# Sprawozdanie - Technologia Mowy

### Temat: Rozpoznawanie mowy za pomocą HTK

*Anna Bartnik*

*Inżynieria Akustyczna, III rok*

*Grupa: 9:30*

## 1. Wstęp

Celem poniższego raportu jest prezentacja wytworzonego przy pomocy Hidden Markov ToolKit rozpoznawcza mowy. Projekt realizowany był w ramach zajęć laboratoryjnych z przedmiotu Technologia Mowy.

## 2. Gramatyka

Wygenerowany rozpoznawacz mowy opiera swój system decyzyjny na uprzednio przygotowanej gramatyce. Składnia ta ma na celu rozpoznanie do którego teatru (miejsce, rząd) i na jaką porę użytkownik chce zamówić bilet (symulacja kasy biletowej). Zarówno słownik jak i gramatyka zastosowana w programie są zgodne z regułami języka polskiego. Podstawowymi zaletami zaproponowanych reguł jest ich prostota wykonania przez użytkownika. Poniżej przedstawiono wygląd gramatyki i słownika zapisanych w plikach tekstowych .txt

W trakcie pracy okazało się jednak, że słownik wymaga poprawek (niekoniecznie zgodnych z regułami zapisu fonetycznego). Była to niejedyna forma zniwelowania błędów, jednak najszybsza. Jej wadą jednak może być inwazyjność, z jaką mogła niekorzystnie wpłynąć na wyniki rozpoznawania.

*gram:*

```
$teatr = stary | bagatela |stu | groteska;

$sala = duZA | maLA | kameralnA;

$rzad = gOrny | dolny;

$miejsce = pierwsze | drugie | trzecie | czwarte |
szOste | siOdme | Osme | dziewiAte | dziesiAte;

$pora = w_piAtek [rano | wieczOr]  |
        w_sobotE [rano | wieczOr]  |
       w_niedzielE [rano | wieczOr];


( SENT-START ( wybieram teatr <$teatr>  sale <$sala
<$rzad>  miejsce <$miejsce> <$pora> ) SENT-END )
```

*dict:*

```
wybieram w y b j e r a m
teatr t e a t r
sale s a l e
rzad r z o n t
miejsce   m j e j s e
starys t a r y
bagatela  b a g a t e l a
stu  s t u
groteska  g r o t e s k a
duZA d u r z o m
maLA m a u o m
kameralnA  k a m e r a l n o m
gOrny     g u r n y
dolny d o l n y
pierwsze  p j e r f s z e
drugie    d r u g j e
trzecie   t s z e i e
czwarte   s z f a r t e
piAte     p j o n t e
szOste    s z u s t e
siOdme    s i u d m e
Osme u s m e
dziewiAte d z i e w j o n t e
dziesiAte d z i e s i o n t e
w_piAtek  w p j o n t e k
w_sobotE  w s o b o t e
w_niedzielE     w n e d z i e l e
rano r a n o
wieczOr   w j e s z u r
SENT-END  [] sil
SENT-START [] sil
```

## 3. Dane treningowe i testowe

Nagranie treningowe, zdeterminowane przez wygenerowany ciąg zdań trenujących (testprompts), zostało wykonane w warunkach domowych, przy względnie dużym stosunku sygnału do szumu, za pomocą mikrofonu zintegrowanego z laptopem. Długość nagrania wynosi **3 minuty i 5 sekund** (słowa izolowane). Wykonana została anotacja zdań, segmentująca wypowiedź na poszczególne części (celem stworzenia pliku mlf).

Nagrania testowe (w liczbie 10) zostały wykonane pod koniec pracy z HTK. Zostały wykonane w tych samych warunkach i na tym samym sprzęcie co nagrania treningowe.

*Zdania z nagrań testowych:*

```
1.wybieram teatr stary sale duZa rzad gOrny miejsce pierwsze w_piAtek rano
2.wybieram teatr stary sale mala rzad gOrny miejsce szOste w_piAtek
3.wybieram teatr stu sale duZa rzad dolny miejsce Osme w_niedziele
4.wybieram teatr bagatela sale kameralnA rzad gOrny miejsce trzecie w_piAtek wieczOr
5.wybieram teatr stary sale kameralnA rzad dolny miejsce czwarte w_sobotE rano
6.wybieram teatr stu sale kameralnA rzad gOrny miejsce Osme w_niedzielE
7.wybieram teatr stary sale maLA rzad dolny miejsce piAte w_niedzielE wieczOr
8.wybieram teatr bagatela sale duZa rzad dolny miejsce pierwsze w_sobote wieczOr
9.wybieram teatr stu sale maLa rzad dolny miejsce trzecie w_niedziele
10.wybieram teatr stary sale maLa rzad gOrny miejsce drugie w_piAtek
```

## 4. Analiza wyników

Poniżej zamieszczono rezultaty dla różnych etapów estymacji (komenda HVite wskazuje który etap jest analizowany). Przedstawiam wyniki tylko dla reestymacji 3-5, te 1-3 i 5-10 były bowiem gorsze. Jak widać najlepszy wynik otrzymujemy z reestymacji **3**.

```
Microsoft Windows [Wersja 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. Wszelkie prawa zastrzeżone.

C:\Users\Mery>cd C:\Users\Mery\Desktop\AB1\htkwindows

C:\Users\Mery\Desktop\AB1\htkwindows>HVite -C config -H hmm3\macros -H hmm3\hmmd
efs -S test.scp -l '*' -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict monophones0

C:\Users\Mery\Desktop\AB1\htkwindows>HResults -I testref.mlf monophones0 recout.
mlf
====================== HTK Results Analysis =======================
  Date: Wed Jan 16 17:45:09 2013
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=10, N=10]
WORD: %Corr=64.42, Acc=45.19 [H=67, D=2, S=35, I=20, N=104]
===================================================================

C:\Users\Mery\Desktop\AB1\htkwindows>HVite -C config -H hmm4\macros -H hmm4\hmmd
efs -S test.scp -l '*' -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict monophones0

C:\Users\Mery\Desktop\AB1\htkwindows>HResults -I testref.mlf monophones0 recout.
mlf
====================== HTK Results Analysis =======================
  Date: Wed Jan 16 17:45:52 2013
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=10, N=10]
WORD: %Corr=63.46, Acc=46.15 [H=66, D=1, S=37, I=18, N=104]
===================================================================

C:\Users\Mery\Desktop\AB1\htkwindows>HVite -C config -H hmm5\macros -H hmm5\hmmd
efs -S test.scp -l '*' -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict monophones0

C:\Users\Mery\Desktop\AB1\htkwindows>HResults -I testref.mlf monophones0 recout.
mlf
====================== HTK Results Analysis =======================
  Date: Wed Jan 16 17:46:13 2013
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=10, N=10]
WORD: %Corr=60.58, Acc=39.42 [H=63, D=2, S=39, I=22, N=104]
===================================================================
```

**Analizując otrzymane wyniki stwierdzam, że rozpoznanie zdań wynosi 0% jednak rozpoznanie słów zawiera się w zakresie 60-65%.**

Czynnikami mającymi wpływ na powyższy wynik rozpoznania mogą być:

- niedokładność w anotacji (błędy słownika),
- zniekształcenia w nagraniach,
- stosunek słów stałych do słów zmiennych (słowa zawsze występujące zajmują ok. 50% wypowiedzi).

Na podstawie powyższej analizy (zwłaszcza podpunktu 3.) można zaobserwować zgodność rozpoznania słów zmiennych, poprzez analizę plików testref i recout.

Gorzej rozpoznawane słowa i frazy to: *bagatela, w niedzielę, w sobotę, wieczór, dużą/małą, numer miejsca.*

Hipoteza ich gorszego rozpoznawania to oczywiście ich stosunkowa rzadkość występowania w całości nagrania oraz niekoniecznie dobre zrealizowanie zapisu fonetycznego (o czym pisałam wyżej). Dodatkowo niektóre z nich brzmią podobnie (szóste/ósme, dużą/małą).

## 5. Analiza metod alternatywnych

Celem sprawdzenia pracy systemu na obcych danych, poprosiłam koleżankę o nagranie tych samych zdań testowych na swoim komputerze. Podejrzewam, że warunki były podobne jak podczas nagrań wykonywanych przeze mnie. Różnił się rodzaj sprzętu i głos osoby nagrywającej.

```
C:\Users\Mery\Desktop\AB1\htkwindows>HVite -C config -H hmm3\macros -H hmm3\hmmd
efs -S test.scp -l '*' -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict monophones0

C:\Users\Mery\Desktop\AB1\htkwindows>HResults -I testref.mlf monophones0 recout.
mlf
====================== HTK Results Analysis =======================
  Date: Wed Jan 16 17:58:08 2013
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=10, N=10]
WORD: %Corr=60.00, Acc=52.38 [H=63, D=6, S=36, I=8, N=105]
===================================================================

C:\Users\Mery\Desktop\AB1\htkwindows>HVite -C config -H hmm4\macros -H hmm4\hmmd
efs -S test.scp -l '*' -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict monophones0

C:\Users\Mery\Desktop\AB1\htkwindows>HResults -I testref.mlf monophones0 recout.
mlf
====================== HTK Results Analysis =======================
  Date: Wed Jan 16 17:58:36 2013
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=10, N=10]
WORD: %Corr=61.90, Acc=57.14 [H=65, D=6, S=34, I=5, N=105]
-------------------------------------------------------------------

C:\Users\Mery\Desktop\AB1\htkwindows>HVite -C config -H hmm5\macros -H hmm5\hmmd
efs -S test.scp -l '*' -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict monophones0

C:\Users\Mery\Desktop\AB1\htkwindows>HResults -I testref.mlf monophones0 recout.
mlf
====================== HTK Results Analysis =======================
  Date: Wed Jan 16 17:58:56 2013
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=10, N=10]
WORD: %Corr=60.95, Acc=56.19 [H=64, D=5, S=36, I=5, N=105]
===================================================================

C:\Users\Mery\Desktop\AB1\htkwindows>HVite -C config -H hmm6\macros -H hmm6\hmmd
efs -S test.scp -l '*' -i recout.mlf -w wdnet -p 0.0 -s 5.0 dict monophones0

C:\Users\Mery\Desktop\AB1\htkwindows>HResults -I testref.mlf monophones0 recout.
mlf
====================== HTK Results Analysis =======================
  Date: Wed Jan 16 17:59:17 2013
  Ref : testref.mlf
  Rec : recout.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=0.00 [H=0, S=10, N=10]
WORD: %Corr=61.90, Acc=57.14 [H=65, D=6, S=34, I=5, N=105]
-------------------------------------------------------------------
```

Jak zauważyłam, **wynik rozpoznawania wahał się w granicy 60-62%.** Najwyższy wynik otrzymałam z reestymacji 4 i 6. Pogorszenie wyniku nie jest znaczne, jest oczywistą konsekwencją nieprzystosowania systemu do testu na obcych danych.