ISSN 1230-7483
e-ISSN 2720-5983

Problems of Forensic Sciences 2025, vol. 142, 143, 157–183

DOI: 10.4467/12307483PFS.25.008.22915

EFFECT OF MUSICAL EDUCATION ON ACCURACY OF IDENTIFICATION IN A VOICE PARADE

Piotr WOLDAN

Trainee advocate at the Kraków Bar Association

Abstract

Voice lineup is a possible form of an identification parade conducted during an investigation. It could occur in cases where the perpetrator's appearance was not observed by the victim or witness, and only the voice was heard. Some of the factors that could impact the accuracy of voice recognition, such as age of the witness, retention interval and whether the voice was heard live (at crime scene) or through a mobile phone, have already been the subject of some studies.

The aim of the study was to examine the impact of the following factors: music education, retention interval and gender. The experiment was conducted with a total of 129 participants, both musically and non-musically educated. The participants were asked to listen to a short voice recording simulating a true event and then (after the retention interval of 1 week or 2 weeks) to listen to a recorded lineup of 6 female voices. Following that, they were asked to try to identify the perpetrator's voice. The results revealed no statistically significant difference between musically educated and non-musically educated participants, and similarly, no significant difference was found based on gender or confidence rating. A nearly significant decrease in accuracy was observed among the participants with a longer retention interval. Nevertheless, the maximum level of accuracy was only 55% which implies low reliability of voice recognition evidence. There results must be taken into consideration by authorities if a voice lineup is used, especially to treat voice recognition evidence with caution.

Keywords

Identification; Voice recognition; Voice parade; Earwitness; Music education; Retention interval.

Received 1 May 2025; accepted 10 July 2025

1. Introduction

A voice parade is a forensic method of identifying persons, with different variants. In Polish law, the voice parade is regulated by Article 173 of the Code of Criminal Procedure (CCP), among other regulations, which *expressis verbis* indicates the identification of a person, likeness or object. This forms the legal basis for conducting a voice parade [1–4]. The possibility to identify a person based on their voice is supported by the individuality and uniqueness of the human voice [5], resulting from differences in the anatomical structures involved in speech production and speech patterns [6]. Speech recognition systems can accurately identify a speaker even with their attempt to mimic another person [5]. However, such identification is, in fact, performed based on speech rather than voice, because the term voice refers to the individual's timbre, tone, and pitch, whereas *speech* also refers to the person's overall manner of speaking, including any articulatory difficulties [7].

The literature underlines that a voice parade is both rare [6, 8, 9] and problematic to conduct. Problems may arise from procedural issues, such as the lack of a uniform voice parade methodology (as is the case in Poland, where the executive legal acts provide for the performance of a speech parade, at most, as an auxiliary tool to a standard identity parade [10]). Nonetheless, attempts have been made to develop this methodology [e.g., 6, 8, 9, 11]. In the UK, there are guidelines providing a set of good practices for investigators conducting a voice parade [12], but the procedure is rarely used there as well [13]. Another important consideration is the factors with an overall effect on one's voice memory, which is relatively weak in most persons [14]. These factors include: the circumstances of hearing a voice for the first time, duration of speech, variability of the voice parameters, a witness's preparation or voice memory, and the time between an event and the voice parade [8], as well as whether a witness was asked beforehand to memorise a voice in order to recall it later [15]. Other studies have also observed difficulties in remembering a voice, specifically noting that familiarising oneself with a voice enough to identify the speaker later takes more time than recognising the faces of new people [16, 17]. However, some researchers have indicated that even in the case of eyewitnesses, the auditory stimulus may dominate over the visual stimulus when attempting to identify someone [11, 18]. On the other hand, the memory imprint of the voice was observed to degrade faster than that of another person's face [19]. Most of the research results obtained to date have also indicated a relatively low accuracy of the voice-based identification of persons that rarely exceeds the chance of randomly choosing the correct speaker (or correctly stating that the speaker is not present) [20–24]. These findings clearly demonstrate the low reliability of a voice parade as evidence, which is lower than that of a standard identity parade [25, 26].

The factors potentially affecting the accuracy of a voice parade that have been assessed to date are the witness's age, including children and teenagers [23, 27], and elderly people [28, 29], visual disability [30, 31], and the voice recorded using a mobile phone [24], among others. Results concerning the relationships between being certain of having identified someone correctly and the identification accuracy are ambiguous and are referred to as unreliable [22, 23] or as having a low diagnostic value [32]; but on the other hand, they are sometimes referred to as significant [33]. The results related to the witness's or the perpetrator's sex are also unclear. To date, studies have observed no difference in the accuracy depending on the perpetrator's sex [34], but other studies have also found that male voices were identified more accurately than female voices [35, 36]. Furthermore, it was found that a female voice was identified with better accuracy by female participants [31]. Participants of the same sex as the perpetrator identified the voice more accurately [35]. Research has also shown that women are better at recognising familiar voices than men [37].

The effect of musical education on the accuracy of voice identification has not been researched in great detail. Musical education has been shown to benefit verbal memory [38, 39], pitch recognition [40], including in spoken language [41], tone recognition in tonal languages [42], linguistic abilities [43–45] and speech perception in noisy surroundings [46].

These findings led to the hypothesis that individuals undergoing musical education and musicians can recognise other people based on their voice more accurately than individuals without musical experience. However, research did not find a statistically significant difference between musicians and non-musicians who were asked to identify the voices of speakers of a language they knew [47]. In both groups, the identification accuracy amounted to 70–75%. On the other hand, musicians recognised the speakers of a language they did not know more accurately than non-musicians.

Even so, the results of those studies cannot be considered reliable for the accuracy of a voice parade, because their participants familiarised themselves with the speakers' voices and identified them on the same day, which is nearly impossible to achieve in judicial practice. Witnesses asked to identify a voice hear the perpetrator for the first time in natural circumstances, typically under stress. Selecting the suspects and organising a voice parade with the accompanying persons is a long-term process that requires the presence of an expert in phonoscopy, profiling the perpetrator's and the suspects' voices, and preparing recordings [12, 48]. As a result, performing a voice parade takes place a long time after an event and after the witness has heard the perpetrator's voice for the first time.

The literature underlines that a voice parade should be prepared and conducted no later than 4–6 weeks after a memory imprint of the perpetrator's voice is created [12, 48]. In order for a voice parade to be considered ecologically viable, this period should be no shorter than 24 hours [24]. The research has so far yielded mixed results with respect to the effect of the length of time between a witness familiarising themself with the perpetrator's voice and their participation in a voice parade on the identification accuracy. Some studies did not observe any difference in the identification accuracy even after eight weeks, compared to an experimental group that was asked to identify a voice after three weeks [22, 49]. Other studies noted a significant decrease in accuracy between a parade performed after two weeks and one performed after three weeks [28, 50], as well as a significant increase in false positives after just a week when a voice parade not including the perpetrator was performed [51].

Due to the above results and other factors, it became necessary to conduct a new study on the effects of musician education as a variable of the identification accuracy in a voice parade. The author of this study also decided to compare the results obtained by persons undergoing and not undergoing musical education at one-week and two-week intervals between the participants familiarising themselves with the perpetrator's voice and attempting to recognise it in a voice parade.

Taking into account the results of earlier investigations addressing the positive effect of musical education on a set of properties and weak voice memory, the following research hypotheses were formulated:
1. Persons undergoing musical education recognise other people based on their voices better than persons not undergoing musical education, a week after familiarising themselves with the perpetrator's voice.
2. Persons undergoing musical education recognise other people based on their voices better than persons not undergoing musical education, two weeks after familiarising themselves with the perpetrator's voice.
3. Persons attempting to identify the perpetrator's voice two weeks after familiarising themselves with it are less accurate than persons attempting to identify it after one week.

Furthermore, despite the ambiguous results, the author of this study decided to investigate the participants' sex and certainty of having provided the correct answer as additional factors.

2. Methods of selecting the voices and participants for a voice parade – methods of assessment and results

This study was constructed according to recommendations provided by forensic researchers, the Polish legal system and the voice parade practices followed in Poland [6, 18, 52, 53]. Thus, the following were included in the analysis: 20 voice samples, voice samples in the form of recordings (indirect parade), selection of voices through elimination assessments, recording duration of about 20 seconds, five fillers (persons other than the perpetrator) in a parade, voice samples containing quotes from the crime scene or the place where the memory imprint was created, texts read aloud and other aspects [27]. At the same time, in order to create conditions resembling those in which the memory imprint of the voice was created as closely as possible, the participants in this study familiarised themselves with the perpetrator's voice by listening to a recording that imitated the course of the event.

2.1. Voice parade selection

2.1.1. General information and selection of voice samples

Voice samples from 20 women aged 22–26 years were collected in the form of recordings with a length of 19–21 seconds. The participants were selected based on the following criteria:
- Age,
- Sex,
- Native speaker of Polish,
- No noticeable speech impediments, such as problems pronouncing the letter /r/ (rhotacism), lisping or stuttering,
- No interference from a regional dialect, such as the alveolar-dental /ł/ (in Poland called the Eastern Borderland /ł/),
- Similar socioeconomic environment and education, i.e., a student of Law or Psychology at the Jagiellonian University in Krakow and living in Krakow.

All samples were recorded using the same device (a built-in recorder in a Google Pixel 6). The recordings were adjusted technically using Audacity 3.5.1 (trimming, cutting out excessive pauses or adding pauses, noise reduction, volume equalisation and normalisation), which did not alter the sound of the voices in any way. Lastly, the recordings were exported as .wav files with a 3072 kb/s bitrate.

A script consisting of the 'perpetrator's' utterances from the crime scene was prepared (Appendix 1).

The twenty participants were asked to read some of the perpetrator's utterances. The participants were informed about the criminal nature of the event and were asked to read the utterances in an emotionally-loaded manner. Next, samples of their voices were recorded consisting of seven utterances from the crime scene (Appendix 2) and a fragment of a civil law textbook (Appendix 3).

After the elimination assessments, six voices were selected for the voice parade and one person was selected to play the role of the perpetrator. The latter was asked to read out the perpetrator's utterances from the crime scene, as realistically as possible.

2.2. Elimination assessments

Three elimination assessments were conducted, in order to select similar voices for the voice parade, with a total of 47 participants. The participants listened to the voice samples twice.

2.2.1. Elimination assessment I

During the first elimination assessment, five participants aged 21–23 years were instructed to listen to all 20 recordings and choose the voices that in their opinion were clearly different from the others. The participants could choose as many voices as they wanted. Six of the twenty voices, which were chosen at least three times, were eliminated (Fig. 1).

2.2.2. Elimination assessment II

In the second elimination assessment, 25 participants aged 15–17 years were instructed to choose the voices that were clearly different from the others with no limit, as with the first assessment. However, this time, the participants chose the voices from the remaining 14 recordings. The voices were ordered by giving them new numbers (Table 1).

Table 1
Voice numbers in Elimination Tests

Voice number in Elimination Test I	Voice number in Elimination Tests II and III	Voice number in Elimination Test I	Voice number in Elimination Tests II and III
I	I	XII	VIII
II	II	XIII	IX
IV	III	XIV	X
V	IV	XVI	XI
VIII	V	XVII	XII
X	VI	XIX	XIII
XI	VII	XX	XIV

As a result of the assessment, six more voices were eliminated, which left eight voices (Fig. 2).

2.2.3. Elimination assessment III

In the third elimination assessment, 17 participants aged 16–17 years, who were different from those in the previous assessment, listened to the voices numbered I, III, IV, VIII, IX, X, XI, and XIII twice. This time, the participants were instructed to choose only one voice out of the eight that was clearly different from the others and to evaluate the similarity of each voice on the following scale:

- +2 – very similar,
- +1 – similar,
- 0 – neither noticeably similar nor different,
- –1 – different,
- –2 – very different

by comparing all the voices to one another.

Due to being chosen the most and their similarity to other voices deviating from 0 (neither noticeably similar nor different), the voices numbered IX and IV (XIII and V in the first elimination assessment) were excluded from the parade. The person with the voice numbered XIII was selected to play the role of the perpetrator due to her voice being rarely chosen as clearly different from the others (Fig. 3) and with a low deviation from 0 (which meant that her voice was

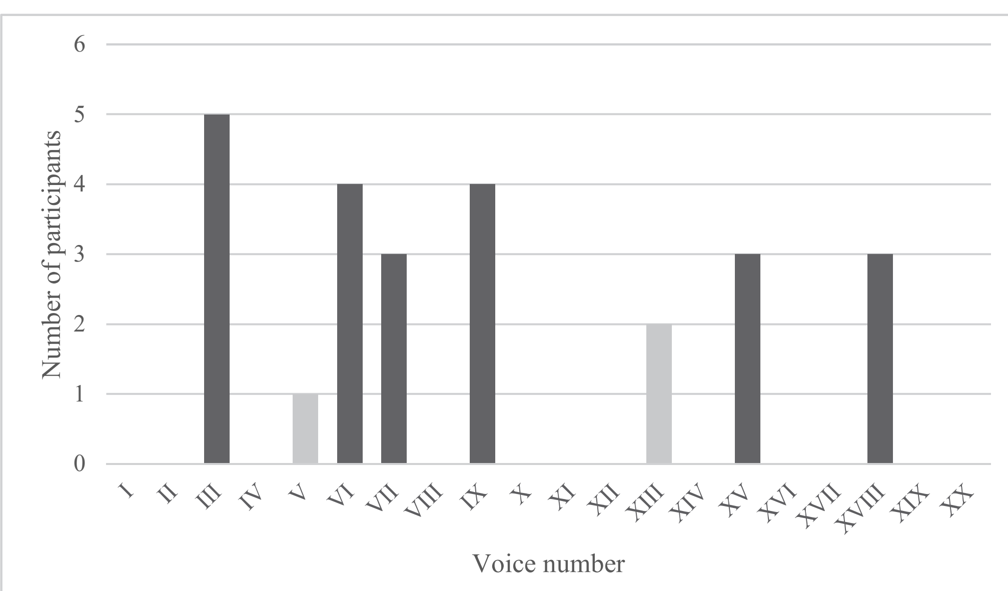


Fig 1. Elimination Test I. Significantly different voices

not markedly distinct; Fig. 4). Because the voice numbered III was chosen more often than XIII in the first part of the third elimination assessment (voice III – 4 times; voice XIII – 1 time), this person was not selected for the role of the perpetrator in order not to make identifying it in the voice parade overly easy. After the elimination assessments, the person with the voice numbered XIII was asked to record lines 'from the crime scene', which were then played back to the participants in the first assessment proper, i.e., when the participants familiarised themselves with the voice of the perpetrator.

2.3. Participants

A total of 129 persons aged 15–22 years participated in the study (mean age 17.35 years, σ = 1.087). All the participants declared not having any auditory disorders. The participants were selected according to the following criteria:

A. Undergoing musical education (n = 55; age 15–22 years; mean age = 16.93 years, σ = 1.359; mean length of the musical education = 9.91 years),
B. Not undergoing musical education (n = 74, age 16–19 years, mean age = 17.66 years, σ = 0.688).

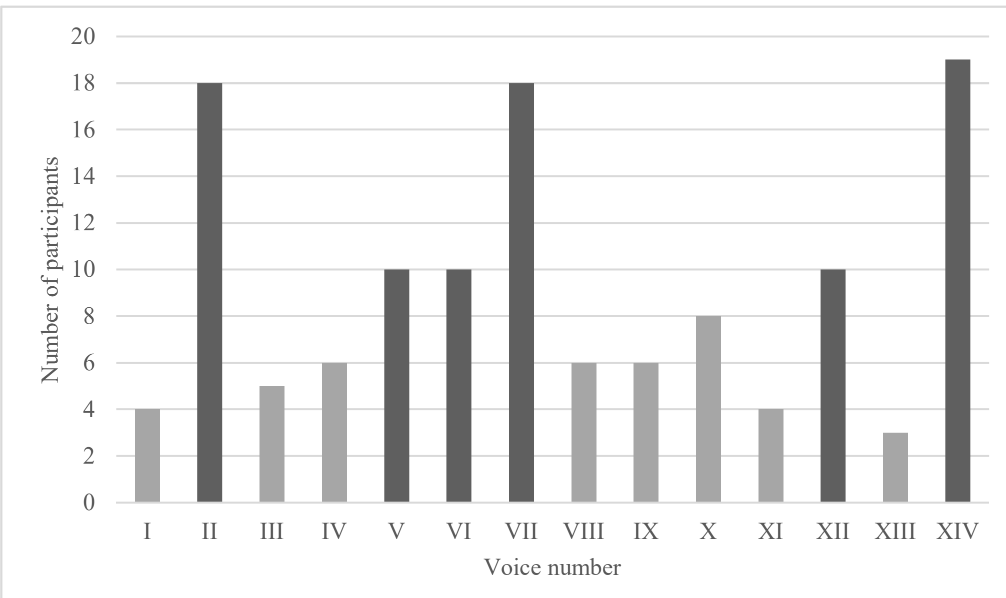


Fig. 2. Elimination Test II. Significantly different voices.

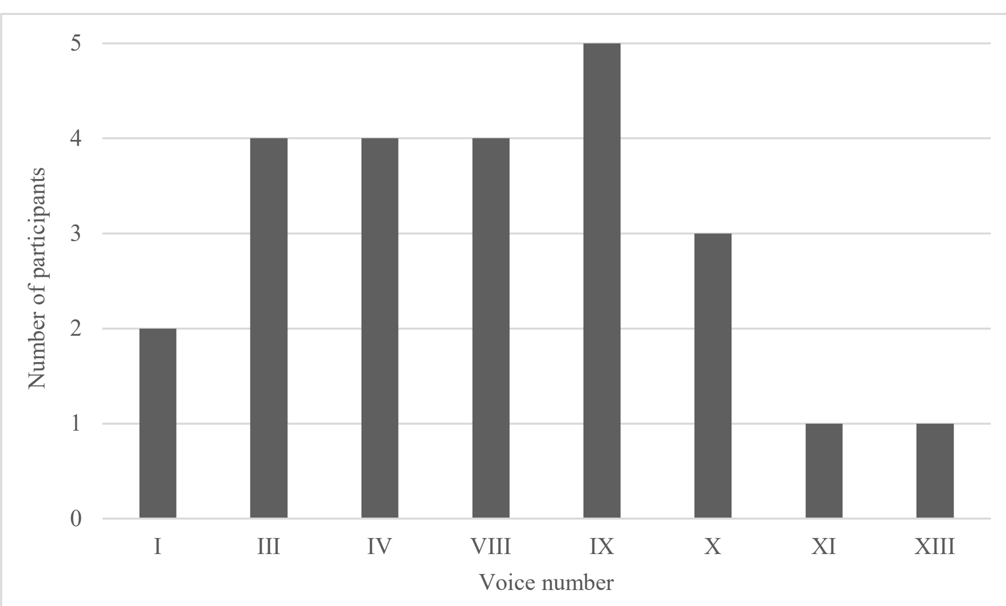


Fig. 3. Elimination Test III. Significantly different voice.

Group A included persons who, at the time of the study, had been undergoing formal musical education (by attending a state music school) for at least six years, played an instrument or sang, and who attended a music school. For comparison, other studies that have examined the effect of musical education classified the following individuals as musicians: amateur instrumentalists with at least nine years of continuous musical training in the Western classical music style, who had attended music classes in the preceding five years and played an instrument at the time of the study [40]; individuals who began musical training before the age of nine years and had practiced at least three times a week for more than an hour for more than 10 years [54]; at least 6 years of musical training on a Western music instrument before the age of 12 years [38]; and amateur instrumentalists or vocalists with at least six years of continuous musical training over their lifetime [47]. The criteria used in this study were a synthesis of the solutions listed above, and resulted from the age of the available study participants and their status as students of music schools who attended curricular classes at an educational institution during the week, including auditory training, instrument classes or participation in an orchestra or a choir, and who were not typically self-taught musicians.

Persons who did not meet the above criteria and had less than four years of formal musical training were assigned to Group B. There was no participant who had not been classified to either Group A or Group B. A similar criterion for the length of formal training was also applied in another study [54]. Other researchers classified individuals as not undergoing musical education based on the following criteria: no more than three years of formal musical training in any instrument, including no training over the preceding five years [40]; no musical training over their lifetime at all [38]; and less than a year of vocal or instrumental musical training over their lifetime [47].

The age of the participants was the result of conducting the study in music schools and secondary schools. To date, publications have not reported any statistically significant differences between the participants' age and the accuracy of voice identification [23]; therefore, this study focused on comparing the results obtained by persons undergoing and those not undergoing musical education. This study was conducted among participants of a similar age, in order to limit the potential effect of their age on the obtained results. The institutions from which the participants were recruited were selected depending on obtaining consent from their authorities and the time availability of the students themselves. The size of Groups A and B depended on the attendance on the day when the study was conducted.

2.4. Assessment proper

2.4.1. Assessment proper I

2.4.1.1. Participants

The first assessment proper was conducted among a total of 70 persons aged 16–21 years (n = 70, age

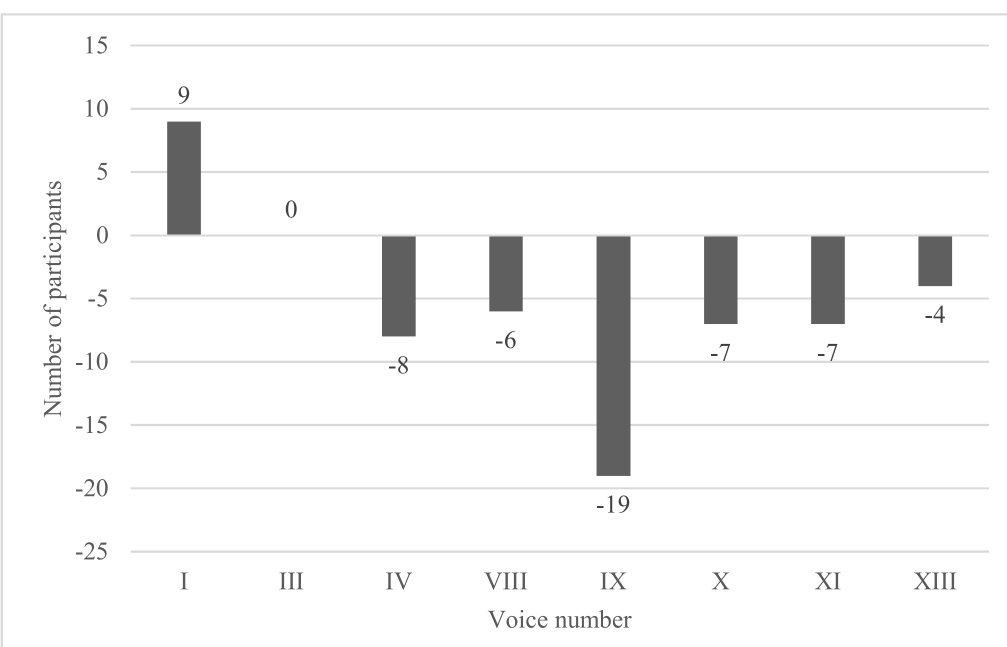


Fig. 4. Voices similarity level.

range 16–21 years, mean age 17.31 years, σ = 0.860), who were divided into Groups A1 (*n* = 29, age range 16–21 years, mean age 17.41 years, σ = 1.211, mean length of musical education: 10.45 years, σ = 1.66017) and B1 (*n* = 41, age range 16–19 years, mean age 17.24 years, σ = 0.489).

2.4.1.2. Procedure

Each participant was informed about the aim of the study and consented to their participation.

In the first stage of the analysis, the participants familiarised themselves with the voice of the perpetrator from the crime scene by listening twice to a recording of utterances from the crime scene, played back using a JBL Go speaker. Before familiarising themselves with the voice, the participants were asked to close their eyes and imagine being in a dark forest, then suddenly hearing a disturbing voice, at which point the recording was played twice. This procedure was used in order to enhance the participants' memorisation processes.

The second stage of the analysis took place a week later. This time, the participants were asked to fill in an assessment sheet, which consisted of two parts: *Before the parade* and *During/after the parade* (Appendix 4). Over the course of the analysis, the participants were informed several times about the possibility to opt out of the six-voice parade aimed at identifying the voice that they had heard a week earlier.

The following information about the participants was collected during the *Before the parade* part:

- Sex,
- Age,
- History of musical education, in particular:
  - Playing an instrument and/or singing,
  - Age at which the participants underwent their musical education, including the age at which they began (and ended, if applicable) learning to play an instrument and/or sing,
  - Form of musical education (e.g., a music school, private lessons, self-taught, or performances),
  - Frequency of practice (number of days per week and duration of practice in hours),
- The certainty of recognising the perpetrator's voice or recognising that the perpetrator's voice was not present in the parade was assessed through answering the question 'Do you think that you will correctly indicate the perpetrator's voice or lack of it in the parade?' on the following scale: 1 – definitely no; 2 – probably no; 3 – probably yes; 4 – definitely yes.

After filling in the first part of the assessment sheet, the participants listened twice to a parade consisting of the six voices selected during the elimination assessments (also played using the JBL Go speaker) and were asked to fill in the *During/after the parade* part of the assessment. The following information was collected in the second part:

- Answer to the question: 'Which of the voices in the parade belongs to the perpetrator?'. The possible answers were: Voice 1, Voice 2, Voice 3, Voice 4, Voice 5, Voice 6 and None (indicating the perpetrator's voice was not present in the parade),
- Reasons for choosing that particular answer,
- Reasons for excluding the other answers,
- Degree of certainty that the perpetrator's voice was identified correctly, or that it was not present in the parade, by answering the question 'Do you think that you have correctly indicated the perpetrator's voice or lack of it in the parade?' on the following scale: 1 – definitely no; 2 – probably no; 3 – probably yes; 4 – definitely yes.

2.4.1.3. Results

2.4.1.3.1. Certainty before listening to the voice parade

Before listening to the parade, the participants from Group A1 felt much more confident about choosing the correct answer. Among them, as many as 93% (27 persons) declared that they would correctly identify the perpetrator's voice or the lack thereof in the parade. The mean certainty in Group A1 was 3.1.

To compare, only 71% (29 persons) of the participants from Group B1 were certain that they would correctly identify the perpetrator's voice or the lack thereof, while as many as 12 persons believed that they would probably not be able to answer correctly. The mean certainty in Group B1 was 2.43 (Fig. 5).

2.4.1.3.2. Accuracy

The overall accuracy amounted to 47.14%. The representatives of Group A1 were more accurate than the representatives of Group B1 (55% vs. 41%; 16 vs. 17 persons), but the difference was not statistically significant. Specifically, a chi-square test of independence showed (chi^2 (I, 70) = 1.281, $df = 1$, $p = 0.258$, $p > 0.05$) that there was no statistically significant relationship between the accuracy of identifying a person based on their voice and the level of musical education, within the week-long interval between hearing the voice for the first time and listening to the parade. At the same time, a similar share of participants in both groups declared that the perpetrator's voice was not present in the parade (15% vs. 17%; randomness level 14%). The results are presented in Fig. 6.

The observed degree of accuracy, as tested using with the chi-square test of independence, disproved Hypothesis 1 formulated in Section I.

2.4.1.3.3. Certainty after listening to the parade

After listening to the parade and choosing an answer, the certainty of a correct answer decreased in Group A1 by over 10 percentage points to nearly 83% (mean certainty 2.97). The certainty among the participants who chose the correct voice increased slightly from 3.13 to 3.19.

The same percentage decreased by almost 5 percentage points in Group B1 (nearly 66% of Group B1 declared that they probably or definitely had chosen the perpetrator's voice, or the lack thereof, in the parade). However, the mean certainty increased to 2.78. Likewise, the percentage of participants in Group B1 who were definitely convinced about being correct increased from 7% to 12%. The certainty among the

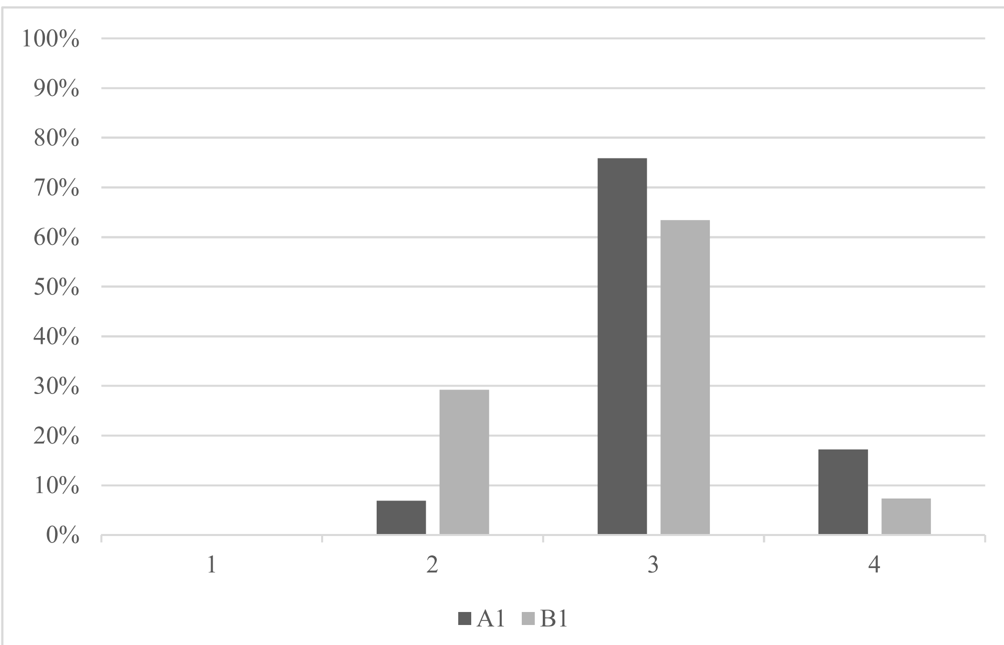


Fig. 5. Study I. Confidence rating before the voice lineup.

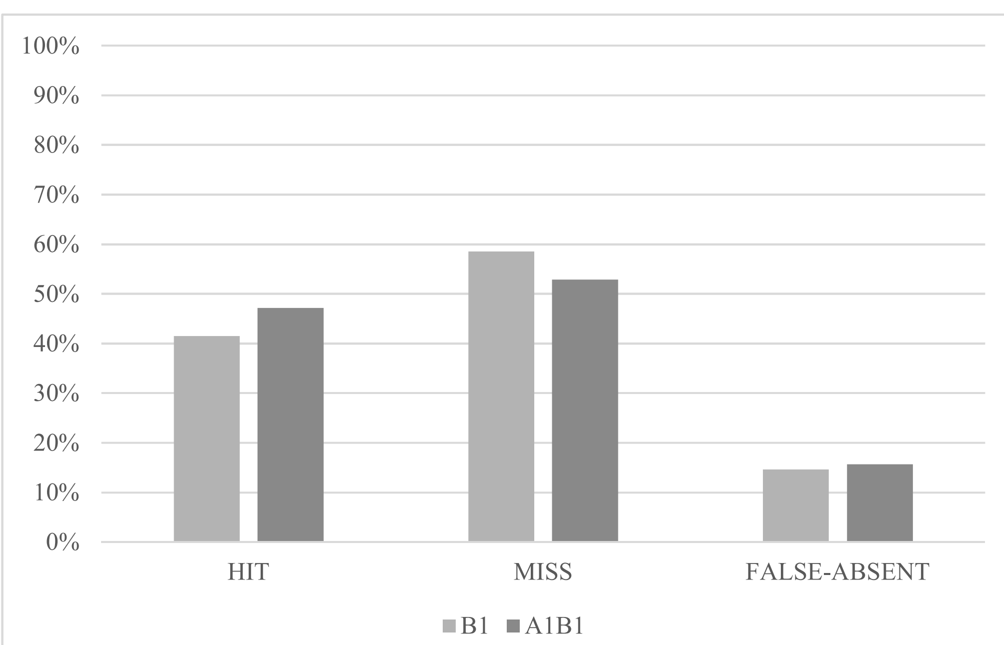


Fig. 6. Study I. Talker identification accuracy.

17 participants who correctly identified the perpetrator's voice decreased slightly, from 2.88 to 2.82 (Fig. 7 and 8).

2.4.1.3.4. Reasons for choosing the correct answer

The participants who correctly identified the perpetrator's voice in the parade gave different reasons for their answers. The participants were asked to provide all of their reasons in an open-ended manner, after which the declared reasons were divided into several categories.

Representatives of Group A1 usually based their answers on the colour of the voice. Relatively common answers were related to the tone and pitch of the voice, and to the closest similarity to the voice heard a week earlier. In Group B1, the main reasons were the timbre of voice and the closest similarity to the voice heard a week earlier (Table 2).

In Group A1, the most common reason for excluding the other voices in the parade was differences in timbre from the perpetrator's voice. Some of the participants also based their answers on differences in

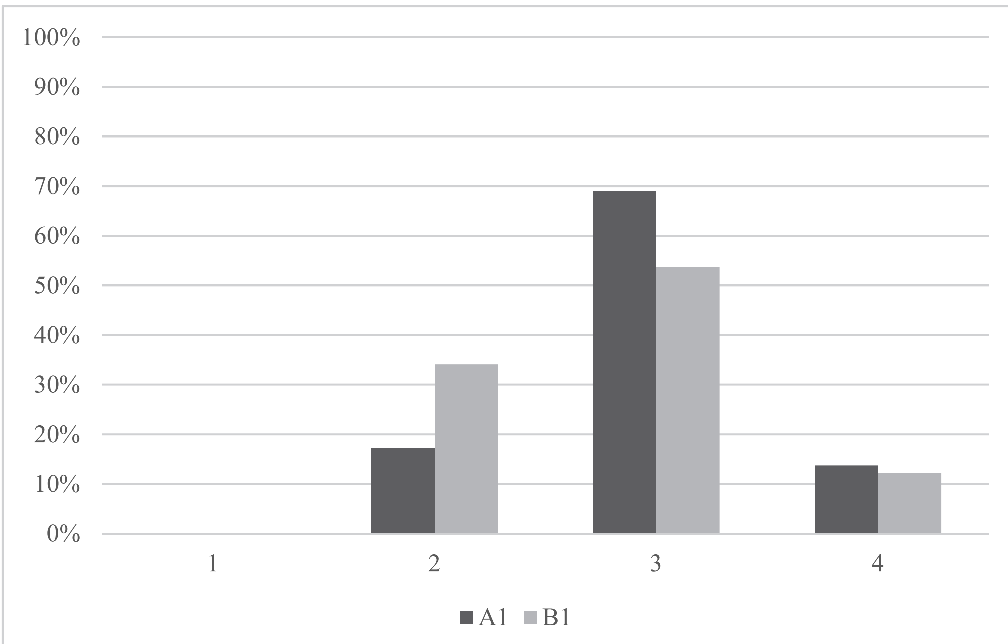
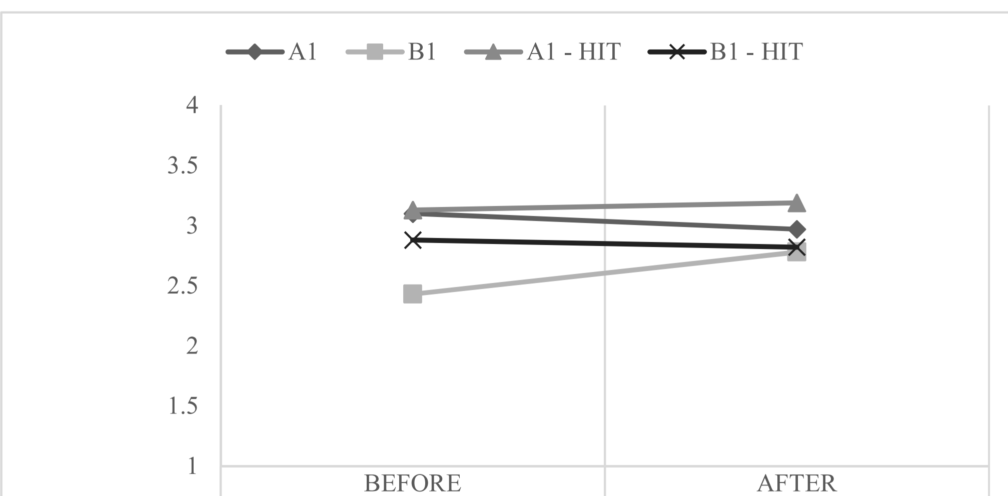


Fig. 7. Study I. Confidence rating after the voice lineup.



	BEFORE	AFTER
A1	3.1	2.97
B1	2.43	2.78
A1 - HIT	3.13	3.19
B1 - HIT	2.88	2.82

Fig. 8. Study I. Confidence rating.

Table 2
Study I. Participants who correctly identified the target – reasons of choosing the voice

Reasons of choice	Mentions	
	A1	B1
Voice timbre	8	6
Most like offender's voice	4	6
Voice pitch	6	1
Sounds pronunciation (including /r/, /s/, /c/)	2	3
"Hey!"	2	2
Accentuation	3	0
General manner of speaking	1	2
Intuition	1	2
Emotion	1	1
General manner of speaking	0	2
Speaking rate	0	2
Like a participant's friend's voice	1	0
Soft consonants at the end of words	0	1
Speaker's age	1	0
Subjective opinion	0	1

tone and pitch, as well as a lack of similarity to the perpetrator's voice. The most common reason for excluding the other voices in Group B1 was a lack of similarity to the perpetrator's voice, which likely indicates problems with justifying their answers. This group of participants justified their responses as differences in timbre much less often (Table 3).

Table 3
Study I. Participants of correctly identified the target – reasons of rejecting the other answers

Reasons of rejecting the other answers	Mentions	
	A1	B1
Voice timbre	10	4
Different voice	4	9
Voice pitch	4	4
Accentuation	1	3
General manner of speaking	0	3
Diction	2	0
Emotion	0	2
Lack of authenticity/imitating	0	2
Lack of speech impediment	0	2
Speaking rate	2	0
Hoarseness	1	0
Intuition	0	1
Not given	1	0
Similar but still different	1	0

2.4.2. Assessment proper II

2.4.2.1. Participants

The second assessment proper was conducted among a total 59 persons (n = 59, age range 15–22 years, mean age 17.39 years, σ = 1.313), who were also divided into two groups: A2 (n = 26, age range 15–22 years, mean age 16.38 years, σ = 1.329; mean length of musical education: 9.31 years, σ = 1.71509) and B2 (n = 33, age range 17–19 years, mean age 18.18 years, σ = 0.528).

2.4.2.2. Procedure

The second assessment proper followed the same procedure as the first one, with the only difference relating to the time period between the participants hearing the perpetrator's voice for the first time and listening to the voice parade, which this time was two weeks.

2.4.2.3. Results

2.4.2.3.1. Certainty before listening to the voice parade

As before, the participants from Group A2 felt more confident before listening to the parade than the participants from Group B2. Over 80% of the participants in Group A2 declared 'probably yes' and 'definitely yes' when asked about their certainty of correctly identifying the perpetrator's voice, or lack thereof, in the parade (mean certainty 2.85). The corresponding share in Group B2 was 58% (mean certainty 2.64). The results are presented in Fig. 9.

2.4.2.3.2. Accuracy

The overall accuracy amounted to 30.5%. The results were nearly identical between both groups (A2 – 30%, B2 – 31%). The chi-square test of independence also showed (chi^2 (I, 59) = 0.001, df = 1, p = 0.969, $p > 0.05$) that in the two-week interval between hearing the voice and listening to the parade, no statistically significant relationship occurred between the accuracy of the identification based on a person's voice and the participant's level of musical education. Both groups declared above the randomness level

(14%) that the parade did not include the perpetrator's voice (A2 – 21%, B2 – 23%). The results are presented in Fig. 10.

The obtained level of accuracy, as tested by the chi-square test, disproved Hypothesis 1 formulated in Section I.

2.4.2.3.3. Certainty after listening to the parade

Most of the participants in Group A2 declared 'probably not' or 'definitely not' when asked about correctly identifying the perpetrator's voice, or lack thereof, in the parade (54%; mean certainty 2.31). The certainty among the eight participants from Group A2 who correctly identified the perpetrator's voice decreased from 2.75 to 2.50.

The certainty of answering correctly among the participants from Group B2 increased to 70% (mean certainty 2.79). The certainty of those participants who correctly identified the perpetrator's voice remained the same (2.7). The obtained results are presented in Fig. 11 and 12.

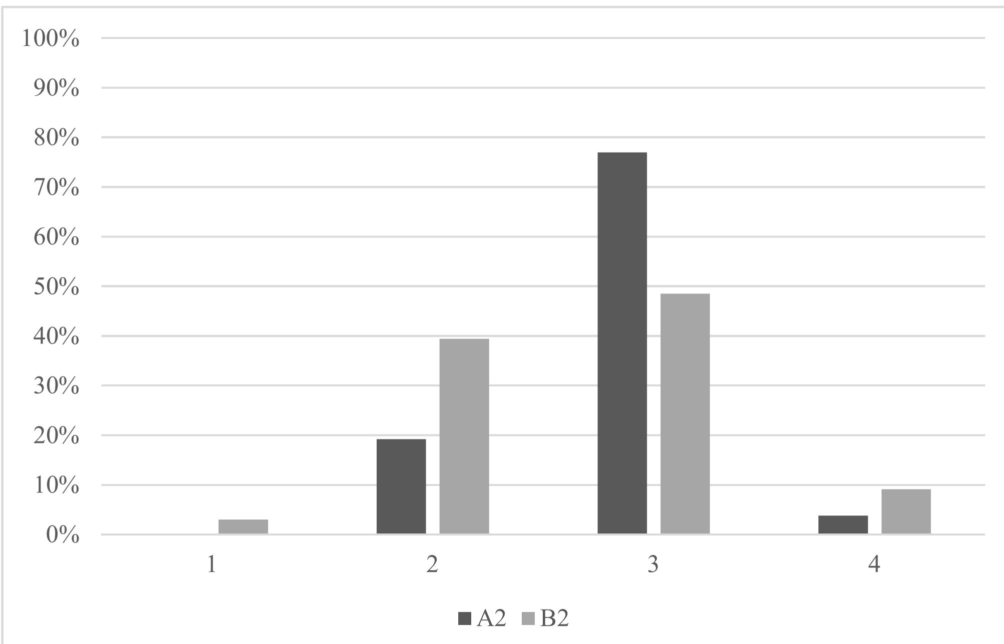


Fig. 9. Study II. Confidence rating before the voice lineup.

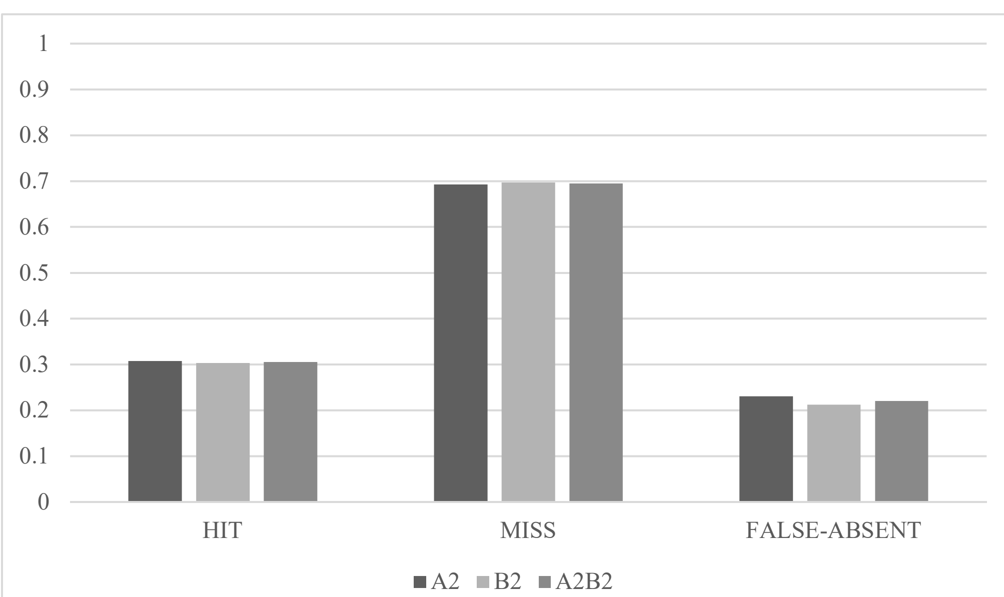


Fig. 10. Study II. Talker Identification Accuracy.

2.4.2.3.4. Reasons for choosing the correct answer

As with the first assessment proper, the participants in the second assessment were able to freely justify their decisions. The responses provided by those participants who correctly identified the perpetrator's voice in the parade were also divided into categories.

In Group A2, there was no single dominant factor affecting the choice of the perpetrator's voice in the parade. The participants usually based their answers on the tone and pitch of the voice, or on the timbre and closest resemblance to the voice they had heard two weeks earlier. Group B2 declared similar reasons. The obtained results are presented in Table 4.

The participants justified their decisions to exclude the other answers in a similar manner. Both groups usually indicated differences in the tone and pitch of the voice (in Group A2, they also mentioned its timbre). The obtained results are presented in Table 5.

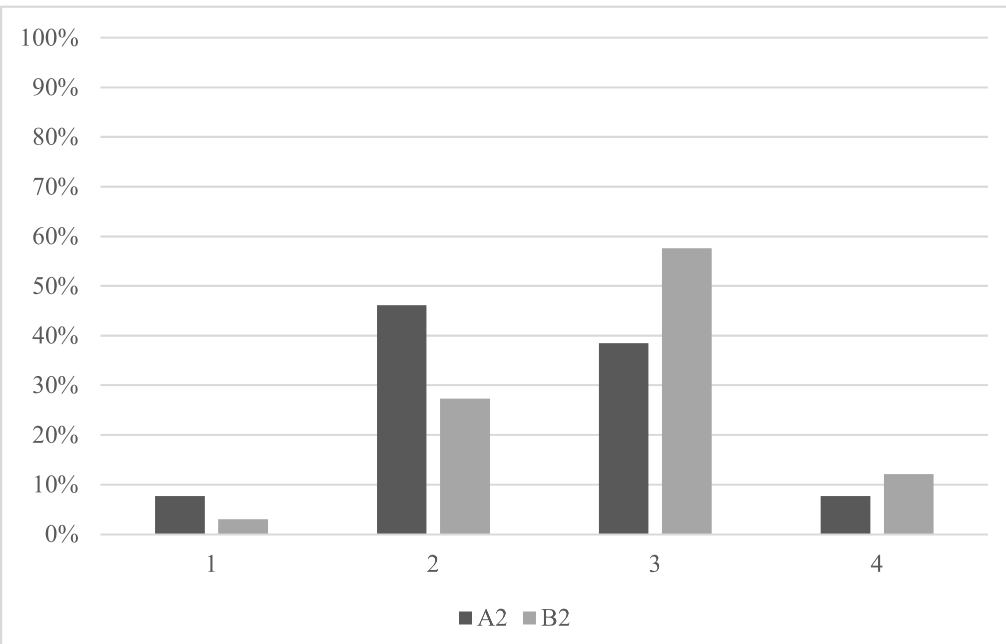
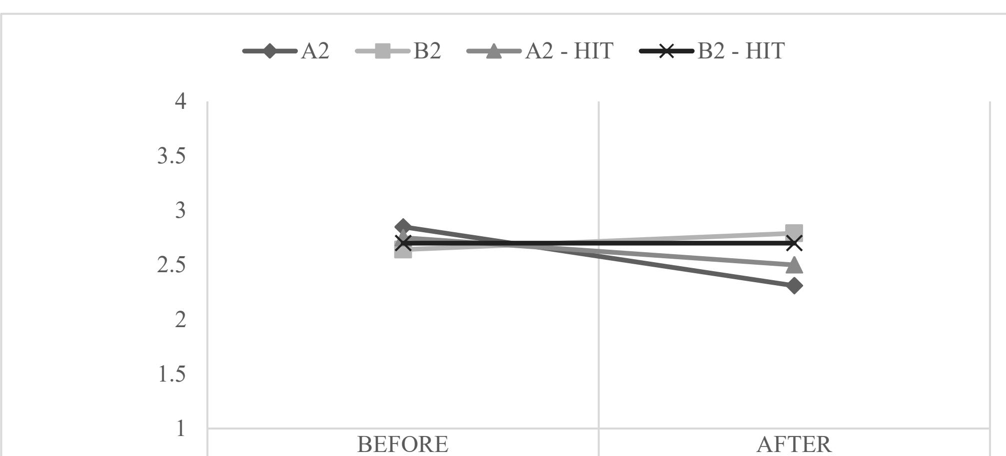


Fig. 11. Study II. Confidence rating after the voice lineup.



	BEFORE	AFTER
A2	2.85	2.31
B2	2.64	2.79
A2 - HIT	2.75	2.5
B2 - HIT	2.7	2.7

Fig. 12. Study II. Confidence rating.

Table 4
Study II. Participants who correctly identified the target – reasons of choice

Reasons of choice	Mentions	
	A2	B2
Voice pitch	4	2
Voice timbre	3	3
Most like offender's voice	3	2
"Hey!"	2	1
General manner of speaking	2	0
Diction	1	0
Speaking rate	1	0
Accentuation	0	1
Pronunciation of word 'understand'	0	1
Sounds pronunciation (including /r/, /s/, /c/)	0	1
Speaker's age	0	1
Subjective opinion	0	1

Table 5
Study II. Participants who correctly identified the target – reasons of rejecting the other answers

Reasons of rejecting the other answers	Mentions	
	A2	B2
Voice timbre	4	2
Voice pitch	4	5
Different voice	3	3
Speaking rate	2	2
Subjective opinion	3	0
General manner of speaking	1	1
Accentuation	0	2
Diction	2	0
Hoarseness	1	0
/R/ pronunciation	0	1

2.5. Comparison of the accuracy of identification based on a person's voice, depending on the period between hearing a voice for the first time and listening to a voice parade

As has already been mentioned, the overall accuracy amounted to 47.14% among the 70 participants in the first assessment proper (Groups A1 and B1) and 30.5% among the 59 participants performing the identification after two weeks (second assessment proper, Groups A2 and B2). The overall accuracy was 39.53%.

The chi-square test of independence showed (chi^2 (I, 129) = 3.706, $df = 1$, $p = 0.054$, $p > 0.05$) that there was no statistically significant relationship between the accuracy of the identification based on a person's voice and listening to the voice parade, although it must be stressed that this result was extremely close to statistical significance.

The obtained results disproved Hypothesis 3 formulated in Section I.

Subsequently, the passage of time was tested for its effect on the identification accuracy in the voice parade, separately among Groups A and B.

The participants in Groups A1 and A2 demonstrated a major difference in their accuracy (nearly 25 percentage points). Even so, the participants undergoing musical education did not show a statistically significant relationship between the accuracy of their identification based on a person's voice and the period between hearing a voice and listening to the parade (chi^2 (I, 55) = 3.319, $p = 0.068$, $p > 0.05$).

Between Groups B1 and B2, the difference in accuracy was slightly over 11 percentage points in favour of Group B1. Likewise, a chi-square test of independence (chi^2 (I, 74) = 0.983, $p = 0.322$, $p > 0.05$) found no statistically significant relationship between the accuracy of their identification based on a person's voice and the period between hearing a person's voice and listening to a voice parade in the participants not undergoing musical education.

2.6. Comparison of the accuracy of identification based on a person's voice depending on certainty

The participants' responses concerning their certainty were divided into two categories: lack of certainty ('definitely no' and 'probably no'); and certainty ('probably yes' and 'definitely yes').

With respect to the first assessment proper, the participants' certainty did not show a statistically significant relationship with the identification accuracy, either before (chi^2 (I, 70) = 0.917, $df = 1$, $p = 0.338$, $p > 0.05$) or after listening to the parade (chi^2 (I, 70) = 0.266, $df = 1$, $p = 0.606$, $p > 0.05$).

Similar results were obtained in the second assessment proper. The certainty before (chi^2 (I. 59) = 0.530, $df = 1$, $p = 0.466$, $p > 0.05$) and after (chi^2 (I, 59) = 0.933, $df = 1$, $p = 0.334$, $p > 0.05$) listening to the voice parade did not have a statistically significant relationship with the accuracy.

The obtained results are shown in Fig. 13.

2.7. Comparison of the identification accuracy based on a person's voice depending on sex

2.7.1. Assessment proper I

The first assessment proper was conducted among 41 women and 29 men. Group A1 consisted of 19 women and 10 men, while Group B1 consisted of 22 women and 19 men.

Men were found to be more accurate than women, both when the groups were considered separately (A1: 60% vs. 53%; B1: 53% vs. 32%) and overall (55% vs. 41%) (Fig. 14).

A chi-square test of independence showed (chi^2 (I, 70) = 1.281, $df = 1$, $p = 0.258$, $p > 0.05$) that there was no statistically significant relationship between the accuracy of the identification based on a person's voice and the participants' sex for the one-week interval between hearing the perpetrator's voice and listening to the voice parade.

A chi-square test of independence also found no statistically significant relationship, both among the participants not undergoing musical education (chi^2 (I, 41) = 1.820, $df = 1$, $p = 0.177$, $p > 0.05$) and those undergoing musical education (chi^2 (I, 29) = 0.144, $df = 1$, $p = 0.705$, $p > 0.05$; with Yates's correction

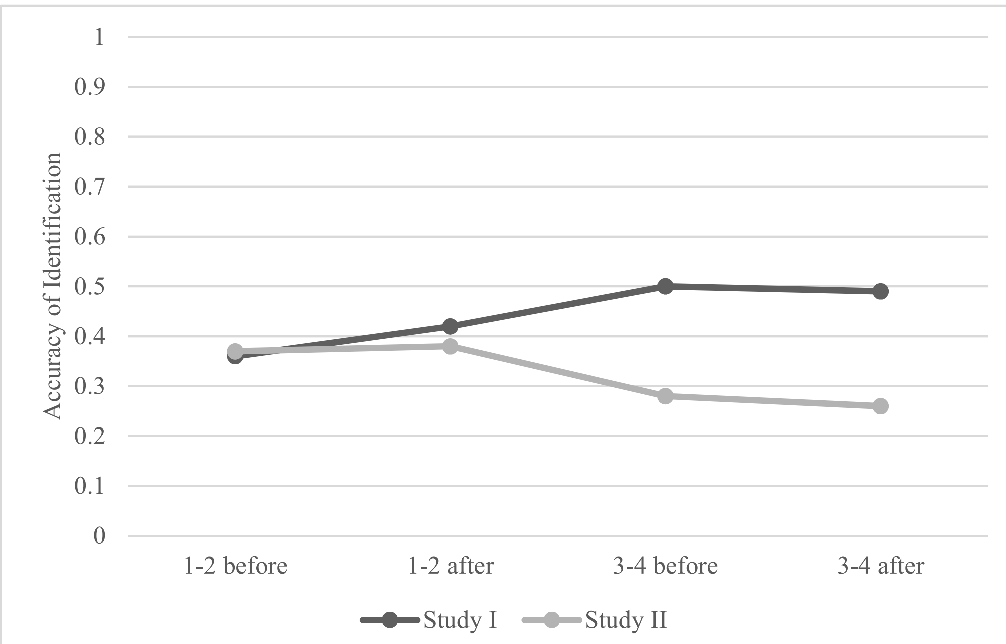


Fig. 13. Confidence rating: 1–2 – not confident; 3–4 – confident.

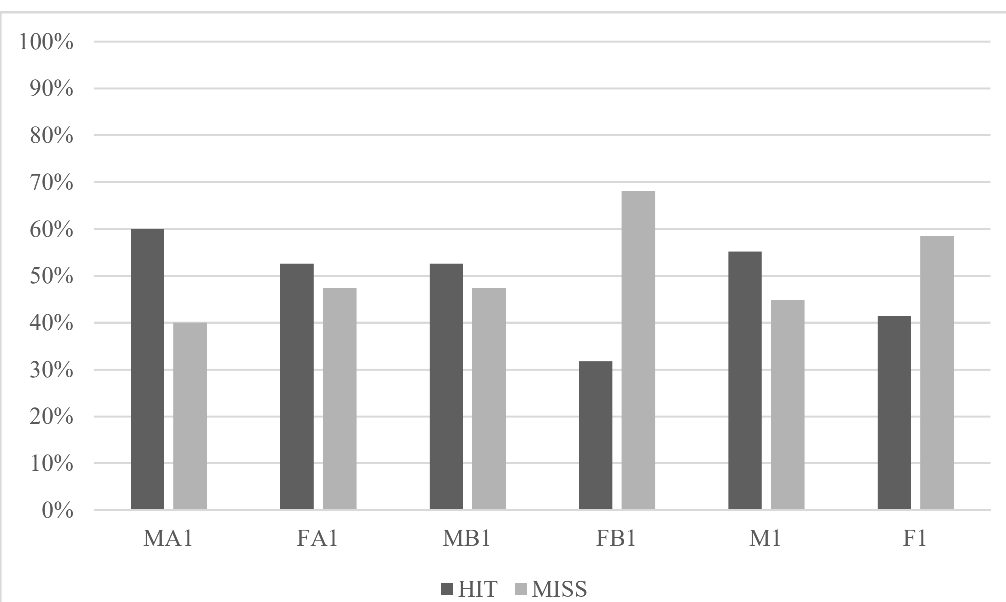


Fig. 14. Study I. Talker identification – gender criteria. M – male; F – female.

for continuity: chi^2 (I, 29) = 0.000, $df = 1$, $p = 1.000$, $p > 0.05$), between the accuracy of their identification based on a person's voice and the participants' sex for the one-week interval.

2.7.2. Assessment proper II

The second assessment proper was conducted among a total of 35 women and 24 men. Group A2 consisted of 19 women and 7 men, while Group B2 consisted of 16 women and 17 men.

In Group A2, the women achieved an accuracy of 42%, whereas all the men failed to correctly indicate the perpetrator's voice. In Group B2, the women achieved an accuracy of 25%, while the men achieved an accuracy of 35%. Overall, the women were more accurate than the men by 9 percentage points (34% vs. 25%). The results are presented in Fig. 15.

A chi-square test of independence showed (chi^2 (I, 59) = 0.579, $df = 1$, $p = 0.447$, $p > 0.05$) that there was no statistically significant relationship between the accuracy of the identification based on a person's voice and the participants' sex for the two-week interval between hearing the perpetrator's voice and listening to the voice parade, which confirmed Hypothesis H_0.

A chi-square test of independence also found no statistically significant relationship (chi^2 (I, 33) = 0.414, $df = 1$, $p = 0.520$, $p > 0.05$; with a necessary Yates's correction for continuity: chi^2 (I, 33) = 0.070, $df = 1$, $p = 0.792$, $p > 0.05$) between the accuracy of the identification based on a person's voice and the participants' sex for the two-week interval among the participants not undergoing musical education. For the participants undergoing musical education, the results appeared to be statistically significant (chi^2 (I, 26) = 4.257, $df = 1$, $p = 0.039$, $p < 0.05$), but because 50% of the cells were smaller than 5, Yates's correction for continuity was used. Ultimately, the effect of sex on the accuracy among the studied group was found to be statistically insignificant (chi^2 (I, 26) = 2.510, $df = 1$, $p = 0.113$, $p < 0.05$).

3. Discussion

The accuracy of the identification based on a person's voice was relatively low. Forty-seven percent of the group who listened to the voice parade after a week correctly identified the perpetrator; for the two-week group, this accuracy amounted to 30.5%. These results match those obtained in earlier studies, in which the identification accuracy also did not exceed 50% [22, 55]. In particular, the results obtained in the second assessment proper were similar to those obtained among children aged 11–13 years who identified a voice after two weeks [23].

It is also worth noting that the difference in accuracy depending on the retention period was not statistically significant. However, as mentioned above, the research results are ambiguous in this respect [22, 28, 49]. This study found a tendency towards statistical significance. Nonetheless, a larger sample size or a longer retention period may have yielded a different outcome.

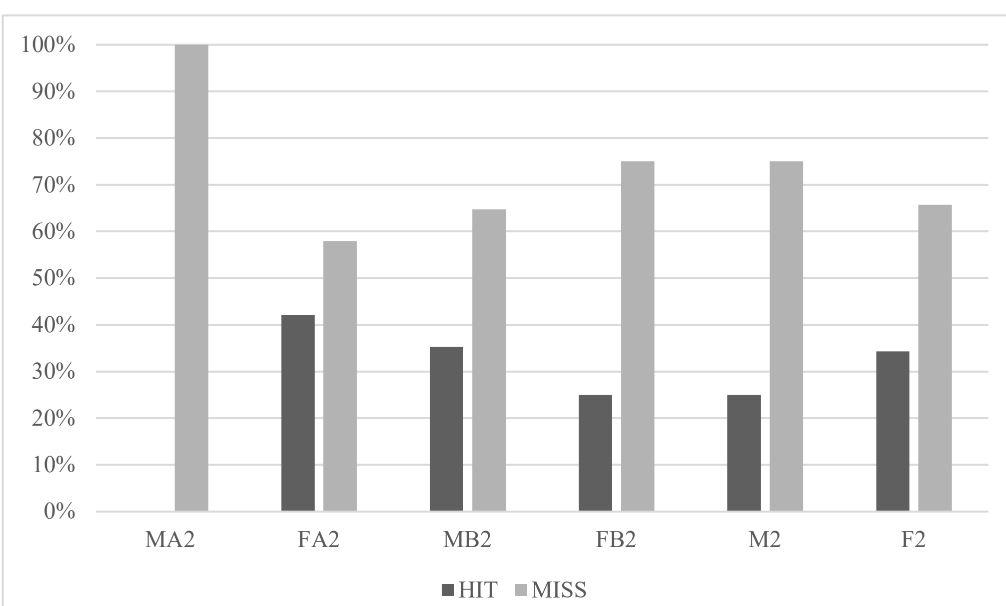


Fig. 15. Study II. Talker identification – gender criteria. M – male; F – female.

Importantly, no statistically significant difference in the identification accuracy was observed between the participants undergoing musical education (55% accuracy after a week and 30% after two weeks) and the participants not undergoing such education (41% accuracy after a week and 31% after two weeks), which corresponds to the results related to recognising speakers of the same language [47]. However, the accuracy achieved in this study was lower, which was likely due to the applied period of retention. At the same time, it should be noted that musical training did not affect the ability to identify individuals based on their voice.

Due to the above, Hypotheses 1, 2 and 3 were not confirmed.

No relationship was also observed between the participants' declared certainty and their identification accuracy based on a person's voice. Therefore, the certainty in this study was a factor with no diagnostic value [22, 23, 32]. Furthermore, no relationship was found between the participants' sex and their accuracy.

In light of the above, this study has confirmed that the results of a voice parade as evidence in criminal proceedings should be approached with scepticism [25, 26]. Due to the observed tendency towards statistical significance with respect to a negative effect of time on the identification accuracy, judicial bodies should conduct voice parades as soon as possible. However, it should be borne in mind that a voice parade requires good preparation, i.e. following the guidelines provided in the literature.

4. Conclusions

The results of this study did not prove that the participants who underwent musical education identified other people based on their voice better than the participants who did not undergo musical education. Furthermore, no statistically significant difference was observed between the participants who listened to the voice parade after a week and those who listened to it after two weeks, in both groups (participants undergoing and those not undergoing musical education). However, it should be noted that a clear tendency towards statistical significance was observed with respect to the effect of the elapsed time on the identification accuracy. Conversely, an effect of the participants' sex on accuracy in the voice parade was excluded. No relationship was also observed between the participants' certainty of choosing the correct answer and the identification accuracy. Nevertheless, the overall results confirmed the hypothesis about a weak voice memory.

There is room for further research in this area. Future studies may involve an extended retention period, attempts at recognising a voice by the same persons after different retention periods, an increased sample size or the participation of persons older than the participants in this study. Another potential avenue of research involves a more restrictive enrolment of the experimental groups, i.e., participation of professional musicians (with higher musical education) and persons with minimal or no musical education and with no ability to sing or play an instrument.

Acknowledgments

This article would not have been possible without the exceptional support and substantive supervision provided by Professor Józef Wójcikiewicz, PhD, Director of the Laboratory of Forensics at the Faculty of Law and Administration of the Jagiellonian University in Krakow, whose knowledge, advice and ideas were key in the preparation of this publication.

I would also like to thank Agata Trawińska, PhD, Director of the Laboratory of Speech and Recording Analysis at the Jan Sehn Institute of Forensic Research for her invaluable remarks and inspiration during the preparation and implementation of the voice parade.

I also wish to extend my thanks to everyone else who contributed to this study by providing voice samples, participating in the elimination assessments and assessments proper, and helping with organisational matters. In particular, I would like to thank the students and teachers at the following schools: VII M. Kopernik School of General Education in Częstochowa, IX Z. Wróblewski School of General Education in Krakow, F. Chopin State Music School of General Education, and W. Żeleński State Music School.

Appendix 1

"Well, well, well… Who's laughing now, huh? Are you proud of yourself, and of how everything turned out? I have one thing to tell you – you wasted three years of my life, you bastard! You get it? How dare you? You're nothing! Hey, don't move! I'll destroy you, I'll ruin you, you're done."

Appendix 2

1. "I'll destroy you", 2. "Who's laughing now?", 3. "Don't move", 4. "You're done", 5. "You wasted three years of my life", 6. "You're nothing", 7. "Hey!"

Appendix 3

"It is of significance that this standard is applied *in abstracto*, that is, irrespective of the individual characteristics and attributes of the perpetrator."

Appendix 4

GENDER: M / F
AGE: …

BEFORE THE PARADE

1. Do you currently play or have you ever played any instrument/sung? Please indicate the instrument(s) and/or singing.
2. At what age did you start learning to play an instrument or learning to sing? Are you still continuing this learning? If you have stopped, please indicate the age you were at that time.
3. What form did the music/singing education take (e.g., music school, private lessons, self-taught, performances)?
4. How often did/do you practice playing an instrument/singing (how many days per week, how many hours per day)?
5. Do you believe you can correctly identify the perpetrator's voice or the absence of the perpetrator's voice in the parade? Please mark your answer with an "x".

1 – definitely not	2 – rather not	3 – rather yes	4 – definitely yes

DURING AND AFTER THE PARADE

1. Which voice is the offender's voice? Please mark your answer with an "x".

Voice 1	Voice 2	Voice 3	Voice 4	Voice 5	Voice 6	Neither

2. Why did you make that decision?
3. Why did you reject the other options?
4. Do you believe that you correctly identified the perpetrator or their absence in the parade? Please mark your chosen answer with an "x".

1 – definitely not	2 – rather not	3 – rather yes	4 – definitely yes

References

1. Postanowienie Sądu Najwyższego z dnia 26 maja 2004 r., sygn. V KK 22/04, LEX nr 121992.
2. Wyrok Sądu Apelacyjnego w Lublinie z dnia 12 marca 2002 r., sygn. II AKa 39/02, LEX nr 76032.
3. Wyrok Sądu Apelacyjnego w Lublinie z 5 lipca 2017 r., sygn. II AKa 116/17, LEX nr 2355880.
4. Karczmarska D. Glosa do postanowienia SN z dnia 26 maja 2004 r., V KK 22/04, Wojskowy Przegląd Prawniczy. 2005;3:140-144.
5. Winiarska K. Efektywność holistycznej metody nauczania emisji głosu. Doctoral dissertation. Katowice: Uniwersytet Śląski; 2021.
6. Gruza E. Okazanie. Problematyka kryminalistyczna. Warszawa: Wydawnictwo Comer; 1995.
7. Kajstura M, Michałek M, Trawińska A. Ekspertyza fonoskopijna. In: Kała M, Wilk D, Wójcikiewicz J, editors. Ekspertyza sądowa. Zagadnienia wybrane. Wyd. 3. Warszawa: Wolters Kluwer; 2023. p. 774-834.
8. Wójcikiewicz J. Okazanie głosu i mowy. Problemy Współczesnej Kryminalistyki. 2002;5:205-209.
9. Tomaszewski T, Rzeszotarski J. Wybrane aspekty okazania mowy. Problemy Kryminalistyki. 2010;268:5-13.
10. Juszka K. Perspektywa doskonalenia systemu wykrywalności sprawców przestępstw. Prokuratura i Prawo. 2009;7-8:219-230.
11. Lisiecki MJ. Metodyka okazania. Studium procesowo-kryminalistyczne. Toruń: TNOiK Dom Organizatora; 2021.
12. Home Office Circular 057/2003. Advice on the use of voice identification parades. London: Crime Reduction and Community Safety Group, Police Leadership and Powers Unit. Available from: https://webarchive.nationalarchives.gov.uk/ukgwa/20130125153221 http://www.homeoffice.gov.uk/about-us/corporate-publications-strategy/home-office-circulars/circulars-2003/057- 2003/
13. Robson J. A fair hearing? The use of voice identification parades in criminal investigations in England and Wales. Crim Law Rev. 2017;36(1):36-50.
14. Clifford BR. Memory for voices: the feasibility and quality of earwitness evidence. In: Lloyd-Bostock SM, Clifford BR, editors Evaluating witness evidence. Recent psychological research and new perspectives. Chichester: John Wiley & Sons; 1983. p. 189-218.

15. Bull R, Clifford BR. Eyewitness testimony. Med Sci Law. 1999;39(2):120-127. https://doi.org/10.1177/002580249903900206
16. Hanley JR, Smith T, Hadfield J. I recognise you but I can't place you. An investigation of familiar-only experiences during tests of voice and face recognition. Q J Exp Psychol. 1998;51A(1):179-195. http://dx.doi.org/10.1080/713755751
17. Damjanovic L, Hanley JR. Recalling episodic and semantic information about famous faces and voice. Mem Cogn. 2007;35:1205-1210. https://doi.org/10.3758/bf03193594
18. Piątkowski K. Okazanie (studium kryminalistyczne). Szczytno: WSO; 1983.
19. Bigelow J, Poremba A. Achilles' ear? Inferior human short-term and recognition memory in the auditory modality. PLoS One. 2014 Feb;9(2):e89914. https://doi.org/10.1371/journal.pone.0089914
20. Yarmey AD, Yarmey AL, Yarmey MJ. Face and voice identification in showups and lineups. Appl Cogn Psychol. 1994 Oct;8(5):453-464. https://doi.org/10.1002/acp.2350080504
21. Olsson N, Juslin P, Winman A. Realism of confidence in earwitness versus eyewitness identification. J Exp Psychol Appl. 1998;4(2):101-18. https://doi.org/10.1037/1076-898X.4.2.101
22. Kerstholt JH, Jansen NJ, Van Amelsvoort AG, Broeders AP. Earwitnesses: effects of accent, retention and telephone. Appl Cogn Psychol. 2006;20(2):187-97. http://dx.doi.org/10.1002/acp.1175
23. Öhman L, Eriksson A, Granhag PA. Overhearing the planning of a crime: do adults outperform children as earwitnesses? J Police Crim Psychol. 2011;26(2):118-27. doi.org/10.1007/s11896-010-9076-5
24. Öhman L, Eriksson A, Granhag PA. Mobile phone quality vs. direct quality: how the presentation format affects earwitness identification accuracy. Eur J Psychol Appli to Leg Context. 2010 July;2(2):161-182.
25. Wójcikiewicz J. Glosa do wyroku Sądu Najwyższego z 9 lipca 2013 r., III K 92/13. Palestra. 2014;3-4:204-208.
26. Sherrin C. Earwitness evidence: the reliability of voice identifications. Osgoode Hall Law J. 2015;52(3):819-862. https://doi.org/10.60082/2817-5069.2957
27. Ferenc K. Głos zza drzwi, czyli okazanie głosu po polsku. In: Moskal D, Wąsik P, editors. Ślady pamięciowe. Aspekty psychologiczne i kryminalistyczne. Kraków: Biblioteka Katedry Kryminalistyki i Bezpieczeństwa Publicznego Uniwersytetu Jagiellońskiego; 2013. p. 59-69.
28. Bull R, Clifford BR. Earwitness voice recognition accuracy. In: Well GF, Loftus EF, editors. Eyewitness testimony: psychological perspectives. Cambridge, MA: Cambridge University Press; 1984. p. 92-123.
29. Biederman-Zaręba A. Przesłuchanie osób w podeszłym wieku. Unpublished Doctoral dissertation. Kraków: Jagiellonian University; 2013.
30. Bull R, Rathborn H, Clifford BR. The voice-recognition accuracy of blind listeners perception. 1983;12(2):223-226. https://doi.org/10.1068/p120223
31. Pang W, Xing H, Zhang L, Shu H, Zhang Y. Superiority of blind over sighted listeners in voice recognition. J Acoust Soc Am. 2020 Aug;148(2):EL208-EL213. https://doi.org/10.1121/10.0001804
32. Olsson N, Juslin P, Winman A. Realism of confidence in earwitness versus eyewitness identification. J Exp Psychol Appl. 1998;4(2):18-118. https://doi.org/10.1037/1076-898X.4.2.101
33. Orchard TL, Yarmey AD. The effects of whispers, voice sample duration, and voice distinctiveness on criminal speaker identification. Appl Cogn Psychol. 1995;9:249-260. https://doi.org/10.1002/acp.2350090306
34. Wilding J, Cook S. Sex differences and individual consistency in voice identification. Percept Mot Skills. 2000;91(2):535-538. https://doi:10.2466/pms.2000.91.2.535
35. Roebuck R, Wilding J. Effects of vowel variety and sample length on identification of a speaker in a line-up. Appl Cogn Psychol. 1993;7(6):475-481. https://doi.org/10.1002/acp.2350070603
36. Thompson CP. Voice identification: speaker identifiability and a correction of the record regarding sex effects. Hum Lear J Pract Res Appl. 1985;4:19-27.
37. Skus VG, Schweinberger SR. Gender differences in familiar voice identification. Hear Res. 2013 Feb;296:131-140. https://doi.org/10.1016/j.heares.2012.11.004
38. Chan AS, Ho YC, Cheung MC. Music training improves verbal memory. Nature 1986;396:128. https://doi.org/10.1038/24075
39. Ho YC, Cheung MC, Chan AS. Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. Neuropsychology. 2003;17(3):439-450. https://doi.org/10.1037/0894-4105.17.3.439
40. Bidelman GM, Gandour JT, Krishnan A. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J Cogn Neurosc. 2010;23(2):425-434. https://doi.org/10.1162/jocn.2009.21362
41. Deguchi C, Boureux M, Sarlo M, Besson M, Grassi M, Schön D, et al. Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: behavioral, electro physiological and psychoacoustic study. Brain Res. 2012;1455:75-89. https://doi.org/10.1016/j.brainres.2012.03.034
42. Burnham D, Brooker R, Reid A. The effects of absolute pitch ability and musical training on lexical tone perception. Psychol Music. 2014 Sep 18;43(6):881-97. https://doi.org/10.1177/0305735614546359

43. Neves L, Correia AI, Castro SL, Martins D, Lima CF. Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. Neurosc Biobehav Rev. 2022 Sep 1;140:104777. https://doi.org/10.1016/j.neubiorev. 2022.104777
44. Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M. Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cereb Cortex. 2008 Oct 1;19(3):712-23. https://doi.org/10.1093/cercor/bhn120
45. Wong PC, Perrachione TK. Learning pitch patterns in lexical identification by native English-speaking adults. Appl Psycholinguist. 28 Sept 2007;28(4):565-85. http://dx.doi.org/10.1017/S0142716407070312
46. Parbery-Clark A, Strait DL, Kraus N. Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. Neuropsychologia. 2011 Oct;49(12):3338-45. https://doi.org/10.1016/j.neuropsychologia.2011.08.007
47. Xie X, Myers E. The impact of musical training and tone language experience on talker identification. J Acoust Soc Am. 2015 Jan;137(1):419-32. https://doi.org/10.1121/1.4904699
48. Broeders AP, van Amelsvoort AG. A practical approach to forensic earwitness identification: constructing a voice line-up. Problems of Forensic Sciences 2001;47:237-245.
49. Broeders AP, Rietveld AM. Speaker identification by earwitnesses. Beiträge zur Phonetik und Linguistik 1995;64:24-40.
50. Clifford BR, Denot H. visual and verbal testimony and identification under conditions of stress. Unpublished manuscript. London: North East London Polytechnic; 1982.
51. Yarmey AD, Matthys E. Voice identification of an abductor. Appl Cogn Psychol. 1992;6:367-377. https://doi.org/10.1002/acp.2350060502
52. Wójcikiewicz J. Temida nad mikroskopem. Judykatura wobec dowodu naukowego 1993-2008. Toruń: TNOiK; 2009.
53. Herbowski P, Słapczyńska D. Okazanie mowy w praktyce śledczej. Palestra. 2016;7-8:49-53.
54. Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. Hear Res. 2010;261(1-2):22-29. https://doi.org/10.1016/j.heares.2009.12.021
55. Kerstholt JH, Jansen NJ, Van Amelsvoort AG, Broeders AP. Earwitnesses: effects of speech duration, retention interval and acoustic environment. Appl Cogn Psychol. 2004;18(3):327-336. http://dx.doi.org/10.1002/acp.974

ORCID
Piotr Woldan https://orcid.org/0009-0009-2931-3063

Corresponding author
Piotr Woldan
ul. Sobieskiego 16
PL 42-270 Zawada
e-mail: piotrjanwoldan@gmail.com

WPŁYW KSZTAŁCENIA SIĘ W KIERUNKU MUZYCZNYM NA TRAFNOŚĆ ROZPOZNANIA PODCZAS OKAZANIA MOWY

1. Wstęp

Okazanie to jedna z kryminalistyczno-procesowych metod identyfikacji człowieka możliwa do przeprowadzenia w różnych wariantach. W polskim systemie prawnym okazanie uregulowane jest m.in. w art. 173 Kodeksu postępowania karnego (k.p.k.), w którym *expressis verbis* wskazuje się okazanie osoby, wizerunku lub rzeczy. Na tej podstawie dopuszcza się w Polsce przeprowadzenie okazania mowy [1–4]. O samej możliwości rozpoznania osoby na podstawie jej głosu świadczą jego indywidualność i niepowtarzalność u każdego człowieka [5], wynikające z różnic w budowie struktur organów anatomicznych mowy, a także używanych wzorców mowy [6]. Nawet przy próbach naśladowania różnica w osobie mówcy zostaje dostrzeżona przez systemy rozpoznawania mówcy [5]. Rozpoznania dokonuje się jednak bardziej na podstawie mowy niż głosu, ponieważ pojęcie „głosu" dotyczy barwy, tonu i wysokości, natomiast „mowa" obejmuje ponadto całokształt sposobu mówienia, w tym wszelkie trudności artykulacyjne [7].

W literaturze podkreśla się, że czynność okazania mowy jest czynnością rzadką [6, 8, 9] i do tego trudną do przeprowadzenia. Trudności te mogą wynikać z kwestii proceduralnych, takich jak brak jednolitej metodyki okazania mowy (taka sytuacja występuje w Polsce, gdzie ustawodawca w aktach wykonawczych przewiduje przeprowadzenie okazania mowy co najwyżej jako element towarzyszący okazaniu sylwetki [10], podjęte zostały także próby opracowania jego metodyki [m.in. 6, 8, 9, 11]; w Wielkiej Brytanii istnieje natomiast okólnik stanowiący zbiór dobrych rad dla śledczych przeprowadzających okazanie mowy [12], jednak również i tam czynność ta jest rzadkością [13]). Nie bez znaczenia pozostają także czynniki wpływające ogólnie na pamięć głosu, która u większości osób jest stosunkowo słaba [14]. Do wspomnianych czynników można zaliczyć: warunki poznania głosu, czas trwania wypowiedzi, zmienność parametrów głosu, nastawienie przygotowawcze czy pamięć głosu świadka, upływ czasu między zdarzeniem a okazaniem [8], a także wcześniejsze uprzedzenie świadka o konieczności zapamiętania i późniejszego rozpoznania głosu [15]. Trudności zapamiętania głosu dowodzą także inne badania, zgodnie z którymi zapoznanie się z głosem w sposób wystarczający, by później przypisać go do właściciela, zajmuje więcej czasu niż rozpoznawanie twarzy nowych osób [16, 17]. Niektórzy autorzy zaznaczają jednak, że nawet w przypadku wzrokowej percepcji zdarzenia bodziec słuchowy może być bardziej dominujący niż wzrokowy przy próbie identyfikacji [11, 18]. Stwierdzono jednak także, że w kontekście śladu pamięciowego głosu jego degradacja następuje szybciej niż śladu pamięciowego twarzy innej osoby [19]. Dotychczasowe wyniki badań w większości również wskazują na stosunkowo niski poziom trafności rozpoznania osoby na podstawie głosu, rzadko przekraczający poziom szansy losowego wskazania poprawnego mówcy (lub jego braku) [20–24]. Powyższe wskazuje jednoznacznie na niski poziom wiarygodności dowodu z okazania mowy – niższy niż dowodu z okazania wyglądu [25, 26].

Wśród dotychczas badanych czynników mogących wpływać na trafność rozpoznania podczas okazania mowy należy wymienić m.in. wiek świadków, w tym dzieci i nastolatków [23, 27] oraz osób starszych [28, 29], niepełnosprawność wzrokową [30, 31], a także nagranie głosu za pomocą telefonu komórkowego [24]. Wyniki dotyczące związku między pewnością co do poprawności odpowiedzi a poziomem trafności są natomiast niejednoznaczne, określane są jako niemiarodajne [22, 23], o niskiej wartości diagnostycznej [32], a jednak czasem jako mające charakter istotnych [33]. Jeśli zaś chodzi o płeć osób badanych lub płeć sprawcy, to również nie można mówić o jednoznacznych wynikach. W dotychczasowych badaniach zarówno nie zaobserwowano różnicy w poziomie trafności w zależności od płci sprawcy [34], jak i dostrzeżono wyższą skuteczność rozpoznawania głosu męskiego niż kobiecego [35, 36]. Stwierdzono także wyższy poziom trafności rozpoznania głosu kobiecego wśród badanych kobiet [31]. Osoby uczestniczące w badaniach lepiej rozpoznawały głos sprawcy tej samej płci co oni sami [35]. Badania wykazały również, że kobiety lepiej rozpoznają znane im głosy [37].

Mniej uwagi poświęcono zbadaniu wpływu wykształcenia muzycznego na trafność rozpoznania osoby na podstawie jej głosu. Dotychczas wykazano, że kształcenie się w kierunku muzycznym wpływa pozytywnie między innymi na pamięć werbalną [38, 39], rozpoznawanie wysokości dźwięku [40] także w języku mówionym [41], rozpoznawanie tonów w językach tonalnych [42], zdolności lingwistyczne [43–45] oraz percepcję mowy w hałasie [46].

Z tych też względów prawdopodobne wydawało się założenie, że osoby kształcące się w kierunku muzycznym i muzycy będą mieli wyższą trafność rozpoznania osób na podstawie ich głosu niż osoby bez doświadczenia muzycznego. Badania nie wykazały jednak statystycznie istotnej różnicy między muzykami i niemuzykami rozpoznającymi na podstawie głosu mówców posługujących się znanym przez badane osoby językiem [47]. Trafność rozpoznania wyniosła w obu grupach ok. 70–75%.

Muzycy w sposób statystycznie istotny lepiej niż niemuzycy rozpoznawali jednak mówców nieznanego im języka.

W przypadku czynności okazania mowy wyniki powyższych badań nie mogą być jednak uznane za miarodajne w zakresie poziomu trafności rozpoznania. Badani zapoznawali się bowiem z brzmieniem głosu mówców i rozpoznawali ich w tym samym dniu. W przypadku okazania mowy przeprowadzanego przez organy procesowe jest to sytuacja praktycznie niemożliwa – świadkowie proszeni o rozpoznanie osoby na podstawie jej głosu zapoznają się bowiem z głosem sprawcy w warunkach naturalnych, często stresowych. Samo wytypowanie podejrzanych oraz stworzenie parady identyfikacyjnej z osobami przybranymi ma miejsce w większym przedziale czasowym, wymaga również m.in. udziału biegłego z zakresu fonoskopii, profilowania mowy sprawcy oraz podejrzanego, a także przygotowania nagrań [12, 48]. Skutkuje to zatem przeprowadzeniem samej czynności okazania mowy dłuższy czas po zdarzeniu i zapoznaniu się przez świadka z głosem sprawcy.

W literaturze podkreśla się, że czynność okazania mowy powinna zostać przygotowana i przeprowadzona w ciągu maksymalnie 4–6 tygodni od powstania śladu pamięciowego głosu sprawcy [12, 48]. By uznać badanie z zakresu okazania mowy za trafne ekologicznie, okres ten nie może być krótszy niż 24 godziny [24]. Dotychczasowe badania skutkowały różnymi rezultatami co do wpływu długości okresu między zapoznaniem się z głosem sprawcy a wysłuchaniem parady głosów na trafność rozpoznania. W niektórych badaniach nie zarejestrowano różnicy w trafności rozpoznania nawet po 8 tygodniach względem grupy badawczej rozpoznającej głos po 3 tygodniach [22, 49]. W innych zaobserwowano znaczący spadek poziomu trafności pomiędzy rozpoznaniem po 2 tygodniach i po 3 tygodniach [28, 50], a także znaczny wzrost fałszywie pozytywnych wskazań już po tygodniu przy wysłuchaniu parady niezawierającej głosu sprawcy [51].

Między innymi ze względu na powyższe okoliczności konieczne stało się przeprowadzenie nowych badań w zakresie sprawdzenia wpływu zmiennej kształcenia się w kierunku muzycznym na trafność rozpoznania podczas okazania mowy. Zdecydowano się także na porównanie rezultatów osiąganych przez osoby kształcące i niekształcące się w kierunku muzycznym w okresach 1 tygodnia oraz 2 tygodni między zapoznaniem się z głosem sprawcy a próbą rozpoznania jego głosu w paradzie identyfikacyjnej.

Biorąc pod uwagę dotychczasowe wyniki badań w zakresie pozytywnego wpływu kształcenia się w kierunku muzycznym na szereg właściwości oraz słabą pamięć głosu, skonstruowano następujące hipotezy badawcze:

1. Osoby kształcące się w kierunku muzycznym lepiej rozpoznają osoby na podstawie ich głosu niż osoby niekształcące się w kierunku muzycznym po tygodniu od zapoznania się z głosem sprawcy.
2. Osoby kształcące się w kierunku muzycznym lepiej rozpoznają osoby na podstawie ich głosu niż osoby niekształcące się w kierunku muzycznym po dwóch tygodniach od zapoznania się z głosem sprawcy.
3. Osoby rozpoznające głos sprawcy po dwóch tygodniach od zapoznania się z nim mają niższy poziom trafności rozpoznania niż osoby rozpoznające głos sprawcy po jednym tygodniu od zapoznania się z nim.

Ponadto mimo niejednoznacznych wyników badań zdecydowano się na zbadanie czynników płci i pewności uczestników co do poprawnego zaznaczenia odpowiedzi.

2. Metody doboru głosów do parady, uczestników, przeprowadzenia badań oraz ich wyniki

Przy konstruowaniu badania uwzględniono zalecenia autorów literatury kryminalistycznej, polskie realia prawne oraz praktykę przeprowadzania okazania mowy przez organy w Polsce [6, 18, 52, 53]. W ten sposób uwzględniono m.in.: 20 próbek głosu, próbki głosu w formie nagrań (okazanie pośrednie), dobór głosów w drodze badań eliminacyjnych, długość nagrań wynoszącą około 20 sekund, 5 głosów przybranych w paradzie, próbki głosu zawierające kwestie zasłyszane z miejsca zdarzenia/powstania śladu pamięciowego oraz tekst czytany [27]. Jednocześnie w celu stworzenia warunków jak najbardziej zbliżonych do sytuacji powstania śladu pamięciowego głosu, uczestnicy badania zapoznawali się z głosem sprawcy poprzez nagranie imitujące przebieg zdarzenia.

2.1. Dobór parady

2.1.1. Informacje ogólne, pobranie próbek głosu

Na potrzeby badania pobrano w formie 19–21-sekundowych nagrań próbki głosu od 20 kobiet w wieku 22–26 lat, które dobrano według następujących kryteriów:

– wiek,
– płeć,
– język polski jako język ojczysty tych osób,
– brak wyraźnych wad wymowy, np. problemy z wymową „r” (reranie), seplenienie, jąkanie się,
– brak naleciałości regionalnych, np. przedniojęzykowo-zębowe „ł” (tzw. ł kresowe),
– podobne środowisko socjoekonomiczne i wykształcenie, tj. studentka Uniwersytetu Jagiellońskiego z kierunku prawo lub psychologia mieszkająca na co dzień w Krakowie.

Wszystkie osoby nagrano tym samym urządzeniem (wbudowany dyktafon Google Pixel 6), a następnie uzyskane w ten sposób nagrania zostały technicznie przystosowane przy użyciu programu Audacity 3.5.1. (dokonano przycięć, wycięć nadmiernych przerw lub dodano je, zredukowano szum, wyrównano głośność, znormalizowano), który nie wpłynął w żaden sposób na zmianę brzmienia głosów. Na końcu nagrania wyeksportowano do plików .wav, szybkość transmisji bitów wynosiła 3072 kb/s.

W ramach badania przygotowano treść wypowiedzi „sprawcy" z miejsca zdarzenia (Załącznik 1).

20 osób poproszono o przeczytanie niektórych kwestii z wypowiedzi sprawcy. Zostały one poinformowane o przestępnym charakterze zdarzenia i poproszone o wypowiedzenie tych kwestii w sposób nacechowany emocjonalnie. Następnie pobrano od nich próbki głosu składające się z siedmiu kwestii zasłyszanych z miejsca zdarzenia (Załącznik 2) oraz fragmentu tekstu neutralnego z podręcznika do prawa cywilnego (Załącznik 3).

Po przeprowadzeniu badań eliminacyjnych i wyłonieniu 6 głosów do parady głosów, wytypowano osobę, która miała odegrać rolę sprawcy. Została ona poproszona o jak najbardziej realistyczne odczytanie treści wypowiedzi sprawcy z miejsca zdarzenia.

2.2. Badania eliminacyjne

W celu odpowiedniego doboru do parady głosów podobnych do siebie przeprowadzono trzy badania eliminacyjne, w których udział wzięło łącznie 47 osób. Osoby te zapoznawały się z nagraniami głosów dwukrotnie.

2.2.1. Badanie eliminacyjne I

W trakcie pierwszego badania eliminacyjnego pięcioro uczestników w wieku od 21 do 23 lat miało za zadanie wskazać wśród wszystkich dwudziestu nagranych głosów te, które ich zdaniem wyraźnie różniły się od innych. Uczestnicy tego badania nie mieli żadnych ograniczeń co do liczby możliwych do wskazania głosów. Ze zbioru dwudziestu głosów wyeliminowano sześć, które uzyskały co najmniej trzy wskazania (Ryc. 1).

2.2.2. Badanie eliminacyjne II

W kolejnym badaniu eliminacyjnym 25 osób w wieku od 15 do 17 lat, podobnie jak poprzednio, miało za zadanie wskazać głosy wyraźnie różniące się, również bez limitu wskazań, jednak tym razem spośród czternastu pozostałych nagrań. Celem uporządkowania głosy zostały na nowo ponumerowane (Tabela 1).

W ten sposób wyeliminowano kolejne sześć głosów. Ich liczba została ograniczona do ośmiu (Ryc. 2).

2.2.3. Badanie eliminacyjne III

W tym badaniu eliminacyjnym 17 innym niż w poprzednim badaniu osobom w wieku od 16 do 17 lat przedstawiono dwukrotnie głosy nr I, III, IV, VIII, IX, X, XI i XIII. Tym razem uczestnicy mieli za zadanie wskazać wyłącznie jeden wyraźnie różniący się od innych z grupy ośmiu głosów oraz ocenić stopień podobieństwa każdego z nich w skali:

- +2 – bardzo podobny,
- +1 – podobny,
- 0 – nie da się stwierdzić podobieństwa lub różnicy,
- –1 – różny,
- –2 – bardzo różny,

porównując je wszystkie do siebie.

Z uwagi na największą liczbę wskazań i stopień podobieństwa odchylony od punktu 0 (nie da się stwierdzić podobieństwa lub różnicy) do parady nie weszły głosy nr IX i IV (w badaniu eliminacyjnym I oznaczone jako XIII i V). Na sprawcę wytypowano osobę o głosie oznaczonym numerem XIII ze względu na małą liczbę wskazań tego głosu jako wyraźnie różniącego się od innych (Ryc. 3) oraz niski poziom odchylenia wyniku ze skali podobieństwa od punktu 0 (co oznacza, że głos nie jest specjalnie wyróżniający się; Ryc. 4). Głos nr III miał większą liczbę wskazań w pierwszej części badania eliminacyjnego III niż głos nr XIII (głos III – 4, głos XIII – 1), dlatego też nie został wybrany na głos sprawcy (możliwe zbyt duże ułatwienie dla uczestników właściwego badania). Po badaniach eliminacyjnych osoba o głosie nr XIII nagrała kwestie „pochodzące z miejsca zdarzenia", które były prezentowane uczestnikom badań właściwych podczas pierwszego spotkania, tj. zapoznania z głosem sprawcy.

2.3. Uczestnicy badania

W badaniu wzięło łącznie udział 129 osób w wieku od 15 do 22 lat (średnia wieku = 17,35, σ = 1,087). Każdy z uczestników oświadczył, że nie boryka się z problemami natury słuchowej. Uczestnicy badania zostali dobrani według następujących kryteriów:

A. kształcenia się w kierunku muzycznym (n = 55, wiek 15–22, średnia wieku = 16,93, σ = 1,359, średni czas kształcenia się w kierunku muzycznym w latach: 9,91),
B. braku kształcenia się w kierunku muzycznym (n = 74, wiek 16–19 lat, średnia wieku = 17,66, σ = 0,688).

Do grupy A zakwalifikowano osoby, które co najmniej od 6 lat w sposób formalny (uczęszczając do państwowych szkół muzycznych) kształciły się w kierunku muzycznym i w momencie przeprowadzenia badania grały na instrumencie lub ćwiczyły śpiew oraz uczęszczały do szkoły muzycznej. Dla porównania w innych badaniach, w których badanym czynnikiem był wpływ wykształcenia muzycznego, jako muzyków kwalifikowano: instrumentalistów amatorów mających za sobą co najmniej 9 lat ciągłego treningu muzycznego w stylu zachodniej muzyki klasycznej, uczęszczających na zajęcia muzyczne w ostatnich 5 latach oraz grających na instrumencie w momencie badania [40]; osoby, które rozpoczęły trening muzyczny przed ukończeniem 9. roku życia i trenujące co najmniej 3 razy w tygodniu dłużej niż godzinę przez więcej niż 10 lat [54]; co najmniej 6 lat treningu muzycznego na instrumencie muzyki zachodniej przed 12. rokiem życia [38]; instrumentalistów lub wokalistów amatorów mających za sobą co najmniej 6 lat ciągłego treningu muzycznego przez okres życia [47]. Przyjęte w niniejszym badaniu kryteria stanowią swoistą syntezę wyżej przytoczonych rozwiązań oraz wynikają z wieku dostępnych do badania osób, a także ich statusu uczniów szkół muzycznych, uczęszczających w ciągu tygodnia na programowe zajęcia w placówce oświatowej, obejmujące m.in. kształcenie słuchu, zajęcia z instrumentem czy udział w zajęciach orkiestry lub chóru, niebędących typowymi samoukami.

Osoby, które nie spełniały powyższych warunków i miały za sobą mniej niż 4 lata formalnego treningu muzycznego, zakwalifikowano do grupy B. Nie było żadnego uczestnika, który nie zostałby zakwalifikowany do żadnej z grup badanych. Podobne kryterium długości formalnego treningu użyto także w innym badaniu [54]. Inni autorzy za kryteria kwalifikacji osób jako niekształcących się w kierunku muzycznym uznali: nie więcej niż 3 lata formalnego treningu muzycznego na jakimkolwiek instrumencie, w tym brak treningu przez ostatnie 5 lat [40]; brak jakiegokolwiek treningu muzycznego w trakcie swojego życia [38]; mniej niż rok treningu muzycznego wokalnego lub instrumentalnego w okresie swojego życia [47].

Wiek uczestników wynikał z decyzji o przeprowadzeniu badania w szkołach muzycznych i szkołach średnich. W dotychchczasowych pracach nie dostrzeżono statystycznie istotnej różnicy między wiekiem osób badanych a trafnością rozpoznania na podstawie głosu [23], dlatego też w niniejszym badaniu skupiono się na porównaniu wyników uzyskanych przez osoby kształcące i niekształcące się w kierunku muzycznym. Celem ograniczenia ewentualnego wpływu wieku uczestników na wyniki badania wzięły w nim udział osoby w podobnym wieku. Wybór placówek, których uczniowie uczestniczyli w badaniu, zależał od zgody dyrekcji oraz dostępności czasowej samych uczniów. Liczebność grup A i B zależna była od frekwencji napotkanej w miejscu przeprowadzania badania w danym terminie.

2.4. Badania właściwe

2.4.1. Badanie właściwe I

2.4.1.1. Uczestnicy

W badaniu właściwym I wzięło udział łącznie 70 osób w wieku 16–21 lat (n = 70, przedział wiekowy 16–21 lat, średnia wieku 17,31, σ = 0,860), którzy zostali podzieleni na grupy A1 (n = 29, przedział wiekowy 16–21 lat, średnia wieku = 17,41, σ = 1,211; średni czas kształcenia się w kierunku muzycznym w latach: 10,45, σ = 1,66017) i B1 (n = 41, przedział wiekowy 16–19 lat, średnia wieku = 17,24; σ = 0,489).

2.4.1.2. Przebieg badania

Każdy z uczestników został poinformowany o celach badania, po czym wyraził świadomą zgodę na udział w tym badaniu.

Pierwszą część stanowiło zapoznanie uczestników badania z głosem sprawcy z miejsca zdarzenia poprzez dwukrotne odtworzenie nagrania zawierającego wypowiedzi z miejsca zdarzenia przy użyciu głośnika JBL Go. Przed zapoznaniem się z głosem uczestnicy badania zostali poproszeni o zamknięcie oczu, wyobrażenie sobie, że są w ciemnym lesie i nagle słyszą niepokojący głos. Po tym dwukrotnie odtworzono nagranie. Motywem tego działania była chęć wzmocnienia procesów zapamiętywania u uczestników.

Druga część badania odbyła się po upływie tygodnia. Tym razem uczestnicy badania zostali poproszeni o wypełnienie karty badań, która składała się z części „Przed paradą" i „Parada/po paradzie" (Załącznik 4). W trakcie badania kilkukrotnie informowani byli o możliwości niewystępowania w 6-głosowej paradzie głosu usłyszanego przed tygodniem.

W części przed paradą zebrano następujące informacje na temat osoby uczestniczącej w badaniu:
- płeć,
- wiek,
- informacje dotyczące historii kształcenia w kierunku muzycznym, tj.:
  - gra na instrumencie i/lub śpiew,
  - wiek, w którym miała/ma miejsce nauka, w tym wiek, w którym rozpoczęto (i ewentualnie zakończono) naukę gry na instrumencie i/lub śpiewu,
  - forma kształcenia muzycznego (np. szkoła muzyczna, zajęcia prywatne, samodzielna nauka, występy),
  - częstotliwość ćwiczeń (liczba dni w tygodniu i czas trwania ćwiczenia ujęty w godzinach),
- stopień pewności co do rozpoznania głosu sprawcy/trafne rozpoznanie braku głosu sprawcy w paradzie poprzez odpowiedź na pytanie „Czy uważa Pan/Pani, że poprawnie wskaże głos sprawcy lub brak głosu sprawcy w paradzie?" w skali: 1 – zdecydowanie nie, 2 – raczej nie, 3 – raczej tak, 4 – zdecydowanie tak.

Po wypełnieniu pierwszej części karty badań uczestnicy badania dwukrotnie wysłuchali parady złożonej z sześciu wyselekcjonowanych podczas badań eliminacyjnych głosów (również odtwarzanej za pomocą urządzenia JBL Go) oraz zostali poproszeni o wypełnienie części „Parada/po paradzie". W części tej zbierane były następujące informacje:

- odpowiedź na pytanie „Który głos z parady należy do sprawcy?”. Możliwe odpowiedzi to: Głos 1, Głos 2, Głos 3, Głos 4, Głos 5, Głos 6, Żaden (brak głosu sprawcy w paradzie),
- powody zaznaczenia wybranej przez siebie odpowiedzi,
- powody odrzucenia innych odpowiedzi,
- stopień pewności poprawnego wskazania głosu sprawcy/jego braku w paradzie poprzez odpowiedź na pytanie „Czy uważa Pan/Pani, że poprawnie zidentyfikował(a) Pan/Pani sprawcę lub jego brak w paradzie?” w skali: 1 – zdecydowanie nie, 2 – raczej nie, 3 – raczej tak, 4 – zdecydowanie tak.

2.4.1.3. Wyniki

2.4.1.3.1. Stopień pewności przed wysłuchaniem parady

Przed wysłuchaniem parady osoby z grupy A1 czuły się znacznie pewniej w odniesieniu do zaznaczenia prawidłowej odpowiedzi. Wśród tych badanych aż 93% (27 osób) stwierdziło, że poprawnie wskaże głos lub brak głosu sprawcy w paradzie. Średni poziom pewności w grupie A1 wyniósł 3,1.

Dla porównania, wśród reprezentantów grupy B1 pewność co do rozpoznania głosu sprawcy lub stwierdzenia jego braku w paradzie wyraziło „tylko” 71% uczestników badania (29 osób), a aż 12 osób zaznaczyło, że raczej nie będzie w stanie wskazać prawidłowej odpowiedzi. Średni poziom pewności w grupie B wyniósł 2,43 (Ryc. 5).

2.4.1.3.2. Poziom trafności

Ogólny poziom trafności wyniósł 47,14%. Reprezentanci grupy A1 istotnie uzyskali lepszą trafność niż przedstawiciele grupy B1 (55% vs. 41%; 16 vs. 17), jednak nie była to różnica statystycznie istotna. Analiza testem chi-kwadrat niezależności wykazała bowiem (chi^2 (I, 70) = 1,281, $df = 1$, $p = 0{,}258$, $p > 0{,}05$), że w tygodniowym przedziale między usłyszeniem głosu a wysłuchaniem parady nie ma istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a kształceniem się w kierunku muzycznym. Jednocześnie podobny odsetek badanych w obu grupach zaznaczył, że w paradzie nie ma głosu sprawcy (15% vs. 17%; poziom losowości – 14%). Wyniki przedstawiono na Ryc. 6.

Wskazany poziom trafności sprawdzony analizą testem chi-kwadrat niezależności obala postawioną w części pierwszej Hipotezę 1.

2.4.1.3.3. Stopień pewności po wysłuchaniu parady

Po wysłuchaniu parady i zaznaczeniu wybranej odpowiedzi stopień pewności w odniesieniu do prawidłowości wskazania odpowiedzi spadł w grupie A1 o ponad 10 punktów procentowych, wynosząc niecałe 83% (średni poziom pewności to 2,97). Wśród uczestników, którzy poprawnie wskazali głos sprawcy, stopień pewności nieznacznie wzrósł z 3,13 do 3,19.

Wśród reprezentantów grupy B1 odsetek ten zmniejszył się o niecałe 5 punktów procentowych (blisko 66% uczestników badania z tej grupy zaznaczyło, że raczej lub na pewno prawidłowo wskazało głos sprawcy lub jego brak w paradzie). Średni poziom pewności wzrósł jednak do 2,78. Również odsetek osób zdecydowanie pewnych trafności swojej odpowiedzi wzrósł wśród reprezentantów grupy B1 z 7% do 12%. Wśród 17 uczestników, którzy poprawnie wytypowali głos sprawcy, stopień pewności nieznacznie spadł z 2,88 do 2,82 (Ryc. 7 i 8).

2.4.1.3.4. Motywy stojące za poprawnym wytypowaniem odpowiedzi

Badani, którzy poprawnie wskazali głos sprawcy w paradzie, różnie argumentowali swój wybór. Uzasadnienia te miały charakter swobodny, badani mieli podać wszystkie powody, którymi kierowali się przy swoim wyborze. Ich odpowiedzi zostały następnie ujęte w kilka kategorii.

Reprezentanci z grupy A1 najczęściej kierowali się barwą głosu. Stosunkowo często pojawiały się odpowiedzi dotyczące tonu i wysokości głosu oraz największego podobieństwa do głosu słyszanego tydzień wcześniej. W grupie B1 głównymi czynnikami były barwa oraz największe podobieństwo do głosu słyszanego tydzień wcześniej (Tabela 2).

Wśród grupy A1 najczęstszym powodem odrzucenia innych głosów z parady były różnice w ich barwie względem głosu sprawcy. Część uczestników powołała się również na różnice w tonie/wysokości głosu oraz brak podobieństwa do głosu sprawcy. W grupie B1 najczęściej przytaczanym powodem odrzucenia innych głosów był ich brak podobieństwa do głosu sprawcy, co świadczy najprawdopodobniej o problemach z uzasadnieniem decyzji. Zdecydowanie rzadziej uczestnicy uzasadniali swoją decyzję różnicą w barwie głosu (Tabela 3).

2.4.2. Badanie właściwe II

2.4.2.1. Uczestnicy

W badaniu właściwym II wzięło udział łącznie 59 osób ($n = 59$, przedział wiekowy 15–22 lata, średnia wieku = 17,39, $\sigma = 1{,}313$), którzy również zostali podzieleni na dwie grupy: A2 ($n = 26$, przedział wiekowy 15–22 lata, średnia wieku = 16,38, $\sigma = 1.329$; średni czas kształcenia się w kierunku muzycznym w latach: 9,31, $\sigma = 1{,}71509$) oraz B2 ($n = 33$, przedział wiekowy 17–19, średnia wieku = 18,18 lat, $\sigma = 0{,}528$).

2.4.2.2. Przebieg badania

Przebieg badania właściwego II był tożsamy z badaniem właściwym I – jedyną różnicę stanowił okres dzielący uczestników od zaprezentowania głosu sprawcy i wysłuchania parady, który tym razem wynosił dwa tygodnie.

2.4.2.3. Wyniki

2.4.2.3.1. Stopień pewności przed wysłuchaniem parady

Badani z grupy A2 ponownie czuli się pewniej przed wysłuchaniem parady niż uczestnicy z grupy B2. W grupie A2 odpowiedzi „raczej tak" lub „zdecydowanie tak" w kontekście stopnia pewności odnośnie do prawidłowego wytypowania głosu sprawcy lub jego braku w paradzie zaznaczyło ponad 80% uczestników (średni poziom pewności: 2,85). W grupie B2 odsetek ten wyniósł 58% (średni poziom pewności: 2,64). Wyniki przedstawiono na Ryc. 9.

2.4.2.3.2. Poziom trafności

Ogólny poziom trafności wyniósł 30,5%. Uzyskane w obu grupach rezultaty były niemal identyczne (A2 – 30%, B2 – 31%). Również analiza testem chi-kwadrat niezależności wykazała (chi^2 (I, 59) = 0,001, df = 1, p = 0,969, $p > 0,05$), że w dwutygodniowym przedziale między usłyszeniem głosu a wysłuchaniem parady nie ma istotnego statystyczne związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a kształceniem się w kierunku muzycznym. W obu grupach powyżej poziomu losowości (14%) wskazywano, że w paradzie nie znajduje się głos należący do sprawcy (A2 – 21%, B2 – 23%). Wyniki przedstawiono na Ryc. 10.

Wskazany poziom trafności sprawdzony analizą testem chi-kwadrat niezależności obala postawioną w części pierwszej Hipotezę 1.

2.4.2.3.3. Stopień pewności po wysłuchaniu parady

Uczestnicy z grupy A2 w większości stwierdzili, że raczej lub z całą pewnością nie wskazali poprawnie głosu sprawcy lub jego braku w paradzie (54%; średni stopień pewności: 2,31). Wśród 8 uczestników z tej grupy, którzy poprawnie rozpoznali głos sprawcy, stopień pewności spadł z 2,75 do 2,5.

Wśród uczestników z grupy B2 nastąpił wzrost stopnia pewności co do udzielenia prawidłowej odpowiedzi – wyniósł on 70% (średni stopień pewności: 2,79). Stopień pewności uczestników, którzy poprawnie wytypowali głos sprawcy, pozostał na tym samym poziomie (2,7). Wyniki przedstawiono na Ryc. 11 i 12.

2.4.2.3.4. Motywy stojące za poprawnym wytypowaniem odpowiedzi

Tak jak w przypadku badania właściwego I, uczestnicy badania właściwego II mieli pełną swobodę w kwestii uzasadnienia motywów swoich decyzji. Odpowiedzi udzielone przez uczestników, którzy poprawnie wskazali głos sprawcy w paradzie, także zostały ujęte w kategorie.

W grupie A2 nie było jednego dominującego czynnika wpływającego na wybór głosu sprawcy w paradzie. Uczestnicy najczęściej kierowali się tonem/wysokością głosu, barwą i największym podobieństwem do głosu prezentowanego dwa tygodnie wcześniej. Podobne uzasadnienia zaobserwowano także w grupie B2. Wyniki przedstawia Tabela 4.

Podobna sytuacja miała miejsce przy uzasadnieniu decyzji stojącej za odrzuceniem innych odpowiedzi. W obu grupach najczęściej wskazywano różnicę w tonie/wysokości głosu (w grupie A2 także barwę). Wyniki przedstawia Tabela 5.

2.5. Porównanie trafności rozpoznania osoby na podstawie jej głosu w zależności od okresu dzielącego usłyszenie głosu a wysłuchanie parady

Jak już zostało wspomniane, ogólna trafność rozpoznania wśród 70 uczestników badania właściwego I (grupy A1 i B1) wyniosła 47,14%, zaś wśród 59 osób rozpoznających po dwóch tygodniach (badanie właściwe II, grupy A2 i B2) – 30,5%. Ogólna trafność wyniosła 39,53%.

Analiza testem chi-kwadrat niezależności wykazała (chi^2 (I, 129) = 3,706, df = 1, p = 0,054, $p > 0,05$), że nie ma istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a okresem między zapoznaniem się z głosem i wysłuchaniem parady, choć zdecydowanie należy podkreślić, że wynik ten jest niezwykle zbliżony do poziomu istotności statystycznej.

W ten sposób obalona została Hipoteza 3 wskazana w części pierwszej.

Porównano następnie, czy upływ czasu wpływa na trafność rozpoznania podczas okazania mowy osobno w grupach A i B.

W przypadku uczestników grup A1 i A2 zauważyć można dużą różnicę w poziomie trafności rozpoznania (blisko 25 punktów procentowych) – mimo tego wśród osób kształcących się w kierunku muzycznym nie zaobserwowano istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a okresem między usłyszeniem głosu i wysłuchaniem parady (chi^2 (I, 55) = 3,319, p = 0,068, $p > 0,05$).

Wśród uczestników grup B1 i B2 różnica w trafności rozpoznania wyniosła niewiele ponad 11 punktów procentowych na korzyść grupy B1. Również i w tym przypadku analiza testem chi-kwadrat niezależności wykazała (*chi*^{2} (I, 74) = 0,983, p = 0,322, p > 0,05), że wśród osób niekształcących się w kierunku muzycznym nie ma istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a okresem między usłyszeniem głosu a wysłuchaniem parady.

2.6. Porównanie trafności rozpoznania osoby na podstawie jej głosu w zależności od stopnia pewności

Odpowiedzi uczestników w zakresie stopnia pewności zostały podzielone na dwie kategorie – brak pewności („zdecydowanie nie", „raczej nie") oraz pewność („raczej tak", „zdecydowanie tak").

W zakresie badania właściwego I stopień pewności uczestników nie miał statystycznie istotnego związku z trafnością rozpoznania osoby, ani przed (*chi*^{2} (I, 70) = 0,917, *df* = 1, p = 0,338, p > 0,05), ani po wysłuchaniu parady (*chi*^{2} (I, 70) = 0,266, *df* = 1, p = 0,606, p > 0,05).

Podobna sytuacja miała miejsce w badaniu właściwym II. Stopień pewności przed (*chi*^{2} (I, 59) = 0,530, *df* = 1, p = 0,466, p > 0,05) i po wysłuchaniu parady głosów (*chi*^{2} (I, 59) = 0,933, *df* = 1, p = 0,334, p > 0,05) nie miał istotnego statystycznie związku z poziomem trafności.

Wyniki zostały przedstawione na Ryc. 13.

2.7. Porównanie trafności rozpoznania osoby na podstawie jej głosu w zależności od płci badanych

2.7.1. Badanie właściwe I

W badaniu właściwym I wzięło udział 41 kobiet i 29 mężczyzn. W grupie A1 znalazło się 19 kobiet i 10 mężczyzn, zaś w grupie B1 – 22 kobiety i 19 mężczyzn.

Zarówno przy podziale na grupy (A1: 60% vs. 53%; B1: 53% vs. 32%), jak i w ogólnym rozrachunku (55% vs. 41%) mężczyźni osiągali trafność na wyższym poziomie niż kobiety (Ryc. 14).

Analiza testem chi-kwadrat niezależności wykazała (*chi*^{2} (I, 70) = 1,281, *df* = 1, p = 0,258, p > 0,05), że w tygodniowym przedziale między usłyszeniem głosu a wysłuchaniem parady nie ma istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a płcią osób badanych.

Analiza testem chi-kwadrat niezależności wykazała również, że zarówno wśród osób niekształcących się w kierunku muzycznym (*chi*^{2} (I, 41) = 1,820, *df* = 1, p = 0,177, p > 0,05), jak i kształcących się w kierunku muzycznym (*chi*^{2} (I, 29) = 0,144, *df* = 1, p = 0,705, p > 0,05, przy zastosowaniu poprawki na ciągłość Yatesa: *chi*^{2} (I, 29) = 0,000, df = 1, p = 1,000, p > 0,05, w tygodniowym przedziale między usłyszeniem głosu a wysłuchaniem parady nie ma istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a płcią badanych.

2.7.2. Badanie właściwe II

W badaniu właściwym II wzięło udział łącznie 35 kobiet i 24 mężczyzn. W grupie A2 znalazło się 19 kobiet i 7 mężczyzn, a w grupie B2 16 kobiet i 17 mężczyzn.

W grupie A2 kobiety osiągnęły trafność rozpoznania na poziomie 42%; żaden z mężczyzn w grupie A2 nie wskazał prawidłowo głosu sprawcy. W grupie B2 mężczyźni osiągnęli trafność na poziomie 35%, a kobiety 25%. W ogólnym rozrachunku to kobiety uzyskały stopień trafności wyższy o 9 punktów procentowych niż mężczyźni (34% vs. 25%). Wyniki przedstawiono na Ryc. 15.

Analiza testem chi-kwadrat niezależności wykazała (*chi*^{2} (I, 59) = 0,579, *df* = 1, p = 0,447, p > 0,05), że w dwutygodniowym przedziale między usłyszeniem głosu a wysłuchaniem parady nie ma istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a płcią osób badanych, co potwierdza hipotezę H_0.

Analiza testem chi-kwadrat niezależności wykazała również, że wśród osób niekształcących się w kierunku muzycznym (*chi*^{2} (I, 33) = 0,414, *df* = 1, p = 0,520, p > 0,05; konieczność zastosowania poprawki na ciągłość Yatesa: *chi*^{2} (I, 33) = 0,070, *df* = 1, p = 0,792, p > 0,05) w dwutygodniowym przedziale między usłyszeniem głosu a wysłuchaniem parady nie ma istotnego statystycznie związku pomiędzy skutecznością rozpoznawania osoby na podstawie jej głosu a płcią badanych. Przy osobach kształcących się w kierunku muzycznym na pierwszy rzut oka istotność ta występowała (*chi*^{2} (I, 26) = 4,257, *df* = 1, p = 0,039, p < 0,05), lecz z uwagi na to, że 50% komórek ma liczebność oczekiwaną mniejszą niż 5, zastosowano poprawkę na ciągłość Yatesa. Ostatecznie więc w badanej grupie wpływ płci na trafność nie jest istotny statystycznie (*chi*^{2} (I, 26) = 2,510, *df* = 1, p = 0,113, p < 0,05).

3. Dyskusja

Poziom trafności rozpoznania osoby na podstawie jej głosu nie był wysoki. W grupie rozpoznającej po tygodniu 47% uczestników poprawnie wytypowało sprawcę. Wśród uczestników badania rozpoznających głos sprawcy po dwóch tygodniach było to 30,5%. Wyniki te korelują z wcześniej przeprowadzonymi badaniami, w których skuteczność rozpoznania również nie przekraczała 50% [22, 55]. Zwłaszcza wyniki uzyskane w badaniu właściwym II są zbieżne z wynikami uzyskanymi wśród dzieci w wieku 11–13 lat rozpoznających głos po dwóch tygodniach [23].

Warto również odnotować, że różnica w poziomie trafności w zależności od okresu retencji nie miała charakteru istotnego statystycznie. Jak już wcześniej wspomniano, wyniki badań w tym zakresie nie są jednak niejednoznaczne [22, 28, 49]. W niniejszym badaniu zaobserwowano jednak trend w kierunku istotności statystycznej. Być może przy większej liczbie uczestników lub wydłużonym okresie retencji wyniki byłyby inne.

Co istotne, w zakresie poziomu trafności rozpoznania nie zaobserwowano statystycznie istotnej różnicy między osobami kształcącymi się (55% trafności po tygodniu, 30% po dwóch tygodniach) a niekształcącymi się w kierunku muzycznym (41% trafności po tygodniu, 31% po dwóch tygodniach), co koresponduje z wynikami uzyskanymi w zakresie rozpoznawania mówców tego samego języka [47]. Poziom trafności uzyskany w niniejszym badaniu był jednak niższy, co zapewne wynika z przyjętego okresu retencji. Jednocześnie należy stwierdzić, że kształcenie się w kierunku muzycznym nie wpływa na zdolność rozpoznania osoby na podstawie głosu.

Powyższe oznacza, że Hipotezy 1, 2 i 3 nie zostały potwierdzone.

Nie stwierdzono również, by stopień pewności uczestników badania miał związek z trafnością rozpoznania osoby na podstawie głosu. W niniejszym badaniu był to więc czynnik niemający wartości diagnostycznej [22, 23, 32]. Nie zaobserwowano również związku pomiędzy płcią uczestników badań a trafnością.

Wyniki badań potwierdzają więc konieczność ostrożnego traktowania w procesie karnym dowodu z okazania mowy [25, 26]. Ze względu na stwierdzony trend w kierunku istotności statystycznej w kontekście negatywnego wpływu upływu czasu na trafność rozpoznania za uzasadnioną należy uznać konieczność niezwłocznego przeprowadzenia przez organy czynności okazania mowy. Powinny one jednak pamiętać o odpowiednim przygotowaniu parady głosów – zgodnie z zaleceniami autorów zawartymi w literaturze.

4. Wnioski

Wyniki badań nie wykazały, by osoby kształcące się w kierunku muzycznym rozpoznawały osoby na podstawie ich głosu lepiej niż te, które nie kształciły się w tym kierunku. Nie stwierdzono również statystycznie istotnej różnicy pomiędzy wynikami uzyskiwanymi przez osoby rozpoznające głos po tygodniu od zapoznania a wynikami osób, które rozpoznawały głos po dwóch tygodniach. Nie wykazano takiego wpływu w obu badanych grupach (osoby kształcące i niekształcące się w kierunku muzycznym). Należy jednak zaznaczyć, że w zakresie wpływu upływu czasu na trafność rozpoznania zaobserwowano wyraźny trend w kierunku istotności statystycznej. Wykluczono natomiast wpływ płci badanych na trafność rozpoznania podczas okazania mowy. Nie zaobserwowano także związku między pewnością uczestników w odniesieniu do zaznaczenia prawidłowej odpowiedzi a trafnością rozpoznania. Ogólnie jednak badania potwierdzają tezę o słabej pamięci głosu.

Istnieje pole do dalszych badań w poruszanym zakresie. Mogłyby one obejmować m.in. wydłużenie okresu retencji, próby rozpoznania głosu przez te same osoby w różnych okresach retencji, zwiększenie liczby uczestników badań lub udział osób starszych niż uczestnicy niniejszego badania. Rozważyć można także bardziej restrykcyjny dobór grup badawczych, tj. udział profesjonalnych muzyków (z wyższym wykształceniem muzycznym) oraz osób, które w ogóle lub w stopniu minimalnym kształciły się w kierunku muzycznym, nie posiadając umiejętności śpiewu lub gry na jakimkolwiek instrumencie.

Podziękowania

Niniejszy artykuł nie mógłby powstać bez wyjątkowego wsparcia i opieki merytorycznej prof. dra hab. Józefa Wójcikiewicza, Kierownika Pracowni Kryminalistyki Wydziału Prawa i Administracji Uniwersytetu Jagiellońskiego, którego wiedza, uwagi i pomysły były kluczowe w procesie przygotowywania niniejszej publikacji.
Chciałbym także podziękować dr Agacie Trawińskiej, Kierownik Pracowni Analizy Mowy i Nagrań Instytutu Ekspertyz Sądowych im. Prof. dra Jana Sehna w Krakowie za wszystkie cenne uwagi oraz inspiracje w zakresie przygotowania i przeprowadzenia okazania mowy.
Podziękowania należą się także wszystkim pozostałym osobom, które przyczyniły się do przeprowadzenia badań poprzez udzielenie próbki głosu, uczestnictwo w badaniach eliminacyjnych, uczestnictwo w badaniach właściwych oraz pomoc w ich organizacji. Szczególnie dziękuję uczniom i nauczycielom następujących instytucji: VII Liceum Ogólnokształcącego im. M. Kopernika w Częstochowie, IX Liceum Ogólnokształcącego im. Z. Wróblewskiego w Krakowie, Państwowej Ogólnokształcącej Szkoły Muzycznej II stopnia im. F. Chopina, Państwowej Szkoły Muzycznej II stopnia im. W. Żeleńskiego.