Instrukcja do laboratorium 7
Fonoskopia

Analiza cech sygnału mowy pod kątem fonoskopii. Formanty

1.Fonemy (głoski)

Fon (dźwięk) jest to najmniejszy element w sygnale mowy. Może mieć różne implementacje określone przez: ton, czas trwania, intonacje. Dwa dźwięki (fony) należą do tej samej kategorii, zwanej fonemem, jeśli są wystarczająco bliskie. Liczba fonemów waha się między 20 a 60 w zależności od języka i dialektu.
Do pracy ze spektrogramami dźwięków podział polskich fonemów dokonuje się na 7 grup (w nadgrupach samogłosek i spółgłosek):

Samogłoski (1-2)

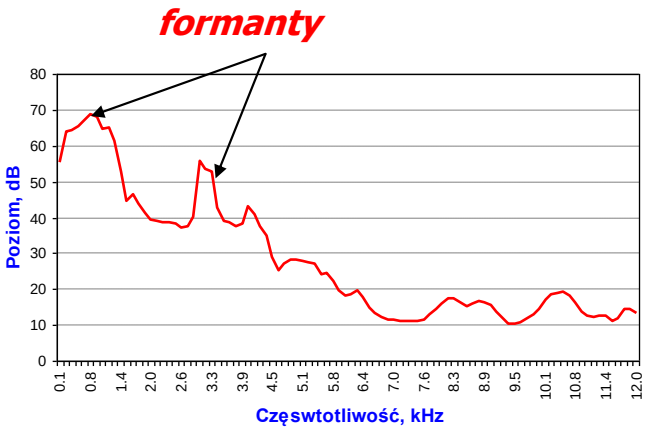
1. Monoftongi (jednogłoskowe samogłoski) – jez. polski: /a/, /^/, /E/, /&/, /i:/, /I/, /o/, /U/, /Y/
2. Dyftongi (dwie samogłoski wymawiane łącznie) – jez. polski: /ą/, /ę/

Spółgłoski (3-7)

3. Ustne (półsamogłoski i sonanty) – jęz. polski: /w/ (dźwięk „ł"), /y/ (dźwięk „j)", /l/, /r/
4. Nosowe – jez. polski: /m/, /n/, /ni/, /ng/
5. Spiranty – jez. polski: /f/, /s/, /v/ (dźwięk „w"), /V/, /z/, /h/, /ś/, /ź/, /sz/, /ż/
6. Zwarte – jez. polski: /p/, /t/, /k/ (bezdźwięczne), /b/, /d/, /g/ (dźwięczne)
7. Afrykaty – jez. polski: /c/, /ć/, /cz/, /dz/, /dź/, /dż/

2. Formanty

Formanty definiujemy jako lokalne maksima obwiedni charakterystyki amplitudo-częstotliwościowej sygnału, a wartości częstotliwości, przy których występują – częstotliwościami formantowymi (oznaczone są kolejno jako F1, F2, F4,...itd.).



Obraz częstotliwościowy mowy (spektogram) przedstawia charakterystyczny rozkład energii drgań o poszczególnych składowych częstotliwościowych. Są one powodowane cyklicznym otwieraniem i zamykaniem fałdów głosowych oraz procesem artykulacji przez elementy traktu głosowego usta i nos. Każdy fonem posiada swój unikatowy wzorzec w postaci spektrogramu. Jednakże odróżnienie tych wzorców wymaga dużego doświadczenia i wiedzy specjalisty.
Przede wszystkim w spektrogramie możemy obserwować charakterystyczne maksima lokalne energii odpowiadające częstotliwości podstawowej i formantom.
Częstotliwość podstawowa drgań (F0) charakteryzuje się największą energią (dominującą) w sygnale mowy. Formanty oznaczają duże koncentracje energii obserwowane na spektrogramie (oznaczane: F1, F2, F3 itd.). Większość głosek (dźwięcznych) można scharakteryzować badając wyłącznie rozkłady tych formantów.

F1 występuje w zakresie od 300 Hz do 1000 Hz, położenie jest głównie zdeterminowane główną „przeszkodą" – ustami. Kiedy otwarcie usta maleje, wtedy F1 również maleje. Ponadto F1 ma tym niższą wartość, im bardziej język zbliżony jest do nasady ust.

F2 występuje w zakresie od 850 Hz do 2500 Hz, jest proporcjonalne do czołowego lub tylnego położenia górnego końca języka. Innymi słowy – odległość od głośni do aktualnego położenia górnego końca języka determinuje położenie F2 – gdy ta odległość rośnie, wówczas położenie F2 również rośnie. Zaokrąglenie ust powoduje obniżenie wartości F2.

Wyższe formanty (F3, F4, F5) występują rzadziej i są pomocne w ocenie samej jakości dźwięku

Przykłady:
/i:/ - samogłoska zamknięta (wysoka) – przednia: małe otwarcie ust, posiada jedno z najniższych F1 ok. 300 Hz, posiada wysokie F2 ok. 2200 Hz, najwyższe ze wszystkich samogłosek, koniec języka wysunięty jest do przodu, a usta nie są zaokrąglone, koncentracja energii w zakresie wysokich częstotliwości > 1800 Hz.

/A/ - samogłoska otwarta (niska) – najwyższe F1 ok. 950 Hz, energia w zakresie 800÷1800 Hz

/u/ - tylna samogłoska – posiada niskie F2 ok. 850 Hz, koniec języka jest cofnięty, a usta są zaokrąglone, prawie cała energia w zakresie niskich częstotliwości < 1000 Hz.

Tabela 1. Zakresy występowania częstotliwości formantowych samogłosek polskich.

Samogłoska	F_1 [Hz]		F_2 [Hz]		F_3 [Hz]		F_4 [Hz]	
	f_d	f_g	f_d	f_g	f_d	f_g	f_d	f_g
a	**680**	**1020**	**1130**	**1570**	**2330**	**2860**	**3100**	**4090**
e	**520**	**630**	**1580**	**2230**	**2470**	**3150**	**3070**	**4030**
i	**190**	**275**	**2080**	**2840**	**2670**	**3430**	**3320**	**4140**
o	**490**	**680**	**790**	**1100**	**2410**	**3030**	**3200**	**3960**
u	**240**	**340**	**560**	**790**	**2270**	**3190**	**2940**	**4058**

Ogólna postać definiująca formanty przedstawia zależność:

$$F_{ij} \equiv f_{ij} \Leftrightarrow \left[\left(\frac{\partial G(t_j, f)}{\partial f} = 0 \right) \wedge \left(\frac{\partial^2 G(t_j, f)}{\partial f^2} < 0 \right) \wedge \left(f_d \le F_{ij} \le f_g \right) \right]$$

gdzie: G(t,f) – widmo czasowo-częstotliwościowe w chwili czasowej t,
i-ty formant i=1,2,3,4,
j-ta chwila czasu t,
f_d,f_g – są odpowiednio w dyskretnej skali częstotliwościami ograniczającymi pasmo poszukiwań formantu (np. wg Tabeli 1.)

3. Podstawowa charakterystyka spektralna grup fonemów

- stabilne (stałe w czasie) formanty występują w monoftongach i głoskach nosowych. Dla głosek nosowych charakterystyczne są obszary względnego „zera" energii między formantami,
- wolnozmienne w czasie formanty występują w dyftongach i głoskach ustnych,
- monoftongi posiadają silne (łatwo wykrywalne) częstotliwości pierwszych dwóch i trzech formantów (F1, F2, F3),
- wszystkie dźwięczne głoski posiadają formanty – chociaż trudniej wykrywalne niż dla monoftongów,
- bezdźwięczne głoski zwykle nie posiadają formantów
- obraz głosek zwartych wygląda jak duże wybuchy energii w całym zakresie częstotliwości, poprzedzone okresem względnej ciszy
- spektrogramy dla afrykatów i spirantów wyglądają jak chmury lub ocean realtywnie ciagłej energii

4. Zadania do wykonania

a) Przy użyciu ***fft*** i ***envelope*** wykreślić obwiednię widmową samogłosek /a/, /e/, /i/, /u/. Wyznaczyć F1 i F2. Porównać z danymi zawartymi w Tabeli 1.
b) Sprawdzić otrzymane wyniki realizując punkt a) przy pomocy funkcji ***lpc*** (więcej: https://www.mathworks.com/help/signal/ug/formant-estimation-with-lpc-coefficients.html)