

# Speech synthesis

## Text-to-Speech (TTS)

Jolanta Bachan

# Speech synthesis

- Automatic system for converting *written* text to *spoken* speech
- Text-to-Speech (TTS)
- TTS should be able to read any text, but in practice it is not so easy to do

# Speech synthesis

- Parametric speech synthesis
  - formant synthesis
  - articulatory synthesis (*VTdemo*, vocal tract demo)
- Concatenative speech synthesis
  - diphone synthesis
  - triphone synthesis
  - unit selection
- HMM-based speech synthesis (Hidden Markov Models)
  - Statistical Parametric Synthesis
- End-to-End (e2e) speech synthesis based on deep learning

# Speech synthesis

- Parametric speech synthesis
  - formant synthesis
  - articulatory synthesis (*VTdemo*, vocal tract demo)
- Concatenative speech synthesis
  - diphone synthesis
  - triphone synthesis
  - unit selection
- HMM-based speech synthesis (Hidden Markov Models)
  - Statistical Parametric Synthesis
- End-to-End (e2e) speech synthesis based on deep learning

# Concatenative speech synthesis

- Concatenative speech synthesis combines smaller units of recorded speech (diphones, triphones, syllables, words) into larger units (words, sentences).
- The system is based on speech recordings. The database is segmented into smaller units (phones, diphones, words), from which utterances are concatenated/glued together.

# What is Close Copy Speech Synthesis?

# Close Copy Speech Synthesis

The CCS synthesis system produces a sound which “repeats an utterance produced by a human speaker with a synthetic voice, while keeping the original prosody” (Dutoit, 1996).

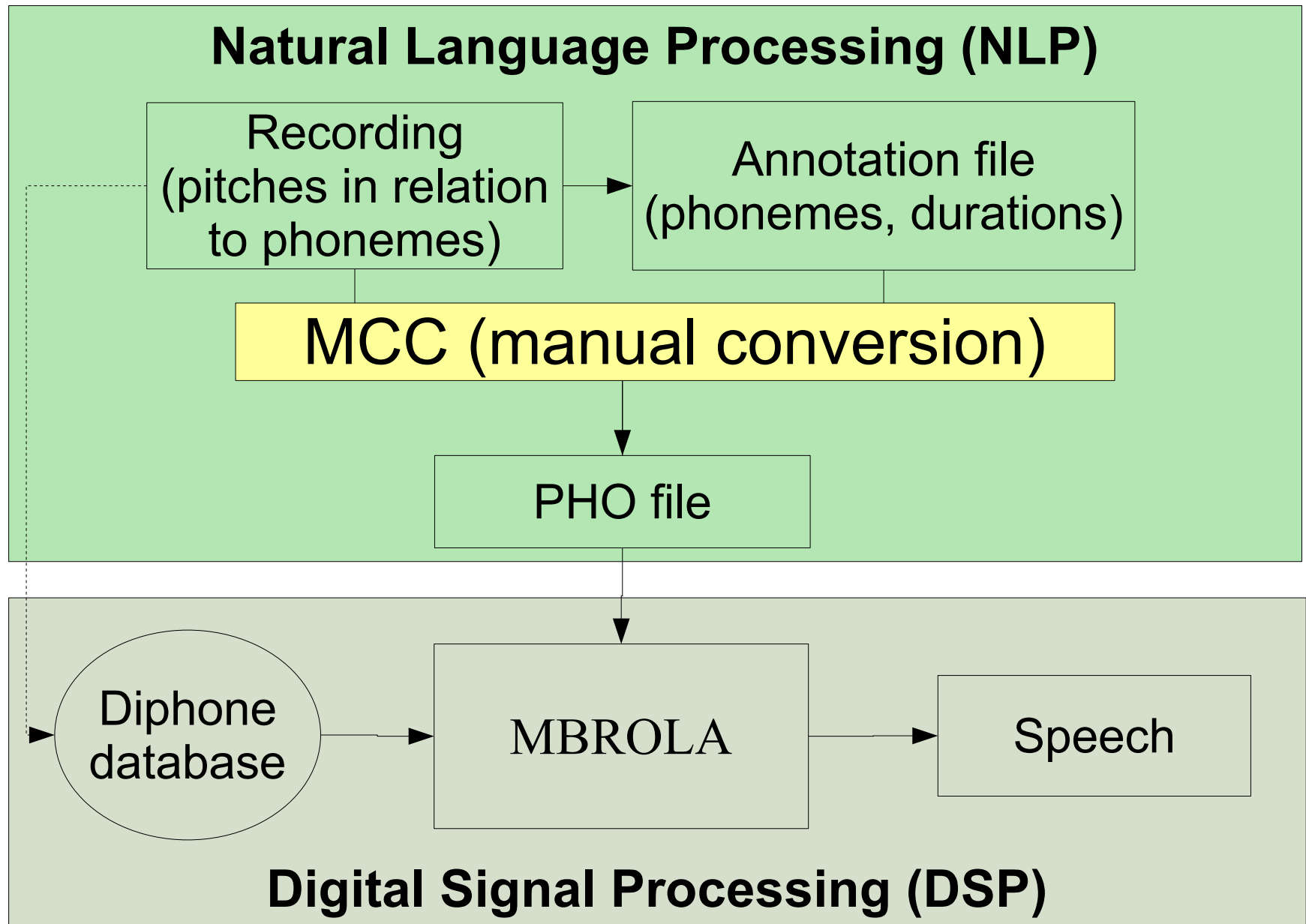
# Manual Close Copy Speech (MCCS) Resynthesis

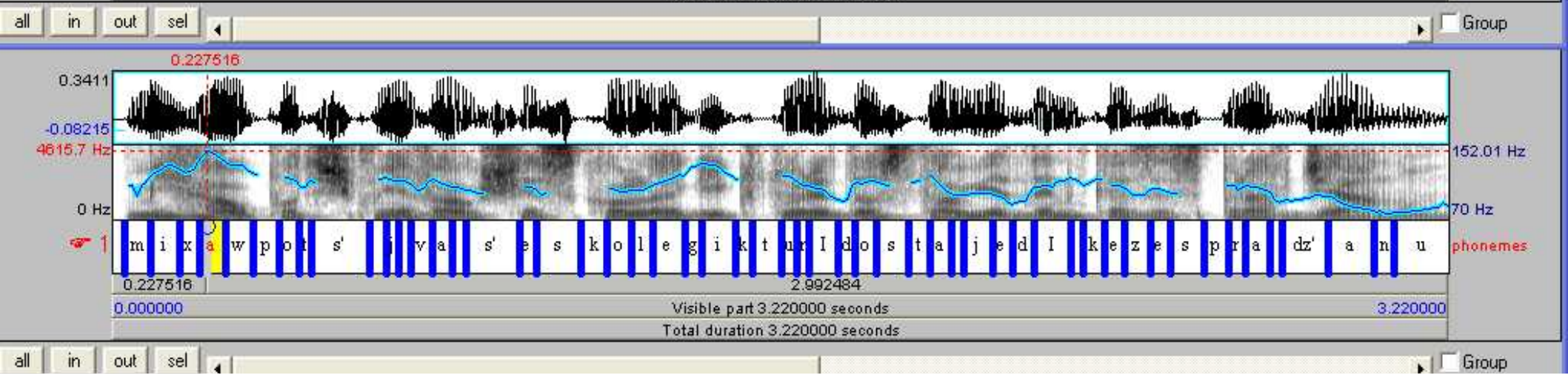
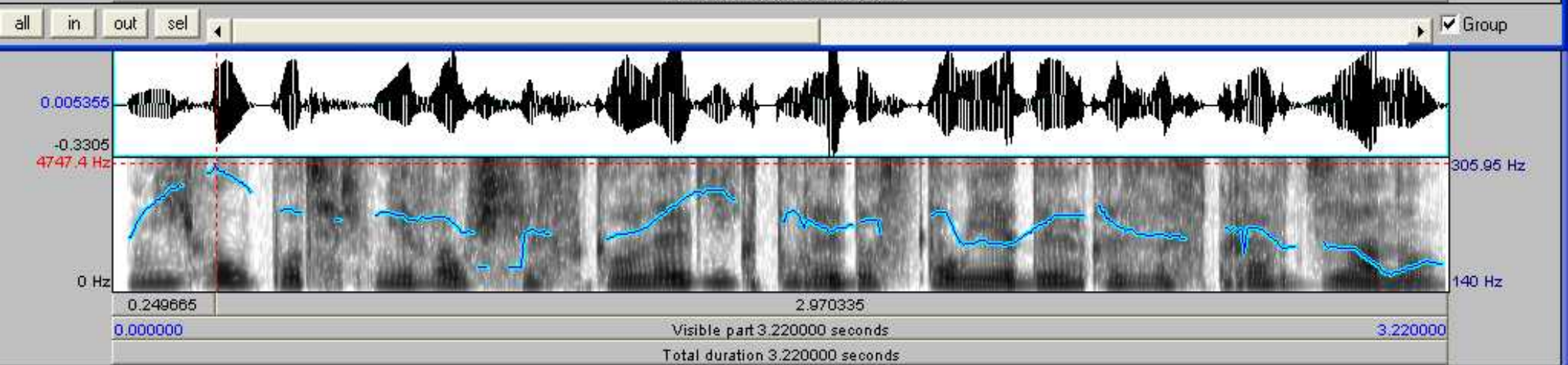
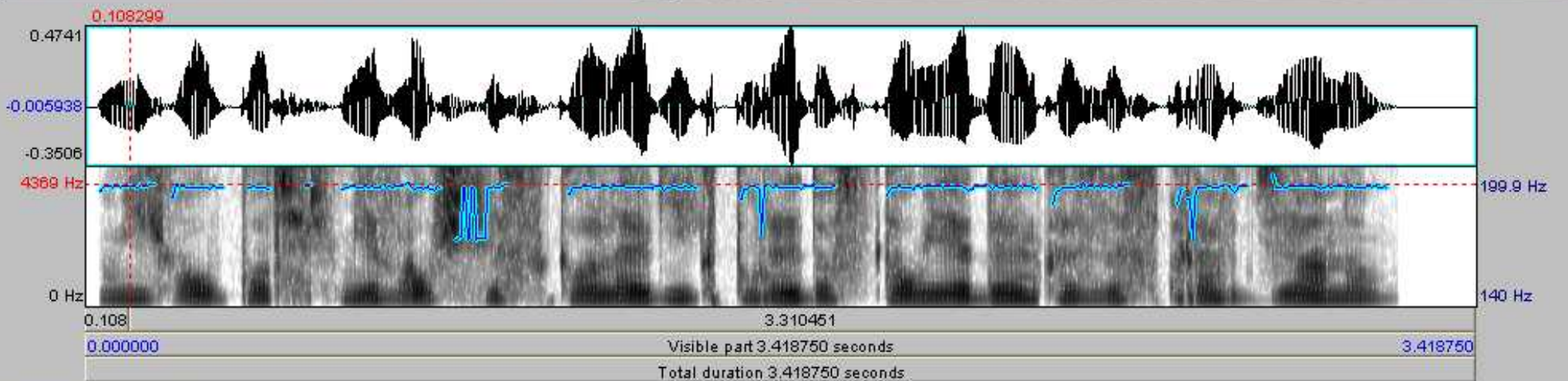


# MCCS components

- Input: Speech
  - speech recordings
  - annotation of speech recordings
- Speech synthesiser (here: MBROLA)
  - diphone database (MBROLA voice)
  - synthesis engine

# MCCS synthesis





# MCCS

- Monotonous

Monotonous

- MCCS synthesis

MCCS

- Original

Original recording

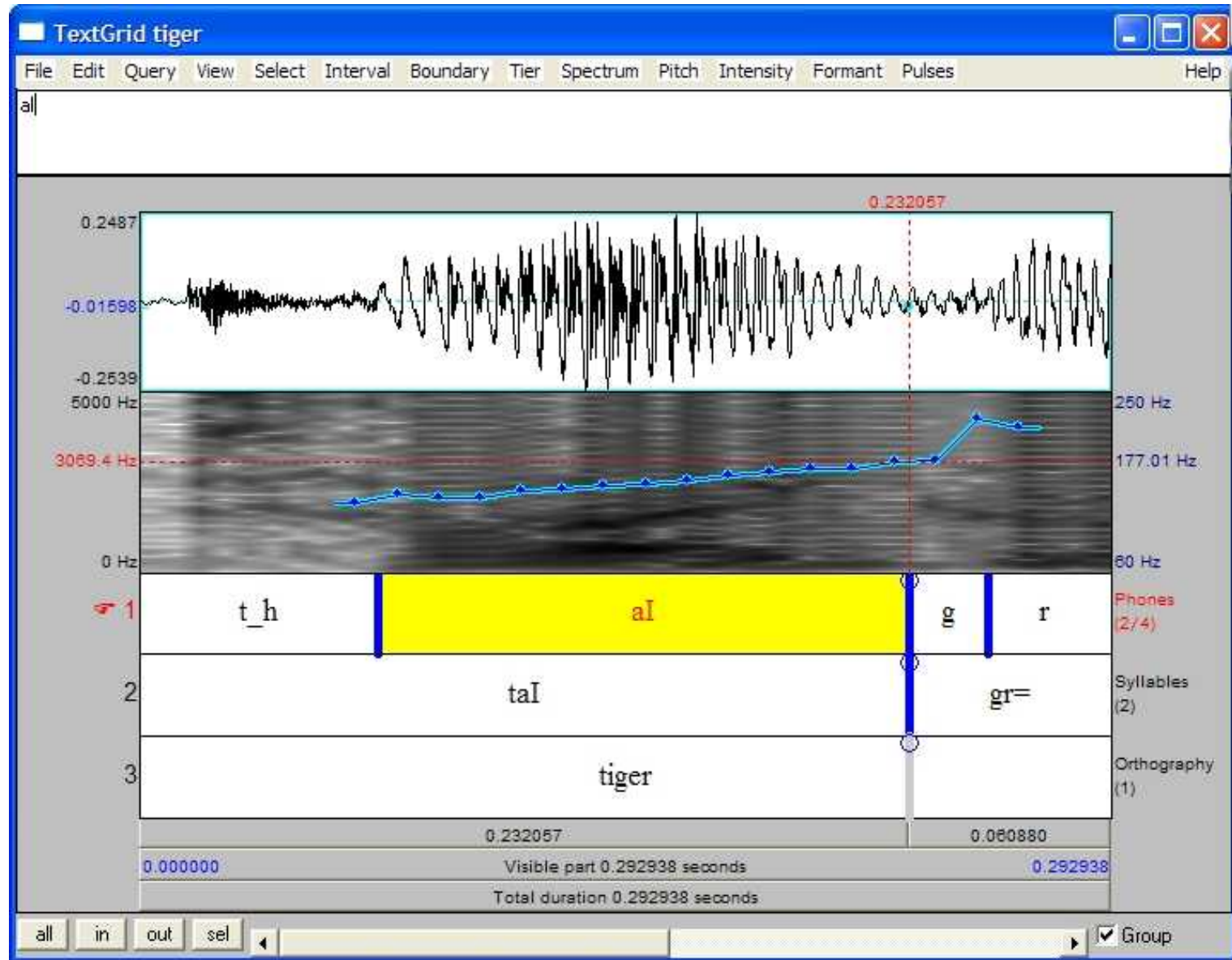
# Applications of MCCS

- Creating stimuli for speech perception tests
- A tool for teaching phonetics
- Experimenting with prosody
- Checking correctness of annotations

# Speech annotation in Praat

# Annotation for CCS synthesis

REQUIRED  
FOR  
SPEECH  
RE-SYNTHESIS



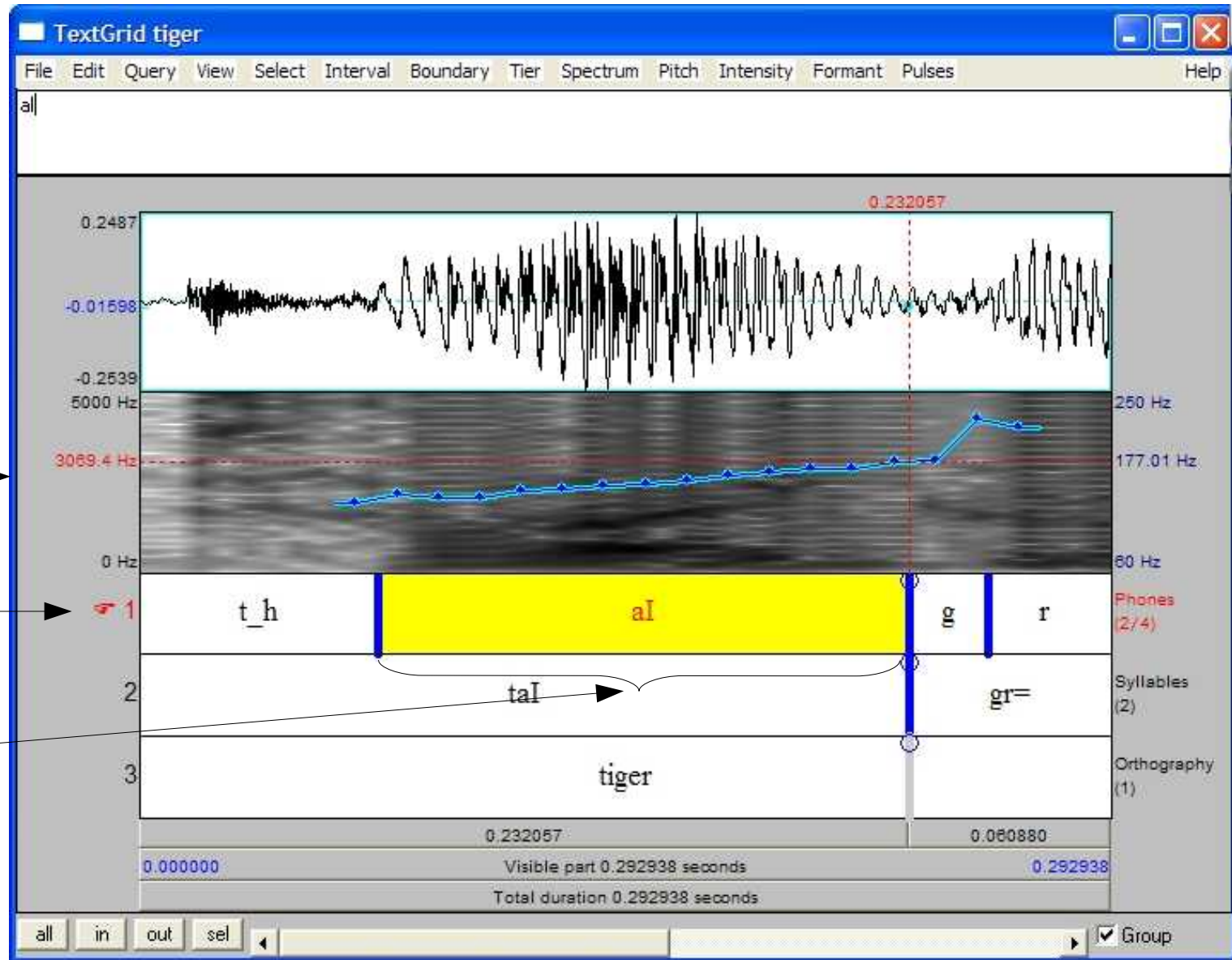
# Annotation for CCS synthesis

REQUIRED  
FOR  
SPEECH  
RE-SYNTHESIS

Pitch

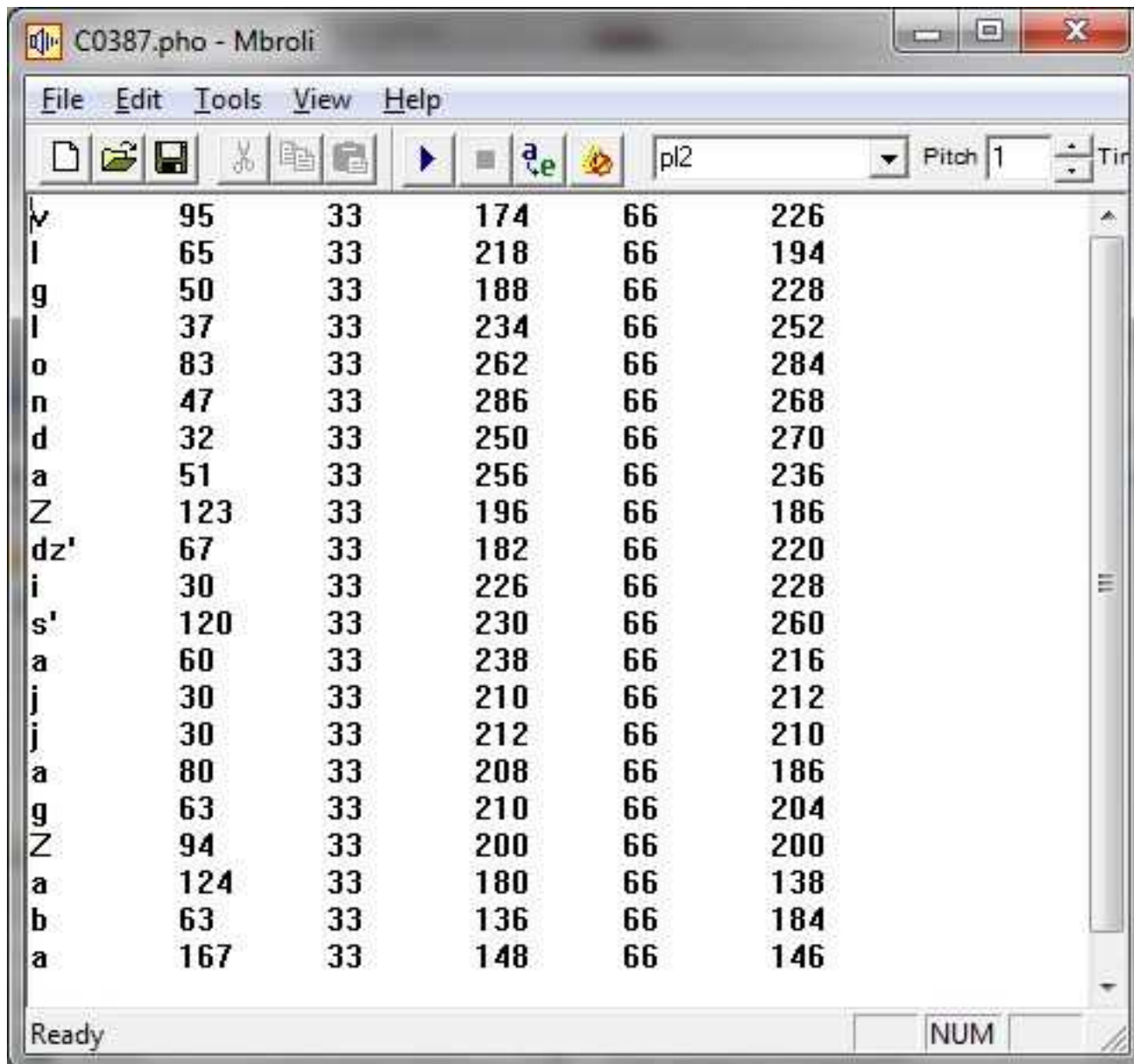
Phones

Durations





# PHO file format



The screenshot shows a window titled "C0387.pho - Mbroli" with a menu bar (File, Edit, Tools, View, Help) and a toolbar. The main area contains a table of phonetic data. The status bar at the bottom shows "Ready" and "NUM".

v	95	33	174	66	226
l	65	33	218	66	194
g	50	33	188	66	228
l	37	33	234	66	252
o	83	33	262	66	284
n	47	33	286	66	268
d	32	33	250	66	270
a	51	33	256	66	236
Z	123	33	196	66	186
dz'	67	33	182	66	220
i	30	33	226	66	228
s'	120	33	230	66	260
a	60	33	238	66	216
j	30	33	210	66	212
j	30	33	212	66	210
a	80	33	208	66	186
g	63	33	210	66	204
Z	94	33	200	66	200
a	124	33	180	66	138
b	63	33	136	66	184
a	167	33	148	66	146

# PHO file format

Phones:  
SAMPA

Durations

F0 positions

F0 values

The screenshot shows a window titled 'C0387.pho - Mbrolé' with a menu bar (File, Edit, Tools, View, Help) and a toolbar. The main area contains a table of phonetic data. The table has 6 columns: Phone, Duration, F0 position 1, F0 position 2, F0 position 3, and F0 position 4. The data is as follows:

v	95	33	174	66	226
l	65	33	218	66	194
g	50	33	188	66	228
l	37	33	234	66	252
o	83	33	262	66	284
n	47	33	286	66	268
d	32	33	250	66	270
a	51	33	256	66	236
Z	123	33	196	66	186
dz'	67	33	182	66	220
i	30	33	226	66	228
s'	120	33	230	66	260
a	60	33	238	66	216
j	30	33	210	66	212
j	30	33	212	66	210
a	80	33	208	66	186
g	63	33	210	66	204
Z	94	33	200	66	200
a	124	33	180	66	138
b	63	33	136	66	184
a	167	33	148	66	146

# Start with MBROLA

- Install **MBROLA**
- Via „Control Panel” in Mbrola Tools add diphone databases **pl1** and **en1**.
- Do MCCS in Mbrola.exe
  - Prepare a PHO file for your annotated speech recording for Polish or English
- Create your own PHO file and synthesise speech

# Speech synthesis evaluation

- intelligibility
- naturalness
- MOS scale – *Mean Opinion Score*, from 1 to 5, where 5 is the highest grade

# Task

- Create your own PHO file in Mbroli.

**See you next week!**