II PRACOWNIA FIZYCZNA

Akustyczno-fonetyczne cechy mowy polskiej

Opis ćwiczenia w ramach II Pracowni Fizycznej

Andrzej Wicher
Aleksander Sęk
Jacek Konieczny

Instytut Akustyki UAM
Poznań, 2005

1. WSTĘP

Jednym z najwydajniejszych sposobów komunikowania się ludzi z otoczeniem jest mówienie i słyszenie. Z fizycznego punktu widzenia ten proces komunikacji polega na generowaniu i odbiorze bodźców akustycznych. Narząd mowy jest wyspecjalizowanym układem umożliwiającym generowanie szerokiej gamy dźwięków. Steruje on strumieniem powietrza wypływającym z płuc, umożliwiając kodowanie użytecznej informacji w postaci zmian chwilowego ciśnienia.

Zmysł słuchu umożliwia odbiór bodźców akustycznych i wyselekcjonowanie z nich użytecznych informacji. W peryferyjnym układzie słuchowym fala akustyczna jest poddana analizie częstotliwościowej i zamianie na ciągi impulsów neuronowych. Impulsy te są następnie interpretowane przez wyższe piętra układu słuchowego na zrozumiałe informacje.

Sposób tej interpretacji w dużym stopniu zależy (choć nie jest to zależność do końca poznana) od rodzaju sygnałów akustycznych, ich cech widmowych, itp.

Niniejsze opracowanie zawiera podstawowe informacje na temat rodzajów sygnałów, możliwości ich analizy, a przede wszystkim zasadnicze informacje dotyczące dźwięków mowy (źródła dźwięków mowy, klasyfikacja dźwięków mowy oraz metody ich analizy).

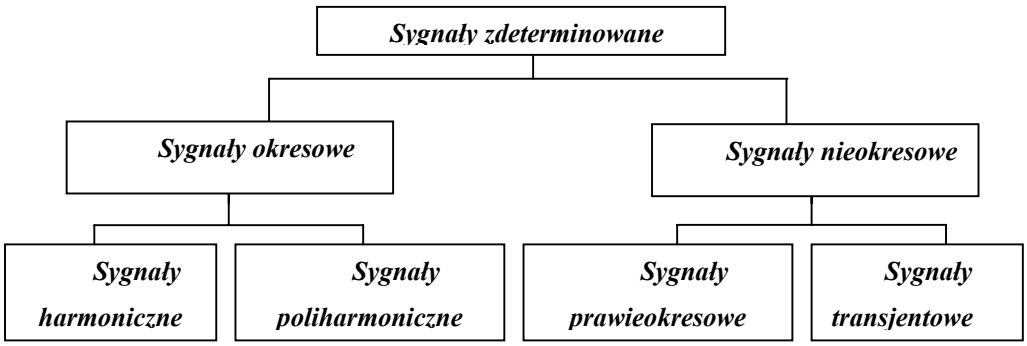
Znaczną część miejsca w niniejszym opracowaniu poświęcono analizie widmowej sygnałów, która jest podstawą analizy dźwięku. Szczególny nacisk położono na analizę sygnałów dyskretnych, z którymi mamy najczęściej do czynienia, w prowadzeniu eksperymentalnej analizy sygnałów mowy za pomocą specjalistycznych pakietów programów.

2. SYGNAŁY ORAZ ICH ANALIZA

Sygnałem, z punktu widzenia analizy sygnałów, jest przebieg w czasie dowolnej wielkości fizycznej (np. przebieg zmian temperatury powietrza, zmian napięcia na odbiorniku elektryczności). Fala akustyczna powstająca na skutek drgań ciała w ośrodku sprężystym (np. powietrzu) jest także przykładem sygnału, ponieważ wprowadza ona chwilowe zmiany ciśnienia akustycznego. W tym też świetle mowa jest sygnałem akustycznym, do którego można zastosować szereg metod analizy sygnałów. Przedstawienie zasadniczych metod analizy sygnałów mowy wymaga uprzedniego wprowadzenia podziału sygnałów na poszczególne klasy/rodzaje. Wiąże się to z tym, że do różnych typów sygnałów stosuje się różnego rodzaju metody analizy.

2.1. Rodzaje sygnałów

Sygnały można ogólnie podzielić na zdeterminowane i niezdeterminowane. Sygnały zdeterminowane to takie, które dają opisać się analitycznie. Sygnały niezdeterminowane, nazywane też sygnałami losowymi, nie dają opisać się zależnościami matematycznymi. Z praktycznego punktu widzenia sygnały, które można odtworzyć warunkach laboratoryjnych, to sygnały zdeterminowane. Sygnały niezdeterminowane nie dają się natomiast odtworzyć w sposób powtarzalny. Klasyfikację sygnałów zdeterminowanych ilustruje Rysunek 2.1.



Rysunek 2.1. Klasyfikacja sygnałów zdeterminowanych

Sygnały okresowe to sygnały, których przebieg powtarza się co jakiś czas T, nazywany okresem. Ścisła definicja sygnału okresowego wymaga, by sygnał ten trwał nieskończenie długo, toteż w praktyce za sygnał okresowy uznaje się każdy sygnał, którego przebieg powtarza się w skończonym przedziale czasu. Najprostszym przykładem sygnału okresowego jest sygnał sinusoidalny, nazywany też sygnałem harmonicznym, którego akustycznym odpowiednikiem jest ton prosty. Sygnał taki w pełni charakteryzują trzy jego parametry: amplituda, częstotliwość i faza początkowa. Znacznie częściej możemy się zetknąć z sygnałami okresowymi złożonymi z wielu sygnałów sinusoidalnych, które nazywane są sygnałami poliharmonicznymi . Sygnał taki składa się często ze składowej stałej A_0 i sumy skończonej lub nieskończonej liczby składowych sinusoidalnych (harmonicznych) o amplitudach A_n, fazach początkowych φ_n i częstotliwościach $f_n=nf_1$, będących wielokrotnościami częstotliwości podstawowej f_1:

$$D(t) = A_0 + \sum_{n=1}^{\infty} A_n \sin\left(2\pi n f_1 t + \varphi_n\right) \tag{2.1}$$

Okresowe sygnały poliharmoniczne mogą powstawać w wyniku sumowania składowych sinusoidalnych, których stosunki częstotliwości są liczbami całkowitymi. Przykładem sygnałów poliharmonicznych są przede wszystkim dźwięki muzyki a nieco

gorszym dźwięczne części mowy (np. samogłoski i spółgłoski dźwięczne), które wytwarzane są przy udziale drgań wiązadeł głosowych.

Jeśli częstotliwości poszczególnych składowych sygnału złożonego nie są całkowitymi wielokrotnościami składowej podstawowej oraz stosunki nie wszystkich możliwych par częstotliwości tych składowych są liczbami wymiernymi, to sygnał taki nazywa się sygnałem prawie okresowym i można go zapisać w postaci:

$$D(t) = \sum_{n=1}^{\infty} A_n \sin\left(2\pi f_n + \varphi_n\right) \tag{2.2}$$

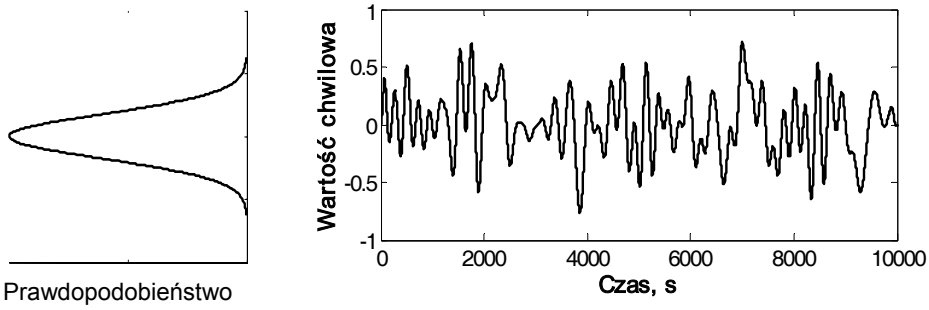
gdzie f_m/f_n nie jest w ogólności liczbą wymierną. Najprostszym przykładem sygnału prawie okresowego jest suma dwóch sinusoid o częstotliwościach np. f_1=10 i f_2=$10^{1/2}$ Hz.

Sygnały transjentowe obejmują szeroką grupę zdeterminowanych sygnałów nieokresowych i można opisać je za pomocą odpowiednich funkcji zmiennych w czasie. Ich charakterystyczną cechą jest zmienna w czasie amplituda o charakterze narastania lub zanikania. Sygnałami takimi są dźwięki muzyki i mowy, w których można wydzielić odcinki o nieustalonym przebiegu w czasie.

Znacznie szerszą klasą sygnałów niż sygnały zdeterminowane jest grupa sygnałów niezdeterminowanych lub losowych. Sygnały odpowiadające losowym zjawiskom są niepowtarzalnymi, jedynymi w warunkach konkretnej obserwacji, i nie można opisać ich analitycznie. Ich losowość oznacza w ogólności, że nie jesteśmy w stanie przewidzieć wartości ich parametrów w żadnej z przyszłych chwil czasu. Jednak dla niektórych typów sygnałów losowych możemy określić prawdopodobieństwo tego, że parametry tych sygnałów osiągną określone wartości. W związku z tym możemy mówić o procesie losowym (lub stochastycznym) oraz o jego realizacji, czyli każdej jego odrębnej obserwacji. Np. zapis przebiegu napięcia na generatorze szumu (typowy przykład sygnału losowego) w skończonym odcinku czasu jest jedną realizacją procesu losowego. Jednym z najczęściej stosowanych sygnałów losowych w badaniach słuchu jest tzw. szum biały. W szumie tym występuje nieskończenie wiele składowych sinusoidalnych, których częstotliwości obejmują cały zakres słyszalny (tj. do ok. 20 kHz). Amplitudy wszystkich składowych są jednakowe a ich fazy początkowe są wartościami przypadkowymi. Nazwa tego szumu jest pewną analogią do światła białego, które jest sumą wszystkich elementarnych barw składowych o różnej częstotliwości z całego zakresu częstotliwości fal widzialnych. Szum biały jest tzw. stacjonarnym sygnałem losowym, ponieważ jego tzw. charakterystyki probabilistyczne (np. wartość średnia, wartość średniokwadratowa) nie zmieniają się w czasie. Jest on nazywany także szumem gaussowskim, ponieważ rozkład jego wartości chwilowych jest opisany za

pomocą rozkładu Gaussa. Rysunek 2.2 przedstawia przebieg czasowy odcinka szumu białego oraz odpowiadający mu rozkład prawdopodobieństwa jego wartości chwilowych.

Warto w tym miejscu dodać, że sygnały mowy nie dają się jednoznacznie zakwalifikować do żadnej z powyższych grup. Jednak niewielkie odcinki czasowe sygnałów mowy, odpowiadające w przybliżeniu poszczególnym głoskom, można w wielu sytuacjach potraktować jako sygnały tego typu. Np. sygnał samogłosek można w przybliżeniu potraktować jako sygnał poliharmoniczny, zaś sygnały odpowiadające spółgłoskom zwarto-trącym ("cz", "c", "dż", "dź") można potraktować jako pasmo szumu.



Rysunek 2.2. Przykładowy przebieg czasowy szumu białego

2.2. Analiza sygnałów okresowych

Analiza sygnału, w najbardziej podstawowym rozumieniu tego słowa, polega na przedstawieniu badanego sygnału za pomocą sumy funkcji elementarnych, tzn. rozłożeniu go na składowe dźwięki elementarne jakimi są tony. Zazwyczaj celem analizy sygnału jest przedstawienie go za pomocą widma, tj. wykresu ilustrującego zależność amplitudy (lub mocy) sygnałów składających się na analizowany dźwięk jako funkcji ich częstotliwości. Dlatego też analizę sygnałów dość często określa się jako analizę widmową.

Analizy widmowej zdeterminowanych sygnałów okresowych dokonuje się wykorzystując matematyczne narzędzie zwane szeregiem Fouriera. Zdeterminowane przebiegi nieokresowe analizuje się z wykorzystaniem przekształcenia (całki) Fouriera.

Rozważmy najpierw analizę sygnałów okresowych. Według twierdzenia Fouriera funkcję okresową *f(t)* można rozłożyć na szereg trygonometryczny postaci:

$$f(t) = A_0 + \sum_{n=1}^{\infty} \left[A_n \cos(\omega_n t) + B_n \sin(\omega_n t) \right] \tag{2.3}$$

Funkcję *f(t)* można więc wyrazić jako sumę sinusoid i cosinusoid o określonych częstotliwościach i amplitudach składowej stałej A_0. W szeregu przedstawionym równaniem (2.3) jedyną zmienną jest czas *t*, a pozostałe parametry są stałe. Częstotliwości sinusoid i

cosinusoid pozostają w stosunku harmonicznym, czyli są wielokrotnościami częstotliwości podstawowej ω_0:

$$\omega_n = n\omega_0 = \frac{2\pi n}{T}. \tag{2.4}$$

Częstotliwość podstawowa jest najmniejszą częstotliwością mogącą wystąpić w szeregu Fouriera (2.3), a fala o tej częstotliwości ma okres $T_0 = \frac{1}{2\pi\omega_0}$ i jest on równy okresowi funkcji *f(t)*. Zasadniczą ideą szeregu Fouriera jest to, że każdą funkcję okresową można przedstawić w postaci sumy sinusoid i cosinusoid, których okresy mieszczą się całkowitą liczbę razy w okresie podstawowym złożonej fali okresowej.

Występujące w równaniu (2.3) współczynniki A_n i B_n są nazywane współczynnikami szeregu Fouriera. Opisują one udział sinusoidy i cosinusoidy o numerze *n* (a więc o częstotliwości *n* razy większej od częstotliwości podstawowej) w sygnale *f(t)*. W ogólności, dowolną funkcję okresową *f(t)* można przedstawić jako sumę nieskończenie wielu składników szeregu Fouriera. Jednak w praktyce do stworzenia najczęściej występujących przebiegów falowych wystarcza skończona liczba składników, czyli można takie przebiegi aproksymować sumą skończonej liczby sinusoid i cosinusoid. Współczynniki A_n i B_n można wyznaczyć za pomocą następujących zależności:

$$\begin{aligned} A_0 &= \frac{1}{T}\int_{-T/2}^{T/2} x(t)dt, \\ A_n &= \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos(\omega_n t)dt, n > 0, \\ B_n &= \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin(\omega_n t)dt, n > 0. \end{aligned} \tag{2.5}$$

Granice całkowania w równaniach (2.5) rozciągają się od *–T/2* do *T/2*. Takie granice całkowania nie zawsze są wygodne i czasem wygodniej jest użyć granic całkowania od 0 do *T*. Wybór granic całkowania do obliczenia współczynników szeregu Fouriera jest w zasadzie dowolny z tym, że przedział całkowania powinien mieć długość jednego okresu, czyli *T*.

Rozwiniecie funkcji *f(t)* w szereg Fouriera w postaci (2.3) można przekształcić, do innej, wygodniejszej formy zakładając, że:

$$\begin{aligned} h_n &= \sqrt{A_n^2 + B_n^2}, \\ \varphi_n &= arctg\frac{B_n}{A_n}. \end{aligned} \tag{2.6}$$

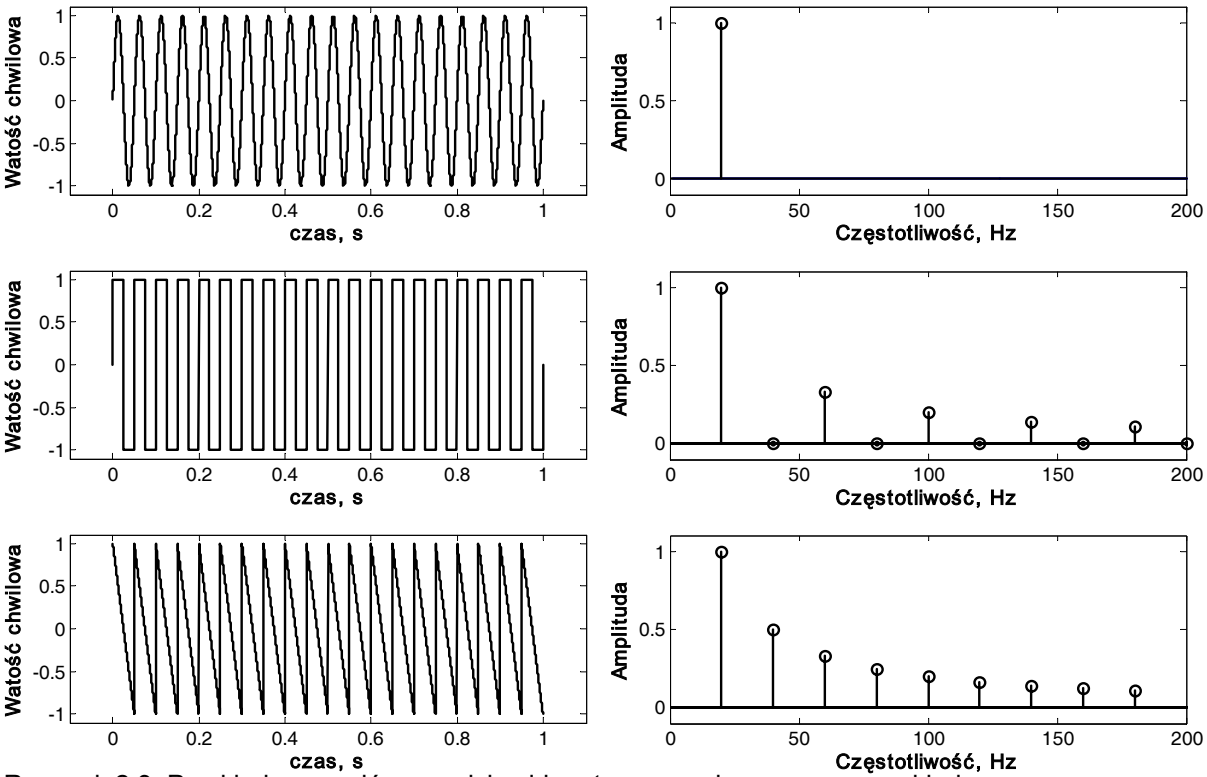
Wartości h_n reprezentują amplitudy kolejnych cosinusoid o numerze n (częstotliwości n razy większej od częstotliwości podstawowej), w związku z czym ich zbiór nazywa się widmem amplitudowym. Analogicznie zbiór wartości φ_n tworzy tzw. widmo fazowe.

Po podstawieniu (2.6) do (2.3) i skorzystaniu z tożsamości trygonometrycznej można napisać:

$$f(t) = A_0 + \sum_{n=1}^{\infty} h_n \cos(n\omega_0 t - \varphi_n). \tag{2.7}$$

Ogólny wyraz $h_n \cos(n\omega_0 t + \varphi_n)$ przedstawia n-tą składową funkcji $f(t)$, zwaną też n-tą harmoniczną.

Należy podkreślić, że rozwinięcie funkcji w szereg Fouriera jest jednoznaczne, tzn., że daną funkcję $f(t)$ można tylko w jeden sposób przedstawić za pomocą szeregu trygonometrycznego. Ponadto warto pamiętać, że suma częściowa szeregu Fouriera jest najlepszym możliwym przybliżeniem funkcji rozwijanej w szereg trygonometryczny. Lepsze przybliżenie można uzyskać jedynie dołączając dalsze wyrazy tego szeregu, a nie przez zmianę współczynników A_n i B_n.



Rysunek 2.3. Przykłady sygnałów oraz ich widm otrzymanych za pomocą rozkładu na szereg Fouriera.

Przykłady widm niektórych sygnałów zdeterminowanych, które można uzyskać za pomocą przedstawienia sygnału za pomocą szeregu Fouriera, przedstawia Rysunek 2.3. Energia dźwięków periodycznych o długim czasie trwania przypada dla pewnych dyskretnych

wartości częstotliwości i widmo takie nazywane jest widmem prążkowym. Pierwsze trzy przykłady z Rysunek 2.3 przedstawiają właśnie widma tego typu. Sygnał sinusoidalny, z definicji, zawiera jedną składową częstotliwościową. Fala prostokątna składa się z nieparzystych harmonicznych składowej podstawowej a amplitudy tych składowych maleją ze wzrostem numeru harmonicznej. Ciąg impulsów powtarzających się co stały przedział czasu zawiera wszystkie składowe harmoniczne składowej podstawowej o równych amplitudach. Ponieważ jednak pojedynczy impuls ma małą energię a składowych harmonicznych jest wiele to składowe te mają małą amplitudę.

W praktyce szereg Fouriera nie może być stosowany do analizy częstotliwościowej sygnałów, ponieważ w rozważaniach teoretycznych zakłada się nieskończony czas trwania przebiegu, podczas gdy zjawiska obserwowane w praktyce mają skończony czas trwania. Ponadto zakłada się okresowość przebiegu *f(t),* a tymczasem przebiegi występujące w rzeczywistości są często nieokresowe.

Pewnym uogólnieniem szeregu Fouriera w powyższej formie jest jego postać zespolona wyrażona następującą formułą:

$$f(t) = \sum_{n=-\infty}^{n=+\infty} X_n e^{jn\omega_0 t} \quad \text{gdzie} \quad X_n = \frac{\omega_0}{2\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} f(t) e^{-jn\omega_0 t} dt \,. \tag{2.8}$$

Dość często zamiast X_n używa się notacji $X(j\omega)$. Jak widać z tego równania dowolną funkcję okresową można zapisać w postaci szeregu, w którym występują ujemne częstotliwości (sumowanie dokonywane jest od n=- ∞ do n=+∞ a wyrażenie $n\omega_0$ oznacza częstotliwość n-tej składowej). Nie ma w tym nic dziwnego, bowiem prosty sygnał cosinusoidalny na mocy równań Eulera można zapisać jako sumę dwóch składników:

$$\begin{aligned} e^{j\beta} &= \cos\beta + j\sin\beta \\ e^{-j\beta} &= \cos\beta - j\sin\beta \\ \frac{\alpha}{2}e^{j\beta} + \frac{\alpha}{2}e^{-j\beta} &= \alpha\cos\beta \\ A\cos(2\pi ft + \theta) &= \frac{A}{2}e^{j(2\pi ft+\theta)} + \frac{A}{2}e^{-j(2\pi ft+\theta)} \end{aligned} \tag{2.9}$$

Z uwagi na tak specyficzne przedstawienie sygnałów analiza Fouriera w przypadku szeregów zespolonych pozwala na wyznaczenie tzw. widm dwustronnego (bo zawiera niezerowe amplitudy składowych o ujemnych częstotliwościach) zarówno widma amplitudowego, jak i fazowego.

Ważną właściwością analizy za pomocą szeregu Fouriera jest również i to, że analiza ta pozwala za rozłożenie złożonej funkcji okresowej na dyskretne składowe, co oznacza, że

energia sygnału skupiona jest tylko w sygnałach o częstotliwościach $n\omega_0$. W przedziałach pomiędzy tymi składowymi nie ma żadnej energii.

Dla rzeczywistej funkcji *x(t)* można zapisać:

$$\begin{aligned}\mathrm{Re}(X_n) &= \mathrm{Re}\left[\frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-j2\pi nf_0 t}dt\right] = \frac{1}{T}\int_{-T/2}^{T/2} x(t)\cos(2\pi nf_0 t)dt \\ \mathrm{Im}(X_n) &= \mathrm{Im}\left[\frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-j2\pi nf_0 t}dt\right] = -\frac{1}{T}\int_{-T/2}^{T/2} x(t)\sin(2\pi nf_0 t)dt\end{aligned} \tag{2.10}$$

Oraz wykazać pewne właściwości symetrii:

$$\begin{aligned}&\mathrm{Re}[X_{-n}] = \mathrm{Re}[X_n] \\ &\mathrm{Im}[X_{-n}] = -\mathrm{Im}[X_n] \\ &X_{-n} = X_n^* \\ &|X_{-n}| = |X_n| \\ &\arg(X_{-n}) = -\arg(X_n)\end{aligned} \tag{2.11}$$

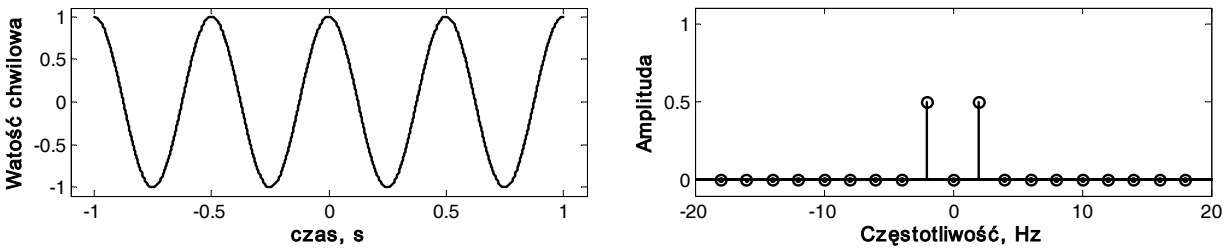
Ponadto, w przypadku sygnału rzeczywistego można zapisać:

$$\begin{aligned}x(t) &= X_0 + \sum_{n=1}^{\infty}\left[X_n e^{j2\pi fnt} + X_{-n}e^{j2\pi f(-n)t}\right] \\ x(t) &= X_0 + \sum_{n=1}^{\infty} 2|X_n|\cos(2\pi ftn + \arg(X_n)\end{aligned} \tag{2.12}$$

co w dalszej kolejności umożliwia wyznaczenie widma jednostronnego (a więc zawierającego tylko składowe o dodatnich częstotliwościach) zarówno amplitudowego, jak i fazowego na podstawie następujących zależności:

$$\begin{aligned}|X_n| &= \sqrt{(\mathrm{Re}(X_n))^2 + (\mathrm{Im}(X_n))^2} \\ Ph((X_n) &= \arctan\frac{\mathrm{Im}(X_n)}{\mathrm{Re}(X_n)}\end{aligned} \tag{2.13}$$

$$x(t) = \cos(\omega_0 t) = \frac{1}{2}\left(e^{j\omega_0 t} + e^{-j\omega_0 t}\right) \qquad X_1 = \frac{1}{2}, \qquad X_{-1} = \frac{1}{2}, \qquad X_k = 0 \quad k \neq \pm 1$$



Rysunek 2.4. Przykładowe widmo dwustronne

2.3. Analiza sygnałów nieokresowych

Analizy spektralnej sygnałów nieokresowych opisanych funkcjami analitycznymi dokonuje się często za pomocą tzw. przekształcenia Fouriera, które zdefiniowane jest następującą zależnością:

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{2.14}$$

gdzie $F(j\omega)$ jest nazywane przekształceniem Fouriera funkcji $f(t)$. Zarówno $F(j\omega)$ jak i $f(t)$ są tu funkcjami ciągłymi, zdefiniowanymi w nieskończonym przedziale swoich argumentów. Ponieważ $F(j\omega)$ jest funkcją ω to o transformacji tej mówi się, że transformuje ona funkcję $f(t)$ z dziedziny czasu do dziedziny częstotliwości. Aby w pełni udokumentować związek wielkości ω z częstotliwością można pokazać, że funkcja $F(j\omega)$ jest swoistym uogólnieniem wielkości h_n (lub c_n), które w przypadku szeregu Fouriera stanowiły zbiór wartości dyskretnych. Jednak w przypadku całki Fouriera $F(j\omega)$ stanowią pewne kontinuum wartości gdy okres funkcji $f(t)$ rośnie do nieskończoności, w rezultacie czego funkcja $f(t)$ staje się aperiodyczna. Aby uzasadnić ten punkt widzenia załóżmy, że okresowa funkcja $f(t)$ może być rozłożona na następujący szereg Fouriera:

$$f(t) = \sum_{n=-\infty}^{n=+\infty} c_n e^{jn\omega_0 t} \quad \text{gdzie} \qquad c_n = \frac{\omega_0}{2\pi}\int_{-\pi/\omega_0}^{\pi/\omega_0} f(t)e^{jn\omega_0 t}dt \tag{2.15}$$

Okres funkcji $f(t)$ jest tu równy $2\pi/\omega_0$ sekund gdy ω_0 wyrażone jest w radianach na sekundę, a skład harmoniczny funkcji $f(t)$ nie jest ograniczony. Każda wartość c_n jest zespoloną składową częstotliwościową funkcji $f(t)$ o częstotliwości kątowej $n\omega_0$. Widmo amplitudowe funkcji $f(t)$ jest symetryczne względem punktu ω=0, ponieważ wartość sprzężona współczynników c_n^* jest równa c_{-n} w związku z czym $|c_n| = |c_{-n}|$.

Różniczkowanie funkcji $F(j\omega)$ zakłada, że wartość ω_0, czyli odstęp częstotliwości pomiędzy dwoma sąsiednimi składowymi w szeregu Fouriera zmierza do 0 tak, że widmo prążkowe wyrażone szeregiem Fouriera (poszczególne składowe co ω_0) staje się ciągłym zbiorem wartości. Aby to osiągnąć nie ogranicza się okresu składowej o częstotliwości podstawowej $2\pi/\omega_0$ (może on rosnąć nieograniczenie) co jednocześnie oznacza, że dopuszcza się istnienie częstotliwości nieskonczenie małych, tj. $\omega_0 \to 0$. Zastępując ω_0 przez $\Delta\omega$, by podkreślić zmniejszanie się różnicy pomiędzy każdymi dwoma sąsiednimi składowymi, oraz

przechodząc z sumowania do całkowania, gdy $\Delta\omega \to 0$, ostanie równanie można zapisać w postaci następujących całek:

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)e^{j\omega t}d\omega \qquad \text{oraz} \qquad F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{2.16}$$

Transformata Fouriera $F(j\omega)$ wyraża więc ciąg współczynników rozwinięcia funkcji $f(t)$ w szereg Fouriera dla wartości $\Delta\omega$ dążącego do nieskończenie małej wartości. Możemy zatem uznać, że wielkość $|F(j\omega)|$ jest widmem amplitudowym i jest teraz funkcją ciągłą w odróżnieniu od dyskretnych wartości h_n. Dwie ostatnie zależności tworzą tzw. parę transformat Fouriera: odwrotną i prostą.

Warunkiem istnienia transformaty Fouriera jest zbieżność analizowanej funkcji tzn.:

$$\int_{-\infty}^{\infty} \left| f(t)dt \right| < \infty \tag{2.17}$$

Oczywiście żadna funkcja periodyczna nie posiada tej właściwości, ale np. pojedynczy zanikający eksponencjalnie impuls, który zanika w skończonym czasie do zera ma taką transformatę, podobnie jak inne funkcje zanikające dość szybko. Jednak przedstawiona para transformat nie pozostawia żadnych wątpliwości: przejście z $f(t)$ do $F(j\omega)$ i z powrotem – nie 'gubi' żadnych informacji o sygnale.

W praktyce eksperymentalnej nie prowadzi się analizy sygnału za pomocą całki Fouriera, lecz za pomocą pewnej jej formy nazwanej Dyskretną Transformatą Fouriera (DFT), realizowanej za pomocą powszechnie uznanego algorytmu szybkiej transformaty Fouriera (FFT). Nim jednak przedstawiona zostanie ta transformata warto zapoznać się z podstawowymi wiadomościami dotyczącymi próbkowania i kwantowania sygnału.

2.4. Analiza sygnałów dyskretnych

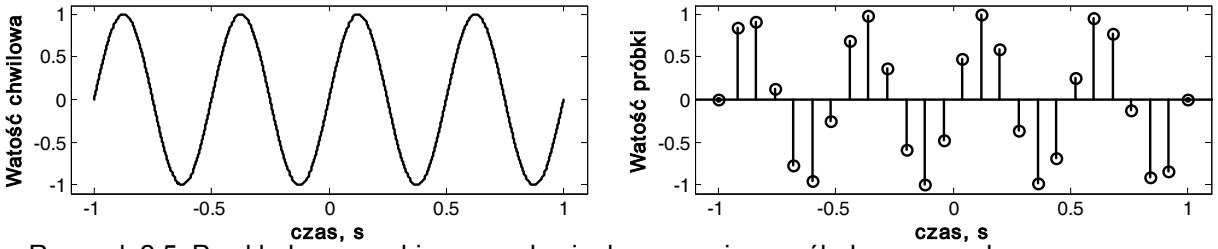
2.4.1. Próbkowanie i dyskretyzacja sygnału

Sygnały są generalnie ciągłe, przyjmujące pewną wartość w każdej chwili czasu. Jednak odczytanie wartości chwilowej sygnału w każdej chwili czasu jest niemożliwe i dlatego też odczyt wartości chwilowej dokonywany jest najczęściej w regularnych odstępach czasu, np. co ΔT. Ten proces odczytywania i zapamiętywania chwilowej wartości sygnału nazywa się próbkowaniem sygnału. Odstęp czasu pomiędzy dwoma sąsiednimi próbkami, ΔT (lub częściej T_s), nazywa się okresem próbkowania a jego odwrotność nosi nazwę szybkości próbkowania:

$$f_s = \frac{1}{\Delta T} = \frac{1}{T_s} \tag{2.18}$$

Dość często wielkości te oznacza się również za pomocą symboli F_s i T_s.

W wyniku próbkowania otrzymujemy sygnał którego wartości są znane tylko w kolejnych, dyskretnych chwilach czasu (co ΔT) i nie wiemy jaką wartość przyjmował sygnał pomiędzy dwoma dowolnymi próbkami. Przykład takiego sygnału przedstawia Rysunek 2.5



Rysunek 2.5. Przykładowy przebieg sygnału ciągłego oraz jego próbek czasowych.

Próbkowanie sygnału jest tylko jednym z elementów przetwarzania analogowo-cyfrowego sygnału. Kolejnym niezmiernie istotnym etapem tego procesu jest tzw. dyskretyzacja, lub kwantowanie sygnału. Dyskretyzacja sygnału polega na przyporządkowaniu wartościom chwilowym sygnału liczb z pewnego, ustalonego z góry zakresu. Zakres ten zależy od tzw. rozdzielczości przetwarzania (np. 8 bitów 16 bitów itd.). Jest to kolejny niezbędny proces, bowiem zapamiętanie każdej wartości amplitudy sygnału ciągłego byłoby znacznie bardziej skomplikowane. Jeśli rozdzielczość analizy wynosi n bitów to chwilowe wartości amplitudy zamienianie są na jedną z 2^n liczb. Jeśli dysponujemy więc przetwarzaniem 16 bitowym to każda chwilowa wartość sygnału zamieniana jest na jedną liczbę całkowitą z przedziału 0-65536. Zatem, gdy analizowany sygnał zmienia się w zakresie ±10 V, to zakres 20 V podzielony zostaje na 65536 jednakowych przedziałów po 0.305 mV każdy (tzw. krok dyskretyzacji), a dowolna chwilowa wartość napięcia zostaje zamieniona na liczbę równa wielokrotności tego pojedynczego przedziału napięcia. Warto jednak dodać, że pierwsza połowa tych liczb (a więc zakres 0-32768) przyporządkowana jest ujemnym wartościom chwilowym, a liczby z przedziału 32769-65536 przyporządkowane są dodatnim napięciom. Zatem napięcie równe 10 V reprezentowane jest przez liczbę 65536 a napięcie –10V przez liczbę 0. W procesie tym nie każda wartość chwilowa znajdzie odpowiadającą mu liczbę. Np. przy kroku dyskretyzacji 0.305 mV napięcie +3.3 powinno zostać zamienione na liczbę 32768+3.3/0.000305=43587.6721311475. Jednak ponieważ przetwarzanie to jest odwzorowaniem wartości ciągłych w wartości dyskretne i całkowite, to w wyniku tej operacji otrzymamy liczbę 43587. Dyskretyzacja niesie więc ze sobą pewne błędy przetwarzania, których wartość zależy od rozdzielczości: im większa rozdzielczość, tym mniejszy błąd.

Błędy te są najczęściej pomijane jeśli tylko dysponujemy przetwarzaniem 16- lub więcej bitowym.

Warto w podsumowaniu stwierdzić, że sygnał dyskretny to taki sygnał którego wartości chwilowe są znane tylko w określonych chwilach czasu oraz, że przyjmują one jedną z dozwolonych wartości.

2.4.2. Analiza widmowa sygnałów dyskretnych

Analizy widmowej sygnałów dyskretnych dokonuje się w oparciu o tzw. Dyskretną Transformatę Fouriera (DFT). Ogólne wyrażenie opisujące tę transformatę jest łudząco podobne do równania przedstawiającego rozwinięcie funkcji periodycznej w zespolony szereg Fouriera i ma postać:

$$X(m) = \sum_{n=0}^{n=N-1} x(n) e^{\frac{-j\omega nm}{N}} \tag{2.19}$$

gdzie *X(m)* oznacza *m*-tą składową dyskretnej transformaty Fouriera, *N* liczbę analizowanych próbek, a *n* i *m* odpowiednio kolejne numery próbek czasowych (wejściowych) i widmowych (wyjściowych). Zasadniczą różnicą jest tu oczywiście zastosowanie dyskretnego sygnału wejściowego *x(n)* (zamiast ciągłego *x(t)* jak to ma miejsce w szeregu Fouriera) oraz wynik w postaci dyskretnych próbek widma na wyjściu tej analizy. Wykładnik potęgi ma też nieco inną postać wynikającą bezpośrednio z faktu zastosowania dyskretnego sygnału na wejściu. Również nieskończona suma szeregu Fouriera zamieniona została na sumę po wszystkich elementach wejściowych. Podobnie jak w przypadku całki Fouriera wartości modułu kolejnych wyrażeń *X(m),* które zazwyczaj są liczbami zespolonymi, tworzą widmo sygnału (a dokładniej widmową gęstość amplitudy sygnału).

Zastosowanie analizy DFT w odniesieniu do sygnału o liczności N daje w efekcie dokładnie tyle samo niezależnych próbek widmowych. Ponieważ analiza ta daje w efekcie ponumerowane jedynie próbki to na postawie wyłącznie wartości tych próbek, trudno jest określić odpowiadające im częstotliwości. Jest to zadanie niewykonalne jeśli nie wiemy z jaką szybkością próbkowany był sygnał. Załóżmy jednak że szybkość próbkowania była równa F_s. Na tej podstawie możemy więc zapisać:

$$\frac{N}{t_{rejestracji}} = F_s = \frac{1}{T_s} \tag{2.20}$$

gdzie $t_{rejestracji}$ oznacza czas trwania analizowanego sygnału. A uproszczoną wersję równania (2.19) pozwalającą na łatwiejsze wyznaczenie częstotliwości *m*-tej próbki widmowej można zapisać następująco:

$$X(m) = \sum_{n=0}^{n=N-1} \left[x(n) \cos\underbrace{\left(\frac{2\pi nm}{N} \right)}_{2\pi f_m t} - jx(n) \sin\left(\frac{2\pi nm}{N} \right) \right] \tag{2.21}$$

Argument funkcji sinus lub cosinus można zapisać w następującej postaci

$$\frac{2\pi nm}{N} = 2\pi f_m t \tag{2.22}$$

gdzie f_m oznacza częstotliwość *m*-tej próbki widmowej, zaś zapisując to równanie dla dwóch kolejnych próbek czasowych ($t=T_s$) otrzymujemy:

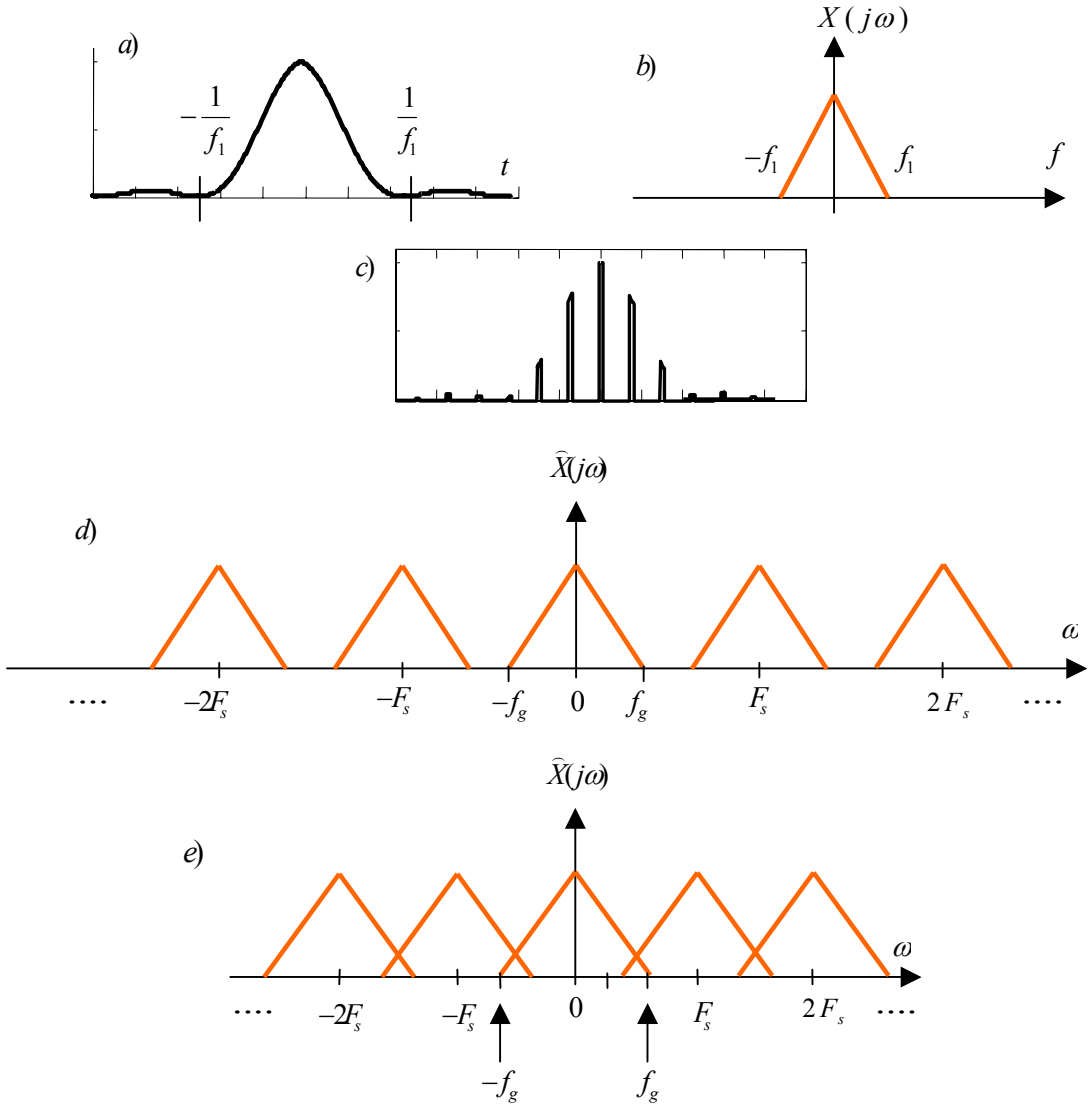
$$\begin{aligned} \frac{(n+1)m}{N} - \frac{nm}{N} &= f_m T_s \\ f_m &= \frac{mf_s}{N} \end{aligned} \tag{2.23}$$

Ponieważ w odpowiedzi na *N* próbek wejściowych otrzymujemy *N* niezależnych próbek wyjściowych w związku z tym pierwsza próbka wyjściowa (m=0) będzie odpowiadała częstotliwości f_m=0, zaś *N*-ta próbka wyjściowa o numerze *N-1* będzie odpowiadała częstotliwości F_s. Wszystkie próbki wyjściowe są równomiernie rozłożone w przedziale od 0 do szybkości próbkowania (F_s).

Warto zauważyć, że równanie (2.19), czy też (2.20) opisuje funkcję periodyczną ze względu na zmienną *m* (jeśli tylko założyć, że *m* może przyjmować dowolną wartość całkowitą) przy czym okres tej funkcji jest równy *N*. Oznacza to, że funkcja ta powtarza się dokładnie co *N* próbek wyjściowych. Ma to ogromne znaczenie dla właściwego odczytania widma sygnału. Załóżmy, że mamy sygnał ciągły, o którym wiemy, że składowa o największej częstotliwości w nim zawarta ma częstotliwość f_1. Załóżmy ponadto, że znamy widmo tego sygnału $F(j\omega)$ wyznaczone na podstawie całki Fouriera co przedstawia Rysunek 2.6.

Załóżmy dalej, że próbkujemy ten sygnał z szybkością F_s, która jest dużo większa (co najmniej dwa razy większa) niż częstotliwość f_1. Ten spróbkowany sygnał przedstawiony jest na rys 2.6c, zaś obliczone próbki widmowe przedstawione są na rys. 2.6d. Rysunek ten pokazuje również możliwe do wyliczenia próbki widmowe w sytuacji, gdy wskaźnik *m* może przybierać dowolne całkowite wartości spoza przedziału (0,N-1). Jak widać z tego rysunku jest to przebieg periodyczny z okresem *N* (lub z okresem F_s), bowiem kolejne widma są

prostymi translacjami widma wyznaczonego na podstawie całki Fouriera. Dzięki założeniu, że $F_s >> f_1$, kolejne translacje tego widma nie zachodzą na siebie, co pozwala przypuszczać, że tak dokonana analiza jest poprawna. Załóżmy jednak teraz, że zmniejszamy szybkość



Rysunek 2.6. Przykładowy sygnał i jego widmo dwustronne obliczone za pomocą całki Fouriera

próbkowania tak, że $2f1 > Fs > f_1$. Widmo tak spróbkowanego sygnału przedstawia rys. 2.6e. Jak widać z tego rysunku dla tak specyficznie dobranej szybkości próbkowania kolejne translacje widma zachodzą na siebie, co czyni wynik analizy niemiarodajny. Zjawisko to nosi nazwę aliasingu i występuje wówczas, gdy szybkość próbkowania jest mniejsza niż podwojona maksymalna częstotliwość występująca w analizowanym sygnale. Aby go uniknąć trzeba dobrać szybkość próbkowania tak, by była ona co najmniej dwa razy większa niż maksymalna częstotliwość występująca w analizowanym sygnale. Jest to kwintesencja twierdzenia (warunku) Nyquista, odgrywającego niezmiernie istotną rolę w analizie sygnałów. Warto też pamiętać, że bezstratne odtworzenie sygnału na podstawie jego widma, w sytuacji gdy występuje aliasing, jest niemożliwe. Dlatego m.in. szybkość próbkowania

sygnałów zapisanych na płytach kompaktowych jest równa 44.1 kHz, bowiem nawet jeśli czułość naszego słuchu sięga 20 kHz, to warunek Nyquista jest i tak spełniony.

Jak widać z rys. 2.5 zamieszczone widma są symetryczne względem całkowitych wielokrotności szybkości próbkowania, lub całkowitej wielokrotności liczby próbek *N* poddanych analizie. Jest to jedna z podstawowych cech wyników dyskretnej transformaty Fouriera. Ponieważ wynik tej analizy jest zbiorem liczb zespolonych, to symetrię tę można zapisać następująco:

$$\begin{aligned} &|X(-m)| = |X(m)| \\ &Arg(X(-m)) = -Arg(X(m)) \\ &X(-m) = X^*(m) \end{aligned} \tag{2.24}$$

oraz:

$$\begin{aligned} &|X(kN-m)| = |X(-m)| = |X(m)| \\ &Arg(X(kN-m)) = Arg(X(-m)) = Arg(X(m)) \\ &X(kN-m) = X^*(m) \end{aligned} \tag{2.25}$$

dla dowolnego całkowitego *k*.

Stwierdzona symetria oraz możliwość wystąpienia alisasingu sugeruje, że widmo zawiera dwa razy więcej próbek niż ich niezbędna liczba. Rzeczywiście, miarodajnym wynikiem analizy widmowej dokonywanej za pomocą dyskretnej transformaty Fouriera jest tylko pierwsza część próbek tj. od próbki zerowej (której wartość wyraża średnią wszystkich próbek czasowych) do próbki o numerze *N/2*.

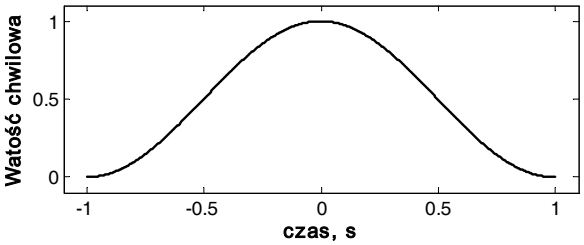
Dlatego też przebieg widmowej gęstości amplitudy wyznaczają kolejne moduły wyrażeń *X(m)* otrzymanych na podstawie DFT ale tych o numerach od 0 do *N/2*. Jednak najczęściej widmo sygnału przedstawia się poprzez wykreślenie mocy (czyli kwadratu modułu) każdej ze składowych w funkcji ich częstotliwości. Zależność ta, będąca funkcją częstotliwości, nosi nazwę widmowej gęstości mocy. Pomimo, że jest to zbiór wartości dyskretnych, przypadających dla ściśle określonych wartości częstotliwości, to każda z nich wyraża moc sygnału przepadającą na pewien przedział częstotliwości, a więc zupełnie inaczej niż w przypadku rozłożenia sygnału za pomocą szeregu Fouriera. Jeśli dla N-punktowej transformaty Fouriera kolejne wartości *X(m)* odległe są o Δf (tzw. rozdzielczość częstotliwościowa analizy), to każda z tych wartości określa moc sygnału w paśmie częstotliwości o szerokości Δf, tj. od $f_m - 0.5\Delta f$ do $f_m + 0.5\Delta f$. Łatwo więc zauważyć, że powiększanie liczby próbek wejściowych prowadzi do coraz lepszej rozdzielczości częstotliwościowej, a więc i do dokładniejszej analizy. Jednak z drugiej strony powiększanie

liczby próbek prowadzi do ‘gubienia’ informacji o zmianach zachodzących w czasie trwania sygnału. Np. jeśli analizie widmowej poddamy sygnał, którego częstotliwość zmieniła się skokowo z f_1 na f_2 to otrzymamy dwa dyskretne prążki. Jest więc to wynik bardzo uśredniony, który też nie jest w pełni miarodajny. Zatem powiększanie liczby próbek bez ryzyka utraty informacji o zmienności w czasie można stosować do sygnałów, które nie zmieniają się. Natomiast w sytuacji np. sygnału mowy zabieg taki prowadzi do nadmiernego uśrednienia.

Powyższe przesłanki legły u podstaw definicji widma długoterminowego i krótkoterminowego. Pierwsze z nich oznacza obliczenie widma na podstawie dużej liczby próbek, podczas gdy drugie (nazywane też niekiedy widmem chwilowym) bazuje na bardzo ograniczonej liczbie próbek. Obliczając widmo chwilowe dzieli się sygnał na niewielkie porcje (np. po 1000 próbek) i dla każdej porcji próbek wyznacza się odrębną DFT. Dodatkowo, aby uniknąć niepożądanego poszerzenie widma związanego z nagłym włączeniem/wyłączeniem sygnału na każdą porcję sygnału zakłada się tzw. okienko czasowe, czyli przemnaża się ją przez sygnał narastający i wybrzmiewający łagodnie. Najczęściej stosuje się w tym zakresie okno Hanninga opisane następującą zależnością:

$$w[k+1] = 0.5\left\{1-\cos\left(2\pi\frac{k}{n-1}\right)\right\} \qquad k = 0,1,2,...,n-1. \tag{2.26}$$

którego przebieg ilustruje Rysunek 2.7.



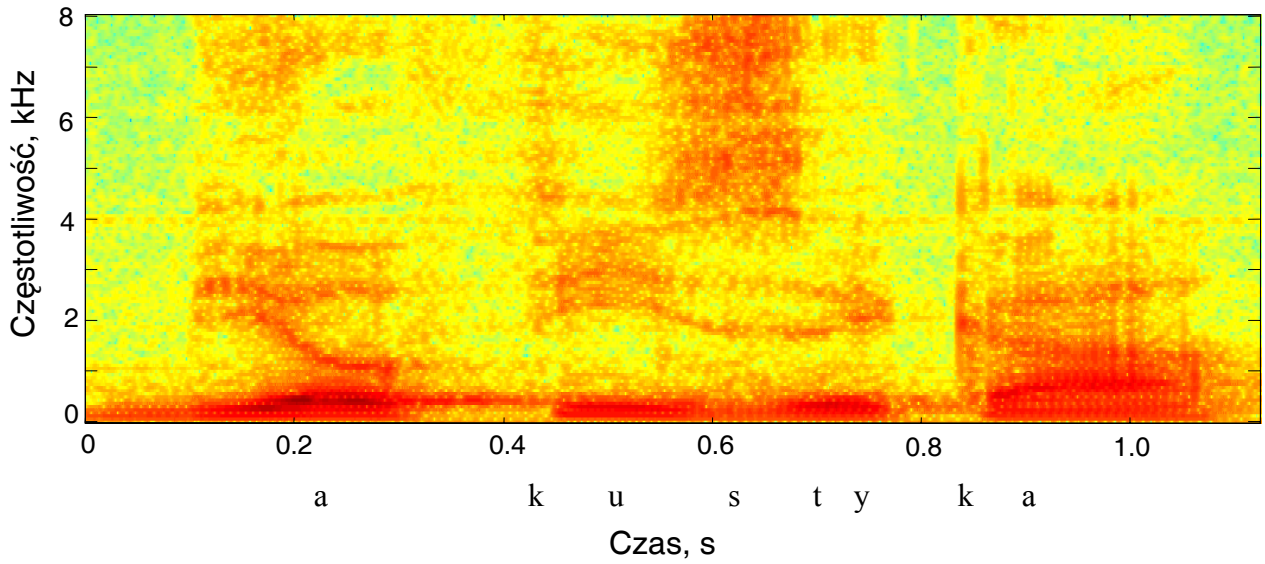
Rysunek 2.7. Przebieg czasowy okna czasowego Hanninga.

Warto też w zakończeniu wspomnieć o tzw. technice nakładkowej. Podział sygnału na porcje czasowe nie oznacza bowiem, że muszą być one rozłączne: następna porcja sygnału nie musi wcale zaczynać się dokładnie z końcem poprzedniej, a może zawierać część próbek (np. połowę) z poprzedniego okna. Taki podział sygnału nosi właśnie nazwę techniki nakładkowej, jako że kolejne okna nakładają się na siebie. Najczęściej stosuje się w tym zakresie 50% nakładanie: każde następne okno zawiera połowę próbek z poprzedniego okna.

2.5. Analiza sygnałów mowy. Spektrogram

Z uwagi na znaczną specyfikę sygnałów mowy zastosowanie konwencjonalnych metod analizy widmowej, opartych o np. wyznaczenie widma długoterminowego, nie przynosi

zadowalających skutków. Dlatego też najważniejszą i najczęściej stosowaną metodą analizy dźwięków mowy jest metoda oparta o wyznaczanie chwilowego widma sygnału oraz przedstawienie ewolucji widm chwilowych w czasie na tzw. spektrogramie. Spektrogram jest trójwymiarowym wykresem pokazującym ilość energii przypadającej dla danej częstotliwości jako funkcję czasu. Aby wyznaczyć spektrogram dzieli się sygnał mowy na krótkie odcinki czasowe i oblicza się widma chwilowe dla tych odcinków. Czas odłożony jest na osi odciętych, częstotliwość na osi rzędnych a stopień zaczernienia (lub odpowiednie kolory) odwzorowuje natężenie (poziom natężenia) dźwięku. Przykładowy spektrogram ilustruje. Rys. 2.8.



Rysunek 2.8. Spektrogram słowa 'akustyka'.

Wyznacza się zazwyczaj dwa typy spektrogramów, tzn. szerokopasmowy i wąskopasmowy, a podział ten wynika ze stosowanej w przeszłości techniki wyznaczania widma sygnałów za pomocą filtrów. W szerokopasmowym spektrogramie najczęściej stosuje się pasmo o szerokości 300 Hz, a uzyskiwana przy tym rozdzielczość czasowa jest wystarczająco dobra, bowiem pozwala obserwować poszczególne impulsy tonu krtaniowego (każdy z impulsów tonu krtaniowego odpowiada jednemu okresowi drgania fałdów głosowych). Nie pozwala jednak na jednoczesną obserwację poszczególnych harmonicznych częstotliwości podstawowej głosu. Szerokopasmowe spektrogramy są często stosowane do analizy przebiegu czasowego i kształtu formantów. Rysunek 2.8 przedstawia właśnie przykład szerokopasmowego spektrogramu. Miejsca oznaczone kolorem czerwonym wskazują na znaczną koncentrację energii dla określonych częstotliwości i w określonych chwilach czasu, podczas gdy miejsca oznaczone kolorem zielonym wskazują na brak energii. Czerwone pasma przebiegające w przybliżeniu poziomo odpowiadają formantom. W

przypadku spektrogramu wąskopasmowego szerokość pasma analizującego jest równa 45 Hz. Jest to zwykle wystarczająca rozdzielczość dla obserwacji zmienności poszczególnych składowych harmonicznych, ale daje też gorszą rozdzielczość czasową, uniemożliwiając obserwację impulsów tonu krtaniowego.

Współczesne metody wyznaczania spektrogramu bazują na wyznaczaniu widm chwilowych i ich odpowiednim przedstawieniu. Szerokość pasma analizującego wynika wprost z rozdzielczości zastosowanej analizy widmowej. Np. środowisko Matlab zawiera funkcję *specgram* umożliwiającą szybkie wyznaczenie spektrogramu dla parametrów analizy.

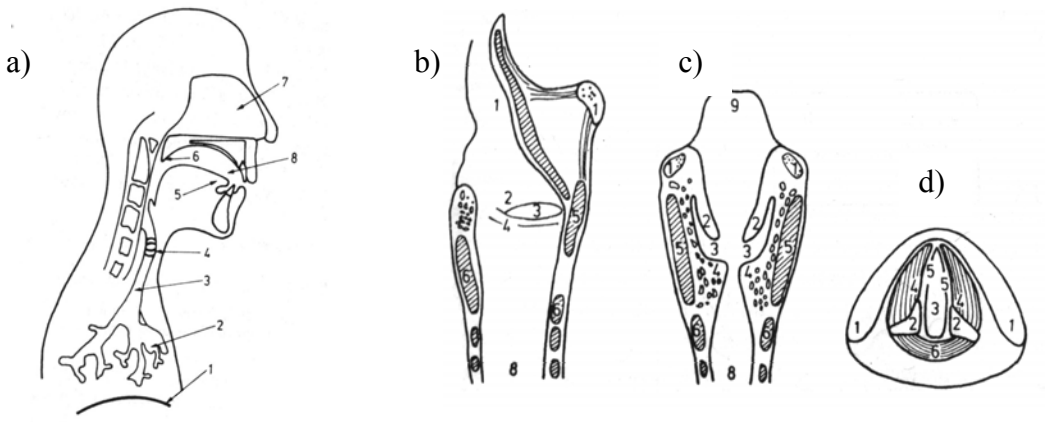
Jednak istnieje wiele specjalizowanych programów do obróbki i analizy sygnałów mowy z których *PRAAT* wydaje się być jednym z najdogodniejszym (http://www.praat.org). Przykładowe wyniki analizy sygnałów mowy uzyskane za pomoca programu *PRAAT* zamieszczono w rozdziale 4.3 oraz 5.1 niniejszego opracowania.

Analizując sygnał mowy wyznacza się też często przebieg zmian częstotliwości tonu krtaniowego na podstawie tzw. przejść sygnału mowy przez zero. Zmiany częstotliwości tego tonu odwzorowują tzw. intonację związaną ze zmianą wysokości głosu (częstotliwości tonu krtaniowego), która często sugeruje ton wypowiedzi oraz stany emocjonalne mówcy.

3. MOWA

3.1. Wytwarzanie dźwięków mowy

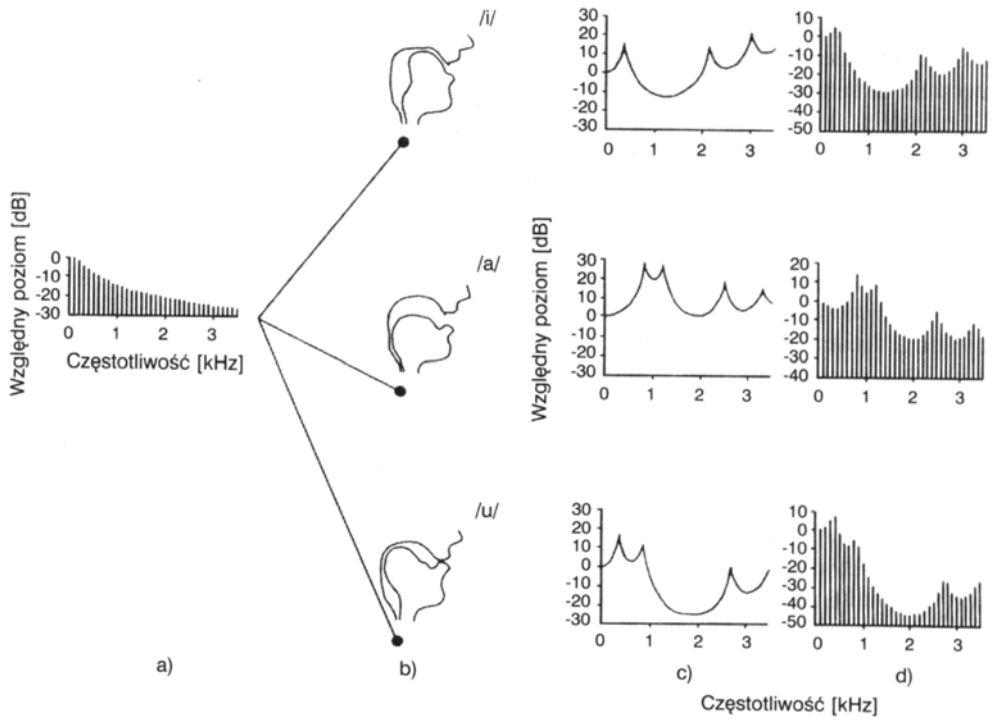
Dźwięki mowy wytwarzane są w tzw. organie mowy, którego przekrój porzeczny przedstawia Rysunek 3.1a. Zasadniczymi jego elementami są płuca, tchawica, krtań, gardło, nos, jama nosowa oraz usta. Część drogi głosowej leżącą powyżej krtani nazywa się kanałem głosowym. Kształt jego przekroju poprzecznego może się znacznie zmieniać pod wpływem ruchów języka, warg i szczęki (tzw. narządów artykulacyjnych) umożliwiając wymawianie (artykulację) różnych głosek. Zasadniczymi elementami krtani, które ilustruje Rysunek 3.1b-d, są tzw. fałdy (wiązadła) głosowe. Przestrzeń pomiędzy fałdami głosowymi nazywa się głośnią. Fałdy głosowe mogą się otwierać i zamykać zmieniając w ten sposób rozmiary głośni, co wpływa na przepływ powietrza z płuc. Dźwięk wytwarzany w trakcie wydostawania się powietrza z płuc przez fałdy głosowe, które wykonują szybkie ruchy (periodyczne lub quasi-periodyczne) zamykające i otwierające głośnię, nazywa się tonem krtaniowym. Dźwięki wytwarzane przy udziale drgań fałdów głosowych nazywają się dźwięcznymi. Tony krtaniowe są dźwiękami periodycznymi o dość małej częstotliwości podstawowej, zawierającymi harmoniczne ze znacznego zakresu częstotliwości.



Rysunek 3.1. Część (a) przedstawia organ mowy: 1- przepona, 2 – płuca, 3 – tchawica, 4 – wiązadła 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 (wewnątrz fałdów głosowych), 5 – wiązadła głosowe, 6 – mięśnie międzynalewkowe. Rysunek na podstawie Basztury (1988).

Częstotliwość podstawowa tego drgania zawiera się w przedziale od 100 do 140 Hz dla głosów męskich, lub od 200 do 250 Hz dla głosów kobiecych. Większa część energii tego dźwięku zawarta jest w składowych o małych częstotliwościach, co ilustruje Rysunek 3.2a.

Ton krtaniowy stanowi jednocześnie pewien pierwotny sygnał wejściowy do kanału głosowego, w którym jego widmo podlega znacznym modyfikacjom. Kanał głosowy zachowuje się jak układ filtrów (rezonatorów) o określonych częstotliwościach rezonansowych (por. Rysunek 3.2b) tak, że widmo tonu krtaniowego po przejściu przez układ tych filtrów charakteryzuje się pewnymi maksimami i minimami lokalnymi. Te lokalne



Rysunek 3.2. Ilustracja sposobu generacji dźwięków niektórych samogłosek. W części (a) przedstawiono widmo dźwięku (tonu krtaniowego) wytwarzanego w wyniku drgań fałdów głosowych. Dźwięk ten składa się z wielu harmonicznych, których poziom maleje ze wzrostem częstotliwości. W części (b) przedstawiono przekroje poprzeczne kanału głosowego w konfiguracjach odpowiadających trzem samogłoskom. Część (c) przedstawia charakterystyki przeniesienia kanału głosowego odpowiadające różnym jego konfiguracjom, właściwym poszczególnym samogłoskom. Część (d) przedstawia widma samogłosek po przejściu tonu krtaniowego (a) przez filtry o charakterystykach przedstawionych w części (c). Rysunek na podstawie Moore (1999).

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. Lokalne minima widma sygnału po przejściu przez kanał głosowy nazywa się często antyformantami, a ich obecność w sygnale ma równie ważne znaczenie jak obecność formantów. Przykładowe widma tonu krtaniowego po przejściu przez kanał głosowy przedstawia Rysunek 3.2c. 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.

Ton krtaniowy, modyfikowany przez narządy artykulacyjne w kanale głosowym, nie jest jedynym dźwiękiem jaki może wygenerować aparat głosowy. Źródłem dźwięku może być np. nagłe uwolnienie powietrza zgromadzonego w pewnym miejscu kanału głosowego lub przepuszczanie strumienia powietrza przez przewężenie. Dlatego też podział dźwięków mowy ze względu na sposób ich wytwarzania przedstawia się następująco:

- aspiraty (samogłoski wypowiadane szeptem), które powstają przez przepuszczanie strumienia powietrza przez przewężenie krtani (szum krtaniowy),
- dźwięczne bezszumowe (samogłoski) powstające przez generowanie tonu krtaniowego,
- dźwięczne szumowe, które powstają przy generowaniu tonu krtaniowego i przepuszczania powietrza przez przewężenie kanału głosowego (/z/ /dz/),
- szumowe (bezdźwięczne), które powstają wyłącznie przez przepuszczanie powietrza przez przewężenie pewnej części kanału głosowego, (/f/, /s/),
- udarowe (zwarte), które powstają na skutek uwolnienia powietrza zgromadzonego w pewnym miejscu kanału głosowego (/b/ lub /p/).

Innego podziału dźwięków mowy można dokonać analizując ich zmienność w czasie. Ustalonymi przebiegami w czasie charakteryzują się samogłoski oraz spółgłoski nosowe, boczne i trące. Grupa spółgłosek zwartych charakteryzuje się przebiegami nieustalonymi zbliżonymi do dźwięków o charakterze impulsowym.

3.2. Głoski, allofony i fonemy

Zasadniczym elementem składowym każdej wypowiedzi są wyrazy, na które składają się głoski. Głoskę intuicyjnie pojmować można jako fonetyczny odpowiednik litery, choć relacje pomiędzy literami pisanymi a ich wymową są w ogólności w wielu językach bardzo zawiłe i nieregularne. Język polski jest pod tym względem jest dość regularny, choć pomiędzy elementami mowy a odpowiadającymi im znakami pisma zachodzą złożone relacje. Zapis w którym zachodzą najprostsze i bezwyjątkowe relacje głoska – litera nazywa się transkrypcją fonetyczną. 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.

Segment fonetyczno-akustyczny to fragment przebiegu czasowego sygnału mowy w obrębie którego parametry fonetyczno-akustyczne są stałe lub zmieniają się w jednym kierunku. Rozróżniamy segmenty zależne i niezależne. Segment niezależny wraz z sąsiednimi segmentami zależnymi stanowi głoskę. Jeśli różnice pomiędzy głoskami mają charakter przypadkowy albo związany z indywidualnymi różnicami pomiędzy mówcami to mówimy, że mamy do czynienia z klasą głosek równoważnych, czyli allofonem akustycznym. W sytuacji gdy na różnice pomiędzy allofonami wpływają bezpośrednio sąsiednie allofony to mamy do czynienia z różnicami kontekstowymi. Natomiast jeśli różnice pomiędzy allofonami są niezależne od sąsiednich allofonów to mówimy o różnicach dystynktywnych. Grupę allofonów pomiędzy którymi nie ma różnic dystynktywnych nazywamy fonemem. Dlatego też fonemy można traktować jako podstawowe elementy alfabetu fonetycznego.

Transkrypcja fonetyczna korzysta z symboli, z których większość ma taki sam sens jak odpowiadające im litery zwykłej pisowni, a inne oznaczane są znakami specjalnymi. „Język" transkrypcji w swym założeniu pomyślany jest jako język międzynarodowy i często zdarza się, że jakaś litera zapisu fonetycznego oznacza zupełnie inną głoskę niż ta, z którą kojarzy się ona w polskiej ortografii. Np. znak fonetyczny /w/ odpowiada głosce pisanej w języku polskim jako Ł a fonetyczne /v/ odpowiada głosce sygnalizowanej przez W. Tak więc wyraz wełna ma w transkrypcji fonetycznej postać /'vewna/ (znak ' oznacza, że następująca po nim sylaba jest akcentowana).

3.3. Samogłoski

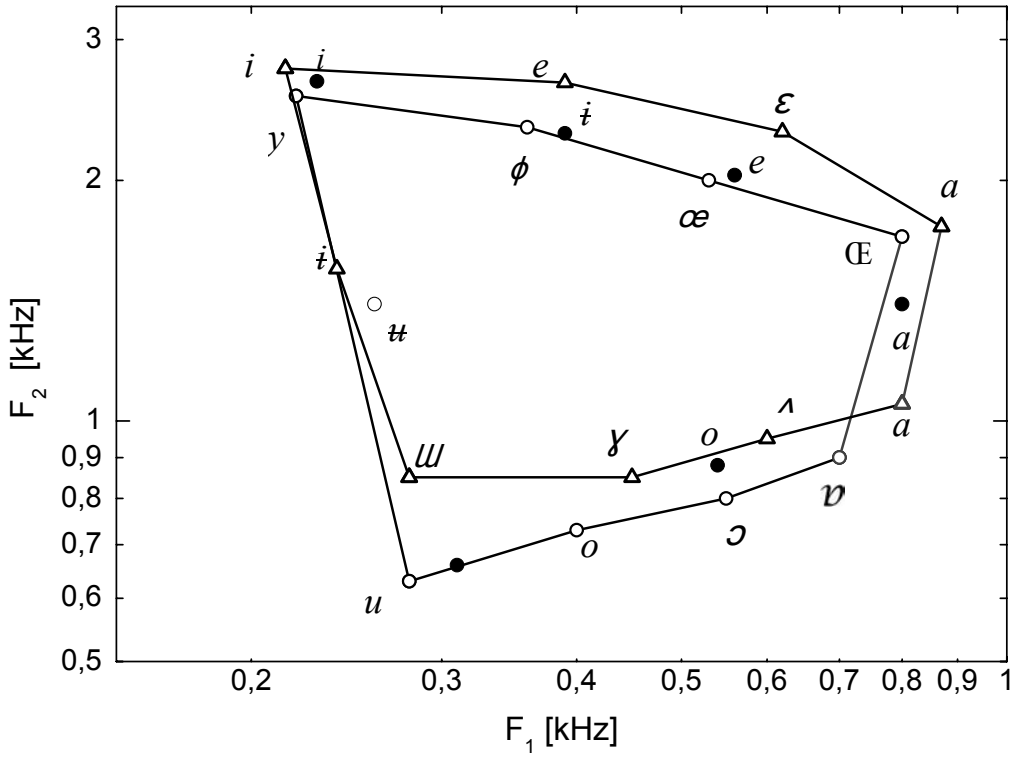
Wśród ustalonych dźwięków mowy samogłoski są tymi dźwiękami, które najłatwiej scharakteryzować. Powstają one w warunkach swobodnego przepływu powietrza wzdłuż języka. Wiązadła głosowe drgają periodycznie lub quasi-periodycznie a podniebienie miękkie jest uniesione do góry blokując dostęp powietrza do 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, o których będzie mowa w dalszej części opracowania.

4. ĆWICZENIE NR 1: ANALIZA WYBRANYCH CECH SAMOGŁOSEK JĘZYKA POLSKIEGO

4.1. Wprowadzenie

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

Rysunek 4.1 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.



Rysunek 4.1. 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), to ich względne wartości nie zależą od indywidualnych cech głosowych. Poziomy poszczególnych formantów, wyrażone względem formantu F1, również wykazują stałe w przybliżeniu wartości. Poziom drugiego formantu jest od 5 (jak w /a/) do 15 dB (jak w /i/) niższy od poziomu pierwszego formantu. Poziom trzeciego formantu jest z kolei o 12 (jak w /a/) do 40 dB (jak w /u/) niższy niż formant F1. Poziom czwartego formantu jest na ogół ponad 20 dB niższy niż formantu podstawowego i nie odgrywa on praktycznie większej roli.

4.2. Cel ćwiczenia

Głównym celem ćwiczenia jest określenie podstawowych widmowych cech samogłosek języka polskiego /i/, /ɨ/, /e/, /a/, /o/, /u/.

Ćwiczenie to polegać ma na :

a) zarejestrowaniu samogłosek izolowanych (wypowiadanych oddzielnie),
b) wyznaczeniu spektrogramu dla samogłosek w przypadku analizy wąsko- i szerokopasmowej,
c) obliczeniu średnich wartości częstotliwości formantów F1, F2, F3 oraz F4 dla poszczególnych samogłosek,
d) wyznaczeniu pętli formantowej dla zarejestrowanych samogłosek,
e) porównaniu uzyskanych wyników z danymi z literatury.

Ponadto wykonujący ćwiczenie mogą dokonać analogicznej analizy samogłosek znajdujących się na początku, w środku i na końcu wyrazu. Również istnieje możliwość dokonania analizy samogłosek w kontekście trybu w jakim wypowiadany jest analizowany fragment sygnału mowy: oznajmujący, rozkazujący i pytający.

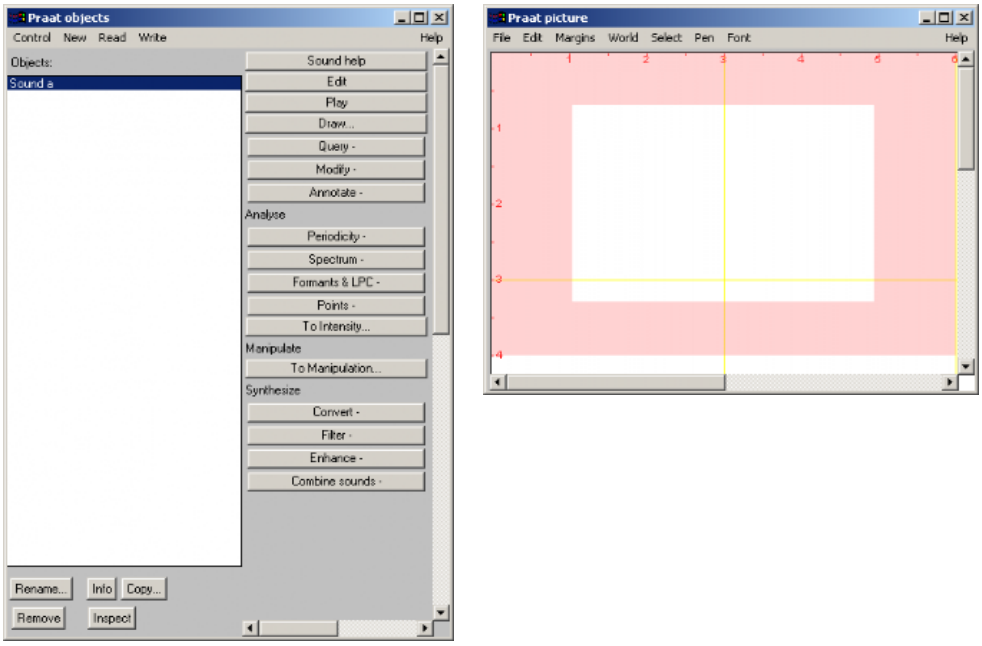
4.3. Przebieg ćwiczenia

Jednym z wielu pakietów software służących do przeprowadzenia analizy dźwięków mowy jest program o nazwie „*PRAAT*”. Program ten został opracowany przez autorów Paula Boersman’a i Davida Weenink’a z Institute of Phonetics Sciences of the University of Amsterdam. Program ten jest bardzo dobrym narzędziem do analizy sygnałów mowy. Zakres dostępnych w nim opcji znacznie wykracza poza zagadnienia omawiane w opracowaniu.

Poniżej przedstawiono przykłady działania programu „*PRAAT*” które są bezpośrednio związane z tematyką tego ćwiczenia.

Po uruchomieniu programu pojawiają się dwa okna z których jedno „*PRAAT* objects” zawiera wszystkie opcje analizy, drugie zaś „*PRAAT* picture” umożliwia tworzenie obiektów graficznych w celu dokumentowania wyników analiz. Przykładowe okna tego programu ilustruje Rysunek 4.2.

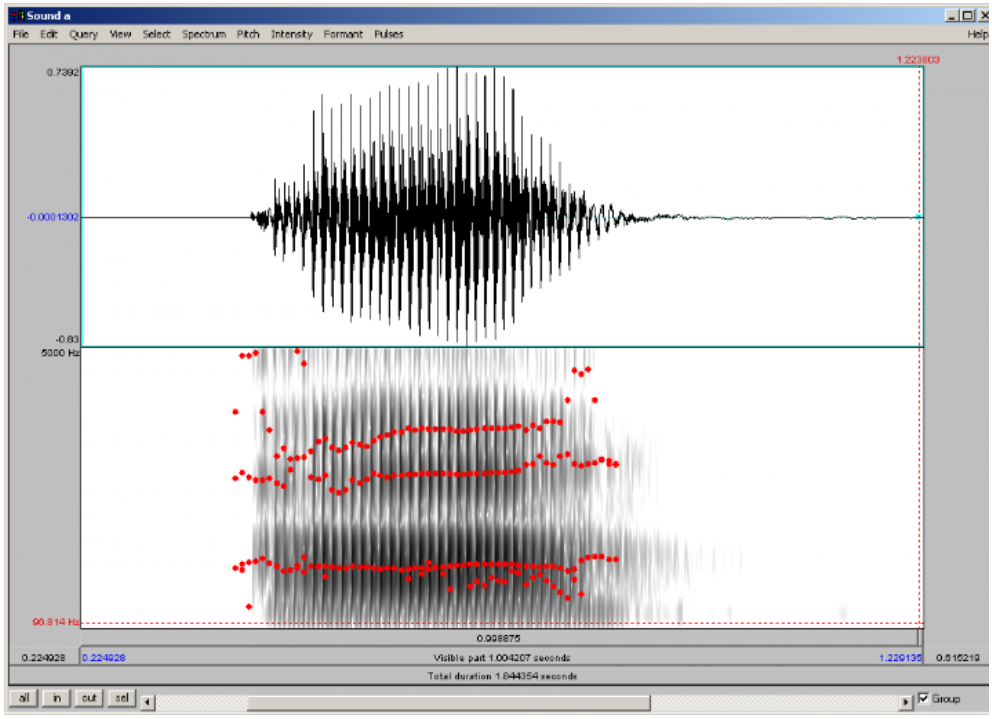
W celu zarejestrowania dźwięku należy z menu głównego „New” wybrać opcję nagrywania. Następnie zarejestrowanemu sygnałowi należy przypisać nazwę. Nazwa ta będzie widoczna w oknie „Objects”. Dla danego obiektu przyporządkowane są dostępne w formie przycisków możliwości analizy. Po wykonaniu dowolnej operacji analizy na zarejestrowanym sygnale program tworzy nowy obiekt, przyporządkowując mu automatycznie nazwę. W celu przeniesienia danego rysunku do okna „Praat picture” należy posłużyć się poleceniem „Draw” dostępnym w oknie głównym programu.



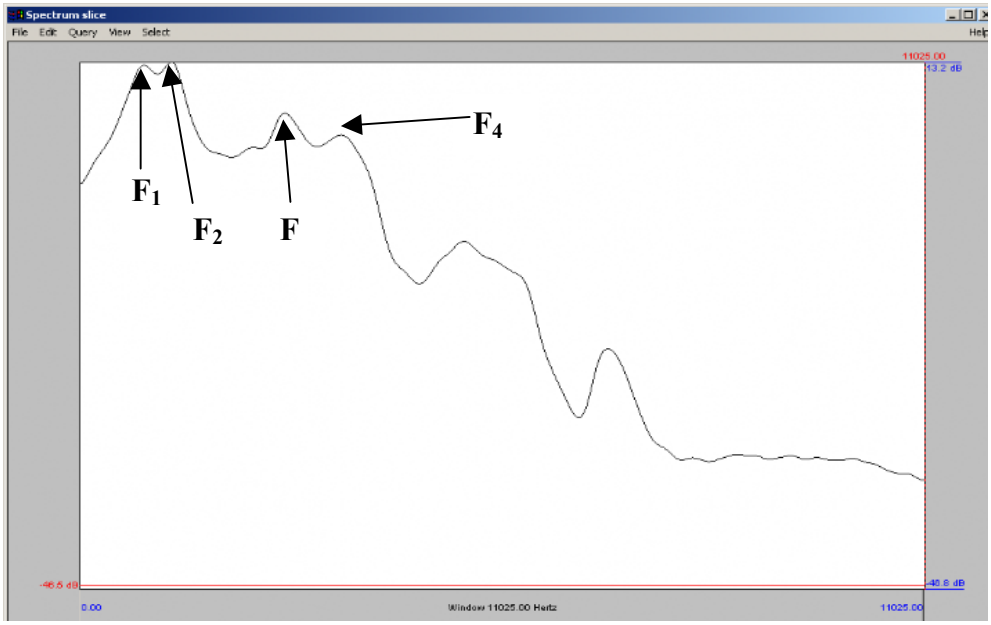
Rysunek 4.2. Główne okna programu „*PRAAT*”.

Rysunek 4.3 przedstawia przykład zarejestrowanej samogłoski /a/. Górny panel tego rysunku przedstawia przebieg czasowy sygnału, natomiast dolny spektrogram. Opcje z menu „Formant” pozwalają na wyznaczenie parametrów poszczególnych formantów, tzn. średnich wartości częstotliwości formantów, Ponadto istnieje możliwość wyznaczenia zależności zmian częstotliwości formantów w funkcji czasu trwania sygnału. Punkty naniesione na spektrogram odpowiadają chwilowym wartościom częstotliwości poszczególnych formantów

F1, F2 oraz F3. Istnieje także możliwość wyznaczenia średniego widma gęstości mocy dla zarejestrowanego sygnału.



Rysunek 4.3. Przebieg czasowy i spektrogram samogłoski /a/.



Rysunek 4.4.Obwiednia widma gęstości mocy dla samogłoski /a/.

Rysunek 4.4 przedstawia przykład obwiedni widma gęstości mocy dla samogłoski /a/. Na rysunku zaznaczono formanty, które odpowiadają lokalnym maksimom obwiedni widma gęstości mocy. W zależności od rozdzielczości analizy widmowej (szerokości okna analizy)

istnieje możliwość „wygładzenia” struktury widmowej, lub też w przypadku, kiedy chcemy uzyskać więcej informacji o subtelnej strukturze widmowej zawężenia okna czasowego analizy.

W celu wyznaczenia pętli samogłoskowej, czyli zależności częstotliwości drugiego formantu F2 od częstotliwości pierwszego formantu F1 dla poszczególnych samogłosek należy dokonać kilkukrotnej rejestracji każdej z samogłosek, a następnie obliczyć średnie wartości częstotliwości formantów F1, F2, F3 i F4. Uzyskane wyniki (wartości średnie i odchylenia standardowe) należy przedstawić w formie wykresu pętli formantowej . Ponadto na wykres należy nanieść dane z literatury (patrz Tabela.1, Załącznik 1).

Uwaga! W przypadku analizy formantów dla głosu żeńskiego zakres analizy należy ustawić do 5.5 kHz, natomiast w przypadku głosu męskiego do 5 kHz.

Tabelę ze średnimi wartościami częstotliwości formantowych samogłosek języka polskiego przedstawiono w Załączniku 1.

5. ĆWICZENIE NR 2: PODSTAWOWE CECHY WIDMOWE SAMOGŁOSEK I WYBRANYCH SPÓŁGŁOSEK JĘZYKA POLSKIEGO

5.1. Wprowadzenie

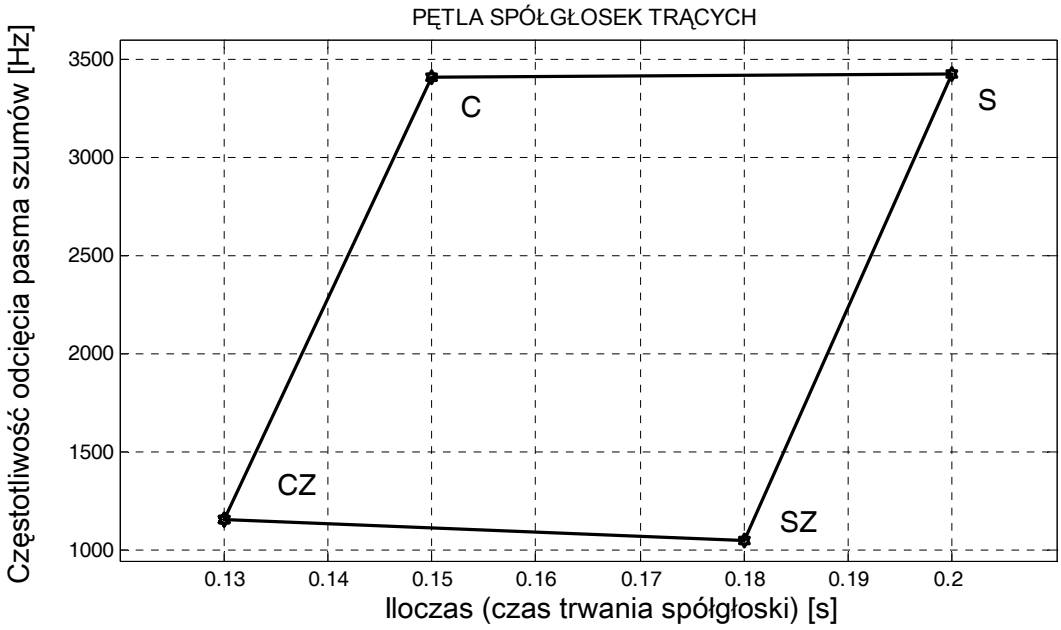
Spółgłoski trące (szczelinowe) są pasmami szumu, natomiast samogłoski są wielotonami (ściślej: wielotonami anharmonicznymi). Spółgłoski trące (i zwarto-trące) to spółgłoski takie jak : *s , sz , z , ż , c , cz, dz , dż.* Spółgłoski trące, jako impulsy szumów pasmowych mogą być charakteryzowane prze dwa parametry : przez czas trwania tej spółgłoski (czyli - przez tzw. "iloczas") oraz przez szerokość pasma szumu, lub równoważnie - przez iloczas oraz przez dolną częstotliwość odcięcia pasma szumu (jeśli stwierdzi się, że górna częstotliwość odcięcia pasma szumu nie różnicuje poszczególnych spółgłosek trących). Wyróżnienie tych dwóch najważniejszych parametrów charakteryzujących spółgłoski trące powoduje, że możemy te spółgłoski klasyfikować ze względu na wartości częstotliwości dwóch pierwszych formantów.

Dla poszczególnych samogłosek częstotliwości pierwszego i drugiego formantu nanosimy w odpowiedniej przestrzeni dwuwymiarowej uzyskując pewną krzywą zwaną "pętlą formantową" (patrz Rysunek 4.1). Jeżeli ograniczaliśmy się wyłącznie do samogłosek przednich i tylnich, to uzyskiwaliśmy w tej przestrzeni cztery wierzchołki, które połączone liniami prostymi stanowiły tzw. "czworobok samogłoskowy". W przypadku samogłosek wzdłuż jednej z osi odkładamy częstotliwości pierwszego z formantów (F1), wzdłuż drugiej osi - częstotliwości drugiego z formantów (F2).

W przypadku spółgłosek trących mamy do czynienia z przestrzenią dwuwymiarową, gdzie wzdłuż jednej osi odkładamy wartości iloczasu a wzdłuż drugiej osi - dolną częstotliwość odcięcia pasma szumu. Zatem przez analogię do "czworoboków samogłoskowych" możemy również konstruować "czworoboki spółgłosek trących oraz zwarto-trących". Te naniesione wartości są współrzędnymi poszczególnych samogłosek lub spółgłosek trących lokalizującymi ich położenie na mapie "głoskowej", jaką stanowi sporządzony w ten sposób wykres. Przykładową pętlę spółgłoskową przedstawia Rysunek 5.1

Niezależnie od szczegółów terminologicznych, poszczególne głoski klasyfikuje się poprzez wyznaczenie względnych odległości pomiędzy nimi w pewnej, konwencjonalnej przestrzeni wielowymiarowej, niż poprzez jednoznaczne określenie ich parametrów.

Klasyfikacja takich czy innych głosek na zasadzie jednoznacznego określenia ich parametrów nie miałaby sensu ze względu na występowanie międzyosobniczych różnic pomiędzy parametrami charakterystycznymi dla poszczególnych mówców.



Rysunek 5.1. Przykładowa pętla spółgłoskowa.

Do spółgłosek zwarto-wybuchowych zaliczamy: *b , p , d , t , g , k.* Spółgłoski te należą do spółgłosek "polisegmentalnych" , czyli do spółgłosek o wyraźnie wyodrębnianych segmentach czasowych. Zarazem należą one do spółgłosek "kontekstowych", czyli ich własności zależą od tego, w jakim sąsiedztwie się znajdują, a także - od tego, na jakiej pozycji znajdują się one w wyrazie:

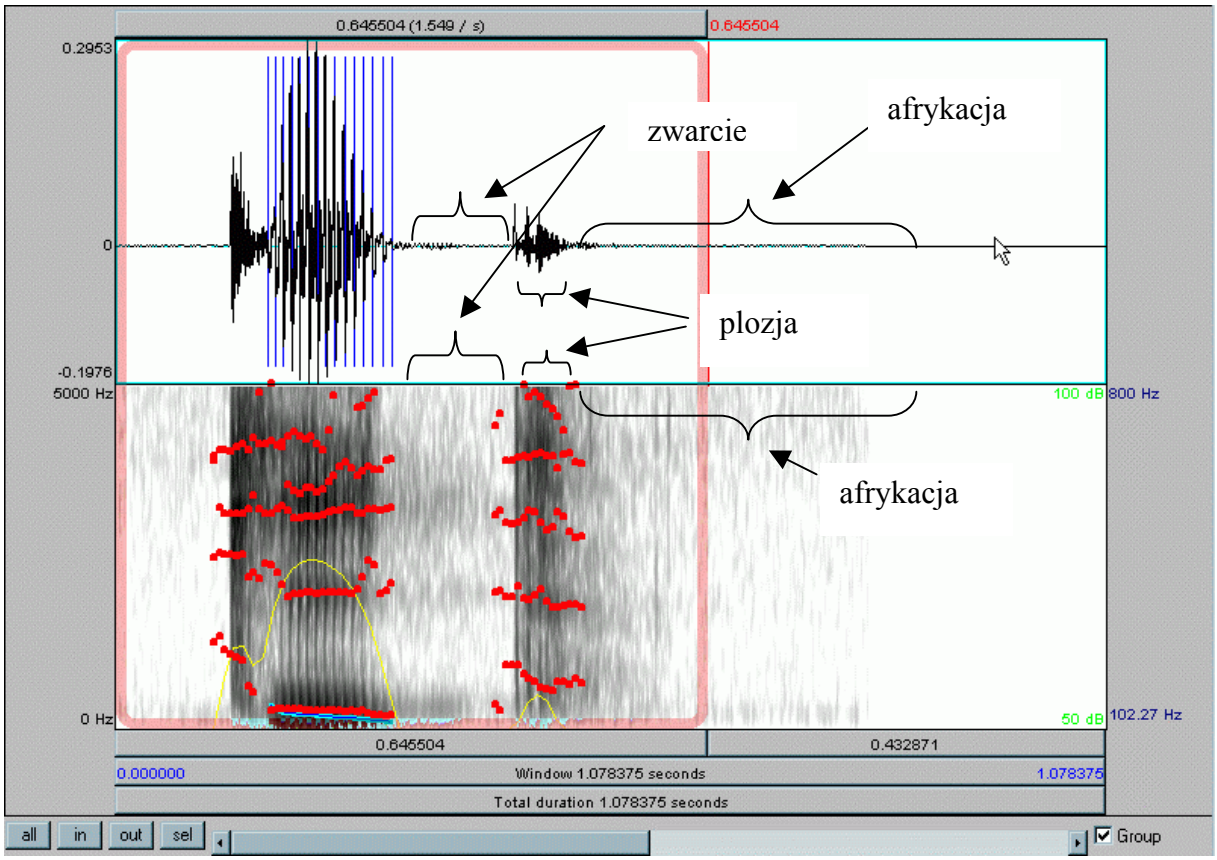
- na początku wyrazu (czyli w tzw. 'nagłosie'),
- w środku wyrazu (czyli w 'śródgłosie'),
- na końcu wyrazu (czyli w tzw. 'wygłosie').

W przypadku spółgłosek zwarto-wybuchowych można wyróżnić 3 charakterystyczne segmenty fonetyczno akustyczne lub tzw. trzy fazy artykulacji tych głosek:

1) Segment zwarcia (fazę zwarcia) - na spektrogramie jest to interwał ciszy (szczególnie dobrze widoczny w przypadku spółgłosek będących w śródgłosie oraz w wygłosie). W przypadku dźwięcznych spółgłosek zwarto-wybuchowych (takich jak: *b , d , g*) na spektrogramie może być widoczny w tej fazie zwarcia ciemny "podest" odpowiedzialny za obecność tonu krtaniowego. Faza zwarcia jest fazą chwilowego zamknięcia przepływu powietrza w torze artykulacyjnym. W tej fazie następuje sprężenie powietrza zamkniętego w torze artykulacyjnym. (Patrz Rysunek 5.2).

2) Segment plozji (fazę plozji) - na spektrogramie jest to krótki, szerokopasmowy impuls ("trzask") zawierający niemalże energie dla wszystkich częstotliwości. Faza plozji jest fazą "wybuchu" - nagłego uwolnienia powietrza sprężonego w poprzedniej fazie zwarcia . (Patrz Rysunek 5.2).
3) Końcowy segment szumowy (tzw. afrykacja lub aspiracja) - wyraźnie widoczny jedynie w przypadku spółgłosek izolowanych (tj. wypowiadanych samodzielnie, bez żadnego kontekstu i sąsiedztwa), bądź w przypadku spółgłosek znajdujących się we "wygłosie absolutnym" (tj. na końcu wyrazów zamykających zdanie lub wypowiadaną frazę). (Patrz Rysunek 5.2).

Te cechy spółgłosek zwarto-wybuchowych można zaobserwować na spektrogramie zawierającym widmo wyrazu "kit", który przedstawia Rysunek 5.2.



Rysunek 5.2. Górna część rysunku zawiera oscylogram, dolna - spektrogram (zarejestrowany przy pomocy programu PRAAT). Końcową spółgłoskę "t" stanowi pionowy ciemny pasek; jest on poprzedzony "jasnym wcięciem" odpowiadającym segmentowi zwarcia. Ten czarny pasek odpowiada segmentowi plozji, natomiast mniej regularne czarne obszary z prawej strony tego paska odpowiadają segmentowi szumowemu (czyli afrykacji lub aspiracji). Ponieważ analizowana spółgłoska "t" znajduje się na końcu wyrazu (czyli w tzw. "wygłosie") wspomniany segment szumowy może być wyraźnie widoczny.

Właśnie ze względu na łatwość obserwacji przy analizowaniu spółgłosek zwarto-wybuchowych bierze się po uwagę czasy trwania dwóch najważniejszych segmentów: segmentu zwarcia i segmentu plozji.

5.2. Cel ćwiczenia

Celem ćwiczenia jest zapoznanie się z ogólnymi sposobami klasyfikacji głosek - sposobami wspólnymi również dla głosek bardzo różniących się widmowo od siebie, np. samogłosek i spółgłosek trących (zwanych również spółgłoskami szczelinowymi), a także dla spółgłosek zwarto-wybuchowych.

5.3. Przebieg ćwiczenia:

1. **Analiza spółgłosek trących**.

a) Zarejestrować przy pomocy programu *PRAAT* izolowane głoski trące. Zadbać o to, aby wśród zestawu zarejestrowanych spółgłosek znalazły się również spółgłoski zwarto-trące, tj. takie jak *c, cz , dz* czy *dż*.

b) Dla każdej z nich wyznaczyć iloczas jako różnice czasów: końcowego oraz początkowego. Wyznaczyć również częstotliwość odcięcia pasma szumu ; jeżeli krawędź odcięcia pasma szumu nie jest równoległa do osi częstotliwości, wówczas jako częstotliwość odcięcia wybrać średnią arytmetyczną z obu częstotliwości "krańcowych". Nanieść znalezione współrzędne (iloczas + częstotliwość odcięcia) na osie "czworoboku spółgłoskowego" i według tych współrzędnych poumieszczać odpowiednio spółgłoski trące.

W tym *czworoboku* nanieść zarówno wartości średnie dla odpowiednich parametrów charakterystycznych (tj. średnią wartość iloczasu dla wszystkich spółgłosek *s* występujących w wyrazach z listy czy średnią wartość) oraz odchylenia standartowe. Wówczas każda spółgłoska trąca będzie na owej *"mapce głosek"* (tj. w *czworoboku* lub na *pętli)* zaznaczona jako prostokąt ; środek owego prostokąta będzie wyznaczał wyliczone wartości średnie, natomiast długości boków owego prostokąta będą równe podwojonym wartościom odpowiednich odchyleń standartowych.

Dopiero takie zobrazowanie pozwoli zorientować się we wzajemnych odległościach pomiędzy poszczególnymi głoskami w owych konwencjonalnych przestrzeniach dwuwymiarowych.

2. **Analiza spółgłosek zwarto-wybuchowych**

Do celów analizy sporządzić listę wyrazów, w których spółgłoski zwarto- wybuchowe występują zarówno w śródgłosie (w środku wyrazów), jak i w wygłosie (na końcu wyrazów) ; zadbać przy tym o to, aby w tych wyrazach występowały zarówno spółgłoski dźwięczne, jak i bezdźwięczne, np.: *baba , papa , dukat , kaduk , otok , potop , atak,* itp. Sporządzić spektrogramy izolowanych spółgłosek dźwięcznych, określić czasy trwania każdego

z tych trzech segmentów dla poszczególnych spółgłosek izolowanych.

Nagrać wyrazy z przygotowanej listy, sporządzić ich spektrogramy i przeanalizować czasy trwania zwarcia i czasy trwania plozji dla spółgłosek znajdujących się w środku wyrazu (w śródgłosie) i na końcu wyrazu (w wygłosie). Określić proporcje obu czasów trwania w śródgłosie i we wygłosie.

Wzajemna proporcja czasów zwarcia i czasów plozji może zależeć również od tego, czy w sąsiedztwie analizowanej spółgłoski zwarto-wybuchowej znajduje się inna spółgłoska zwarto-wybuchowa. Aby sprawdzić wpływ tego kontekstu (sąsiedztwa) należy przygotować dwa zestawy wyrazów zawierających "zbitki" spółgłosek zwarto-wybuchowych. W pierwszym z tych zestawów sąsiadujące ze sobą spółgłoski zwarto-wybuchowe powinny być takie same, natomiast w drugim z tych zestawów powinny ze sobą sąsiadować różne spółgłoski zwarto-wybuchowe. Zatem pierwszy zestaw mógłby przykładowo składać się z takich wyrazów, jak: *lekko, motto, attyka, Mekka*; natomiast drugi zestaw mógłby się przykładowo składać z takich wyrazów, jak : *apteka, oktet, optyka, takt, Bogdan, okpić, dukt,* itp. Sporządzić spektrogramy tych wyrazów. Sprawdzić, jak układają się proporcje czasów zwarcia i czasów plozji w przypadku spółgłosek zwarto-wybuchowych sąsiadujących ze sobą (przeprowadzić tę analizę osobno dla sąsiedztwa spółgłosek identycznych i osobno dla sąsiedztwa spółgłosek różniących się między sobą).

6. ĆWICZENIE NR 3: PROZODYCZNE CECHY MOWY

6.1. Wprowadzenie

Mechanizm artykulacji mowy ludzkiej nie jest "jednorodny" w tym sensie, że za powstawanie pewnych charakterystycznych składników tej mowy odpowiadają różne elementarne źródła dźwięku. Jednym z takich elementarnych źródeł dźwięku są tzw. więzadła głosowe, zwane potocznie - strunami głosowymi. Wzbudzane w nich drgania odpowiadają za generację tonu podstawowego lub tonu krtaniowego. Ton ten jest najniższym (na skali częstotliwości) składnikiem widma sygnału mowy. Towarzyszy on zawsze wypowiadanym samogłoskom, natomiast w przypadku spółgłosek towarzyszy on jedynie niektórym spółgłoskom - tzw. spółgłoskom dźwięcznym. W przypadku artykulacji spółgłosek bezdźwięcznych ton krtaniowy jest nieobecny. Częstotliwość tego tonu odpowiada przede wszystkim za skalę głosu ludzkiego, ton ten różnicuje przede wszystkim głosy męskie od żeńskich. W skali głosów żeńskich ton ten różnicuje soprany, mezzosoprany i alty; natomiast w skali głosów męskich różnicuje basy, barytony i tenory. Częstotliwość tonu krtaniowego ulega zmianie w trakcie trwania artykulacji, przy czym sposób tych zmian zależy od zamierzonej przez mówcę intonacji swej wypowiedzi. Jeśli chodzi o związek akcentu z przebiegiem częstotliwości tonu krtaniowego, to zdania na temat są podzielone wśród językoznawców - fonetyków. Część z nich (zwłaszcza tzw. "szkoła poznańska", wywodząca się od prof. W. Jassema) uważa, że w percepcji mowy najważniejszy jest tzw. akcent melodyczny związany z występowaniem maksimów częstotliwości tonu krtaniowego. Z kolei inni językoznawcy - fonetycy uważają, że ważniejszy jest tzw. akcent dynamiczny związany z występowaniem maksimów chwilowej mocy sygnału mowy.

6.2. Cel ćwiczenia

Celem ćwiczenia jest określenie wpływu intonacji na zmiany częstotliwości chwilowej tonu krtaniowego i zmiany wartości chwilowej poziomu natężenia w wypowiadanej frazie.

6.3. Przebieg ćwiczenia

Przygotować jakąś frazę, którą można nagrać zarówno w intonacji pytającej, jak i rozkazującej. Frazą tą może być zarówno całe zdanie, jak i pojedynczy wyraz. Pojedynczym wyrazem mogłoby być np. jakieś imię dwusylabowe. Imię to należałoby wypowiedzieć i nagrać zarówno w intonacji pytającej, jak i w intonacji rozkazującej. Przykładowo : "Taadek ?" (zdziwienie połączone z pytaniem) oraz "Tadek !" (przywołanie owego Tadka w trybie

rozkazującym). Do analizy najlepiej użyć programu *PRAAT*. Zaznaczyć (na spektogramie) cały wyraz jako blok i na górnej listwie wybrać opcję 'Pitch', a następnie opcję 'Pitch listing'. Otworzy się wówczas dodatkowe okno z dwiema kolumnami liczb. Lewa kolumna liczb będzie określać czas bieżący, druga – wartości tonu krtaniowego oznaczanego symbolicznie jako *F0.*

W niektórych fragmentach frazy ton krtaniowy zanika i wówczas w drugiej kolumnie pojawia się napis *undefined.* Ten układ dwóch kolumn można zaznaczyć jako blok i przenieść do jakiegoś tekstowego edytora ASCII, zamieniając po drodze napis *undefined* na cyfrę 0. Tak przygotowany dwukolumnowy plik liczbowy można następnie wczytać do EXCEL`a (lub innego programu pozwalającego sporządzać wykresy) i sporządzić wykres wartości F0, czyli tzw. *intonogram.*

Aby z kolei zbadać akcent dynamiczny należy w programie *PRAAT* wybrać opcję "Intensity", a następnie - "Intensity listening" Utworzy się wówczas analogiczne okno z dwiema kolumnami liczb. Druga kolumna nowo utworzonego okna zawiera tym razem chwilowe wartości poziomu natężenia. Ten układ dwóch kolumn można analogicznie przenieść do pliku tekstowego i sporządzić wykres.

We wnioskach do opracowania odpowiednio skojarzyć oba przebiegi z przebiegiem intonacji w obu wariantach wypowiedzi.

LITERATURA

Basztura, C. (**1988**). *Źródła, sygnały i obrazy akustyczne* (Wydawnictwo Komunikacji i Łączności, Warszawa).
Jassem, W. (**1973**). *Podstawy fonetyki akustycznej* (PWN, Warszawa).
Moore, B. C. J. (**1999**). *Wprowadzenie do psychologii słyszenia* (PWN, Poznań).

ZAŁĄCZNIK

Tabela 1. Częstotliwości formantów spółgłosek języka polskiego dla dziesięciu różnych mówców

/i/										
	AS	WJ	HN	ZM	TL	KD	BK	HK	BK	HS
F_1	188	196	203	208	212	226	235	215	275	244
F_2	2360	2620	2450	2130	2090	2370	2080	2420	2840	2690
F_3	3070	3180	2790	2690	2890	3290	2670	3020	3420	3430
F_4	3510	3630	3520	3580	3320	3650	3400	3410	4140	4130
/i̯/										
	AS	WJ	HN	ZM	TL	KD	RK	HK	BK	HS
F_1	269	391	262	345	316	371	262	371	343	314
F_2	2220	2210	2060	1800	1870	2190	1670	1950	2360	2310
F_3	2680	2770	2590	2450	2530	2750	2420	2510	3150	2940
F_4	3430	3630	3410	3570	3420	3500	3240	3120	4230	4200
/e/										
	AS	WJ	HN	ZM	TL	KD	BK	HK	BK	HS
F_1	530	564	561	567	524	526	557	604	620	630
F_2	1720	2030	1850	1670	1710	1950	1580	1840	2190	2230
F_3	2470	2720	2520	2580	2560	2610	2500	2550	3150	2810
F_4	3590	3580	3340	3610	3640	3260	3220	3060	4030	3890
/a/										
	AS	WJ	HN	ZM	TL	KD	RK	HK	BK	HS
F_1	744	819	757	687	718	790	683	816	1020	978
F_2	1320	1350	1230	1190	1130	1290	1300	1260	1520	1570
F_3	2650	2760	2400	2390	2520	2540	2510	2450	2860	2650
F_4	3480	2980	3280	3660	3630	2950	3160	3220	4100	3760

/o/										
	A8	WJ	HN	ZM	TL	KD	BK	HK	BK	HS
F_1	538	534	529	564	493	520	531	614	603	679
F_2	889	880	893	853	788	862	963	937	1060	1100
F_3	2750	2680	2530	2450	2540	2620	2620	2410	3030	2560
F_4	3370	3290	3330	3450	3410	3380	3190	3200	3950	3810
/u/										
	AS	WJ	HN	ZM	TL	KD	RK	HK	BK	HS
F_1	297	314	246	338	253	264	243	256	317	330
F_2	603	557	573	730	644	704	693	789	679	782
F_3	2490	2420	2440	2450	2280	2660	2520	2270	3190	3090
F_4	3140	3170	3410	3310	3050	3390	3220	2940	4060	3820