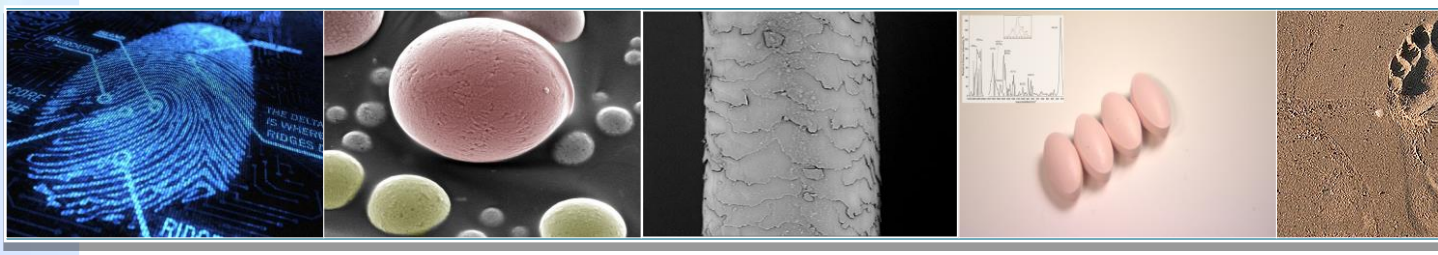
Instytut Fizyki Doświadczalnej
Wydział Matematyki, Fizyki i Informatyki
UNIWERSYTET GDAŃSKI

DLF
DYDAKTYCZNE
LABORATORIUM
FIZYCZNE

Ćwiczenie 13

Analiza sygnałów mowy



I. Cel ćwiczenia

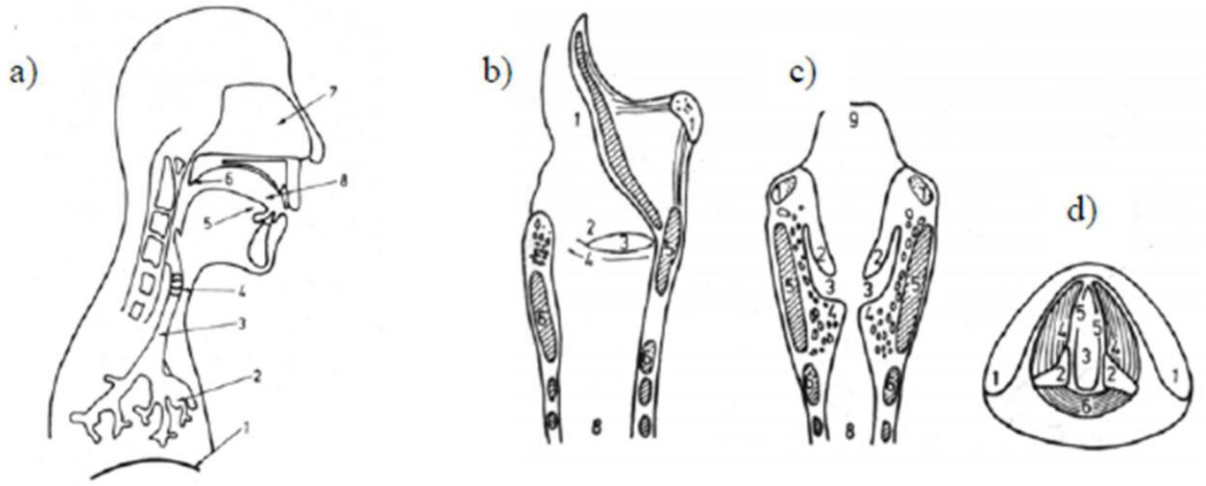
Celem ćwiczenia jest:

- przybliżenie zagadnienia pomiaru cech charakterystycznych fonemów-samogłosek, procesu generowania sygnału mowy, tonu podstawowego, zależności pomiędzy narządami artykulacji a formantami w postaci czworoboku formantowego, obecności cech dystynktywnych, osobniczych na tle danych bazowych dla danego języka,
- poznanie programów do analizy sygnałów dźwiękowych: Praat i Audacity.

II. Wstęp teoretyczny

Proces komunikacji z fizycznego punktu widzenia polega na generowaniu oraz obiorze bodźców akustycznych. Narząd mowy człowieka jest układem, który umożliwia generowanie szerokiej gamy dźwięków poprzez sterowanie strumieniem powietrza wypływającego z płuc. Daje nam to możliwość kodowania użytecznej informacji w postaci chwilowych zmian ciśnienia. Narząd słuchu umożliwiający odbiór tych informacji odpowiada za analizę częstotliwościową docierających do niego fal akustycznych i zmianę ich na ciągi impulsów neuronowych. Z punktu widzenia analizy sygnałów, przebieg w czasie dowolnej wielkości fizycznej może być traktowany jako sygnał. Przykładem sygnału jest fala akustyczna powstająca na skutek drgań ciała w ośrodku sprężystym, ponieważ wprowadza ona chwilowe zmiany ciśnienia akustycznego. Mowa zatem jest sygnałem akustycznym, do którego można zastosować szereg metod analizy sygnałów. Sygnały można podzielić na zdeterminowane dające się opisać analitycznie oraz niezdeterminowane zwane losowymi, które nie dają się opisać zależnościami matematycznymi. W sygnałach okresowych przebieg powtarza się co jakiś czas T zwany okresem. Najprostszym sygnałem okresowym jest sygnał sinusoidalny (harmoniczny). Znacznie częściej mamy do czynienia z sygnałami okresowymi złożonymi z wielu sygnałów sinusoidalnych, które nazywane są sygnałami poliharmonicznymi. Przykładem takich sygnałów są dźwięki muzyki, a także dźwięczne części mowy (samogłoski i spółgłoski dźwięczne) wytwarzane przy udziale drgań wiązadeł głosowych. Sygnały transjentowe, czyli takie, które można opisać za pomocą odpowiednich funkcji zmiennych, obejmują szeroką grupę sygnałów nieokresowych. Charakteryzują się one zmienną w czasie amplitudą o charakterze narastania lub zanikania. Przykładami takich sygnałów są dźwięki muzyki i mowy, które pozwalają na wydzielenie odcinków o nieustalonym przebiegu w czasie. Kolejną grupę sygnałów tworzą sygnały niezdeterminowane (losowe). Ich losowość w ogólności oznacza, że nie da się przewidzieć wartości ich parametrów w żadnej przyszłej chwili czasu. Jednym z najczęściej spotykanym i stosowanym sygnałem losowym (np. w badaniach słuchu) jest tak zwany szum biały. W sygnale tym występuje nieskończenie wiele składowych sinusoidalnych, których częstotliwości obejmują cały zakres słyszalny (do około 20 kHz). Dzięki występowaniu niewielkich odcinków czasu w sygnałach mowy, które odpowiadają w przybliżeniu poszczególnym głoskom, można w wielu sytuacjach traktować je jako sygnały tego typu. Głoskę intuicyjnie pojmować można jako fonetyczny odpowiednik litery. Podstawowymi elementami alfabetu fonetycznego są tzw. fonemy, czyli najmniejsze segmenty (odcinki czasowe) sygnału mowy pozwalające na odróżnienie znaczenia. Fonemy stanowią zatem podstawowe elementy wypowiedzi, a ich zbiory układają się w wyrazy i zdania.

Sygnał mowy jest splotem parametrów układu artykulacji – traktu głosowego (jama ustna, nosowa, język itp.) i tonu krtaniowego (charakterystyczna częstotliwość pobudzenia F_0). Zależność poziomu ciśnienia sygnału mowy od częstotliwości oznacza, ze w określonych zakresach częstotliwości, składowe widma dźwięków mowy przybierają wartości znacznie wyższe niż w pozostałych zakresach częstotliwości. Te zakresy częstotliwości, w których składowe widma przyjmują maksymalne wartości nazywa się formantami, zaś odpowiadające im częstotliwości – częstotliwościami formantowymi.
Za wytwarzanie mowy odpowiedzialny jest organ mowy, którego przekrój czynny został pokazany na Rysunku 1. Głównymi elementami tego organu są płuca, tchawica, krtań, gardło, nos, jama nosowa oraz usta. Leżąca powyżej krtani cześć drogi głosowej, nazywa się kanałem głosowym. Na skutek ruchów języka, warg i szczęk kształt przekroju poprzecznego kanału głosowego może się zmieniać, dzięki czemu możliwe jest wymawianie różnych głosek. Przestrzeń pomiędzy fałdami głosowymi, które otwierać się i zamykać nazywana jest głośnią. Dźwięk, który wytwarzany jest w trakcie wydostawania się powietrza z płuc przez fałdy głosowe, nazywa się tonem krtaniowym. Dźwięki takie nazywają się dźwięcznymi, są dźwiękami periodycznymi o dość małej częstotliwości podstawowej, zawierającymi harmoniczne ze znacznego zakresu częstotliwości.

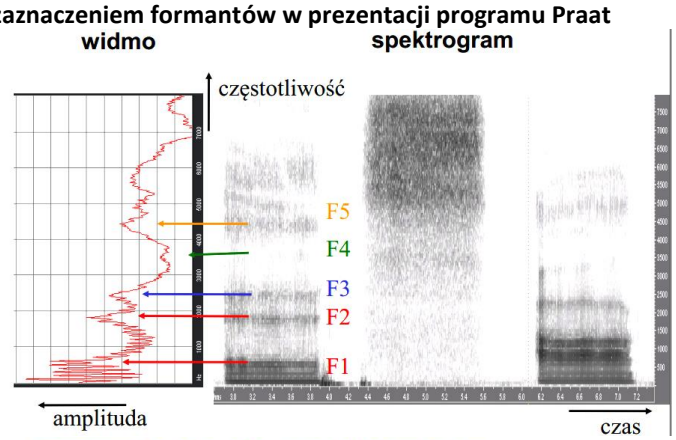


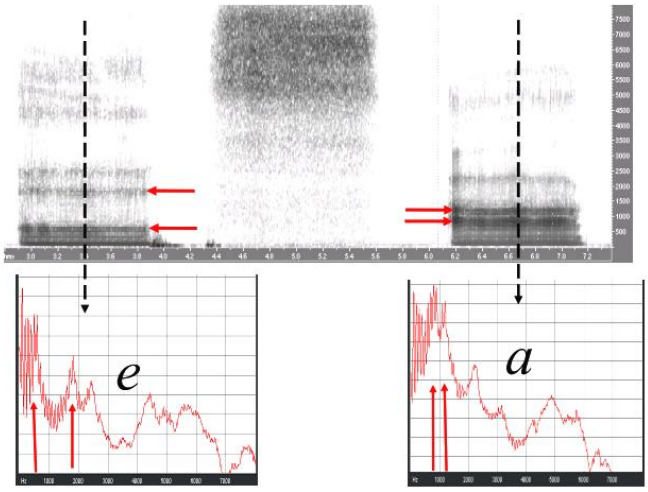
Rysunek 1. Część a) przedstawia organ mowy: 1-przepona, 2-płuca, 3-tchawica, 4- więzadła głosowe, 5-język, 6-języczek, 7-jama nosowa, 8-kanał głosowy. W części b) przedstawiono przekrój profilowy krtani: 1-nagłośnia, 2-fałd kieszonki krtaniowej Morganiego, 4-fałd głosowy, 5-chrząstka tarczowa, 6-chrząstka pierścieniowa, 7-kość gnykowa, 8-tchawica. W części c) przedstawiono przekrój czołowy krtani (oznaczenia jak w części b). Część d) przedstawia przekrój poziomy krtani: 1-chrząstka tarczowa, 2-chrząstki nalewkowe, 3-głośnia, 4-mięśnie głosowe, 5-wiązadła głosowe, 6-mięśnie międzynalewkowe.

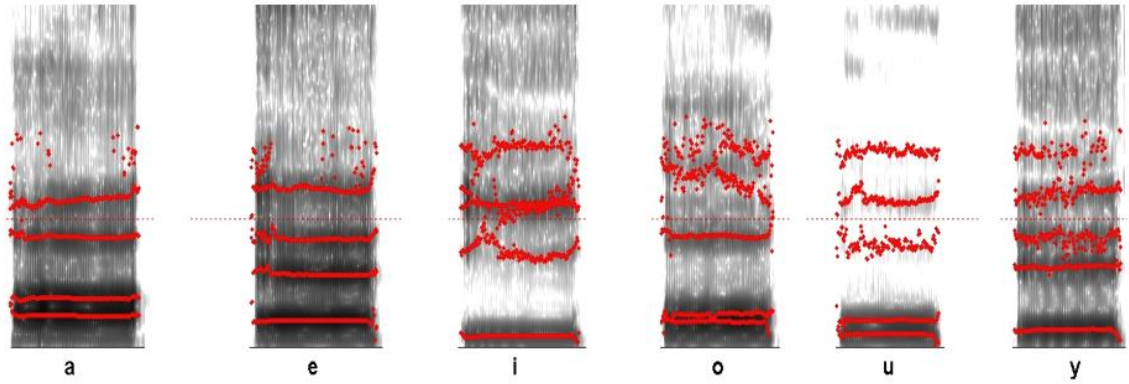
Na Rys.2 zobrazowano sposób w jaki zostają wygenerowane fonemy ’i’, ’a’ oraz ’u’. W krtani wytwarzany jest strumień o zadanej wysokości dźwięku, co przedstawia się w postaci widma tonu krtaniowego. Widać na nim ton podstawowy F_0 oraz jego harmoniczne. Następnie strumień przechodzi przez uformowany kanał głosowy. Formowanie to jest filtrem i można je przedstawić w postaci charakterystyki częstotliwościowej. Na wyjściu otrzymuje się fonem, z którego charakterystyki można wyczytać zarówno ton podstawowy jak i charakterystykę traktu głosowego.
Kanał głosowy zachowuje się jak układ filtrów (rezonatorów) o określonych częstotliwościach rezonansowych tak, że widmo tonu krtaniowego po przejściu przez układ tych filtrów

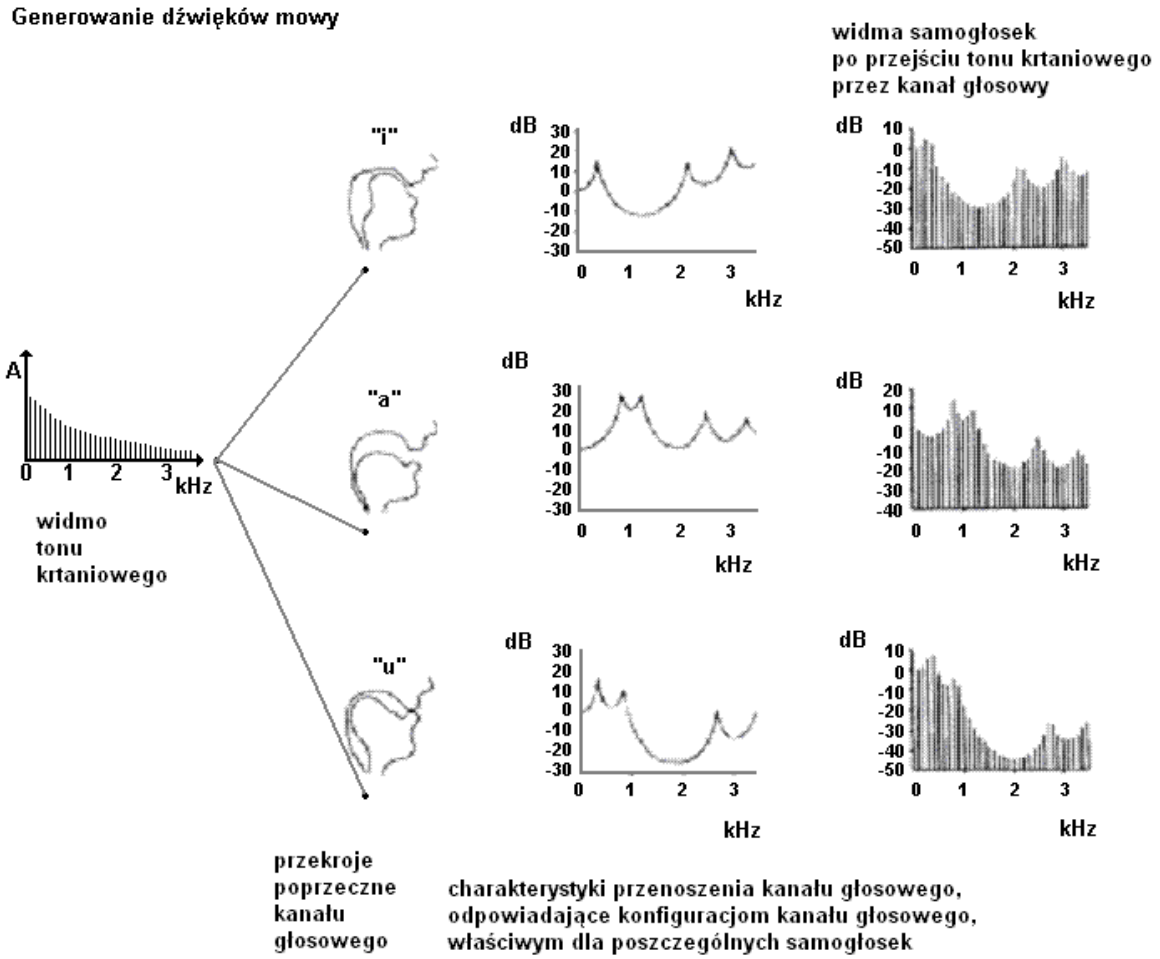
charakteryzuje się pewnymi maksimami i minimami lokalnymi. Te lokalne maksima nazywane są formantami i oznaczane są zwykle za pomocą litery F z liczbą oznaczającą kolejny numer formantu. Formant o najmniejszej częstotliwości jest nazywany pierwszym formantem (F1), następny drugim formantem (F2) itd. Częstotliwość środkowa każdego z formantów jest inna i ściśle związana z kształtem kanału głosowego.

Widmo i spektrogram z zaznaczeniem formantów w prezentacji programu Praat



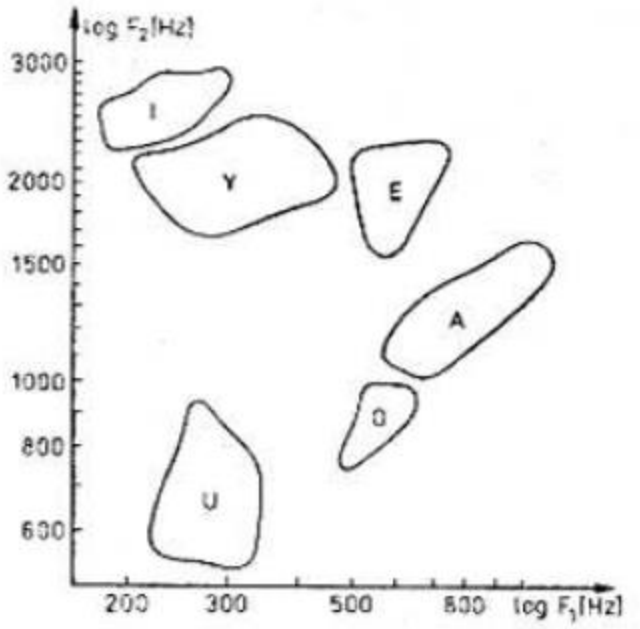






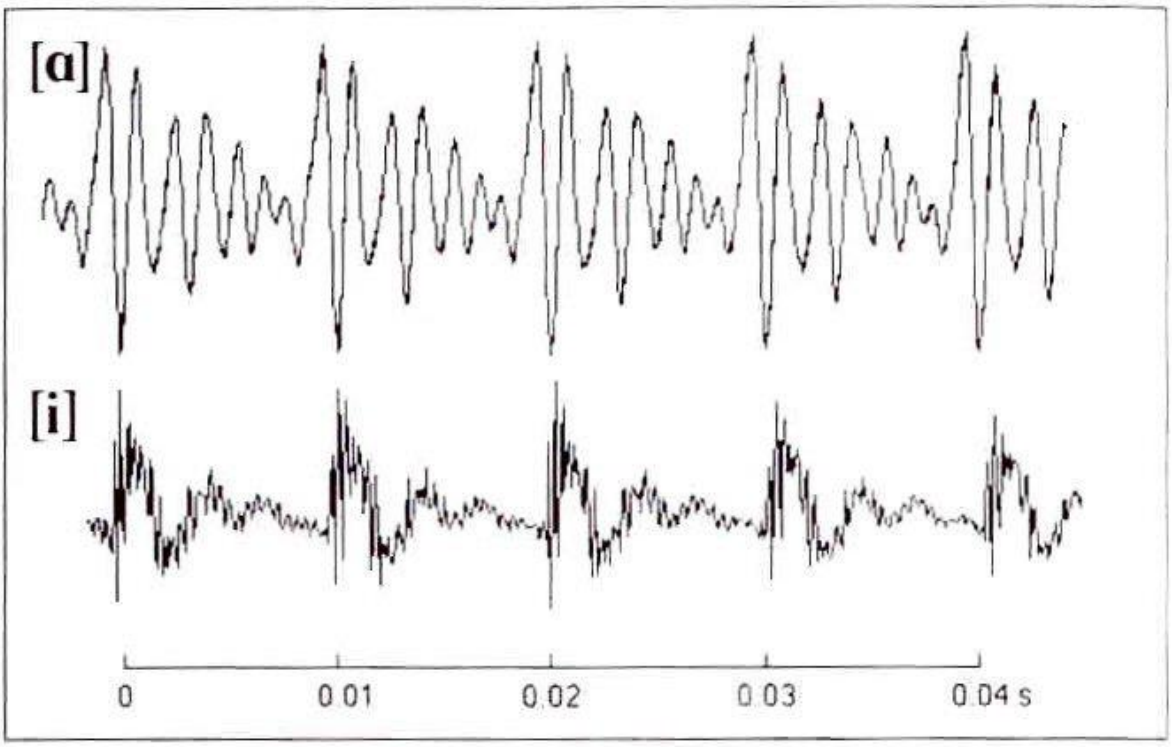
Rysunek 2. Generacja mowy.

W przypadku samogłosek, częstotliwość F1 jest wprost proporcjonalna do stopnia otwarcia jamy ustnej, zaś częstotliwość F2 rośnie w miarę przesuwania się miejsca najwyższego wzniesienia języka w kierunku warg (czyli od /u/ do /i/). Bezwzględne wartości częstotliwości poszczególnych formantów wykazują dość duże zróżnicowanie międzyosobnicze. Jednak ich względne wartości nie zależą od indywidualnych cech głosowych. Również poziomy poszczególnych formantów (wyrażone względem formantu F1) wykazują stałe w przybliżeniu wartości. Gdy na jednej płaszczyźnie umieścimy punkty odpowiadające formantowi F1 i F2 (osie odpowiednio OX i OY) samogłosek, okazuje się że ich umiejscowienie nie jest przypadkowe. Punkty odpowiadające formantom tej samej głoski zlokalizowane są blisko siebie i tworzą odseparowane od innych głosek obszary Rys.3.



Rysunek 3. Obszary lokalizacji poszczególnych głosek jezyka polskiego na płaszczyznie pierwszego i drugiego formantu (wartości w Tabeli Uzupełnienia).

Oprócz formantów, obraz spektrograficzny samogłosek charakteryzuje się obecnością **pionowych prążków** obrazujących drgania strun głosowych. Odstępy między dwoma kolejnymi prążkami odpowiadają **częstotliwości podstawowej** danego drgania złożonego (Rys. 4).



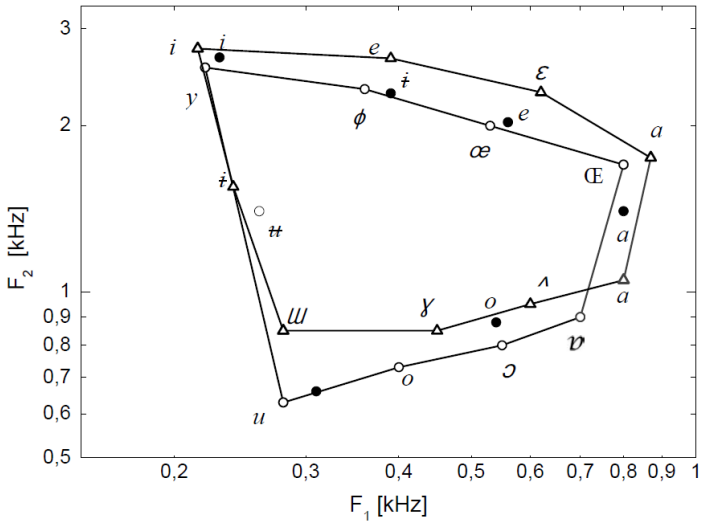
Rys. 4. Samogłoska /a/ i /i/ wypowiedziane z tą samą wysokością tonu (F_0 = 700 Hz). Wyraźnie widoczne są tony składowe w przebiegu okresowym obu samogłosek.

Samogłoski

W studiach fonetycznych bada się zazwyczaj trzy główne aspekty fali dźwiękowej: natężenie (głośność), iloczas (czas trwania czyli długość) i własności widmowe, czyli częstotliwościowe (brzmienie– jakość/barwę dźwięku). W opisie samogłosek najważniejsze są właśnie cechy widmowe; w niektórych sytuacjach bada się też długość. Wśród ustalonych dźwięków mowy samogłoski są tymi dźwiękami, które najłatwiej scharakteryzować. Powstają one w warunkach swobodnego przepływu powietrza wzdłuż języka. Wiązadła głosowe drgają periodycznie lub quasi-periodycznie a podniebienie miękkie jest uniesione do góry blokując dostęp powietrza do nosa. O tym jaką samogłoskę wypowiadamy decyduje położenie środkowej i tylnej części języka względem podniebienia twardego, gdyż przód języka jest bierny. Dlatego też możemy mówić o samogłoskach przednich i tylnych oraz o wysokich i niskich. Samogłoski są stosunkowo łatwymi obiektami do badań z uwagi na dość długi ich czas trwania (powyżej 100 ms) i wyraźnie zarysowany stan ustalony umożliwiający wyznaczenie częstotliwości tzw. formantów. Formantem nazywamy lokalne maksimum energii dźwięku mowy.

Parametry samogłosek omawia się dość często w kontekście tzw. pętli formantowej, tj. wykresu ilustrującego zależność częstotliwości formantu F1 od częstotliwości formantu F2.

Rysunek 5. ilustruje pętle formantowe samogłosek zaokrąglonych oraz niezaokrąglonych. Samogłoski niezaokrąglone to te, które są wymawiane bez zaokrąglania ust (tj. przy możliwie najstaranniejszej wymowie i przy jak najszerszym otwieraniu ust), natomiast zaokrąglone powstają podczas wypowiadania samogłosek z zaokrąglonymi (nie w pełni otwartymi) ustami.

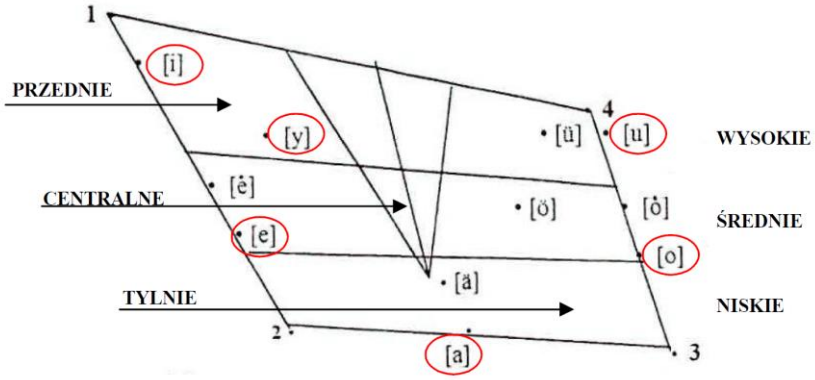


Rys. 5. Pętla formantowa samogłosek zaokrąglonych (do której należy np. samogłoska /u/) oraz pętla formantowa samogłosek niezaokrąglonych (do których należy np. samogłoska /a/), Jassem, 1973.

Na podstawie analizy pętli formantowej można stwierdzić np. że im samogłoska jest bardziej otwarta (tylna część języka usytuowana jest w najniższym z możliwych położeń względem podniebienia miękkiego) tym większa jest częstotliwość formantu F1. Ponadto duża wartość F2 jest charakterystyczna dla samogłosek przednich (tj. takich w czasie wymawiania których tylna część języka jest wysunięta możliwie najdalej do przodu). Choć bezwzględne wartości częstotliwości poszczególnych formantów wykazują dość duże zróżnicowanie międzyosobnicze (płeć, wiek).

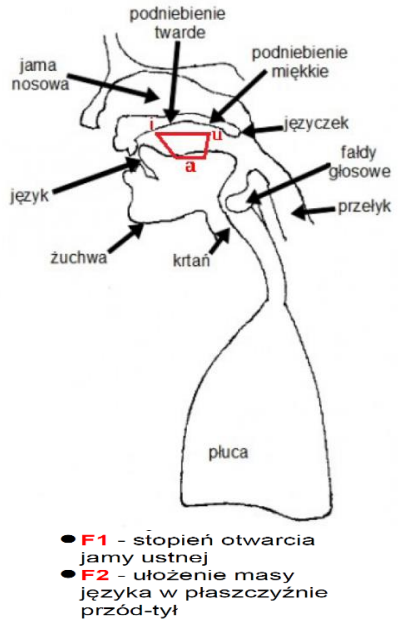
Czworobok samogłoskowy

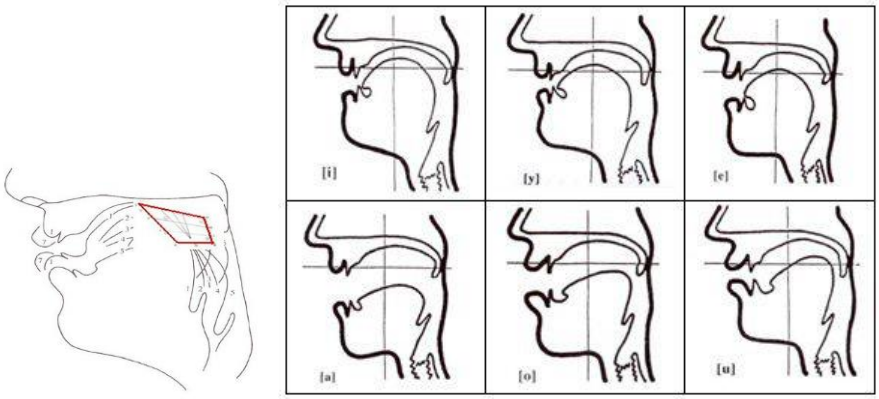
Jest to sposób graficznej reprezentacji samogłosek ze względu na artykulację (położenie języka w płaszczyźnie poziomej i pionowej) oraz wynikające z niej właściwości akustyczne: częstotliwości formantów F1 i F2 (Rys. 6).



Rys. 6. Czworobok samogłoskowy samogłosek polskich (J. Wiśniewski, Fonetyka i fonologia współczesnego języka polskiego).

Formanty a artykulacja samogłosek

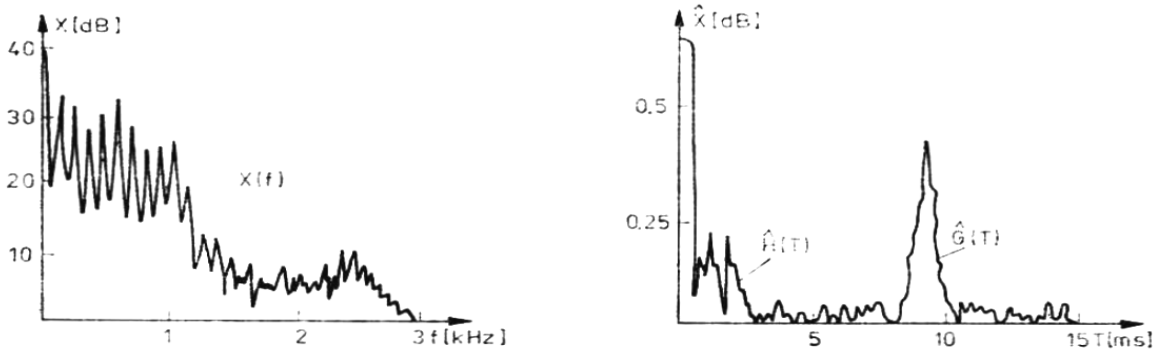




Rys. 7. Przekroje artykulacyjne samogłosek polskich: /i/ (wysoka przednia), /y/ (wysoka, przednia cofnięta), /e/ (średnia przednia), /a/ (niska centralna), /o/ (średnia tylna), /u/ (wysoka tylna).

Ton krtaniowy – bogaty widmowo (widmo harmoniczne) dźwięk o nieokreślonej barwie, zawiera różne częstotliwości, wśród nich również wszystkie częstotliwości samogłosek niezbędne do ich wytworzenia. Ton krtaniowy jest dźwiękiem pierwotnym w aparacie głosowym widocznym w funkcji

CEPSTRUM sygnału jako maksimum ↓



WIDMO **Fonem "a"** **CEPSTRUM**

4-79. Widmo głoski *a* oraz obliczone na jego podstawie cepstrum (po prawej stronie rysunku), w którego przebiegu wyraźnie zaznacza się składowa pochodząca od pobudzenia krtaniowego $\hat{G}(T)$ oraz, w okolicy $T = 0$, składowa pochodząca od procesu artykulacji $\hat{H}(T)$. Rozdzielenie tych dwu składowych jest teraz łatwe, a wynik — w postaci wyodrębnionego przebiegu $\hat{H}(T)$ — jest bardzo przydatny przy rozpoznawaniu mowy

III. Literatura

1. Tadeusiewicz R.: Sygnał mowy, WKiŁ, Warszawa 1988.

2. Basztura Cz.: Źródła, sygnały i obrazy akustyczne, WKiŁ, Warszawa 1988.

3. Jassem W.: Podstawy fonetyki akustycznej, PWN, Warszawa 1973.

4. Moore, B. C. J.: Wprowadzenie do psychologii słyszenia, PWN, Poznań 1999.

5. Halliday,D., Resnick,R., Walker J.: Podstawy fizyki, PWN, Warszawa 2003.

6. Praca zbiorowa pod red. Jaroszyka, F.: Biofizyka – podręcznik dla studentów, Wydawnictwo Lekarskie PZWL, Warszawa 2001.

IV. Zagadnienia do opracowania

1. Ogólne wiadomości o falach, podział fal ze względu na kierunek drgań, fala biegnąca, fala stojąca, rezonans akustyczny.
2. Prędkość rozchodzenia się fal w powietrzu, zależność prędkości od temperatury powietrza.
3. Tony, dźwięki, szumy.
4. Widmo fal akustycznych, próg słyszalności, próg bólu.
5. Prawo Webera-Fechnera.
6. Sposoby wyrażania głośności (poziom natężenia, ciśnienie akustyczne), jednostki (dB, fony).

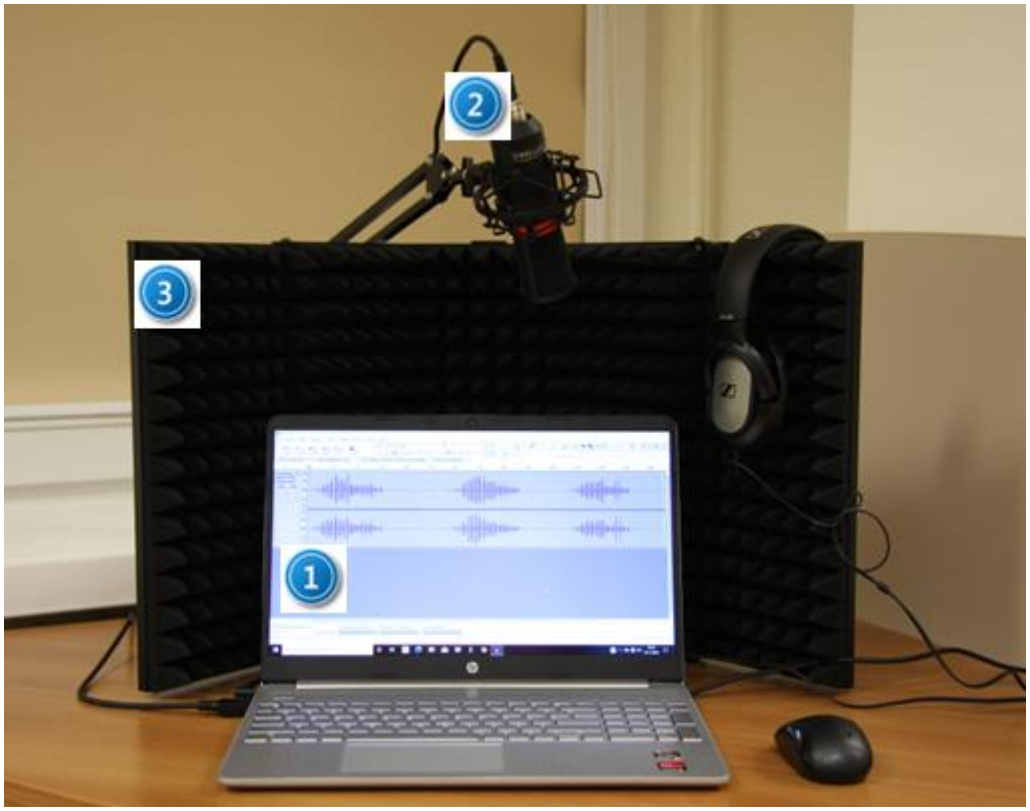
7. Budowa i zasada działania słuchawki, mikrofonu.
8. Budowa układu mowy i słuchu u człowieka.
9. Czworobok samogłoskowy.

V. Zestaw przyrządów.

1. Zestaw komputerowy z oprogramowaniem.
2. Mikrofon.
3. Ekran akustyczny.

VI. Wykonanie doświadczenia i opracowanie wyników

1. Podłączyć mikrofon do komputera, ustawić ekran akustyczny.



Zdjęcie: 1. Komputer; 2– mikrofon; 3 -ekran akustyczny.

1. Komputer wyposażony w mikrofon otoczyć ekranem akustycznym.
2. Uruchomić program PRAAT. Po uruchomieniu programu pojawiają się dwa okna, z których jedno „PRAAT objects" zawiera wszystkie opcje analizy, drugie okno „PRAAT picture" umożliwia tworzenie obiektów graficznych w celu dokumentowania wyników analiz. Używając dostępnego mikrofonu zarejestrować kilkukrotnie samogłoski /i/, /e/, /a/, /o/, /u/; czas trwania nie dłużej niż 10 s zwykle 1-2 s. W tym celu w oknie Praat Objects klikamy New->Record mono Sound.
3. Po zarejestrowaniu wszystkich samogłosek należy zmierzyć średnie wartości częstotliwości formantów F1-F4 dla każdej z nich. Z listy Objects wybieramy poszczególne nagrania i klikamy View & Edit w celu dokonania analizy. Po ukazaniu się okna z przebiegiem czasowym oraz spektrogramem wybranej wcześniej samogłoski przechodzimy do obliczania średnich wartości częstotliwości formantów. W tym celu otwieramy okno przedstawiające obwiednię widma gęstości poprzez kliknięcie w oknie z przebiegiem czasowym Spectrum->View spectra slice. W nowo otwartym oknie obliczamy wartości częstotliwości formantów F1-F4 dla każdej z samogłosek.
4. Sporządzić tabelę ze średnimi wartościami częstotliwości formantowych wszystkich zarejestrowanych samogłosek.

Przykładowa tabela danych:

Samogłoska "I"				
l.p.	Formant F1, Hz	Formant F2, Hz	Formant F3, Hz	Formant F4, Hz
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
Wartość średnia				
Odchylenie standardowe wartości średniej				

5.Tak samo postąpić dla samogłoski „e", „a", „o":„u":

6. Wyznaczyć pętlę formantową dla zarejestrowanych samogłosek (tzw. czworobok samogłosek),określającą uśrednione położenie samogłosek na płaszczyźnie pierwszego i drugiego formantu (F1 i F2). Uzyskane wyniki porównać z danymi literaturowymi.

7. Wyznaczyć częstotliwości tonu krtaniowego F_0 dla każdej z samogłosek korzystając z przebiegu cepstrum używając programu Audacity.

Uruchomić program Audacity. Otworzyć w oknie zarejestrowane przebiegi samogłosek. Z opcji narysuj widmo wybrać cepstrum. W przebiegu znaleźć występujące maksimum najeżdżając kursorem i odczytując odpowiadająca mu częstotliwość tonu krtaniowego F_0. Porównać wartości F_0 otrzymane dla rejestrowanych samogłosek z danymi literaturowymi.

Samogłoski	**Częstotliwość tonu krtaniowego F_0, Hz**
„i"	
„e"	
„a"	
„o"	
„u"	

8. Sporządzić opinię dla organu procesowego.

Sprawozdanie ma mieć charakter przykładowej opinii biegłego tzn.: zawierać takie elementy jak:

- imię, nazwisko, stopień i tytuł naukowy, specjalność i stanowisko zawodowe biegłego;
- imiona i nazwiska oraz pozostałe dane innych osób, które uczestniczyły w przeprowadzeniu ekspertyzy, ze wskazaniem czynności dokonanych przez każdą z nich, w przypadku opinii instytucji - także pełną nazwę i siedzibę instytucji;
- czas przeprowadzonych badań oraz datę wydania opinii;

- szczegółowy opis nadesłanego materiału dowodowego, porównawczego (sposób zabezpieczenia podczas transportu, opis opakowania, jego cechy ogólne i indywidualne) i cytowane pytania organu procesowego;
- informację o zastosowanych technikach i metodach;
- sprawozdanie z przeprowadzonych badań i obserwacji;
- interpretację wyników i wnioski;
- podpisy wszystkich biegłych, którzy uczestniczyli w wydaniu opinii.

Opinia powinna być napisana zrozumiałym językiem, a wnioski powinny być bardzo czytelne, zwłaszcza dla organów procesowych (dla prokuratury i sędziego).

Dodatek:

1) Formanty F1 i F2 polskich samogłosek:

	F_1 [Hz]	F_2 [Hz]
i	188 - 275	2090 -2840
y	262 - 391	1670 - 2360
e	524 - 630	1580 - 2230
a	683- 1020	1130 - 1570
o	493 - 679	788 - 1100
u	243 - 338	573 - 789

2) Wielkością najbardziej prawdopodobną dla skończonej liczby *N* pomiarów jest **średnia arytmetyczna** dana zależnością:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

gdzie: i jest numerem pomiaru a x_i wynikiem i-tego pomiaru.

Miarą niepewności wartości średniej $\bar{x}$ jest **odchylenie standardowe średniej arytmetycznej** serii pomiarów wyrażone zależnością:

$$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N(N-1)}}.$$