



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Hands-on tutorial:

Using Praat for analysing a speech corpus

Mietta Lennes

12.-13.8.2005

Palmse, Estonia

Department of Speech Sciences
University of Helsinki





Objectives

- **Lecture:** Understanding what speech annotation means
 - efficient annotation
 - theoretical pitfalls

- **Exercises:** Learning to use Praat for annotating speech
 - basic techniques and analysis displays
 - incremental annotation

- **Exercises:** Using simple Praat scripts to analyse a small annotated speech corpus
 - understanding basic acoustic analyses
 - running and editing scripts



Annotation

- Annotation generally means describing, classifying and organizing (speech) material by systematically adding symbolic labels to its parts.
- The analyses you will be able to perform are restricted by the accuracy and types of annotations you have for your corpus.
- Up to date, no automatic speech segmentation or recognition tool exists for any language that can perform as well as a human annotator.



**Transcripts are not annotations as such.
Annotations and transcripts are not data.**



DATA
(audio)

...a lot of listening and typing...

One day Tim went shopping.
He bought yellow boots and a
funny hat.

TRANSCRIPT

TextGrid semmonen_juttu

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help

I

0.05374
-0.0002495
-0.08542
5000 Hz
0 Hz

1.901341

+ 1 s e / l i s E n m o n e n & j u x t l u x m phone

2 se o li sem mo nen jut tu syllable

3 se oli semmonen juttu word

4 se oli semmonen juttu utterance

5 ? H 0 ? ? Hv voice quality

6 .bro breath

7 paralinguistic

8 ok ready?

9 ok Checked by Mieta

10 school topic

11 school comment

12 school comparison

13 Lv intonation unit by He

14 Lv prominence by Hann

1.901341

0.000000 Visible part 1.958729 seconds 1.958729

Total duration 1.958729 seconds

all in out sel Group



Prerequisites for annotating and analysing a speech corpus

- Signal files in a format readable by the annotation tool (Praat: WAV, AIFF, AIFC, Next/Sun, NIST; 16- or 8-bit)
- Sufficiently high signal-to-noise ratio
- Different speakers should preferably be separated into different audio files (crosstalk is difficult to annotate).
- High acoustic quality is required for complex acoustic analyses (e.g., formant modeling).
- If studying speech and interaction, there should be a common timeline for all audio/video/other signal files.



Planning an annotation project

- Annotation is boring and time-consuming
-> you should make sure it is worth all the work!
- **Annotation should help to run analyses automatically and to reduce the need for manually browsing through your corpus.**
- Explore and practise with a small material, then complete your annotations.
- What are you aiming to study?

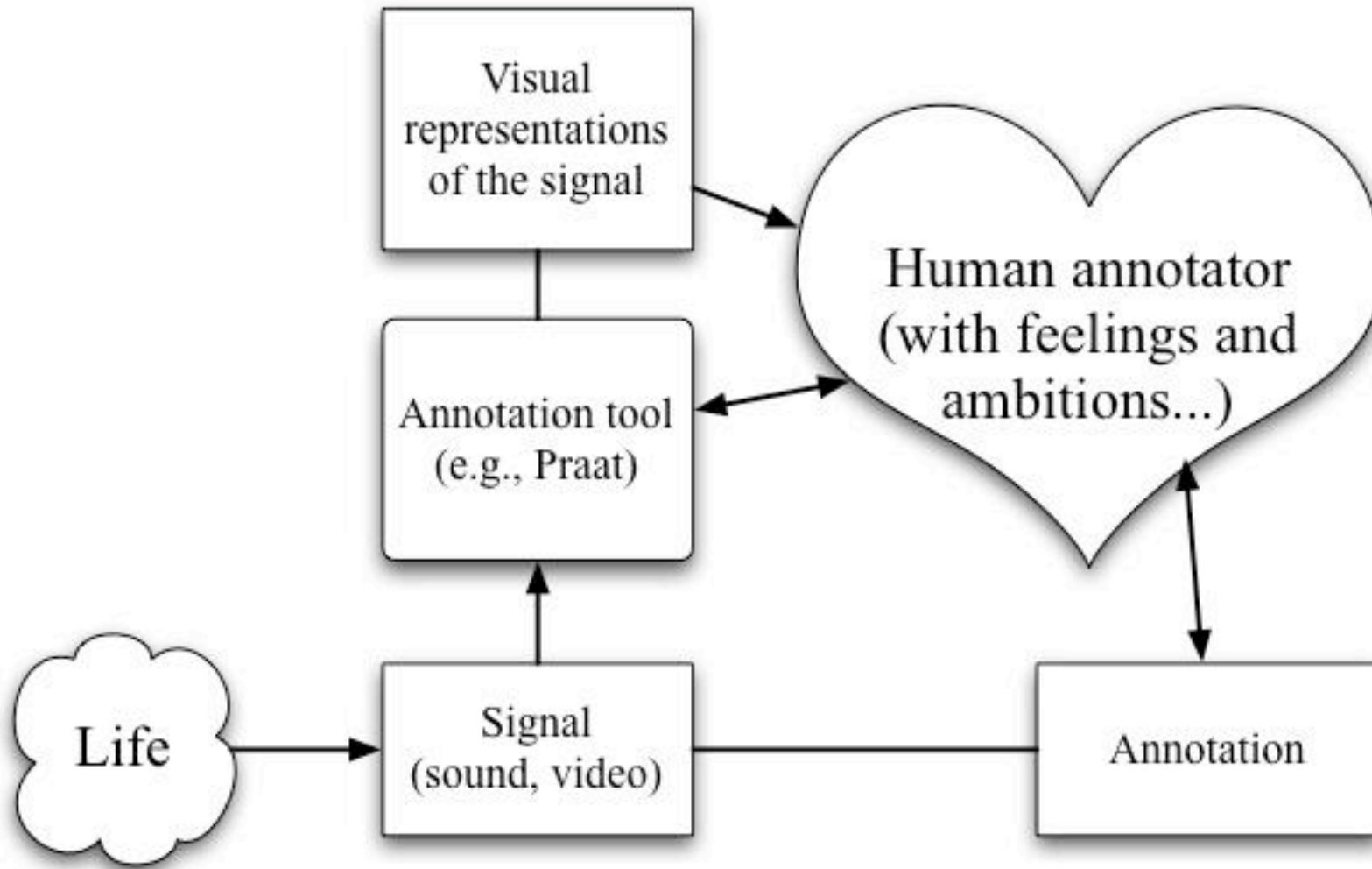


Remember...

- **Speech communication is much more than an "acoustic form of writing".**
- Writing things down in a specific notation and carefully classifying them does not make these things nor the categories any more real.
- All units that you plan to annotate tend to be "fuzzy" when you try to find them in real speech: the temporal boundaries are unclear, the different categories are sometimes difficult to separate, etc.



Annotation and the Human Factor...



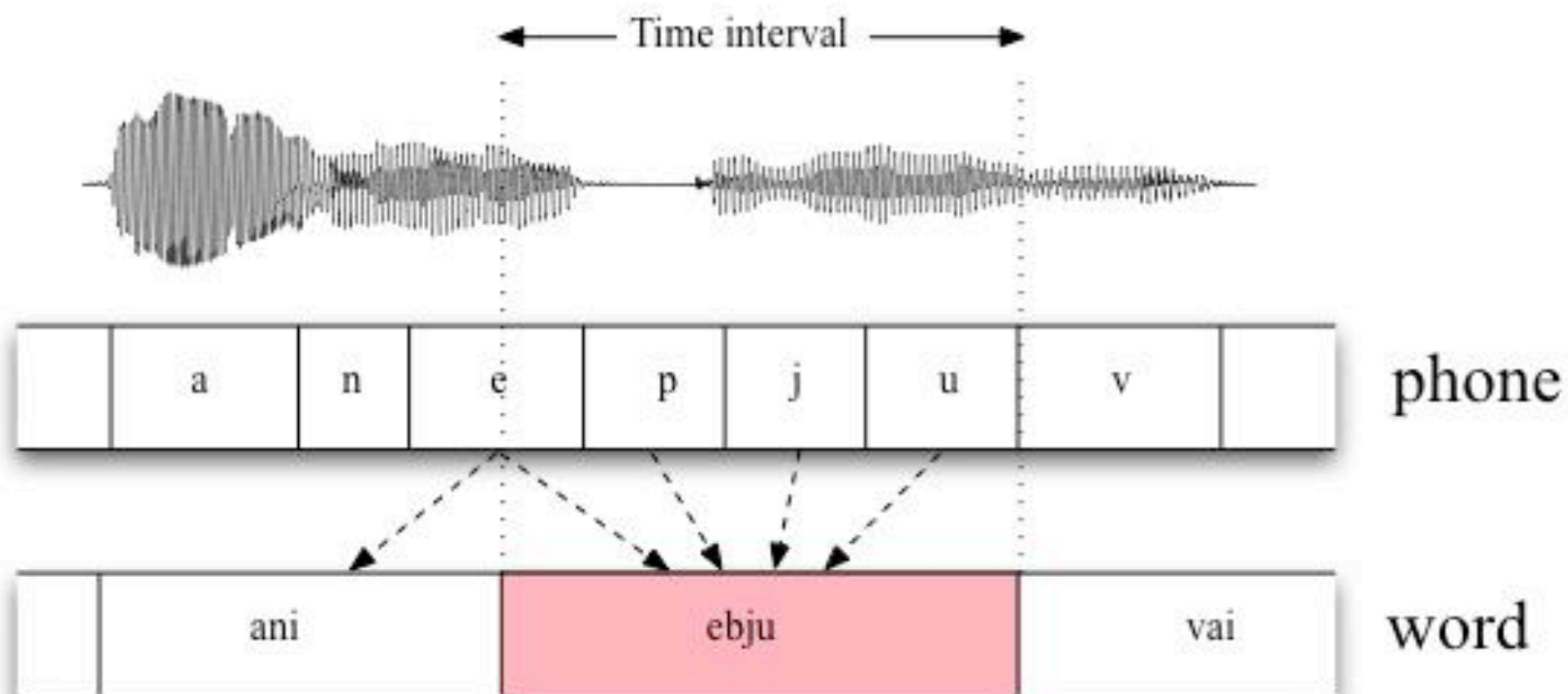


Defining your annotation structure

- List your **units**:
 - what kind of **labels** are allowed?
 - What kind of **properties** do your units have?
 - Which **values** are allowed for the properties?
 - How many layers (tiers) of annotation do you need?

- You should understand how the use of these units, labels and tiers can help you to automatically analyse your material in a consistent way.

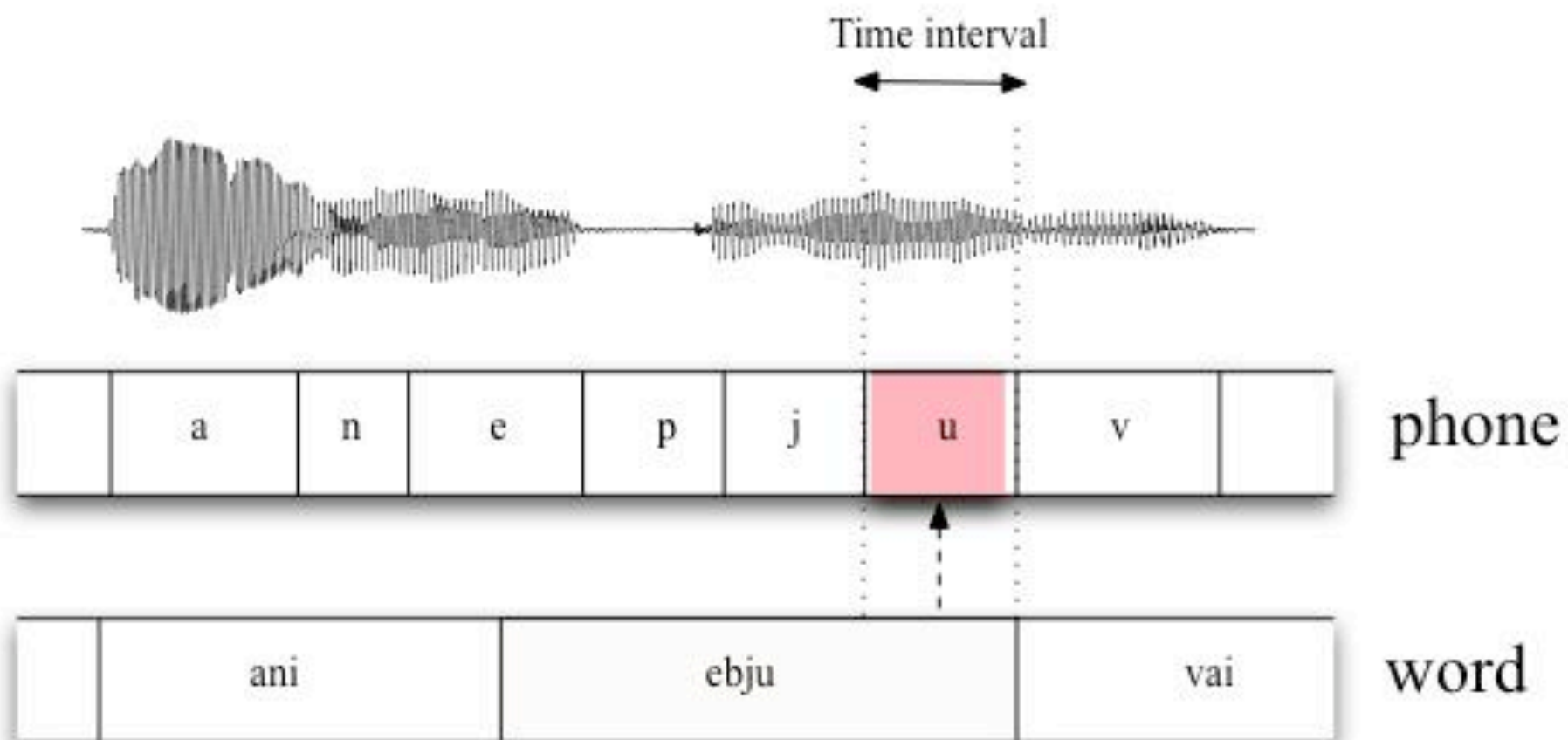
- Do not waste time labeling things that can be automatically measured! (e.g. labeling pause durations into a TextGrid)



Focus = word:

Interval	Label	Duration	Segments	MaxPitch
57	ani	0.327987	a,n,e	271.35837
58	ebju	0.500873	e,p,j,u	157.65027

etc...



Focus = phone:

Interval	Label	Duration	Word	MaxPitch
478	j	0.100368	ebu	153.50399
479	u	0.102878	ebu	157.65027

etc...



Metadata

- It is important to gather sufficiently detailed metadata about the speech material (speakers and their background, recording conditions, etc.)
- Metadata can also be used when analysing the corpus! E.g., the speakers' sex and age are factors that tend to affect their linguistic behaviour. (If a speech database system is not available, you can encode information about the speakers, e.g., into the filenames.)



Why choose Praat for analysing your corpus?

- Widely used, well known, well maintained
- Easily installed on multiple platforms
- Scriptable
- All Praat scripts and files can be made fully portable from one system to another.
- With Praat, you can use your corpus almost anywhere!



Why not to use Praat

- Video annotation must be done with another tool.
- Praat does not include a proper database system as such, so searching a speech corpus with Praat must be implemented through Praat scripts (which can become painfully slow).
- Recommended: If your corpus is large, use Praat (scripts) to dump your annotations and acoustic analysis results to a suitable format and do the searching and statistics somewhere else.



Links

- Praat: <http://www.praat.org>
- Praat scripts: <http://www.helsinki.fi/~lennes/praat-scripts/>
- Linguistic annotation (tools and formats):
<http://www ldc.upenn.edu/annotation/>
- Annotation guide (in Finnish; a "public draft" version):
<http://www.helsinki.fi/~lennes/nimikointiopas.html>
- An RDF/XML Schema for formally defining your annotation structure, e.g., in your own applications:
<http://www.csc.fi/kielipankki/projektit/sapuhe/>