Phonetics: a short course

A brief introduction to Praat
Review of Phonetics
Praat Scripting
TGA: Analysis of Praat TextGrids

Dafydd Gibbon

Bielefeld-Abidjan Project: Language Documentation

November 2015

Phonetics 1

A Brief Introduction to Praat

Dafydd Gibbon

Bielefeld-Abidjan Project: Language Documentation

November 2015

Objectives

At the end of the course, participants should be able to

- extract duration and fundamental frequency information from a speech recording, using Praat;
- analyse speech timing with *Time Group Analyser*.

Your task:

- prepare a report on this course and send it to me for comments.

Part I: Annotation

- Basic knowledge of signal analysis with Praat.
- Basic knowledge of syllable annotation with Praat.
- Practical application to the recorded data of course participants.

Part II: Syllable timing analysis

- Examination information in Praat annotation file.
- Analysis of Praat annotation file with Time Group Analyser.

Praat

Praat is a phonetic workbench application developed in Amsterdam by Paul Boersma and David Weenink. The name means ‘talk’ in Dutch.

The basic functionality of Praat includes:

- Input: speech recordings
- Methods:
  - analysis of properties of speech signals such as spectral analysis, pitch analysis, annotation of signals with transcription labels
- Outputs:
  - files with information about the speech signal

The annotation information files which Praat produces

- can be re-formatted, and analysed with other means,
- e.g. with the online tool *Time Group Analyser*, for efficient analysis of timing relations in the speech signal.

Praat Input

Praat Input

Pre-recording phase:

- definition of purposes for which the data will be used
- scenario: domain, activities, speakers
- equipment and technical operator:
  - general: digital audio (recorder / laptop), digital video
  - specialised: laryngograph, etc

Recording phase:

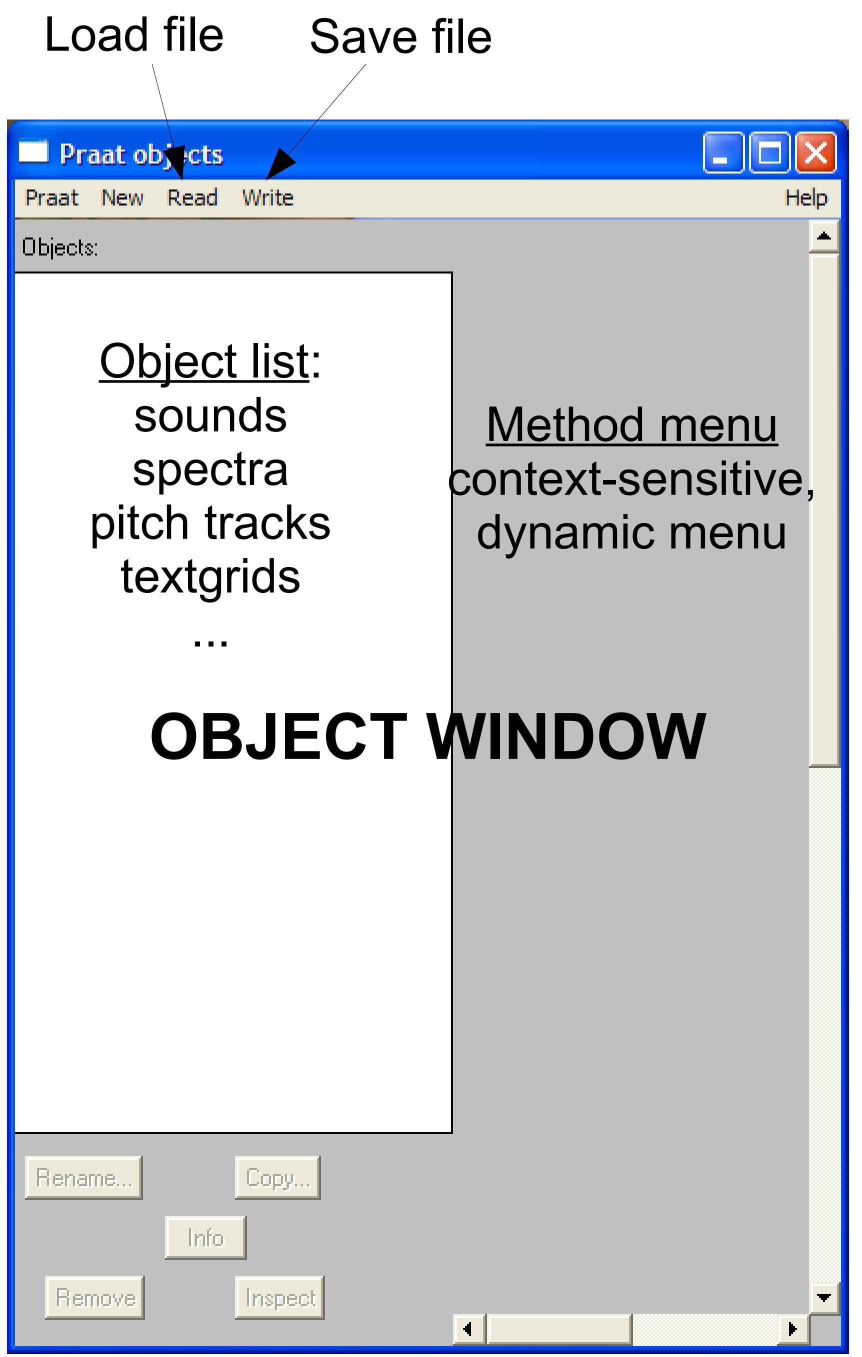
- negotiate scenario with chiefs, elders, speakers
- ensure the recording location is quiet
- if possible ensure the microphones, video tripod etc. can be stably positioned

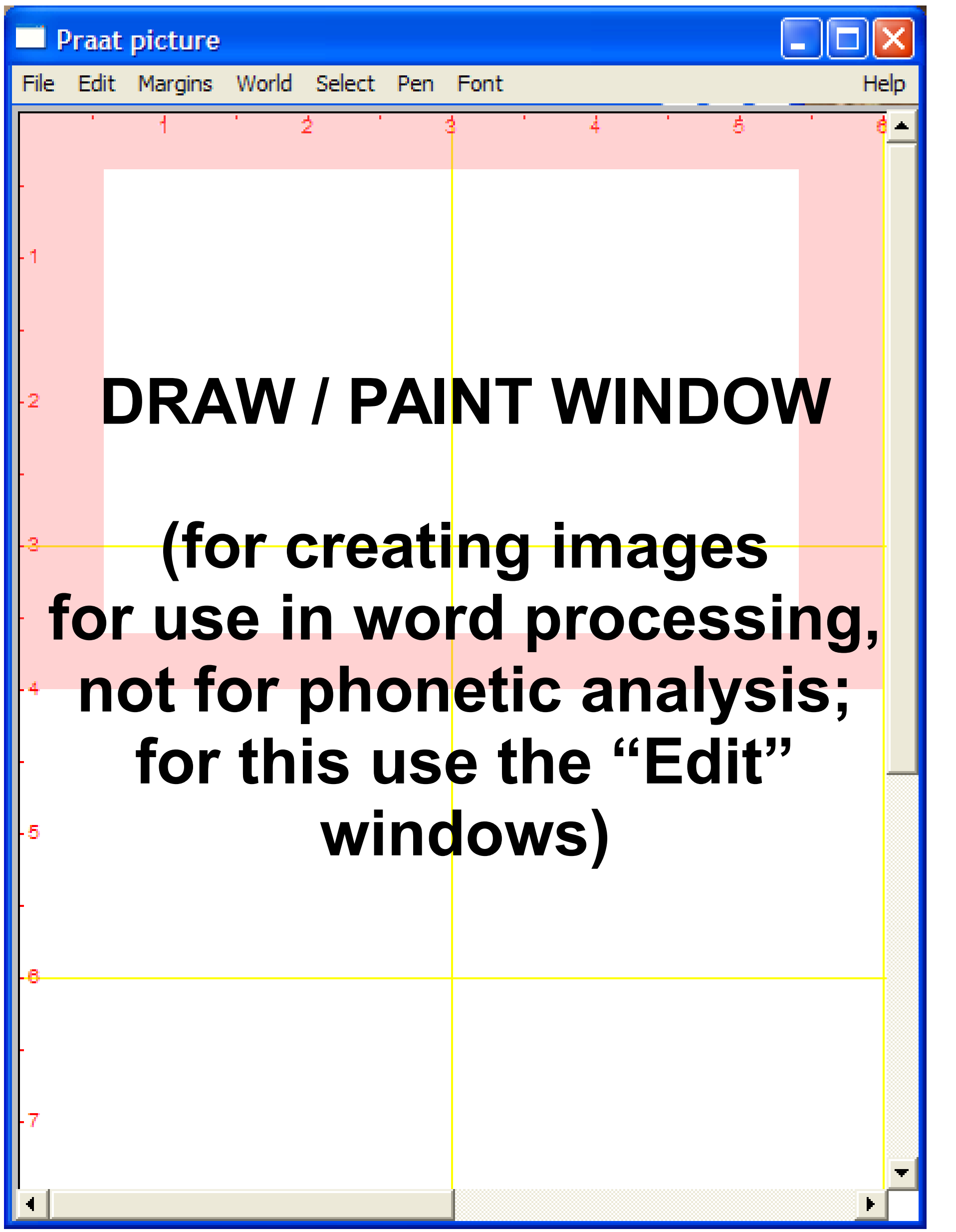
Post-recording phase:

- provide recordings with metadata immediately
- label the data media immediately
- make safety copies immediately

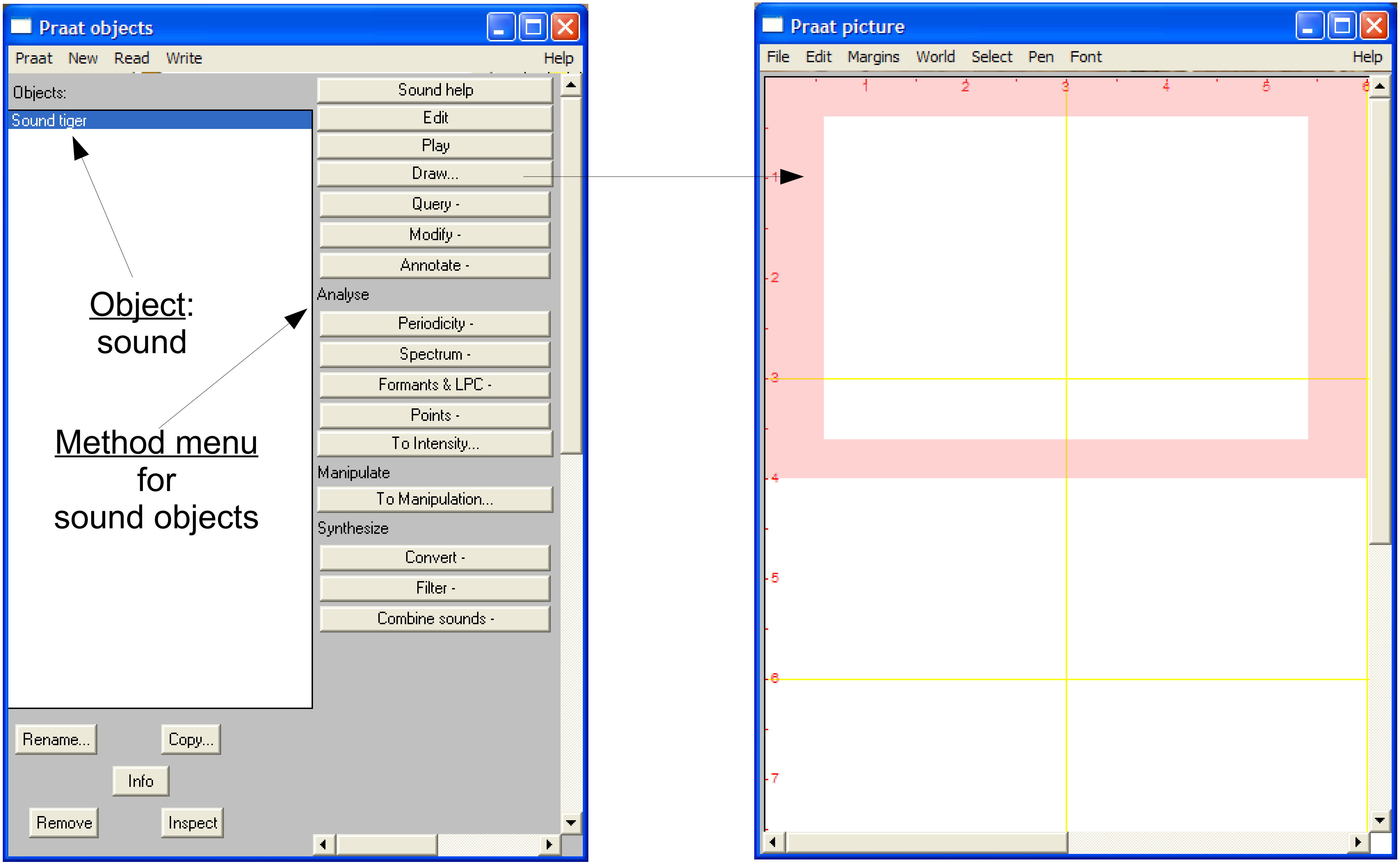
Basic Praat Methods and Operation

Praat Windows

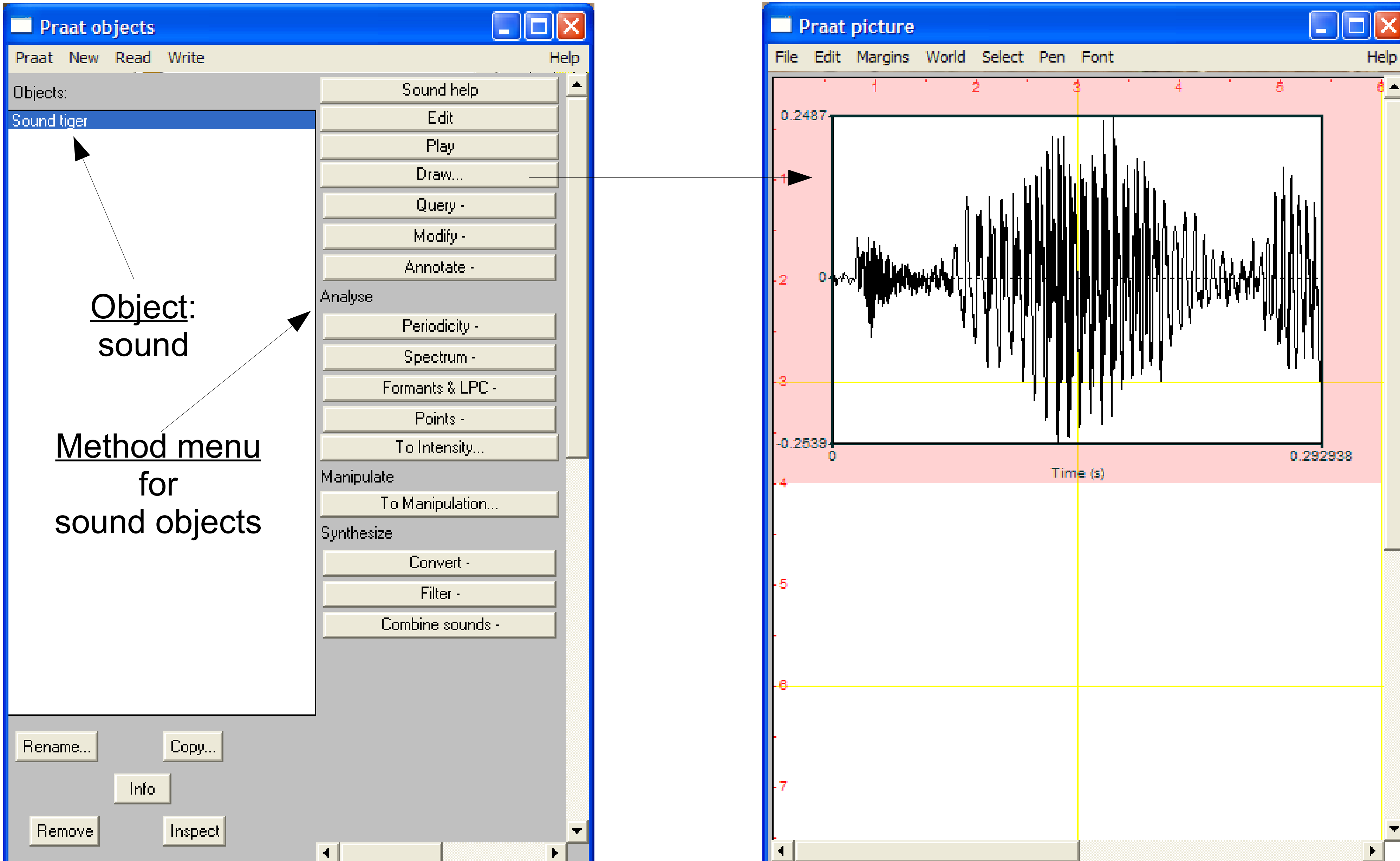




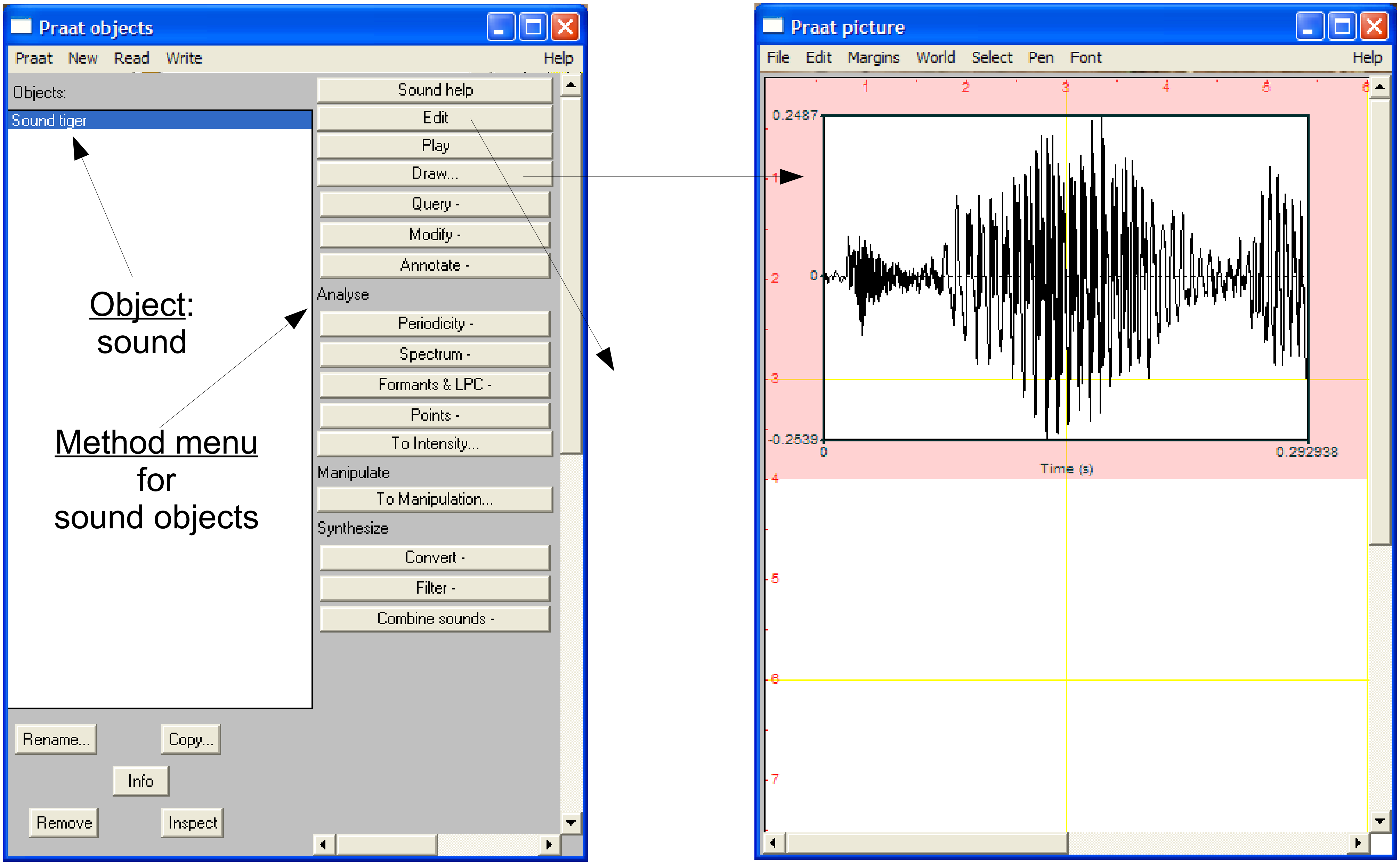
Processing Praat Objects



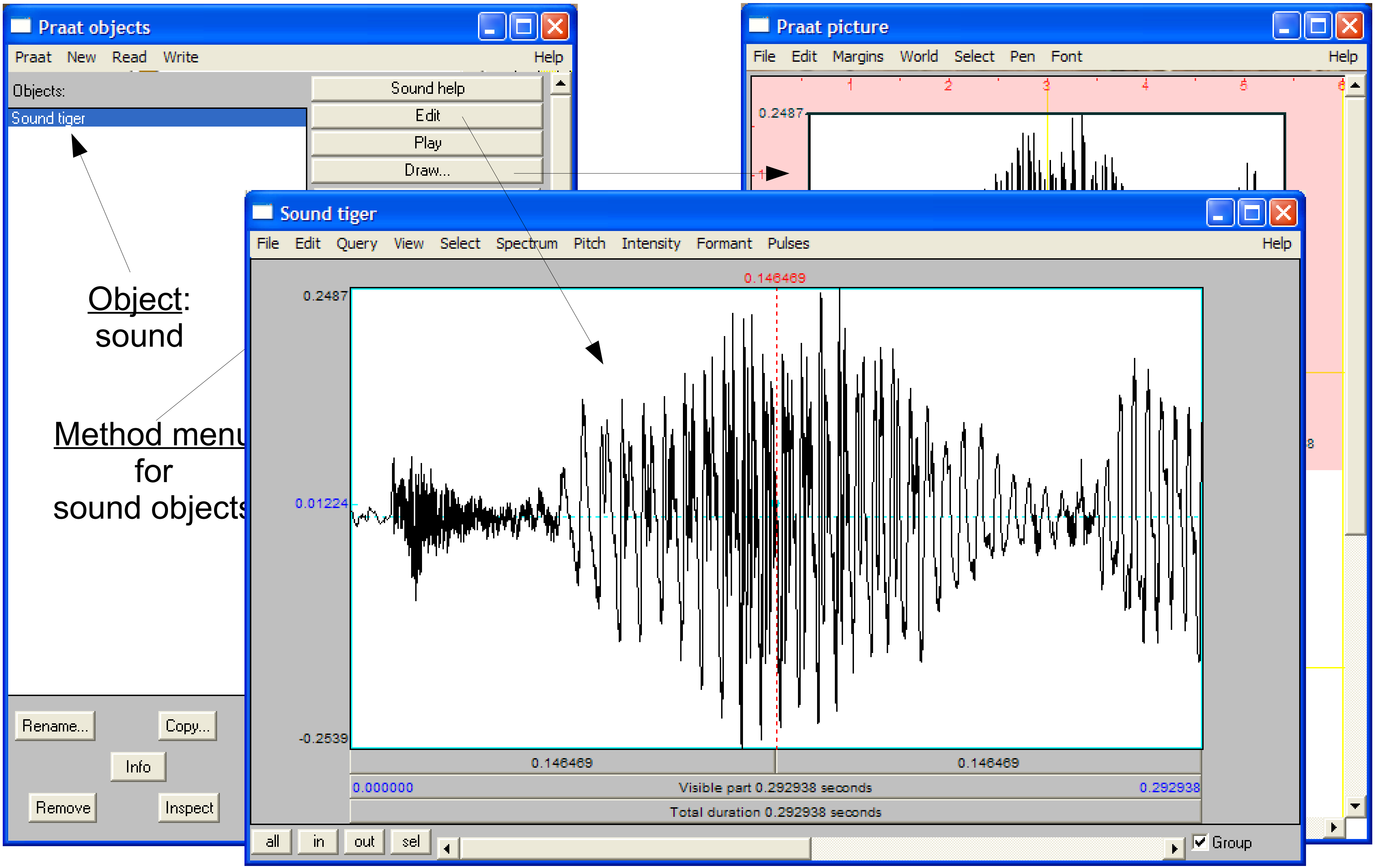
Processing Praat Objects



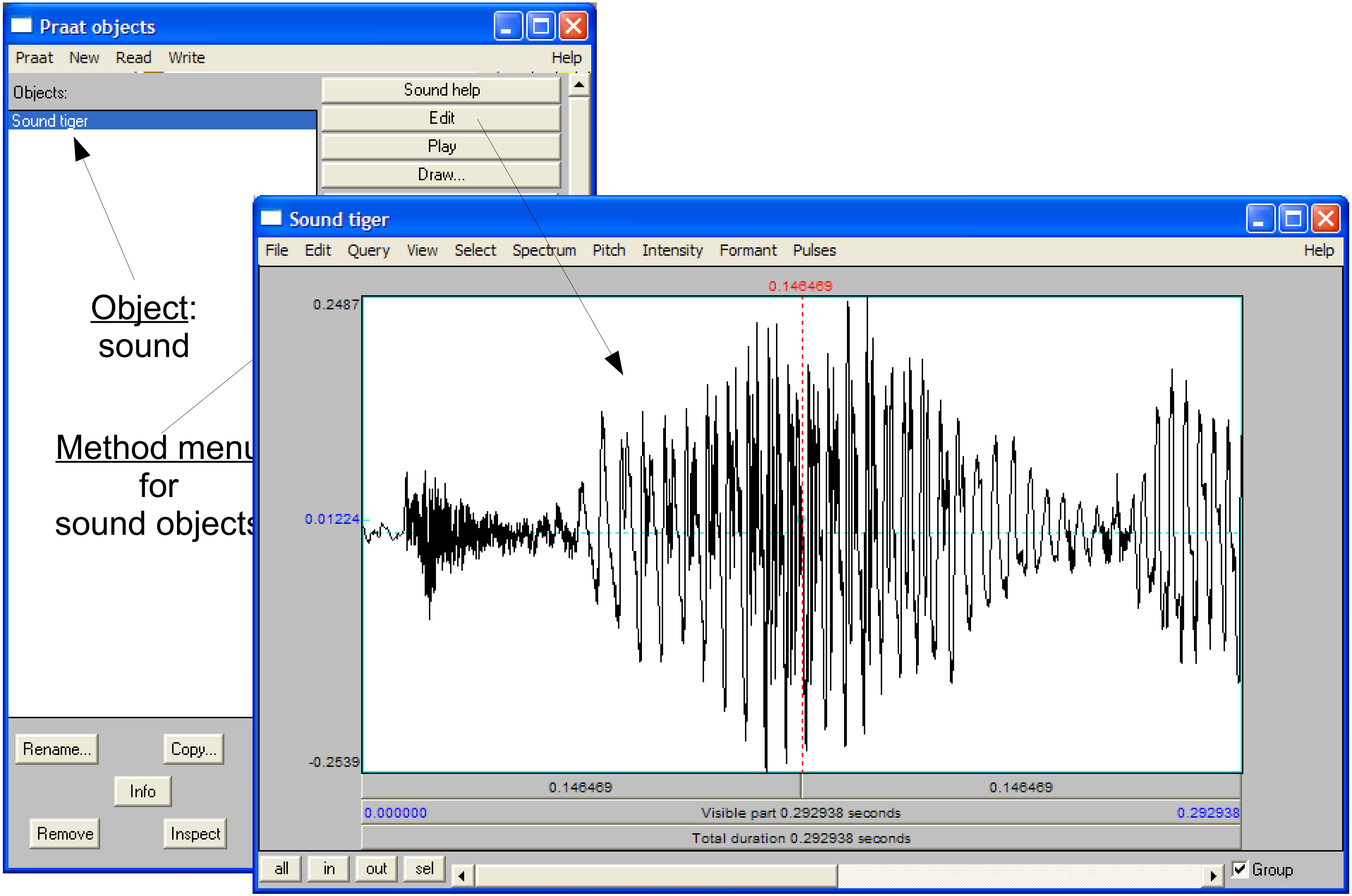
Processing Praat Objects



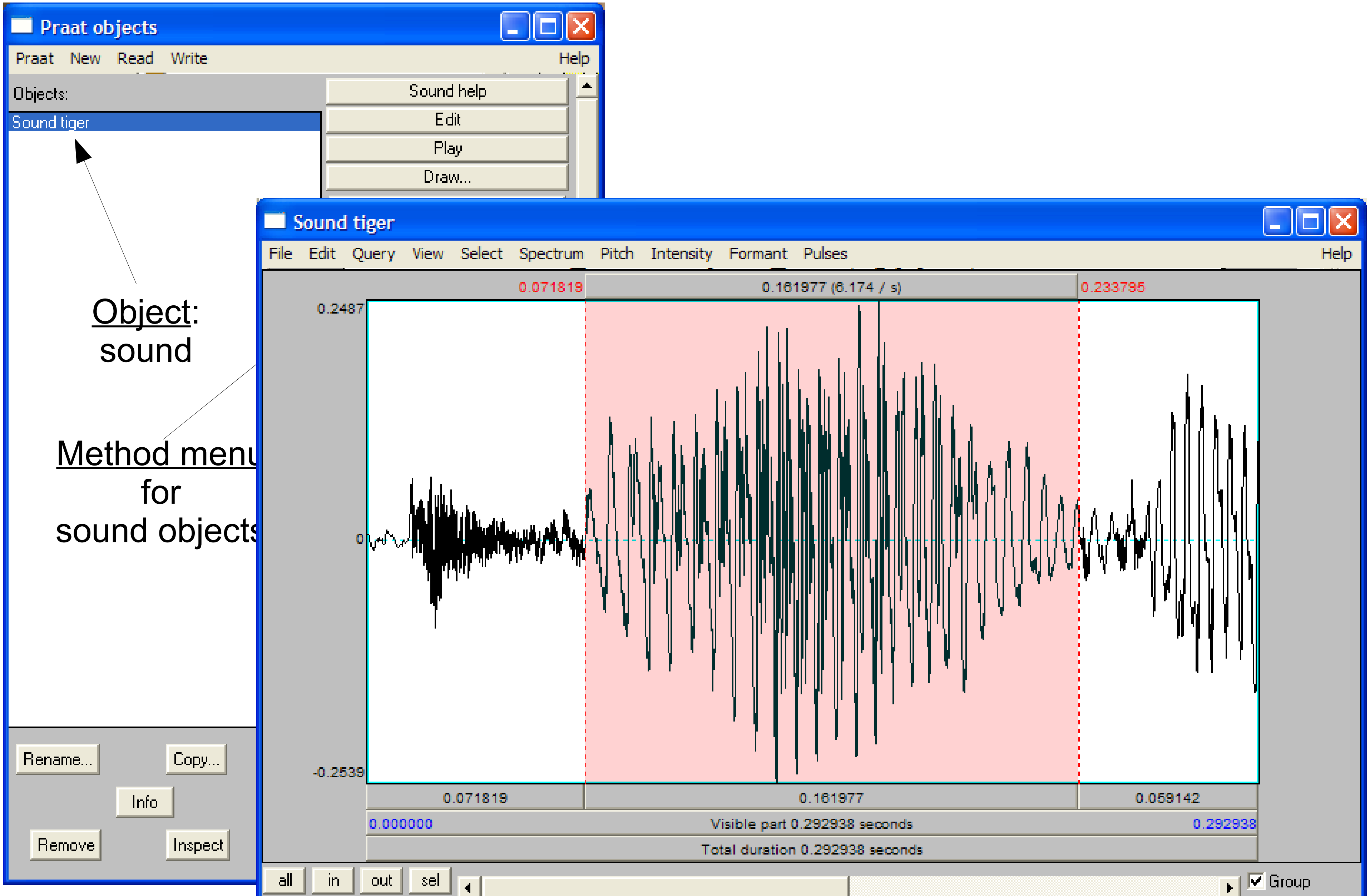
Processing Praat Objects



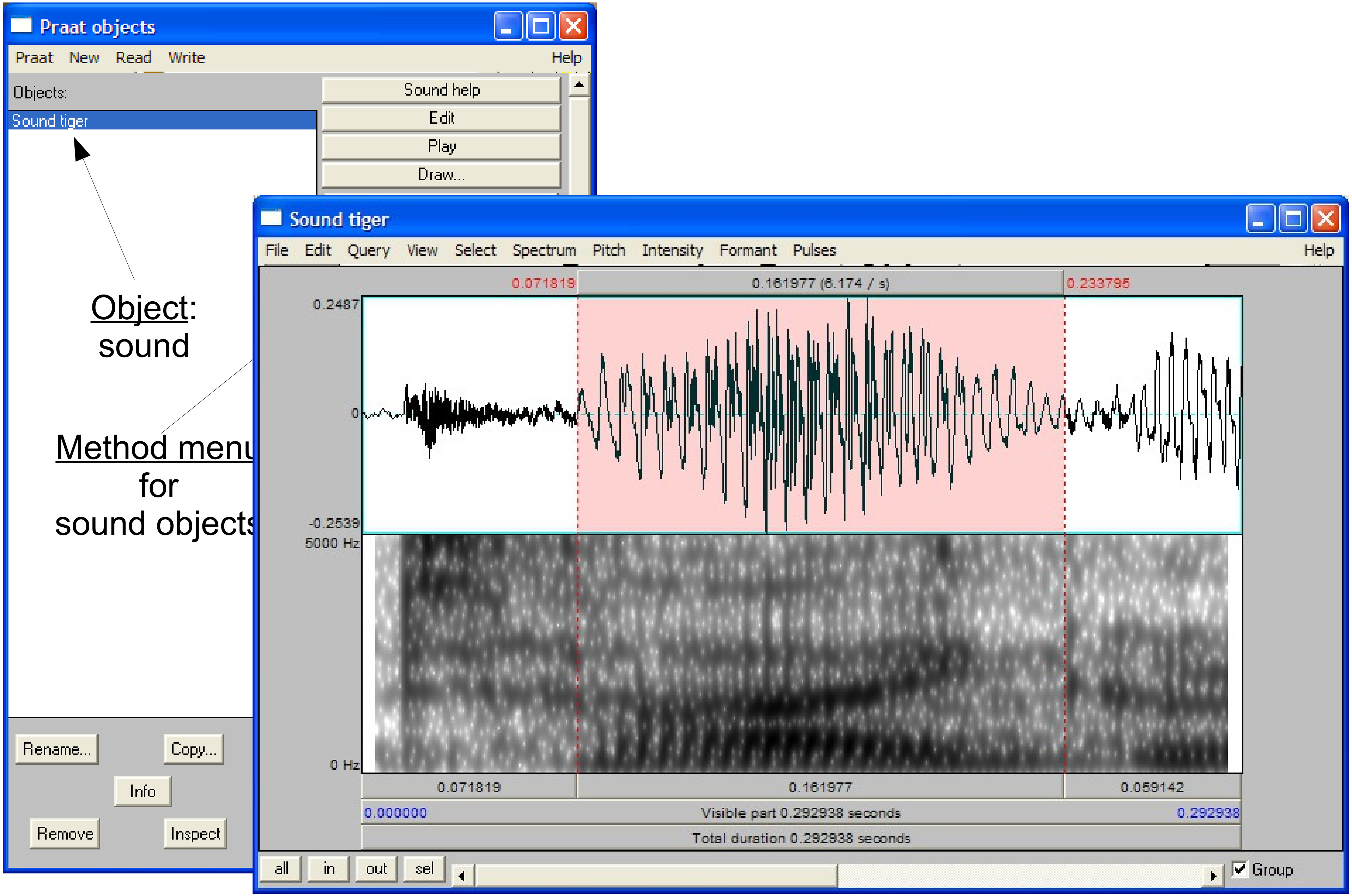
Processing a Sound Object



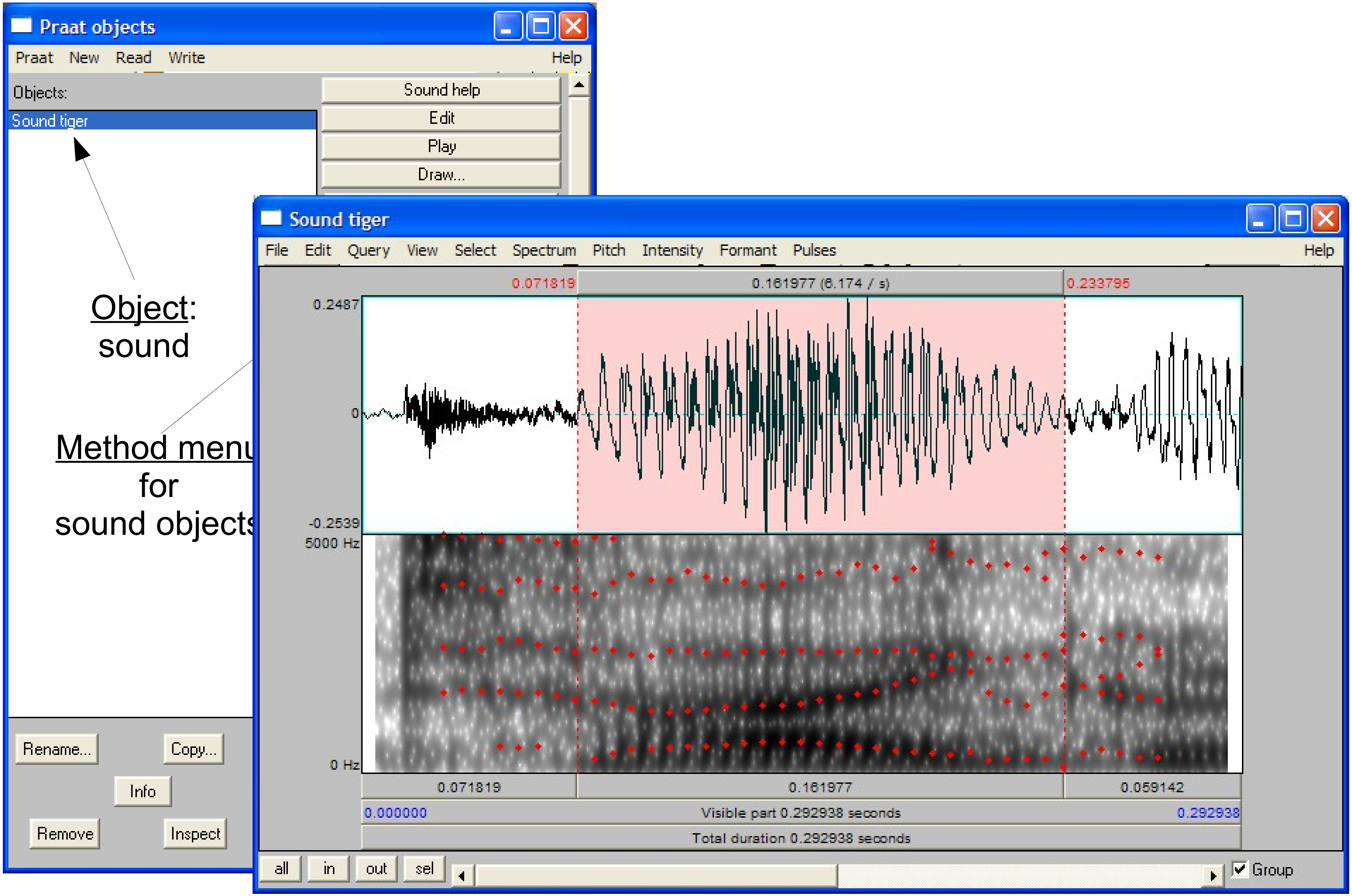
Selecting Part of a Sound Object



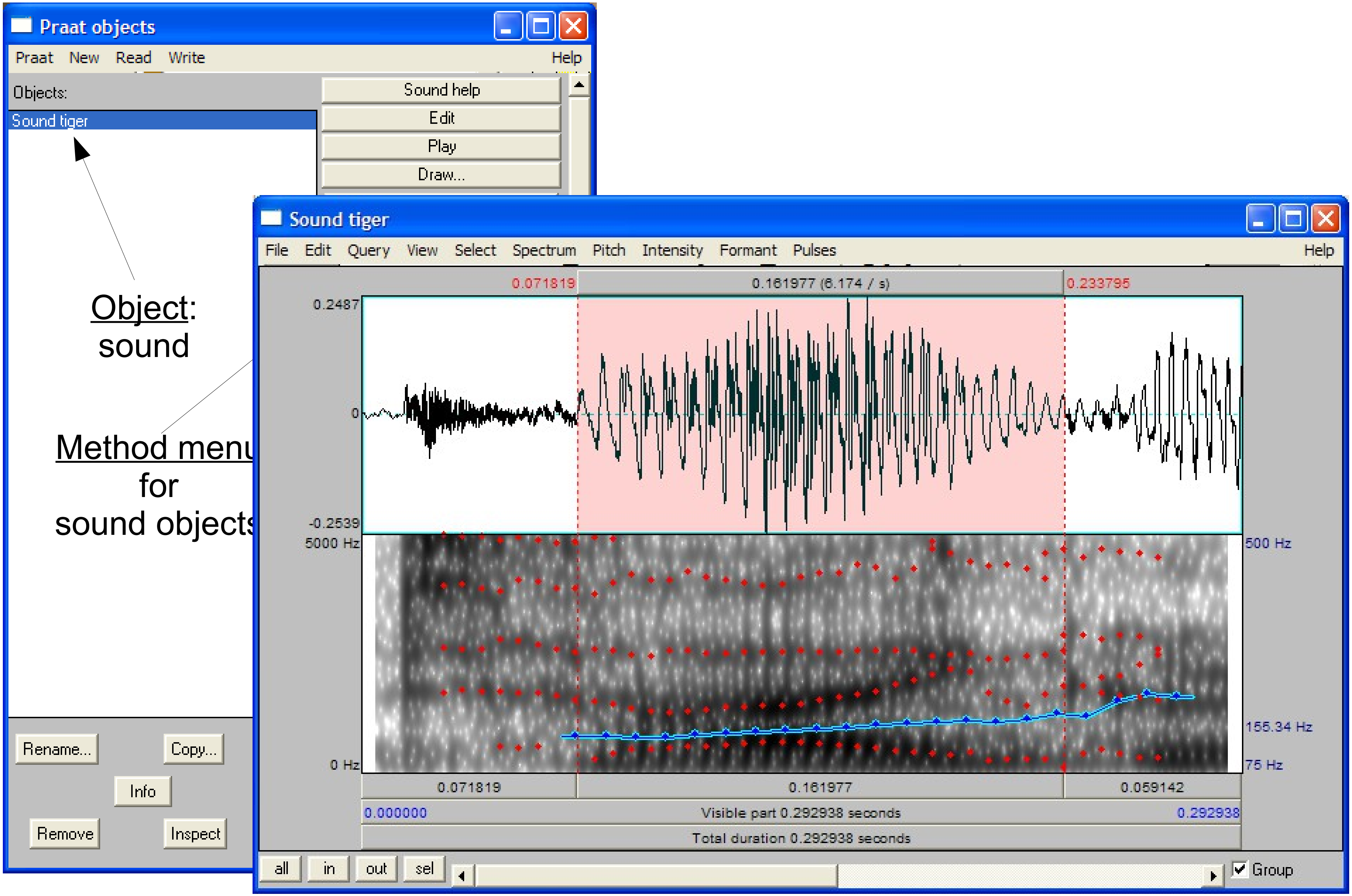
Displaying More Properties of a Sound Object



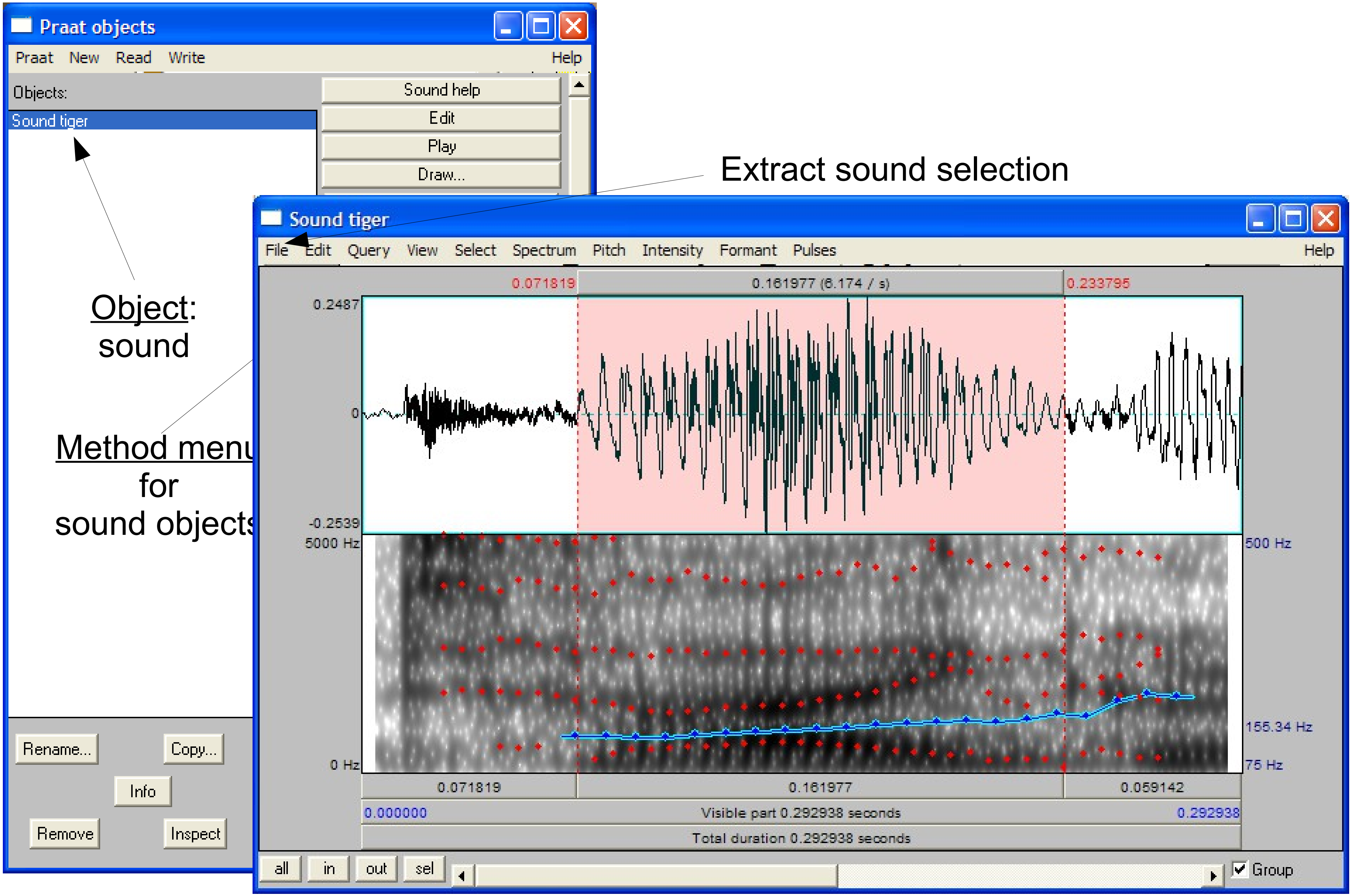
Displaying More Properties of a Sound Object



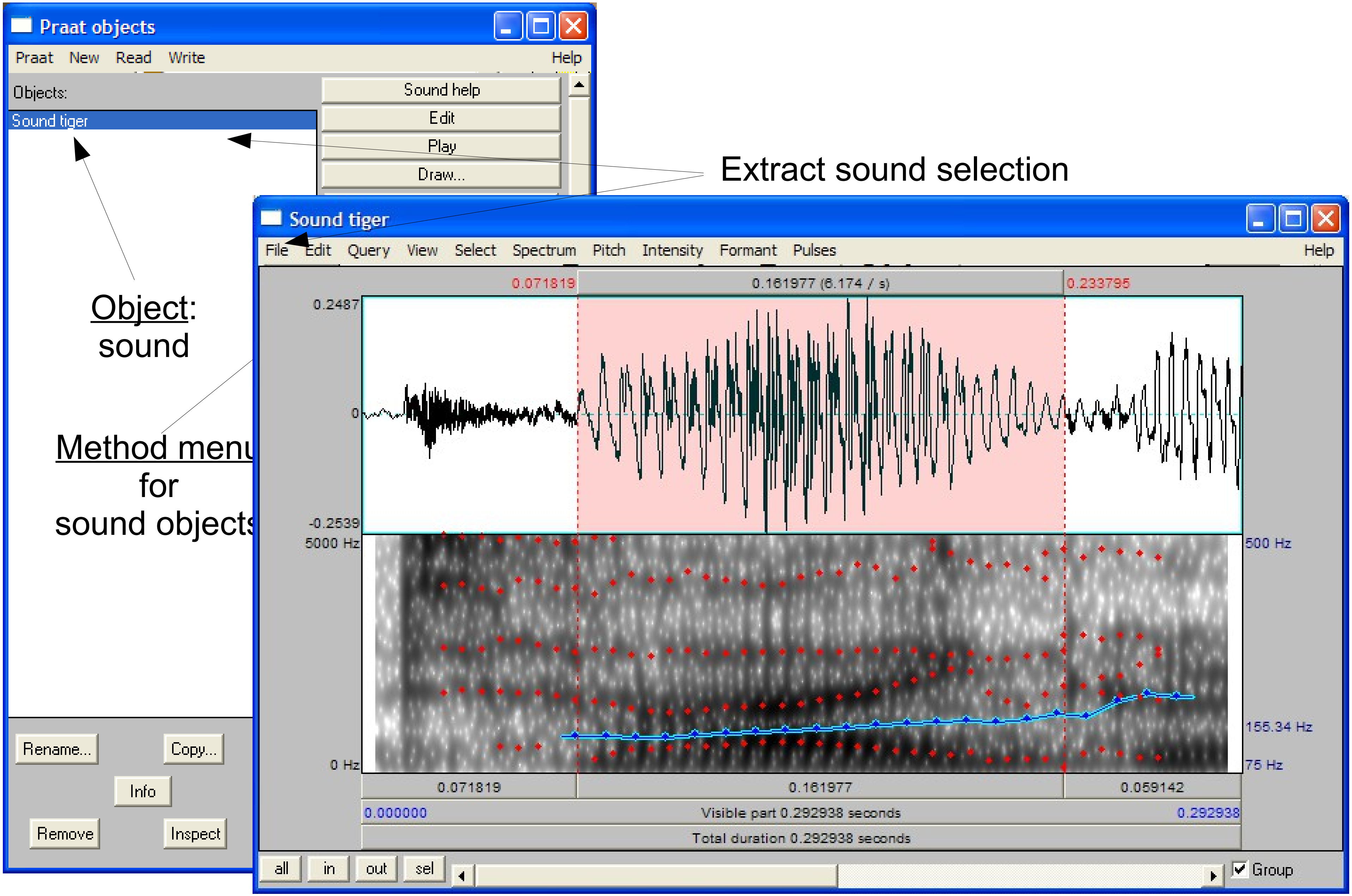
Displaying More Properties of a Sound Object



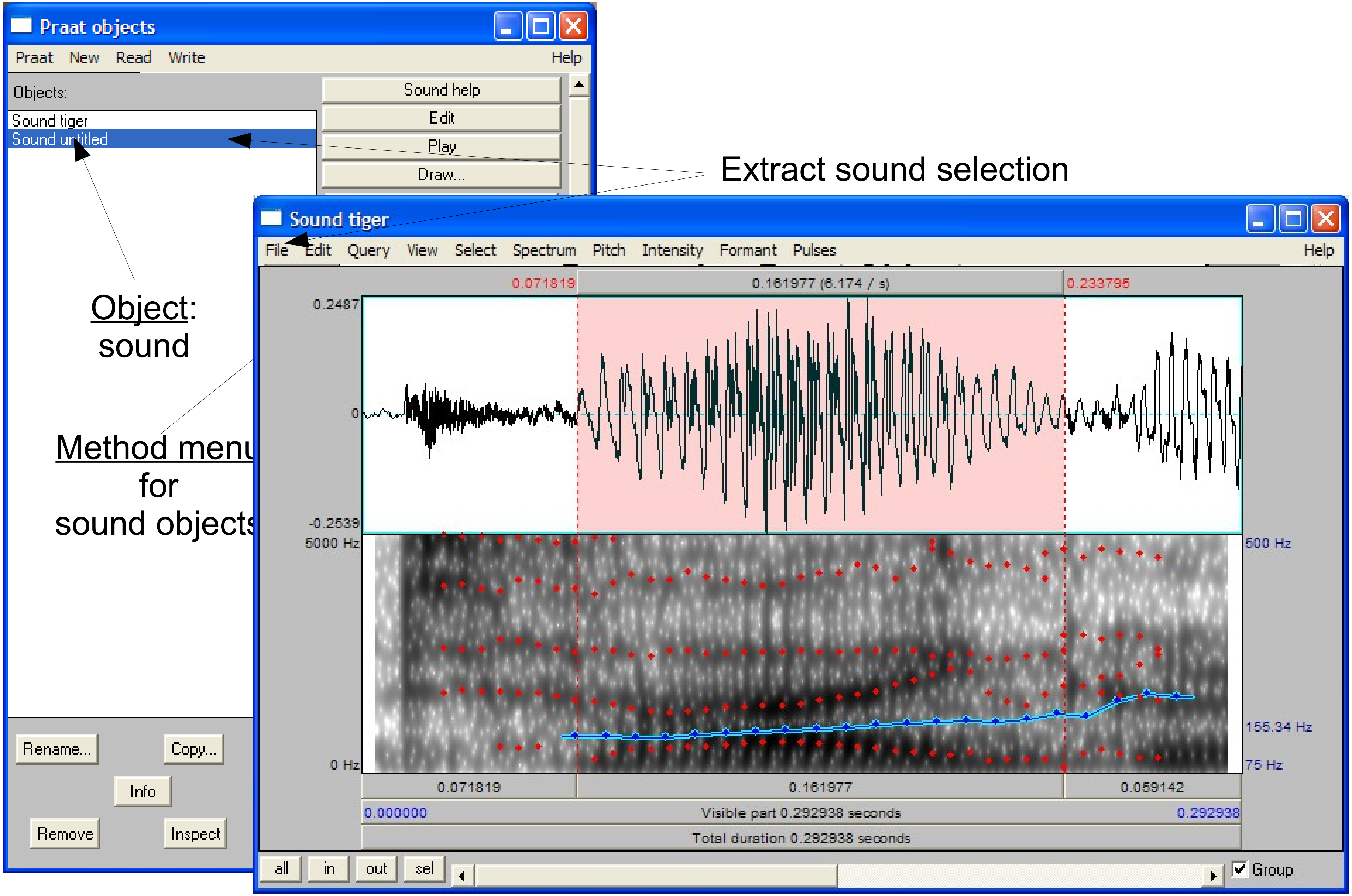
Creating a New Object



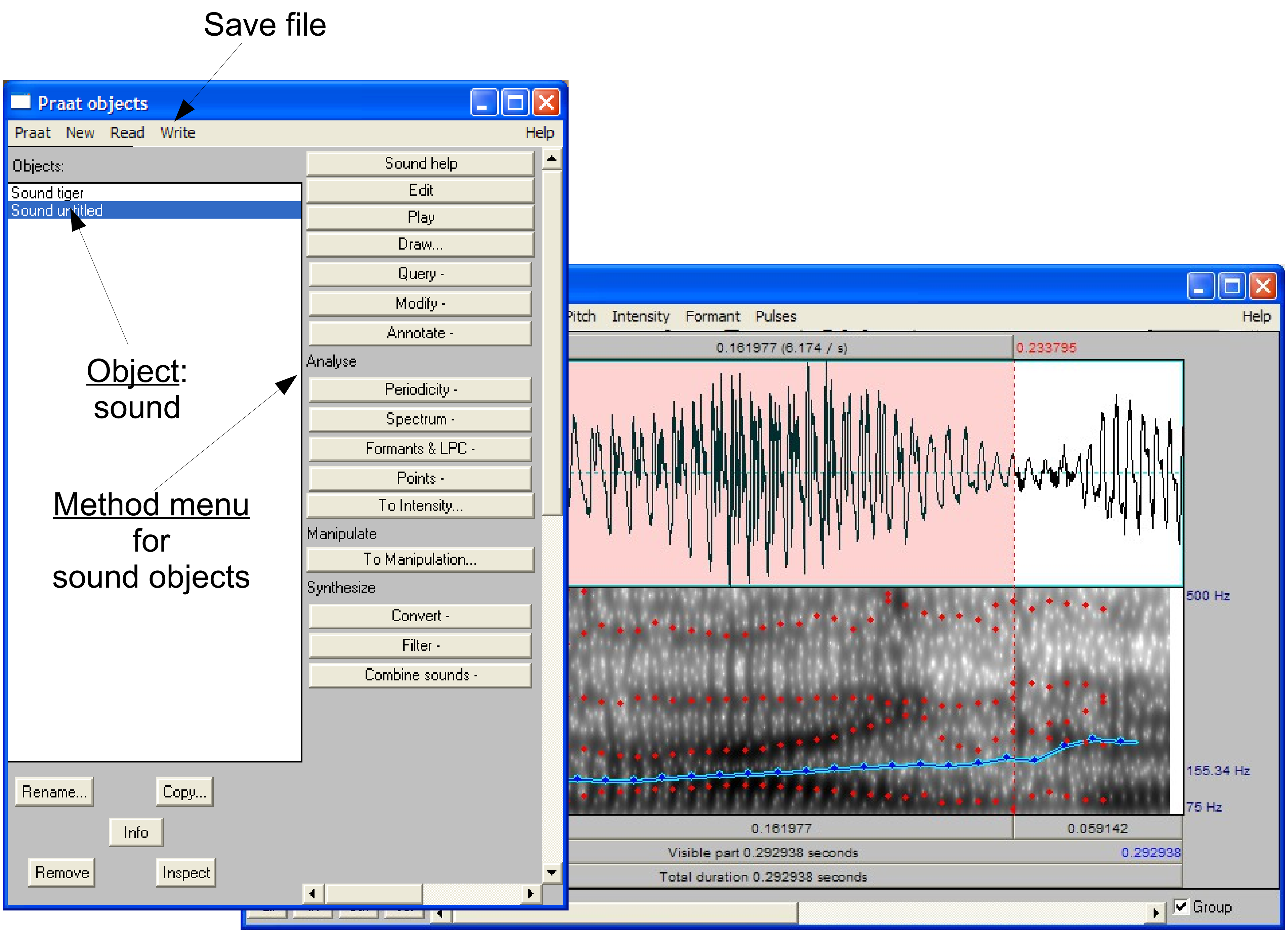
Creating a New Object



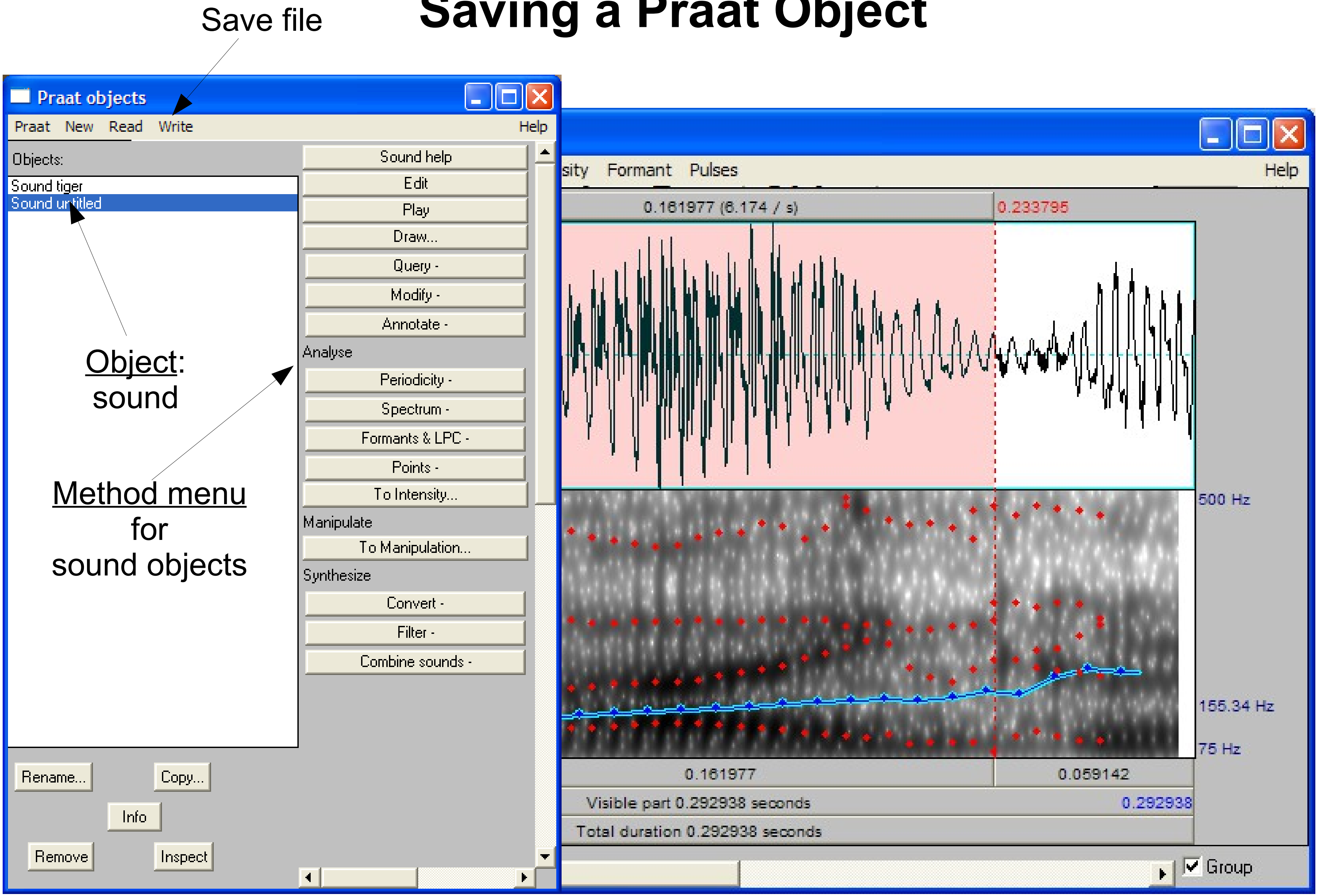
Creating a New Object



Saving a Praat Object



Saving a Praat Object



Extracting information about the speech signal

Operations:

- Read a speech file into Praat
- Open an “Edit” window
- Select sections of the signal and listen carefully for tones
- Experiment with the menu options to show
  - pitch
  - spectrum
  - …
- Experiment with pitch:
  - modify the analysis range for pitch

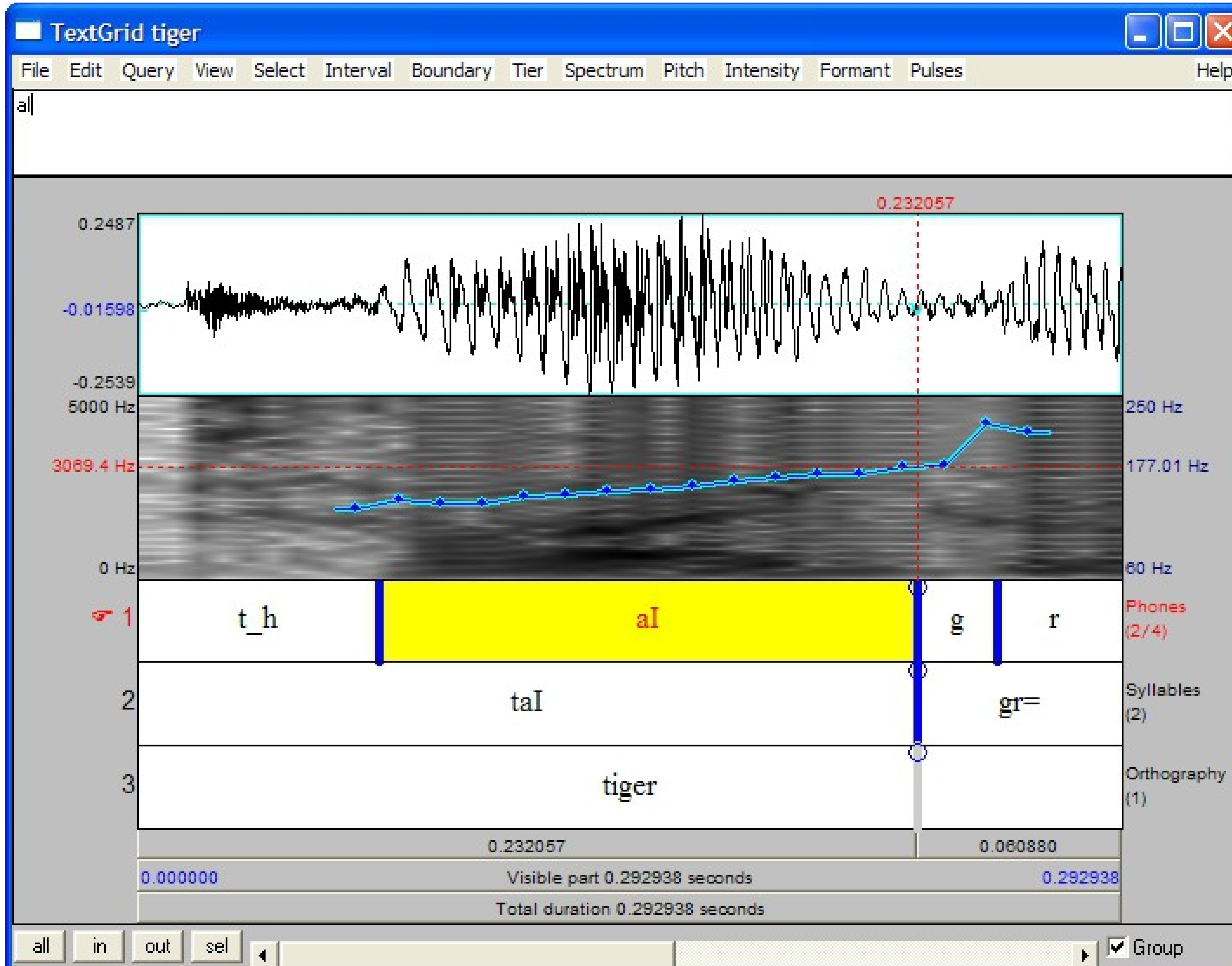
Assigning information to the speech signal: annotation

Operations:

- Read a speech file into Praat
- Open an annotation window (‘Annotate’)
- Delete the pre-defined tier names
- Enter a tier name ‘Word’

- An annotation window should appear
    - with a single tier ‘Word’
- Select a word in the speech signal
    - check the word by listening
- Small circles should appear on the annotation tier
    - at the beginning and the end of the selection
- Click each of these circles to create an event annotation
    - consisting of a label and an interval
- Save the file

Annotation with Praat



Annotation with Praat

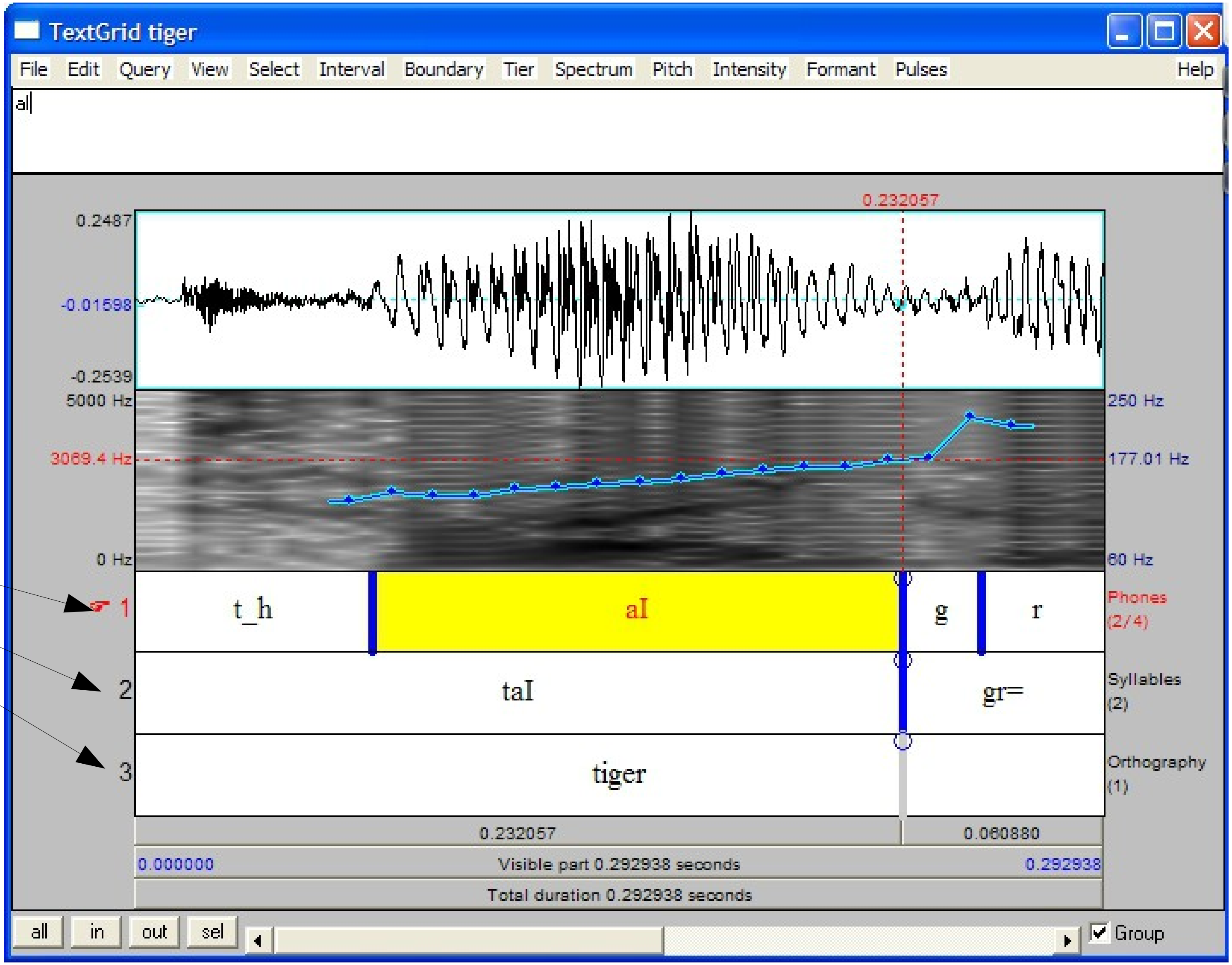
Annotation:

Tier 1: Phones
Tier 2: Syllables
Tier 3: Orthography

Event

Label

Interval



Annotation with Praat

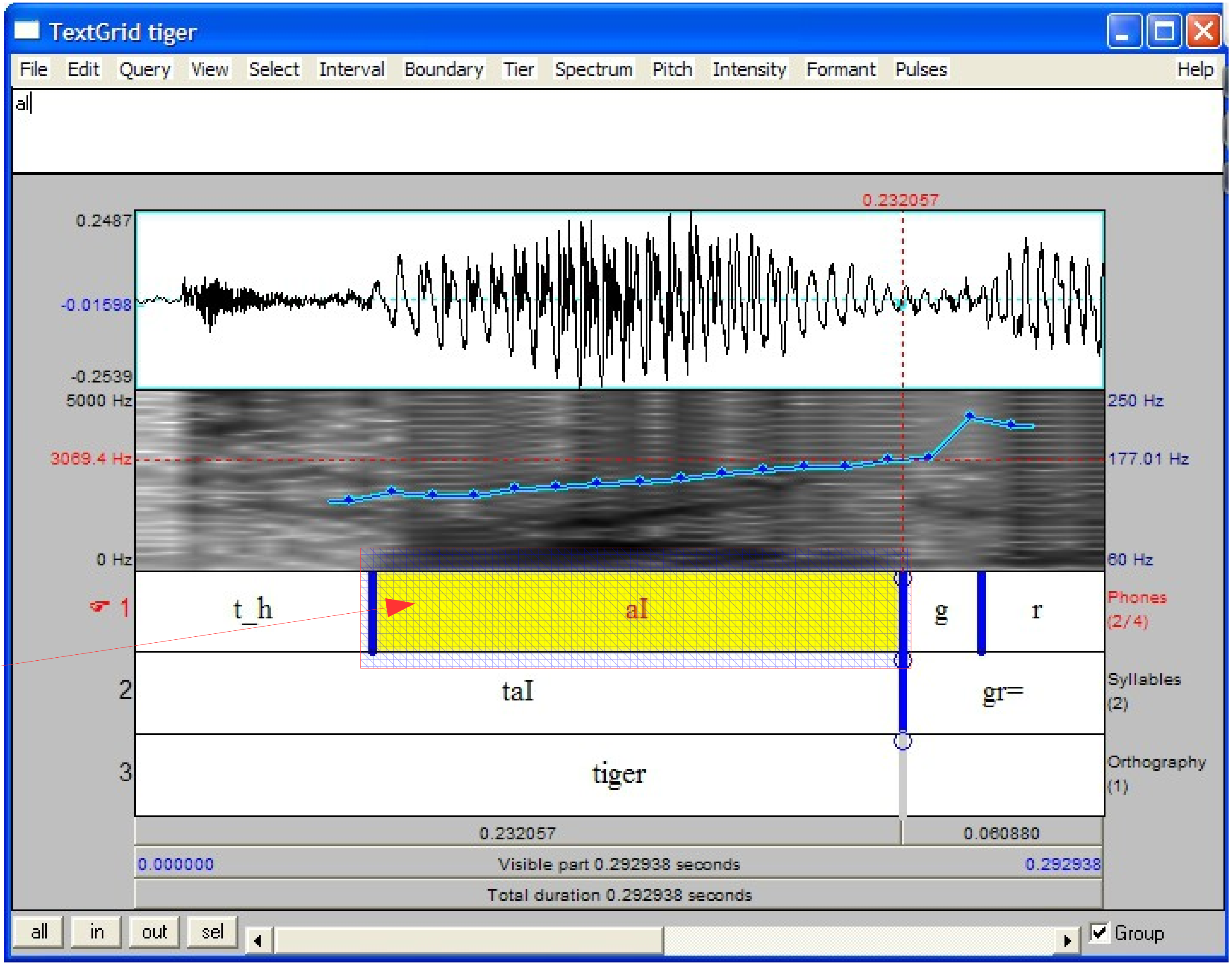
Annotation:

Tier 1: Phones
Tier 2: Syllables
Tier 3: Orthography

Event

Label

Interval



Annotation with Praat

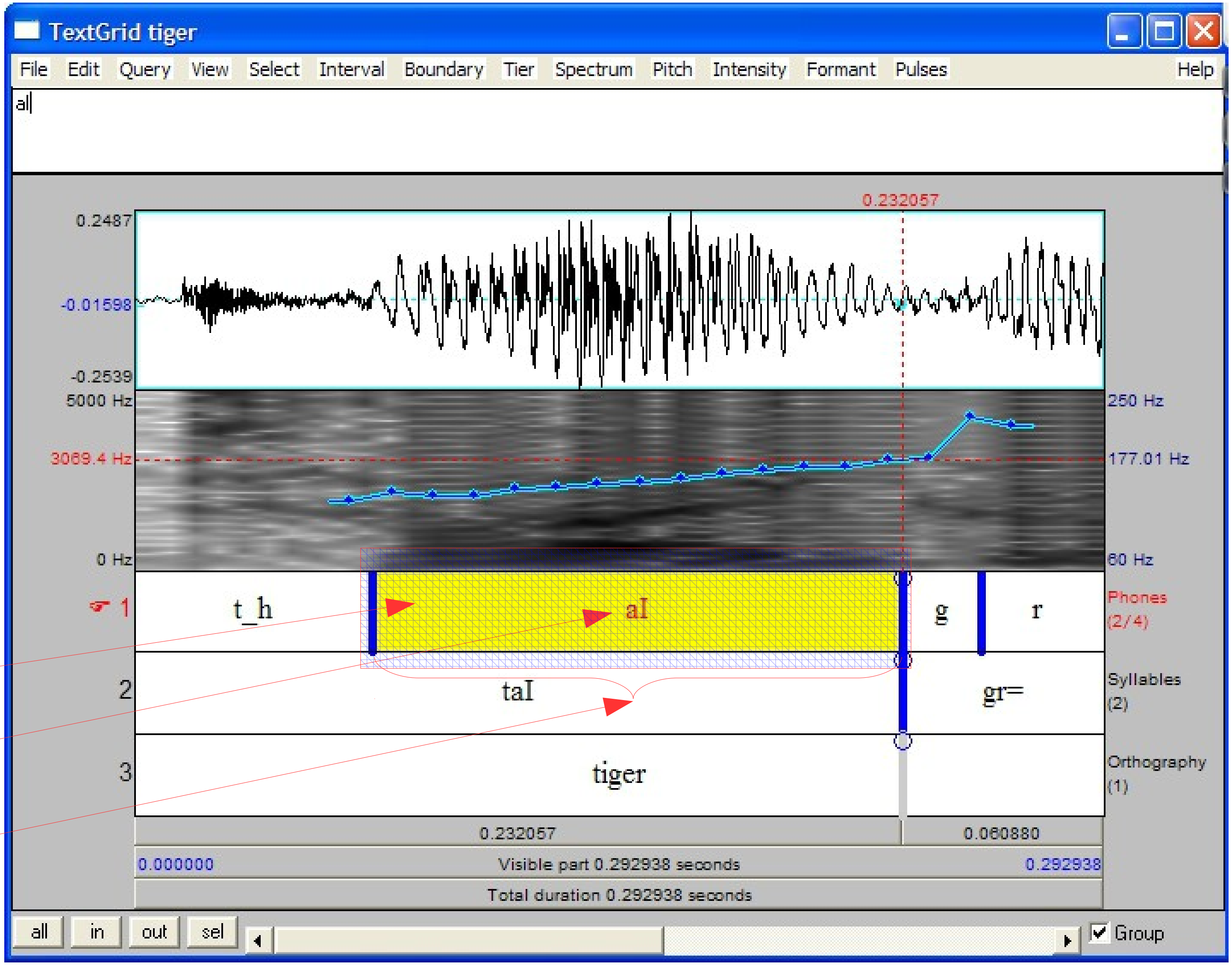
Annotation:

Tier 1: Phones
Tier 2: Syllables
Tier 3: Orthography

Event

Label

Interval



Annotaton with Praat: TextGrids

This is what the contents of your annotation file look like.

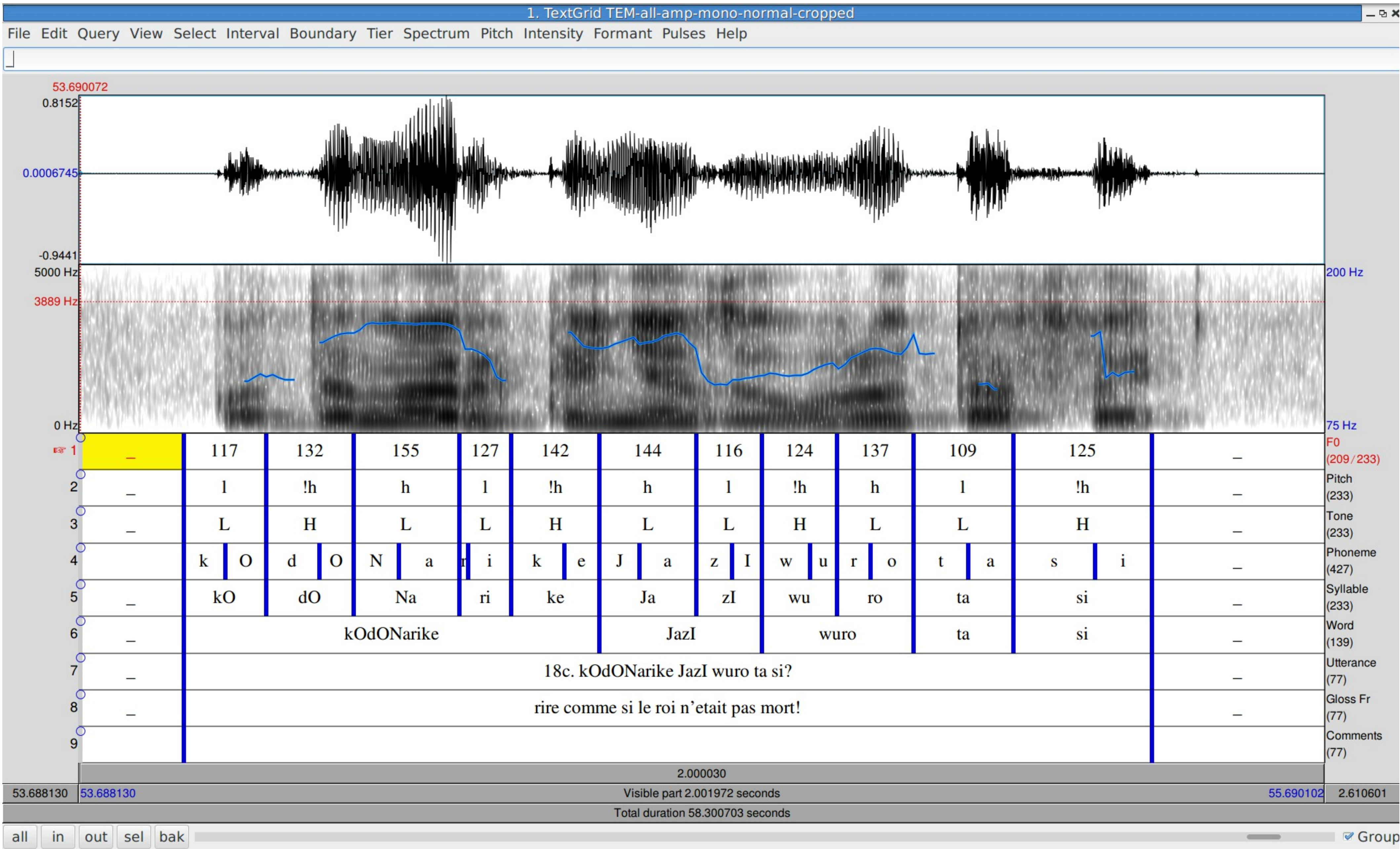
```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 0.2929375
tiers? <exists>
size = 1
item []:
    item [1]:
        class = "IntervalTier"
        name = "Phones"
        xmin = 0
        xmax = 0.2929375
        intervals: si       4
```

The last section of this slide set shows one way of analysing Praat speech annotations.

```
        intervals [1]:
            xmin = 0
            xmax = 0.0718
            text = "t_h"
        intervals [2]:
            xmin = 0.0718
            xmax = 0.2323
            text = "aI"
        intervals [3]:
            xmin = 0.2323
            xmax = 0.25579
            text = "g"
        intervals [4]:
            xmin = 0.2557
            xmax = 0.2929
            text = "r\`="
```

Annotating a speech signal



Praat Output

Praat Output

- Praat produces information which can be stored in a file.
- The file contents are not normally seen by the user.
- However, they can be seen and used for further analysis:
  - Excel, OpenOffice Calc
  - Praat scripting
  - Shell, Perl, Python scripting
  - Time Group Analyser online tool

```
intervals [1]:
    xmin = 0
    xmax = 0.3559744193778952
    text = "_"
intervals [2]:
    xmin = 0.3559744193778952
    xmax = 0.500147057910385
    text = "ta"
intervals [3]:
    xmin = 0.500147057910385
    xmax = 0.614452757446077
    text = "la"
intervals [4]:
    xmin = 0.614452757446077
    xmax = 0.8853950267508599
    text = "sin"
intervals [5]:
    xmin = 0.8853950267508599
    xmax = 1.096059981756913
    text = "Ge"
intervals [6]:
    xmin = 1.096059981756913
    xmax = 1.5079951315848832
    text = "_"
```

Exercises

- Load a short speech signal (less than 5 seconds) into Praat.
- Listen to individual words and syllables.
- Experiment with switching the different displays (spectrogram, pitch track).
- Examine the highest, lowest and average frequencies in the signal.
- Annotate the syllables in the signal with a ‘TextGrid’ annotation. annotation (a ‘TextGrid’) of syllables, and save the file.
- Using a text editor, examine the file which has been saved and calculate the durations of some of the intervals in the file.

Phonetics 2

Review of Acoustic Phonetics

Dafydd Gibbon

Bielefeld, November 2015

Overview: Acoustic Phonetics

- The Domains of Phonetics: the Phonetic Cycle
- The Articulatory Domain
  - The IPA (A = Alphabet / Association)
  - The Source-Filter Model of Speech Production
- The Acoustic Domain
  - The Speech Wave-Form
  - Basic Speech Signal Parameters
  - The Time Domain: the Speech Wave-Form
  - The Frequency Domain: simple & complex signals
    - Fourier Analysis: the Spectrum
    - Pitch extraction
  - Analog-to-Digital (A/D) Conversion
- The Auditory Domain: Anatomy of the Ear

The Domains of Phonetics

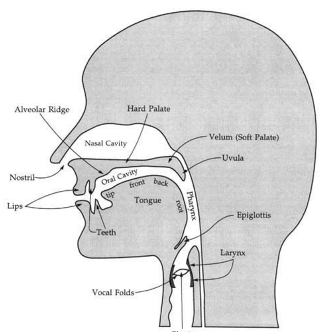
- Phonetics is the scientific discipline which deals with
  - speech production (articulatory phonetics)
  - speech transmission (acoustic phonetics)
  - speech perception (auditory phonetics)
- The scientific methods used in phonetics are
  - direct observation (“impressionistic”), usually based on articulatory phonetic criteria
  - measurement
    - of position and movement of articulatory organs
    - of the structure of speech signals
    - of the mechanisms of the ear and perception in hearing
  - statistical evaluation of direct observation and measurements
  - creation of formal models of production, transmission and perception

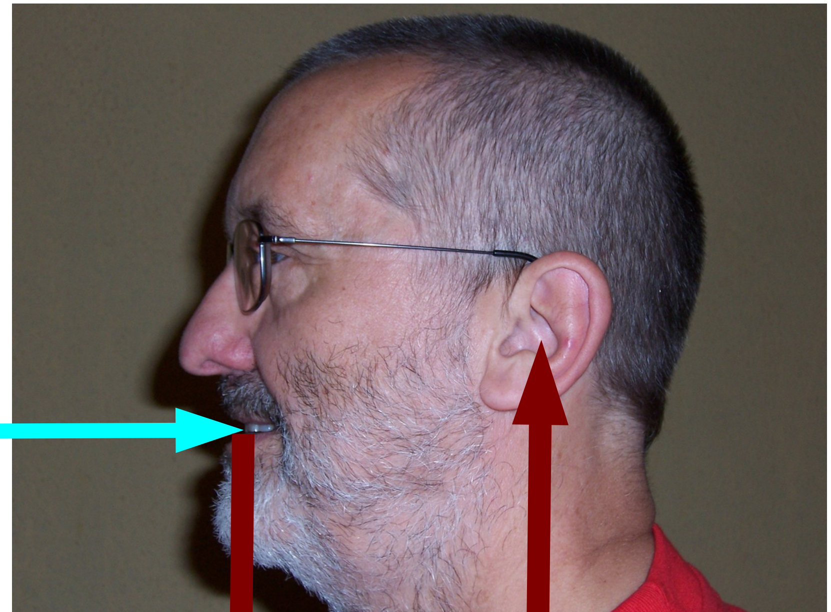
The Domains of Phonetics: the Phonetic Cycle

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

The Domains of Phonetics: the Phonetic Cycle

Articulatory Phonetics

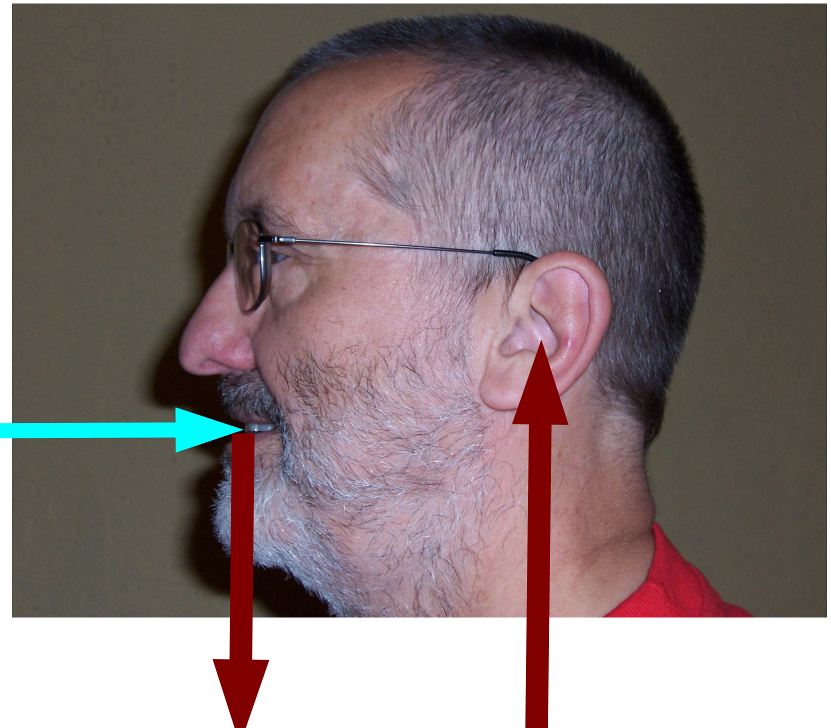


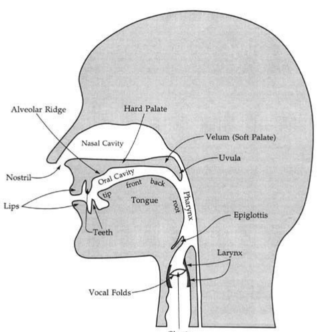


A tiger and a mouse were walking in a field...

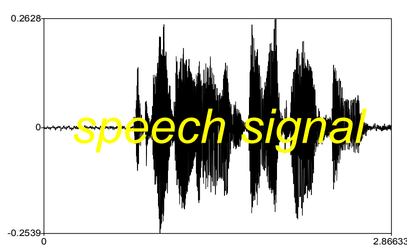
The Domains of Phonetics: the Phonetic Cycle

Articulatory Phonetics





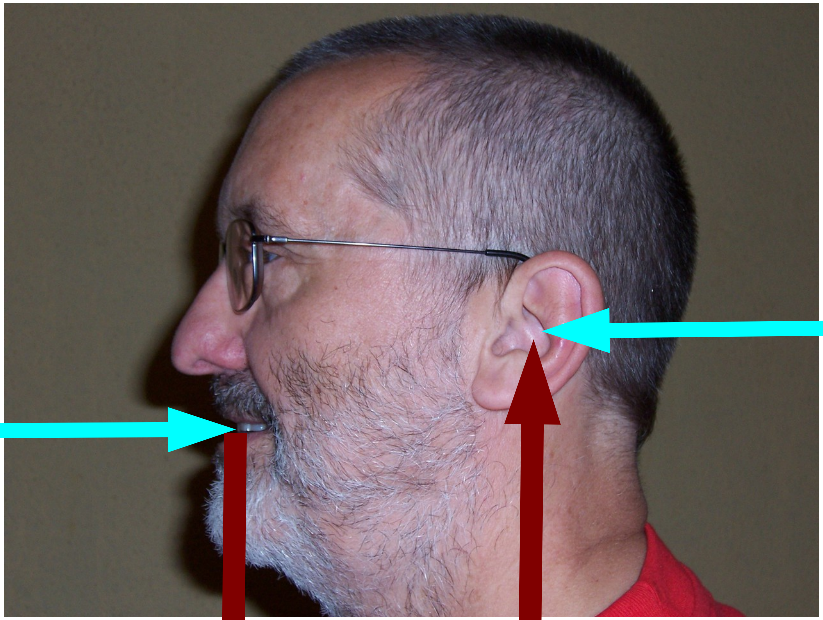
A tiger and a mouse were walking in a field...

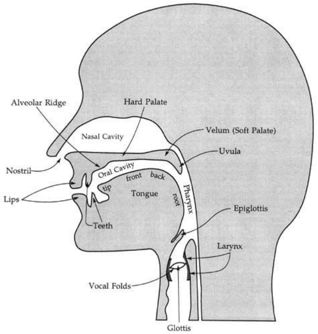


Acoustic Phonetics

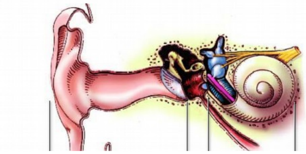
The Domains of Phonetics: the Phonetic Cycle

Articulatory Phonetics





Auditory Phonetics



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

Acoustic Phonetics

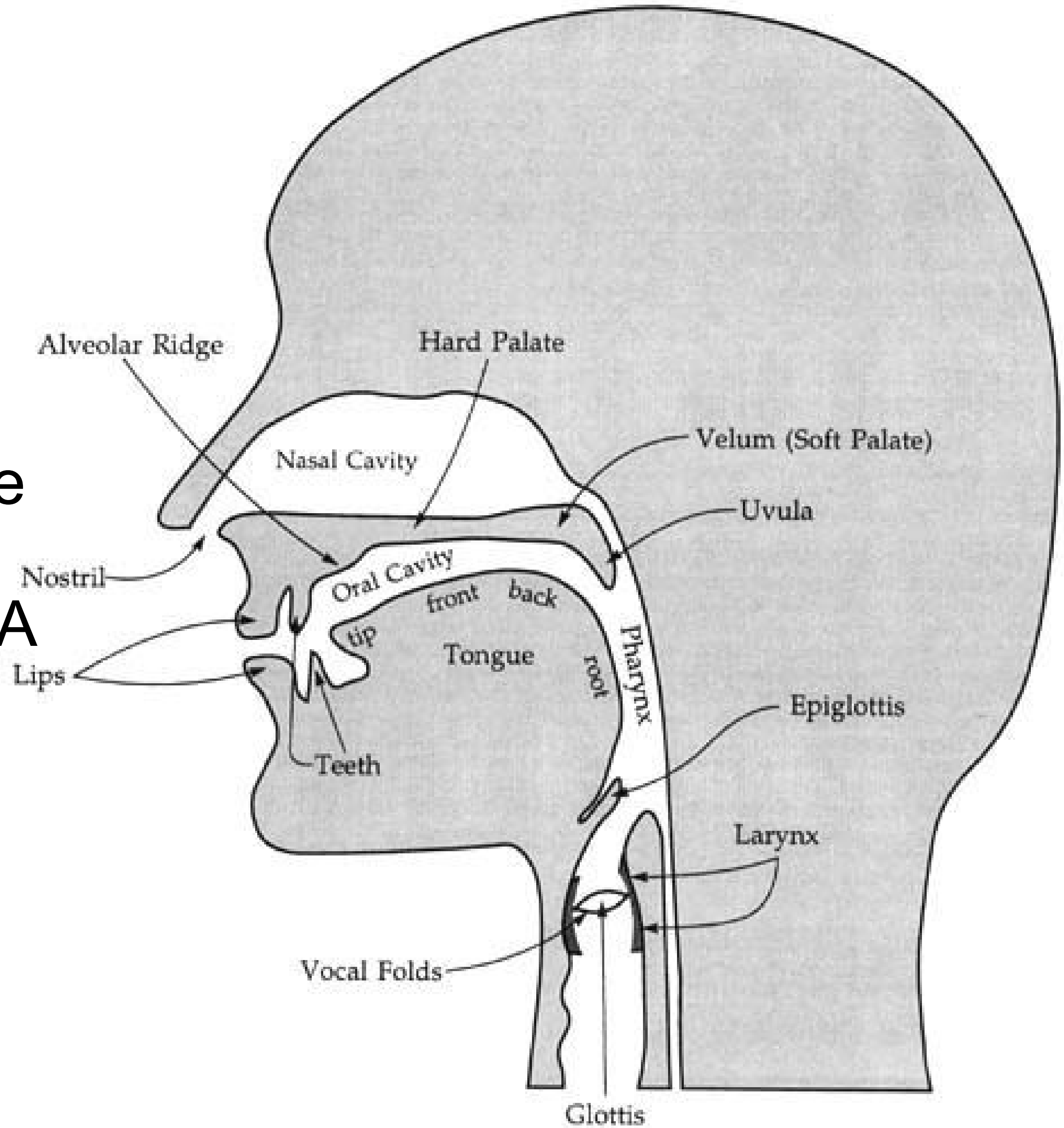
Quiz on the Phonetic Cycle

- Define each of the following:
  - articulatory phonetics?
  - acoustic phonetics?
  - auditory phonetics?
- Which parts of the head are they associated with?
- What is the “phonetic cycle”?

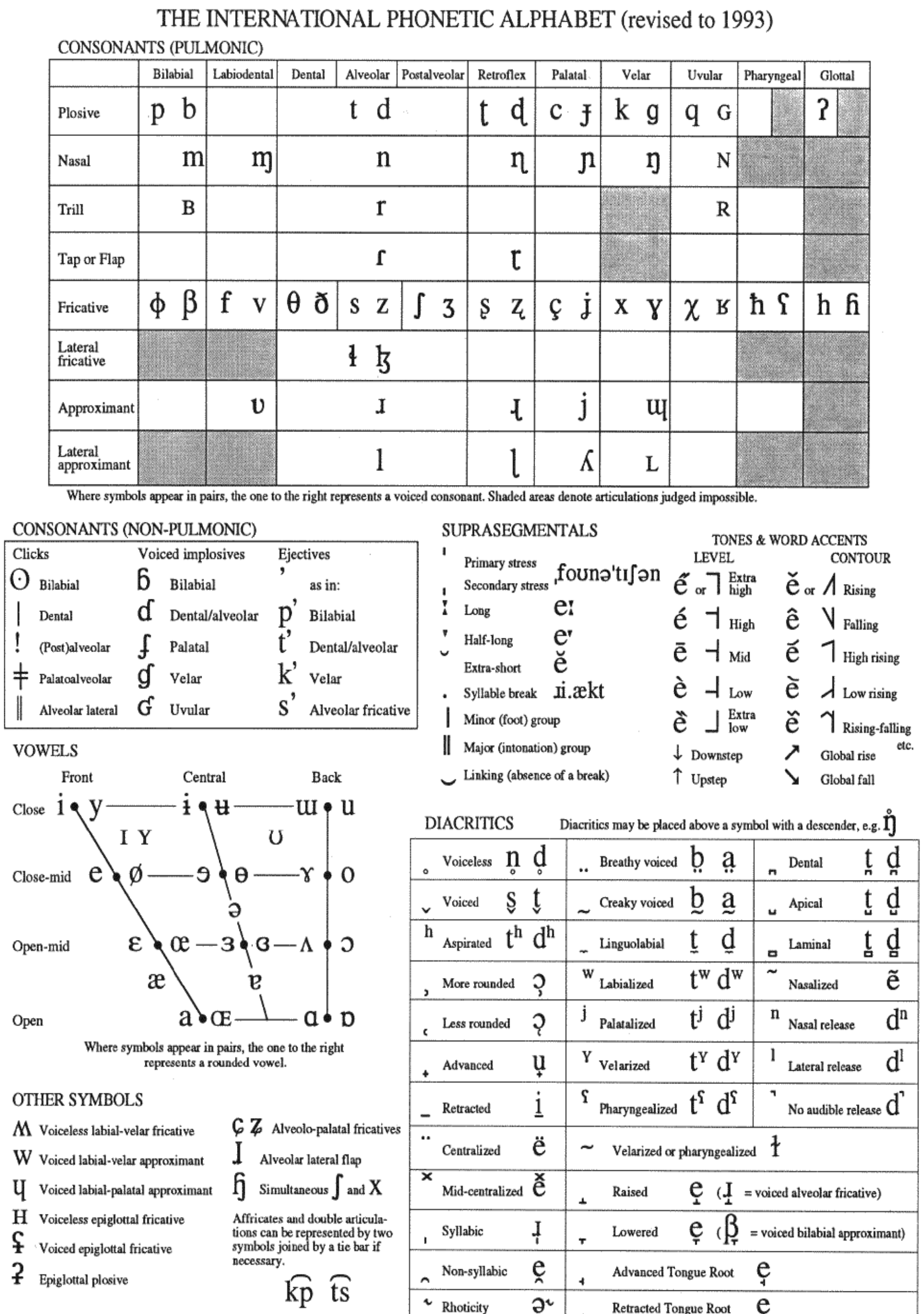
Articulatory Phonetics (Speech Production)

The articulatory domain

- Domain of speech production
- Articulatory organs are relatively easily observable
- Domain of reference for phonetic categories of the IPA
- Investigated via
  - corpus creation
  - experiment paradigm



The IPA (A = Alphabet / Association)



THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

	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				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.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	ʼ	as in:
ǀ	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
ǃ	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

VOWELS

Front Central Back

Close i • y — ɨ • ʉ — ɯ • u

ɪ ʏ ʊ

Close-mid e • ø — ɘ • ɵ — ɤ • o

ə

Open-mid ɛ • œ — ɜ • ɞ — ʌ • ɔ

æ ɐ

Open a • ɶ — ɑ • ɒ

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

OTHER SYMBOLS

ʍ Voiceless labial-velar fricative
w Voiced labial-velar approximant
ɥ Voiced labial-palatal approximant
ʜ Voiceless epiglottal fricative
ʢ Voiced epiglottal fricative
ʡ Epiglottal plosive
ɕ ʑ Alveolo-palatal fricatives
ɺ Alveolar lateral flap
ɧ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

SUPRASEGMENTALS

ˈ Primary stress ˌfoʊnəˈtɪʃən
ˌ Secondary stress
ː Long eː
ˑ Half-long eˑ
˘ Extra-short ĕ
. Syllable break ɹi.ækt
| Minor (foot) group
‖ Major (intonation) group
‿ Linking (absence of a break)

TONES & WORD ACCENTS

LEVEL		CONTOUR	
e̋ or ˥	Extra high	ě or ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e᷄ ˧˥	High rising
è ˨	Low	e᷅ ˩˧	Low rising
ȅ ˩	Extra low	e᷈ ˧˦˧	Rising-falling etc.
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

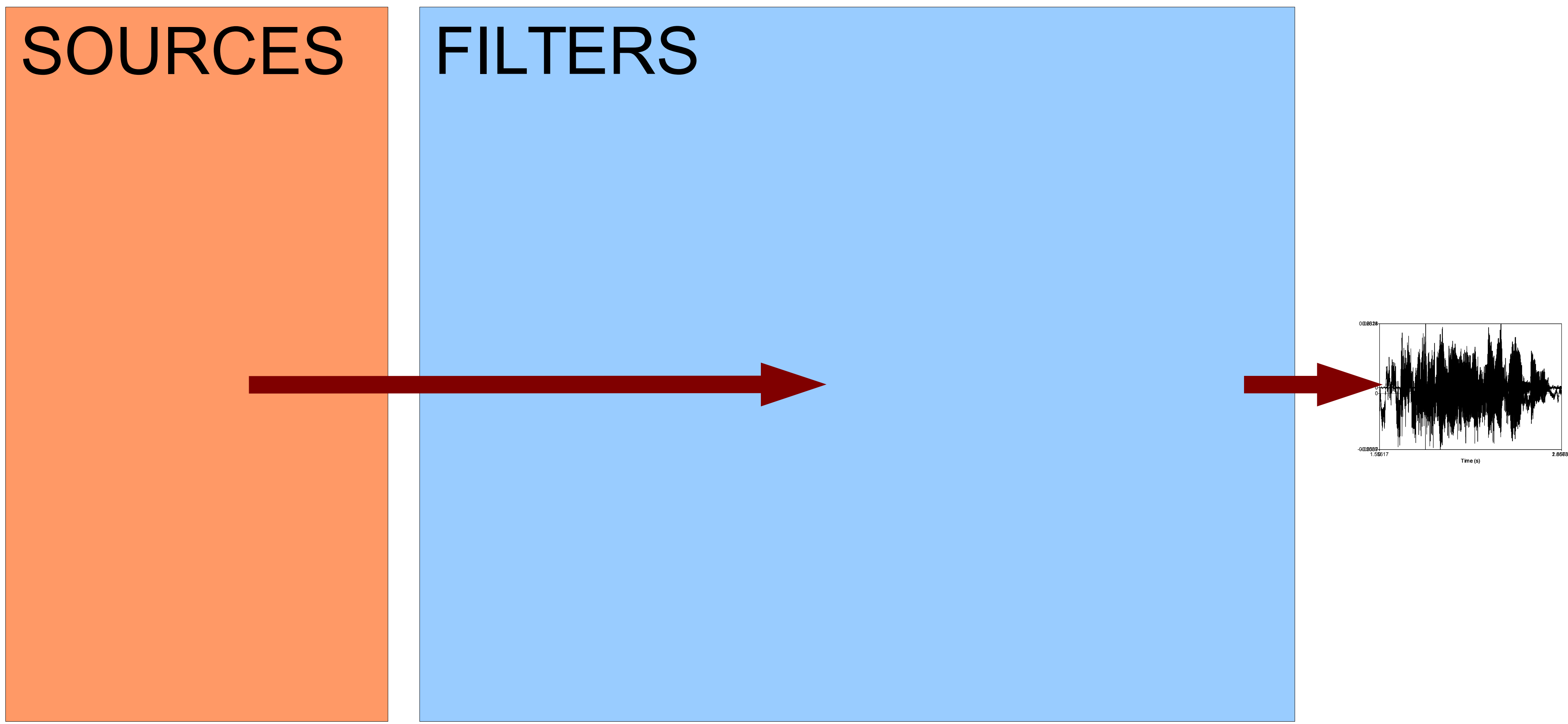
̥ Voiceless	n̥ d̥	̤ Breathy voiced	b̤ a̤	̪ Dental	t̪ d̪
̬ Voiced	s̬ t̬	̰ Creaky voiced	b̰ a̰	̺ Apical	t̺ d̺
ʰ Aspirated	tʰ dʰ	̼ Linguolabial	t̼ d̼	̻ Laminal	t̻ d̻
̹ More rounded	ɔ̹	ʷ Labialized	tʷ dʷ	̃ Nasalized	ẽ
̜ Less rounded	ɔ̜	ʲ Palatalized	tʲ dʲ	ⁿ Nasal release	dⁿ
̟ Advanced	u̟	ˠ Velarized	tˠ dˠ	ˡ Lateral release	dˡ
̠ Retracted	i̠	ˤ Pharyngealized	tˤ dˤ	̚ No audible release	d̚
̈ Centralized	ë	̴ Velarized or pharyngealized ɫ			
̽ Mid-centralized	e̽	̝ Raised e̝ (ɹ̝ = voiced alveolar fricative)			
̩ Syllabic	n̩	̞ Lowered e̞ (β̞ = voiced bilabial approximant)			
̯ Non-syllabic	e̯	̘ Advanced Tongue Root e̘			
˞ Rhoticity	ə˞	̙ Retracted Tongue Root e̙			

- IPA: 120 years old
- regularly re-examined and revised by Association
- based on articulatory categories
- designed to capture the phonemes of all languages of the world: i.e. phonetic distinctiveness of the corresponding sound in a language of the world is one key criterion for adopting a symbol

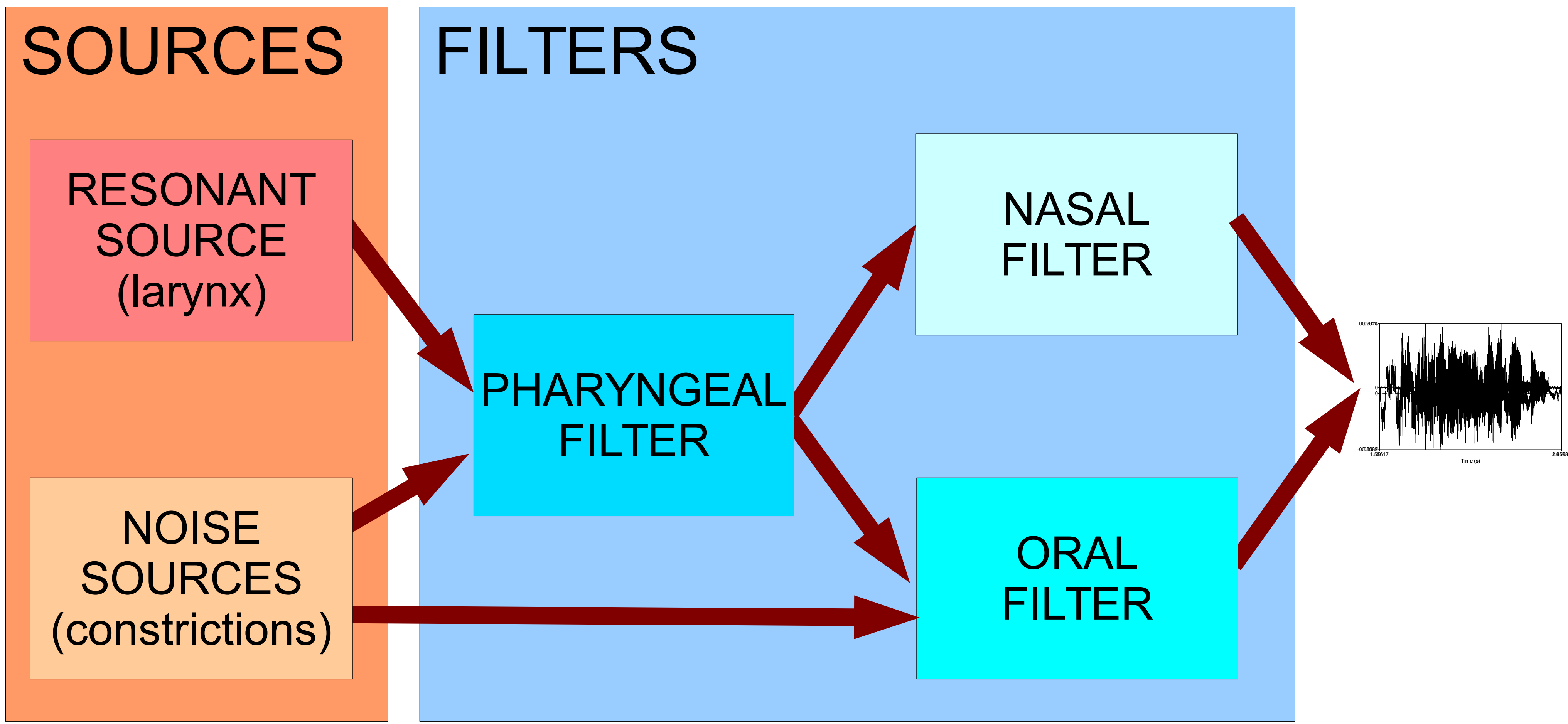
The Source-Filter Model of Speech Production

- A “model” is a simplified representation of relevant features of reality (but it also adds its own artefacts)
- In the Source-Filter Model of speech production, the sound is generated by the SOURCE and modified by the FILTER
- The Source-Filter Model represents the speech production process in two phases:
  - The SOURCE of the sound:
    - LARYNX (for resonant, voiced sounds)
    - CONSTRICTION OF THE ORAL CAVITY (for noisy sounds such as obstruents)
  - The FILTER through which the sound has passed:
    - the PHARYNGEAL CAVITY
    - the ORAL CAVITY
    - the NASAL CAVITY

The Source-Filter Model of Speech Production



The Source-Filter Model of Speech Production



Quiz on Articulatory Phonetics

- What are the main articulators involved in
  - vowel production?
  - consonant production?
  - tone production?
- Produce the following consonants, followed by the vowel [a]:
  - voiceless bilabial fricative
  - voiced alveolar affricate
  - voiced palatal stop
  - voiceless labial-velar stop
  - implosive velar stop
  - velar nasal
- What is the source-filter model?
  - Illustrate this, referring to the difference in sound between speaking in a tiled bathroom and in the open air.

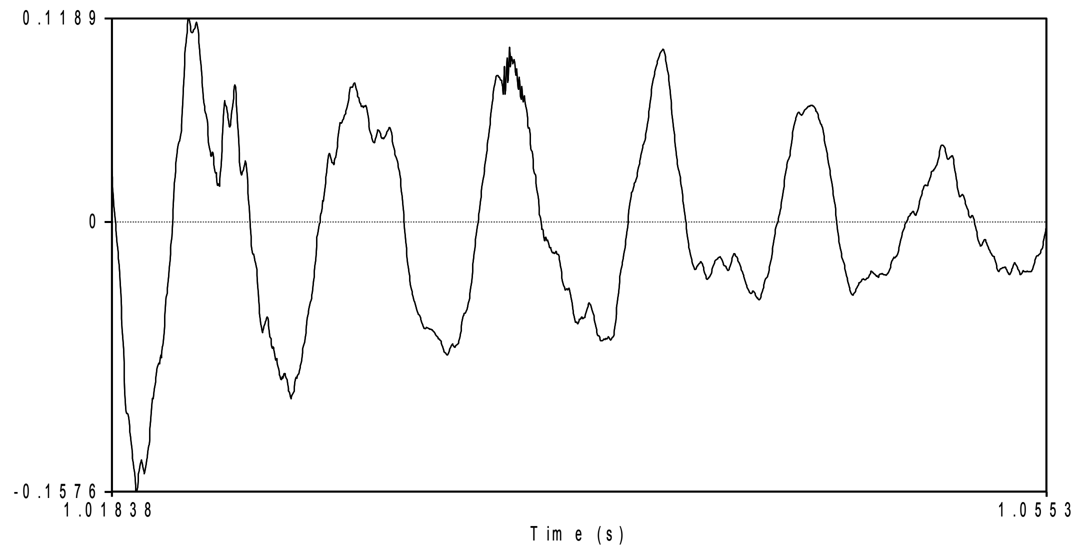
Acoustic Phonetics
(Speech Transmission)

The acoustic domain

- Acoustic phonetics is concerned with investigating the transmission of speech signals through
  - gases such as air, other substances (e.g. bone, tissue)
  - electronic amplification and storage
- The basic parameters of the speech signals are
  - amplitude
  - time (duration)
- The main derived parameters of speech signals are
  - intensity
  - noise vs. resonance (voicing)
  - frequency and formants
- The methods used to analyse speech signals are:
  - analog-to-digital (A/D) conversion
  - mathematical definitions of filters and transformations

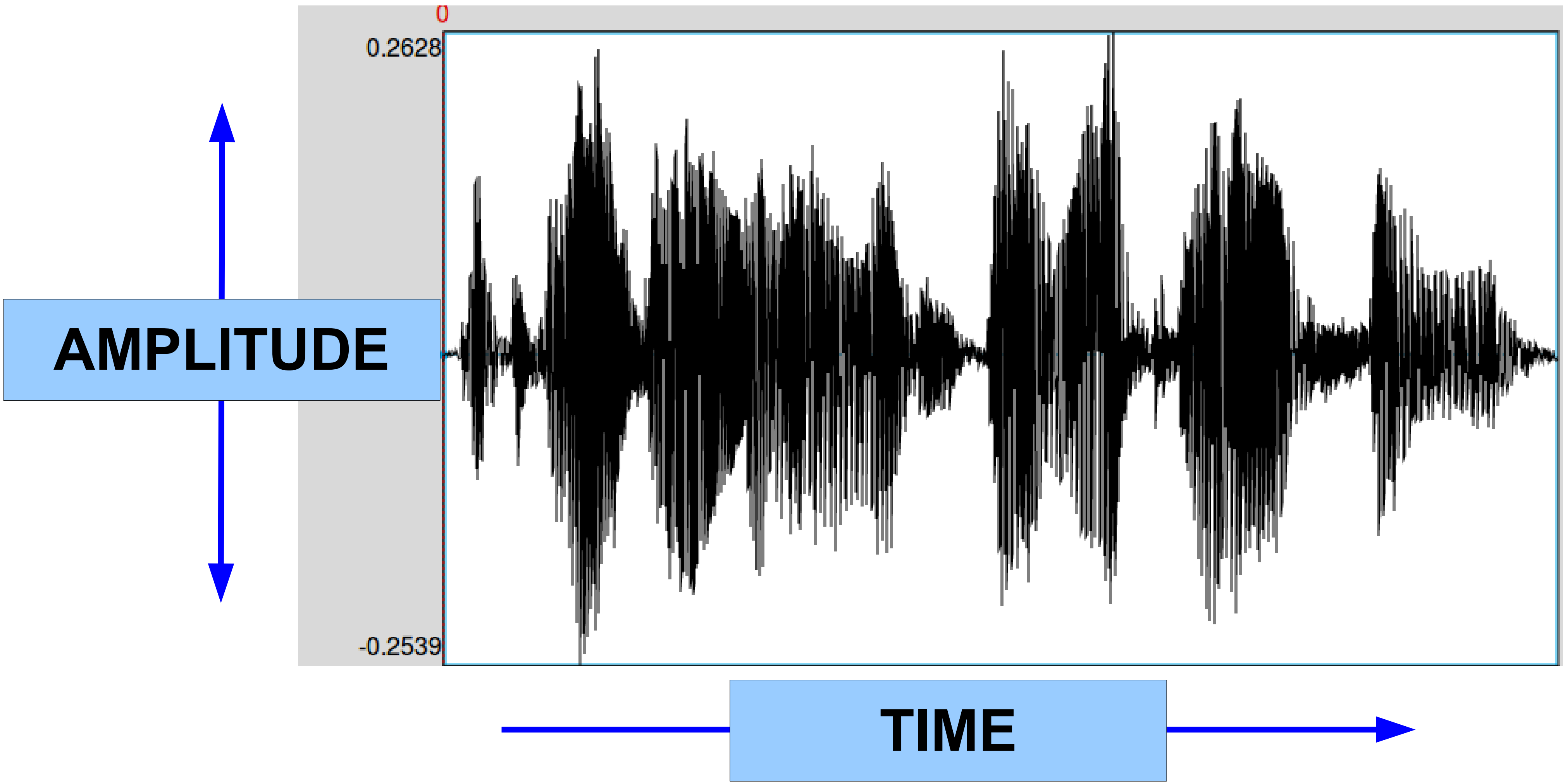
The Speech Wave-Form

- Speech is transmitted through air (and other substances) as a regular wave of pressure changes:



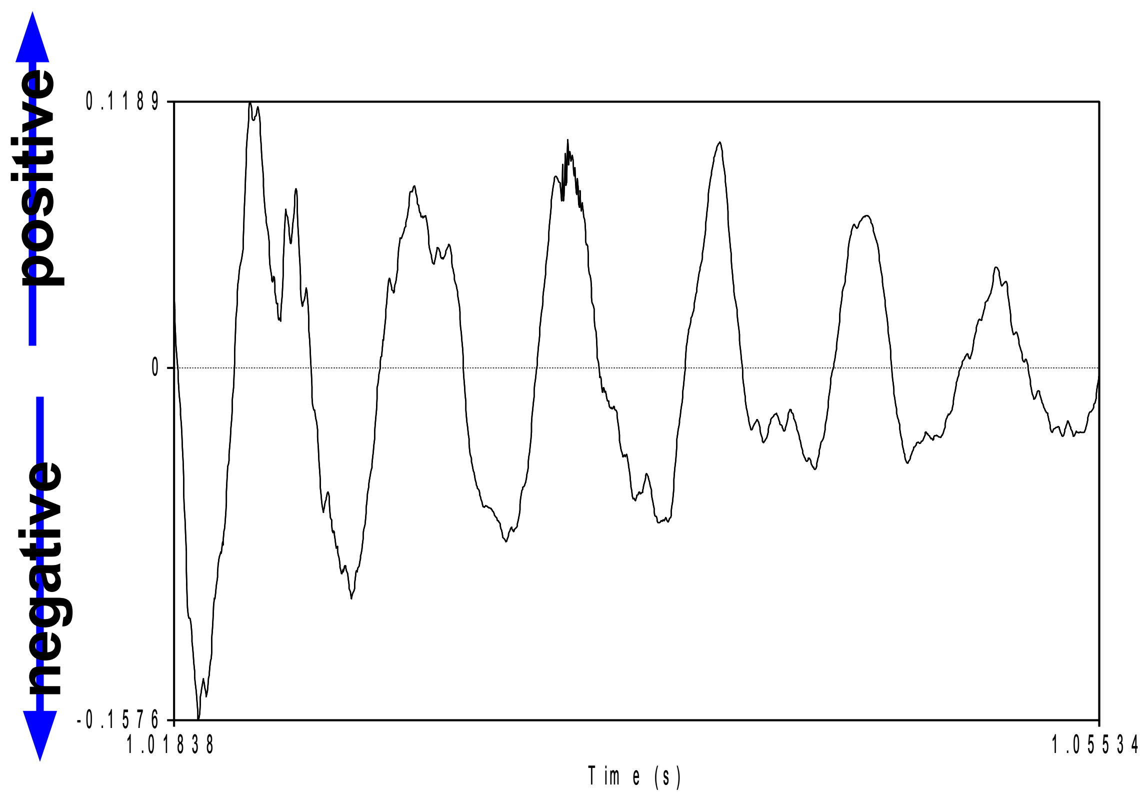
- The changes in air pressure
  - but can be heard
  - and cannot be seen (unlike the waves on the ocean)
  - but can be measured (like the waves on the ocean)
  - and the measurements can be visualised and used for calculating statistical models of the structure of speech

Basic Speech Signal Parameters



The Time Domain: the Speech Wave-Form

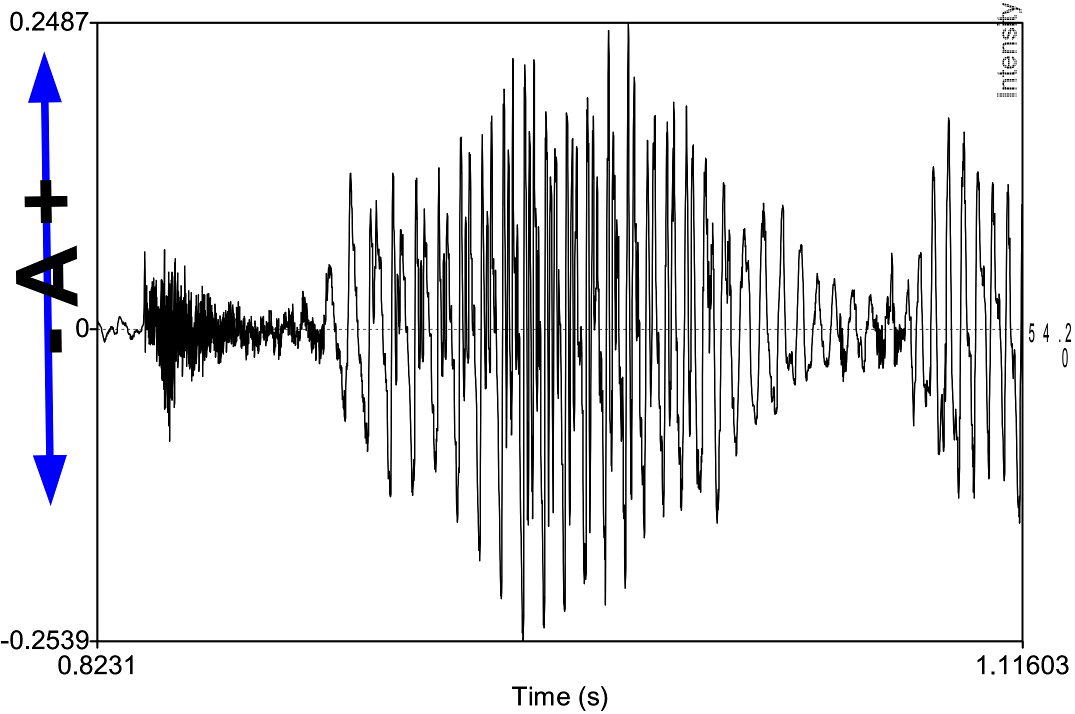
- The *positive* or *negative amplitude* **A** of the speech signal at any given point in time is the *distance* of the wave from zero at this point in time.

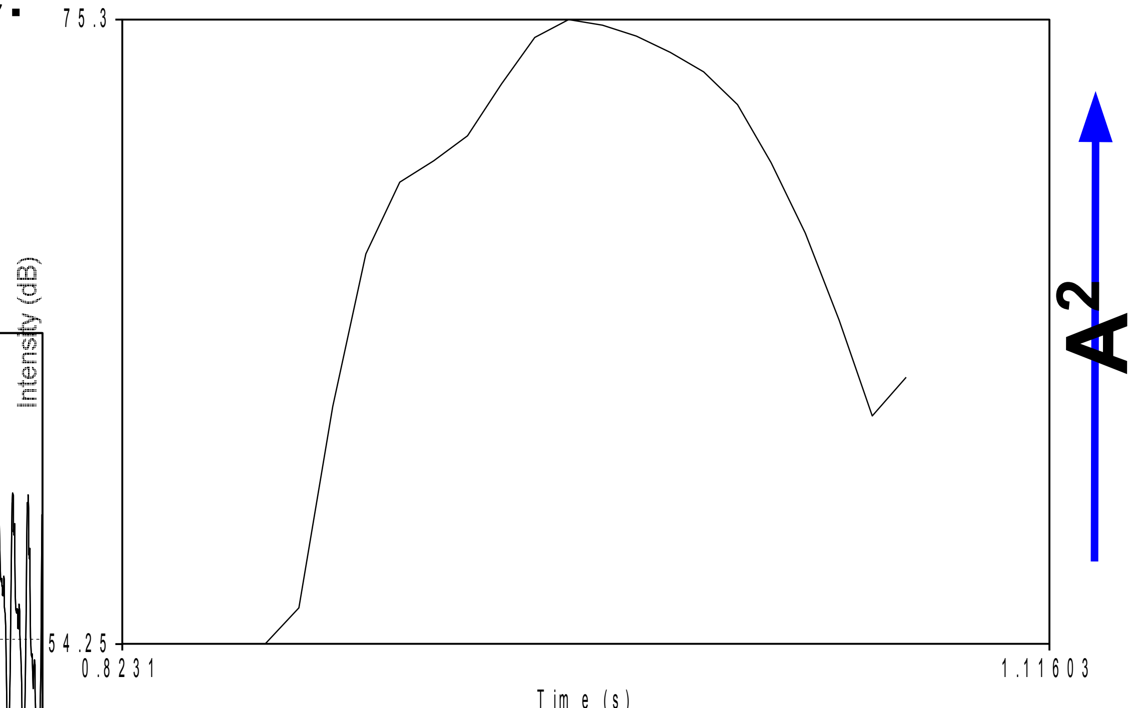


Derived parameter *INTENSITY*

- The *intensity* of the speech signal at any given point in time is the *square of the amplitude* of the wave from zero at this point in time:

$I = A^2$





tiger

Derived parameter *ENERGY*

- The energy E (root-mean-square energy) is
  - the square root of the mean of a sequence of intensity values $I_1, ..., I_n$ (remember: intensity is amplitude squared)

$$E = \sqrt{\frac{\sum_{i=1...n} A(x_i)^2}{n}}$$

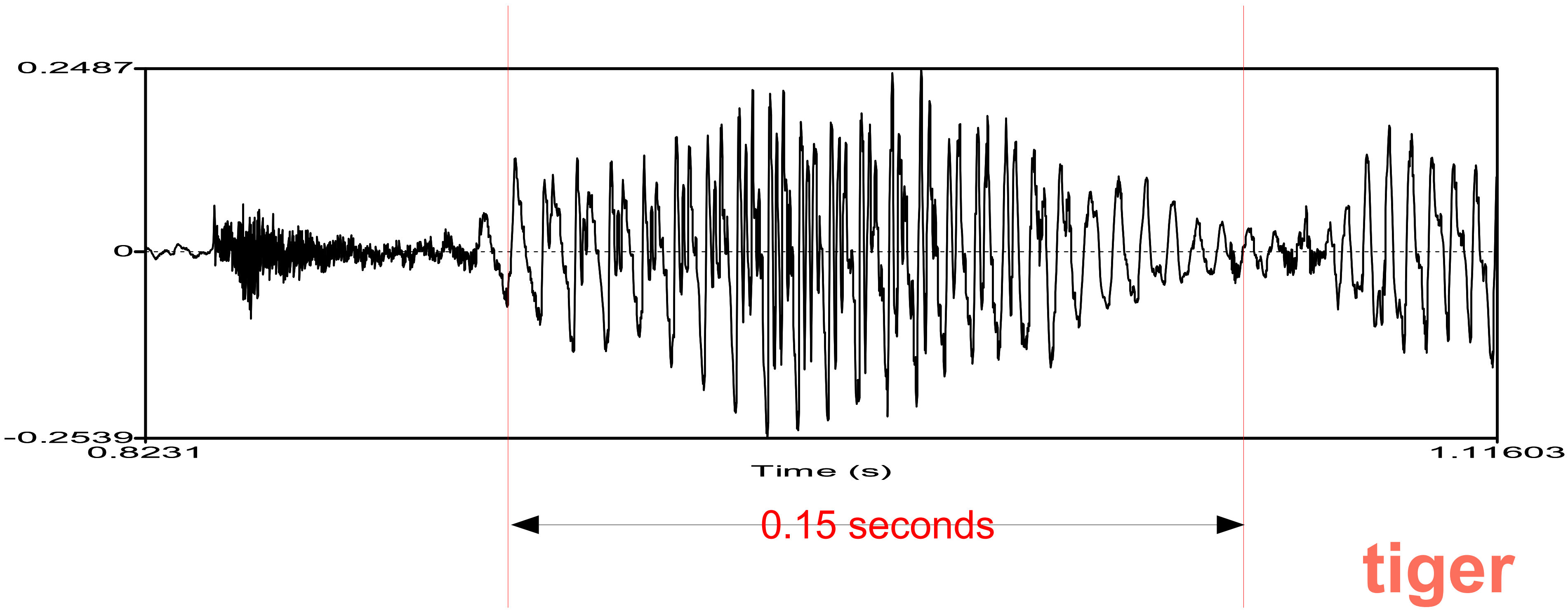
- Energy is therefore intensity averaged over time
  - In fact, intensity measurements are, in practise, energy measurements over very short periods of time
- Compare other measurement units per time unit:
  - miles per hour
  - kilowatts per hour

Derived parameters *PERIOD* & *WAVELENGTH*

- The *period* or *interval* of a single wave in a speech signal is the duration of this single wave.
  - A *resonant signal* is a signal whose periods are regular, i.e. even in duration.
  - A signal is *noisy* if the periods are irregular, i.e. uneven in duration
  - The average period of a speech signal
- The *wavelength* λ (lambda) in cm of a speech signal is the speed of sound in cm/sec divided by the number of periods per second.
  - *You can forget the definition of wavelength...*
  - *A task for the very interested:*
    - *What is the speed of sound?*
    - *What is the wavelength of a sound with 100 periods per second?*

The Frequency Domain: simple & complex signals

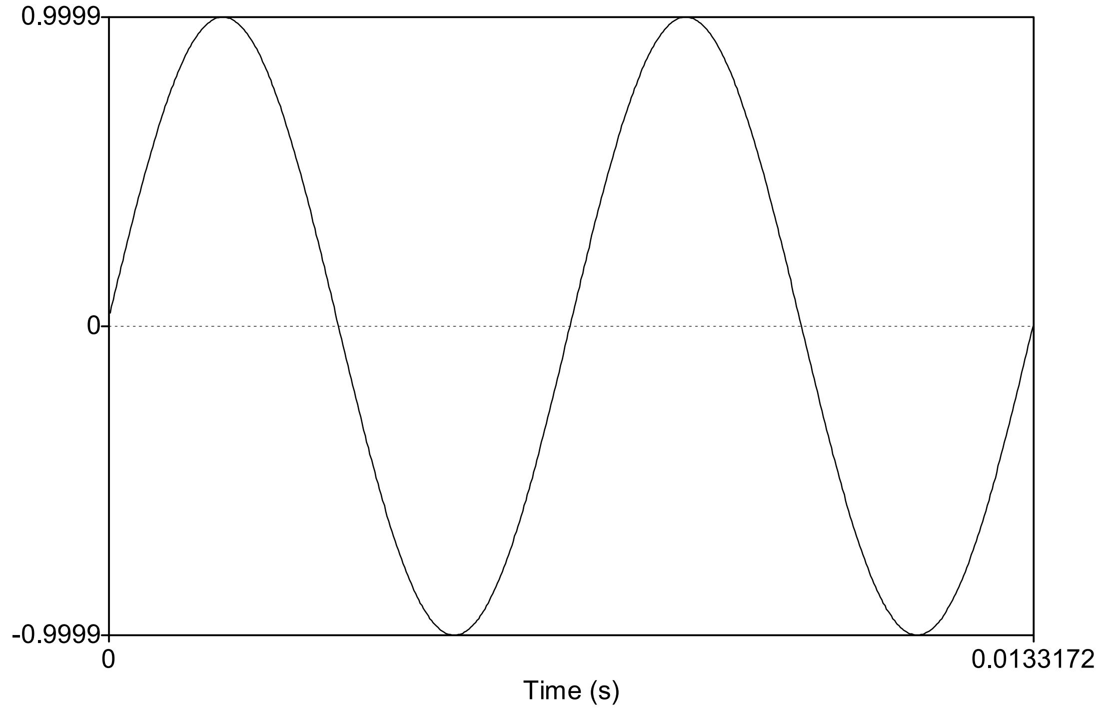
- The *frequency* of a speech signal is the number of waves (periods) per second in the waveform



- Question:
  - Ignoring the irregularities in the waveforms: what is the average frequency of the segment between the red lines?

The Simplest Sources produce Sine Waves

- A sine wave of frequency F is produced by an evenly swinging pendulum (a very slow sine wave, of course).



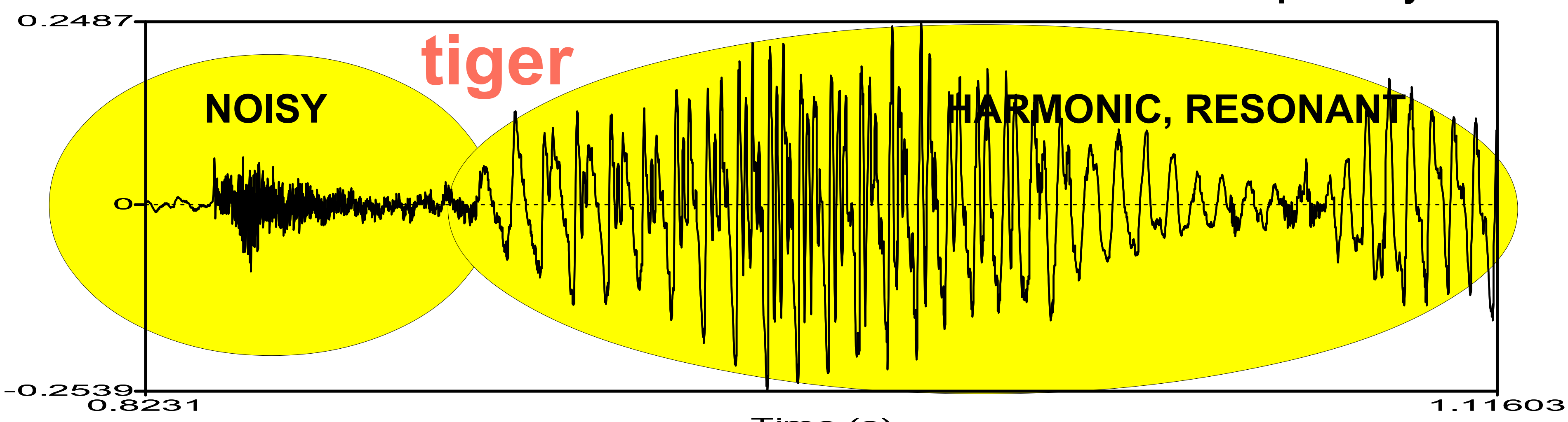
- The speech signal is not a simple sine wave, however, but a complex signal.

The Frequency Structure of Speech

- The SOURCE
  - for harmonic, voiced sounds
  - is the larynx.
- The larynx produces
  - a complex waveform, consisting of
    - a fundamental frequency
      - about 80 Hz - 150 Hz for men
      - about 160 Hz - 300 Hz for women
    - many overtones, which are audible up to about 20 kHz
    - different intensities of the overtones, relative to each other, determines the overall waveform, and therefore the kind of sound which the source produces
    - during voicing, the larynx generates a waveform which is rather like a “sawtooth” sequence

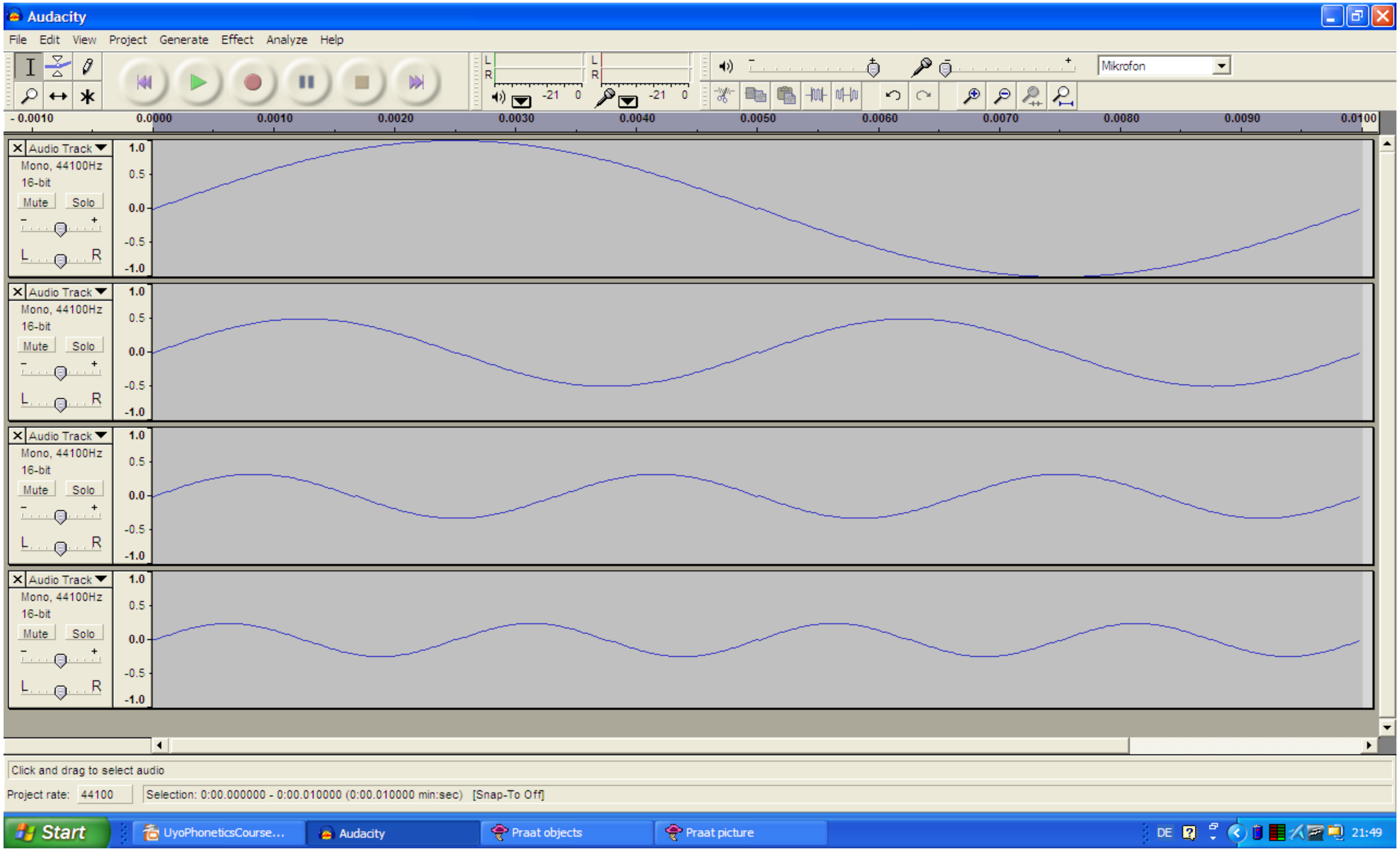
Complex Sources: noisy & harmonic signals

- If many sine waves of arbitrary frequencies occur together, the result is NOISE.
- If many sine waves occur together, with each being an integer multiple of some lowest frequency,
  - the resulting overall wave is a HARMONIC wave:
  - the lowest frequency of a harmonic waveform is the *fundamental frequency*, F0 (f-zero, f-nought)
  - the higher frequencies in a harmonic waveform are called the *harmonics* or *overtones* of the fundamental frequency



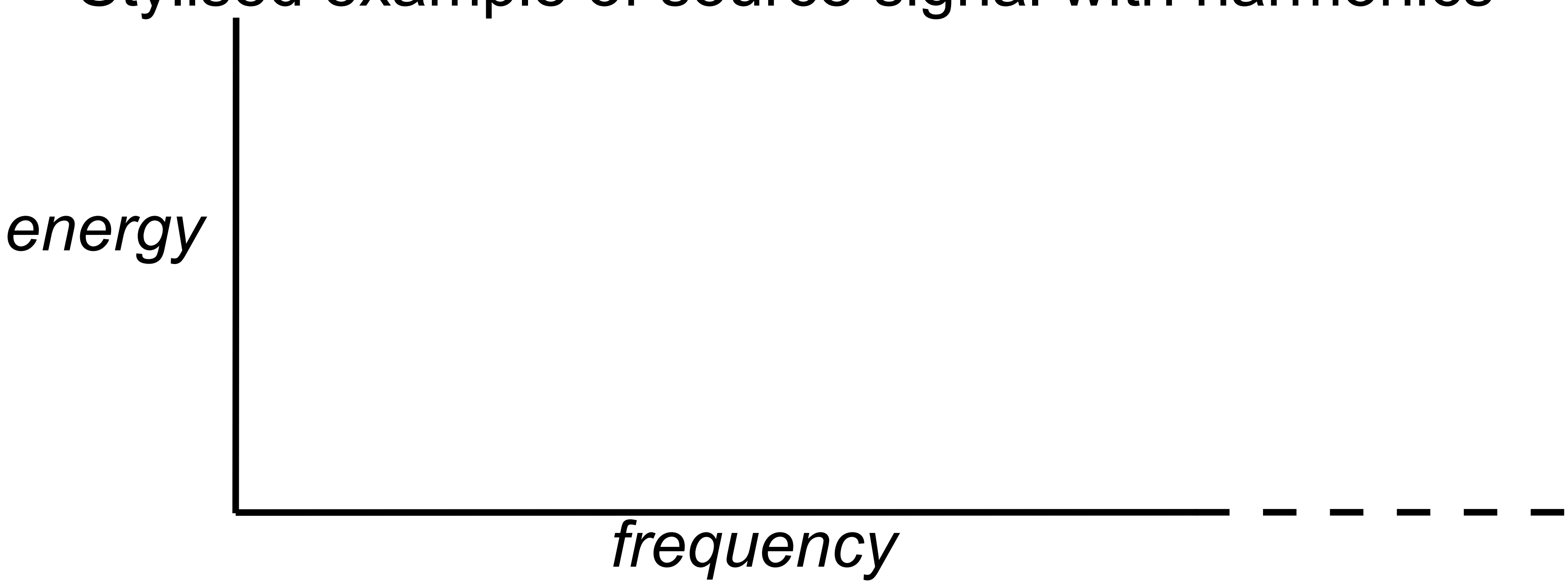
Sources with Integer Multiples of Sine Waves

- Harmonic, resonant frequencies are created by adding several sine waves together, point by point
- The larynx sound source is a special case of this



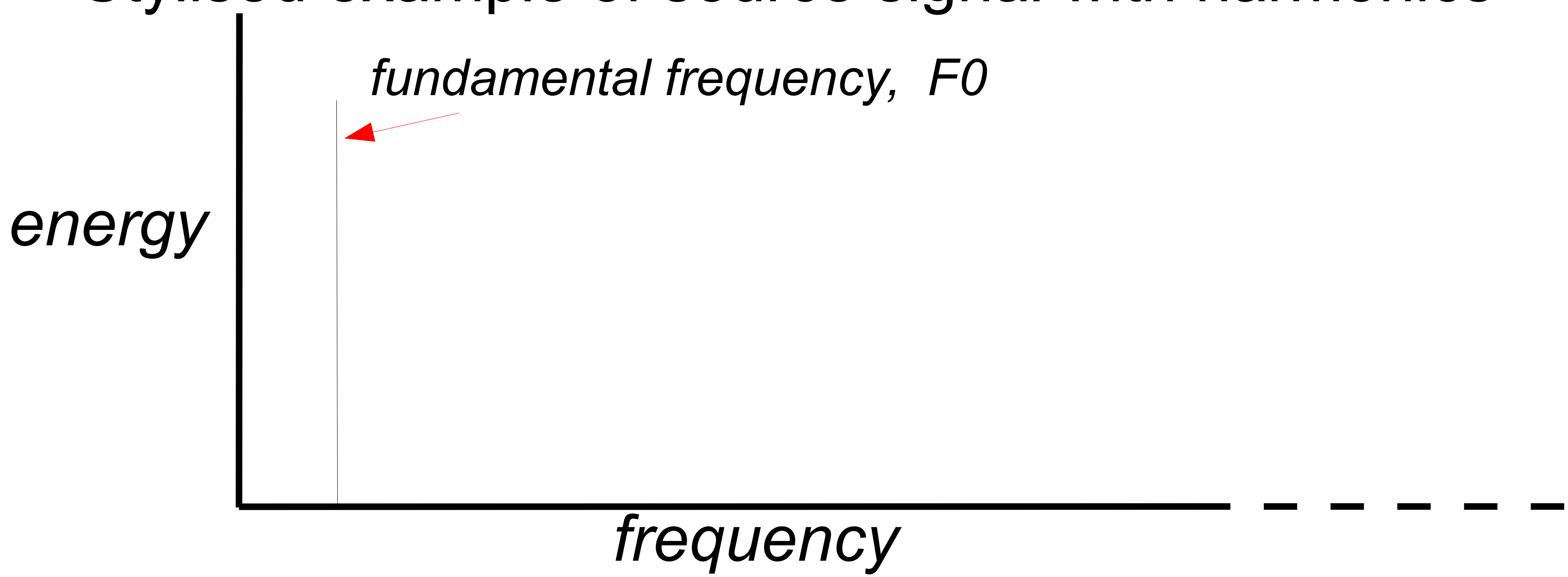
Harmonics / overtones in complex signals

- If a complex signal consists of
  - a series of sine waves with frequencies of *f*, 2*f*, 3*f*, ..., *nf*
    - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
  - then the signal is a resonant signal
  - and *f* is the *fundamental frequency* F0
  - while 2*f*, 3*f*, ..., *nf* are harmonics of the fundamental frequency
- Stylised example of source signal with harmonics



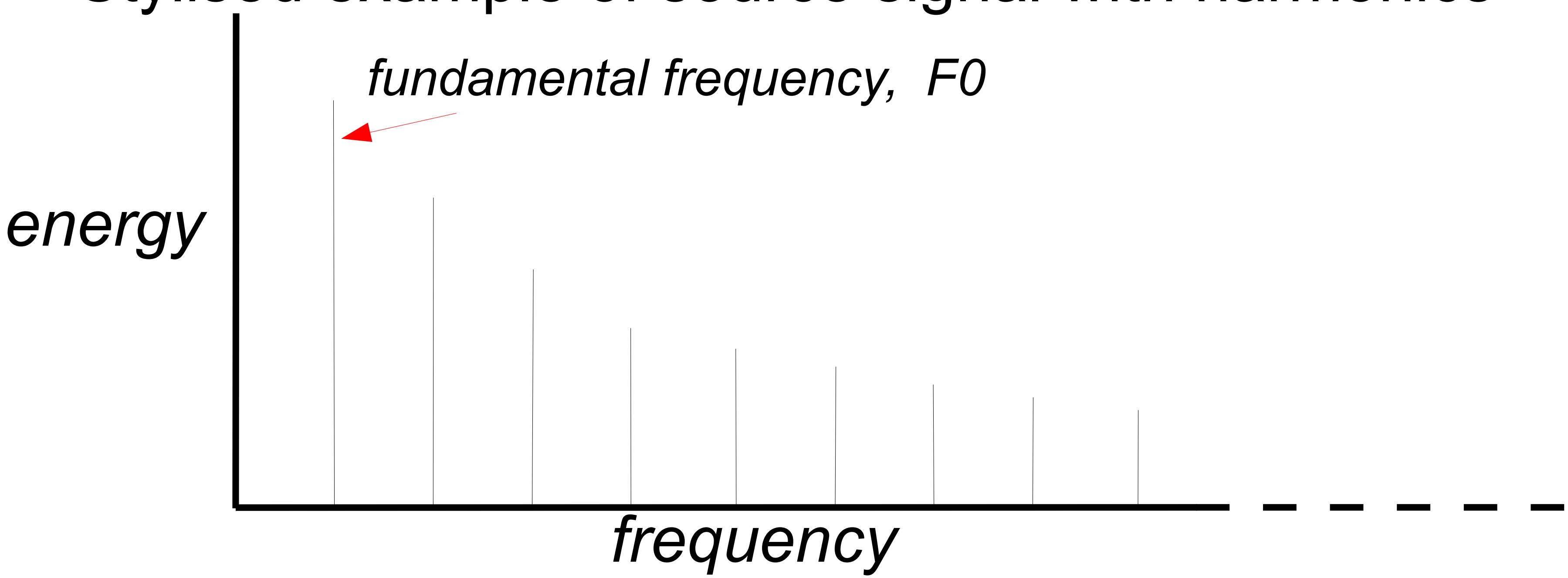
The Spectrum of Complex Signals

- If a complex signal consists of
  - a series of sine waves with frequencies of *f*, *2f*, 3*f*, ..., *nf*
    - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
  - then the signal is a resonant signal
  - and *f* is the *fundamental frequency* F0
  - while 2*f*, 3*f*, ..., *nf* are *harmonics* of the fundamental frequency
- Stylised example of source signal with harmonics



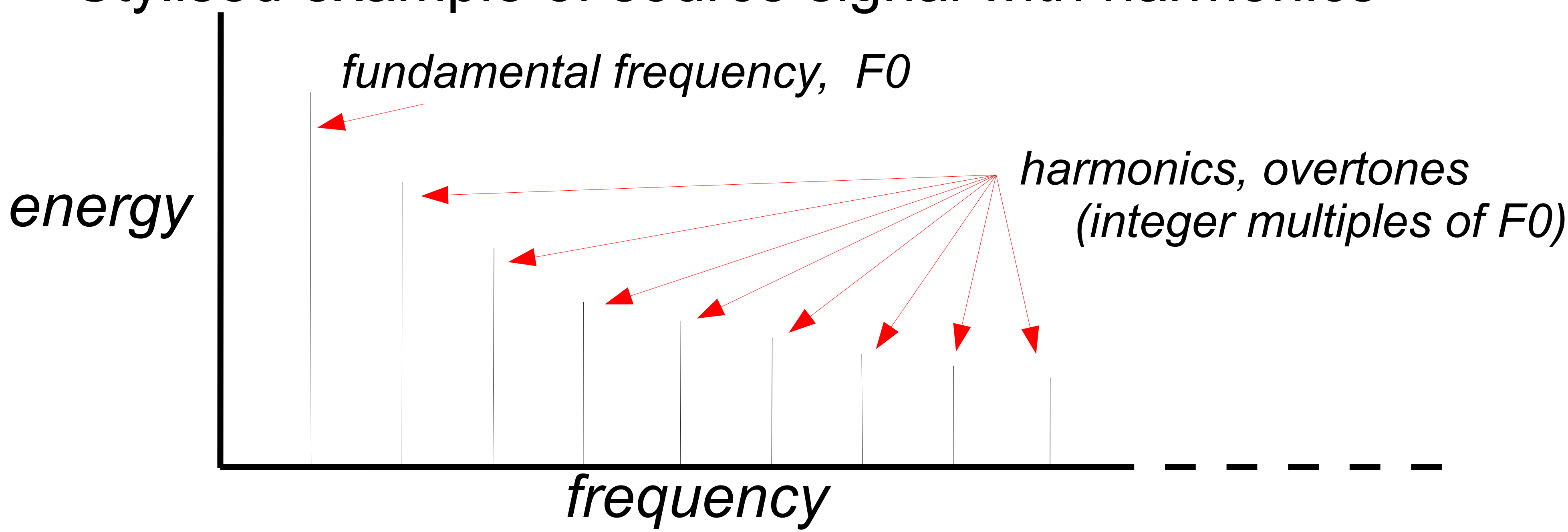
The Spectrum of Complex Signals

- If a complex signal consists of
  - a series of sine waves with frequencies of *f*, 2*f*, 3*f*, ..., *nf*
    - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
  - then the signal is a resonant signal
  - and *f* is the *fundamental frequency* F0
  - while 2*f*, 3*f*, ..., *nf* are harmonics of the fundamental frequency
- Stylised example of source signal with harmonics



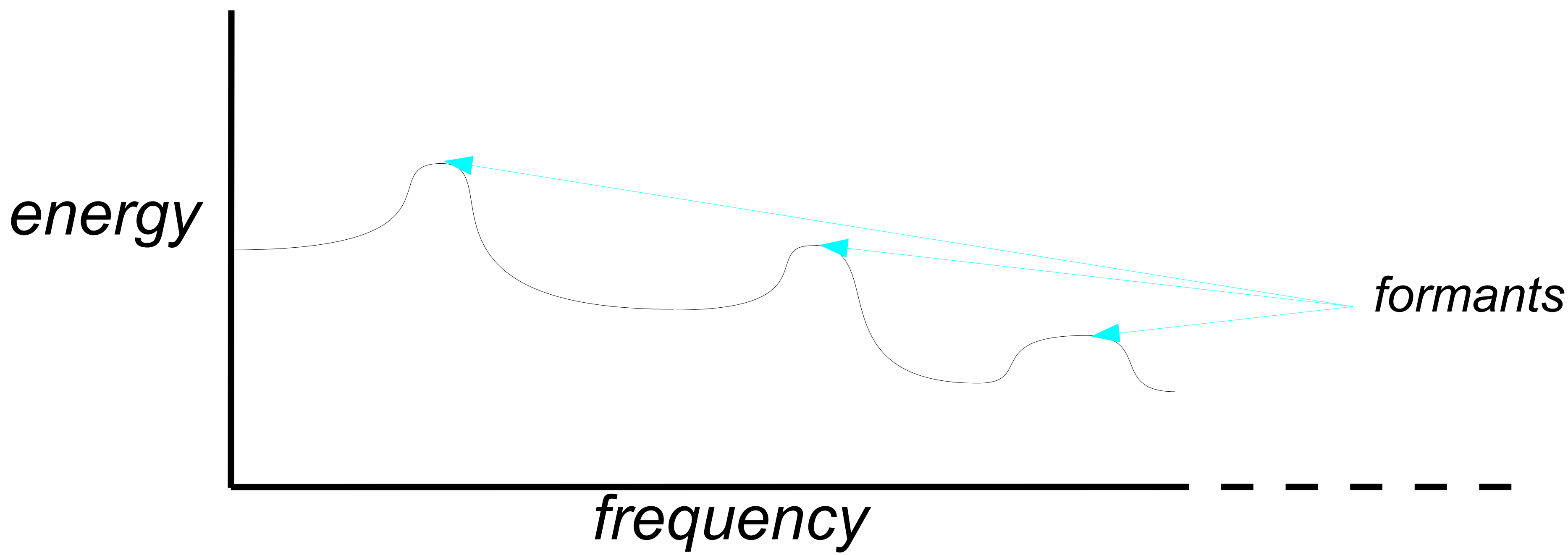
The Spectrum of Complex Signals

- If a complex signal consists of
  - a series of sine waves with frequencies of *f*, 2*f*, 3*f*, ..., *nf*
    - e.g. frequencies of 150 Hz, 300 Hz, 450 Hz, 600 Hz, ..
  - then the signal is a resonant signal
  - and *f* is the *fundamental frequency* F0
  - while 2*f*, 3*f*, ..., *nf* are harmonics of the fundamental frequency
- Stylised example of source signal with harmonics



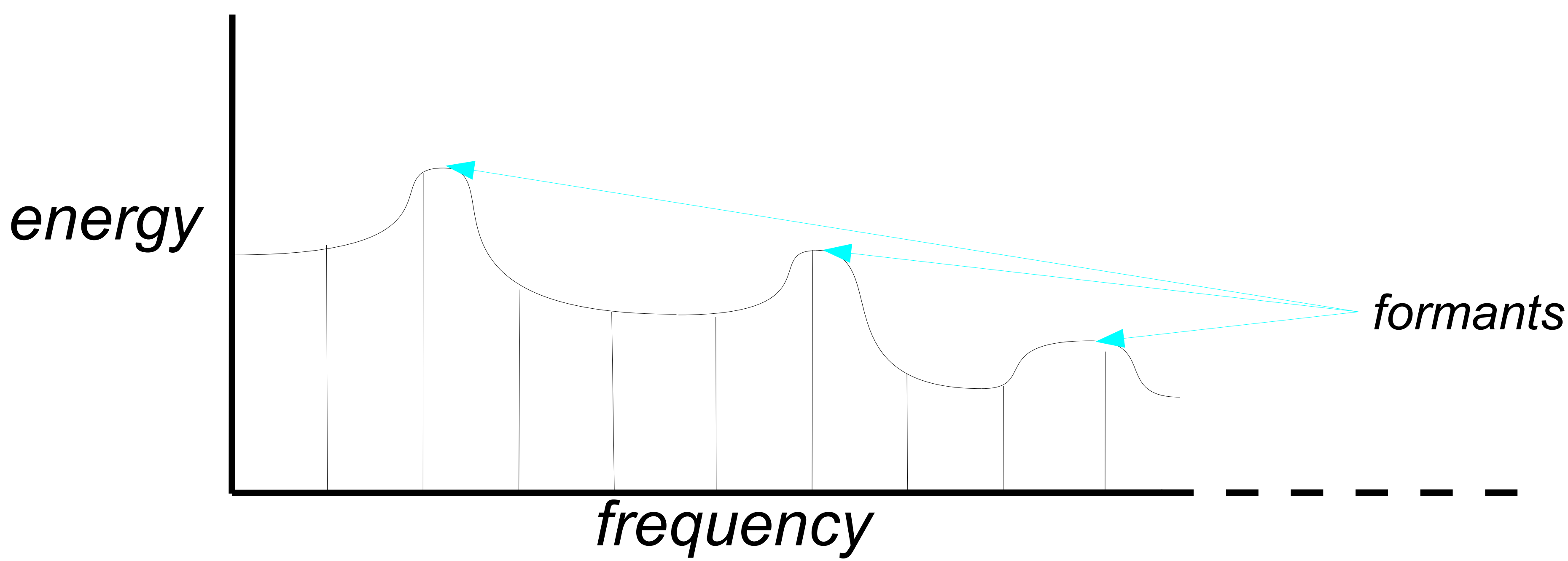
The Spectrum of Complex Signals

- The filter system consists of the pharyngeal, nasal and oral cavities, which have cavities have specific resonant frequencies
- These filter frequency bands are called *formants*
- Formant frequencies of the oral cavity can be modified by the variable filters (articulators *tongue* and *lips*)



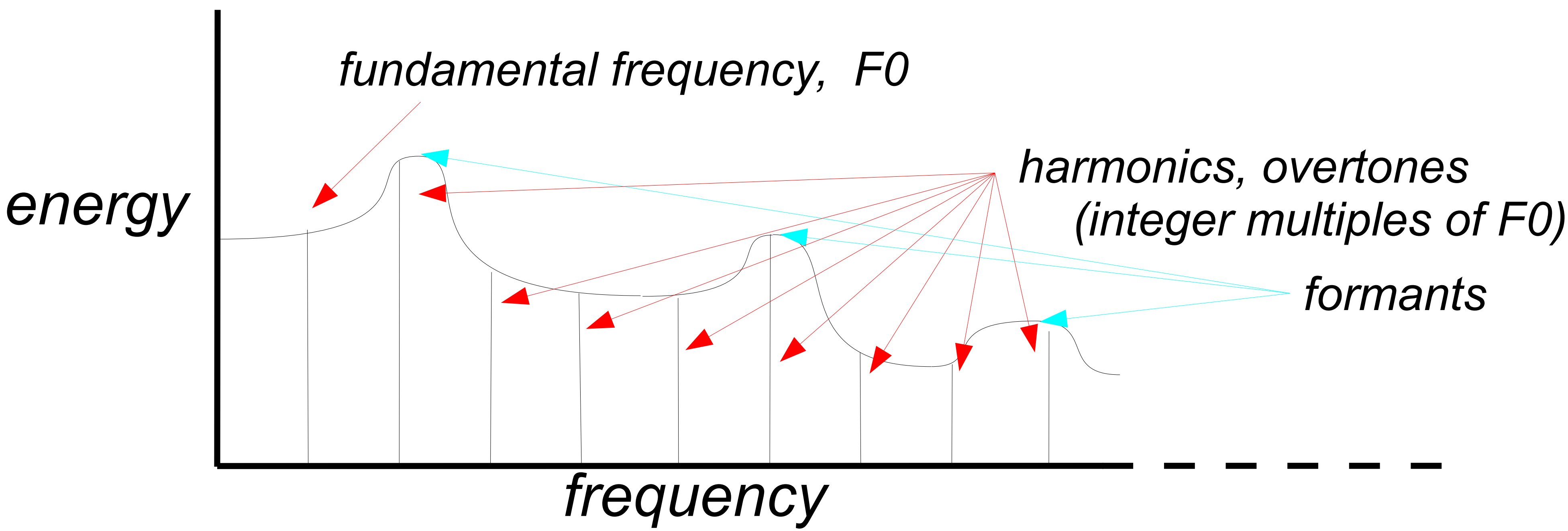
The Spectrum of Complex Signals

- The filter system consists of the pharyngeal, nasal and oral cavities
- The cavities have specific resonant frequencies
- The frequencies of the oral cavity can be modified by the variable filters (the articulators *tongue* and *lips*)
- This means that the energy of the *harmonics* is modified



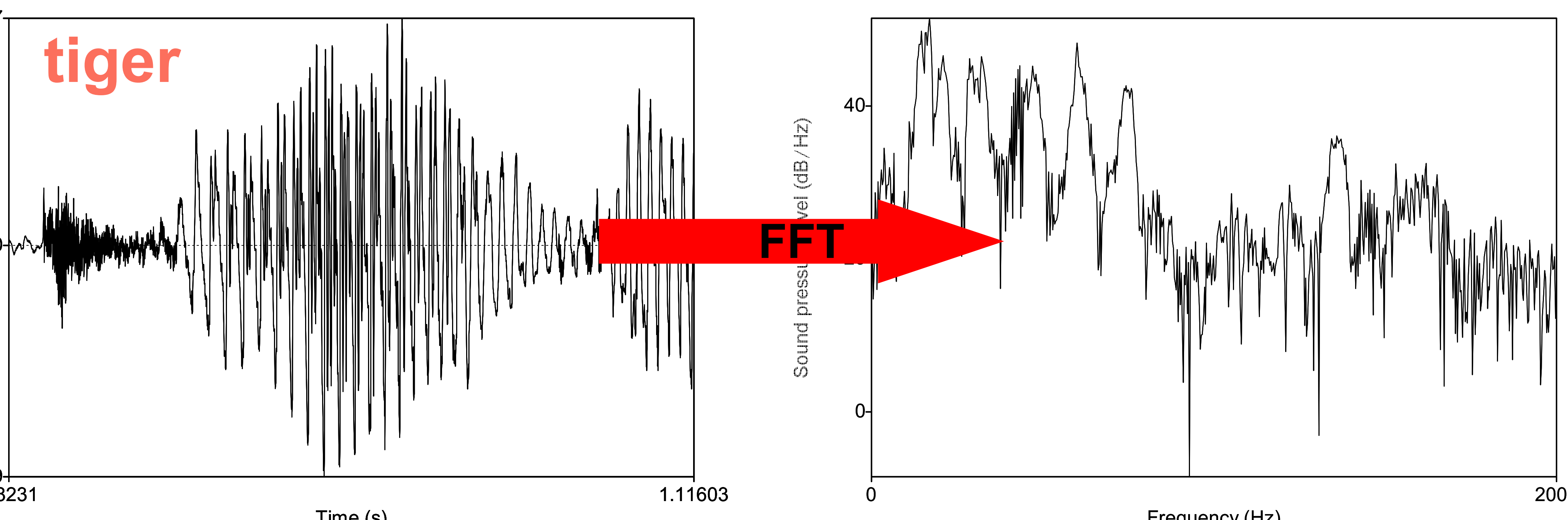
The Spectrum of Complex Signals

- The filter system consists of the pharyngeal, nasal and oral cavities
- The cavities have specific resonant frequencies
- The frequencies of the oral cavity can be modified by the variable filters (the articulators *tongue* and *lips*)
- This means that the energy of the *harmonics* is modified



Fourier Analysis: the Spectrum

- Complex waveforms can be analysed as sums of sine waves (Joseph Fourier, *Fourier Analysis*):
    - the mathematical operation is the *Fourier Transform (FT)*
    - the *Discrete Fourier Transform (FFG)* applies to digitised signals
    - the *Fast Fourier Transform (FFT)* is an optimised version
  - The spikes (harmonics) are generated by the SOURCE, and the peaks (formants) are generated by the FILTER:



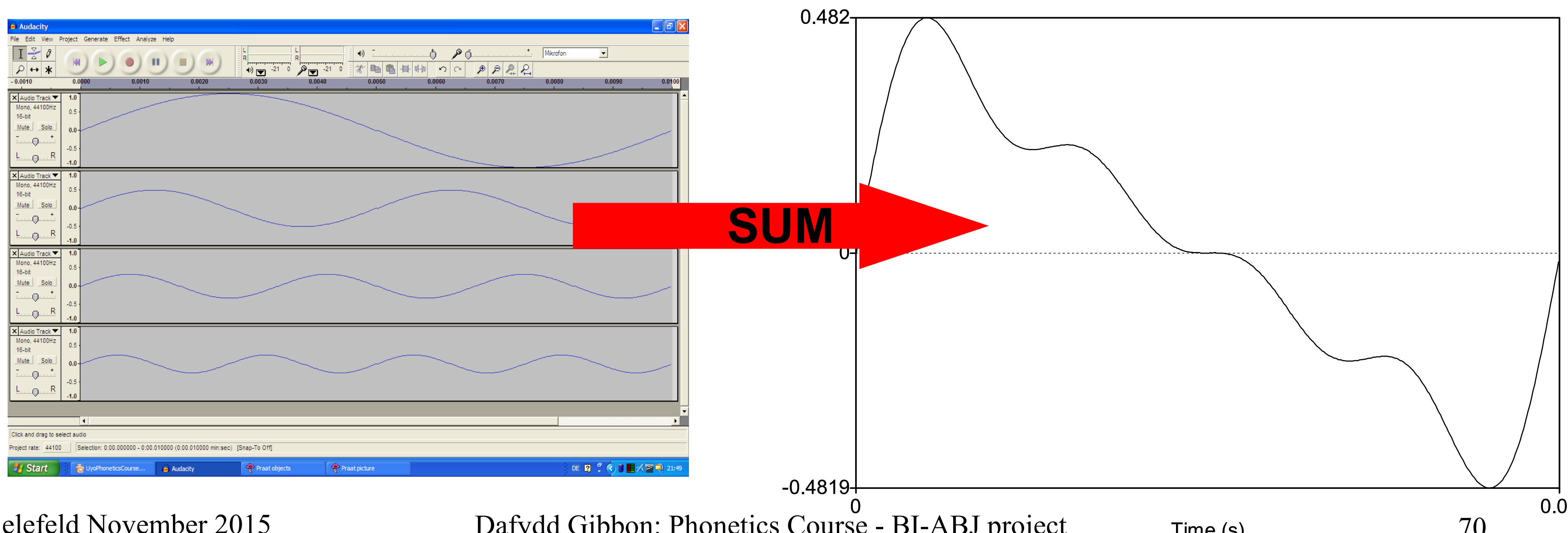
The Speech Sound Source: sawtooth waveforms

- The sum of harmonics which are integer multiples, with A inverse to F, creates a sawtooth waveform:

$$\text{For } x = x_1 \ldots x_n : x_i = \Sigma_{h=1\ldots\mathrm{m}} \frac{\sin(i \times h)}{h}$$

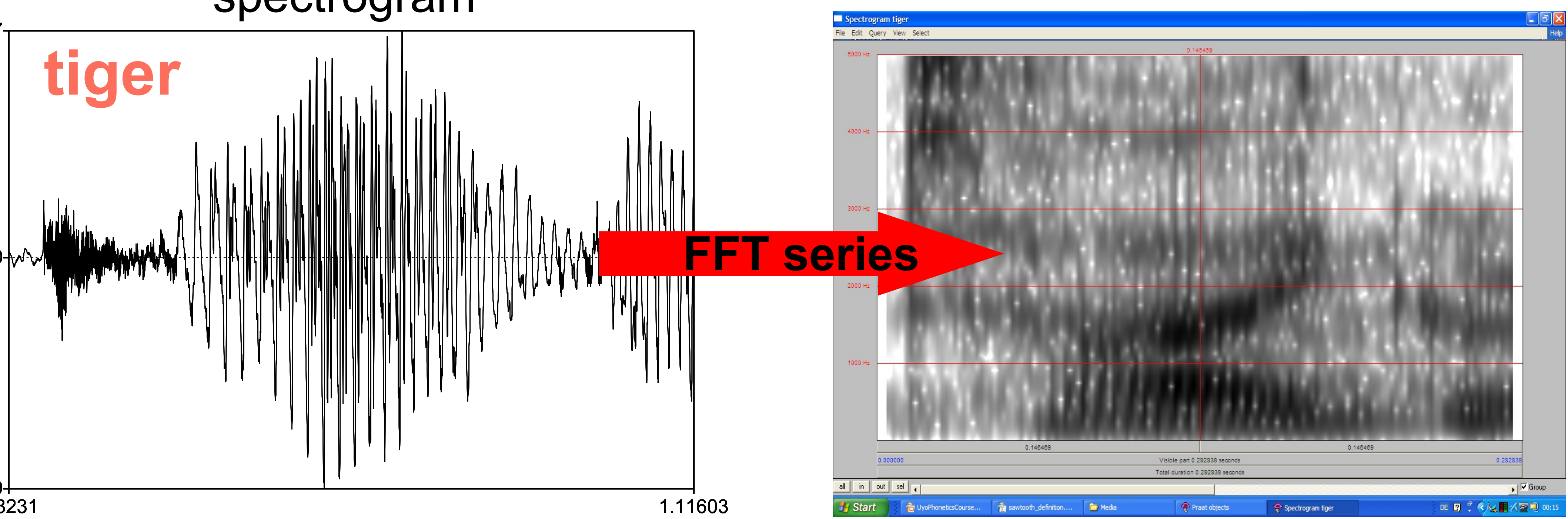
- This example illustrates the sum of four sine waves:

  100 Hz + 200 Hz + 300 Hz + 400 Hz

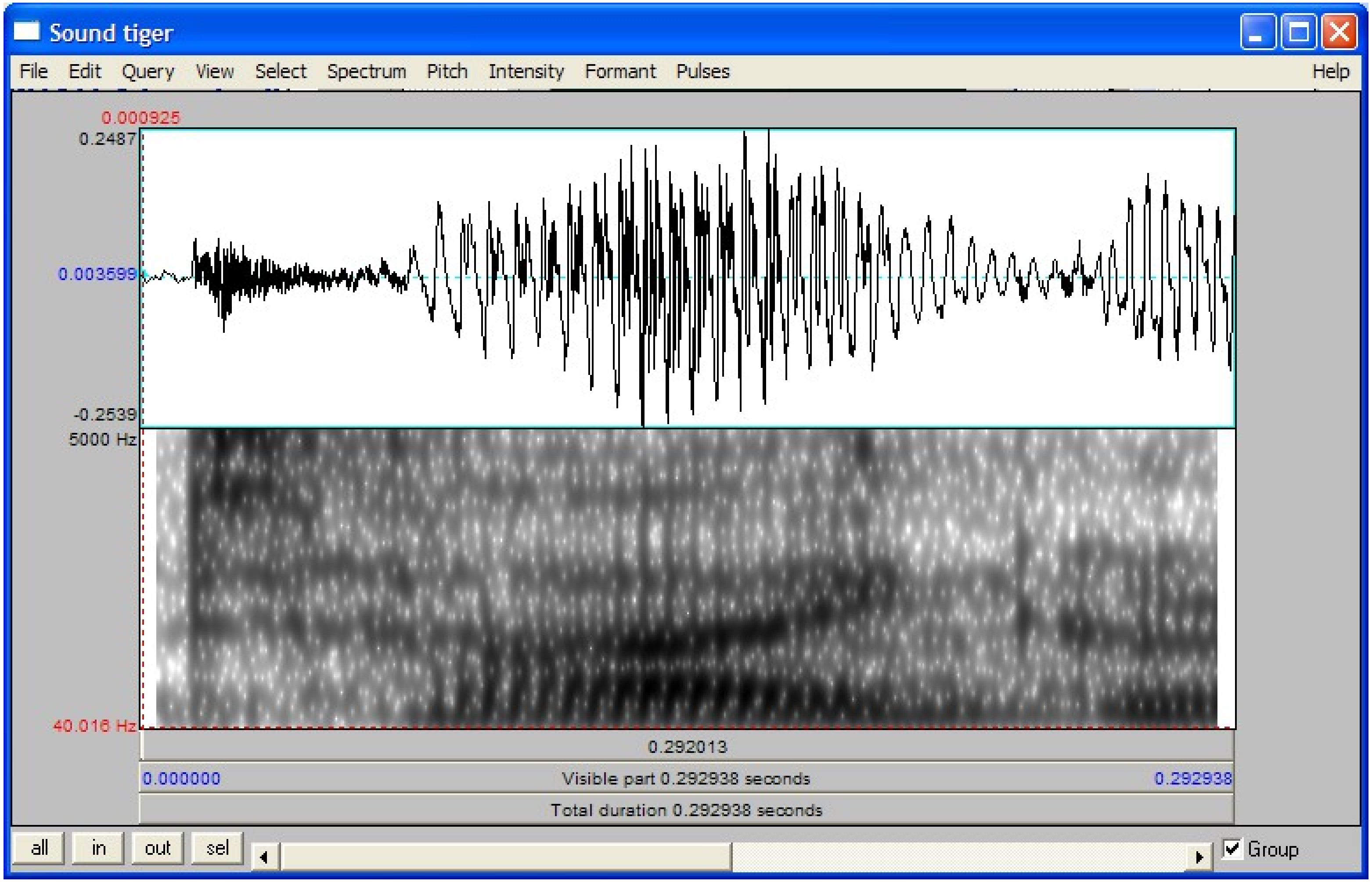


Fourier Analysis: the Spectrogram

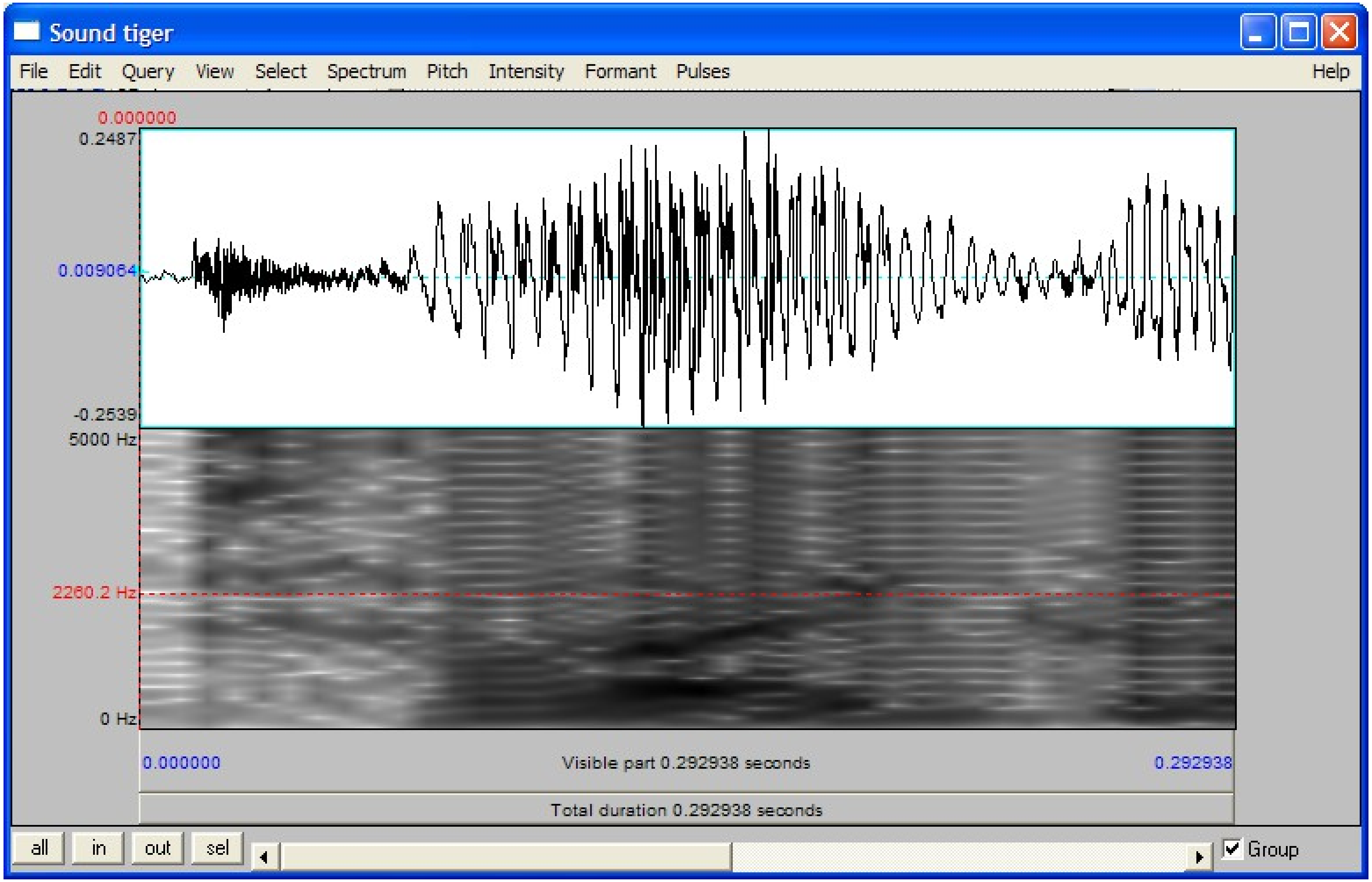
- A single spectral analysis of part of a speech signal, yielding a spectrum, requires a sequence of several periods:
  - In order to track the changing structure of a speech signal, a sequence of spectra is needed.
  - A representation of a sequence of spectra is called a spectrogram



Broad band spectrogram



Narrow band spectrogram

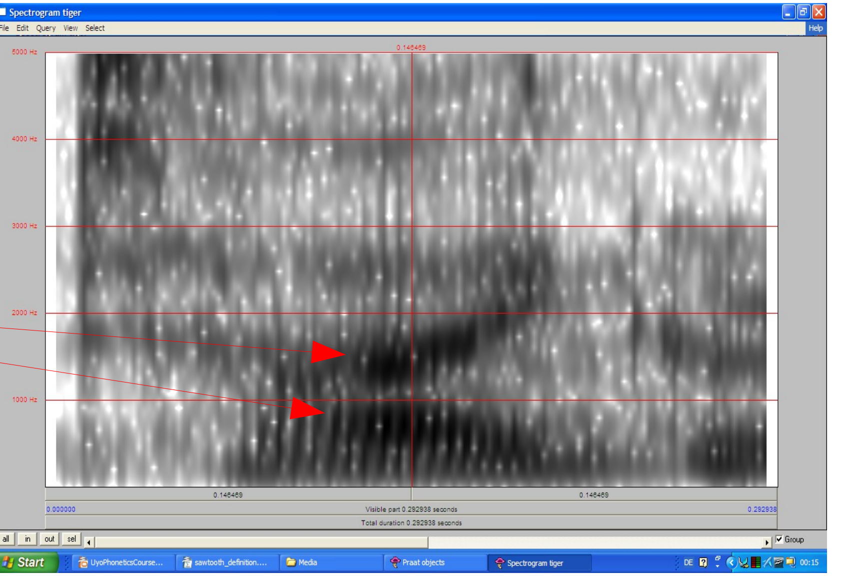


Spectrogram Filtering: Formants

- The FILTER which modifies the SOURCE signal consists of the pharyngeal, nasal and oral cavities. Formants are frequency bands in a spectrogram which differ in intensity from other frequency bands
  - harmonics in these areas are differ in strength
  - formants sonorant sounds (vowels, liquids, nasals, approximants)

tiger

VOWEL FORMANTS

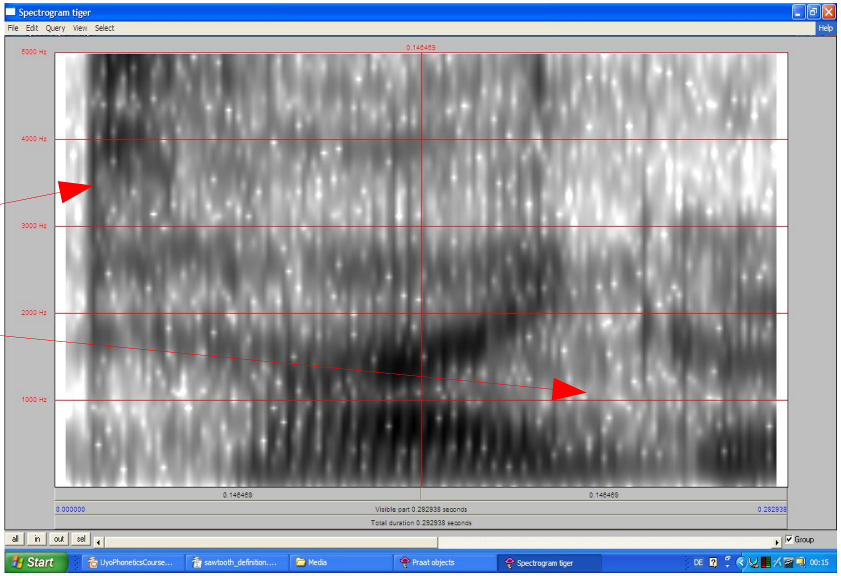


Spectrogram Filtering: Consonantal Noise

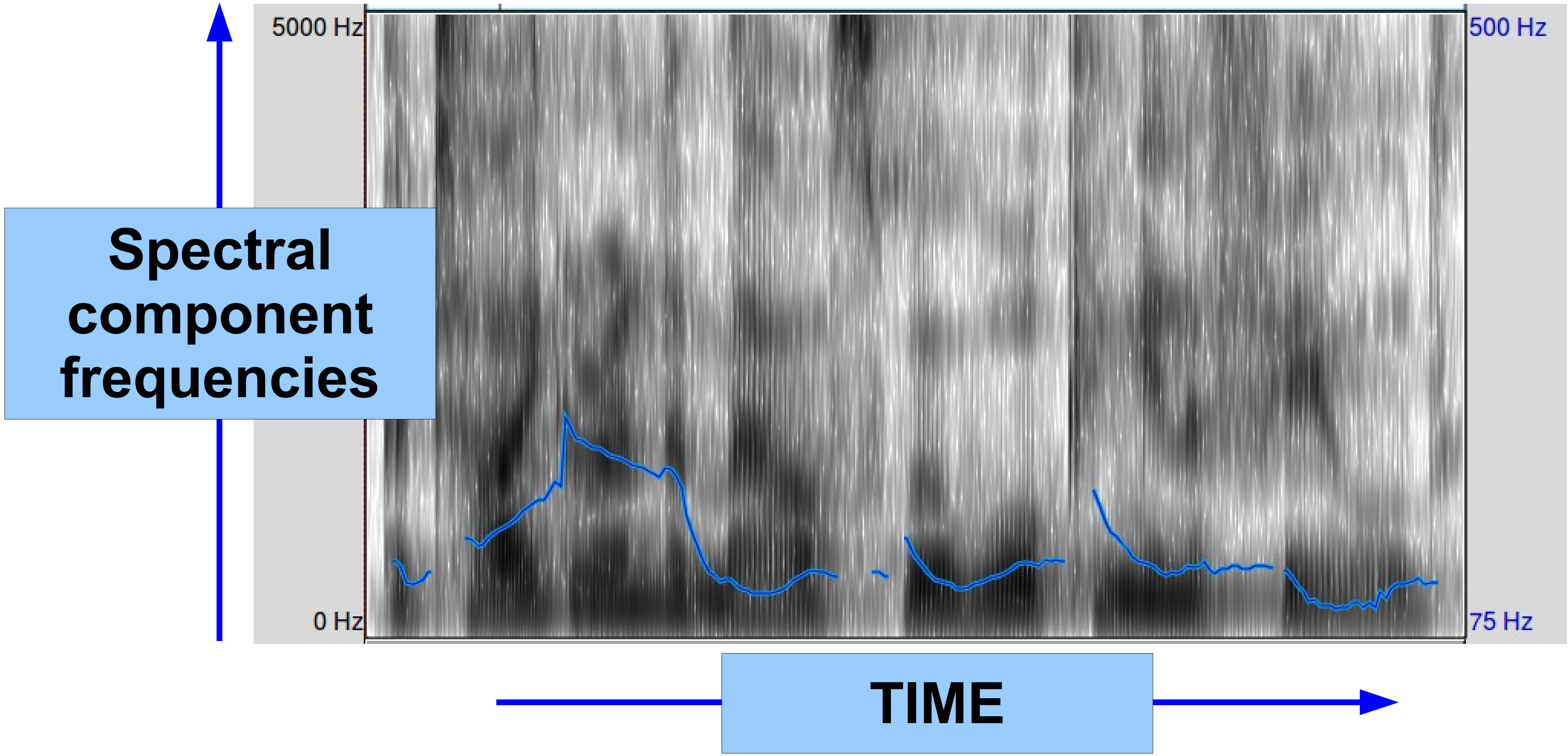
- Obstruent consonants involve
  - obstruction in the oral tract which causes noise
    - stops: closure of (oral and nasal) tracts, followed by noise burst
    - fricatives: near-closure of oral tract (and closure of nasal tract) causing noise

tiger

CONSONANTS

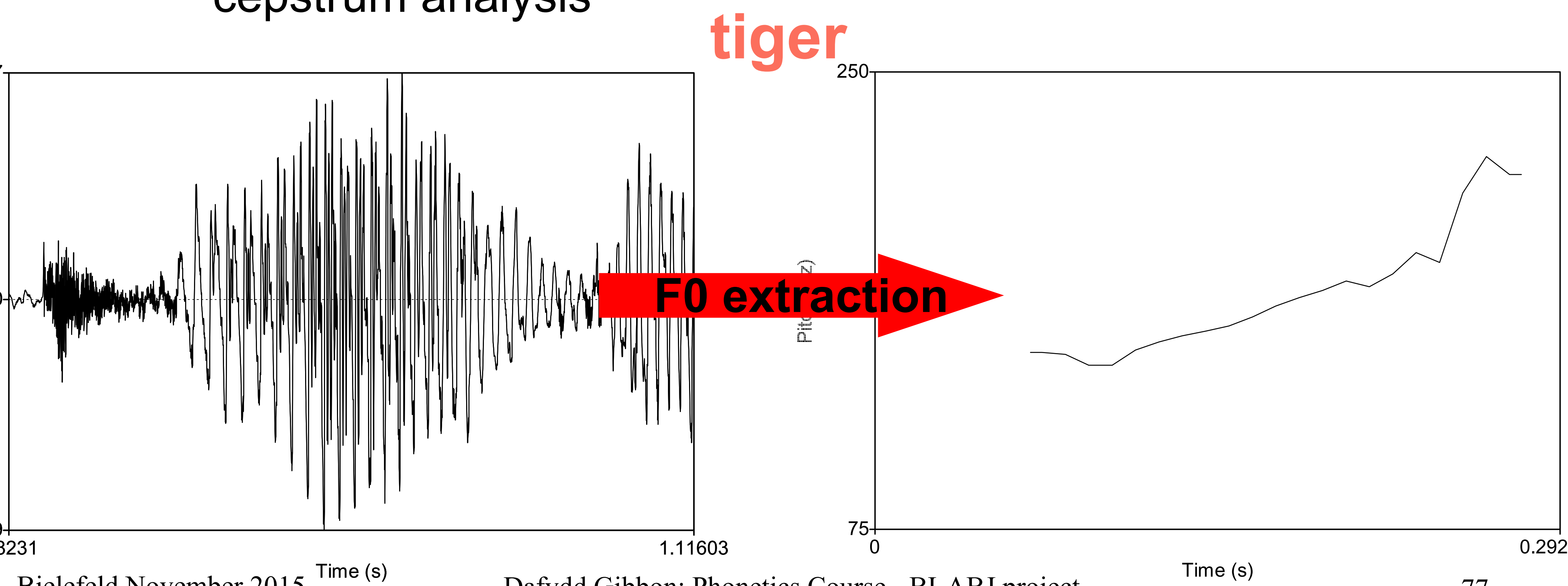


Basic Speech Signal Parameters



Pitch extraction

- Separation of F0 from harmonics is *pitch extraction*
- Methods of pitch extraction are:
  - counting zero-crossings in the same direction
  - counting peaks in the signal
  - auto-correlation
  - cepstrum analysis

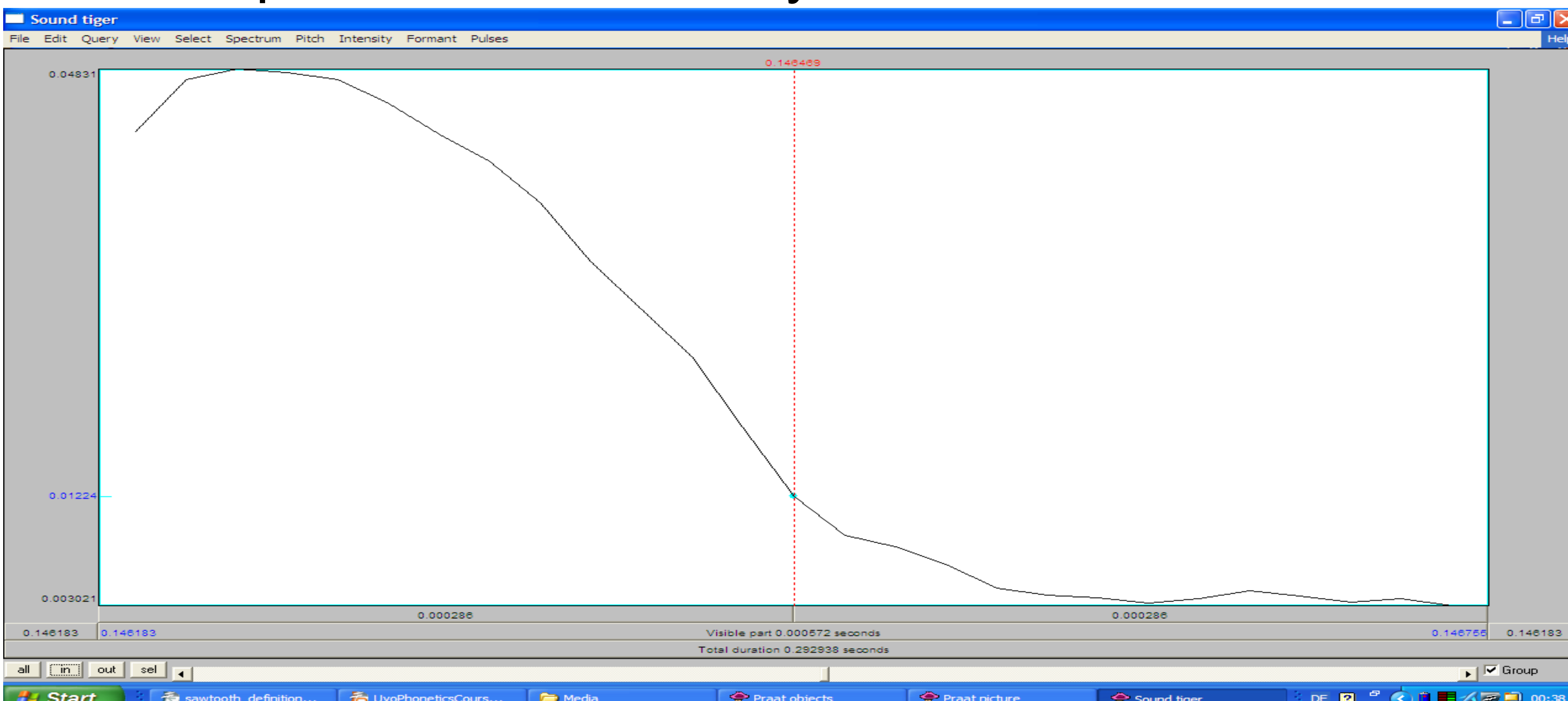


Analog-to-Digital (A/D) Conversion

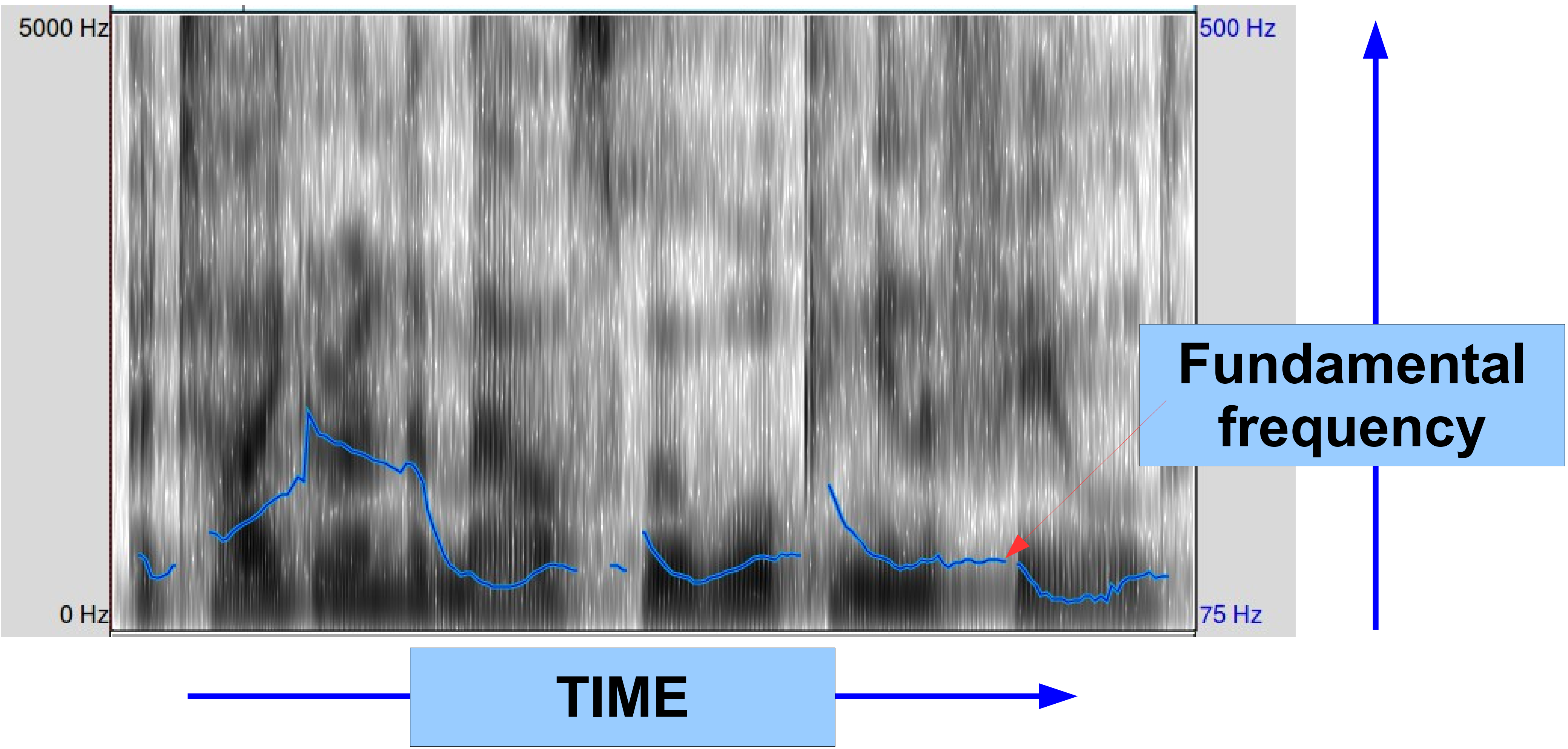
- In order to enter a speech signal into a computer it is digitised:
  – the signal is sampled regularly and the amplitude of the sample is measured automatically
  – the speed with which the measurements are made is called the *sampling rate*
  – standard sampling rates are:
    - 44.1 kHz (CDs) = 2 x 2 x 3 x 3 x 5 x 5 x 7 x 7
    - 48 kHz (DAT tapes)
    - 22.05 kHz (laboratory recordings)
    - .... (other sampling rates are also found)
- The minimum sampling rate is twice the frequency of the highest harmonic in the signal (Nyquist theorem), otherwise false measurements are made and “aliasing” occurs (ghost frequencies)

Analog-to-Digital (A/D) Conversion

- The corners in the visualisation represent measuring points
- The measuring points are joined by straight lines to give an impression of continuity



Basic Speech Signal Parameters

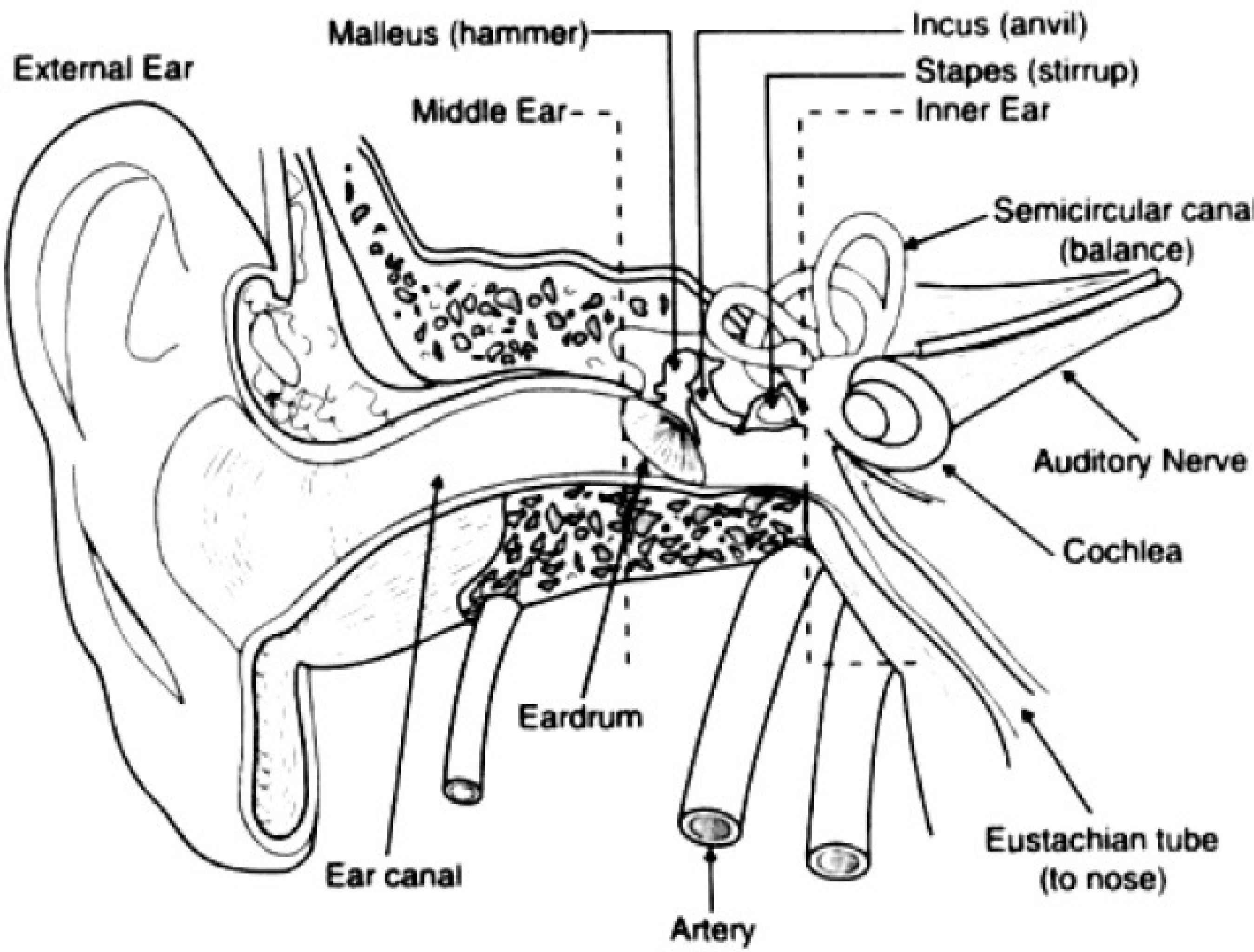


Quiz on Acoustic Phonetics

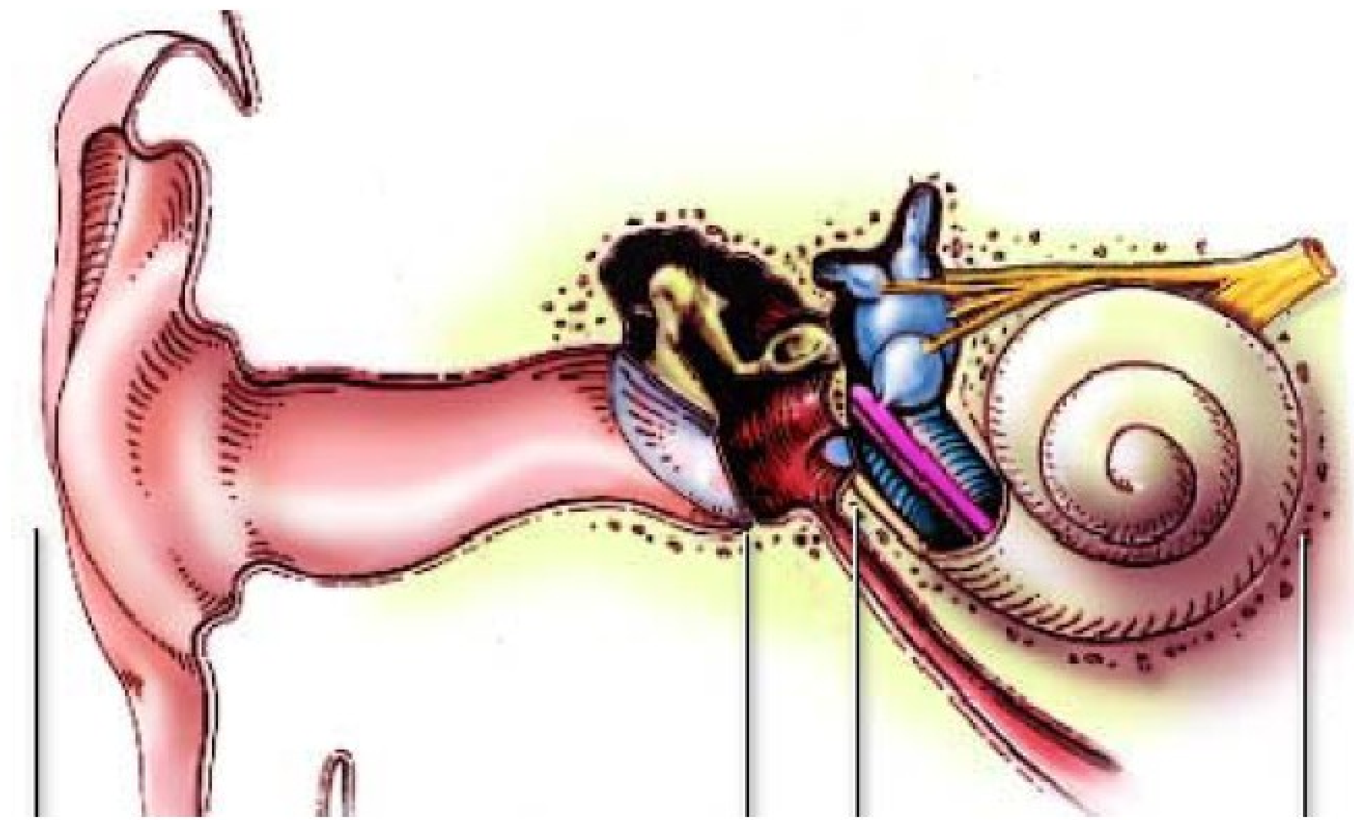
- What are the basic parameters of the speech signal?
- Define the following terms:
  - amplitude
  - intensity
  - energy
- How are time-domain representations of speech signal converted to frequency domain representations?
- Define the following terms:
  - spectrum
  - spectrogram
  - fundamental frequency, F0, pitch
  - harmonic
  - formant
  - analog-to-digital conversion

Auditory Phonetics (Speech Perception)

The Auditory Domain: Anatomy of the Ear



The Auditory Domain: Anatomy of the Ear



outer ear

inner ear

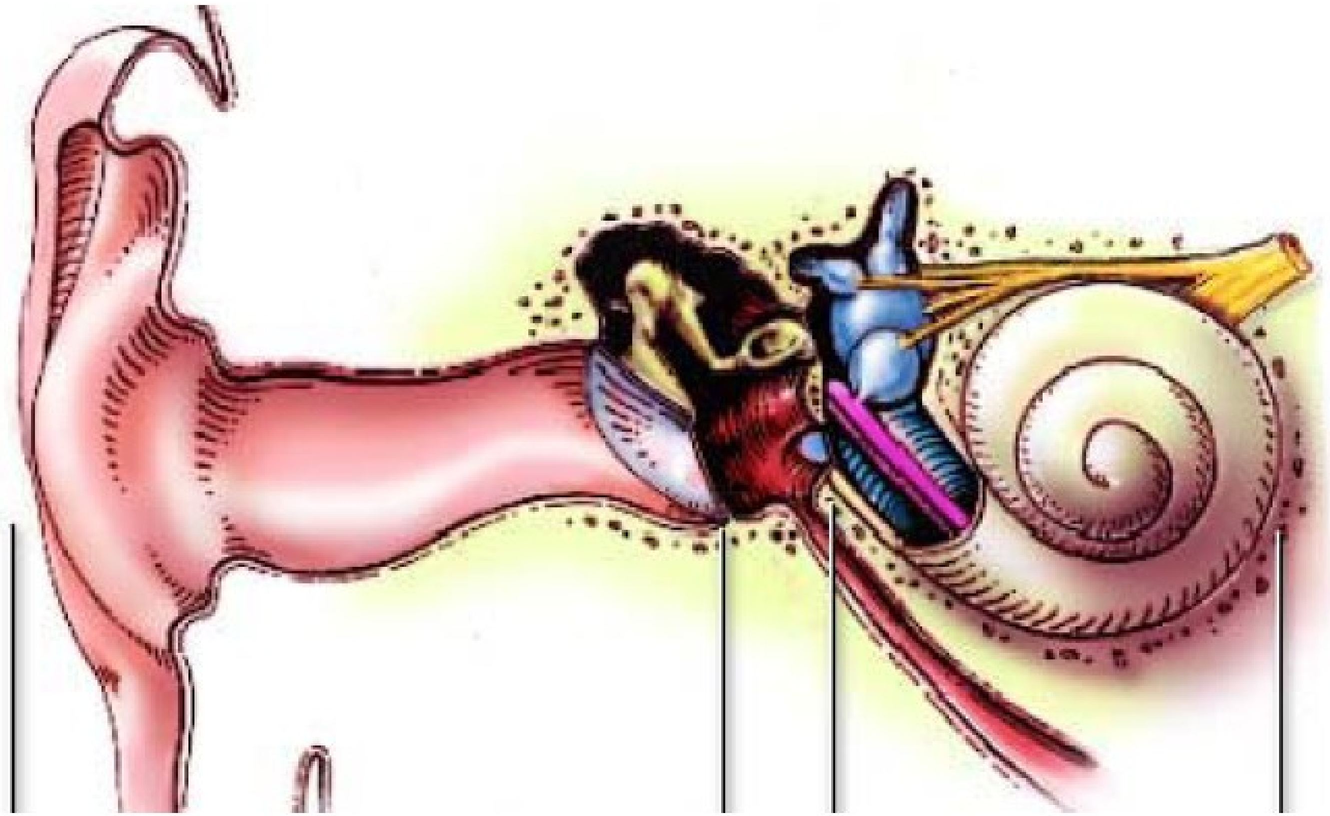
middle ear

The Auditory Domain: Anatomy of the Ear

microphone

amplifier

Fourier transform



outer ear

middle ear

inner ear

Quiz on Auditory Phonetics

- What are the functions of
  - the outer ear?
  - the middle ear?
  - the inner ear?
- What are
  - the ossicles?
  - the oval window?
  - the cochlea?
  - the basilar membrane?

Final Remarks

After the first unit

- you should have learned the basic theoretical foundations on which phonetic activities with Praat are based
- you should be able to use a Praat TextGrid file with the TGA online timing analysis tool

After the second unit

- you should thoroughly understand what you are doing with Praat

Phonetics 3

Scripting with Praat

Dafydd Gibbon

Bielefeld, November 2015

Overview

- What is Scripting?
- Scripting languages
- Praat scripting
- Example of Praat Scripting (1): The Task
- Example of Praat Scripting (2): The History
- Praat Scripting in Language Documentation
- More Information on Praat Scripting
- Praat Scripting Tutorial
- “My First Script”
- Praat Scripting Introduction
- Summary

What is Scripting?

- In the computer world of Documentary Linguistics, a *script* is
  - a list of actions to be performed by a computer
  - to solve a particular problem, such as
    - automatically editing a text
    - extracting words from a text to make a dictionary or concordance
    - calculating statistics about frequencies of words in texts
    - normalising annotation formats
    - doing statistical analyses of durations of segments in annotations
    - ...
  - therefore a kind of “to do list” for a computer
    - and a very basic kind of computer program

Scripting languages

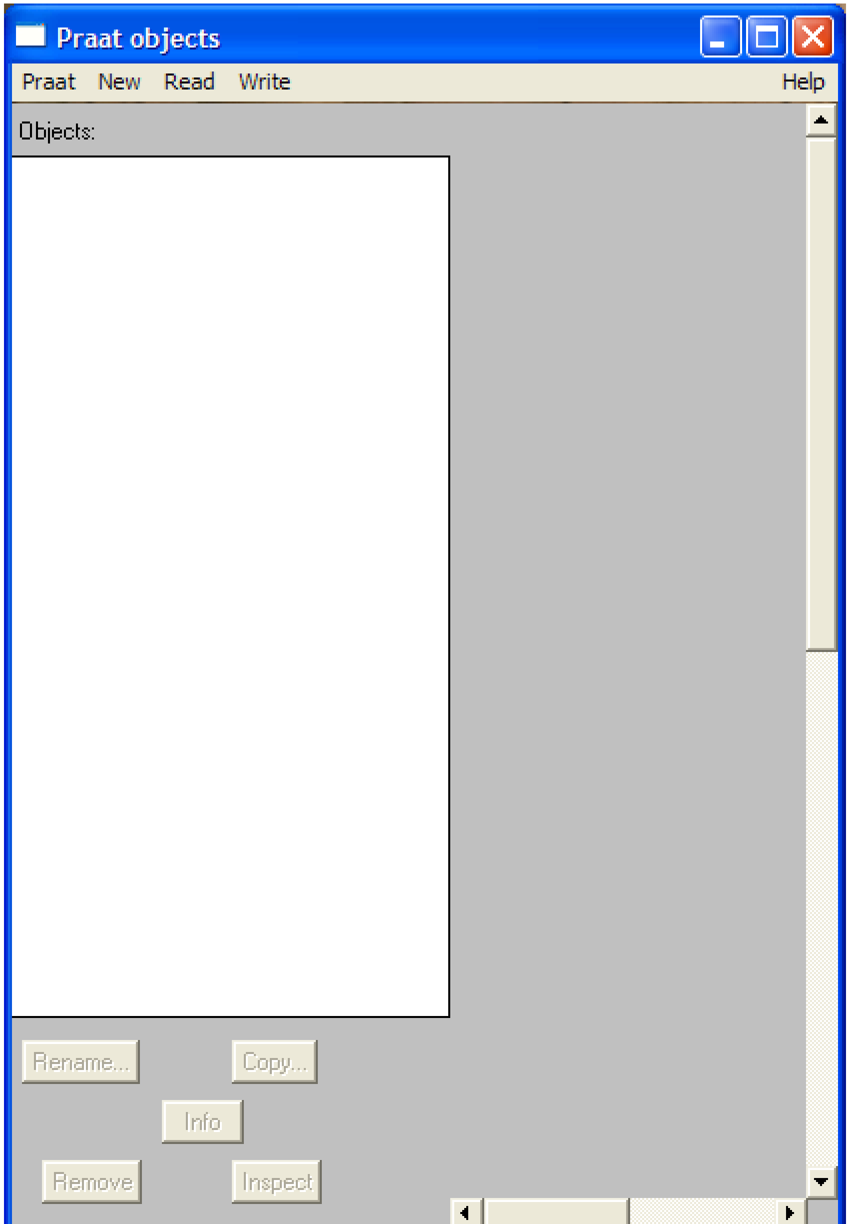
- Examples of scripting languages are
  - general scripting languages such as
    - UNIX shell
    - awk
    - Perl
    - Python
  - special scripting languages such as
    - Praat scripting language (for performing complex and time-consuming operations in Praat automatically)
- An example of a Unix shell script:
  - simple editing for use in format conversions:

```
#!/bin/sh
sed “
    s/I:/I/g
    s/o[^:]/O/g
    “
```

Praat scripting

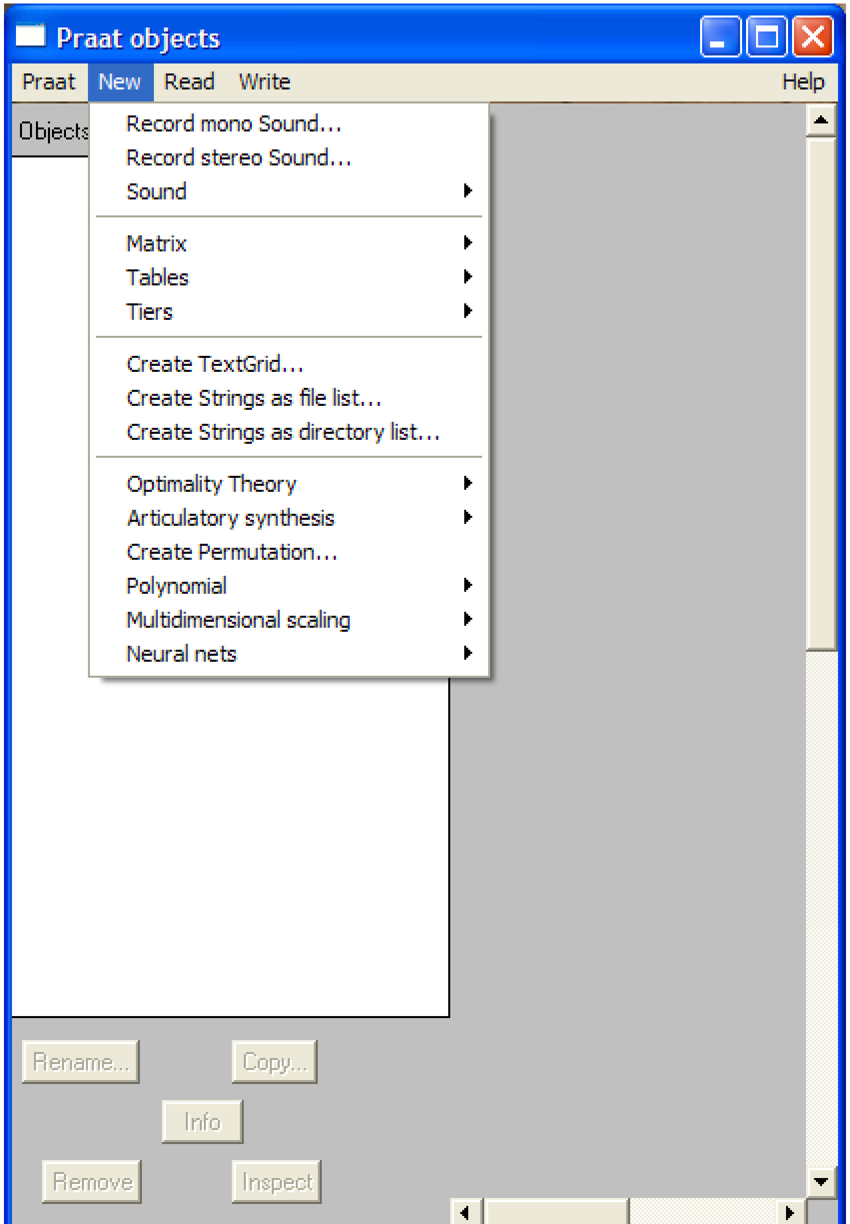
- The Praat program contains a scripting language with which sequences of operations with Praat can be reproduced automatically.
- There are two ways of producing a Praat script:
  - History mechanism:
    - Tracking and recording manual operations, which are then written into a script file which can be run / executed later.
    - Whenever Praat buttons are pressed,
      - a Praat operation is performed
      - the operation is recorded in the “History”
    - The history can then be pasted into the Script Editor and the sequence will be performed automatically
  - Script editor:
    - Writing the file manually, which presupposes familiarity with the Praat scripting language.

Example of Praat Scripting (1): The Task



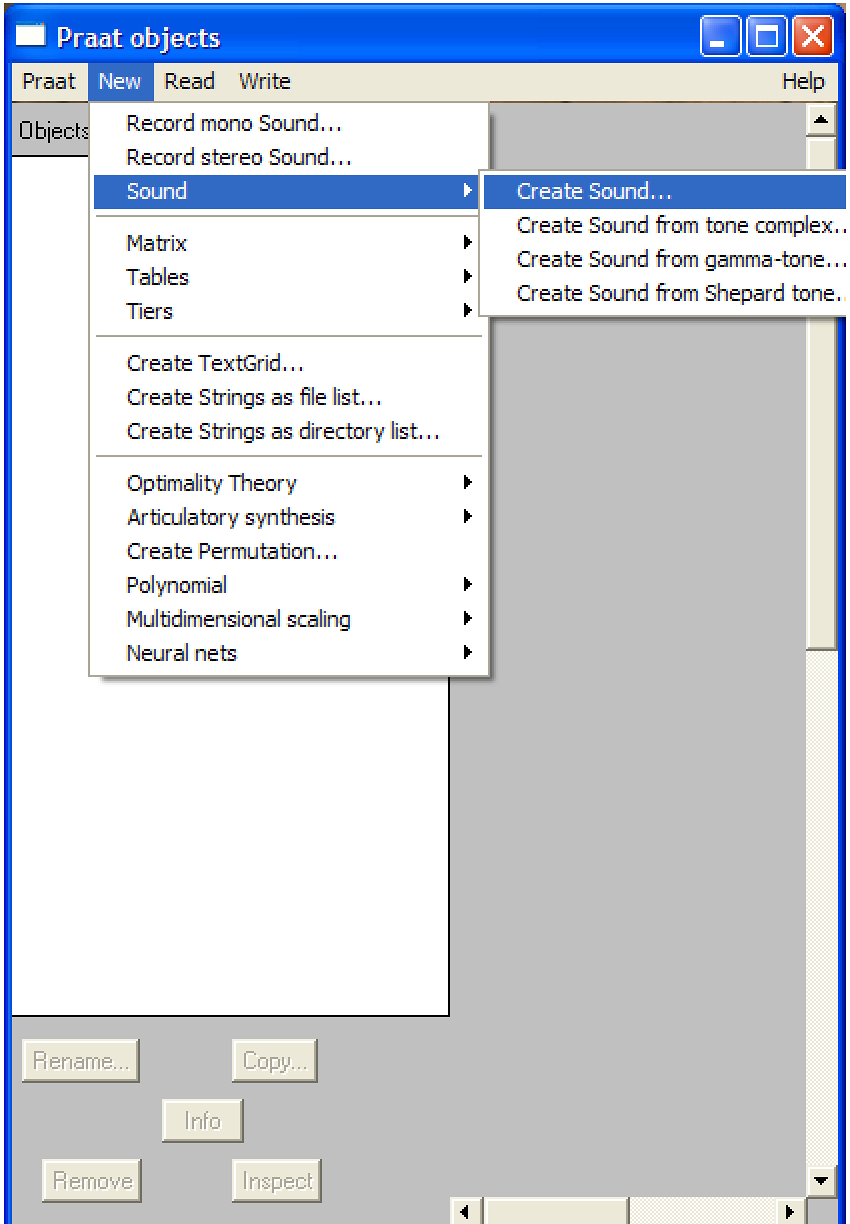
- To generate a sine wave tone

Example of Praat Scripting (1): The Actions



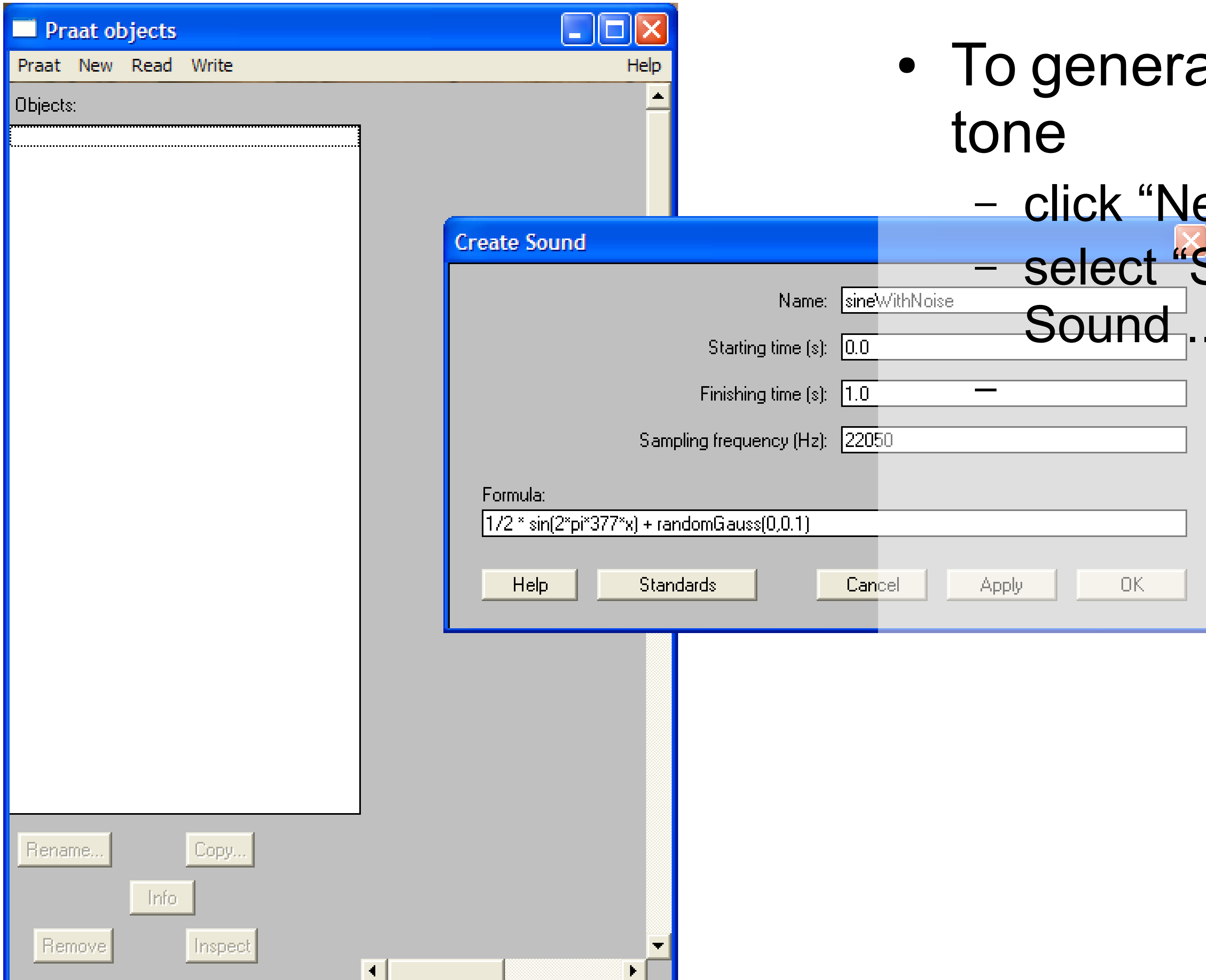
- To generate a sine wave tone
  - click “New”

Example of Praat Scripting (1): The Actions



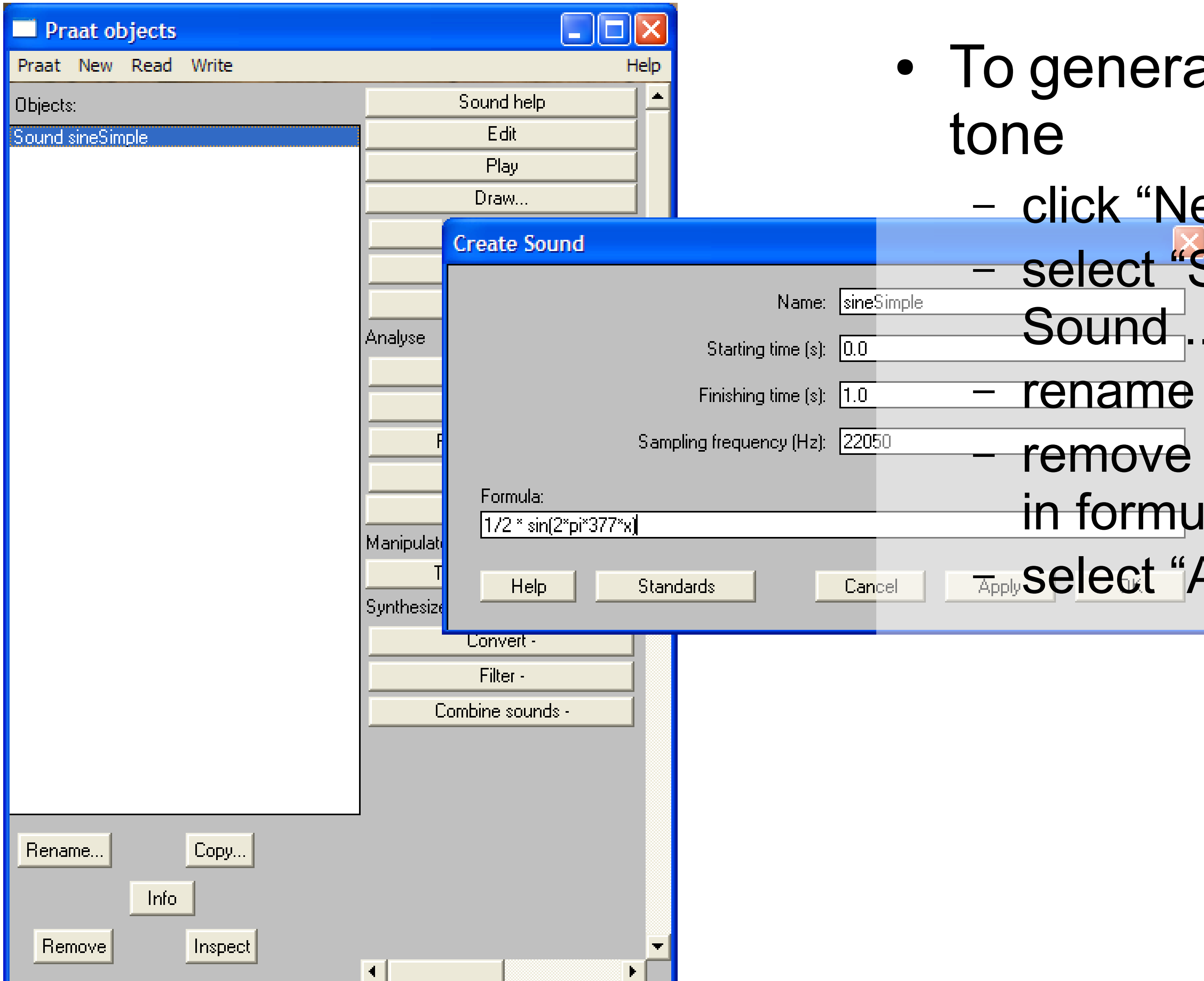
- To generate a sine wave tone
  - click “New”
  - select “Sound > Create Sound ...”
  -

Example of Praat Scripting (1): The Actions



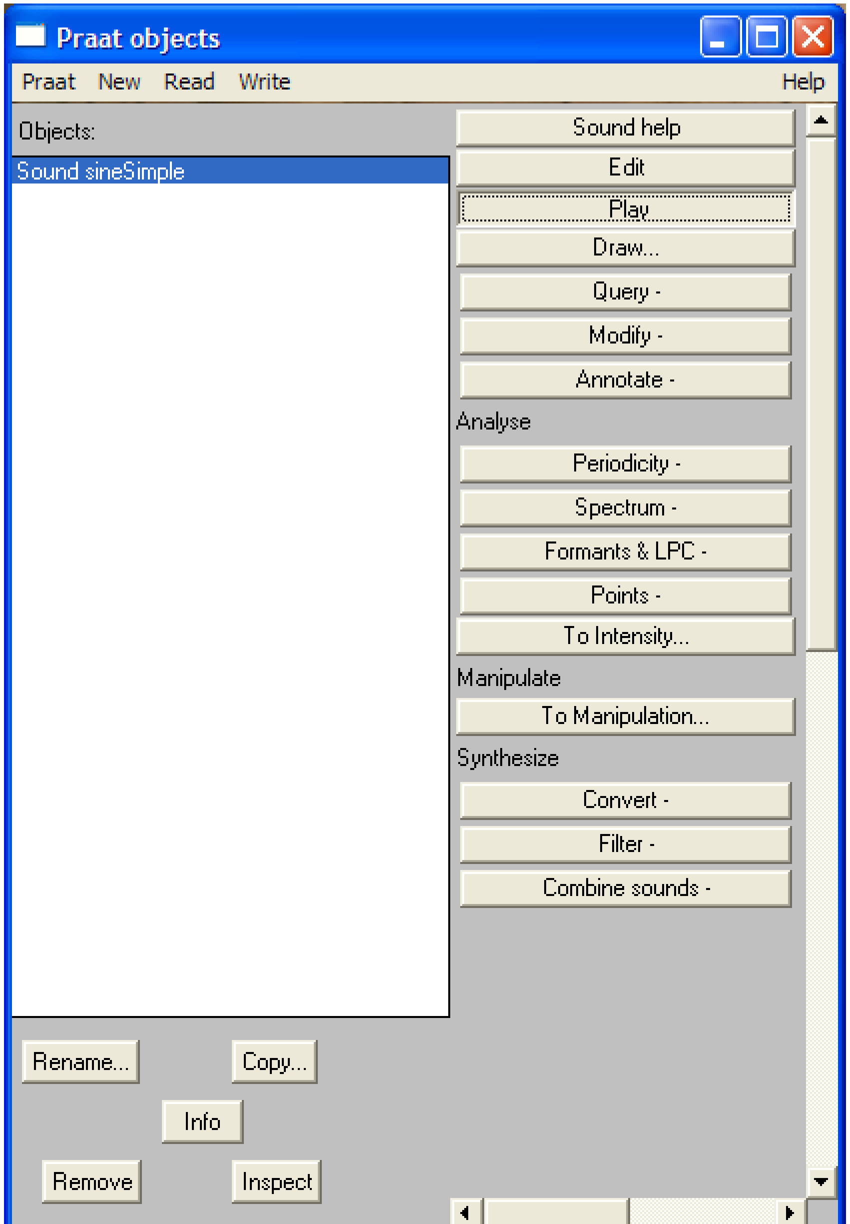
- To generate a sine wave tone
  - click “New”
  - select “Sound > Create Sound ...”

Example of Praat Scripting (1): The Actions



- To generate a sine wave tone
  - click "New"
  - select "Sound > Create Sound ..."
  - rename "soundSimple"
  - remove Gaussian element in formula
  - select "Apply"

Example of Praat Scripting (1): The Actions



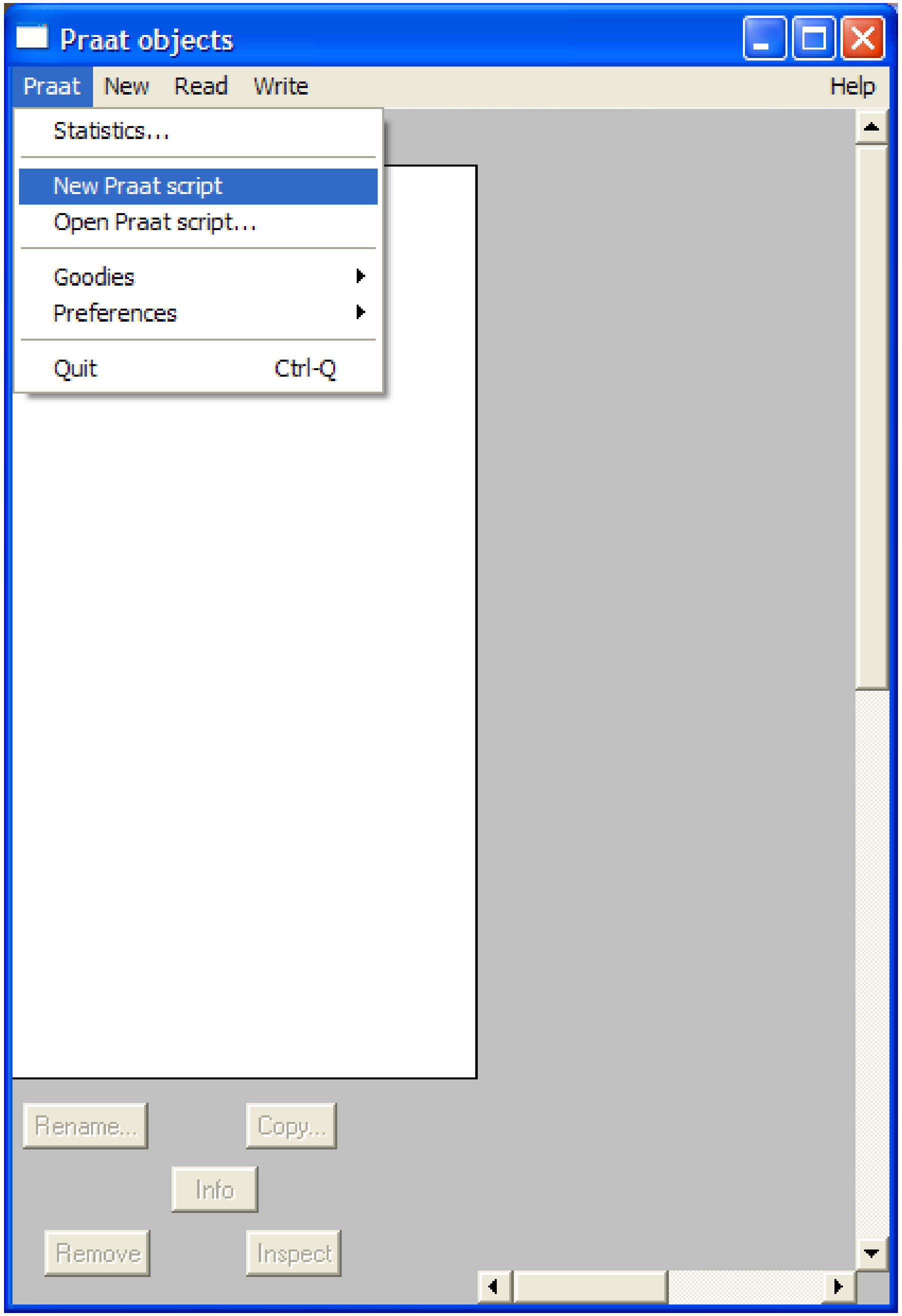
- To generate a sine wave tone
  - click “New”
  - select “Sound > Create Sound ...”
  - rename “soundSimple”
  - remove Gaussian noise element in formula
  - select “Ok”
  - select “Play”
  - select “Remove”

Example of Praat Scripting (2): The History

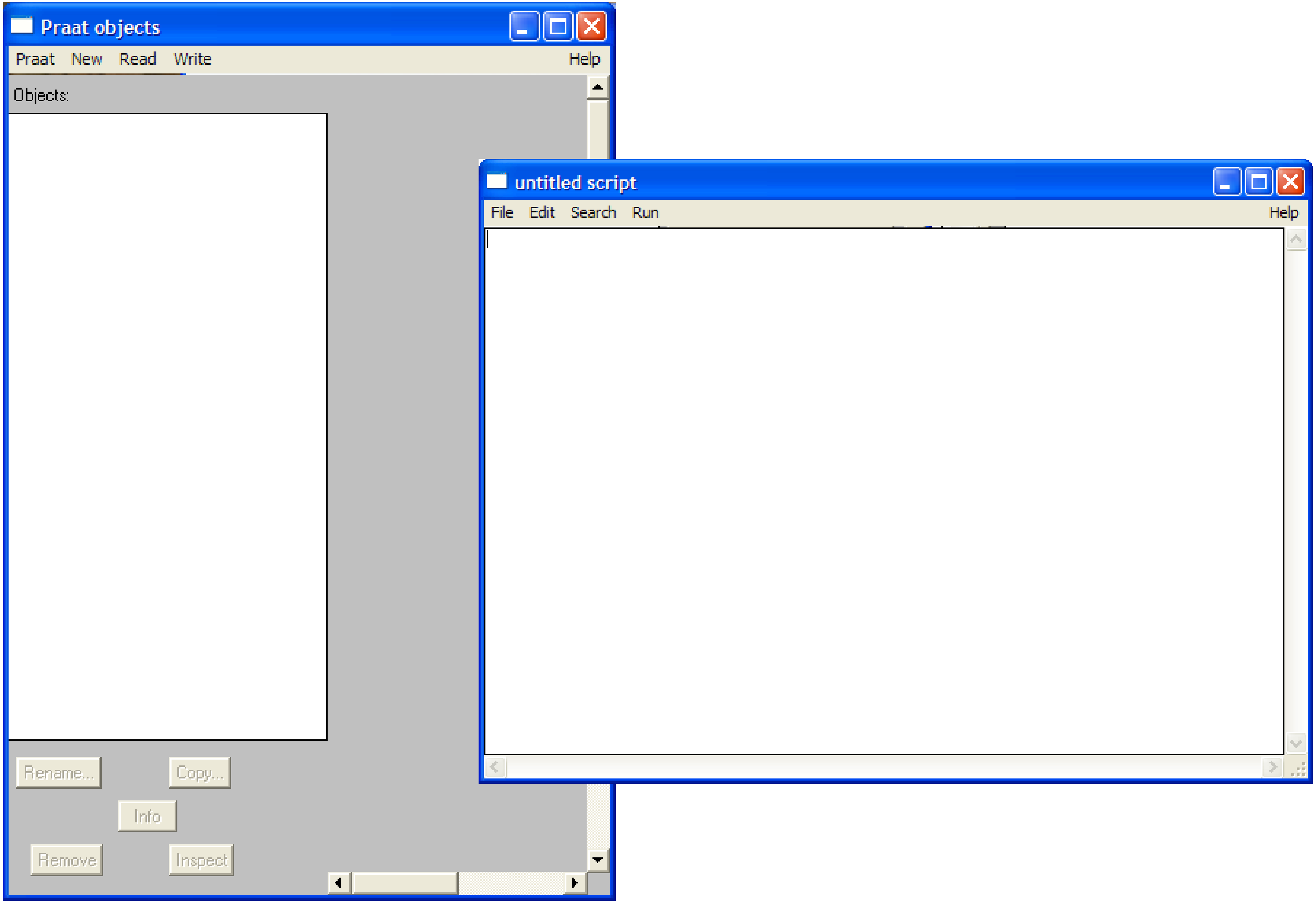
- Whenever a button is selected
  - which causes a selection or an action
  - Praat records the selection or action in a *log* or *history*
- The recorded selection or action
  - can be copied into an editor
- The result is...
  - a Praat script
- The script for the preceding set of actions is...

```
Create Sound... sineSimple 0 1 22050 1/2 * sin(2*pi*377*x)
Play
Remove
```

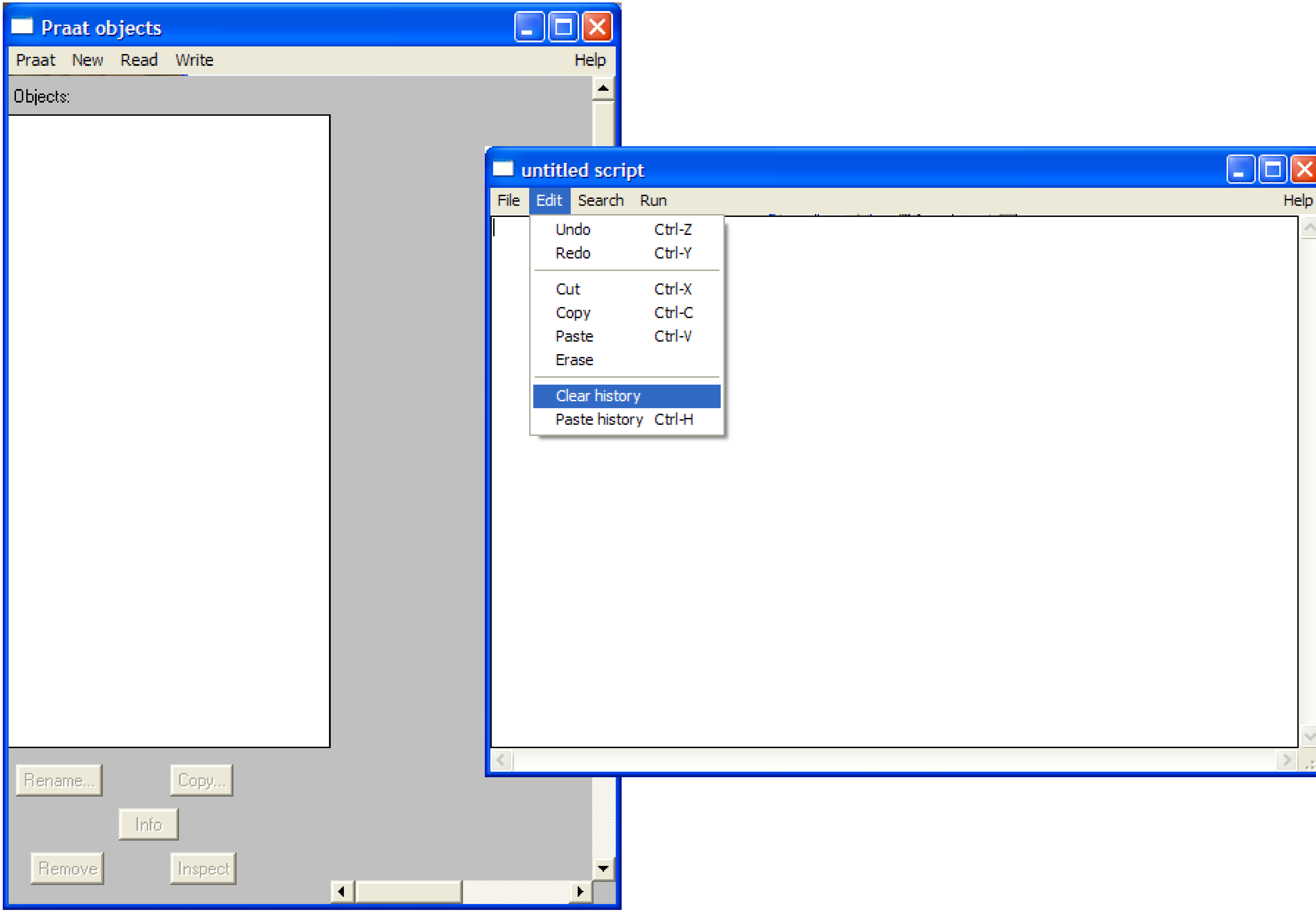
Example of Praat Scripting (2): The History



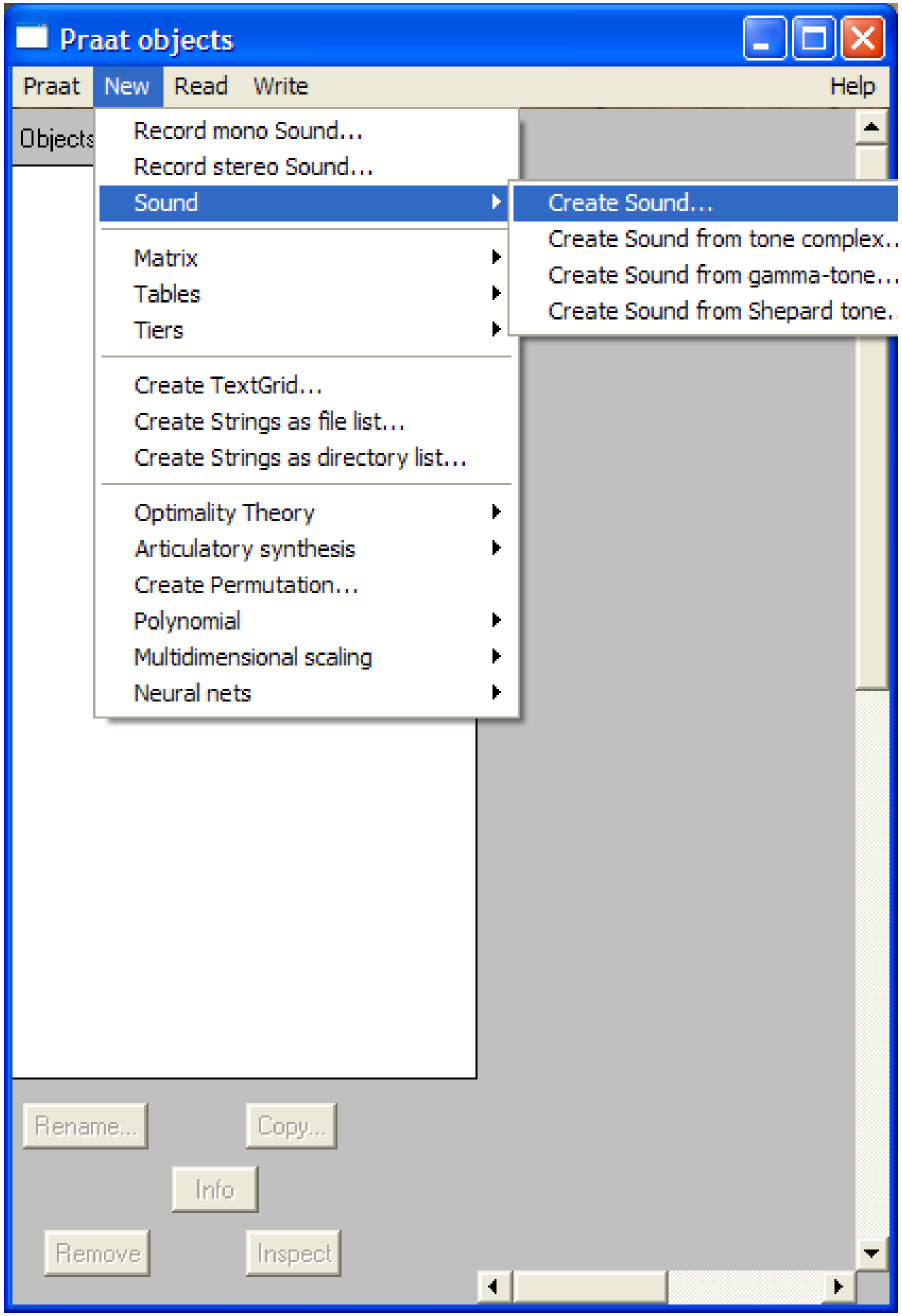
Example of Praat Scripting (2): The History



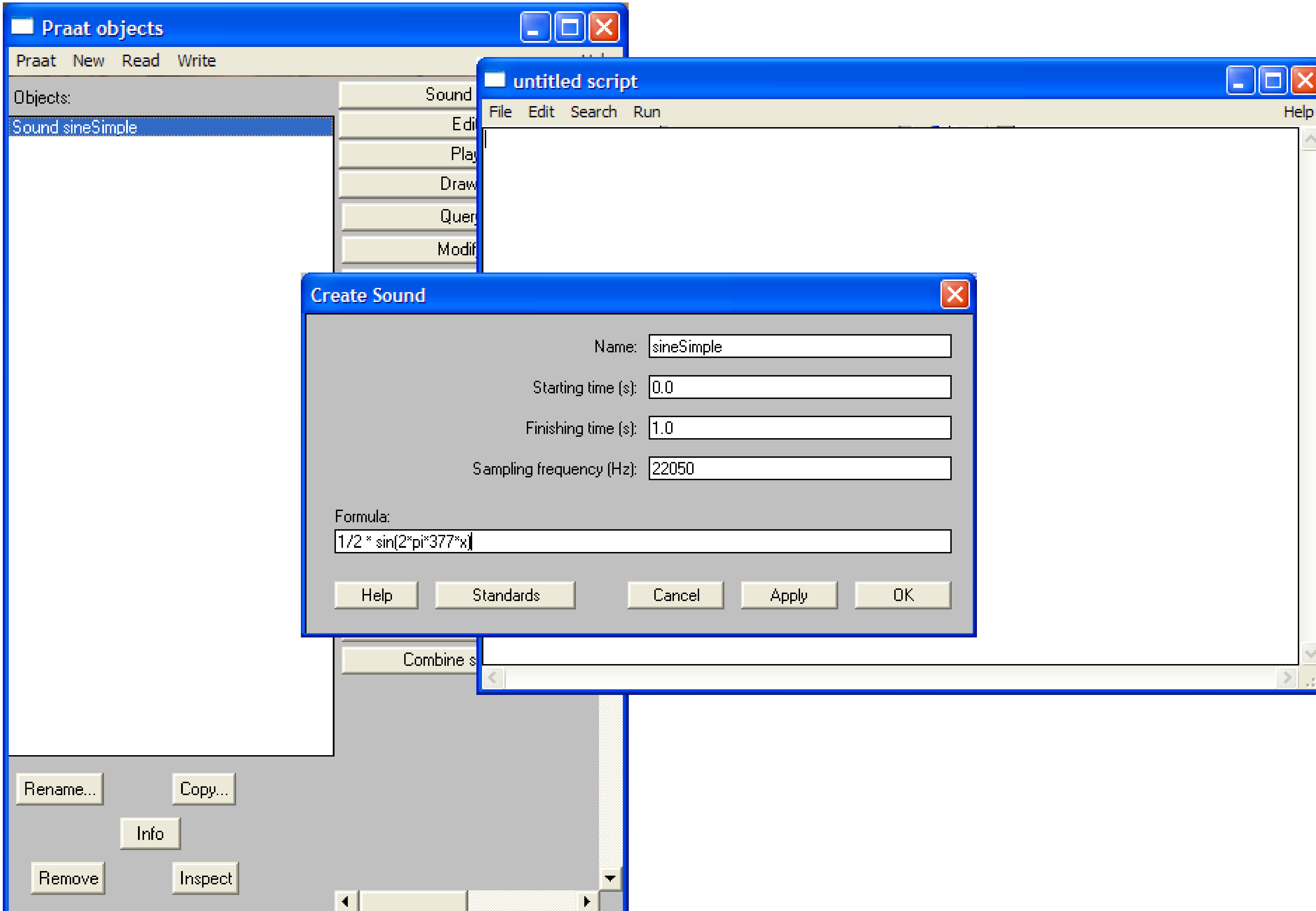
Example of Praat Scripting (2): The History



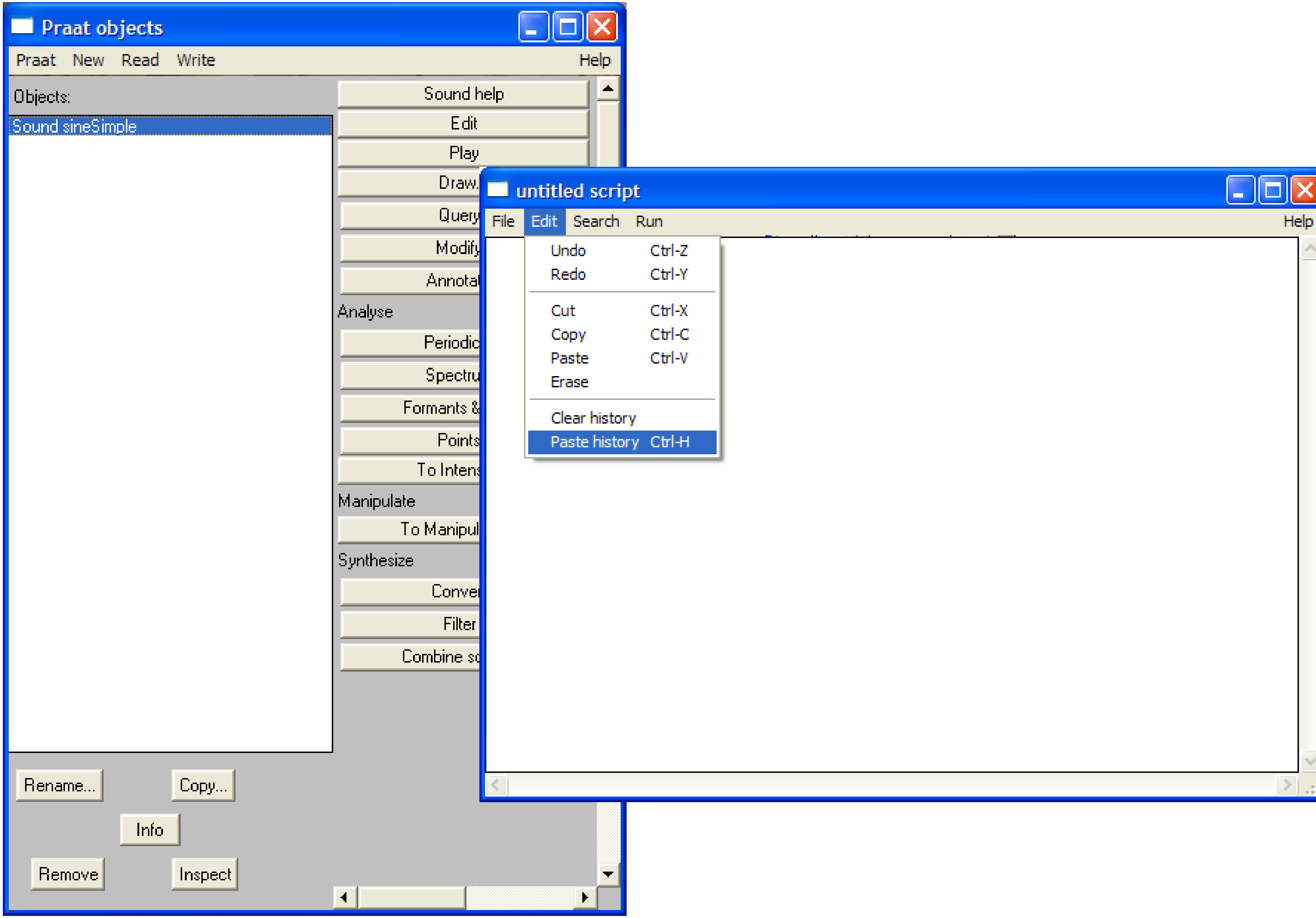
Example of Praat Scripting (2): The History



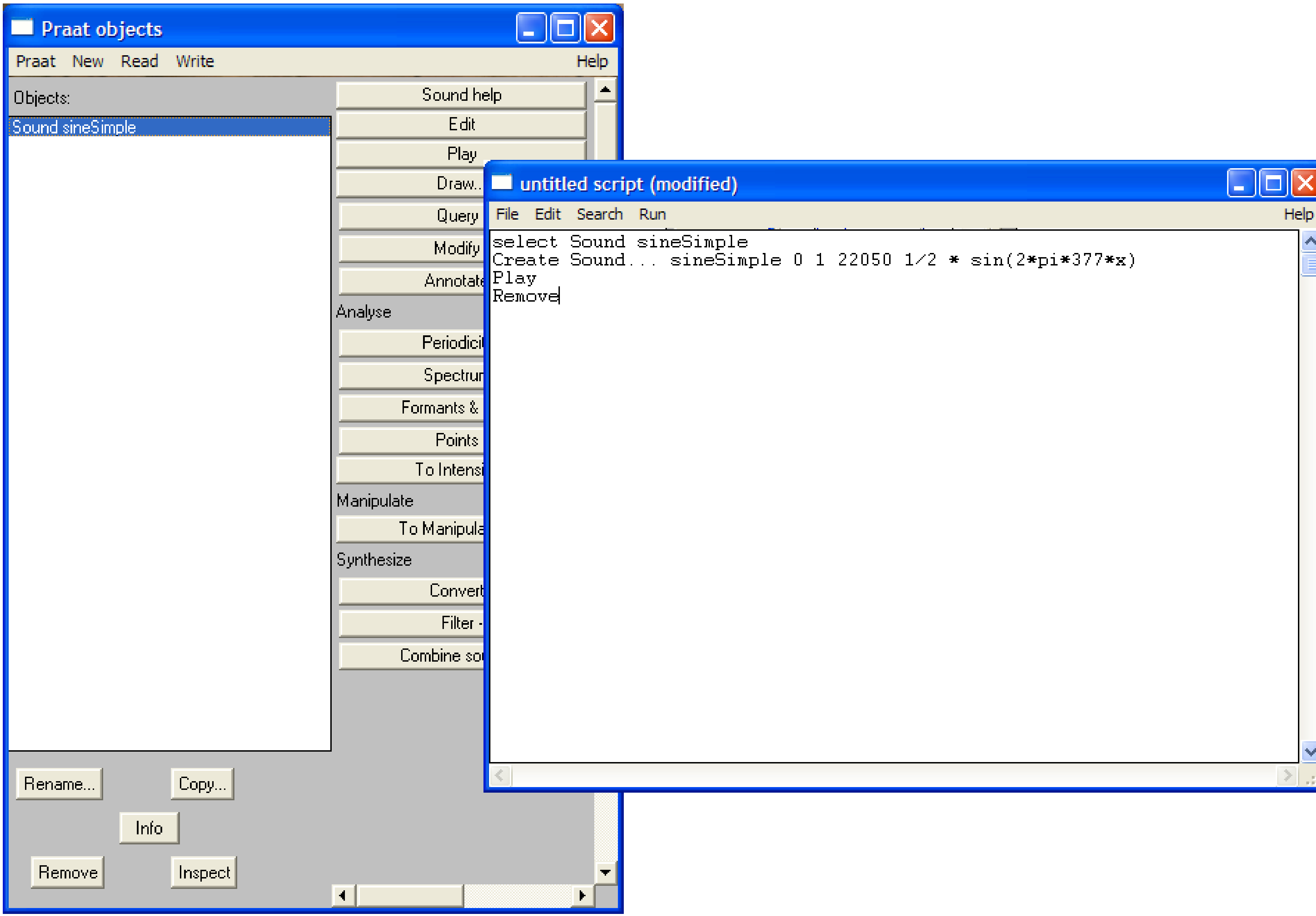
Example of Praat Scripting (2): The History



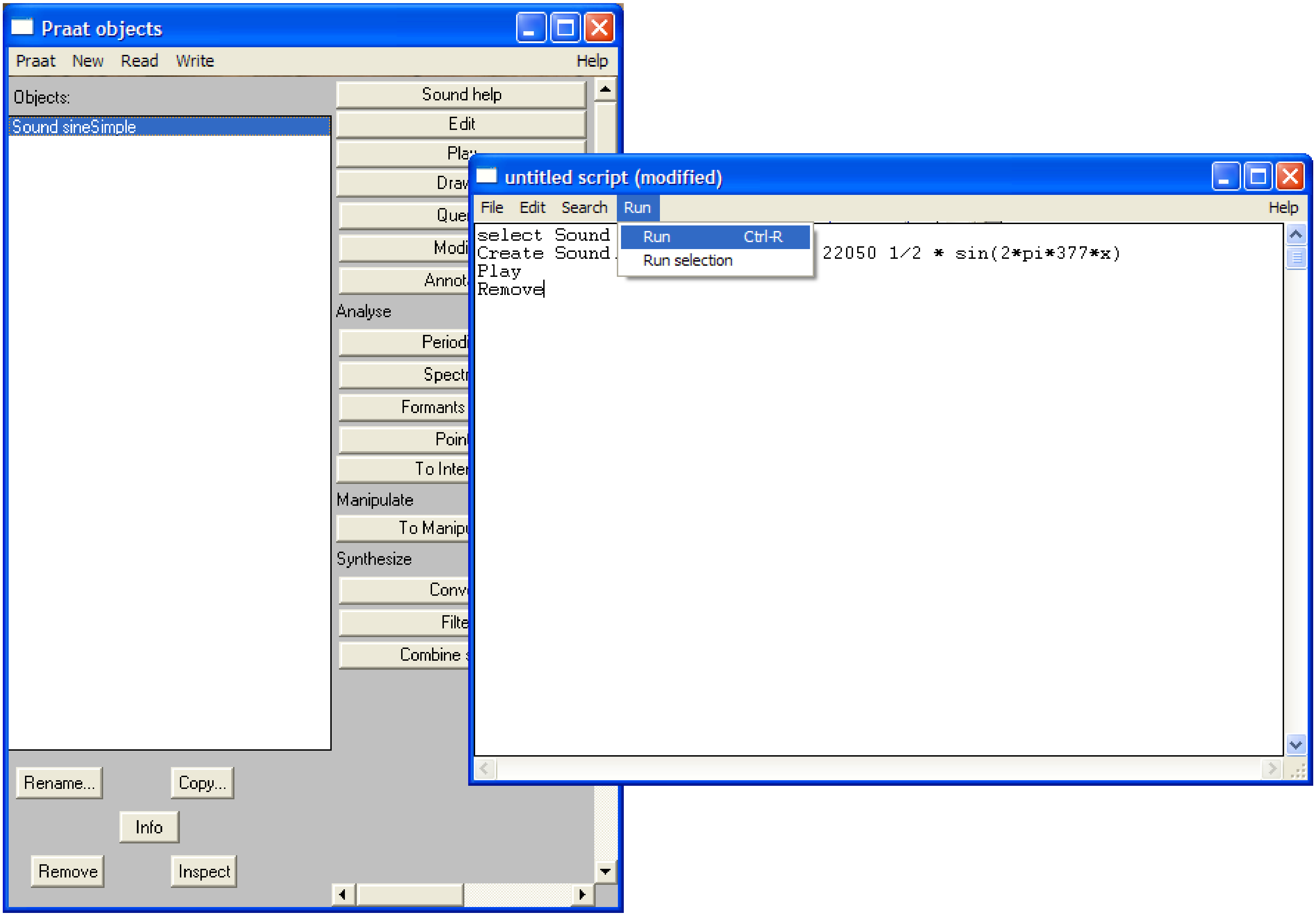
Example of Praat Scripting (2): The History



Example of Praat Scripting (2): The History



Example of Praat Scripting (2): The History



Praat Scripting in Language Documentation

- Example: create graphics files from sound files
  - procedure:
    - open a new scripting window
    - erase history
    - perform the actions

Praat Scripting in Language Documentation

- Example: create graphics files from sound files
  - procedure:
    - open a new scripting window
    - erase history
    - perform the actions
      1. load a sound file into Praat
      2. draw the oscillogram for the sound file
      3. save the oscillogram to Windows metafile, *.emf file
      4. run the script

Praat Scripting in Language Documentation

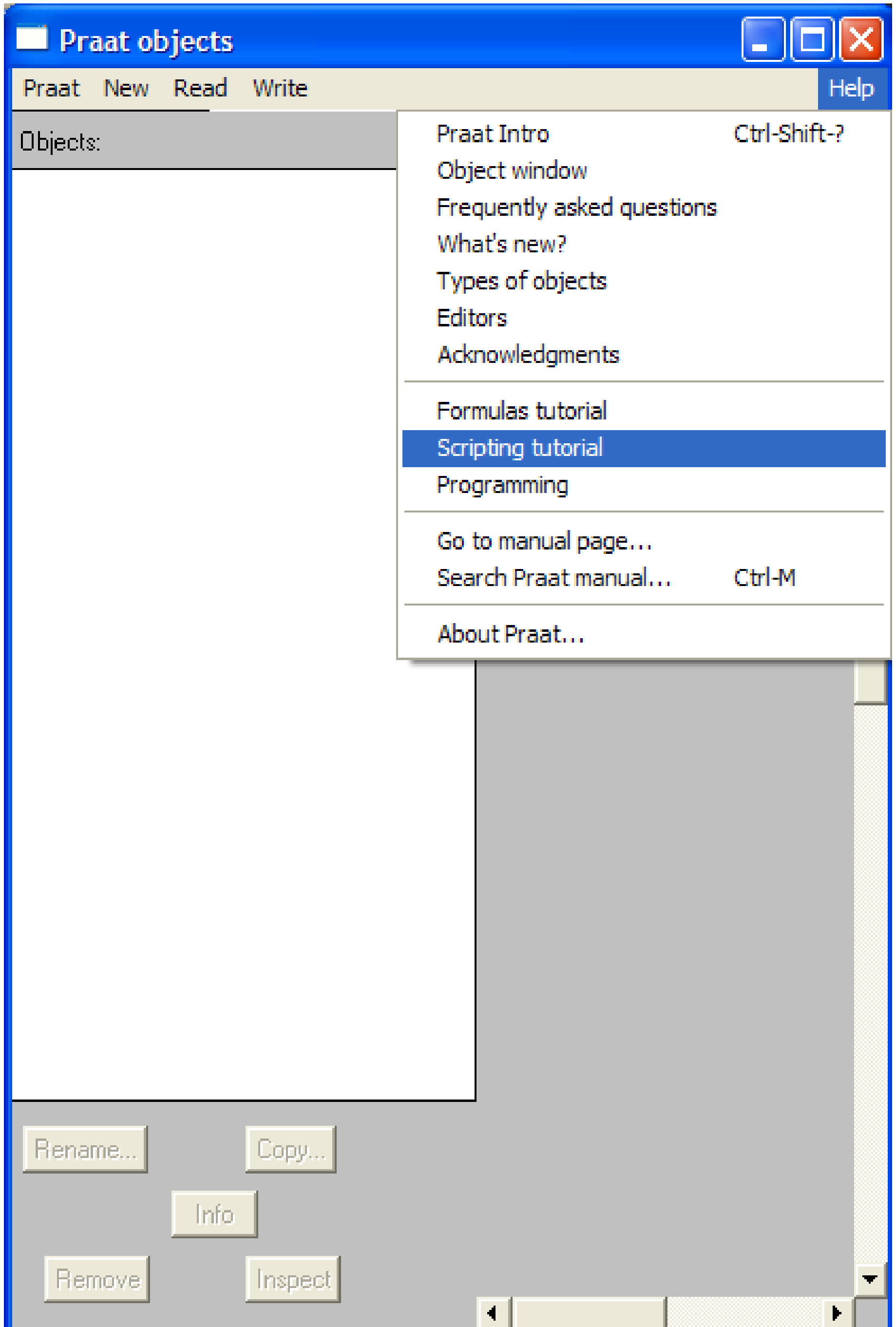
- Example: create graphics files from sound files
  - procedure:
    - open a new scripting window
    - erase history
    - perform the actions
      1. load a sound file into Praat
      2. draw the oscillogram for the sound file
      3. save the oscillogram to Windows metafile, *.emf file
      4. run the script
  - script:

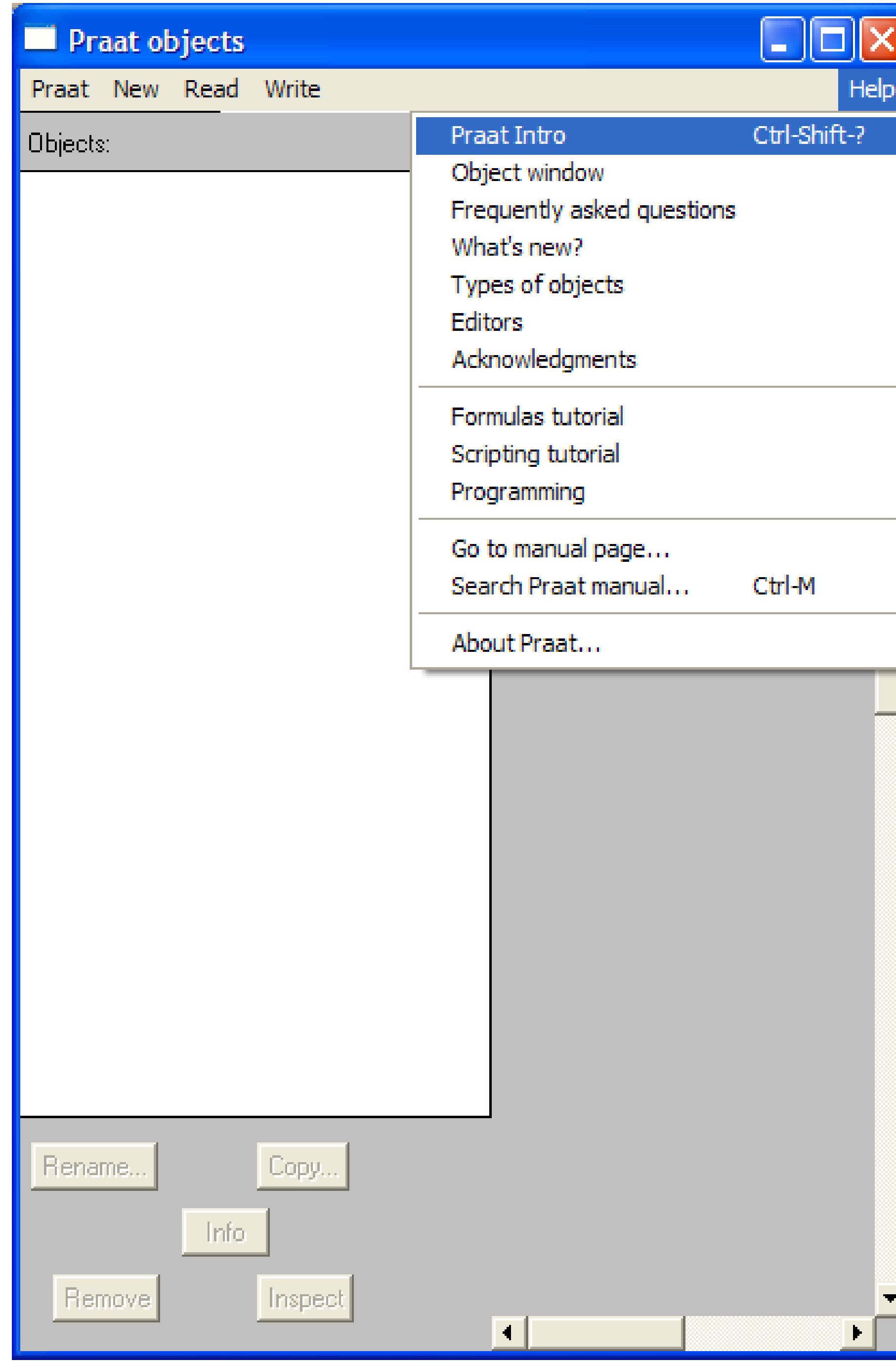
```
Read from file... tiger.wav
Play
Draw... 0 0 0 0 yes curve 0 0 0 0 yes curve
Write to Windows metafile... tiger_osc.emf
```

Praat Scripting for Power Users ...

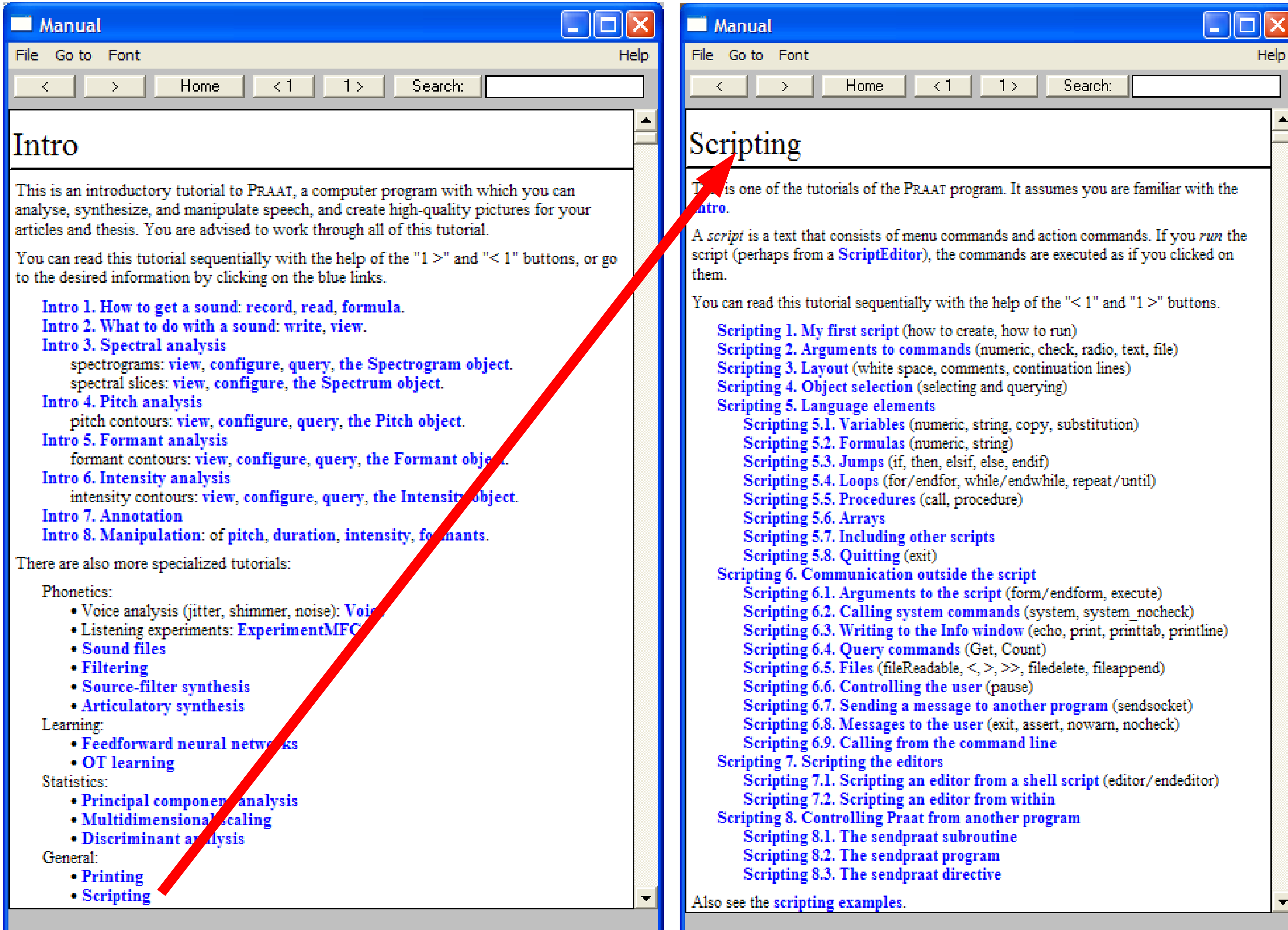
- The Praat Scripting Langage interpreter is very powerful, and allows full programming facilities such as:
  - Operations:
    - strings
    - numbers
  - Loops:
    - for, while, until
  - Conditional branching
  - File operations
  - System calls, calls from system
- We have used Praat scripting to create graphics and cut sound files for an internet audio concordance
  - Thurid Spiess
  - Thorsten Trippel

More Information on Praat Scripting





Praat Scripting Tutorial



"My First Script"

Manual

File Go to Font Help

< > Home < 1 1 > Search:

Scripting 1. My first script

Suppose that you want to create a script that allows you to play a selected Sound object twice. You first create an empty script, by choosing **New Praat script** from the Praat menu. A **ScriptEditor** will appear on your screen. In this editor, you type

```
Play
Play
```

Now select a **Sound** in the object menu. As you expect from selecting a Sound, a **Play** button will appear in the dynamic menu. If you now choose **Run** in the ScriptEditor, the sound will be played twice.

What commands can I put into my script?

In the above example, you could use the "Play" command because that was the text on a button currently available button in the dynamic menu. Apart from these selection-dependent (dynamic) commands, you can also use all fixed commands in the menus of the **Object window** and the **Picture window**. For how to proceed with commands that need *arguments* (i.e. the commands that end in "..." and present a settings window), see **§2**.

Faster ways of getting a script text

Instead of manually typing the command lines, as described above, it may be easier to create a script text from a *macro recording*, with the **History mechanism**. This allows you to create script texts without typing. For instance, if you choose **Clear history** in your ScriptEditor, then click the ***Play*** button twice (after selecting a Sound), and then choose **Paste history**, your new script text contains exactly two lines that read "Play".

To edit a script that is already contained in a file on disk, use **Open Praat script...**.

How to run a script

You can run scripts from the **ScriptEditor**. If you will have to use the script very often, it is advisable to create a button for it in the fixed menu or in a dynamic menu. See the **ScriptEditor** manual page.

On Unix and Windows, you can also run scripts from the command line. See **Scripting 6.9. Calling from the command line**.

Summary of Praat scripting

- It is possible to automatise operations with the Praat scripting language
- Operations which can be sensibly automatised within the context of language documentation include
  - production and storage of visualisations of sound files
  - cutting of sound files into segments using annotations
  - and many others…
- There are very many Praat scripts which are available at different sites on the internet.
- In principle, it is also possible to process annotation files with the Praat scripting language (e.g. extracting word lists, calculating durations, etc.)
  - However, this is best done with a modern general scripting language like Python

Phonetics 4

Using annotations to analyse speech timing

Dafydd Gibbon

Bielefeld, November 2015

Prosody and the architecture of language

Rank	*Functions: Prosodic Hierarchy*
Grammar – compositionality	
DISCOURSE	*discourse intonation/rhythm*
TEXT	*text intonation/rhythm*
SENTENCE CLAUSE PHRASE	*sentence, clause, phrase intonation/rhythm*
COMPOUND WORD DERIVED WORD LEXICAL ROOT MORPHEME	*morphological tone/accent/stress* *in inflection & derivation*
(MORPHO)PHONEME	*phonological segment/tone/accent/stress*

LEXICON – holistic properties, opacity

Prosody and the architecture of language

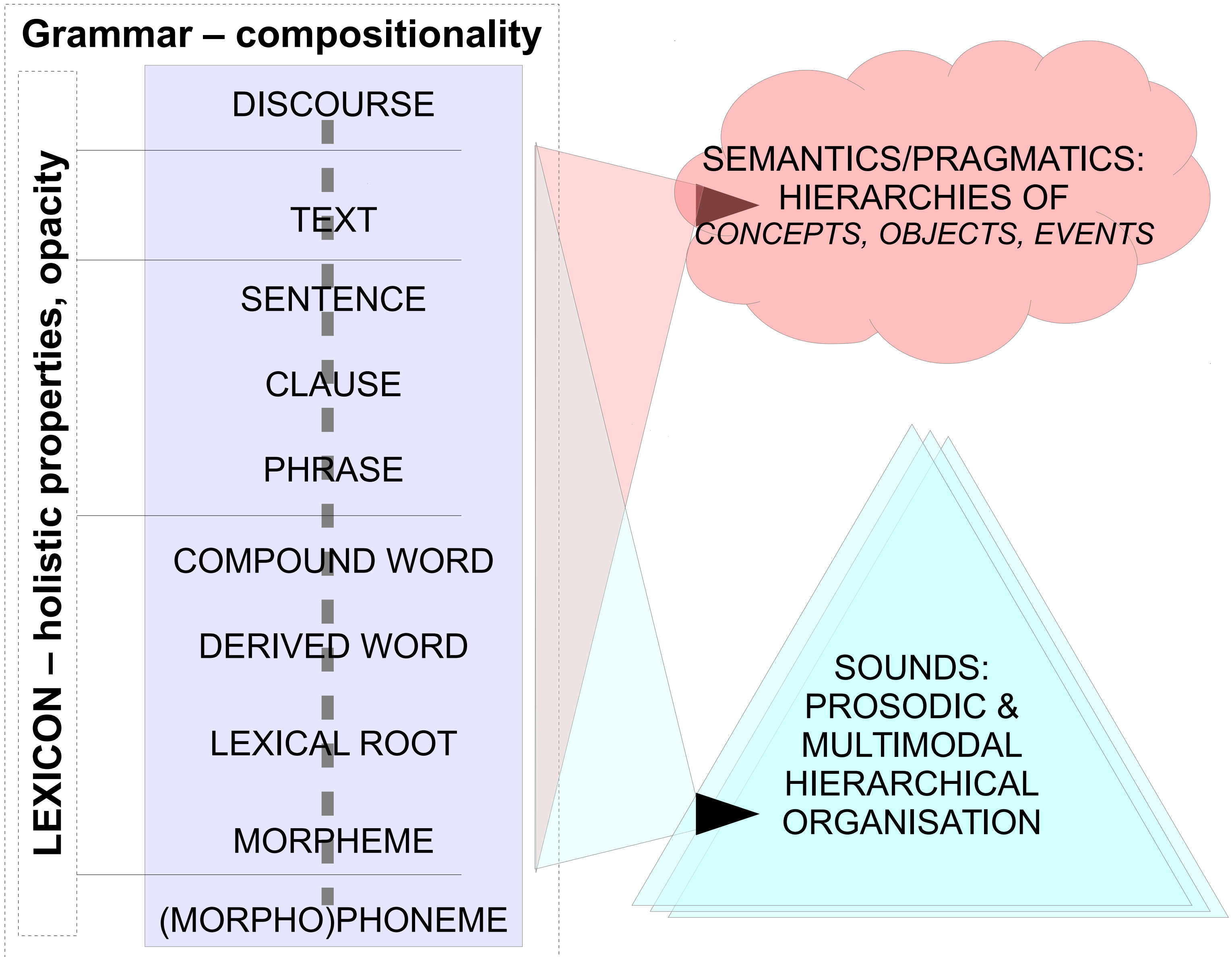
Rank	*Sounds: Phonetic Interpretation*
DISCOURSE	*discourse sounds: intonation/rhythm*
TEXT	*utterance sounds: intonation/rhythm*
SENTENCE CLAUSE PHRASE	*sentence, clause, phrase sounds: intonation/rhythm*
COMPOUND WORD DERIVED WORD LEXICAL ROOT MORPHEME	*sounds of inflection and derivation and word formation:* *phonetics of tone/accent/stress*
(MORPHO)PHONEME	*phonetics of segment/tone/accent/stress*

Grammar – compositionality

LEXICON – holistic properties, opacity

Prosody and the architecture of language

Rank Interpretation Model



Prosody: rhythm and melody

Analysing time:

1. *Empirical resources* – recording, transcription, annotation
2. *Time Group Analysis* – an online tool for studying speech timing and rhythm

Speech data

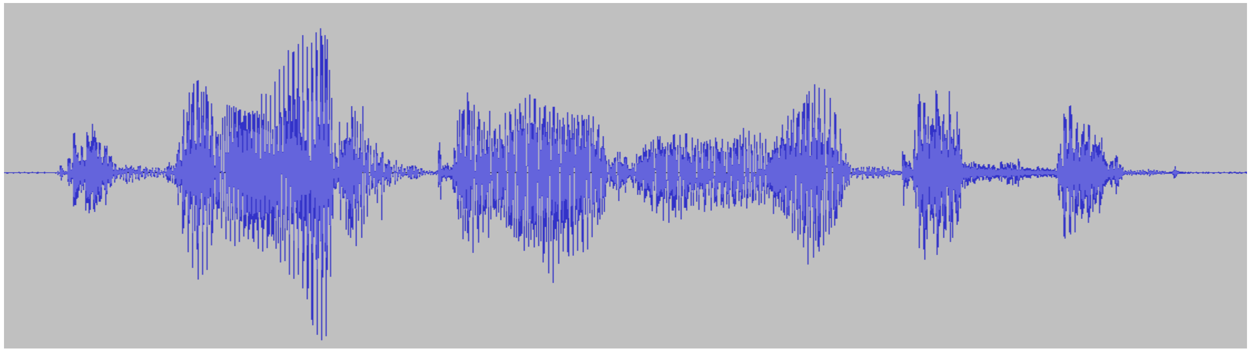
- Pre-recording phase:
  - definition of purposes for which the data will be used
  - scenario: domain, activities, speakers
  - equipment and technical operator:
    - general: digital audio (recorder / laptop), digital video
    - specialised: laryngograph, etc
- Recording phase:
  - negotiate scenario with chiefs, elders, speakers
  - ensure the recording location is quiet
  - if possible ensure the microphones, video tripod etc. can be stably positioned
- Post-recording phase:
  - provide recordings with metadata immediately
  - label the data media immediately
  - make safety copies immediately

A case study on rhythm and melody: Tem (Gur; ISO 639-2 kth)

Empirical resources 1: recording

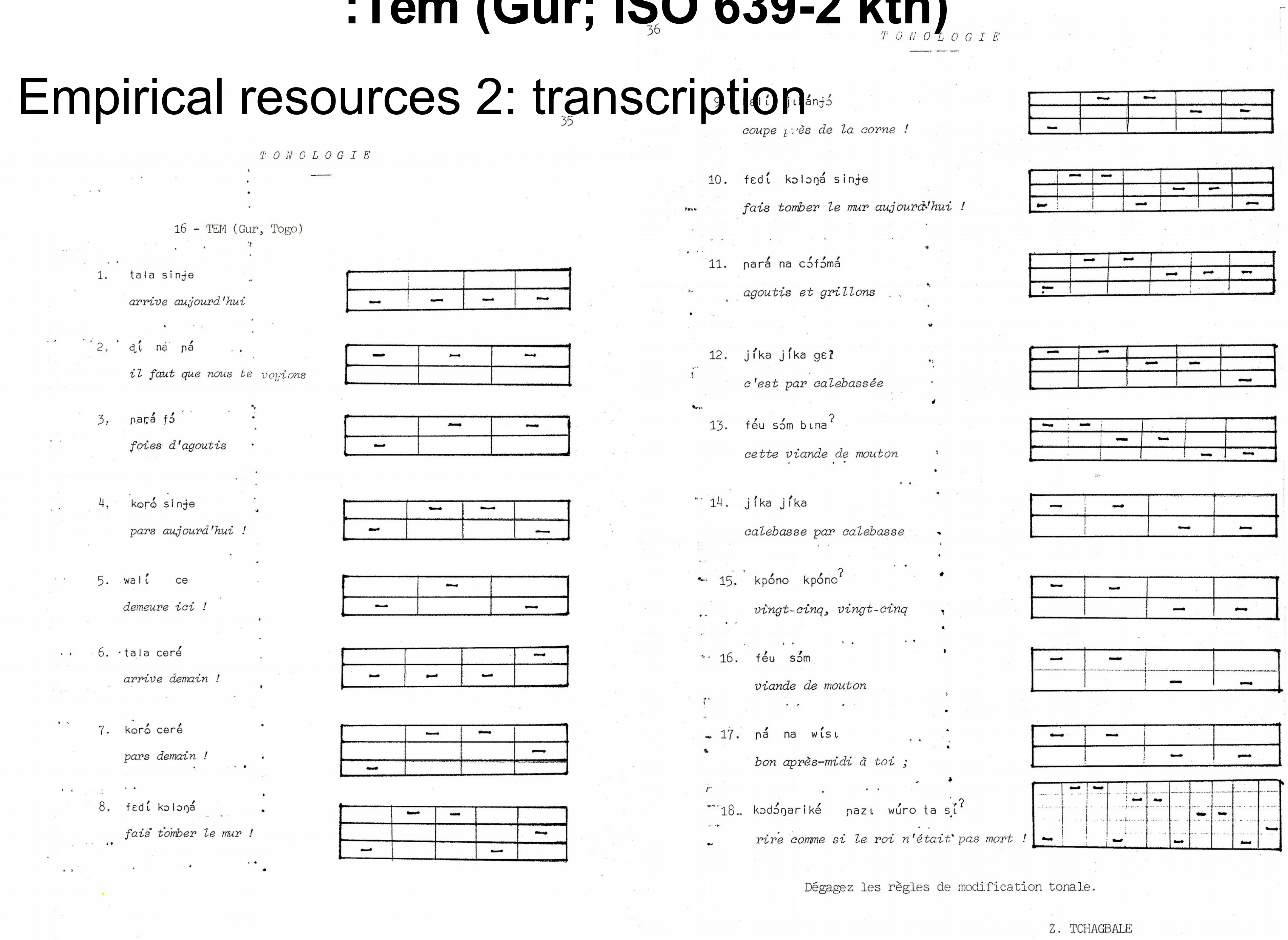
‘Facebook Fieldwork’: ☺

- Zakari Tchagbale’s “Exercises et corrigés” (1984)
  - resources for many languages, including Tem (1984)
  - DG had already developed a tone model for Tem (1987)
- DG asks ZT via internet for recording of Tem data (2012)
- recording by ZT with mobile phone (2012)
- ZT emails recording to DG (2012)



A case study on rhythm and melody :Tem (Gur; ISO 639-2 kth)

Empirical resources 2: transcription



35

T O N O L O G I E

16 - TEM (Gur, Togo)

1. tala sinje
   arrive aujourd'hui
2. ɖí ná ɲá
   il faut que nous te voyions
3. ɲaɽá fɔ́
   foies d'agoutis
4. koró sinje
   pars aujourd'hui !
5. walí ce
   demeure ici !
6. tala ceré
   arrive demain !
7. koró ceré
   pars demain !
8. fɛdí kɔlɔŋá
   fais tomber le mur !

36

T O N O L O G I E

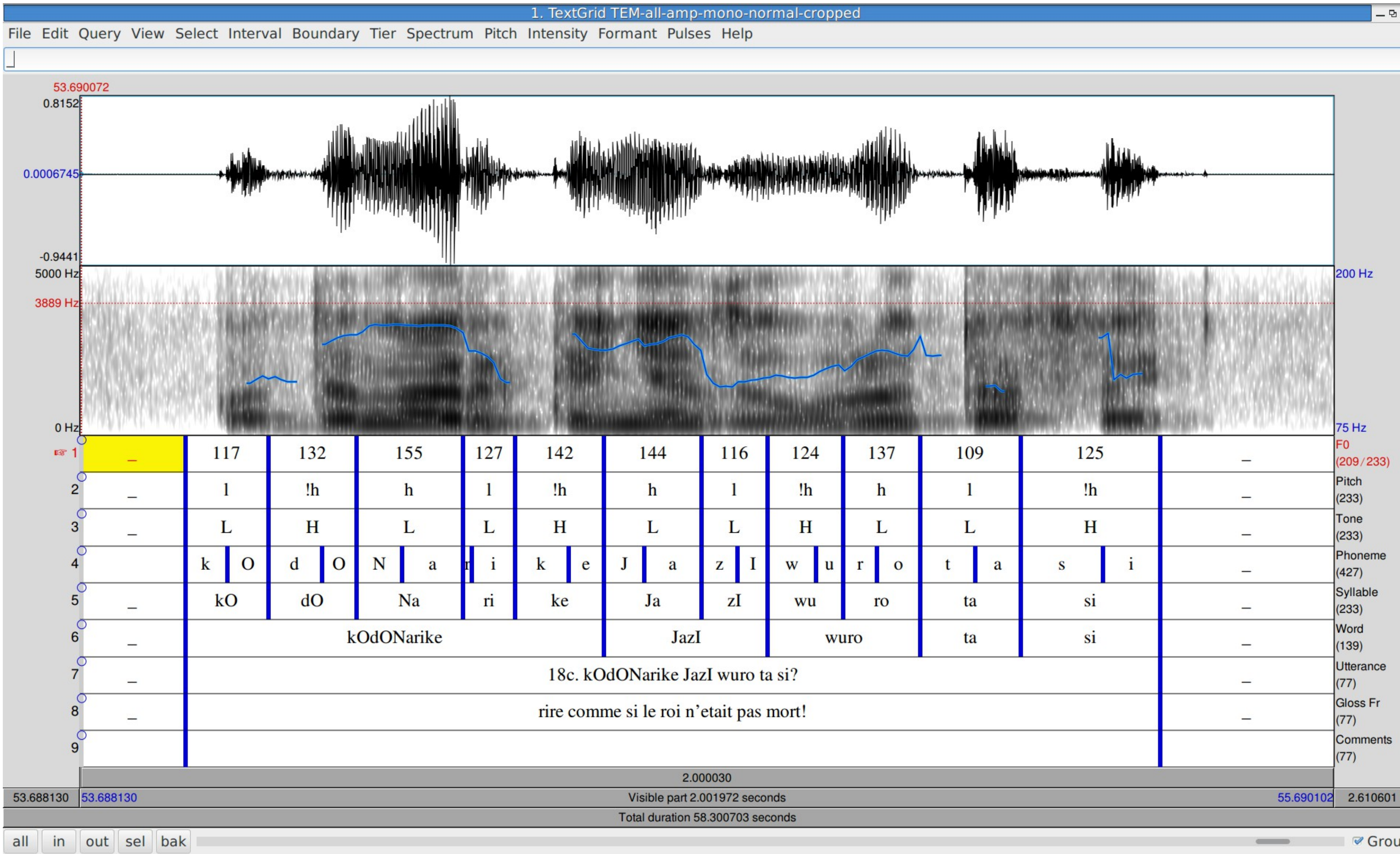
9. [illegible] ánjó
   coupe près de la corne !
10. fɛdí kɔlɔŋá sinje
    fais tomber le mur aujourd'hui !
11. ɲará na cɔ́fɔ́má
    agoutis et grillons
12. jíka jíka gɛʔ
    c'est par calebassée
13. féu sɔ́m bɩna?
    cette viande de mouton
14. jíka jíka
    calebasse par calebasse
15. kpóno kpóno?
    vingt-cinq, vingt-cinq
16. féu sɔ́m
    viande de mouton
17. ɲá na wísɩ
    bon après-midi à toi ;
18. kɔdɔ́ŋariké ɲazɩ wúro ta sí?
    rire comme si le roi n'était pas mort !

Dégagez les règles de modification tonale.

Z. TCHAGBALE

A case study on rhythm and melody: Tem (Gur; ISO 639-2 kth)

Empirical resources 3: annotation



Time Group Analysis

(Interpausal Group Analysis)

Time Group Analysis – Rhythm?

Rhythm is an emergent property of timing, determined multiple, typologically varying factors:

- functional factors – these are well investigated:
  - Lexical: contrastive duration (2 or 3 values)
  - Phrasal: relations between stressed-unstressed items
  - Discoursal: rhetorical pause

- phonetic factors – these are highly controversial:
  - Inherent Consonant and Vowel duration
  - Balance of Consonant and Vowel duration in syllables
  - Language specific compositional units of timing:

    mora – syllable – foot

→ 'Rhythm metrics'
→ *TGA – an online tool for speech timing analysis*

Timing: temporal relations – rhythm metrics

$PIM(I_{1,\ldots n})$	=	$\sum_{i \neq j} \lvert \log \frac{I_i}{I_j} \rvert$
$PFD(foot_{1\ldots n}) =$		$\frac{100 \times \sum \lvert MFL - len(foot_i) \rvert}{len(foot_{1\ldots n})}$ where MFL = 'mean foot length'
$rPVI(d_{1\ldots m})$	=	$\sum_{k=1}^{m-1} \lvert d_k - d_{k+1} \rvert / (m-1)$
$nPVI(d_{1\ldots m})$	=	$100 \times \sum_{k=1}^{m-1} \lvert \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \rvert / (m-1)$

'Rhythm metrics' of relative ('fuzzy', 'sloppy') isochrony:

- measures of regularity...irregularity of timing units
  - *PIM*: Pairwise Irregularity Measure
  - *PFD*: Pairwise Foot Difference
  - *rPVI*, *nPVI*: raw and normalised Pairwise Variability Index
- not rhythm, though: they ignore rhythmic alternation

A case study on rhythm: Tem (Gur; ISO 639-2 kth)

N	dur	rate	mean	median	stdev	npvi	median npvi	inter-cept	slope
11	1795	6.13	163.18	162.00	36.57	24	19	132.27	6.18

Above: example of TGA information output for one interpausal unit.

Below: reformatted information.

N:		17
duration (ms):		95
syllable rate:		6.13
mean syllable duration:		163.18
median syllable duration:		162.00
standard deviation of syllable duration:		36.57
nPVI:		24
median-nPVI:		19
linear regression	intercept:	132.27
	slope:	6.18

A case study on rhythm: Tem (Gur; ISO 639-2 kth)

Example of TGA information output for a sequence of interpausal units.

Duration properties (syllables)

Attributes	Values	Attributes	Values
n:	194	intercept:	185.235
min:	78	slope:	-0.19
max:	350	std:	49.979
mean:	166.8	nPVI:	35
median:	162.0	rPVI:	60
total:	32360	100*rPVI/med:	37
range:	272	nPVI*med/100:	57

A case study on rhythm: Tem (Gur; ISO 639-2 kth)

TGA output sample: statistics

Overall duration:	32360	Overall raw longer, ms:	5804	Overall raw shorter, ms:	5711
Overall min:	78.00	Overall max:	350.00	Overall range:	272.00
Valid Time Groups:	38	Overall rate/sec:	01.06.00		

Overall mean:	166.80	Overall median:	162.00	Overall SD:	49.98
Overall npvi:	35.00	Overall intercept:	185.23	Overall slope:	-0.18

Mean of means:	171.36	Median of means:	167.60	SD of means:	18.02.14
Mean of medians:	164.57	Median of medians:	168.25	SD of medians:	21.61
Mean of SDs:	45.61	Median of SDs:	41.03	SD of SDs:	20.41

mean::TGdur:	-0.403	median::TGdur:	-0.201	SD::TGdur:	-0.140
nPVI::TGdur:	-0.199	slope::TGdur:	-0.373	intercept::TGdur:	-0.113
nPVI::mean:	0.242	slope::mean:	0.718	intercept::mean:	0.003
nPVI::median:	-0.053	slope::median:	0.358	intercept::median:	0.266
nPVI::SD:	0.798	slope::SD:	0.840	intercept::SD:	-657

A case study on rhythm: Tem (Gur; ISO 639-2 kth)

TGA online tool: visualisation of syllable time relations

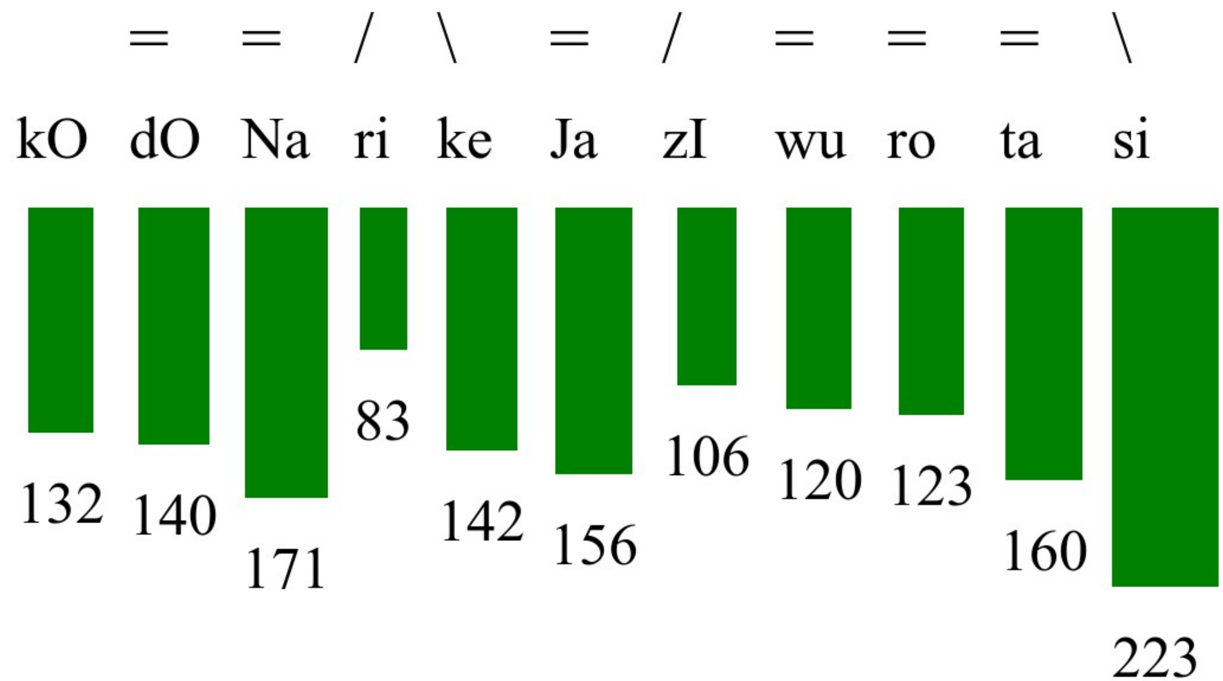
http://wwwhomes.uni-bielefeld.de/gibbon/TGA/

Duration difference tokens:

Keyboard friendly transcription:

2D visualisation of durations:

Syllable durations:



Duration difference tokens for utterance #37:

- pos, neg, equal differences between neighbours: /, \, =
- difference threshold: 40ms.
- clear indication of syllable isochrony

A case study on rhythm: Tem (Gur; ISO 639-2 kth)

n-grams	10ms		20ms		30ms		40ms		50ms		60ms		70ms		80ms	
	%	seq	%	seq	%	seq	%	seq	%	seq	%	seq	%	seq	%	seq
Uni 232	35	\	31	\	30	\	27	=	36	=	43	=	49	=	51	=
	24	/	21	/	19	=	25	\	20	\	17	\	16	+	16	+
	16	+	16	+	19	/	16	+	16	+	16	+	16	#	16	#
Di 194	20	/\	15	/\	14	/\	13	=\	19	==	24	==	31	==	35	==
	15	\/	14	\/	12	\/	11	\#	15	+=	16	+=	19	+=	19	+=
	12	\#	12	\#	11	\#	11	==	13	=\	13	=\	13	=\	12	=\
Tri 156	12	+/\	9	+/\	14	/\	8	=\#	10	===	15	===	19	===	24	===
	11	\/\	8	\/\	12	\/	8	+=\	10	+==	11	+==	14	+==	15	+==
	10	/\/	8	/\/	11	\#	6	+/\	8	+=\	8	+=\	10	==#	11	==#
Quad 118	12	+/\/	9	+/\/	9	+/\/	7	+=\/	8	+=\/	8	====	10	====	14	====
	7	/\/#	6	+=\/	6	+=\/	7	+/\/	5	=\/#	8	+===	10	+===	14	+===
	6	\/\\	5	\/\#	4	\/\#	4	=\/#	5	====	7	=\=#	9	===#	11	===#

Duration difference tokens for utterance #37:

- *n*-grams: unigrams, digrams, trigrams, quadgrams
- thresholds 10...80 ms

A case study on rhythm: Tem (Gur; ISO 639-2 kth)

Visualisation of duration difference relation hierarchy:

Iambic grouping:

Greater-than:

10ms: (((kO (dO Na)) (((ri ke) Ja) (((zI wu) (ro ta)) si)))

40ms: ((kO (dO (Na ((ri ke) (Ja (zI (wu (ro (ta si))))))))))

Greater-than-or-equal:

10ms: (kO dO) Na ri ke Ja zI (wu ro)

40ms: ((kO dO) Na) ri ((ke Ja) (((zI wu) ro) ta))

Trochaic:

10ms: (kO (dO (Na ri))) (ke (Ja zI)) wu ro ta si

40ms: (kO (dO (Na ri))) ke (Ja zI) wu ro ta si

Duration difference tokens for utterance #37

Note that unidirectional branching is effectively linear.

A case study on rhythm: Tem (Gur; ISO 639-2 kth)

TGA output sample:

Visualisation of interpausal units

(screenshot)

31	4	683	5.86	170.75	173.00	31.19	37	40	155.30	10.30	\ / \ kpo no kpo no 150 196 131 206	kpo:150 no:196 kpo:131 no:206 **PAUSE**:766 # **iambicTTgt:** ((kpo no) ((kpo no) **PAUSE**)) **iambicTTgte:** kpo no kpo no **PAUSE** **trochaicTTlt:** kpo (no kpo) no **PAUSE** **trochaicTTlte:** kpo no kpo no **PAUSE**
32	4	617	6.48	154.25	127.50	53.00	47	57	137.60	11.10	= \ / fe u sO m 117 137 245 118	fe:117 u:137 sO:245 m:118 **PAUSE**:782 # **iambicTTgt:** ((fe (u sO)) (m **PAUSE**)) **iambicTTgte:** (fe u) sO m **PAUSE** **trochaicTTlt:** fe u (sO m) **PAUSE** **trochaicTTlte:** fe u sO m **PAUSE**
33	4	635	6.30	158.75	141.00	48.85	56	64	157.40	0.90	/ \ / fe u sO m 160 115 238 122	fe:160 u:115 sO:238 m:122 **PAUSE**:619 # **iambicTTgt:** ((fe (u sO)) (m **PAUSE**)) **iambicTTgte:** fe u sO m **PAUSE** **trochaicTTlt:** (fe u) (sO m) **PAUSE** **trochaicTTlte:** fe u sO m **PAUSE**
34	4	642	6.23	160.50	151.50	41.26	22	22	106.50	36.00	= = \ Ja na wI sI 115 135 168 224	Ja:115 na:135 wI:168 sI:224 **PAUSE**:714 # **iambicTTgt:** ((Ja (na (wI sI))) **PAUSE**) **iambicTTgte:** ((Ja na) wI) sI **PAUSE** **trochaicTTlt:** Ja na wI sI **PAUSE** **trochaicTTlte:** Ja na wI sI **PAUSE**
35	4	713	5.61	178.25	173.50	18.21	13	7	167.30	7.30	= = \ Ja na wI sI 180 167 159 207	Ja:180 na:167 wI:159 sI:207 **PAUSE**:672 # **iambicTTgt:** (Ja (na ((wI sI) **PAUSE**))) **iambicTTgte:** Ja na wI sI **PAUSE** **trochaicTTlt:** Ja na wI sI **PAUSE** **trochaicTTlte:** ((Ja na) wI) sI **PAUSE**
36	11	1839	5.98	167.18	164.00	51.98	35	37	119.86	9.46	/ \ / \ \ = = / = \ kO dO Na ri ke Ja zI wu ro ta si 155 114 176 87 143 185 185 204 131 164 295	kO:155 dO:114 Na:176 ri:87 ke:143 Ja:185 zI:185 wu:204 ro:131 ta:164 si:295 **PAUSE**:476 # **iambicTTgt:** ((kO ((dO Na) (((ri ke) Ja) (zI (wu (ro (ta si))))))) **PAUSE**) **iambicTTgte:** kO dO Na ri ke ((Ja zI) wu) (ro ta) si **PAUSE** **trochaicTTlt:** (kO dO) (Na ri) ke (Ja (zI (wu ro))) ta si **PAUSE** **trochaicTTlte:** kO dO Na ri ke Ja zI wu ro ta si **PAUSE**
37	11	1795	6.13	163.18	162.00	36.57	24	19	132.27	6.18	= \ / \ = = = = = \ kO dO Na ri ke Ja zI wu ro ta si 152 118 162 103 185 181 182 170 149 145 248	kO:152 dO:118 Na:162 ri:103 ke:185 Ja:181 zI:182 wu:170 ro:149 ta:145 si:248 **PAUSE**:549 # **iambicTTgt:** ((kO ((dO Na) ((ri ke) (Ja (zI (wu (ro (ta si)))))))) **PAUSE**) **iambicTTgte:** kO dO Na ri ke (Ja zI) wu ro ta si **PAUSE** **trochaicTTlt:** kO dO (Na ri) ke Ja zI wu ro ta si **PAUSE** **trochaicTTlte:** (kO dO) Na ri ((((ke Ja) (zI wu)) ro) ta) si **PAUSE**

Timing: temporal relations - rhythm

Summary of utility of timing analysis:

- Studies in prosodic typology of timing
  - e.g. mora, syllable, foot timing (depending on annotation)
  - Check papers and examples of automatic annotation analysis here:
    - http://wwwhomes.uni-bielefeld.de/gibbon/TGA
- Studies in musicology
  - e.g. annotated music performances
- Software development for
  - measuring foreign language phonetic proficiency
  - diagnosis and therapy in speech pathology
  - benchmarking
    - duration models in natural speech synthesis
    - designing disambiguation models in speech recognition