

Analiza głosu - rozpoznawanie samogłosek.

Edward Pótrolniczak, dr inż., Katedra Systemów Multimedialnych

Rozpoznawanie mowy

- Jednym z elementów systemów rozpoznawania mowy jest rozpoznawanie fonemów, które z kolei składają się dalej na głoski
- Przypomnijmy, że:
 - 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)
 - Formant - obszar koncentracji energii w widmie danego dźwięku mowy lub inaczej: taki zakres widma, którego obwiednia zawiera maksimum; **formanty składają się na fonemy**

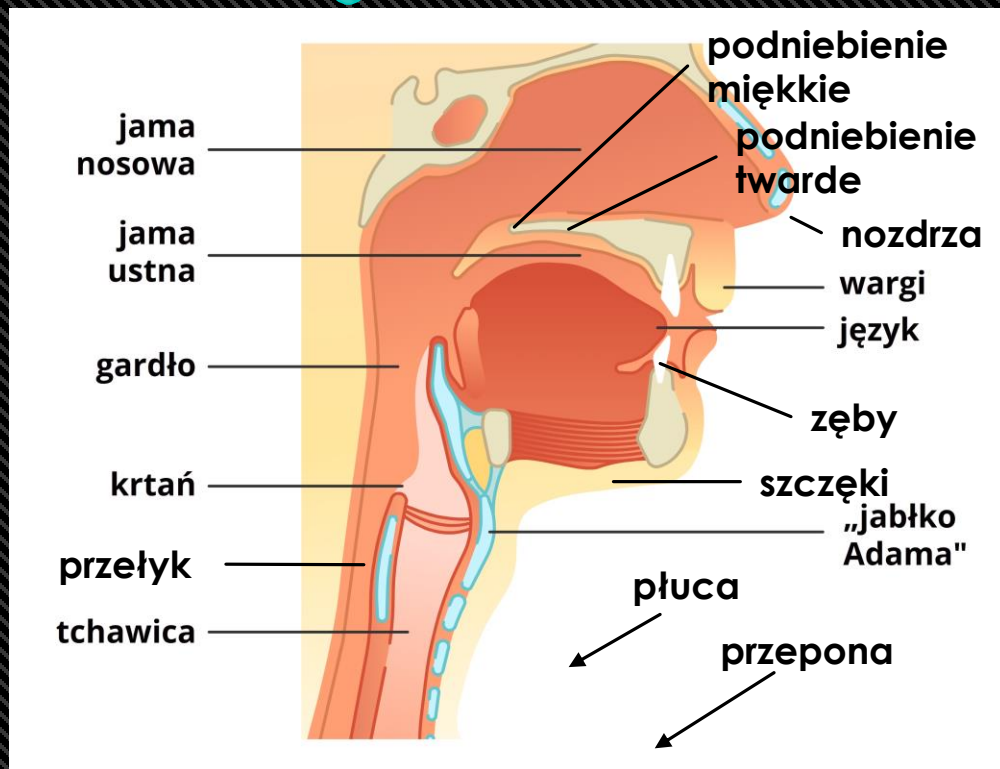
Rozpoznawanie mowy

- W języku polskim literatura wyróżnia 37 fonemów + 2 samogłoski nosowe. Można je sklasyfikować za pomocą binarnych cech dystynktywnych:
 - spółgłoski - samogłoski,
 - ponadkrtaniowe - krtaniowe,
 - nosowe - ustne,
 - łagodne - raptowne,
 - skupione - rozproszone,
 - jasne - ciemne,
 - niskotonowe - wysokotonowe,
 - długie - krótkie,
 - dźwięczne - bezdźwięczne.

Cel zadania

- Rozpoznawanie dźwięków o charakterze quasiperiodycznym, którymi są samogłoski sylabiczne (a, e, i, o, u, y) z wykorzystaniem formantów

Trakt głosowy



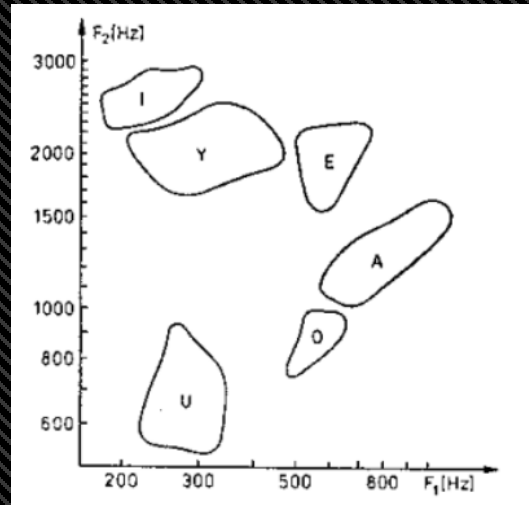
- Trakt głosowy jest konstrukcją złożoną z wielu „elementów” wyszczególnionych na rysunku obok
- Formanty wykrywane na niniejszych laboratoriach generowane są w końcowym etapie produkcji głosu
- Dysfunkcje tego odcinka traktu głosowego prowadzi do wad wymowy i trudności w zrozumieniu przez słuchacza (również sprawia trudności systemowi analizy mowy)

Plan zadania

- Wykorzystując metody zawarte w plikach dostarczonego pakietu VowelRecognition należy zbudować własny zbiór uczący, na który składają się wartości formantów dla głosek a e i o u y oznaczonych kolejno etykietami 1 2 3 4 5 6 (wzór znajduje się w pliku data_for_training.csv)
- Do stworzeniu ciągu uczącego należy dokonać analizy 6 powtórzeń każdej z głosek czyli 36 głosek
- Nagrania należy zrealizować możliwie najlepszym mikrofonem, w warunkach otoczenia pozbawionych hałasów i pogłosu (w razie potrzeby zamknąć okna i zasłonić zasłony) (proszę posłuchać czy jakieś urządzenie w pobliżu nie szumi lub nie warczy – jeśli tak to proszę je wyłączyć lub przenieść się poza zasięg zakłócenia)
- Proszę dokonać analizy skuteczności rozpoznawania z wykorzystaniem zbudowanego samodzielnie ciągu uczącego, proszę opisać czy wskazania pokryły się z rzeczywistością (określoną „na ucho”); proszę przeprowadzić również 36 prób (po 6 na każdą głoskę) i wyznaczyć współczynnik poprawnego rozpoznawania (błąd fałszywego rozpoznania – FRR (False Recognition Rate) – ile razy na zadaną liczbę prób głoska została źle rozpoznana, współczynnik poprawnego rozpoznawania – TRR – True Recognition Rate (ile razy była właściwie rozpoznana); współczynniki można podać dla każdej głoski osobno i wypadkowo dla całego systemu
- Opcjonalnie: Można stworzyć kilka modeli i je ze sobą porównać; można stworzyć model dla kilku osób i sprawdzić czy zwiększenie uniwersalności pogarsza jakość rozpoznawania (można pociągnąć dane od osób z grupy – ważne, żeby etykiety się zgadzały!)

Plan zadania

- Jako pomoc do wykonania zadania dostarczony jest pakiet VowelRecognition, plik CSV zawiera wartości formantów dla głosek, ale należy je skorygować – dla języka polskiego powinny (choć niekoniecznie muszą) mieścić się w przedziałach
- Proszę zwrócić uwagę, że formanty w przestrzeni cech układają się mniej więcej jak na rysunku poniżej



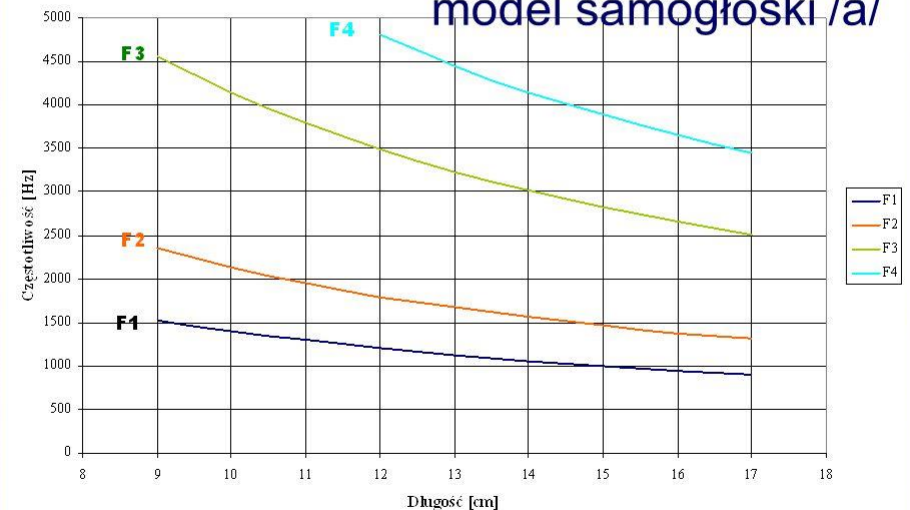
Samogloska	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
/i/	188-275	2078-2836	2670-3432	3316-4144
/y/	262-391	1689-2362	2424-3146	3124-4226
/e/	524-630	1580-2228	2468-3146	3064-4034
/a/	683-1021	1132-1566	2328-2860	3098-4088
/o/	493-679	788-1100	2410-3026	3194-3954
/u/	242-338	558-789	2266-3188	2942-4058

Plan zadania

- Z wykonania zadania można sporządzić sprawozdanie zawierające opis zadania, opis pozyskanych próbek, opis uzyskanych wyników, wnioski
- Wyniki, dane i skrypty należy umieścić w systemie MTeams w odpowiedzi na zadanie wygenerowane przez prowadzącego zajęcia

Wpływ długości toru głosowego na rozkład częstotliwości formantowych

model samogłoski /a/



Podsumowanie

- Przypomnienie najważniejszych pojęć

Dziękuję za uwagę