

Akustyka mowy wprowadzenie



Opracował:
dr inż. Piotr Suchomski



Kontakt

Katedra Systemów Multimedialnych
Wydział ETI

dr inż. Piotr M. Suchomski, pok. EA 730

e-mail: *pietka@sound.eti.pg.gda.pl*

tel. 23-01

<http://www.sound.eti.pg.gda.pl>



Wprowadzenie

- Wykład obejmuje prezentację szeregu podstawowych pojęć z zakresu szeroko rozumianej akustyki mowy.
- Akustyka mowy obejmuje takie zagadnienia jak badanie sposobu wytwarzania dźwięków mowy, sposoby rozumienia mowy, metody analizy i przetwarzania sygnałów mowy oraz metody syntezy mowy.



Plan wykładu

1. Wprowadzenie, podstawowe wiadomości na temat sygnału mowy i traktu głosowego
2. Teoria wytwarzania dźwięków mowy, modelowanie mechanizmów wytwarzania dźwięków mowy
3. Metody analizy sygnału mowy
4. Parametryzacja sygnału mowy, perceptualne skale częstotliwości
5. 5. Kodowanie i komprimowanie sygnału mowy, standardy μ -law i A-law



Plan wykładu

6. Przetwarzanie sygnału mowy
7. Transformacja głosu
8. Podstawy syntezy mowy
9. Podstawy automatycznego rozpoznawania mowy



Laboratorium

1. Metody detekcji aktywności głosowej w sygnale mowy
2. Badanie formantowości sygnału mowy
3. Algorytm poprawy zrozumiałości mowy obarczonej zakłóceniami addytywnymi
4. Synteza mowy
5. Badanie systemów kodowania mowy
6. Analiza sygnału mowy z zastosowaniem techniki predykcji liniowej – rozpoznawanie elementów mowy



Laboratorium

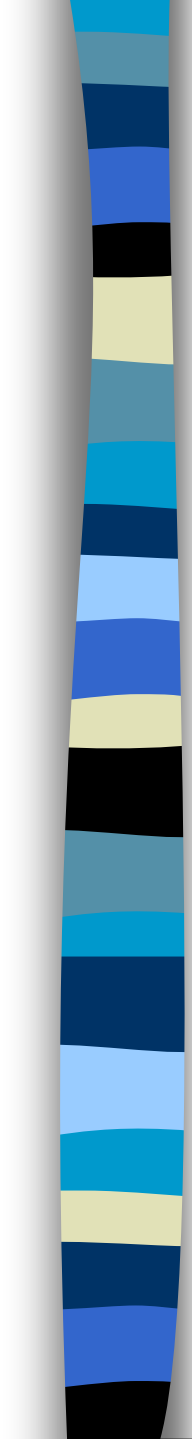
- Wtorek, czwartek lab. EA 630, godz. 14-16

Nr ćwiczenia	terminy
wprowadzenie	22.02, 24.02
1	8.03, 10.03
2	22.03, 24.03
3	05.04, 07.04
4	12.04, 14.04
5	10.05, 12.05
6	24.05, 26.05



Zaliczenie

- Wynik kolokwium (na koniec semestru)
+ ocena z laboratorium (50% +50%)



Podstawowe informacje o wytwarzaniu mowy i sygnale mowy



Procesy wytwarzania mowy

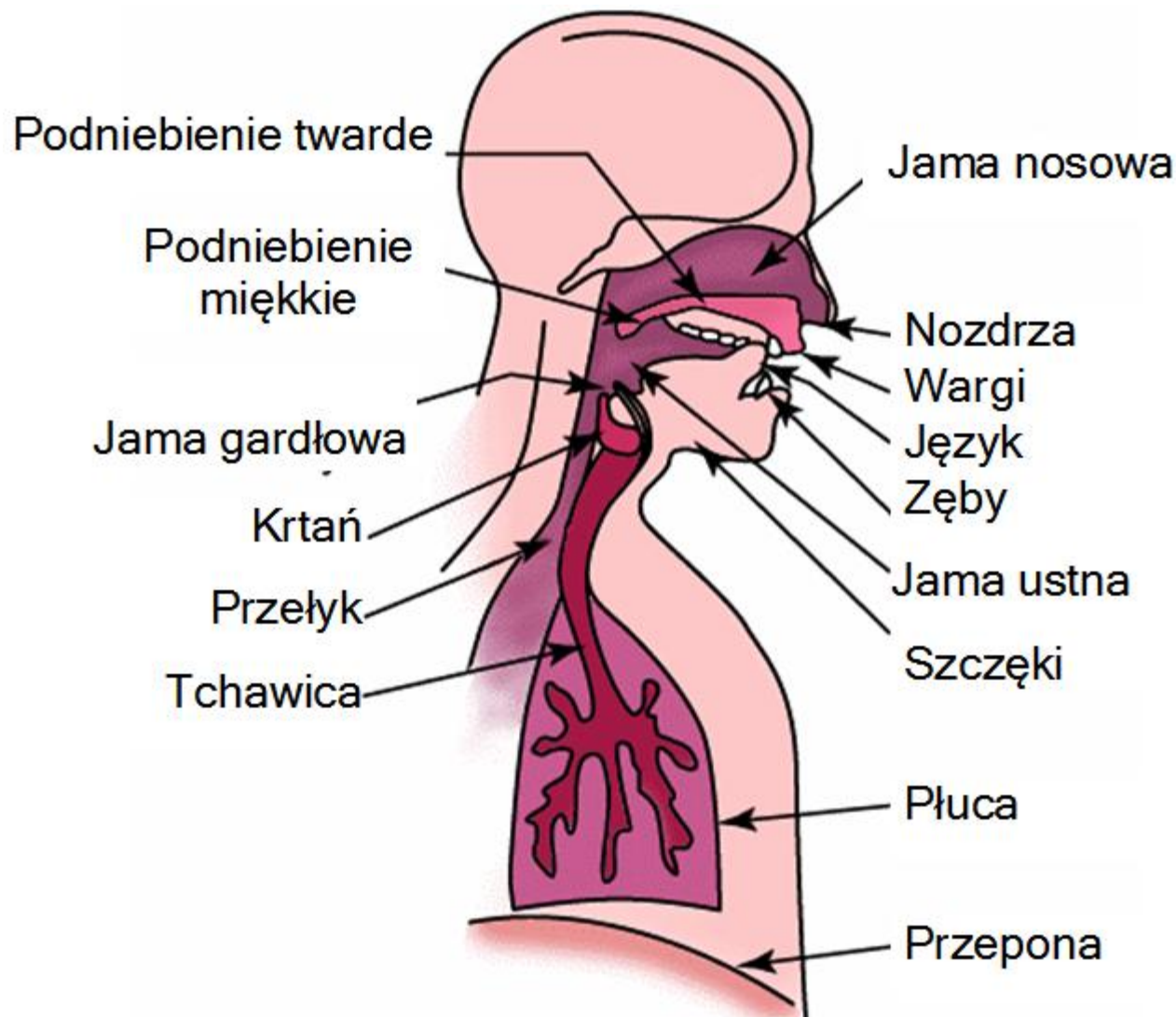
- Proces psychologiczny – przygotowanie informacji do przekazania (konceptualizacja),
- Proces neurologiczny – pobudzenie w ośrodkowym układzie nerwowym oraz na drodze eferentnej mięśni narządu mowy,
- Proces fizjologiczny – artykulacja,
- Proces aerodynamiczny – przepływ powietrza i generowanie drgań o złożonej strukturze widmowo-czasowej.



Zdolność mówienia

- Mimo, że budowa narządu mowy u wszystkich naczelnych jest podobna to zdolność mówienia posiadał tylko człowiek.

Narząd mowy – trakt głosowy

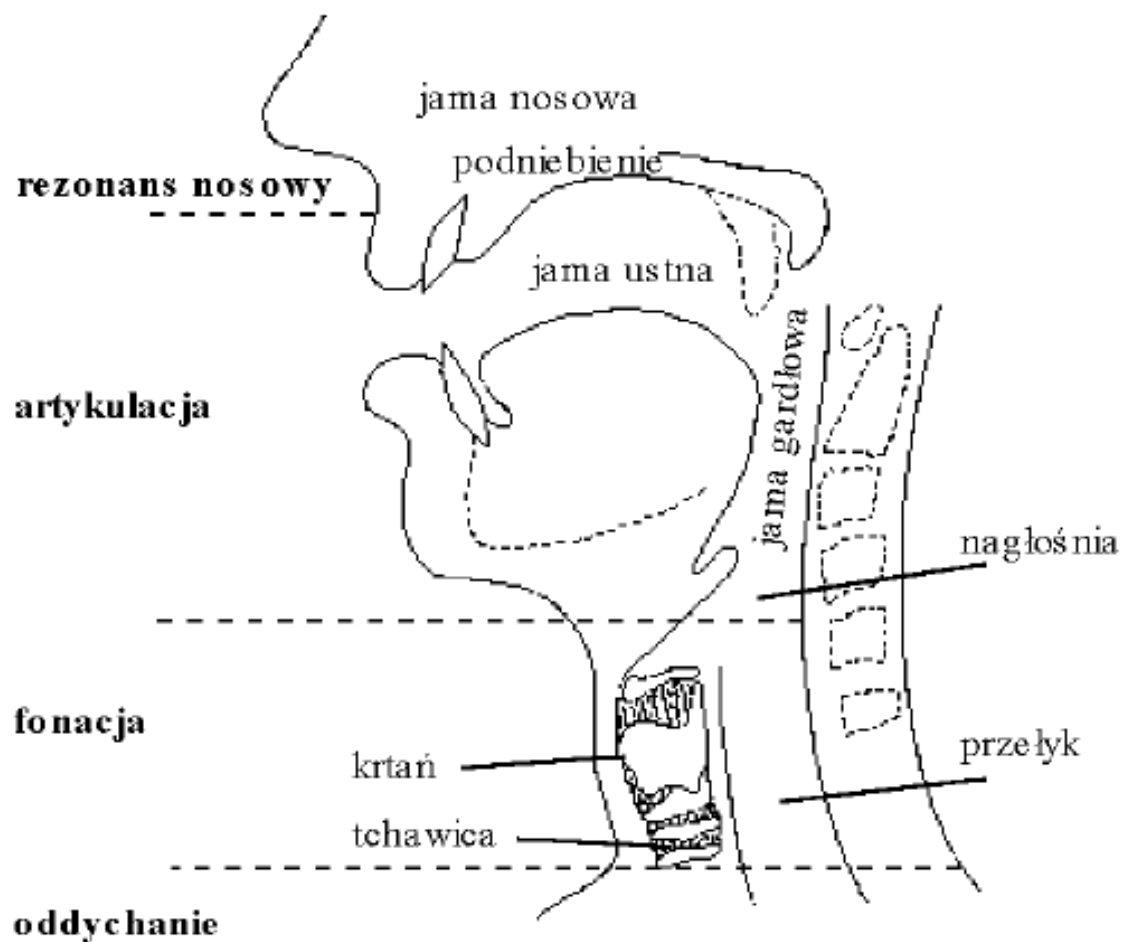




Wytwarzanie mowy

- Płuca, oskrzela i tchawica tworzą drogę doprowadzającą powietrze do krtani.
- W krtani znajdują się fałdy głosowe (pot. struny głosowe), które pod ciśnieniem powietrza zaczynają drgać i wytwarzają tzw. ton krtaniowy.
- Wygenerowany ton następnie jest „filtrowany” w dalszej części traktu głosowego. Na sposób filtracji wpływają takie narządy jak: język, języczek, podniebienie, zęby, usta, jama nosowa.

Wytwarzanie mowy





Sygnał mowy - struktura

- Sygnał mowy powstaje w wyniku splotu tonu krtaniowego (fonacja) i odpowiedzi traktu głosowego (artykulacja).
- Fonacja odpowiada za proces wytwarzania energii dźwięku, natomiast artykulacja kształtuje charakterystykę częstotliwościową.
- Niestety proces rozplotu nie jest operacją trywialną



Cechy charakterystyczne

- Dwie warstwy wpływające na charakterystykę mowy:
 - Fizyczna – wynikająca z anatomicznych właściwości elementów traktu głosowego,
 - Psychiczna – indywidualny, wyuczony sposób fonacji i artykulacji.
- O charakterystycznej barwie głosu w dużej mierze decyduje ton krtaniowy, którego częstotliwość zależy od długości fałd głosowych w krtani.



Cechy mowy

- Mowę można analizować na kilku poziomach:
 - semantyczny – treść mowy,
 - osobniczy – cechy pozwalające zidentyfikować mówcę,
 - emocjonalny – poznanie stanu emocjonalnego, stanu zdrowia, pozycji społecznej itp.
 - prozodyczny – związany z akcentem, intonacją itp..



Sygnał mowy

- Sygnał mowy jest splotem parametrów układu artykulacji – traktu głosowego (jama ustna, nosowa, język itp.) i tonu krtaniowego (charakterystyczna częstotliwość pobudzenia).
- Zależność poziomu ciśnienia sygnału mowy od częstotliwości oznacza, że w określonych zakresach częstotliwości, składowe widma dźwięków mowy przybierają wartości znacznie wyższe niż w pozostałych zakresach częstotliwości. Te zakresy częstotliwości, w których składowe widma przyjmują maksymalne wartości nazywa się formantami, zaś odpowiadające im częstotliwości – częstotliwościami formantowymi.



Sygnał mowy

- Sygnał mowy charakteryzuje duża grupa parametrów akustyczno-fonetycznych, przy czym nie wszystkie parametry biorą bezpośredni udział w procesie percepcji. Ta redundancja danych pozwala percypować dźwięk nawet w trudnych warunkach akustycznych.
- Tony podstawowe mowy są w zakresie od 74 Hz do 1056 Hz. Istotne znaczenie dla zrozumiałości mowy mają częstotliwości nawet do 10 kHz.
- Poziom dźwięku mowy – 34 dB – 94 dB.



Poziomy percepcji mowy

- Proces percepcji mowy jest procesem złożonym i składa się z kilku faz:
 - Aerodynamiczna – drgania powietrza w przewodzie słuchowym,
 - Akustomechaniczna – przenoszenie drgań od błony bębenkowej do ślimaka,
 - Neurologiczna – przenoszenia i przetwarzanie impulsów w ośrodkowym układzie nerwowym,
 - Psychologiczna – rozpoznanie i zrozumienie przekazanej informacji.



Specyfika percepcji mowy

- Percepcja dźwięków mowy nie zależy wyłącznie od ich struktury czasowo-widmowej, ale również od treści lingwistycznej.
- Nie wiadomo jakie elementy mowy (np. fonemy, sylaby, wyrazy itp.) są podstawową jednostką percepcji mowy.



Specyfika percepcji mowy

- Badania neurofizjologiczne pokazują, że u większości osób percepcja mowy odbywa się w lepszym stopniu za pomocą ucha prawego zaś w przypadku muzyki jest odwrotnie (dźwięki mowy są lepiej dekodowane przez lewą półkulę mózgu)



Wyrazistość a zrozumiałość

- Pojęcie wyrazistości dotyczy tych elementów fonetycznych mowy, które nie mają określonego znaczenia semantycznego (głoski, zgłoski, logatomy), natomiast zrozumiałość dotyczy elementów mowy, które mają określone znaczenie semantyczne.
- Zrozumiałość jest złożoną funkcją wyrazistości.
- Jako miarę wyrazistości lub zrozumiałości mowy przyjmuje się stosunek liczny poprawnie odebranych elementów fonetycznych do całkowitej liczby wszystkich zaprezentowanych elementów fonetycznych.



Zrozumiałość mowy

- Zrozumiałość mowy może być analizowana w pasmach częstotliwości. Największy udział w zrozumiałości mowy ma wąskie pasmo wokół częstotliwości 1900 Hz. „Odfiltrowanie” sygnału mowy poniżej 200 Hz lub powyżej 6000 powoduje brak zrozumiałości mowy.
- Miarą zrozumiałości jest współczynnik artykulacji AI. Pasmo mowy dzielone jest na 20 pasm, z których każde wnosi do wypadkowej zrozumiałości 5%. Wartość współczynnika AI jest z zakresu od 0 do 1.



Wpływ na zrozumiałość mowy

■ Zrozumiałość mowy utrudnia:

- Szum (dla $\text{SNR} = 0 \text{ dB}$ zrozumiałość jest na poziomie 50%, dopiero gdy $\text{SNR} > 6 \text{ dB}$ zrozumiałość jest zadowalająca),
- Pogłos – im dłuższy czas pogłosu, tym słabsza zrozumiałość mowy,
- Zniekształcenia nieliniowe,
- Zniekształcenia fazowe,
- Zniekształcenia amplitudowe (np. obcięcie sygnału)
- zniekształcenia



Język naturalny a sygnał mowy

- Aby dźwięki mowy miały określone znaczenie musi istnieć wzajemne przyporządkowanie między strukturą akustyczną sygnału mowy a przekazywaną informacją.
- Badaniem struktury dźwiękowej języka naturalnego zajmuje się fonetyka. Nauka ta zajmuje się selekcją dźwięków elementarnych, z których przez złożenie, powstają określone formy językowe.



Dźwięki elementarne mowy

- Fonem - minimalny segment dźwiękowy mowy, który może odróżniać znaczenie, lub inaczej klasa dźwięków mowy danego języka o różnicach wynikających wyłącznie z charakteru indywidualnej wymowy lub kontekstu.
- Alofon - wariant fonemu odróżniający się od innego alofonu cechami fonetycznymi a nie funkcją.
- Difon (diafon) –przejście (złączenie) dwóch fonemów.
- Mikrofonem – jednostka sygnału mowy o stałej długości (20-40 ms.)



Fonemy

- W języku polskim można wyróżnić 37 fonemów + 2 samogłoski nosowe
- Fonemy języka polskiego można sklasyfikować za pomocą binarnych cech dystynktywnych:
 - Spółgłoski – samogłoski,
 - Ponadkrtaniowe – krtaniowe,
 - Nosowe – ustne,
 - Łagodne-raptowne
 - Skupione-rozproszone
 - Jasne – ciemne,
 - Niskotonowe-wysokotonowe,
 - Długie-krótkie,
 - Dźwięczne - bezdźwięczne



Fonemy

- dźwięki o charakterze quasiperiodycznym:
 - 1) samogłoski sylabiczne (a, e, i, o, u, y)
 - 2) samogłoski niesylabiczne (j, ł)
 - 3) spółgłoski nosowe (m, n, ń, ą, ę)
 - 4) spółgłoski boczne (l)
- dźwięki o charakterze przebiegów nieperiodycznych - szumowych:
 - 1) spółgłoski bezdźwięczne trące (f, s, sz, ś, h)
 - 2) spółgłoski bezdźwięczne zwarto-trące (c, ć, cz)
- dźwięki o charakterze przebiegów nieperiodycznych - quasi-impulsowych:
 - 1) spółgłoski zwarte dźwięczne (b, d, g)
 - 2) spółgłoski zwarte bezdźwięczne (p, t, k)



Fonemy

- dźwięki o charakterze przebiegów będących superpozycją quasiperiodycznych i nieperiodycznych:
 - 1) spółgłoski trące dźwięczne (w, z, \hat{S} , \hat{Z})
 - 2) spółgłoski zwarto-trące dźwięczne (dz, $d\hat{S}$, $d\hat{Z}$)



Inne ważne pojęcia

- Formant (częstotliwość formantowa) - obszar koncentracji energii w widmie danego dźwięku mowy lub inaczej: taki zakres widma, którego obwiednia zawiera maksimum.
- Cechy dystynktywne – cechy obiektów, na podstawie których można je rozróżniać.
- Ekstrakcja parametrów - procedura wydzielania z sygnału cech reprezentowanych przez wartości liczbowe.



Inne ważne pojęcia

- Wokodery - urządzenia służące do ograniczania objętości informacyjnej sygnału mowy metodą ekstrakcji parametrów i następnie po przesłaniu parametrów przez kanał telekomunikacyjny dokonujące resyntezy tego sygnału.



Dziękuję za uwagę