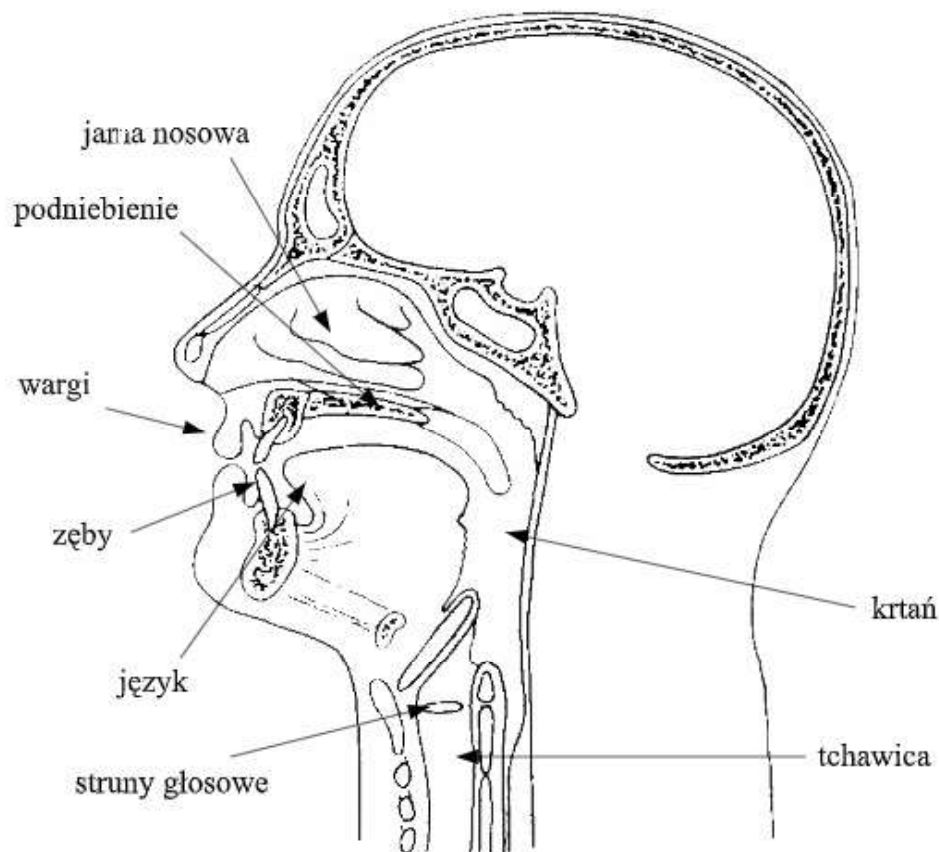


Laboratorium 8

1. Liniowe Kodowanie Predycyjne (LPC) i zastosowania

Tor oddechowy kształtuje widmo sygnału krtaniowego powstającego przy przechodzeniu fali dźwiękowej przez struny głosowe (Rys. 1). Dokonując estymacji parametrów opisujących trakt głosowy jako układ filtrujący sygnał pobudzenia, można wyznaczyć jego charakterystykę amplitudowo–częstotliwościową.



Rys. 1. Schemat traktu głosowego

Liniowe kodowanie predycyjne (LPC, *ang. Linear Predictive Coding*), zastosowali po raz pierwszy Saito i Itakura w 1966 roku. Model ten, oparty jest na analizie budowy narządu mowy człowieka. Transmitancja filtru predykcji liniowej jest wyrażona następującym wzorem:

$$H(z) = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (1)$$

Przetwarzanie sygnałów cyfrowych

gdzie p jest liczbą współczynników, a_k jest współczynnikiem filtra predykcji liniowej o indeksie $k = 1, \dots, p$.

Otrzymane moduły współczynników filtra predykcji liniowej $|a_1|, \dots, |a_p|$ są **cechami** sygnału dla metody LPC. Liniowe kodowanie predycyjne oparte jest na kombinacji liniowej p próbek. Na podstawie p poprzednich wartości sygnału akustycznego $s(n)$ przewiduje się wartości kolejne:

$$s'(n) = -\sum_{k=1}^p a_k \cdot s(n-k) \quad (2)$$

gdzie $s'(n)$ jest prognozą wartości sygnału w chwili n -tej.

Błąd między bieżącą próbką i przewidywaną można wyrazić jako:

$$err(n) = s(n) - s'(n) \quad (3)$$

Przyjmując, że N jest liczbą próbek w oknie, należy znaleźć optymalne wartości współczynników a_1, \dots, a_p takie, aby średni błąd prognozy był jak najmniejszy. Najczęściej jako kryterium jakości predykcji wybiera się błąd średniokwadratowy, zdefiniowany następująco:

$$J = \sigma^2 = \frac{1}{N-p} \sum_{n=p}^{N-1} err^2(n) = \frac{1}{N-p} \sum_{n=p}^{N-1} \left[s(n) + \sum_{j=1}^p a_j s(n-j) \right]^2 \quad (4)$$

Aby wyznaczyć optymalne współczynniki a_1, \dots, a_p , należy obliczyć pochodną cząstkową J względem każdego z tych współczynników i przyrównać ją do zera. Otrzymujemy w ten sposób układ p równań z p niewiadomymi, który posiada następujące rozwiązanie:

$$\mathbf{a} = -\mathbf{R}^{-1} \mathbf{r} \quad (5)$$

gdzie

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix}, \mathbf{r} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{bmatrix} \quad (6)$$

$$r(k) = \frac{1}{N-p} \sum_{n=p}^{N-1} s(n)s(n+k) \quad (7)$$

Znając współczynniki predykcji, można z równania (4) obliczyć jej błąd $J_{min} = \sigma_{min}^2$:

$$J_{\min} = \sigma_{\min}^2 = r(0) + a^T r = r(0) + \sum_{j=1}^p a_j r(j) \quad (8)$$

Obliczenie współczynników filtra predykcji liniowej ze wzoru (5) wymaga wyznaczenia macierzy odwrotnej \mathbf{R}^{-1} dla każdej „ramki” sygnału akustycznego. Rozwiązaniem tego zadania jest zastosowanie algorytmu Durбина-Levinsona opisanego w literaturze. Algorytm ten jest skuteczny, ponieważ potrzebuje tylko M^2 mnożeń, aby wyliczyć współczynniki filtra predykcji liniowej a_1, \dots, a_p (złożoność obliczeniowa wynosi $O(N^2)$). Współczynniki te tworzą **wektor cech** $[|a_1|, |a_2|, \dots, |a_p|]$, który będzie używany w etapie klasyfikacji (zobacz wzór 1). W literaturze autorzy stosują liczby współczynników filtra predykcji liniowej od 8 do 20.

Metody z rodziny liniowego kodowania predykcyjnego – LPC. Polegają na podziale cyfrowego sygnału mowy na krótkie segmenty, które są parametryzowane. Kompresja metodą LPC polega na reprezentacji i transmisji sygnału oryginalnego w postaci współczynników filtra analizy $|a_1|, \dots, |a_p|$. Rekonstrukcja sygnału wykonywana jest przy pomocy filtra syntezy mowy o charakterystyce odwrotnej do filtra analizy. W literaturze metoda LPC często określa się mianem parametrycznej reprezentacji mowy, gdzie filtr modeluje właściwości narządu mowy, który jest pobudzany prostym sygnałem syntetycznym.

Metoda LPC jest zwykle używana do analizy mowy i resyntezy. Jest stosowana jako forma kompresji głosu przez firmy telekomunikacyjne, na przykład w standardzie GSM. Jest również używana do bezpiecznego połączenia bezprzewodowego, w którym głos musi być zaszyfrowany i wysłany przez wąski kanał. Metody LPC można również użyć dla ekstrakcji cech.

Większość źródłowych kodeków mowy opartych jest na liniowym kodowaniu predykcyjnym (LPC – Linear Predictive Coding). Algorytmy kodeków źródłowych wykorzystują model wytwarzania sygnału mowy przez człowieka. Sygnał mowy powstaje przez przefiltrowanie pobudzenia przez trakt głosowy. W dostatecznie krótkich przedziałach czasu sygnał mowy może być traktowany jako sygnał stacjonarny, zatem możliwe jest wyznaczenie parametrów filtra cyfrowego, odpowiadającego transmitancji traktu głosowego. Wtedy zamiast przysyłać próbki sygnału mowy, wystarczy przesłać współczynniki filtra, wraz z dodatkowymi informacjami niezbędnymi do odtworzenia sygnału w dekodерze. Operację taką należy wykonać dla każdej z ramek czasowych sygnału. Algorytm LPC wykorzystywany jest do takiego doboru współczynników filtra, aby uzyskać transmitancję filtra najlepiej dopasowaną do transmitancji traktu głosowego. Aby zminimalizować błąd predykcji, najczęściej stosowany jest algorytm Levinsona-Durбина.

- Analiza w dziedzinie czasu, odwzorowująca rezonansową strukturę traktu głosowego
- Ramkowanie, okienkowanie i autokorelacja między ramkami sygnału wejściowego
- Aproksymacja każdej kolejnej próbki jako liniowa kombinacja N poprzednich próbek

$$\hat{s}[n] = \sum_{k=1}^p a_k s(n-k)$$