

Sławomir MICHALAK, Alicja MALANOWICZ, Tomasz PRUSIK, Sławomir PASZEK, Krzysztof ROGALA, Grzegorz STĘPIEŃ, Maciej CHYBOWSKI  
*Air Force Institute of Technology (Instytut Techniczny Wojsk Lotniczych)*

## SELECTED ASPECTS OF PHONOSCOPIC EXAMINATIONS OF AIRCRAFT RECORDS. PART II

### Wybrane aspekty fonoskopijnych badań nagrań lotniczych. Część II

**Abstract:** *The article presents crucial aspects related to examinations of records from cockpit voice recorders (CVR) and the possible effects of the application of digital correction of these type of records. It also presents the resulting limitations concerning phonoscopic recognition examinations of speakers and determining the content of their utterances. A method for the real reduction of the described limitations was suggested. Part II of the article describes research experiments conducted at Air Force Institute of Technology.*

**Keywords:** cockpit voice recorders, flight safety, investigation of the reasons for aircraft incidents

**Streszczenie:** *W artykule omówiono istotne problemy związane z badaniami nagrań pochodzących z pokładowych rejestratorów rozmów, a także możliwe skutki stosowania cyfrowej korekcji tego typu nagrań. Przedstawiono także wynikające stąd ograniczenia fonoskopijnych badań identyfikacyjnych mówców oraz ustalania treści ich wypowiedzi. Zaproponowano sposób realnego zmniejszenia opisanych ograniczeń tych badań. Część II artykułu przedstawia opis eksperymentów badawczych przeprowadzonych w Instytucie Technicznym Wojsk Lotniczych.*

**Słowa kluczowe:** pokładowe rejestratory rozmów, bezpieczeństwo lotów, badania przyczyn zdarzeń lotniczych

## **1. Introduction**

The rapid development of digital techniques in the world resulted in the popularisation of both specialist, professional noise processing systems as well as easily accessible, but also relatively advanced edition and correction algorithms. However, for a record to be considered useful in explaining the circumstances of an event, it should be determined what was uttered and by whom. The level of this reliability depends mainly on the technical quality of records and a character of a particular event [1].

The article addressed the study of the impact of the most popular activities aimed at investigating forensic speeches, especially on enabling to reliably recognize the speakers (i.e. indicating a particular speaker in each utterance without any doubts) conducted at the Avionics Division of the Air Force Institute of Technology under the supervision of experts from the Aviation Laboratory of Speech and Audio Analysis [6]. The first part of the article presented the overview of recorders used by the Armed Forces of the Republic of Poland, characteristics of records preserved by cockpit voice recorders (CVC) [7] as well as the assumptions and research methodology [8].

## **2. Description of research experiments**

### **2.1. Experiment No. 1**

This experiment consisted in simulating a standard nature of records registered by CVRs in the channel intended for signal retention from cockpit area microphones located in the vicinity of crew members. These are the records registered by cockpit voice recorders of the high quality of digital parameters, and in spite of that, some parts of the pertinent audio data present considerable difficulties with the understanding of the retained utterances or recognition of speakers. It is mostly due to significant distorting the utterances, which are then subject to examinations of intensive background sounds, even though the utterances are usually spoken directly into the microphone or reach it from close proximity.

This effect was achieved by mixing the previously prepared undistorted referential records with record simulating the distortions, but the relation of the referential speech signal to distortion degree was selected in this way so that the prepared investigated records do not allow to quickly understand the recorded utterances or raise doubts as to who is telling the particular words. According to assumption 1, the intensity of distortions was diversified in the case of utterances of every speaker.

#### **Characteristics of the prepared records**

- 1) Test referential records (references) and simultaneously subject to further conversions in experiment No. 2.

The foregoing records were registered in the isolated closed rooms with similar acoustics, without the participation of background sounds. They include the simultaneous utterances of two men and one woman, spoken directly into the microphone of a recording device. The same device recorded the statements of each speaker with CD parameter and stored in separate digital files.

2) Test distorting records.

They were recorded by the same device as referential records, with the same digital parameters.

These records include various distorting sounds: mainly excerpts from statements of different persons distant from the microphone and various noises. Environmental conditions ensured the longer duration of an echo than in referential records, and therefore, the sounds of diverse character underwent more uniform mixing.

3) Test records simulating the distorted aircraft records.

Test referential records of each speaker were mixed with the test distorting record. In this way, a test distorting record was created. In the case of two speakers (women K and man A), the intensity degree of distortions was chosen so as to cause a loss of intelligibility of a different degree, but in both cases only in fragments and concerning the utterance of the third speaker (man B), the complete loss of intelligibility was achieved, but the slight audibility was provided. Thus, the situation was created of perceptible and various difficulties in understanding the speech and recognising the speakers.

### **Preparation of voice samples for investigation**

Test referential records and test distorted records were used to make referential and test voice samples, which in this article were denoted as follows:

A-1 – first referential sample (undistorted) of the voice of the first man (A)

B-1 – first referential sample (undistorted) of the voice of the second man (B)

K-1 – first referential sample (undistorted) of the voice of the woman (K)

A-1z – first test sample (distorted) of the voice of the first man (A)

B-1z – first test sample (distorted) of the voice of the second man (B)

K-1z – first test sample (distorted) of the voice of the woman (K).

### **Computer comparisons of prepared voice samples**

All previously prepared referential and test voice samples were processed by a specialist SIVE programme [4, 5], applied since many years in phonoscopic recognition examinations of persons by forensic laboratory experts formulating opinions in this field. The primary objective was the comparison of sample pairs in view of their similarity within four physical voice parameters, mentioned in [8] and their total similarity.

**Results of comparisons**

Table 1 presents the results of comparisons of referential pairs (undistorted, references) of voice samples of both men and a woman. These are the high-quality comparisons of samples, which give an outline on real differences in voices of the speakers that were contrasted.

**Table 1**

**Comparisons of undistorted samples**

Physical voice parameters	Voice samples			
		A-1	B-1	K-1
F0	A-1	1,000	0,513	0,412
H F0		1,000	0,576	0,596
VT		1,000	0,964	0,574
T		1,000	0,469	0,675
----		----	----	----
ALL		1,000	0,591	0,513
F0	B-1	0,513	1,000	-0,038
H F0		0,576	1,000	0,981
VT		0,964	1,000	0,630
T		0,469	1,000	0,140
----		----	----	----
ALL		0,591	1,000	0,273

The results demonstrated in table 1 indicate that there should be no risk of confusing none of the speakers, in spite of the e.g., high probability of voice sonorosity parameter of both men. The auditory events (auditory analysis) confirmed both this similarity as well as the lack of the possibility of false recognition of both men. The investigation confirmed the correct calculation of maximum compatibility of comparing the same samples (1,000) done by software. Whereas, table 2 exhibited the results of mutual comparisons of some combinations of voice test samples (in this experiment-distorted).

**Table 2**

**Comparisons of undistorted samples with distorted samples**

Physical voice parameters	Voice samples						
		A-1z	A-1	B-1	B-1z	K-1	K-1z
F0	A-1z	1,000	0,495	0,144	0,527	0,590	0,510
H F0		1,000	0,643	0,943	0,727	0,973	0,966
VT		1,000	0,761	0,879	0,906	0,399	0,908
T		1,000	0,495	0,400	-0,048	0,352	0,307
----		----	----	----	----	----	----
ALL		1,000	0,564	0,442	0,527	0,582	0,618
F0	B-1z	0,527	-	-0,080	1,000	0,438	0,584
H F0		0,727	-	0,892	1,000	0,819	0,791
VT		0,906	-	0,806	1,000	0,277	0,995
T		-0,048	-	0,231	1,000	0,078	0,025
----		----	----	----	----	----	----
ALL		0,527	-	0,281	1,000	0,415	0,594
F0	K-1z	0,510	-	-	0,584	0,895	1,000
H F0		0,966	-	-	0,791	0,989	1,000
VT		0,908	-	-	0,995	0,268	1,000
T		0,307	-	-	0,025	0,307	1,000
----		----	----	----	----	----	----
ALL		0,618	-	-	0,594	0,708	1,000

The results presented in table 2 confirmed that indeed, the diversification of intensity of the same distorting signal causes a different effect, which is observed in the case of different speakers, exceptionally clear in the case of the speaker (B). Besides, the auditory analysis of distorted samples shows the highest voice signal degradation concerning the sample of a man (B) – both in terms of the intelligibility of his speech and the recognition of the speaker. Unfortunately, this analysis does not confirm the possibility to confuse the voice of a man (A) with the voice of a woman (K) in distorted samples, which would suggest lower similarity of sample (A-1z) to its own (A-1), than to both samples of a woman voice [distorted (K-1z) and undistorted (K-1)], which was shown in table 2. Not to mention that it should be emphasised that in line with the assessments of results of calculations made by the applied software (included previously), the results of all compiled lists [(A-1z):(B-1z) and (A-1z):(K-1z)], and even (A-1z):(K-1) – line (A-1z) of table 2] can/should be construed with the same probability that all of these samples belong to different persons and the same person as well!

The situation referred to has the following practical effects:

The first consists in the lack of clear results of the comparison of a distorted voice of one man with the distorted voice of the second man and with the distorted (or even undistorted) voice of the woman, based on the presented subject examinations.

The second one lies in the possibility to determine this on the basis of the auditory analysis, but only in the case of differences between the contrasted speakers, evident for such assessment (e.g., as in the majority of cases between a man and a woman).

The third one consists in the fact that strong distortions and deformations present in the records may be the reason of such intensive deformations of speech and voice features, that the speakers, whose are hard to confuse under normal acoustic conditions [see table 1, especially the lists of men samples: (A-1):(B-1)], are no longer so clearly distinguishable, both in the presented subject examinations [as in the comparison (A-1z):(B-1z) from table 2], as well as and in the auditory analysis (loss of intelligibility and significant audibility loss of utterance of man (B) in his distorted speech sample (B-1z)].

## **2.2. Experiment No. 2**

In this experiment, the distorted utterances (from experiment No. 1) were digitally corrected.

To this end, one used the algorithms available in commercial AVS Editor Audio software utilised for sound processing and correction, defined as moderately advanced ('auto correction' was used, which implements the sum of reduction of the dynamic range, normalisation and sound smoothing) and highly advanced ('noise reduction' with the algorithm of its profiling, this function can seem a 'learning' function).

The conversions made in this experiment imitate common behaviours of listeners convinced that this way will restore the original sound to the distorted (or deformed) speech signal. The effects of such conduct are illustrated, as previously, in the calculations of values of the same physical voice parameters and the additional list of sonograms (graphs of the distribution of the signal frequency in time, with their intensity) and the envelope of signals, at particular stages of conversions.

### **Preparation of voice samples**

In this experiment, samples of woman voice (K) were used, prepared for experiment No. 2, i.e.:

K-1 – first referential sample (undistorted) of the voice of the woman (K)

K-1z – first test sample (distorted) of the voice of the woman (K)

A-1 – first referential sample (undistorted) of the voice of the man (A)

A-1z – first test sample (distorted) of the voice of the first man (A)

B-1 – first referential sample (undistorted) of the voice of the second man (B)

B-1z – first test sample (distorted) of the voice of the second man (B).

Then, a distorted sample (K-1z) was corrected by creating:

K-1kor – second test sample (corrected) of the voice of the woman (K), made after using 'auto correction' function twice relative to sample (K-1z)

K-1f – third test sample (filtered) of the voice of the woman (K), made after using 'noise reduction' function relative to sample (K-1z)

A-1kor – second test sample (corrected) of the voice of the first man (A), made after using ‘auto correction’ function twice relative to sample (A-1z)

B-1kor – second test sample (corrected) of the voice of the second man (B), made after using ‘auto correction’ function twice relative to sample (B-1z).

**Computer comparisons of the prepared voice sample**

A referential sample and all three test samples of the voice of the woman (K) were subjected to the same calculations as in experiment No.1, and then, the obtained results were compared. Also, for each sample, the sonogram was produced, and for two samples the graph of the signal envelope was made.

**Results of comparisons**

Table 3 demonstrated the results of comparisons of the distorted sample (K-1z) with samples made after they were subjected to two different corrections.

**Table 3**

**Comparisons of the distorted sample, undistorted sample and corrected samples**

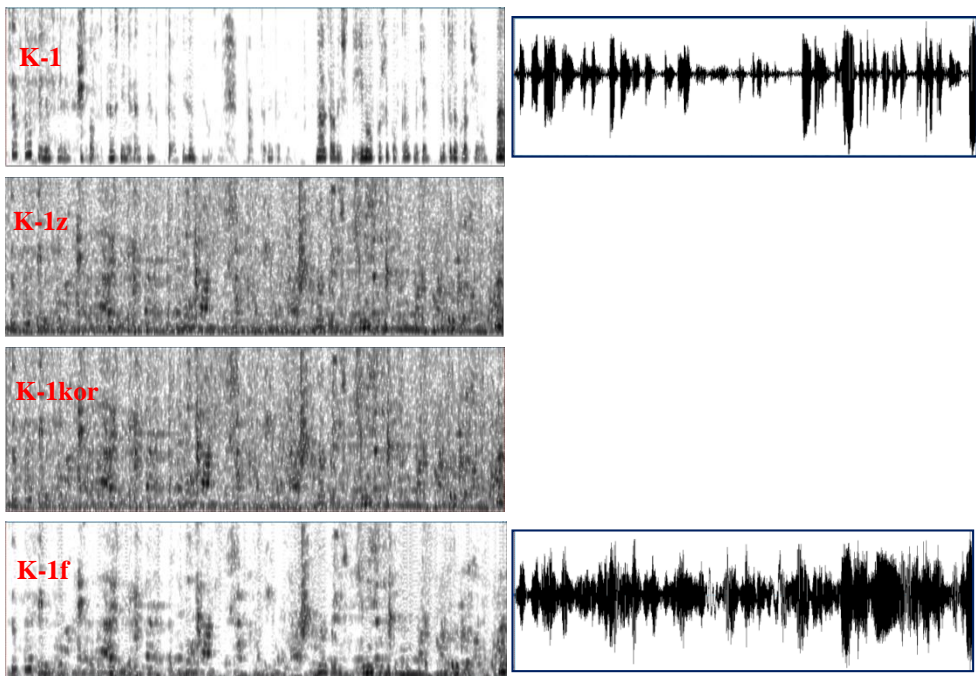
Physical voice parameters	Voice samples				
		K-1	K-1z	K-1kor	K-1f
F0	K-1	1,000	0,895	0,875	0,928
H F0		1,000	0,989	0,950	0,995
VT		1,000	0,268	0,180	0,895
T		1,000	0,307	0,700	0,663
-----		-----	-----	-----	-----
ALL		1,000	0,708	0,742	0,890
F0		K-1z	0,895	1,000	0,974
H F0	0,989		1,000	0,955	0,983
VT	0,268		1,000	0,981	0,611
T	0,307		1,000	0,549	0,484
-----	-----		-----	-----	-----
ALL	0,708		1,000	0,901	0,826
F0	K-1f		0,928	0,960	0,976
H F0		0,995	0,983	0,959	1,000
VT		0,895	0,611	0,527	1,000
T		0,663	0,484	0,624	1,000
-----		-----	-----	-----	-----
ALL		0,890	0,826	0,840	1,000

High and very high correlation of a distorted sample with both corrected samples [line (K-1z) of the table – above 0,9 and above 0,8] shall indicate the lack of a significant improvement of the situation and high probability of effects of both types of corrections. The results of comparisons between both corrected voice samples of this person [third line, comparison (K-1f) with (K-1kor)] suggest the similar conclusions. However, the sonograms

observed in fig. 1 and the results of the auditory analysis of these samples contradict the above conclusions.

Under the influence of the analysis of sonograms one may have an impression that the function of ‘noise reduction’ with its profiling restores the original quality of signal to a large extent and in this scope confirms the results from table 3 highlighted in blue [line (K-1), list (K-1):(K-1f)]. Unfortunately, this impression is only apparent, due to the fact that the auditory analysis of the sample (K-1f) indicated the presence of significant signal distortions, with the distortion degree of the sound of a woman’s voice (K) and the lack of intelligibility of her speech comparable with the level of these parameters already in the sample (K-1z). The importance of the differences in the quality of signal between the sample (K-1) and (K-1f) illustrate the graphs of signal envelope in fig. 1 more extensively than the results in table 3.

Furthermore, the conclusion on the small impact of distortions on the sample of woman’s voice (K-1), which could follow from the relatively slight signal degradation caused by its distortion [first line, comparing (K-1) with (K-1z)], is not confirmed neither by sonograms (fig. 1) nor in the auditory analysis of these samples.



**Fig. 1.** Results of samples’ analyses of the voice of the woman converted during experiment No. 2 (sonograms on the left, envelopes on the right)



Table 4 shows the results of comparisons of voice samples of both men (A) and (B), which were firstly distorted [by making samples (A-1z) and (B-1z)], and then they were subjected to ‘auto correction’ twice.

**Table 4**

**Comparisons of undistorted samples, distorted samples and corrected samples of the man (A) and (B)**

Physical voice parameters	Voice samples			Voice samples		
	A-1	A-1z	A-1kor	B-1	B-1z	B-1kor
<b>F0</b>	1,000	0,495	<b>0,071</b>	1,000	-0,080	<b>-0,148</b>
<b>H F0</b>	1,000	0,643	<b>0,440</b>	1,000	0,892	<b>0,261</b>
<b>VT</b>	1,000	0,761	<b>0,647</b>	1,000	0,806	<b>0,660</b>
<b>T</b>	1,000	0,495	<b>0,226</b>	1,000	0,231	<b>-0,129</b>
----	----	----	----	----	----	----
<b>ALL</b>	1,000	0,564	<b>0,254</b>	1,000	0,281	<b>0,058</b>

Results presented in table 4, in case of both speakers, show not only the lack of the expected improvement of similarity of samples after correction to the undistorted samples, but also its advancing reduction. Thus, the correction with the aid of ‘noise reduction’ function was not implemented.

### 3. Summary

The conducted experiments, especially the scope of their elaboration, do not fully reflect the complexity of issues concerning both the phonoscopic examinations of aircraft records as being distorted and deformed [2], but also the correction of sound records. They illustrate however the nature of difficulties existing despite using the most advanced technologies of digital signal processing, incomparably more advanced than the one used in experiments discussed in this article. The results of these experiments showed that:

1. Recognition examinations of speakers in distorted and deformed records involve the risk of obtaining ambiguous or unreliable results, providing that the findings will be based only on the results of computer calculations of presented voice parameters.
2. The solution to problems referred to in point 1 can be the use of the auditory analysis of the features of continuous speech and voice, but it also requires a satisfactory level of clarity and intelligibility of speech, and these parameters are limited by degree of distortions of nature of deformations.
3. The functions of digital corrections applied in experiment No. 2, to improve the quality of distorted records, are conversions, which are easy to use, but may cause a misleading view of their impacts. This fact suggests some caution when using them, and in extreme circumstances, to totally resign from them, even in phonoscopic examinations.

4. Phonoscopic speech and audio analyses and recognition examinations involve i.a., the interpretation of the final and partial results accumulated at each stage. The factor, which reduces the risk of committing errors in this field, is the expertise of research personnel, resulting from personal predispositions, extensive knowledge and rich experience [3].

The foregoing findings let us formulate the following conclusions:

1. The specificity of records registered by cockpit voice recorders and the specificity of digital processing of speech signal present a high risk of making errors in recognising the speakers and determining the content of their speech. Therefore, it is necessary to adopt the following principles enabling to minimise the above mentioned risk:
  - in phonoscopic speech and audio analyses: limited and conscious application of signal correction, selected individually to the case and the analysing person,
  - in recognition examinations: withdrawal from the correction of the analysed signal,
  - in both:
    - to narrow the scope of categorical findings only to unquestionable cases (unambiguous),
    - to allow the non-categorical findings (established with a specific probability level),
    - to deviate from findings, even non-categorical, in case of too many potential variant results,
    - to conduct research by the appropriately selected, trained and experienced staff, with the use of a specialist high-quality equipment, in specially adapted acoustic rooms.
2. Irrespective of the application of rules referred to in point 1, to increase the scope of using the cockpit voice recorder, it is necessary to:
  - provide a signal of the appropriate quality, which reaches the inputs of cockpit voice recorders, by the analysis and limitation of the negative impact of the elements of sound channels and ambient factors,
  - reduce the harmful conversion of signal in the recorders themselves and applications copying the sound data from recorders.

## **4. References**

1. Błasikiewicz S., Wójcik W.: Podstawowe zasady pobierania materiału porównawczego do badań identyfikacyjnych mowy. Problemy Kryminalistyki nr 80-81, Wydawnictwo Zakład Kryminalistyki KGMO, Warszawa 1969.
2. Błasikiewicz S.: Metoda odsłuchu szeptu i mowy intensywnie zakłóconej. Problemy Kryminalistyki nr 90, Wydawnictwo Zakład Kryminalistyki KGMO, Warszawa 1971.

3. Malanowicz A., Koźmiński L.: *Fonoskopia*, Zakład Kryminalistyki Szkoły Policji w Pile, Wydawnictwo SP, Piła 2009.
4. Malanowicz A., Kowalczyk A., Šalna B.: *Pakiet programowy SIVE, jako narzędzie wspomagające kryminalistyczne badania fonoskopijne z perspektywy praktyki eksperckiej – cz. I. Problemy Kryminalistyki nr 261*, Wydawnictwo CLK KGP, Warszawa 2008.
5. Malanowicz A., Kowalczyk A., Šalna B.: *Pakiet programowy SIVE, jako narzędzie wspomagające kryminalistyczne badania fonoskopijne z perspektywy praktyki eksperckiej – cz. II. Problemy Kryminalistyki nr 284*, Wydawnictwo CLK KGP, Warszawa 2010.
6. Malanowicz A., Prusik T.: *Laboratorium Fonoskopii Lotniczej*. Journal of KONBiN, 1(17) 2011.
7. Michalak S., Malanowicz A., Prusik T.: *System Badań Fonoskopijnych w Lotnictwie*. Prace Naukowe – Transport z. 92, Oficyna Wydawnicza Politechniki Warszawskiej, 2013.
8. Michalak S. et al.: *Selected aspects of phonoscopic examinations of aircraft records. Part I*. Journal of Konbin, Vol. 49, Iss. 4, 2019, DOI 10.2478/jok-2019-0082.

## **WYBRANE ASPEKTY FONOSKOPIJNYCH BADAŃ NAGRAŃ LOTNICZYCH. CZĘŚĆ II**

### **1. Wprowadzenie**

Błyskawiczny rozwój techniki cyfrowej na świecie doprowadził do upowszechnienia zarówno specjalistycznych, profesjonalnych systemów przetwarzania dźwięku, jak i łatwo dostępnych, choć także stosunkowo zaawansowanych algorytmów edycyjnych i korekcyjnych. Jednak żeby nagranie było przydatne do wyjaśnienia okoliczności zdarzenia, należy wiarygodnie ustalić, co zostało powiedziane i przez kogo. Poziom tej wiarygodności uwarunkowany jest przede wszystkim jakością techniczną nagrań oraz charakterem konkretnego zdarzenia [1].

W artykule opisano badania wpływu najbardziej popularnych działań na sygnale dźwiękowym przede wszystkim na zachowanie pełnych możliwości wiarygodnej identyfikacji mówców (tzn. pozbawionego wątpliwości wskazania konkretnego mówcy w każdej wypowiedzi), przeprowadzone w Zakładzie Awioniki Instytutu Technicznego Wojsk Lotniczych pod kierunkiem ekspertów Laboratorium Fonoskopii Lotniczej [6]. Przegląd rejestratorów użytkowanych przez Siły Zbrojne RP, charakterystykę nagrań utrwalanych przez pokładowe rejestratory rozmów [7] oraz założenia i metodę badawczą eksperymentów przedstawiono w części pierwszej artykułu [8].

### **2. Opis eksperymentów badawczych**

#### **2.1. Eksperyment nr 1**

W eksperymencie tym zasymulowano standardowy charakter nagrań utrwalanych przez rejestratory pokładowe, w kanale przeznaczonym do zapisu sygnału z mikrofonów otwartych, rozlokowanych w pobliżu członków załogi. Przejawia się on tym, że są to nagrania utrwalane przez pokładowe rejestratory o wysokiej jakości parametrów cyfrowych, a mimo to stwarzają – we fragmentach – znaczne kłopoty ze zrozumieniem utrwalonych w nich wypowiedzi lub rozpoznaniem mówców. Najczęściej jest to spowodowane istotnym zakłócaniem wypowiedzi podlegających późniejszym badaniom intensywnymi odgłosami tła, mimo że wypowiedzi te na ogół kierowane są wprost do mikrofonu lub dochodzą do niego z bliskiej odległości.

Efekt ten uzyskano poprzez zmiksowanie przygotowanych wcześniej niezakłóconych nagrań referencyjnych z nagraniami symulującymi zakłócenia, przy czym stosunek referencyjnego sygnału mowy do poziomu zakłóceń dobrano w taki sposób, aby przygotowane nagrania badawcze nie pozwalały na łatwe zrozumienie utrwalonych w nich wypowiedzi lub budziły wątpliwości co do tego, kto wypowiada konkretne słowa. Zgodnie z założeniem 1., zróżnicowano intensywność zakłóceń w przypadku wypowiedzi każdego z rozmówców [8].

### **Charakterystyka przygotowanych nagrań**

1) Testowe nagrania referencyjne (odniesienia) i równocześnie poddawane późniejszym przekształceniom w eksperymencie nr 2.

Nagrania te zostały utwalone w izolowanych pomieszczeniach zamkniętych, o zbliżonej akustyce, bez udziału sygnałów tła. Zawierają bezpośrednie wypowiedzi spontaniczne dwóch mężczyzn i jednej kobiety, skierowane wprost do mikrofonu urządzenia rejestrującego. Wypowiedzi każdego z mówców zostały utwalone przez to samo urządzenie rejestrujące, z parametrami płyt CD, i zapisane w odrębnych plikach cyfrowych.

2) Testowe nagrania zakłócające.

Zostały zarejestrowane przez to samo urządzenie co nagrania referencyjne, z takimi samymi parametrami cyfrowymi. W nagraniach tych zarejestrowano zróżnicowane odgłosy zakłócające: głównie fragmenty wypowiedzi różnych osób oddalonych od mikrofonu urządzenia rejestrującego oraz zróżnicowane hałasy. Warunki otoczenia zapewniały dłuższy czas pogłosu niż w nagraniach referencyjnych, przez co odgłosy o zróżnicowanym charakterze uległy bardziej jednolitemu wymieszaniu.

3) Testowe nagrania symulujące zakłócone nagrania lotnicze.

Testowe nagranie referencyjne każdego rozmówcy zmiksowano z testowym nagraniem zakłócającym, tworząc w ten sposób testowe nagrania zakłócone. W przypadku dwojga mówców (kobiety K i mężczyzny A) poziom intensywności zakłóceń dobrano tak, aby spowodował w różnym stopniu utratę zrozumiałości wypowiedzi, choć w obu przypadkach tylko we fragmentach. Natomiast w przypadku wypowiedzi trzeciego mówcy (mężczyzny B) doprowadzono do praktycznie całkowitej utraty zrozumiałości, przy nieznacznym zachowaniu słyszalności. W ten sposób stworzono sytuację odczuwalnych i zróżnicowanych trudności w rozumieniu mowy i rozpoznawaniu mówców.

### **Przygotowanie próbek głosów do badań**

Z testowych nagrań referencyjnych oraz z testowych nagrań zakłóconych utworzono referencyjne i badawcze próbki głosów, które w niniejszym artykule oznaczono następująco:

- A-1 – pierwsza próbka referencyjna (niezakłócona) głosu pierwszego mężczyzny (A)
- B-1 – pierwsza próbka referencyjna (niezakłócona) głosu drugiego mężczyzny (B)
- K-1 – pierwsza próbka referencyjna (niezakłócona) głosu kobiety (K)
- A-1z – pierwsza próbka badawcza (zakłócona) głosu pierwszego mężczyzny (A)
- B-1z – pierwsza próbka badawcza (zakłócona) głosu drugiego mężczyzny (B)

K-1z – pierwsza próbka badawcza (zakłócona) głosu kobiety (K).

### Komputerowe porównania przygotowanych próbek głosu

Wszystkie przygotowane wcześniej referencyjne i badawcze próbki głosów poddano przetwarzaniu przez specjalistyczny pakiet programowy SIVE [4, 5], od wielu lat wykorzystywany w fonoskopijnych badaniach identyfikacyjnych osób przez laboratoria kryminalistyczne sporządzające w tym zakresie opinie sądowe. Głównym celem było porównanie między sobą par próbek pod kątem ich podobieństwa w obrębie czterech fizycznych parametrów głosu, wymienionych w [8], oraz ich podobieństwa całkowitego.

### Wyniki porównań

W tabeli 1 przedstawiono wyniki wzajemnych porównań par referencyjnych (niezakłóconych, odniesienia) próbek głosów obu mężczyzn i kobiety. Są to porównania próbek wysokiej jakości, dające pogląd na rzeczywiste różnice głosów porównywanych mówców.

Tabela 1

### Porównania próbek niezakłóconych

Fizyczne parametry głosu	Próbki głosu			
		A-1	B-1	K-1
F0	A-1	1,000	0,513	0,412
H F0		1,000	0,576	0,596
VT		1,000	0,964	0,574
T		1,000	0,469	0,675
----		----	----	----
ALL		1,000	0,591	0,513
F0	B-1	0,513	1,000	-0,038
H F0		0,576	1,000	0,981
VT		0,964	1,000	0,630
T		0,469	1,000	0,140
----		----	----	----
ALL		0,591	1,000	0,273

Wyniki przedstawione w tab. 1 wskazują, że nie powinno zachodzić ryzyko pomylenia ze sobą żadnego z mówców, mimo np. wysokiego podobieństwa parametru dźwięczności głosu obu mężczyzn. Zarówno to podobieństwo, jak i brak możliwości błędnego rozpoznania obu mężczyzn potwierdziły wrażenia słuchowe (analizy audytywne). Badanie potwierdziło także fakt prawidłowego obliczenia przez oprogramowanie maksymalnej zgodności porównywania ze sobą tych samych próbek (1,000).

Natomiast w tab. 2 przedstawiono wyniki wzajemnych porównań niektórych kombinacji próbek badawczych głosu (w tym eksperymencie – zakłóconych).

Tabela 2

## Porównania próbek niezakłóconych z próbkami zakłóconymi

Fizyczne parametry głosu	Próbki głosu						
		A-1z	A-1	B-1	B-1z	K-1	K-1z
F0	A-1z	1,000	0,495	0,144	0,527	0,590	0,510
H F0		1,000	0,643	0,943	0,727	0,973	0,966
VT		1,000	0,761	0,879	0,906	0,399	0,908
T		1,000	0,495	0,400	-0,048	0,352	0,307
----		----	----	----	----	----	----
ALL		1,000	0,564	0,442	0,527	0,582	0,618
F0	B-1z	0,527	-	-0,080	1,000	0,438	0,584
H F0		0,727	-	0,892	1,000	0,819	0,791
VT		0,906	-	0,806	1,000	0,277	0,995
T		-0,048	-	0,231	1,000	0,078	0,025
----		----	----	----	----	----	----
ALL		0,527	-	0,281	1,000	0,415	0,594
F0	K-1z	0,510	-	-	0,584	0,895	1,000
H F0		0,966	-	-	0,791	0,989	1,000
VT		0,908	-	-	0,995	0,268	1,000
T		0,307	-	-	0,025	0,307	1,000
----		----	----	----	----	----	----
ALL		0,618	-	-	0,594	0,708	1,000

Wyniki przedstawione w tab. 2 potwierdziły, że faktycznie zróżnicowanie intensywności tego samego sygnału zakłócającego wywołuje różny skutek, co jest widoczne na przykładzie różnych mówców, a w największym stopniu w przypadku mówcy (B). Równocześnie w ocenie audytywnej próbek zakłóconych najwyższa degradacja sygnału mowy dotyczy próbki męczyzny (B) – tak w zakresie zrozumiałości jego wypowiedzi, jak i rozpoznania mówiącego. Jednak analiza ta nie potwierdza możliwości pomylenia w próbkach zakłóconych głosu męczyzny (A) z głosem kobiety (K), co sugerowałyby widoczne w tab. 2 niższe podobieństwo próbki (A-1z) do jego własnej (A-1), niż do obu próbek głosu kobiety [zakłóconej (K-1z) i niezakłóconej (K-1)]. Warto przy tym podkreślić, że wg kryteriów oceny rezultatów obliczeń wykonywanych przez zastosowany identyfikacyjny pakiet (zamieszczonych wcześniej), wyniki wszystkich wykonanych tu czterech zestawień [(A-1z):(B-1z) i (A-1z):(K-1z), a nawet (A-1z):(K-1) - wiersz (A-1z) tab. 2] mogą/powinny być interpretowane jako świadczące z równym prawdopodobieństwem o tym, że wszystkie te próbki należą do różnych osób, jak i do tej samej! Przedstawiona sytuacja ma następujące skutki praktyczne:

Pierwszy polega na braku jednoznacznych wyników porównania zakłóconego głosu jednego męczyzny z zakłóconym głosem drugiego i z zakłóconym (a nawet niezakłóconym) głosem kobiety, na gruncie zaprezentowanych badań instrumentalnych.

Drugi polega na możliwości dokonania tego rozstrzygnięcia na gruncie oceny audytywnej, lecz jedynie w przypadku różnic występujących pomiędzy porównywanymi mówcami, ewidentnych dla takiej oceny (jak np. w większości przypadków występujących pomiędzy kobietą i mężczyzną).

Trzeci natomiast polega na tym, że intensywne zakłócenia i zniekształcenia występujące w nagraniach mogą być przyczyną tak znacznych zniekształceń cech mowy i głosu, że mówcy, których nie da się ze sobą pomylić w normalnych warunkach akustycznych [patrz tab. 1, szczególnie zestawienia próbek mężczyzn: (A-1):(B-1)], przestają być tak jednoznacznie rozróżnialni, zarówno w przedstawionych badaniach instrumentalnych [jak w porównaniu (A-1z):(B-1z) z tabeli 2], jak też w ocenie audytywnej (utrata zrozumiałości i znaczna utrata słyszalności wypowiedzi mężczyzny (B) w jego próbce zakłóconej (B-1z]).

## **2.2. Eksperyment nr 2**

W eksperymencie tym wypowiedzi zakłócone (z eksperymentu nr 1) poddano cyfrowym korekcjom.

W tym celu zastosowano algorytmy dostępne w komercyjnym oprogramowaniu AVS Editor Audio, służącym do przetwarzania i korekcji dźwięku, określane jako: średnio zaawansowane (wykorzystano funkcję „auto korekcja”, realizującą sumę redukcji zakresu dynamiki, normalizacji i wyrównania dźwięku) oraz wysoko zaawansowane (wykorzystano funkcję „usuwanie szumu” z algorytmem jego profilowania; funkcja ta może sprawiać wrażenie „uczającej się”).

Przekształcenia wykonane w tym eksperymencie naśladują częste zachowania słuchaczy pozostających w przekonaniu, że sposób ten przywraca zakłóconemu (lub zniekształconemu) sygnałowi mowy jego rzeczywiste brzmienie. Ilustracją skutków takiego postępowania są, jak poprzednio, obliczenia wartości tych samych fizycznych parametrów głosów, a dodatkowo zestawienie sonogramów (wykresów rozkładu częstotliwości sygnału w czasie, z uwidocznioną ich intensywnością) i obwiedni sygnałów, na poszczególnych etapach podjętych przekształceń.

### **Przygotowanie próbek głosów**

W tym eksperymencie wykorzystano próbki głosu kobiety (K), przygotowane na potrzeby eksperymentu nr 1, tj.:

K-1 – pierwszą próbkę referencyjną (niezakłóconą) głosu kobiety (K)

K-1z – pierwszą próbkę badawczą (zakłóconą) głosu kobiety (K)

A-1 – pierwszą próbkę referencyjną (niezakłóconą) głosu pierwszego mężczyzny (A)

A-1z – pierwszą próbkę badawczą (zakłóconą) głosu pierwszego mężczyzny (A)

B-1 – pierwszą próbkę referencyjną (niezakłóconą) głosu drugiego mężczyzny (B)

B-1z – pierwszą próbkę badawczą (zakłóconą) głosu drugiego mężczyzny (B).

Następnie próbkę zakłóconą (K-1z) poddano korekcji cyfrowej, tworząc:



- K-1kor – drugą próbkę badawczą (skorygowaną) głosu kobiety (K), powstałą po dwukrotnym zastosowaniu funkcji „auto korekcji” względem próbki (K-1z)
- K-1f – trzecią próbkę badawczą (odfiltrowaną) głosu kobiety (K), powstałą po zastosowaniu funkcji „usuwanie szumu” względem próbki (K-1z)
- A-1kor – drugą próbkę badawczą (skorygowaną) głosu pierwszego mężczyzny (A), powstałą po dwukrotnym zastosowaniu funkcji „auto korekcji” względem próbki (A-1z)
- B-1kor – drugą próbkę badawczą (skorygowaną) głosu drugiego mężczyzny (B), powstałą po dwukrotnym zastosowaniu funkcji „auto korekcji” względem próbki (B-1z).

### Komputerowe porównania przygotowanych próbek głosu

Próbkę referencyjną i wszystkie trzy próbki badawcze głosu kobiety (K) poddano takim samym obliczeniom jak w eksperymencie nr 1, a następnie porównano ze sobą uzyskane wyniki. Dodatkowo, dla każdej próbki wykonano sonogram, a dla dwóch próbek wykonano wykres obwiedni sygnału.

### Wyniki porównań

W tab. 3 przedstawiono wyniki porównań próbki zakłóconej (K-1z) z próbkami powstałymi po jej poddaniu dwóm różnym korekcjom.

Tabela 3

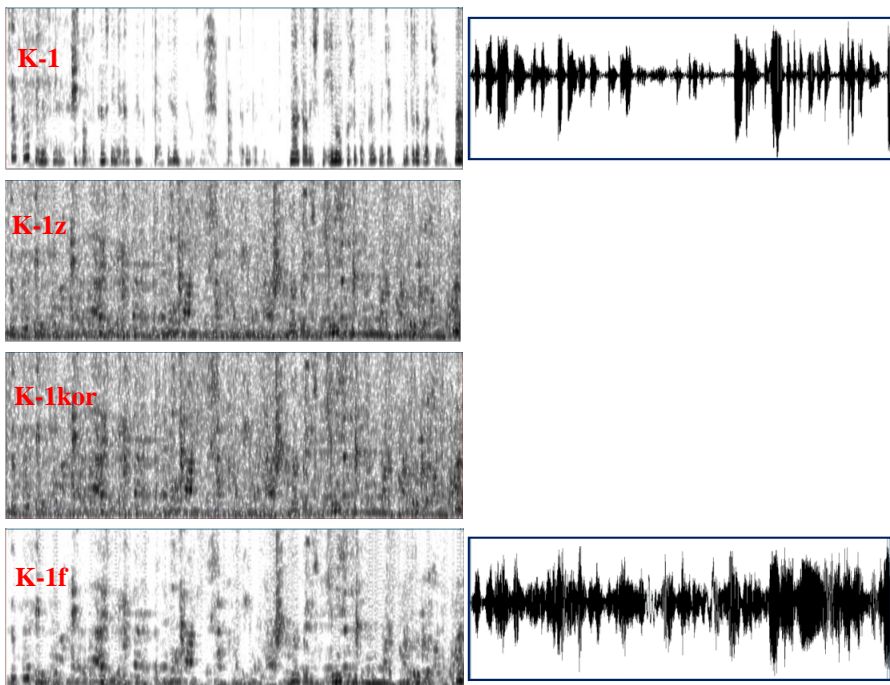
### Porównania próbki zakłóconej, niezakłóconej i próbek skorygowanych

Fizyczne parametry głosu	Próbki głosu				
		K-1	K-1z	K-1kor	K-1f
F0	K-1	1,000	0,895	0,875	0,928
H F0		1,000	0,989	0,950	0,995
VT		1,000	0,268	0,180	0,895
T		1,000	0,307	0,700	0,663
----		----	----	----	----
ALL		1,000	0,708	0,742	0,890
F0	K-1z	0,895	1,000	0,974	0,960
H F0		0,989	1,000	0,955	0,983
VT		0,268	1,000	0,981	0,611
T		0,307	1,000	0,549	0,484
----		----	----	----	----
ALL		0,708	1,000	0,901	0,826
F0	K-1f	0,928	0,960	0,976	1,000
H F0		0,995	0,983	0,959	1,000
VT		0,895	0,611	0,527	1,000
T		0,663	0,484	0,624	1,000
----		----	----	----	----
ALL		0,890	0,826	0,840	1,000

Wysoka i bardzo wysoka korelacja próbek zakłóconej z obiema próbkami skorygowanymi [wiersz (K-1z) tabeli – ponad 0,9 i ponad 0,8] powinna wskazywać na brak istotnego polepszenia sytuacji oraz duże podobieństwo skutków działania obu rodzajów korekcji. Podobne wnioski sugerują wyniki porównań pomiędzy obiema skorygowanymi próbkami głosu tej osoby [wiersz trzeci, porównanie (K-1f) z (K-1kor)]. Przeczą temu jednak sonogramy widoczne na rys. 1, a także wyniki oceny audytywnej tych próbek.

Pod wpływem analizy sonogramów można ulec wrażeniu, że funkcja „usuwanie szumu”, z jego profilowaniem, w dużym stopniu przywraca wyjściową jakość sygnału i w tym zakresie potwierdza wyniki z tab. 3 wyróżnione kolorem niebieskim [wiersz (K-1), zestawienie (K-1):(K-1f)]. Jest to jednak pozorne, gdyż analiza audytywna próbki (K-1f) wykazała obecność wyraźnych zniekształceń sygnału tła, przy poziomie zniekształcenia brzmienia głosu kobiety (K) oraz braku zrozumiałości jej mowy porównywalnym z poziomem tych parametrów już w próbce (K-1z). Istotność różnic w jakości sygnału pomiędzy próbkami (K-1) i (K-1f), w większym stopniu niż wyniki tab. 3, ilustrują wykresy obwiedni sygnału na rys. 1.

Równocześnie wniosek o niewielkim wpływie zakłóceń na próbkę głosu kobiety (K-1), który mógłby wynikać ze stosunkowo niewielkiej degradacji sygnału wywołanej jego zakłóceniem [pierwszy wiersz, porównanie (K-1) z (K-1z)], nie znajduje potwierdzenia ani w sonogramach (rys. 1), ani w ocenie audytywnej tych próbek.



Rys. 1. Zestawienie wyników analiz próbek głosu kobiety przekształczanych podczas eksperymentu nr 2 (z lewej sonogramy, z prawej obwiednie)

W tab. 4 przedstawiono wyniki porównań próbek głosów obu mężczyzn (A) i (B), które najpierw zakłócono [tworząc próbki (A-1z) i (B-1z)], a następnie dwukrotnie poddano działaniu funkcji „auto korekcji”.

Tabela 4

Porównania próbek niezakłóconych, zakłóconych i skorygowanych mężczyzny (A) i (B)

Fizyczne parametry głosu	Próbki głosu			Próbki głosu		
	A-1	A-1z	A-1kor	B-1	B-1z	B-1kor
F0	1,000	0,495	0,071	1,000	-0,080	-0,148
H F0	1,000	0,643	0,440	1,000	0,892	0,261
VT	1,000	0,761	0,647	1,000	0,806	0,660
T	1,000	0,495	0,226	1,000	0,231	-0,129
----	----	----	----	----	----	----
ALL	1,000	0,564	0,254	1,000	0,281	0,058

Wyniki przedstawione w tab. 4, w przypadku obu mówców, pokazują nie tylko brak spodziewanej poprawy podobieństwa próbek po korekcji do próbek niezakłóconych, ale wręcz istotnie pogłębiający się jego spadek. Z tej przyczyny pominięto korekcję funkcją „usuwanie szumu”.

### 3. Podsumowanie

Przeprowadzone eksperymenty, a szczególnie zakres ich omówienia, z konieczności nie oddają w pełni złożoności problematyki związanej zarówno z badaniami fonoskopijnymi nagrań lotniczych jako nagrań zakłóconych i zniekształconych [2], jak i samej korekcji zapisów dźwiękowych. Ilustrują one jednak charakter trudności istniejących mimo stosowania najnowocześniejszych technologii cyfrowego przetwarzania sygnałów, nieporównanie bardziej zaawansowanych od wykorzystanych w eksperymentach omówionych w niniejszym artykule. Wyniki tych eksperymentów wykazały bowiem, że:

1. Badania identyfikacyjne rozmówców w nagraniach zakłóconych i zniekształconych wiążą się z ryzykiem uzyskania niejednoznacznych lub niewiarygodnych wyników rozróżnienia osób, jeśli poczynione ustalenia będą opierać się wyłącznie na wynikach komputerowych obliczeń zaprezentowanych parametrów głosu.
2. Rozwiązaniem problemów, o których mowa w punkcie 1, może być zastosowanie audytywnej analizy cech mowy ciągłej i głosu, jednak i ona wymaga zachowania zadowalającego poziomu wyrazistości i zrozumiałości mowy, a parametry te także są limitowane poziomem zakłóceń i charakterem zniekształceń.
3. Funkcje cyfrowych korekcji zastosowane w eksperymencie nr 2, w celu poprawy jakości nagrań zakłóconych, są przekształceniami prostymi w użyciu, lecz

mogącymi wywołać mylący obraz skutków ich działania. Fakt ten skłania do ostrożności w ich wykorzystaniu, a w skrajnych przypadkach do całkowitej z nich rezygnacji, nawet w badaniach odsłuchowych.

4. Fonoskopijne badania odsłuchowe i identyfikacyjne wymagają m.in. interpretacji rezultatów końcowych, a także cząstkowych, uzyskiwanych na każdym ich etapie. Czynnikiem minimalizującym w tym zakresie ryzyko popełnienia błędów są kompetencje personelu badającego, wynikające z predyspozycji, rozległej wiedzy i doświadczenia [3].

Powyższe ustalenia nasuwają następujące wnioski praktyczne:

1. Specyfika nagrań utrwalanych przez pokładowe rejestratory dźwięku oraz specyfika cyfrowego przetwarzania sygnału mowy są – łącznie – źródłem wysokiego ryzyka popełnienia błędów w identyfikowaniu mówiących oraz ustalaniu treści ich wypowiedzi. Dlatego też konieczne jest przyjęcie zasad pozwalających na minimalizację tego ryzyka, polegających na:
  - w badaniach odsłuchowych: ograniczonym i świadomym stosowaniu korekcji sygnału, dobieranej indywidualnie do przypadku i osoby badającej,
  - w badaniach identyfikacyjnych: całkowitej rezygnacji z korekcji badanego sygnału,
  - w obu rodzajach badań:
    - zawężaniu zakresu ustaleń kategorycznych tylko do przypadków bezsprzecznych (jednoznacznych),
    - dopuszczaniu rozstrzygnięć niekategorycznych (ustalonych z określonym stopniem prawdopodobieństwa),
    - odstępowaniu od jakichkolwiek ustaleń, nawet niekategorycznych, w przypadku zbyt dużej możliwej wariantowości wyników,
    - wykonywaniu badań przez odpowiednio wyselekcjonowany, wyszkolony i doświadczony personel, z wykorzystaniem specjalistycznej aparatury wysokiej jakości, w przystosowanych akustycznie pomieszczeniach.
2. Niezależnie od stosowania zasad przedstawionych w punkcie 1, w celu zwiększenia zakresu wykorzystania pokładowych rejestratorów rozmów należy:
  - zapewnić odpowiedniej jakości sygnał docierający do wejść rejestratorów pokładowych, poprzez analizę i ograniczenie negatywnego oddziaływania elementów torów fonicznych i czynników otoczenia,
  - zredukować szkodliwe przekształcenia sygnału w samych rejestratorach oraz aplikacjach kopiujących dane dźwiękowe z rejestratorów.

## **4. Literatura**

1. Błasikiewicz S., Wójcik W.: Podstawowe zasady pobierania materiału porównawczego do badań identyfikacyjnych mowy. *Problemy Kryminalistyki*, nr 80-81, Wydawnictwo Zakład Kryminalistyki KGMO, Warszawa 1969.
2. Błasikiewicz S.: Metoda odsłuchu szeptu i mowy intensywnie zakłóconej. *Problemy Kryminalistyki* nr 90, Wydawnictwo Zakład Kryminalistyki KGMO, Warszawa 1971.
3. Malanowicz A., Koźmiński L.: *Fonoskopia*, Zakład Kryminalistyki Szkoły Policji w Pile, Wydawnictwo SP, Piła 2009.
4. Malanowicz A., Kowalczyk A., Šalna B.: Pakiet programowy SIVE, jako narzędzie wspomagające kryminalistyczne badania fonoskopijne z perspektywy praktyki eksperckiej – cz. I. *Problemy Kryminalistyki*, nr 261, Wydawnictwo CLK KGP, Warszawa 2008.
5. Malanowicz A., Kowalczyk A., Šalna B.: Pakiet programowy SIVE, jako narzędzie wspomagające kryminalistyczne badania fonoskopijne z perspektywy praktyki eksperckiej – cz. II. *Problemy Kryminalistyki*, nr 284, Wydawnictwo CLK KGP, Warszawa 2010.
6. Malanowicz A., Prusik T.: Laboratorium Fonoskopii Lotniczej. *Journal of KONBiN*, 1(17), 2011.
7. Michalak S., Malanowicz A., Prusik T.: System Badań Fonoskopijnych w Lotnictwie. *Prace Naukowe – Transport*, z. 92, Oficyna Wydawnicza Politechniki Warszawskiej, 2013.
8. Michalak S. i in.: Selected aspects of phonoscopic examinations of aircraft records. Part I. *Journal of Konbin*, Vol. 49, Iss. 4, 2019, DOI 10.2478/jok-2019-0082.

