System automatycznej konwersji mowy polskiej na tekst LVCSR

Grażyna Demenko

Zakład Fonetyki
UAM Poznań
Laboratorium Zintegrowanych Systemów Przetwarzania Języka i Mowy PCSS

Activities

- Large Vocabulary Continuous Speech Recognition
- Speaker recognition
- Speech data collection
- Segmentation and annotation of speech
- Natural language processing
- Text to Speech
- Para and Extralinguistic. Vocal communication
- Speech and hearing pathology

Past Projects

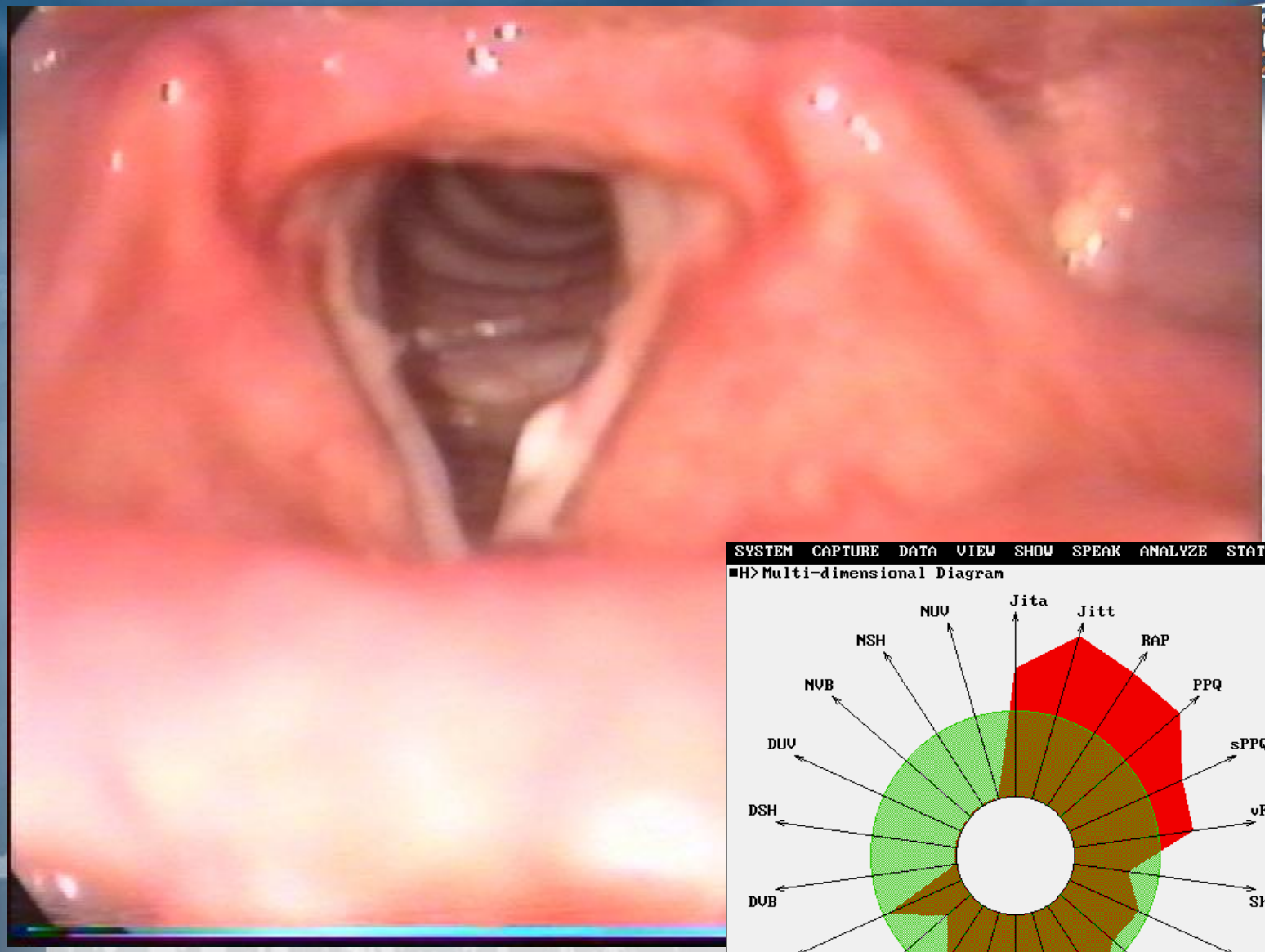
Phoniatrics and Audiology
CALL
Speech synthesis
LVCSR

Phoniatrics and Audiology

The Goals

Acoustic Speech Protocol Goals

- Verbal communication assesment
- Diagnosis
  - functional
  - organic
  - others
- Rehabilitation
- Speech organs prothesis



SYSTEM CAPTURE DATA VIEW SHOW SPEAK ANALYZE STATS GRAPH EDIT MACROS

■H>Multi-dimensional Diagram

Fo = 201.9 Hz

Current Values

Jita =142.6 us
Jitt = 2.88 %
RAP = 1.68 %
PPQ = 1.81 %
sPPQ = 1.69 %
vFo = 2.21 %
ShdB = 0.32 dB
Shim = 3.69 %

Threshold Values

Jita = 94.6 us
Jitt = 1.44 %
RAP = 0.90 %
PPQ = 0.97 %
sPPQ = 1.16 %
vFo = 1.57 %
ShdB = 0.50 dB
Shim = 4.15 %

Curr Print

OK

CALL
APLICATIONS IN EDUCATION

Education and Culture DG

Lifelong Learning Programme

AUDIOVISUAL FEEDBACK FOR FOREIGN LANGUAGE LEARNING

Grażyna Demenko[1], Agnieszka Wagner[1], Natalia Cylwik[1], Oliver Jokisch[2], Uwe Koloska[3], Diane Hirschfeld[3]

([1]) Adam Mickiewicz University, Institute of Linguistics, Department of Phonetics, Poznań, Poland
([2]) Dresden University of Technology, Laboratory of Acoustics and Speech Communication, Dresden, Germany
([3]) voice INTER connect GmbH, Research and Development, Dresden, Germany
[1]{lin, wagner, nataliac}@amu.edu.pl, [2]oliver.jokisch@ias.et.tu-dresden.de,
[3]{koloska,hirschfeld}@voiceinterconnect.de

the AzAR software provides a multimodal feedback – it includes visual and audio modules in the form of curriculum recordings by a reference voice and the visualization of the speech signal under the transcribed and phonemically segmented reference utterances.

The software uses HMM-based speech recognition and speech signal analysis on the learner‟s input which makes a visual and aural comparison of user‟s own performance with that of the reference voice possible. Most importantly, the system also performs an automatic error detection on the phonemic level.

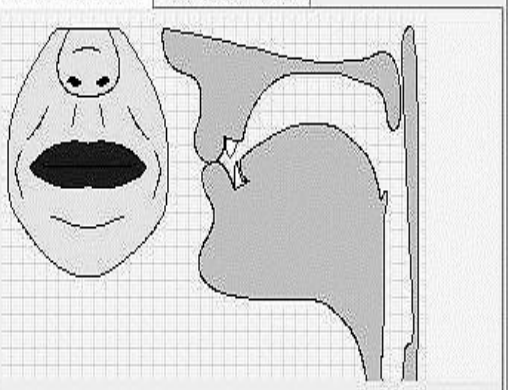
Inhalt | Artikulation | Glossar | Auswertung

Aufgabenstellung

Hören Sie bitte die Satzbeispiele, achten Sie auf die Aussprache der Graphemfolgen <ng> und <nk>! Lesen Sie bitte anschließend selbst, vergleichen Sie Ihre Aufzeichnung mit dem Muster!

Wir brauchen dringend etwas zu trinken!

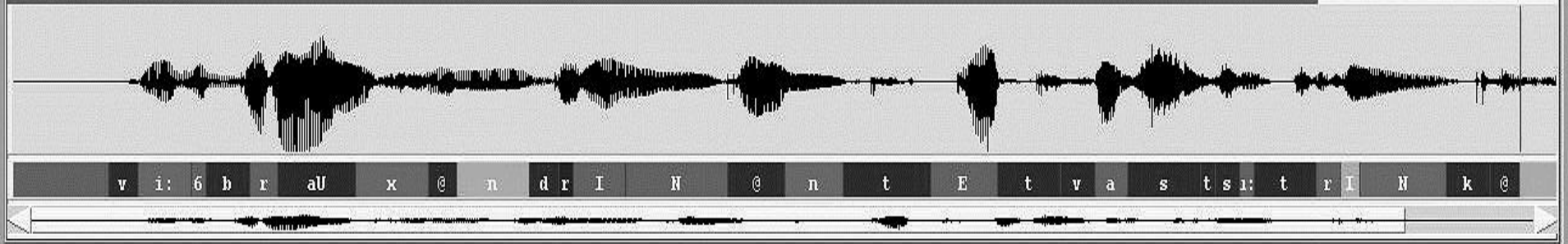
Animation | Formanten



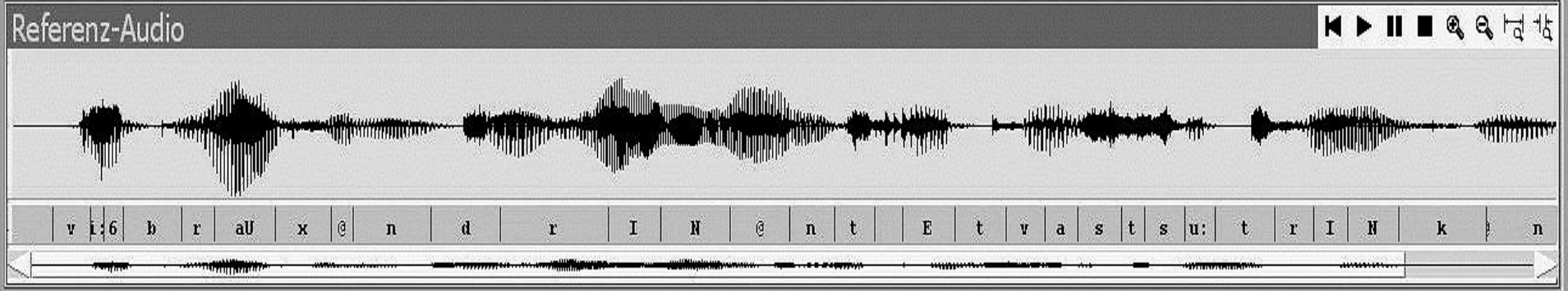
Schlecht — Gut

Aufnahme

Benutzer-Audio



Referenz-Audio



Speech Synthesis

The BOSS TTS system is an open source architecture for concatenative speech synthesis, especially for unit selection.

BOSS was originally developed for German but the latest version has seen significant changes to software design and architecture that makes it easily extensible to be used in a multilingual context.

Several of the system components have been generalized to accommodate other languages, and TTS development for Polish has served as a testbed for the language-independent applicability of the BOSS architecture

The speech corpus used for Polish speech

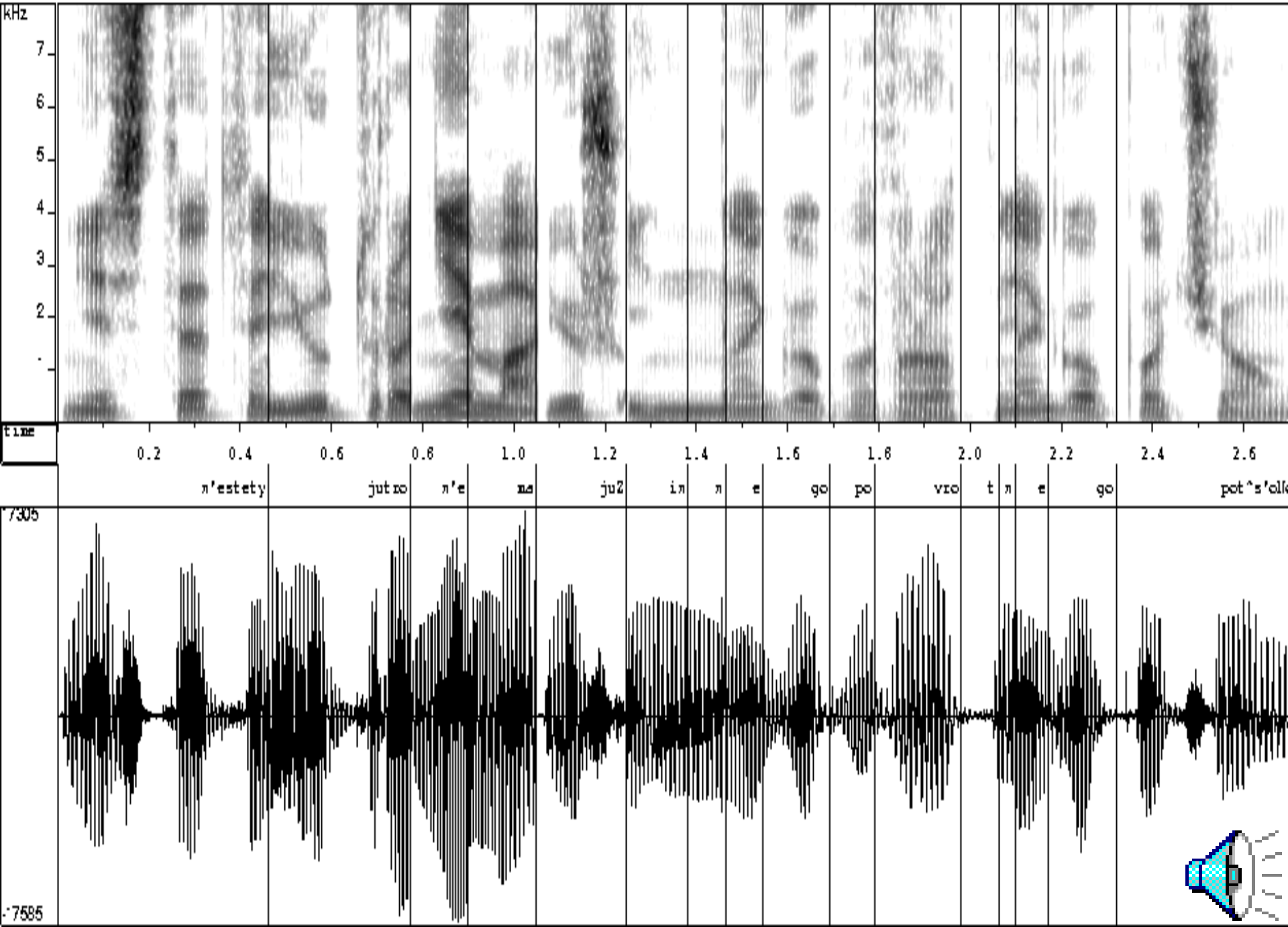
• consonant clusters corpus - phrases with most frequent consonant structures, 258 consonant clusters of various types were used.

• triphone coverage corpus - phrases for covering all VC and CV spectral transitions (only CVC combinations in logatoms – approx. 6000 units).

• diphone coverage corpus - diphones in various phonetic (aprox. 5000 units). The database contains several instances of each diphone, taken from different meaningful phrase contexts (phonetic, rhythmic, morphological) and

• short 5 - 8 syllable phrases with typical suprasegmental structures realized on the most frequent syllabic clusters (200 short meaningful phrases, for covering the most important prosodic features),

• continuous read speech containing elements of dialogue phrases, monologue and newspaper language in short text paragraphs – approx. 15 minutes of speech.


kHz
7
6
5
4
3
2
time
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
n'estety
jutro
n'e
ma
ju2
in
n
e
go
po
vro
t
n
e
go
pot^s'olle
7305
-7585

Current Project

Speaker recognition & characterization

STRESS DETECTION

Assessment of whether or not a speaker is under stress is of importance in many civilian and military applications.

Automatic detections of vocal stress is also becoming increasingly important in the field of multilingual communication, security systems, banking and law enforcement, specifically since emergency call centers and police departments all over the world are overloaded with different kinds of calls, only some of which represent real danger and need immediate response.

Our study focuses on the analysis of vocal stress produced in response to the occurrences in the people's surroundings, perceived by them as unusual **AND IMPOSSIBLE TO CONTROL.** We analyzed third order stressors, the psychological ones, which have their effect at the highest level of speech production

Speech Corpus

The 997 - Emergency Calls Database is a spontaneous speech recordings collection that comprises crime/offence notifications and police intervention requests. All recordings were automatically grouped into sessions according to the phone number from which the call was made, in all comprising over 8 000 sessions.

The annotation included:
(1)background acoustics,
(2) types of dialog act,
(3) suprasegmental features such as speech rate, loudness), intonation
(4) context (threat, complain and depression)
(5) time (passed, immediate and potential)
(6) emotional coloring (up to 3 categorical labels and values for 3 dimensions: potency, valency, arousal;

where potency is the level of control that a person has over the situation causing the emotion, valency states whether the emotion is positive or negative and arousal refers to the level of intensity of an emotion

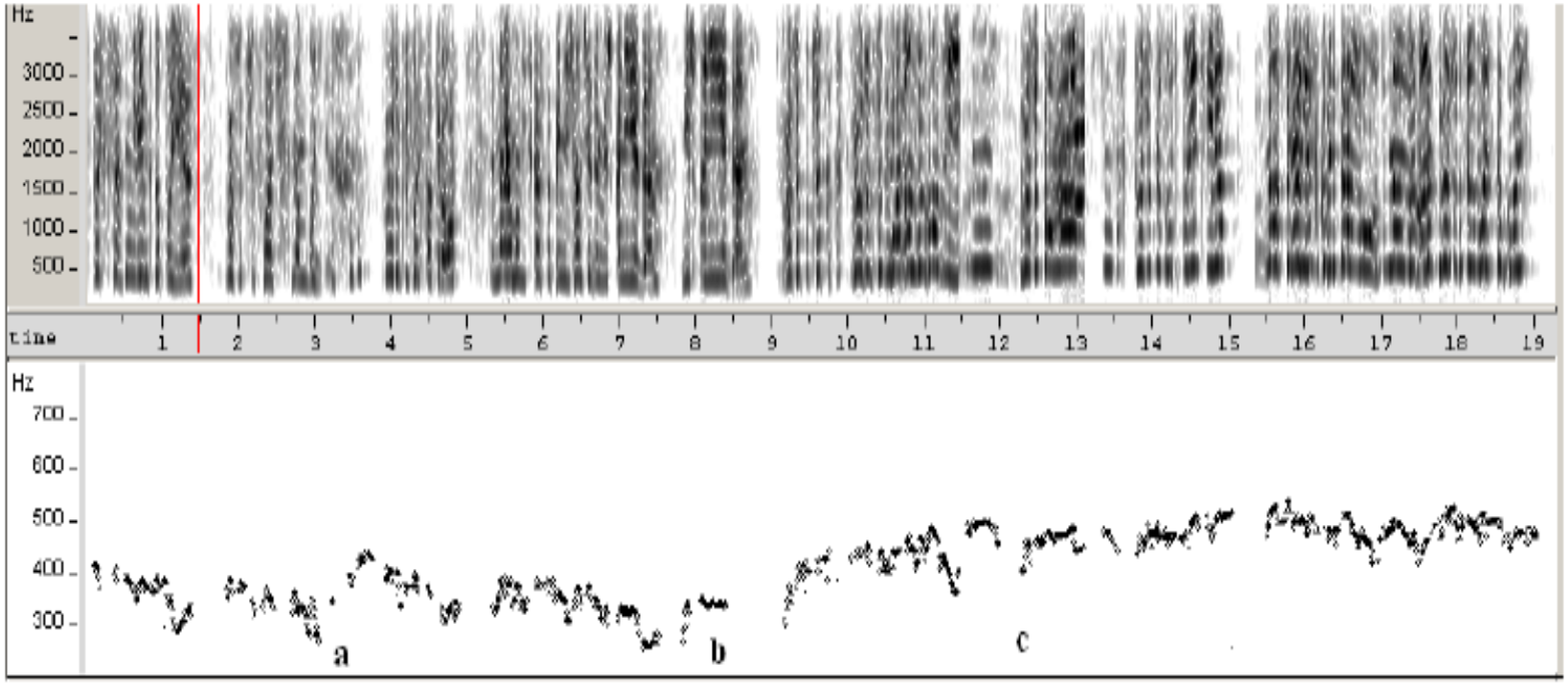


Fig.4. A gradual increase in stress in the utterances: a) someone is entering the apartment (F_{min} =315 Hz), b) someone is entering the apartment, he's masked, (F_{min} =350 Hz), c) he's leaving the room - he is somewhere [here] - direct threat (F_{min} =400 Hz).

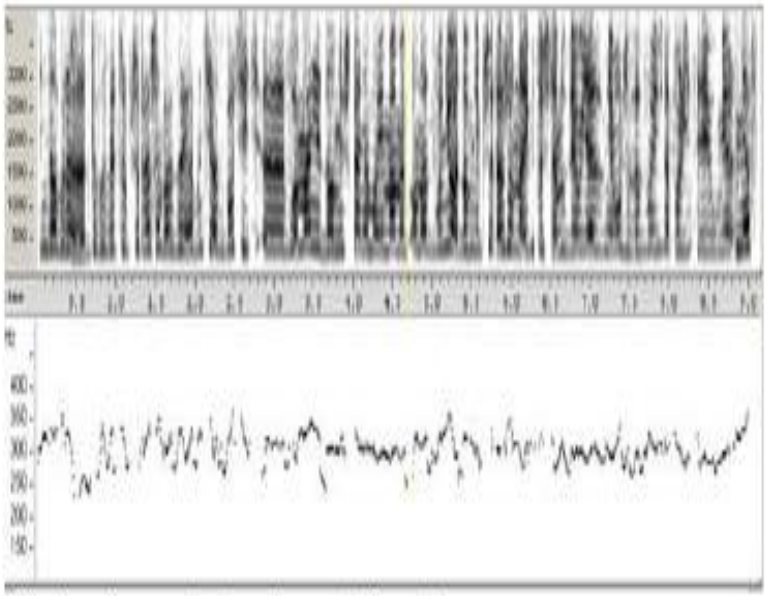


Fig.6a. A constant stress in the utterance:
Please, come over, there's a house-breaking, she's called. She's skirt to death. (F_{min} =240 Hz, F_{max}=352 Hz).

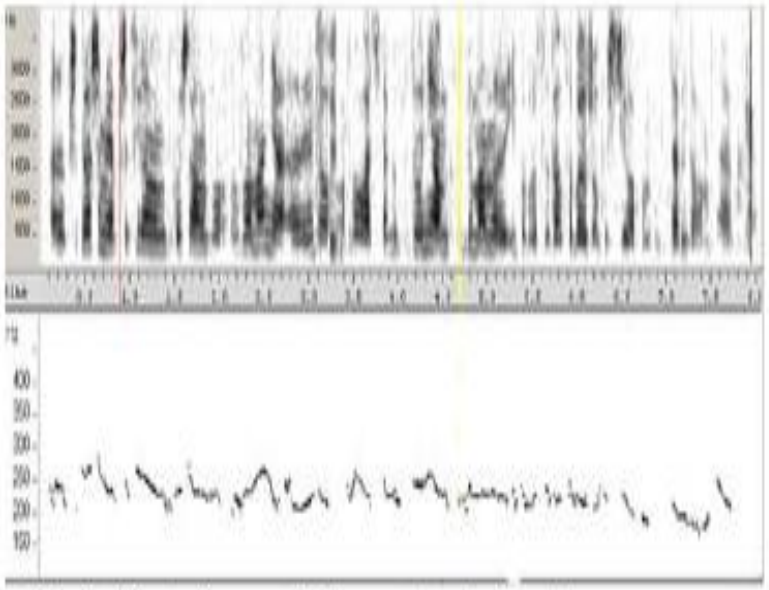


Fig.6b. Neutral speech. Fo contour in the utterance :
I called one hour ago, I want to call off he intervention. He's already left (F_{min} =167 Hz, F_{max}=264 Hz).

Speech recognition
LVCSR

OR00006707 Project:

Integrated automatic conversion of the Polish language for text-based language model created in the environment analysis and circulation of legal documents on internal security

project funded by the Ministry of Education of Poland under Decision No 0067/R/T00/2009 / 07

System projektowany jest dla:

- instytucji wymiaru sprawiedliwości

  łatwiejsze i szybsze przygotowywanie dokumentów, opracowywanie stenogramów

- organów odpowiedzialnych za bezpieczeństwo państwa i obywateli

  usprawnienie technik śledczych oraz dochodzeniowych na różnych jego etapach:

  w postępowaniu przygotowawczym i dowodowym, procesie karnym, tajnym działaniu operacyjnym, prewencji, centrum zarządzania sytuacjami kryzysowymi

sporządzanie wszelkiego rodzaju notatek, raportów, protokołów, orzeczeń, uzasadnień i sprawozdań, których ręczne spisywanie jest pracochłonne, a ponadto może być źródłem pomyłek, a nawet celowych manipulacji.

.

Polish Speech Recognition as an Open Scientific Problem

Complexity of Lexicon
Language Model
Ambiguities

Leksykon

Słownik ok.3 mln jednostek (otagowany gramatycznie i fonetycznie
Leksykon ARM ok.350.000 jednostek (otagowany gramatycznie, fonetycznie oraz prozodycznie)

Korpus tekstowy LEX
Korpus tekstowy Gazety Wyborczej
Raporty Policyjne
Raporty Straży Granicznej

1. i	8148355	i	i	i	i	i	i	i	i
2. na	7095141	n a	n a	n a	n a	n a	n a	n a	n a
3. do	6693193	d o	d o	d o	d o	d o	d o	d o	d o
4. nie	5749216	n' e	n' "e	n' e	n' "e	n' e	n' "e	n' e	n' "e
5. nie	5749216	n' e	n' e	n' e	n' e	n' e	n' e	n' e	n' e
6. o	5213853	o	o	o	o	o	o	o	o
7. że	3796050	Z e	Z e	Z e	Z e	Z e	Z e	Z e	Z e
8. się	3311326	s' e	s' e	s' e	s' e	s' e	s' e	s' e	s' e
9. jest	3099154	j e s t	j "e s t	j e z d	j "e z d	j e s t	j "e s t	j e z d	j "e
z 10. przez	3058114	p S e s	p S e s	p S e z	p S e z	p S e s	p S e s	p S e z	p S e z
11. od	2637359	o t	o t	o d	o d	o t	o t	o d	o d
12. z	2529073	s	s	z	z	s	s	z	z
13. roku	2444120	r o k u	r "o .k u	r o k u	r "o .k u	r o k u	r "o .k u	r o k u	r "o .k u
14. a	2325906	a	a	a	a	a	a	a	a
	z a	z a	z a	z a	z a	z a	z a		

JURISDICT

Baza systemu JURISDICT:

mowa spontaniczna, semispontaniczna, czytana, frazy izolowane

- Fonetyczna reprezentacja (zdania na pokrycie trifonowe)
- Dyktowanie opisu
- Calls (kierowane rozmowy telefoniczne)
- Wyroki, raporty policyjne
- Wyrażenia prawnicze, akronimy, frazy aplikacyjne
- Liczby, jednostki, adresy internetowe

The general assumptions for the structure of database:

1. Semantic structure
1.1. Legal lexicon: 30%
1.2. Common words: 50%
1.3. Proper names: 20%

2. Syntactic factors
2.1. Simple (one phrase) and complex sentences (more phrases).
2.2. Variable concatenation of phrases

3. Grammatical and phonetic factors
3.1. Statistical covering: most frequent triphones, specially CVC triphones in context of sonorants in accented/not accented position.
3.2. Statistical covering of the consonant clusters

1. Dictation in the court (by a judge)
2. Dictation in the legal/notary's/prosecutor's office (by a lawyer)
3. Dictation in the police station (by a policemen)

JURISDICT – Polish Speech Database

Corpus	Sub corpus	Time	Description (number of items per speaker)
A. Semi-Spontaneous Descriptive	A1	10 min	Free semi-spontaneous speech
	A2	5 min	Elicited spontaneous speech
B. Read speech Grammatically and phonetically controlled structure	B1	10 min	Syntactically complex sentences 70
	B2	10 min	Syntactically simple sentences 50
	B3	5 min	Special lexical phrases (words) 5
C. Read speech Core words and application phrases, texts	C1	5 min	Core words
	C2	15 min	Semantically controlled structure

SemiSpontaneous Speech - Corpus A

Sub-corpus 1A. Spontaneous Dictation (legal, police, court vocabulary)

This sub-corpus contains formal speech (dictation on various application topics).

Typical tasks are: dictation of any kind of legal texts (areas: judicial, disciplinary, criminal, divorce) in court, police reports (different topics, e.g. a description of a theft, burglary using common vocabulary, etc.).

The number of the recorded topics varies between speakers

Sub-corpus 2A. Spontaneous Dictation (common topics)

This sub-corpus contains informal speech (dictation on various common topics).

Typical tasks are: a description of a birthday, giving directions, giving an excuse, a description of holidays, etc.

The speaker is requested to be speak in a neutral style following instructions such as: *Imagine that you are calling your friend/father/boss and telling them something/excusing yourself/deciding on something*, etc.
The number of the recorded topics varies between speakers.

Sub-corpus 3A. Elicited Dictation (Answering questions)

The aim of sub-corpus 3A is to obtain some semantically important, frequent items such as birth dates, relative dates, times of day, city names, proper names, age, money amounts, currencies, sequences of digits and numbers, telephone numbers, mathematical operations as well as answers like yes/no/maybe, etc. and education, profession, etc. (27 categories).

Read Speech. Grammatically and Phonetically Controlled Structure - Corpus B

Sub-corpus 1B. Phonetically controlled structure.

Syntactically complex sentences.
By ‘syntactically complex’ we mean:
a) variable concatenation of phrases,
b) variable phrase length.

By ‘phonetically controlled’ we mean:
adequate coverage of triphones, triphones in the final position of a word/phrase.

The aim of the Corpus B was also to obtain:

CVC triphones in context of sonorants in a chosen accented/unaccented position. The number of ccented positions depends on a particular word's frequency, e.g. for triphone: jem

(I eat/I am eating) we have 4 prosodic positions e.g. *Łososia dziś jemy?* (Eng. *Are we eating salmon today?*).

The voiced context for the accented triphones was chosen because of a strong influence of ccent on acoustic features of the triphone (especially the sonorant-vowel connection is extremely context dependent).

Sub-corpus 2B. Phonetically controlled structure. Syntactically simple sentences

We expect that 90 short sentences will be provided by each speaker with the explicit intention of obtaining an adequate coverage for the chosen consonant clusters, short bigrams and triphones both in the accented and unaccented position

Sub-corpus 3B. Special lexical phrases (words)

The sub-corpus with more than 400 short one- or two-word includes special words like modulants, greetings, jargon/vulgar expressions.

It was constructed manually based on dictionaries and other resources for Polish.

At least 7 items are provided by one speaker.

The whole sub-database consists of approx. 2000 sentences with the controlled bigrams (e.g. two conjunctions, conjunction and reposition, etc.) in initial position and in the middle of a phrase for the most frequent bigrams.

The short (one- or two-syllable) words are most difficult to recognize for ASR systems.

The absolute frequency of different bigrams in Polish is based on the analysis of twenty million words taken from newspaper texts).

Read Speech. Semantically Controlled Structure – Corpus C

Sub-corpus 1C. General purpose words and phrases

Within this group utterances are divided into: general words/phrases and general-purpose commands.
The general-purpose words/phrases include 33 categories, among them: isolated digits, numerals, measures, letters, special keyboard characters, special legal acronyms, emails, web addresses.

Sub-corpus 2C. Application-specific short texts for users' needs

Texts extracted from original police reports and professional legal documents (up to 100 sentences).

Triphone coverage and statistics

triphones within word: 10593,
triphones containing an accented vowel: 8492, unaccented triphones 10650,
triphones in phrase final position: 4495.

Triphone lists serving as reference for the purpose of manual preparation of the B text corpus were created as follows: 2 million words were randomly selected from a corpus of texts including about 10 million words.

Baza JURISDICT składa się z 2.219 sesji nagraniowych. Zawiera łącznie 704.520 wypowiedzi o całkowitej długości trwania przekraczającej 1.200 godzin mowy (3000 mówców)

Annotation
Procedure

General procedure

Verification of the annotation by expert phoneticians

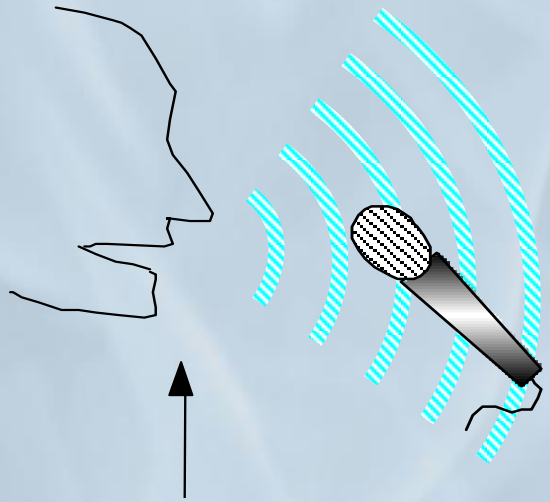
Automatic parsing of the annotation files, synchronization with lexicon.

Final verification by expert phoneticians.

Labelling by trained MA and PhD students from the Institute of Linguistics in Poznań

TECHNOLOGY

Trudności systemów ASR



Szum
Inni mówcy
Szum otoczenia

Zniekształcenia
Szum
Echo

TRANSMISJA

ASR

Mówca
- Jakość głosu
- Fo
- Płeć
- Dialekt

Styl
- Stres/Emocje
- Szybkość wypowiedzi

Kontekst
- Człowiek-komputer dialog
- Dyktowanie
- Konwersacja
- Wywiad

Mikrofon
- szum
- charakterystyka

Model akustyczny

Parametryzacja sygnalu

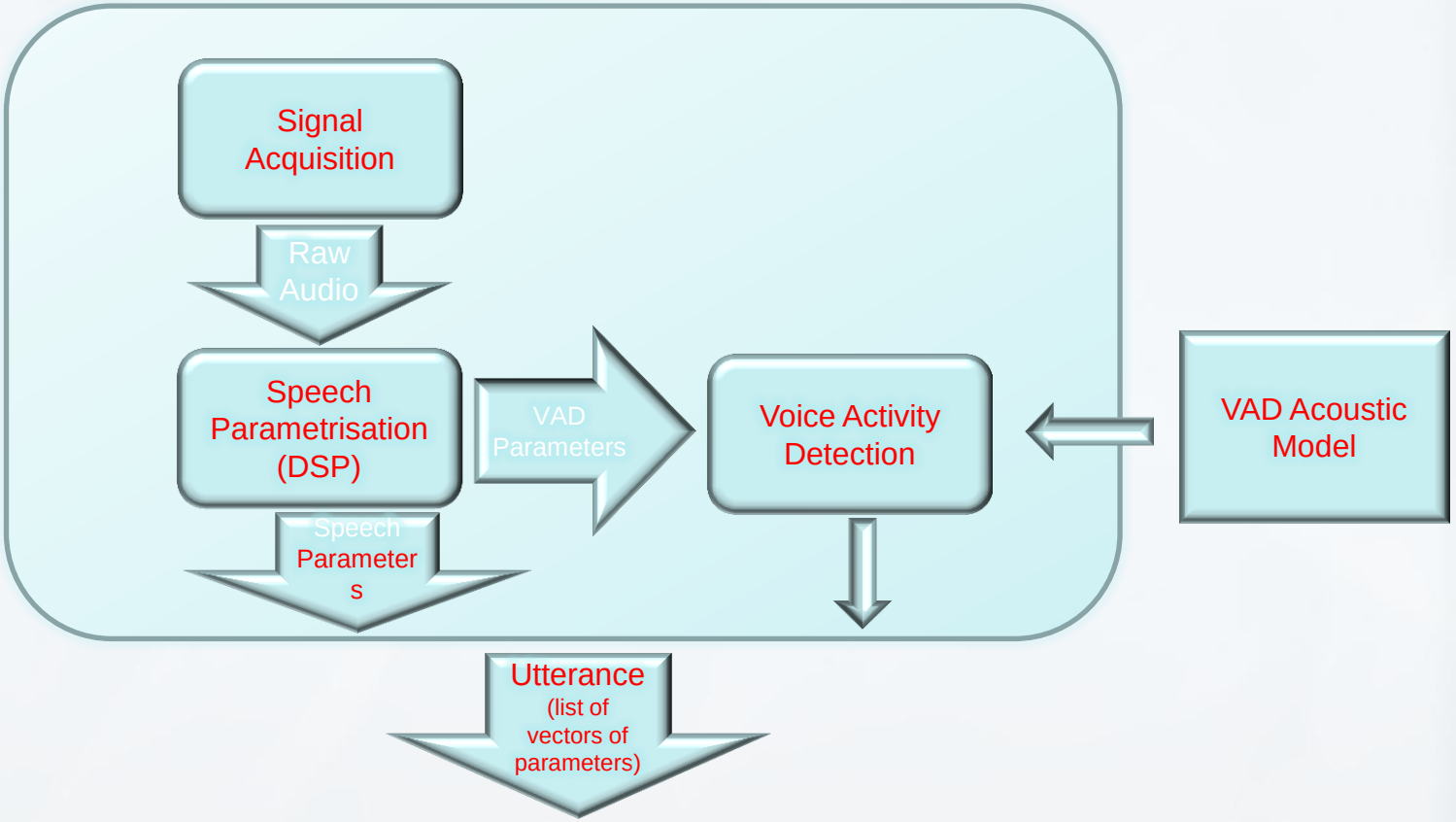
Klasyfikacja Dekodowanie

Prawdopodobna hipoteza

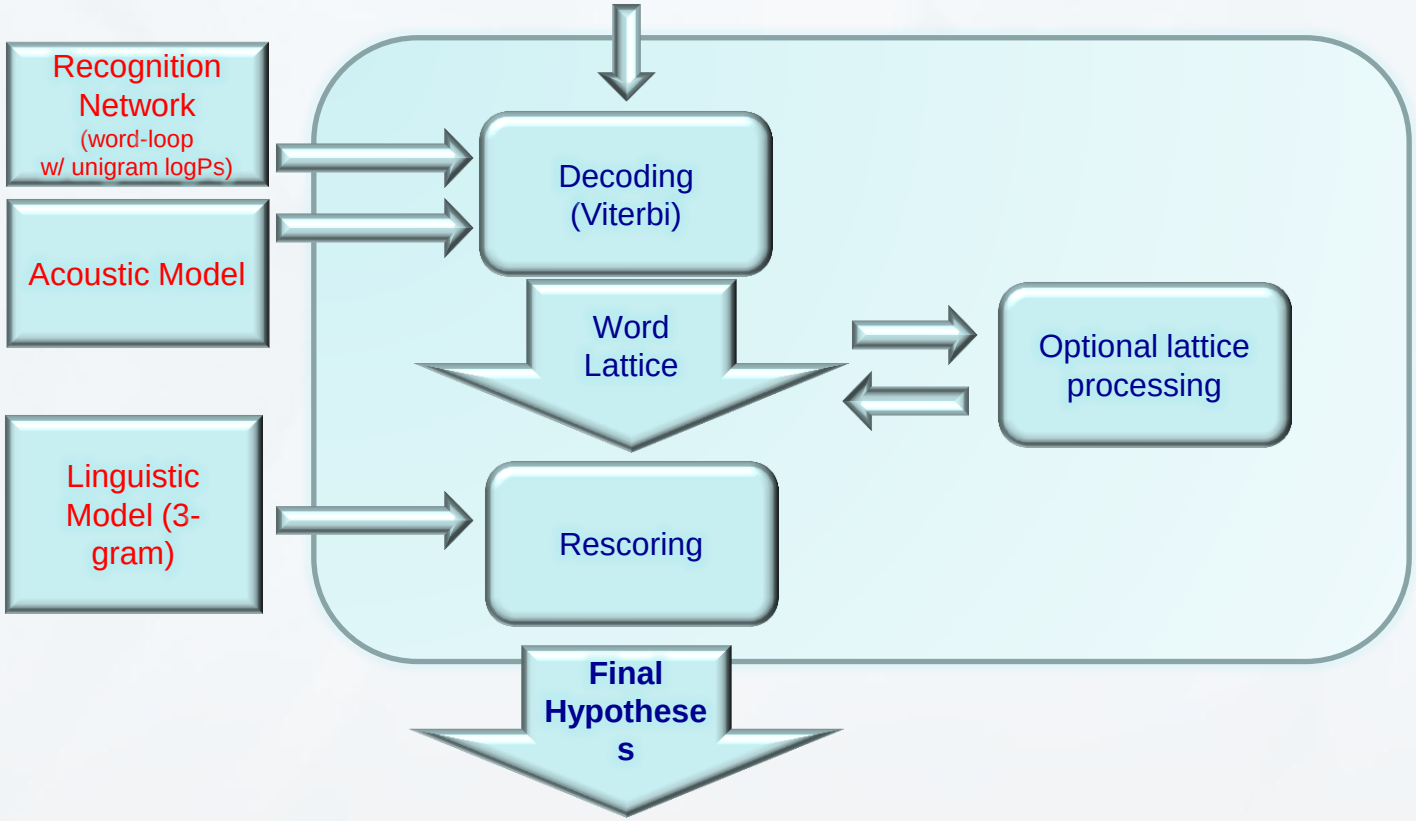
Leksykon

Modele lingwistyczne

DSP + VAD


Signal Acquisition
Raw Audio
Speech Parametrisation (DSP)
VAD Parameters
Voice Activity Detection
VAD Acoustic Model
Speech Parameters
Utterance
(list of vectors of parameters)

Recognizer (Decoding + Rescoring)



Evaluation

The program sclite is a tool for scoring and evaluating the output of speech recognition systems. Sclite is part of the NIST SCTK Scoring Tookit. The program compares the hypothesized text (HYP) output by the speech recognizer to the correct, or reference (REF) text. After comparing REF to HYP, (a process called alignment), statistics are gathered during the scoring process and a variety of reports can be produced to summarize the performance of the recognition system.

The National Institute of Standards and Technology (NIST)

The categories tallied are:

sclite - score speech recognition system output

Percent of correct words	**=**	**# Correct words / # Reference words**	*** 100**
Percent of substituted words	**=**	**# Substituted words / # Reference words**	*** 100**
Percent of inserted words	**=**	**# Inserted words / # Reference words**	*** 100**
Percent of deleted words	**=**	**# Deleted words / # Reference words**	*** 100**
Percent of sentence errors	**=**	**# incorrect ref and hyp pairs / # ref and hyp pairs**	*** 100**

SCLITE

```
CONFUSION PAIRS                          Total                          (3378)
                                         With >=   1 occurances         (3378)

   1:    13   ->   nejman ==> najman
   2:    13   ->   o ==> od
   3:    13   ->   opinie ==> opinię
   4:    12   ->   zamykam ==> zamyka
   5:    10   ->   wezwani ==> wezwanie
   6:     9   ->   była ==> gdyby
   7:     9   ->   dwa ==> dla
   8:     9   ->   mieszkania ==> mieszkanie
   9:     9   ->   nie ==> posiłkowanie
  10:     8   ->   i ==> psychicznej
  11:     8   ->   otwieram ==> otwiera
  12:     8   ->   powrÓt ==> powÓd
  13:     8   ->   pytanie ==> napytanie
  14:     8   ->   sprawy ==> sprawie
  15:     8   ->   tysiące ==> tysięcy
  16:     8   ->   ustalono ==> nieustaloną
  17:     8   ->   w ==> z
  18:     7   ->   /m ==> m
  19:     7   ->   i ==> ich
  20:     7   ->   karany ==> ukarany
  21:     7   ->   ma ==> na
  22:     7   ->   na ==> dla
  23:     7   ->   oskarżony ==> oskarżonej
  24:     7   ->   polskiej ==> polski
  25:     7   ->   pytania ==> pytanie
  26:     7   ->   sprawiał ==> sprawie
  27:     7   ->   zszokowany ==> zszokowane
  28:     7   ->   życie ==> należycie
```

POWINNO BYĆ	ROZPOZNANO
w	z
z	w
z	za
t	p
i	iż
n	m
c	s
czy	trzy
po	pod
ich	i
przez	przed
sąd	są
złoty	złotych
iż	niż
od	o
przy	przez
stanie	wstanie
a	o
ja	jak
że	ż
do	to
i	ich
l	r
o	od
przez	przy

POWINNO BYĆ	ROZPOZNANO
artykułu	artykuł
strony	stronę
pracę	pracą
które	który
pierwszej	pierwszy
pozwany	pozwanej
drugi	drugiej
który	które
dniu	dnia
oskarżony	oskarżonej
przepisu	przepisów
uzasadnienie	uzasadnienia
zamykam	zamykamy
decyzję	decyzją
ono	ona
pracodawcę	pracodawcą
pracy	pracę
uchylenie	uchylenia
złote	złotych
cel	celu
podstawy	podstawę
powinno	powinna

POWINNO BYĆ	ROZPOZNANO
wynagradzania	wynagrodzenia
określanego	określonego
określanie	określenie
określonych	określanych
rozpoznania	rozpoznawania
zapewniania	zapewnienia
zarzucanego	zarzuconego
zarzuconego	zarzucanego
zaspokajania	zaspokojenia
zatrudnieniu	zatrudnianiu
zgłaszanego	zgłoszonego

wiąże	wiążę
informację	informacje
tę	te
partie	partię
pracę	prace
premię	premie
chcę	chce
podtrzymuję	podtrzymuje
relację	relacje
sprawcą	sprawcom
uznaję	uznaje
Bożęcki	Borzęcki
konstrukcje	konstrukcję
kwalifikację	kwalifikacje
najemcą	najemcom
podpisuję	podpisuje
przyznaję	przyznaje

Recognition Results

(Acc - mean percentage of correctly recognized minus inserted words**,**
T - recognition time percentage, 100% = the real recognition time**)**.

Quality level	Acc[%]	T[%]
The highest	88,7	745,14
Higher	88,1	370,42
High	87,3	197,5
Mean	**86**	**117,1**
Low	84,1	74,95
Lower	82,2	56,9
The lowest	79,6	45,42

Test set: 97 speakers (total duration ca. 6 hour).
Quality level depends on internal decoder parameters.

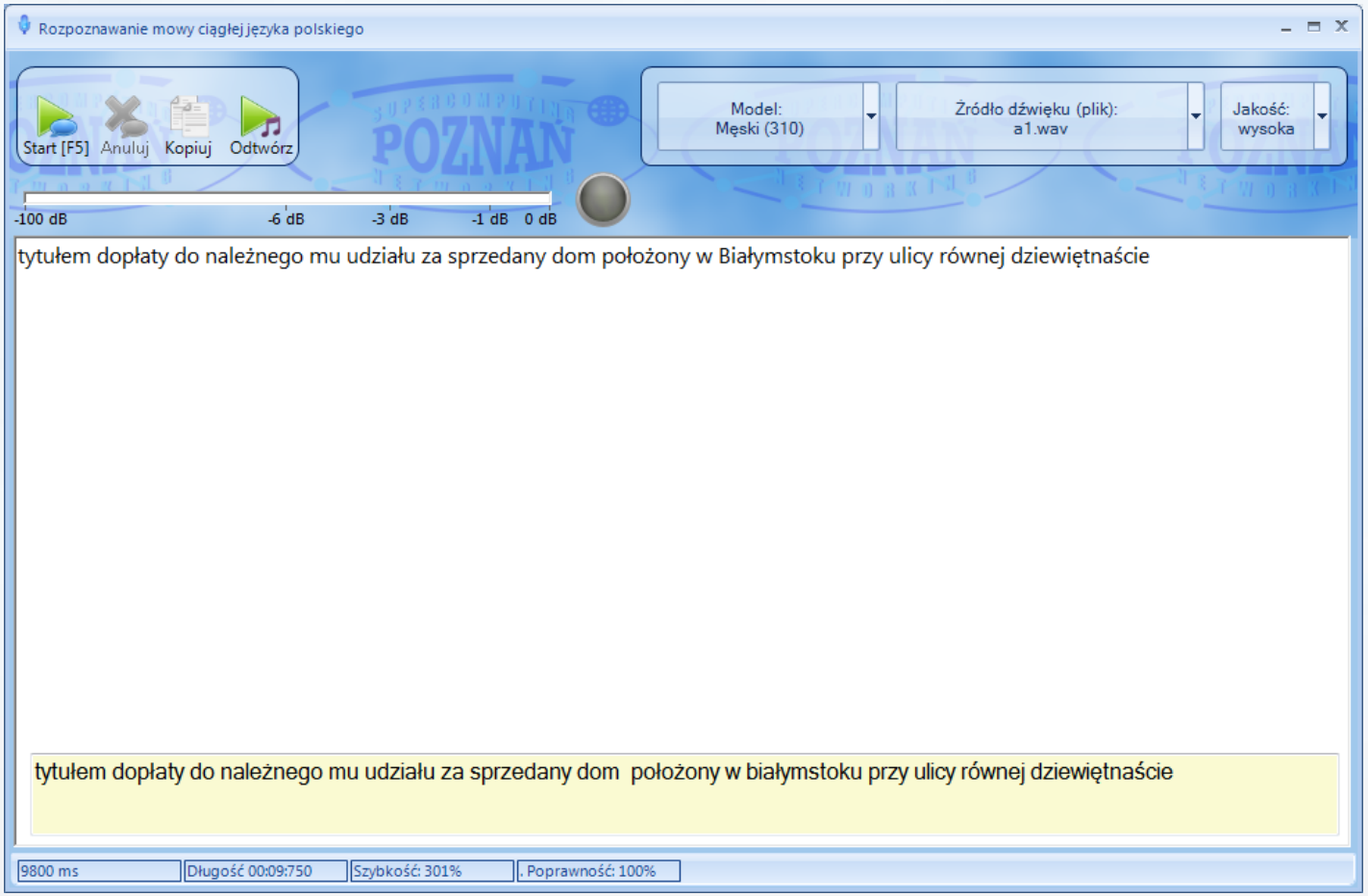
Speaker Adaptation Results

(Acc - mean percentage of correctly recognized minus inserted words**,**
T - recognition time percentage, 100% = the real recognition time**)**.

Quality level	without adaptation		with adaptation	
	Acc[%]	T[%]	Acc[%]	T[%]
The highest	90,7	500,09	93	297,08
Higher	90,4	254,89	92,6	161,06
High	89,7	145,51	92,2	97,6
Mean	**87,9**	**82,83**	**90,8**	**52,72**
Low	85,6	56,95	89,2	39,51
Lower	83,5	45,6	88	33,17
The lowest	81,2	37,66	85	28,2

Test set: 13 speakers (1.911 sentences, duration of ca. 3 hour).
Quality level depends on internal decoder parameters.

End User Application

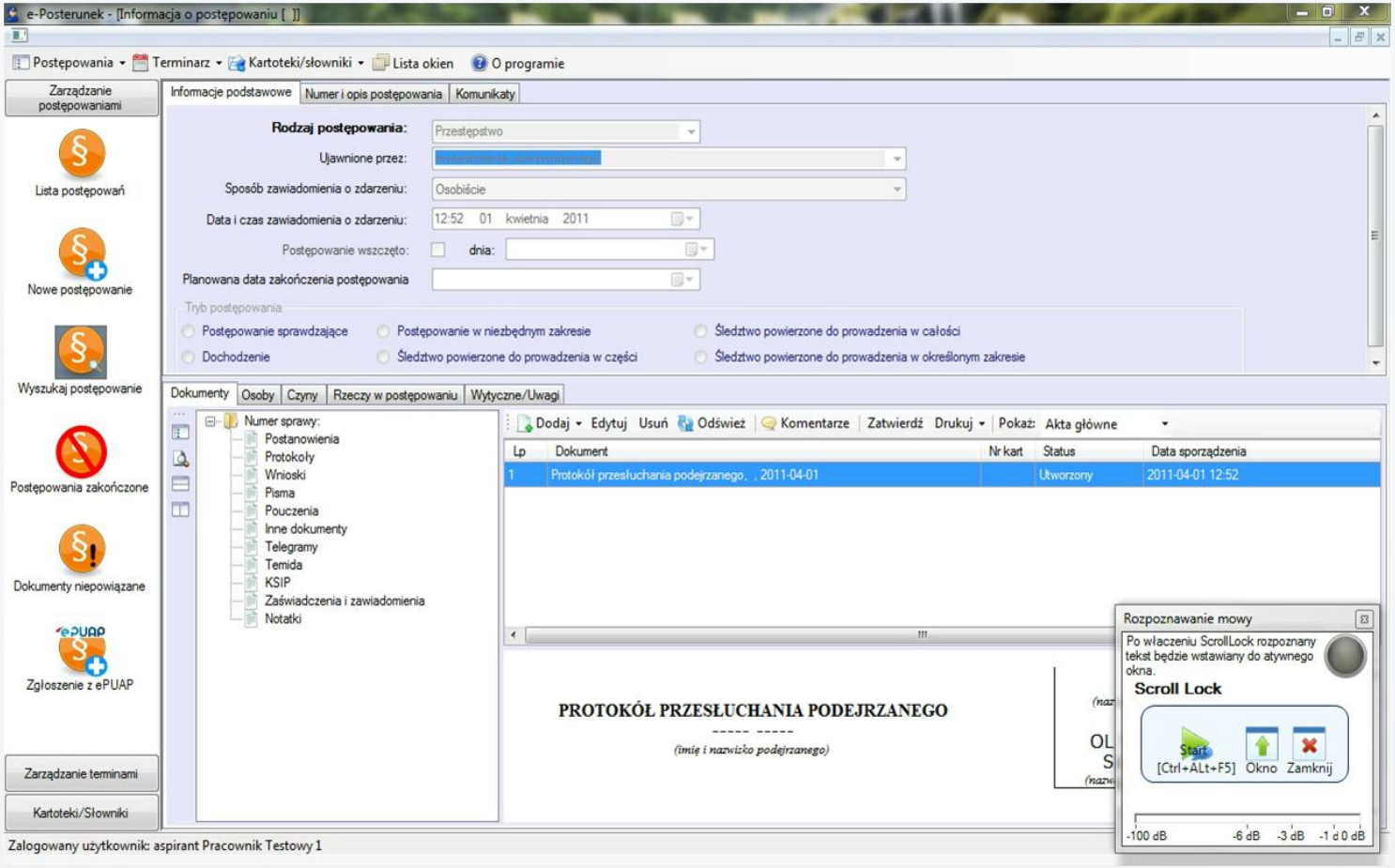


TESTY SYSTEMU

- Policja
- Sądownictwo
- Straż graniczna
- Wojsko
- Kancelarie prawne

Poszczególne służby i instytucje państwowe testują system: Policja, Straż Graniczna, Centralne Biuro Antykorupcyjne, Izba Celna, Wymiar Sprawiedliwości, a zainteresowane są Żandarmeria Wojskowa i Prokuratura.

e-Posterunek



Wyniki testów ARM (zad. A)

JEDNOSTKA		Średni zysk z wprowadzania tekstu przy użyciu ARM w stosunku do pisania na komputerze [%]		Zaangażowanie czasowe (czas przepisania dokumentu) [min.]
		Czasowy	Ilości znaków/min.	
KWP Poznań	1	33,3	53,3	24,0
	2	37,3	61,9	67,0
	3	11,7	31,6	22,7
	4	53,3	59,4	39,3
KWP Warszawa	5	16,4	22,2	128,0
	6	**-0,6**	**-0,5**	**83,5**
	7			
KWP Wrocław	8	**31,3**	**52,0**	**96,6**
	9	-1,3	-1,3	157,5
	10	13,4	19,0	35,0
KWP Szczecin	11	-5,8	-3,6	52,0
	12	-4,3	0,0	34,5
	13	-7,7	-7,1	13,0
KWP Kraków	14	**-57,3**	**-36,4**	**19,0**
	15			
	16			

Średni zysk czasowy ważony: **12%**, zysk w ilości znaków/min: **21%**

Porównanie średnich czasów wprowadzania tekstu przy użyciu ARM w stosunku do pisania za pomocą komputera (MAN). Tabela pokazuje procentowy zysk czasowy i zysk na szybkości pisania (ilości znaków/min.) oraz zaangażowanie czasowe poszczególnych osób testujących. Uwzględniono czas korekty tekstu.

Ankieta Oceniająca system ARM – wyniki [%]

W ankiecie wzięło udział 13 osób (3 kobiety, 10 mężczyzn).

1. Czy łatwo jest uruchamiać program ARM i rozpocząć w nim pracę?
2. Czy interfejs użytkownika (wygląd okna, informacje programu o postępie itp.) jest zrozumiały i jednoznaczny w interpretacji?
3. Czy nawigacja w programie jest łatwa? (czy czytelne są ikonki, menu itp.)
4. Czy jest zawsze zrozumiałe dla użytkownika, w jakim punkcie programu się znajduje, co się w danym momencie dzieje w programie?
5. Czy użytkownikowi łatwo jest zorientować się, że popełnił błąd w posługiwaniu się programem lub sprzętem (mikrofon itp.)?
6. Czy użytkownik może w prosty, szybki sposób wyjść z programu lub wyłączyć daną opcję programu (np. gdy użytkownik musi nagle wykonać inną pracę na tym samym stanowisku komputerowym, jeśli rozpoznawanie trwa zbyt długo)?

nr	TAK	RT	RN	NIE
1.	69	31	0	0
2.	92	8	0	0
3.	85	15	0	0
4.	77	23	0	0
5.	31	46	8	15
6.	23	46	15	15

7. Czy język i nazewnictwo użyte w menu, w pliku pomocy są zrozumiałe dla użytkownika?
8. Czy szybkość rozpoznawania mowy jest wystarczająca (czy czas, po którym program wyświetla wynik rozpoznawania jest odpowiedni)?
9. Czy w rozpoznawaniu mowy występują powtarzające się błędy, takie które są szczególnie uciążliwe dla użytkownika (jeśli tak, to proszę je opisać w polu Komentarz)?
10. Czy informacja o postępie podczas rozpoznawania jest wystarczająca?
11. Czy sposób przedstawienia wyniku rozpoznawania jest wystarczający (czy użytkownik oczekiwałby większej liczby wyników do wyboru itp.)?
12. Czy zastosowany słownik jest wystarczający (prosimy o uwagi w polu Komentarz, podanie brakujących słów)?
13. Czy rozpoznawanie mowy w obecnej wersji mogłoby już być użyteczne w pracy użytkownika lub w wybranych obszarach pracy, np. przy dyktowaniu tylko niektórych typów tekstów)?
14. Czy mikrofon dołączony do zestawu z programem *ARM* jest wygodny w użyciu?

nr	TAK	RT	RN	NIE
7.	69	23	8	0
8.	46	39	15	0
9.	46	31	15	8
10.	39	46	15	0
11.	54	39	8	0
12.	8	46	15	31
13.	15	69	0	15
14.	54	39	0	8

Właściwości systemu ARM

- **Jakość mowy dyktowanej ma wpływ na czas i poprawność rozpoznawania**: lepszy mówca wyższy procent rozpoznanych słów i krótszy czas dekodowania.
- **Adaptacja** do głosu mówcy i zastosowanego mikrofonu **zwiększa poprawność rozpoznawania** o ok. 5% dla dobrych mówców, dla niestarannych ten wpływ może wynosić nawet 30%. Adaptacja **skraca czas dekodowania**.
- Szybkość procesora komputera ma wpływ na szybkość rozpoznawania, nie ma wpływu na jakość.
- **Dedykowany mikrofon** zapewnia odpowiednią jakość działania systemu.
- Nie należy dyktować „jednym ciągiem”. Konieczne jest robienie pauz. Niezalecane jest jednak izolowanie pojedynczych wyrazów.
- System wymaga pozytywnego nastawienia i spokoju!

Main Features

- Designed tu run on Windows Operating Systems
- Based on Microsoft .NET 4.0 Platform
- Written in C# Language
- Use maximum of hardware resources
- Real time recognition (depending on quality preset)
- „Off-line” recognition
- Processing large amount of data
  - Up to 500 000 words in state network (currently 320 000)
  - 1 GB size of linguistic model
- Support for adaptation
- User friendly interface

Kontakt

Laboratorium Zintegrowanych Systemów Przetwarzania Języka i Mowy
Poznańskie Centrum Komputerowo-Sieciowe
ul. Zwierzyniecka 20, 61-612 Poznań
tel. +48 (61) 6682151, wewn. 31 lub +48 (61) 6682150, wewn. 32
mail.: speechlabs@speechlabs.pl
strona internetowa: *www.speechlabs.pl*

Międzynarodowe Targi Techniki i Wyposażenia Służb Policyjnych oraz Formacji Bezpieczeństwa Państwa Europoltech 2013, które odbędą się w dniach **17-19 kwietnia 2013 roku w hali EXPO XXI w Warszawie**.