Politechnika Częstochowska

INTELIGENTNE SYSTEMY UWIERZYTELNIANIA

dr hab. inż. Mariusz Kubanek, prof. PCz
mariusz.kubanek@icis.pcz.pl
Katedra INFORMATYKI

Politechnika Częstochowska

KI

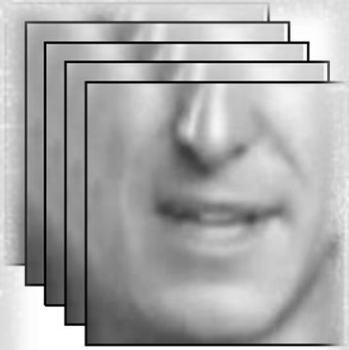
Wykład 5
Uwierzytelnianie na podstawie głosu

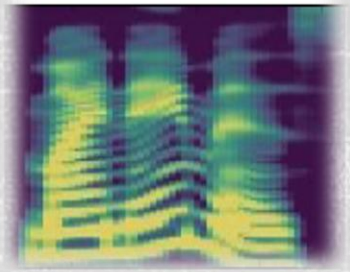
MOWA CZŁOWIEKA

- Percepcja ludzkiej mowy jest z natury wielo-modalnym procesem, w którym wykorzystuje się analizę sygnału akustycznego, polegającą na analizie gramatycznej, semantycznej i pragmatycznej.
- Dodatkowo wiadomo, że człowiek posiada zdolność czytania mowy poprzez analizę ruchu ust mówcy, czyli tzw. zdolność czytania z ruchu warg.
- Do tej pory wiele badań prowadzono na temat automatycznego rozpoznawania.

MOWA CZŁOWIEKA

- Obecnie główne wysiłki skierowane są na tworzenie systemów odpornych na negatywnie wpływające czynniki zewnętrzne.
- Zaczęto poszukiwać sposobów ograniczenia wpływu zakłócenia na właściwą pracę systemów.
- Jednym z takich sposobów może być zastosowanie w dołączenia do rozpoznawanej audio mowy, mowy wideo, będącej elementem ograniczającym wpływ negatywnych czynników zewnętrznych na skuteczność rozpoznawania.





MOWA CZŁOWIEKA

- **Z uwagi na możliwość kojarzenia mowy na podstawie ruchu warg możliwe jest połączenie informacji audio i wideo w podjęciu decyzji o treściowym wyniku wypowiedzi, specjalnie w zakłóconym środowisku audio mowy.**

Politechnika Częstochowska

MOWA CZŁOWIEKA

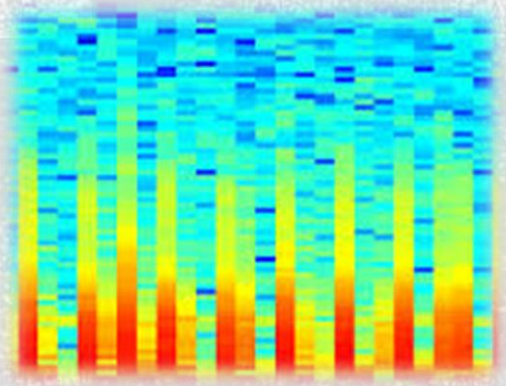
- Zastosowanie rozpoznawania audio mowy w zakłóconym otoczeniu prowadzi często do błędnych wyników, spowodowanych nieprawidłową interpretacją fonemów o bliskim brzmieniu.
- Wideo mowa również może być błędnie interpretowana.
- Wideo sygnał nie niesie wystarczającej informacji, zawiera jednak kilka uzupełniających informacji do audio sygnału.

MOWA CZŁOWIEKA

- Powyższe fakty wywarły duży wpływ na sfery rozpoznawania audio-wideo.
- Prace w tej dziedzinie prowadzone są w celu polepszenia zakresu rozpoznawania automatycznej mowy poprzez ekstrakcję cech z obszaru ust mówcy i połączenie z tradycyjną mową akustyczną.
- Takie osiągnięcie zysku jest szczególnie imponujące w hałaśliwym środowisku, gdzie tradycyjna metoda rozpoznawania audio mowy wypada niezbyt korzystnie.

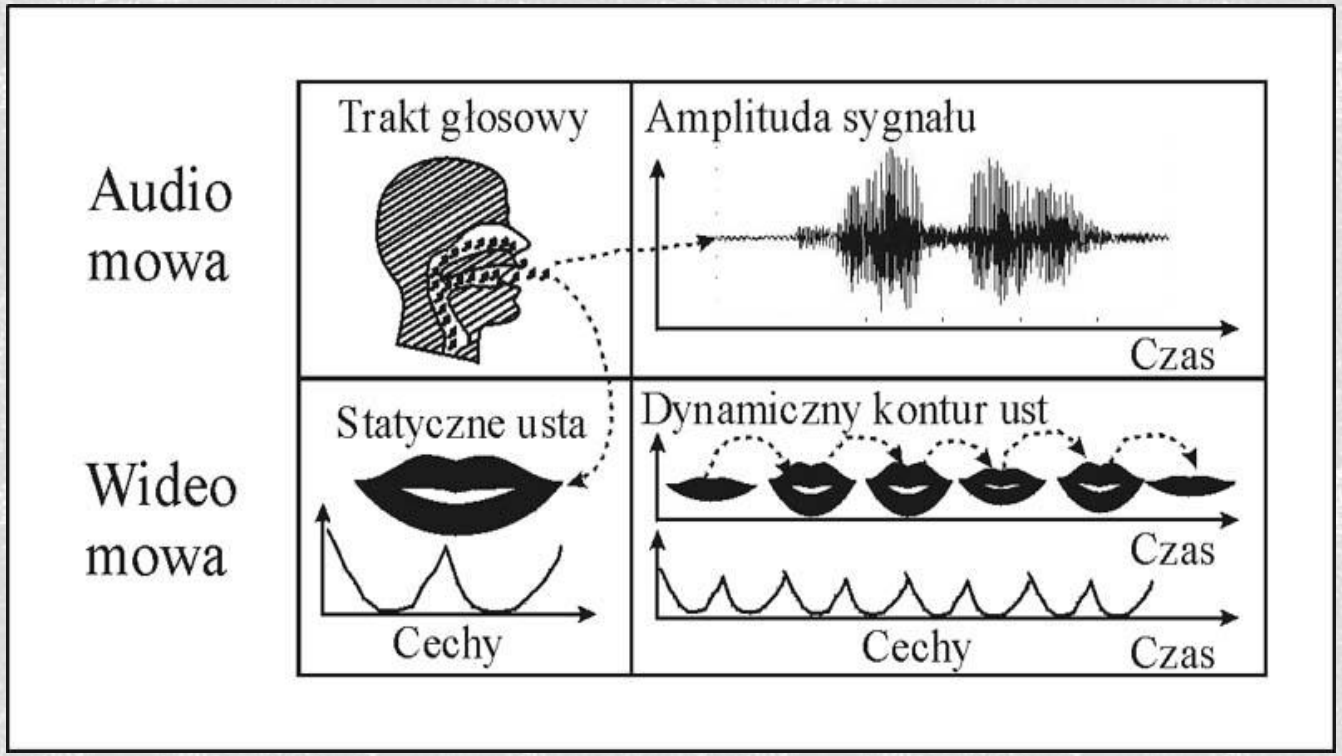
MOWA CZŁOWIEKA

- **Przy rozpoznawaniu audio-wideo mowy należy rozwiązać cztery podstawowe zagadnienia:**
  - **identyfikacji i ekstrakcji określonych charakterystyk audio,**
  - **identyfikacji i ekstrakcji określonych charakterystyk wideo,**
  - **racjonalnej integracji (fuzji) i synchronizacji audio-wideo sygnałów,**
  - **wyboru i realizacji aparatu realizującego uczenie i rozpoznawanie sygnałów mowy.**



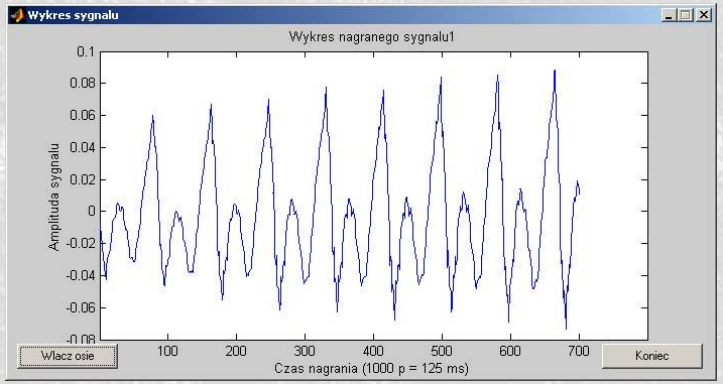
Politechnika Częstochowska

SFORMUŁOWANIE PROBLEMU ROZPOZNAWANIA



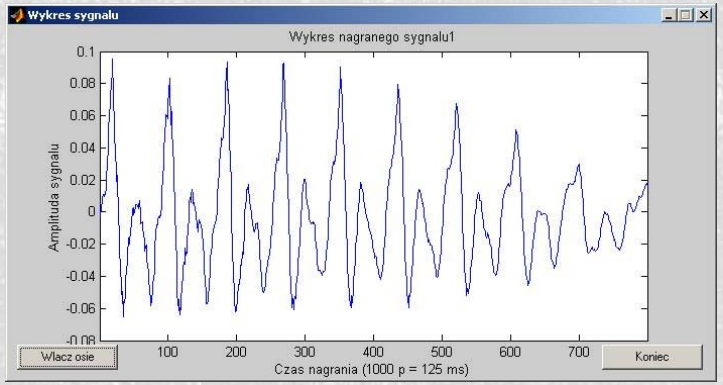
Politechnika Częstochowska

SPECYFICZNE CECHY MOWY



Widok fonemu *m* audio

Widok fonemu *m* video

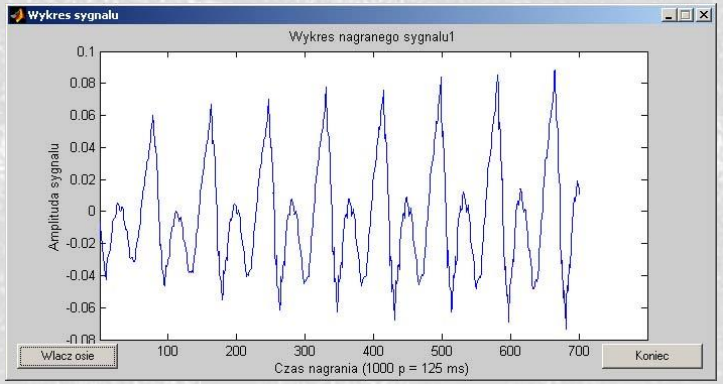


Widok fonemu *n* audio

Widok fonemu *n* video

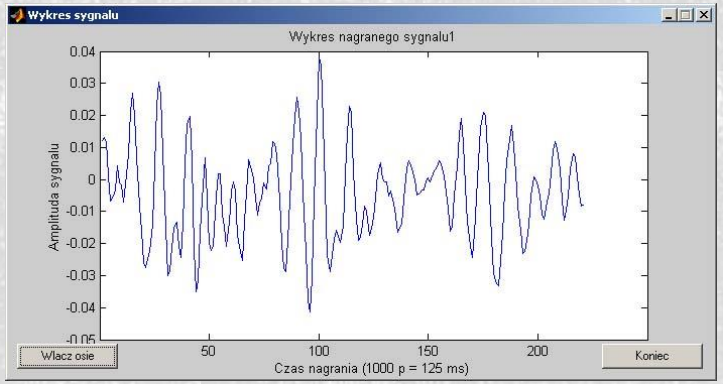
Politechnika Częstochowska

SPECYFICZNE CECHY MOWY



Widok fonemu *m* audio

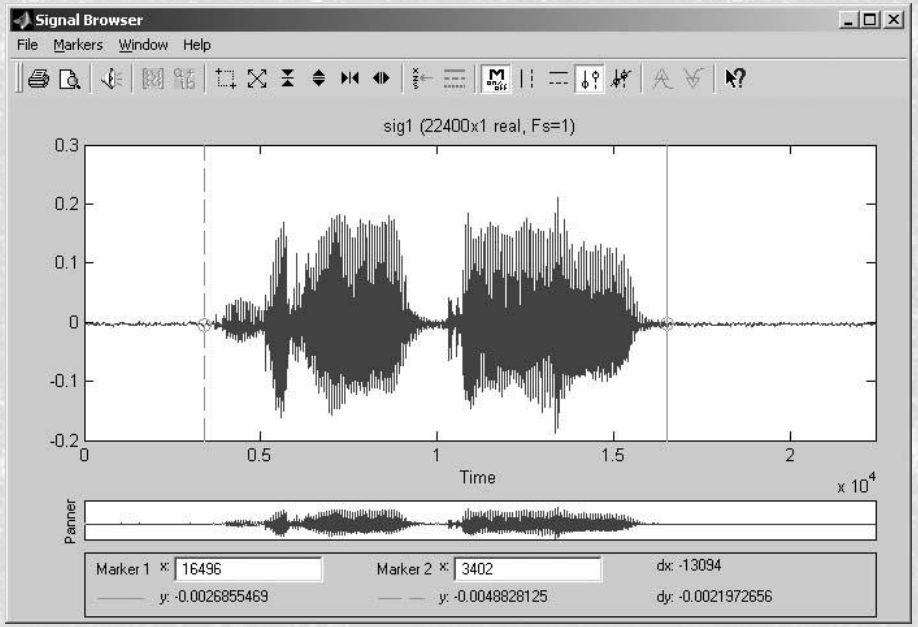
Widok fonemu *m* video



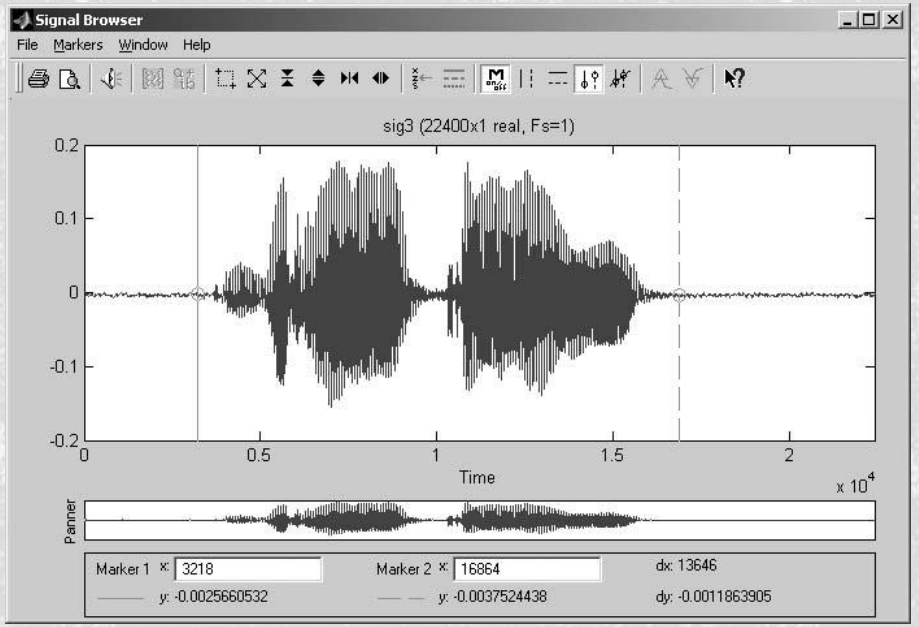
Widok fonemu *p* audio

Widok fonemu *p* video

TWORZENIE WEKTORÓW OBSERWACJI



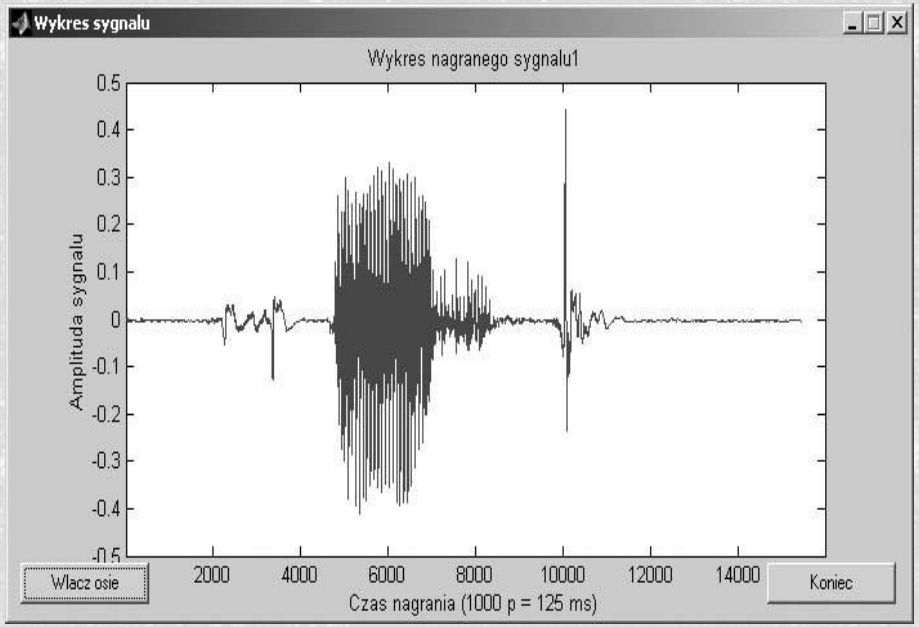
przed filtracją



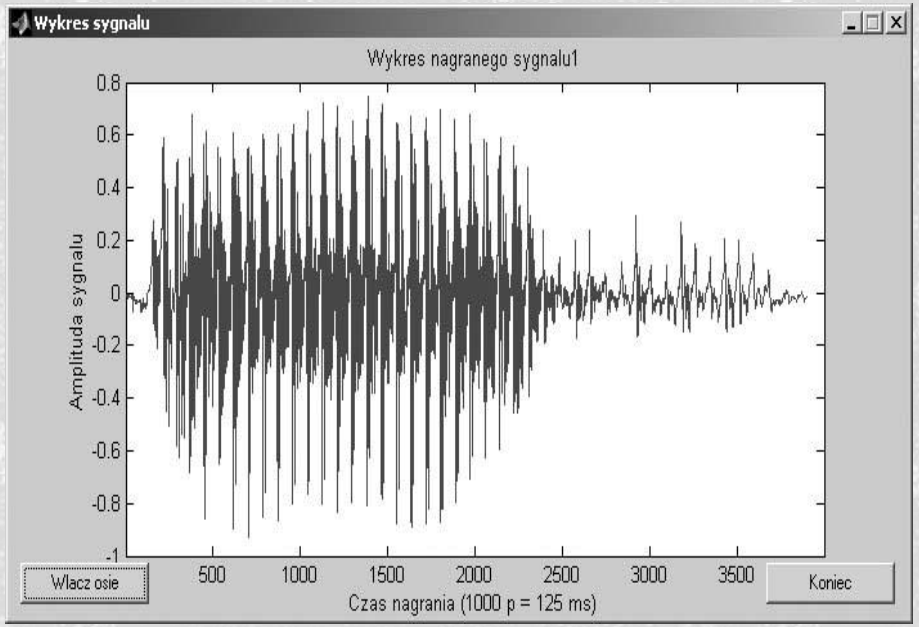
po filtracji

Politechnika Częstochowska

IZOLOWANIE SŁÓW
ENERGIA



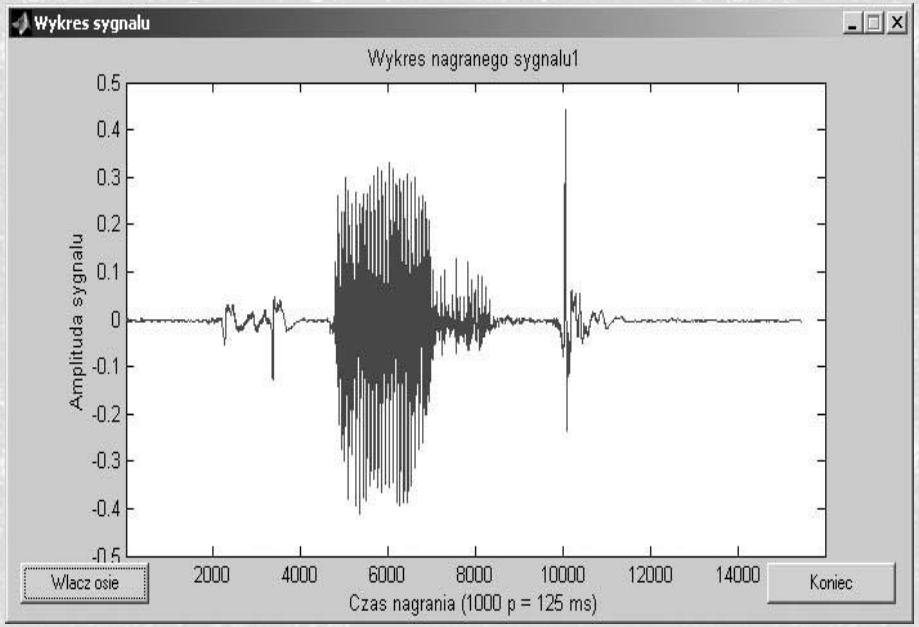
przed izolacją



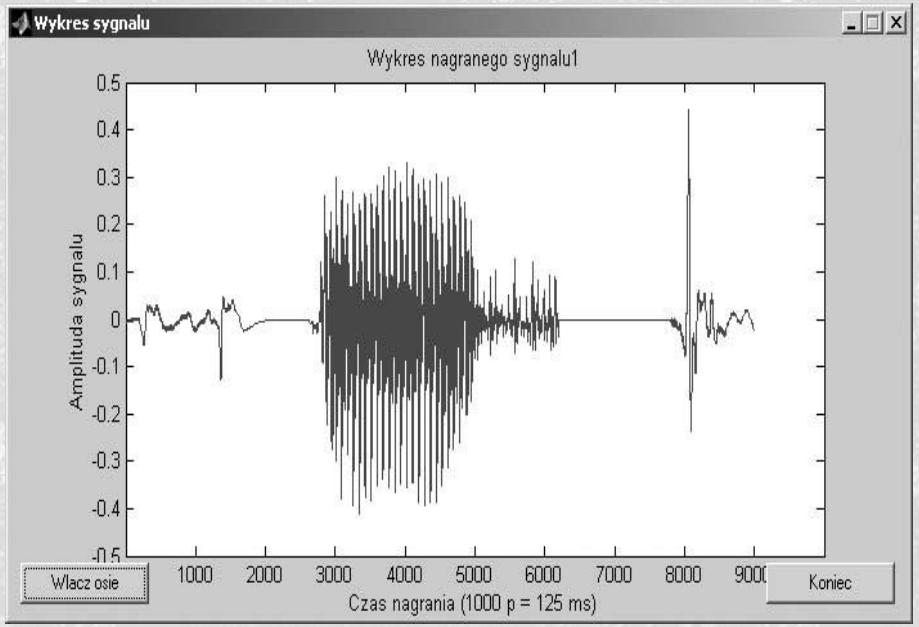
po izolacji

Politechnika Częstochowska

IZOLOWANIE SŁÓW CZĘSTOTLIWOŚĆ



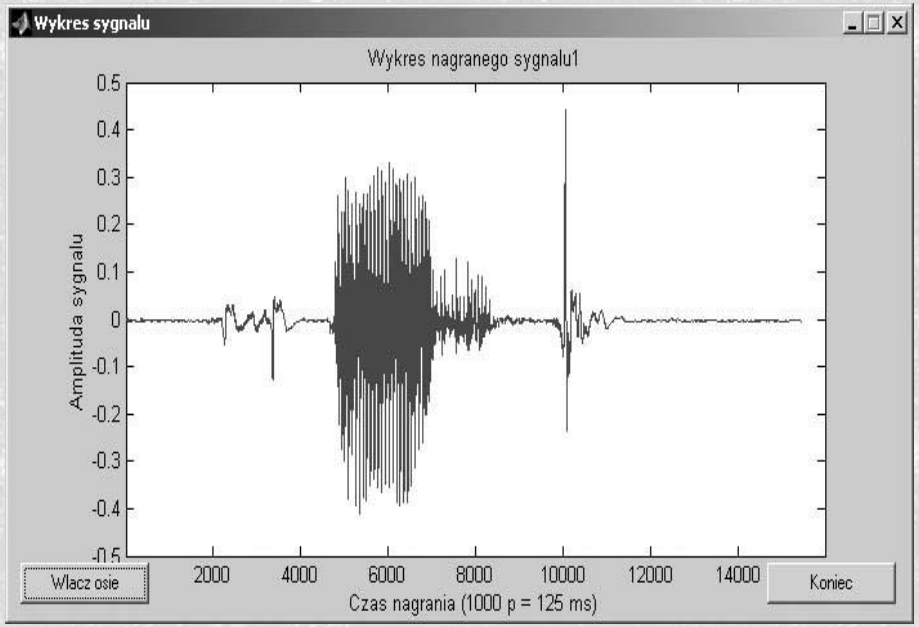
przed izolacją



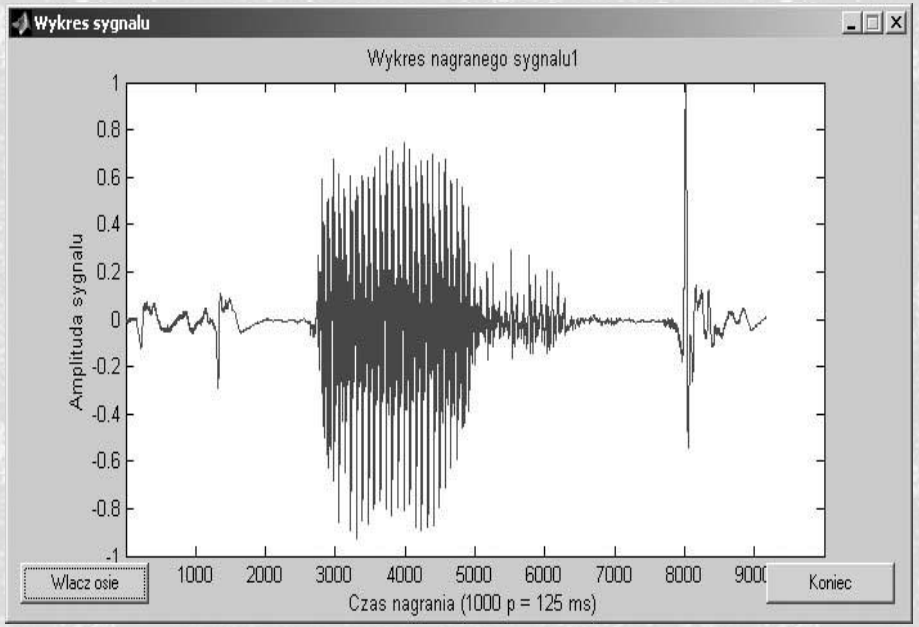
po izolacji

Politechnika Częstochowska

IZOLOWANIE SŁÓW
ENERGIA I CZĘSTOTLIWOŚĆ



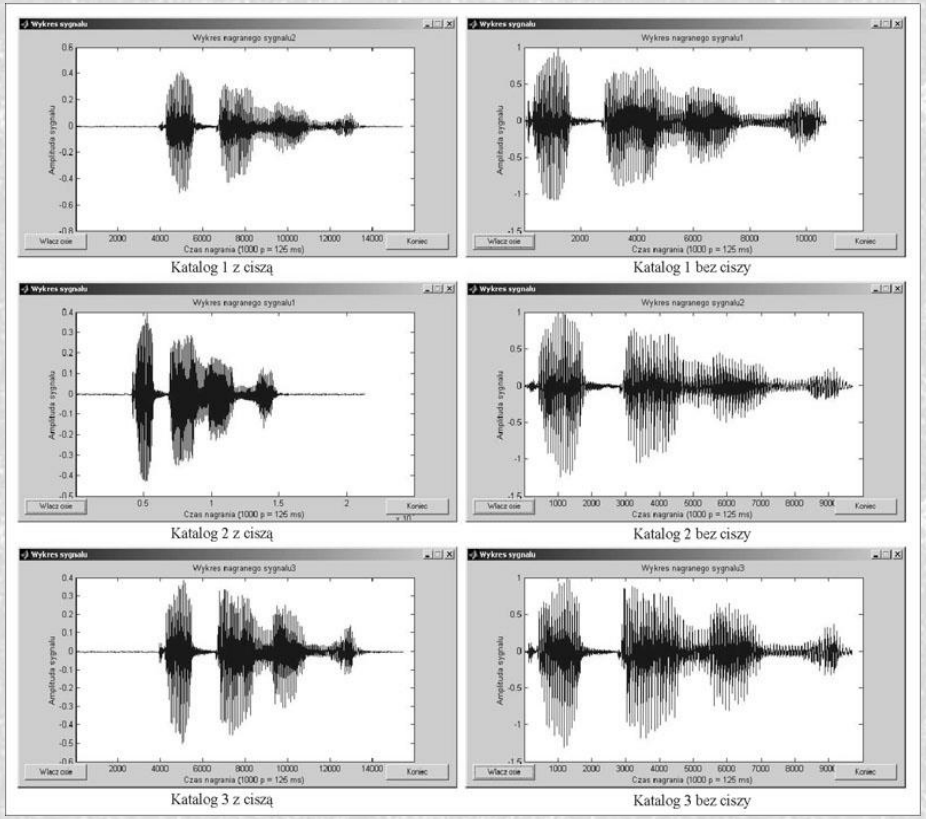
przed izolacją

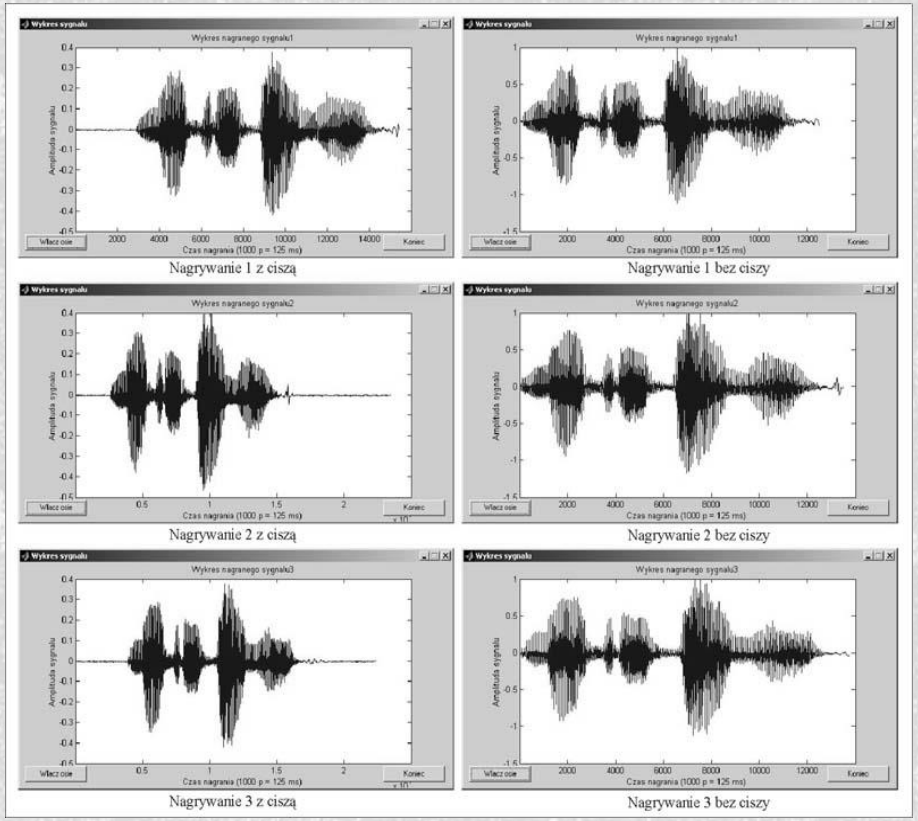


po izolacji

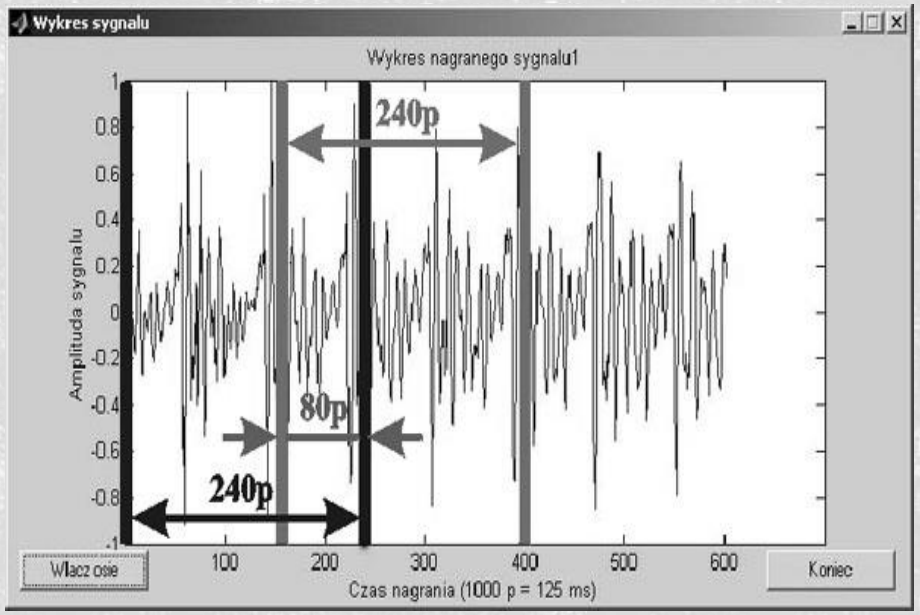
Politechnika Częstochowska

IZOLOWANIE SŁÓW
PRZYKŁADY





PODZIAŁ SYGNAŁU NA STACJONARNE RAMKI



ZALEŻNOŚĆ HAMMINGA

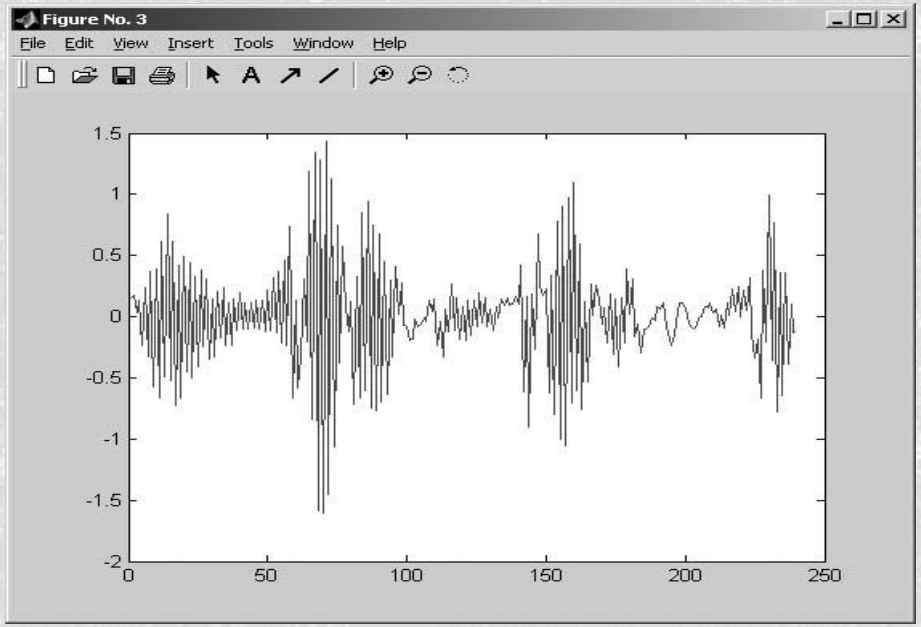
$$w[k+1] = 0.54 - 0.46\cos\left(2\pi\frac{k}{n-1}\right)$$

k **= 0,1,…,*n*-1; *n* – rozmiar okna Hamminga (240)**

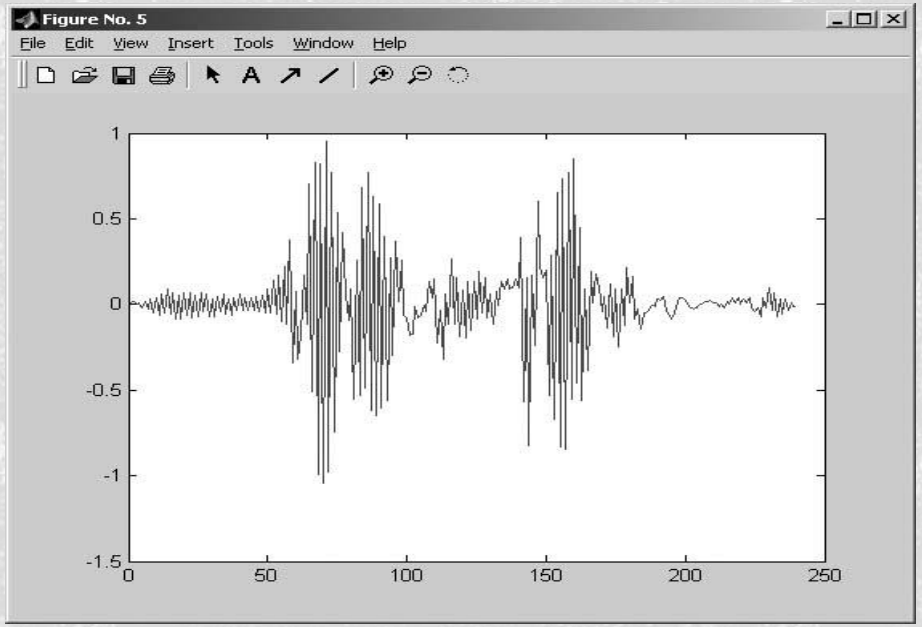
Sposób podziału i sztucznego opóźnienia kolejnych nakładających się ramek

Politechnika Częstochowska

TŁUMIENIE SKRAJNYCH PRÓBEK



przed tłumieniem



po tłumieniu

KODOWANIE SYGNAŁU WSPÓŁCZYNNIKI CEPSTRUM

- Każdą stacjonarną ramkę można kodować przy pomocy szybkiej transformaty Fouriera. W procesie percepcji sygnałów mowy ucho ludzkie dokonuje nieliniowej w dziedzinie częstotliwości analizy widma tego sygnału. W celu dostosowania charakterystyk filtrów skalę częstotliwościową zamienia się na skalę mel za pomocą następującej zależności:

$$f_{mel} = 2595 \log_{10}(1 + f_{Hz} / 700)$$

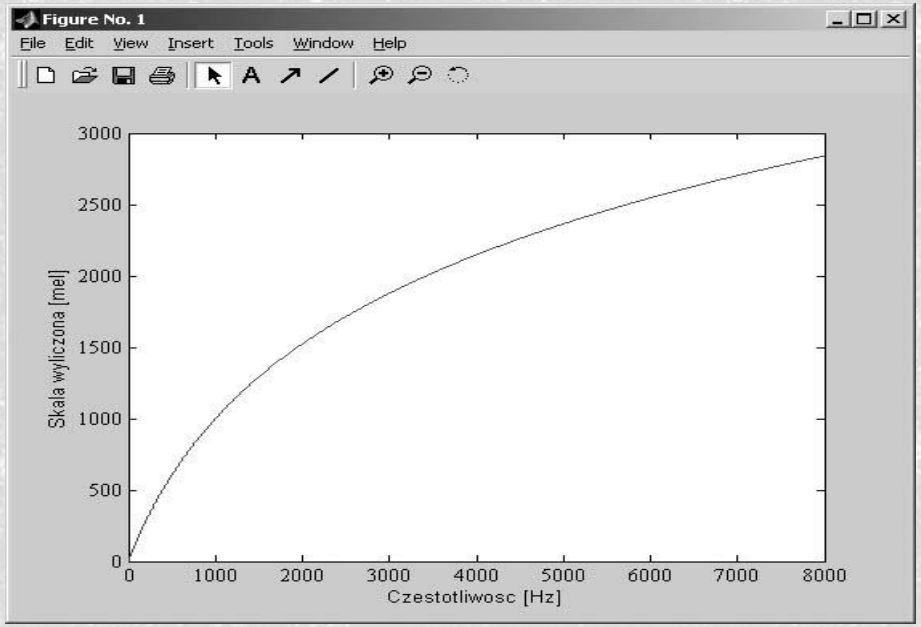
- Przejście ze skali mel do skali częstotliwości opisuje następująca zależność:

$$f_{Hz} = 700\left(10^{\frac{f_{mel}}{2595}} - 1\right)$$

Politechnika Częstochowska

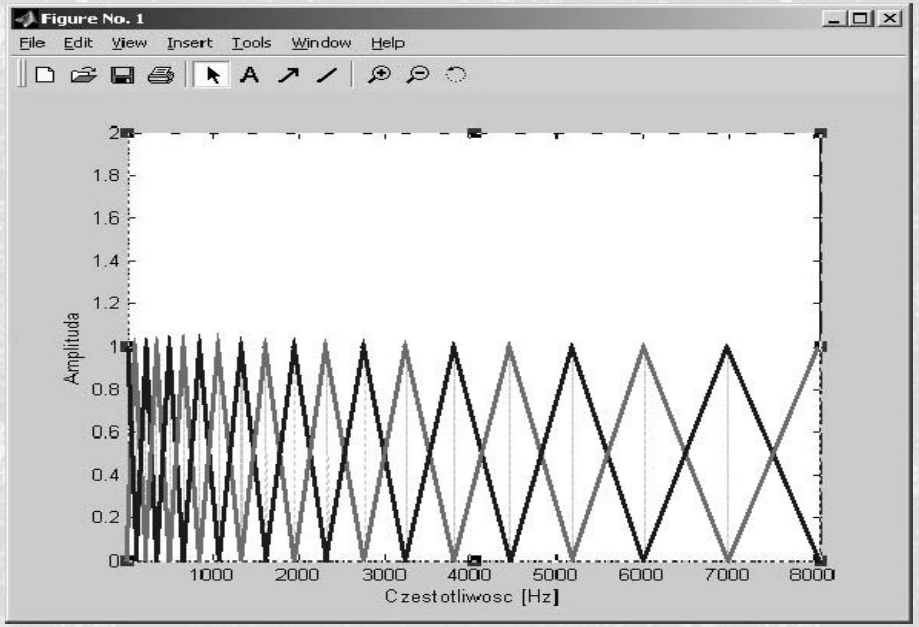
KODOWANIE SYGNAŁU WSPÓŁCZYNNIKI CEPSTRUM

mel = f (Hz)



Związek pomiędzy skalą częstotliwości, a skalą mel

BANK FILTRÓW



Bank 20 filtrów o szerokości 300 meli przesuniętych względem siebie o 150 meli

KODOWANIE SYGNAŁU WSPÓŁCZYNNIKI CEPSTRUM

- **Korzystając z widma mocy można wyznaczyć elementy każdego filtru poprzez pomnożenie jego amplitudy oraz średniego widma mocy odpowiedniej częstotliwości dźwięku wejściowego. Wyznaczone elementy filtrów sumuje się. Całą operację opisuje poniższa zależność:**

$$S_k = \sum_{n=0}^{(N/2)-1} (P_n \cdot A_{k,n})$$

N **– całkowita liczba próbek w ramce;**

Sk **– współczynniki widma mocy**

KODOWANIE SYGNAŁU WSPÓŁCZYNNIKI CEPSTRUM

- **Poprawienie jakości rozpoznawania można uzyskać poprzez zastosowanie przekształcenia cepstralnego parametrów banku filtrów, polegającego na wyznaczeniu współczynników cepstralnych w skali mel (ang. Mel-Frequency Cepstral Coefficients, MFCC) według zależności:**

$$MFCC_n = \sum_{k=1}^{K} (\log S_k) \cos\left[n(k-0.5) \frac{\pi}{K+1} \right], \quad dla\ n = 1 \ldots N,$$

K – liczba wymaganych współczynników cepstrum;

N – liczba filtrów w banku filtrów

POLI TECHNIKA Politechnika Częstochowska

WERYFIKACJA BARWY GŁOSU

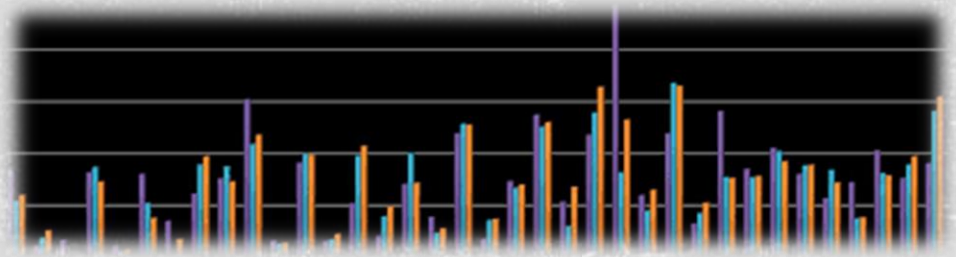
[KI]

- Rozpoznawanie mówcy na podstawie mowy niezależnej polega najczęściej na analizie dość długiego nagrania (najczęściej kilka zdań) pod kątem cech częstotliwościowych wynikających z barwy danego głosu.
- Wymogiem w tego typu systemach jest zarejestrowanie ciągu słów, wypowiadanych w sposób płynny.
- Pomocnym rozwiązaniem może być udostępnianie użytkownikowi do odczytu losowo wygenerowanych sekwencji zdań.

Politechnika Częstochowska

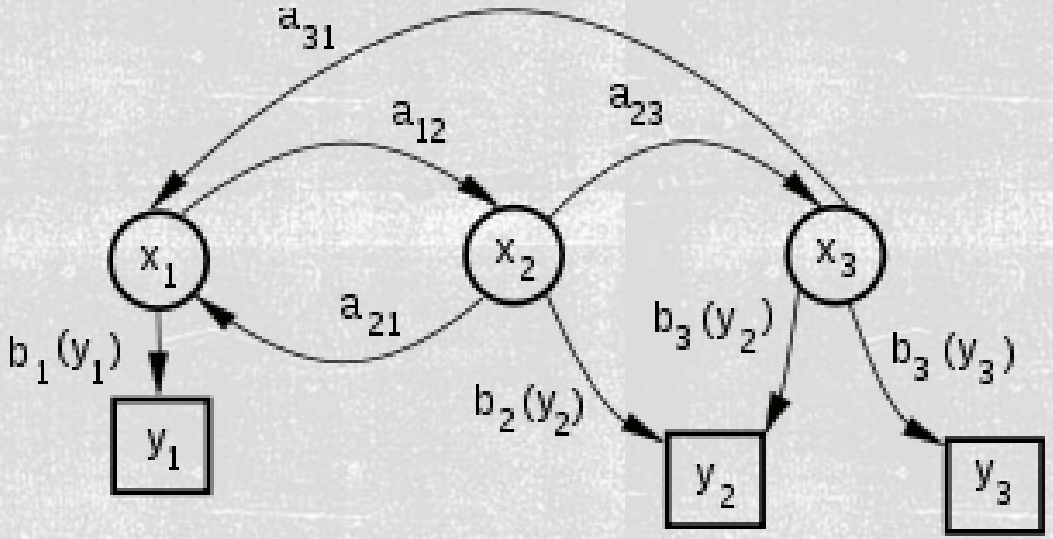
WERYFIKACJA BARWY GŁOSU

- Analiza cech barwy głosu polega na rozkładzie częstotliwościowym danego użytkownika dla każdego z wypowiadanych fonemów.
- Sekwencje zdań powinny być tak dobrane, aby w całości pokryć tzw. przestrzeń akustyczną weryfikowanej osoby.
- Weryfikacja polega na ocenie zgodności poszczególnych cech głosu na różnych poziomach częstotliwościowych.



WERYFIKACJA BARWY GŁOSU

- **W celu udowodnienia przydatności stosowania audio-wideo mowy dla celów weryfikacyjnych, można zbudować taki system, który wykorzysta przydatność sprzężonych ukrytych modeli Markowa.**



Politechnika Częstochowska

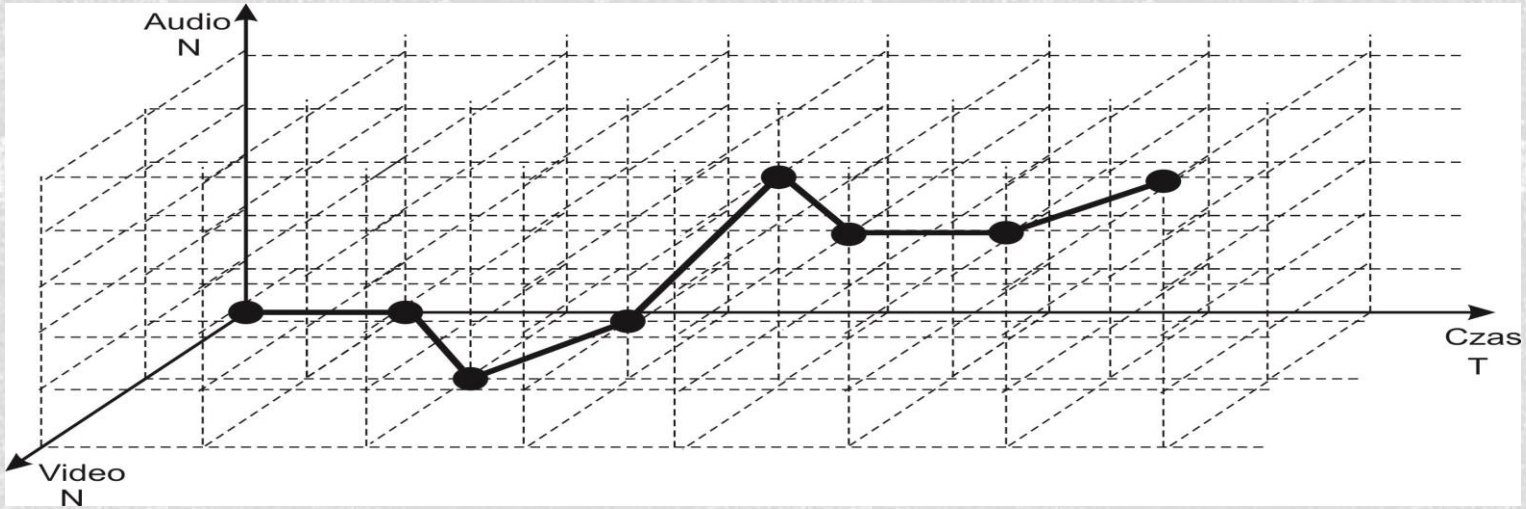
WERYFIKACJA BARWY GŁOSU

- Proste rozwiązania porównujące nagrania tej samej treści polegają na porównywaniu zarejestrowanego hasła z weryfikowanym na poziomie pliku.

- Bardziej zaawansowana jest analiza tej samej wypowiadanej treści, ale przez różnych użytkowników.

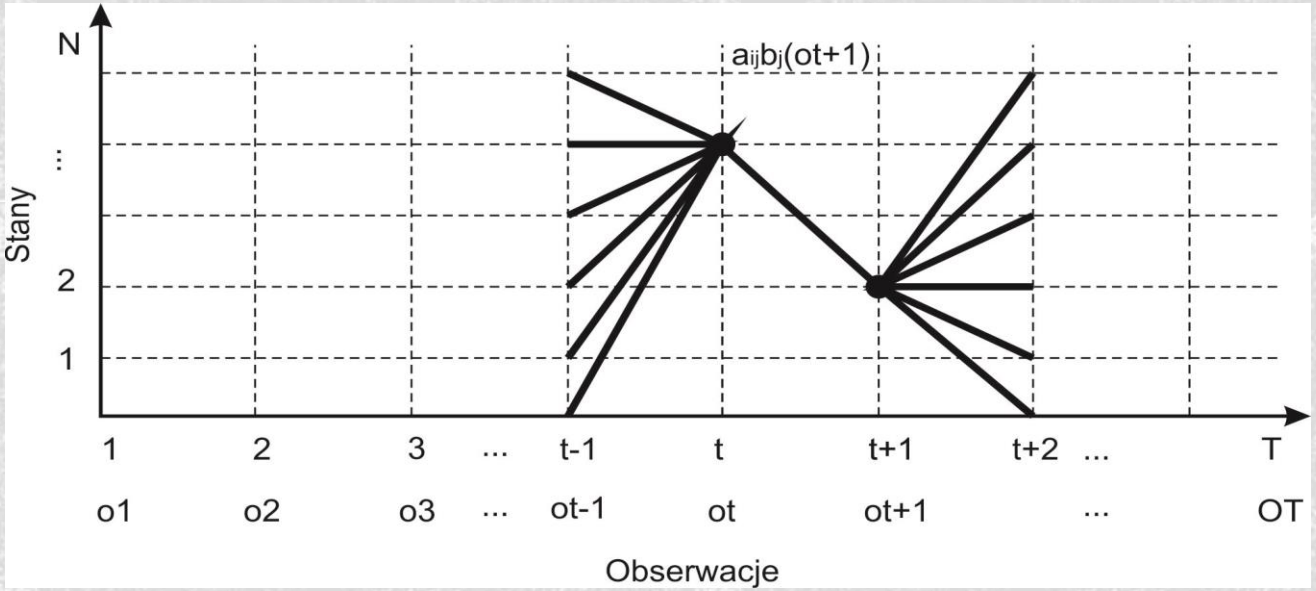
WERYFIKACJA BARWY GŁOSU

- **Takie podejście pozwala na optymalne wykorzystanie sprzężonych ukrytych modeli Markowa oraz na udowodnienie przydatności stosowania wideo mowy również podczas weryfikacji tożsamości, niekoniecznie przy zakłóconym sygnale audio mowy.**



WERYFIKACJA BARWY GŁOSU

- **W tym miejscu należy zaznaczyć, iż tego typu systemu weryfikujące dedykowane są najczęściej dla niewielkiej liczby użytkowników, ze względu na sporą zmienność w czasie analizowanych cech.**

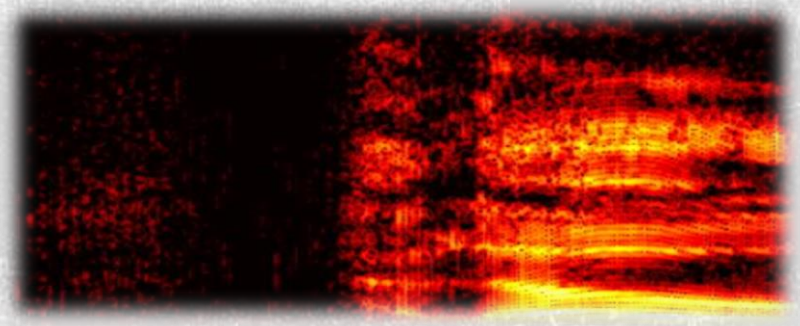


Politechnika Częstochowska

WERYFIKACJA BARWY GŁOSU

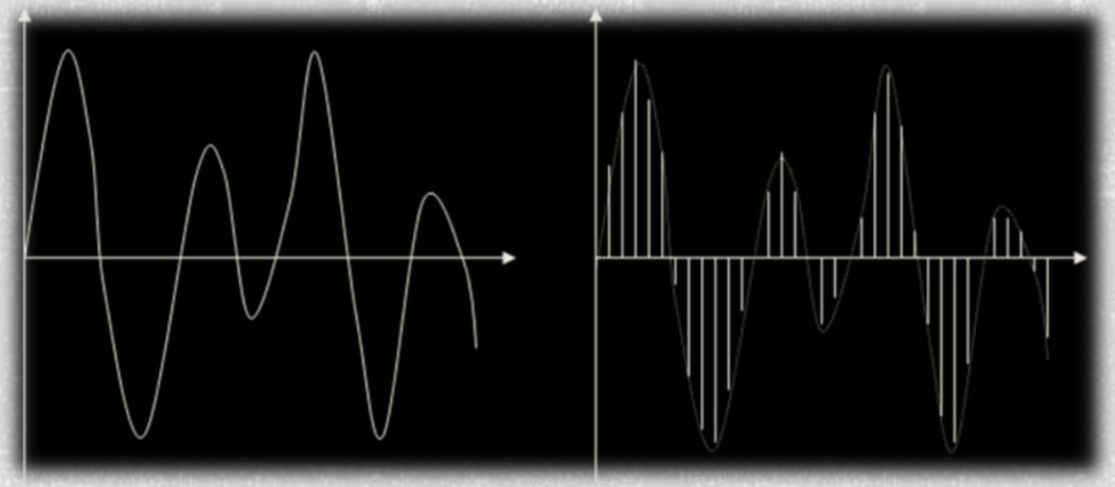
- Metoda weryfikacji tożsamości na podstawie znanej treściowo mowy audio i wideo polega na zarejestrowaniu wypowiadanej tej samej komendy przez wszystkich użytkowników.

- Następnie na podstawie wszystkich nagrań tworzona jest przestrzeń akustyczna, wspólna dla rejestrowanych w bazie użytkowników.



WERYFIKACJA BARWY GŁOSU

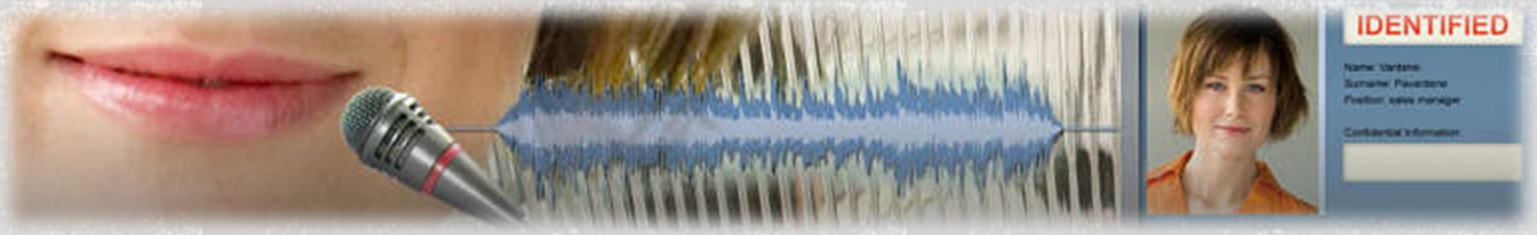
- Użytkownik podczas rejestracji wypowiada podaną komendę kilka razy, z czego jedno z nagrań trafia do puli nagrań tworzących książkę kodową, a pozostałe do nauczenia sprzężonych ukrytych modeli Markowa.



Politechnika Częstochowska

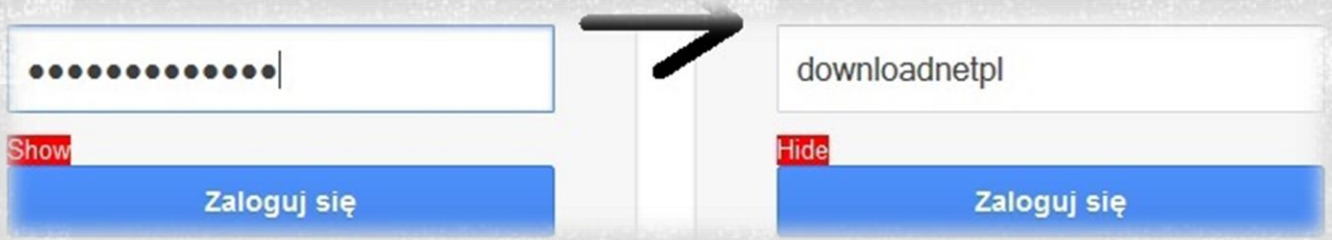
WERYFIKACJA BARWY GŁOSU

- Dla każdego użytkownika budowany jest oddzielny model.
- Po nauczeniu, podczas już normalnej pracy z systemem każdorazowo przy próbie dostępu, następuje weryfikacja tożsamości zainteresowanego użytkownika.



WERYFIKACJA BARWY GŁOSU

- Zwycięski model wskazuje na poprawną (bądź też nie) weryfikację.
- W celu zwiększenia wiarygodności można wymusić próbę podwójnego logowania i dla obu zgodności przydzielać dopiero status poprawnej weryfikacji.



Politechnika Częstochowska

Projekt finansowany w ramach programu Ministra Nauki i Szkolnictwa Wyższego pod nazwą „Regionalna Inicjatywa Doskonałości" w latach 2019 - 2023 nr projektu 020/RID/2018/19 kwota finansowania 12 000 000 PLN

Politechnika Częstochowska

KI

Dziękuję za uwagę

dr hab. inż. Mariusz Kubanek, prof. PCz
mariusz.kubanek@icis.pcz.pl
Katedra INFORMATYKI