

Waldemar Maciejko
Jacek Rzeszotarski

SYSTEM JAKOŚCI W LABORATORIUM FONOSKOPIJNYM. WALIDACJA METOD BADAWCZYCH (część I)

WSTĘP

Zlecniodawcy kryminalistycznych badań nagrań dźwięku – prokuratorzy i sędziowie - zwracają się do biegłych fonoskopii – pomocników procesowych – w sytuacji, gdy ocena zabezpieczonego dowodu wymaga stosownej wiedzy specjalnej¹. Zlecniodawca określa cel badań stawiany przed biegłym, natomiast biegły, jako dysponent wiedzy specjalnej, niezależnie stwierdza, czy jest w stanie spełnić oczekiwania zlecniodawcy (czy dysponuje odpowiednim oprzyrządowaniem i niezbędną fachowością), a także określa, jakimi metodami będzie dochodził do ustalenia prawdy².

Mimo że każda opinia biegłych podlega ocenie sądu³, to faktycznie zlecniodawca badań nie dysponuje narzędziami merytorycznej oceny wyborów dokonanych przez biegłego. W polskim orzecznictwie brakuje jednolitych kryteriów oceny naukowości dowodu. Ogólne wymagania mówią, że ekspertyza powinna opierać się na jednolitej i powszechnie przyjętej metodzie badawczej⁴ oraz że do jej wydania należy wykorzystywać wyniki najnowszych badań naukowych⁵. Niezbędne jest zatem precyzyjne udzielenie zlecniodawcy informacji na temat zastosowanych metod badawczych. Wprowadzający do procesu karnego dowód z opinii biegłego powinien być bowiem pewien, że badania zostały wykonane najlepiej jak to tylko możliwe, z wykorzystaniem odpowiednich metod.

¹ Art. 193 kpk.

² Postanowienie SN z 8.01.2007 r. w sprawie o sygn. akt II KK 194/06. OSNwSK 2007/1/86.

³ Wyrok SN z 12.10.2006 r. w sprawie o sygn. akt IV KK 236/06. OSNwSK 2006/1/1958.

⁴ Orzeczenie SN z 12.12.1974 r., III CRN 305/74, LEX nr 7630.

⁵ Orzeczenie SN z 25.02.1975 r., II CR 898/74, LEX nr 7663; orzeczenie SN z 6.11.2000 r., IV KKN 477/99, Prokuratura i Prawo 2001/4/9.

Wymagania w stosunku do laboratoriów badawczych, które chcą wykazać, że uzyskiwane przez nie wyniki są miarodajne, określa norma międzynarodowa PN-EN ISO/IEC 17025:2005. Zgodnie z normą wynik miarodajny to wynik, który jest:

- wiarygodny – uzyskany za pomocą metody o właściwej wartości diagnostycznej. W przypadku analizy ilościowej pożądane jest, oprócz oszacowania wartości mierzonej, podanie wartości niepewności. W przypadku niektórych badań fonoskopijnych lub szerzej kryminalistycznych podanie niepewności nie zawsze daje czytelny wynik lub wręcz jest niemożliwe; kluczowe jest zidentyfikowanie i kontrolowanie źródeł niepewności;
- użyteczny – wynik został poprawnie zinterpretowany, a wnioski są zrozumiałe nie tylko dla specjalistów z danej dziedziny, ale przede wszystkim dla organów zlecających wykonanie badań. W tym pojęciu zamyka się również dążenie do dostarczania wyników w rozsądnym czasie, takim, aby otrzymanie opinii nie wpływało na nieuzasadnione przedłużanie postępowania⁶;
- rzetelny – laboratorium w trakcie badań postąpiło zgodnie z dobrą praktyką laboratoryjną⁷, a wszystkie czynności zostały przeprowadzone z wszelką starannością.

Biegli fonoskopii stają więc przed wyzwaniem - podobnie jak eksperci innych dziedzin kryminalistyki – sprostania wymaganiom stawianym przez normę ISO 17025.

Niniejszy artykuł stanowi swojego rodzaju studium prac naukowych, norm międzynarodowych, poradników organizacji międzynarodowych (EA, ILAC, AES) oraz publikowanych sprawozdań grup roboczych (ENFSI, SWGDE-IT) związanych z systemowym podejściem do zarządzania jakością w laboratorium fonoskopijnym, ujednoczeniem metod badawczych i walidacją metod. Następnie w oparciu o tą dokumentację zaproponowano ogólny schemat walidacji metod stosowanych przez biegłych fonoskopii.

WALIDACJA METOD WG POLSKIEJ NORMY PN-EN ISO/IEC 17025

Zgodnie z normą ISO 17025, laboratoria powinny wykorzystywać metody badawcze opublikowane w normach międzynarodowych, regionalnych lub krajowych. Ze względu na charakter badań, które w dużej mierze należą do grupy badań jakościowych, nie istnieją spójne metody szacowania niepewności pomiaru. Szacowanie niepewności metod jakościowych w ogólności jest problemem dotąd nierozwiązanym. W poradniku EA (ang. *European co-operation for Accreditation*) znaleźć można następującą informację: *Obecnie bierze się pod uwagę tylko niepewność pomiaru w badaniach ilościowych. Strategia dotycząca postępowania z wynikami badań jakościowych powinna zostać opracowana przez społeczność*

⁶ M. Pekała, E. Marciniak, *Pojęcie jakości we współczesnej technice kryminalistycznej*, Problemy Kryminalistyki nr 260/2008, s. 45-54.

⁷ PN-EN ISO/IEC 17025:2005 Ogólne wymagania dotyczące kompetencji laboratoriów badawczych i wzorcujących.

naukową⁸. Podobne stanowisko reprezentuje ILAC (ang. *International Laboratory Accreditation Cooperation*): *Niektóre badania są czysto jakościowe i nadal jeszcze rozważa się jak w takich przypadkach stosować niepewność pomiaru... Uważa się, że kwestia szacowania niepewności pomiaru w odniesieniu do wyników jakościowych jest obszarem, gdzie potrzebne są dalsze wytyczne*⁹.

Norma dopuszcza wykorzystanie metod nieznormalizowanych, opublikowanych przez renomowane organizacje techniczne lub w odpowiednich czasopiśmie naukowych, jednakże metody nieznormalizowane, przed zastosowaniem, powinny być właściwie zwalidowane.¹⁰ Walidacja (inaczej uwierzytelnienie) jest potwierdzeniem, przez zbadanie i przedstawienie obiektywnego dowodu, że zostały spełnione wymagania dotyczące konkretnego zamierzonego zastosowania¹¹ danej metody analitycznej. To określenie stawia proces walidacji jako jeden z najważniejszych czynników systemu jakości opartego o normę ISO 17025. Wyróżnia się dwa zasadnicze cele walidacji:

- ustalenie sprawności działania i ograniczeń metody i dzięki temu dokonanie identyfikacji czynników, które mogą wpłynąć na otrzymany wynik i w jakim zakresie, oraz jaki poziom precyzji i dokładności można osiągnąć,
- sprawdzenie, czy metoda jest przydatna do określonych celów.

Polska Norma ISO 17025 przewiduje pięć podstawowych technik określania możliwości metody:

1. wzorcowanie z wykorzystaniem wzorców odniesienia,
2. porównanie wyników z wynikami uzyskanymi innymi metodami,
3. porównania międzylaboratoryjne,
4. ocena czynników wpływających na wynik,
5. ocena niepewności wyników oparta na naukowym rozumieniu podstaw teoretycznych metody i doświadczeniu praktycznym.

Proces walidacji ma obejmować wszystkie etapy badań, włączając w to – zgodnie z definicją metody analitycznej – czynności wstępne oraz zabiegi następujące po oznaczeniu cech nagrania aż do opracowania wyników i sformułowania wniosków z przeprowadzonych badań. Walidacja zatem powinna obejmować procedury przechowywania materiałów badań, postępowanie z nimi oraz właściwy proces badawczy. Zaleca się, aby każde laboratorium określiło konkretne wymagania dotyczące możliwości realizacyjnych poszczególnych celów badawczych i zachowało dane z walidacji w celu

⁸ Wytyczne EA dotyczące wyrażania niepewności w badaniach ilościowych, EA-04/16 Grudzień 2003.

⁹ *Wprowadzenie problematyki niepewności pomiaru w badaniach w związku z wejściem do stosowania normy ISO/IEC 17025*, ILAC-G17:2002, Australia 2002.

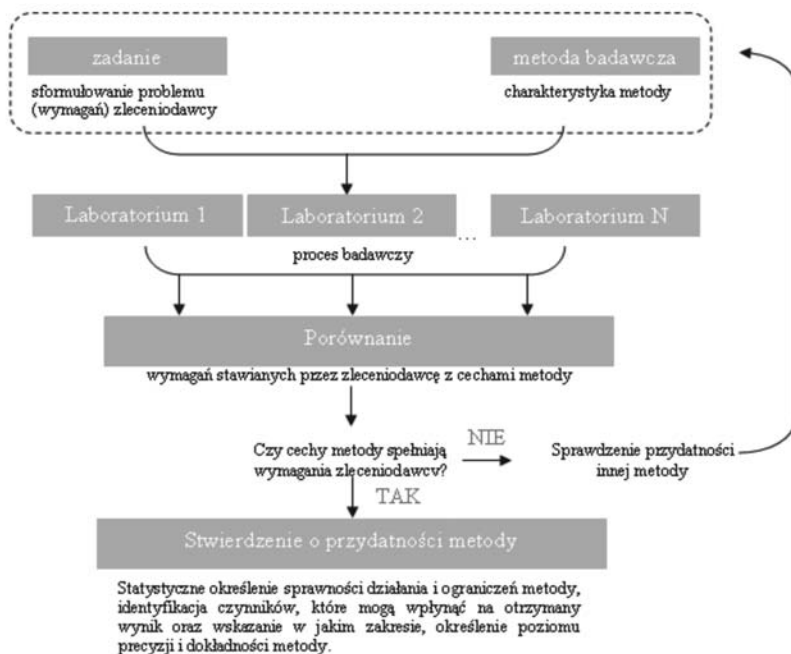
¹⁰ PN-EN ISO/IEC 17025 op. cit.

¹¹ PN-EN ISO 9000:2006, *Systemy zarządzania jakością. Podstawy i terminologia*.

udowodnienia, że metoda spełnia wymagania. W zależności od stosowanej metody, następujące czynniki mają szczególne znaczenie:

- odtwarzalność lub powtarzalność,
- próg wykrywalności,
- porównanie z innymi metodami,
- badania międzylaboratoryjne.¹²

Niektóre metody badawcze stosowane w fonoskopii należą do grupy badań sensorycznych (spisanie treści, identyfikacja w obrębie materiału dowodowego), inne wykorzystują do wyznaczenia cech środki techniczne (przydźwięk sieciowy, spektrograf), jednakże ich wynik jest również typowo jakościowy (np. wniosek o autentyczności nagrania). Uniemożliwia to ściśle, metrologiczne i statystyczne określenie niepewności wyniku. W takich przypadkach jej oszacowanie może być oparte np. o odtwarzalność. Wydaje się, że najprostszym sposobem walidacji tych metod jest wykorzystanie porównań międzylaboratoryjnych. Odpowiednia organizacja i przeprowadzenie takich badań umożliwi wyznaczenie parametrów analitycznych metod oraz pomoże zidentyfikować składowe niepewności wpływające na wynik badań. Ogólny algorytm określania możliwości analitycznych metod za pomocą porównań międzylaboratoryjnych przedstawiono na schemacie poniżej.



Rys. 1. Etapy postępowania walidacyjnego w oparciu o wyniki testów międzylaboratoryjnych.

¹² Akredytacja laboratoriów badań sensorycznych, EA-4/09 2003.

Każda z metod wykorzystywanych przez ekspertów fonoskopii wymaga odrębnego podejścia w trakcie organizacji, realizacji oraz podsumowania badań międzylaboratoryjnych. Z punktu widzenia walidacji niewątpliwie najistotniejszym jest etap ostatni, tzw. statystyka programu testów. Wynikiem tej części powinna być próba określenia niepewności otrzymywanych wyników.

TESTY BIEGŁOŚCI JAKO NARZĘDZIE WYZNACZENIA PARAMETRÓW ANALITYCZNYCH METOD BADAWCZYCH

Porównania międzylaboratoryjne stosowane w odniesieniu do badań fonoskopijnych, w zależności od celu badań, można podzielić na dwa podstawowe typy przedstawione poniżej.



Rys. 2. Podstawowa klasyfikacja badań międzylaboratoryjnych.¹³

Badania międzylaboratoryjne wykonywane w celu sprawdzenia metod badawczych charakteryzują się tym, że wszyscy uczestnicy przy użyciu tych samych metod prowadzą badania z wykorzystaniem identycznego materiału badań.

Badanie kompetencji laboratorium ma na celu skontrolowanie jakości wyników oraz określenie kompetencji pracowników w zakresie stosowania konkretnej metody, wskazanej przez organizatora porównań międzylaboratoryjnych jako właściwej dla uzyskania miarodajnego wyniku.

¹³ L. Cortez, *Use of LRM in quality control: interlaboratory testing – EC Growth Projects*, TRAP-LRM/TRAP-NAS, 2001.

Wymagania ogólne dla materiałów badawczych wykorzystywanych w międzylaboratoryjnych testach biegłości laboratoriów kryminalistycznych następująco określiło ENFSI:

- treść i jakość nagrań powinna być maksymalnie zbliżone do rzeczywistych,
- szczegóły badania są przedstawione w protokole przygotowanym przez organizatora,
- materiał badawczy powinien być jednolity, odpowiednio scharakteryzowany i udokumentowany¹⁴.

Zespół dokonujący badania powinien zostać ściśle określony wraz z rolą i zakresem odpowiedzialności. W przypadku badań jakościowych to właśnie zespół laboratorium stanowi rzeczywiste narzędzie pomiarowe, a wyniki zależą od jego członków. Oznacza to, że niewłaściwy dobór członków zespołu może obciążyć wyniki - odpowiedzialność za właściwy dobór kompetentnego zespołu spoczywa na kierownictwie laboratorium.

ORGANIZATOR BADAŃ BIEGŁOŚCI

Organizatorem badań biegłości może być jednostka, która spełnia wymagania stawiane tego typu podmiotom przez organizacje krajowe i międzynarodowe. ILAC opracował „Wymagania dotyczące kompetencji organizatorów programów badań biegłości”¹⁵. Z kolei właściwym dokumentem PCA w tym zakresie są wymagania szczegółowe w zakresie akredytacji organizatorów badań biegłości¹⁶. Podstawowe wymagania stawiane organizatorowi badań biegłości zostaną przedstawione poniżej.

Organizator badań biegłości jest zobowiązany do uzyskania akredytacji. Ma on prawo prowadzić tylko takie programy, co do których wykaże kompetencję. Oczekiwany sposobem wykazania kompetencji jest zrealizowanie programu badania biegłości co najmniej raz w ciągu trzech lat przed złożeniem wniosku o akredytację¹⁷. Spełnienie tego wymogu w przypadku badań fonoskopijnych jest obecnie w Polsce niemożliwe – porównania międzylaboratoryjne jak dotychczas nie były prowadzone. Wydaje się zatem, że akredytacja organizatora badań biegłości będzie możliwa dopiero po przeprowadzeniu pierwszych badań biegłości i wykazaniu, że organizator programu spełnił oczekiwania i wywiązał się z roli jednostki dostarczającej właściwe próbki, określającej adekwatną metodykę badawczą, czuwającej nad poprawnością przebiegu programu i wiarygodnie analizującej uzyskane wyniki.

¹⁴ *Standing Committee for Quality and Competence Guidance on the Conduct of Proficiency Tests and Collaborative Exercises within ENFSI*, 18-02-2005 ENFSI.

¹⁵ *Wytyczne ILAC - Wymagania dotyczące kompetencji organizatorów programów badań biegłości*, ILAC-G13:08/2007.

¹⁶ *Akredytacja organizatorów badań biegłości. Wymagania szczegółowe*, DAPT-01.

¹⁷ DAPT-01, s. 4.

ILAC zwraca uwagę na warunki, jakie powinien spełniać organizator badań biegłości¹⁸. Najważniejsze z nich to:

- niezależność od innych podmiotów, którą trzeba rozumieć jako brak powiązań z jakimkolwiek laboratorium uczestniczącym w programach oraz brak angażowania się w działalność, która mogłaby być interpretowana jako mogąca skutkować stronniczością,
- zdolność prowadzenia procedur w sposób zapewniający poufność. Błędy powinny być opisywane w sposób zapewniający anonimowość wszystkich uczestników porównań międzylaboratoryjnych,
- dysponowanie wykwalifikowanym (także w zakresie możliwości badawczych fonoskopii) i świadomym rangi realizowanych zadań personelem na każdym szczeblu, od kierownictwa po personel techniczny,
- wdrożenie i utrzymywanie właściwego systemu zarządzania jakością, co gwarantuje, że podmiot ten zapewnia najwyższą z możliwych jakości usług w deklarowanym zakresie działalności¹⁹.

Nie jest celem niniejszej publikacji wskazywanie jednostki, która mogłaby pełnić funkcję organizatora badań biegłości. Należy jednak zauważyć, że podobne próby były już czynione w XX wieku, jednak wówczas osiągnięcie kompromisu pomiędzy poszczególnymi laboratoriami badań kryminalistycznych nie było możliwe²⁰.

Na organizatorze badań biegłości spoczywa obowiązek określenia celu programu, który będzie determinował dalsze procedury, od przygotowania próbek aż do oceny jego wyników. Cel musi być odpowiedni dla specyfiki badawczej fonoskopii. Najlepiej, gdyby wskazywał ekspertom sposób prowadzenia badań w przypadku, który nie został jeszcze jednoznacznie zaakceptowany przez środowisko ekspertów – na organizatorze spoczywa więc rola kreatora, wyznaczającego nowe kierunki badań w niespotykanych dotychczas w praktyce opiniodawczej przypadkach. Organizator powinien także pełnić rolę doradcą, zwłaszcza dla ośrodków o mniejszej praktyce w zakresie wydawania opinii fonoskopijnych.

Na organizatorze spoczywa obowiązek dostarczenia uczestnikom wszelkich danych niezbędnych do właściwego postępowania z próbką testową. W szczególności ma określić realny termin zakończenia badań (uwzględniający czasochłonność poszczególnych analiz), określić metodę właściwą do rozwiązania postawionego problemu (gdy jest to niezbędne, organizator powinien dostarczyć próbki referencyjne, np. wahań prądu sieciowego w przypadku badań autentyczności nagrań cyfrowych).

¹⁸ ILAC-G13:08/2007, s. 9-10.

¹⁹ Opis poszczególnych elementów systemu nie stanowi przedmiotu niniejszego artykułu.

²⁰ T. Tomaszewski, J. Rzeszutarski, *Weryfikacja kwalifikacji biegłych wydających opinie kryminalistyczne (na przykładzie opinii fonoskopijnych)*, Kryminalistyka i Nauki Penalne Wobec Przemocności. Księga Pamiątkowa dedykowana Profesorowi Mirosławowi Owocowi, Poznań 2008, s. 255-265.

Po odpowiednim przeprowadzeniu badań biegłości (zgodnym z zaleceniami ILAC) organizator musi przeprowadzić właściwą i poprawną analizę statystyczną otrzymanych wyników. Ma on uwzględniać m.in. oczekiwaną niepewność wyników oraz liczbę uczestników programu.

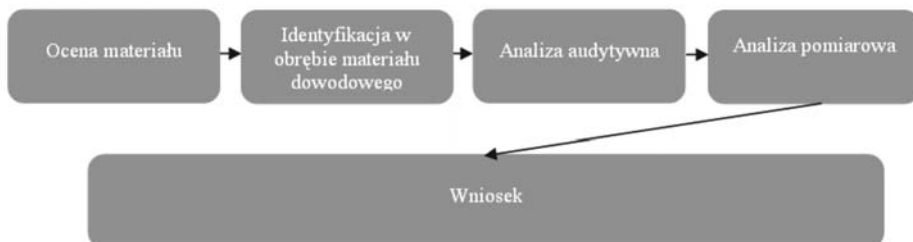
Szczegółowe wymagania techniczne, które powinien spełniać organizator programów, zostały zawarte w cytowanych dokumentach²¹.

WALIDACJA METOD IDENTYFIKACJI MÓWCÓW

Złożoność przedmiotu badań identyfikacji mówców sprawia, że w badaniach zalecane jest podejście holistyczne, uwzględniające analizy audytywne oraz parametryczne²². W Polsce podejście takie stosuje metoda językowo-pomiarowa, której jeden z wariantów jest opisany w polskiej literaturze²³. Celem tych badań jest:

1. identyfikacja mówców w obrębie materiału dowodowego (przypisanie poszczególnym mówcom konkretnych wypowiedzi),
2. ustalenie, czy wypowiedzi mówcy, od którego uzyskano materiał porównawczy, zostały utrwalone w materiale dowodowym.

Na wynik analiz językowo-pomiarowych składa się szereg czynności:



Rys. 3. Etapy wykorzystania metody językowo-pomiarowej. Pierwsza czynność - ocena materiału, to analiza opisowa nagrań, w której dokonywana jest ocena ich przydatność pod kątem dalszej analizy porównawczej, wskazuje się również wymagania co do materiału porównawczego (czas trwania, rodzaj wypowiedzi, kanał transmisji).

²¹ DAPT-01, ILAC-G13:08/2007.

²² T. Tomaszewski, J. Rzeszotarski, *Holistyczne ujęcie analizy jakościowej i ilościowej w opinio-waniu biegłych na przykładzie badań pismoznawczych i fonoskopijnych*, Problemy Kryminalistykinr 264/2009, s. 5-10.

²³ J. Dolecki, J. Rzeszotarski, *Zastosowanie metody językowo-pomiarowej do identyfikacji osób w badaniach fonoskopijnych*, Z Zagadnień Nauk Sądowych”, t. LII, 2002, s. 108-123.

Analiza pomiarowa

Analiza parametryczna (czyli oparta na pomiarach parametrów głosu) dodatkowo dzieli się na metodę techniczną²⁵ oraz automatyczną²⁶. W badaniu technicznym to ekspert wskazuje segment, z którego zostanie przeprowadzona ekstrakcja cech głosu (w praktyce analizie poddawane są m.in. częstotliwości formantowe i częstotliwość podstawowa tonu krtaniowego). Najdłużej w aplikacjach kryminalistycznych wykorzystuje się najlepiej udokumentowane częstotliwości formantowe²⁷.

Obecnie coraz częściej wykorzystuje się bardziej skomplikowane cechy głosu, jak np. cepstralne współczynniki częstotliwościowe skali melowej (MFCC)²⁸, współczynniki liniowego kodowania perceptualnego (PLP)²⁹. Wykorzystanie tych czynników wymaga jednak dysponowania odpowiednim oprogramowaniem, które prowadzi m.in. normalizację cech oraz ze względu na ilość danych automatyzuje kolejne etapy porównania.

Walidacja pomiarowych metod identyfikacji mówców polega na sprawdzeniu, jakie wartości charakteryzujące stopień podobieństwa (współczynnik korelacji, iloraz wiarygodności, punkty (ang. *scores*) itp.) zostanie otrzymany, gdy w warunkach danego laboratorium zastosowano daną metodę. W celu takiego sprawdzenia konieczne jest dysponowaniem bazą głosów testowych. Do takich celów wykonuje się specjalne bazy, przykładowe przedstawia poniższa tabela.

²⁵ P. Rose, *Technical forensic speaker recognition: Evaluation, types and testing of evidence*, Computer Speech and Language 20 (2006), 159-191.

²⁶ D.A. Reynolds, T.F. Quatieri, R.B. Dunn, *Speaker verification using adapted gaussian mixture models*, MIT Lincoln Labs, Digital Signal Processing 10, 19-41 (2000).

²⁷ P. Rose, *Forensic Speaker Identification*, New York 2002.

²⁸ D.A. Reynolds, T.F. Quatieri, R.B. Dunn, *Speaker verification using adapted gaussian mixture models*, Digital Signal Processing 10, 19-41 (2000).

²⁹ H. Hermansky, B.A. Hanson, H. Wakita, *Perceptually based linear predictive analysis of speech*, Speech Technology Laboratory, Santa Barbara 1985.

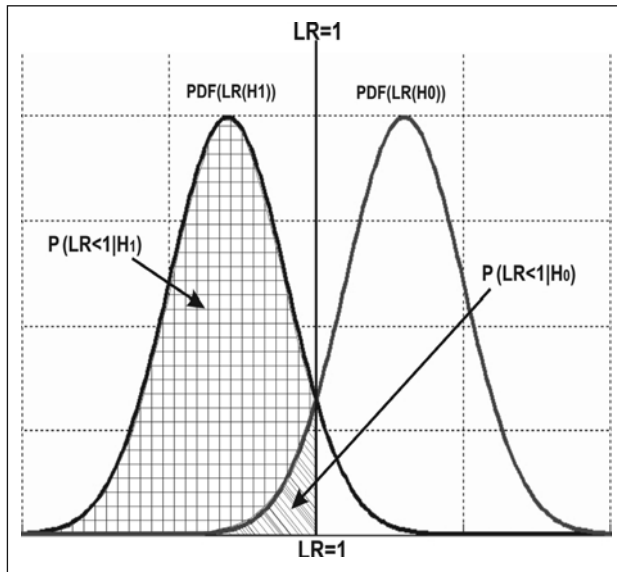
Rok	Corpus	Rozmiar	Rodzaj wypowiedzi
1990/1991	Switchboard I	543 mówców, 2400 rozmów	Wypowiedzi w języku amerykańskim angielskim na określony temat
1996	Switchboard II phase 1	657 mówców, 3638 rozmów	Dialekt amerykańskiego angielskiego, rozmowy przez telefon
1997	Switchboard II phase 2	679 mówców, 4472 rozmowy	Dialekt amerykańskiego angielskiego, rozmowy przez telefon
1997/1998	Switchboard II phase 3	640 mówców, 2728 rozmów	Dialekt amerykańskiego angielskiego, rozmowy przez telefon
1999/2000	Switchboard cellular p1	254 mówców, 1309 rozmów	Rozmowy przez telefon komórkowy GSM, rozmowy w amerykańskim angielskim
2000	Switchboard cellular p2	419 mówców, 2020 rozmów	Rozmowy przez telefon komórkowy w amerykańskim angielskim
2003	MIXER p1 oraz p2	600 mówców po 10 rozmów telefonicznych, 200 rozmów ze zmiennym kanałem transmisji	Rozmowy w języku różnym od angielskiego
2005	MIXER p3	1867 mówców z minimalnie 15 rozmowami przez telefon	Rozmowy zawierają wypowiedzi w 19 językach
2007	MIXER p4	200 mówców po 10 rozmów telefonicznych ze zmiennym kanałem transmisji	Rozmowy w języku angielskim
2007	MIXER p5	300 mówców przeprowadza 6 wywiadów oraz 10 rozmów przez telefon	Rozmowy oraz wypowiedzi czytane

Tab. 1. Najpopularniejsze korpusy wykorzystywane do oceny systemów identyfikacji mówców.³⁰

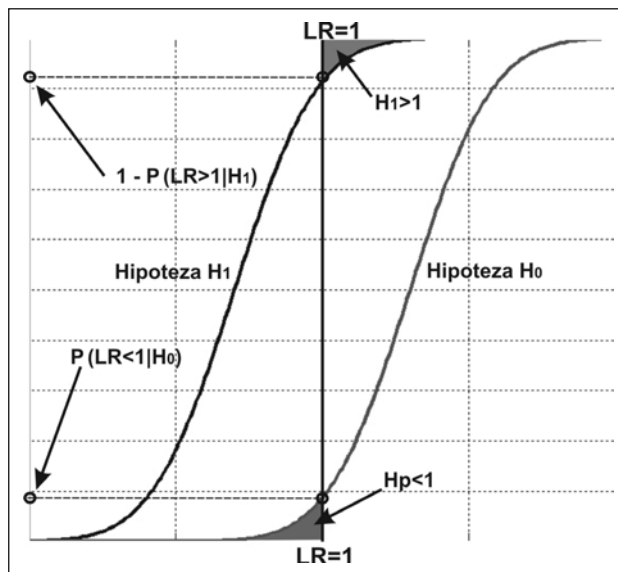
Obecnie na rynku nie ma korpusów, które by uwzględniały specyfikę kryminalistyczną. Przedstawione powyżej bazy były wykorzystywane w różnych okresach przez Amerykański Instytut Normalizacyjny (NIST) w badaniach międzylaboratoryjnych.

Za pomocą tych głosów tworzy się scenariusze porównań. Jeżeli ich liczność będzie odpowiednio duża, możliwe jest obliczenie rozkładów otrzymanych wyników dla poszczególnych scenariuszy. Poniżej przedstawiono abstrakcyjne rozkłady gęstości dla przypadku, gdy do określania stopnia różnicy głosów wykorzystany jest bayesowski iloraz wiarygodności. Charakterystyki te tworzy się poprzez porównywanie głosów w dwóch seriach: porównanie tego samego mówcy (przyjmijmy, że jest to hipoteza H_0) oraz porównanie różnych mówców (hipoteza H_1). Na podstawie otrzymanych wyników obliczana są rozkłady gęstości oraz dystrybuanty.

³⁰ Na podstawie danych Linguistic Data Consortium Corpus Catalog www ldc.upenn.edu



Wykres 1. Funkcja rozkładu gęstości prawdopodobieństwa poszczególnych wartości LR dla dwóch serii porównań H_0 – porównania tych samych mówców, H_1 – porównania różnych mówców.



Wykres 2. Przykładowa dystrybucja rozkładu gęstości prawdopodobieństwa poszczególnych wartości LR dla dwóch serii porównań H_0 – porównania tych samych mówców, H_1 – porównania różnych mówców.

Za pomocą powyższych krzywych wyznacza się prawdopodobieństwo popełnienia jednego z dwóch błędów:

- tego, że iloraz wiarygodności osiągnie wartość większą od 1, gdy prawdziwa jest hipoteza H_1 ,
- tego, że LR osiągnie wartość mniejszą od 1, gdy prawdziwa jest hipoteza H_0 .

Podstawowe źródła mające wpływ na niepewność dla metody językowej i pomiarowej są w zasadzie te same. Są to:

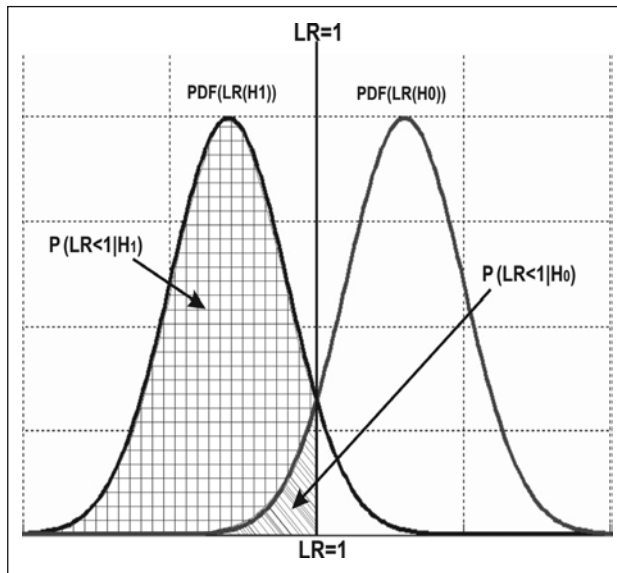
- język, którym mówcy się posługują,
- czas trwania i obszerność wypowiedzi,
- jakość nagrania i jakość wypowiedzi,
- kanał transmisji (szczególnie istotne w przypadku metod pomiarowych),
- stan emocjonalny mówcy,
- intoksykacja mówcy.

Badanie międzylaboratoryjne są podstawową metodą walidacji identyfikacji mówców metodą audytywną. Za pomocą takich testów możliwe jest oszacowanie odtwarzalności wyników oraz innych parametrów metody, takich jak wymagany czas trwania wypowiedzi, wpływ zakłóceń i inne. Statystyka badań może zostać opracowana w oparciu o formularz (rys. 1). Odpowiedni scenariusz testów pozwoli określić konkretne wymagania dotyczące możliwości realizacyjnych danego problemu dla podstawowych źródeł niepewności.

Metoda pomiarowa – techniczna – podczas badań międzylaboratoryjnych może zostać włączona jako element analizy audytywnej. W takim przypadku w całości metoda językowo-pomiarowa zostanie poddana uwierzytelnieniu. Jeżeli badania są prowadzone zgodnie z formularzem (rys. 4), dopuszcza on ocenę takich parametrów, jak wysokość głosu (reprezentowany przez częstotliwość tonu krtaniowego) oraz barwa głosu (reprezentowana przez częstotliwości formantowe), które mogą zostać określone pomiarowo. Warunkiem włączenia metody technicznej do testów międzylaboratoryjnych jest uzgodnienie tego ze wszystkimi laboratoriami.

Badania międzylaboratoryjne mają szczególny wymiar w przypadku systemów automatycznej identyfikacji mówcy. Jest to dziedzina, której rozwój od kilku lat ulega istotnemu przyspieszeniu. Najpopularniejsze testy porównań międzylaboratoryjnych (w rzeczywistości są to porównania efektywności działania systemów, którymi te laboratoria się posługują) są organizowane przez Narodowy Instytut Standaryzacji (NIST-SRE – ang. *National Institute of Standards and Technology Speaker Recognition Evaluations*). Udział w nich biorą przede wszystkim jednostki badawczo-rozwojowe. Pierwsze badania przeprowadzono w roku 1997.

Do prezentacji wyników porównań NIST wykorzystuje tzw. charakterystyki DET (ang. *Detection Error Tradeoff*)³¹, które pokazują zależność pomiędzy prawdopodobieństwami wystąpienia jednego z dwóch rodzajów błędów (błędnej akceptacji – FA, oraz błędnego odrzucenia - FR). Przykład charakterystyki DET przedstawiono poniżej.



Wykres 3. Przykładowa charakterystyka DET czterech systemów. System reprezentowany przez kolor zielony jest najbardziej efektywny. Czarnym okręgiem zaznaczono wartość DCF.

Interpretacja charakterystyki jest następująca: każda krzywa reprezentuje działanie jednego systemu. Im krzywa jest położona bliżej lewego dolnego rogu, tym wyniki zwracane przez system są obciążone mniejszym prawdopodobieństwem błędnej decyzji i tym samym ten system uważany jest za lepszy w danych warunkach pomiarowych. Oprócz charakterystyk DET wykorzystywany jest również współczynnik DCF (ang. *Decision Cost Function*), który stanowi liniową kombinację prawdopodobieństwa wystąpienia dwóch rodzajów błędów oraz uwzględnia tzw. koszty poszczególnych błędów³².

³¹ Por. A.J. Mansfield, J.L. Wayman, *Best Practices in Testing and Reporting Performance of Biometric Devices*, Teddington 2002; A. Martin, G.R. Doddington, T.Kamm, M. Ordowski, M.A. Przybocki, *The DET curve in assessment of detection task performance*, Proceedings of Eurospeech, Rhodes, Greece, 1997, 1895-1898.

³² A. Martin, M. Przybocki, *The NIST 1999 Speaker Recognition Evaluation – An Overview*, Digital Signal Processing 10, 1–18 (2000).

WALIDACJA METOD BADAWCZYCH Z ZAKRESU SPISANIA TREŚCI ROZMOWY³³

Celami badawczymi spisanie treści są:

- odtworzenie i spisanie treści,
- identyfikacja mówców w obrębie materiału badawczego.

Środkiem służącym do osiągnięcia tych celów jest niekiedy poprawa jakości nagrań poprzez usunięcie dźwięków maskujących (tzw. filtracja lub też korekcja nagrań). Nierzadko korekcja stanowi odrębny zakres badań wymieniony w postanowieniu.

Identyfikacja w obrębie materiału badawczego jest elementem wiążącym spisanie treści rozmowy z identyfikacją mówców z wykorzystaniem materiału porównawczego. Dlatego do tego elementu odnoszą się również uwagi zawarte w części dotyczącej walidacji metod identyfikacji mówców.

Ograniczeniami dla metody spisanie treści to:

- czas trwania i obszerność wypowiedzi,
- jakość nagrania i jakość wypowiedzi,
- kanał transmisji,
- stan emocjonalny mówcy,
- intoksykacja mówcy.

W warunkach polskich test biegłości badający poprawność spisywania treści rozmów powinien być opracowany z wykorzystaniem wypowiedzi polskojęzycznych, zarejestrowanych w sposób możliwie najbliższy realiom pracy ekspertów fonoskopii.

Nagranie testowe powinno zostać przygotowane z uwzględnieniem następujących wymogów:

1. rejestracji bezpośredniej,
2. cyfrowego utrwalenia i przetwarzania zapisu testowego,
3. dokonania nagrania testowego – niskiej jakości i jednocześnie nagrania referencyjnego – możliwie wysokiej jakości,
4. utrwalenia rozmowy kilku osób, które niekiedy mówią równocześnie,
5. zastosowanie zmiennego tempa wypowiedzi i wykorzystanie - w części rozmowy testowej – zjawisk utrudniających odsłuch, np. równoczesnego mówienia, przekrzykiwania się czy mówienia szeptem.

Ad. 1. W praktyce fonoskopijnej nagrania utrwalane bezpośrednio charakteryzują się zwykle najgorszą jakością. Czynnikiem spowalniającymi ich odsłuch są m.in. zakłócenia związane z kamuflażem, niski odstęp sygnał-zakłócenia, ograniczone pasmo sygnału mowy. Spisanie treści

³³ Ta część artykułu została oparta na informacjach opublikowanych wcześniej w t. XII Problemów Współczesnej Kryminalistyki, s. 297-302: J. Rzeszutowski, W. Witkowska-Pawlak, O. Komosa, *Test biegłości ekspertów fonoskopii spisujących treść rozmów*.

tych nagrań jest najbardziej czasochłonne, wymaga stosowania odpowiednich technik³⁴ i prowadzi do powstawania błędnych skojarzeń³⁵.

- Ad. 2.** Miniaturowe rejestratory cyfrowe coraz powszechniej są wykorzystywane do utrwalania rozmów dowodowych, czynnikiem, który często komplikuje prowadzenie badań nad takimi nagraniami, jest silna kompresja lub niskie parametry rejestracji (mała częstotliwość próbkowania, niska rozdzielczość)
- Ad.3.** Nagranie powinno zostać dokonane z jednoczesnym wykorzystaniem dwóch mikrofonów: jednego umieszczonego blisko źródła mowy i drugiego, który powinien być zakamuflowany³⁶. Nagranie drugie, niższej jakości, posłuży jako materiał testowy, podczas gdy to lepsze – do bezspornego ustalenia wzorca spisaną treść rozmowy.
- Ad. 4.** Nagranie rozmowy kilku osób (zamiast zainscenizowanego monologu) dodatkowo skomplikuje test, gdyż wprowadzi element identyfikacji mówców w obrębie materiału dowodowego. Ten etap badań fonoskopijnych jest powszechnie marginalizowany. Tymczasem błędy popełnione w trakcie przypisania poszczególnym osobom konkretnych wypowiedzi mogą zostać powielone w badaniach identyfikacyjnych z wykorzystaniem materiału porównawczego³⁷, co ma znaczenie w przypadku opinii kompleksowych.
- Ad. 5.** Test powinien zawierać zarówno wypowiedzi wyraźne i dobrze zrozumiałe (przy rejestracji bezpośredniej te walory posiadają wypowiedzi mówcy, który kamufluje mikrofon pod swoją odzieżą), jak i treści o niepełnej słyszalności (wypowiedzi mówców oddalonych od mikrofonu).

Test spełni zadanie weryfikacyjne, jeżeli eksperci będą spisywać nagranie około 5-minutowe. Organizator testu powinien określić limitu czasu na jego rozwiązanie.

Praktyka w zakresie badań fonoskopijnych pozwala przypuszczać, że w każdej z nadesłanych do organizatora testu wersji spisanej treści pojawią się rozbieżności

³⁴ Por. S. Błasikiewicz, *Metoda odsłuchu szeptu i mowy intensywnie zakłóconej*, Problemy Kryminalistyki nr 90/1971, s. 159-183.

³⁵ Zjawisko zostało opisane w tekście: J. Rzeszotarski, *Błędne skojarzenia jako element spisaną treść rozmów dowodowych*, Problemy Kryminalistyki nr 250/2005, s. 22-26.

³⁶ Tak dokonane nagranie nabiera waloru praktycznego poprzez wierne odtworzenie warunków dokonywania nagrań zakamuflowanym dyktafonem. Laboracyjne „pogarszanie” jakości dobrego nagrania zawsze niesie ryzyko niepełnego odzwierciedlenia specyfiki bezpośrednich nagrań dowodowych.

³⁷ Odpowiednią rangę identyfikacji w obrębie materiału dowodowego nadano w tekście: J. Rzeszotarski, W. Witkowska-Pawlak *Identyfikacja mówców w obrębie materiału dowodowego*, Problemy Kryminalistyki nr 246/2004, s. 15-18.

w stosunku do treści spisanego nagrania referencyjnego. Tak bowiem dzieje się w przypadku nagrań dowodowych - powstawanie rozbieżności odsłuchanego tekstu to nieuchronne zjawisko, towarzyszące spisaniu treści nagrań niskiej jakości. Te realia nakładają na organizatora programu obowiązek opracowania adekwatnego sposobu analizy wyników badań jakościowych. W praktyce kryminalistycznej i sądowej jak dotychczas nie jest znany skuteczny sposób ustalania jedynej wersji tekstu w sytuacji, gdy eksperci przedstawiają wersje rozbieżne. Obecnie sąd powołuje kolejnego biegłego (który przedstawia swoją wersję tekstu) albo nagranie jest odsłuchiwane w trakcie rozprawy (co przy sprzeczności niskiej jakości i warunkach akustycznych sal sądowych zazwyczaj nie przynosi spodziewanego rozstrzygnięcia).

Jednym z rozwiązań, które umożliwiłoby reprezentatywną analizę wyników testów, może być pytanie o udzielenie odpowiedzi, ile razy w treści przedmiotowego nagrania występują konkretne sylaby. Należy jednak zauważyć, że samo podanie wyniku badań, dopuszczalne w testach międzylaboratoryjnych, jest niezgodne z intencją wyrażoną w orzecznictwie. Sąd Najwyższy podkreśla bowiem, że biegli oprócz ostatecznej konkluzji obowiązani są wskazać drogę, jaką doszli do wniosków, oraz opisać metody i sposób przeprowadzonych badań³⁸.

WALIDACJA METOD BADANIA AUTENTYCZNOŚCI NAGRAŃ

Badanie autentyczności to proces składający się z kilku uzupełniających się metod badawczych. W przypadku nagrań analogowych mogą być to:

- analiza audytywna,
- analiza czasowej prezentacji sygnału w poszukiwaniu śladów ruchu głowic,
- analiza spektralna w celu określenia charakterystyki przenoszenia toru elektroakustycznego rejestratora,
- analiza spektrograficzna pozwala na analizę spektrogramów pod kątem obecności nieciągłości nagrania,
- analiza wizualizacyjna z wykorzystaniem specjalnej aparatury optycznej, pozwalającej na oglądanie z dużą rozdzielczością struktury taśmy ferromagnetycznej³⁹.

W przypadku nagrań cyfrowych wachlarz dostępnych metod jest węższy. Analiza audytywna oraz spektrograficzna i spektralna nie mogą być traktowane jako izolowane (należy łączyć te rodzaje analizy, choć i wyniki takiego połączenia w realiach cyfrowych nie zawsze są wiarygodne). Inne metody wymienione

³⁸ Por. wyrok SN z 13.10.1998 roku w sprawie o sygn. akt II KKN 225/96, Prokuratura i Prawo, 1999/3/13.

³⁹ A.J. Cooper, *Detection of copies of digital audio recordings for forensic purposes*, The Open University.

wyżej nie mają zastosowania. Obecnie wydaje się, że najskuteczniejszym narzędziem badania autentyczności nagrań cyfrowych jest analiza przydźwięku sieciowego.

Analiza z wykorzystaniem powyższych metod pozwala na sformułowanie problemów badawczych w odniesieniu do nagrań cyfrowych i analogowych. Wspólnym celem podstawowym jest ustalenie, czy nagranie jest autentyczne. Różnice dotyczą celów pośrednich. Dla nagrań analogowych tymi celami pośrednimi są:

- ustalenie, czy jest oryginalne,
- ustalenie, czy nagranie zostało zarejestrowane przy użyciu jednego urządzenia.

W przypadku nagrań cyfrowych cele pośrednie to dodatkowo:

- ustalenie daty i czasu rejestracji,
- ustalenie, czy nagranie mogło zostać zarejestrowane za pomocą tego typu urządzenia, jak to przekazane do badań.

Każda z metod badania autentyczności nagrań powinna zostać poddana procesowi walidacji. Grupa robocza Kryminalistycznej Analizy Mowy i Nagrań (FSAAWG) działająca w ramach ENFSI⁴⁰, jak dotychczas stworzyła wytyczne tylko dla najlepszej praktyki laboratoryjnej badania autentyczności metodą analizy przydźwięku sieciowego. Dokument ten jest zbieżny z innymi poradnikami⁴¹ dotyczącymi m.in. walidacji i może zostać przyjęty jako wiążący dla innych metod badania autentyczności. Główne punkty to:

- identyfikacja czynników mających wpływ na proces badawczy,
- sprawdzenie, czy metody obliczeń, urządzenia oraz oprogramowanie wykorzystywane w badaniach jest odpowiednie,
- metoda badawcza jest całkowicie udokumentowana przez laboratorium,
- otrzymanywane rezultaty przez laboratorium są wiarygodne oraz odtwarzalne, wszystkie czynniki mające wpływ na wynik każdorazowo należy zidentyfikować i udokumentować,
- technika ekstrakcji oraz procedura stosowana przez laboratorium zostały ocenione przez niezależnych specjalistów,
- każdy członek zespołu laboratoryjnego prowadzący badania tą metodą powinien potwierdzić swoje kompetencje.

Źródłami mającymi wpływ na niepewność badań autentyczności zapisu są:

- czas trwania i obszerność wypowiedzi (istotna dla audytywnych badań sensu logicznego utrwalonych wątków rozmowy),

⁴⁰ Forensic Speech and Audio Analysis Working Group ENFSI *Best practice guidelines for ENF analysis in forensic authentication of digital evidence*, wyd. 2.06.2009.

⁴¹ Np. *Best practices for forensic audio (Vresja 1.0)*, Scientific Working Group on Digital Evidence (SWGDE), styczeń 2008 lub *Guidelines for forensic science laboratories*, ILAC-G19:2002, Australia.

- jakość nagrania (zbyt niska jakość zapisu może utrudniać interpretację występujących efektów dźwiękowych i ustalenie przyczyn ich wgrania),
- konwersja analogowo-cyfrowa.

Według FSAAWG ENFSI, jeżeli istnieje laboratorium, które przeprowadziło już walidację tej metody, laboratorium starające się o wykonanie walidacji powinno wykazać, że otrzymuje zbieżne wyniki.

Testy międzylaboratoryjne, o których mowa również w tekście Najlepszej Praktyki Laboratoryjnej wg FSAAWG, ze względu na jakościowy charakter badań autentyczności, wydają się jedyną techniką uwierzytelnienia tych metod.

Pierwszym i zasadniczym problemem planowania i realizacji testów międzylaboratoryjnych z zakresu badania autentyczności nagrań jest przygotowanie materiału testowego. Nagrania powinny zostać tak opracowane, aby z jednej strony uprawnione było sformułowanie możliwie wszystkich problemów badawczych zdarzających się w praktyce (patrz powyżej), z drugiej natomiast możliwa była analiza statystyczna otrzymywanych wyników. Materiał powinien zatem składać się z kilkudziesięciu odrębnych plików dźwiękowych. Wśród nich oprócz nagrań autentycznych powinny znaleźć się nagrania, w których autentyczność dokonano ingerencji.

W praktyce zdarzają się następujące techniki naruszenia autentyczności nagrania:

- a) dla nagrań cyfrowych:
 - chwilowe zatrzymanie rejestracji,
 - dokonanie montażu z wykorzystaniem komputerowego edytora dźwięku,
- b) dla nagrań analogowych:
 - wstrzymanie rejestracji,
 - cyfryzacja nagrania, a następnie dokonanie montażu z wykorzystaniem komputerowego edytora dźwięku i powtórne przegranie na ten sam lub inny nośnik.

Możliwości montażu z wykorzystaniem komputerowego edytora dźwięku jest wiele, np. wykasowanie, zamiana kolejności zdarzeń, zaszumienie nagrania w taki sposób, aby zamaskować jego fragmenty itp.

W przypadku badania autentyczności nagrań proces badawczy rozpoczyna się w momencie kopiowania nagrań na twardy dysk (z wyjątkiem badań z wykorzystaniem wizualizatora stanu namagnesowania warstwy aktywnej taśm magnetofonowych oraz nagrań cyfrowych w postaci plików dźwiękowych). W przypadku nagrań analogowych etap ten może okazać się kluczowy. Dlatego konieczne jest wykonanie odpowiedniej ilości identycznych kopii nagrań na rzeczywistych nośnikach. Ich dostateczna jednorodność pozwoliłaby uniknąć podejrzania, że wyniki uznane za ekstremalne⁴² wynikają z niejednorodności materiału badawczego.

⁴² Wynik ekstremalny to wartość znacznie odstająca od pozostałych wartości zbioru danych.

Odrębnym zagadnieniem od formy materiału testowego jest opracowanie statystyczne wyników. W typowych testach międzylaboratoryjnych analizy ilościowej za wartość referencyjną, do której odnosi się wyniki pomiarów uczestników, przyjmuje się tzw. wartość przypisaną⁴³, która zwykle stanowi średnią ze wszystkich wyników poszczególnych laboratoriów. W przypadku badań autentyczności nagrań, pomimo że wykorzystywane są oprócz metody audytywnej również metody instrumentalne (z wykorzystaniem urządzeń np: spektrografia, przydźwięk, zjawisko Faradaya), prowadzą one zwykle do wniosków jakościowych (autentyczne, nieautentyczne, nie można stwierdzić). W każdym razie autorom nie są znane metody szacowania niepewności otrzymywanych wyników badań autentyczności nagrań (wyjątkiem może okazać się np. współczynnik korelacji lub inna miara podobieństw przebiegów przydźwięku sieciowego nagrań dowodowego i referencyjnego, jednakże obecnie brak jest naukowych publikacji w tym zakresie). Wobec tego analiza porównawcza mogłaby się oprzeć na trafności i odtwarzalności wskazywanych miejsc ingerencji przez różne laboratoria.

ZAKOŃCZENIE

Walidacja metod badawczych w fonoskopii nie jest celem samym w sobie. Rada Unii Europejskiej – dostrzegając potrzebę ochrony danych osobowych obywateli krajów Unii oraz w trosce o wysoką jakość usług kryminalistycznych na potrzeby procesu karnego - przyjęła *Decyzję w sprawie akredytacji dostawców usług kryminalistycznych wykonujących czynności laboratoryjne*⁴⁴, na mocy której standaryzacji podlegać będą badania profili DNA i daktyloskopijne. Zapewne podobne wytyczne zostaną niebawem przekazane krajom członkowskim w odniesieniu do kolejnych dziedzin kryminalistyki, w tym do fonoskopii. Celem powyższych rozważań jest pokazanie, że eksperci fonoskopii starają się przygotować na ten moment, tak aby płynnie wejść w etap standaryzacji badań w każdym z państw członkowskich Unii Europejskiej, zgodnie z zalecaną normą ISO 17025⁴⁵.

⁴³ Wartość przypisana – wartość przypisana wielkości określonej i uznana, niekiedy umownie, jako wartość wyznaczona z niepewnością akceptowalną w danym zastosowaniu.

⁴⁴ Decyzja Ramowa Rady 2009/905/WSiSW z dnia 30 listopada 2009 r.

⁴⁵ Ibidem. Termin wprowadzenia zaleceń jest stosunkowo odległy: dla badań profili DNA – do 30.11.2013, dla badań daktyloskopijnych – do 30.11.2015.