

## Korpusy mowy i narzędzia do ich przetwarzania

---

**Danijel Korzinek, Krzysztof Marasek**

Polsko-Japońska Akademia Technik Komputerowych

Katedra Multimediów

[kmarasek@pjwstk.edu.pl](mailto:kmarasek@pjwstk.edu.pl)

[danijel@pjwstk.edu.pl](mailto:danijel@pjwstk.edu.pl)

2015-05-18

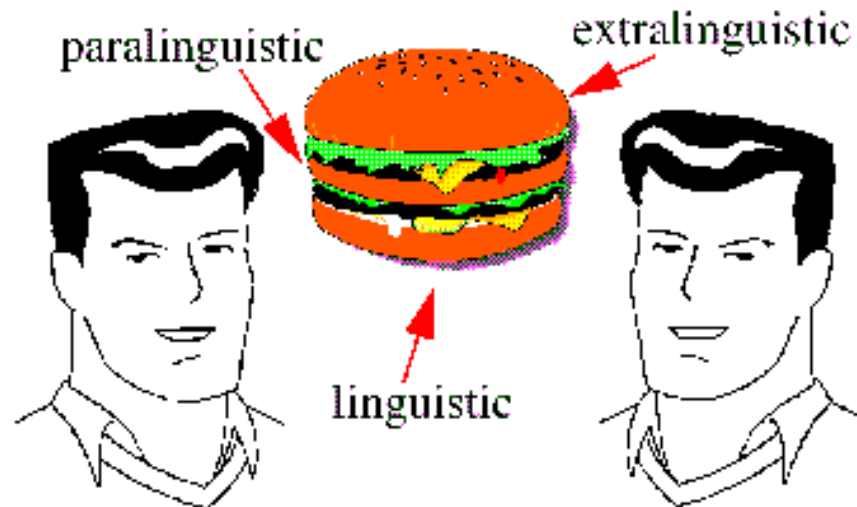


- Fonetyka
  - Badanie dźwięków mowy od strony ich artykulacji
    - Cechy akustyczne dźwięków mowy i ich percepcja
    - Cechy prozodyczne, jakość głosu
- Fonologia
  - Nauka o systemach dźwiękowych
    - segmentalna – segmenty/fonemy
    - Suprasegmentalna – dłuższe jednostki i prozodia
- Lingwistyka komputerowa
  - Wykorzystanie narzędzi komputerowych w analizie mowy
  - Technika
    - Automatyczne rozpoznawanie mowy, rozumienie mowy
    - Synteza mowy, naturalność syntezy
    - Korpusy mowy, ich analiza

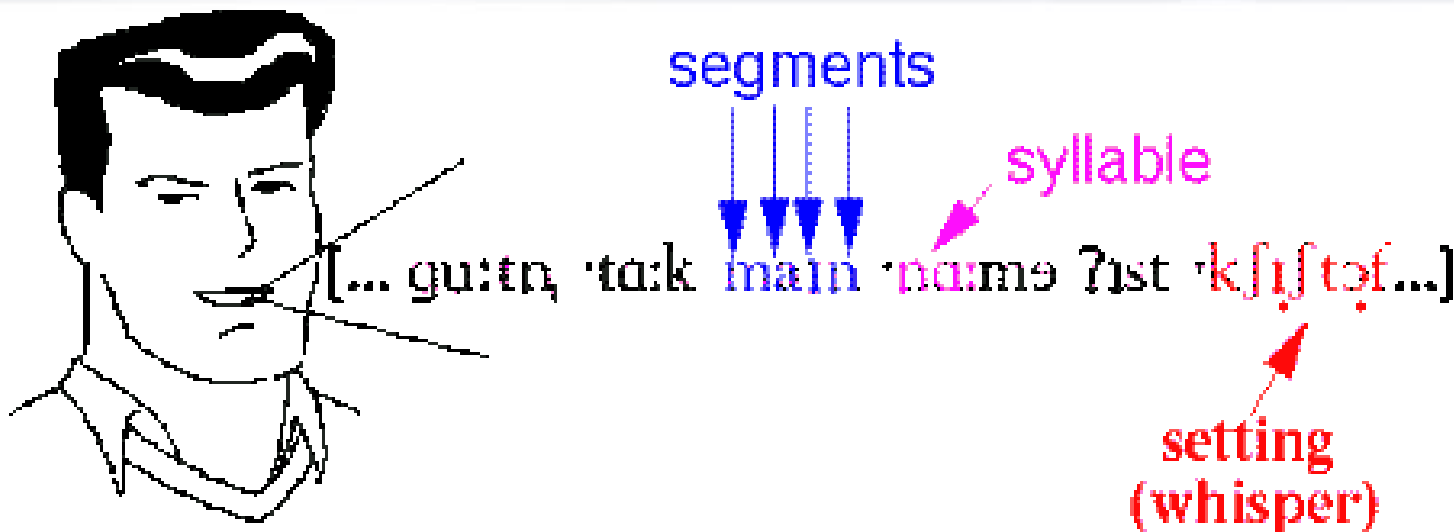
# Warstwowa struktura komunikacji



- Warstwa lingwistyczna – tekst wypowiedzi
- Warstwa paralingwistyczna – nasz stosunek do tego tekstu
- Warstwa ekstralingwistyczna – cechy mówcy, jego status społeczny, to co nas identyfikuje



# Segmentalna struktura mowy



- Feature – cecha odróżniająca od innych segmentów
- Segment – część o stałych cechach
- Syllable – sylaba, jednostka organizacji wypowiedzi
- Setting – cecha paralingwistyczna
- utterance – fraza wypowiedzi
- speaking-turn – cała wypowiedź

# Głoska, segment, fonem

---



- Głoska: najmniejsza, niepodzielna cząstka formy dźwiękowej języka.
- Cecha dystynktywna
  - odróżniająca dźwięki (głoski, segmenty) danego języka, przy czym różnica między segmentami pociąga za sobą różnicę w znaczeniu wyrazów (opozycja fonologiczna). Cechy dystynktywne funkcjonują zawsze w określonym języku
- Fonem
  - abstrakcyjna jednostka, będąca zbiorem cech dystynktywnych głoski. Fonem pełni w mowie rolę podobną do liter alfabetu.
  - Głoska jest to konkretna realizacja dźwiękowa danego fonemu. W przypadku regularnych kontekstowych zmian realizacji dźwiękowej danego fonemu taką realizację głoski nazywamy allofonem.

# Fonemy w języku polskim

---



- Analiza akustyczna mowy polskiej daje podstawy do przyjęcia, że w języku polskim mamy
  - 6 samogłosek ustnych
  - 2 samogłoski nazalizowane
  - 29 spółgłosek
- W sumie 37 fonemów

# Transkrypcja fonetyczna

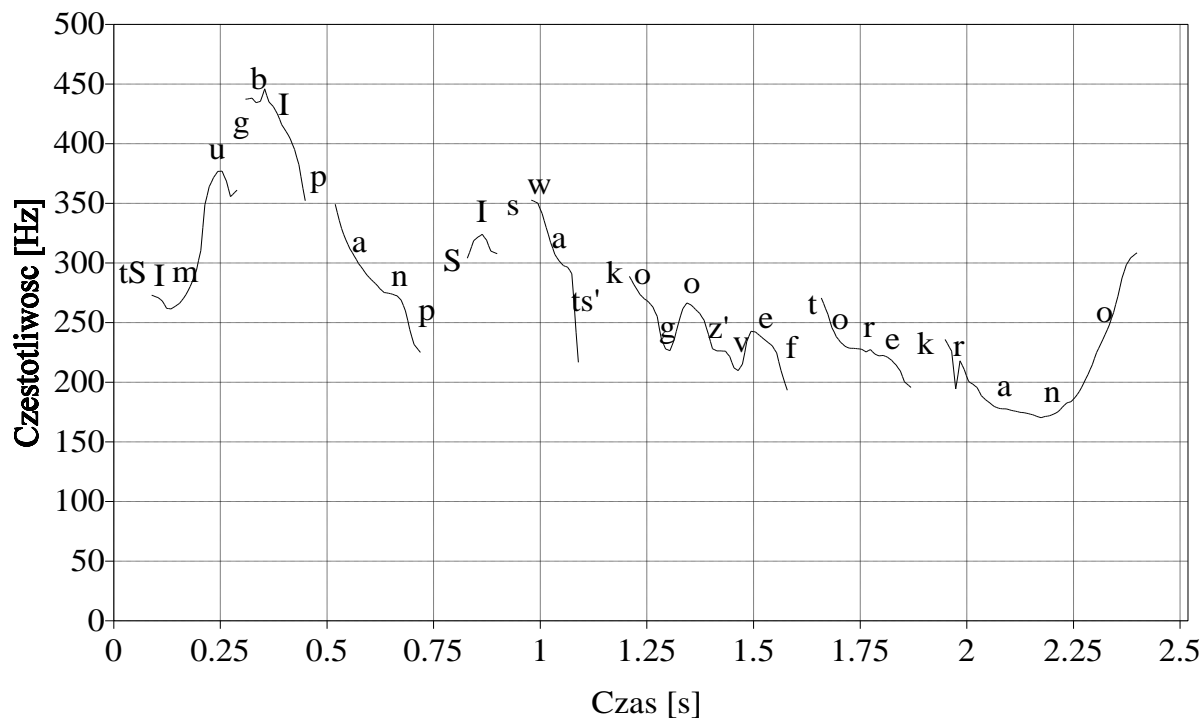
- Konwersja zapisu ortograficznego na fonetyczny
  - Różne alfabety (IPA, SAMPA..)
- Problemy transkrypcji
  - Warianty wymowy
  - Zjawiska fonetyczne
    - koartykulacja, upodobnienia pod względem dźwięczności, nosowości, miękkości, miejsca i sposobu artykulacji
    - Asymilacje (perseweracje i antycypacje)
    - Niedbała wymowa, regionalizmy

Symbol	Word	Transcription
PLOSIVES		
p	pat	pat
b	bat	bat
t	test	test
d	dym	dIm
k	kat	kat
g	gen	gen
AFFRICATES		
ts	coś	tsos'
dz	dzwon	dzvon
ts'	ćwicz	ts'fitS
dz'	dźwięk	dz'vje~k
tS	czyn	tSIn
dZ	dżin	dZIn
FRICATIVES		
f	fin	fin
v	waga	vaga
s	syk	sIk
z	zez	zes
S	szyk	SIk
Z	żyto	ZItO
s'	świt	s'fit
z'	źle	z'le
x	hak	hak
NASALS		
m	mak	mak
n	nasz	naS
n'	koń	kon'
N	gong	goNg
LATERAL		
l	luk	luk
APPROXIMANTS		
r	rak	rak
w	łuk	wuk
j	jak	jak
VOWELS		
i	tik	tik
l	typ	tIp
e	test	test
a	pat	pat
o	pot	pot
u	puk	puk
e~	tę	te~
o~	ta	to~

# Organizacja wypowiedzi

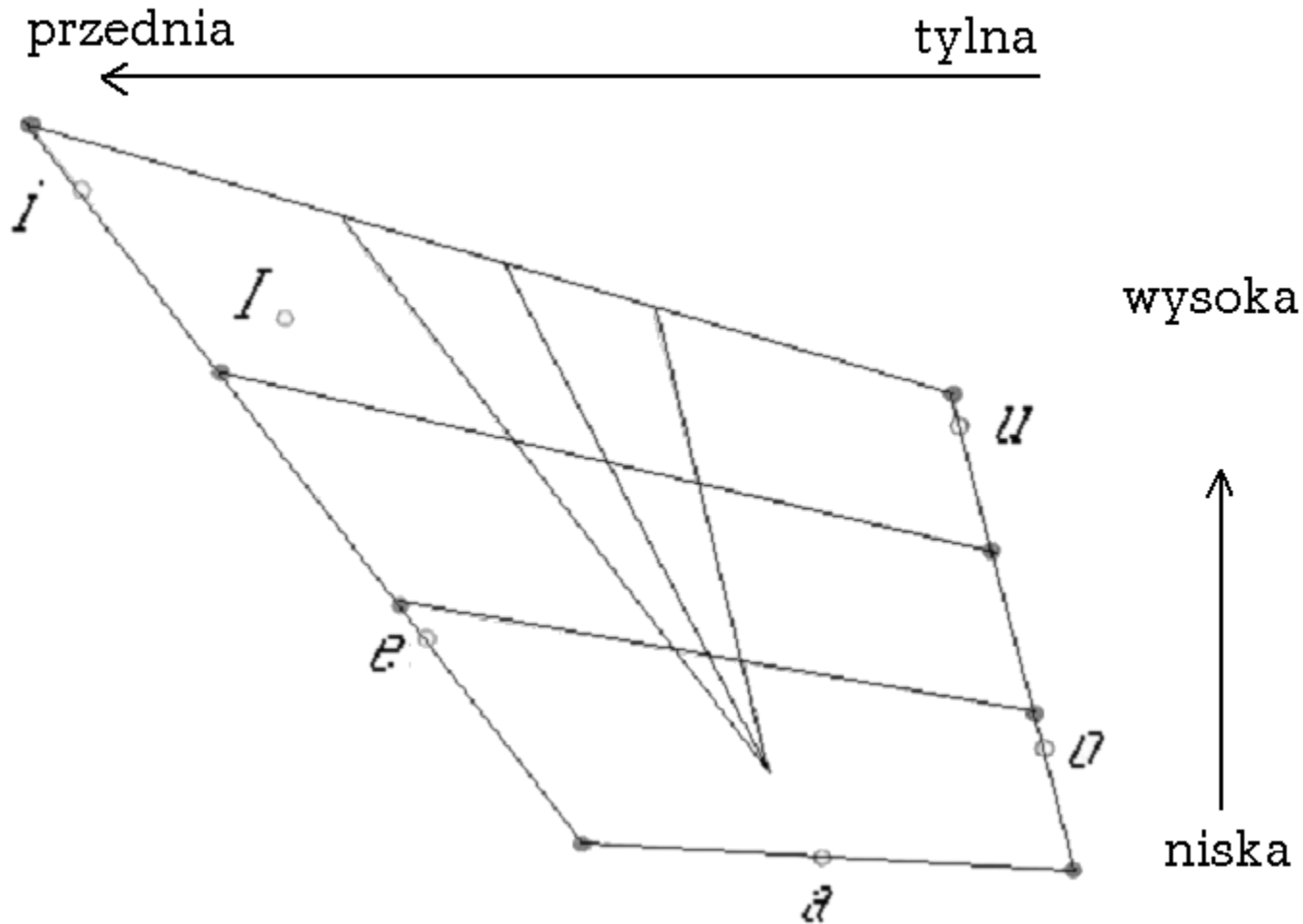


- Struktura płytka – opis akustyczny, składnia
  - Cechy segmentów akustycznych
- Struktura głęboka – semantyka
  - Prozodia – intonacja, głośność, sposób wypowiedzi





# Opis sygnału mowy: samogłoski



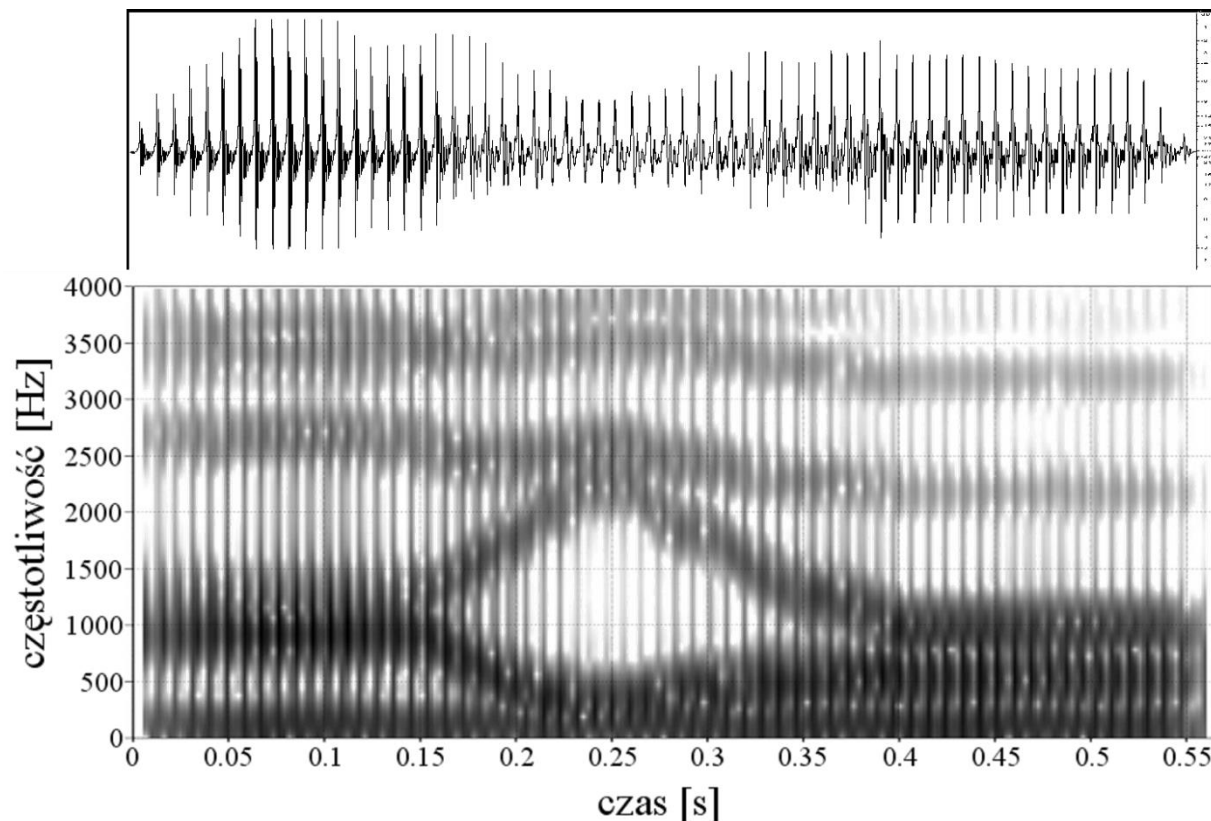
# Opis sygnału mowy: spółgłoski



# Analiza sygnału mowy

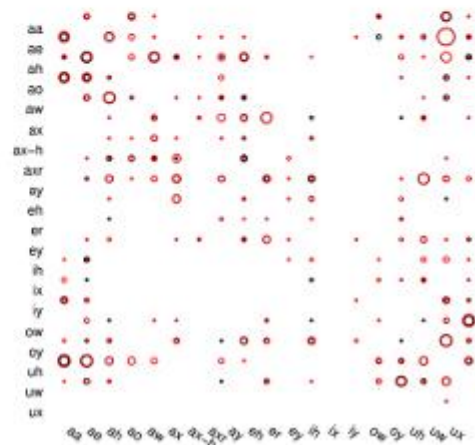
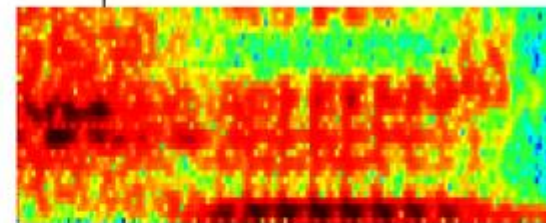
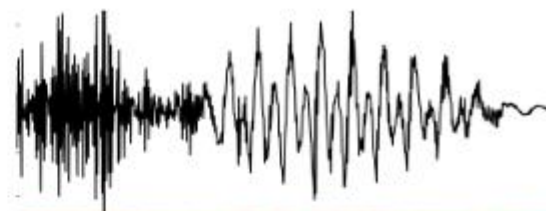
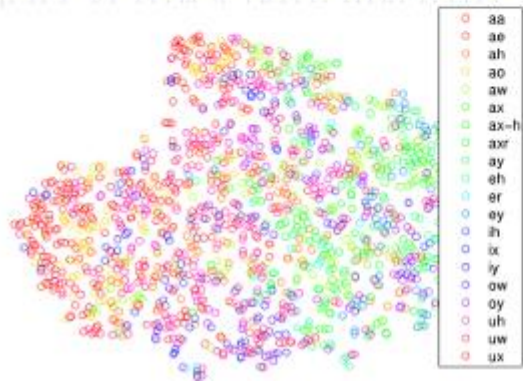
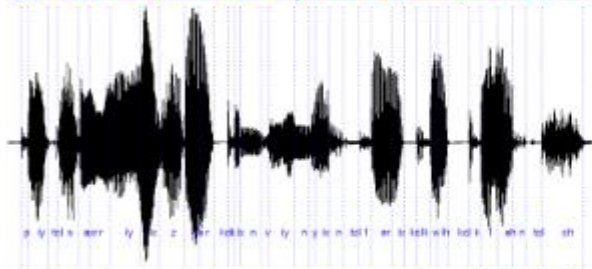
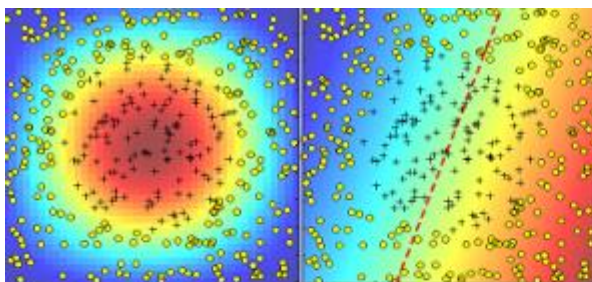
- Intensywność (głośność)
- Rodzaj pobudzenia – dźwięki harmoniczne, trące, zwarte
- Barwa dźwięku
- Iloczas

- Praat



# Mowa a inne dźwięki

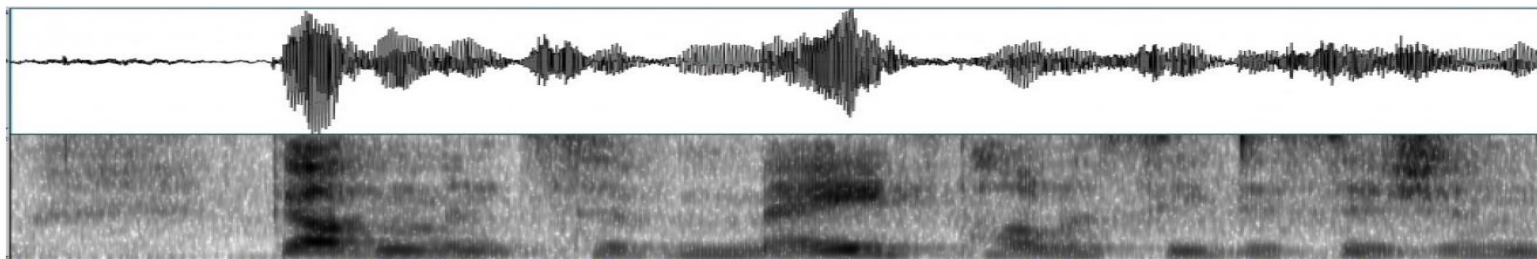
- Parametryzacja sygnału mowy
- Klasyfikatory statystyczne
- Uczenie maszynowe
- Podejście hierarchiczne
  - Detekcja mowy
  - Rozpoznawanie innych dźwięków



# Dopasowanie czasowe

- Dopasowanie tekstu do nagrań
  - Znany tekst, znane nagranie
  - Wykorzystanie technologii ASR
    - Tekst->transkrypcja fonetyczna->parametryzacja sygnału->wyszukanie najlepszych modeli->dopasowanie do sygnału

With transcription:  
Already know exactly what is in the audio.



↑ Align

sp	G	AH1	M	N	S	V	M	EY1	D	P	A1	L	S	IY0	D	I	H	S	I	ZH	N	Z
sp	GOVERNMENTS					HAVE		MADE		POLICY				DECISIONS								

# W przygotowaniu

---



- Wyszukiwanie słów kluczowych
- Fonetyczna baza danych

# CLARIN

Common Language Resources and Technology Infrastructure

---



Dziękuję za uwagę

---

# Diaryzacja mówców

- Podział nagrania na poszczególnych mówców i ich ew. rozpoznawanie
- Aktualne rozwiązanie bazuje na LIUM Speaker diarization toolkit (FR)
  - Własne modele mówców
  - Wyszukiwanie tego samego mówcy w nagraniach
  - Przy dużej bazie mówców konieczne duże zasoby i dużo cierpliwości

