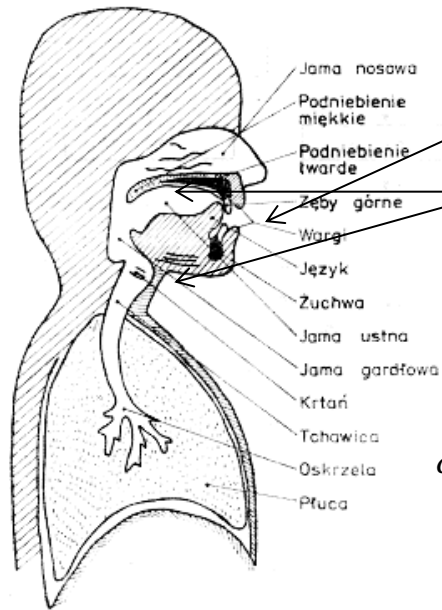


# **Ukryte Modele Markowa w rozpoznawaniu słów**



# Parametryzacja sygnału mowy



$$X(k) = G(k) \cdot H(k)$$

Widmo

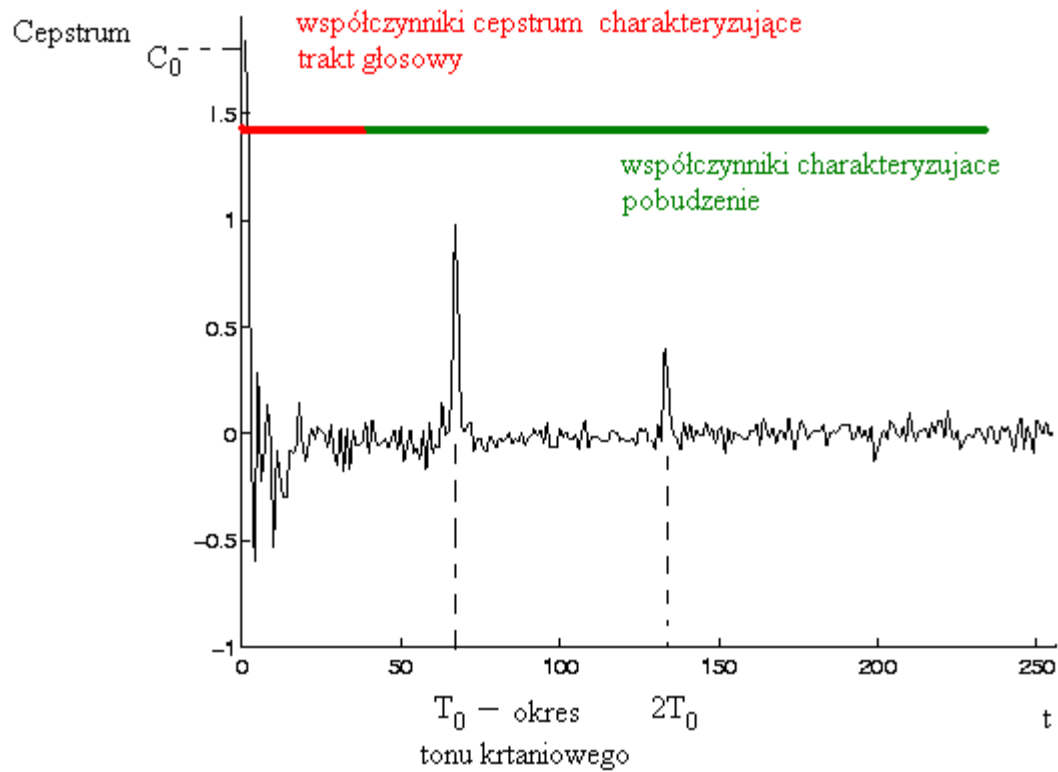
$$\log |X(k)| = \log |G(k)| + \log |H(k)|$$

Widmo ampl. [dB]

$$c(n) = \mathfrak{F}^{-1} \log |X(k)| = \mathfrak{F}^{-1} \log |G(k)| + \mathfrak{F}^{-1} \log |H(k)|$$

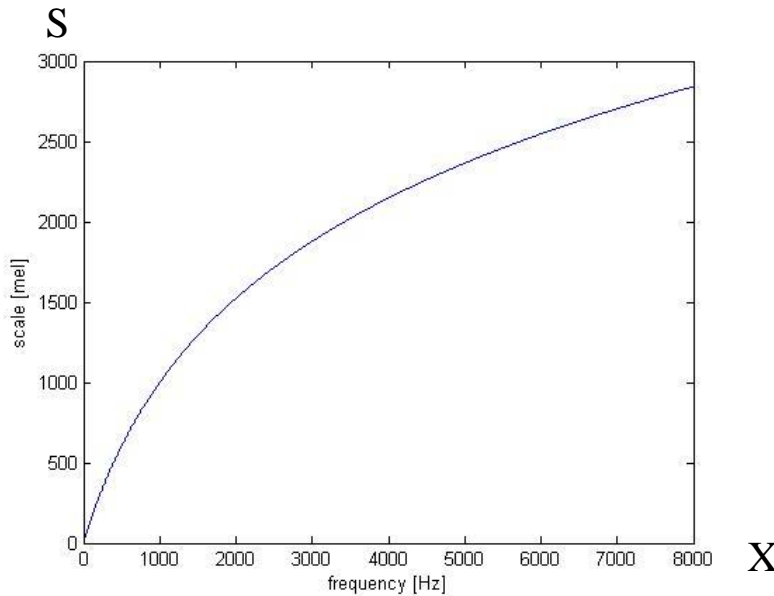
cepstrum

# Parametryzacja sygnału mowy - cepstrum



# Parametryzacja sygnału mowy – mel-cepstrum

„przepróbkowanie” widma:  $X(k) \rightarrow S(k)$  (skala melowa)



Mel-cepstrum:  $MFCC(n) = F^{-1} [\log |S(k)|]$

Co 5-10 ms mierzymy kilkanaście współczynników MFCC:

Tworzą one wektory obserwacji

$O_1, O_2, \dots, O_T$

# Modele obserwacji

## Model ciągły (np. gaussowski)

$$b_i(o) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_i|}} \exp\left[-\frac{1}{2}(o - \mu_i)' \Sigma_i^{-1} (o - \mu_i)\right]$$

gdzie  $o$  - wektor obserwacji,  $n$  - wymiar wektora obserwacji  $o$

$\mu_i$  -  $n$ -wymiarowa wektorowa wartość średnia charakteryzująca obserwacje w stanie  $S_i$ ,

$\Sigma_i$  - macierz kowariancji (o wymiarze  $n \times n$ ) charakteryzująca obserwacje w stanie  $S_i$

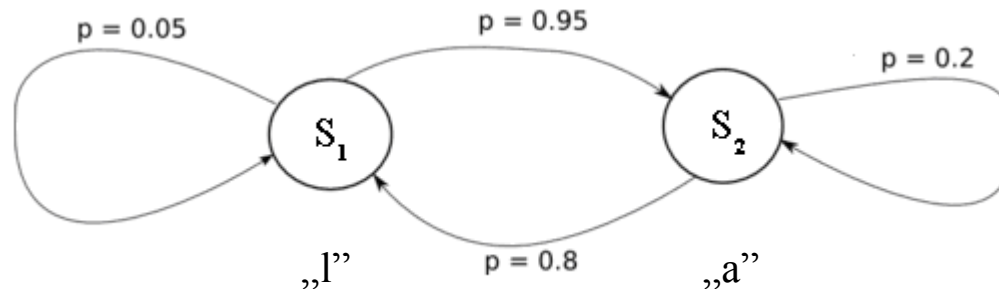
**Model dyskretny:** kwantujemy wyniki obserwacji (kwantyzator wektorowy) i podajemy prawdopodobieństwa wystąpienia każdego z nich:

$$b_i(o^j)$$

„ $i$ ” jest indeksem modelu (np. określoną głoską), a „ $j$ ” identyfikuje jeden ze skwantowanych wektorów obserwacji

# HMM – dynamiczny model ciągu obserwacji

Np. „lalalalala...”



Kolejne obserwacje dokonane podczas wypowiedzania danej frazy są przypisywane określonym stanom, tworząc ukryty ciąg stanów:

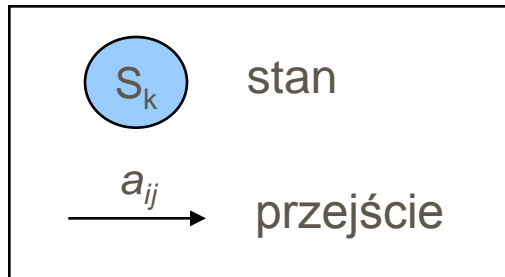
$$Q = [q(1), q(2), \dots, q(t), \dots, q(T)]$$

gdzie  $q(t)$  jest jednym z  $N$  stanów:  $S_1, S_2$  lub  $S_N$

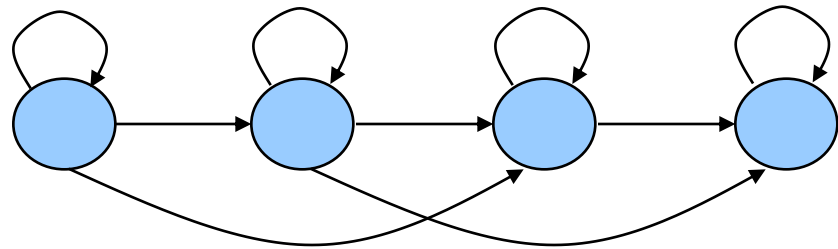
Np. Dla „lalala”,  $T=6$ ,  $N=2$ ,  $Q=[S_1, S_2, S_1, S_2, S_1, S_2]$

HMM  $\lambda$  jest opisany następująco:  $\lambda = [A, B, \pi]$ , gdzie  $A$  – macierz przejść między stanami,  $B$  – modele (po jednym dla każdego stanu),  $\pi$  – prawdopodobieństwa startu (po jednym dla każdego stanu)

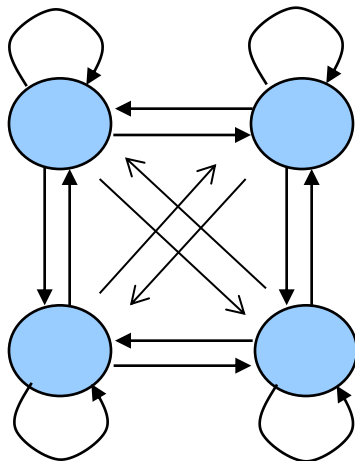
# Topologie HMM



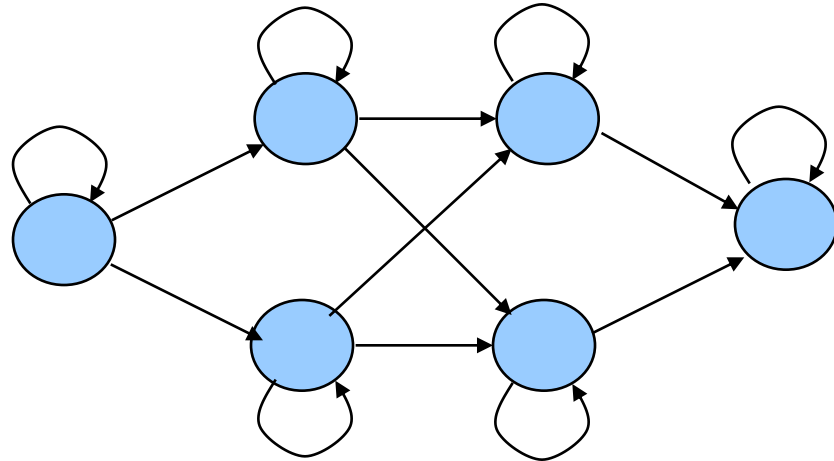
Bakisa



ergodyczny

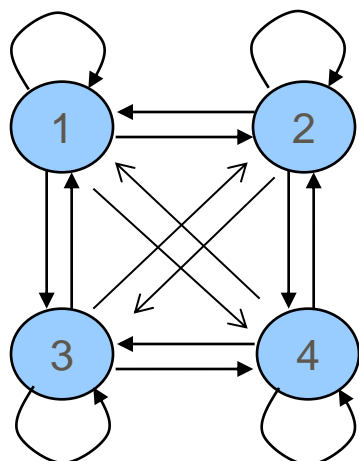


równoległo-szeregowy



# Opis HMM

ergodyczny



$A$  – macierz prawdopodobieństw przejść między stanami

$B$  – macierz prawdopodobieństw wystąpienia zdarzenia w danym w stanie

$\Pi$  – wektor prawdopodobieństwa inicjalizacji

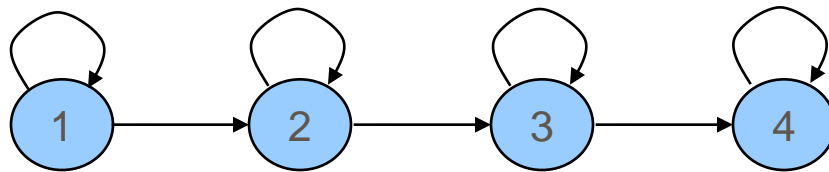
$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \quad \Sigma = 1$$

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \\ b_{51} & b_{52} & b_{53} & b_{54} \end{bmatrix} \quad \Sigma = 1$$

$$\Pi = \underbrace{[\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4]}_{\Sigma = 1}$$



# HMM szeregowy (Bakisa 1-go rzędu)



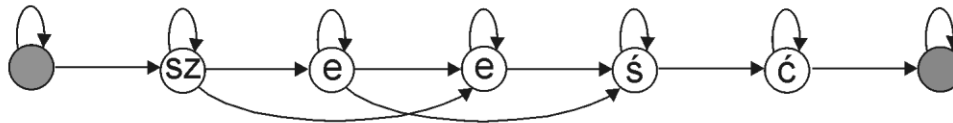
$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \\ b_{51} & b_{52} & b_{53} & b_{54} \end{bmatrix}$$

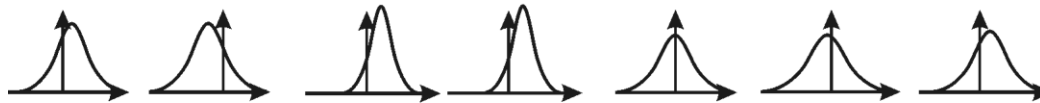
$$\Pi = [1 \ 0 \ 0 \ 0]$$

# HMM w rozpoznawaniu słów

stany i przejścia



modele obserwacji (CHMM) - ciągłe modele gaussowskie



kryterium rozpoznawania:

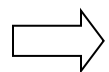
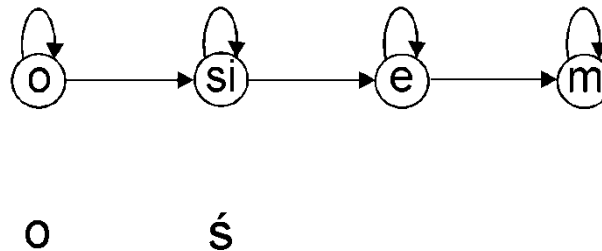
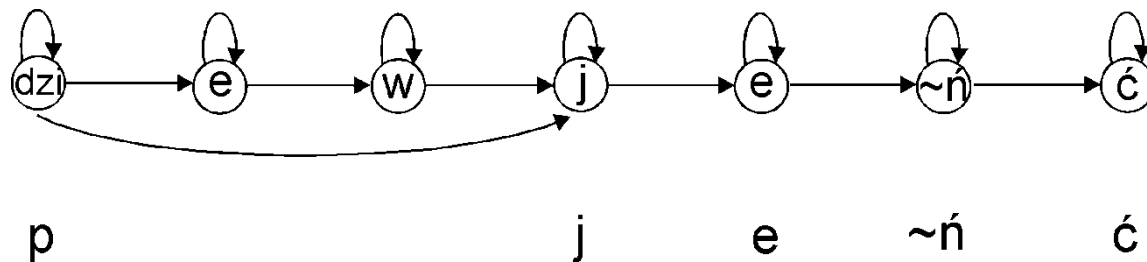
$$\max_i p[X|\text{HMM}(i)]$$

X - ciąg obserwacji

kryterium projektowania HMM(i):

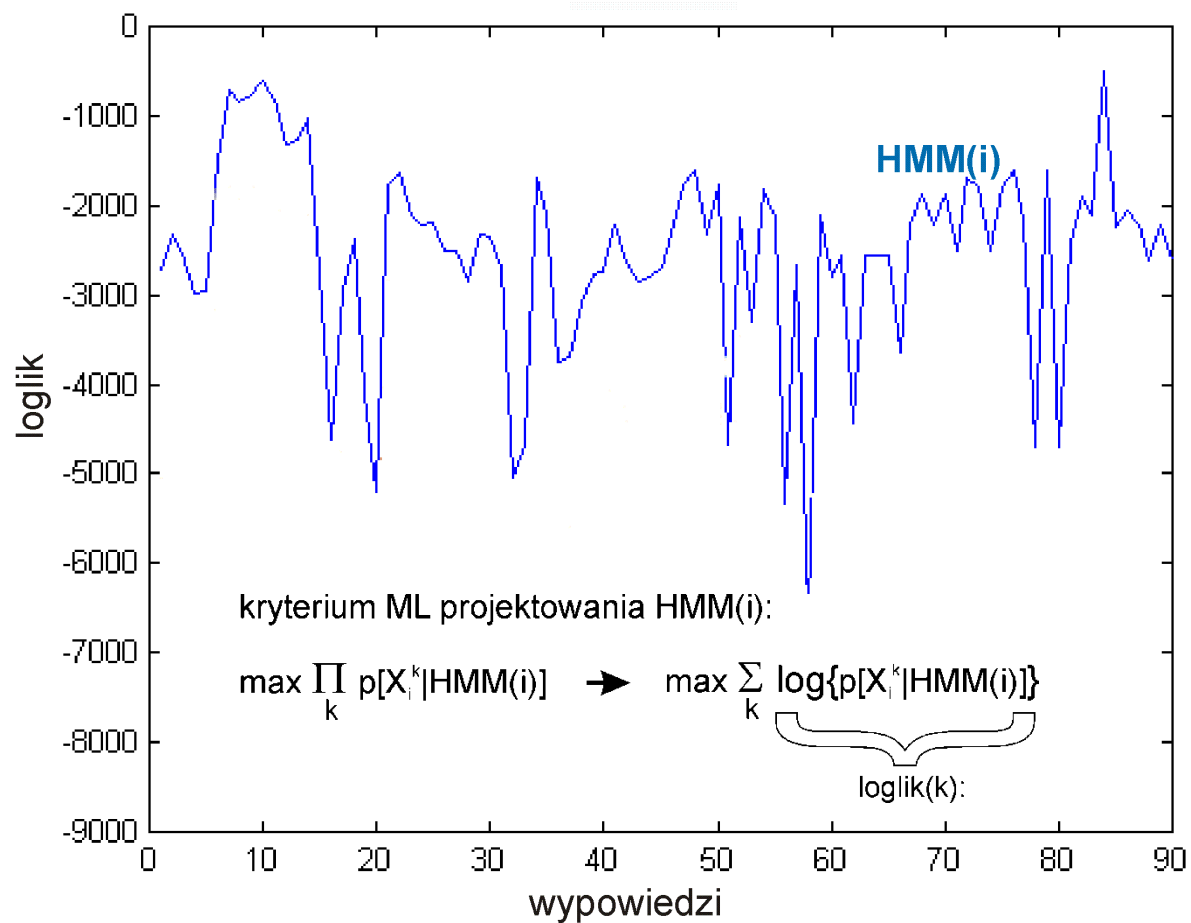
$$\max \prod_k p[X_i^k|\text{HMM}(i)] \rightarrow \max \sum_k \log\{p[X_i^k|\text{HMM}(i)]\}$$

# Wpływ struktury HMM na rozpoznawanie słów

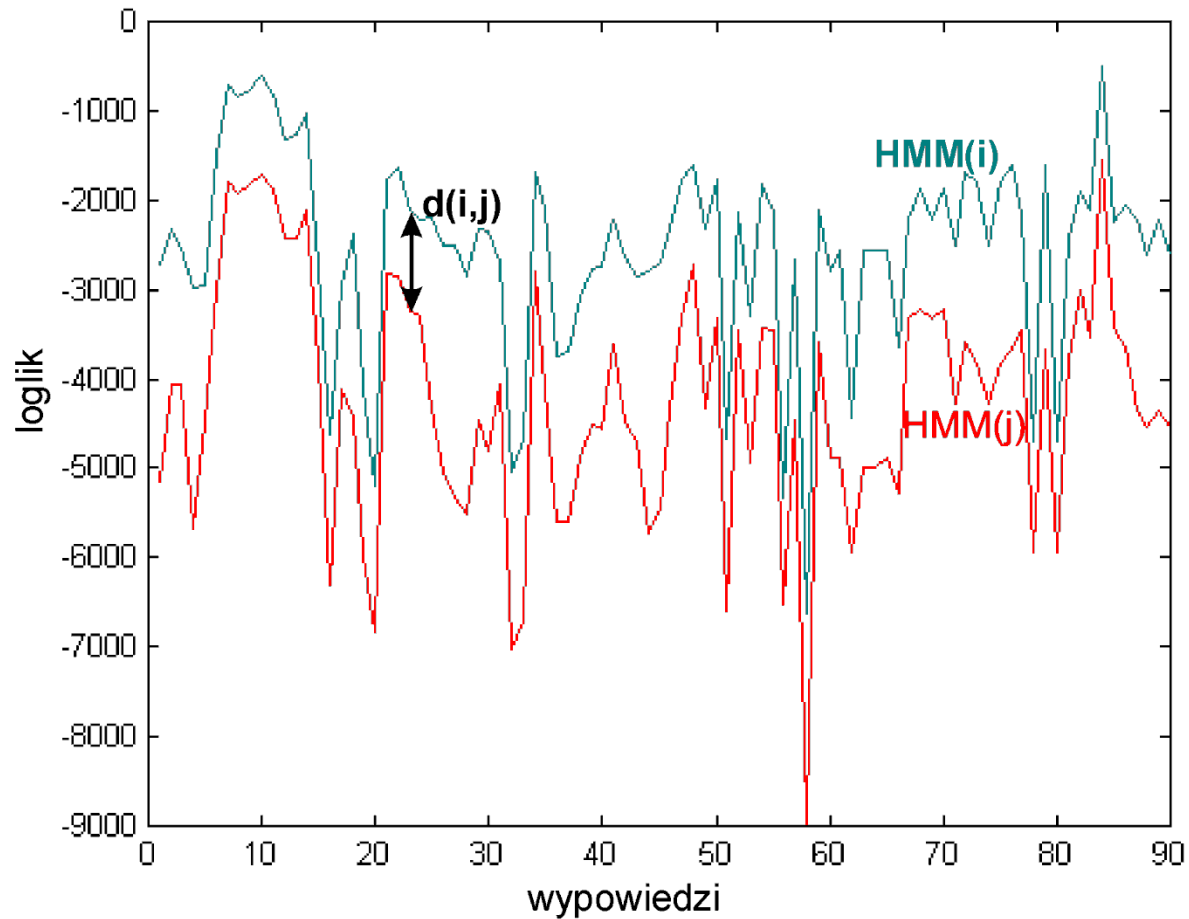


Dobre właściwości dyskryminacyjne ma struktura szeregową z wymuszeniem dojścia do ostatniego węzła

# Projektowanie HMM – kryterium ML (maximum likelihood)

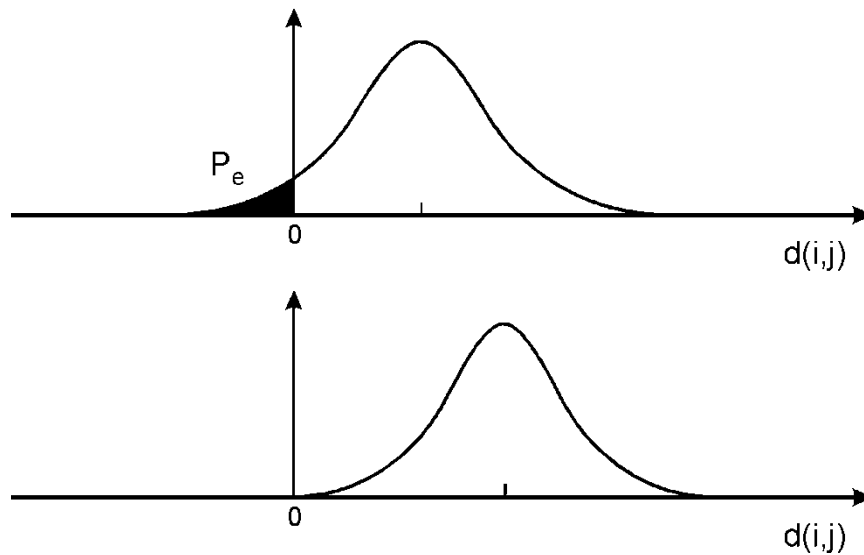


# Właściwości dyskryminacyjne HMM



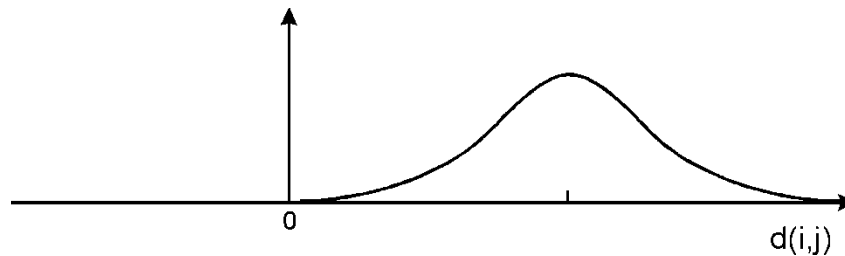
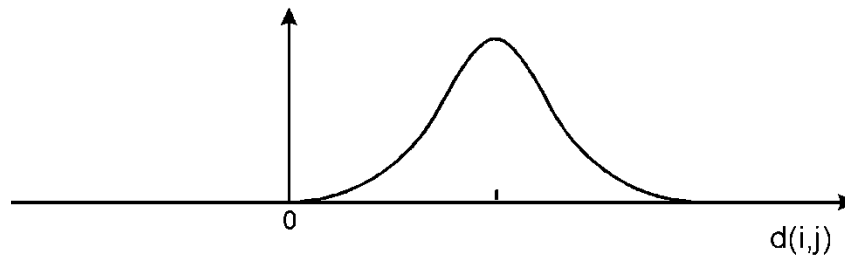
# Dyskryminacyjne metody projektowania HMM

## Minimum błędów klasyfikacji



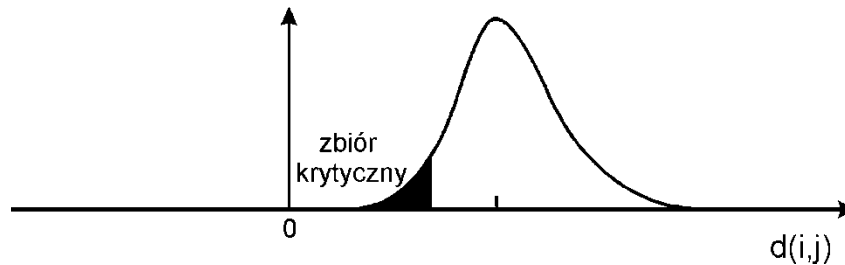
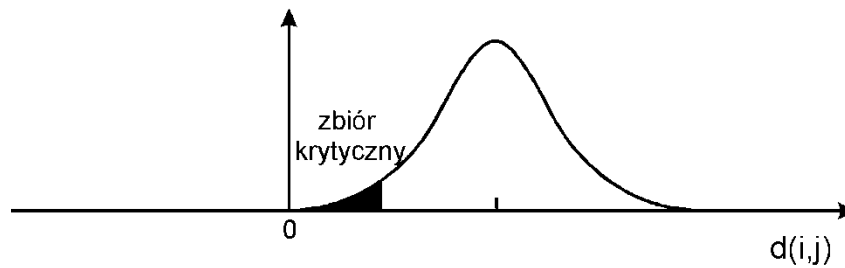
# Dyskryminacyjne metody projektowania HMM

Maksimum prawdopodobieństwa warunkowego



# Dyskryminacyjne metody projektowania HMM

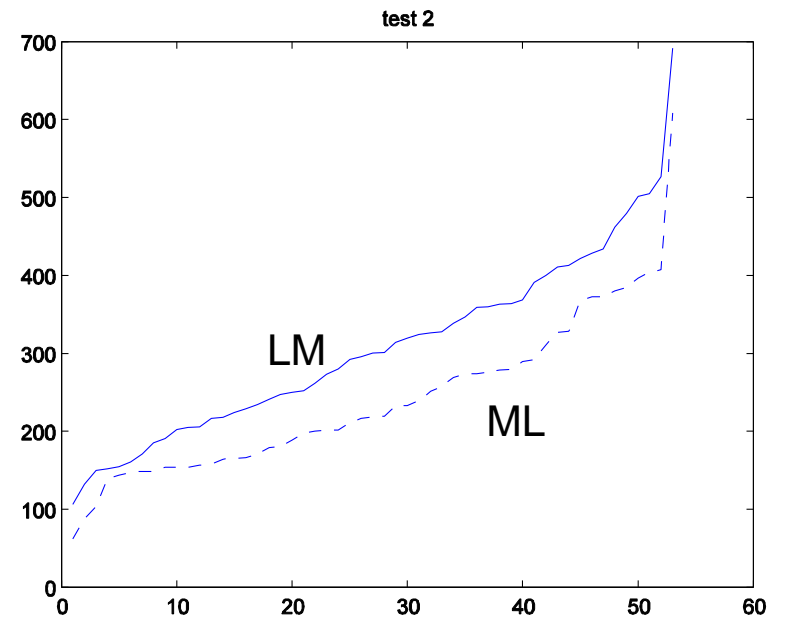
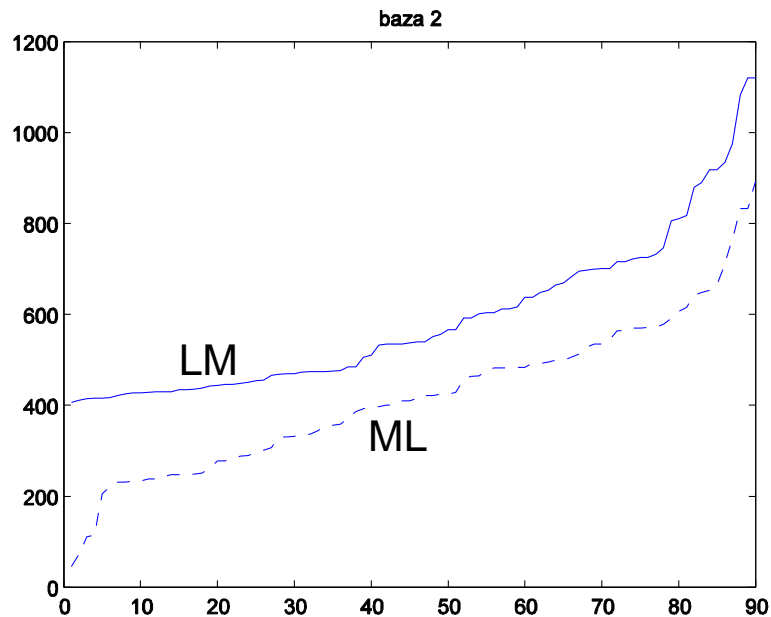
Maksimum odstęp (LM – large margin)





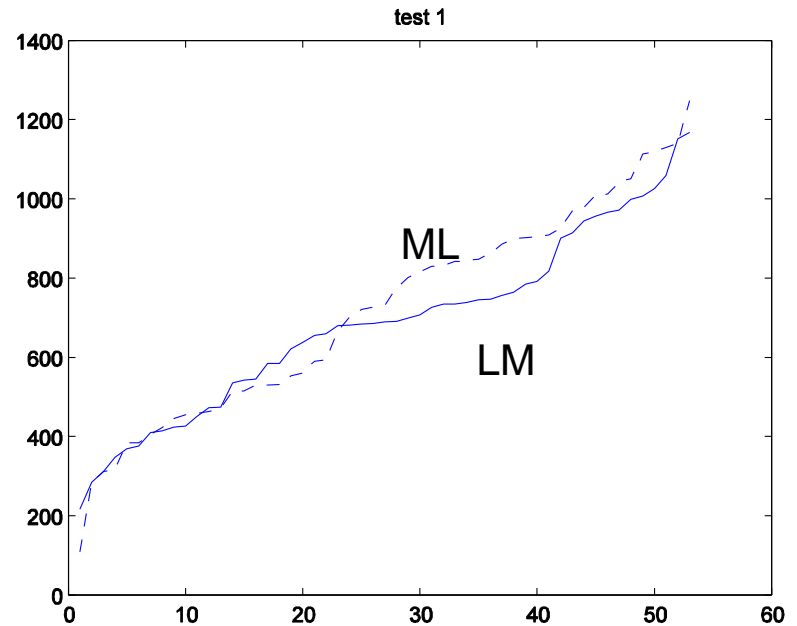
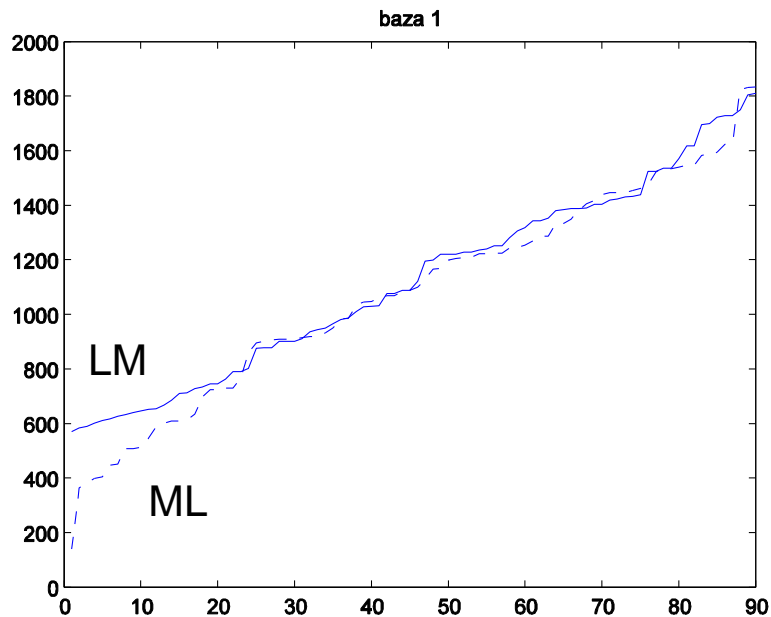
# Porównanie metod ML i LM

Wypowiedzi słowa „pięć”, uporządkowane rosnąco względem  $d(i,j)$ ,  
 $i = \text{'pięć'}$ ,  $j = \text{'dziewięć'}$



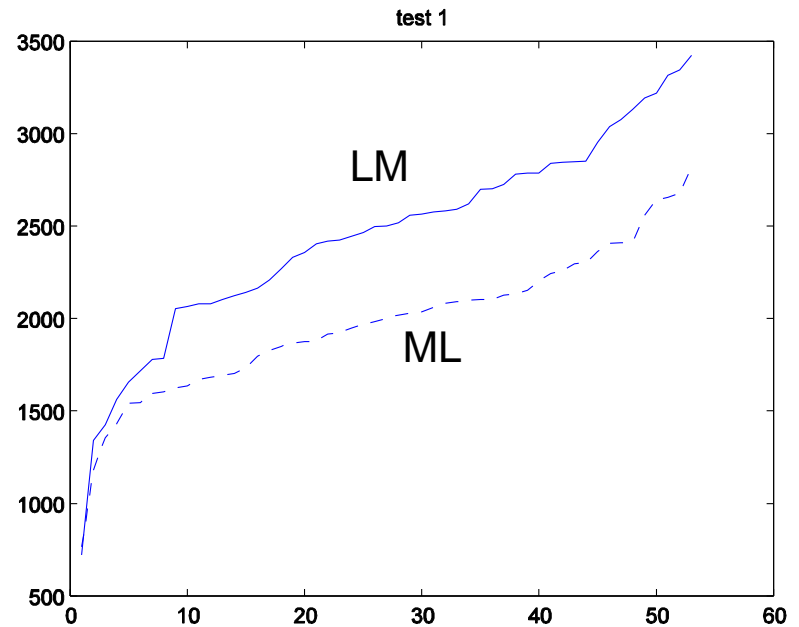
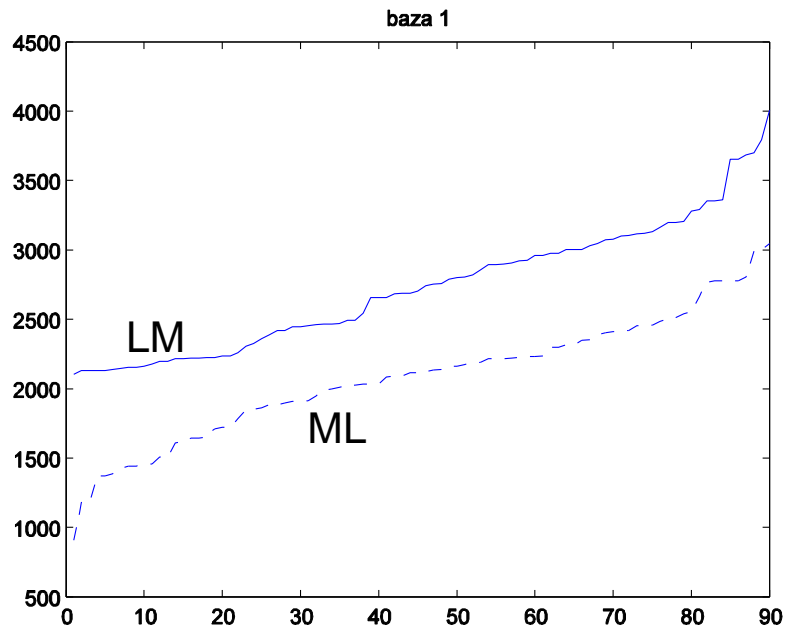
# Porównanie metod ML i LM

Wypowiedzi słowa „dziewięć”, uporządkowane rosnąco względem  $d(i,j)$ ,  
 $i = \text{'dziewięć'}$ ,  $j = \text{'pięć'}$



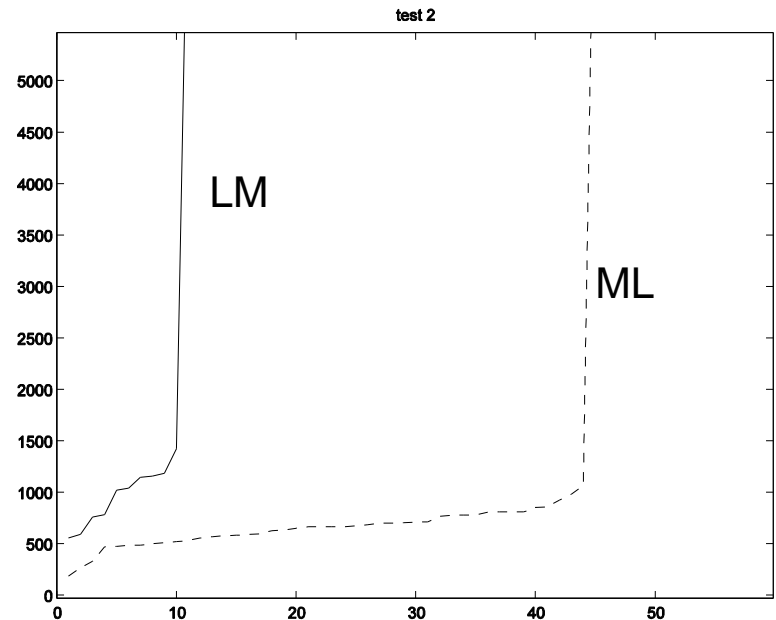
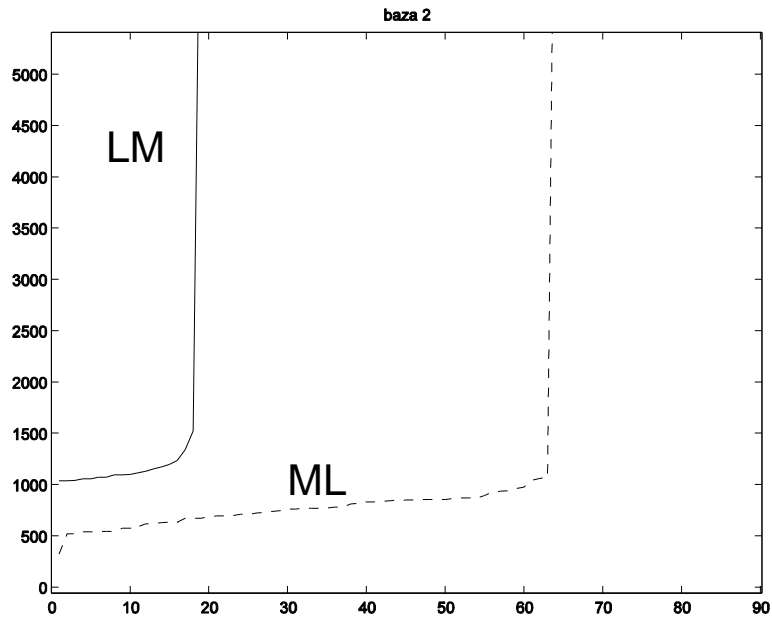
# Porównanie metod ML i LM

Wypowiedzi słowa „osiem”, uporządkowane rosnąco względem  $d(i,j)$ ,  
 $i = \text{'osiem'}$ ,  $j = \text{'oś'}$



# Porównanie metod ML i LM

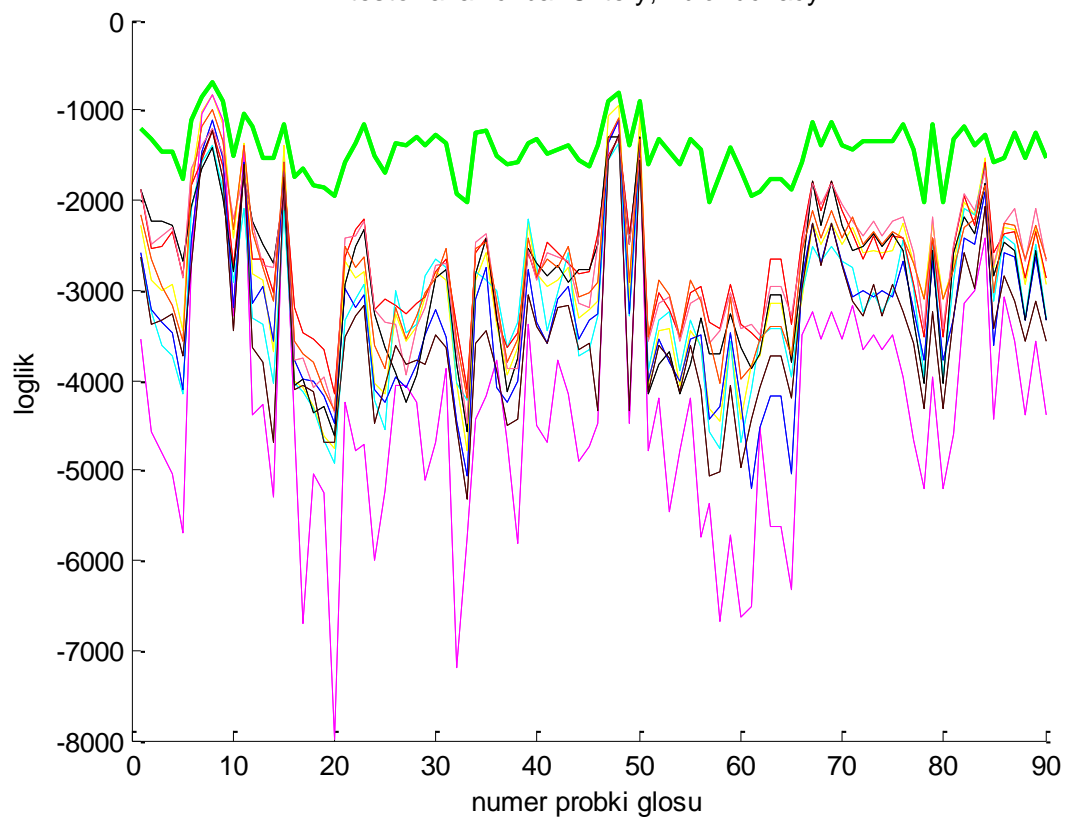
Wypowiedzi słowa „oś”, uporządkowane rosnąco względem  $d(i,j)$ ,  
 $i = \text{'oś'}, j = \text{'osiem'}$



# Rozpoznawanie wypowiedzianych cyfr

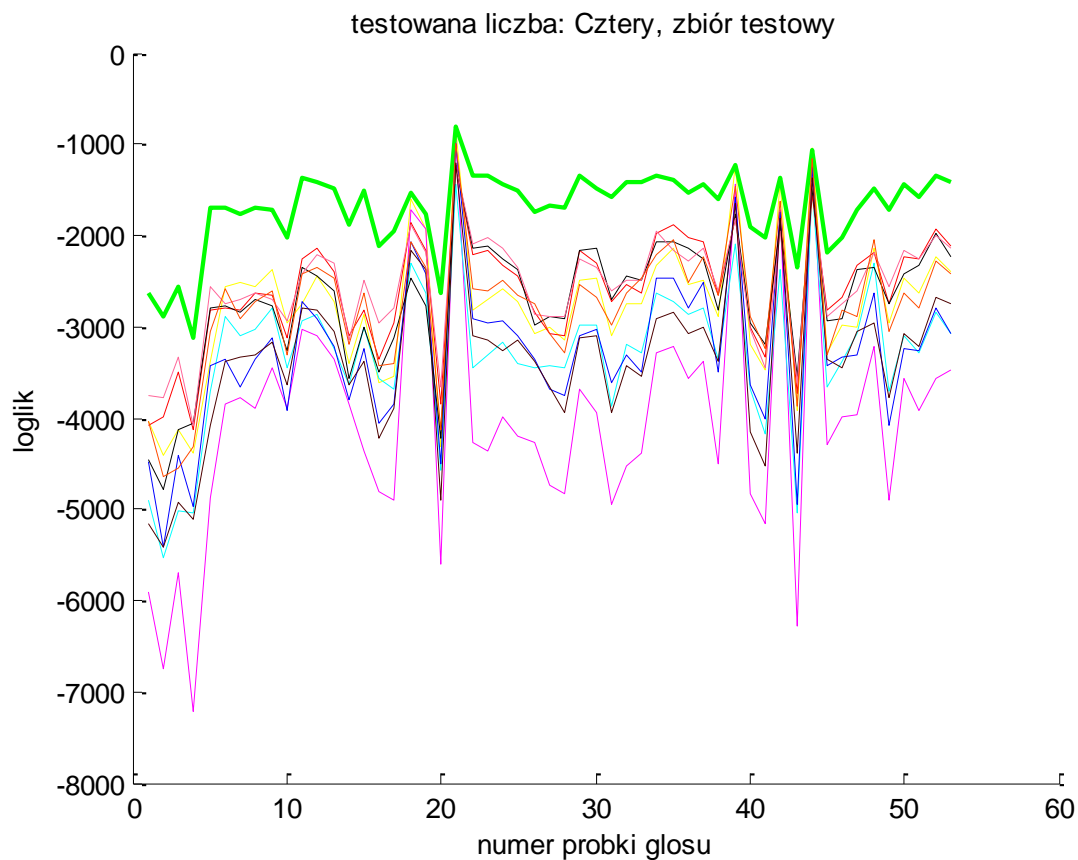
baza ucząca

testowana liczba: Cztery, zbiór uczacy



# Rozpoznawanie wypowiedzianych cyfr

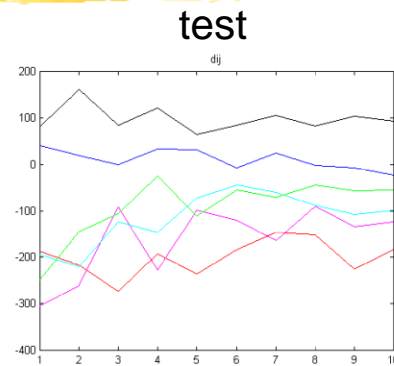
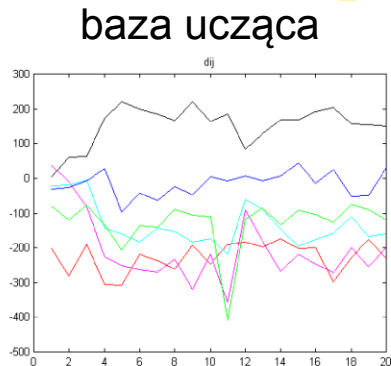
test



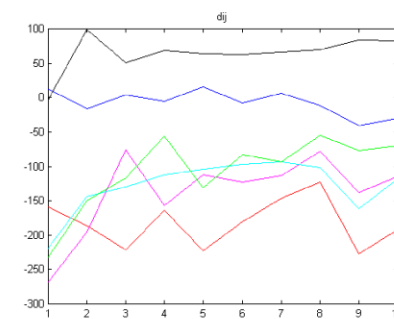
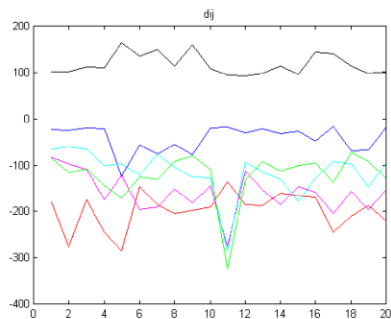
# Rozpoznawanie mówców

## wypowiedzi mówcy M3

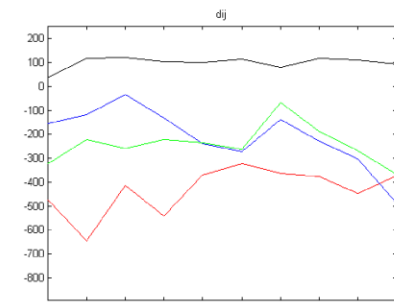
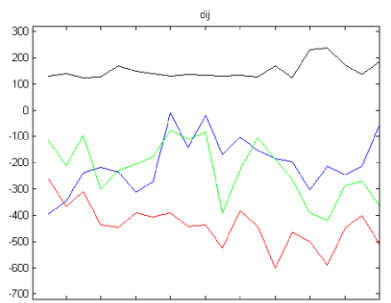
**ML HMM  
18 CC**



**LM HMM  
18 CC**



**LM HMM  
18 CC +T0**



← **HMM(M3)**

← **HMM(M3)**

← **HMM(M3)**

Praca mgr. Marta Korbin