

Wpływ typu rejestratora na rozmieszczenie formantów sygnału mowy

Rozwój myśli technicznej w dziedzinie nagrywania i odtwarzania dźwięku z wykorzystaniem nośników magnetycznych, zapoczątkowany przez Valdemara Poulsena ponad sto lat temu [1], był jedną z przesłanek powstania pół wieku później działu kryminalistyki o nazwie „fonoskopia” (łac. *fono* – dźwięk, *skopeo* – patrzeć). Przedmiotem badań fonoskopijnych są dzisiaj analizy wszelkich zapisów sygnałów akustycznych bez względu na typ nośnika i rejestratora.

Utworzenie pierwszej w Polsce placówki fonoskopii niemal zbiegło się w czasie ze skonstruowaniem w 1963 r. przez firmę Philips pierwszego wzoru kasyety magnetofonowej typu Compact [4]. Był to krok milowy w upowszechnianiu rejestracji magnetycznej dźwięku, który przez lata stanowił najbardziej popularny nośnik, obecnie wypierany przez generację cyfrową.

Zarówno w przypadku techniki analogowej, jak i cyfrowej rynek zaakceptował cały szereg nośników, począwszy od wspomnianej kasyety typu Compact, przez mikrokasetę (różnych formatów, zależnie od producenta), liczne typy i standardy kaset wideo rejestrujących ścieżkę dźwiękową, do taśm DAT, dyktafonów cyfrowych z wbudowanymi lub wymiennymi pamięciami półprzewodnikowymi flash, komputerowych dysków twardych itd. Przedmiotem zleconych badań fonoskopijnych mogą być nagrania dokonane na wszystkich typach dostępnych nośników dźwięku. W zakres takich badań najczęściej wchodzi spisanie treści rozmowy dowodowej, badanie autentyczności nagrania i identyfikacja mówców.

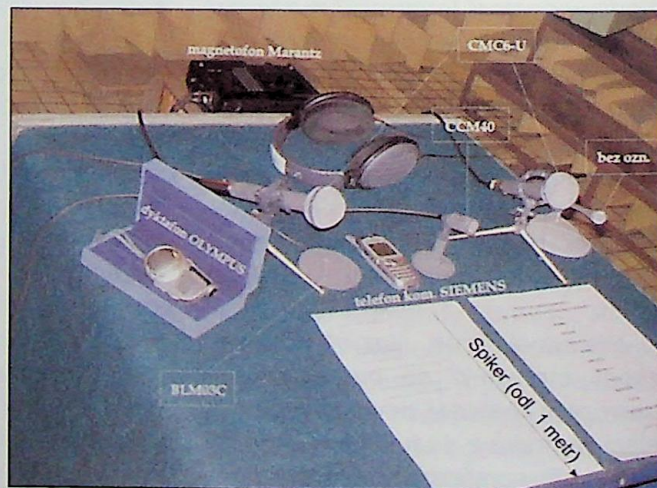
Jednym z etapów identyfikacji mówców jest analiza rozmieszczenia

formantów w widmie sygnału mowy. Ich układ jest bowiem zależny m.in. od anatomicznego ukształtowania struktur zaangażowanych w mówienie [3]. Silna korelacja rozmieszczenia formantów w wypowiedziach dowodowych i porównawczych to jedno z kryteriów (przy identyfikacji mówcy oprócz analizy formantowej przeprowadza się także analizę struktury językowej wypowiedzi), pozwalające na sformułowanie wniosku, że autorem obydwu porównywanych wypowiedzi jest ten sam mówca. W praktycznych zastosowaniach analizy formantowej korelacja ta nigdy jednak nie osiąga wartości zupełnej (współczynnik korelacji $r=1$). Wpływa na nią bowiem szereg czynników, takich jak subtelne różnice pomiędzy poszczególnymi realizacjami słownymi tych samych ciągów alofonów, różnica jakości nagrania dowodowego i porównawczego czy błąd pomiaru [5]. W praktyce badań fonoskopijnych do utrwalenia wypowiedzi porównawczych wykorzystuje się profesjonalne urządzenia rejestrujące mowę, ponadto czynność ta najczęściej odbywa się w warunkach zbliżonych do studyjnych. Zapisy dowodowe realizowane są z użyciem różnych typów rejestratorów analogowych i cyfrowych, przy czym większość rejestratorów cyfrowych wykorzystuje zjawisko kompresji, w tym także tzw. kompresji stratnej (w celu wydłużenia czasu nagrywania), która w zależności od zastosowanego stopnia wpły-

wa na jakość utrwalanego sygnału mowy. Powstała zatem konieczność rozstrzygnięcia wątpliwości, czy stosowanie różnych rejestratorów ma wpływ na rozmieszczenie formantów w widmie sygnału dowodowego i porównawczego. Celem wyjaśnienia niniejszego problemu przeprowadzono stosowny eksperyment w warunkach izolacji akustycznej.

Materiał i metody

Materiał do badań stanowiły wypowiedzi mężczyzny, utrwalane równocześnie przez pięć rejestratorów. Dobór warunkowy był chęcią przesłania ewentualnych zmian rozkładu formantów – w odniesieniu do zapisu analogowego – przy różnych typach rejestratorów cyfrowych. Wykorzystano:



Ryc. 1. Rozmieszczenie mikrofonów urządzeń rejestrujących
Fig. 1. Placement of microphones connected to recording devices

- Analogowy magnetofon reporterski firmy Marantz, model CP 430; prędkość przesuwu taśmy – 4,75 cm/s, taśmę kompaktową firmy TDK typu D60. Do urządzenia podłączono dwa mikrofony pojemnościowe firmy

Schoeps, symbol CMC6U (P12-48), z końcówką MK-4 o charakterystyce kardioidalnej. Para rozstawiona (ustawienie A-B). Wykorzystano przedwzmacniacz firmy SONOSAX typu SX-M2.

- Dysk twardy komputera. Zapis na dysku odbywał się w następujący sposób: sygnał z mikrofonu stojącego firmy Creative (bez oznaczeń, charakterystyka kierunkowa) poprzez zewnętrzną kartę dźwiękową (USB) Exdigy Sound Blaster firmy Creative transmitowany był z użyciem sieci internetowej i komputerowego oprogramowania o nazwie „Scope”, służącego do internetowych rozmów telefonicznych w standardzie VoIP (ang. *Voice over Internet Protocol*). Następnie na drugim zestawie komputerowym, poprzez wewnętrzną kartę AC97 firmy ASUS, rejestrowano sygnał zwrotny, wykorzystując oprogramowanie Cool Edit Pro w wersji 2.0 firmy Syntrillium.

- Cyfrowy dyktafon firmy Olympus, model DW-90, tryb SP (Normal), częstotliwość próbkowania 11 kHz. Urządzenie rejestruje z kompresją bezstratną w postaci plików WAVE.

- Portadisc (MiniDisc) firmy HHB, model MDP500. Urządzenie zapisuje dźwięk na nośnikach minidisc w postaci PCM, częstotliwość próbkowania – 44,1 kHz, rozdzielczość 16 bitów, z wykorzystaniem kodu ATRAC 4.5. Jest to system stratnej kompresji dźwięku, w którym dane ulegają redukcji w proporcji 5:1 z wykorzystaniem fizjologicznych własności słyszenia (np. maskowanie tonów cichych przez współbrzące tony głośnie o sąsiednich częstotliwościach), co jest wykorzystywane przez liczne algorytmy kompresji stratnej [2]. Do urządzenia podłączono dwa mikrofony pojemnościowe firmy Schoeps; pierwszy o symbolu CCM40 i charakterystyce kardioidalnej, a drugi o symbolu BLM03C i charakterystyce dookólnej.

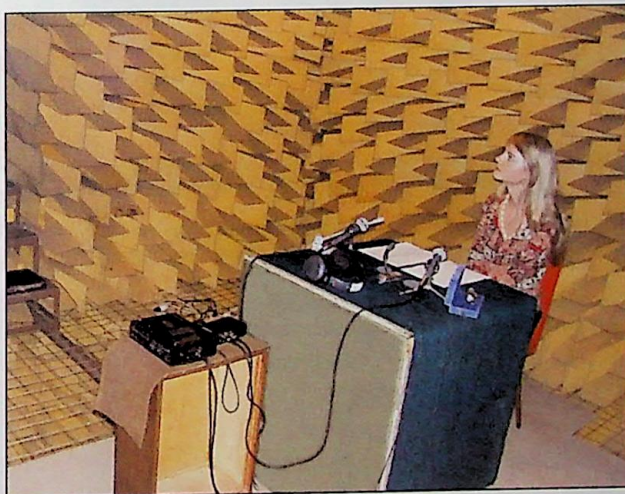
- Telefon komórkowy firmy Siemens, model SL45i (urządzenie nie było połączone z siecią telefonii komórkowej – usunięto kartę SIM). Dane audio zapisywane są w pamięci telefonu, z własną kompresją stratną, do plików w formacie .vmo.

Przebieg eksperymentu

Spiker (mężczyzna, 30 lat) znajdował się w komorze bezchowej, będącej w wyposażeniu Akademii Górniczo-Hutniczej w Krakowie. Właściwości komory: objętość 60 m³, poziom ciśnienia akustycznego (SPL) ok. 16 dB w charakterystyce A. Przed mówcą, w odległości ok. 1 m od ust, na wysiękanym stoliku umieszczono mikrofony podłączone do opisanych

nijny ład”. W celu eliminacji przypadkowych błędów językowych, samogłoski i kwestie zdaniowe powtórzono trzykrotnie. Zgromadzony materiał dźwiękowy został następnie zarchiwizowany na dysku twardym komputera w postaci plików .wav z kodowaniem Windows PCM, o częstotliwości próbkowania 44,1 kHz i rozdzielczości 16 bitów.

Kolejnym etapem eksperymentu była analiza widmowa dla wszystkich nagrań pod kątem wyznaczenia położenia poszczególnych formantów w odpowiadających sobie alofonach. Na wstępie, z wykorzystaniem komputerowego oprogramowania Cool Edit Pro w wersji 2.1. firmy Syntrillium, przeprowadzono analizę spektralną struktury zarejestrowanego sygnału. Badanie wykonano na tym samym dla wszystkich próbek odcinku czasowym, zawierającym ciszę. Następnie przystąpiono do pomiarów sygnału mowy. Przeprowadzono je dwuetapowo. Najpierw za pomocą komputerowego oprogramowania Praat (wersja 4.2.04) zostały zmierzone wartości formantów dla pojedynczych samogłosek oraz dla samogłosek w otoczeniu fonetycznym. Dobór i wartości



Ryc. 2. Widok pomieszczenia badawczego – komory bezchowej
Fig. 2. Examination room

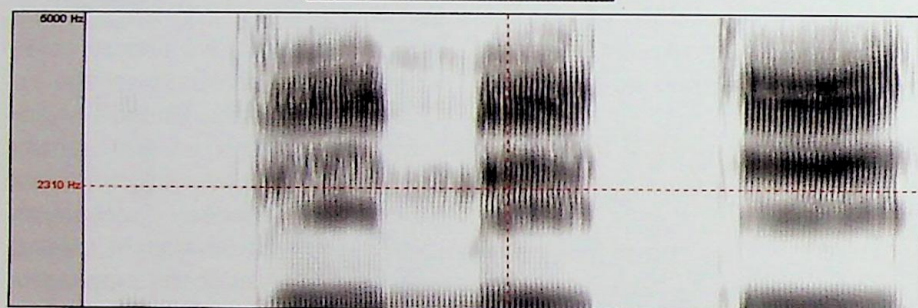
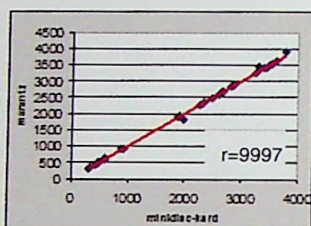
powyżej rejestratorów. Ze względu na poziom generowanych zakłóceń zestaw komputerowy znajdował się na zewnątrz pomieszczenia badawczego.

Eksperyment rozpoczęto od odczytania samogłosek w kolejności: a, e, y, o, u, i. Następnie mówca wypowiedział wcześniej dobrane słowa i zdania: „mama”, „byłem”, „borsuk”, „kino”, „umizgi”, „wymyty”, „bolo”, „bergen”, „lufa”, „gaderypoluki” oraz „Jaka ładna dziś pogoda”, „Duży dom stał daleko”, „Muzyka buduje harmo-

położenia formantów realizowano tzw. sposobem ręcznym. Drugim etapem była analiza spektralna sygnału, a do jej przeprowadzenia wykorzystano wyżej opisane oprogramowanie Cool Edit Pro (ustawienia: FFT-16384, okno Hanninga). W tej części eksperymentu zliczano maksima amplitud zaznaczone uprzednio w tych samych miejscach położenia samogłosek i o tej samej długości czasowej dla wszystkich analizowanych próbek. Uzyskane wyniki badań zestawiono w formie tabelarycznej i na

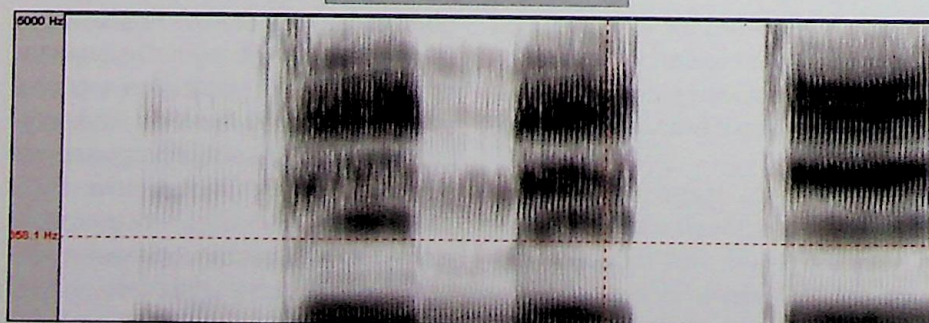
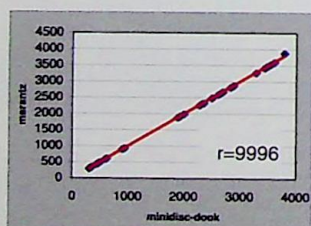


Ryc. 3. Spektrogram wypowiedzi testowej utrwalonej z użyciem rejestratora Marantz
Fig. 3. Spectrogram of test utterance recorded with Marantz device



Ryc. 4. Stopień korelacji wypowiedzi utrwalonych z użyciem rejestratorów Marantz i portadisc (mikrofon kardioidalny). Poniżej spektrogram wypowiedzi testowej utrwalonej z użyciem rejestratora portadisc (mikrofon kardioidalny).

Fig. 4. Correlation of utterances recorded with Marantz devices and portadisc (cardioid microphone). Below: spectrogram of test utterance recorded with portadisc (cardioid microphone)



Ryc. 5. Stopień korelacji wypowiedzi utrwalonych z użyciem rejestratorów Marantz i portadisc (mikrofon dookólny). Poniżej spektrogram wypowiedzi testowej utrwalonej z użyciem rejestratora portadisc (mikrofon dookólny).

Fig. 5. Correlation of utterances recorded with Marantz devices and portadisc (surround microphone). Below: spectrogram of test utterance recorded with portadisc (surround microphone)

ich podstawie przeprowadzono analizę.

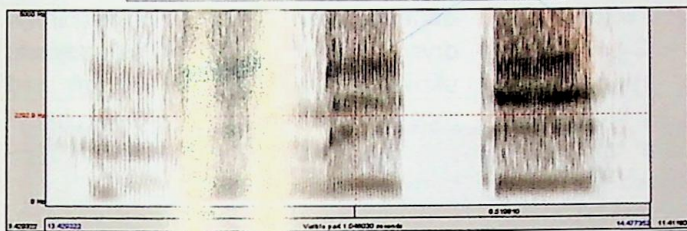
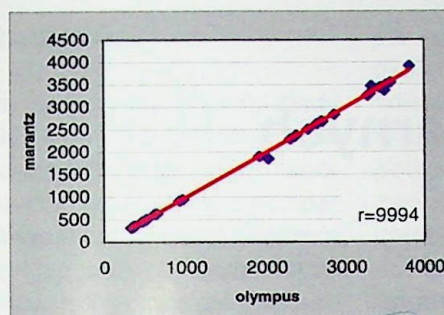
Wyniki i dyskusja

Analiza korelacji położenia formantów została przeprowadzona w odniesieniu do zapisu analogowego, uzyskanego z użyciem magnetofonu firmy Marantz, model CP 430. Prezentowane poniżej wykresy korelacji to odniesienie uzyskanych wyników dla nagrań cyfrowych do pomiarów rozkładu formantów w zapisie analogowym.

Zaprezentowano również spektrogramy przedstawiające rozkład energii w tej samej wypowiedzi, utrwalonej z wykorzystaniem sześciu torów transmisji.

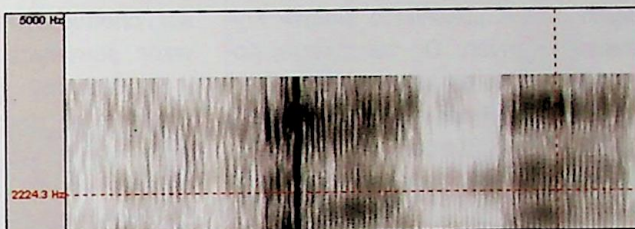
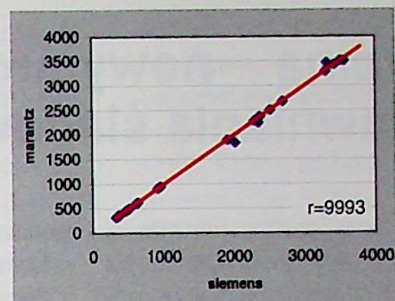
W wyniku analizy położenia formantów każdorazowo uzyskano bardzo wysoki, zbliżony do 1, współczynnik korelacji. Ma to istotny wpływ na praktykę fonoskopijną, gdyż wykorzystanie tak różnych urządzeń rejestrujących nie wprowadziło istotnych różnic w rozkładzie maksimów lokalnych w widmie sygnału mowy, uwzględnianym w procesie identyfikacji mówców. Dla biegłych i techników kryminalistyki konkluzja ta oznacza, że do utrwalenia wypowiedzi porównawczych nie jest wymagane użycie takiego samego rejestratora jak dowodowy. Zakres przeprowadzonego eksperymentu nie uprawnia do wnioskowania o wpływie na położenie formantów transmisji telefonicznej czy innych nietestowanych urządzeń.

Odwwołanie się do praktyki fonoskopijnej pozwala dostrzec, że uzyskana wysoka korelacja położenia formantów w obrębie rejestratorów użytych do eksperymentu jest silniejsza niż w realiach kryminalistycznych. W nagraniach dowodowych dodatkowo pojawiają się bowiem czynniki, które nie wystąpiły w eksperymencie, jak np. odmienny stan emocjonalny mówcy w trakcie wypowiedzi dowodowych i porównawczych,



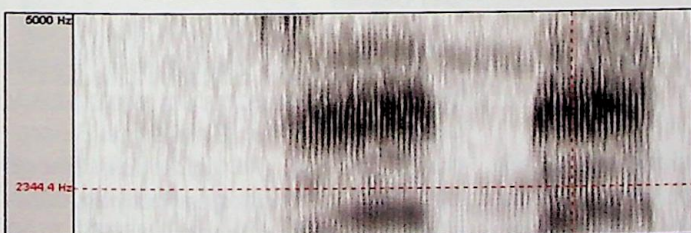
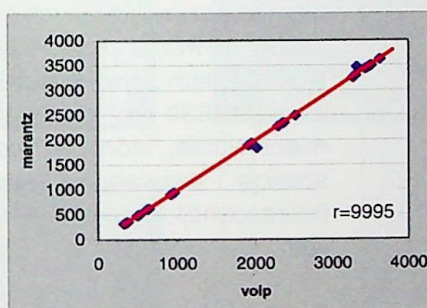
Ryc. 6. Stopień korelacji wypowiedzi utrwalonych z użyciem rejestratorów Marantz i Olympus. Poniżej: spektrogram wypowiedzi testowej utrwalonej z użyciem rejestratora Olympus.

Fig. 6. Correlation of utterances recorded with Marantz and Olympus devices. Below: spectrogram of test utterance recorded with Olympus recorder



Ryc. 7. Stopień korelacji wypowiedzi utrwalonych z użyciem rejestratorów Marantz i Siemens. Poniżej: spektrogram wypowiedzi testowej utrwalonej z użyciem rejestratora Siemens.

Fig. 7. Correlation of utterances recorded with Marantz and Siemens devices. Below: spectrogram of test utterance recorded with Siemens recorder



Ryc. 8. Stopień korelacji wypowiedzi utrwalonych z użyciem rejestratorów Marantz i VoIP. Poniżej: spektrogram wypowiedzi testowej utrwalonej z użyciem rejestratora VoIP.

Fig. 8. Correlation of utterances recorded with Marantz and VoIP devices. Below: spectrogram of test utterance recorded with VoIP recorder

nie zawsze pełna zgodność ciągów alofonów w badanych samogłoskach.

Odrębne zagadnienie stanowi błąd pomiaru położenia formantów. Należy przyjąć, że jest to jeden z czynników, który bez względu na

czenie formantów w widmie sygnału mowy;

– identyfikacja mówców w oparciu o analizę formantową nie wymaga wiernego odwzorowania specyfiki zapisu dowodowego w nagraniu porównawczym;

sposób rejestracji i dobór próbek zawsze wpływa na osłabienie korelacji rozkładu formantów, tak że nie jest ona równa 1.

Wnioski

Na podstawie przeprowadzonych badań sformułowano następujące wnioski:

– sposób zapisu rejestratorów cyfrowych użytych w eksperymencie nie wpływa w istotny sposób na rozmieszczenie

– uogólnienie uzyskanych wyników na nagrania utrwalone z wykorzystaniem transmisji telefonicznej wymaga przeprowadzenia kolejnego eksperymentu.

**Grzegorz Ślusarczyk
Jacek Rzeszotarski
Tadeusz Sikora
Jacek Wierzbicki**
zdj.: autorzy

BIBLIOGRAFIA

1. Blair Benson K.: Audio Engineering Handbook, McGraw-Hill Book Company, 1988.
2. Haines R.: Cyfrowe przetwarzanie dźwięku, Wydawnictwo MIKOM, Warszawa 2002.
3. Jassem W.: Podstawy fonetyki akustycznej, Państwowe Wydawnictwo Naukowe, Warszawa 1973.
4. Kominek M.: Zaczęło się od fonografu, Polskie Wydawnictwo Muzyczne, Kraków 1986.
5. Rzeszotarski J., Witkowska-Pawlak W.: Identyfikacja mówców w obrębie materiału dowodowego, „Problemy Kryminalistyki” 2004, nr 246, s. 15–18.