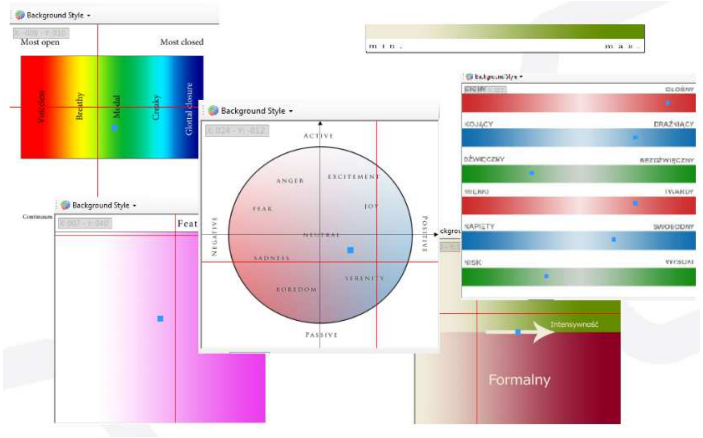
Annotation Pro Software
Speech signal visualisation, part 1

klessa@amu.edu.pl
katarzyna.klessa.pl

Katarzyna Klessa



Topics of the class

1. Introduction: annotation of speech recordings
2. *Annotation Pro*
    - Graphical representation of the feature space
    - Annotation: multiple layers (tiers) and operations on segments
    - Perception test interface
    - Import - Export options
3. Visualisations of the speech signal: waveform vs. spectrogram

The goals and general assumptions

- What is **annotation** of speech recordings?
- What can we annotate?

The goals and general assumptions

- What is **annotation** of speech recordings?
- What can we annotate?

orthography phonetic transcription

information about speaker(s)

environment dialect interlocutors

gesture

emotions voice quality

health condition **language**

The goals and general assumptions

- What is **annotation** of speech recordings?
- What can we annotate? - Categorisations, eg.:

  linguistic vs. non/para-linguistic features

  data vs. metadata

State of the art

- Why another annotation software?
- State of the art. A wide range of annotation software available

The goals and general assumptions

- Some reasons & assumptions for creating new software:
  - continuous features & rating scales
  - easy access to perception test options
  - easy to operate and start with
  - universal character (non task-specific)
  - extendable by users

Annotation Pro

- Please check whether the software is available at your PC (classroom)

Basic information

- Download: annotationpro.org/download
- Documentation forthcoming at: annotationpro.org
- Licence: freeware for research and education
- How to start?
- New versions of the software can be updated at launch

.....see how it works.

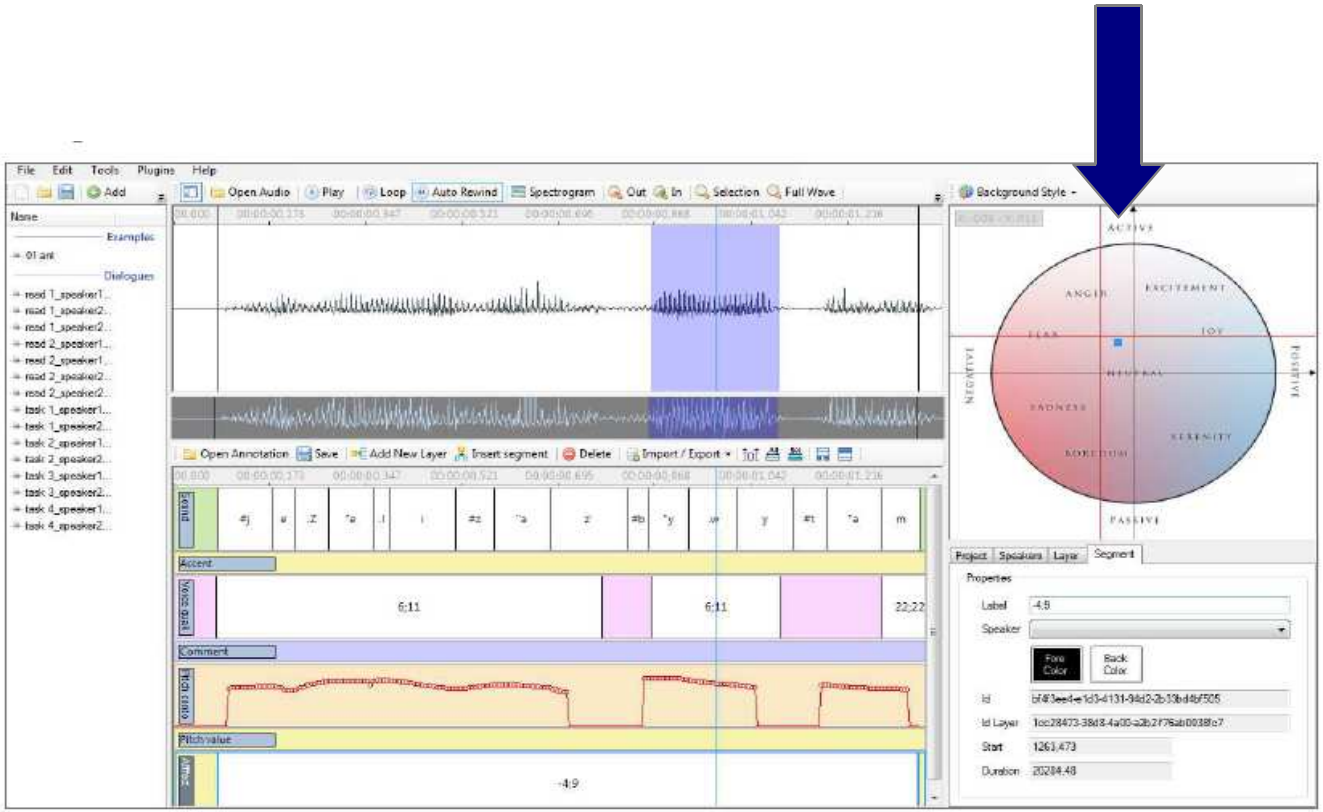
Basic information

- Download: annotationpro.org/download
- Documentation forthcoming at: annotationpro.org
- Licence: freeware for research and education
- How to start?
- New versions of the software can be updated at launch

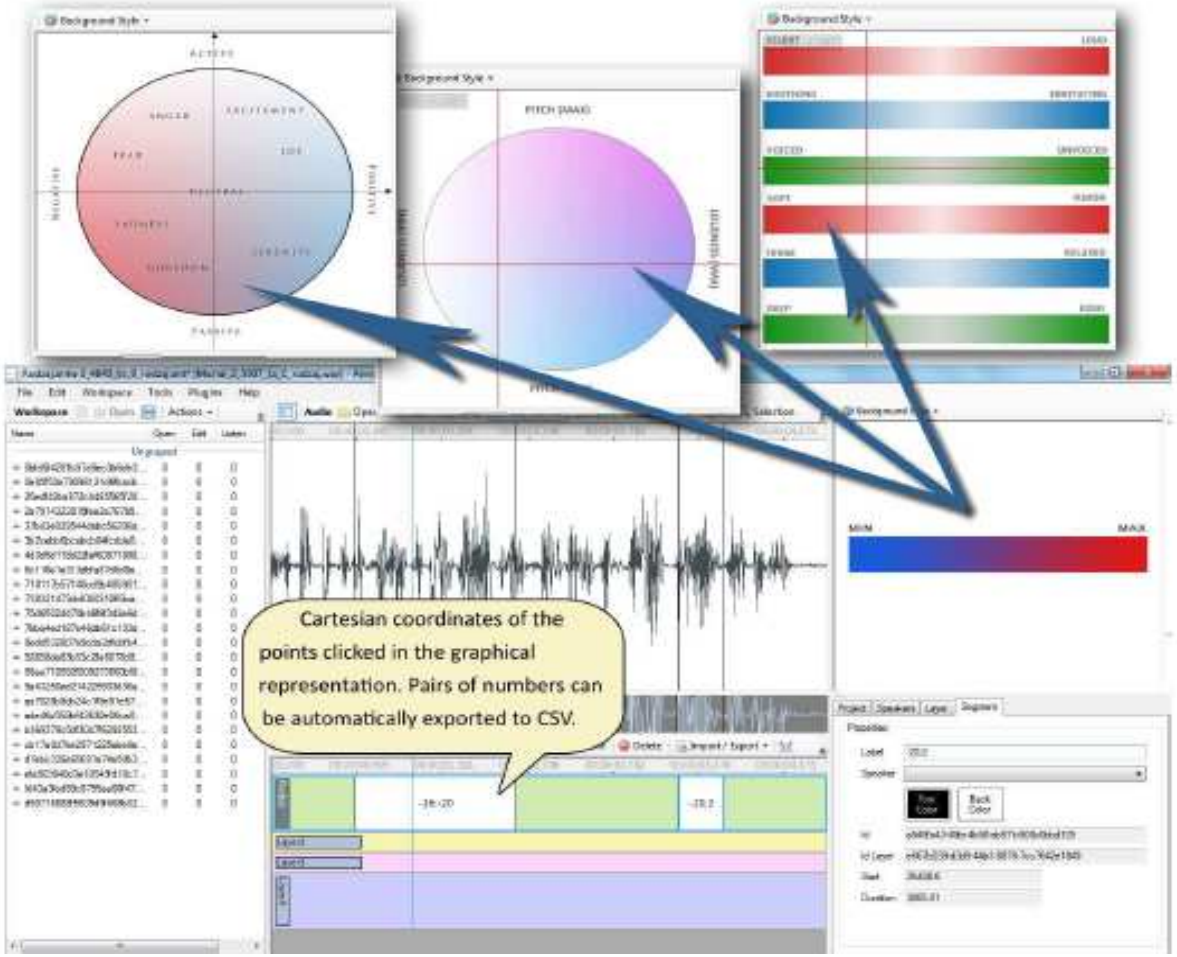
.....see how it works.

The user interface

Graphical respresentation of feature space

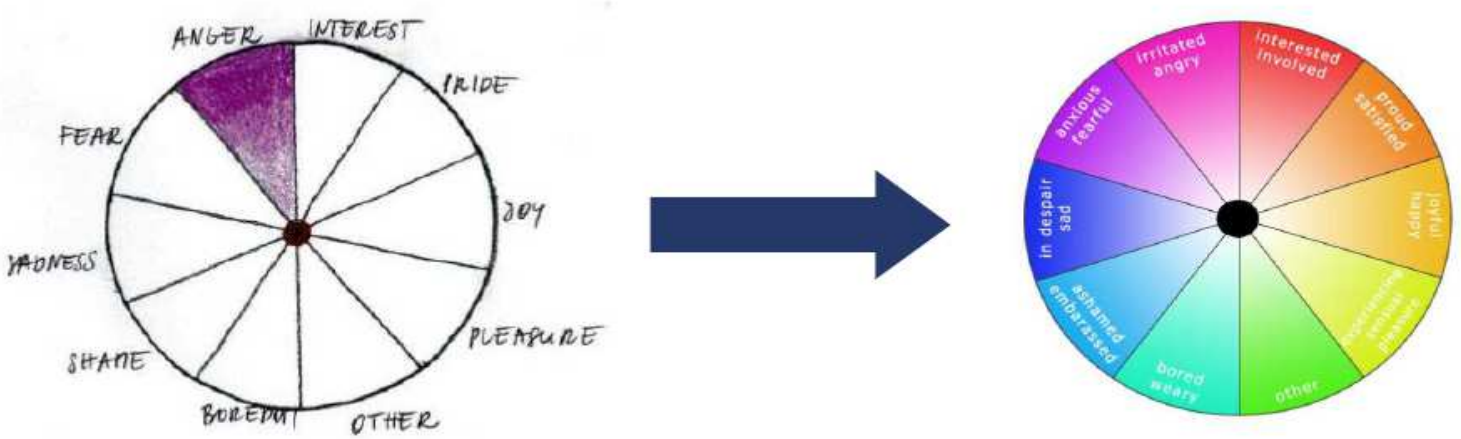


Graphical representation of the feature space



Graphical representation of the feature space

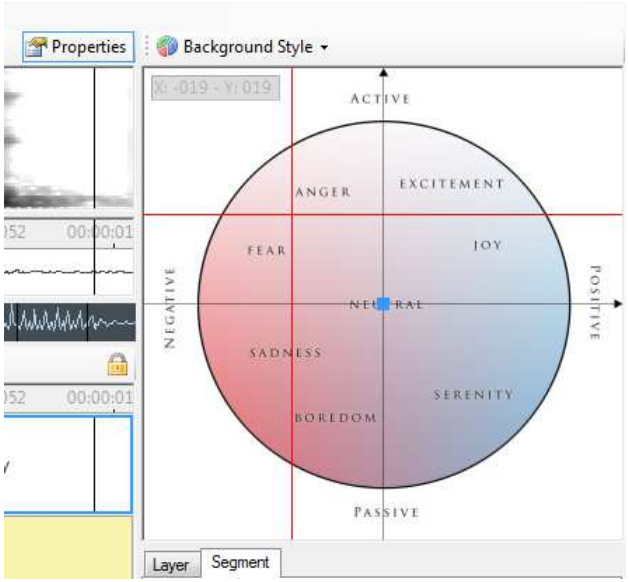
- Create your own feature space,
- or upload an existing picture from your disk.



.....see how it works.

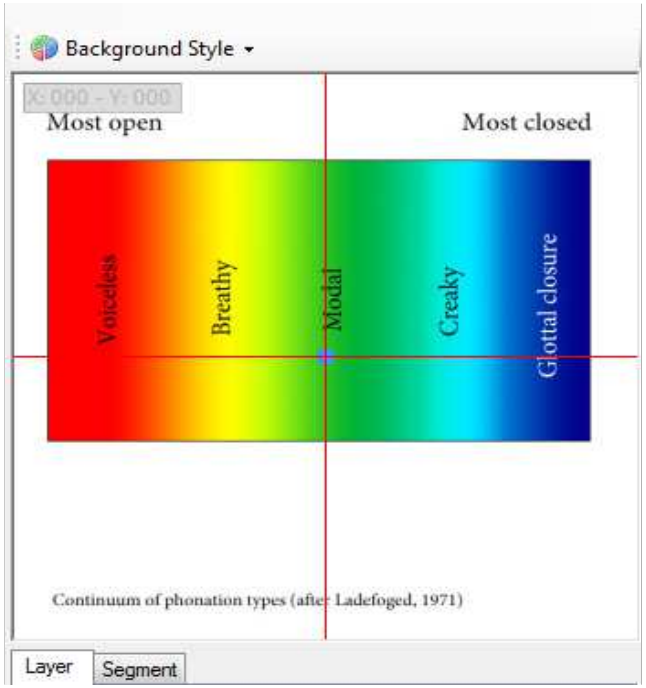
Graphical representation of the feature space - examples

- Relatively low number of emotion categories in most studies - it might be useful to apply several classifications or domains
- Vague categorisations
- Possibility to discover new categories, tendencies by observing clusters using continuous feature spaces



Graphical representation of the feature space - examples

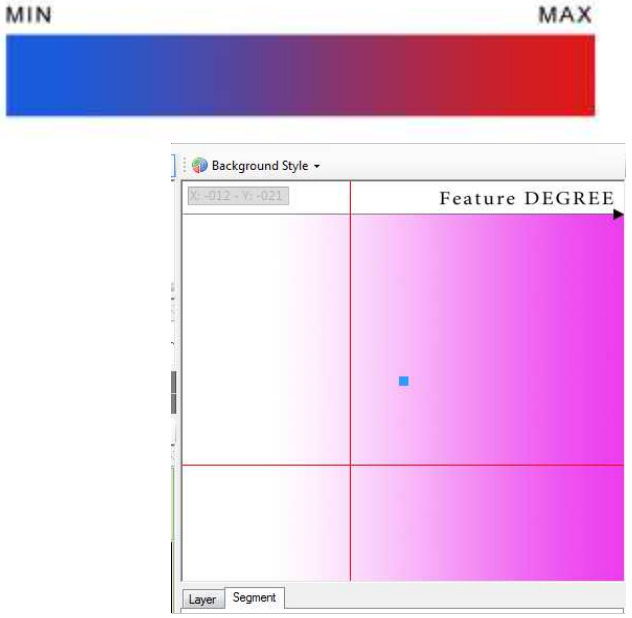
- Applying, verifying existing representations
- Phonation types continuum (e.g. after P. Ladefoged, 1971)
- Flexibility of interpretation, defining related continua, etc.



Graphical representation of the feature space - examples

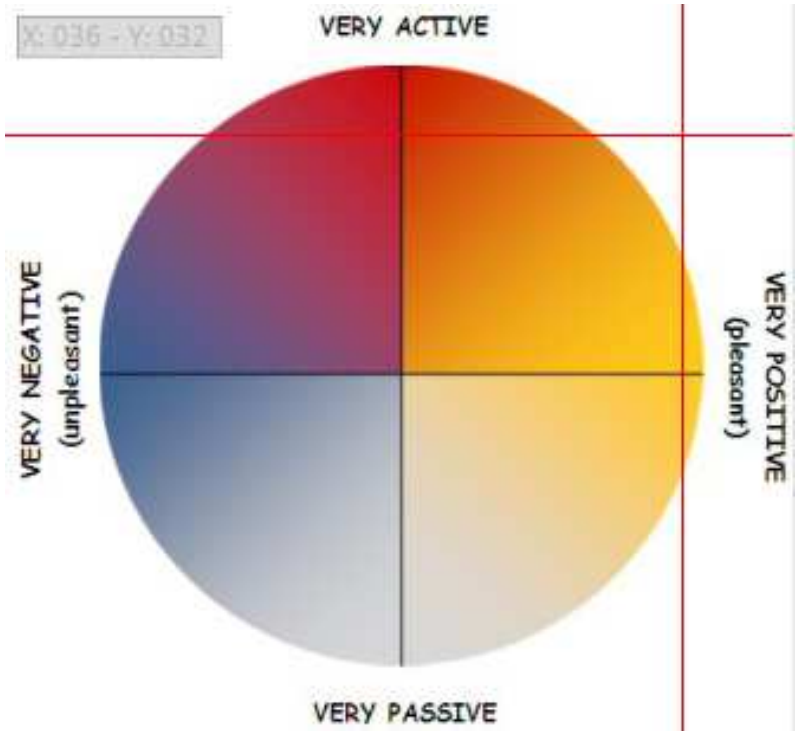
User-defined feature spaces

- speaker noises
- environment noises
- voice quality
- speaker specificity
- conversation characteristics

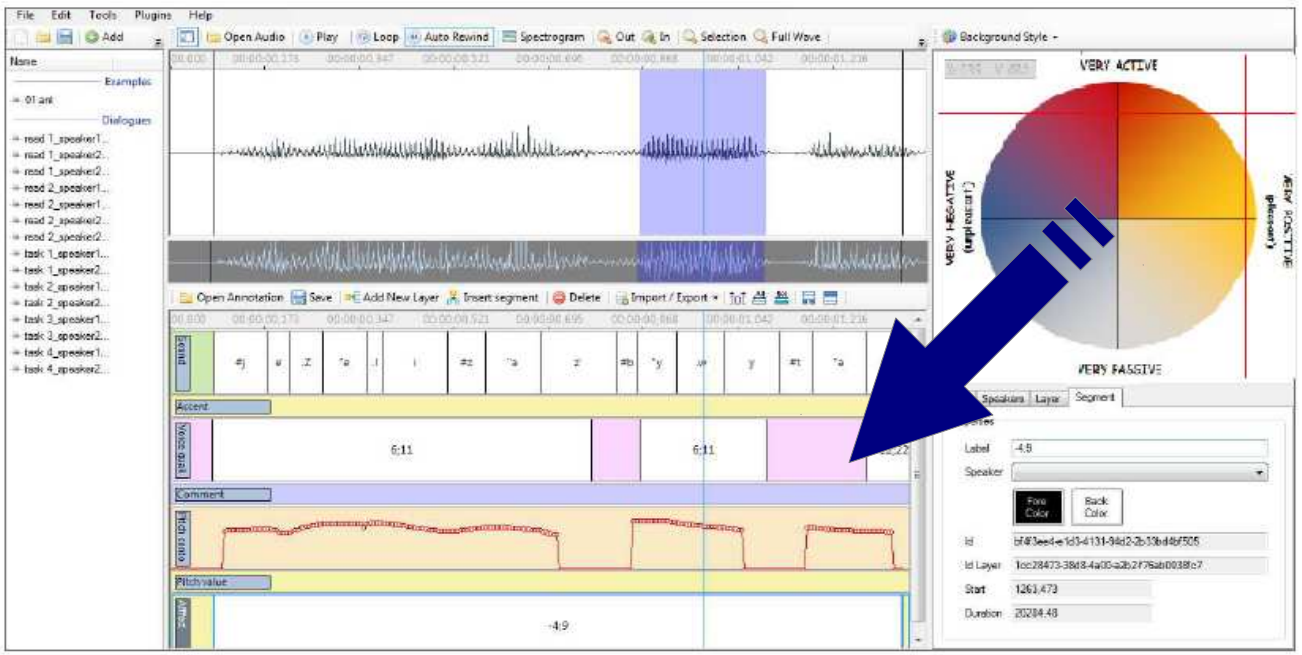


Graphical representation of the feature space - annotation of emotions

- Study material: emotionally marked speech from 3 speakers, monologues, high quality recordings
- Participants: students of III, IV grade of linguistics
- Task: perceptually assess the utterances using the dimensions: positive/negative, active/passive by clicking on continuous feature space.



Graphical representation of the feature space - annotation of emotions



- Cartesian coordinates as a result of clicking
- Numbers or graphs on layer

.....see how it works.

Graphical representation of the feature space - annotation results

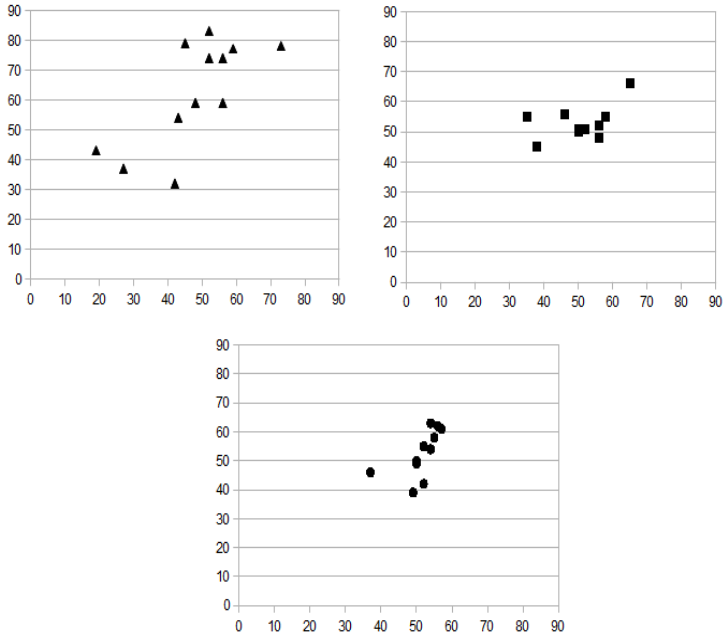
Export to CSV ->
to a spreadsheet

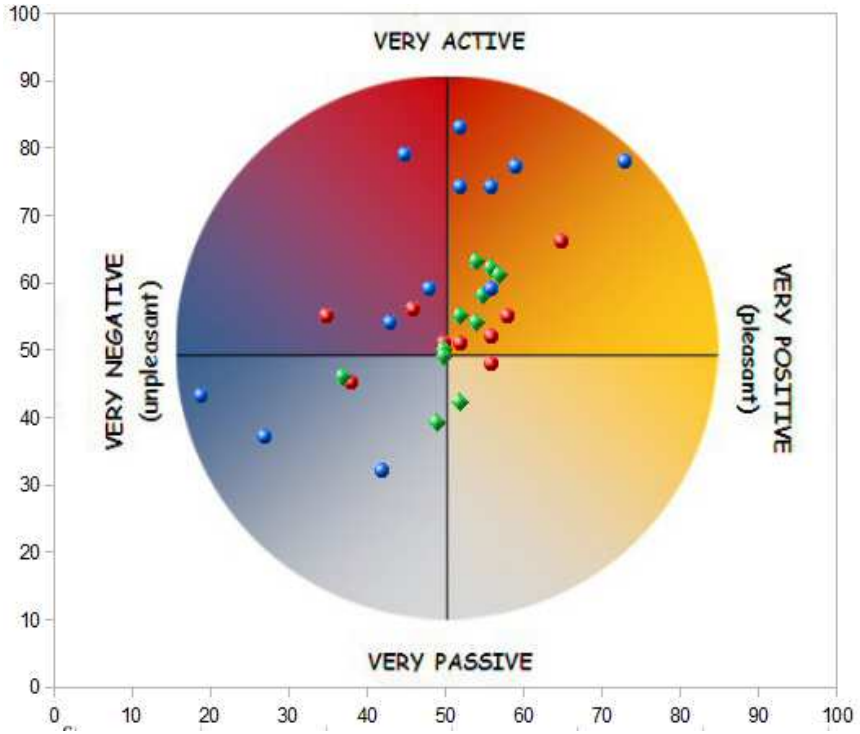
	A	B	C	M	N
1	Labeler	File	File name	Label 1 (0-100)	Label 2 (0-100)
2	1	007.ant	7	52	74
3	1	009.ant	9	43	54
4	1	010.ant	10	73	78
5	1	011.ant	11	45	79
6	1	012.ant	12	27	37
7	1	014.ant	14	56	59
8	1	015.ant	15	19	43
9	1	016.ant	16	59	77
10	1	017.ant	17	42	32
11	1	018.ant	18	48	59
12	1	019.ant	19	56	74
13	1	026.ant	26	52	83
14	2	007.ant	7	56	48
15	2	009.ant	9	46	56
16	2	010.ant	10	65	66
17	2	011.ant	11	50	50
18	2	012.ant	12	38	45
19	2	014.ant	14	58	55
20	2	015.ant	15	50	50
21	2	016.ant	16	56	52
22	2	017.ant	17	35	55
23	2	018.ant	18	52	51
24	2	019.ant	19	50	50
25	2	026.ant	26	50	51
26	3	007.ant	7	56	62
27	3	009.ant	9	49	39
28	3	010.ant	10	57	61
29	3	011.ant	11	50	50
30	3	012.ant	12	37	46
31	3	014.ant	14	52	55

Arkusz1 Arkusz2 Arkusz3

Graphical representation of the feature space - annotation results

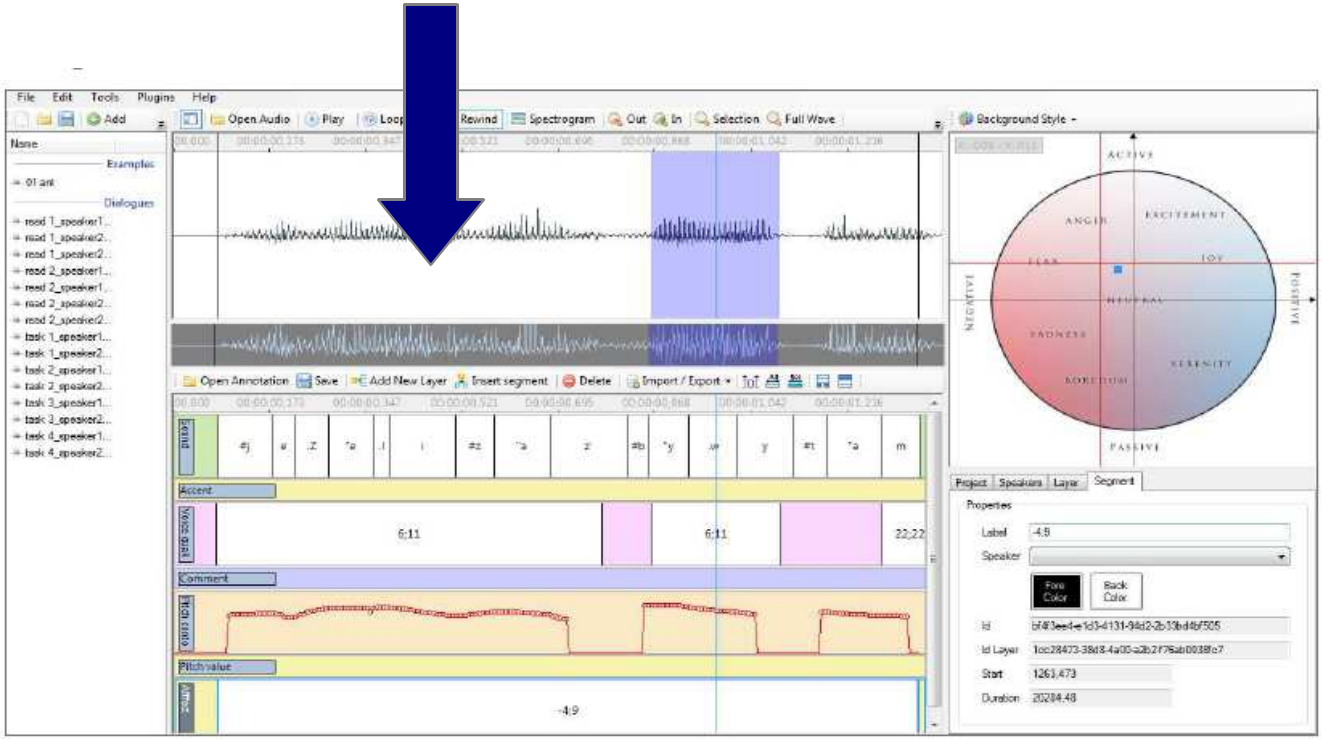
- Create graphs, calculate statistics.





The user interface

“Traditional”
annotation layers



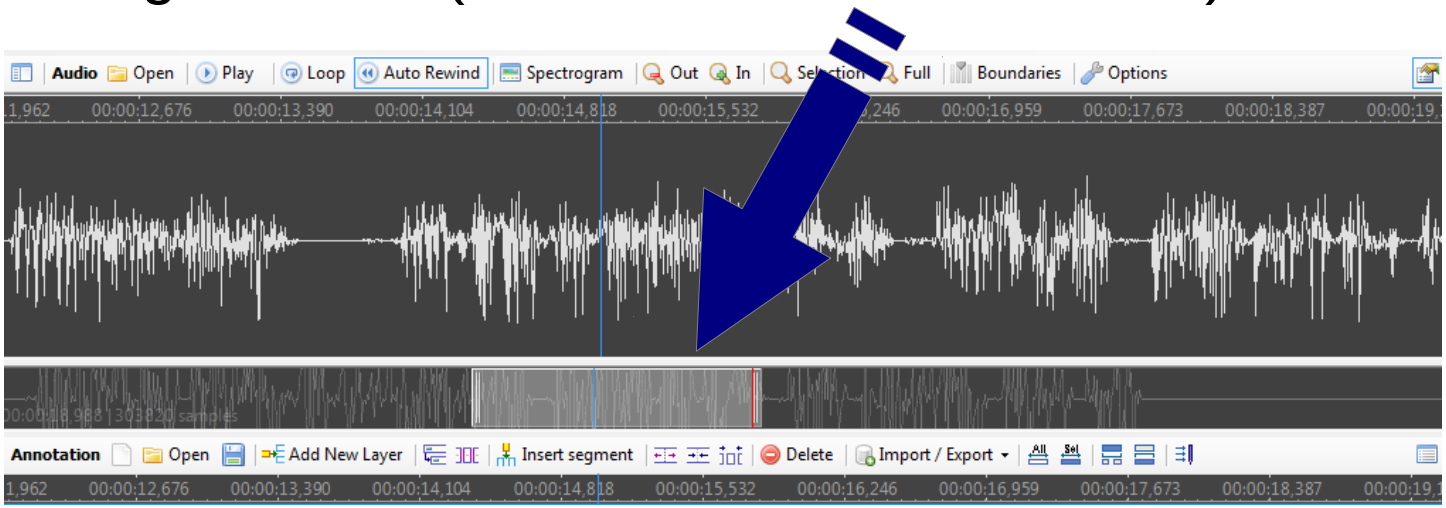
TASK 1

1. Open the “DzienDobry.wav” file
2. Create two segments on the annotation layer, each for one word
3. Transcribe the sound orthographically
4. Save annotation to disk
5. Create two new layers
6. Name the annotation layers: *Orhography, Phonetic, Emotions*, respectively
7. Choose Emotions layer and then select the “Valence-Activation” background as picture and mark your subjective judgment of emotional load of the utterance

- Remember to save the file often.

User interface
- layers and segments

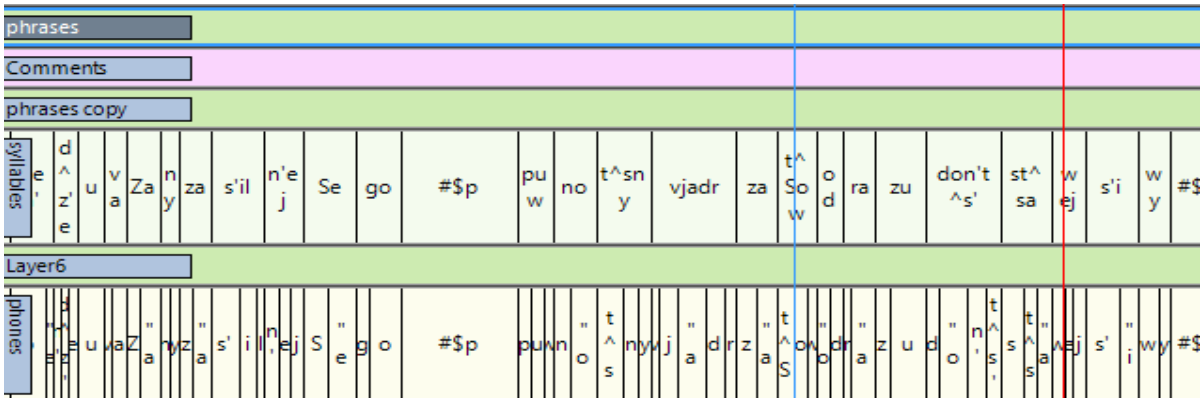
- Sound signal **visualisation** - waveform, spectrogram
- **Navigation** - zoom - mouse scroll or buttons, navigation bar (move, resize visible frame)



.....see how it works.

User interface
- layers and segments

- **layers** - any number of layers, options to duplicate, copy, hide, lock, export layers
- **Segments** - the basic units in a layer, options to resize, move, duplicate, many font families available

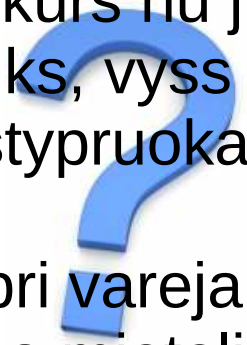


.....see how it works.

Take a guess: what is the story about? - what's the language?

Puorsoka - **Zimels i Saule**

Tys nutyka vacus laikus. Saule i Zimels guoja pa celu i idami runuoja sova storpa, kurs nu jus stypruoks.
Te pretim guoja celiniks, vyss sasatins sylta mieteli.
Ji nuspride, ka pats stypruokais ir tys, kurs liks celinikam numaukt mieteli.
Zimels pyute, cik stypri vareja, bet ku vaira jis pyute, tu celiniks vaira sasatyna mieteli, cikom jau Zimels mete miru.
Niu givuos Saule sildeit gaisu ar sovim syltajim spaitim i jau piec eisa laika celiniks nuvylka sovu mieteli.
Tai Zimelam daguoja atzeit, ka Saule par ju stypruoka.



The sound: http://www.youtube.com/watch?v=FLIMBZQeUfc&feature=youtu.be

Answer: Latgalian version of *North Wind and the Sun*

Puorsoka - **Zimels i Saule**

Tys nutyka vacus laikus. Saule i Zimels guoja pa celu i idami runuoja sova storpa, kurs nu jus stypruoks.
Te pretim guoja celiniks, vyss sasatins sylta mieteli.
Ji nuspride, ka pats stypruokais ir tys, kurs liks celinikam numaukt mieteli.
Zimels pyute, cik stypri vareja, bet ku vaira jis pyute, tu celiniks vaira sasatyna mieteli, cikom jau Zimels mete miru.
Niu givuos Saule sildeit gaisu ar sovim syltajim spaitim i jau piec eisa laika celiniks nuvylka sovu mieteli.
Tai Zimelam daguoja atzeit, ka Saule par ju stypruoka.

The sound: http://www.youtube.com/watch?v=FLIMBZQeUfc&feature=youtu.be

The North Wind and the Sun

The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak.
They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other.
Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

The sound, e.g.: http://www.ua.ac.be/main.aspx?c=.EDINBURGHIPA&n=35607

Wiatr Północny i Słońce

For the analysis of the Polish IPA, and text & transcript of *North Wind...* refer to:

Jassem., W. (2003) Illustrations of the IPA: Polish. *Journal of the International Phonetic Association, 33*(01), 103-107.

TASK 1

1. Open the “DzienDobry.wav” file
2. Create two segments on the annotation layer, each for one word
3. Transcribe the sound orthographically
4. Save annotation to disk
5. Create two new layers
6. Name the annotation layers: *Orhography, Phonetic, Emotions*, respectively
7. Write phonetic transcriptionof Dzień Dobry to the *Phonetic* layer
8. Choose Emotions layer and then select the “Valence-Activation” background as picture and mark your subjective judgment of emotional load of the utterance

- Remember to save the file often.

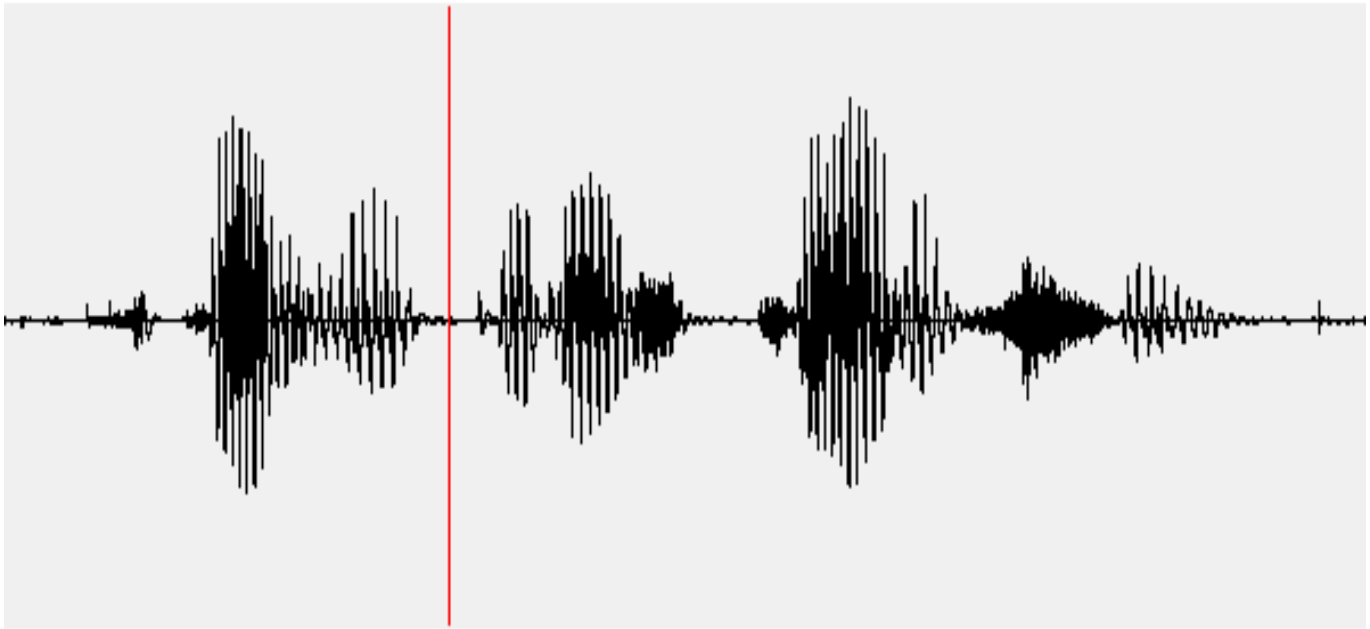
Annotation procedures - examples

Procedures followed so far:

1. Preliminary **listening** to the recording (preferably using headphones) and verifying the script
2. **Importing** the orthographic transcription to Annotation Pro or typing it directly into the layer
3. **Adjusting** the boundaries of segments
4. **Duplicating** layer and transforming ortography to phonetic transcription on the syllable & phone level

.....see how it works.

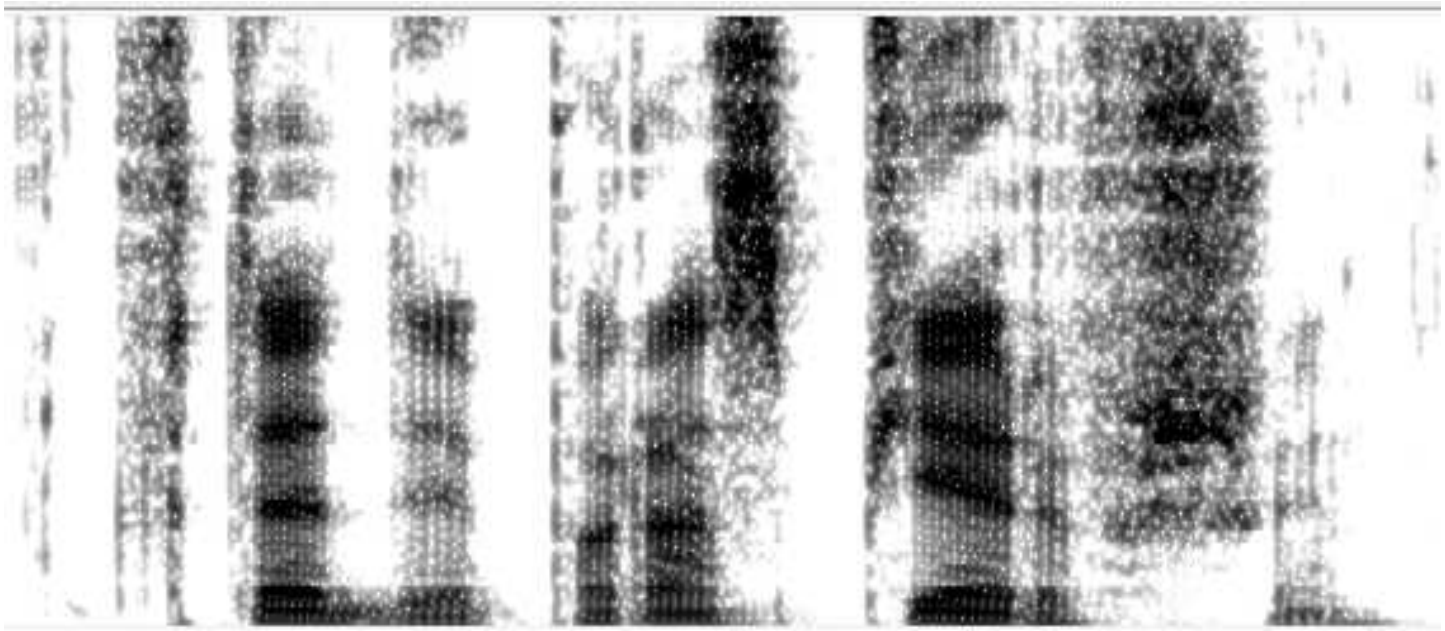
Speech sound visualisation: waveform



Waveform: mainly intensity & time

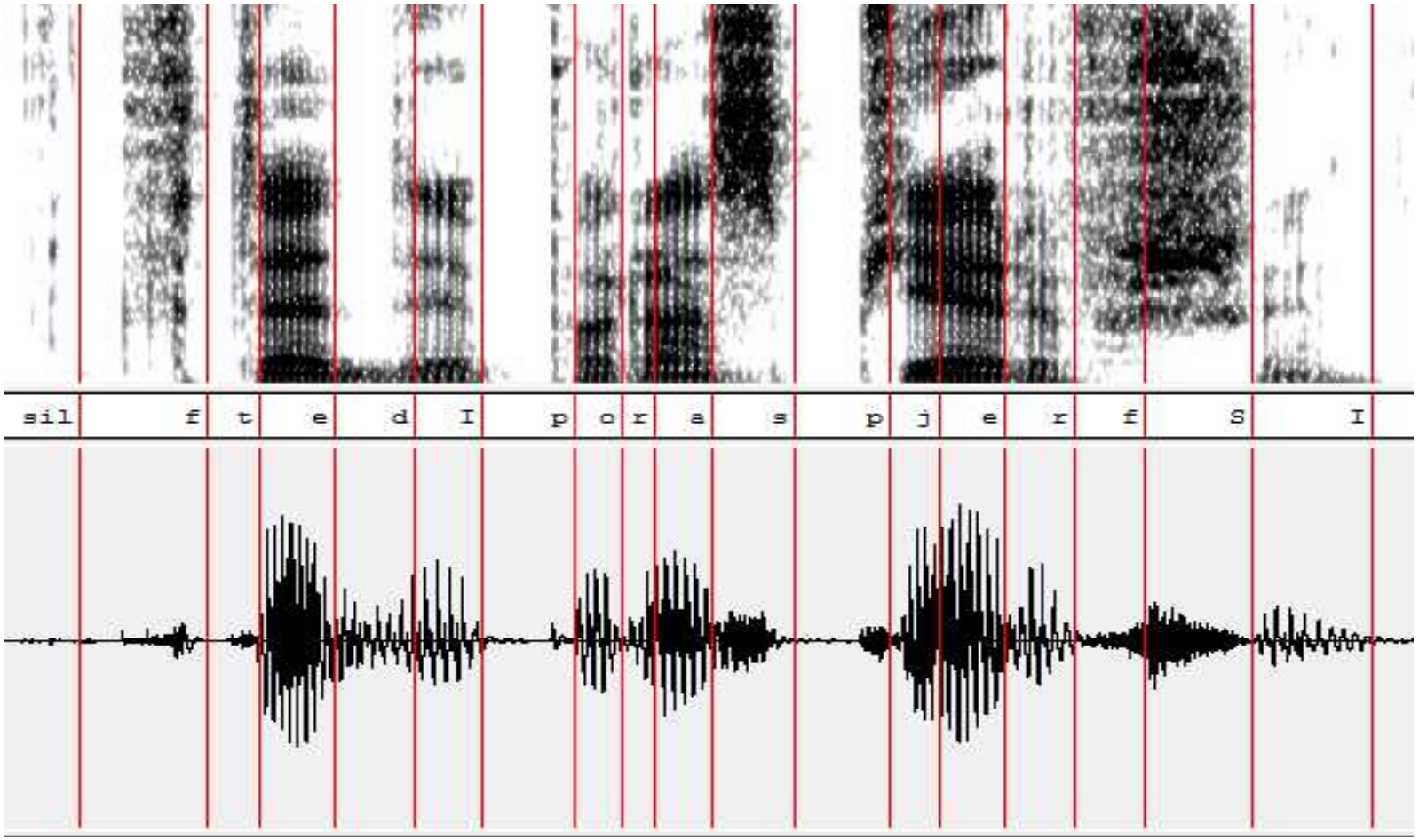
Wtedy po raz pierwszy

Spectrogram: three dimensions - time, intensity, frequency

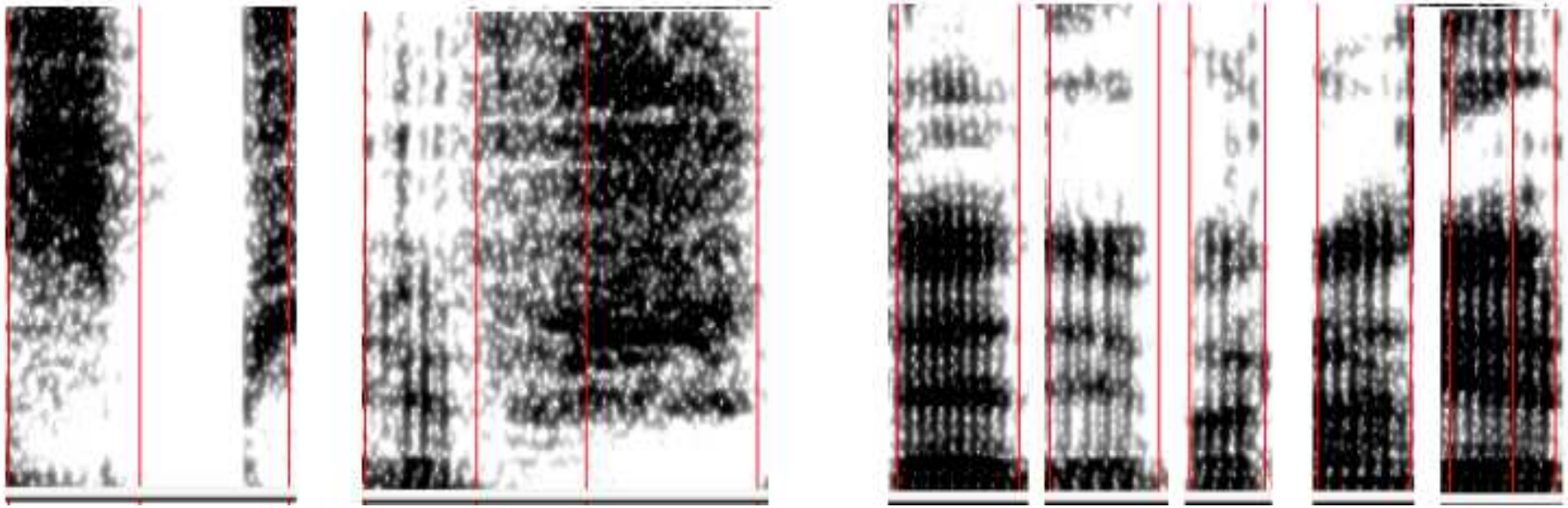


Wtedy po raz pierwszy
EN.Then for the first time

Segmentation into speech sounds

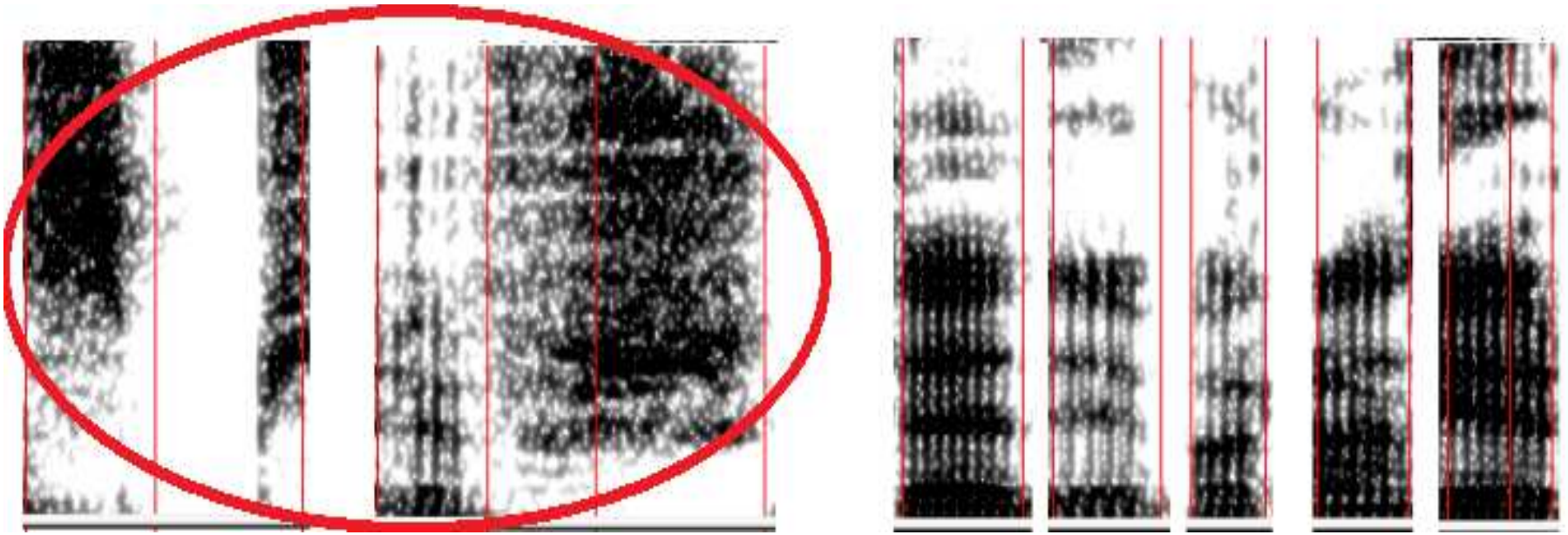


What kind of sounds are these?

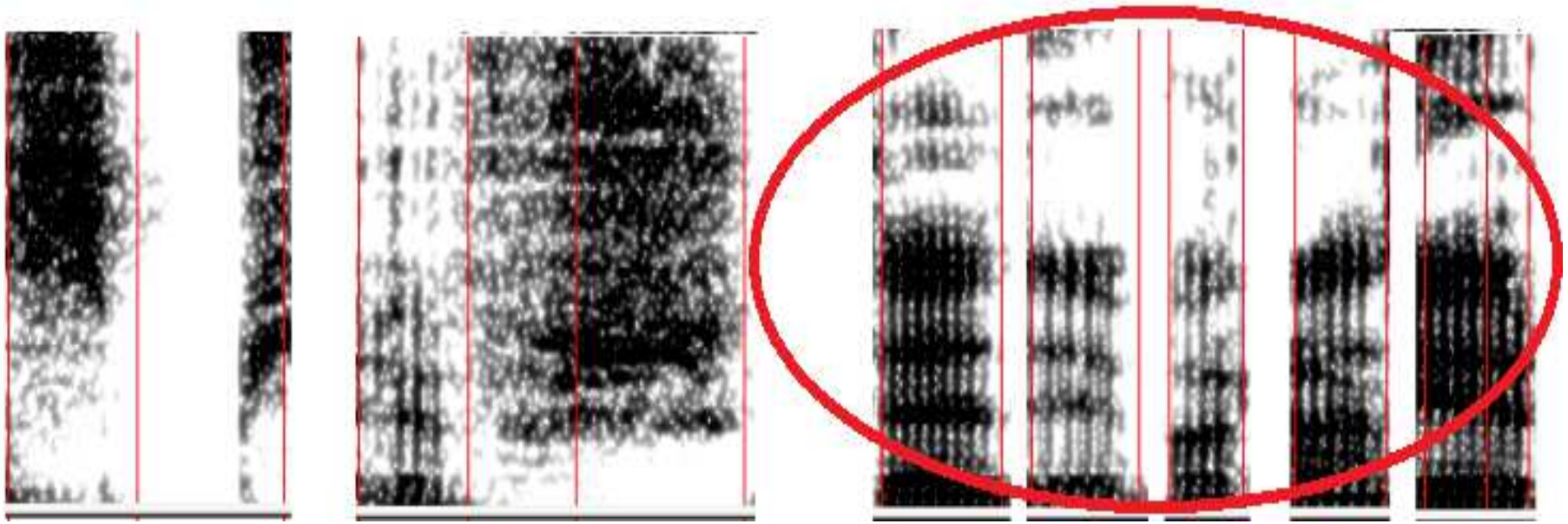


What speech sounds types? What specific sounds?

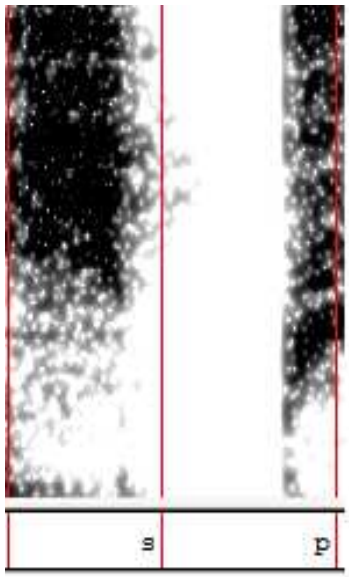
What kind of sounds are these?

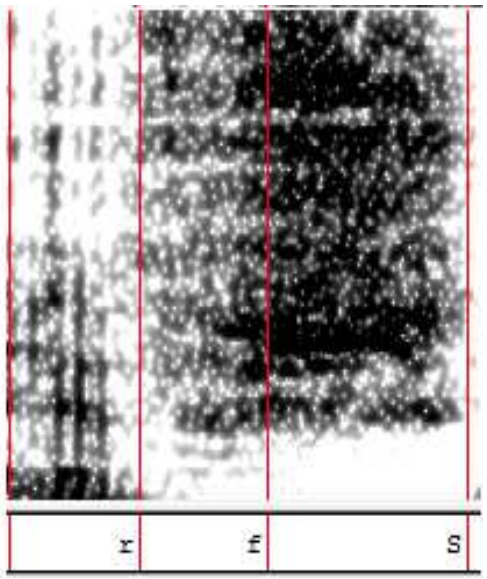


What kind of sounds are these?

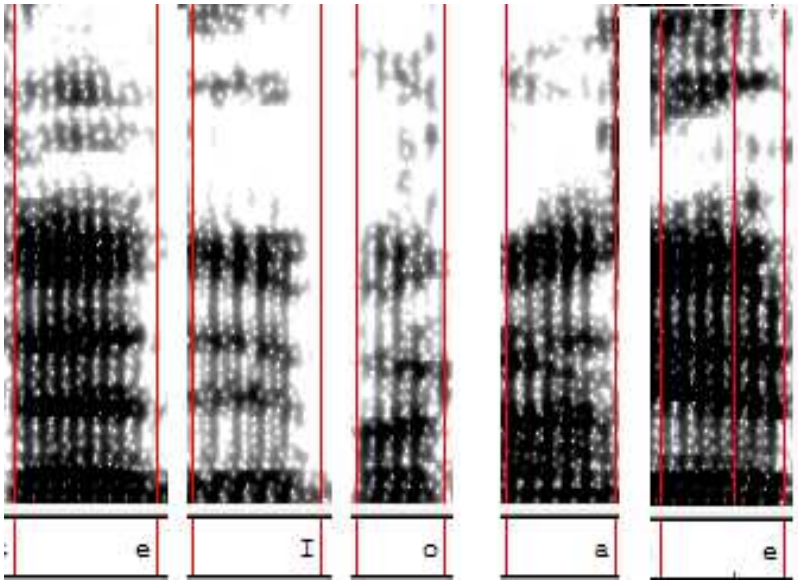


Noises (vowels) vs. consonants vs. vowels





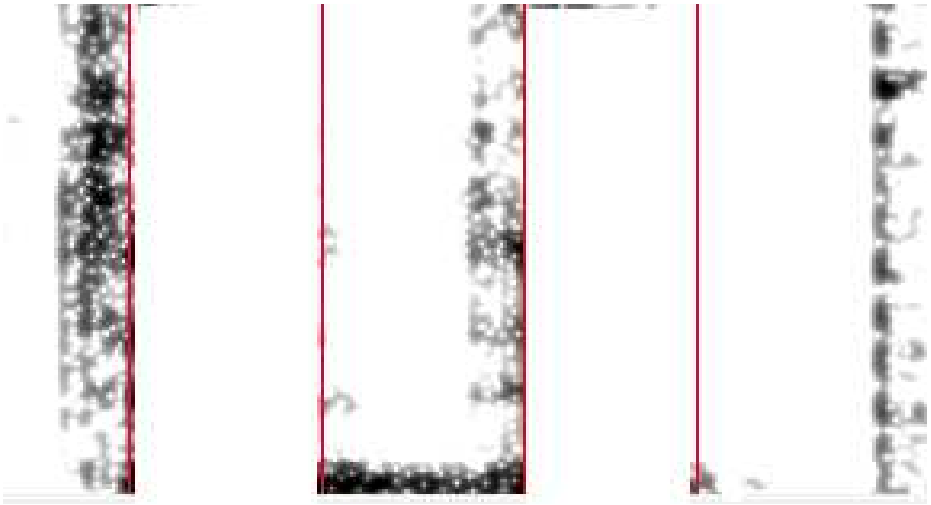
realisations of: s, p, r, f, S



realisations of: e, y, o, a, e

How is voicing demonstrated?

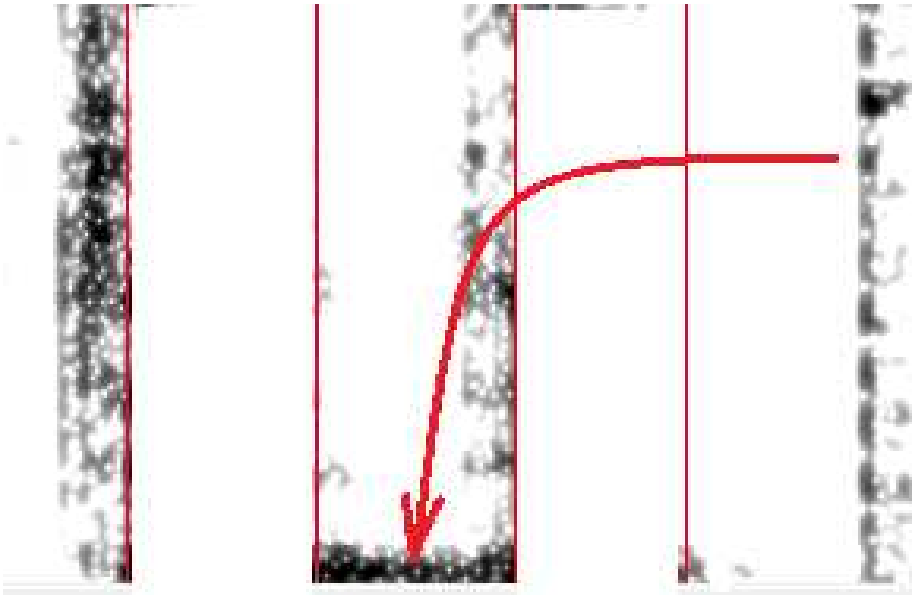
- The vocal cords vibrate at lower frequencies during production of voiced sounds - this is visible on a spectrogram, here: stop sounds:



How is voicing demonstrated?

- The vocal cords vibrate at lower frequencies during production of voiced sounds - this is visible on a spectrogram, here: stop sounds:

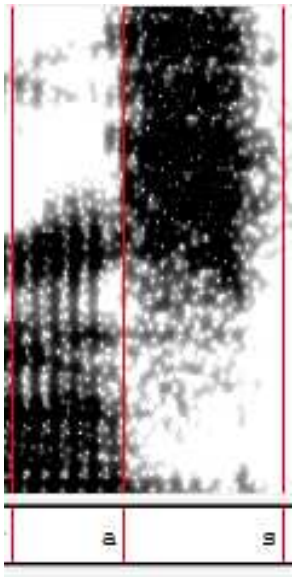
t, d, p

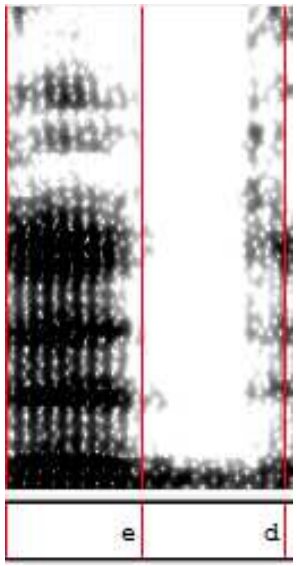


Segmentation into speech sounds

Boundaries:

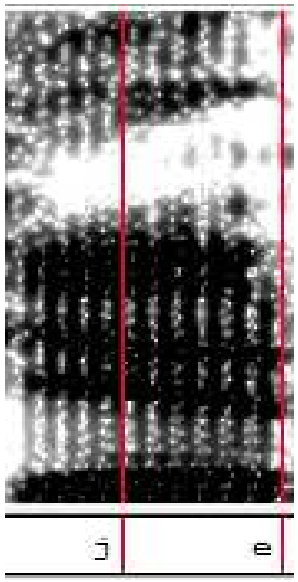
- rather clear

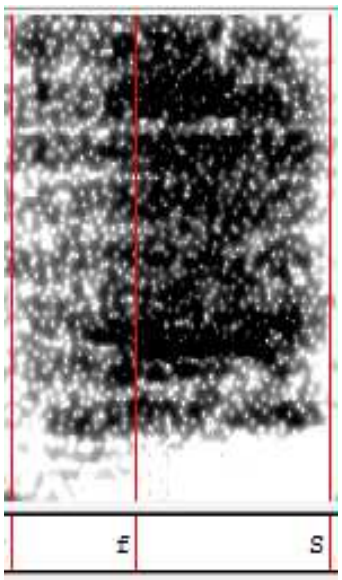




vowel/fricative,
vowel/stop

Boundaries:
continuous, ambiguous

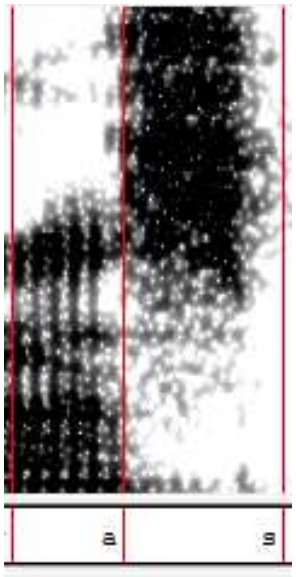


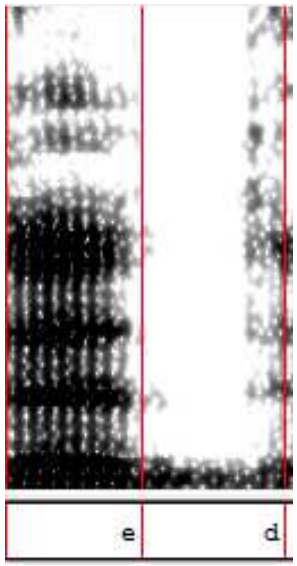


vowel/sonorant /j/
fricative/fricative

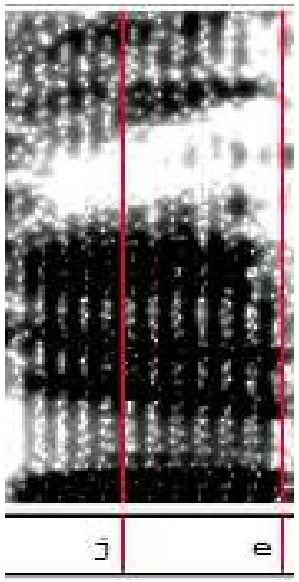
Segmentacja sygnału mowy na głoski

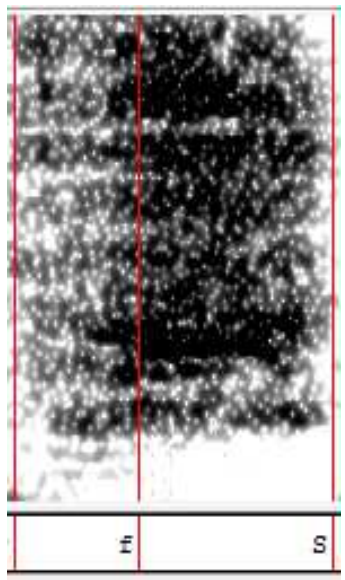
Granice
względnie jednoznaczne





Granice “ciągłe”,
“płynne”





Phonetic transcription: IPA

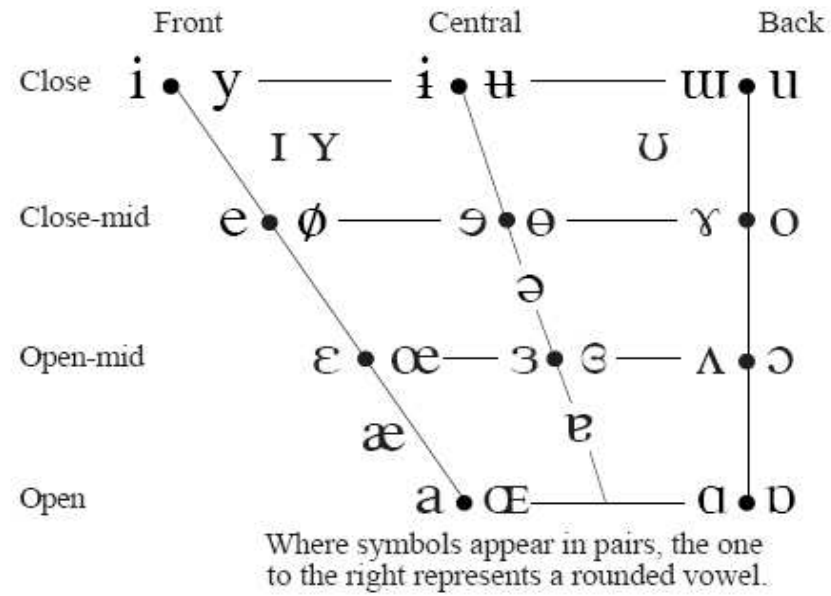
THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

Phonetic transcription
SAMPA - IPA

SAMPA - Speech Assessment Methods Phonetic Alphabet

SAMPA - no need for special fonts

SAMPA for Polish:
http://www.phon.ucl.ac.uk/home/sampa/polish.htm

TASK 1b

Please transcribe the „DzienDobry.wav” file using SAMPA phonetic alphabet.

TASK 2

1. Please find *the North Wind and the Sun* fable in your own language (a recording in wave format and a script if possible). If that's not possible, please use an English or Polish version (PL ver. available from the teacher)
2. Import or paste annotations to Annotation Pro
3. Adjust the annotations so that they match the recording

Thank you!

1. Contact e-mail: klessa@amu.edu.pl
2. Website: katarzyna.klessa.pl