

Pakiet programowy SIVE jako narzędzie wspomagające kryminalistyczne badania fonoskopijne z perspektywy praktyki eksperckiej. Część I

**Pomiar wartości podstawowych
fizycznych parametrów głosu
w badaniach identyfikacyjnych osób**

Zakres badań fonoskopijnych

Fonoskopia jest stosunkowo młodą dziedziną kryminalistyki i nie we wszystkich krajach rozwijała się w tym samym czasie, tempie i w ten sam sposób. Wynika to przede wszystkim z odmiennych tradycji – determinowanych przez np. ustrój danego kraju – które z kolei mają bezpośredni związek z obowiązującymi w określonym państwie wcześniej i obecnie uwarunkowaniami prawnymi i proceduralnymi. Można przyjąć, że badania fonoskopijne najwcześniej zafunkcjonowały jako badania kryminalistyczne w państwach demokracji ludowej, w tym w Polsce, w której system prawny dopuścił wykorzystanie w postępowaniu sądowym nagrań magnetofonowych w charakterze dowodów już w roku 1960. Równocześnie zostały określone warunki, aktualne do dziś, pod jakimi nagrania takie można było za dowodowe uznać. Należą do nich:

- ustalenie treści wypowiedzi zarejestrowanych w tym nagraniu,
- potwierdzenie, że przedstawione nagranie jest wiernym odzwierciedleniem zarejestrowanej w nim rzeczywistości,
- ustalenie tożsamości osób, których wypowiedzi są utrwalone w nagraniu przedstawionym sądowi.

Określenie tych warunków wymusiło powstanie nowej, samodzielnej dziedziny kryminalistyki – nazwanej fonoskopią – w której wyodrębniły się trzy podstawowe rodzaje badań, tj.:

- badania odsłuchowe, mające na celu spisanie treści dowodowych wypowiedzi,
- badania autentyczności nagrań, mające na celu ustalenie ich ciągłości i wykluczenie jakiegokolwiek ingerencji w nie, w szczególności montażu,
- badania identyfikacyjne osób.

Stosowane metodyki

Badania identyfikacyjne osób przeprowadzane są na podstawie analizy mowy i/lub głosu. Metodyki tych badań, stosowane w poszczególnych krajach, są

obecnie dosyć zróżnicowane. Najbardziej kompleksowymi i pozwalającymi na przeprowadzenie analiz w najszerszym zakresie – a co za tym idzie, często na wydawanie opinii kategoriycznych – są metodyki opierające się na szczegółowych badaniach cech lingwistycznych, fonetycznych i akustycznych, uzupełnianych pomiarami np. podstawowych fizycznych parametrów sygnału mowy i/lub analizą struktur formantowych porównywanych wypowiedzi. Od kilkudziesięciu lat – z wieloma sukcesami – metodyka taka stosowana jest również przez ekspertów fonoskopii CLK KGP w Warszawie. Jest ona jednak bardzo czasochłonna. W znacznej mierze opiera się na wieloaspektowej audytywnej analizie nagrań oraz audytywnym ekstrahowaniu cech w poszczególnych ich grupach, w związku z czym wymaga od eksperta rozległej wiedzy i dużego doświadczenia. Siłą rzeczy długi jest okres kształcenia samodzielnego eksperta, wykorzystującego tego typu metodykę. Równocześnie na przeciwnym biegunie znajdują się metodyki przeprowadzania fonoskopijnych badań identyfikacyjnych osób w sposób w pełni zautomatyzowany, wymagające od operatorów opracowanych systemów jedynie krótkiego przeszkolenia w zakresie wprowadzania danych. Takie metodyki pozwalają na wydawanie opinii niekategoriycznych, często niskiego stopnia prawdopodobieństwa. Ponadto, w dobie znacznego wzrostu liczby osób poddawanych fonoskopijnym badaniom identyfikacyjnym, żadna z przedstawionych powyżej metod nie jest w stanie sprostać zmieniającej się rzeczywistości. Środowiska ekspertów fonoskopii różnych krajów podjęły zatem starania na rzecz bądź opracowania nowych metod badawczych, bądź stworzenia nowych narzędzi pracy. Najszybciej pojawił się efekt w postaci specjalistycznych programów komputerowych, służących analizom głosu – jako zjawiska akustycznego – oraz szerokim analizom językowym. Powstałe w tym kontekście pakiety programowe są w różnym stopniu rozbudowane i zazwyczaj obciążone wadą przywiązania do języka kraju, w którym je stworzono i wdrożono. Przykładami takich pakietów są systemy: Dialect – stosowany w pracowniach fonoskopii MSW Federacji Rosyjskiej, Phonograph – stosowany na Ukrainie lub Idem – stosowany we Włoszech. Za ich zaletę uznać należy, że – poza funkcją

podstawową – konstrukcja tych systemów uwzględnia także potrzebę korzystania z nich w celu szybkiej weryfikacji wyników badań identyfikacyjnych uzyskiwanych przez ekspertów ośrodków terenowych podlegających stałemu nadzorowi merytorycznemu, sprawowanemu przez laboratoria centralne. Jednak stosunkowo szybko praktyka wykazała, że konieczność znacznego rozbudowania części fonetycznej i lingwistycznej tego typu systemów nie pozwala sprostać kryteriom stawianym programom uniwersalnym. Ich zastosowanie wymaga bowiem tworzenia dla poszczególnych krajów odrębnych i obszernych baz leksykalnych (rzędu 100–200 tysięcy słów), uwzględniających wzorce ortofoniczne oraz wariantowość dialektalną i regionalną danych języków. Choć teoretycznie jest to możliwe, to w praktyce wymaga znacznych nakładów finansowych oraz kilkuletniej pracy licznej grupy doświadczonych lingwistów, szczególnie zainteresowanych tego typu problematyką. Należy jednak zdawać sobie sprawę, że przydatność takich baz do zastosowań procesowych uwarunkowana jest nie tylko ich konstrukcją – obiektywną i ściśle odpowiadającą realiom – ale także rzetelnością i stałym ich aktualizowaniem. W czasach wzmożonej migracji ludności, mieszania się kultur, intensywnego oddziaływania mass mediów oraz interferencji wielu kultur i języków, sprostanie tak ostrym wymogom staje się praktycznie niemożliwe. Natomiast opieranie wyników fonoskopijnych badań identyfikacyjnych osób na metodzie wykorzystującej bazy nie w pełni odpowiadające wspomnianym wyżej kryteriom, grozi nie tylko niską efektywnością samych badań, ale przede wszystkim wysokim ryzykiem popełnienia błędu.

Pakiet programowy SIVE

W kontekście tego uzasadnione zainteresowanie wzbudzają propozycje wykorzystujące analizę innych zjawisk występujących w mowie, niewymagających gromadzenia obszernych baz wzorców odniesienia. Optymalne wydaje się zastąpienie żmudnej pracy eksperta obliczeniami komputerowymi wszędzie tam, gdzie jest to możliwe bez szkody merytorycznej, czyli stworzenie informatycznych systemów półautomatycznych. Jedną z najlepiej poznanych tego typu propozycji – sprawdzonych i zoptymalizowanych – stanowi pakiet programowy SIVE, opracowany przez dr. Bernardasą Šalnę, doświadczonego eksperta fonoskopii, od wielu lat naczelnika pracowni fonoskopii litewskiego Instytutu Ekspertyz Sądowych w Wilnie. System ten obecnie z powodzeniem stosowany jest już w siedmiu krajach europejskich, a w niniejszym artykule przedstawiona jest jego nowa wersja. Wyróżnia się prostotą obsługi oraz brakiem konieczności dosto-

sowywania go do specyfiki języka, którym posługują się identyfikowane osoby i – jak dotąd – jest jednym z nielicznych systemów wskazujących stopień zbieżności wartości poszczególnych porównywanych parametrów. Jeden z modułów systemu SIVE służy obliczaniu wartości kilku fizycznych parametrów głosu oraz statystycznej obróbce uzyskanych wyników i ustaleniu wartości korelacji zestawianych ze sobą wielkości. Fizyczne parametry głosu, których wartości w prezentowanym systemie podlegają pomiarom, należą do najbardziej podstawowych i są to: częstotliwość podstawowa tonu krztaniowego (tzw. F_0), rozkład harmonicznego tonu krztaniowego, dźwięczność i barwa głosu. Wartości tych parametrów (oraz zakres ich zmienności osobniczej) są charakterystyczne dla danego mówcy, gdyż zależą od anatomicznej budowy i sprawności jego narządu mowy. Fakt ten wykorzystywany jest nie tylko w kryminalistycznych badaniach identyfikacyjnych osób. Każdy człowiek w swoim codziennym życiu indywidualności tych parametrów zawdzięcza możliwość audytywnego rozpoznawania swoich rozmówców. Zazwyczaj bowiem wyodrębniamy słuchem obecność takich cech głosów, jak ich barwa czy wysokość.

Wcześniej, na użytek badań identyfikacyjnych osób, pomiary wartości tylko jednego z tych parametrów – F_0 – przeprowadzane były metodą ręczną, wykorzystującą obrazy sonograficzne, pozwalające obliczyć składowe harmoniczne sygnału mowy badanych mówców. Istotnym więc mankamentem tej metody była jej wyjątkowa czasochłonność, w praktyce na ogół niepozwalająca na pełne przeanalizowanie dłuższych materiałów, a mająca znaczący wpływ na koszty wydawanych opinii. Dlatego stworzenie programowego i wiarygodnego narzędzia umożliwiającego dokonanie pomiaru nie tylko F_0 , ale również kilku innych ważnych fizycznych parametrów głosu oraz statystyczne porównanie wyników i określenie stopnia ich zbieżności, w sposób wyraźnie odczuwalny skraca procesy analiz głosu i pozwala na szybkie wnioskowanie na podstawie otrzymanych wyników. Prezentowany system jest więc wygodnym narzędziem wspomagającym pracę ekspertów fonoskopii.

Składa się z on trzech głównych bloków programowych analizujących sygnał mowy z wykorzystaniem różnych procedur. Dotychczas – jego wcześniejsze wersje – przedstawione zostały między innymi podczas międzynarodowego seminarium w Nałęczowie, w październiku 2002 roku oraz w roku 2004 na międzynarodowym zjeździe ekspertów fonoskopii w Ankarze, odbywającym się w ramach prac grupy roboczej ENFSI, zrzeszającej pracowni fonoskopii laboratoriów kryminalistycznych resortów spraw wewnętrznych i sprawiedliwości krajów członkowskich.

Moduł służący do analiz podstawowych fizycznych parametrów głosu

Na pierwszy z trzech bloków systemu składają się dwa programy omawiane w niniejszym artykule, które początkowo funkcjonowały samodzielnie (wymagając wstępnej obróbki sygnałów, dokonywanej przez inne programy), a które dopiero po latach modyfikacji weszły w skład szerszego pakietu programowego, jakim jest SIVE. W aktualnej wersji są to:

- program mierzący wartości kilku podstawowych fizycznych parametrów głosu mówców, w tym m.in. częstotliwość podstawową tonu krtaniowego F_0 , tj. program o nazwie Pitch,
- tzw. program statystyczny, obliczający m.in. stopień zbieżności wartości poszczególnych parametrów zmierzonych z użyciem poprzedniego programu, używany w różnych wariantach porównawczych, tj. program o nazwie Pitchcom.

Przykładowe wydruki ilustrują przejrzystość graficzno-numerycznego sposobu prezentowania otrzymywanych wyników.

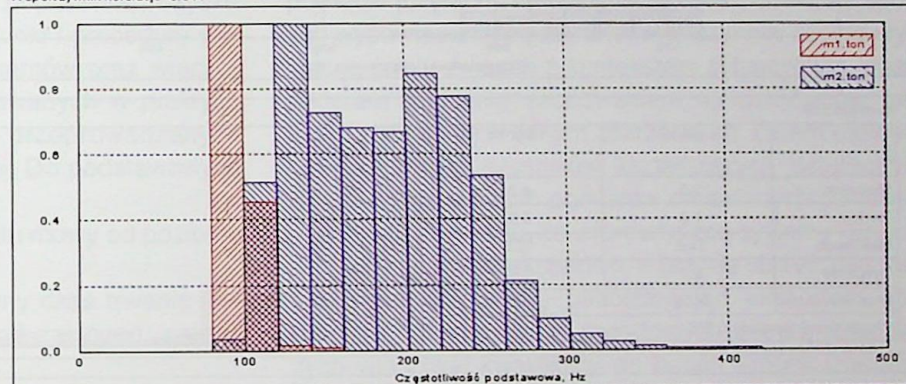
Powstanie obu programów jest wynikiem około pięcioletniej pracy – teoretycznej i praktycznej – grupy specjalistów Instytutu Ekspertyz Sądowych w Wilnie oraz Instytutu Matematyki i Informatyki Akademii Nauk Litwy. Programy te wykorzystują metodę spektralno-korelacyjną, uznającą w drodze doświadczałnej za optymalną dla obliczania parametru F_0 . Na uwagę zasługuje fakt, że obecnie istnieje co najmniej kilkadziesiąt programów umożliwiających pomiar częstotliwości podstawowej tonu krtaniowego F_0 , jednak dokładność uzyskiwanych z ich użyciem wyników jest nie do zaakceptowania w kryminalistycznych badaniach identyfikacyjnych osób, w których materiałem badaw-

czym najczęściej są nagrania stosunkowo niskiej jakości. Podstawowym tego powodem jest fakt, iż na ogół programy te pierwotnie powstają jako narzędzia dla innego typu badań, stawiając analizowanym materiałom m.in. warunek co najmniej 30 dB odstępu sygnału mowy od poziomu towarzyszących mu zakłóceń. W przypadku np. zastosowań medycznych na ogół nie ma trudności ze spełnieniem tak określonego warunku, co równocześnie gwarantuje stosunkowo wysoką dokład-

Częstotliwość podstawowa

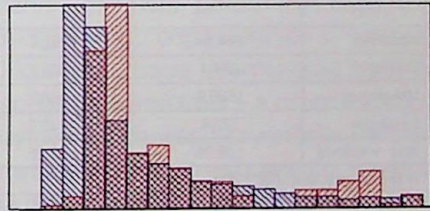
	Minimum	Średnia	Maksimum	Dyspersja	Wsp. wariancji
m1.ton	84.0 Hz	98.2 Hz	121.9 Hz	9.9 Hz	10.13 %
m2.ton	107.9 Hz	180.3 Hz	207.8 Hz	51.3 Hz	27.52 %

Współczynnik korelacji: -0.044



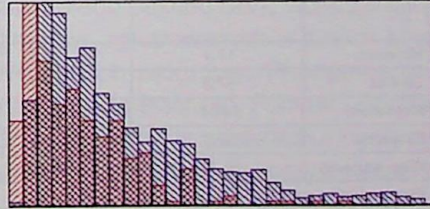
Harmoniczne

	m1.ton	m2.ton
Minimum	3.5	2.0
Średnia	11.5	6.6
Maksimum	28.2	20.7
Dyspersja	8.2	4.9
Wsp. wariancji	71.69	73.40
Współczynnik korelacji	0.521	



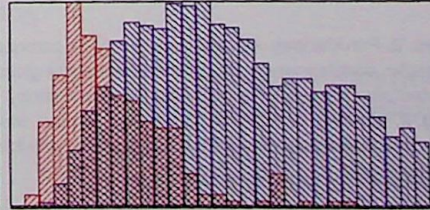
Dźwięczność głosu

	m1.ton	m2.ton
Minimum	40.6	74.0
Średnia	251.8	453.5
Maksimum	838.1	1736.6
Dyspersja	202.1	408.5
Wsp. wariancji	80.26	90.08
Współczynnik korelacji	0.780	



Barwa głosu

	m1.ton	m2.ton
Minimum	11.1	25.5
Średnia	37.0	102.9
Maksimum	96.3	263.5
Dyspersja	19.8	62.3
Wsp. wariancji	53.46	60.53
Współczynnik korelacji	-0.003	



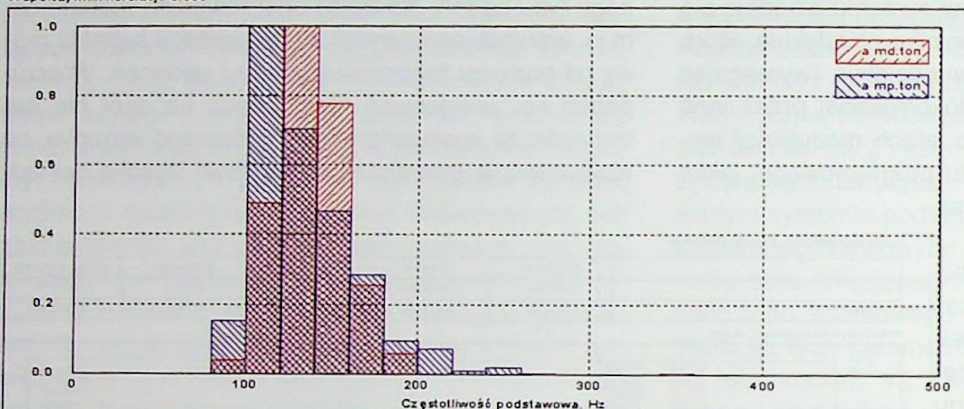
Ryc. 1. Przykładowy wydruk, uzyskiwany za pomocą programu Pitchcom, zestawiający wyniki porównania wartości kilku podstawowych fizycznych parametrów głosu z dwóch próbek, świadczące o niskim stopniu prawdopodobieństwa, że próbki te pochodzą od jednej osoby.

Fig. 1. Example of Pitchcom generated printout, which collates results of comparison of values of several basic physical voice parameters from two samples providing for low probability of these samples originating from the same person

Częstotliwość podstawowa

	Minimum	Średnia	Maksimum	Dyspersja	Wsp. wariancji
a md.ton	103.4 Hz	137.9 Hz	181.1 Hz	19.1 Hz	13.87 %
a mp.ton	98.8 Hz	133.9 Hz	216.3 Hz	30.1 Hz	22.45 %

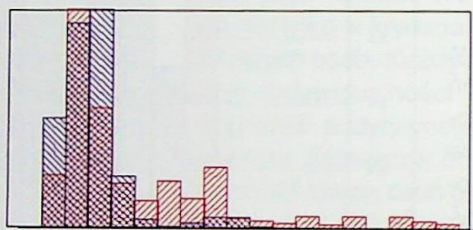
Współczynnik korelacji: 0.855



Harmoniczne

	a md.ton	a mp.ton
Minimum	2.0	1.9
Średnia	7.3	3.9
Maksimum	24.5	10.7
Dyspersja	6.1	2.0
Wsp. wariancji	84.50	51.41

Współczynnik korelacji: 0.889



Dźwięczność głosu

	a md.ton	a mp.ton
Minimum	44.8	50.4
Średnia	274.4	222.1
Maksimum	1598.5	566.7
Dyspersja	373.4	140.0
Wsp. wariancji	136.08	63.05

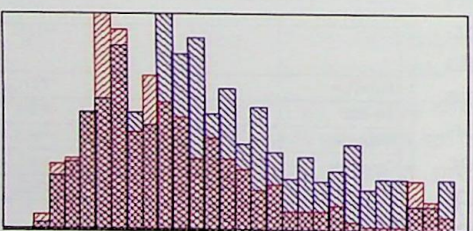
Współczynnik korelacji: 0.909



Barwa głosu

	a md.ton	a mp.ton
Minimum	17.4	20.2
Średnia	57.3	80.6
Maksimum	146.4	206.5
Dyspersja	32.3	47.6
Wsp. wariancji	56.47	59.00

Współczynnik korelacji: 0.737



Średni współczynnik korelacji: 0.850

Ryc. 2. Przykładowy wydruk, uzyskiwany za pomocą programu Pitchcom, zestawiający wyniki porównania wartości kilku podstawowych fizycznych parametrów głosu z dwóch próbek, świadczące o wysokim stopniu prawdopodobieństwa, że próbki te pochodzą od jednej osoby.

Fig. 2. Example of Pitchcom generated printout, which collates results of comparison of values of several basic physical voice parameters from two samples providing for high probability of these samples originating from the same person

ność dokonywanych pomiarów. Jednak materiały dźwiękowe najczęściej spotykane w praktyce kryminalistycznej w większości charakteryzują się znacznie niższymi parametrami. Ponadczterdziestoletnie doświad-

Bardzo często też dominującym sposobem weryfikacji wiarygodności tworzonych systemów jest testowanie ich głównie na nagraniach modelowych, siłą rzeczy znacząco odbiegających swoją specyfiką od zapisów

czenia ekspertów Pracowni Fonoskopii CLK KGP w Warszawie wykazują, że spośród wszystkich nadsyłanych tu do badań dowodowych nagrań, w ok. 80 procentach badany sygnał mowy przewyższa poziom zakłóceń o 10–15 dB. Fakt ten zmusza więc do korzystania z oprogramowania umożliwiającego poprawne przeprowadzanie żądanych pomiarów i obliczeń w odniesieniu do takich materiałów dźwiękowych. Kryterium 30-decybelowego odstępu sygnału mowy od poziomu zakłóceń spełnia zaledwie ok. 15 procent. Z tego względu oba omawiane tu programy są przystosowane do pracy z materiałami średniej i niskiej jakości, co stanowi jedną z ich głównych zalet.

Warto również zaznaczyć, iż pełne wdrożenie tego systemu poprzedził długi – bo blisko dziesięcioletni – okres jego weryfikacji w warunkach praktyki kryminalistycznej, przeprowadzanej m.in. w CLK KGP. Fakt ten wymaga podkreślenia choćby z tego powodu, że mimo iż wydaje się, że długotrwała procedura potwierdzenia wiarygodności powinna być standardowym wymogiem wobec metodyk i narzędzi stosowanych w codziennej pracy podmiotów wydających opinie na potrzeby postępowań procesowych, to jednak zdarzają się przypadki podjęcia w tym charakterze zaledwie kilku do-
rażnych eksperymentów.

poddawanych badaniom w laboratoriach kryminalistycznych.

Wśród ośrodków zagranicznych weryfikujących przydatność programów Pitch computing i Pitch comparing (nazwy pierwotne) jednym z pierwszych było CLK KGP w Warszawie, gdzie eksperci fonoskopii – od roku 1995 – testowali je, nie tylko przeprowadzając eksperymenty z udziałem własnym i osób im znanych, ale przede wszystkim wykorzystując materiały pochodzące ze spraw na bieżąco nadsyłanych do badań. Po przeprowadzeniu około 1000 testów, zakończonych wynikami o wysokim stopniu wiarygodności, a także po ustaleniu kryteriów, jakie powinny spełniać badane nagrania, gwarantujących poprawność uzyskiwanych efektów, opracowano warunki i procedury wykorzystania przedmiotowych programów oraz włączono je do rutynowych metod stosowanych w praktyce badań identyfikacyjnych osób, przeprowadzanych na podstawie analizy mowy ciągłej. Do podstawowych kryteriów zalicza się:

- minimum 10 dB odstęp sygnału mowy od poziomu zakłóceń,
- minimum 20 sekundowy łączny czas trwania wyselekcjonowanych, wolnych od zakłóceń, próbek mowy,
- zbliżony stan emocjonalny porównywanych próbek mowy,
- analogiczny charakter toru fonicznego.

Wnioski

Podsumowując, stwierdzić należy, że ponaddziesięcioletnie doświadczenie ekspertów CLK KGP w pracy z tymi programami pozwala im uznać je za narzędzie ważne, użyteczne i wygodne. Jednak przekonanie, że wyniki otrzymywane z jego zastosowaniem – szczególnie liczbowe – mogą stanowić jedyną, a jeśli nie, to co najmniej decydującą, podstawę dla formułowanych wniosków końcowych jest z gruntu błędne, chociaż bardzo rozpowszechnione wśród osób niebędących zawodowymi, doświadczonymi ekspertami fonoskopii oraz wśród przedstawicieli organów procesowych. Należy przy tym wyraźnie podkreślić, iż brak możliwości takiego wykorzystania uzyskiwanych wyników nie jest skutkiem zastosowania w omawianych programach niewłaściwych algorytmów, lecz wynika ze specyfiki i charakteru samych – podlegających badaniom – fizycznych parametrów głosu, jak też z ich podatności na fałszujące działanie czynników zewnętrznych, takich jak np. odgłosy tła wypowiedzi badanych, równoczesne wypowiedzi innych rozmówców itp. I choć – o czym była już mowa wcześniej – fizyczne parametry głosu, które analizuje się podczas fonoskopijskich badań identyfikacyjnych osób, są cechami indywidualnymi mówców, gdyż

ich wartości (a także zakres zmienności osobniczej) są uwarunkowane anatomiczną budową narządu mowy każdego człowieka, to jednak właściwy dobór próbek badawczych oraz stosowna do ich specyfiki interpretacja wyników uzyskiwanych w pomiarach i obliczeniach dokonywanych z użyciem opisywanych tutaj programów, nie może być oderwana od wyników analiz cech mowy (lingwistycznych, fonetycznych, akustycznych). Dość powiedzieć, że np. znaczna osobnicza zmienność wartości fizycznych parametrów mowy zależna jest choćby od takich czynników, jak: stan emocjonalny mówców, stopień zmęczenia głosu, krótkotrwałe stany chorobowe wypowiadających się osób, charakter wypowiedzi (czytanie, monolog, rozmowa), chwilowy sposób wypowiadania się itd. Z kolei te czynniki często wskazują ścisły związek z kontekstem sytuacyjnym, charakterem zdarzenia, zachowaniem się rozmówców, roli mówcy pełnionej w danym zdarzeniu itp. Zatem wielość i różnorodność czynników kształtujących ostateczne wyniki dokonywanych pomiarów omawianych parametrów głosu powodują, że informatyczne systemy nie powinny zastępować ekspertów, a jedynie służyć jako narzędzia obliczeniowe, pozostawiając ekspertom konieczność dokonywania świadomych decyzji dotyczących właściwej kwalifikacji do badań próbek izolowanych z materiału dowodowego, uzyskiwania odpowiedniego merytorycznie materiału porównawczego oraz poprawnej ekstrakcji z niego odpowiednich próbek porównawczych. Także prawidłowa interpretacja wyników pomiarów wartości tych parametrów – a przede wszystkim wartości otrzymywanych współczynników korelacji wzajemnej zestawianych próbek wypowiedzi – wymaga dużej wiedzy i doświadczenia eksperta.

Poza zaprezentowanymi powyżej programami Pitch i Pitchcom, w skład całego pakietu SIVE wchodzi także kilka innych programów, zgrupowanych w dwóch kolejnych blokach. Wśród nich programy: Phonemos (nazwa wersji wcześniejszej to Phonos), Fonvek i Fondist, stanowią narzędzia przydatne w analizach innych zjawisk towarzyszących mowie, które wykorzystywane są w dalszych etapach fonoskopijskich badań identyfikacyjnych osób. Ich omówienie przewidziano w ramach odrębnej publikacji.

Alicja Malanowicz
Aleksander Kowalczyk
 Ryciny – wydruk uzyskany
 za pomocą programu Pitchcom
 – autorzy

Autorzy dziękują doktorowi Bernardasowi Šalnie za współpracę przy opracowaniu niniejszego tekstu.