

Instrukcja do laboratorium 6

Fonoskopia

Analiza cech sygnału mowy pod kątem fonoskopii. F0.

1.Wprowadzenie

Dźwięk wytworzony przez głośnie nazywamy *tonem krtaniowym* lub *częstotliwością podstawową (tonem podstawowym)* i wyrażony jest wzorem:

$$F_0 = \frac{1}{2\pi} \sqrt{\frac{s}{m}} \quad (1)$$

gdzie: m – masa drgających strun [kg], s – współczynnik sprężystości strun [N/m]

Częstotliwość tonu krtaniowego, dla populacji polskiej, w zależności od rodzaju głosu mówcy zawierają się w następujących zakresach (Tab.1).

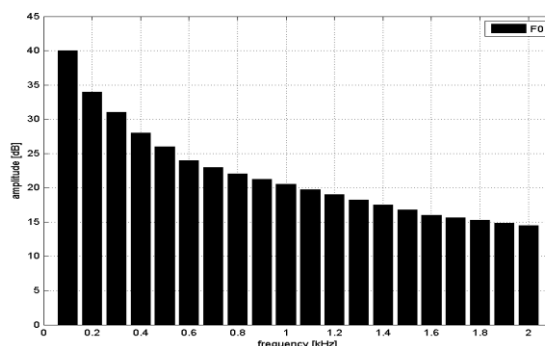
Tabela 1. Zakres tonu podstawowego (krtaniowego)

Nazwa głosu	Zakres częstotliwości [Hz]
bas	80 ÷ 320
baryton	100 ÷ 400
tenor	120 ÷ 480
alt	160 ÷ 640
mezzosopran	200 ÷ 800
sopran	240 ÷ 960

Powstałe fale ulegają następnie przekształceniu (wzmacnianiu) w układzie rezonansowym tj. gardle, jamie ustnej i nosowej. Wyemitowany sygnał akustyczny mowy w postaci ciśnienia akustycznego $p(t)$ można opisać analitycznie w dziedzinie czasu jako funkcję splotu przebiegu czasowego sygnału źródła $p_g(t)$ i odpowiedzi impulsowej kanału głosowego $h(t)$:

$$p(t) = \int_0^T h(t - \tau) p_g(\tau) d\tau \quad (2)$$

W przypadku artykulacji fonemów generowanych w źródle krtaniowym (pobudzenie dla głosek dźwięcznych) charakter widmowy fali głosowej związany jest z częstotliwością tonu podstawowego F_0 drgań fałdów głosowych człowieka. Modelowe widmo tonu podstawowego charakteryzuje się malejącą wykładniczo amplitudą tak, że wyższe harmoniczne są tłumione z nachyleniem około - 12 dB/oktawę, ale mimo to wyraźnie widoczne są harmoniczne o częstotliwości trzydziestokrotnie wyższej od częstotliwości podstawowej (rys.1).



Rys.1. Widmo tonu krtaniowego

2. Metody wyznaczania F0

Dzięki rozwojowi profesjonalnych metod rejestracji i przetwarzania sygnałów akustycznych, powstało szereg algorytmów wyznaczania tego tonu. W akustycznym sygnale mowy $p(t)$ właściwości źródła (w tym przypadku głośni) i właściwości kształtującego dźwięk kanału głosowego są ze sobą ściśle powiązane. Zatem zasadniczym problemem w wyznaczeniu parametrów źródła jest brak bezpośredniego dostępu do sygnału akustycznego, wytwarzanego przez to źródło.

Akustyczne metody wyznaczania tonu podstawowego:

- analiza przejść przez zero (*Zero Crossing Measure*),
- funkcja autokorelacji,
- analiza cepstralna,
- szacowanie okresu funkcji metodą opartą o filtr skończonej odpowiedzi impulsowej (*FIR - Finite Impulse Response*),
- algorytm wyznaczania stosunków częstotliwości harmoniczných do podharmoniczných (*sub-harmonic to harmonic ratio*)
- spektralna analiza wyższych rzędów (*Higher Order Spectra Analysis*)

2.1 Analiza przejść przez zero

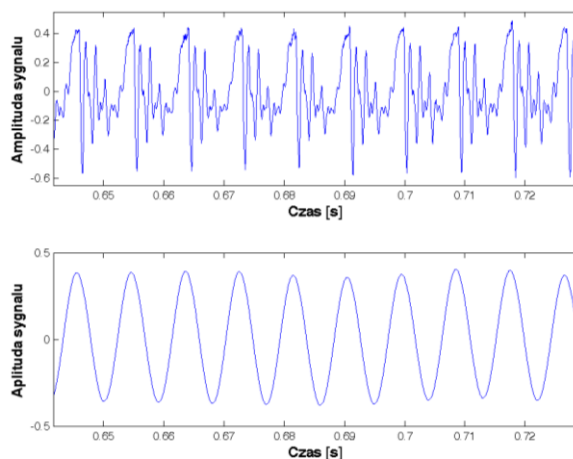
Analiza przejść przez zero jest wykonywana na sygnale $p(t)$ w dziedzinie czasu. Polega na wyznaczeniu punktów przecięcia się tego sygnału z osią czasu t , czyli takich punktów t_j , dla których $p(t)=0$. W praktyce, z punktu widzenia cyfrowego przetwarzania sygnału opiera się ona na wyznaczeniu wartości funkcji *signum* kolejno po sobie występujących próbek badanego sygnału - $p(i)$, zgodnie z zależnością:

$$\rho(i) = \frac{|sgn(p(i) - sgn(p(i-1)))|}{2}$$

gdzie:

$$sgn(i) = \begin{cases} +1 & \text{gdy } p(i) \geq 0 \\ -1 & \text{gdy } p(i) < 0 \end{cases}$$

W celu wyznaczenia tonu krztaniowego F_0 , dokonujemy na czasowym sygnale mowy $p(i)$ filtracji dolnoprzepustowej (z odpowiednio dobraną częstotliwością graniczną filtra), a następnie wykonujemy analizę zgodnie z założeniem równania uzyskując funkcję gęstości przejść przez zero $\rho(i)$. Znajomość statystycznych właściwości rozkładów interwałów czasowych między kolejnymi przejściami przejść przez zero pozwala wyznaczyć średnią częstotliwość tonu krztaniowego wraz z odchyleniem standardowym oraz jej wartości z okresu na okres.



Rys. 2. Czasowy przebieg sygnału mowy $p(t)$ – dźwięk /a/ o przedłużonej fonacji przed i po filtracji dolnoprzepustowej. Głos męski.

Na rysunku 2 pokazano czasowy sygnał mowy $p(i)$. W górnej części tego rysunku pokazano „surowy” sygnał mowy $p(i)$, natomiast w dolnej jego części ten sam sygnał mowy po zastosowaniu filtracji dolnoprzepustowej. Znajomość statystycznych właściwości tych rozkładów interwałów w czasie pozwala wyznaczyć średnią częstotliwość tonu krztaniowego oraz jej wartości z cyklu na cykl.

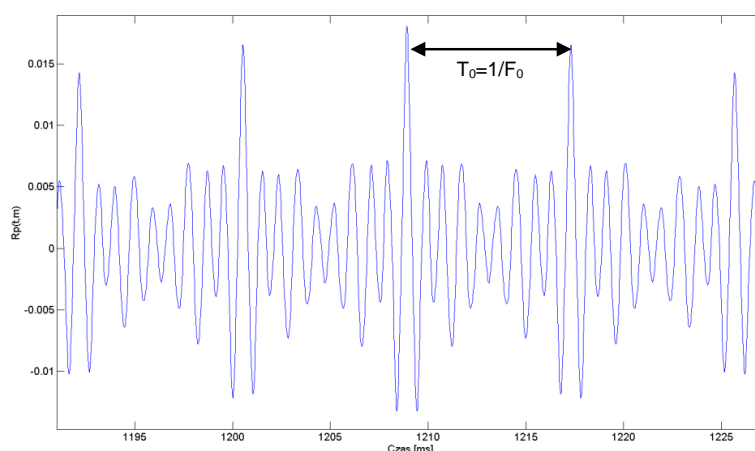
2.2 Funkcja autokorelacji

Popularną metodą estymacji F_0 w dziedzinie czasu jest metoda oparta o funkcję korelacji własnej i dla dowolnego sygnału definiowana jest zgodnie z zależnością:

$$R_{xx}(k) = \frac{1}{N - |k|} \sum_{n=0}^{N-1-|k|} x_n \cdot x_{n-k}^*, \quad -N+1 \leq k \leq N-1$$

Funkcja autokorelacji jest wykorzystywana do badania „powtarzalności” (okresowości) sygnału mowy, ponieważ przyjmuje ona wartości maksymalne dla wartości przesunięcia n równego wielokrotności okresu. Największe maksimum funkcji występuje dla opóźnienia zerowego ($n = 0$). Położenie następnego maksimum daje estymatę okresu, a jego wysokość wskazuje okresowość sygnału. Na rysunku 3 pokazano funkcję autokorelacji sygnału mowy, samogłoska „a” o przedłużonej fonacji, tej samej osoby co na rysunku 2.

Zarówno w funkcji korelacji wzajemnej, jak i autokorelacji przed znakiem sumy stosuje się czynnik normalizacyjny $1/N$ (estymator obciążony) lub $1/(N-k)$ (estymator nieobciążony).



Rys. 3. Funkcja korelacji własnej sygnału mowy. Dźwięk /a/ o przedłużonej fonacji. Głos męski.

3. Zadania do wykonania

- Dla wybranej metody: algorytm przejść przez zero lub algorytm autokorelacji wyznaczyć częstotliwość podstawową F_0 samogłosek o przedłużonej fonacji (/aaa/, /eee/, /iii/, /uuu/) dla sygnału akustycznego oraz sygnału EGG. Porównać otrzymane wyniki w ramach jednej samogłoski.
- Sprawdzić otrzymane wyniki realizując punkt a) przy pomocy funkcji wbudowanej **pitch**
- zakwalifikować mówcę do grupy głosu zgodnie z Tab. 1.