

Hermann J. Künzel

Kryminalistyczne badanie mowy

„Kriminalistik” 2003, nr 11, s. 676–684

W ostatnich trzydziestu latach kryminalistyczne badanie mowy przekształciło się w nową dziedzinę nauk kryminalistycznych. Na skalę międzynarodową pisze sesja ENFSI na ten temat, która odbyła się w 1997 r. Niniejszy artykuł stanowi przegląd podstawowych zagadnień i prowadzonych prac naukowo-badawczych w dziedzinie kryminalistycznego badania mowy oraz przedstawia możliwości i ich granice z uwzględnieniem potrzeb użytkowników, czyli policji i wymiaru sprawiedliwości.

W celu przedstawienia w sposób najbardziej zbliżony do warunków rzeczywistych najważniejszych elementów kryminalistycznego badania mowy autor posłużył się przykładem z przypadku dowodowego.

Przypadek dowodowy, część 1

Siedmioletni chłopczyk, mieszkający w jednej z zamożniejszych dzielnic Kolonii, został uprowadzony w drodze do szkoły. Gdy jego matka około godz. 14 wróciła do domu z pracy, otrzymała telefon od nieznanego mężczyzny (dalej nazywanego anonimem). Przekazany tekst brzmiał następująco: „Przejęliśmy opiekę nad waszym chłopcem na drodze do szkoły. Czuje się on dobrze. Jeżeli spełnicie nasze żądania, w ciągu kilku dni wasz syn wróci do domu. Przygotujcie 500 000 euro i czekajcie na dalsze wskazówki. Nie wciągajcie w to policji, bo więcej go nie zobaczycie. Wkrótce się odezwie-

my”. Matka chłopca była tak wstrząśnięta tym co usłyszała, że nie pomyślała o zarejestrowaniu tej rozmowy. Natychmiast powiadomiła męża, który jest właścicielem średniej wielkości przedsiębiorstwa. Już około godz. 18 ojciec dziecka odebrał kolejny telefon, który tym razem został zarejestrowany na automatycznej sekretarce. Niestety było to tanie urządzenie, którego kasetka na skutek długoletniego używania była mocno podniszczona. Poza tym przewód telefoniczny był uszkodzony tak, że na nagranie mowy nałożyły się zakłócenia w postaci przydźwięku na częstotliwości sieci i pojedynczych tonów harmonicznym. Porywacz udzielił wskazówek na temat nominałów banknotów okupu i pojemnika, w którym miały być dostarczone. Przekazanie miało nastąpić samochodem. Ojciec dziecka nie był pewny czy wszystkie wskazówki dobrze zrozumiał, ponieważ anonim mówił cicho, a w tle słychać było szum pojazdów, jakby na ruchliwej ulicy. Mimo ostrzeżeń ze strony porywacza rodzice natychmiast zawiadomili policję, która założyła profesjonalny podsłuch telefoniczny.

Od tego momentu przed ekspertem fonoskopii otwierają się trzy możliwe obszary działania: analiza głosu, fonetyczna analiza tekstu i elektroniczne oczyszczanie nagrania mowy.

Analiza głosu

Jeżeli przedmiotem badania jest rozmowa anonimowa lub pseudo-

anonimowa, można podjąć tzw. analizę głosu. To znaczy, przy uwzględnieniu ilości i jakości technicznej nagranego materiału uzyskanie możliwie najbardziej szczegółowego opisu nawyków głosowych mówcy oraz wszystkich innych cech, które mogą służyć do identyfikacji lub lokalizacji. Jeżeli istnieje więcej zarejestrowanych rozmów, należy przeprowadzić porównanie także między nimi, ponieważ dla policji ważna jest informacja, czy należy się liczyć z większą liczbą sprawców. Sam fakt, że porywacz powiedział „my”, nie ma większego znaczenia, ponieważ jak wynika z praktyki wielu pojedynczo działających sprawców, w celu zwiększenia psychicznego nacisku używa liczb mnogiej. W zależności od obfitości materiału nagrania ekspertyza może dotyczyć następujących cech:

● Płeć

W przypadkach gdy nie zachodzi maskowanie głosu określenie płci mówcy z reguły nie przysparza kłopotu. Może być jednak inaczej, gdy tzw. średni poziom głosu mówiącego, będący szczególnie ważnym wskaźnikiem płci jest bardzo wysoki lub bardzo niski. Jeżeli podstawowa częstotliwość średniego pasma głosu, czyli przeciętna częstotliwość na sekundę z jaką drgają przeciwległe – umieszczone w krtani struny głosowe jest bardzo wysoka u mężczyzny lub bardzo niska u kobiety, to pokrywają się one wyraźnie z rozkładem częstotliwości dla płci przeciwnej. Ponieważ

większość badanych kryminalistycznie mówców to mężczyźni, odpowiednio częściej pojawiają się problemy z wysokimi niż z niskimi głosami. W niewielkiej liczbie przypadków, zwłaszcza gdy występują zniekształcenia głosu, jak szept, przytłumienie lub mówienie falsetem nie jest możliwe określenie płci z całkowitą pewnością.

• Wiek

Określenie wieku mówcy może być potrzebne przede wszystkim wtedy, gdy nie mieści się on w statystycznie przeciętnym zakresie między wiekiem 20 i 40 lat. Przy tym trzeba jednak pamiętać, że określenie z większą dokładnością niż do $\pm$ pięciu lat nawet dla doświadczonego eksperta fonoskopii kryminalistycznej nie jest możliwe, a mianowicie z następującego powodu: tak samo jak ludzie mogą na swój wiek chronologiczny wyglądać młodziej lub starzej, tak samo ich głosy z różnych powodów mogą brzmieć młodziej lub starzej. Innymi słowy, wiek biologiczny i chronologiczny mogą się różnić. Wprawdzie istnieje na ten temat obszerna literatura fachowa oparta na licznych badaniach empirycznych, jednak całkowita liczba cech ważnych dla określenia wieku wciąż jeszcze nie jest zamknięta, a tym bardziej nie jest sprecyzowane ich wzajemne oddziaływanie.

Pewne jest jednak, że głosy kobiece na ogół trudniej ocenić niż męskie, co przede wszystkim spowodowane jest przez z reguły wyższe ustawienie głosu mowy i związane z tym większe odległości między nadającymi głosowi brzmienie tonami harmonicznymi. Można by wykazać, że palenie papierosów (jedna paczka lub więcej dziennie) może przeciętnie postarzyć głos o około 6 lat. Z tego powodu przy szacunkowym określaniu wieku

wprowadzono zastrzeżenie, że jeżeli badany mówca jest chronicznym palaczem, może on być nawet do dziesięciu lat młodszy. Empirycznie udo- wodniono przy tym, że ocena wieku zależy także od stanu zdrowia mówcy, analogicznie do młodego lub staro- go wyglądu człowieka. Ogólnie można powiedzieć, że przy obecnym stanie wiedzy szacowanie wieku jest procesem subiektywnym i badania akustyczne niewiele tu pomagają. Na podstawie nowszych wyników można nawet stwierdzić, że naukowe badania fonoskopijne, a nawet specjalistyczne przygotowanie kryminalistyczne nie są tak miarodajne, jak codzienne doświadczenie w badaniu komunikacji werbalnej.

• Pochodzenie regionalne

Na obszarach, na których występują wyraziste dialekty, jak w przypadku języka niemieckiego, przede wszystkim w Niemczech, Austrii i Szwajcarii doświadczenie pokazuje, że w zależności od materiału nagrania dokładne opisanie dialektu, a w związku z tym pochodzenia regionalnego mówcy w bardzo wielu wypadkach w decydujący sposób przyczynia się do identyfikacji, względnie wykluczenia podejrzanego. W praktyce oznacza to najbardziej znaczący wynik kryminalistycznej analizy głosu. Pochodzenie regionalne odnosi się przy tym niekoniecznie do miejsca urodzenia, raczej do dialektologicznie zdefiniowanej okolicy, podczas pobytu w której rozstrzyga się fazę kształtowania nawyków mowy. Faza ta przeważnie odpowiada latom szkolnym. Godny uwagi jest też fakt, że podatność językowa na wpływy dialektu utrzymuje się tym dłużej, im dłuższy jest okres uczęszczania do szkoły i szkół wyższego stopnia. Przy dostatecznej ilości materiału do badania pojawia się nawet

możliwość wykrycia u tego samego mówcy większej liczby kompleksowych cech dialektalnych, co wskazuje na dłuższy pobyt w różnych obszarach regionalnych, dzięki czemu w dochodzeniu zyskuje się dodatkowe informacje.

W większości przypadków mówca badany fonoskopijnie nie mówi dialektem w ścisłym naukowym sensie, a jedynie wykazuje mniej lub bardziej silne naleciałości do dialektu w mowie potocznej. Zadaniem eksperta polega na identyfikacji i klasyfikacji tych regionalnych elementów związanych z płaszczyzną językową. Oczywiście ważna jest nie tylko fonetyka, ale także gramatyka, dobór słów i budowa zdań, które mogą wykazywać regionalną specyfikę. Technika budowania regionalnej charakterystyki mówcy opiera się na klasycznym zastosowaniu tradycyjnej fonetyki, która obecnie może być wspierana przez nowoczesne systemy komputerowe w taki sposób, że subiektywna wiedza eksperta współdziała z obiektywną podstawą techniczną, dzięki czemu ekspertyza osiąga wyższą wartość dowodową. Specjalnego rodzaju narzędziem jest tzw. system ekspercki utworzony w Federalnym Urzędzie Kryminalnym. Jest to wspomagany komputerowo bank danych regionalnej niemieckiej mowy potocznej (DRUGS), za pomocą którego stwierdzone fonetyczne cechy brzmienia poszczególnych dialektów można przyporządkować konkretnym regionom niemieckiego obszaru językowego. Podstawową ideą utworzenia DRUGS była świadomość, że nawet najbardziej biegły specjalista od dialektów nie może opanować wszystkich regionalnych odmian języka niemieckiego, a w każdym razie niejednakowo dobrze. Przy udziale specjalistów BKA i urzędów kryminalnych w landach bank danych zbudowano

w taki sposób, aby użytkownik nie musiał sam dysponować wysoką wiedzą dialektologiczną, a jedynie był obeznany z audytywno-fonetyczną metodą analityczną. Wynik takiej analizy otrzymuje się w postaci symboli wymowy fonetycznej (nie do pomylenia ze zwykłymi znakami alfabetu), a na ekranie ukazuje się mapa niemieckiego obszaru językowego z barwnym oznakowaniem regionu, w którym występuje dany sposób wymowy. Podczas przeszukiwania pod kątem różnych kombinacji brzmieniowych metoda rastru pozwala na szybkie określenie poszukiwanego regionu.

● Obcy język ojczysty

Określenie obcego języka ojczystego nieznanego mówcy może mieć duże znaczenie dla jego identyfikacji. Przeprowadza się je podobnie jak określanie pochodzenia regionalnego metodą audytywno-fonetyczną. Przy tym trzeba zwrócić uwagę na pewien szczególny fakt znany z doświadczenia, a mianowicie że w około połowie przypadków, w których mówca – według opinii policji, ale także ofiar i świadków – ma obcy akcent, po wykonaniu analizy okazuje się, że było to fałszerstwo.

Jest to wartościowa informacja dla śledztwa. W ten sposób na przykład anonimowy szantażysta może udawać, że jest członkiem mafii, aby nadać większą wagę swoim pogrożkom i w tym celu zmienia akcent na taki, który jego zdaniem sugeruje włoskie pochodzenie. W większości wypadków ekspert bez większego trudu może stwierdzić, czy obcy akcent jest prawdziwy, czy udawany, jeżeli tylko zbada tak zwane fonetyczne i gramatyczne interferencje między dwoma oddziałującymi na siebie językami. Szczególne znaczenie ma przy tym wpływ brzmienia języka oj-

czystego na obcy. O udawanym obcym akcencie świadczy właśnie brak interferencji w miejscach, w których można by ich oczekiwać jako wyniku oddziaływania obcego języka ojczystego na niemiecki. Warunkiem wykonania takiej analizy jest oczywiście dobra znajomość systemu fonetycznego danego języka u eksperta. W związku z tym bardzo przydatny okazał się bank danych z próbkami 40 języków. Każdy tekst w próbce przeczytany jest w obcym języku przez native speakera, a następnie przez tę samą osobę po niemiecku. Pierwsza część przedstawia więc wymowę głosek i ich kombinacji oraz melodykę języka obcego, a druga będąca ich następstwem interferencje w języku niemieckim.

● Indywidualne cechy charakterystyczne

Na podstawie trzech przedstawionych powyżej głównych dziedzin analizy głosu możliwe są kolejne badania, mające na celu indywidualizację cech. Jest zrozumiałe, że cecha jest tym bardziej indywidualizująca, im rzadziej występuje. Przede wszystkim chodzi o patologiczne cechy głosu, mowy i sposobu mówienia, wśród których najczęstszą jest seplenienie, czyli nieprawidłowa wymowa głoski „s”. Za pomocą odpowiednich metod instrumentalnych można takie cechy przedstawić obiektywnie, np. seplenienie za pomocą analizy spektralnej, a chrypkę za pomocą tzw. analizy Jitter-Shimmer. Oddech mówcy może również objawiać cechy indywidualne. W tym wypadku analizuje się częstotliwość, głębokość (czas trwania), a przede wszystkim towarzyszące mu dźwięki (skład spektralny), przy czym pewne ograniczenia wyznacza wysokość amplitudy drgań w nagraniu telefonicznym. W zakresie sposobu mó-

wienia interesujące są przede wszystkim tzw. stereotypy, to znaczy przede wszystkim przesadnie częste używanie słów, frazesów lub całych zdań w miejscach nie wynikających z treści wypowiedzi. Te nawyki językowe (habits) mogą być wykryte audytywnie i akustycznie zakładając, że materiał nagrania jest dostatecznie długi, aby mogły się one nagrać. Należą do nich także tak zwane nawyki niejęzykowe, np. pociąganie nosem (charakterystyczne dla konsumentów kokainy) lub cmokanie przy otwieraniu ust. Najczęstszą przyczyną tego są źle dopasowane protezy zębowe i niektóre leki zmieniające konsystencję śliny (np. beta-blokery).

Pożytek z tych cech dla eksperta polega przede wszystkim na tym, że mówca sam sobie z nich nie zdaje sprawy i nie próbuje ich wyeliminować. Cechy te występują głównie u osób na niższym poziomie wykształcenia, ale nie tylko. Dla eksperta mogą one stanowić pewną wskazówkę. Znaczenie ma też słownictwo używane przez mówcę, które może wskazywać na przynależność do pewnej grupy społecznej lub zawodowej.

W zależności od złożoności zadania, szczególnie od ilości i jakości nagrań materiału, doświadczony ekspert potrzebuje zwykle nie więcej niż jeden dzień na sporządzenie analizy głosu na piśmie. Fakt, że w początkowym stadium porwania lub szantażu rejestracja głosu sprawcy jest często jedynym bezpośrednim śladem, powoduje, że jego znaczenie kryminalistyczne jest oczywiste.

Fonetyczna analiza tekstu

W opisywanym przypadku dowodowym ojciec porwanego dziecka nie był pewien, czy dobrze zrozumiał słowa anonima, ponieważ ten mówił

cicho, a w tle słychać było szum ulicy. Do tego doszły zakłócenia na skutek uszkodzonego kabla. W tym stadium ekspertyzy trzeba więc odtworzyć wypowiedziane słowa. Może to wykonać tylko osoba wykształcona, która miała już do czynienia z nagraniami mowy sporządzonymi w szczególnych okolicznościach, a więc najlepiej, aby był to ekspert fonoskopii kryminalistycznej. Do tego zadania, będącego w zasadzie rozpoznaniem mowy w języku niemieckim, wprowadza się pojęcie fonetycznej analizy tekstu, która często związana jest z elektronicznym poprawianiem jakości nagrania.

Podczas pierwszego odsłuchania całości nagranych tekstu ekspert najpierw odbiera ogólne wrażenie jego charakterystyki akustycznej i językowej. Następnie materiał nagrania zostaje przetworzony za pomocą cyfrowej akustycznej techniki edytorskiej. Szczególnie chodzi o możliwość skoncentrowania się na niejednoznacznych i potencjalnie kontrowersyjnych miejscach. Jeżeli nie osiąga się jednoznaczności, ekspert sporządza listę wszystkich możliwych pod względem fonetyczno-fonologicznym opcji. Czynność ta wymaga nie tylko wyszkolonego ucha, ale także podstawowej wiedzy filologicznej i znajomości danego języka. Z reguły transkrypcja słowna inkryminowanego tekstu sporządzana jest zgodnie ze zwykłymi zasadami ortografii, stosowanymi oczywiście przez policję i sądy. Dla własnych potrzeb biegły często sporządza przynajmniej częściową transkrypcję fonetyczną, która dokumentuje szczegóły brzmienia mowy. Dzięki temu powstaje możliwość stworzenia podstaw dla własnej odrębnej opinii w przypadku różnic zdań między biegłymi.

Szczególny problem podczas spisywania tekstu z nagrania stanowią

nazwy miejscowości, nazwiska osób i inne nazwy własne, a więc elementy mowy, których w znacznej części trzeba się domyślać z kontekstu – najczęściej w złych warunkach akustycznych. W praktyce trzeba tu korzystać z pomocy prowadzącego dochodzenie, zorientowanego w sprawie i znającego okolicę.

Kolejnym problemem kryminalistycznej rejestracji mowy jest równoczesna rozmowa kilku osób. Zdarza się to często, gdy warunki akustyczne są zdecydowanie niekorzystne, np. rejestracja przebiega z ukrytego mikrofonu w pomieszczeniu akustycznie twardym, w którym jest pogłos (np. w celi więziennej), a mówiący zbliża się do mikrofonu i oddala od niego. W takim wypadku ekspert musi się koncentrować każdorazowo tylko na jednym głosie, a wszystkie inne świadomie wyciszać, notując wypowiadane przez ten głos słowa, a następnie powtarzać tę samą procedurę kolejno dla wszystkich pozostałych głosów. Najlepszym sposobem dokumentacji równoczesnej rozmowy kilku mówców jest notacja stopadła wzajemnego następstwa głosów, podobnie jak w partyturze orkiestrowej. Dzięki temu czytelnik otrzymuje nie tylko ciągły zapis przebiegu rozmowy (poziomo), ale także może poznać czasowe odstępy między wypowiedziami poszczególnych mówców (pionowo).

Elektroniczne czyszczenie zapisu mowy

Od początku lat 80. ubiegłego wieku pojawiają się coraz nowsze metody techniczne – początkowo analogowe, później cyfrowe – poprawiania zakłóconego sygnału mowy. Szczególnie przydatne okazało się to dla badań kryminalistycznych. Liczne techniki, przede wszystkim takie, któ-

re znajdują się w kręgu zainteresowania przemysłu fonograficznego, obecnie są dostępne w postaci względnie niedrogich pakietów programów. Niektóre z najbardziej skutecznych metod są jednak – o ile w ogóle znajdują się w handlu – wciąż bardzo drogie i dlatego zastrzeżone dla laboratoriów państwowych. Chociaż większość algorytmów pod względem czysto technicznym nie jest skomplikowana, stosuje się jednak pewne środki ostrożności w celu ich ochrony. Ogólnie można stwierdzić, że istniejący algorytm poprawiania zapisu mowy nie istnieje, tak samo jak istniejący lek w medycynie; niepożądane działania uboczne są nie do uniknięcia. Nie jest możliwe usunięcie wszystkich zakłóceń sygnału i jednocześnie pozostawienie przy tym miejsc niezakłóconych w stanie nienaruszonym. W praktyce trzeba zawsze szukać kompromisu między dwiema skrajnościami, przy czym najważniejszy jest cel przedsięwzięcia: dla niewyszkolonych słuchaczy, jak policjanci, świadkowie lub członkowie składu sędziowskiego, odsłuchanie dłuższego nagrania mowy zniekształconego przez głośnie brzęczenie sieci stanowi wielkie obciążenie, w krótkim czasie powodujące zmęczenie i obniżenie uwagi. W takim wypadku proste usunięcie zakłóceń ułatwia odsłuchiwanie, ale słuchacze powinni mieć świadomość, że brzmienie nagranych głosów uległo zmianie.

Takie ostrzeżenie jest szczególnie ważne wtedy, gdy zastosuje się silnie inwazyjne algorytmy, jak filtrowanie adaptatywne lub substrakcja spektralna. W przypadku gdy nie dokonuje się fonetycznej analizy tekstu, a więc rozpoznania mowy, a nadrzędnym celem jest rozpoznanie mówcy, wówczas wykwalifikowany ekspert w celu zachowania niezafałszowanego

brzmienia analizowanego głosu powinien zawsze przy jednym uchu mieć słuchawkę z oryginalnym, zakłóconym sygnałem mowy, a przy drugim odfiltrowany sygnał i w ten sposób próbować audytywnie wyeliminować zakłócenia.

Wynika z tego, że również techniczne, pozornie obiektywne elektroniczne poprawianie sygnału mowy zawiera znaczące komponenty subiektywne, poczynając od podjęcia decyzji, czy w danym przypadku w ogóle wskazane jest „oczyszczenie” sygnału mowy. Wprawdzie pracuje się tylko na jednym algorytmie filtra, dokładnie i zawsze w ten sam sposób, ale wszystkie wartości progowe czułości, właściwości filtra, stałe czasowe itp. są określane przez eksperta na podstawie jego wiedzy i doświadczenie. Podsumowując trzeba powiedzieć, że w przypadku takich zakłóceń dźwiękowych, jak: hałas pojazdów, szumy sieci, muzyka, impulsy GSM oraz – co jest najgorsze – inne zakłócające głosy ludzkie w nagraniu (mono) rozmowy telefonicznej lub z mikrofonu – nawet jeśli się wykorzystuje najnowocześniejsze elektroniczne metody oczyszczania sygnału mowy, nie należy oczekiwać cudów ani efektów typu wszystko albo nic. Dokonanie skoku jakościowego z kategorii całkowitej niezrozumiałości oryginału do elektronicznie ulepszonej kopii całkowicie zrozumiałej i bez zakłóceń – pozostaje w sferze iluzji.

Przypadek dowodowy, część 2

Następnego dnia po otrzymaniu taśmy z nagraniem głosu anonimowego porywacza ekspert fonoskopii przekazał policji wyniki analizy głosu. Po zastosowaniu elektronicznego czyszczenia mowy można było zro-

zumieć bez wątpliwości słowa nagranej wiadomości. Opracowany profil fonograficzny anonimusa był następujący: męczyzna, język ojczysty niemiecki z naleciałościami dialektu nadreńskiego. Wiek 20–40 lat, jak na dorosłego mężczyznę dość wysoka średnia wysokość głosu. Na podstawie melodyki i rytmu mowy można przypuszczać, że tekst został odczytany. Nic nie wskazuje jednak, aby tekst przegrano z kieszonkowego dyktafonu lub walkmana (żadnych impulsów spowodowanych przez włączanie lub wyłączenie, żadnej redukcji charakterystyki częstotliwościowej, żadnych szumów przesuwającej się taśmy). Ponieważ wiadomość nie była nagrana wolnym tekstem, trudno było ocenić możliwości werbalne (elokwencję) mówcy, w każdym razie budowa zdań i formy gramatyczne były bez zarzutu. Jediną zwracającą uwagę cechą była szczególna wymowa głoski -sz-, która powtórzona była trzykrotnie, a w dwóch wypadkach towarzyszył jej lekki poświsł, wykrywalny na spektrogramie w postaci rezonansu przy 2,8 kHz. Najczęstszą fizjologiczną przyczyną takiej cechy jest luka między dwoma przednimi górnymi siekaczami. W opisywanym przypadku nie ma znaczenia, czy cecha ta jest odchyleniem od normy, czy też można ją uważać za defekt wymowy – w każdym razie jest ona stosunkowo rzadka i dlatego ma charakter indywidualny.

Zestawienie analizy głosu z wynikami innych badań kryminalistycznych wykazało, że mówca jest mężczyzną w wieku 20–40 lat, pochodzącym z okolic Kolonii, na południe od Bonn, ale nie dalej na północ niż Düsseldorf, posiadającym wiele szczegółowych informacji na temat ofiary i jej rodziny. Ojciec dziecka jako jednego z możliwych sprawców

wymienił K., jednego z dawniejszych klientów jego firmy z Bonn, który z powodu niewywiązywania się ze zobowiązań finansowych został podany do sądu przez firmę i przegrał proces, za co później wielokrotnie groził zemstą. Już pierwsza rozmowa telefoniczna K., podczas której zamówił on pizzę, dostarczyła materiału porównawczego. Ponadto Policja przekazała ekspertom próby głosu trzech innych osób, A, B i C, które w związku z innymi śladami zostały wytypowane do sprawdzenia zgodności głosu.

Porównywanie głosu (rozpoznawanie mówcy)

Mając na uwadze najważniejsze i pod względem naukowym najbardziej wymagające zadanie identyfikacji osoby na podstawie mowy (rozpoznanie mówcy, identyfikacja mówcy) trzeba na wstępie zaznaczyć, że zgodnie z zasadą *nemo tenetur* obwiniony może odmówić współdziałania przy pobieraniu od niego próbki porównawczej głosu. Odnosi się to także do opisaną poniżej identyfikacji przez laików, czyli konfrontacji akustycznej. Na szczęście w praktyce istnieją też inne możliwości uzyskania mniej lub bardziej odpowiedniego materiału porównawczego z innych źródeł, np. z podsłuchu telefonicznego. Trudności praktyczne powstają także wówczas, gdy mówca dostarcza tylko część potrzebnej próbki, a więc na przykład chętnie opowiada o swoim hobby lub czyta tekst z gazety, ale odmawia przeczytania na głos lub wypowiedzenia tekstu nagranego przez sprawcę. W takich wypadkach nie jest możliwe porównywanie „słowo do słowa”, które z naukowego punktu widzenia uważa się za najbardziej efektywne i niezależne od tego, jaką skalę prawdopo-

PRZEKŁADY

dobieństwa się zastosuje. Stanowi ono dla eksperta najlepsze z możliwych potwierdzenie końcowych wniosków. W niewielkiej mierze (jeżeli ilość materiału dowodowego jest bardzo mała) utrudnienie to odnosi się także do niezależnej od znaczenia słów w próbie mowy, czysto akustycznej metody automatycznej, wspomaganej komputerowo identyfikacji mówcy.

W postępowaniu karnym głos ludzki w pewnych okolicznościach ma jedyną w swoim rodzaju zaletę w stosunku do innych środków dowodowych, włącznie z odciskami palców lub śladami DNA, a mianowicie wiąże on w sposób niewątpliwy osobę mówcy z popełnionym czynem. Inaczej niż przy np. teoretycznie możliwym pozostawieniu odcisku palca na miejscu zdarzenia przez człowieka nie mającego związku z popełnionym przestępstwem, rozmowa związana treściowo z wydarzeniem, zawiadomienie lub – jak w omawianym przypadku – żądanie okupu zakłada co najmniej współwzajemność. Decydującą kwestią – a równocześnie znacznym upośledzeniem w stosunku do odcisków palców lub śladów DNA – jest tylko stopień pewności, z jaką można dokonać identyfikacji indywidualnej na podstawie głosu i mowy. Problem rozpoznawania sprawców inspiruje językoznawców i inżynierów na całym świecie do ulepszania istniejących i wynajdywania nowych metod badania. W związku z tym od lat 70. prowadzi się liczne badania mające na celu opracowanie wspomaganej komputerowo, w znacznym stopniu zautomatyzowanej metody, która umożliwi wyekstrahowanie z materiału dowodowego i porównawczego możliwie najbardziej specyficznych cech akustycznych i ocenę ich prawdopodobieństwa statystycznego. Metody te znaj-

dują zastosowanie głównie komercyjne – w biometrycznych systemach kontroli dostępu (chodzi o specjalnie chronione obiekty, jak ośrodki oblężeniowe, składy broni, ostatnio także dostęp do kont bankowych), jednak do niedawna jeszcze nie sprawdziły się w fonoskopii kryminalistycznej. Główną tego przyczyną okazuje się nie tyle niedostateczny potencjał

indywidualizacji cech akustycznych ani niedoskonałość algorytmów użytych do ich oceny, lecz niemożność kontrolowania niezależnych warunków środowiska, wywierających bezpośredni wpływ na przekaz telefoniczny (np. szum pojazdów, zamierzona zmiana brzmienia głosu przez mówcę, bardzo krótki lub złej jakości dźwiękowy materiał dowodowy i brak

Cechy specyficzne mówców w badaniu porównawczym głosu

Tabela

Cechy fonetyczne, lingwistyczne, fizjologiczne	Korelaty akustyczne, fizyczne, statystyczne
<u>Głos</u> – średni poziom głosu mowy, – intonacja (melodyka), – jakość głosu.	podstawowa częstotliwość (F_0), matematyczny wymiar rozrzutu podstawowej częstotliwości, parametry drgań strun głosowych, np. vocal jitter (chrypka), shimmer (wahania głośności) itp.
<u>Mowa</u> – dialekt: rodzaj i stopień naleciałości, – obcojęzyczny akcent, – idiolekt (suma typowych cech indywidualnych), – socjolekt (suma specyficznych cech grupowych; języki profesjonalne, żargon itp.).	suma odchyśleń od standardowych norm wymowy, suma odchyśleń od norm wymowy języka obcego dla mówcy, suma odchyśleń od reguł języka standardowego, odchylenia od leksykalnej, w pewnych sytuacjach także gramatycznej płaszczyzny danego języka.
<u>Sposób mówienia</u> – rytm mowy (następujące po sobie sylaby akcentowane i nieakcentowane), – intonacja, – szybkość artykulacji (tempo mówienia), – sposób oddychania.	rozdział czasu trwania poszczególnych dźwięków, intensywność („głośność”), wzór częstotliwości podstawowej, częstość sylab (liczba sylab na sek.), częstość artykulacji (liczba sylab na sek. po odjęciu pauz), częstość, długość, spektrum oddechu.

We wszystkich cechach oprócz dialektu mogą występować formy patologiczne i niezgodności z normą, które powodują wzrost wartości dowodowej specyficznej cechy mówcy (zaburzenia wymowy, głosu i mowy)

współpracy ze strony mówcy). W ostatnich latach większość problemów przekazów telefonicznych została rozwiązana dzięki tzw. channel equalizing (normalizacji kanału), możliwe więc, że nastąpi przełom w systemach automatycznej identyfikacji mówców i skorzysta z tego także fonoskopia kryminalistyczna. Obecnie w skali światowej testowanych jest do najmniej pół tuzina metod, niektóre weszły już do praktyki, o czym będzie mowa w dalszej części artykułu. Najpierw jednak krótkie wprowadzenie do stosowanych obecnie metod konwencjonalnych.

Używana dawniej w wielu krajach tzw. metoda odcisku głosu (Voiceprint) na szczęście dziś już pozostaje daleko za nami. Polega ona na wizualnym porównywaniu przedstawionych w określony sposób trójwymiarowych spektrogramów, z których sporządza się próbki porównawcze mowy w odcinkach co 2,5 sek. Następnie lepiej lub gorzej przeszkolony obserwator, np. policjant po dwutygodniowym kursie, określa podobieństwa i różnice w postaci różnic szarości na specjalnym papierze przedstawiającym amplitudę częstotliwości głosek i ich kombinacji w próbce dowodowej i porównawczej i ocenia je według subiektywnej skali. Obecnie takie spektra o wysokiej jakości technicznej sporządza się za pomocą cyfrowego przetwarzania sygnału, ale nie zmienia to faktu, że metoda ta obciążona jest tak licznymi wadami, że w praktycznie jednogłośnej opinii naukowców jest całkowicie nieprzydatna do identyfikacji mówców. Dowiedziano tego już w 1979 r. w USA na podstawie studium naukowego przeprowadzonego na zlecenie ministra sprawiedliwości oraz wielu innych badań. Zwłaszcza wskazuje się na brak potencjału indywidualizującego cech specyficznych mówcy: rzekome ce-

chy indywidualne mówcy różniły się nieraz bardziej w szerszej skali w porównaniu z cechami innych mówców. Aby uniknąć nieporozumień: widmo szerokopasmowe jest samo w sobie od chwili jego wynalezienia przed 50 laty niezastąpionym instrumentem analizy sygnału mowy i innych sygnałów akustycznych (jak tony serca, szum przepływu krwi w naczyniach, głosy zwierząt), ale nie może służyć jako podstawowa metoda rozpoznawania mówców.

W Niemczech wyciągnięto dość wcześnie wnioski ze wspomnianych amerykańskich badań i zaczęto szukać nowych dróg metodycznych. W połowie lat 80. powstała najczęściej stosowana obecnie także w skali międzynarodowej fonetyczno-instrumentalna metoda kombinacyjna. Polega ona na ekstrakcji specyficznych cech mówcy z trzech obszarów języka mówionego: głosu (w sensie zawężonym), mowy i sposobu mówienia. Suma zgodności i różnic oceniana jest według skali priorytetów. Dla niektórych cech istnieją dane statystyczne na temat ich rozdziału w populacji, które w pojedynczych przypadkach pozwalają na kwantyfikację proporcji. Ważną właściwością metody w kontekście wykorzystania środka dowodowego jest podwójna analiza każdej z cech; z jednej strony, za pomocą klasycznych metod fonetyki lub diagnostyki medycznej, a z drugiej, w postaci przedstawienia akustycznego z wykorzystaniem nowoczesnych cyfrowych metod analizy sygnału. Możliwa jest przy tym kontrola dokonywana zarówno przez wykonujących analizy ekspertów, jak też przez innych ekspertów, a do pewnego stopnia również przez laików (sąd). Tabela zawiera pobieżny przegląd oznaczonych tą metodą najczęściej spotykanych specyficznych cech mówcy. Słabością jej jest z jed-

nej strony, niemożność uniknięcia subiektywizmu, który wiąże się przede wszystkim z oceną zgodności, względnie różnic między cechami. Z drugiej zaś, trzeba stwierdzić, że wyekstrahowane cechy mają różny potencjał indywidualizacji i w konkretnym przypadku również nie wszystkie poddają się ocenie, jak na przykład wspomniana wyżej średnia częstotliwość głosu, jeżeli mówca mówi szeptem.

Ogólnie można uznać, że stosowana przez kompetentnych ekspertów fonetyczna metoda kombinacyjna jest wartościowym środkiem identyfikacji mówców w trudnych warunkach. Jednak na ogół – pomijając rzadkie przypadki – nie pozwala ona na identyfikację z prawdopodobieństwem graniczącym z pewnością. Obecnie prawie wszyscy eksperci w Niemczech wykonujący tego rodzaju ekspertyzy wykorzystują tradycyjną skalę prawdopodobieństwa zapożyczoną z dziedziny kryminalistycznej identyfikacji pisma ręcznego, która w istocie napotyka na takie same problemy metodyczne i rzeczowe oraz imponderabilia.

Jak już wspomniano, od niedawna stosuje się także automatyczne metody kryminalistycznego rozpoznawania mówców. Amerykański Instytut Normalizacji NIST organizuje regularne testy jakościowe, które pozwalają na ewaluację systemu na olbrzymim materiale nagrań mowy (z mikrofonu i telefonu). W Hiszpanii, Francji i Szwajcarii metody te znalazły już miejsce w codziennej pracy sądów, w USA jeden z systemów jest testowany przez FBI. Ponieważ wciąż jeszcze istnieją podstawowe i metodyczne braki i niewyjaśnione pytania trzeba przyjąć założenie, że w przewidywalnym czasie nie mogą one zastąpić konwencjonalnych metod, a jedynie je uzupełniać. W przyszłości

według wszelkich przewidywań ekspertyza fonoskopijna dotycząca rozpoznania mówcy będzie się składała z dwóch komponentów: konwencjonalnego i automatycznego porównywania głosu.

Metody automatyczne różnią się pod względem parametrów akustycznych, zaplecza w postaci bazy danych i algorytmów statystycznych. Ich wspólną właściwością są:

- znaczne ograniczenie komponentów subiektywnych, wprowadzonych przez użytkownika systemu,
- automatyczna ekstrakcja cech akustycznych pochodzących z właściwości odbicia w jamie nosowo-gardłowej,
- budowanie modeli statystycznych dla badanego mówcy na tle populacji światowej,
- obliczanie prawdopodobieństwa, że między modelem inkryminowanego mówcy i mówców porównawczych występuje policzalna różnica także na tle populacji. Im mniejsze jest to prawdopodobieństwo, tym większe prawdopodobieństwo identyczności.

Na podstawie faktu, że cechy analizowane metodami automatycznymi nie rozdzielają się tak jak w metodzie fonetyczno-instrumentalnej na elementy lingwistyczne, takie jak słowa, sylaby i głoski i, jak już wspomniano, ukierunkowane są na właściwości rezonansowe pochodzące od kształtujących mowę struktur w ustach i jamie nosowo-gardłowej, zostały one nazwane systemami Total-Voice. W tym badaniu nie jest ważne, w jakim języku wypowiedziany został tekst i ekspert obsługujący system wcale nie musi go znać. Fakt ten jest niezwykle korzystny dla pracy policji, a także dla sądu.

Można wyobrazić sobie następujący przypadek: w pewnym postępowaniu zabezpieczono bardzo wiele podsłuchanych rozmów w całkowicie egzotycznym języku, dla którego trudno było znaleźć tłumacza. Wiadomo było, że mówca mógł się znajdować wśród około tuzina osób, od których uzyskano nagrania porównawcze o niekwestionowanej autentyczności. System automatyczny może w ciągu kilku dni lub najwyżej tygodni określić, która z tych osób z największym prawdopodobieństwem uczestniczyła w nagraniach dowodowych.

Następnie można te rozmowy przetłumaczyć, ze względu na koszty i czas ograniczając się tylko do tych rozmów, w przypadku wystąpiła zgodność – jaka oszczędność pracy, czasu i pieniędzy. Ta funkcja systemów automatycznych nazywana jest wewnętrzną klasyfikacją.

Następujący przykład ilustruje zastosowanie automatycznego systemu rozpoznawania mówców na podstawie porównywania głosu. Chodzi o system Identi-Vox opracowany na Politechnice w Madrycie na zlecenie hiszpańskiej policji, z którego autor artykułu w ramach wspólnego projektu badawczego korzystał, równolegle z fonetyczną metodą kombinacyjną, podczas sporządzania ekspertyz. We wszystkich opracowywanych w ten sposób przypadkach z Niemiec i jednym z Norwegii osiągnięto całkowitą kompatybilność wyników. Z powodu różnych skali prawdopodobieństwa ilościowa porównywalność obu metod nie jest możliwa, ale można stwierdzić, że w żadnym z przypadków nie ujawniły się tendencje przeciwnostawne.

W przypadku dwóch nagrań telefonów z pogrózkami, skierowanych do pewnej firmy, które razem trwały tylko 18 sekund, podejrzanym był je-

den z pracowników A. Fonetyczna ekspertyza porównania głosów wykazała wysokie prawdopodobieństwo identyczności A i szantażysty, m.in. z powodu charakterystycznej, zauważalnej nawet dla laików na sali sądowej niewyraźnej wymowy głosek f przez obu mówców, co ujawniło się w analizie spektralnej w postaci utraty częstotliwości powyżej 2,5 kHz. Kuriozalną przyczyną u podejrzanego A okazał się brak wszystkich czterech górnych sił głosu, co jest rzadką i dlatego silnie widoczną cechą. Analiza za pomocą Identi-Vox wykazała w tym przypadku prawdopodobieństwo identyczności 0,1 do 1. Wynik ten widząc, statystycznie jest całkowicie niezależny od wyniku ekspertyzy fonetycznej, ponieważ tamten oparty był na badaniu zupełnie innych cech w zakresie częstotliwości niższym niż ten, w którym ujawniła się wymieniona rzadka cecha. W ten sposób uzyskano znaczące podwyższenie pewności identyfikacji. Widać jest więc, że propagowane przez niektórych inżynierów przede wszystkim we Francji proste zastąpienie subiektywnej metody fonetycznej przez obiektywną automatyczną nie byłoby rozsądne.

W dużym skrócie Identi-Vox został w tym przypadku zastosowany w następujący sposób: najpierw pobrano reprezentatywne próbki mowy od 100 mówców niemieckich płci męskiej jako tło populacji. Po znormalizowaniu używanych kanałów telefonicznych wyekstrahowano cechy akustyczne każdego z mówców (38 tzw. współczynników Mel Cepstrum). Na tej podstawie przeliczono modelowe nagranie 100 mówców oraz wyznaczono różnice między nimi. Następnie część próbki mowy podejrzanego A przeliczono również na model mówcy. Zmienność, czyli osobisty zakres odchyleń jego cech akustycz-

nych, został obliczony z wydzielonej części jego próbki mowy (tzw. próbki kontrolnej). Na końcu przeanalizowano akustycznie nagranie szantażysty. Na podstawie próbek tła populacji identyczność głosu szantażysty i podejrzanego jest 381,7 razy bardziej prawdopodobna niż nieidentyczność. Czas potrzebny na przygotowanie próbek tła populacji i właściwe porównywanie głosów z wykorzystaniem nowoczesnego komputera (procesor IV Pentium) wynosi kilka dni.

Przypadek dowodowy, część 3

Dalszy przebieg badania przypadku dowodowego był następujący. Profesjonalnie zarejestrowane próbki porównawcze mowy czterech mówców miały dobrą jakość techniczno-akustyczną. Ilość nagranych materiału i jego wydajność była w każdym przypadku inna. Tylko nieliczne słowa użyte w tekście anonima powtarzały się w próbkach porównawczych; dlatego można było wykonać niewiele porównań typu „słowo do słowa”.

U mówców A i B ujawniły się znaczne różnice i żadnych znaczących zgodności. Szczególnie duże różnice dotyczyły naleciałości dialektalnych, które są tzw. kryterium zabójczym, tzn. wykluczają identyczność. Nagranie mówcy C wykazywało podobieństwa w zakresie takich cech, jak naleciałości regionalne, ale też niezgodność innych cech. Mówca K. w swoim nagraniu nie tylko ujawnił znaczącą zgodność z nagraniem anonima w zakresie tych cech, które mimo krótkości nagrania udało się określić, ale także zgodność rzadko spotykanego sposobu wymowy głoski -sz-.

Porównawcza analiza spektralna ujawniła nawet zgodność ilościową

(częstotliwość). Ekspert ocenił wszystkie cztery wyniki na skali prawdopodobieństwa. Gdyby posłużono się metodą automatyczną, w tym miejscu znajdowałyby się odpowiednie ilorazy prawdopodobieństwa. Wyniki ekspertyzy (A i B z bardzo dużym prawdopodobieństwem nie są identyczni z anonimem, C prawdopodobnie nieidentyczny, K. z dużym prawdopodobieństwem identyczny) pozwoliły policji na wykluczenie pierwszych dwóch mówców, zaniechano też dochodzenia wobec C, koncentrując się na osobie K. Założono nadzór telefoniczny jego telefonu sieciowego i komórkowego w celu uzyskania nowego materiału do dalszych badań porównawczych głosu, a także bezpośredniego dowodu sprawstwa (ponowne żądanie okupu).

Konfrontacja akustyczna (identyfikacja mówcy przez laików)

Na zakończenie trzeba jeszcze krótko omówić nie występujące w opisanym przypadku dowodowym rozpoznanie mówcy, które dość często wykorzystuje się w badaniu napałów na banki, zgwałceń lub innych przestępstw, podczas których sprawca jest zamaskowany lub jego identyfikacja wizualna nie jest możliwa z innych przyczyn (ciemność, atak od tyłu). Ponieważ nie dysponuje się wówczas żadnym nagraniem głosu sprawcy, rozpoznanie mówcy przez eksperta nie jest możliwe. Korzysta się wówczas z tzw. konfrontacji, która naukowo określana jest jako rozpoznawanie mówców przez laików (ofiara, świadek). Analogia do dobrze znanej konfrontacji wizualnej jest oczywista, ale identyfikacja na podstawie głosu ma poważną niedogodność, a mianowicie tranzytoryczny charakter sygnału mowy, który w przeciwieństwie do zdjęcia lub wi-

doku osoby jest ograniczony w czasie. Badania wykazały, że identyfikacja w obszarze wizualnym funkcjonuje lepiej niż w audytywnym – przynajmniej tak wynika z eksperymentów laboratoryjnych. Wprawdzie w ostatnich latach wykonano wiele badań w dziedzinie konfrontacji akustycznej i opracowano wytyczne do przeprowadzania testów, jednak wiele problemów naukowych pozostało niewyjaśnionych, np. jeśli chodzi: o rodzaj i czas trwania pierwszego kontaktu z inkryminowanym głosem (familiarizacja), odstęp czasu między kwestionowanym wydarzeniem i powtórzeniem testu, rola strachu i stresu, forma przedstawienia ocenianego głosu (jednokrotne czy wielokrotne). W każdym razie w środowisku profesjonalistów panuje zgodność co do tego, że akustyczna konfrontacja nie może być brana pod uwagę jako jedyny środek identyfikacji. Pojawia się ona jednak w codziennej pracy kryminalistycznej jako element dochodzenia („Czy to jest głos mężczyzny, który dokonał napadu?”) ponieważ stwarza szansę wykrycia fałszywej identyfikacji i przy tym pokazuje, w jakim stopniu oceniający jest w stanie dokonać przybliżonej identyfikacji mówcy. Podczas organizowania konfrontacji akustycznej, podobnie jak w przypadku wizualnej, muszą być spełnione określone warunki. Ze względu na złożoność koncepcji, interpretację wyniku rozpoznania, a szczególnie strukturę mowy i głosu wybranego wzoru referencyjnego w celu uniknięcia dodatkowych błędów przeprowadzanie testu powinno być zastrzeżone dla eksperta.

Oprac. Anna Henschke