Speech Synthesis for Linguists:

An Introduction to MBROLA

Dafydd Gibbon

Universität Bielefeld
May 2007

OVERVIEW

Speech Synthesis

- Definition:
  - Speech Synthesis is communication using software which implements an artificial voice
- Types:
  - Text-To-Speech synthesis (TTS)
  - Concept-To-Speech synthesis (CTS)
  - Close Copy Speech synthesis (CCS)
- Inverse:
  - Automatic Speech Recognition (ASR)

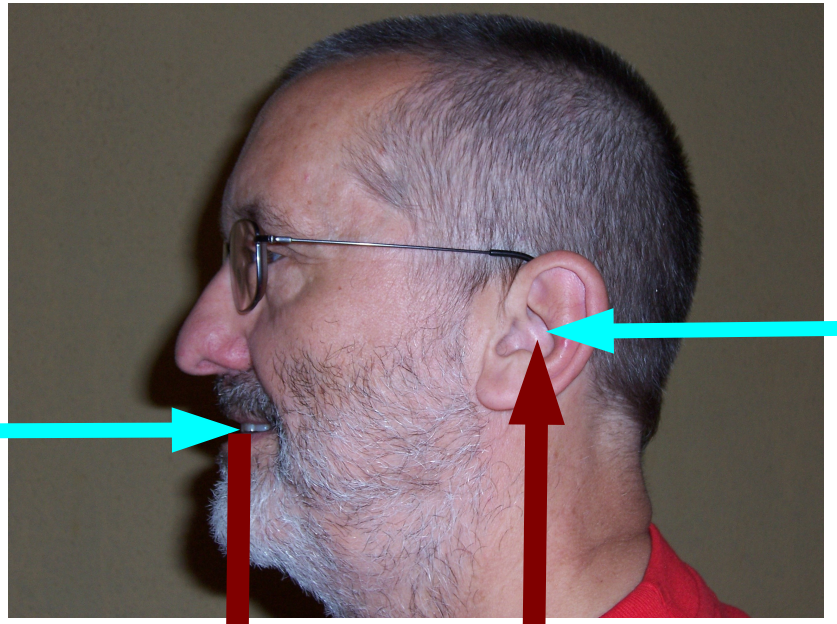
Speech Synthesis

- Uses:
  - Reading software for the blind
  - Computer output in visually difficult situations
  - Readback in dictation software
  - Linguistic research and language teaching
- Illustration:
  - Microsoft Sam
  - ...
- Background information:
  - Speech Synthesis - Wikipedia
  - The MBROLA project

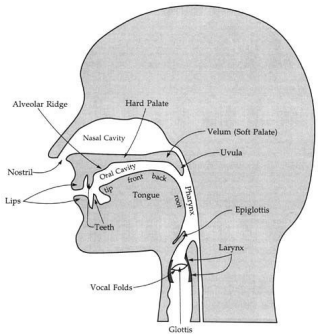
SPEECH COMMUNICATION

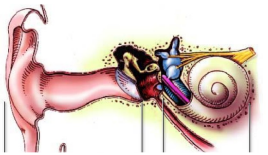
Natural speech cycle

Articulatory organs

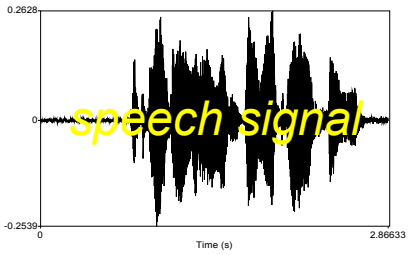


Auditory organs





A tiger and a mouse were walking in a field...



Acoustic channel

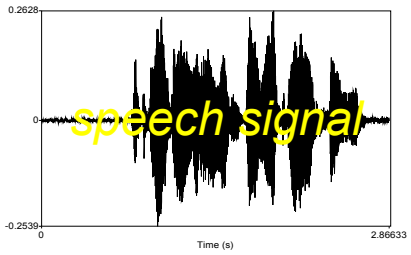
Artificial speech cycle

Speech synthesis

Speech Technology Software

ASR

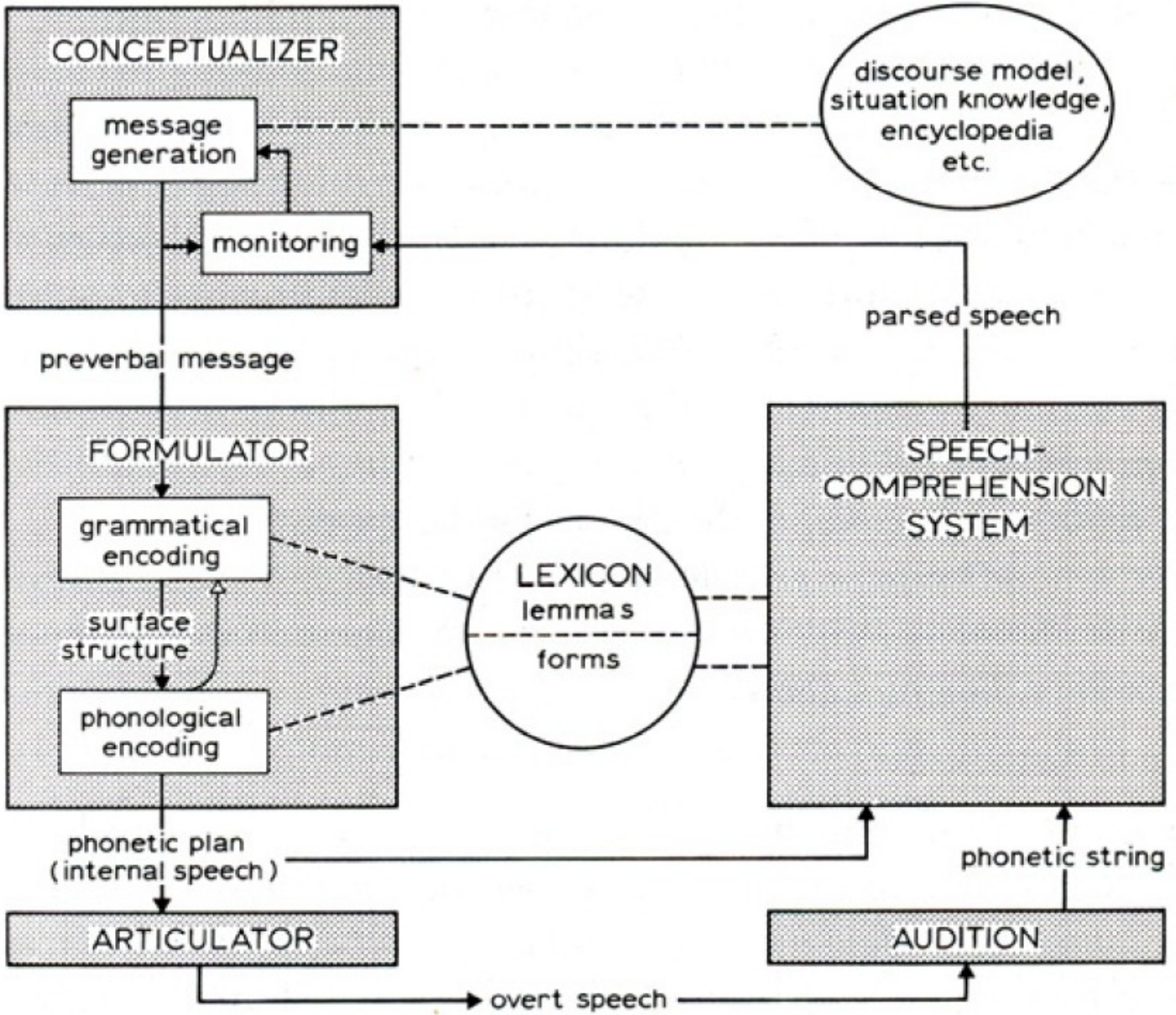
It would be a considerable invention...



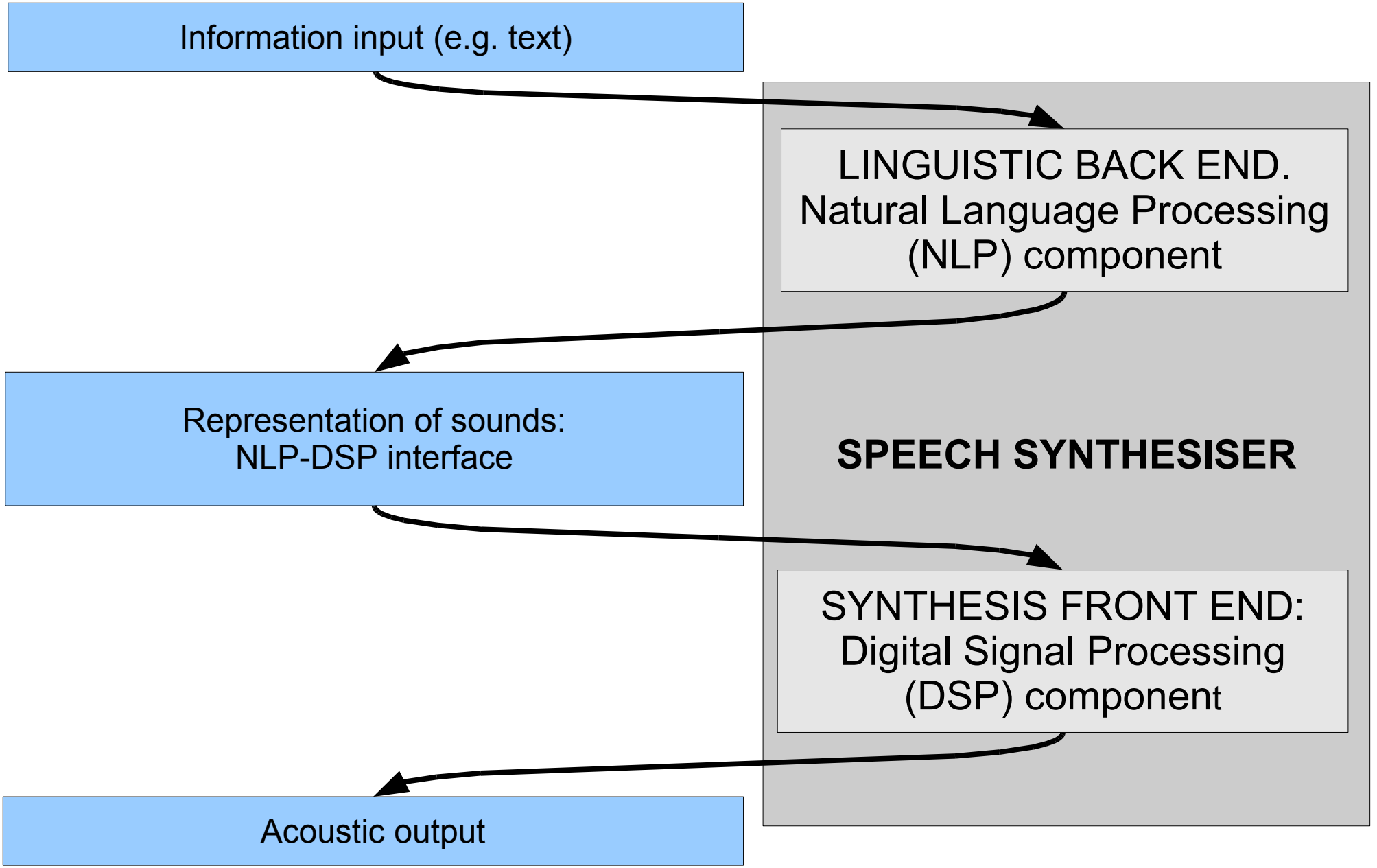
Acoustic channel

SPEECH PROCESSING

Natural speech processing



Artificial speech processing

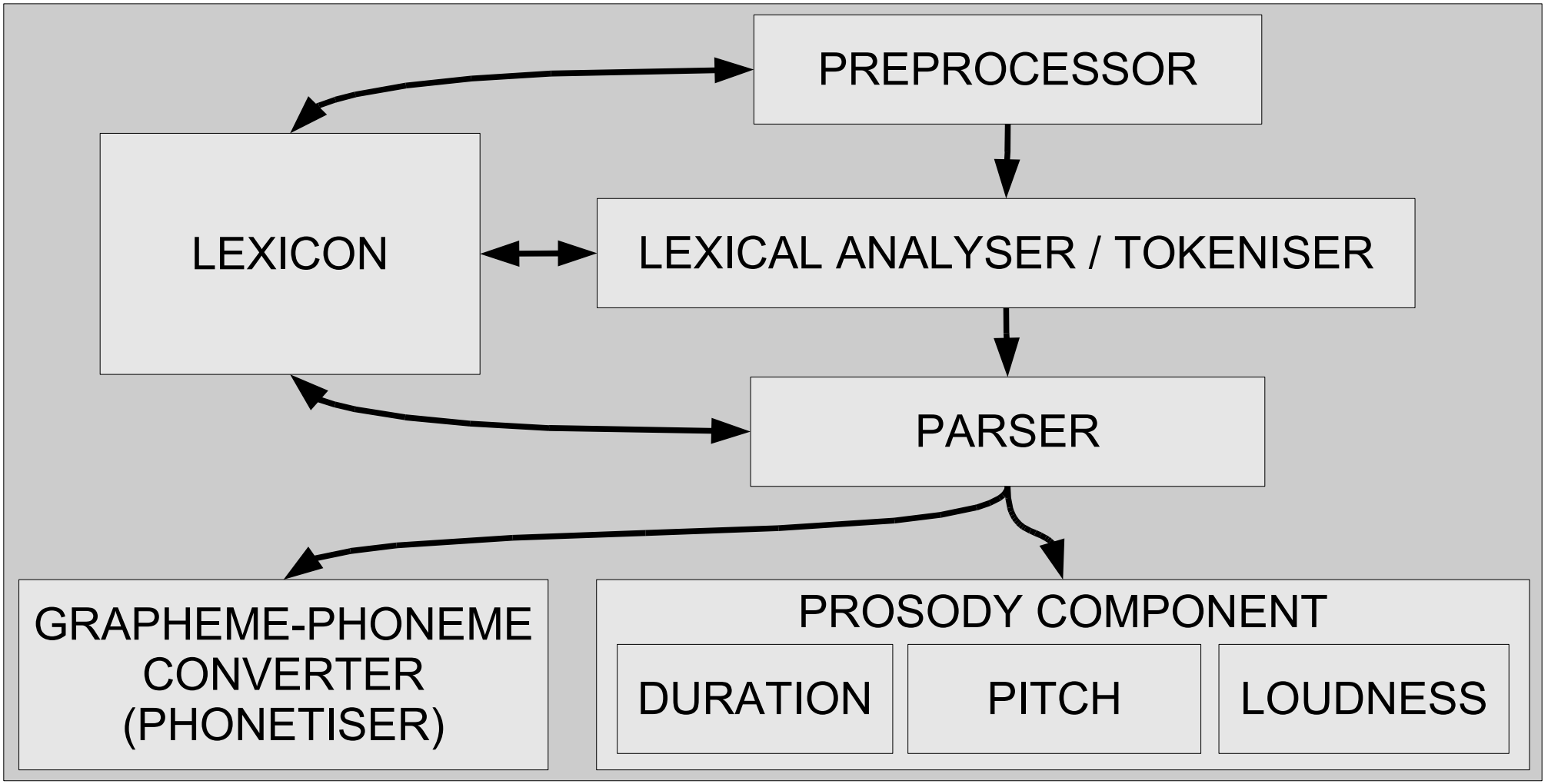


Text-To-Speech NLP Component

- Preprocessing
  - Abbreviations
  - Numbers
- Lexical analysis, tokenisation
  - Identification of words (word boundaries, lexicon)
- Parsing
  - Identification of parts of speech
  - Identification of phrases (grammar)
  - Identification of stress, focus, emphasis positions
- Phonetisation
  - Grapheme-phoneme conversion
  - Prosodic analysis
    - Pitch assignment (accentuation, intonation)
    - Duration assignment (tempo, phrasing, rhythm)

TTS NLP Components

TEXT INPUT



NLP-DSP INTERFACE TO SYNTHESIS ENGINE

NLP-DSP INTERFACE

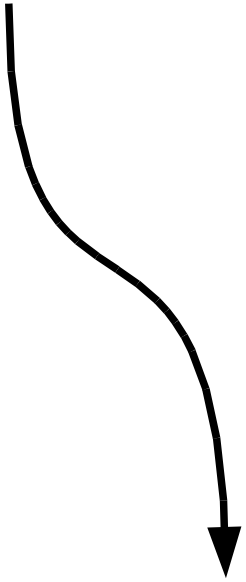
SIMPLIFIED NLP-DSP INTERFACE

MBROLA

General Information about MBROLA

- The MBROLA Project
- DSP front end only
  - you have to find or make the NLP back end:
    - TTS, CCS, ...
- Voice:
  - diphone database
  - MBROLA timing and pitch normalisation algorithm
  - normalised intensity
- Development
  - there are many MBROLA voices for many languages
  - anyone can make an MBROLA voice:
    - recording-annotation-diphone splitting - conversion
    - conversion by MBROLA team
    - voice is then in the public domain

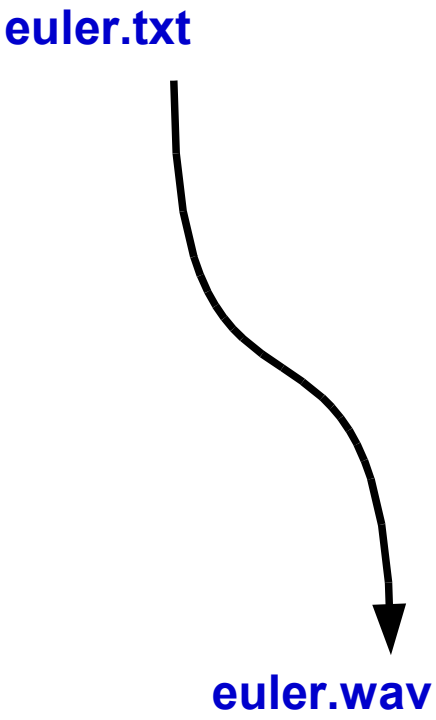
Example



euler.wav

Example

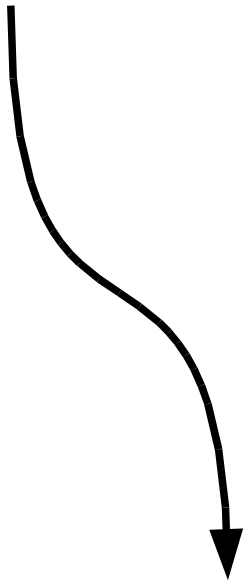
“It would be a considerable invention indeed, that of a machine able to mimic our speech, with its sounds and articulations. I think it is not impossible.”
Leonard Euler, 1761



Example

"It would be a considerable invention indeed, that of a machine able to mimic our speech, with its sounds and articulations. I think it is not impossible."
Leonard Euler, 1761

euler.txt

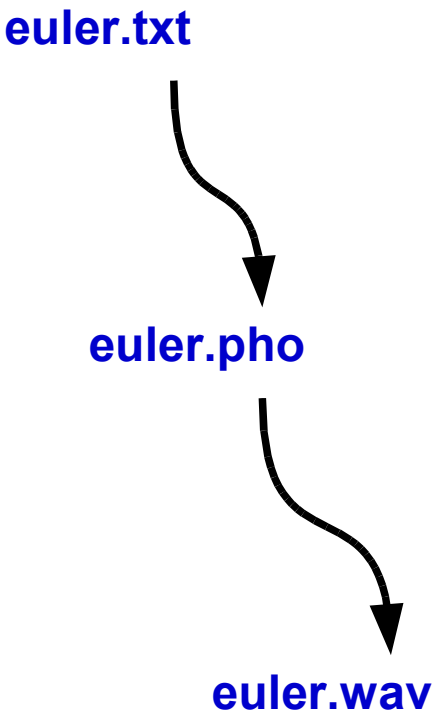


euler.wav

Phoneme (SAMPA)	Duration (ms)	Pitch place/value pairs (% Hz)
I	60	75 109
n	80	
v	40	
e	90	0 109 50 153 75 177
n	30	25 195
S	80	
@	70	
n	30	

Example

"It would be a considerable invention indeed, that of a machine able to mimic our speech, with its sounds and articulations. I think it is not impossible."
Leonard Euler, 1761



Phoneme (SAMPA)	Duration (ms)	Pitch place/value pairs (% Hz)
I	60	75 109
n	80	
v	40	
e	90	0 109 50 153 75 177
n	30	25 195
S	80	
@	70	
n	30	

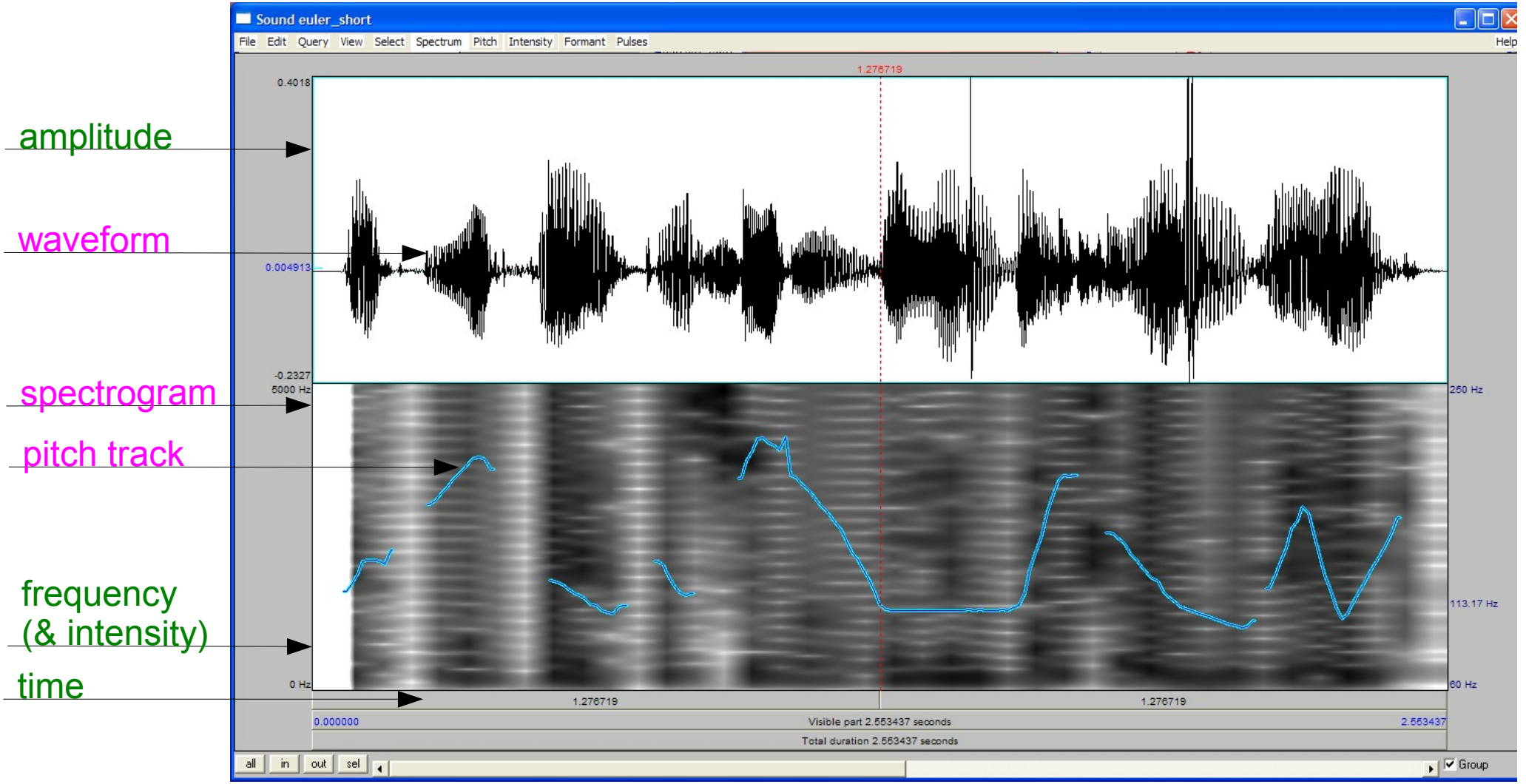
Output

euler.pho **euler_short.pho**

euler.wav **euler_short.wav**

Praat

euler_short.TextGrid



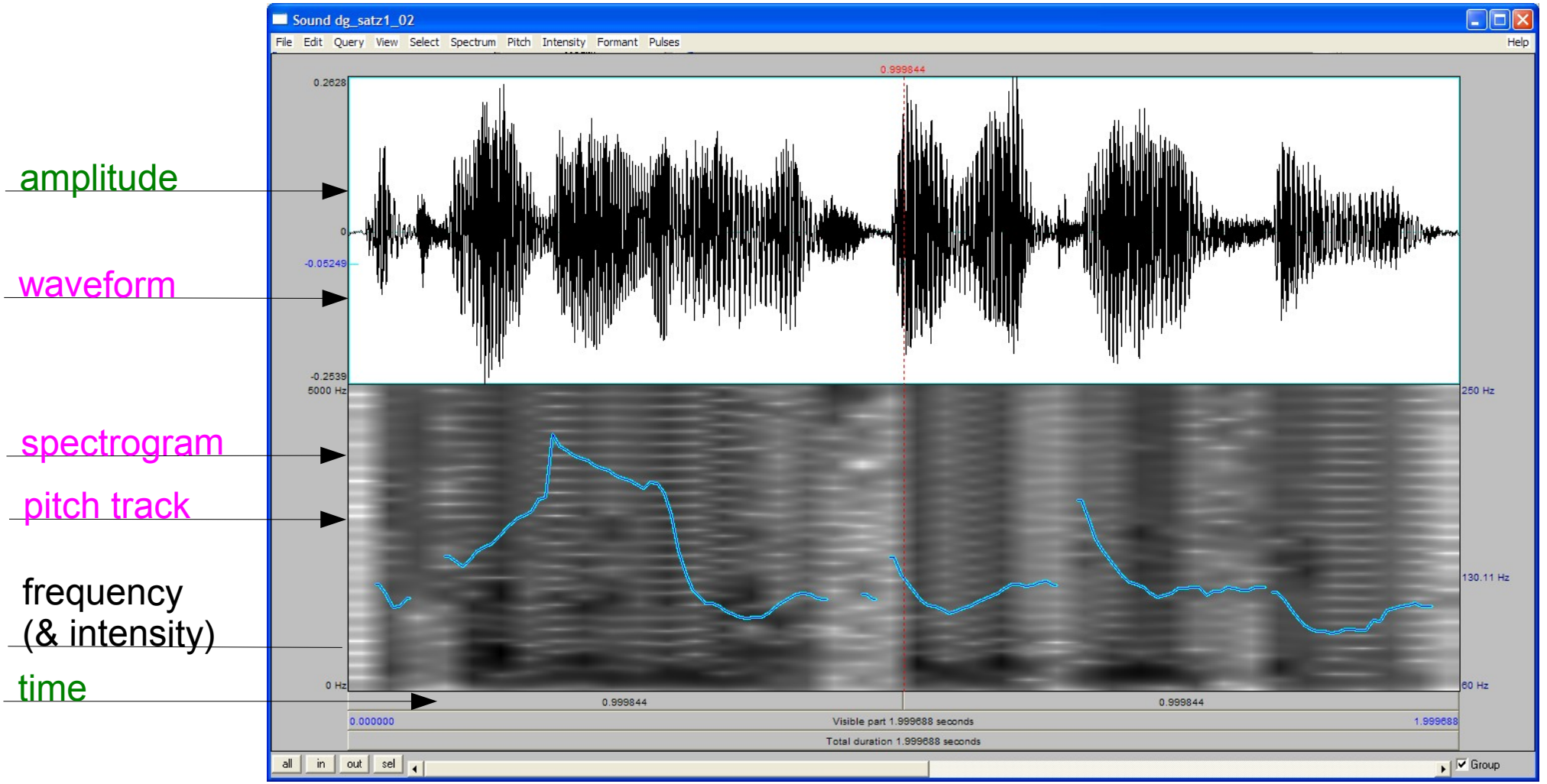
PRACTICAL WORK WITH MBROLA

Starting up with MBROLA

- Go to the MBROLA website
  - http://tcts.fpms.ac.be/synthesis/mbrola.html
- Download
  - the MBROLA binary file for your operating system
  - an MBROLA voice
- Install the MBROLA binary and the voice
  - follow the instructions
- Find a .pho file
  - double-click, which opens the MBROLI user interface
  - go ahead...

MAKING A VOICE: FIRST STEPS

Recorded speech data



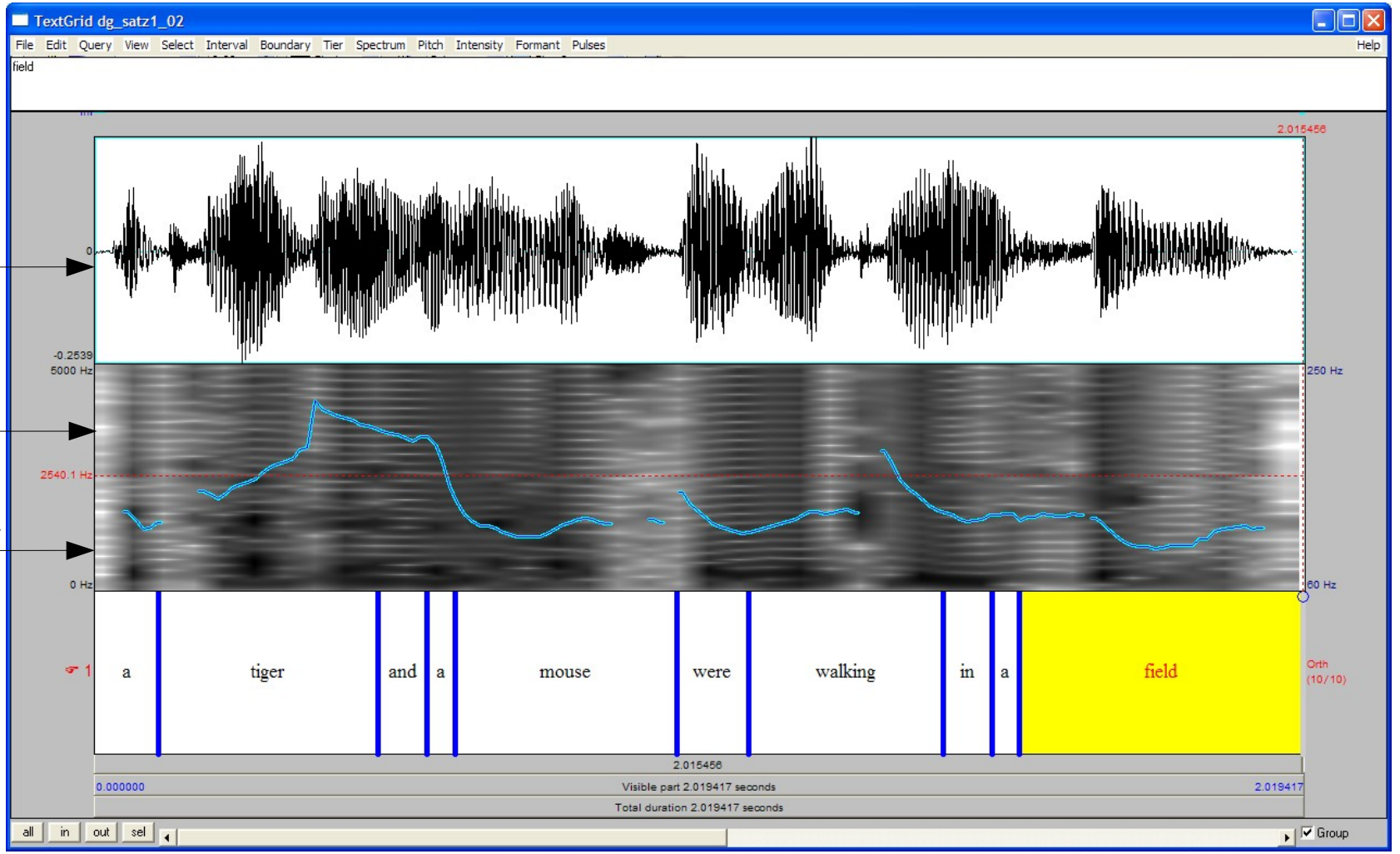
A tiger and a mouse were walking in a field...

Annotated speech data

oscillogramme

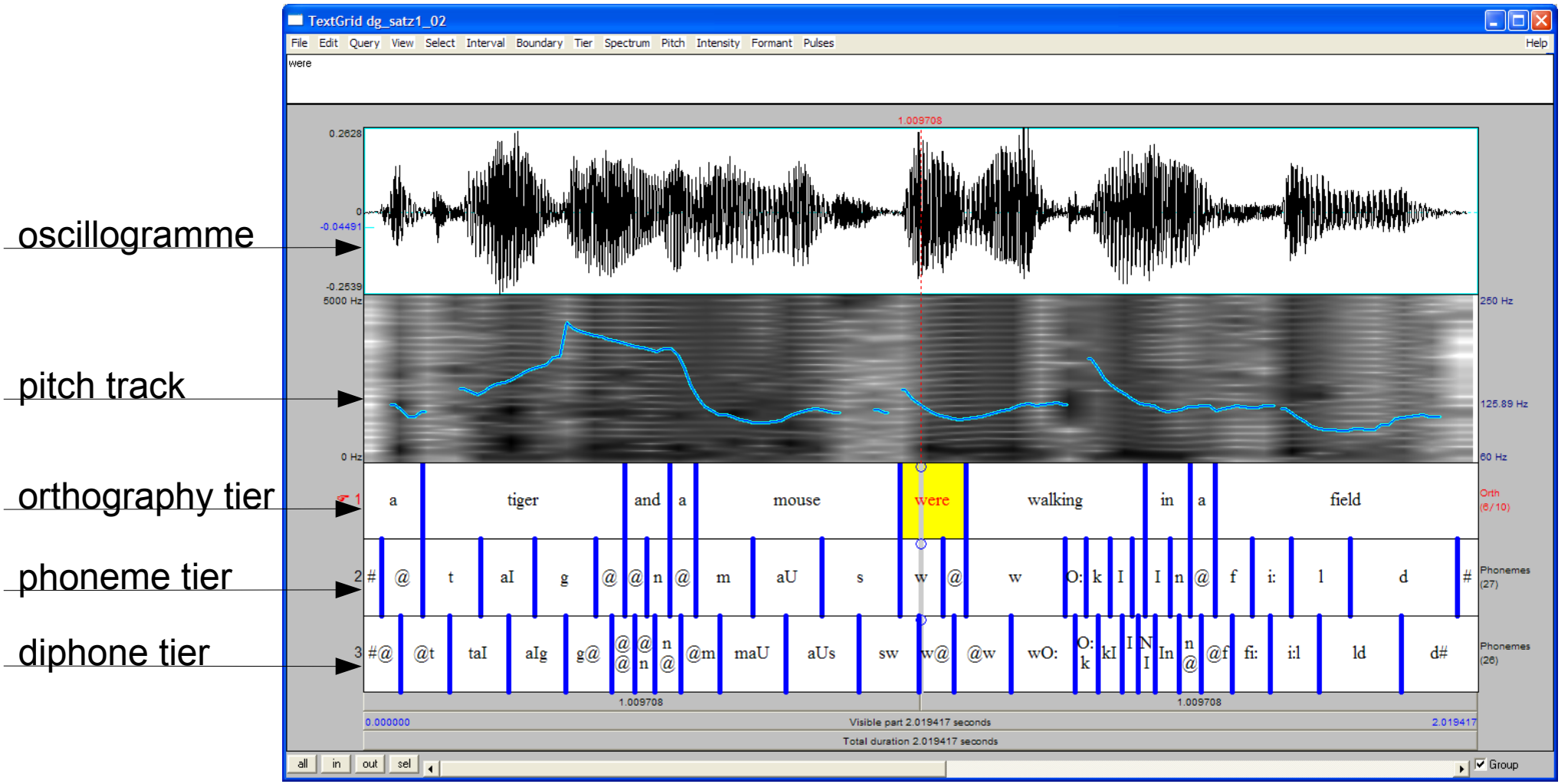
pitch track

orthography tier



A tiger and a mouse were walking in a field...

Creating diphones



A tiger and a mouse were walking in a field...

Creating the diphone database

- Instructions for creating voices are to be found on the MBROLA project page:
  - cut diphones out of the annotation file(s)
  - organise as a diphone database according to the instructions
  - submit the database to the MBROLA team
- The voice (normalised diphone database) will be returned to you.
- Test the voice in the manner illustrated in these slides.
- Create an NLP front end
  - CCS
  - or TTS, which is an entirely different story...

And then have your picture
put on
the MBROLA project site...