Kodeks Postępowania Karnego], 2016). Auditory assessment is also used in linguistic-measurement (BŁASIKIEWICZ,

1971; DOLECKI, RZESZOTARSKI, 2002) and auditory-spectral (ALEXANDER *et al.*, 2005; BEGAULT, POZA, 2005; MCDERMOTT, OWEN, 1996; ROSE, 2002) methods. In these methods, at the auditory assessment stage, attention is paid, among other things, to the sound of voices being compared, the manner of accentuation, the rate of speech, pronunciation defects and the manner of utterance are analysed. Psychologists, phonoscopy specialists, and researchers in the field of auditory perception and forensic science more broadly describe many cases where this testimony has resulted in a wrongful conviction (ELMORE, 2020; POSSLEY, 2018).

Humans are able to recognise speakers based on their voice with varying degrees of effectiveness. Many factors affect the reliability of this method, namely familiarity with the speaker, duration of the speech sample, context, emotion, pronunciation defects, etc. (DEFFENBACHER, 1989; HOLLIEN, SCHWARTZ, 2000; YARMEY *et al.*, 2001). Professor Frances McGehee, through the events of 1935 (Conviction of Bruno Hauptmann in the case of the kidnapping and mur-

der of Charles Lindbergh Junior (HOLLIEN, 1990; The State of New Jersey v. Bruno Richard Hauptmann, 1935; VAN WYK, 1953)), conducted an experiment in which she wanted to prove the thesis that a person is unable to recognise an unfamiliar voice after a considerable lapse of time from its first hearing.

To this day, there are still expert claims that after 29 months, as in the case of Charles Lindbergh's identification of Hauptmann, auditory identification of the other person is impossible. McGehee, with her research, showed that the speaker's voice identification on the next day was quite high (83%), but that the effectiveness of the identification gradually declines over time, reaching only 13% after five months. Many people including Harry Hollien point out that Frances McGehee in her experiment did not take into account several important aspects such as the fact that the accuracy of identification can be affected by the appearance of an additional stimulus when hearing the speaker – emotional involvement or the ability of different people to remember the voice (ELMORE, 2020; HOLLIEN, 1990; 2002).

Later studies (BRICKER, PRUZANSKY 1966; HOLLIEN, 2002; PRUZANSKY, 1963) confirmed the conclusions of McGehee's research. However, these studies did not faithfully reproduce the experiment performed by McGehee.

The aim of the experiment presented in this paper was to investigate, under the Polish language conditions, how the passage of time affects the correct identification of men and women when the recogniser does not know the speaker's voice and when the speaker's voice is well known.

2. Study by Frances McGehee

The conviction of Bruno Hauptmann for the kidnapping and murder of little Charles Lindbergh Junior, and more specifically the fact that it was based on evidence of voice identification, initiated a series of experiments aimed at confirming the ability of humans to remember the voice of a speaker over the long term. One of the most famous experiments was an experiment performed by Frances McGehee in 1937 (McGEHEE, 1937; HOLLIEN, 1990). The study involved 740 students (554 men, 186 women), while there were 49 speakers (31 men, 18 women). The study participants were divided into 15 groups. Each group was assigned a number of days until the next listening day, i.e., a lapse of time from 1 day to 5 months. The listeners' task was to recognise and indicate which of the five voices they had heard previously. The speakers presented to the listeners were selected from 49 people.

The thesis she put forward in her research can be formulated as follows: "Humans are unable to recognise an unfamiliar voice after a significant lapse of time from when they first hear it". To confirm this thesis, McGehee performed an experiment consisting of two parts. In the first part of the experiment, listeners heard a 56-word text read by a single speaker sitting behind an opaque screen. After a set amount of time, the group members heard the same sequence read in random order by five speakers (one identified and four whom the listeners had never heard). The listeners' task was to write down the number of the speaker they thought they had originally heard. McGehee repeated this experiment with the difference that the speakers presented to the listeners were prerecorded on tape. In both experiments, the results were very similar, with the effectiveness of identification decreasing as time passed. The correct identification after 1 and 2 days was 83% and after 7 days 81%. A noticeable decrease was found after 2 weeks, when the correct identification was 69%, dropping to 51% after 3 weeks and 35% after 3 months. The last period studied was after 5 months when the correct identification dropped to 13% (McGEHEE, 1937; HOLLIEN, 1990).

McGehee's research showed that the correct speaker identification depends on the time which elapsed between hearing the voice of the person being identified and attempting to recognise the speaker, as well as the listener's ability to remember the voice pattern.

3. Experiment I – Recognition of an unknown speaker

The test material was a passage from the book "Norse Mythology" by Neil GAIMAN (2017) read by five female and five male speakers. The speakers were selected based on a subjective assessment of the similarity of the voice tone, as well as the value of the laryngeal tone and the first four formants. The utterances were recorded on a digital recorder at the sampling rate of 44.100 samples/s and a resolution of 16 bits in PCM (wav) format. Before recording, each speaker practised reading the text to ensure fluency. The text was read out in an even, calm voice. The recordings were made in a home environment, in a quiet room isolated from external distractions. The utterances, i.e., the text being read out, were recorded five times. From all the recordings, one with the best-sounding utterances, without stammers, repetitions or uncontrolled artefacts, was selected for each reader (HUS, 2022).

Due to the state of epidemic emergency prevailing both nationally and internationally caused by the COVID-19 coronavirus (SARS-CoV-2 virus), the entire research process took place online from sending the message about the start of listening to sharing the evidence and comparison recording and receiving feedback with the listener's response.

The female voice recognition study involved 100 participants (65 male, 35 female) who were randomly divided into 11 smaller study groups. In contrast,

150 participants (88 male, 62 female) took part in the male voice recognition study, who were randomly divided into 11 smaller study groups of 10 participants each. The ages of the study participants ranged from 21 to 26 years for the female voice identification and 20–30 years for the male voice identification. The speaker identification group was formed from the candidates who passed the so-called zero test. This consisted of each candidate being presented with the statements of all five speakers (the male voice to the candidates in the male identification group and the female voice to the female identification group) 30 seconds after hearing the speech of the person being identified. Each candidate had to correctly identify the speaker being recognised. If the speaker identification was not correct then the candidate was not included in the study group.

The recording of the identified speaker was presented on the same day and at the same time to all study participants, with a female and a male identification group separately. After listening to the recording, it was deleted from the folder provided to the listeners. The participants in the experiment were therefore not given the opportunity to listen again to the recording read by the identified speaker. After the time set for the group had elapsed, the group members were informed that the listening window had been opened and the speaker's identification should be made. Each group member listened to a prerecorded text read by five speakers. After listening to all the voices, he or she indicated the number of the speaker whose voice, in his or her opinion, corresponded to the voice heard for the first time. In addition, each listener provided a degree of confidence in the identification. The information provided by the listener was automatically entered into the measurement form. All the information entered automatically updated the results table, which included:

– information about the correct identification of the suspect for a given listener;
– the total number of listeners who correctly identified the suspect;
– the number of people who correctly identified the suspect after a certain time;
– information about which speaker the people tested pointed to most often;
– information specifying the number of people not tested.

The recognition of the female voice was performed after: 1, 2, 3, and 7 days, 2 and 3 weeks, and 1 month after the first hearing of the recording of the recognised speaker's utterances, and the male voice additionally after 2, 3, and 4 months.

Figure 1 presents the results obtained from the experiment presented and reported by McGehee. The recognition performance of the female voice is more dependent on the passage of time than that of the male voice. After one day of hearing the female voice, 90% of the listeners made the correct identification, whereas for the male voice, all the listeners correctly recognised the person being identified. As the time passed, the recognition success rate decreased, so that after seven days the female voice was 60% and the male voice 90%; McGehee's figures were 81%. On the other hand, after
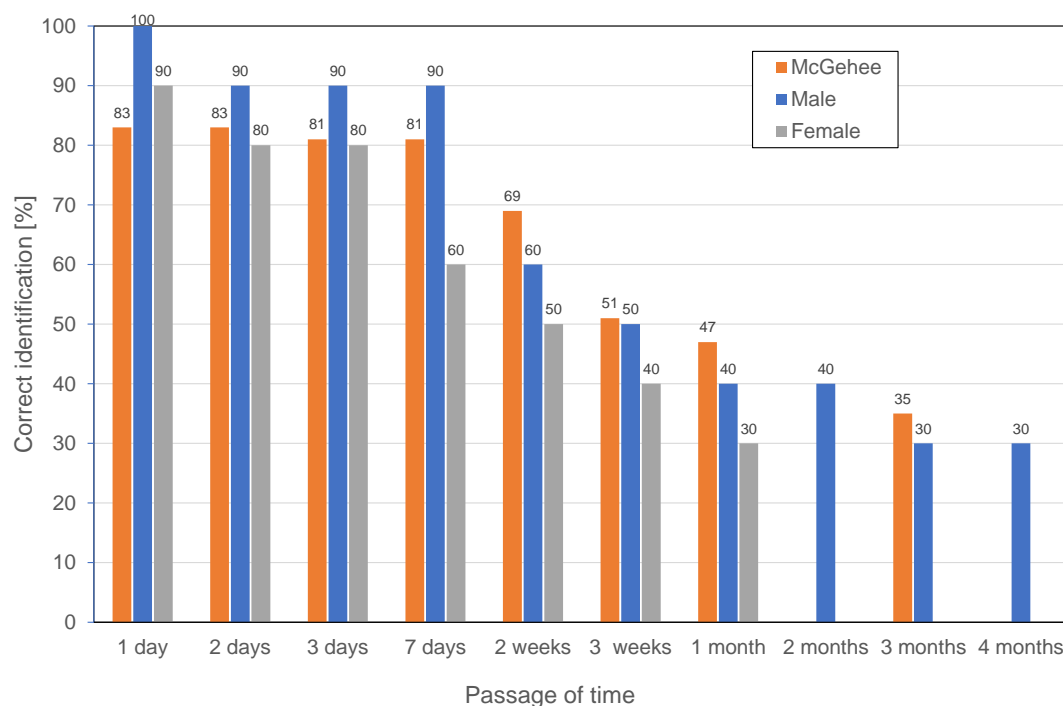

Fig. 1. Effectiveness of identifying an unknown person as a function of the passage of time.

one month, the recognition rate decreased very significantly, with the female voice at 30% and the male voice at 40%. The speaker recognition performance 1 month after the first hearing of the male voice was similar to McGehee's result (40% and 47%, respectively).

Analysing the summary results of the speaker identification by all the listeners, it was found that the correct identification of the male speaker was marginally better than that of the female speaker (Fig. 2). The correct identification of the female was 58% and that of the male was 62%, a difference of 4% in favour of the male voice.
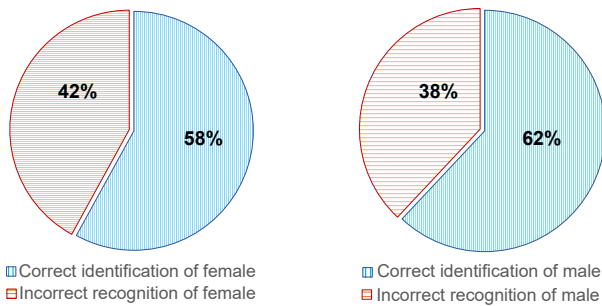

Fig. 2. Percentage of correct and incorrect identifications of female (a) and male (b) voices.

The results were analysed to see what effect the gender of the person identifying the speaker has on the correct identification. As a result of this analysis, it was found that there is no significant difference whether the speaker is recognised by a woman or a man (Fig. 3). A difference of 2% in the effectiveness of the person identification is within the statistical error range.
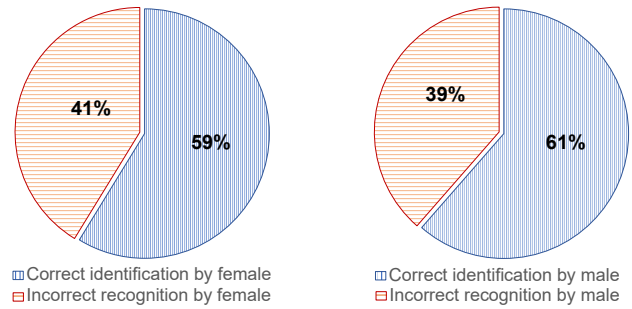

Fig. 3. Percentage of correct and incorrect voice identifications by female (a) and male (b).

4. Experiment II – Recognition of a known speaker

The second part of the experiment concerned the recognition of a speaker whose voice was previously known to the listeners. Hundred participants (54 men, 46 women), the age range 15 to 55 years, took part in the recognition of the female voice. Each person in this group had contact with an identified female at least once every fortnight. Ninety people (63 men, 27 women) participated in the male recognition, the age range 20 to 45 years. As in the female recognition group, also in this case each person in this group had contact with the male suspect at least once every fortnight.

This part of the experiment used the same test material as in the first part of the study, i.e., the recognition of the unknown speaker. The speaker recognition procedure was the same as in experiment I.

The results of this part were statistically analysed and the results are shown in Fig. 4. This part of the
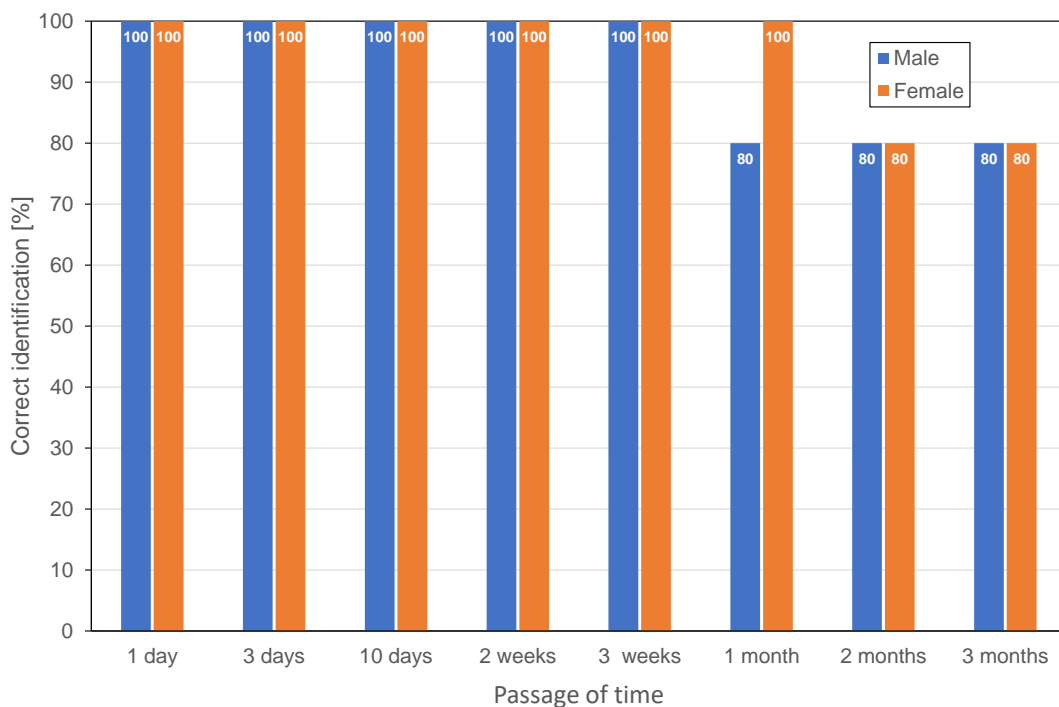

Fig. 4. Effectiveness of identifying a known person as a function of the passage of time.

experiment showed that if the voice of a well-known speaker is identified, even after 1 month it is possible to correctly identify both male and female speakers. It is only after 2 months that the identification efficiency of the familiar speaker drops to 80%. Analysing the results of the speaker identification carried out by all the listeners, it was found that the correct identification is not influenced by the gender of the person being identified (Fig. 5). When the speaker is known, the difference between female and male identification is only 1%, which is within the statistical error range.
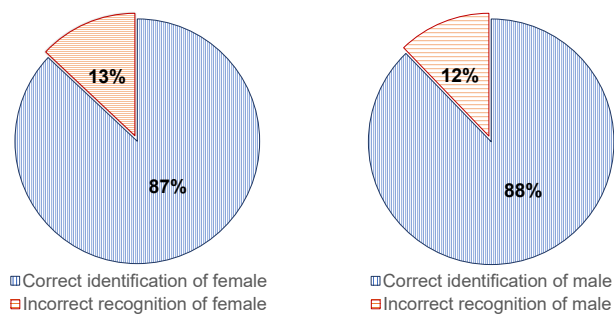

Fig. 4. Percentage of correct and incorrect identifications of voice female (a) and male (b).

In experiment II, taking into account the results of experiment I, the results were not analysed for the effect of the gender of the person identifying the known speaker on the correctness of the identification.

5. Conclusions

The research confirmed the conclusions of McGehee's experiment that as the time passes after hearing a speaker's voice, the speaker's recognition efficiency decreases rapidly. Comparing the results obtained in experiment I with those of Frances McGehee's study, it can be concluded that under the Polish language conditions the identification efficiency of the female voice decreases faster, while that of the male voice at a comparable rate. In the first three days, the correct identification of both female and male voices exceeded 80%. The very high efficiency of male voice identification persisted for one week (90%), but after two weeks there was a decrease to 60% (69% in McGehee's study), and after one month it decreased to 50% (47% in the McGehee study).

The resulting convergence of the results was expected in the light of Ebbinghaus' research on memory. In his classic work, he presented quantitative data on the decay of stored material over time (EBBINGHAUS, 1885). The conclusion that the number of remembered items decreases with time has been confirmed by other researchers (FALKOWSKI, 2004; IWANICKA, 2020). According to the Ebbinghaus curve, also known as the forgetting curve, which shows the relationship between the amount of information stored in memory and the

time elapsed since hearing it, a person is able to reconstruct a limited number of units heard, e.g., after 5 days only 25% of the units heard, and after 30 days 20% (EBBINGHAUS, 1885). Stressful circumstances can affect learning and memory processes. However, the nature of the effect of stress on memory is not fully understood, as both memory-enhancing and memory-impairing effects have been reported (SCHWABE *et al.*, 2012). The memory curve is language-independent and can be adapted to many branches of learning related to perception (EBBINGHAUS, 1885; FALKOWSKI, 2004). Thus, it can be assumed that the Ebbinghaus curve also applies to the ability to remember the sound of the voice, including the auditory identification of the speaker. It should be noted, however, that in the case of remembering the sound of the voice, the curve falls much more slowly than the Ebbinghaus curve. Humans are able to remember the artefacts of the speaker's voice for longer than learned speech units, and especially in situations of emotional involvement. It can be assumed that auditory identification of a speaker does not depend on a language when it is made by speakers of the same language as the person being identified.

The results of experiment II showed that if the recognised voice was previously known to the identifier then 100% correct identification of the female voice was maintained over a period of 1 month. The high value of correct identification was still maintained after 3 months and was 80% for both female and male voices.

The results obtained in both experiments I and II presented here, as well as in the McGehee study, cast considerable doubt on the validity of identifying a person solely on the basis of a voice two years after hearing the voice for the first time (as in the case of Charles Lindbergh). However, retention, i.e., the ability to remember, especially under conditions of threat or personal involvement, cannot be overlooked. In Lindbergh's case, the production of a significant dose of adrenaline, which sharpened the hearing and also enhanced the ability to remember the voice for a long time, may have played an indiscernible role in remembering the voice of the abductor. Hollien's research shows that in a stressful situation, a person can remember the sound of a voice for a very long time. Therefore, it cannot be ruled out that in individual cases it is possible to recognise a voice even after 2 years, as was the case with Lindbergh.

In the process of auditory speaker identification, the technique of remembering the speaker's voice is also important. In the experiment in question, one listener used the technique of associating the voice he heard with the voice of a person he knew well. When listening to recordings of five people, including the identified person, the listener looked for the speaker whose voice best reproduced that of the person close to him. This listener correctly identified the suspect after 40 days and rated his confidence of identification at 10.

In summary, it can be said that speaker identification deteriorates very quickly when it is made by people who do not know the speaker and are not emotionally involved in the event, whereas it persists for a longer period of time in people who know the speaker.

In addition to further research into the effect of the passage of time on the effectiveness of speaker recognition by listeners, future work will focus on factors studied so far for automatic methods. These are factors masking the personal parameters of the speaker's voice, such as the influence of the speech coding and transmission techniques used (Jarina *et al.*, 2017), voice disguise techniques or the speaker's state or condition (Staroniewicz, 2021).

References

1. Alexander A., Dessimoz D., Botti F., Drygajlo A. (2005), Aural and automatic 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.

2. Begault D.R., Poza F. (2005), Voice identification and elimination using aural-spectrographic protocols, [in:] *Audio Engineering Society Conference: 26th International Conference: Audio Forensics in the Digital Age*, www.aes.org/tmpFiles/elib/20220803/13256.pdf (access 03.08.2022).

3. Błasikiewicz S. (1971), The listening method for whispering and intensely distorted speech [in Polish: Metoda odsłuchu szeptu i mowy intensywnie zakłóconej], *Problemy Kryminalistyki*, **90**: 159–183.

4. Bricker P., Pruzansky S. (1966), Effect of stimulus content and duration on talker identification, *Journal of the Acoustical Society of America*, **40**(6): 1441–1450, doi: 10.1121/1.1910246.

5. Deffenbacher K.A. *et al.* (1989), Relevance of voice identification research to criteria for evaluating reliability of an identification, *Journal of Psychology*, **123**(2): 109–119, doi: 10.1080/00223980.1989.10542967.

6. Dolecki J., Rzeszotarski J. (2002), Use of the language-measurements method for the identification of people in phonoscopic examinations, *Problems of Forensic Sciences,* **LII**: 108–123, www.forensicscience.pl/pfs/52_dolecki.pdf (access 03.08.2022)

7. Ebbinghaus H. (1885), *Memory: A contribution to Experimental Psychology* [in German: *Uber das gedachtnis: Untersuchungen zur experimentallen psychologie*], Duncker and Humboltd.

8. Elmore T.L. (2020), *The effect of cued memory recognition strategies on word and speaker identification*, Ph.D. Thesis, University of London.

9. Falkowski A. (2004), Memory and knowledge in the context of scientific cognition development [in Polish: Pamięć i wiedza w kontekście rozwoju poznania naukowego], *Nauka*, **2**: 105–124.

10. Gaiman N. (2017), *Norse Mythology* [in Polish: *Mitologia Nordycka*], MAG.

11. Hollien H. (1990), *The Acoustics of Crime: The New Science of Forensic*, Kluwer Academic – Plenum Publishers, New York.

12. Hollien H. (2002), *Forensic Voice Identification*, Academic Press, London – San Diego.

13. Hollien H., Didla G., Harnsberger J.D., Hollien K.A. (2016), The case for aural perceptual speaker identification, *Forensic Science International*, **269**: 8–20, doi: 10.1016/j.forsciint.2016.08.007.

14. Hollien H., Schwartz R. (2000), Aural-perceptual speaker identification: Problems with noncontemporary samples, *Forensic Linguistics*, **7**(2): 199–211, doi: 10.1558/sll.2000.7.2.199.

15. Hus B. (2022), *The impact of the passage of time on auditory identification of a person* [in Polish: *Analiza wpływu czasu na audytywną identyfikację osoby w akustyce kryminalistycznej*], MA Thesis, Wroclaw University of Science and Technology.

16. Iwanicka K. (2020), Psychological basis of working with a group, with a particular emphasis on the issue of learning styles [in Polish: Psychologiczne podstawy pracy z grupą ze szczególnym uwzględnieniem zagadnienia stylów uczenia się] [in:] *Pedagogiczne aspekty edukacji filozoficznej z uwzględnieniem systemów penitencjarnych*, Lipski D., Malinowski M. [Eds.], pp. 105–113, Wydawnictwo DiG, https://swws.edu.pl/wp-content/uploads/2021/05/3.-Pedagogiczne-aspekty-edukacji-filozoficznej.pdf#page=105 (access 13.08.2022).

17. Jarina R., Polacky J., Pocta P., Chmulik M. (2017), Automatic speaker verification on narrowband and wideband lossy coded clean speech, *IET Biometrics*, **6**(4): 276–281, doi: 10.1049/iet-bmt.2016.0119.

18. McDermott M.C., Owen T. (1996), *Voice identification: The aural/spectrographic method*, Frank M. McDermott, Ltd., Owl Investigations, Inc. (access 03.08.2022).

19. McGehee F. (1937), The reliability of the identification of the human voice, *Journal of General Psychology*, **17**(2): 249–271, doi: 10.1080/00221309.1937.9917999.

20. Possley M. (2018), *Wilber Jones*, The National Registry of Exonerations, www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=5392 (access 01.08.2022).

21. Pruzansky S. (1963), Pattern-matching procedure for automatic talker recognition, *The Journal of the Acoustical Society of America*, **35**(3): 354–358, doi: 10.1121/1.1918467.

22. Rose P. (2002), *Forensic Speaker Identification*, Taylor & Francis Inc, London – New York, doi: 10.1201/9780203166369.

23. Schwabe L., Joëls M., Roozendaal B., Wolf O.T., Oitzl M.S. (2012), Stress effects on memory: An update and integration, *Neuroscience & Biobehavioral Reviews*, **36**(7): 1740–1749, doi: 10.1016/j.neubiorev.2011.07.002.

24. Staroniewicz P. (2021), Effect of deliberate and non-deliberate natural voice disguise on speaker recognition performance, *Acoustics, Acoustoelectronics and Electrical Engineering*, **888**: 312–325.

25. *The Code of Criminal Procedure* [in Polish: *Kodeks Postępowania Karnego*] (2016), Dziennik Ustaw, 1749.

26. *The State of New Jersey v. Bruno Richard Hauptmann* (1935), Famous Trials by Professor Douglas O. Linder, https://famous-trials.com/hauptmann/1393-hauptmann appeal (access 28.07.2022).

27. Van Wyk H.S.J. (1953), State of New Jersey v. Bruno Richard Hauptmann, *The South African Law Journal*, **70**: 182, https://heinonline.org/HOL/LandingPage?handle=hein.journals/soaf70&div=30&id=&page= (access 16.07.2022).

28. Yarmey A.D., Yarmey A.L., Yarmey M.J., Parliament L. (2001), Commonsense beliefs and the identification of familiar voices, *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, **15**(3): 283–299, doi: 10.1002/acp.702.