Leszek Grad[1]

KOMPRESJA STRATNA DŹWIĘKU

Streszczenie

W artykule przedstawione zostały elementarne wiadomości z zakresu kompresji stratnej dźwięku. Przedstawiony został liniowy model predykcji, wykorzystywany w kompresji dźwięku w paśmie telefonicznym oraz opisane zostały transformacje ortogonalne i model psychoakustyczny człowieka mające podstawowe znaczenia w kompresji dźwięku wysokiej jakości.

Abstract

This paper presents an overview of the basic information on sound loose compression. Linear predictive coding, which is used in voice compression in the phone band, as well as orthogonal transforms and psychoacoustic model, which are very important in high level sound compression (voice, speech), are revised.

Kompresja stratna polega na zmniejszaniu objętości danych w taki sposób, że po dekompresji dopuszcza się zniekształcenie sygnału w stosunku do pierwotnego, lecz jego percepcja przez człowieka (ucho, oko) jest taka jak oryginału lub do niego zbliżona (lub odbiega, ale świadomie się na to godzimy). Zaletą kompresji stratnej jest uzyskiwanie wysokiego stopnia kompresji, wyższego niż w metodach kompresji bezstratnej (osiąga się wyniki poniżej entropii).

Metody kompresji stratnej dźwięku bazują zasadniczo na trzech podstawach:

– wykorzystują metodę predykcji liniowej;

– wykorzystują ortogonalne transformacje liniowe (DFT, DCT) przekształcające sygnał do nowego układu współrzędnych, gdzie energia sygnału jest zgromadzona w punktach w początku układu współrzędnych;

– wykorzystują znany psychoakustyczny model słuchu człowieka, co pozwala

[1] Dr inż. Leszek Grad pracuje w Warszawskiej Wyższej Szkole Informatyki i w Instytucie Teleinformatyki i Automatyki Wojskowej Akademii Technicznej.

na usunięcie lub kodowanie z mniejszą dokładnością elementów, które ucho ludzkie nie słyszy lub słyszy słabo.

Metoda predykcji liniowej ma zastosowanie w systemach transmisji sygnału mowy w paśmie telefonicznym, w szczególności w telefonii GSM.

Transformacje ortogonalne oraz model psychoakustyczny narządu słuchu człowieka są wykorzystywane do kompresji sygnału dźwiękowego wysokiej jakości (muzyka, systemy nagłaśniania kina itp.).

W dalszej części artykułu omówione zostaną szczegółowo poszczególne zagadnienia.

1. LINIOWY MODEL PREDYKCJI – LPC

Idea metody LPC (*Linear Predictive Coding*) polega na przybliżeniu wartości sygnału kombinacją liniową wartości sygnału z chwil poprzednich. Oznaczając przez $u(k)$ wartość sygnału w chwili k, a przez $z(k)$ wynik predykcji sygnału, możemy zapisać:

$$z(k)=\sum_{i=1}^{p} a_i u(k-i) \qquad k > p \tag{1.1}$$

gdzie a_i są współczynnikami predykcji, a p jest rzędem predykcji.

Oznaczmy przez $w(k)$ różnicę sygnałów $u(k)$ i $z(k)$:

$$w(k)=u(k)-z(k) \qquad k > p \tag{1.2}$$

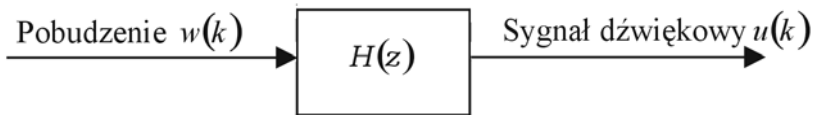
Sygnał $w(k)$ jest błędem predykcji, bywa też często nazywany sygnałem szczątkowym. Przedstawmy inaczej zależność (1.2):

$$u(k)=z(k)+w(k)=\sum_{i=1}^{p} a_i u(k-i)+w(k) \tag{1.3}$$

Transmitancja $H(z)$ układu (1.3) jest następująca:

$$H(z)=\frac{U(z)}{W(z)}=\frac{1}{1-\sum_{i=1}^{p} a_i z^{-i}} \tag{1.4}$$

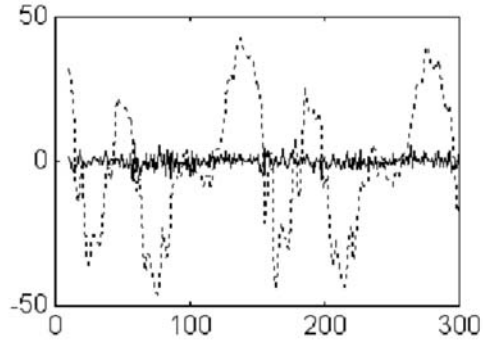
W odniesieniu do modelowania mowy układ o transmitancji (1.4) opisuje tor głosowy człowieka, a sygnał $w(k)$ jest wówczas pobudzeniem tego układu (rys. 1.1) W przypadku mowy dźwięcznej jest to sygnał tworzony przez struny i więzadła głosowe, a dla mowy bezdźwięcznej – szum przepływającego powietrza.



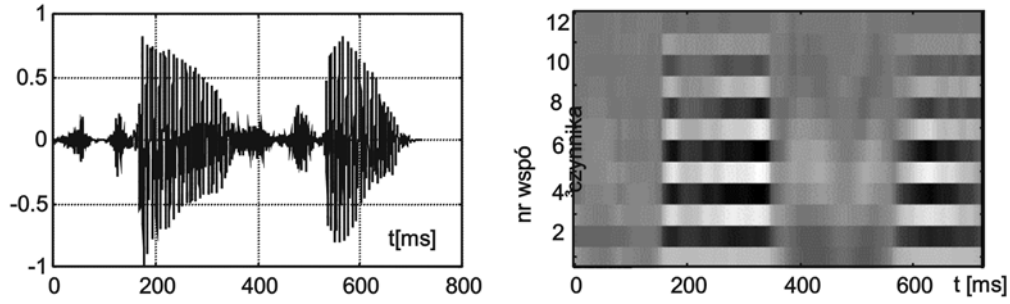
Rys. 1.1. Układ syntezy dźwięku

Analiza LPC ma w dziedzinie przetwarzania dźwięku dwa zasadnicze zastosowania. Po pierwsze jest wykorzystywana do kodowania i kompresji sygnału. Drugie, nie mniej ważne zastosowanie, znajduje metoda LPC w rozpoznawaniu mowy. Tutaj współczynniki są wykorzystywane do opisu ramek sygnału mowy.

W obu zastosowaniach istotnym jest, aby dokładność predykcji była jak największa. Można ją oceniać na podstawie wielkości sygnału szczątkowego (jego amplitudy i energii). Na rys. 1.2 przedstawiono fragment sygnału mowy (linia przerywana) i sygnał szczątkowy (linia ciągła). Na rysunku 1.3 przedstawiono przebiegi zmienności współczynników LPC słowa analizowanego metodą okien czasowych.



Rys. 1.2. Kodowanie LPC, p=9, sygnał wejściowy (linia przerywana), sygnał szczątkowy (linia ciągła), częstotliwość próbkowania 11kHz, kwantyzacja 8-bitowa



Rys. 1.3. Przebieg czasowy i odpowiadająca mu mapa współczynników LPC słowa „szczęście”

Poniżej przedstawiono opis sposobu wyznaczania współczynników predykcji dla spróbkowanego i skwantowanego sygnału dźwiękowego metodą najmniejszych kwadratów.

Wyznaczanie współczynników LPC metodą najmniejszych kwadratów

Wyznaczenia współczynników liniowego modelu predykcji można dokonywać kilkoma sposobami. Znane są metody: najmniejszych kwadratów, kowariancyjna, korelacyjna i Durbina. Przedstawiona w tym punkcie szeroko stosowana metoda najmniejszych kwadratów zapewnia najmniejszy błąd predykcji dla sygnału na podstawie, którego wyznaczane są współczynniki LPC.

Dany jest sygnał $\mathbf{u} = [u(1),u(2),...,u(U)]$. Oznaczmy: $\mathbf{u}_i = \begin{bmatrix} u(i) \\ ... \\ u(i+p-1) \end{bmatrix}$

Przy tak przyjętych oznaczeniach oszacowanie (1.1) przyjmuje postać:

$$z(k) = \mathbf{u}_\mathbf{k}^{'}\mathbf{a} \qquad k = 1,2,...,N-p \tag{1.5}$$

gdzie $\mathbf{a} = [a_1\ a_2\ ...\ a_p]'$. Utwórzmy z sygnału **a** macierz **U** następująco:

$$\mathbf{U} = \begin{bmatrix} u(1) & u(2) & ... & u(N-p) \\ u(2) & u(3) & ... & u(N-p+1) \\ ... & ... & ... & ... \\ u(p) & u(p+1) & ... & u(N-1) \end{bmatrix} \tag{1.6}$$

Wyrażając macierz **U** przy pomocy wektorów $\mathbf{u}_i$, mamy:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \mathbf{...} & \mathbf{u_{N-p}} \end{bmatrix} \tag{1.7}$$

Oznaczmy przez **x** wektor:

$$\mathbf{x} = \begin{bmatrix} u(p+1) & u(p+2) & ... & u(N) \end{bmatrix} \tag{1.8}$$

Stosując powyższe oznaczenia, układ równań (1.5) można zapisać jako:

$$\mathbf{z} = \mathbf{U}^{'}\mathbf{a} \tag{1.9}$$

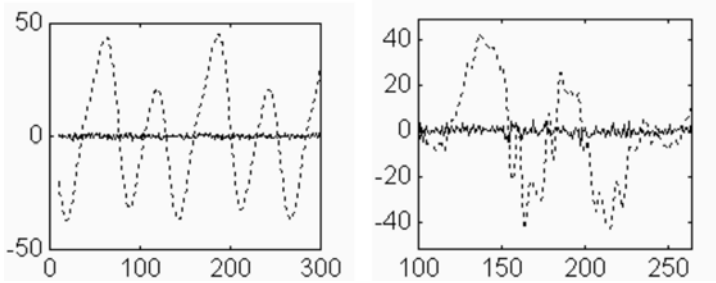
a sygnał szczątkowy:

$$\mathbf{w} = \mathbf{z} - \mathbf{x} = \mathbf{U}^{'}\mathbf{a} - \mathbf{x} \tag{1.10}$$

Wektor współczynników **a** należy wyznaczyć w taki sposób, aby energia sygnału szczątkowego $\|\mathbf{w}\|^2$ była minimalna. Współczynniki spełniające warunek minimalizacji błędu średniokwadratowego wyznacza się z następującego wzoru:

$$\mathbf{a} = \left(\mathbf{U}\mathbf{U}'\right)^{-1}\mathbf{U}\mathbf{x} \tag{1.11}$$

Aby uzyskać rozwiązanie, macierz $\mathbf{U}\mathbf{U}'$ musi być nieosobliwa.



Rys. 1.4. Przykładowe sygnały szczątkowe (linia ciągła) na tle sygnałów analizowanych (linia przerywana) wyznaczone metodą najmniejszych kwadratów, częstotliwość próbkowania 11kHz, kwantyzacja 8-bitowa

Określenie rzędu predykcji

Zdefiniujmy dwa wskaźniki określające stosunek sygnału szczątkowego do sygnału analizowanego. Pierwszym będzie stosunek energii sygnału szczątkowego **w** do energii sygnału oryginalnego **u**:

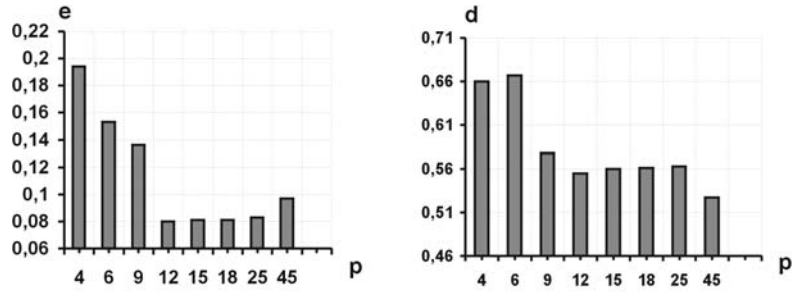
$$e = \frac{\sqrt{\sum_{i=p+1}^{N} w(i)^2}}{\sqrt{\sum_{i=p+1}^{N} u(i)^2}} \tag{1.12}$$

Jako drugi przyjmijmy stosunek zakresu zmienności sygnału szczątkowego **w** do zakresu zmienności sygnału oryginalnego **u**:

$$d = \frac{w_{\max} - w_{\min}}{u_{\max} - u_{\min}} \tag{1.13}$$

gdzie: $w_{\max} = \max_i |w(i)|$, $w_{\min} = \min_i |w(i)|$, $u_{\max}$ i $u_{\min}$ – analogicznie.

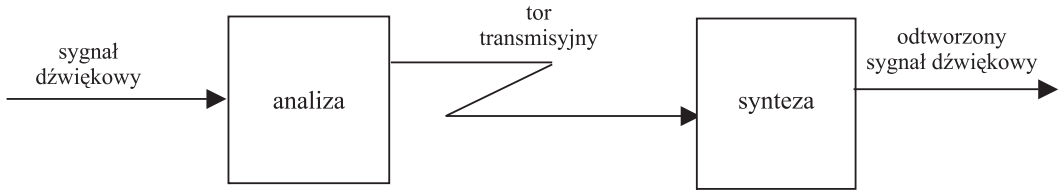
Na rys. 1.5 przedstawiono wykresy zmienności wskaźników e i d wyznaczonych przy zmiennym rzędzie predykcji. Na podstawie tego, jak i wielu innych eksperymentów, można stwierdzić, że 10-14 współczynników LPC dobrze opisuje sygnał i dalsze zwiększanie tej liczby przynosi niezauważalną poprawę jakości aproksymacji sygnału.



Rys. 1.5. Wykresy zależności wskaźników e oraz d w funkcji rzędu predykcji p

Standardy kodowania sygnału mowy oparte na LPC

Predykcja liniowa jest obecnie szeroko stosowana w cyfrowych układach transmisji. Idea metod kompresji stratnej opartych na LPC polega na zastąpieniu sygnału jego parametrycznym opisem. Ów parametryczny opis, jak już wspomniano, zawiera parametry filtru modelującego tor głosowy człowieka oraz informacje o sygnale pobudzenia (na etapie analizy nazywanym szczątkowym). Na rysunku 1.6 przedstawiono ogólny schemat działania układu transmisji wykorzystującego predykcję liniową. Układ taki składa się z: nadajnika, w którym przeprowadzana jest analiza sygnału (wyodrębnienie istotnych cech sygnału), kanału transmisyjnego oraz odbiornika, w którym następuje odtworzenie sygnału. Przejdźmy do krótkiego omówienia dwóch mających obecnie szerokie zastosowanie standardów kodowania wykorzystujących omawianą technikę: LPC-10 oraz CELP.



Rys. 1.6. Schemat działania układów transmisji wykorzystujących predykcję liniową

Standard LPC-10

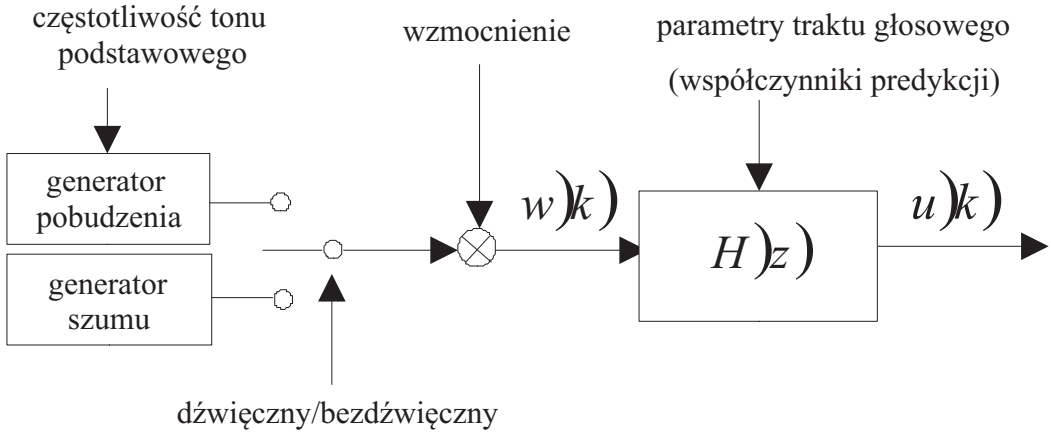
Pierwszy opis metody pojawił się w 1979 r., a za standard została uznana w 1984 r. pod nazwą „*US federal standard 1015*". Do transmisji mowy w tym standardzie wystarcza przepustowość 2,4 kb/s. Dyskretny sygnał dźwiękowy poddawany jest analizie w tzw. ramkach. Próbkowanie sygnału mowy odbywa się z częstotliwością 8 kHz. Każda ramka zawiera 180 próbek, (44,4 ramki na sekundę). Na podstawie

tej próby wyznacza się 10 współczynników predykcji dla mowy dźwięcznej, a 4 dla faz bezdźwięcznych. Współczynniki są wyznaczane jako tzw. współczynniki odbicia[2] (ang. *reflection coefficients*), z czego pierwsze dwa dla zmniejszenia amplitudy przedstawia się w postaci logarytmicznej tzw. LARs (ang. *Log-Area Rations*). Ponadto wyznacza się częstotliwośc tonu krtaniowego (dla ramek dźwięcznych) oraz współczynnik wzmocnienia. W tabeli 1.1 przedstawiono liczbę bitów potrzebnych do zakodowana poszczególnych elementów wektora parametrów transmitowanych w standardzie LPC-10.

Po stronie odbiornika następuje odtworzenie sygnału. Stosuje się dwa typy filtracji: długoterminową i krótkoterminową. W ramach filtracji krótkoterminowej (rys. 1.6) odbywa się rekonstrukcja pojedynczej ramki. Na wejście filtru syntezującego podawany jest sztucznie wygenerowany sygnał, pobudzenie (rys. 1.7), przeskalowany zgodnie z częstotliwością tonu podstawowego oraz wzmocnieniem dla danej ramki. W przypadku generowania faz bezdźwięcznych na wejście układu podawany jest szum. Filtracja długoterminowa pozwala na wygładzanie sygnału.

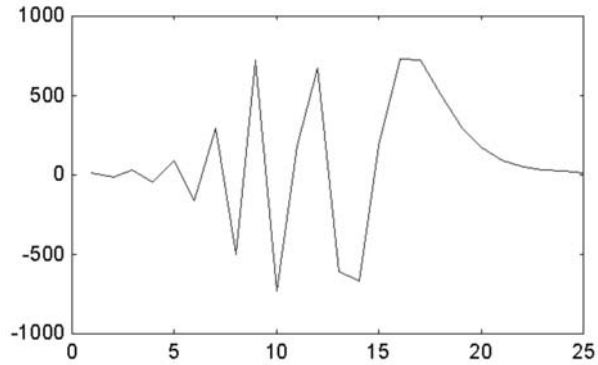
Tabela 1.1

Element	Liczba bitów
10 współczynników odbicia (4 w przypadku fazy bezdźwięcznej)	41
częstotliwość tonu podstawowego	7
dźwięczność	1
współczynnik wzmocnienia	5
Razem	54



Rys. 1.7. Schemat syntezy krótkoterminowej w standardzie LPC-10

[2] Współczynniki pośrednie wyznaczane w metodzie Durbina.



Rys. 1.8. Pobudzenie wykorzystywane do generowania mowy dźwięcznej w standardzie LPC-10

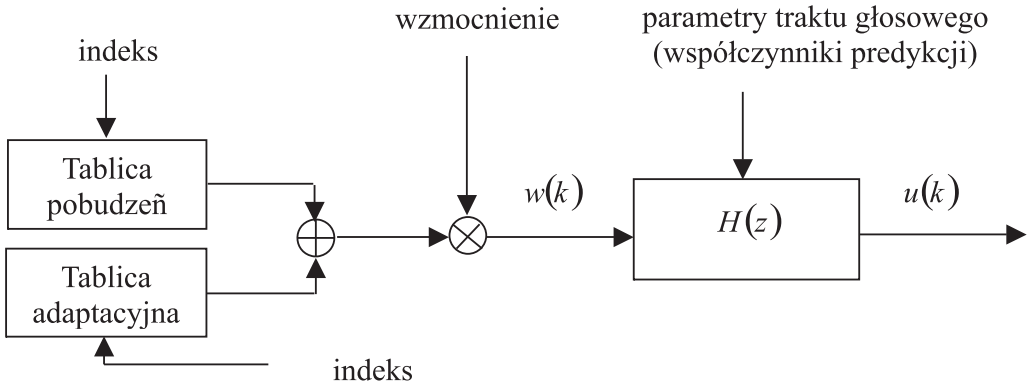
Standard CELP

Metoda CELP działa na podobnej zasadzie jak metoda LPC-10, z tą różnicą, że w standardzie CELP wykorzystuje się nie sztucznie generowany sygnał pobudzenia, lecz jeden z sygnałów zgromadzonych w banku (książce kodowej). Transmitowany jest jedynie kod sygnału najlepiej dopasowanego do danej ramki. Stąd jego nazwa *Code Excited Linear Prediction* Standard CELP został przyjęty w 1991 roku (figuruje pod nazwą „*US federal standard 1016"*). Wymagana przepustowość kanału transmisji dlatego standardu wynosi 4,8 kb/s. Na wejściu układu sygnał dźwiękowy jest próbkowany z częstotliwością 8kHz. Długość ramki wynosi 240 próbek (30ms). Część parametrów jest transmitowana częściej i wyróżnia się dodatkowo 4 krótsze ramki (ang. s*ubframe*) w ramach ramki 30 ms. Zarówno ten jak inne bardziej szczegółowe zabiegi w czasie analizy i syntezy sygnału mowy mają na celu jak najwierniejsze odtworzenie sygnału wejściowego podczas syntezy w odbiorniku. Schemat syntezy w ramach pojedyńczej ramki (filtracji krótkoterminowej) został przedstawiony na rys. 1.9. W tabeli 1.2 przedstawiono parametry podlegające transmisji w standardzie CELP w ramach jednej ramki oraz liczby bitów potrzebnych do ich zapisu.

Tabela 1.2

Element	Liczba bitów
współczynniki predykcji liniowej (10)	34
częstotliwość tonu podstawowego	28 (4x7)
współczynnik wzmocnienia	20 (4x5)
indeks do tablicy pobudzeń	36 (4x9)
indeks do tablicy adaptacyjnej	20 (4x5)

korekcja błędu	4
synchronizacja	1
bit ekspansji	1
Razem	144



Rys. 1.9. Schemat syntezy krótkoterminowej w standardzie CELP

Obydwa standardy: LPC-10 oraz CELP wykorzystywane są do transmisji sygnałów leżących w paśmie telefonicznym. Te oraz inne metody oparte na modelu LPC znalazły szerokie zastosowanie w telefonii komórkowej GSM.

2. TRANSFORMACJE ORTOGONALNE

W punkcie tym omówione zostaną wybrane transformacje ortogonalne mające podstawowe znaczenie w metodach kompresji stratnej sygnałów cyfrowych. Zalicza się do nich transformaty: DFT, DCT, Haara. Ostatnia ma duże znaczenia zwłaszcza w przetwarzaniu obrazów cyfrowych.

Przedmiotem rozważań będzie przekształcenie liniowe postaci:

$$\mathbf{y} = \mathbf{T}\mathbf{u} \tag{2.1}$$

gdzie **u** i **y** są sygnałami postaci: $\mathbf{u} = \begin{bmatrix} u_0 \\ u_1 \\ \dots \\ u_{N-1} \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \dots \\ y_{N-1} \end{bmatrix}$,

a macierz przekształcenia **T** jest macierzą nieosobliwą o wymiarach *N*x*N* postaci:

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_0^{'} \\ \mathbf{t}_1^{'} \\ \cdots \\ \mathbf{t}_{N-1}^{'} \end{bmatrix} \tag{2.2}$$

Przekształcenie (2.1) nie zmienia energii sygnału. Oznacza to, że energia sygnału przed i po transformacji jest taka sama, powinien, więc być spełniony warunek:

$$\mathbf{y}^{'}\mathbf{y} = \mathbf{u}^{'}\mathbf{u} \tag{2.3}$$

Warunek ten będzie spełniony, jeżeli macierz **T** będzie macierzą ortogonalną, tzn. taką, że $\mathbf{T}'\mathbf{T} =$ (**I** jest macierzą jednostkową), gdyż:

$$\mathbf{y}^{'}\mathbf{y} = \left[\mathbf{Tu}\right]\mathbf{Tu} = \mathbf{u}^{'}\mathbf{T}^{'}\mathbf{Tu} \tag{2.4}$$

Warunek ortogonalności dla macierzy **T** można zapisać następująco:

$$\mathbf{t}_i^{'}\mathbf{t}_j = \begin{cases} 1 \text{ dla } i = j \\ 0 \text{ dla } i \neq j \end{cases} \tag{2.5}$$

O wektorach $\mathbf{t}_i$ takich, że $\mathbf{t}_i'\mathbf{t}_i = 1$ mówimy, że są ortonormalne.

Z warunku $\mathbf{T}'\mathbf{T} = \mathbf{I}$ wynika także, iż $\mathbf{T}^{-1} = \mathbf{T}'$ (macierz odwrotna jest równa macierzy transponowanej). Upraszcza to wyznaczanie transformaty odwrotnej gdyż:

$$\mathbf{u} = \mathbf{T}^{'}\mathbf{y} \tag{2.6}$$

Podstawę transformacji ortogonalnych stanowi rozwinięcie w bazie funkcji ortogonalnych. Sposób wyznaczania macierzy **T** dla konkretnych transformat zależy od zastosowanej bazy funkcji ortogonalnych. Przekształcenia ortogonalne transformują sygnał do nowego układu współrzędnych, w którym to układzie energia sygnału rozkłada się nierównomiernie ze zdecydowaną przewagą początkowych współrzędnych. Stanowi to o dużym znaczeniu tego typu przekształceń w stratnej kompresji sygnałów.

Transformata DFT

Dyskretną transformatą Fouriera (DFT – *Discrete Fourier Transform*) nazywamy odwzorowanie sygnału (skończonego ciągu liczbowego)

$$\mathbf{u} = \begin{bmatrix} u_0 \\ u_1 \\ \cdots \\ u_{N-1} \end{bmatrix} \text{ w ciąg liczb zespolonych: } \mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \cdots \\ y_{N-1} \end{bmatrix},$$

zgodnie ze wzorem:

$$y_n = \sum_{k=0}^{N-1} u_k w_N^{-kn} \ , \ n = 0,1,2,...,N-1 \ \ , \ w_N = e^{j\frac{2\pi}{N}} , \tag{2.7}$$

a przekształcenie odwrotne (IDFT – *Inverse Discrete Fourier Transform)*:

$$u_k = \frac{1}{N} \sum_{n=0}^{N-1} y_n w_N^{kn} \ , \ \ k = 0,1,2,...,N-1 \ , \ w_N = e^{j\frac{2\pi}{N}} . \tag{2.8}$$

Przekształcenia DFT i IDFT można zapisać w postaci macierzowej:

$$\mathbf{y} = \mathbf{M} \cdot \mathbf{u} \tag{2.9}$$

$$\mathbf{u} = \mathbf{W} \cdot \mathbf{y} \tag{2.10}$$

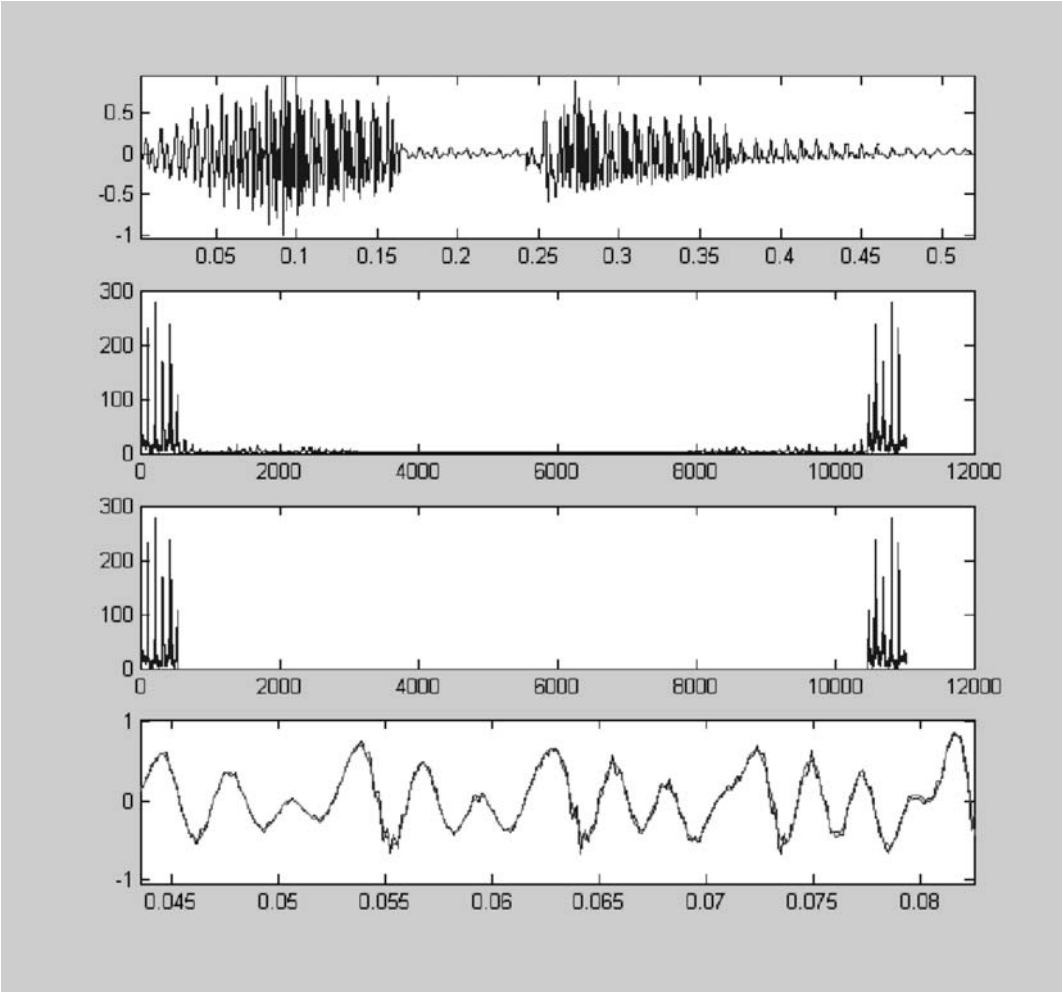
gdzie:

$$\mathbf{M} = \begin{bmatrix} w_N^{-0*0} & w_N^{-1*0} & \dots & w_N^{-(N-1)*0} \\ w_N^{-0*1} & w_N^{-1*1} & \dots & w_N^{-(N-1)*1} \\ \dots & \dots & \dots & \dots \\ w_N^{-0*1} & w_N^{-1*1} & \dots & w_N^{-(N-1)*(N-1)} \end{bmatrix}, \quad \mathbf{W} = \frac{1}{N} \begin{bmatrix} w_N^{0*0} & w_N^{1*0} & \dots & w_N^{(N-1)*0} \\ w_N^{0*1} & w_N^{1*1} & \dots & w_N^{(N-1)*1} \\ \dots & \dots & \dots & \dots \\ w_N^{0*1} & w_N^{1*1} & \dots & w_N^{(N-1)*(N-1)} \end{bmatrix}$$

Macierz **M** nie spełnia warunku ortogonalności gdyż jej wiersze są wektorami otrogonalnymi, ale nie ortonormalnymi. Warunek, jaki spełnia jest następujący: $\mathbf{M}'\mathbf{M} = N\mathbf{I}$. Aby przekształcenie DFT było ortogonalnym konieczne jest unormowanie macierzy postaci:

$$\mathbf{T} = \frac{1}{\sqrt{N}} \mathbf{M} \tag{2.11}$$

Na rys. 2.1 przedstawiono właściwości kompresyjne transformaty DFT. Dla sygnału dźwiękowego przedstawionego na pierwszym od góry przebiegu obliczono transformatę DFT (jej moduł przestawiony został na wykresie drugim od góry). Następnie z widma sygnału usunięto 95% końcowych próbek (o najmniejszej amplitudzie, wykres trzeci od góry) oraz wykonana została transformata odwrotna IDFT. Wynik tej operacji przedstawiony został na wykresie dolnym, gdzie na tle fragmentu sygnału pierwotnego (linia niebieska) przedstawiono przebieg sygnału odtworzonego z obcięcia transformaty DFT. Przebieg odtworzony dość dobrze aproksymuje sygnał pierwotny.

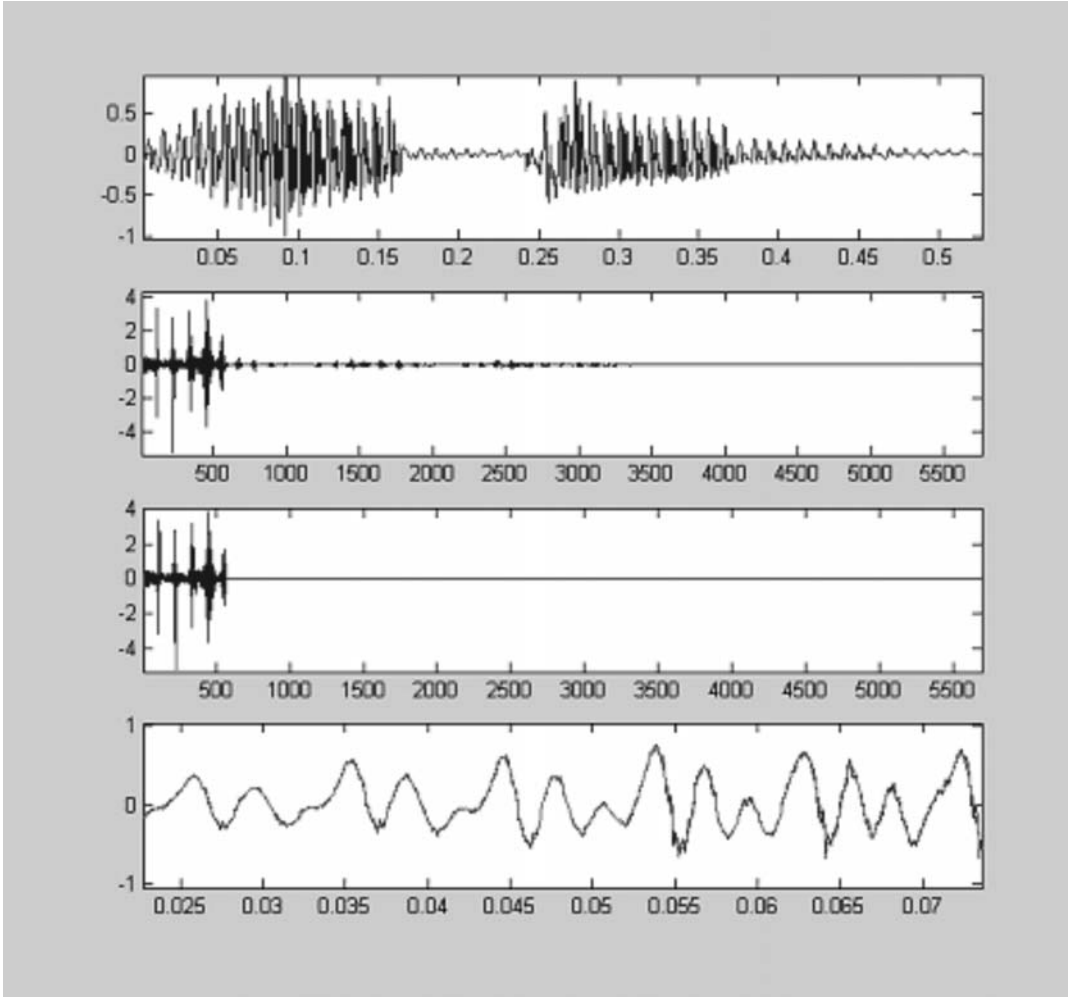


Rys. 2.1. Transformata DFT sygnału dźwiękowego (słowo jeden). Na kolejnych od góry wykresach przedstawiono: przebieg czasowy sygnału, widmo amplitudowe (moduł transformaty DFT), widmo amplitudowe po usunięciu 95% próbek transformaty, fragment przebiegu z wykresu pierwszego (linia niebieska) oraz sygnał odtworzony z widma przedstawionego na wykresie trzecim (linia czerwona)

Transformata DCT

Dyskretna transformata kosinusowa (DCT – *Discrete Cosine Transform*) wykorzystuje rozwinięcie sygnału w bazie funkcji ortogonalnych zbudowanej z wielomianów Czebyszewa.

Macierz przekształcenia kosinusowego jest tworzona na drodze dyskretyzacji wielomianów Czebyszewa.



Rys. 2.1. Transformata DCT sygnału dźwiękowego (słowo „jeden"). Na kolejnych od góry wykresach przedstawiono: przebieg czasowy sygnału, transformatę DCT sygnału, transformatę DCT po usunięciu 90% końcowych jej próbek, fragment przebiegu z wykresu pierwszego (linia niebieska) oraz sygnał odtworzony z obciętej transformaty przedstawionej na wykresie trzecim (linia czerwona)

Postać unormowanej macierzy przekształcenia kosinusowego jest następująca:

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}'_0 \\ \mathbf{t}'_1 \\ \dots \\ \mathbf{t}'_{N-1} \end{bmatrix}, \text{ gdzie}$$

$$\mathbf{t}'_k = \begin{cases} \frac{1}{\sqrt{N}}\left[\cos\left(\frac{k(2\cdot 0+1)\pi}{2N}\right), \cos\left(\frac{k(2\cdot 1+1)\pi}{2N}\right), ..., \cos\left(\frac{k(2\cdot(N-1)+1)\pi}{2N}\right)\right] & \text{dla } k=0 \\ \sqrt{\frac{2}{N}}\left[\cos\left(\frac{k(2\cdot 0+1)\pi}{2N}\right), \cos\left(\frac{k(2\cdot 1+1)\pi}{2N}\right), ..., \cos\left(\frac{k(2\cdot(N-1)+1)\pi}{2N}\right)\right] & \text{dla } k\neq 0 \end{cases} \tag{2.12}$$

Zatem przekształcenie kosinusowe można zapisać w postaci macierzowej:

$$\mathbf{y} = \mathbf{Tu} \tag{2.13}$$

lub po rozpisaniu :

$$y_k = \sqrt{\frac{2}{N}} \sum_{m=0}^{N-1} u_m \cos\frac{k(2m+1)\pi}{2N}, \quad k = 1,2,..., N-1 \tag{2.14}$$

$$y_k = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} u_m, \quad k = 0 \tag{2.15}$$

Na rys. 2.2 przedstawiono właściwości kompresyjne transformaty DCT w sposób analogiczny jak w przypadku DFT. Dla sygnału dźwiękowego przedstawionego na pierwszym od góry przebiegu obliczono transformatę DCT (wykres drugi od góry). Następnie z transformaty sygnału usunięto 90% końcowych próbek (o najmniejszej amplitudzie, wykres trzeci od góry) oraz wykonana została transformata odwrotna IDCT. Wynik tej operacji przedstawiony został na wykresie dolnym, gdzie na tle fragmentu sygnału pierwotnego (linia niebieska) przedstawiono przebieg sygnału odtworzonego z obcięcia transformaty DCT. Przebieg odtworzony również dość dobrze aproksymuje sygnał pierwotny.

Transformata Haara

Transformata Haara jest przekształceniem realizującym rozwinięcie sygnału w bazie ortogonalnych funkcji mających postać impulsów prostokątnych. Funkcje te przedstawia się w postaci ciągu indeksowanego parą liczb. Wartość funkcji o indeksie (*r,m*), gdzie $r \geq 0$, $1 \leq m \leq 2^r$ w punkcie $t \in R^1$ jest oznaczana: haar (*r,m,t)*.

Funkcje Haara wyznacza się z zależności (2.16) i (2.17):

$$\text{haar}(0,0,t) = 1, \qquad t \in [0,1) \tag{2.16}$$

$$\text{haar}(r,m,t)=\begin{cases} 2^{r/2}, & \text{dla } \dfrac{m-1}{2^r} \le t < \dfrac{m-\frac{1}{2}}{2^r} \\ -2^{r/2}, & \text{dla } \dfrac{m-\frac{1}{2}}{2^r} \le t < \dfrac{m}{2^r} \\ 0, & \text{poza tym} \end{cases} \tag{2.17}$$

Przekształcenie Haara definiowane jest następująco:

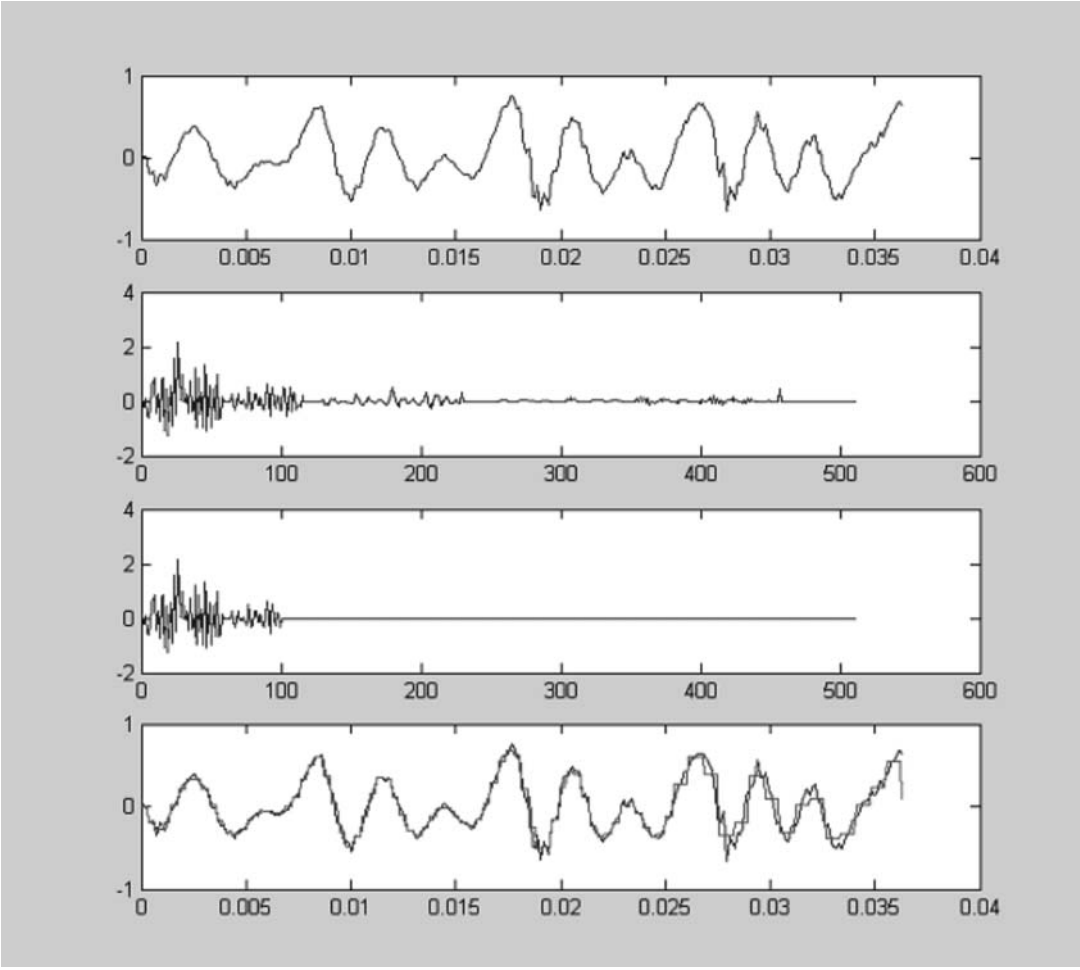
$$\mathbf{y} = \frac{1}{\sqrt{N}}\mathbf{H}(n)\mathbf{u} \tag{2.18}$$

a odwrotne wzorem:

$$\mathbf{u} = \frac{1}{\sqrt{N}}\mathbf{H}(n)'\mathbf{y} \tag{2.19}$$

Macierz $\mathbf{H}(n)$ jest macierzą NxN otrzymaną na drodze dyskretyzacji funkcji Haara. Zależności (2.18) i (2.19) wskazują na to, iż macierz transformaty, aby spełniała warunek ortogonalności, musi zostać unormowana czynnikiem $\frac{1}{\sqrt{N}}$ (podobnie jak w przypadku transformat DFT i DCT). Wymiar macierzy transformacji Haara musi być potęgą liczby 2 tak, aby $2^n = N$.

Na rys. 2.3 przedstawione zostały własności kompresyjne transformaty Haara na przykładzie sygnału dźwiękowego w sposób analogiczny do przykładów dla DFT i DCT (patrz opis rysunku).



Rys. 2.3. Transformata Haara sygnału dźwiękowego (fragment słowa „jeden"). Na kolejnych od góry wykresach przedstawiono: przebieg czasowy sygnału, transformatę Haara sygnału, transformatę Haara po usunięciu 80% końcowych jej próbek, sygnał odtworzony z obciętej transformaty przedstawionej na wykresie trzecim (linia czerwona) na tle oryginału (linia niebieska)

3. MODEL PSYCHOAKUSTYCZNY SŁUCHU CZŁOWIEKA – KODOWANIE PERCEPTUALNE

Drugim po transformatach ortogonalnych, lecz równie ważnym w kompresji stratnej sygnału dźwiękowego jest model psychoakustyczny narządu słuchu człowieka. Znajomość pewnych właściwości słuchu pozwala na oszczędniejsze kodowanie niektórych tonów lub zupełne ich pomijanie w zakodowanym strumieniu wyjściowym. Kodowanie z uwzględnieniem modelu psychoakustycznego nazywane jest kodowaniem perceptualnym.

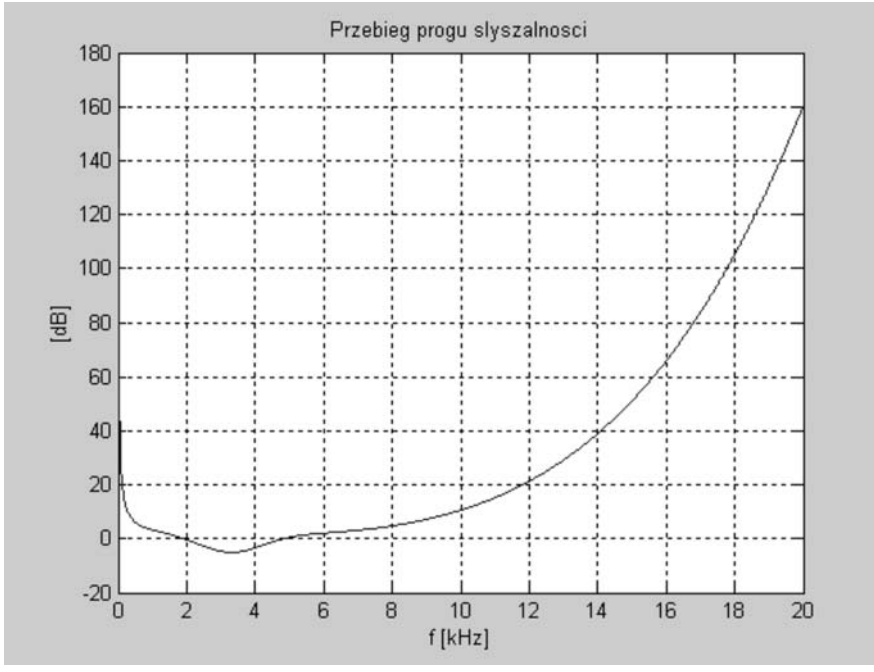
Na psychoakustyczny model słuch składają się trzy podstawowe zagadnienia: przebieg funkcji absolutnego progu słyszenia w funkcji częstotliwości, pasmowa analiza dźwięku realizowana przez zmysł słuchu oraz tak zwane maskowanie tonów przejawiające się tym, że tony o dużym natężeniu potrafią zagłuszyć całkowicie zaraz po nich występujące (lub przed) tony o niższym natężeniu. W dalszej części tego punktu zostaną krótko omówione poszczególne zagadnienia.

Absolutny próg słyszenia

Na rys. 3.1 przedstawiony został przebieg absolutnego progu słyszalności w funkcji częstotliwości wg modelu Terharda:

$$T = 3{,}64 f^{-0{,}8} - 6{,}5 e^{-0{,}6(f-3{,}3)^2} + 10^{-3} f^4 \tag{3.1}$$

Z przebiegu wynika, że dźwięki o tym samym natężeniu w zależności od ich częstotliwości mogą być słyszalne bardzo dobrze lub wcale. Wykorzystanie przebiegu absolutnego progu słyszalności w procesie kodowania dźwięku polegać będzie na pominięciu tych składowych harmonicznych widma, dla których natężenie jest mniejsze od wartości progu słyszalności.



Rys. 3.1. Wykres funkcji absolutnego progu słyszalności wg modelu Terharda

Pasma krytyczne

Wyniki badań psychoakustycznych wykazały również, że system słuchowy człowieka przetwarza dźwięki w pewnych podpasmach, zwanych *pasmami krytycznymi*. Każdemu pasmu krytycznemu odpowiada odcinek na błonie podstawowej ślimaka (ok. 1,3 mm). Oznacza to, że system słuchowy może być modelowany jako zestaw filtrów pasmowoprzepustowych o szerokościach równych szerokościom odpowiednich pasm krytycznych. Szerokości poszczególnych pasm nie są jednakowe. Są stałe do częstotliwości 500Hz i wynoszą ok.100Hz, następnie ich szerokość wzrasta o 20% w stosunku do poprzedniego pasma krytycznego.

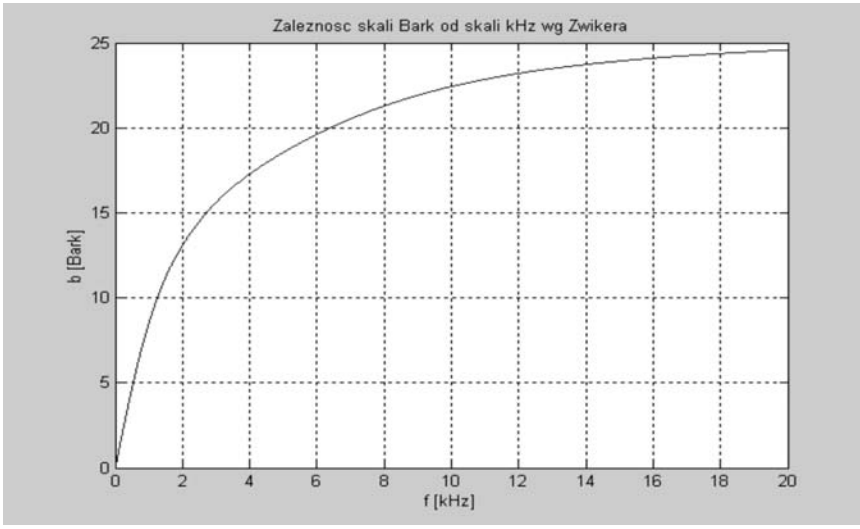
Skalę, na której odłożone są liniowo numery kolejnych pasm krytycznych nazywa się skalą w Barkach, 1 Bark – jedno pasmo krytyczne. Oszacowania szerokości pasma krytycznego można dokonać posługując się wzorem:

$$\Delta f = 25 + 75\left(1 + 1{,}4 f^2\right)^{0{,}69} \tag{3.2}$$

gdzie f jest częstotliwością środkową pasma.

Na rys. 3.2 przedstawiona została zależność skali w Barkach od skali w Hz wg modelu Zwikera:

$$Bark = 13 arctg\left(0{,}76 f\right) + 3{,}5 arctg\left[\left(\frac{f}{7{,}5}\right)^2\right] \tag{3.3}$$

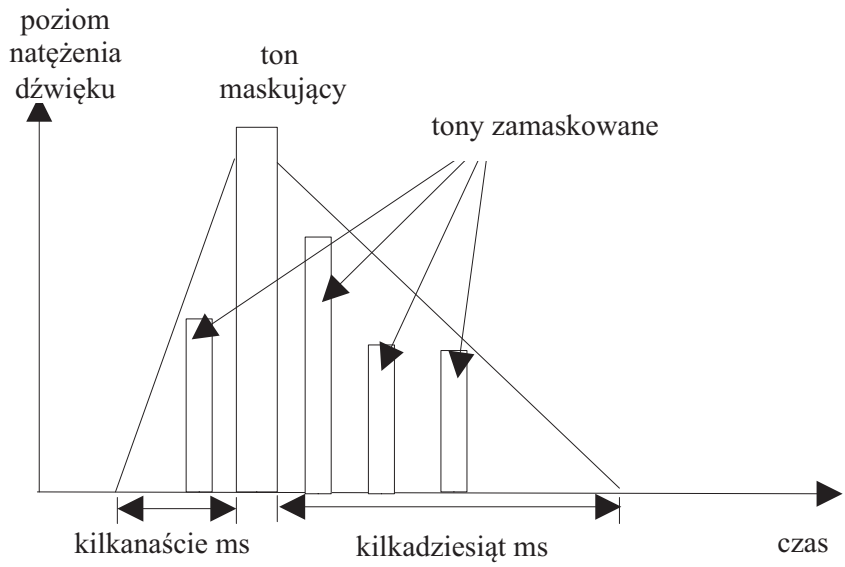


Rys. 3.2. Zależność częstotliwości w skali Bark od częstotliwości w skali Hz

Wykorzystanie wiedzy o analizie podpasmowej realizowanej przez ucho człowieka skutkuje możliwością dokonywania redukcji informacji w widmie jedynie do opisu poszczególnych pasm krytycznych np. średnią ważoną amplitud harmonicznych wchodzących w skład pasma.

Maskowanie tonów

Narząd ludzkiego słuchu zachowuje się jak równoległy analizator widma o ograniczonej rozdzielczości widmowej i czasowej (niejednorodna podatność błony podstawnej i ograniczona liczba komórek nerwowych narządu Cortiego). Wynikiem tego jest zjawisko tzw. maskowania dźwięków przez tony głośne i to dźwięków występujących zarówno przed (premaskowanie) jak i po tonie głośnym (postmaskowanie). Zjawisko maskowania dźwięków w sposób schematyczny przedstawione zostało na rys. 3.3. W rzeczywistości przebieg progu słyszenia w obrębie tonu maskującego jest funkcją nieliniową jednakże w systemach kompresji, ze względu na szybkość obliczeń, stosuje się aproksymację liniową. Należy również pamiętać o tym, że w zależności od częstotliwości tonów maskowanych wartości progu słyszenia będą różne (krzywa absolutnego progu słyszalności).



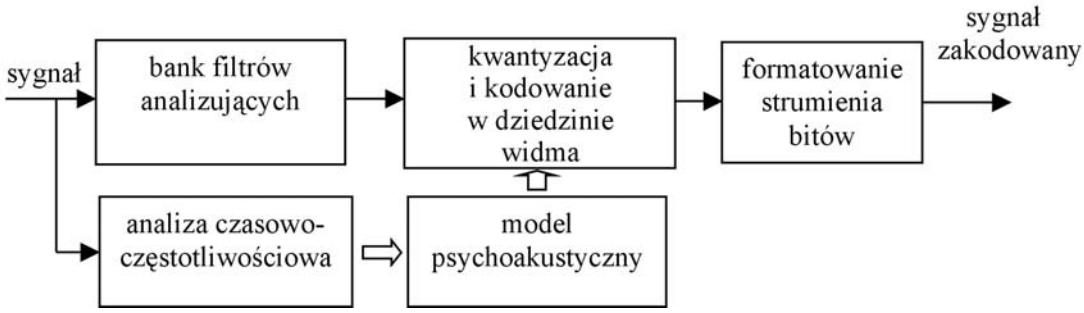
Rys. 3.3. Maskowanie tonów

Uwzględniając, zatem zjawisko maskowania tonów w procesie kompresji stratnej można w strumieniu wyjściowym pomijać tony zamaskowane.

Ogólny schemat kompresji stratnej dźwięku

W formie podsumowania zagadnień transformat ortogonalnych oraz modelu psychoakustycznego, na rys. 3.4 przedstawiony został ogólny schemat kodera stratnego sygnału dźwiękowego wysokiej jakości.

Sygnał dźwiękowy poddawany jest równolegle analizie pasmowej oraz analizie czasowo-częstotliwościowej. W bloku analizy pasmowej wyznaczane są widma dla poszczególnych pasm. Na podstawie spektrogramów z analizy czasowo-częstotliwościowe wyznaczany jest model psychoakustyczny (wartości progu słyszalności). Następnie odbywa się kwantyzacja widma z uwzględnieniem modelu psychoakustycznego oraz kodowanie i formowanie strumienia wyjściowego.



Rys. 3.4. Ogólny schemat kodera stratnego dźwięku

Literatura

[1] Basztura Cz. i inni, Metody parametryzacji sygnału mowy do automatycznego rozpoznawania głosów. Prace Naukowe ITiA Politechniki Wrocławskiej, nr 31, 1990.
[2] Heines R., Cyfrowe przetwarzanie dźwięku, Mikom, Warszawa 2002.
[3] Czyżewski A., Dźwięk cyfrowy, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2001.
[4] Grad L., Badanie możliwości rozpoznawania mówcy na podstawie reprezentacji LPC sygnału mowy. Biuletyn IAiR nr 13, 2000.
[5] Grad L., Badania porównawcze zastosowania liniowego i nieliniowego modelu predykcji w analizie sygnału mowy. Biuletyn IAiR nr 10, 1999.
[6] Kwiatkowski W., Wstęp do cyfrowego przetwarzania sygnałów, Instytut Automatyki i Robotyki WCY WAT, Warszawa 2003.
[7] Wiśniewski A. M., Analiza pasmowa sygnałów mowy. Biuletyn IAiR WAT, nr 8, 1997.