

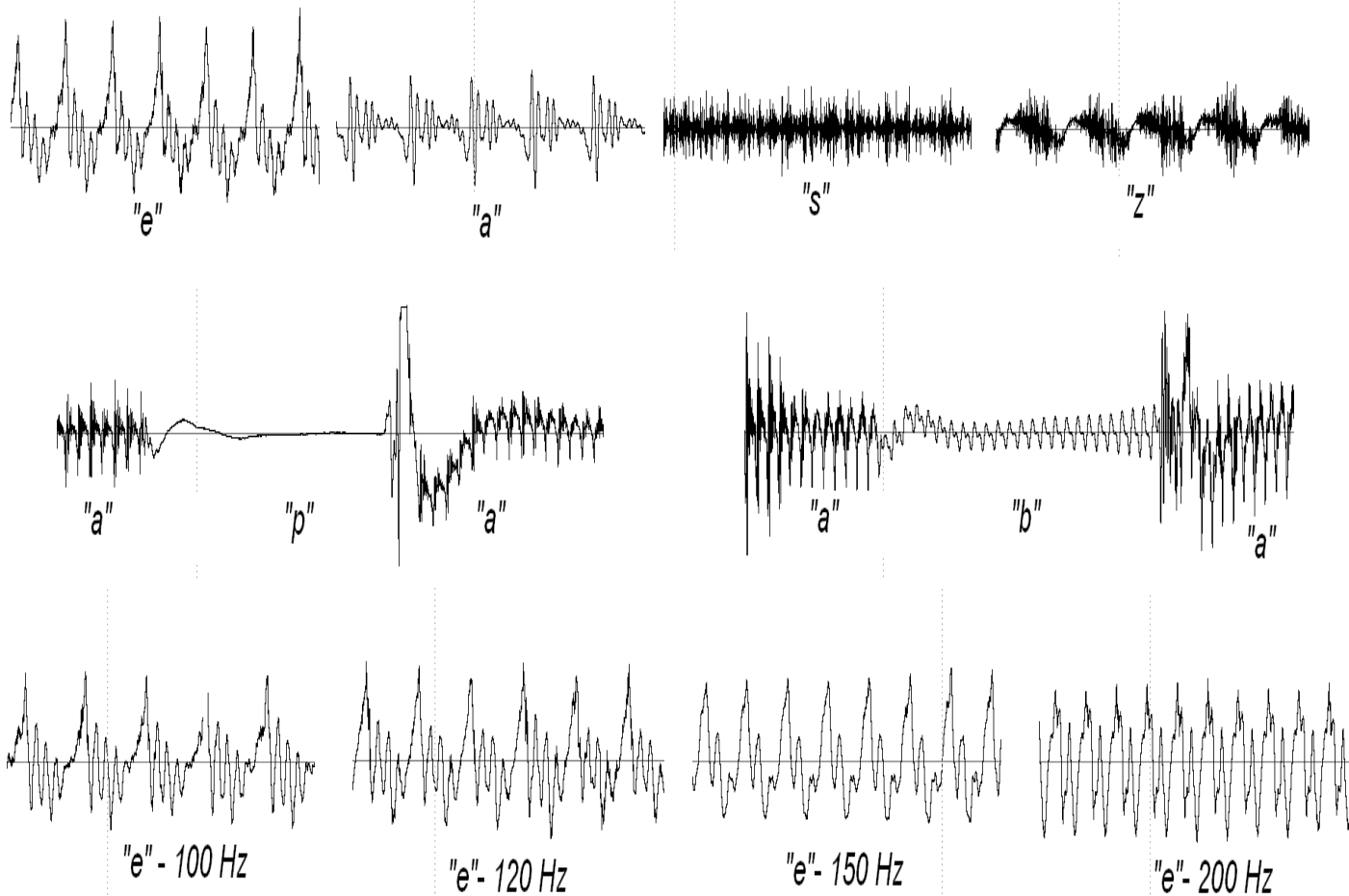


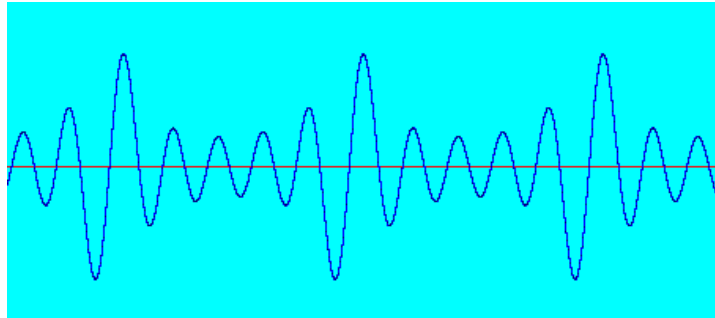
Cz. 2. Obraz sygnału mowy

*co i jak można odczytać z reprezentacji
graficznych – przypomnienie*



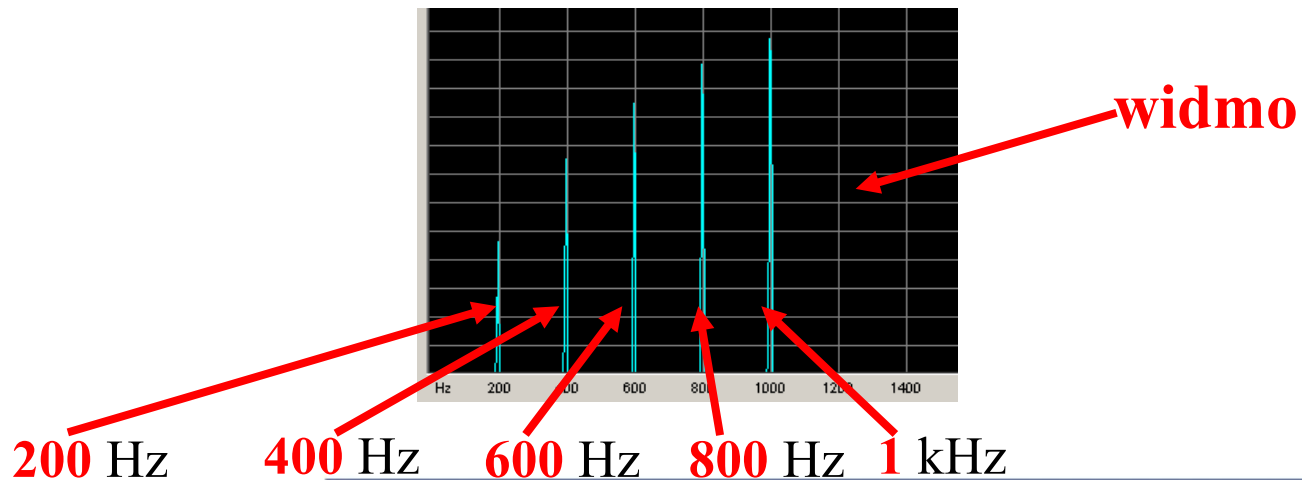
Jak wyglądają przebiegi fal głosowych różnych głosek





200 Hz + 400 Hz + 600 Hz + 800 Hz + 1 kHz

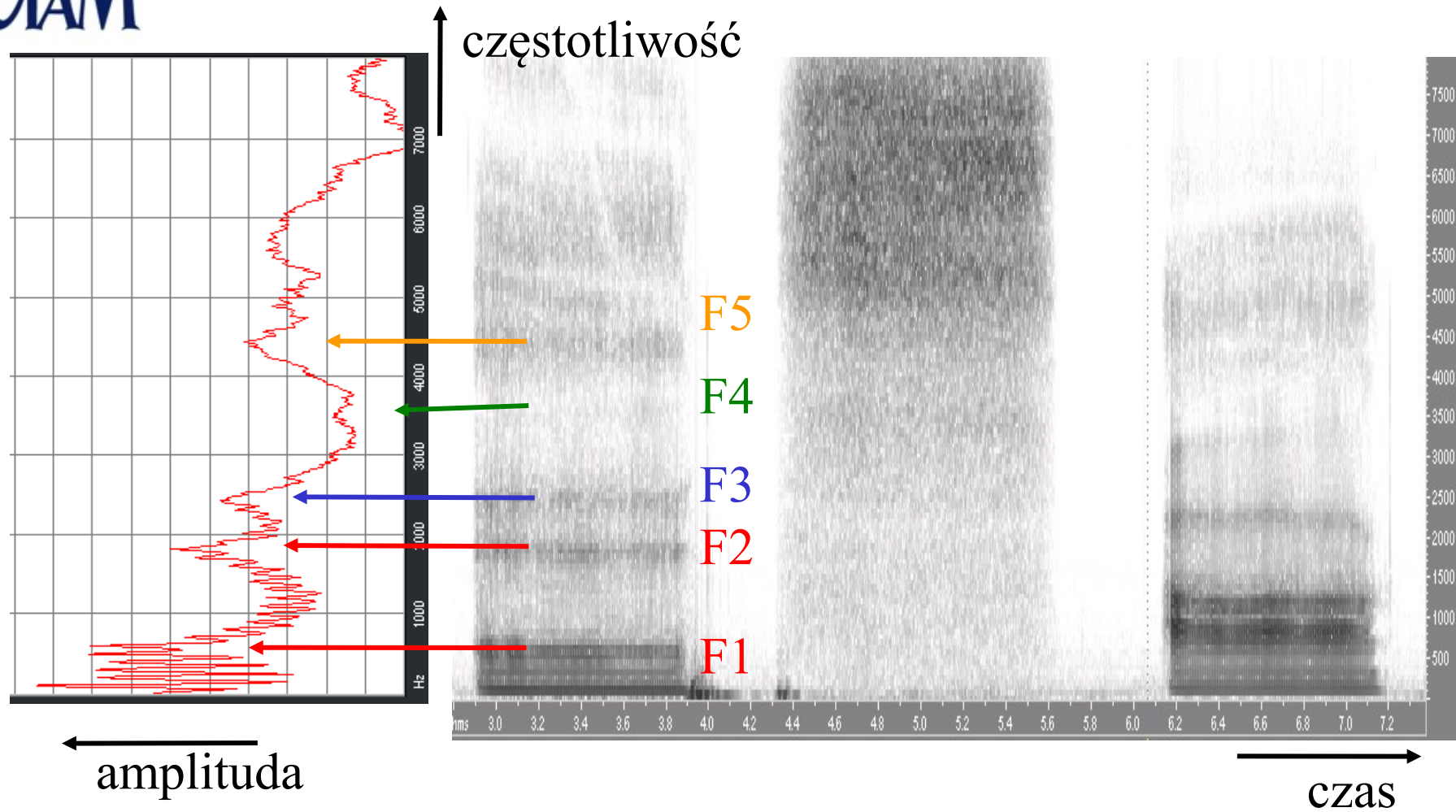
Czy istnieje sposób, aby można było
„**zobaczyć skład częstotliwościowy**” sygnału mowy



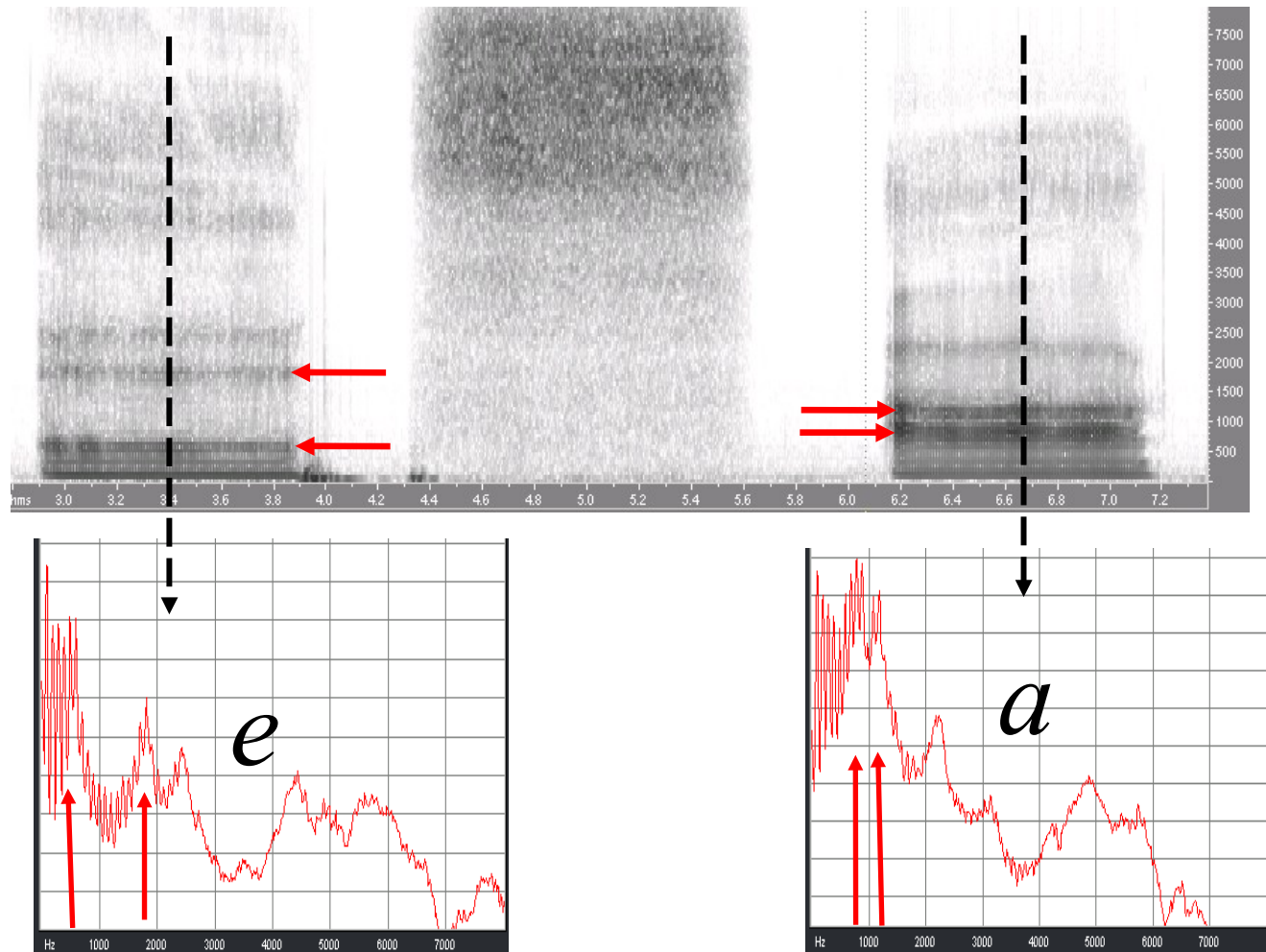


widmo

spektrogram

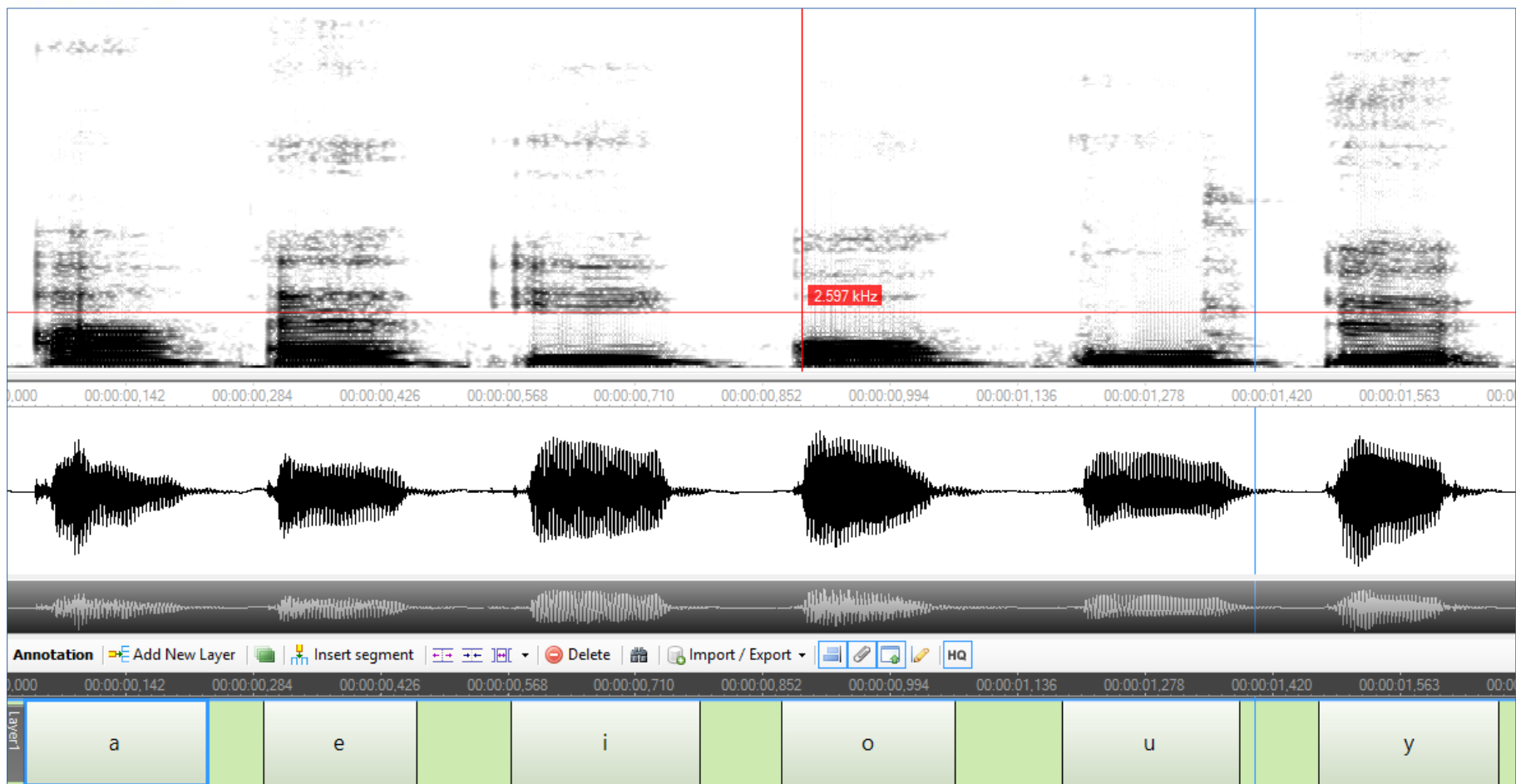


Czym różnią się spektrogramy różnych głosek?

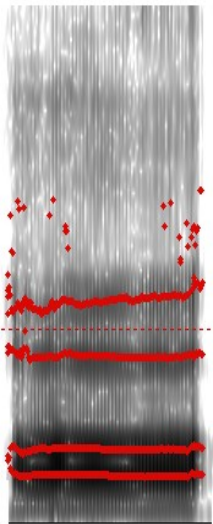




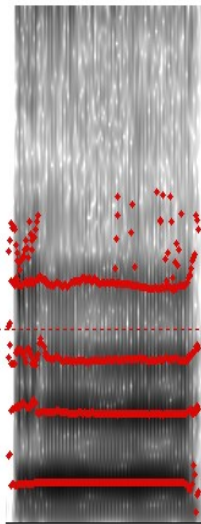
Spektrogramy samogłosek polskich



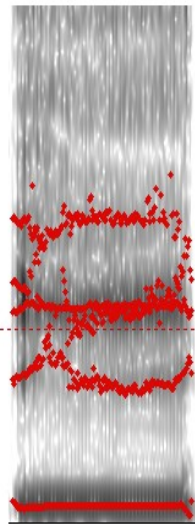
Spektrogramy samogłosek polskich



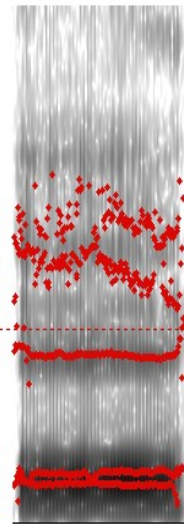
a



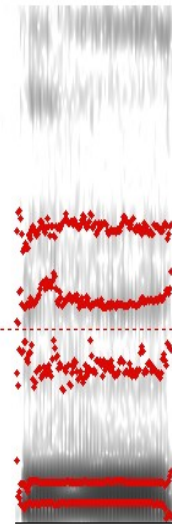
e



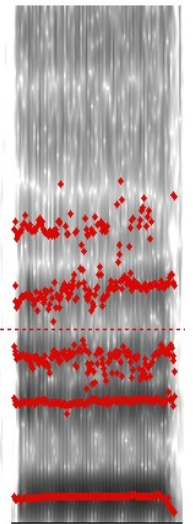
i



o



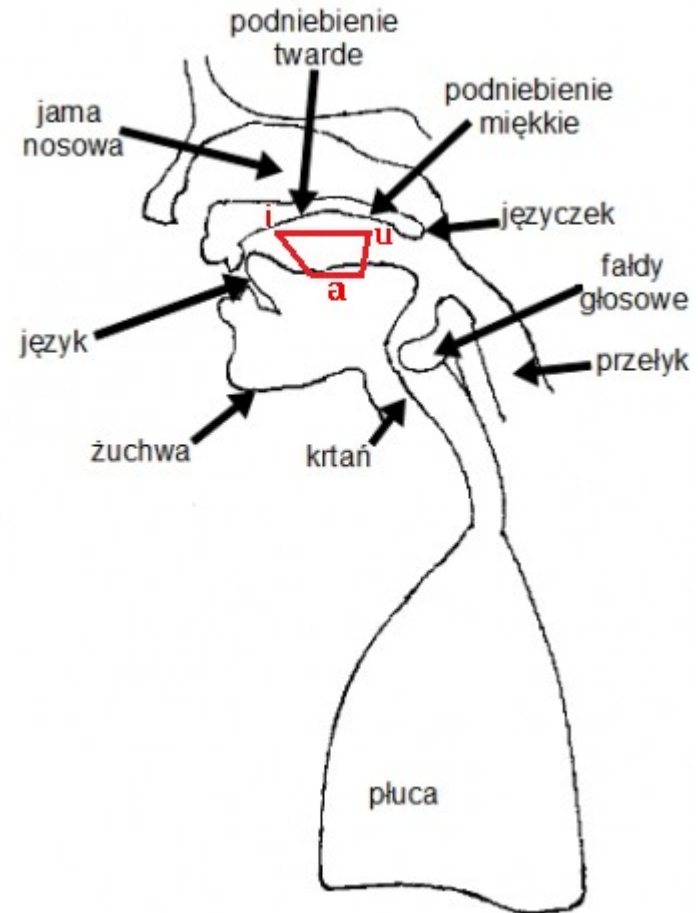
u



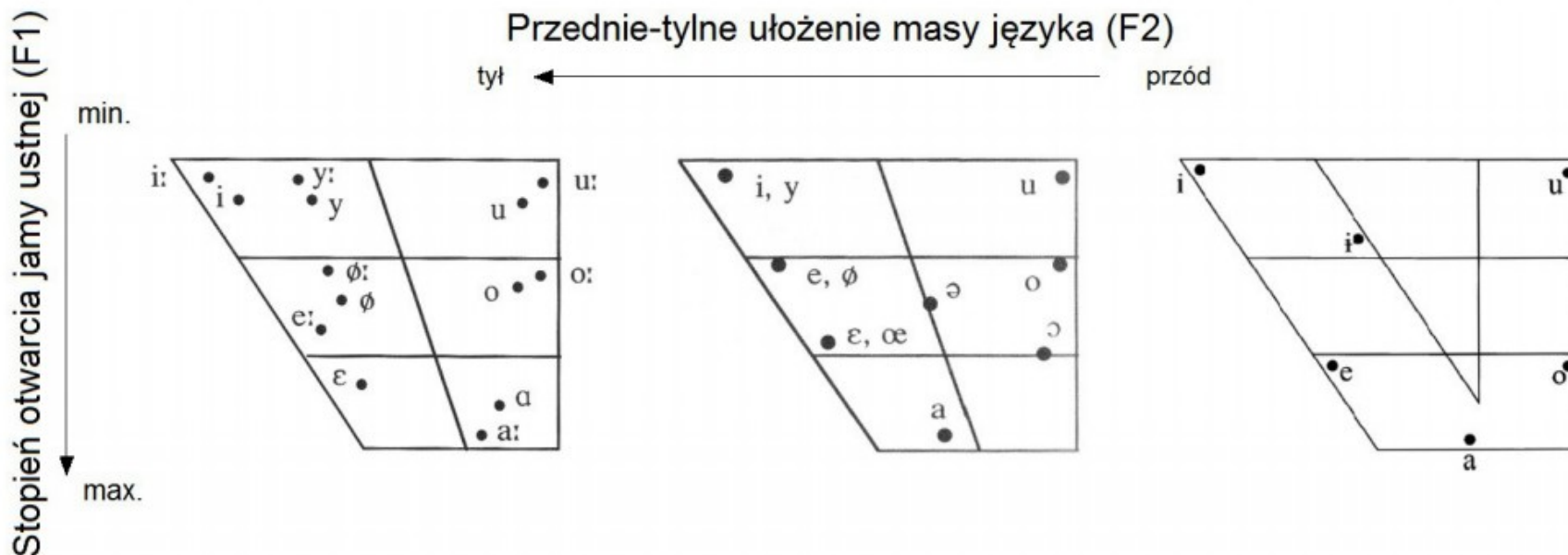
y

Percepcja samogłosek - **F1 i F2** - czworoboki

- Formanty a artykulacja (produkcja) samogłosek
- **F1** - stopień otwarcia jamy ustnej
- **F2** - ułożenie masy języka w płaszczyźnie przód-tył



IPA – inne inwentarze samogłosek

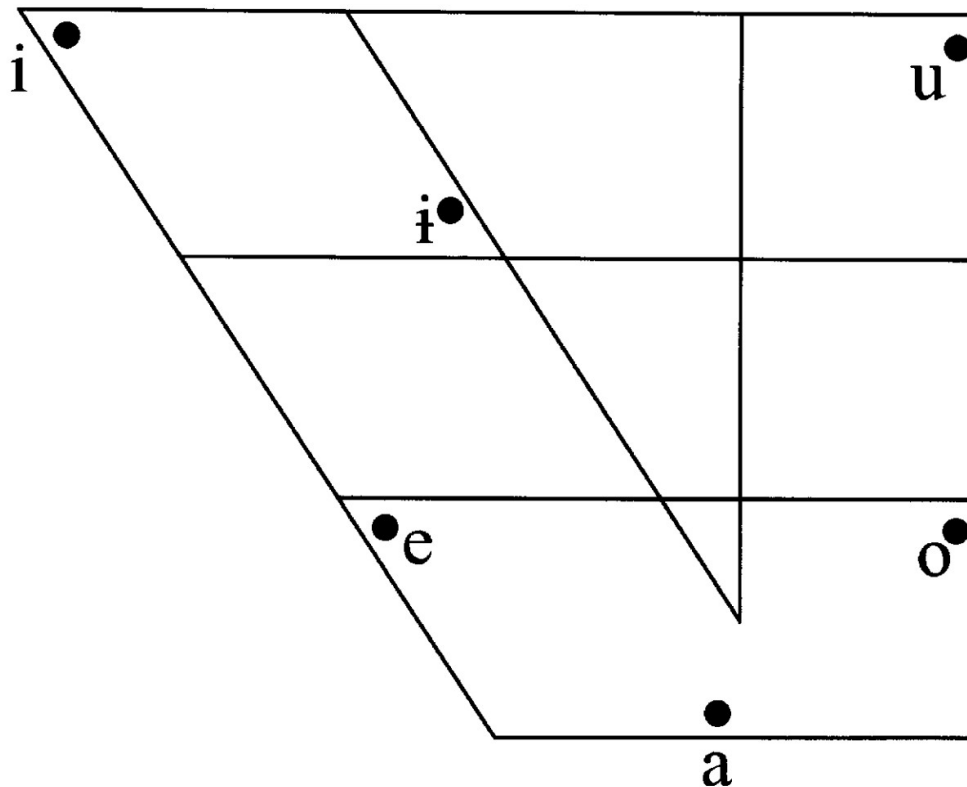


Ilustracja 1 Czworoboki samogłoskowe języków: węgierskiego (z lewej), francuskiego (środek), polskiego (z prawej); za: Illustrations of the IPA: French (Fougeron, Smith, 1999), Hungarian (Szende, 1999), Polish (Jassem, 2003).



Transkrypcja IPA

Międzynarodowy Alfabet fonetyczny (PL, Jassem 2005)





Transkrypcja, segmentacja, anotacja

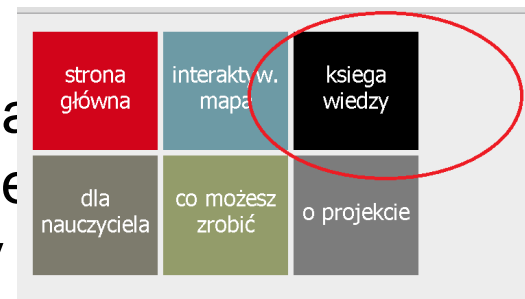
Aby dokonać pomiaru iloczasu, należy przeprowadzić segmentację sygnału mowy na głoski / sylaby i in., zależnie od potrzeb. Przy mniejszej ilości danych można to wykonać ręcznie, większa ilość danych wymaga automatyzacji procedur.

Jak można definiować te terminy?

-> zob. też np. Księga wiedzy (language)

-> -> rozdz. 10, sekcja 'Przetwarzanie'

-> -> rozdz. 4, sekcja 'Dźwięki mowy'

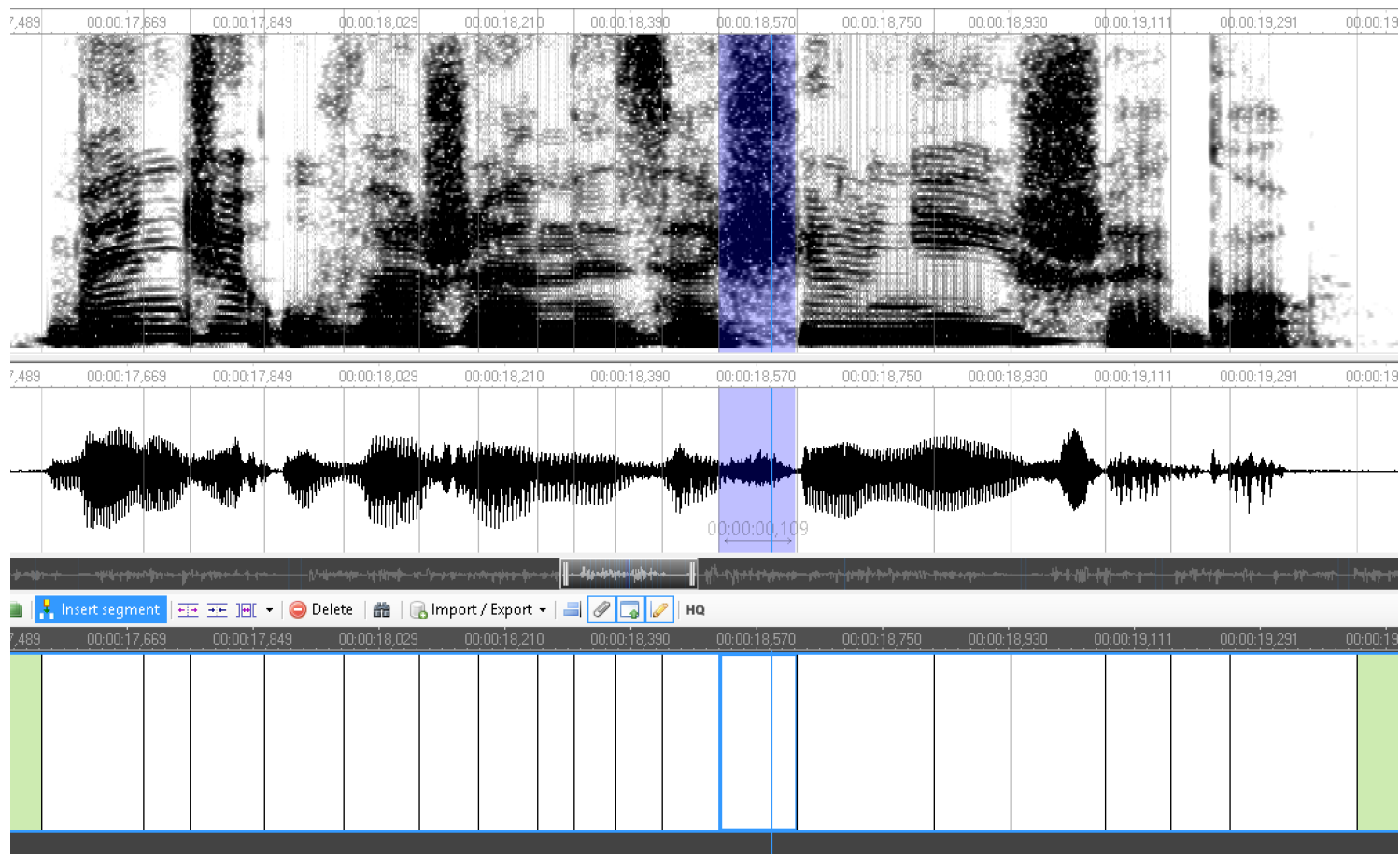




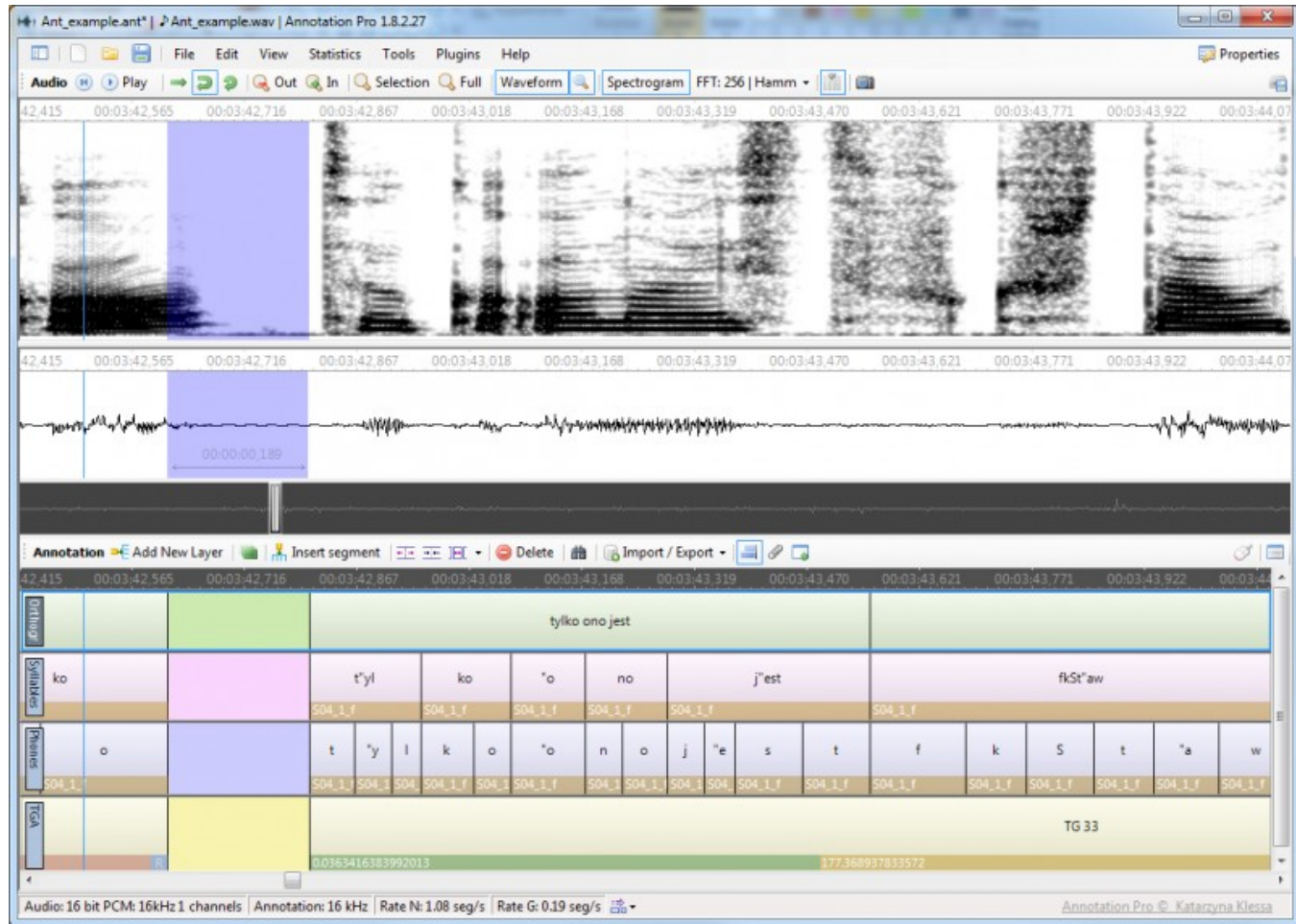
Transkrypcja, segmentacja, anotacja

pe`vnego ,razu || puw`notsni °vjatr i,swontse || spje>tʃaliçe || `ktoznix
jest cil,nejʃi || `vwaçne pʃe°xodʒiw °drogoŵ jaciç >tʃwovjek || ovi`penti
°ftçepwi ,pwaʃtʃ || `umu ,viliçe °vjents || zetenznix °kturi `pjerʃi zmuçi
pʃexodzon,tsego || abi`zdjow o`kritçe || `bejdze uva°zani zaçilnej,sego ||
puw`notsni °vjatr °zatʃow od`razu °doptç s`tsawej ,çiwi || aleim`motsnej
>dow || tim`çilnej po°družni o°tulawçe ,fpwaʃtʃ || `vreʃtçe puw°notsni
°vjatr daw,,spokuj || `ftedi °swontse za`tʃewo pʃi,gzevaç, || af'xfile
'pužnej po'družni zdjow,pwaʃtʃ || 'ften >sposup || puw`notsni °vjatr `muçaw
>pʃiznatç || ze`swontse jestçil°nejʃe od,,nego

Transkrypcja, segmentacja, anotacja



Transkrypcja, segmentacja, anotacja





Transkrypcja IPA

Międzynarodowy Alfabet fonetyczny (PL, Jassem 2005)

	Labial	Labiodental	(Post)dental	Alveolar	Alveolo-palatal	Palatal	Velar
Plosive	p b		t d			c ɟ	k ɡ
Fricative		f v	s z	ʃ ʒ	ç ʒ		x
Affricate			ts dz	tʃ dʒ	tɕ dʑ		
Nnasal	m		n		ɲ		ŋ
Lateral			l				
Flap/Trill				r			
	Front			Back			
Approximant	j (ĩ)			w (ũ)			



Cz. 3. Iloczas i rytm



Czas trwania dźwięków mowy, iloczyn

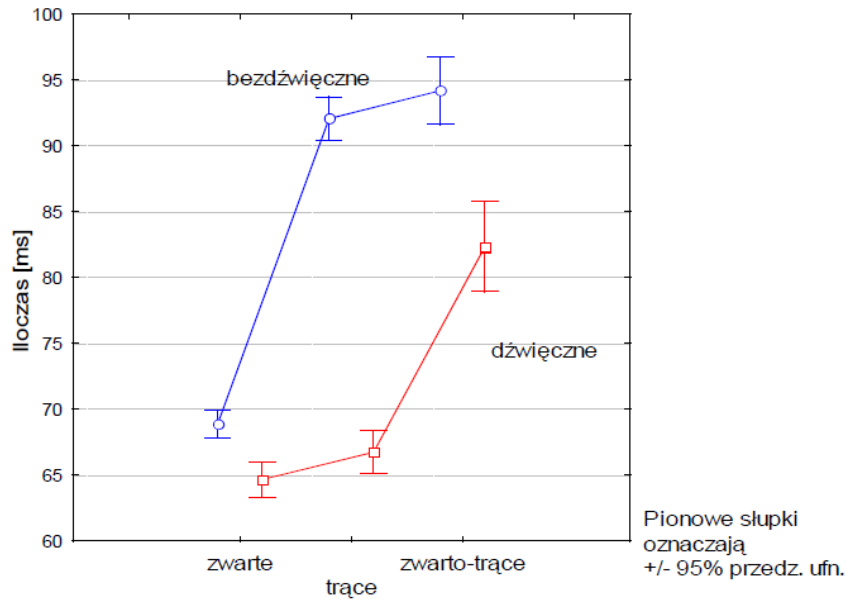
Czas trwania dźwięków mowy – elementy językowe vs. inne dźwięki

Iloczyn głoskowy – średnie iloczyny głosek

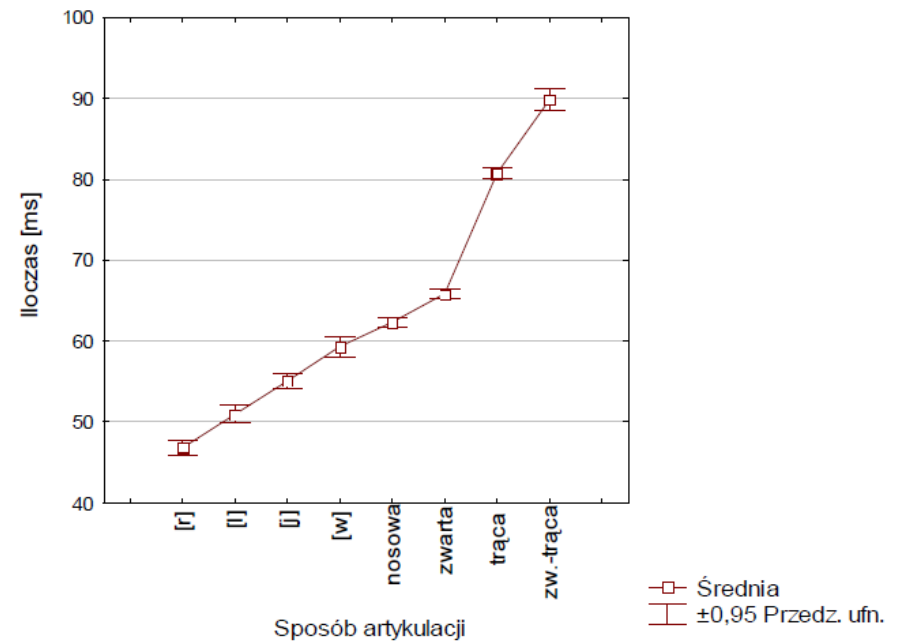
Iloczyn właściwy – rozróżnianie znaczeń

Iloczyn a rytm mowy

Czas trwania dźwięków mowy, iloczyn

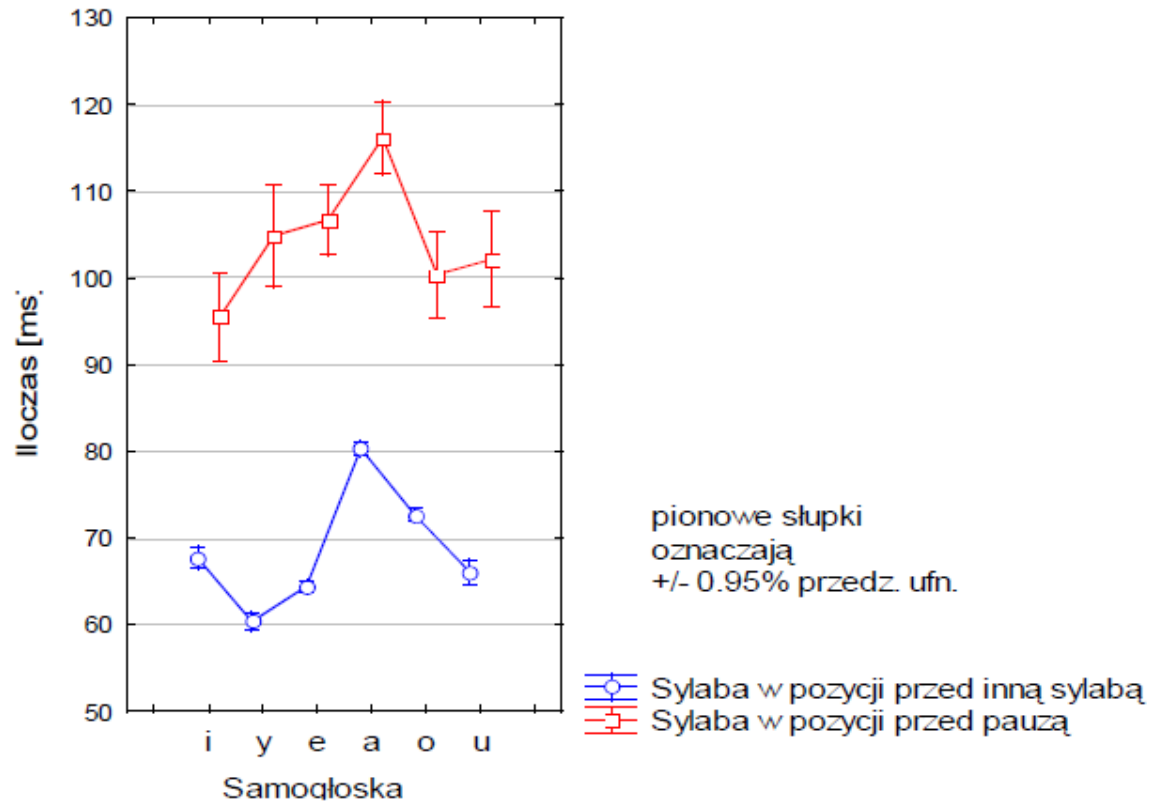


Rycina 4.6. Wartości średnie iloczasu głoskowego dla spółgłosek zwartych, trączych oraz zwarto-trączych skategoryzowanych ze względu na dźwięczność



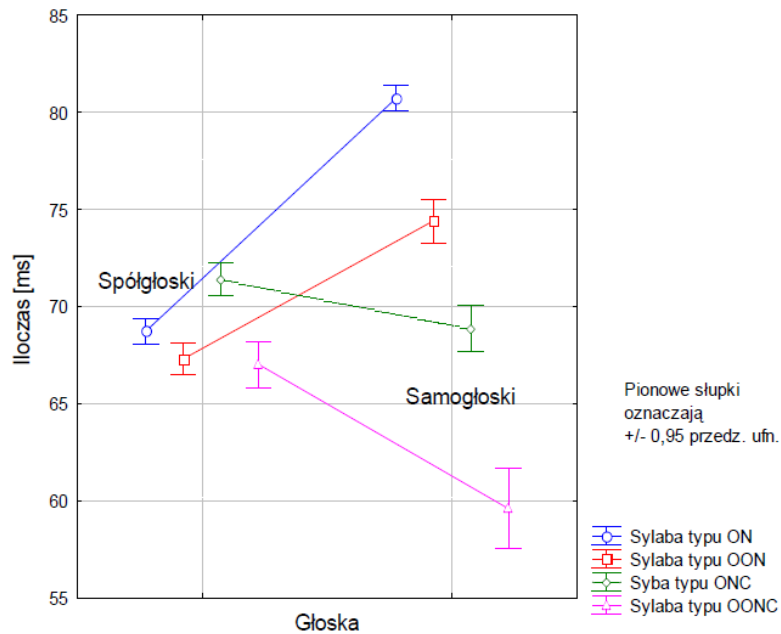
Rycina 4.8. Wykres średnich wartości iloczasu głoskowego dla spółgłosek polskich w zależności od sposobu artykulacji

Czas trwania dźwięków mowy, iloczyn

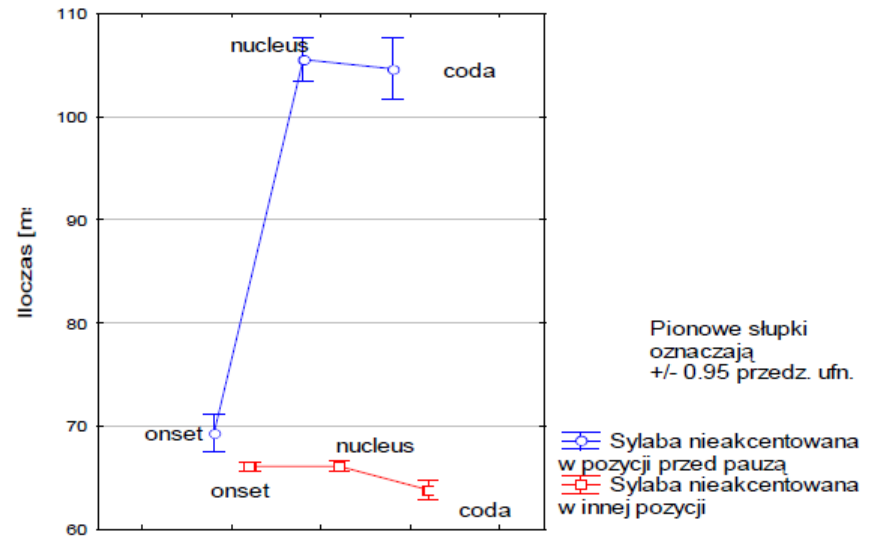


Rycina 4.10. Wartości średnie iloczasu samogłosek głównych w zależności umiejscowienia w stosunku do pauzy

Czas trwania dźwięków mowy, iloczas

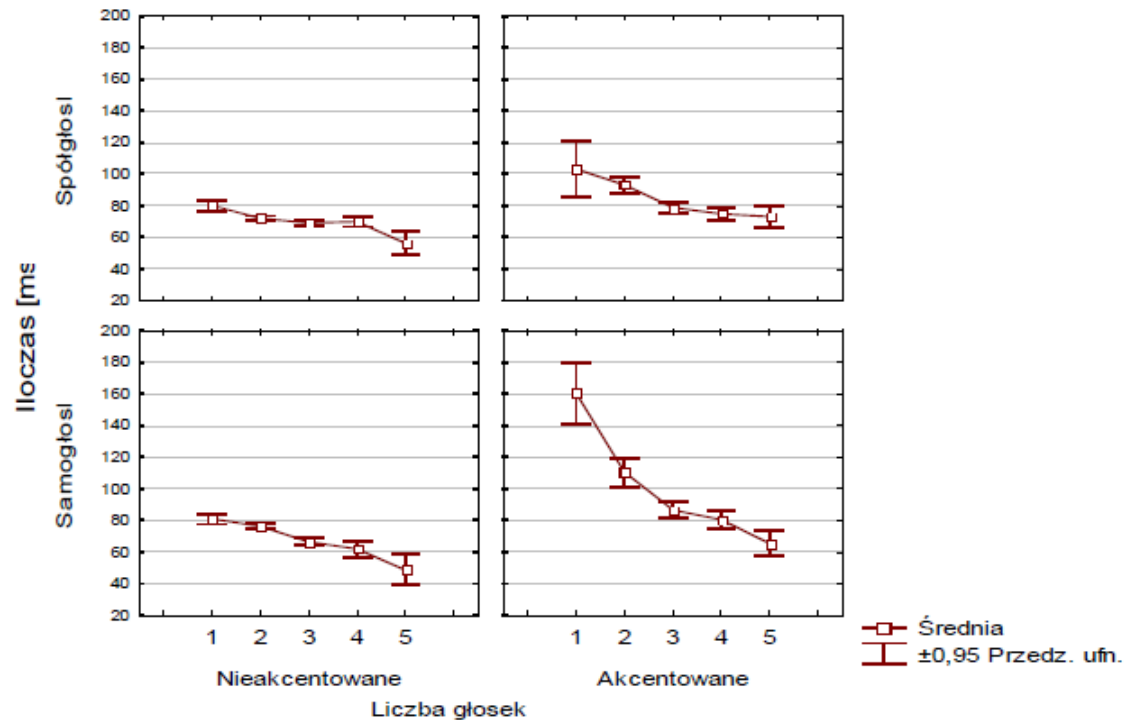


Rycina 4.16: Zmienność iloczasu samogłosek i spółgłosek w zależności od rodzaju sylab (cztery najczęstsze rodzaje sylab)



Rycina 4.21. Wpływ położenia sylaby nieakcentowanej w pozycji przed pauzą na iloczasy jej elementów składowych (w nagłosie - onset, jądrze - nucleus oraz wygłosie sylaby - coda)

Czas trwania dźwięków mowy, iloczyn



Rycina 4.26. Wartości iloczasu głoskowego w zależności od czynników: akcent, rodzaj segmentu i długość wyrazu w głosekach w wyrazach jednosylabowych



Zjawiska iloczynowe a cechy (para)językowe mowy - przykłady

Istotnie wydłużone czasy trwania dla pauz wypełnionych w porównaniu z czasem trw. głosek, sylab.

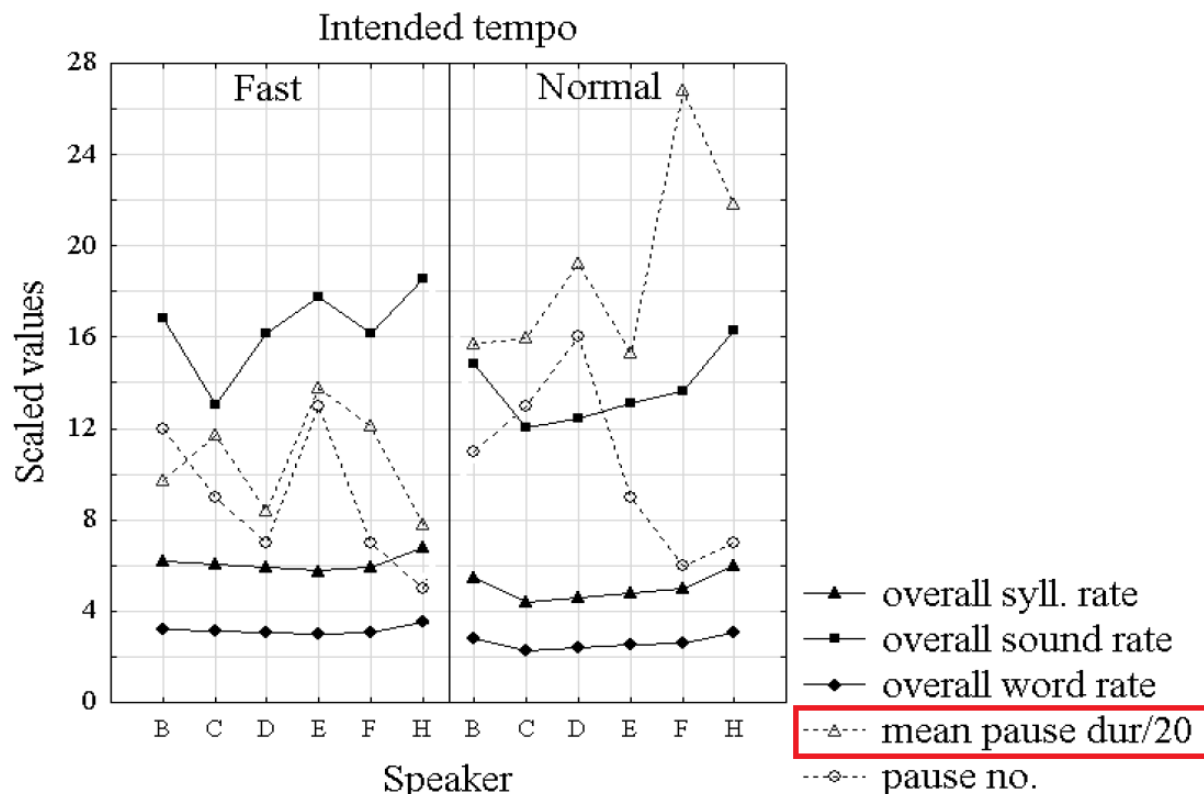
Tempo mowy a percepcja (nie) pewności w głosie – wolniejsze tempo, większa percypowana pewność; ale: skrajnie wolne – niepewność.

Przyspieszanie/zwalnianie w obrębie grup iloczynowych jako cecha różnych stylów wypowiedzi.

Skłonność do synchronizacji tempa mowy u rozmówców w dialogach.

Charakterystyka rytmiczna języków w oparciu o własności iloczynowe – różne “metra” opisujące rytm. Podział języków na syllable-timed, stress-timed. J. polski – problematyczny (wys. złożoność fonotaktyczna jak w stress-timed, por. angielski, niemiecki, ale wykazuje stosunkowo niewiele redukcji – jak w syllable-timed, por. włoski, francuski).

Zjawiska iloczynowe a cechy (para)językowe mowy - przykłady



Mean values of: syllable, sound, and word rates, number and duration of pauses for six speakers in fast and normal intended speech tempo.

Zjawiska iloczynowe a cechy (para)językowe mowy - przykłady

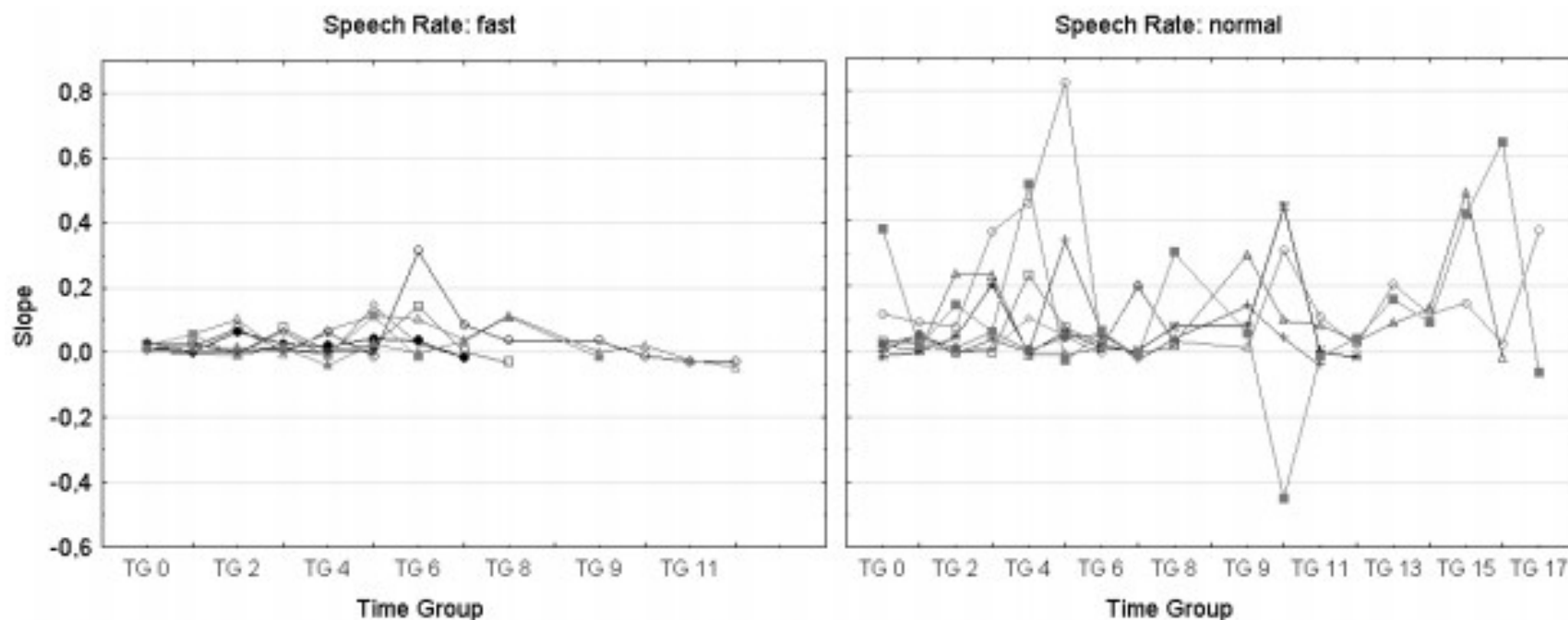


Figure 7: *Slope* variability in Polish read speech for 8 speakers in two tempi. Each line chart represents *slope* variability for interpausal groups realized by one speaker (Data: Paralingua, Emo sub-corpus; Klessa et al., 2013b).

Zjawiska iloczynowe a cechy (para)językowe mowy - przykłady

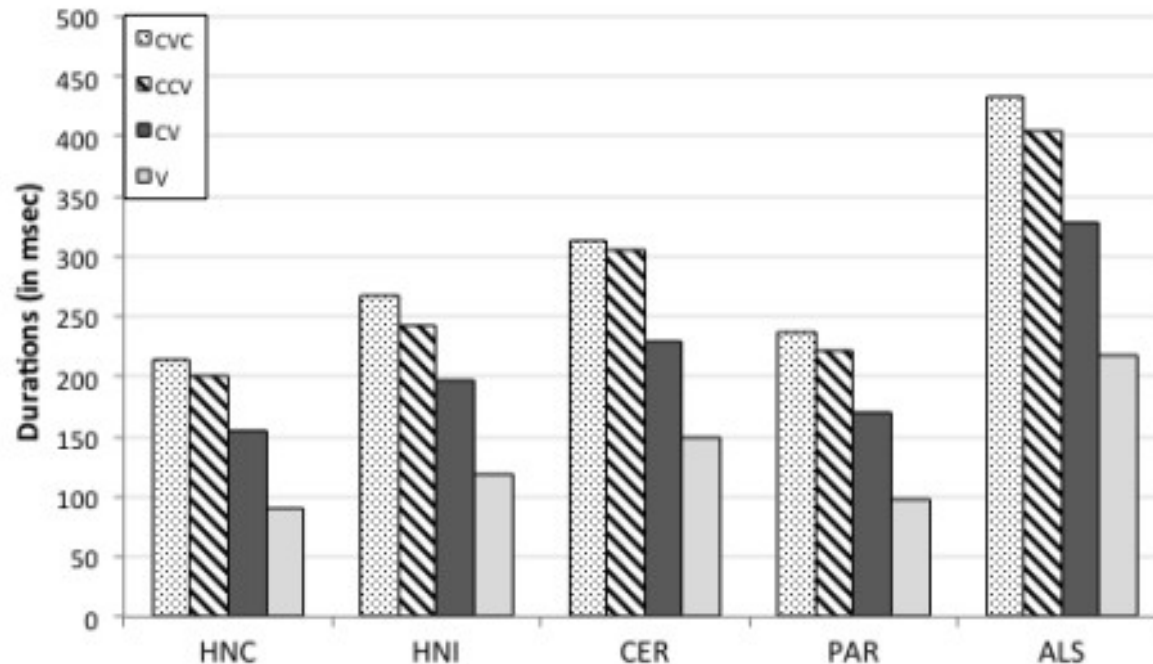


Figure 3: *Durations of syllables structure (CVC, CCV, CV, V) according to populations considered in this study (HNC, HNI, CER, PAR, ALS), in msec.*



Badanie iloczasu wymaga segmentacji nagrań

Aby dokonać pomiaru iloczasu, należy przeprowadzić segmentację sygnału mowy na głoski / sylaby i in., zależnie od potrzeb.

Przy mniejszej ilości danych można to wykonać ręcznie, większa ilość danych wymaga automatyzacji procedur.

Przykładowe narzędzia dostępne darmowo: SPPAS (Annotation Pro importuje / eksportuje pliki XRA wygenerowane w SPPAS), narzędzie on-line: WebMaus



Ćwiczenia

Segmentacja sygnału mowy na głoski na podstawie wizualnej analizy spektrogramu (Praat, Annotation Pro).

(Transkrypcja fonetyczna z wykorzystaniem SAMPA:
<https://www.phon.ucl.ac.uk/home/sampa/polish.htm>,
optymalnie: wersja rozszerzona).



Cz. 6. Wybrane typy analiz w oknach czasowych

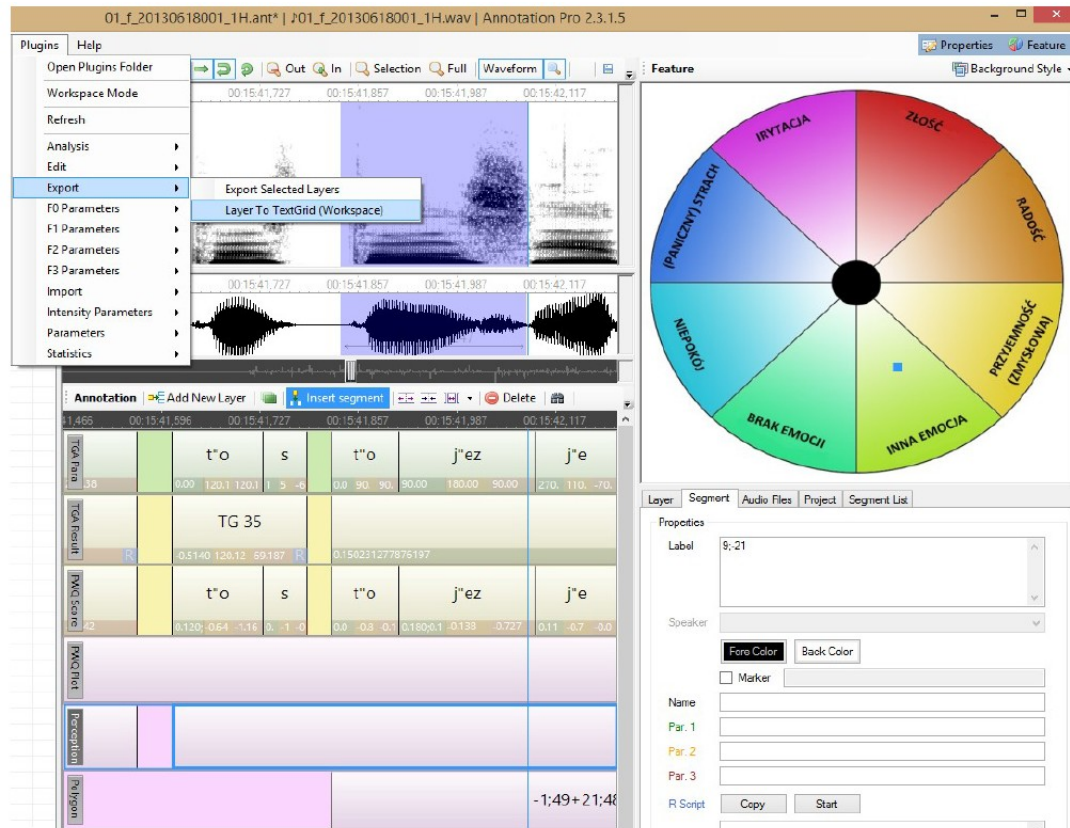
Pluginy Annotation Pro

SRMA – synchronizacja tempa w dialogach

TGA – zmienność iloczynowa w grupach czasowych



Annotation pro – pluginy (<http://annotationpro.org/plugins/>)



Annotation Pro: Plugins menu & user interface with example plugin result layers.



Annotation Pro – szablon pluginu

```
using System;
using System.Windows.Forms;
using AnnotationPro.Plugin;
using AnnotationPro.Logic;
using AnnotationPro.Presentation;

namespace AnnotationPro.Plugin
{
    public class AnnotationPlugin : IAnnotationPlugin
    {
        public void Run(AnnotationEditor editor)
        {
            // BEGIN USER CODE

            editor.Cursor = Cursors.WaitCursor;

            MessageBox.Show("Hello World! I'm your new plugin! I'm in 'Documents\\Annotation
Pro\\Plugins' folder.", "Annotation Pro Plugin");

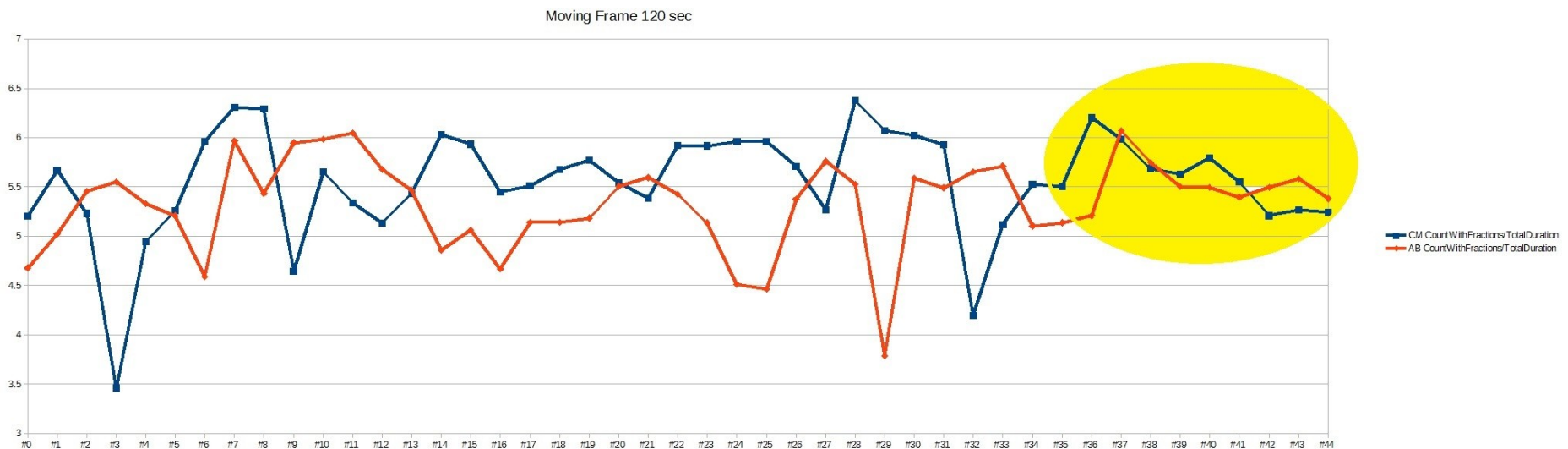
            editor.Cursor = Cursors.Default;

            // END USER CODE

        }
    }
}
```



SRMA – Speaking Rate Moving Average – synchronizacja tempa w dialogach



approx. 15-20 min



SRMA – Speaking Rate Moving Average – synchronizacja tempa w dialogach

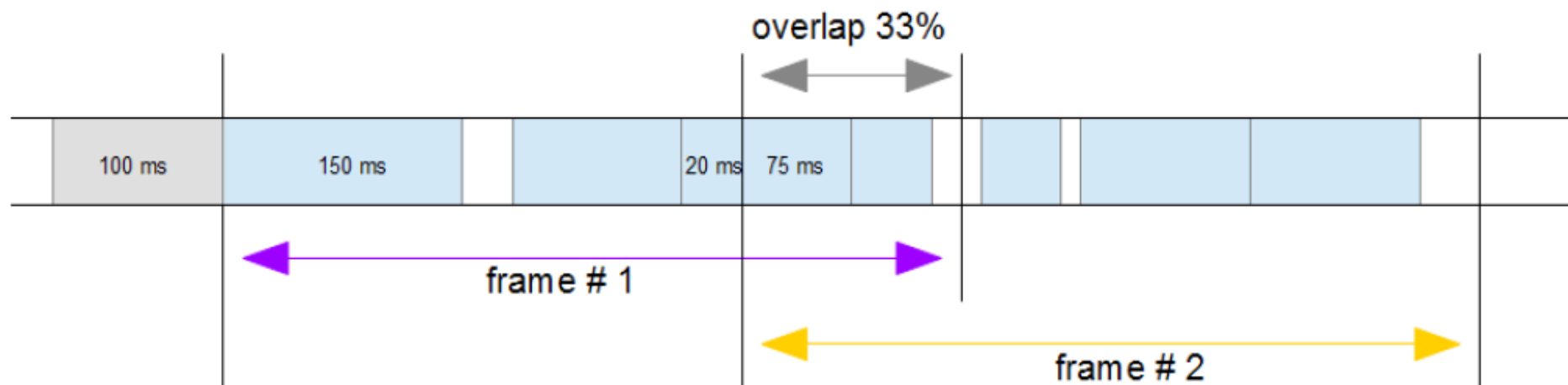


Figure 1. Segment rate moving average scheme

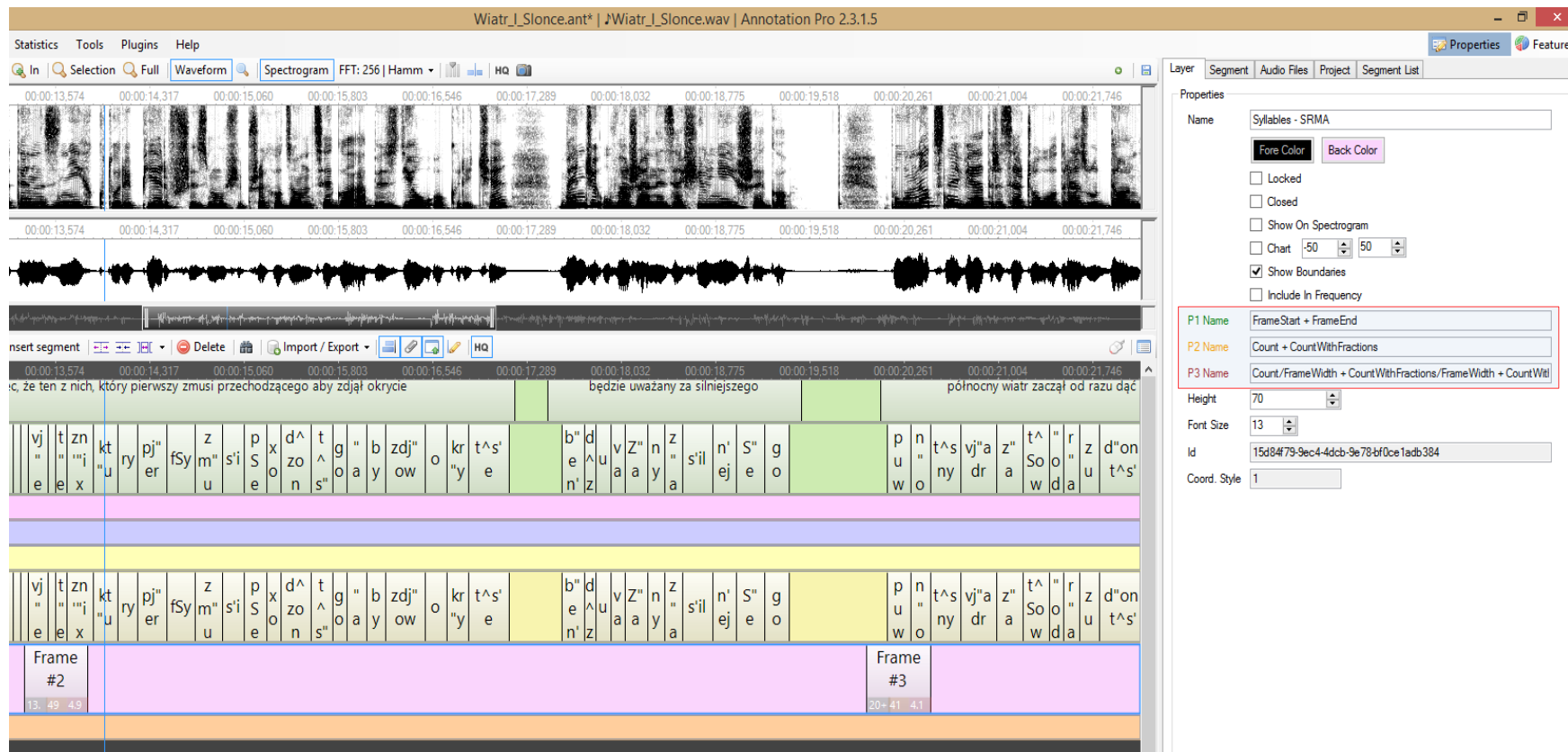


Annotation Pro – SRMA plugin - modyfikacja

```
6
7 namespace AnnotationPro.Plugin
8 {
9     public class AnnotationPlugin : IAnnotationPlugin
10     {
11         public void Run( AnnotationEditor editor )
12         {
13             //BEGIN USER CODE
14
15             string layerName = "Syllables";
16             float frameWidth = 10;
17             float frameStep = frameWidth * 2 / 3; //33%
18
19             Annotation annotation = editor.Synchronizer.DataServer.Annotation;
20             Audio audio = editor.Synchronizer.DataServer.Audio;
21
22
```



Annotation Pro – SRMA plugin - wynik





Annotation Pro – SRMA plugin - wynik

- (1) COUNT - **the number of segments within a moving frame** (the segments that are only partly included in the frame range are treated as whole),
- (2) COUNT with fractions - the number of segments within a moving frame (the segments that are only partly included in the frame range are treated as fractions),
- (3) **the ratio of the segments COUNT to the duration of the frame,**
- (4) the ratio of the segments COUNT with fractions to the duration of the frame,
- (5) the ratio of the segments COUNT with fractions to the total duration of all segments within the frame.

P1 Name	FrameStart + FrameEnd
P2 Name	Count + CountWithFractions
P3 Name	Count/FrameWidth + CountWithFractions/FrameWidth + CountWithFractions/FrameWidth + CountWithFractions/FrameWidth



Annotation Pro – SRMA plugin – eksport i analiza

P1 Name

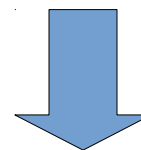
FrameStart + FrameEnd

P2 Name

Count + CountWithFractions

P3 Name

Count/FrameWidth + CountWithFractions/FrameWidth + CountWithFractions/FrameWidth

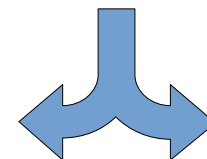


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	FileName	Start	End	Duration	Start (s)	End (s)	Duration	Label	Feature	Language	Name	Parameter1	Parameter2	Parameter3
2	Wiatr	0	8000	8000	0	500	500	Frame #0	Count: 37, Count (FRACT):			0+10	37+36.18357	3.7+3.618357+4.467002
3	Wiatr	1E+05	1E+05	8000	6667	7167	500	Frame #1	Count: 49, Count (FRACT):			6.666667+16.666667	49+48.03701	4.9+4.8037+5.522025
4	Wiatr	2E+05	2E+05	8000	13333	13833	500	Frame #2	Count: 49, Count (FRACT):			13.333333+23.333333	49+48.51208	4.9+4.851209+5.566542
5	Wiatr	3E+05	3E+05	8000	19970	20470	500	Frame #3	Count: 41, Count (FRACT):			20+30	41+40.35496	4.1+4.035496+4.623229
6	Wiatr	4E+05	4E+05	8000	26667	27167	500	Frame #4	Count: 32, Count (FRACT):			26.666667+36.666667	32+31.93337	3.2+3.193337+3.955735
7	Wiatr	5E+05	5E+05	8000	33333	33833	500	Frame #5	Count: 27, Count (FRACT):			33.333333+43.333333	27+26.54634	2.7+2.654634+3.898449
8	Wiatr	6E+05	6E+05	8000	40000	40500	500	Frame #6	Count: 6, Count (FRACT):			40+50	6+5.088397	0.6+0.5088397+5.119796



Annotation Pro – SRMA plugin – eksport i analiza

P1 Name	FrameStart + FrameEnd
P2 Name	Count + CountWithFractions
P3 Name	Count/FrameWidth + CountWithFractions/FrameWidth + CountWithFractions/TotalDuration

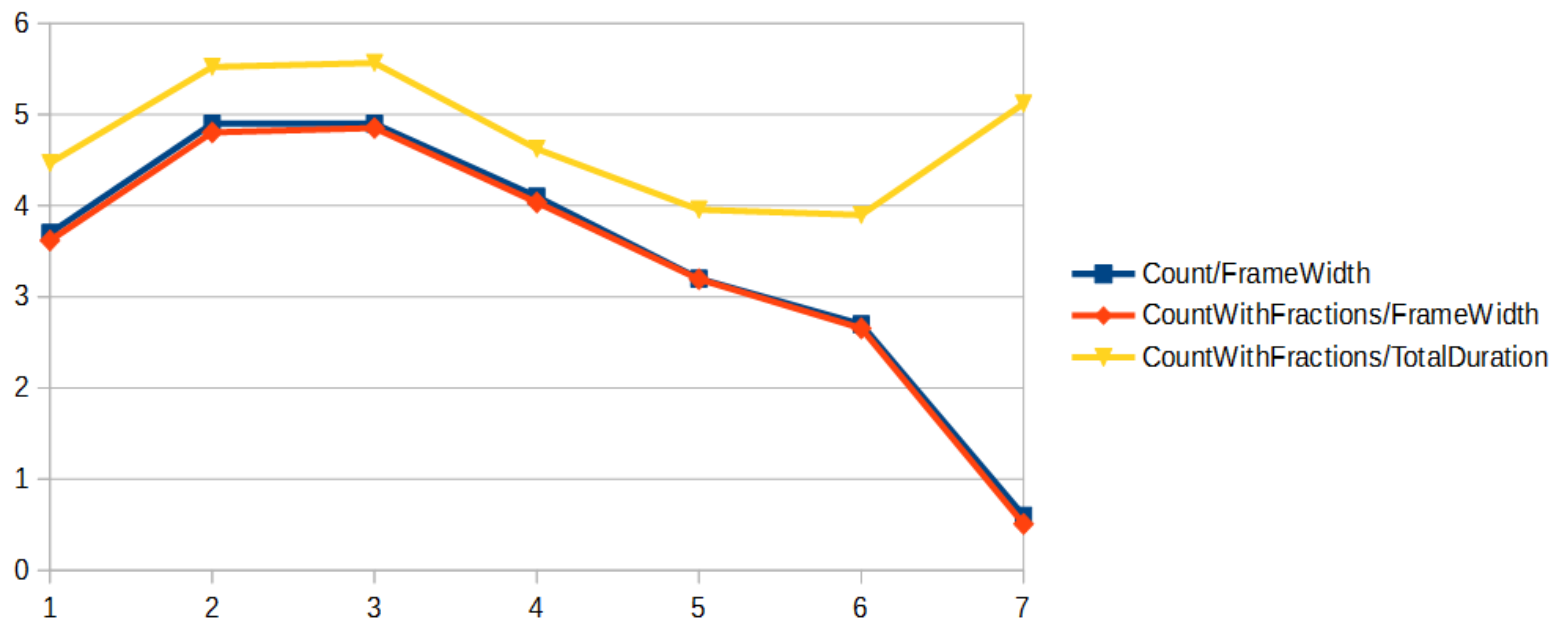


Label	Feature	Language	Name	FrameStart + FrameEnd	Count + CountWithFractions	Count/FrameWidth + CountWithFractions/FrameWidth + CountWithFractions/TotalDuration	Count/FrameWidth	CountWithFractions/FrameWidth	CountWithFractions/TotalDuration
0	Frame #0	Count: 37, Count (FRACT)	0+10		37+36.18357	3.7+3.618357+4.467002	3.7	3.618357	4.467002
0	Frame #1	Count: 49, Count (FRACT)	6.666667+16.666		49+48.03701	4.9+4.8037+5.522025	4.9	4.8037	5.522025
0	Frame #2	Count: 49, Count (FRACT)	13.333333+23.333		49+48.51208	4.9+4.851209+5.566542	4.9	4.851209	5.566542
0	Frame #3	Count: 41, Count (FRACT)	20+30		41+40.35496	4.1+4.035496+4.623229	4.1	4.035496	4.623229
0	Frame #4	Count: 32, Count (FRACT)	26.666667+36.666		32+31.93337	3.2+3.193337+3.955735	3.2	3.193337	3.955735
0	Frame #5	Count: 27, Count (FRACT)	33.333333+43.333		27+26.54634	2.7+2.654634+3.898449	2.7	2.654634	3.898449
0	Frame #6	Count: 6, Count (FRACT)	40+50		6+5.088397	0.6+0.5088397+5.119796	0.6	0.5088397	5.119796



Annotation Pro – SRMA plugin – eksport i analiza

P1 Name	FrameStart + FrameEnd
P2 Name	Count + CountWithFractions
P3 Name	Count/FrameWidth + CountWithFractions/FrameWidth + CountWithFractions/TotalDuration





TGA – Time Group Analysis **- m.in. zwolnienia / przyspieszenia w grupach czasowych**

TGA: Time Group Analyser - Speech Annotation Data Mining

Dafydd Gibbon (Universität Bielefeld)

Analiza regresji:

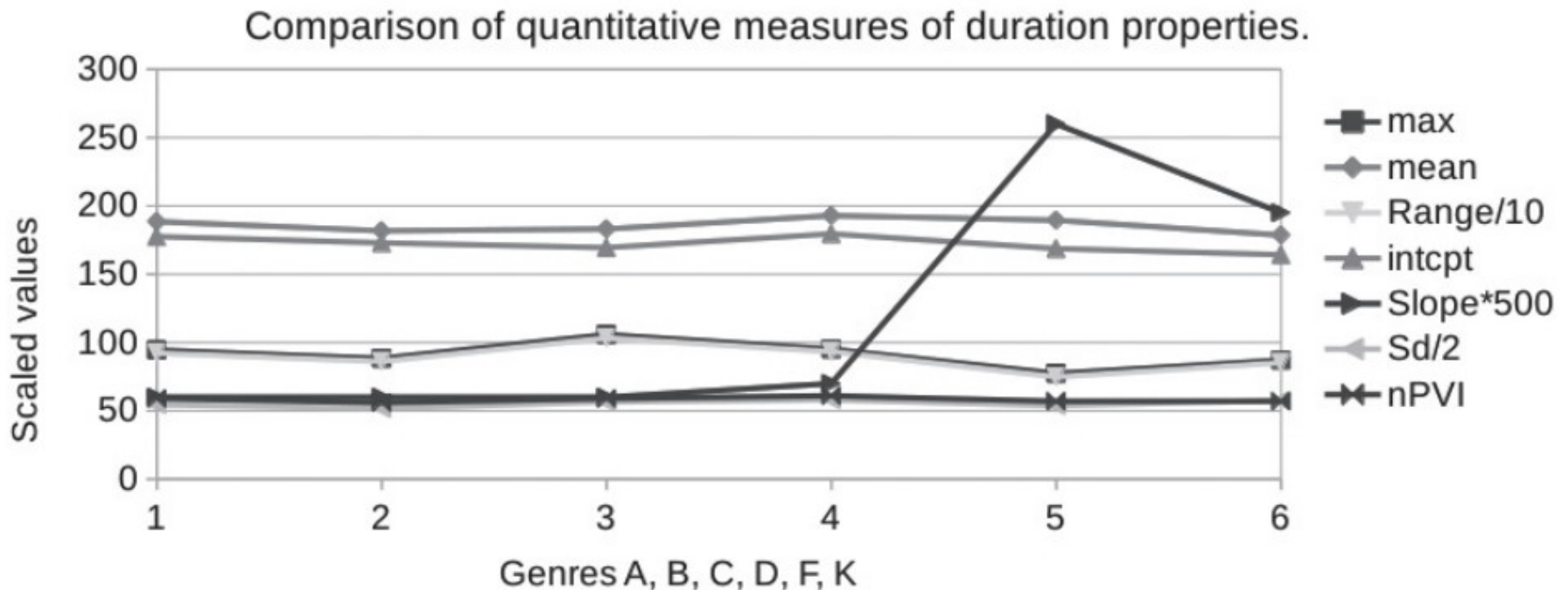
- parametry: timestamp, duration, duration difference
- wyniki: slope, intercept

nPVI:

- pairwise duration variability index (Grabe & Low, 2002)

TGA – Time Group Analysis

- m.in. zwolnienia / przyspieszenia w grupach czasowych



TGA – Time Group Analysis - m.in. zwolnienia / przyspieszenia w grupach czasowych

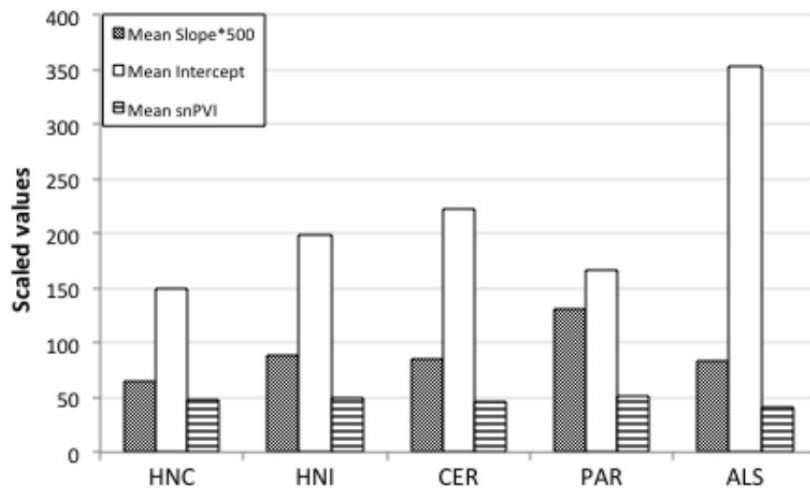


Figure 4: TGA results. A comparison of quantitative measures for healthy and dysarthric speakers (mean duration difference slopes, mean intercepts, mean syllable nPVI).

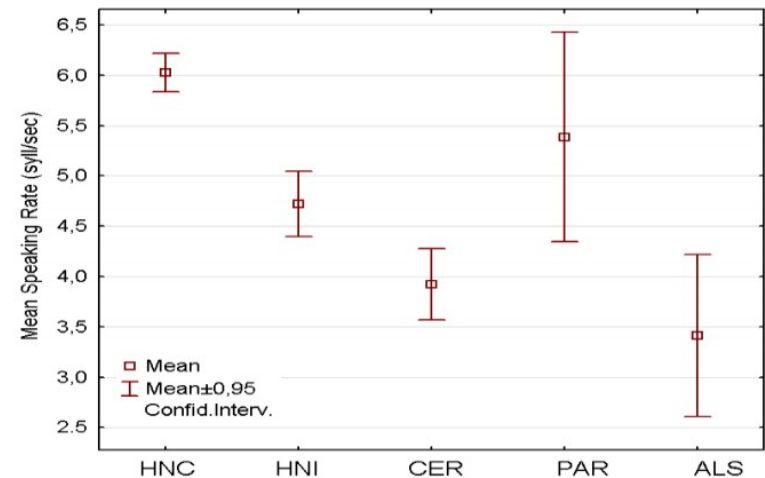


Figure 5: Mean speaking rate (number of syll. per sec.) for each population.



Plugin Annotation Pro + TGA - modyfikacja pluginu

```
21 using System.Collections.Generic;
22 using AnnotationPro.Statistics;
23 using System.Text;
24
25
26 namespace AnnotationPro.Plugin
27 {
28     public class AnnotationPlugin : IAnnotatio
29     {
30
31         public void Run(AnnotationEditor editor)
32         {
33             /// <summary>
34             /// runs TGA for the layer as defined by layerName
35             /// </summary>
36
37             string layerName = "syllables";
38
```

```
73
74     /// define items to ignore in the analysis
75     List<string> pauses = new List<string>();
76     pauses.Add("#");
77     pauses.Add("#$p");
78     pauses.Add("#00v");
79     pauses.Add("#$noise_spk");
80     pauses.Add("");
```

Plugin Annotation Pro + TGA - parametry i wyniki

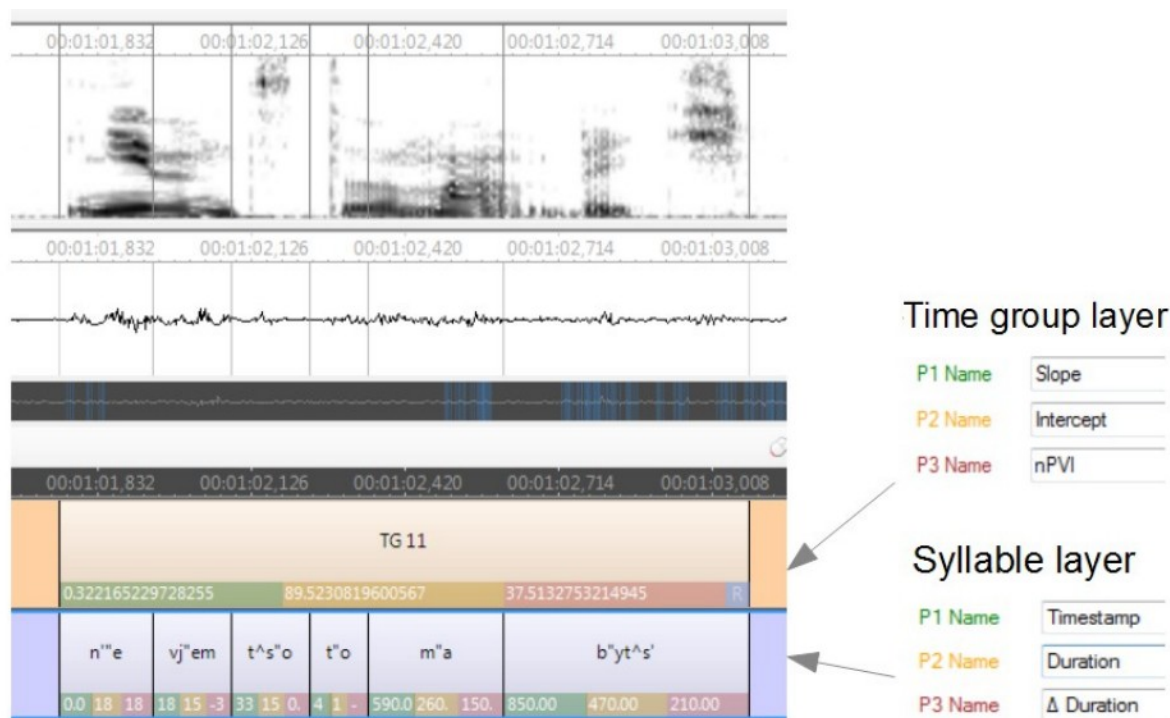


Figure 4: Example *TGA* results as displayed in the *Annotation Pro* interface for a Polish utterance: *Nie wiem co to ma być*. The obtained values are visible as segment parameters (P1, P2, P3) with coloured background at the bottom of each segment (a time group segment or a syllable segment respectively).



Plugin Annotation Pro + TGA - wykresy

The screenshot displays the Annotation Pro software interface. The main window shows a phonetic transcription timeline with segments labeled TG 8, TG 9, TG 10, TG 11, and TG 12. The transcription is color-coded by segment. A red box highlights the 'Start' button in the TGA plugin settings.

The TGA plugin settings are visible on the right side of the interface. The 'Properties' tab is selected, showing the following information:

- Label: TG 11
- Speaker: (empty)
- Fore Color: (empty)
- Back Color: (empty)
- Marker: (empty)
- Name: (empty)
- Par. 1: 0.0185601347812714
- Par. 2: 233.635907681479
- Par. 3: 40.0322259689264
- R Script: (empty)

The 'Start' button is highlighted with a red box. Below the 'Start' button, the 'Duration' is set to 2820.

The R Script section shows the following code:

```
# TGA Linear Regression
x <- c(0,240,480,750,1010,1170,1420,1830,1960,2200,2400)
y <- c(240,250,260,260,160,250,410,130,240,200,420)
l <- c("ft","en","sp","o","sup","puw","n","o","t","any","vj","adr",
      "par(mfrow=c(2,1))
plot(x,y,ylim=range(0,y),type="s", pch=9, lwd=1,col="black",
      title(main="Time Group Analysis, Slope: 0.019, Intercept: 233
```

Plugin Annotation Pro + TGA - wykresy

Time Group Analysis, Slope: 0.060, Intercept: 154.274

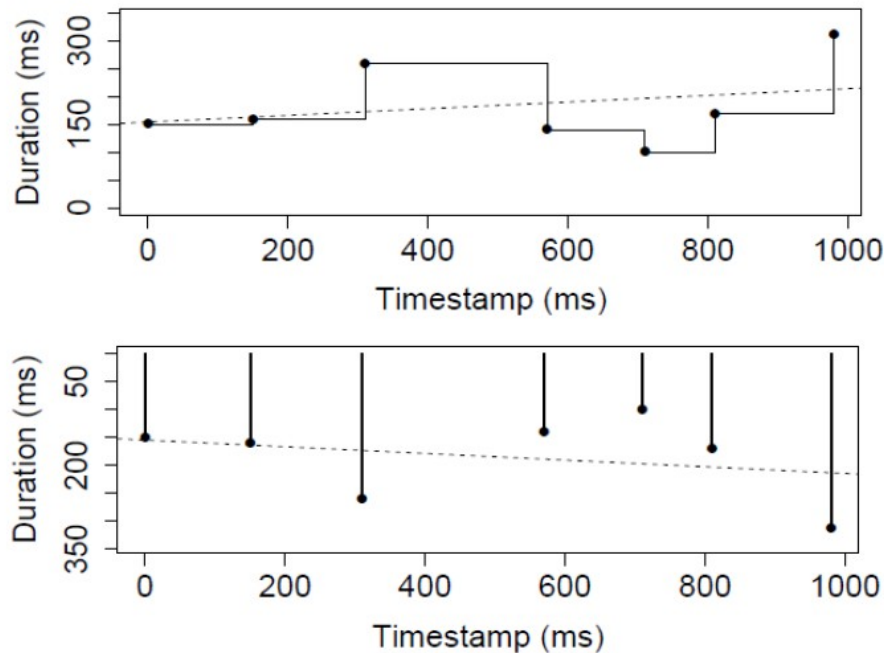


Figure 5: TGA results for a single Time Group with increasing (deceleration) duration tendencies shown as a duration differences chart (top) and a top-suspended bar chart (bottom).



Plugin Annotation Pro + TGA - eksport i dalsza analiza

Wiatr_I_Slonce.ant | Wiatr_I_Slonce.wav | Annot

File Edit View Statistics Tools Plugins Help

Layers + Audio Play Out In Selection Full Waveform Spectrogram FFT: 256 | Hamm HQ

ORTO +
AutoSeg +
ManualSeg + -
ORTO_ENG -
SAMPA + -
Syllables
Syllables - SRMA
COMMENTS -
syllables +
TGA Parameters +
TGA Result +

Annotation Add New Layer Insert segment Delete Import / Export

- Import CSV File...
- Import Layers From CSV (Label, Start, Stop)...
- Import Layers From Text Files...
- Import Layers From BLF Files...
- Import Layers From TRS Files...
- Import Layers From TextGrid Files...
- Import Layers From XRA Files...
- Import Layers From EAF Files...
- Import Praat F0 File...
- Import Praat Formant File...
- Export Layer To CSV File...
- Export Layer To Text File...
- Export Layer To Text & Audio Files...
- Export To CSV...
- Export To TextGrid...
- Export To XRA...
- Export To EAF...
- Export To ANT & Audio Files...

ORTO
AutoSeg s'il n"ej Sy vw" a
ManualSeg
SAMPA
Syllables s'il n"ej Sy vw" a
TGA Para s'il n"ej Sy vw" a
TGA Result 883218 0.04123867238

Import/export – opcje
wbudowane vs.
pluginy do importu



Plugin Annotation Pro + TGA - eksport i dalsza analiza

Layer Segment Audio Files Project Segment List

Properties

Name TGA Result

Fore Color Back Color

☐ Locked

☐ Closed

☐ Show On Spectrogram

☐ Chart -50 50

☒ Show Boundaries

☐ Include In Frequency

P1 Name Slope

P2 Name Intercept

P3 Name nPVI

Height 70

Font Size 13

Id edf83cc2-1eb0-4ce5-a4c5-4ad5e5f921c2

Coord. Style 1



Plugin Annotation Pro + TGA

- eksport i dalsza analiza

P1 Name

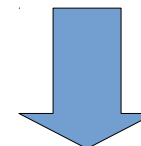
Slope

P2 Name

Intercept

P3 Name

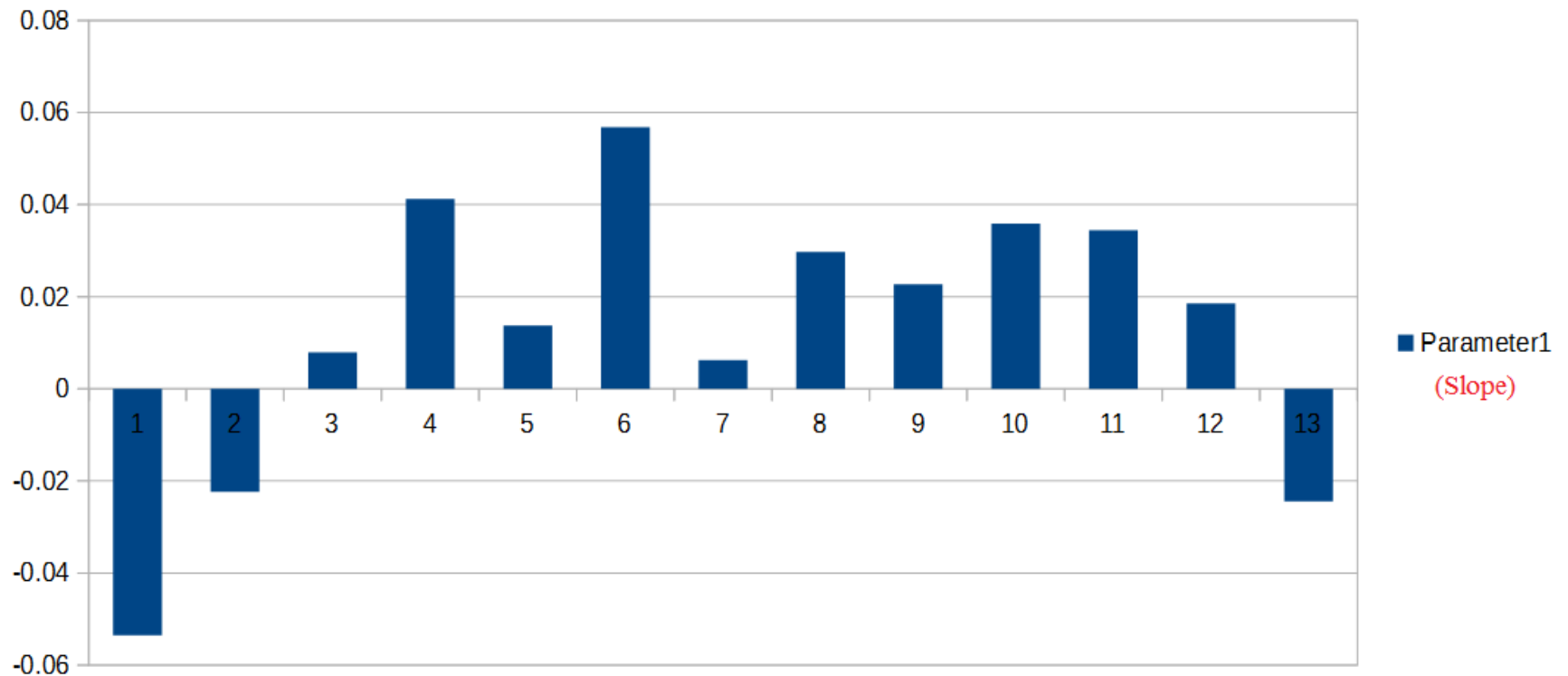
nPVI



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	FileName	Start	End	Duration	Start (ms)	End (ms)	Duration (ms)	Label	Featu	Langu	Name	Parameter1	Parameter2	Parameter3
2	Wiatr_1_Slor	15185.37	32305.37	17120	949.0853	2019.085	1070	TG 0				-0.05353639	237.877231	24.0427928
3	Wiatr_1_Slor	34041.14	62681.14	28640	2127.572	3917.572	1790	TG 1				-0.02239413	272.637847	79.6960403
4	Wiatr_1_Slor	64492.14	105452.1	40960	4030.759	6590.759	2560	TG 2				0.0079311	247.109233	32.2939872
5	Wiatr_1_Slor	117116.9	178556.9	61440	7319.807	11159.81	3840	TG 3				0.04123867	147.626382	45.0979693
6	Wiatr_1_Slor	188920.4	274406.4	85486	11807.52	17150.4	5342.875	TG 4				0.01376429	134.151231	41.8477128
7	Wiatr_1_Slor	281044.3	309684.3	28640	17565.27	19355.27	1790	TG 5				0.05681779	121.663503	29.7030767
8	Wiatr_1_Slor	322075	368795	46720	20129.69	23049.69	2920	TG 6				0.00623655	200.321362	32.8343364
9	Wiatr_1_Slor	370327.5	426007.5	55680	23145.47	26625.47	3480	TG 7				0.02975334	160.793451	37.8146488
10	Wiatr_1_Slor	439204.4	475044.4	35840	27450.27	29690.27	2240	TG 8				0.02269659	227.377565	45.136385
11	Wiatr_1_Slor	478580.2	511700.2	33120	29911.26	31981.26	2070	TG 9				0.03586562	175.868644	43.1880566
12	Wiatr_1_Slor	514604.7	563084.7	48480	32162.79	35192.79	3030	TG 10				0.03441222	259.778252	47.8172203
13	Wiatr_1_Slor	574943.8	620063.8	45120	35933.99	38753.99	2820	TG 11				0.01856013	233.635908	40.032226
14	Wiatr_1_Slor	621821.9	655901.9	34080	38863.87	40993.87	2130	TG 12				-0.02443149	237.358199	39.9365567



Plugin Annotation Pro + TGA - eksport i dalsza analiza





Literatura

Jassem, W. 2005. Illustrations of the IPA: Polish. *Journal of the International Phonetic Association* 33(1): 103-107.

Karpiński, M., Klessa, K. Prozodia niepewności. [w druku w:] **Sens i brzmienie**, preprint version (in Polish) available for download at Academia.edu.

Karpiński, M., Klessa, K., Czoska, A. (2014). Local and global convergence in the temporal domain in Polish task-oriented dialogue, *Proceedings of the 7th Speech Prosody Conference*, 20-23 May 2014, Dublin, Ireland. ISSN: 2333-2042.

Madelska, L., Witaszek-Samborska, M. Zapis fonetyczny. Zbiór ćwiczeń, str. 17-21 (alfabety, ich opisy, ćw. 1), oraz str. 26-31

Jassem, W. Podstawy fonetyki akustycznej, str. 139-145 (o transkrypcji fonetycznej)

Karpiński, M. Dźwięki mowy, *Book of Knowledge of Languages in Danger* (PL). Ćwiczenia:

Dukiewicz, L., Sawicka, I. Fonetyka i fonologia, str. 49-95 (od rozdziału: Istota i struktura fal głosowych do rozdz. Częstość występowania elementów mowy).

Ryalls, J.H. (1996) A basic introduction to speech perception. Rozdziały 3-6 oraz 11.

Moore, B.C.J. (1999) Wprowadzenie do psychologii słyszenia. Rozdział 8 (Percepcja mowy).



Dziękuję za uwagę!

