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.

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

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

kuka jossa esimerkkejä monenmoisista
annotaatiokerroksista

se / ll

sem monen

ju

tu

se oli semmonen

juttu

bro

ok

ok

school

Ly
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:

<table>
<thead>
<tr>
<th>Interval</th>
<th>Label</th>
<th>Duration</th>
<th>Segments</th>
<th>MaxPitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>ani</td>
<td>0.327987</td>
<td>a,n,e</td>
<td>271.35837</td>
</tr>
<tr>
<td>58</td>
<td>ebju</td>
<td>0.500873</td>
<td>e,p,j,u</td>
<td>157.65027</td>
</tr>
</tbody>
</table>

etc...
Multiple annotation layers:

Phone units in search focus

Time interval

```
| ane | p | j | u | v |
```

```
| ani | ebju | vai |
```

**Focus = phone:**

<table>
<thead>
<tr>
<th>Interval</th>
<th>Label</th>
<th>Duration</th>
<th>Word</th>
<th>MaxPitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>478</td>
<td>j</td>
<td>0.100368</td>
<td>ebju</td>
<td>153.50399</td>
</tr>
<tr>
<td>479</td>
<td>u</td>
<td>0.102878</td>
<td>ebju</td>
<td>157.65027</td>
</tr>
</tbody>
</table>

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