

Sprawozdanie z laboratoriów HTK

1. Przeznaczenie tworzonego systemu

Celem było stworzenie systemu służącego do sterowania samochodem. Zaimplementowane zostały podstawowe funkcję silnika, okien, radia i klimatyzacji.

2. Gramatyka

Model gramatyki został oparty o naturalne komendy, które mogłyby być wydane samochodowi z zainstalowanych systemem rozpoznawania mowy. Każde polecenie rozpoczyna się od kluczowego elementu, typu silnik, okno, klimatyzacji i definiuje obiekt, którym będziemy sterować. Poniżej zamieszczona została gramatyka systemu:

```
$uruchamianie = wLAcz | wyLAcz |  
uruchom | zgaS;  
$ktOre_okno = kierowcy | pasaZera |  
tylne_lewe | tylne_prawe | przednie |  
tylne | wszystkie;  
$okna = otwOrz | zamknij | uchyl;  
$drzwi = zablokuj | odblokuj;  
$radio = wLAcz | wyLAcz | zmieN_stacje  
| uruchom_pLyte_CD;  
$Swiatla = wLAcz | wyLAcz | dlugie |  
krOtkie | drogowe | przeciwmgielne |  
przeCiwmglowe;  
$klimatyzacja = wLAcz | wyLAcz;  
  
( SENT-START ( (silnik|samochOd)  
$uruchamianie | okno $ktOre_okno $okna  
| drzwi $drzwi | Swiatla $Swiatla |  
klimatyzacja $klimatyzacja | radio  
$radio) SENT-END )
```

3. Słownik

W słowniku umieszczonych zostało 30 słów, które posłużyć miały do stworzenia systemu.

```
silnik si i l ni i k sp  
samochOd s a m o h u t sp  
wLAcz w l _ a _ cz sp  
wyLAcz w y l _ a _ cz sp  
uruchom u r u h o m sp  
zgaS z g a si sp  
okno o k n o sp  
kierowcy k j e r o f c y sp  
pasaZera p a s a r z e r a sp  
tylne_lewe t y l n e l e w e sp  
tylne_prawe t y l n e p r a w e sp  
przednie p s z e d n i e sp  
tylne t y l n e sp  
wszystkie f s z y s t k j e sp  
otwOrz o t f u s z sp  
zamknij z a m k ni i j sp  
uchyl u h y l sp  
drzwi d r z w i sp  
zablokuj z a b l o k u j sp  
odblokuj o d b l o k u j sp  
radio r a d i o sp  
zmieN_stacje z m j e ni s t a c j e sp  
uruchom_pLyte_CD u r u h o m p l _ y t e  
si di sp  
Swiatla si f j a t l _ a sp  
dlugie d l _ u g j e sp  
krOtkie k r u t k j e sp  
drogowe d r o g o w e sp  
przeciwmgielne p s z e ci i w m g j e l n e sp  
przeCiwmglowe p s z e ci i w m g l _ o w e sp  
klimatyzacja k l i m a t y z a c j a sp  
SENT-START  
SENT-END
```

4. Nagrania treningowe

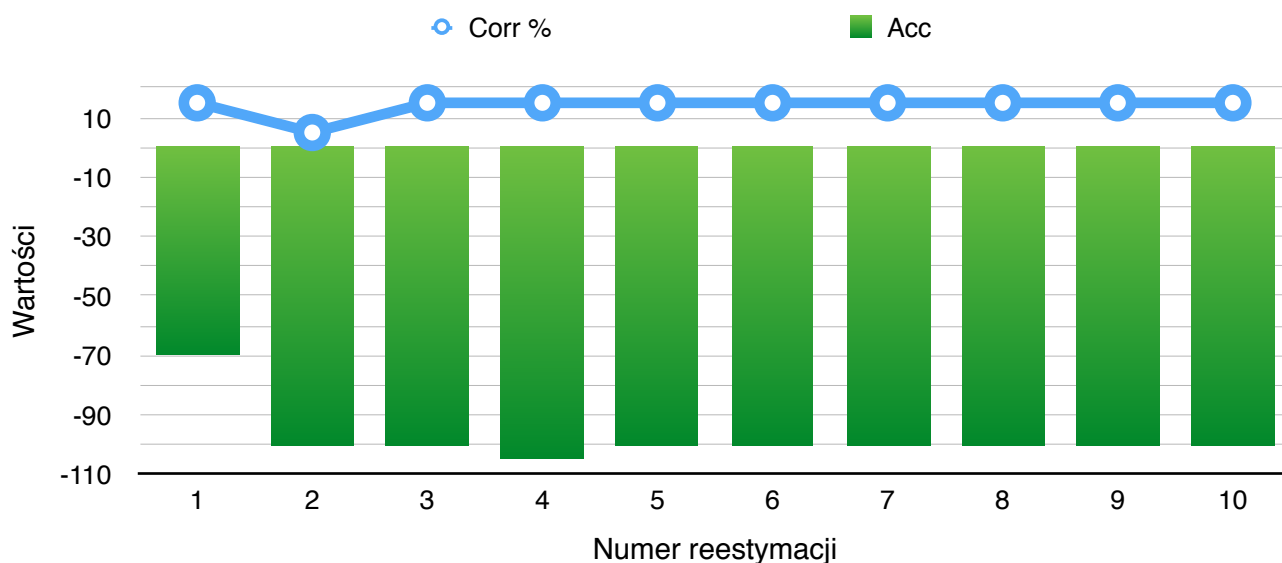
Do zrealizowania nagrań posłużył mikrofon wbudowany w laptopie i program Audacity. Nagrania zostały wykonane w 16 bitach, 16 kHz i mono. Łączny czas nagrań wyniósł 2:50. W nagraniach użyto wszystkich możliwych do wystąpienia kombinacji zdać. Wykonane zostały w pokoju, w którym panowała cisza.

5. Wyniki

Numer reestymacji	Corr %	Acc
1	15	-70
2	5	-100
3	15	-100
4	15	-105
5	15	-100
6	15	-100
7	15	-100
8	15	-100
9	15	-100
10	15	-100

Tabela 1. Wyniki uzyskane podczas rozpoznania nagrań testowych

W tabeli przedstawione zostały wyniki uzyskane podczas procesu rozpoznawania mowy. Jak można zauważyć żaden z wyników nie przekroczył progu 15% procent dobrze rozpoznanych słów. Dla sentencji wynik zawsze wynosił 0%. Wynik najwyższy uzyskany został przy pierwszej reestymacji. Poprawność na poziomie 15% i celność wynosząca tylko -70.



Rysunek 1. Wyniki uzyskane podczas rozpoznawania nagrań testowych

6. Analiza wyników

Niestety uzyskana rozpoznawalność mowy jest bardzo niezadowolająca i wyniki uzyskane w trakcie pracy z programem są marnej jakości. Wynikać to może z budowy gramatyki, która mimo że nie jest skomplikowana mogła przysporzyć HTK sporo problemów. Innym istotnym problemem, który wpłynąć mógł na poziom otrzymanych wyników, mogły być nagrania testowe, a dokładniej zawarte w nich zdaniach, które pozwoliły „wytrenować” niektóre fonemy lepiej od innych. Po spróbowaniu wszystkich znanych mi metod poprawy wyników i marnych rezultatów tychże prób postanowiłem spróbować od początku i naprawić każdą część systemu począwszy od gramatyki.

7. Gramatyka

W poprzedniej wersji zauważone zostały, że niektóre komendy mają zbyt wiele możliwości i dlatego w późniejszych etapach pracy z HTK możliwe było powielanie się niektórych instrukcji i problem z rozpoznawaniem konkretnych zdań dlatego postanowiłem edytować gramatykę i otrzymałem plik o następującej treści:

```
$uruchamianie = wLAcz | wyLAcz | uruchom | zgaS;  
$ktOre_okno = kierowcy | pasaZera | tylne_lewe |  
tylne_prawe;  
$okna = otwOrz | zamknij;  
$drzwi = zablokuj | odblokuj;  
$SwiatLa = wLAcz | wyLAcz | dLugie | krOtkie | drogowe |  
przeciwmglowe;  
$klimatyzacja = wLAcz | wyLAcz;  
$radio = wLAcz | wyLAcz | zmieN_stacje | podgLoSnij |  
wycisz;  
$bagaZnik = otwOrz | zamknij;  
$wycieraczki = wLAcz | wyLAcz | uruchom;
```

```
( SENT-START ( (silnik|samochOd) $uruchamianie | okno  
$ktOre_okno $okna | drzwi $drzwi | SwiatLa $SwiatLa |  
klimatyzacja $klimatyzacja | radio $radio | багаZnik  
$bagaZnik | wycieraczki $wycieraczki ) SENT-END )
```

8. Słownik

Słownik również został poprawiony, a dodatkowo wprowadzone zostały do niego pozostałe możliwe transkrypcje fonetyczne występujących w nim słów.

```
silnik si i l ni i k  
silnik si l ni k  
samochOd s a m o h u t  
samochOd s a m o h u d  
wLAcz w l _ a _ cz  
wLAcz w l _ o n cz  
wyLAcz w y l _ a _ cz  
wyLAcz w y l _ o n cz  
uruchom u r u h o m  
zgaS z g a s i  
okno o k n o  
kierowcy k j e r o f c y  
kierowcy k i e r o f c y  
pasaZera p a s a r z e r a  
tylne_lewe t y l n e l e w e  
tylne_prawe t y l n e p r a w e  
otwOrz o t f u s z  
zamknij z a m k ni i j  
zamknij z a m k ni j  
drzwi d r z w i  
zablokuj z a b l o k u j  
odblokuj o d b l o k u j  
radio r a d i o  
zmieN_stacje z m j e ni s t a c j e  
podgLoSnij p o d g l _ o si ni j  
wycisz w y ci s z  
SwiatLa si f j a t l _ a  
dLugie d l _ u g j e  
dLugie d l _ u g i e  
krOtkie k r u t k j e  
krOtkie k r u t k i e  
drogowe d r o g o w e  
przeciwmglowe p s z e ci i w m g l _ o w e  
klimatyzacja k l i m a t y z a c j a  
bagaZnik b a g a r z ni k  
bagaZnik b a g a r z ni g  
wycieraczki w y ci e r a cz k i  
SENT-START  
SENT-END
```

9. Nagranie treningowe

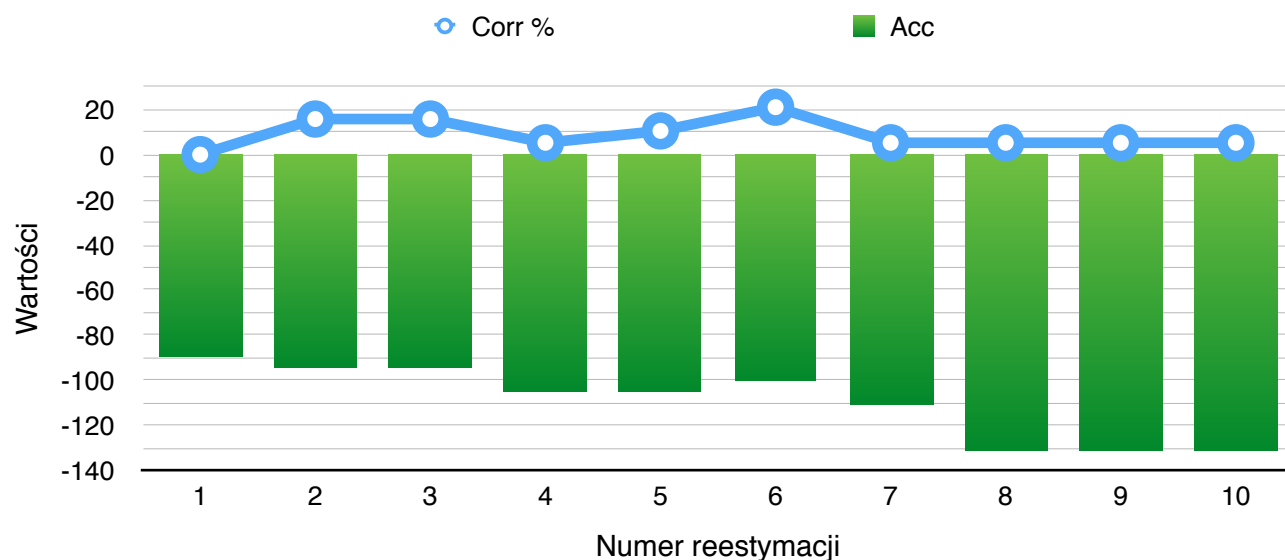
Nowa gramatyka i nowy słownik wymagały nowych nagrań treningowych, które i w tym wypadku zostały nagrane na mikrofon w laptopie i w programie Audacity. Nagrania zostały wykonane w 16 bitach, 16 kHz i mono. Łączny czas nagrań wyniósł 2:36. W nagraniach użyto wszystkich możliwych do wystąpienia kombinacji zdań i każde z nich powtórzone dwukrotnie. Wykonane zostały w pokoju, w którym panowała cisza.

10. Wyniki

Numer reestymacji	Corr %	Acc
1	0	-89.47
2	15.79	-94.74
3	15.79	-94.74
4	5.26	-105.26
5	10.53	-105.26
6	21.05	-100
7	5.26	-110.53
8	5.26	-131.58
9	5.26	-131.58
10	5.26	-131.58

Tabela 2. Wyniki uzyskane podczas rozpoznania nagrań testowych dla zmienionej gramatyki i słownika

W powyżej tabeli widać, że niestety wyniki uzyskane po edycji gramatyki i słownika nie są zauważalnie lepsze, przy niektórych reestymacjach są dużo niższe!



Rysunek 2. Wyniki uzyskane podczas rozpoznania nagrań testowych dla zmienionej gramatyki i słownika

11. Analiza wyników

Wyniki otrzymane podczas drugiej próby również nie są zadowalające. Tylko szóstej reestymacji udało się przekroczyć 20 procent rozpoznawalności. Wynik ten jest o zaskakujące, ponieważ inne reestymacje oscylują w granicach 5 procent. Mimo poprawionej gramatyki i słownika wyniki nie są na tyle satysfakcjonujące, aby mogły być używane w życiu codziennym. 21.05% rozpoznanych słów czyli najlepszy wynik z całego zadania jest to niewiele ponad $\frac{1}{5}$, co czyni mój program bardzo nieudolnym, należy dodać że cały czas mam na myśli rozpoznawanie słów. Zaimplementowane dane nie były w stanie rozpoznać ani jednego zdania poprawnie. Problem rozpoznawaniem mojej mowy mógł wynikać z budowy gramatyki, niedokładnej adnotacji, słabej jakości mikrofonu lub po prostu jakości mojej wypowiedzi i artykułowanych przeze mnie dźwięków.

Wyrażam zgodę na dołączenie moich nagrań do korpusu mowy AGH. Nagrania mogą być odtwarzane ale wyłącznie bez podawania tożsamości mówcy (np. w celu prezentacji jakości, rodzaju nagrań itd.).