

**OPINIA**  
**W ZAKRESIE FONETYCZNO-AKUSTYCZNEJ ANALIZY**  
**ZAPISÓW DŹWIĘKOWYCH Z SAMOŁOTU**  
**Tu-154M NR 101**

**na rzecz Zespołu Biegłych w sprawie Po. Śl. 54/10**

**Opracowała:**

**Grażyna Demenko**

---

**WARSZAWA 2015**

## Spis treści

Spis załączników.....	4
1. Uwagi wstępne.....	5
2. Przedmiot i zakres opinii.....	6
3. Założenia i przebieg badań.....	8
3.1. Założenia techniczne .....	8
3.2. Przebieg badań.....	9
3.2.1. Tryb przeprowadzenia analiz .....	9
3.2.2. Odszumianie sygnału.....	11
3.2.3. Parametryzacja i weryfikacja pomiarów .....	12
3.2.4. Dokumentacja analiz.....	12
4. Ogólna analiza treści językowych oraz ogólna weryfikacja poprawności transkrypcji.....	13
4.1. Ogólna analiza.....	14
4.2. Analiza uzupełniająca.....	17
4.2.1. Wypowiedzi nieprzetranskrybowane (ostatnie minuty lotu).....	17
4.2.2. Inne wypowiedzi nieprzetranskrybowane.....	17
5. Ogólna ocena ekspresywności wypowiedzi.....	18
6. Ocena możliwości weryfikacji i rozpoznawania głosów.....	25
6.1. Uwarunkowania rozpoznawania mówcy.....	25
6.1.1. Źródła zmienności wpływające na sygnał mowy.....	25
6.1.2. Kategorie parametrów formujących wektory cech osobniczych.....	27
6.1.3. Ograniczenia związane z materiałem eksperymentalnym.....	29
6.2. Analizy profilu akustycznego mówcy.....	30
6.2.1. Uwagi ogólne. Analizy treści językowych i pozajęzykowych.....	30
6.2.2. Analiza profilu akustycznego określonych mówców.....	30
6.3. Wnioskowanie.....	31
6.3.1. Uwagi ogólne.....	31
6.3.2. Weryfikacja głosów już przypisanych.....	33
6.3.3. Wypowiedzi niezidentyfikowane.....	34
7. Podsumowanie opinii.....	41

Bibliografia .....	43
Załączniki.....	46

## Spis załączników

### Załącznik 1.

1. Dowód 1. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP1-SP9)
2. Dowód 2. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP10-SP14)
3. Dowód 3. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP15-SP17)
4. Dowód 4. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP18-SP22)
5. Dowód 5. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP23-SP25, SP25a)
6. Dowód 6. Tabela 1. Pomiary parametru  $F_0$ , Spektrogramy SP26-SP 27)
7. Dowód 7. Analizy porównawcze (Spektrogram 28, LTA 1-5)
8. Dowód 8. Analizy porównawcze (Spektrogramy SP29, 29a, 30, 31, 32, LT6, LT7, DTW1)
9. Dowód 9. Analizy porównawcze (Spektrogram SP33, LT8 - LT12)
10. Dowód 10. Obliczanie odległości między głosami metoda DTW (DTW2)

### Załącznik 2

Nagrania wykorzystane do analiz w niniejszej opinii umieszczono w załączniku 1

### Załącznik 3

Nagrania dowodowe materiału badawczego

### Załącznik 4

Nagrania dowodowe materiału porównawczego

## 1. Uwagi wstępne

Niniejszą opinię wykonałam na polecenie Prokuratury na podstawie pisma z dnia 31 grudnia 2012 r., w oparciu o dokumentację z Wojskowej Prokuratury Okręgowej w Warszawie z dnia 5.01.2013 r. Otrzymałam wyselekcjonowany materiał dowodowy w postaci:

- 1) Cyfrowych kopii aktów śledztwa (tom 251, 253, 252, 254, 269, 270, 271, 272, 273, 274, 275, 322)
- 2) Raport końcowy Tu-154M, z załącznikiem nr.8 (odpis korespondencji pokładowej)
- 3) Zestawienia transkrypcji ortograficznej wykonanej przez Międzypaństwowy Komitet Lotniczy (MAK) oraz transkrypcji wykonanej przez Centralne Laboratorium Kryminalistyczne CLK (według raportu Tu-154-101 PLF101 z dnia 16-01-2011) oznaczonego w niniejszej opinii skrótem MAK\_CLK.
- 3) Kopie nagrań dowodowych zarchiwizowanych na płytach CD, opisanych, jako dowody o numerach: 6, 12, 25, 26, 29, 30, 27, 28, 31, 33, 34, 9, 10, 11, 35, 36, 37, 15, 16, 17, 18, 19, 20, 21, 22, 23 i 24 oraz zapis typu Verbatim stanowiący kopie dowodowego nagrania.
- 4) Nagrania porównawcze wybranych głosów opisane jako dowody o numerach: 10, 11, 35, 36 oraz 37.

Wstępne analizy wymienionych w pkt. 3) powyżej nagrań dowodowych opisanych jako dowody o numerach 25, 26, 27, 29, 30, dostarczonych mi przez Prokuraturę w dniu 17.06.2013 r. (zrealizowanych z częstotliwością próbkowania 11 KHz) wykazały potrzebę nagrań ponownych z właściwą częstotliwością próbkowania, (co najmniej 22 KHz). Częstotliwość próbkowania 11 KHz jest zdecydowanie niewystarczająca dla analizy sygnału mowy.

Do przeprowadzenia ponownych nagrań powołany został pan mgr Andrzej Artymowicz jako biegły w zakresie akustycznych zapisów sygnału dźwiękowego, który w marcu 2014 r. podczas wyjazdu do Moskwy Zespołu w składzie: płk Karol Kopczyk, płk Robert Latkowski,

prof. Grażyna Demenko, przeprowadził w Moskwie ponowny zapis dźwiękowych na podstawie oryginalnych nagrań z czarnej skrzynki samolotu Tu-154M nr 101 udostępnionych przez prokuraturę rosyjską.

Niniejszą opinię sporządziłam na podstawie nagrań wykonanych w marcu w 2014 r. w Moskwie, w załączniku 3 umieszczono nagrania porównawcze (wymienionych powyżej jako dowody o numerach 10 oraz 11), dalszych nagrań porównawczych wybranych głosów udostępnionych mi przez Zespół Biegłych (ZB) we wrześniu w 2014 r. (umieszczonych w załączniku 4) oraz zestawienia MAK\_CLK zawierającego transkrypcję ortograficzną treści językowych i identyfikacji głosów członków Załogi samolotu Tu-154M nr 101.

Analizę przeprowadziłam wykorzystując odizolowane stanowisko badawcze w odrębnym prywatnym pomieszczeniu. Ochronę danych zapewniło umieszczenie kopii nagrań w prywatnym sejfie i analizy na komputerze odizolowanym od Internetu.

Na podstawie przeglądu literaturowego oraz wyników prac własnych w zakresie transkrypcji ortograficznej i fonetycznej sygnału mowy zaszumionego, przetwarzania mowy ekspresywnej oraz rozpoznawania głosów w szumie (w systemie niezależnym od tekstu i mówcy) opracowałam metodologię badań na potrzeby niniejszej opinii obejmującą założenia dla specyfikacji technicznych, analiz fonetyczno-akustycznych oraz selekcji treści językowych i pozajęzykowych.

## **2. Przedmiot i zakres opinii**

Ocenie poddałam fragmenty wypowiedzi określone wstępnie w piśmie do Prokuratury Wojskowej z dnia 27 października 2012 r. i które zostały bezpośrednio wskazane przez Przewodniczącego Zespołu Biegłych (ZB) po ponownym wykonaniu kopii zapisów akustycznych w Moskwie w marcu w 2014 r. Wstępnie sformułowane przez ZB zadanie ekspertyzy obejmowało:

- 1) ogólną ocenę poprawności transkrypcji ortograficznej według MAK\_CLK oraz przeprowadzenie transkrypcji fragmentów niezidentyfikowanych według MAK\_CLK,
- 2) ustalenie możliwości identyfikacji bądź weryfikacji głosów na podstawie percepcji słuchowej lub analizy akustycznej.

Ponieważ przeprowadzenie relatywnie wiarygodnych badań opartych na percepcji słuchowej związane byłoby z koniecznością uwzględnienia wielu czynników wpływających na odsłuchowy odbiór sygnału mowy m.in.: stopienia znajomości dziedziny lub tematyki,

kompetencji językowej, ogólnej możliwości słuchowego odbioru sygnału akustycznego (np. słyszenia określonych zakresów częstotliwości), metody odsłuchu (np. wąskiej, szerokiej uwzględniającej zróżnicowany kontekst lingwistyczny), technicznych warunków odsłuchu i w związku z tym miałyby charakter w dużej mierze subiektywne, niniejszą opinię opracowałam na podstawie wyników analiz fonetyczno-akustycznych umożliwiających fizyczną parametryzację sygnału, która jest podstawą obiektywnych badań.

Szczególną uwagę poświęciłam zadaniom zasygnalizowanym przez członków Zespołów Biegłych: płk. Ryszarda Michałowskiego, prof. Adama Tarnowskiego podczas kilku spotkań roboczych, które odbyły się w okresie obejmującym ostatnie miesiące 2013 r. do stycznia 2015 r. Te dodatkowe zadania obejmowały:

- 1) analizę ekspresywności wypowiedzi (wydobycia treści pozajęzykowych) dla określenia profilu fonetyczno-akustycznego charakteryzującego poszczególnych mówców na podstawie wypowiedzi wytwarzanych pod wpływem stresu,
- 2) ustalenie możliwości identyfikacji bądź weryfikacji głosów nie rozpoznanych i nieoznaczonych w transkrypcji MAK\_CLK oraz
- 3) ustalenie możliwości transkrypcji fonetycznej kilku oznaczonych wypowiedzi (nieprzetranskrybowanych według dotychczasowej dokumentacji, a zasygnalizowanych podczas spotkań ZB (w grudniu 2014 r. i w styczniu 2015 r.) jako potencjalnie istotnych dla interpretacji czynności wykonywanych przez Załogę w ostatnich minutach lotu (wskazanych w załączniku 1, dowód 5, SP23-Sp25).

W podsumowaniu przedmiotem niniejszej opinii jest:

- 1) Ogólna ocena i fonetyczno-akustyczna weryfikacja transkrypcji wybranych wypowiedzi ustalonych z Przewodniczącym ZB (**punkt 4 opinii**).
- 2) Ocena stylu ekspresywności wybranych wypowiedzi i ich emfazy. Poprzez emfazę (w akcentuacji i frazowaniu) odzwierciedlającą się w strukturze prozodycznej mówcy podkreśla ten element wypowiedzi, który chce przekazać słuchaczowi, jako szczególnie ważny (**punkt 5 opinii**). Ten aspekt badań uznałam, jako istotny, ponieważ warunkuje on poprawną interpretację wyników analiz fonetyczno-akustycznych (tzn. taką interpretację treści pozajęzykowych, która wskazuje na intencję mówcy). Analizie poddałam również fonetyczno-akustyczne cechy głosu charakteryzujące poszczególnych mówców na podstawie wypowiedzi wytwarzanych pod wpływem stresu (w różnych fazach lotu, por. punkt 4 opinii).

- 3) Ocena możliwości identyfikacji i weryfikacji wybranych głosów na drodze fonetyczno-akustycznych analiz (**punkt 6 opinii**). W szczególności podjęłam próbę odpowiedzi na pytania ZB dotyczące oceny możliwości weryfikacji głosów niezidentyfikowanych w raporcie MAK\_CLK i przypisania ich głosom członków Załogi oraz głosom osób spoza Załogi, w szczególności dyr. Mariusza Kazany oraz gen. Andrzeja Błasika.

### 3. Założenia i przebieg badań

#### 3.1. Założenia techniczne

Do badań wykorzystałam najnowsze oprogramowanie umożliwiające efektywne odszumianie sygnału oraz analizy fonetyczno-akustyczne. Z uwagi na problemy ekstrakcji parametrów sygnału mowy w warunkach silnego zaszumienia, trudności efektywnego odszumiania (tzn. niepowodującego uszkodzenia struktury widmowej sygnału mowy), niedoskonałości oprogramowania dedykowanego analizie mowy mocno zaszumionej wykorzystywałam różne możliwe platformy sprzętowo-programistyczne. Niniejszą opinię przygotowałam wykorzystując następujące elementy środowiska sprzętowo-programistycznego.

- **Oprogramowanie do analiz fonetyczno-akustycznych:** Praat, Kay, Wavesurfer, WinPitch
- **Oprogramowanie do przetwarzania sygnału (odszumiania):** Izotope RX3, Goldwave, Audacity
- **Oprogramowanie do weryfikacji mówców: DTW- Dynamic Time Warping** (autorskie własne, wykonane na Politechnice Poznańskiej w 2000 r., autor: prof. Stefan Grochowski, założenia analiz, testy fonetyczne i ocena prof. Grażyna Demenko)
- **Komputery:** MacBookPro (wypożyczony w dniu 7.04.2014 r. przez Instytut Techniczny Wojsk Lotniczych w Warszawie ITWL) oraz prywatny komputer stacjonarny.



### 3.2. Przebieg badań

Podstawowe założenia odnośnie obiektywizacji badań akustycznych związane są z wyborem materiału słownego, rodzajem analizy, metodami przetwarzania sygnału, weryfikacją wyników oraz dokumentacją analiz. Założenia te mają charakter bardzo ogólny, dlatego też na potrzeby niniejszej opinii przedstawiłam bardziej szczegółowo wybrane aspekty badań akustycznych związane z transkrypcją treści językowych oraz identyfikacją i weryfikacją głosów z sygnału mowy wytworzonego w warunkach zaszumienia i specyficznej sytuacji komunikacyjnej (w punkcie 4, 5 oraz 6). Przyjęłam podstawowe założenia odnośnie:

- 1) trybu przeprowadzenia analiz,
  - 2) metody przetwarzania (odszumiania sygnału),
  - 3) weryfikacji pomiarów
- oraz
- 4) sposobu dokumentacji wyników.

#### 3.2.1. Tryb przeprowadzenia analiz

1. Wstępnej, kilkunastokrotnej analizie odsłuchowej poddałam cały materiał (umieszczony w załączniku 3) dla ogólnej oceny treści językowych w celu wyłonienia relatywnie jednorodnych kontekstów komunikacji językowej. Transkrypcje występujące głównie w początkowym fragmencie analizowanego zapisu akustycznego odnoszą się w dużej części do rozmów niezwiązanych z przebiegiem lotu (np. komentarze na tematy ogólne) oraz rozmowy prywatne członków załogi. Wypowiedzi te nie były przedmiotem szczegółowych analiz fonetyczno-akustycznych.

2. Kilkudziesięciokrotną analizę odsłuchową przeprowadziłam natomiast na wybranych, frazach z dalszych fragmentów zapisu sygnału akustycznego odnoszących się bezpośrednio do pilotażu, w celu wstępnego ustalenia możliwości oceny nie tylko treści językowych, ale również pozajęzykowych (stresu, emfazy) na drodze obiektywnych pomiarów parametrów akustycznych (dowody 1-4).

3. W szczególności przyjąłam następujący materiał językowy do analiz fonetyczno-akustycznych:

- a) Materiał określony przez ZB (uzupełniony o dodatkowe analizy fragmentów mowy wskazanych w styczniu w 2015 r. umieszczony w punkcie 4.2).
- b) Materiał wykazujący rozbieżności w istniejących transkrypcjach MAK\_CLK. Szczegółnej weryfikacji poddałam 11 końcowych minut lotu.
- c) Fragmenty wypowiedzi subiektywnie uznane przeze mnie za potencjalnie istotne dla oceny treści językowych i pozajęzykowych bezpośrednio związane z przebiegiem lotu (np. formalne stwierdzenia, komendy lub wypowiedzi wskazujące na emocje, stres). Analizom poddałam również frazy lub ich fragmenty nieprzetranskrybowane lub nieprzetranskrybowane jednoznacznie przez ekspertów (umieszczone w otrzymanych transkrypcjach MAK\_CLK).
- d) Wybrane frazy służące do oceny poprawności zidentyfikowanych przez ekspertów głosów członków Załogi i dodatkowych osób przebywających podczas lotu w kabinie samolotu (na podstawie raportu MAK\_CLK). Jako podstawowe kryteria wyboru przyjąłam uwarunkowania fonetyczno-akustyczne (związane głównie z wyszukaniem określonych kontekstów segmentalnych i suprasegmentalnych). Przykładowo dla porównania struktur formantowych konkretnych głosek, widma sygnału konieczne są te same konteksty fonetyczne (lewo i prawostronne, co najmniej w zakresie głoski poprzedzającej i następującej).
- e) Nagrania porównawcze głosów (gen. Andrzeja Błasika, por. Artura Ziętka, kpt. pil. Arkadiusza Protasiuka, chor. Andrzeja Michalaka, dyr. Mariusza Kazany) otrzymane we wrześniu w 2014 r. Z uwagi na fakt, że nagrania te przeprowadzono nie podczas lotu (w kabinie samolotu), a w innych zróżnicowanych kontekstach sytuacyjnych (np. podczas przemówienia gen. Andrzeja Błasika, rozmów rodzinnych w domu np. por. Artura Ziętka, wywiadu dyr. Mariusza Kazany) to porównania głosów wyżej wymienionych osób z wypowiedziami osób podczas lotu znajdującymi się w kabinie samolotu mogą mieć według obecnego stanu wiedzy charakter wyłącznie pomocniczy (por. punkt 5 opinii).

Jako podstawową metodę analizy przyjąłam metodę fonetyczno-akustyczną bazującą zarówno na cechach segmentalnych (odnoszących się do struktury widmowej określonej

przykładowo wartościami formantów, szerokościami pasm formantowych, rozkładem energii w różnych pasmach częstotliwości), jak i suprasegmentalnych (odnoszących się do struktur prozodycznych utworzonych akustycznie głównie poprzez zmiany tempa, częstotliwości podstawowej, intensywności w obrębie sylaby bądź ciągu sylab). Przeprowadziłam manualnie segmentację fonetyczną badanych fraz (wyzaczyłam przybliżone granice występowania poszczególnych głosek przypisanych określonym fonemom) i suprasegmentalną odnośnie typu konturu intonacyjnego i miejsca wystąpienia akcentu. Poprawność segmentacji (relatywnie dokładnej, tj. w granicach 20 – 45 ms) zapewniła dość żmudna manualna analiza widmowa sygnału. Oczywiście, w niektórych wypowiedziach segmentacja na fonemy była absolutnie niemożliwa (z uwagi na zaszumienie sygnału, fonetyczne redukcje, asymilacje głosek, zmiany intensywności sygnału), w tych przypadkach orientacyjnie wyznaczyłam granice wystąpienia całego wyrazu. W kilku fragmentach nagrań nie było możliwe nawet orientacyjne określenie granic całej wypowiedzi (z uwagi na poziom szumu, nakładające się wypowiedzi, zmienny w dużym zakresie poziom nagrania sygnału mowy, redukcje artykulacyjne wynikające z różnych przyczyn).

### 3.2.2. Odszumianie sygnału

Odszumianie sygnału przeprowadzałam na badanych fragmentach nagrań (rzędu długości frazy, co najmniej jednak w zakresie pojedynczej jednostki akcentowej). Stopień i charakter odszumiania uzależniłam od chwilowego widma szumu estymowanego bezpośrednio przed analizowaną wypowiedzią. Odszumianiu poddawałam również dłuższe ciągi fraz, tak aby uchwycić relacje między zmianami amplitudy w poszczególnych fragmentach kilku wypowiedzi. W wielu przypadkach efekt odszumiania uznałam jako niezadawalający, za nieefektywne uznałam np. odszumianie, które nie dawało możliwości obiektywnego ustalenia przybliżonej struktury widmowej analizowanego całościowego fragmentu mowy lub przynajmniej struktury samogłosek. Niniejszą opinię oparłam na tych fragmentach mowy, które po odszumianiu charakteryzowały się relatywnie czytelną strukturę widmową dającą podstawę segmentacji widmowej i pomiarom parametrów akustycznych.

### 3.2.3. Parametryzacja i weryfikacja pomiarów

Pomiary powtarzałam wielokrotnie dla uzyskania jak najbardziej dokładnej ekstrakcji parametrów przy wykorzystaniu różnych algorytmów (Praat, Wavesurfer, WinPitch).

Dla podstawowej parametryzacji widma przyjąłam współczynniki cepstralne: Mel-frequency cepstral coefficients (MFCC), stosowane obecnie powszechnie w parametryzacji sygnału mowy. Założenia techniczne odnośnie parametryzacji (liczby współczynników, rodzaje okna, filtry, próbkowanie, itp.) ilustrują zrzuty z ekranu oprogramowania DTW (dowód 8 i 10).

Większość pomiarów wykonałam manualnie (np. odczyty częstotliwości podstawowej tzw. parametru F0, formantów na podstawie analizy fourierowskiej), stosowałam również manualne korekty na bazie krótko i długoterminowych analiz widmowych. W przypadku wątpliwości np. co do wiarygodności pomiaru częstotliwości podstawowej na końcu wypowiedzi (z uwagi na niską intensywność sygnału, zniekształcenia), danych tych w analizach fizycznych nie uwzględniałam.

### 3.2.4. Dokumentacja analiz

Dokumentację wyników badań stanowią spektrogramy, tabele pomiarów częstotliwości podstawowej, oscylogramy oraz wyniki pomiarów formantów wybranych parametrów, widma LTAS (Long Term Average Spectrum).

W części opinii odnoszącej się do weryfikacji mówców umieściłam również zrzuty z ekranu ilustrujące wyniki weryfikacji mówcy metodą DTW (Dynamic Time Warping). Oprogramowanie tej metody zostało wykonane na Politechnice Poznańskiej przez prof. Stefana Grocholewskiego i testowane przeze mnie w ramach różnych badań identyfikacyjnych.

Analizowane próbki mowy zanotowałam według standardów ogólnie istniejących i transkrypcji fonetycznej dla języka polskiego SAMPA (opracowanie własne, por. Bibliografia). W niniejszej opinii umieściłam wyłącznie wyniki, które okazały się najbardziej reprezentatywne dla analizowanego problemu. Nie umieściłam w opinii wyników analiz nie wnoszących istotnych nowych informacji (redundantnych z punktu widzenia analizy językowej) oraz niepewnych związanych np. z brakiem możliwości wiarygodnej ekstrakcji parametrów akustycznych spowodowanych silnym zniekształceniem sygnału mowy bądź też niską efektywnością odszumiania (niszczącego strukturę widmową mowy).

W załączniku 2 umieściłam również wybrane akustyczne sygnały, które wykorzystywałam do akustycznych analiz.

#### **4. Ogólna analiza treści językowych oraz ogólna weryfikacja poprawności transkrypcji**

Treści językowe w analizowanym materiale związane są:

a) z wykonywaniem czynności i procedur dotyczących pilotażu,

b) z rozmowami prywatnymi członków Załogi

oraz

c) z elementami dyskursowymi, tzw. operatorami metatekstowymi (grupą wyrazów i wyrażen bez specjalnego znaczenia, niewchodzących w związki składniowe z innymi jednostkami językowymi, lecz tworzących samodzielne struktury składniowe) ułatwiającymi interpretację wypowiedzi, sygnalizującymi kontakt między rozmówcami (np. *no, wiesz, znaczy*); otwierającymi, zamykającymi wypowiedź, wprowadzającymi temat, (przekazującymi często treści parajęzykowe, np. ekspresywne).

Ponieważ na sposób wypowiedzi (w języku technicznym – na jej fonetyczno-akustyczny sposób realizacji) wpływa wiele czynników (por. np. punkt 5 i 6 niniejszej opinii), to konieczna jest interpretacja treści słownych biorąca pod uwagę kontekst sytuacyjny. Istotne jest określenie relatywnie spójnych kontekstów sytuacyjnych (komunikacyjnych – zależnych od określonych uwarunkowań językowych i technicznych związanych np. ze zmieniającym się poziomem szumu) wpływających zarówno na treści językowe, jak i pozajęzykowe. Porównując wypowiedzi dla tej samej osoby należy odnieść się, czy te wypowiedzi były realizowane w względnie stałych kontekstach sytuacyjnych. W niejszym materiale wydzielono 5 faz lotu, istotność wydzielenia tych faz związanych ze zmianą parametrów akustycznych sygnału mowy spowodowanych przykładowo ze stresem (objawiającym się akustycznie m.in. zwiększeniem tempa wypowiedzi i podniesieniem wysokości głosu) omówiono w punkcie 5 opinii.

W punkcie 5 opinii odniosłam się również bardziej szczegółowo do interpretacji wybranych analiz treści językowych i pozajęzykowych.

#### 4.1. Ogólna analiza

Przeprowadziłam analizy fonetyczne całego zapisu akustycznego, jednakże w dowodach 1 – 4 umieściłam tylko wyniki analiz uznane subiektywnie przeze mnie, jako reprezentatywne dla danej fazy lotu. Ocena badanych treści słownych (tj. językowych i pozajęzykowych) oraz analiza ich ekspresywności pozwoliła wydzielić następujące etapy lotu (według kodu czasowego podanego w transkrypcji MAK\_CLK).

##### 1) 8:02:53,5 – 8:11:11,0

W fazie tej (trwającej około 9 min wg początkowych zapisów na taśmie) przeważały głównie rozmowy prywatne członków załogi, swobodne komentarze, żarty oraz wypowiedzi związane ze standardowymi procedurami związanymi z przebiegiem lotu.

Wstępne analizy odsłuchowe wykazują, że zarówno percepcyjnie jak i fonetycznie treści słowne są w tym fragmencie relatywnie czytelne. Również treści pozajęzykowe nie wnoszą informacji, które można byłoby zinterpretować jako nietypowe, szczególnie charakterystyczne. Stopień ekspresywności wypowiedzi można uznać jako nie odbiegający od statystycznych norm (m.in. – tempo wypowiedzi przeciętne, zakres zmian wysokości tonu niewielki, nie przekraczający oktawy), poziom stresu nie mający większego wpływu na artykulację. Jeżeli Prokuratura uzna taką potrzebę, można zweryfikować transkrypcje wypowiedzi dokonane na drodze odsłuchu zawartej w raporcie MAK\_CLK z tej części lotu poprzez odpowiednie analizy fonetyczno-akustyczne. W kilku przypadkach stwierdziłam błędne transkrypcje (np. *Nadwiślanska* zamiast *Franciszkańska*).

##### 2) 8:11:11,5 – 8:26:18,0

Następna faza lotu (trwająca około 15 minut) pozwala wyodrębnić sygnalizowne zaniepokojenie członków Załogi (głównie kpt. pil. Arkadiusza Protasiuka i mjr. pil. Roberta Grzywiny) warunkami pogodowymi, mgłą. Zaniepokojenie to wyraża się językowo (sformułowania bezpośrednie) oraz pozajęzykowo (związane jest ono z akcentowaniem tych fragmentów wypowiedzi, które odnoszą się do bezpieczeństwa lotu (np. wyrazów *wylądujemy*, *mgła*, *tragedii*). W tej fazie lotu padają również

wypowiedzi związane ze standardowymi czynnościami wykonywanymi przez członków Załogi (neutralne językowo i pozajęzykowo).

**Dowód 1:** istniejące transkrypcje ortograficzne rozmów (MAK\_CLK), uzupełniające obiektywne analizy fonetyczno akustyczne niniejszej opinii (spektrogramy 1 – 9).

- SP1) 8:11:11,5 *nie no ziemię widać coś tam widać może nie będzie tragedii*  
SP2) 8:11:49 *ja ci powiem temperaturę... ziiimno*  
SP3) 8:16:42,5 *ale dziesiąta i mgła*  
SP4) 8:17:43,0 *Basiu jest nieciekawie wyszła mgła i nie wiadomo czy wylądujemy...*  
SP5) 8:18:15,5 *a jak nie wylądujemy to co? To odejdziemy*  
SP6) 8:18:36 *afery będzie* (wypowiedź bez transkrypcji według dokumentacji MAK\_CLK)  
SP7) 8:21:22,0 *ty patrz po kierunku - Arek – wysokość po odległości czy czytać?*  
SP8) 8:22:38,5 *i w dół panie Arek*  
SP9) 8:22:42,5 *mały gaz- mały gaz*

Transkrypcja ortograficzna i fonetyczna tych fragmentów nie budzi wątpliwości.

### 3) 8:26:18,5 – 8:30:38,5

Następna faza lotu (trwająca około 5 minut) wyodrębnia formalny meldunek dowódcy o braku możliwości lądowania, stanowi czytelne potwierdzenie oceny ryzyka związanego z lądowaniem, sygnalizuje brak możliwości lądowania i oczekiwanie na decyzje.

**Dowód 2:** istniejące transkrypcje rozmów (MAK\_CLK) uzupełniające obiektywne analizy fonetyczno-akustyczne niniejszej opinii (spektrogramy 10 – 14).

- SP10) 8:26:18,5 *panie dyrektorze – wyszła mgła i w tej chwili i w tych warunkach, które są obecnie nie damy rady osiąść*  
SP11) 8:26:27,5 *spróbujemy podejść – zrobimy jedno zajście- ale prawdopodobnie nic z tego nie będzie. Tak, że proszę już myśleć nad decyzją, co będziemy robili*  
SP12) 8:26:39,5 *paliwa nam tak dużo nie starczy żeby do skutku*  
SP13) 8:26:45 *no to mamy problem* (dalszy ciąg odpowiedź: *możemy pół godziny powisieć i odchodzimy na zapasowe*)  
SP14) 8:30:33 *na razie nie ma decyzji prezydenta, co dalej robimy*

Transkrypcja językowa tych fragmentów nie budzi również żadnych wątpliwości

**4) 8:30:44,5 – 8:33:01**

Faza ta związana jest ze standardowymi przygotowaniem do lądowania i decyzją o próbie lądowania.

**Dowód 3:** istniejące transkrypcje rozmów (MAK\_CLK) uzupełniające obiektywne analizy fonetyczno-akustyczne niniejszej opinii (spektrogramy 15 – 17).

SP15) 8:30:44,5 *do wysokości kręgu, do wysokości kręgu*

SP16) 8:30:59 *do pięćset metrów zejdziemy*

SP17) 8:32:58 *w przypadku nieudanego podejścia odchodzimy w automacie*

Transkrypcja językowa tych fragmentów nie budzi również wątpliwości

**5) 8:33:01,5 – 8:41:07,5**

Faza ta obejmująca końcowe minuty lotu związana jest z wypowiedziami padającymi podczas próby lądowania i ostatecznej katastrofy.

**Dowód 4:** istniejące transkrypcje rozmów ( MAK\_CLK), uzupełniające obiektywne analizy fonetyczno-akustyczne niniejszej opinii (spektrogramy 18 – 22).

SP18) 8:36:48,5 *witamy*

SP19) 8:40:05 *czteryście metrów*

SP20) 8:40:23 *trzysta!*

SP21) 8:40:53 *dyktowanie wysokości: dziewięćdziesiąt, osiemdziesiąt, siedemdziesiąt, sześćdziesiąt, pięćdziesiąt, czterdzieści, trzydzieści, dwadzieścia)*

SP22) 8:41:03,0 *kurwa mać aaa kurwa...*

Transkrypcja językowa tych fragmentów nie budzi również wątpliwości



## 4.2. Analiza uzupełniająca

### 4.2.1. Wypowiedzi nieprzetranskrybowane (ostatnie minuty lotu)

Dodatkowo poddałam szczegółowej analizie wybrane, następujące fragmenty sygnału akustycznego nieprzetranskrybowane w raporcie MAK\_CLK i uznane przez ZB w styczniu 2015 r. jako potencjalnie istotne dla interpretacji sytuacji w ostatnich minutach lotu: 5 sygnałów akustycznych oznaczonych punktami czasowymi: 1) 8:26:37, 2) 8:39:12, 3) 8:39:33,7, 4) 8:40:21,6, 5) 8:40:48,4

#### **Dowód 5:** Uzupełniająca analiza wypowiedzi nieprzetranskrybowanych

Dla następujących sygnałów otrzymano interpretację fonetyczną:

- 1) 8:26:37,0     *będziemy próbować do skutku* (SP23)
- 2) 8:39:12,0     *klapy są pod skrzydłami* (SP24)
- 5) 8:40:48,4     *nic nie widać* (SP25)

Dla sygnału 3 i 4 przeprowadzenie transkrypcji nie było możliwe – nie znaleziono odpowiedniego przypisania fragmentów sygnału określonym ciągom słownym. Sposób odszumiania może prowadzić do zniekształcenia sygnału i również percepcyjnie prowadzić do niewłaściwych przypisań. Fraza *nic nie widać* jest poprzedzona prawdopodobnie tą samą frazą wypowiedzianą przez tego samego mówcę kilka sekund wcześniej (8:39): *nic nie widać* (SP25a), struktura jednakże tej wypowiedzi nie jest czytelna – relatywnie interpretowalne samogłoski to *i* oraz *a*.

### 4.2.2. Inne wypowiedzi nieprzetranskrybowane

Szczegółnej analizie poddałam również inne fragmenty zapisu potencjalnie będących sygnałem mowy. Analizy wykazały, że niektóre fragmenty można przypisać ciągom słownym np. *witamy*, *nic nie widać*, *afera będzie* (spektrogramy SP6, SP18, SP25).

## 5. Ogólna ocena ekspresywności wypowiedzi

Dokonując oceny ekspresywności wypowiedzi należy wziąć pod uwagę aspekty językowe i pozajęzykowe. Treść językowa wypowiedzi ma istotny wpływ na jej strukturę fonetyczno-akustyczną. Duże zróżnicowania w parametrach fizycznych wykazują np. wypowiedzi pochodzące ze spontanicznego dialogu (różny przedmiot rozmów) i wypowiedzi stanowiące np. komendy (bazujące na ściśle określonych tematycznie frazach). Ponadto należy uwzględnić uwarunkowania pozajęzykowe np. narastający stres w kolejnych fazach lotu. Problemy metodologiczne związane z taksonomią stresorów i aspekty automatycznego wykrywania stresu przedstawiono w kilku pracach. Najbardziej przejrzystą klasyfikację stresorów dla aplikacji w automatycznych systemach przetwarzania mowy podaje np. Hansen, wyróżniając stresory: 1) *fizyczne*, 2) *fizjologiczne*, 3) *percepcyjne* oraz 4) *psychologiczne*, mające najważniejszy wpływ na strukturę sygnału mowy (Hansen et al., 2000).

1) **Zerowego stopnia *fizyczne stresory*** obejmują wpływ fizycznych wibracji, zaburzenia układu oddechowego, które mają niekorzystny wpływ na funkcje narządu głosu. Są to efekty najłatwiejsze do zbadania i znalezienia relacji między bodźcem a wywoływanymi przez niego skutkami.

Zerowe stresory z powodu np. wibracji samolotu, mogły mieć na członków Załogi stały wpływ podczas całego lotu. Przykładowo, uwidaczniać się one mogą akustycznie w głosie w zmiennym sposobie artykulacji.

W analizowanym materiale brak jest danych, które mogłyby stanowić podstawę do analizy tego czynnika.

2) **Pierwszego stopnia *fizjologiczne stresory*** obejmują szeroki zakres bodźców wywoływanych środkami farmakologicznymi. Fizjologiczne zmiany w sterowaniu narządem głosu ujawniają się różnymi nietypowymi fonetyczno-akustycznymi strukturami mowy.

W analizowanym materiale brak jest danych, które mogłyby stanowić podstawę do analizy tego czynnika.

**3) Drugiego stopnia *percepcyjne stresory*** związane są z oceną dyskomfortowej sytuacji (np. mowa w szumie, efekt Lombarda, złe warunki toru przenoszącego informację).

Te stresory miały bardzo istotny wpływ na głosy Załogi w zróżnicowanym stopniu w poszczególnych fazach lotu. Głośność wypowiedzi oraz częstotliwość podstawowa wprost proporcjonalnie jest zależna od poziomu hałasu. Czynniki ten w analizowanym materiale można skorelować z poziomem hałasu w kabinie, zależnie od fazy lotu (start, wznoszenie się, zniżanie się samolotu, lądowanie).

**4) Trzeciego stopnia *psychologiczne stresory*** odnoszą się do najszerszego zakresu bodźców, które mogłyby być uważane za niemal nieograniczone i interpretowane w świetle indywidualnych przekonań. Niekontrolowany stan emocjonalny mówiącego spowodowany różnymi czynnikami stanowiącymi zagrożenie dla zdrowia lub mienia, również różnego rodzaju obawy (np. osobisty konflikt z bliską osobą, problemy spowodowane wykonywaną czynnością, przepracowanie), powoduje wytworzenie mowy ekspresywnej, wymagającej znacznie bardziej zaawansowanych metod niż obecnie dostępne tradycyjne, bazujące na mowie mało ekspresywnej, nieafektywnej. Stres wytwarzany w odpowiedzi na zdarzenia postrzegane przez człowieka, jako niezwykle i niemożliwe do opanowania (np. zagrożenie życia) należy do kategorii stresorów o najwyższej wadze (Hansen et al., 2000).

Tego rodzaju czynniki stresogenne mają również najważniejszy wpływ na zaburzenie fonetyczno-akustycznych struktur mowy na wszystkich poziomach językowej i pozajęzykowej analizy.

Bez wątpliwości miały istotny wpływ (uwarunkowany indywidualnie) na strukturę fonetyczno-akustyczną wypowiedzi członków Załogi.

W związku z tym, analizując wypowiedzi z poszczególnych faz lotu należy mieć na uwadze, że w głosach członków Załogi pojawi się od początku lotu zróżnicowany osobniczo wzrastający stres dochodzący do maksymalnego w ostatniej fazie lotu. Akustycznie dość przejrzysto uwidacznia się on w zwiększeniu energii spektralnej w wyższych pasmach częstotliwości, zwiększeniu wartości  $F_{\min}$  oraz intensywności i dynamice wypowiedzi. Te uwarunkowania są ewidentne i relatywnie czytelne w analizach akustycznych. Oprócz uwarunkowań związanych ze stresem szczególną uwagę należy zwrócić na ekspresywność wypowiedzi związaną z podkreśleniem ważnych do przekazania przez mówcę treści (por. spektrogramy: SP1-SP22).

W niniejszej opinii umieszczono, więc analizy ekspresji określonych fragmentów wypowiedzi, akcentu emfatycznego (objawiającego się fizycznie zwiększeniem wartości parametru F0, amplitudy, iloczasu i zmianami struktury widmowej). Analiza akcentu emfatycznego pozwoliła wyodrębnić najważniejsze fragmenty wypowiedzi, które za pomocą środków prozodycznych podkreślali mówcy (dowody 1, 2, 3, 4 oraz 5 przedstawione w początkowym fragmencie punktu 4 niniejszej opinii). Poniżej przedstawiono krótką prozodyczną charakterystykę wybranych wypowiedzi.

### **Dowód 1:**

SP1) *nie no ziemie widać coś tam widać może nie będzie tragedii*

Wypowiedź wskazująca na wątpliwość (kontur intonacyjny z całej frazy wskazujący wahanie i zaniepokojenie). Tempo wypowiedzi dość wolne. Zdecydowane zaakcentowanie wyrazu **tragedii** (bardzo wyraźne istotne zwiększenie energii sygnału, wzrost parametru F0 na sylabie akcentowanej **ge**).

SP2) *ja ci powiem temperature... ziiimno*

Wypowiedź neutralna, fragmentarycznie wskazująca dyskomfort (wyraźne przedłużenie samogłoski *i* w wyrazie *zimno*).

SP3) *ale dziesiąta i mgła*

Wypowiedź wskazująca dość jednoznacznie zdziwienie (zaakcentowany wyraz *dziesiąta* poprzez kontur intonacyjny oraz wyraz *mgła* – wzrost amplitudy, przedłużenie samogłoski *a*, charakterystyczny wzrost częstotliwości podstawowej na wyrazie *mgła*)

SP4) *Basiu jest nieciekawie wyszła mgła i nie wiadomo czy wylądujemy... tak?*

Wypowiedź z wyraźnym zaakcentowaniem wyrazu **mgła** (zinterpretowana jako czytelne stwierdzenie (określone układem akcentów frazowych, wysokim tonem frazowym na wyrazie *mgła*, znacznie poniżej wypowiedziana została fraza *i nie wiadomo czy wylądujemy*).

SP5) *a jak nie wyladujemy to co? To odejdziemy*

Wypowiedź wskazująca wyraźne pytanie mjr. pil. Roberta Grzywny i w odpowiedzi pada zaakcentowany wyraz **odejdziemy** przez kpt. pil. Artura Protasiuka. Ton frazowy wysoki na wyrazach *to co*, następna fraza *to odejdziemy* wypowiedziana również dość wysoko.

SP6) *afera będzie*

Wyraźne, zdecydowane zaakcentowanie wyrazu **afera**

SP7) *ty patrz po kierunku - Arek –wysokość po odległości ci czytać?*

Kontur intonacyjny wskazujący na polecenie (*ty patrz po kierunku* i pytanie – *wysokość po odległości ci czytać?*).

SP8) *i w dół panie Arek*

Zdecydowane zaakcentowanie wyrazu *w dół* (sygnalizowane wzrostem intensywności, częstotliwości podstawowej, iloczasu) oznacza zachętę podjęcie akcji.

SP9) *mały gaz - mały gaz*

Wypowiedzi neutralne ilustrujące różnicowania osobnicze w zakresie tempa wypowiedzi, w przypadku wypowiedzi pierwszej *mały gaz* przez kpt. pil. Artura Protasiuka tempo dwukrotnie szybsze niż wypowiedzi *mały gaz* przez chor. Andrzeja Michalaka.

## **Dowód 2:**

SP10) *panie dyrektorze – wyszła mgła i w tej chwili i w tych warunkach które są obecnie nie damy rady usiąść*

Wyraźne określenie przyczyny braku możliwości lądowania – zaakcentowanie wyrazu *mgła* (zwiększenie energii, częstotliwości podstawowej), również zaakcentowanie frazy *nie damy*. Zdecydowane stwierdzenie faktu.

SP11) *spróbujemy **podejść** – zrobimy jedno zejście – ale prawdopodobnie nic z tego nie będzie. Tak że proszę już myśleć nad decyzją co będziemy robili.*

Zaakcentowanie frazy: *spróbujemy **podejść***, sygnalizujące gotowość podjęcia trudnego zadania. Formalny meldunek: treściowo i prozodycznie sygnalizowany odpowiednimi frazowymi konturami intonacyjnymi).

SP12) *paliwa nam tak dużo nie **starczy** żeby do skutku*

Zdecydowane stwierdzenie niemożliwości tego manewru (wyrazne zaakcentowanie frazy: *nie starczy*).

SP13) *no to **mamy** problem ..... możemy pół godziny powisieć i odejść*

Stwierdzenie zdecydowane (zaakcentowany wyraz **mamy**), wypowiedź dyr. Mariusza Kazany. Odpowiedź kpt. pil. Arkadiusza Protasiuka: *możemy pół godziny powisieć i odchodzimy na zapasowe* ma treściowo i prozodycznie charakter formalny, można zinterpretować ją jako propozycję rozwiązania sytuacji.

SP14) *na razie nie **ma** decyzji prezydenta co **dalej** robimy*

Typowe przekazanie informacji. Zdecydowane zaakcentowanie wyrazu **nie ma** oraz **dalej**. Podkreśla brak decyzji.

### **Dowód 3**

SP15) *do wysokości kręgu – do wysokości kręgu* (kpt. pil. Arkadiusz Protasiuk oraz mjr pil. Robert Grzywna)

Wypowiedzi neutralne kpt. pil. Arkadiusza Protasiuka oraz mjr pil. Roberta Grzywny. Wypowiedzi ilustrujące przekaz informacji. Tempo wypowiedzi dowódcy nadającego informację znacznie szybsze niż drugiego pilota odbierającego informację.

S16) *do **pięćset** metrów zejdziemy*

Bardzo wyraźnie zaakcentowany wyraz **pięćset** przez mjr. pil. Roberta Grzywnę wskazujący tę wartość jako graniczną.

SP17) *w przypadku nieudanego podejścia odchodzimy w automacie*

Zaakcentowany wyraz *w przypadku* (zasygnalizowanie alternatywy) oraz wyrazu *podejścia*. Przejrzysta struktura akcentuacyjna w całej frazie wyraźnie sygnalizująca alternatywny manewr przez kpt. pil. Arkadiusza Protasiuka.

#### **Dowód 4:**

SP18) *witamy* (brak identyfikacji głosu).

Wypowiedź neutralna, raczej rzeczowa, formalna (kontur intonacyjny płaski, tempo dość szybkie, prawidłowo zaakcentowana sylaba *ta*).

SP19) *czterysta metrów*

Wyraźne zaakcentowanie wyrazu *czterysta* przez por. Artura Ziętka (wyraz *metrów* dopowiedziany bez emfazy). Pewność przekazywanej informacji.

SP20) *Trzysta!*

Wyraźne akcentowanie przez por. Artura Ziętka, emfaza sylaby *trzy*. Duża zmiana parametru F0 wskazująca na pewność przekazywanej informacji.

SP21) dyktowanie wysokości *odchodzimy (dziewięćdziesiąt, osiemdziesiąt, siedemdziesiąt, sześćdziesiąt, pięćdziesiąt, czterdzieści, trzydzieści, dwadzieścia)*  
nałożenie wypowiedzi: *odchodzimy*

Wysokości dyktowane bardzo szybko, jedna po drugiej (frazy o długości 600 ms – 100 ms). Ekstremalne zaakcentowanie daje się zauważyć w odczycie przez Artura Ziętka ostatniej wartości: *dwadzieścia!*

SP22) *Kurwa mać aaa kurwa* ( Krzyk )

W tej wypowiedzi (a) wartość parametru F0 wynosi około 280 Hz, we fragmencie (b) wartość parametru F0 dochodzi do 400 Hz (krzyk), we fragmencie (c) obejmującym 2 samogłoski częstotliwość podstawowa waha się w zakresie 360–390 Hz. Z danych literaturowych wynika, że w głosie męskim częstotliwość podstawowa w krzyku może osiągnąć wartość około 260 Hz. W powyższym przykładzie odnotowano zmiany parametru F0 w chwili ekstremalnego pobudzenia ofiary katastrofy przekraczające

4 oktawy. Jakkolwiek wartość parametru F0 około 280 Hz w głosie męskim jest wartością wskazującą na bardzo wysoki stres (w minimalnym stopniu tylko uświadomianym), tak wartość  $F_{0min}$  ponad 360 Hz wskazuje stres absolutnie ekstremalny związany z bezwarunkowymi reakcjami fizjologicznymi (kompletnie niekontrolowanymi przez mowę).

### **Dowód 5:**

SP23) 8:26:37 *będziemy próbować do skutku.*

Wypowiedź neutralna, najprawdopodobniej dyr. Mariusza Kazany. Prawidłowy akcent na frazie: *pró**o**wać* (F0=150Hz). Sugeruje podejmowanie prób.

SP24) 8:39:12 *kłapy są pod skrzydłami*

Neutralne stwierdzenie, dość niski głos ok. 90 Hz, prawidłowy akcent na frazie *skrzydłami*. Mówca o niskim głosie.

SP25) 8: 40: 48,4 *nic nie widać*

Fraza bardzo szybko wypowiedziana, zaakcentowana sylaba *nic*, fraza bardzo niestaranna artykulacyjnie, raczej komentarz skierowany do siebie. Wskazuje na to poziom sygnału, struktura prozodyczna.

SP25a) 8:40:46,5 prawdopodobnie fraza *nic nie widać* wypowiedziana przez tego samego mówcę bezpośrednio wcześniej. Struktura widmowa trudno czytelna.



## 6. Ocena możliwości weryfikacji i rozpoznawania głosów

### 6.1. Uwarunkowania rozpoznawania mowy

Ocena możliwości rozpoznawania mowy musi uwzględniać następujące elementy: (Adami et al., 2003; Aronowitz, 2014; Campbell J, 1997; Doddington et al., 2000; Gonzalez-Rodriguez, 2014; Hansen et al., 2004; Hansen, Varadarajan, 2009; Mary, Yegnanarayana, 2008; Nolan, 1980; Reynolds et al., 2003; Rose, 2006; Shriberg, 2007; Shriberg et al., 2005):

- Źródła zmienności wpływające na sygnał mowy
- Kategorie parametrów formujących wektory cech osobniczych
- Ograniczenia związane z danym materiałem

#### 6.1.1. Źródła zmienności wpływające na sygnał mowy

Analiza zmienności mowy na poziomie (A) *fizjologicznym i neurofizjologicznym*, (B) *psychologicznym i psycholingwistycznym* oraz (C) *lingwistycznym* powinna dostarczyć ważnych informacji o mówcy.

#### A. *Poziom fizjologiczny i neurofizjologiczny*

##### 1. Cechy uwarunkowane fizjologicznie i anatomicznie

- a) *Wzrost, waga, wiek, biomechaniczne cechy narządu mowy, płeć.* Cechy sygnału mowy wynikające z powyższych uwarunkowań są uznawane raczej za relatywnie stabilne, przykładowo wpływ wzrostu, wagi, wieku, płci mówcy na częstotliwość podstawową potwierdziło wielu badaczy, z reguły osoby drobno wzrostu, o drobnej posturze wytwarzają sygnał mowy z wyższą częstotliwością niż osoby wysokie, zwłaszcza korpulentnej
- b) *Nawyki artykulacyjne.* Występują głównie w odniesieniu do wykorzystywania określonego zakresu zmian parametru F0 lub specyficznych wzorców intonacyjnych, ale również ubezdźwięczniania/udźwięczniania, redukcji zbitek spółgłoskowych, charakterystycznej artykulacji niektórych głosek-indywidualne.

Są to czynniki istotne w rozpoznawaniu mówcy, jednakże należy dysponować statystycznie i fonetycznie-akustyczną odpowiednią próbką wypowiedzi, które mogą dać podstawę do obiektywnej oceny określonego nawyku artykulacyjnego.

W analizowanym materiale takiej próbki nie można skonstruować, a wszelkie hipotezy związane z sporadycznym występowaniem określonego nawyku są na poziomie subiektywnych ocen.

## 2. Czynniki fizyczne zewnętrzne, środowiskowe

*Hałas.* Hałas jest uważany za najbardziej istotny czynnik środowiskowy wpływający na wytwarzanie mowy. Aby mówić w hałasie, potrzebny jest dodatkowy wysiłek: mówienie głośniej, wyżej, z bardziej staranną artykulacją. Zmieniają się relacje między sygnałem pobudzającym a charakterystyką toru głosowego. Zmienia się struktura widmowa sygnału (tzw. efekt Lombarda). Różnice pomiędzy mową wytworzoną w środowisku z przeciętnym poziomem hałasu (typowym szumem otoczenia) i w warunkach wysokiego poziomu hałasu (np. przy szumie samochodu, samolotu, maszyny budowlanej) związane są ze wzrostem częstotliwości, przesunięciem energii z pasma niskich częstotliwości w kierunku pasm częstotliwości średnich i wysokich, wzrostem natężenia dźwięku, wzrost czasu trwania samogłosek, nachylenie spektrum (ang. *Spectral tilting*). Dla uwzględnienia tego czynnika konieczne jest porównywanie wypowiedzi wytworzonych w różnorodnych warunkach akustycznych oraz wypowiedzi zlokalizowanych blisko siebie, wytworzonych w podobnych warunkach akustycznych (Ikeno, Hansen, 2007)

## 3. Uwarunkowania sytuacyjne, techniczne

Fizyczna odległość między mówcą a słuchaczem, ewentualnie mikrofonem, bezpośrednio wpływa na podwyższenie poziomu głosu mówcy. Czynnikiem ten również związany jest z efektem Lombarda (opisanego punkt wyżej) i ma charakter uniwersalny, a jego wpływ na strukturę mowy jest dobrze udokumentowany w literaturze.

## **B. Poziom psychologiczny i psycholingwistyczny**

Obiektywna ocena zmian w danym głosie może być wykonywana, jeśli nagrania wypowiedzi tego samego mówcy, realizowanych zarówno w warunkach stresowych, jak bezstresowych, są dostępne (Banziger, Scherer, 2005; Koolagudi, Rao, 2012; Lee et al., 2001; Scherer et al., 2001; Schuller et al., 2005; Schuller et al., 2004; Squartini et al., 2012). Na tym poziomie należy również uwzględnić takie czynniki jak relacje mówcy przekazującego wiadomość do rozmówcy odbierającego komunikat, lingwistyczny kontekst komunikacyjny, cechy osobowościowe rozmówców. Próbką mowy powinna być reprezentatywna statystycznie i fonetyczno-akustycznie, uwzględniać możliwie jak najwierniej powyższe uwarunkowania.

Dokonując oceny możliwości rozpoznania głosów, będącej przedmiotem niniejszej opinii, zwrócono uwagę na wymienione powyżej źródła zmienności i uwzględniono te czynniki, które były możliwe do analizy.

### 6.1.2. Kategorie parametrów formujących wektory cech osobniczych

W charakterystyce i rozpoznawaniu mówcy przydatny jest także kompleksowy zbiór różnego typu parametrów: (1) długo- i krótkoterminowych akustycznych, uwarunkowanych fizjologicznie (poprzez anatomię i fizjologię narządu głosu), (2) zdeterminowanych fonetycznie (na podstawie czynników językowych i pozajęzykowych), (3) gramatycznych (określonych regułami gramatyki dla określonego języka i mówcy), (4) wizualnych cech wynikających z analizy obrazu, Demenko, 2015).

#### 1) Parametry akustyczne uwarunkowane fizjologicznie (nielingwistycznie):

- a) szybkość mówienia (liczba słów na oddechu, liczba słów na minutę),
- b) sposób oddychania (podczas mowy, spoczynku),
- c) poziom wysokości głosu (przeciętna częstotliwość podstawowa, zakres zmian),
- d) poziom głośności (zakres zmian w mowie i szeptcie),
- e) dynamika głośności (zmiany intensywności, tonu i rytmu),
- f) staranność artykulacji (ewentualne redukcje sygnału),
- g) rodzaje i rozkład zaburzeń płynności (lokalizacje w wypowiedzi, wzajemne lokalizacje),

- h) barwa głosu i jej zmiany (w kategoriach subiektywnych i obiektywnych),
- i) cechy tonu krtaniowego (współczynniki kształtu, częstotliwość, cykl zamknięcia i otwarcia głośni).

Potencjalnie wszystkie te cechy mogłyby być wykorzystane w rozpoznawaniu głosów w niniejszych badaniach. Niektóre jednakże są niedostępne np. parametry związane z analizą mikroprozodii wymagającą szczegółowego precyzyjnego pomiaru częstotliwości podstawowej.

2) Parametry akustyczne uwarunkowane fonetycznie:

- a) rytm,
- b) melodia wypowiedzi, jej wariantywność, typowe przebiegi, prominencje (realizacja akcentów, ich zróżnicowanie itd.),
- c) pauzy ciche, pauzy wypełnione (również ich realizacja – „zawartość” akustyczna, role w wypowiedzi), zawahania, powtórzenia itd., rozkład zaburzeń płynności (jako cecha idiosynkratyczna, również w powiązaniu z budową wypowiedzi na poziomie słownictwa i składni itd.),
- d) skład widmowy,
- e) mikroprozodia,
- f) różne zjawiska fonetyczne (np. redukcje, ubezdźwięcznienia, koartykulacja, upodobnienia).

Potencjalnie wszystkie te cechy mogłyby być wykorzystane w rozpoznawaniu głosów w niniejszych badaniach, praktycznie jednak nie możliwe do wykorzystania z uwagi na ograniczenia opisane poniżej w punkcie 3 niniejszej opinii.

3) Cechy gramatyczne

- a) zawartość leksykalna wypowiedzi, charakter stosowanego słownictwa, profile frekwencyjne, cechy składniowe,
- b) typowe konstrukcje składniowe, powtarzalne błędy,
- c) struktury semantyczne (sposób przekazywania informacji/znaczenia),
- d) charakterystyki frazy (przeciętna długość, budowa),

e) łączne specyfikacje („profil”) wypowiedzi na poziomie leksykalnym, frazeologicznym i składniowym.

Z uwagi na procedury związane z określoną, narzuconą strukturą gramatyczną i leksykalną głównie wypowiedzi Załogi, te cechy praktycznie nie mogły być uwzględnione w rozpoznawaniu głosów w niniejszych badaniach.

Z powodów wymienionych w punkcie 6.1.3 (ograniczenia związane z materiałem eksperymentalnym) analizowano w praktyce na potrzeby niniejszej opinii tylko wybrane, dostępne parametry: zmienność częstotliwości podstawowej, iloczas, intensywność, wartość i szerokość formantów, tempo wypowiedzi, cechy spektralne.

#### 6.1.3. Ograniczenia związane z materiałem eksperymentalnym

Negatywne czynniki, które wpływały na przygotowanie reprezentatywnego materiału i wykonanie na potrzeby niniejszej opinii obiektywnych pomiarów, pozwalających na przeprowadzenie prawidłowego wnioskowania są następujące:

- a) Zła jakość nagrania związana ze specyfiką odczytu z taśm magnetycznych.
- b) Wysoki stopień szumu o zmiennym nasileniu (zależnie od fazy lotu).
- c) Różne warunki komunikacji słownej. Porównywanie wypowiedzi osób z Załogi oraz pasażerów podczas lotu z ich wypowiedziami nagramymi w innych warunkach akustycznych nie daje pewności właściwego wniosku (może zmieniać się np. styl wypowiedzi).
- d) Bardzo ograniczone konteksty fonetyczno-akustyczne, liczebność próbek jest niewielka (krótkie frazy, często jedno- dwu- trzy- czterosylabowe).
- e) Różne warunki sytuacyjne – odległości mówców od mikrofonów (może zmieniać się wysiłek głosowy wpływający na artykulację).
- f) Spontaniczny charakter wypowiedzi – wypowiedzi się nakładają, często mają charakter bardzo ekspresywny.

## 6.2. Analizy profilu akustycznego mówcy

### 6.2.1. Uwagi ogólne. Analizy treści językowych i pozajęzykowych

Analizy spektrograficzne w dowodach ilustrują:

- a) treści językowe zapisane według transkrypcji ortograficznej,
- b) treści pozajęzykowe (stopień emfazy określonych fragmentów wypowiedzi, zilustrowany zmianami iloczasu, amplitudy i parametru F0).

Oszacowania tempa mowy dokonano na podstawie liczby sylab na sekundę (jako jednej z możliwych miar szybkości artykulacji). W niektórych analizach podpisano wynik transkrypcji ortograficznie, w innych fonetycznie (z uwagi na wybrane elementy wymowy). W kilku przypadkach analizowano widmo długoterminowe LTAS. W przypadku badania cech osobniczych zbior parametrów dla oprogramowania DTW określono na podstawie istniejących standardów. Wybrane analizy umieszczono w załączniku 1.

### 6.2.2. Analiza profilu akustycznego określonych mówców

Przeprowadzono ocenę profilu akustycznego zidentyfikowanych głosów (na podstawie dokumentacji MAK\_CLK).

#### **Dowód 5. Wypowiedzi zidentyfikowane**

W tabeli nr 1 (zawartej w załączniku 1) umieszczono poglądowo fragmenty pomiaru częstotliwości podstawowej z wybranych fraz (zapewniających wiarygodność pomiaru). Poniżej podano krótką charakterystykę wybranych głosów.

Osoby z Załogi:

- kpt. pil. Arkadiusz Protasiuk

Bardzo wysoki głos (prawdopodobnie uwarunkowany posturą mówcy). Tempo wypowiedzi szybkie. Styl wypowiedzi zdecydowany. Przeciętny zakres zmian parametru F0 (120–260 Hz).

- mjr pil. Robert Grzywna  
Również stosunkowo wysoki głos. Tempo wypowiedzi stosunkowo szybkie. Styl wypowiedzi w większości przypadków zdecydowany. Przeciętny zakres zmian parametru F0 (90–240 Hz).
- por. Artur Ziętek  
Głos dość wysoki. Zmienny zakres parametru F0. Styl wypowiedzi związany z podkreśleniem wagi odczytu pomiaru, zwłaszcza odległości (emfaza, korekty). Przeciętny zakres zmian parametru F0 (90–220 Hz).
- chor. Andrzej Michalak  
Głos relatywnie niski. Mało próbek wypowiedzi. Głos raczej monotony (90–140 Hz), tempo dość wolne.

Wypowiedzi scharakteryzowane na podstawie nagrań porównawczych (osoby nie z Załogi):

- gen. Andrzej Błasik  
Niski głos. Na podstawie analizy nagrania porównawczego (SP26). Przeciętny zakres zmian parametru F0 (70–180 Hz).
- dyr. Mariusz Kazana  
Przeciętny zakres zmian parametru F0 (100–180 Hz). Na podstawie nagrania porównawczego (SP27).

## 6.3. Wnioskowanie

### 6.3.1. Uwagi ogólne

System rozpoznawania mówcy ma dwie możliwości – weryfikacji mówcy lub jego identyfikacji. W systemie weryfikacji, mówca musi najpierw podać swoją tożsamość, a następnie system sprawdza czy jest ona prawidłowo rozpoznana. W przypadku weryfikacji system zwraca odpowiedź TAK/NIE – wejściowy sygnał głosowy jest porównywany

do modelu powiązanego z tożsamością weryfikowanego użytkownika. Oznacza to, że system rozpoznający mowę wykonuje jedno porównanie, a następnie podejmuje jedną decyzję bazującą na wyniku tego porównania.

W systemie identyfikacji, mówca nie musi udowadniać tożsamości. Zamiast tego, gdy mówca wprowadza sygnał głosowy do systemu, algorytm decyduje, który model mówcy jest najbardziej zbliżony do sygnału wejściowego. W tym przypadku system przeprowadza  $N$  porównań, gdzie  $N$  jest liczbą modeli mówców w bazie systemu. Każde porównanie tworzy ocenę wiarygodności, z których system może wybrać model powiązany z tożsamością najbardziej prawdopodobnego mówcy. Przykładem decyzji może być „mówca  $i$ ”, gdzie  $i=1 \dots N$ .

Oczywistym jest, że problem identyfikacji mówcy jest bardziej złożony niż jego weryfikacji, dlatego też wyniki uzyskane w systemach identyfikacji mówcy są z reguły gorsze niż w weryfikacji.

Należy zwrócić uwagę, iż przeprowadzenie na potrzeby niniejszej opinii identyfikacji głosów na drodze analiz fonetyczno-akustycznych z uwagi na uwarunkowania podane w punktach wcześniejszych niniejszej opinii (w szczególności z uwagi na brak dostępu do nagrań głosów wszystkich pasażerów samolotu) jest NIEMOŻLIWE. **W opinii zajmuję się, więc wyłącznie problemem weryfikacji** (tzn. problemem przypisania niezidentyfikowanych wypowiedzi do zbioru głosów Załogi oraz wskazanych przez ZB głosów gen. Andrzeja Błasika oraz dyr. Mariusza Kazany).

Z uwagi na dalsze uwarunkowania podane w punktach wcześniejszych opinii (np. związane z technicznymi trudnościami akustycznej analizy, ograniczeniem materiału słownego) ograniczę się do wskazania bądź wykluczenia określonego głosu z potencjalnie statystycznym istotnym prawdopodobieństwem (tzn. wynikającym z analiz, co najmniej kilku parametrów fonetyczno-akustycznych).

Do tego celu optymalne jest wykorzystanie metody tzw. Dynamic Time Warping (DTW), wymagającej jednakże, aby analizowane sygnały zawierały taką samą treść lingwistyczną. Rozpoznawanie polega na specyficznym porównywaniu wektora parametrów, opisujących sygnał (ze zbiorem takich wektorów będących wektorami wzorcowymi – opisującymi wcześniej nagrane wypowiedzi).

Ponieważ dany wyraz można wypowiedzieć wolniej lub szybciej, proste porównanie sygnałów, ramka do ramki, nie przyniosłoby oczekiwanych efektów. Dlatego stosuje się specjalny algorytm wykorzystujący programowanie dynamiczne, pozwalające znajdować najlepsze dopasowanie pomiędzy dwoma rozciągniętymi w czasie sygnałami. Dopasowanie



to polega na „zawijaniu” miejsc rozciągniętych w czasie w stosunku do wzorca i rozciąganiu miejsc, które są nazbyt skrócone - jest to znajdowanie korespondujących ze sobą obszarów pomiędzy dwiema seriami danych. Algorytm DTW znajduje optymalną ścieżkę „zawinięcia czasu”, tak, aby porównywane dwa sygnały miały najmniejszą odległość względem siebie (różniły się najmniej).

W podejściu tym należy wyliczyć macierz odległości (w sensie Euklidesa) między każdą ramką sygnału badanego, a ramką sygnału wzorca. Następnie trzeba wyliczyć macierz skumulowanej odległości między sygnałami poruszając się od elementu macierzy oznaczający odległość pierwszych ramek sygnałów, potem dodając najmniejszy z sąsiednich elementów macierzy, potem najmniejszy z elementów sąsiednich właśnie dodanego elementu itd. Powstaje ścieżka optymalnego dopasowania, otrzymujemy także minimalną odległość między sygnałem badanym, a danym wzorcem. Następnie za pomocą DTW należy obliczyć odległości między wszystkimi wzorcami, a badanym sygnałem. Sygnał klasyfikujemy, jako ten, do wzorca, którego, odległość jest minimalna. Dodatkowo, można założyć pewien próg, poniżej którego ta minimalna odległość musi zejść, by uznać, że wypowiedzany sygnał, to jeden z tych, które znajdują się w bazie.

Z uwagi na ograniczenia materiału badawczego wymienione w punkcie 3 niniejszej opinii, przeprowadzona w niniejszej opinii analiza cech osobniczych może mieć więc wyłącznie charakter poszlakowy, ponieważ nie ma możliwości oceny statystycznej istotności stwierdzonych różnic w parametrach, a zwłaszcza ich weryfikacji.

### 6.3.2. Weryfikacja głosów już przypisanych

Analizy fonetyczno-akustyczne nie dały podstawy do zakwestionowania przypisań w zakresie identyfikacji mówcy, przeprowadzonych przez ekspertów na drodze analiz odsłuchowych i umieszczonych w aktualnych transkrypcjach (MAK\_CLK): tzn. wyniki analiz wykazały, iż parametry wyekstrahowane z wybranych wypowiedzi z głosów przypisanych kpt. pil. Arkadiuszowi Protasiukowi, mjr. pil. Robertowi Grzywnie, por. Arturowi Ziętkowi, chor. Andrzejowi Michalakowi i dyr. Mariuszowi Kazanie tworzą fonetyczno-akustycznie uzasadnione wektory cech osobniczych. Nie było, więc żadnych podstaw, żeby obiektywnie kwestionować przypisania dokonane przez ekspertów, umieszczone w transkrypcji (MAK\_CLK).

### 6.3.3. Wypowiedzi niezidentyfikowane

Wnioskowanemu poddano również wypowiedzi głosów niezidentyfikowanych na drodze percepcyjnej przez ekspertów (były to wypowiedzi między innymi: *dwieście trzydzieści metrów, sto metrów, nic nie widać*).

W szczególności postawiono trzy pytania.

1. Czy **niezidentyfikowane** wypowiedzi podające w ostatnich 11 minutach lotu należą do tej samej osoby?

To jest problem nieposiadający jednoznacznego rozwiązania. Padające wypowiedzi niezidentyfikowanych mówców mają różną strukturę fonetyczno-akustyczną (por. punkt 4 niniejszej opinii) nie można dokonać, więc adekwatnych dla badań porównawczych pomiarów formantów, czy współczynników cepstralnych, ponadto są bardzo mocno zaszumione (ich odszumianie prowadzi do zniekształcenia sygnału mowy). Należy jednakże stwierdzić, że **nie można wykluczyć**, iż niektóre wypowiedzi mogły należeć do tej samej osoby. Przykładowo: wyraz *metrów* (pojawiający się we frazie *dwieście trzydzieści metrów* oraz we frazie *sto metrów* mógł należeć z bardzo wysokim prawdopodobieństwem do tej samej osoby. Wskazuje na to podobna struktura widmowa tych wypowiedzi (por. spektrogram SP28). Frazy *dwieście trzydzieści metrów* oraz *sto metrów* to frazy informacyjne, wypowiedziane pewnie z przekonaniem o prawidłowości odczytu. Wskazuje na to właściwa akcentuacja *dwieście trzydzieści metrów* oraz *sto metrów*, brak nieciągłości w wypowiedziach, iloczasy sylabiczny w normie mowy płynnej. Szczegółowo podobieństwo na fragmentach frazy *metru* z obu wypowiedzi ilustrują widma LTAS: LT1 oraz LT2. Aby odpowiedzieć, czy fraza *nic nie widać* (również niezidentyfikowany głos, spektrogram SP25) została zrealizowana przez mówcę, który wypowiedział frazę *dwieście trzydzieści metrów* oraz mówcę, który wypowiedział frazę *sto metrów* przeanalizowano LTAS na samogłosce *e* (było to jedyne możliwe porównanie, przy założeniu relatywnego ustalonego kontekstu fonetycznego) z fraz: *dwieście trzydzieści metrów*, *sto metrów* oraz *nic nie widać* (widma LT3, LT4, LT5). Zauważyć można istotne podobieństwa struktury widmowej samogłoski *e* w analizowanych frazach LT3 oraz LT4. Widmo samogłoski *e* z frazy *nic nie widać* wykazuje podobieństwa i różnice, jej przypisanie z relatywnie wysokim stopniem prawdopodobieństwa określonemu mówcy wiąże się bardziej ze spekulacją niż dowodem

naukowym. Brak jest możliwości bardziej szczegółowego porównania cech widmowych z cechami widmowymi w wypowiedziach *dwieście trzydzieści metrów* oraz *sto metrów*, zwłaszcza zweryfikowanej oceny statystycznej podobieństwa.

### **Dowód 7:**

Spektrogramy SP28 oraz widma LT1, LT2, LT3, LT4, LT5.

2. Czy *niezidentyfikowane* wypowiedzi podające w ostatnich 11 minutach lotu mogły należeć do członków załogi?

Niektóre wypowiedzi ze stopniem prawdopodobieństwa graniczącym z pewnością nie należały do członków Załogi z powodu różnego typu ograniczeń (fizjologicznych i fizycznych).

Przykładowo frazy: *sto metrów* i *sto* zostały wypowiedziane przez 2 różne osoby. Cechy prozodyczne, artykulacyjne tych wypowiedzi są zupełnie odrębne (pierwsze *sto metrów* wypowiedziane powoli, z dalszego planu akustycznego, z niską częstotliwością podstawową, drugie *sto* padające kilkaset milisekund później ma wysoką amplitudę, inny wzorzec intonacyjny. Jest NIEMOŻLIWE, aby mówca w tak krótkim interwale czasowym zdołał przestawić ustawienia swojego toru artykulacyjnego (i powtórzył wypowiedź *sto* kilkaset milisekund później z odmienną strukturą fonetyczno-akustyczną. Wypowiedź *sto* została zrealizowana przez por. Artura Ziętka (według transkrypcji MAK\_CLK). Nie można, więc z całą pewnością wypowiedzi frazy *sto metrów* jemu przypisać.

Również, praktycznie zerowe prawdopodobieństwo istnieje, że frazę *sto metrów* wypowiedział kpt. pil. Arkadiusz Protasiuk lub mjr pil. Robert Grzywna. Głos kpt. pil. Arkadiusz Protasiuka charakteryzuje się znacznie wyższą częstotliwością podstawową niż głos osoby, która wypowiedziała *sto metrów*. Praktycznie jest nieprawdopodobne, aby nagle wystąpiło obniżenie częstotliwości podstawowej w głosie kpt. pil. Arkadiusza Protasiuka (pod koniec lotu można byłoby się spodziewać raczej jeszcze podwyższenia  $F_{\min}$  (związanego ze stresem ekstremalnie występującym w tej fazie lotu) niż jego obniżenia.  $F_{\min}$  jest silną cechą osobniczą (dość odporną na szum) i w tym przypadku należy głos kpt. pil. Arkadiusza Protasiuka wykluczyć.

W przypadku analizy głosu mjr. pil. Roberta Grzywiny można również przyjąć, że prawdopodobieństwo wypowiedzenia przez niego frazy *sto metrów* jest praktycznie zerowe. Wskazuje na to kilka czynników:

- a) Fraza *dwa-osiem-zero* została wypowiedziana (przez mjr. pil. Roberta Grzywnę, według transkrypcji MAK\_CLK) z bliższego planu akustycznego (na co wskazuje wyraźna wyższa intensywność sygnału) niż następująca około 4 sekundy dalej fraza *dwieście trzydzieści metrów* oraz o około 20 sekund dalej fraza *sto metrów* (o jeszcze niższej intensywności).
- b) Dodatkowo tempo wypowiedzi frazy *dwa – osiem - zero* zrealizowanej przez mjr. pil. Roberta Grzywnę jest znacznie szybsze (praktycznie około 2 razy szybsze) niż dwóch następujących wypowiedzi (wypowiedź *dwa - osiem - zero* została zrealizowana z tempem około 150 ms/na sylabę, wypowiedź *dwieście trzydzieści metrów* z tempem około 270 ms/na sylabę, wypowiedź *sto metrów* z tempem około 290 ms/na sylabę, por. spektrogram SP29).
- c) Widma LTAS fragmentów fraz *metru* uzyskane dla fraz *dwieście trzydzieści metrów* (LT1) i *sto metrów* (LT2) oraz fragmentu *metru* (LT7) z frazy *pięciuset metrów* (wypowiedzianej przez mjr. pil. Roberta Grzywnę w 8:30:59 minucie lotu) znacznie różnią się (por. również spektrogram SP32). Analizowano fragmenty */metru/* a nie */metruf/* z uwagi na fakt, że spółgłoska */ff/* była silnie zredukowana artykulacyjnie i dodatkowo zaszumiona.

Dodatkowo, dla porównania umieszczono widmo z fragmentu *metru* (LT6) uzyskane z poprzedniej wypowiedzi por. Artura Ziętka.

- d) Dodatkowo obliczono odległości między sylabami: 1) *me* (z głosu niezidentyfikowanego z frazy *sto metrów*, oznaczonego, jako X\_me (100), 2) *me* (z głosu niezidentyfikowanego z frazy *dwieście trzydzieści metrów*, oznaczonego, jako X\_me (230), 3) *me* (z wypowiedzi mjr. pil. Roberta Grzywiny z frazy *pięciuset metrów*, oznaczonego jako S\_me), 4) *me* z wypowiedzi por. Artura Ziętka (z frazy *pięćset*, oznaczonego jako N\_me). Wyniki DTW wskazują na odległość między wypowiedzią X\_me (100) a wypowiedzią X\_me (230) równą 42,8. Odległość X\_me (100) od próbki N\_me=92,6, odległość X\_me (100) od próbki S\_me wynosi 100,33.

Nie umieszczono w zbiorze wypowiedzi porównywanych frazy *me* dla kpt. pil. Arkadiusza Protasiuka, ponieważ nie było dla tej sylaby właściwego kontekstu fonetycznego z wypowiedzi tego mowcy.

W związku z powyżej wymienionymi różnicami, prawdopodobieństwo, że frazę *sto metrów* zrealizował mjr pil. Robert Grzywna należy uznać jako nikłe, praktycznie bliskie zeru.

Nie możliwe było dalsze jeszcze bardziej szczegółowe porównanie innych wypowiedzi mjr. pil. Roberta Grzywny i kpt. pil. Arkadiusza Protasiuka z wypowiedziami nieprzypisanych określonym członkom Załogi, głównie z powodu braku znalezienia odpowiednich kontekstów fonetyczno-akustycznych.

Jednakże biorąc pod uwagę wszystkie wyżej wymienione czynniki, prawdopodobieństwo, że w głosie mjr. pil. Roberta Grzywny drastycznie zmieniło się tempo, barwa, częstotliwość podstawowa, amplituda w ciągu 4 sekund należy uznać praktycznie (z przyczyn niewytłumaczalnych) jako zerowe. Potencjalnie można by raczej oczekiwać zmian związanych z podwyższeniem tempa wypowiedzi, zwiększeniem wartości parametru  $F_{\min}$ , amplitudy sygnału wywołanych wzrastającym stresem. Należy zaznaczyć, iż tempo wypowiedzi również por. Artura Ziętki jest stosunkowo szybkie (por. SP30 i SP31).

Również nikłe prawdopodobieństwo istnieje, że frazę *sto metrów* wypowiedział dyr. Mariusz Kazana. Jak wskazuje analiza wypowiedzi *no to mamy problem*, głos dyr. Mariusza Kazany ma dość znacznie wyższą częstotliwość podstawową niż mówca wypowiadający frazę *sto metrów*.

### **Dowód 8**

- Spektrogram SP29, ilustruje wypowiedzi:

*dwa osiem zero, trzysta!, dwieście trzydzieści metrów, dwieście pięćdziesiąt*

Wypowiedź: *dwa osiem zero*, zrealizowana została z szybkim tempem w sposób wyraźnie akcentowany (przez mjr. pil. Roberta Grzywnę), wypowiedź: *trzysta!*, (przez por. Artura Ziętki) również wyraźnie została zaakcentowana, wypowiedź: *dwieście trzydzieści metrów* z wolnym tempem, monotonicznie została zrealizowana przez niezidentyfikowanego mówcę, wypowiedź: *dwieście pięćdziesiąt* (przez por. Artura Ziętki) koryguje poprzednie *dwieście trzydzieści* (po wypowiedzi *dwieście* – zawahanie – podniesienie głosu, krótka przerwa i stwierdzenie *pięćdziesiąt*).

- Spektrogramy SP30 i SP31, ilustrują wypowiedzi:

*Dwieście, sto pięćdziesiąt, sto metrów, sto!* w dłuższej frazie.

Niezidentyfikowany mówca wypowiedzi: *sto metrów* zrealizował tą frazę również z wolnym tempem, monotonicznie.

- Spektrogram SP32, ilustruje wypowiedzi: *do pięćset metrów zejdziemy* por. Artura Ziętka oraz *do pięćuset metrów zejdziemy* mjr pil. Roberta Grzywiny

- Widma sylab **me**: LT6 i LT7 z fraz *dwieście trzydzieści metrów* i *sto metrów*

- Wynik analizy DTW1

Odległości między sylabami: 1) **me** (z głosu niezidentyfikowanego z frazy *sto metrów*, oznaczonego jako X\_me(100), 2) **me** (z głosu niezidentyfikowanego z frazy *dwieście trzydzieści metrów*, oznaczonego, jako X\_me (230), 3) **me** (z wypowiedzi mjr. pil. Roberta Grzywiny z frazy *pięćuset metrów*, oznaczonego jako S\_me), 4) **me** z wypowiedzi por. Artura Ziętka (z frazy *pięćset metrów*, oznaczonego jako N\_me).

3. Czy można wykluczyć, iż **niezidentyfikowane** wypowiedzi padające w ostatnich minutach lotu należały do osób spoza Załogi?

Nie. Przykładowo nie można wykluczyć głosu gen. Andrzeja Błasika – z nagrania porównawczego przemówienia gen. Andrzeja Błasika wynika, że częstotliwość podstawowa  $F_{\min}$  Jego wypowiedzi jest podobna jak w wypowiedzi *sto metrów* ( $F_{\min}$ : 70–80 Hz,  $F_{\max}$ : 140–150 Hz), podobne parametry wykazuje również struktura formantowa, podobne wartości formantu trzeciego (niosącego dużo informacji osobniczych) i jego szerokości. Analizę przeprowadzano na tej samej samogłosce *o* wyekstrahowanej z frazy: *historii* (samogłoska akcentowana) z przemówienia z nagrania porównawczego oraz z analizowanej frazy: *sto metrów* (również samogłoska akcentowana) nagrania dowodowego. Przeanalizowano tę frazę, jako przykładową, ponieważ udało się znaleźć sylabę *sto* również w nagraniu porównawczym gen. Andrzeja Błasika we frazie *historii*. Przeanalizowano również widma, LTAS na samogłosce *o* z obu wymienionych powyżej fraz oraz dla porównania z frazy *sto* z głosu dyr. Mariusza Kazany, mjr. pil. Roberta Grzywiny oraz por. Artura Ziętka (dowód 9).

**Dowód 9**

- Spektrogram sylaby *sto* z frazy *historii* (z nagrania porównawczego gen. Andrzeja Błasika, SP33)
- LT8: z samogłoski *o* z frazy *historii*
- LT9: z samogłoski *o* z frazy *sto metrów*
- LT10: z samogłoski *o* z frazy *co* (dyr. Mariusz Kazana)
- LT11: z samogłoski *o* z frazy *sto dwudziesty ósmy poziom* (mjr pil. Robert Grzywna)
- LT12: z samogłoski *o* z frazy *sto* (por. Artur Ziętek)

Ostatecznie podjęto próbę porównania odległości sylaby *sto* wyekstrahowanej z frazy wypowiedzianej przez nieznanego mówcę (*sto metrów*) do 10 sylab *sto* zapisanych w nagraniach rozmów w kabinie samolotu (3 sylaby wypowiedziane przez mjr. pil. Roberta Grzywnę, 3 sylaby wypowiedziane przez kpt. pil. Arkadiusz Protasiuka, 3 sylaby wypowiedziane przez por. Artura Ziętka i sylaby *sto* wypowiedzianej przez gen. Andrzeja Błasika z frazy *historii* (według tabeli 2).

**Tabela 2**

Sylaby *sto* analizowane z różnych fragmentów wypowiedzi mjr. pil. Roberta Grzywny, kpt. pil. Arkadiusza Protasiuka, por. Artura Ziętka, gen. Andrzeja Błasika oraz sylaby *sto* wypowiedzianej przez weryfikowanego mówcę z frazy *sto metrów*.

Lp.	kod	Fraza	Kod mówcy
1	8:08:59,	<i>sto pięćdziesiąt cztery</i>	S_sto(154)
2	8:10:15,5	<i>sto dwudziesty ósmy poziom</i>	S_sto(128)
3	8:12:41.0	<i>do dziewięć sto</i>	A_sto(do9-sto)
4	8:13:46.5	<i>Dziesięć stopni</i>	A_sto (10stopni)
5	8:16:53.5	<i>o ktorej te uroczystosci</i>	S_sto(uroczystości)
6	8:27:20,5	<i>dwa stopnie siedem cztery</i>	A_sto(2stopnie)

7	8:27:24,5	<i>dwa stopnie siedem cztery piec</i>	I_sto(2stopnie)
8	8:40:40	<i>sto pięćdziesiąt</i>	N_sto(150)
9	8:40:45,5	<i>sto</i>	N_sto
10	8:40:51	<i>sto (last)</i>	N_stolast
11	Historii	<i>historii</i>	B_sto

### **Dowód 10**

Algorytm DTW wskazał jako najbardziej bliskie sylaby *sto* z frazy *sto metrów* frazie *sto* wyekstrahowanej z *historii* (z głosu gen. Błasika). Na tej podstawie nie można jednak z pewnością potwierdzić, że frazę *sto metrów* (oraz *dwieście trzydzieści metrów*) wypowiadał gen. Andrzej Błasik. Trzeba uwzględnić następujące czynniki:

- 1) Analiza dotyczy wyłącznie jednej sylaby. Algorytm DTW funkcjonuje tym skuteczniej im dłuższe są wypowiedzi.
- 2) Istnieją błędy w segmentacji (nie we wszystkich przypadkach udało się np. wyekstrahować spółgłoski *s* oraz *t*. W związku z tym miary odległości są nieprecyzyjne.
- 3) We wszystkich sygnałach istnieje również duża energia szumu.

Podobieństwa i różnice w obu widmach mogą mieć charakter **wyłącznie** orientacyjny. Niestety nie było możliwości większej liczby porównań.

Obliczenie tych odległości miało, za zadanie jedynie odpowiedzieć czy można wykluczyć, że frazę *sto metrów* wypowiadał gen. Andrzej Błasik. Nie można wykluczyć, iż niezidentyfikowany głos wypowiadający pod koniec lotu frazę *sto metrów* (czy też frazę *dwieście trzydzieści metrów*) mógł należeć do gen. Andrzeja Błasika lub innej osoby o podobnym głosie.



## 7. Podsumowanie opinii

Zgodnie z punktem 1 niniejsza opinia obejmuje:

1. Fonetyczno-akustyczną weryfikację transkrypcji wypowiedzi.

Analizy fonetyczno-akustyczne przedstawione w niniejszej opinii, poza wskazanymi w załączniku wyjątkami, potwierdziły ogólnie transkrypcje przeprowadzoną przez (MAK\_CLK). Dały one podstawę do zakwestionowania transkrypcji językowej dokonanej na drodze odsłuchów przez ekspertów MAK\_CLK kilku wypowiedzi (umieszczonych w załączniku 1).

Szczegółowa analiza fonetyczna innych niezidentyfikowanych wypowiedzi (wskazanych według kodu czasowego przez ZB) nie była możliwa ponieważ po ich odszumieniu (różnymi procedurami) okazało się, że sygnał mowy jest zbyt zniekształcony, aby dokonać relatywnie poprawnej ekstrakcji parametrów, a interpretacja treści jest niejednoznaczna, zależna od metody odszumiania i sposobu odsłuchu. Wnioskowanie na podstawie nie jednoznacznych danych pomiarowych oczywiście nie jest możliwe.

2. Dodatkowo opinia obejmowała ocenę stylu, ekspresywności wybranych wypowiedzi i ich emfazy.

Analiza treści językowych i pozajęzykowych wykazała, iż zaniepokojenie związane z pilotażem samolotu pojawiło się po ocenie sytuacji pogodowej. Dał temu wyraz kpt. pil. Arkadiusz Protasiuk meldując formalnie dyr. Mariuszowi Kazanie o niemożności lądowania. Dalsze fazy lotu ujawniają wzrastające zaniepokojenie załogi związane z podjęciem pilotażowej próby lądowania. Charakterystyczne jest emfatyczne akcentowanie wyrazów: *tragedii*, *mgła* (kilkakrotnie), *nie damy rady*. Ekstremalny stres daje się zauważyć w odczycie przez por. Artura Ziętka wartości: ***dwadzieścia!*** Niekontrolowany ekstremalny stres w głosie prawdopodobnie jednego z członka Załogi występuje w ostatnich sekundach katastrofy podczas krzyku.

3. Ocenę możliwości identyfikacji i weryfikacji głosów na drodze fonetyczno-akustycznych analiz.

W kabinie samolotu niewątpliwie przebywały inne osoby, z pewnością można przyjąć, że niektóre frazy należały do dyr. Mariusza Kazany (który wypowiada się po meldunku kpt. pil Arkadiusza Protasiuka). Nie można wykluczyć, iż niezidentyfikowany głos wypowiadający pod koniec lotu frazę *sto metrów* (czy też frazę *dwieście trzydzieści metrów*) mógł należeć do gen. Andrzeja Błasika lub innej osoby o podobnym głosie. Z dużym stopniem prawdopodobieństwa, graniczącym z pewnością, należy stwierdzić, iż frazy te nie były wypowiedziane przez członków Załogi. Inne niezidentyfikowane wypowiedzi z bardzo dużym prawdopodobieństwem nie należały do głosów Załogi, jednakże trudno wskazać obiektywne przypisania.

UNIwersytet IM. ADAMA MICKIEWICZA W POZNANIU  
Wydział Neofilologii  
Instytut Językoznawstwa  
al. Niepodległości 4, 61-874 Poznań  
tel. 61 829 36 63, fax 61 829 36 62  
REGON 000001293 NIP 7770006350  
(1)

Wicedyrektor  
Instytutu Językoznawstwa ds. nauki  
prof. dr hab. inż. Grażyna Demenka



## Bibliografia

- Adami, A. G., Mihaescu, R., Reynolds, D. A., Godfrey, J. J. (2003). *Modeling prosodic dynamics for SPeaker recognition*. Proceedings of the ICASSP'03, vol. IV, pp. 788-791.
- Alexander, A., Dessimoz, D., Botti, F., Drygajlo, A. (2005). Aural and automatic forensic SPeaker recognition in mismatched conditions. *International Journal of Speech, Language and the Law*, 12(2), pp. 214-234.
- Aronowitz, H. (2014). *Inter dataset Variability compensation for SPeaker recognition*. Proceedings of the IEEE International Conference on Acoustics, SPEech and Signal Processing (ICASSP) pp. 4002-4006.
- Banziger, T., Scherer, K. R. (2005). The role of intonation in emotional expressions. *SPEech Communication*, 46(3-4), pp. 252-267. doi: 10.1016/j.SPecom.2005.02.016
- Campbell J, P. (1997). SPeaker recognition: A tutorial. *Proceedings of the IEEE*, 85(9), pp. 1437-1462.
- Demenko, G., Jastrzebska, M., Izdebski, K., Yan, Y. (2012). *Visualizations by acoustics of voice stress: Is there an optical mucośal wave correlate*. Proceedings of the SPIE, vol. 8207, pp. 82072M.
- Demenko, G., Wypych, M., & Baranowska, E. (2003). *Implementation of grapheme-to-phoneme rules and extended SAMPA alphabet in Polish text-to-SPEech synthesis*. *SPEech and Language Technology*, 7, 79-97.
- Demenko, G., Jastrzębska, M. (2012a). Analysis of natural SPEech under stress. *Acta Physica Polonica 121*(1 ), pp. A92-A95.
- Demenko, G., Jastrzębska, M. (2012b). *Analysis of Voice Stress in Call Centers Conversations*. Proceedings of the SPEech Prosody 2012, pp. 183-186.
- Demenko, G. (2015) Korpusowa analiza mowy, wyd. Exit.
- Doddington, G. R., Przybocki, M., A, Martin, A., F, Reynolds, D. A. (2000). The NIST SPeaker recognition evaluation—overview, methodology, systems, results, perSPective. *SPEech Communication*, 31(2), pp. 225-254.
- Gonzalez-Rodriguez, J. (2014). Evaluating Automatic SPeaker Recognition systems: An overview of the NIST SPeaker Recognition Evaluations (1996-2014). *Loquens*, 1(1), pp. e007.

- Hansen, E., G. Slyh, R., Anderson, T., R. (2004). *SPeaker recognition using phoneme-Specific GMMs*. Proceedings of the ODYSSEY04-The SPeaker and Language Recognition Workshop.
- Hansen, J. H., Swail, C., South, A. J., Moore, R. K., Steeneken, H., Cupples, E. J., . . . Verlinde, P. (2000). The impact of SPeech under 'stress' on military SPeech technology *Nato Project* (vol. 4). Dallas.
- Hansen, J. H. L., Varadarajan, V. (2009). Analysis and Compensation of Lombard SPeech Across Noise Type and Levels With Application to In-Set/Out-of-Set SPeaker Recognition. *IEEE Transactions on Audio SPeech and Language Processing*, 17(2), pp. 366-378. doi: 10.1109/tasl.2008.2009019
- Ikeno, A., Hansen, J., H. L. (2007). *Lombard SPeech Impact on Perceptual SPeaker Recognition*. Proceedings of the InterSPeech 2007: 8th Annual Conference of the International SPeech Communication Association, vol. 1-4, pp. 2025-2028.
- Koolagudi, SP., Rao, K. SP. (2012). Emotion recognition from SPeech: a review. *International Journal of SPeech Technology*, 15(2), pp. 99-117. doi: 10.1007/s10772-011-9125-1
- Lee, C. M., Narayanan, SP., Pieraccini, R. (2001). *Recognition of negative emotions from the SPeech signal*. Proceedings of the Automatic SPeech Recognition and Understanding, 2001. ASRU'01, pp. 240-243.
- Mary, L., Yegnanarayana, B. (2008). Extraction and representation of prosodic features for language and SPeaker recognition. *SPeech communication*, 50(10), pp. 782-796.
- Nolan, F. D. (1980). *The phonetic bases of SPeaker recognition*. University of Cambridge.
- Reynolds, D., Andrews, W., Campbell, J., Navratil, J., Peskin, B., Adami, A., . . . Xiang, B. (2003). *The SuperSID project: Exploiting high-level information for high-accuracy SPeaker recognition*. Proceedings of the Acoustics, SPeech, and Signal Processing (ICASSP'03), vol. 4, pp. 784-787.
- Rose, P. (2006). Technical forensic SPeaker recognition: Evaluation, types and testing of evidence. *Computer SPeech and Language*, 20(2-3 SPEC. ISS.), pp. 159-191.
- Scherer, K. R., Banse, R., Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), pp. 76-92. doi: 10.1177/0022022101032001009
- Schuller, B., Müller, R., Lang, M. K., Rigoll, G. (2005). *SPeaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles*. Proceedings of the InterSPeech, pp. 805-808.

- Schuller, B., Rigoll, G., Lang, M. (2004). *SPeech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture*. Proceedings of the Acoustics, SPeech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, vol. 1, pp. I-577-580 vol. 571.
- Shriberg, E. (2007). *Higher-level features in SPeaker recognition* (Springer Ed. Vol. ). Berlin Heidelberg.
- Shriberg, E., Ferrer, L., Kajarekar, SP., Venkataraman, A., Stolcke, A. (2005). Modeling prosodic feature sequences for SPeaker recognition. *SPeech Communication*, 46(3-4), pp. 455-472.
- Squartini, SP., Schuller, B., Hussain, A. (2012). Cognitive and Emotional Information Processing for Human-Machine Interaction. *Cognitive Computation*, 4(4), pp. 383-385. doi: 10.1007/s12559-012-9180-1

## **Załącznik 1**

Przedstawiono następujące analizy:

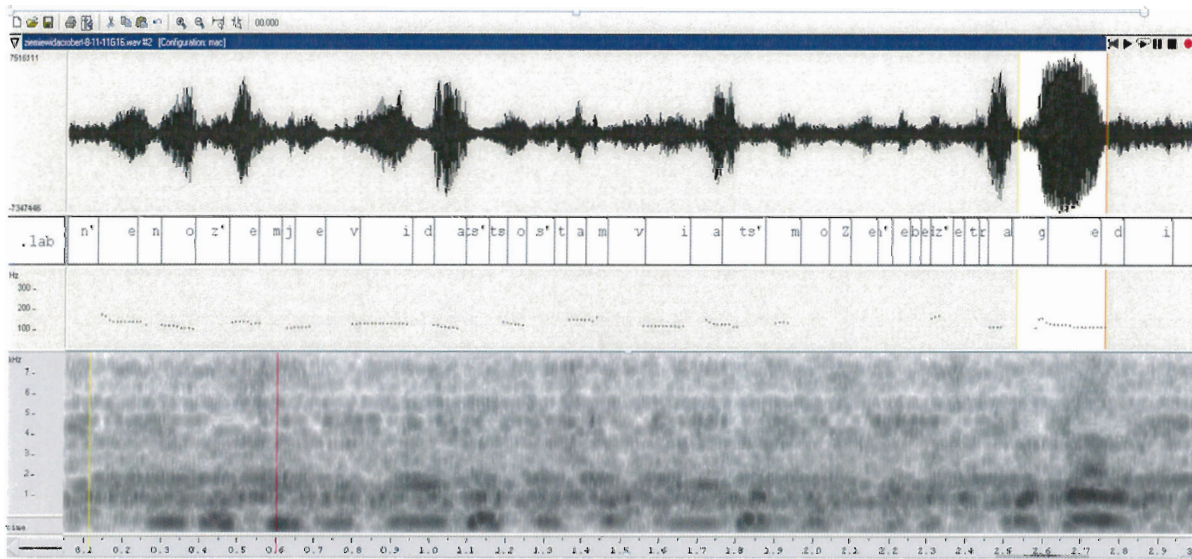
- Trójwymiarowe Spektrogramy (oś pozioma-wymiar czasu, oś pionowa –wymiar częstotliwości i stopień zaciemnienia związany z energią sygnału)
  - Oscylogramy wypowiedzi (zmiany amplitudy w funkcji czasu)
  - Pomiar i wykresy przebiegu częstotliwości podstawowej
  - Wykresy widma długoterminowego LTAS
  - Wyniki identyfikacji mówcy DTW
- 
1. Dowód 1. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP1-SP9)
  2. Dowód 2. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP10-SP14)
  3. Dowód 3. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP15-SP17)
  4. Dowód 4. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP18-SP22)
  5. Dowód 5. Transkrypcja fonetyczna i ortograficzna (Spektrogramy SP23-SP25, SP25a)
  6. Dowód 6. Tabela 1. Pomiar parametru  $F_0$ , Spektrogramy SP26-SP 27)
  7. Dowód 7. Analizy porównawcze (Spektrogram 28, LTA 1-5)
  8. Dowód 8. Analizy porównawcze (Spektrogramy SP29, 29a, 30, 31, 32, LT6, LT7, DTW1)
  9. Dowód 9. Analizy porównawcze (Spektrogram SP33, LT8 - LT12,)
  10. Dowód 10. Obliczanie odległości między głosami metoda DTW (DTW2)

## Dowód 1

W nawiasach podano przypisanie wypowiedzi poszczególnym osobom według transkrypcji (MAK\_CLK )

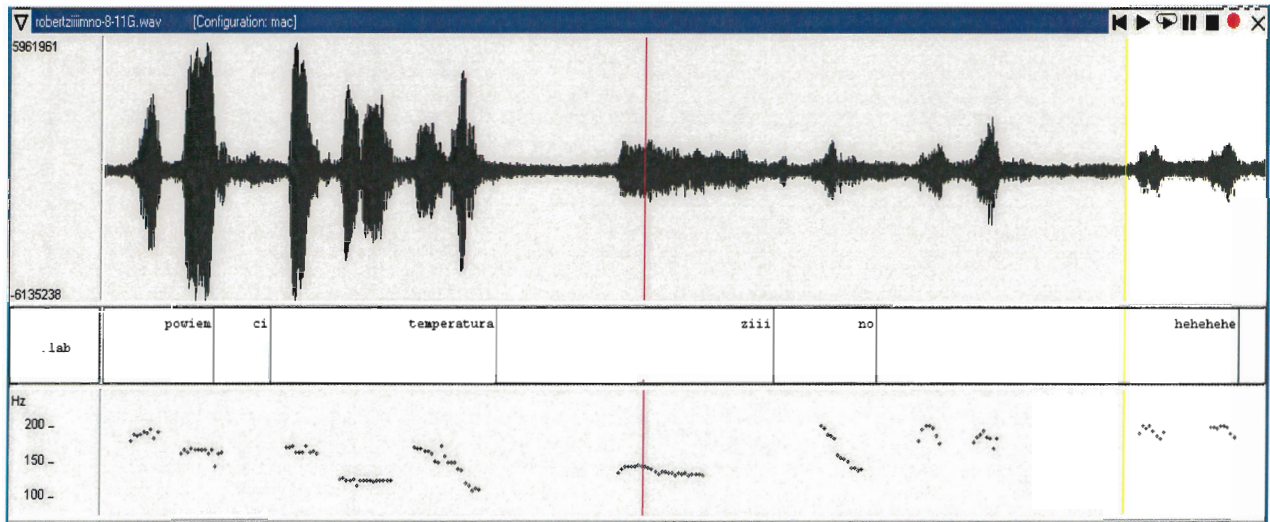
SP1) *nie no ziemie widać coś tam widać może nie będzie tragedii* ( mjr pil. Robert Grzywna)

Wypowiedź wskazująca na wątpliwość (kontur intonacyjny z całej frazy wskazujący wahanie i zaniepokojenie). Tempo wypowiedzi dość wolne. Zdecydowane zaakcentowanie wyrazu **tragedii** (bardzo wyraźne istotne zwiększenie energii sygnału, wzrost parametru F0 na sylabie akcentowanej **ge**).



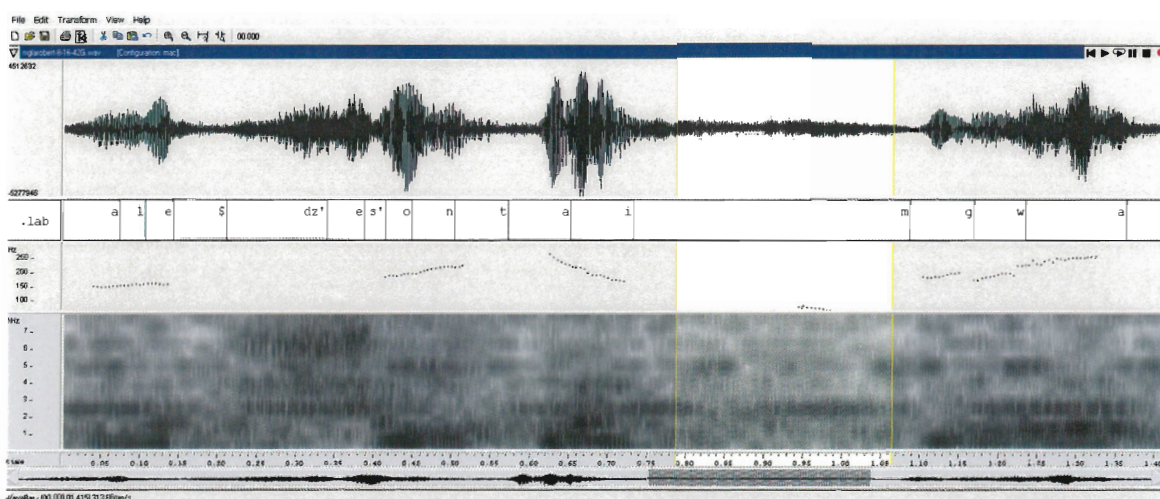
SP2) *ja ci powiem temperaturę ziiimno...* (mjr pil. Robert Grzywna)

Wypowiedź neutralna, fragmentarycznie wskazująca dyskomfort (wyrazne przedłużenie samogłoski *i* w wyrazie *zimno*)



SP3) *ale 10 ta i mgła* (mjr pil. Robert Grzywna)

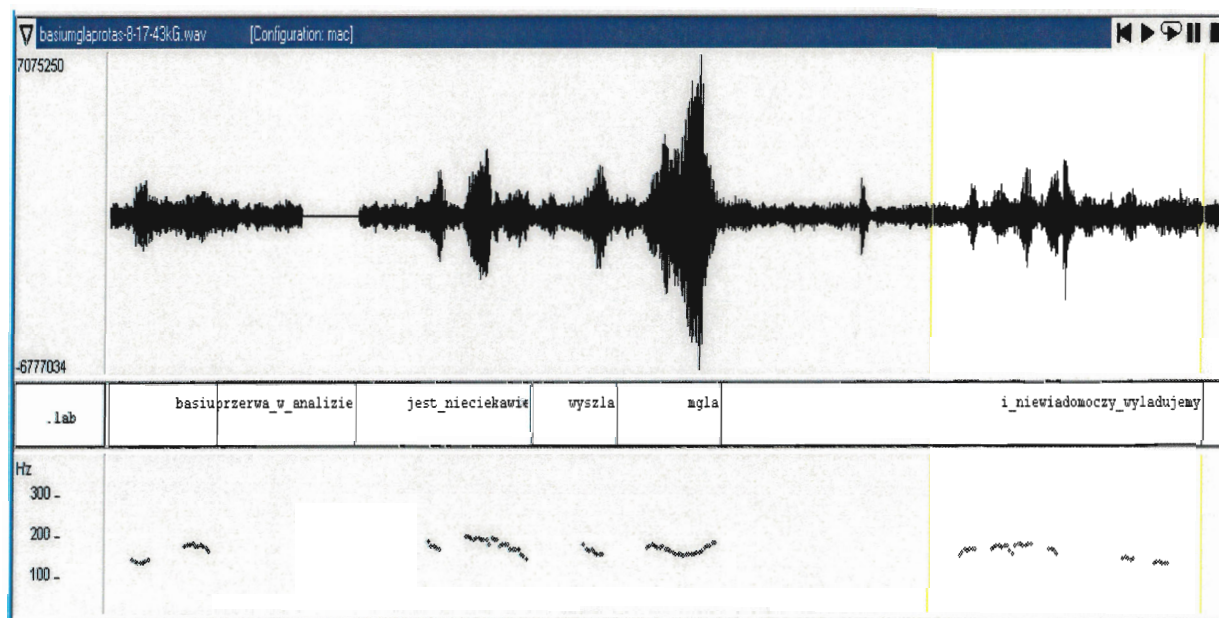
Wypowiedź wskazująca dość jednoznacznie zdziwienie (zaakcentowanie intonacyjnie oraz wzrostem amplitudy wyrazu *dziesiąta* oraz iloczynowo wyrazu *mgła*, przedłużenie samogłoski *a*, charakterystyczny wzrost częstotliwości podstawowej w wyrazie *mgła*).





**SP4) Basiu jest nieciekawie wyszła *mgła* i nie wiadomo czy wylądujemy** (kpt. pil. Arkadiusz Protasiuk)

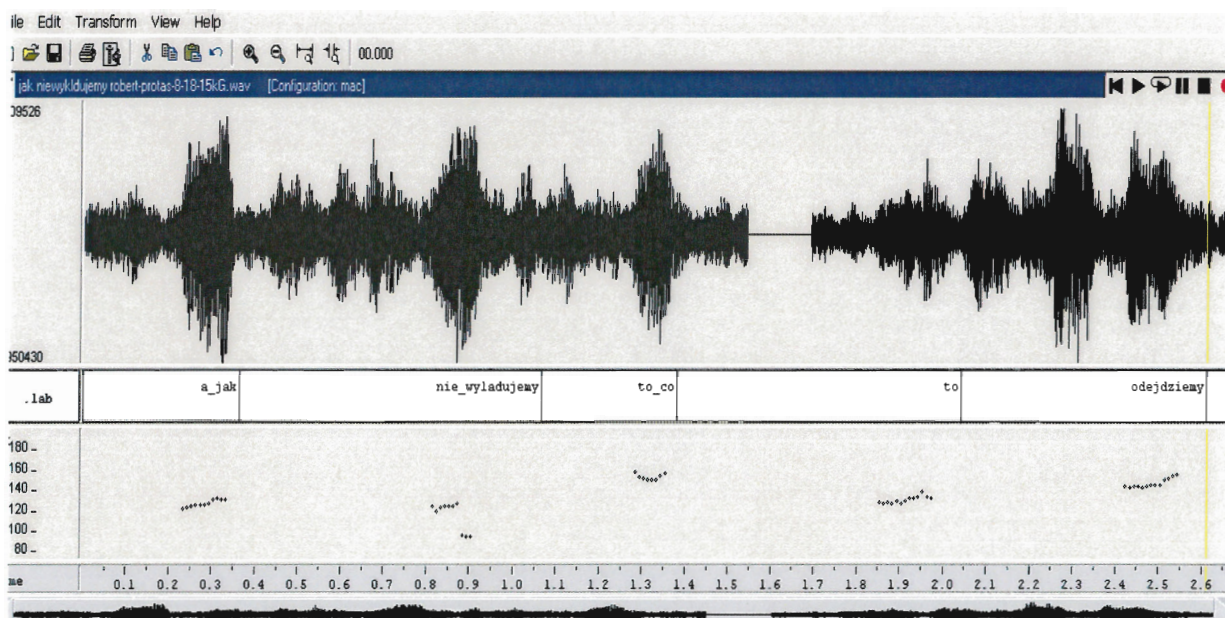
Zdecydowane zaakcentowanie wyrazu *mgła*. Struktura prozodyczna wskazująca wahanie (ton średni na wyrazie *mgła*, spadek F0 na końcu wypowiedzi, wydłużenie sylab końcowych).



**SP5) a jak nie wylądujemy to co?, (mjr pil. Robert Grzywna)**

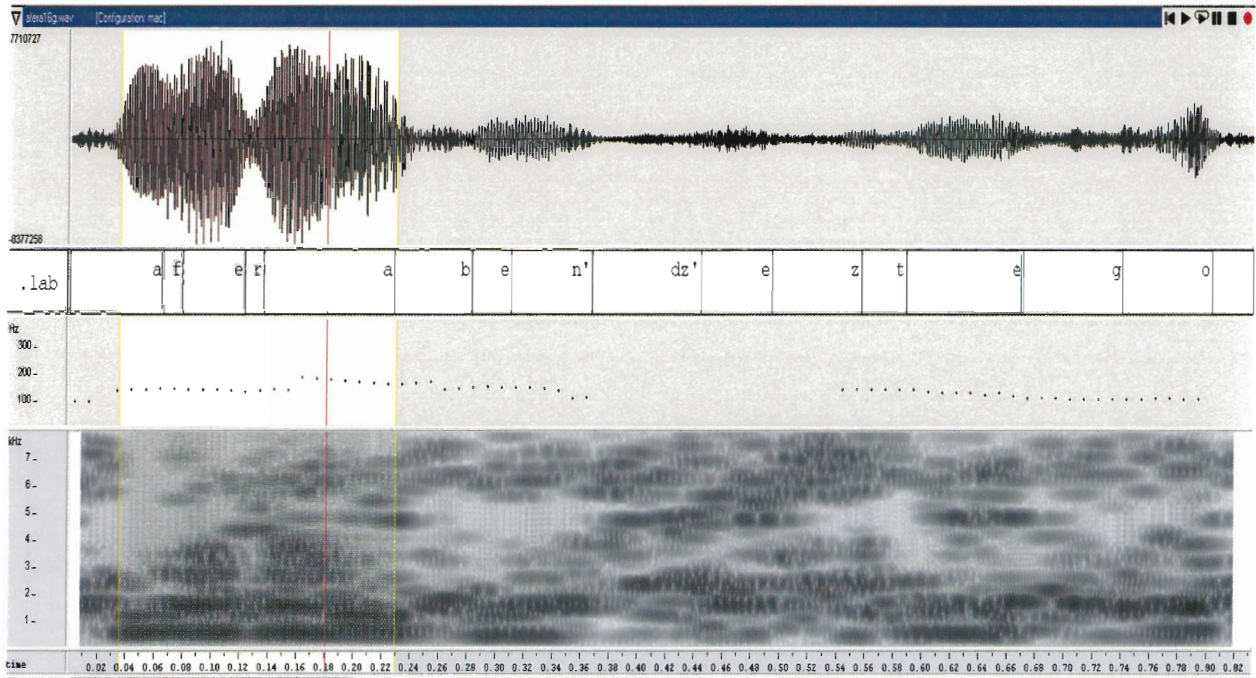
*To odejdziemy, (kpt. pil. Arkadiusz Protasiuk)*

Wypowiedź wskazująca wyraźne pytanie mjr. pil. Roberta Grzywny i odpowiedź – zaakcentowanie wyrazu *odejdziemy* przez kpt. pil. Arkadiusza Protasiuka.



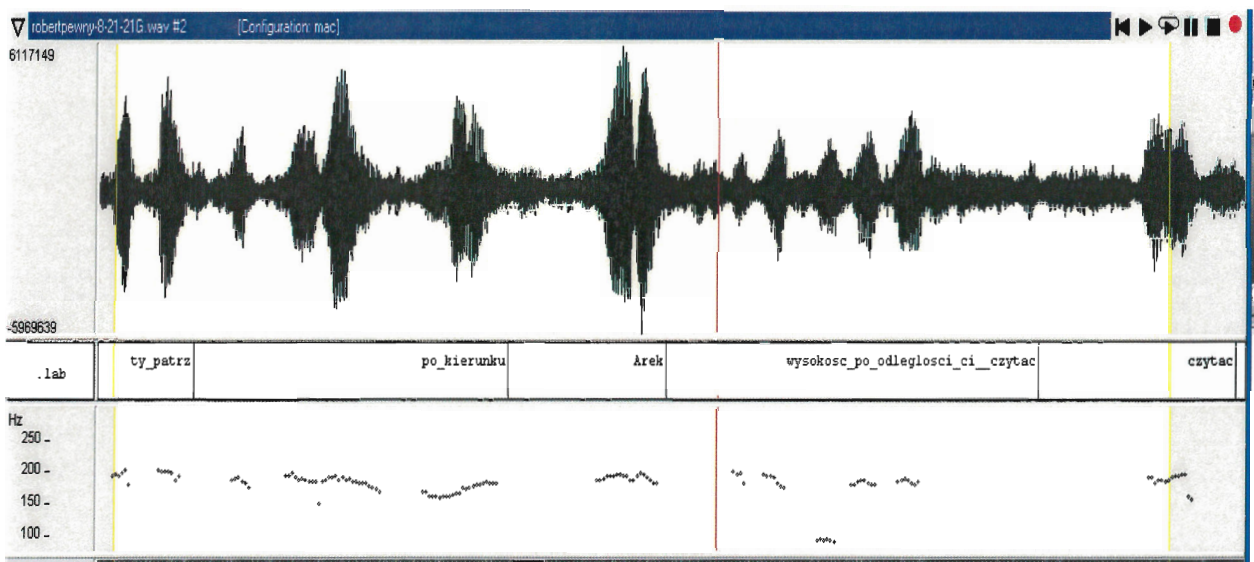
SP6) *afera* będzie (brak transkrypcji w przekazanej dokumentacji)

Zaakcentowanie wyrazu *afera*.



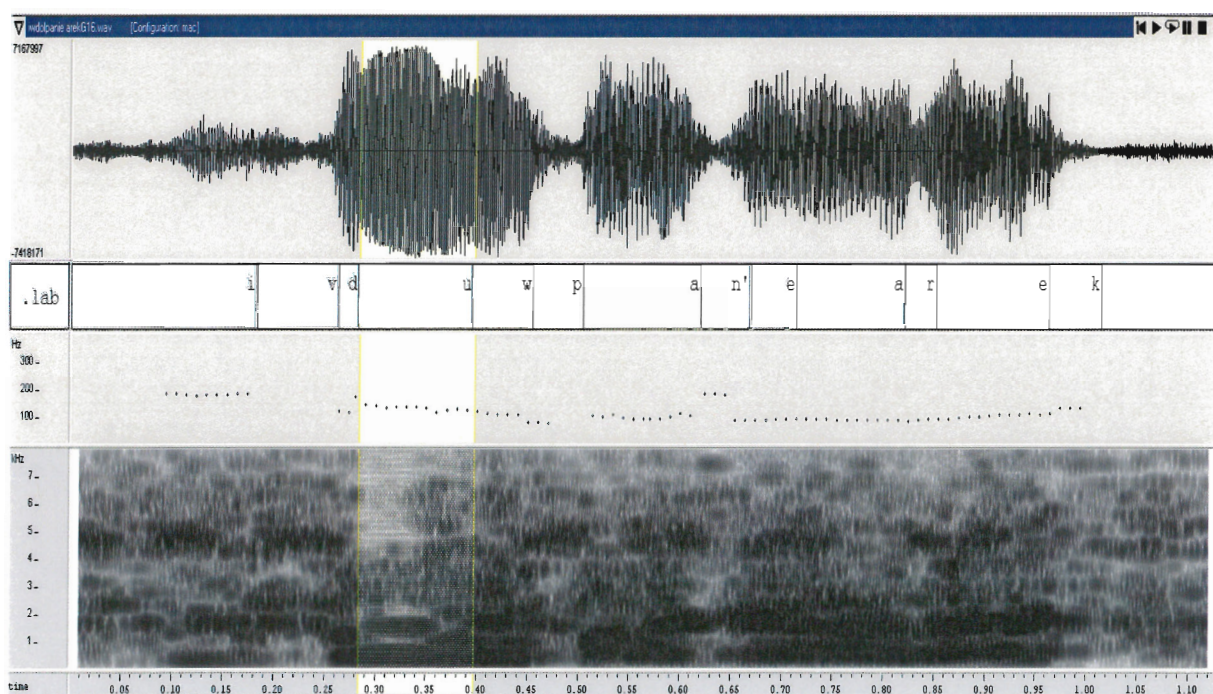
SP7) *ty patrz po kierunku - Arek – wysokość po odległości ci czytać?*(mjr pil. Robert Grzywna)

Wypowiedź neutralna. Kontur intonacyjny wskazujący na polecenie (*ty patrz po kierunku* i pytanie: *wysokość po odległości ci czytać?*),



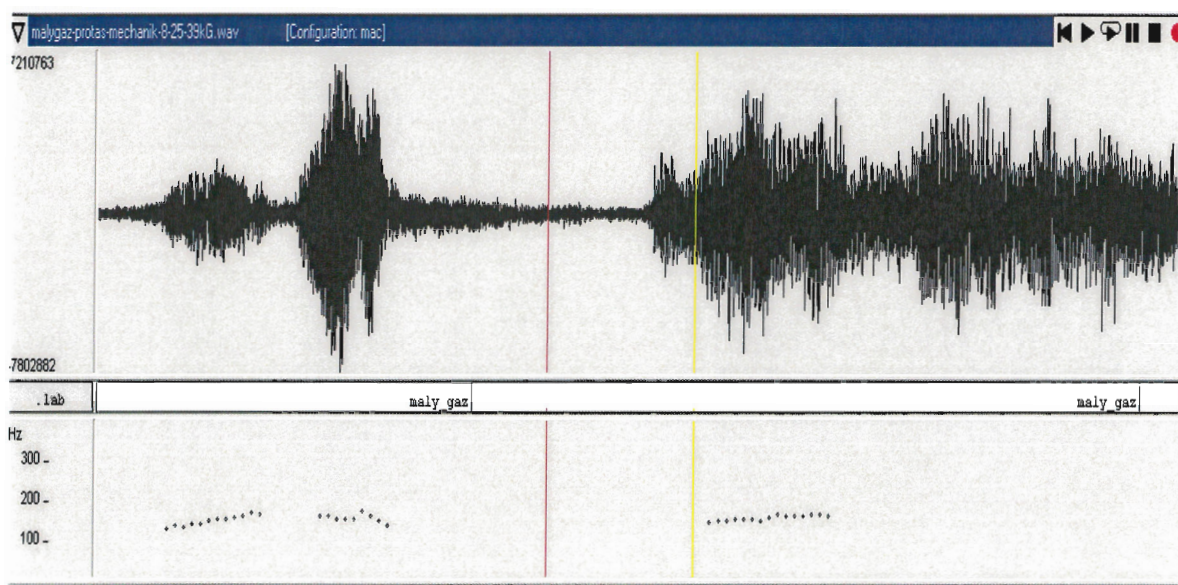
SP8) i w **dół** panie Arek mały gaz, (Mjr pil. Robert Grzywna)

Zdecydowane zaakcentowanie wyrazu w **dół** (sygnalizowane wzrostem intensywności, częstotliwości podstawowej, iloczasu) oznacza zachetę podjęcie akcji.



**SP9) mały gaz** (kpt. pil. Arkadiusz Protasiuk) - *mały gaz* (chor. Andrzej Michalak)

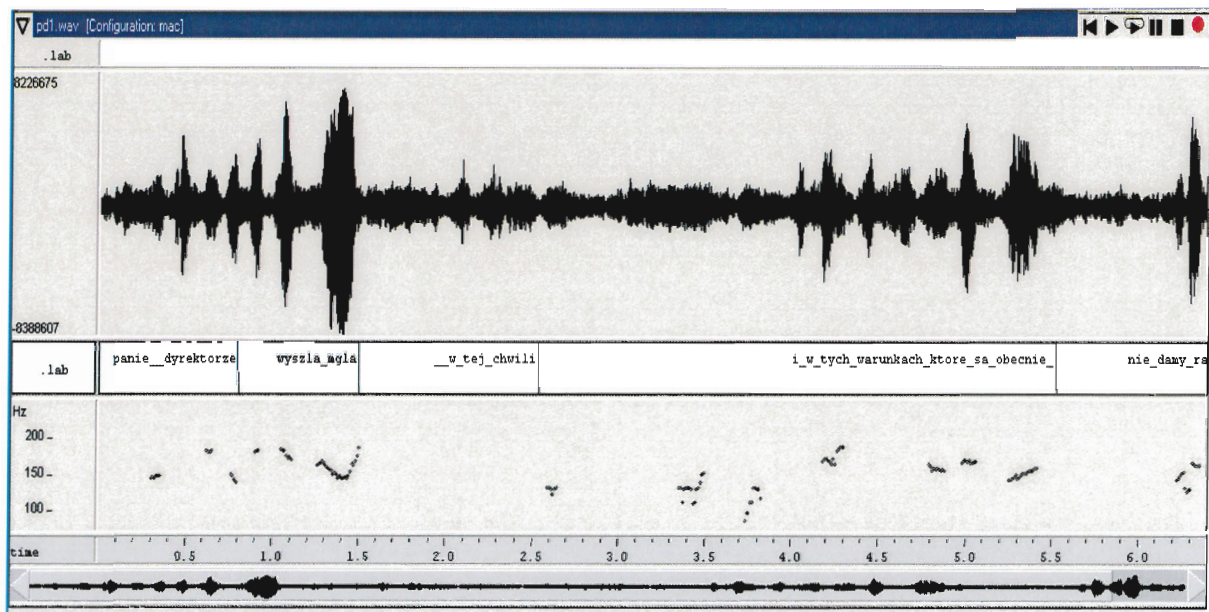
Wypowiedzi neutralne ilustrujące różnicowania osobnicze w zakresie tempa wypowiedzi, w przypadku wypowiedzi pierwszej (*mały gaz*) przez dowódcę tempo dwukrotnie szybsze niż wypowiedzi *mały gaz* przez mechanika chor. Andrzeja Michalaka.



## Dowód 2

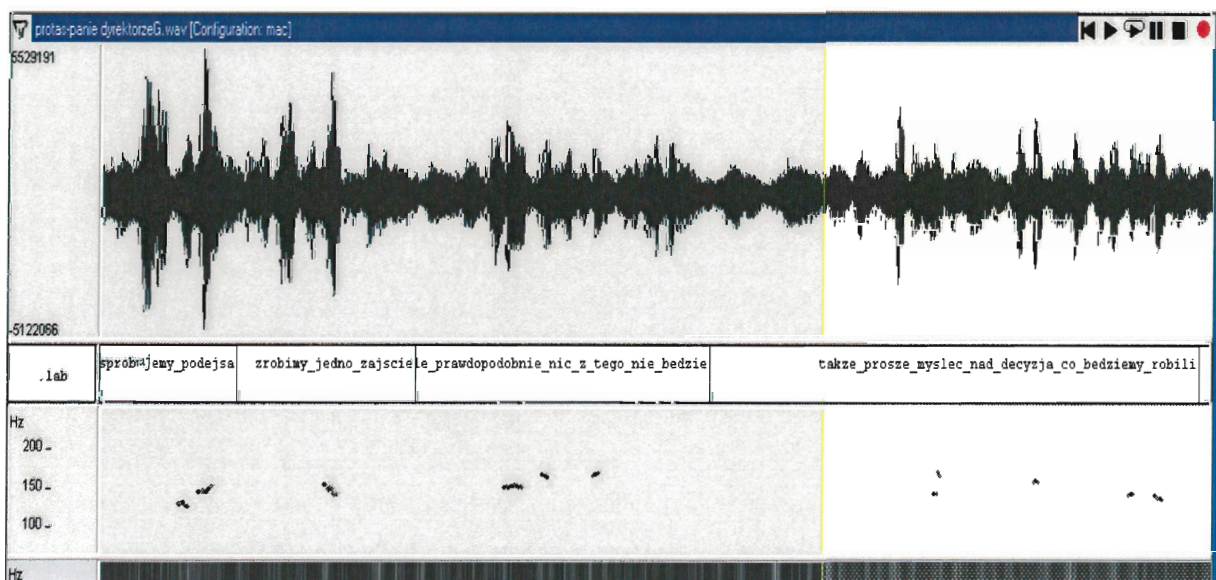
S10) *panie dyrektorze – wyszła mgła i w tej chwili i w tych warunkach które są obecnie nie damy rady usiąść* (kpt. pil. Arkadiusz Protasiuk)

Wyraźne zaakcentowanie wyrazu *mgła*, *nie damy* rady, zdecydowane stwierdzenie faktu.



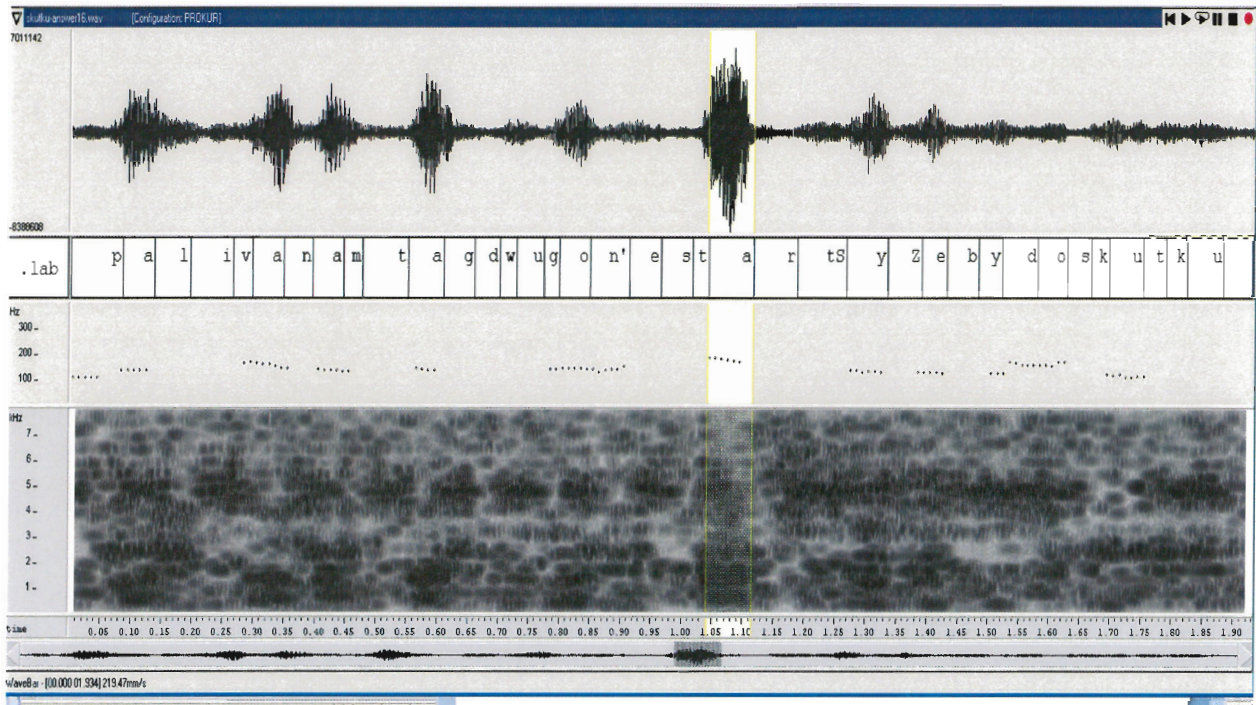
S11) *Spróbujemy podejść – zrobimy jedno zejście- ale prawdopodobnie nic z tego nie będzie. Tak że proszę już myśleć nad decyzją co będziemy robili* (kpt. pil. Arkadiusz Protasiuk)

Formalny meldunek: i treściowo i prozodycznie sygnalizowany odpowiednimi frazowymi konturami intonacyjnymi).



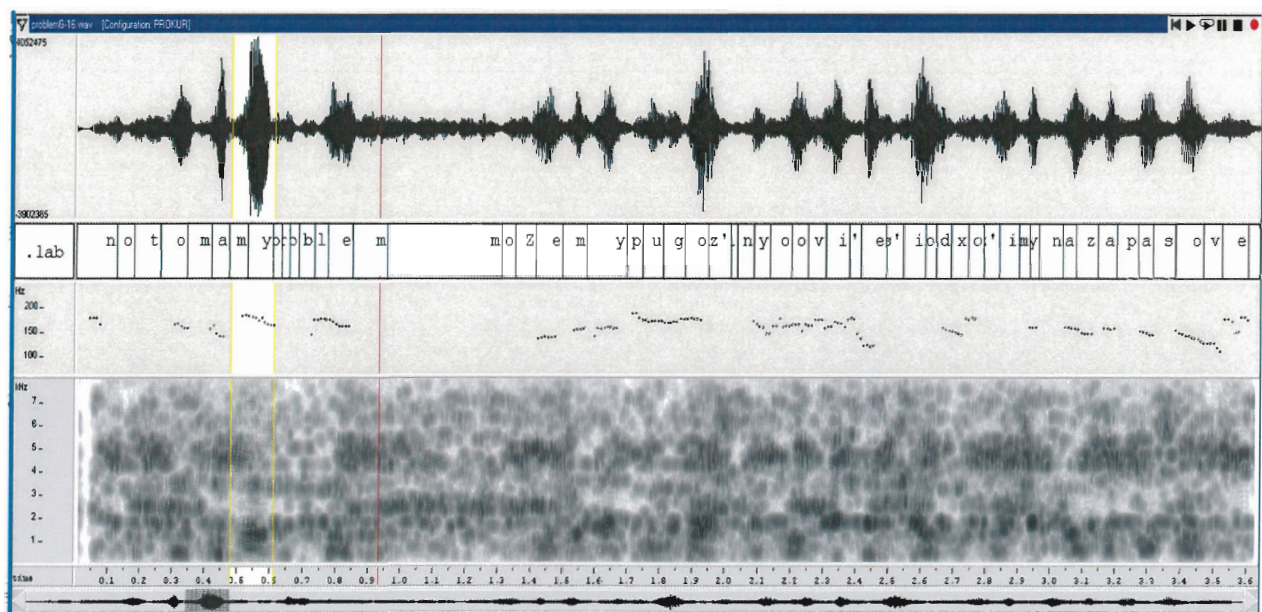
**S12)** paliwa nam tak dużo nie *starczy* żeby do skutku (kpt. pil. Arkadiusz Protasiuk)

Zdecydowane stwierdzenie niemożliwości tego manewru (zaakcentowanie frazy: *nie starczy*)



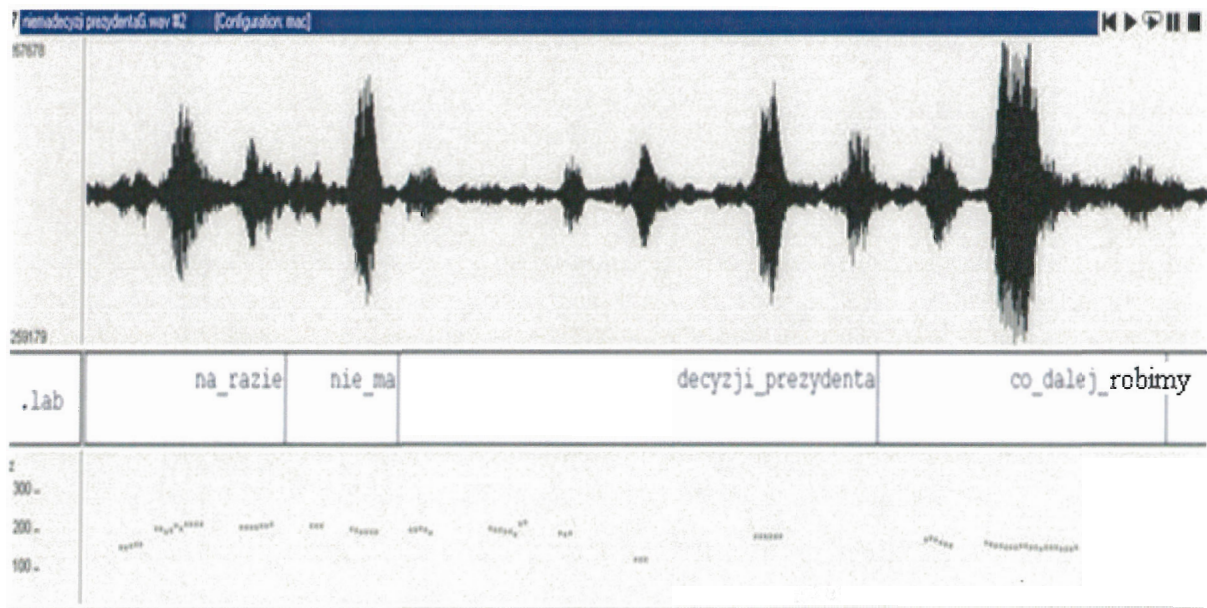
**S13)** *no to mamy* problem (dyr. Mariusz Kazana), odpowiedź: *możemy pół godziny powiesić i odchodzimy na zapasowe*, (kpt. pil. Arkadiusza Protasiuka)

Zdecydowane stwierdzenie (zaakcentowany wyraz *mamy*)



**S14)** *na razie nie ma decyzji prezydenta co dalej robimy* (dyr. Mariusz Kazana)

Wypowiedź informacyjna. Zaakcentowanie wyrazu *nie ma* oraz *dalej* (wzrost amplitudy, iloczasu).

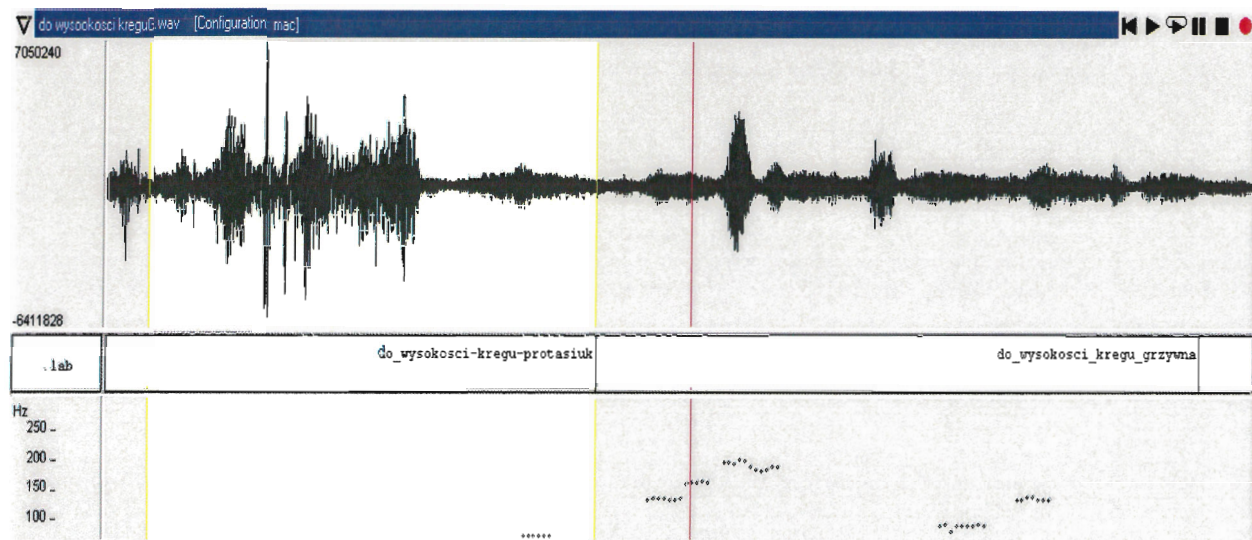




### Dowód 3

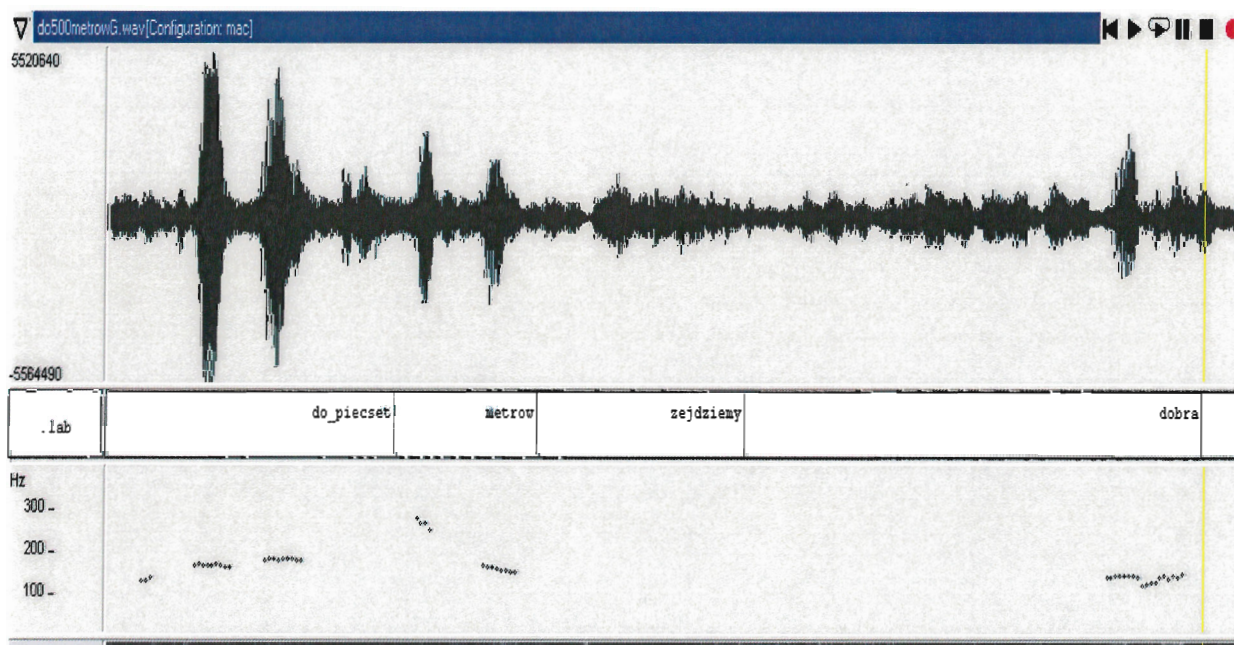
SP15) do wysokości kręgu – do wysokości kręgu (kpt. pil. Arkadiusz Protasiuk oraz mjr pil. Robert Grzywna)

Wypowiedzi neutralne kpt. pil. Arkadiusza Protasiuka oraz mjr. pil. Roberta Grzywny. Tempo wypowiedzi dowódcy znacznie szybsze niż drugiego pilota.



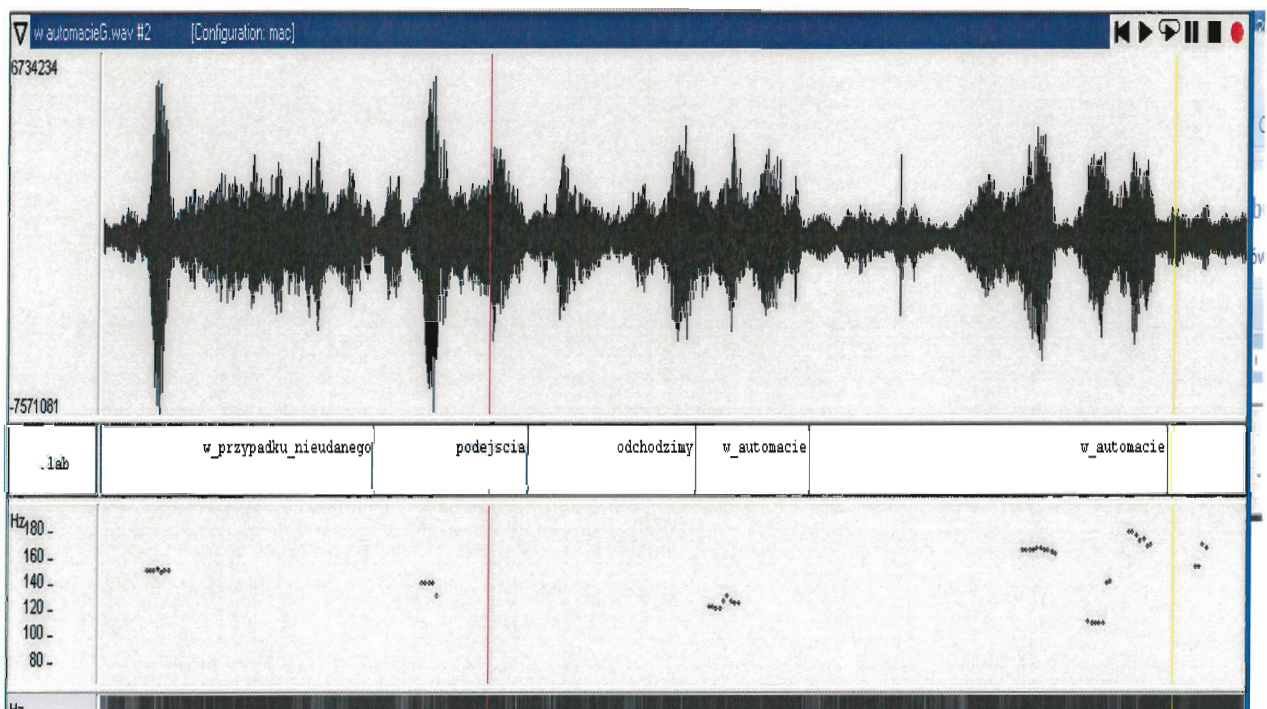
SP16) do pięćset metrów zjedziemy (mjr pil. Robert Grzywna)

Zaakcentowany wyraz *pięćset* wskazujący tą wartość jako graniczną.



**SP17) w przypadku nieudanego podejścia odходимy w automacie** (kpt. pil. Arkadiusz Protasiuk)

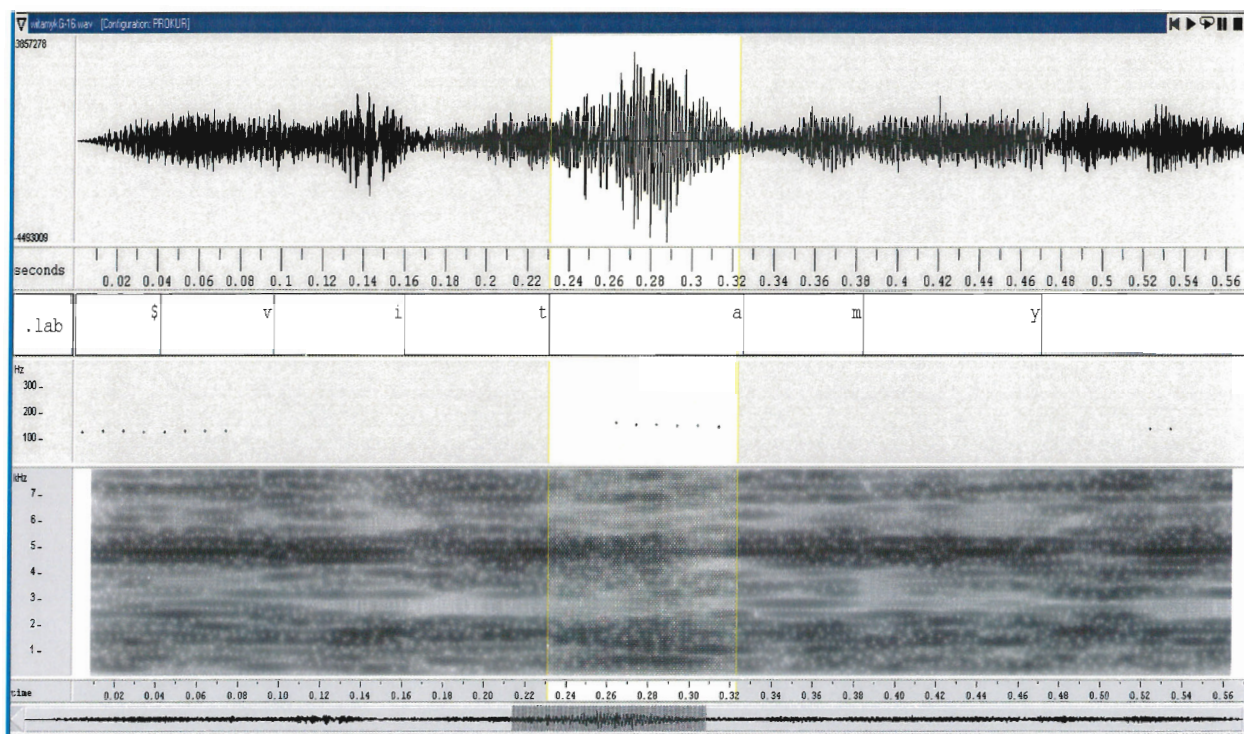
Zaakcentowany wyraz **w przypadku** (zasygnalizowanie alternatywy) oraz wyrazu **podejścia**. Przejrzysta struktura akcentuacyjna w całej frazie wyraźnie sygnalizująca możliwość alternatywnego manewru. Odpowiedź: *w automacie*.



## Dowód 4

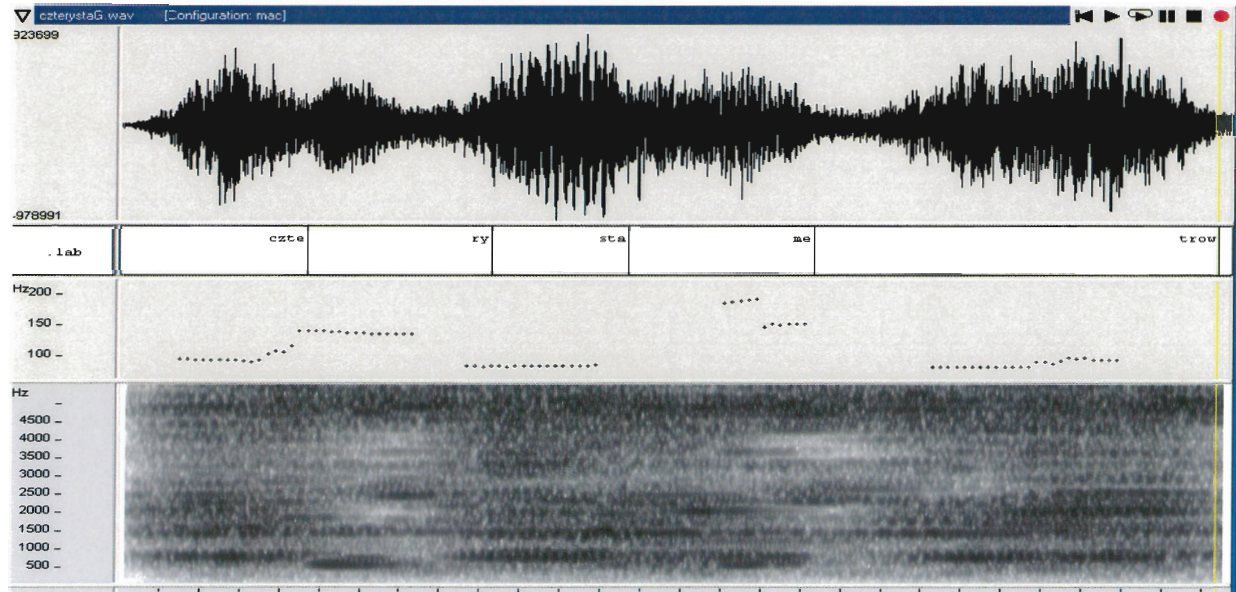
**SP18) *witamy*** (brak identyfikacji głosu)

Wypowiedź neutralna, raczej rzeczowa, formalna (kontur intonacyjny płaski, tempo dość szybkie, prawidłowo zaakcentowana sylaba *ta*). Długość wypowiedzi 430 ms, sylaba akcentowana *ta* – 90 ms.



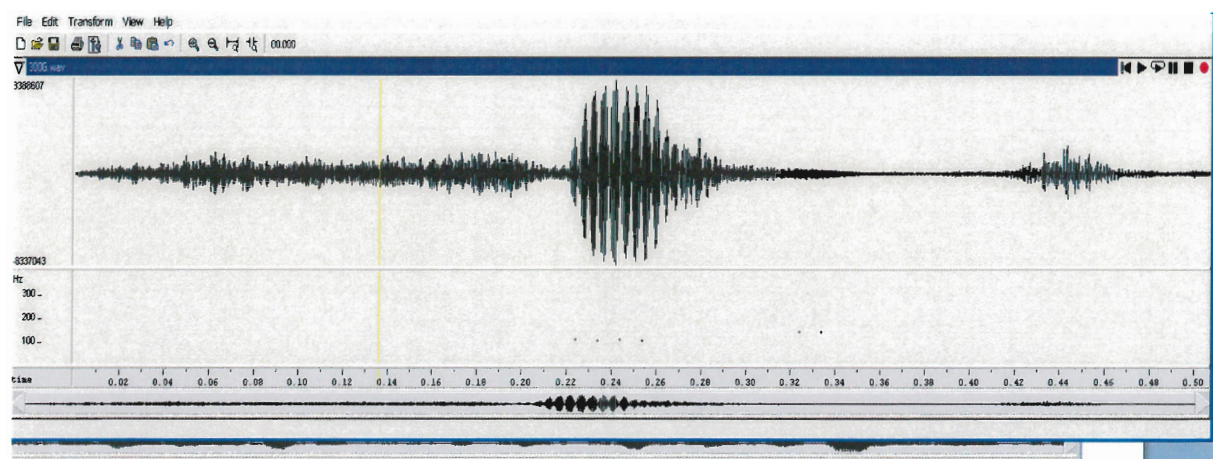
SP19) *czterysta metrów* (por. Artur Ziętek)

Wyraźne akcentowanie wyrazu *czterysta* (wyraz metrów wypowiedziany bez emfazy).



SP20) *Trzysta!* (por. Artur Ziętek)

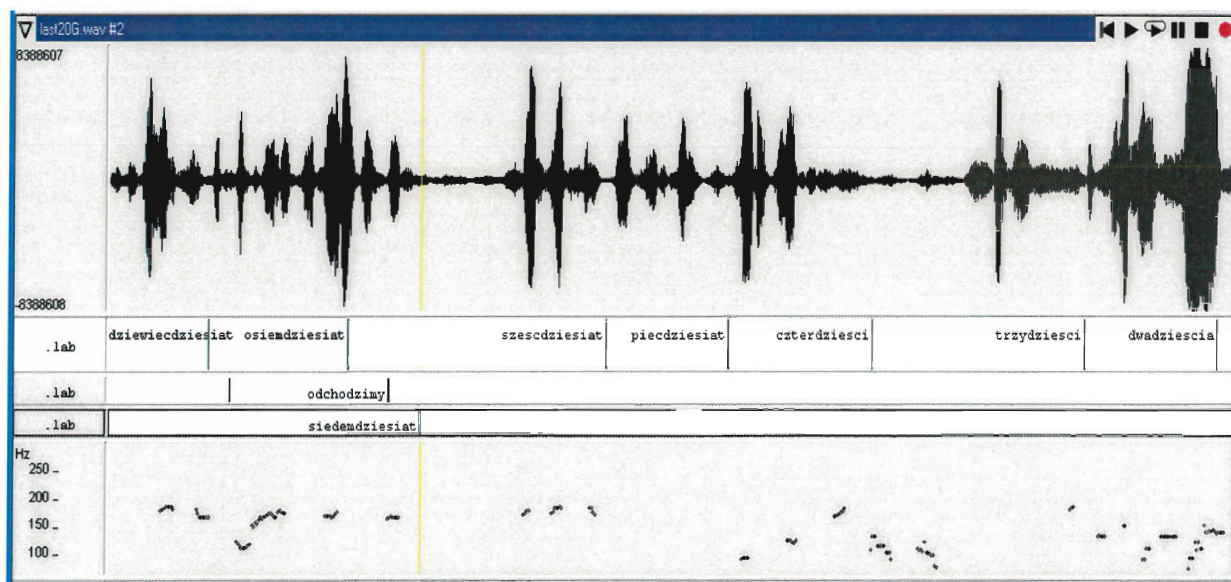
Wyraźne akcentowanie, emfaza sylaby *trzy*. Duża zmiana parametru F0 wskazująca na pewność przekazywanej informacji.



SP21) dyktowanie wysokości: *dziewięćdziesiąt, osiemdziesiąt, siedemdziesiąt, sześćdziesiąt, pięćdziesiąt, czterdzieści, trzydzieści, dwadzieścia*

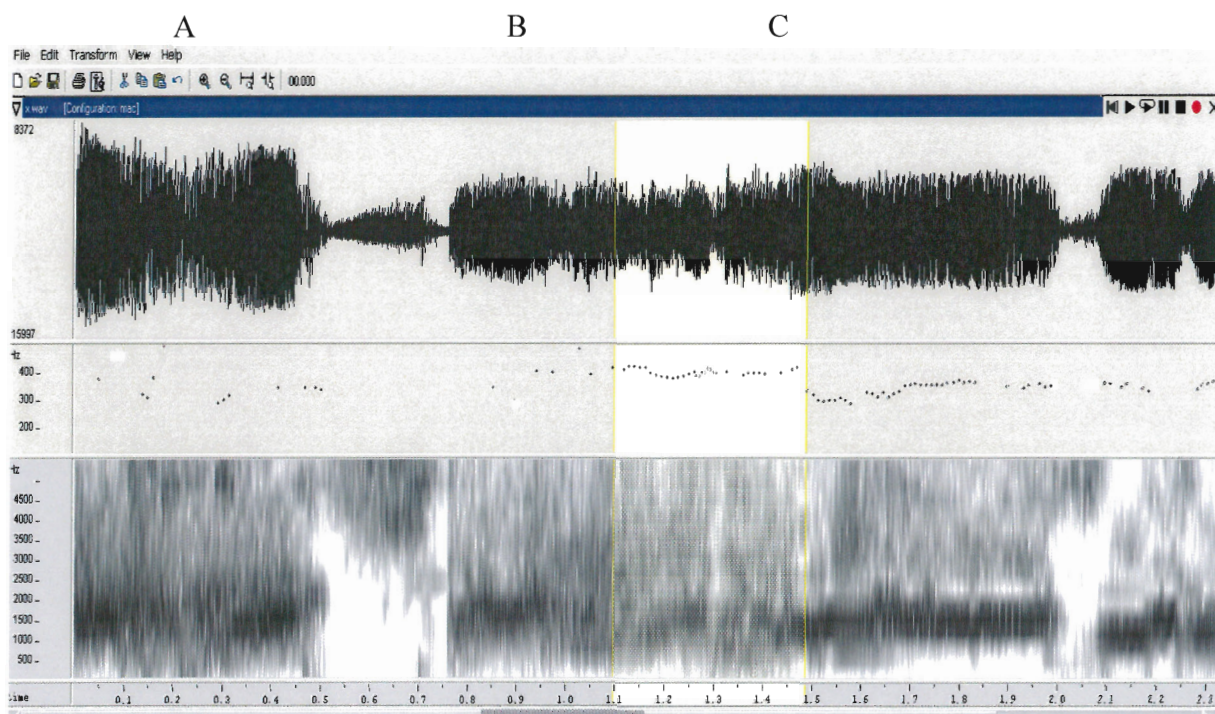
(por. Artur Ziętek)

Dyktowane wysokości bardzo szybko, jedna po drugiej (frazy o długości 600ms – 100ms). Ekstremalne zaakcentowanie daje się zauważyć w odczycie przez por. Artura Ziętka ostatniej wartości: **dwadzieścia!**



**SP22) Kurwa mać aaa kurwa ( Krzyk )**

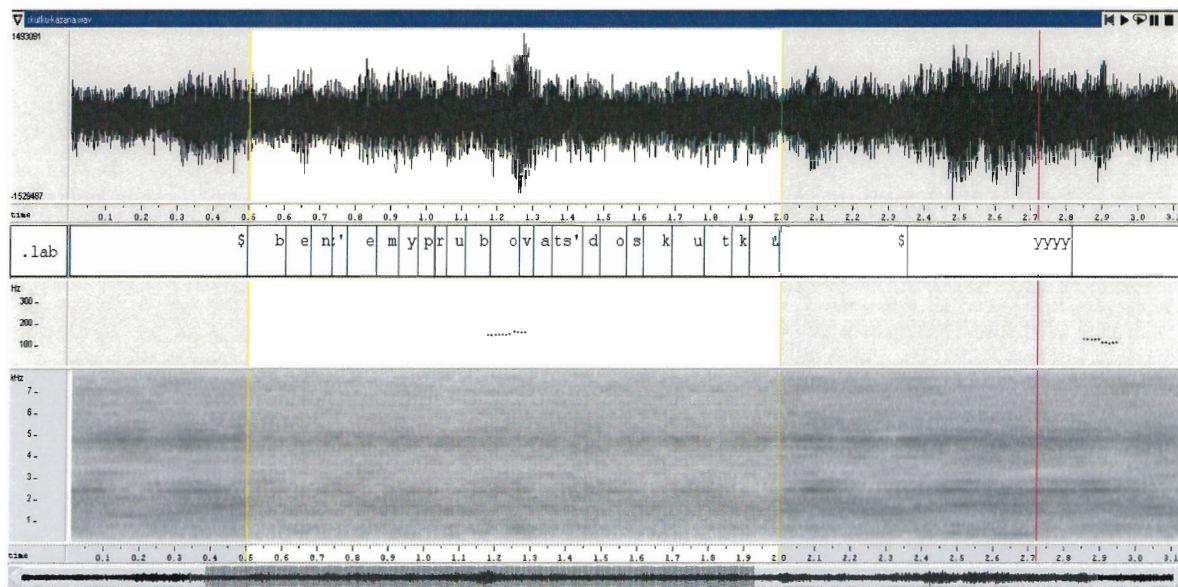
We fragmencie (a) wartość parametru F0 wynosi około 280 Hz, we fragmencie (b) wartość parametru F0 dochodzi do 400 Hz (krzyk), we fragmencie (c) obejmującym 2 samogłoski częstotliwość podstawowa waha się w zakresie 360–390 Hz.



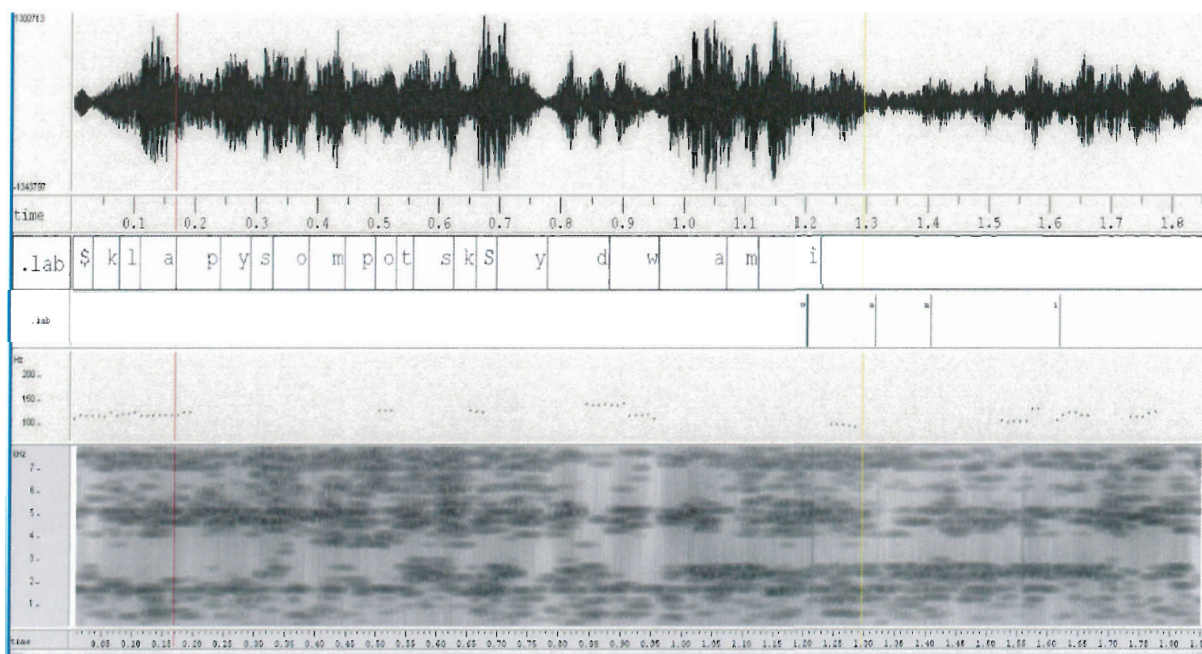
## Dowód 5

SP23) 8:40:48 *będziemy próbować do skutku.*

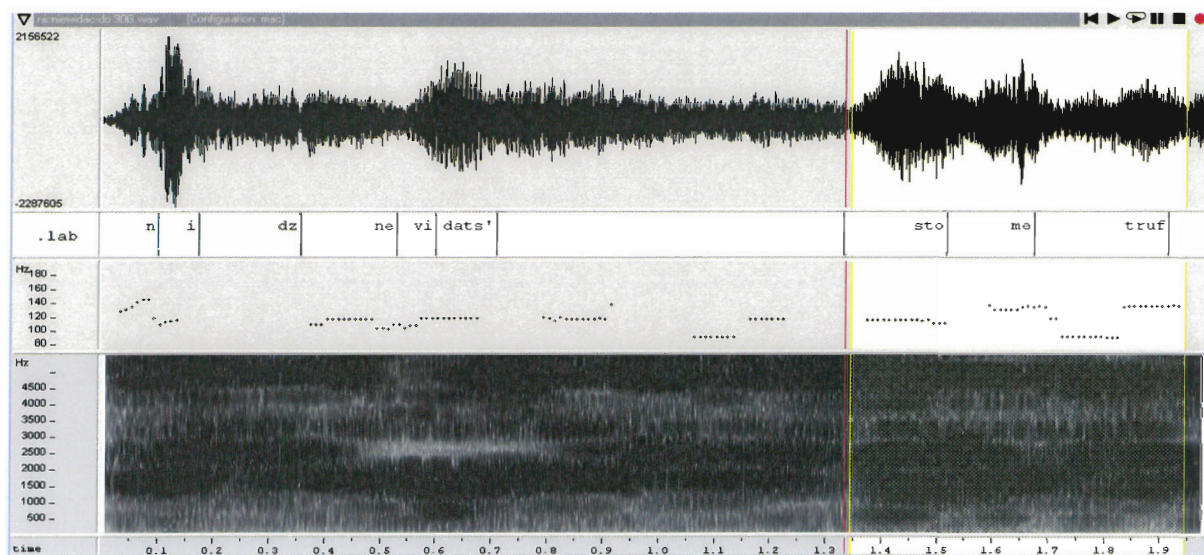
Odpowiedź: yyy... ..wyrażna wątpliwość kpt. pil. Arkadiusza Protasiuka



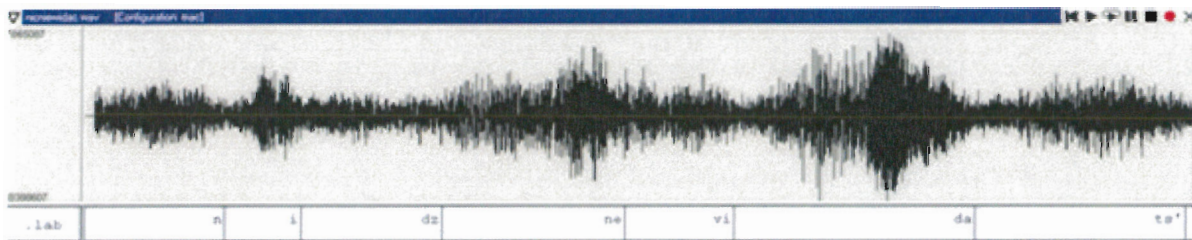
SP24) 8:39:12 *klapy są pod skrzydłami*



SP25) fraza: *nic nie widać*



SP25 a)





## Dowód 6

### Wypowiedzi zidentyfikowane.

- kpt. pil. Arkadiusz Protasiuk  
Bardzo wysoki głos (prawdopodobnie uwarunkowany posturą mówcy). Tempo wypowiedzi szybkie.
- mjr pil. Robert Grzywna – barwa  
Również stosunkowo wysoki głos. Tempo wypowiedzi również dość szybkie.
- por. Artur Ziętek  
Zmienny zakres parametru F0 przeciętny. Tempo wypowiedzi bardzo zróżnicowane. Styl wypowiedzi związany z podkreśleniem wagi odczytu pomiaru, zwłaszcza odległości (emfaza, korekty).
- chor. Andrzej Michalak  
Mało próbek wypowiedzi. Głos raczej monotony, przynajmniej na podstawie dostępnych nagrań.

### Tabela 1

Pomiar częstotliwości podstawowej, w kolumnie 1- wartości parametru F0 z wypowiedzi kpt. pil. Arkadiusza Protasiuka (fragment wypowiedzi: *Basiu wyszła mgla.*), w kolumnie 3- wartości parametru F0 z wypowiedzi mjr. pil. Roberta Grzywny (fragment wypowiedzi: *i w dół panie Arek....*), w kolumnie 5- wartości parametru F0 z wypowiedzi por. Artura Zietka (fragment wypowiedzi: *czterysta.*), w kolumnie 7- wartości parametru F0 z wypowiedzi chor. Andrzeja Michalaka (fragment wypowiedzi: *mały gaz.*).

1	2	3	4	5	6	7	8
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	154.715301514	1.0	0.0	0.0	0.0	0.0
0.0	0.0	159.599594116	1.0	0.0	0.0	148.628692627	1.0
0.0	0.0	161.198684692	1.0	0.0	0.0	152.259017944	1.0
0.0	0.0	160.942886353	1.0	0.0	0.0	152.988204956	1.0
0.0	0.0	163.751937866	1.0	0.0	0.0	156.481719971	1.0
0.0	0.0	163.957107544	1.0	0.0	0.0	155.97454834	1.0
0.0	0.0	164.481155396	1.0	0.0	0.0	154.966049194	1.0
0.0	0.0	164.80456543	1.0	0.0	0.0	153.784118652	1.0
0.0	0.0	164.77835083	1.0	184.469650269	1.0	161.075134277	1.0



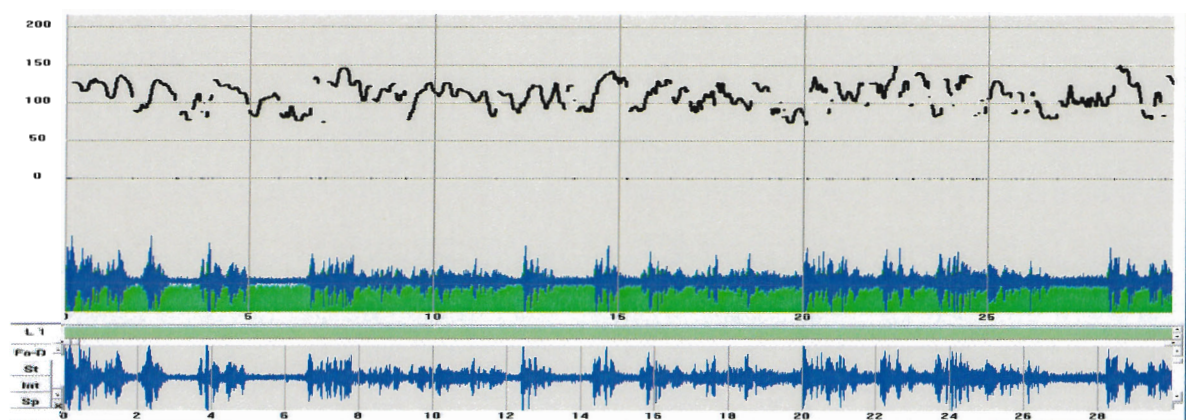


*Opinia Zespołu Biegłych w sprawie Po. Śl. 54/10*

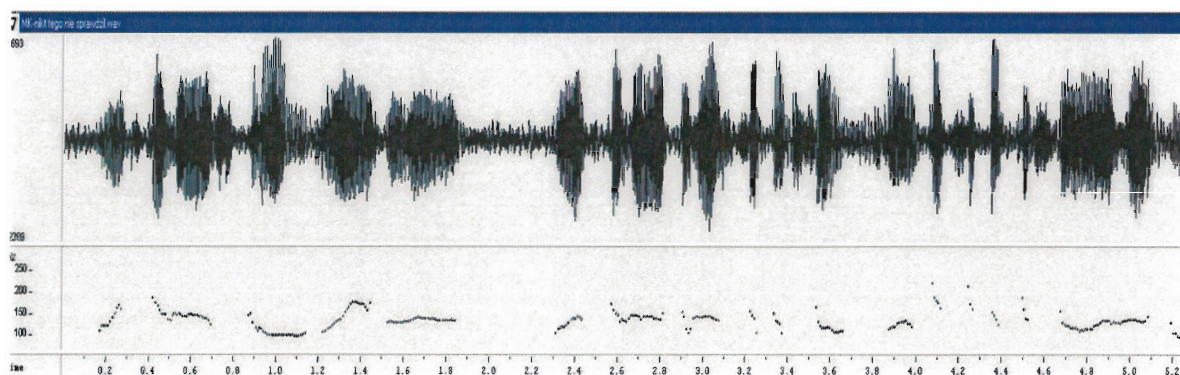
---

0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
182.331451416	1.0	0.0	0.0
175.330230713	1.0	0.0	0.0
173.447357178	1.0		
171.277160645	1.0		
167.724685669	1.0		

SP 26) Nagrania porównawcze głosu gen. Andrzeja Błasika. Wartość parametru  $F_{0min}=69\text{Hz}$



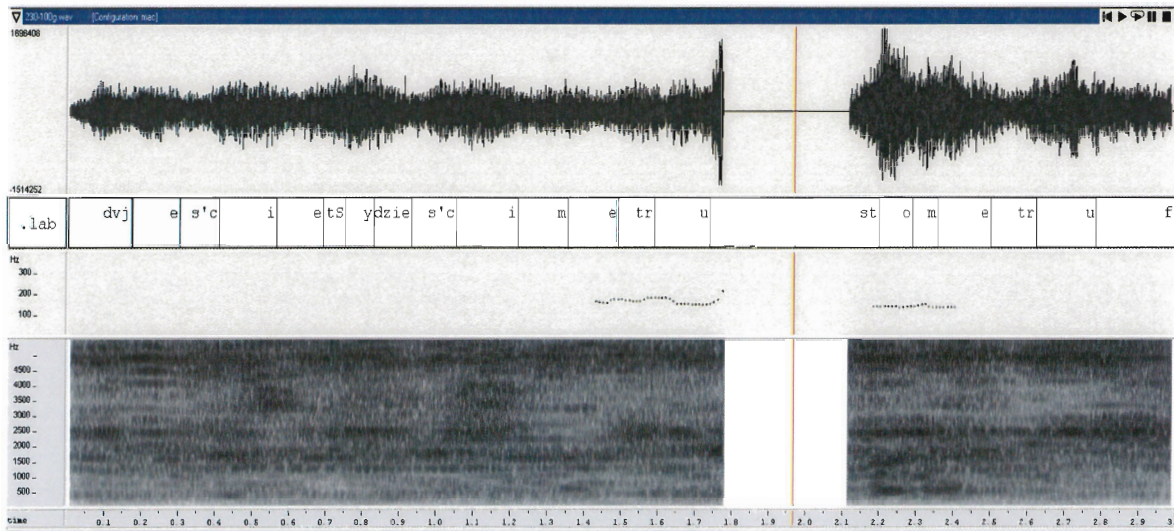
SP27) Nagranie porównawcze. Fraza wypowiedziana przez dyr. Mariusza Kazanę: *Nikt tego nie sprawdził i nikt nie chce tego potwierdzić, ale tak się faktycznie mówi..*



## Dowód 7

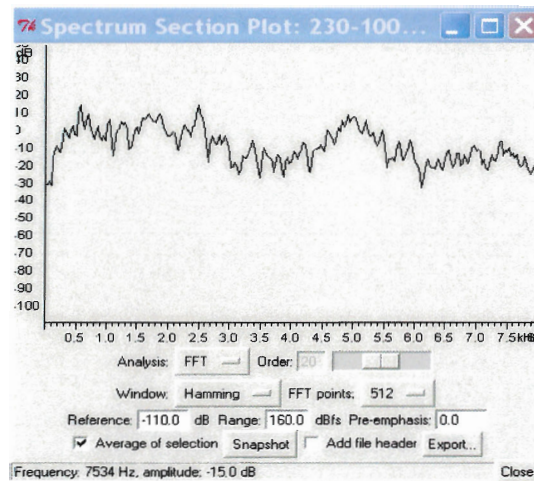
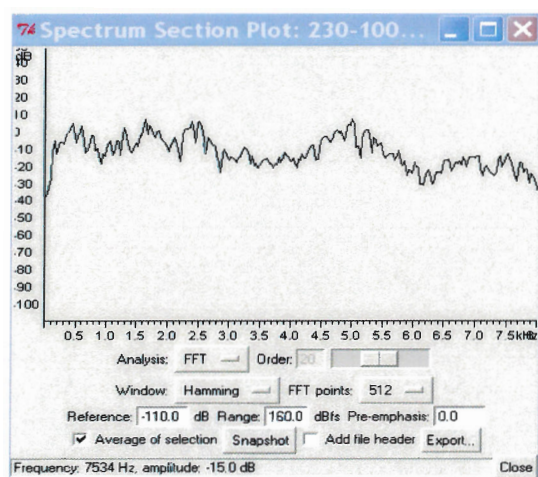
SP28) 8:40:26 *dwieście trzydzieści metrów - sto metrów.*

Frazy informacyjne, wypowiedziane pewnie z przekonaniem o prawidłowości odczytu. Wskazuje na to właściwa akcentuacja *dwieście trzydzieści metrów* oraz *sto metrów*, brak nieciągłości w wypowiedziach, iloczasy sylabiczny w normie mowy płynnej.



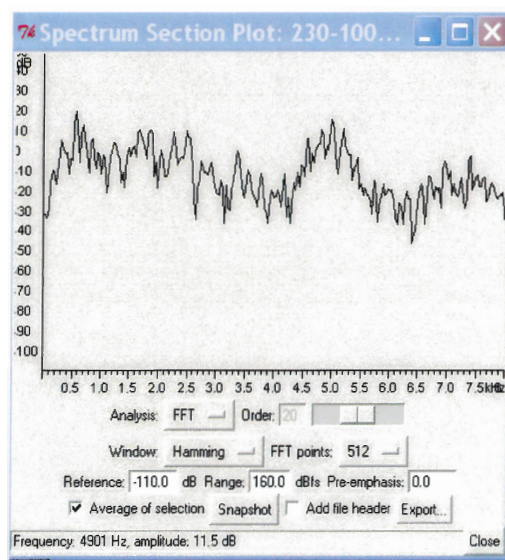
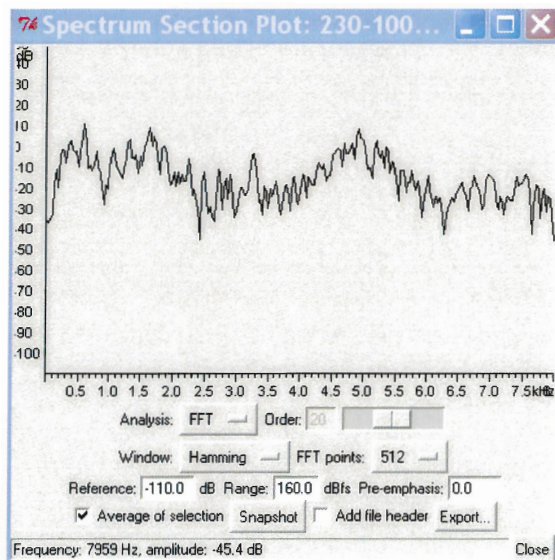
LT1: wyraz *metrów* z frazy (230 m)

LT2: wyraz *metrów* (z frazy 100 m)

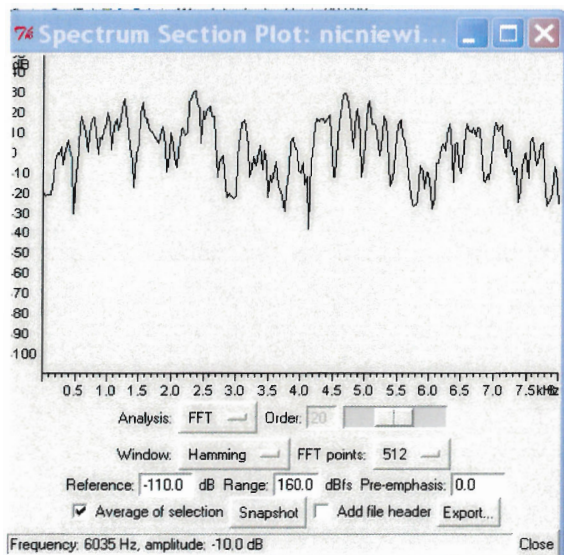


Widma LTAS z smogłoski *e*

**LT3:** samogłoska *e* z frazy *dwieście trzydzieści metrów*    **LT4:** samogłoska *e* z frazy *sto metrów*

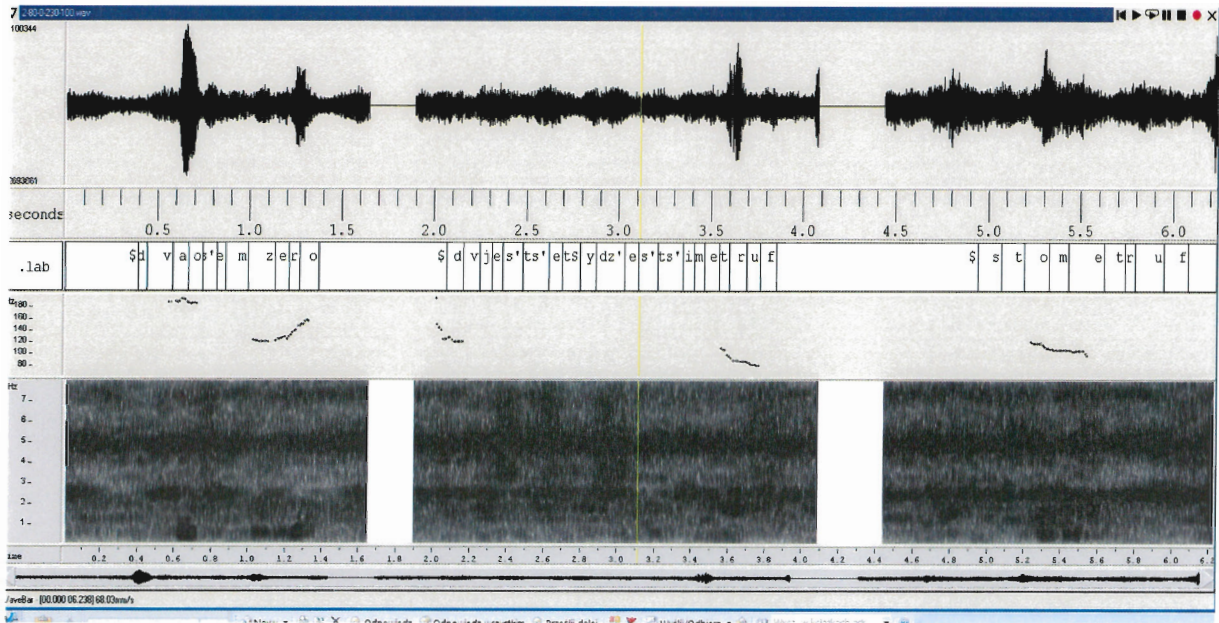


**LT5:** samogłoska *e* z frazy *nic nie widać*

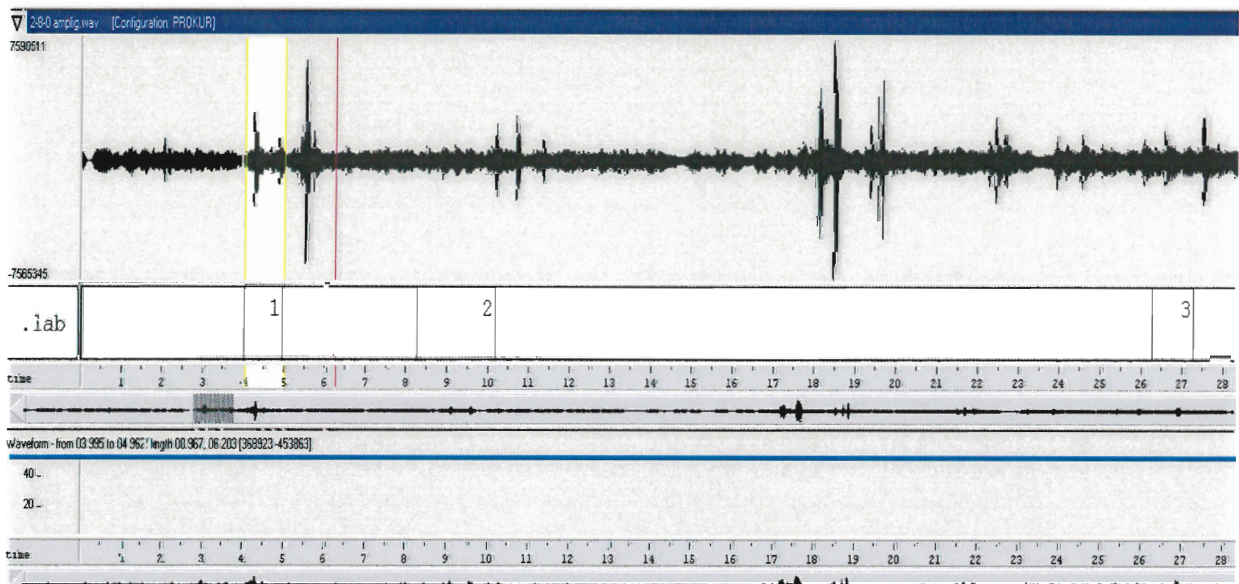


## Dowód 8

SP29) Frazy: *dwa osiem zero, dwieście trzydzieści metrów, sto metrów*

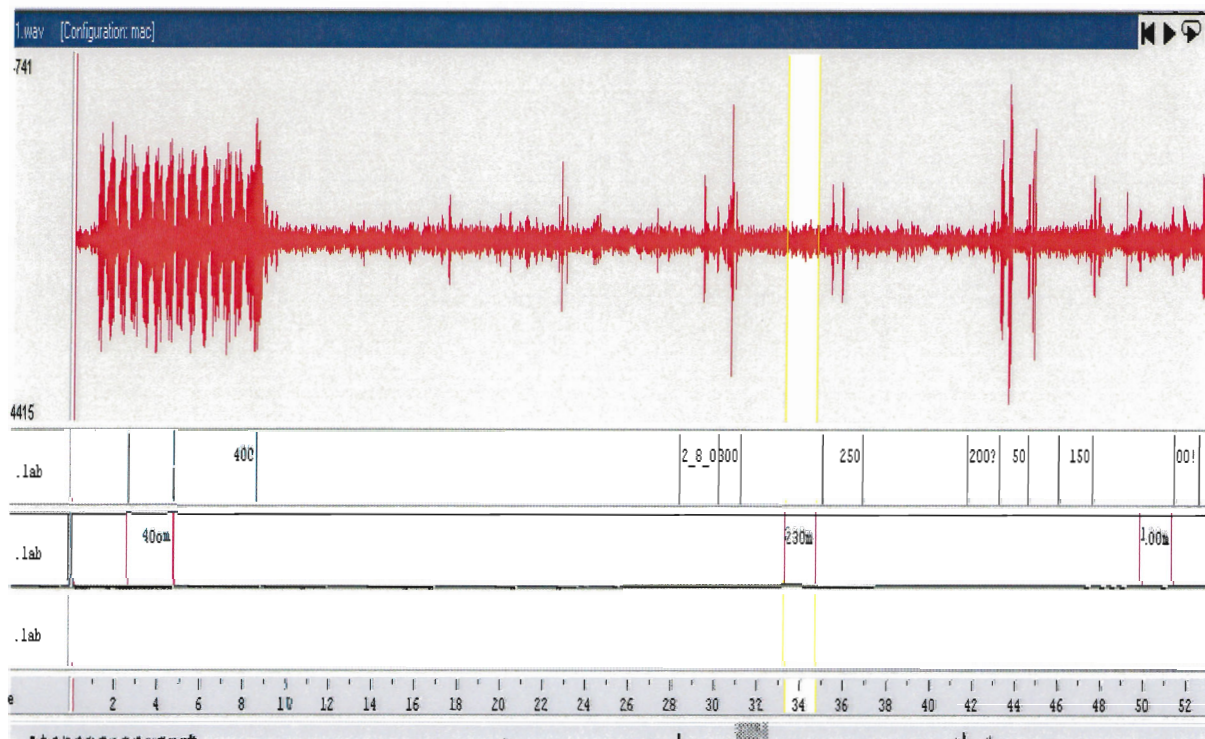


SP29a) Frazy ze spektrogramu 29 w całym ciągu wypowiedzi: 1) *dwa osiem zero*, 2) *dwieście trzydzieści metrów*, 3) *sto metrów*.

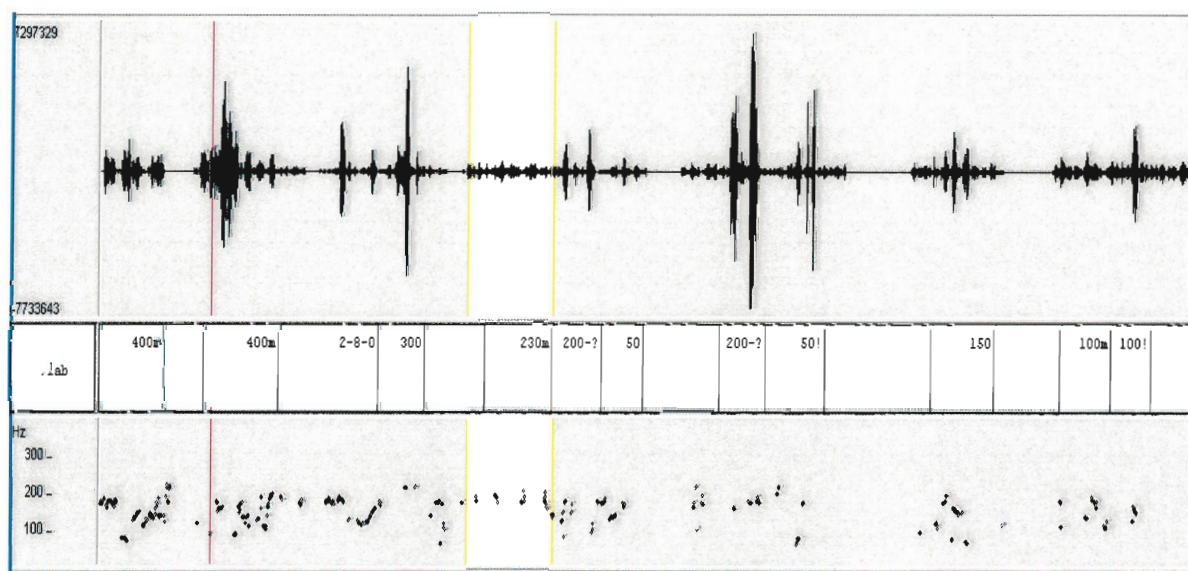




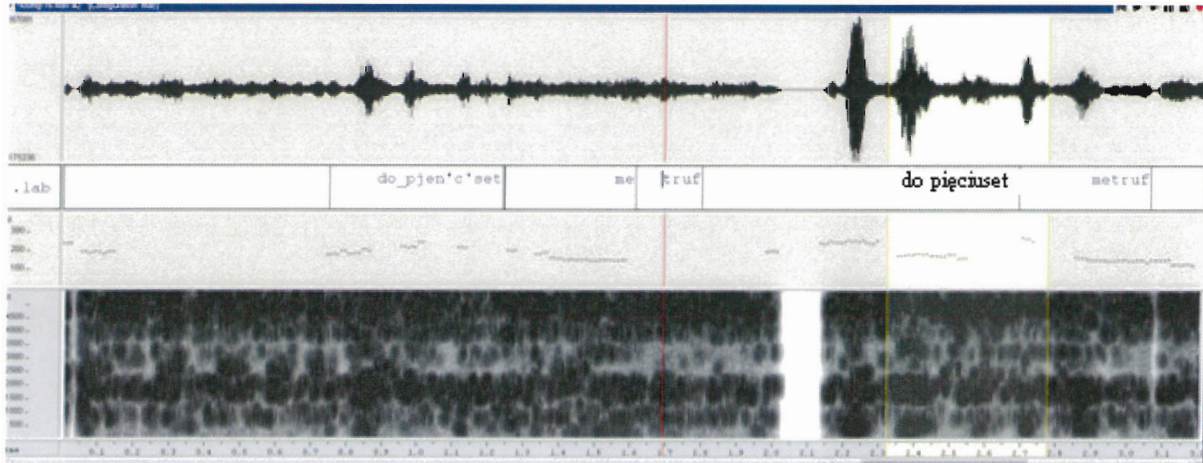
SP30) odczytywanie wysokości



S31) *sto!* Zdecydowane



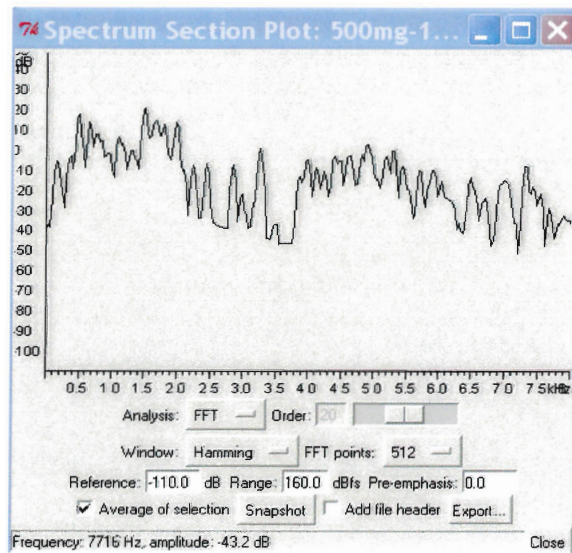
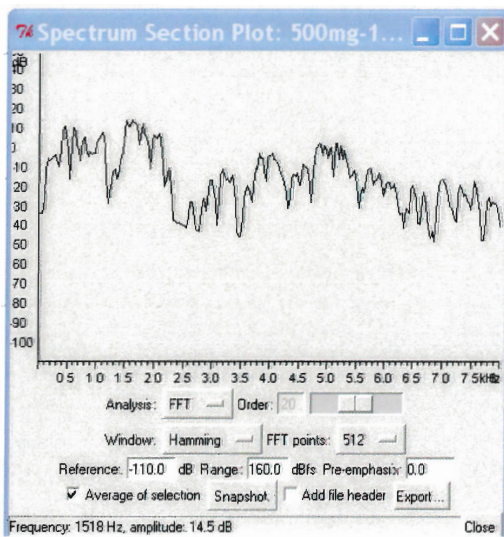
SP 32) frazy *do pięćset metrów zjedziemy* (por. Artur Ziętek) i *do pięciuset metrów* (mjr pil. Robert Grzywna)



## Widma LTAS

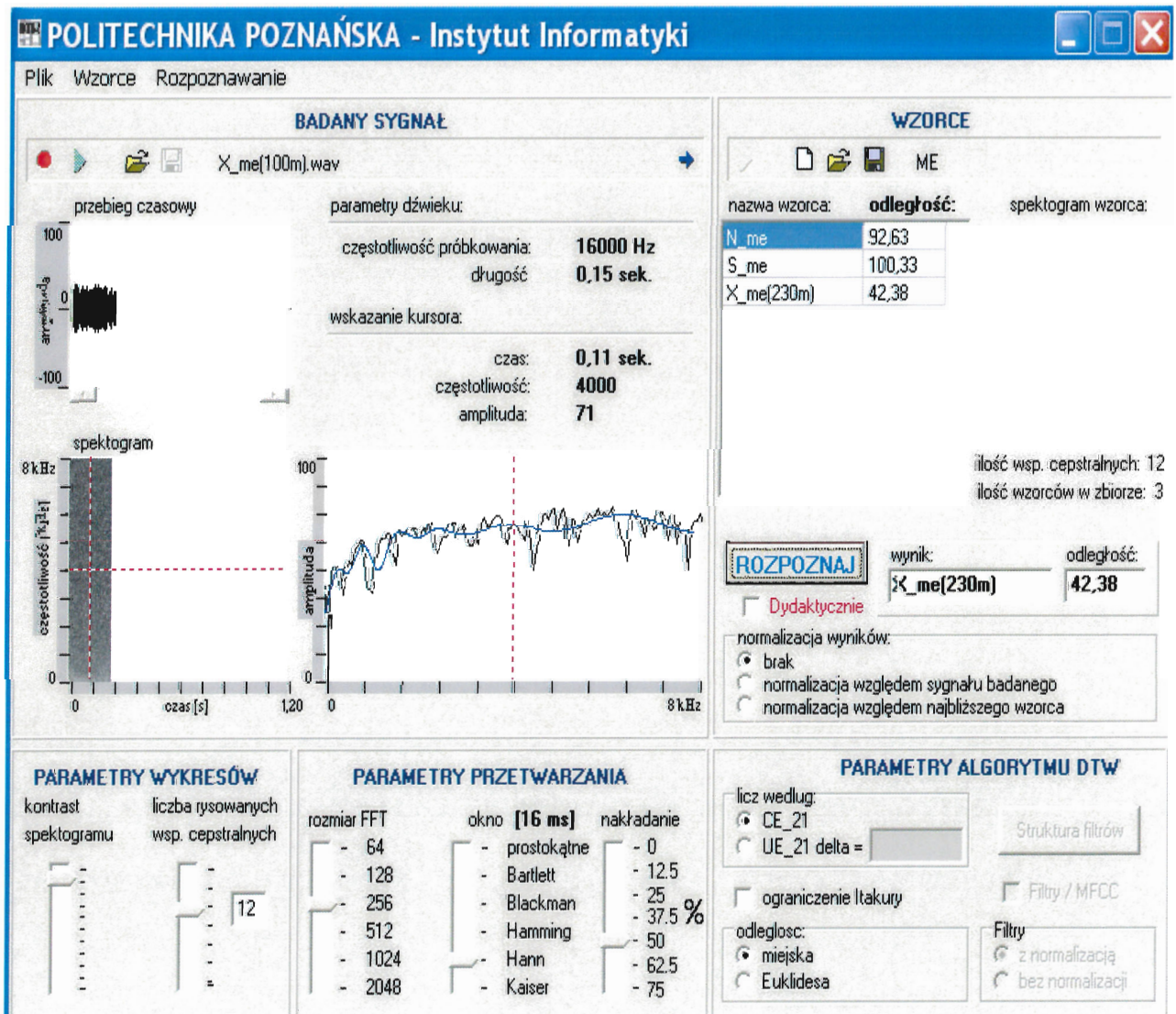
LT6: z frazy *metrów*, por. Artur Ziętek

LT7: z frazy *metrów*, mjr pil. Robert Grzywna



**Wynik DTW1: porównanie**

- 1) *me* (z głosu niezidentyfikowanego z frazy *sto metrów*, oznaczonego jako X\_me (100),
- 2) *me* (z głosu niezidentyfikowanego z frazy *dwieście trzydzieści metrów*, oznaczonego jako X\_me (230),
- 3) *me* (z wypowiedzi mjr pil. Roberta Grzywny z frazy *pięciuset metrów*, oznaczonego jako S\_me),
- 4) *me* z wypowiedzi por. Artura Ziętka (z frazy *pięćset metrów* oznaczonego, jako N\_me).



## Badany sygnał

1) X\_me(100m)

## Wzorce:

1) X\_me(230)

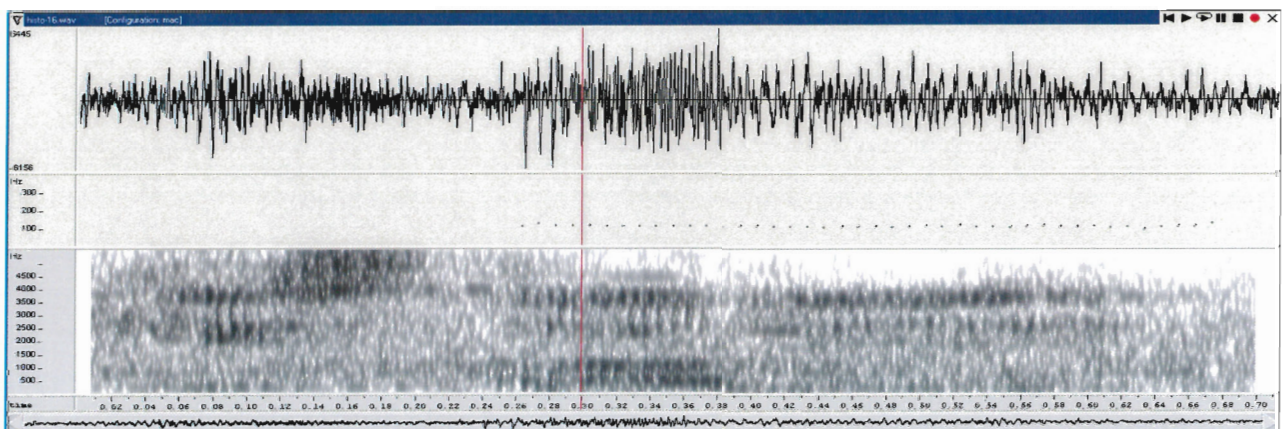
2) S\_me

3) N-me

**Odległość minimalna:** między sygnałami: X\_me (230m) i X\_me (100m)

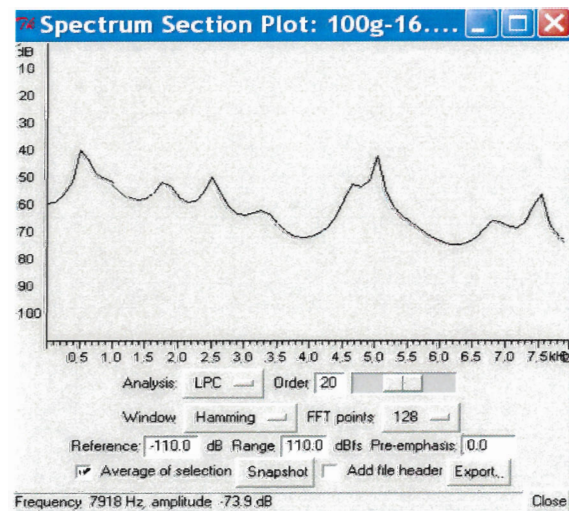
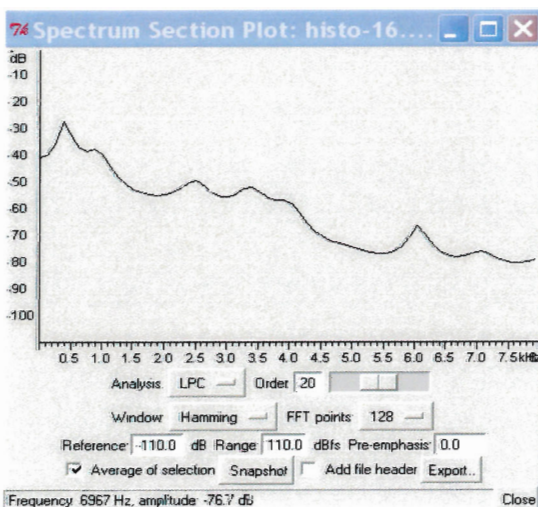
## Dowód 9

**SP33).** Spektrogram sylaby *sto* z frazy *historii* (z nagrania porównawczego gen. Andrzeja Błasika

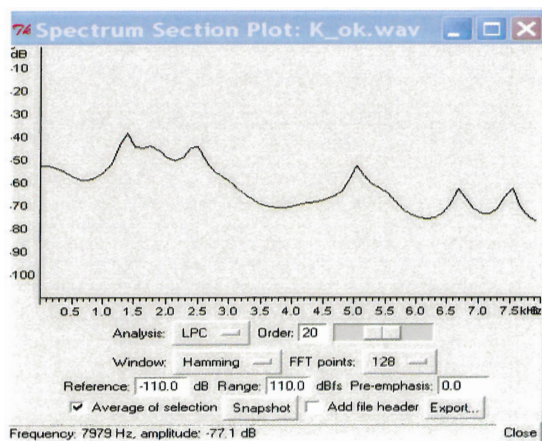


**LT8:** z samogłoski *o* z frazy *historii*

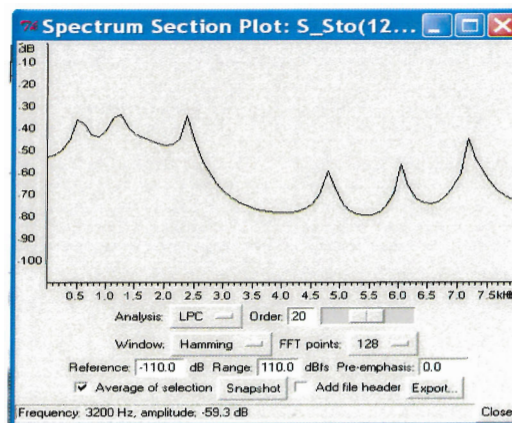
**LT9:** z samogłoski *o* z frazy *sto metrów*



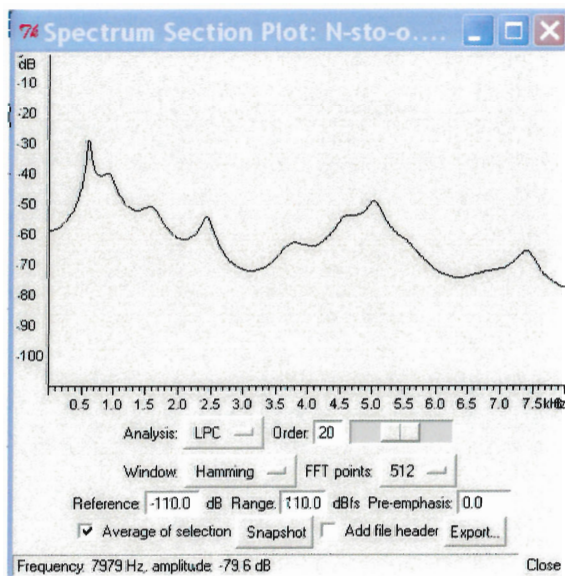
LT10: z samogłoski *o* z frazy *co*  
(dyr. Mariusza Kazany)



LT11: z samogłoski *o* (głos mjr pil. Roberta  
Grzywny z frazy *sto dwudziesty ósmy poziom*)



LT12: z samogłoski *o* z frazy *sto* (por. Artura Ziętka)



## Dowód 10

DTW2: Obliczanie odległości sylab *sto* (z fraz i głosów poniżej wymienionych).

8:08:59	sto dwudziesty ósmy poziom	S_sto (128)
8:12:41.0	do dziewięć sto	A_sto (do9-sto)
8:13:46.5	dziesięć stopni	A_sto (10stopni)
8:16:53.5	o której te uroczystości	S_sto (uroczystości)
8:27:20,5	dwa stopnie siedem cztery	A_sto (2stopnie)
8:27:24,5	dwa stopnie siedem cztery pięć	I_sto (2stopnie)
8:40:40	sto pięćdziesiąt	N_sto (150)
8:40:45,5	sto	N_sto
8:40:51	stolast	N_stolast
historii	sto	B-sto

**POLITECHNIKA POZNAŃSKA - Instytut Informatyki**

Plik Wzorce Rozpoznawanie

**BADANY SYGNAŁ**

przebieg czasowy

parametry dźwięku:

częstotliwość próbkowania: **16000 Hz**  
długość: **0,21 sek.**

wskazanie kursora:

czas: **0,10 sek.**  
częstotliwość: **2562**  
amplituda: **75**

spektrogram

**WZORCE**

nazwa wzorca: **odległość:** spektrogram wzorca:

A_sto(2stopnie)k1.v	81,14
A_sto(10stopni)k.w	65,20
A_sto(do9-sto)k.wa	95,18
B-sto.wav	51,70
I_-sto(2stopnie)k2.v	78,92
N_sto(150).wav	62,88
N-sto.wav	60,09
N-sto-last.wav	52,96
S_Sto(128)k1.wav	95,54
S_sto(154)k.wav	linfinitv

ilość wsp. cepstralnych: 12  
ilość wzorców w zbiorze: 10

fp=16000 Hz | t=0,31 sek. | FFT=256 | n=50 % | o=Hann

**ROZPOZNAJ**    wynik:    odległość:

Dydaktycznie    **B-sto.wav**    **51,70**

normalizacja wyników:

brak  
 normalizacja względem sygnału badanego  
 normalizacja względem najbliższego wzorca

**PARAMETRY WYKRESÓW**

kontrast spektrogramu    liczba rysowanych wsp. cepstralnych

    12

**PARAMETRY PRZETWARZANIA**

rozmiar FFT    okno [16 ms]    nakładanie

- 64	- prostokątne	- 0
- 128	- Bartlett	- 12,5
- 256	- Blackman	- 25
- 512	- Hamming	- 37,5 %
- 1024	- Hann	- 50
- 2048	- Kaiser	- 62,5
		- 75

**PARAMETRY ALGORYTMU DTW**

licz według:

CE\_21    Struktura filtrów

UE\_21 delta =

ograniczenie Itakury     Filtry / MFCC

odległość:

miejska    Filtry

Euklidesa     z normalizacją

    bez normalizacji