Prosody: speech rhythms and melodies

3. Acoustic Phonetic Basics

Dafydd Gibbon

Summer School
Contemporary Phonology and Phonetics
Tongji University 9-15 July 2016

Contents

The Domains of Phonetics

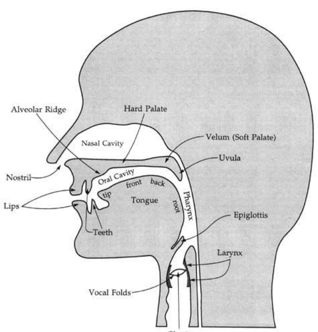
- Phonetics is the scientific discipline which deals with
  - speech production (articulatory phonetics)
  - speech transmission (acoustic phonetics)
  - speech perception (auditory phonetics)
- The scientific methods used in phonetics are
  - direct observation ("impressionistic"), usually based on 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

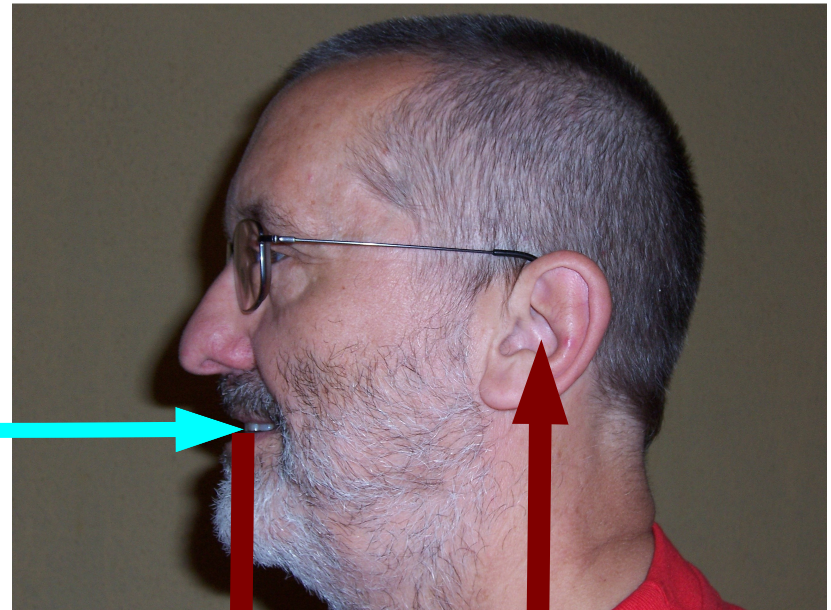
The Domains of Phonetics: the Phonetic Cycle

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

The Domains of Phonetics: the Phonetic Cycle

Sender:
Articulatory
Phonetics

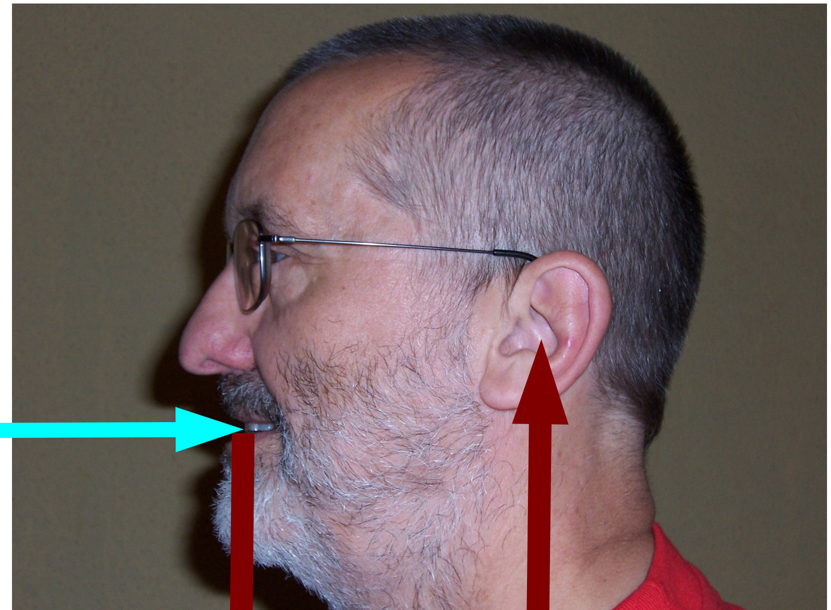


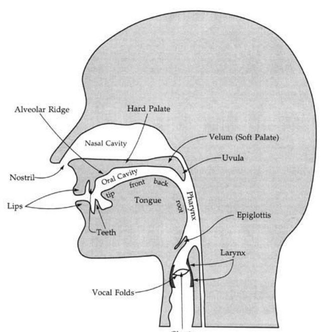


A tiger and a mouse were walking in a field...

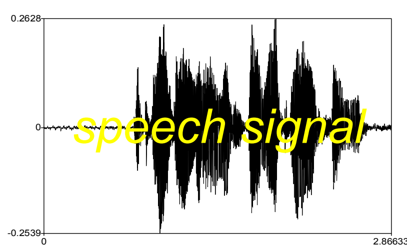
The Domains of Phonetics: the Phonetic Cycle

Sender:
Articulatory
Phonetics



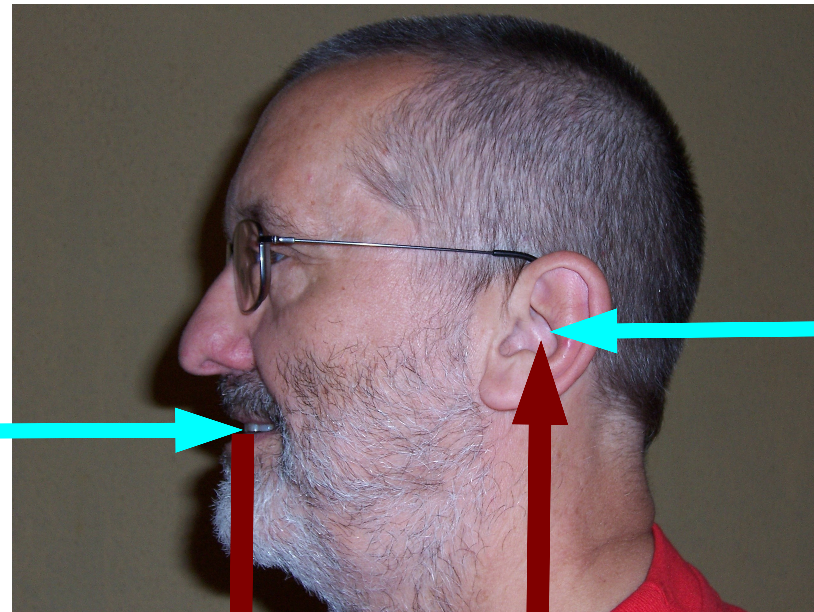


A tiger and a mouse were walking in a field...

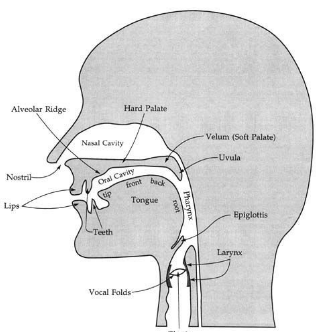


Channel:
Acoustic
Phonetics

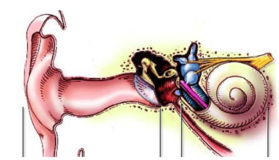
The Domains of Phonetics: the Phonetic Cycle



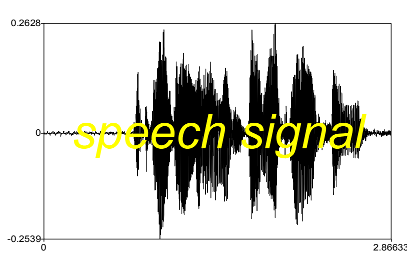
Sender:
Articulatory
Phonetics



Receiver:
Auditory
Phonetics



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



Channel:
Acoustic
Phonetics

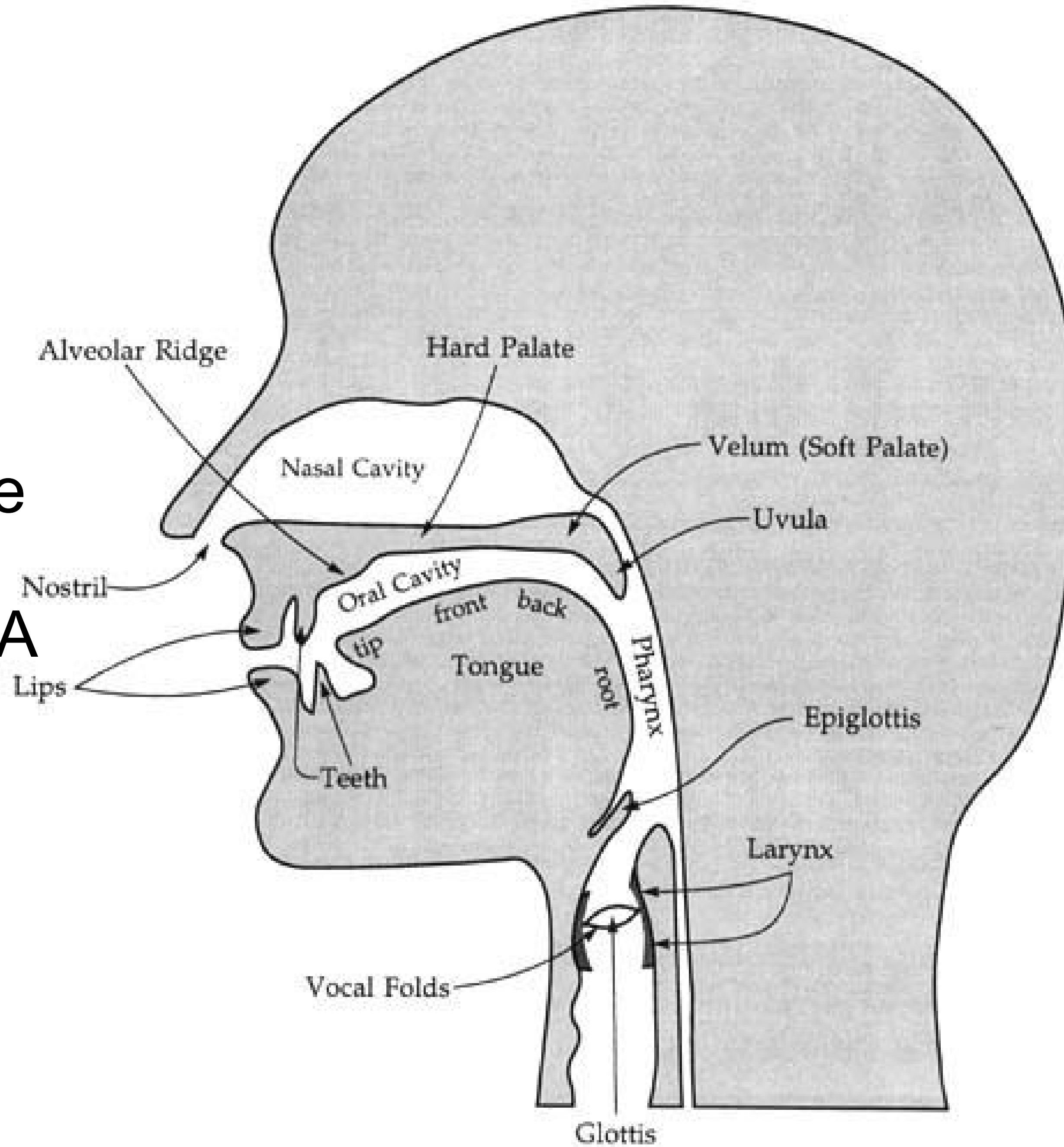
Quiz on the Phonetic Cycle

- 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”?

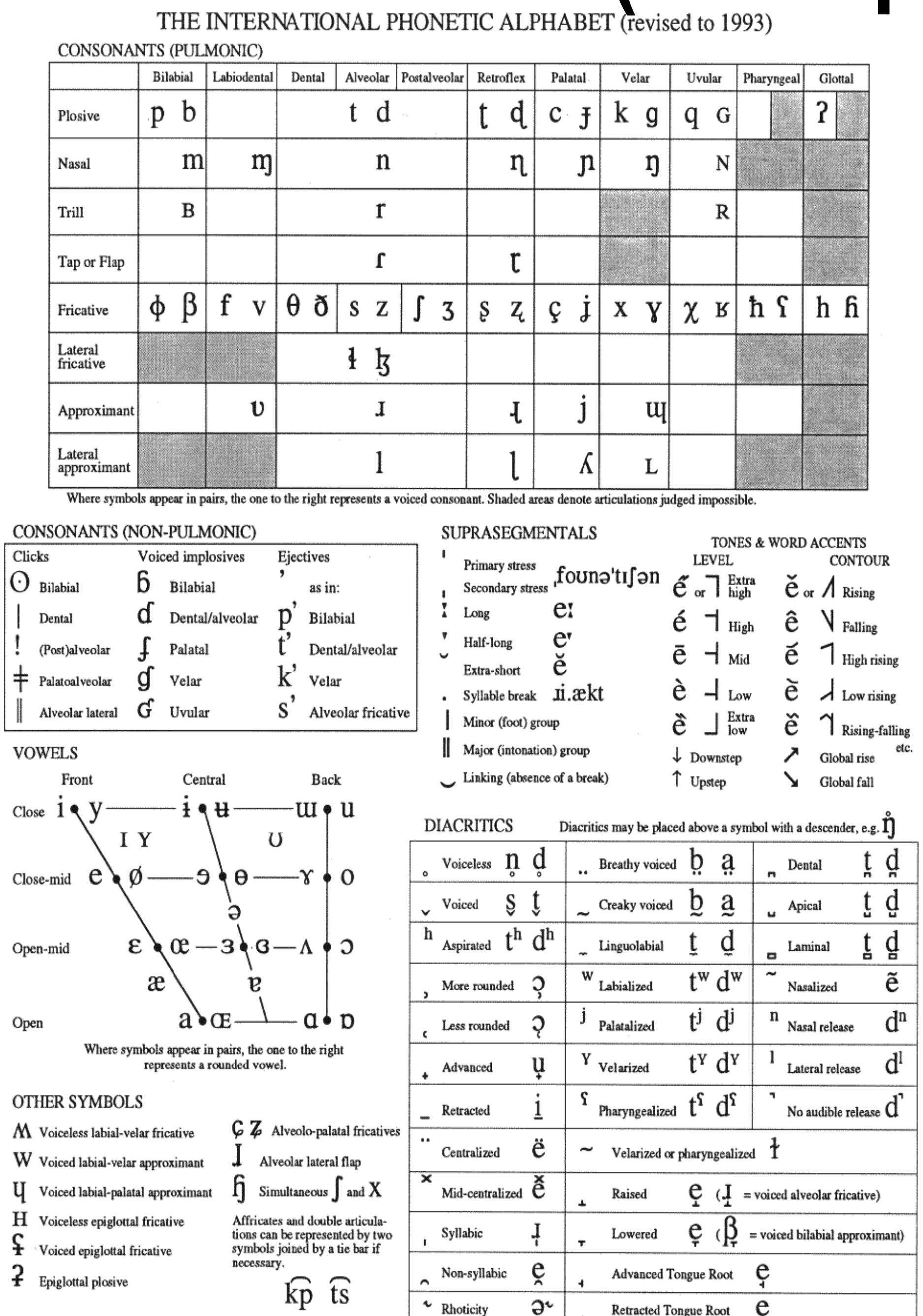
Articulatory Phonetics
(Speech Production)

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 ɡ	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

ˌ Secondary stress ˌfoʊnəˈtɪʃən

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

e̋ or ˥ Extra high

é ˦ High

ē ˧ Mid

è ˨ Low

ȅ ˩ Extra low

↓ Downstep

↑ Upstep

CONTOUR

ě or ˩˥ Rising

ê ˥˩ Falling

e᷄ ˧˥ High rising

e᷅ ˩˧ Low rising

e᷈ ˧˦˧ Rising-falling etc.

↗ Global rise

↘ Global fall

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

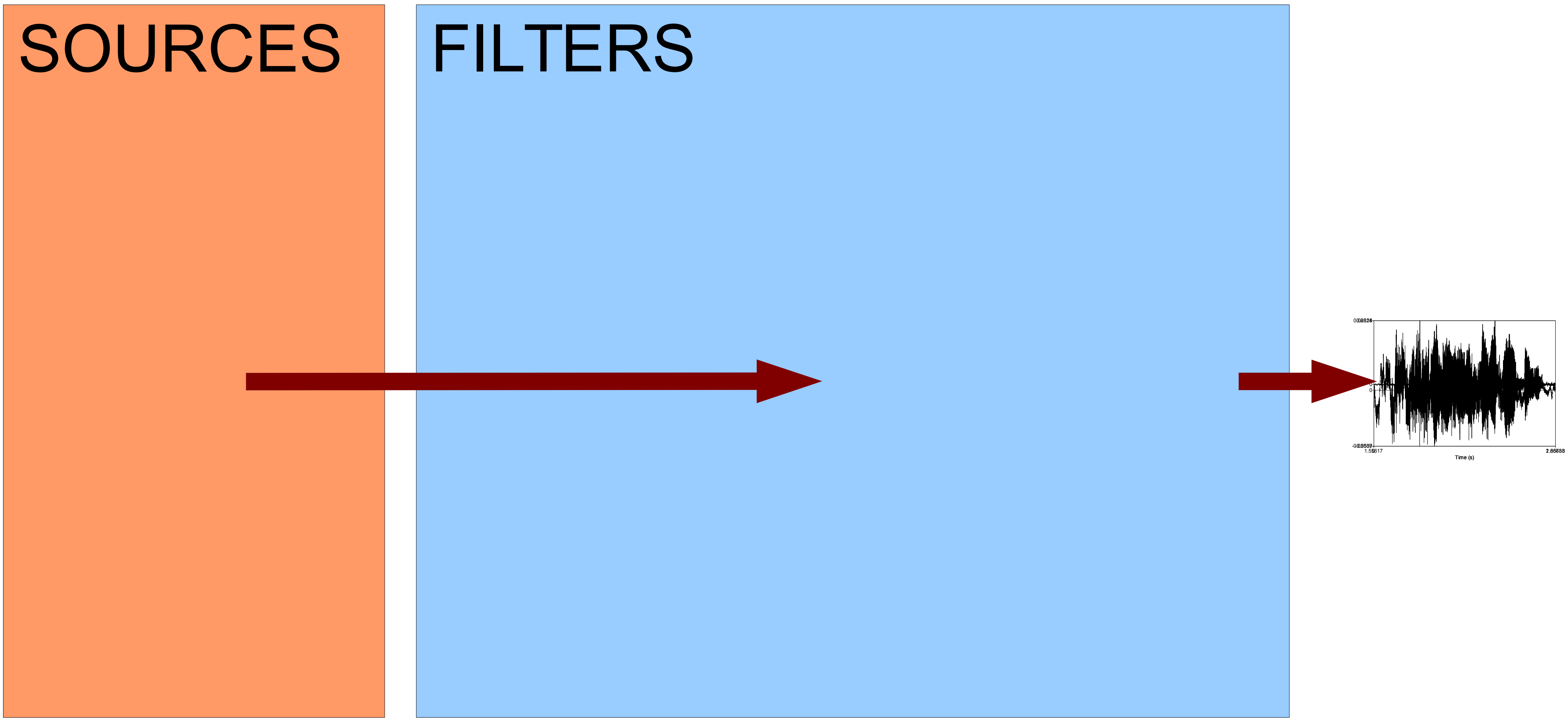
̥ Voiceless	n̥ d̥	̤ Breathy voiced	b̤ a̤	̪ Dental	t̪ d̪
̬ Voiced	s̬ t̬	̰ Creaky voiced	b̰ a̰	̺ Apical	t̺ d̺
ʰ Aspirated	tʰ dʰ	̼ Linguolabial	t̼ d̼	̻ Laminal	t̻ d̻
̹ More rounded	ɔ̹	ʷ Labialized	tʷ dʷ	̃ Nasalized	ẽ
̜ Less rounded	ɔ̜	ʲ Palatalized	tʲ dʲ	ⁿ Nasal release	dⁿ
̟ Advanced	u̟	ˠ Velarized	tˠ dˠ	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	ɹ̩	̞ Lowered	e̞ (β̞ = voiced bilabial approximant)		
̯ Non-syllabic	e̯	̘ Advanced Tongue Root	e̘		
˞ Rhoticity	ɚ	̙ 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

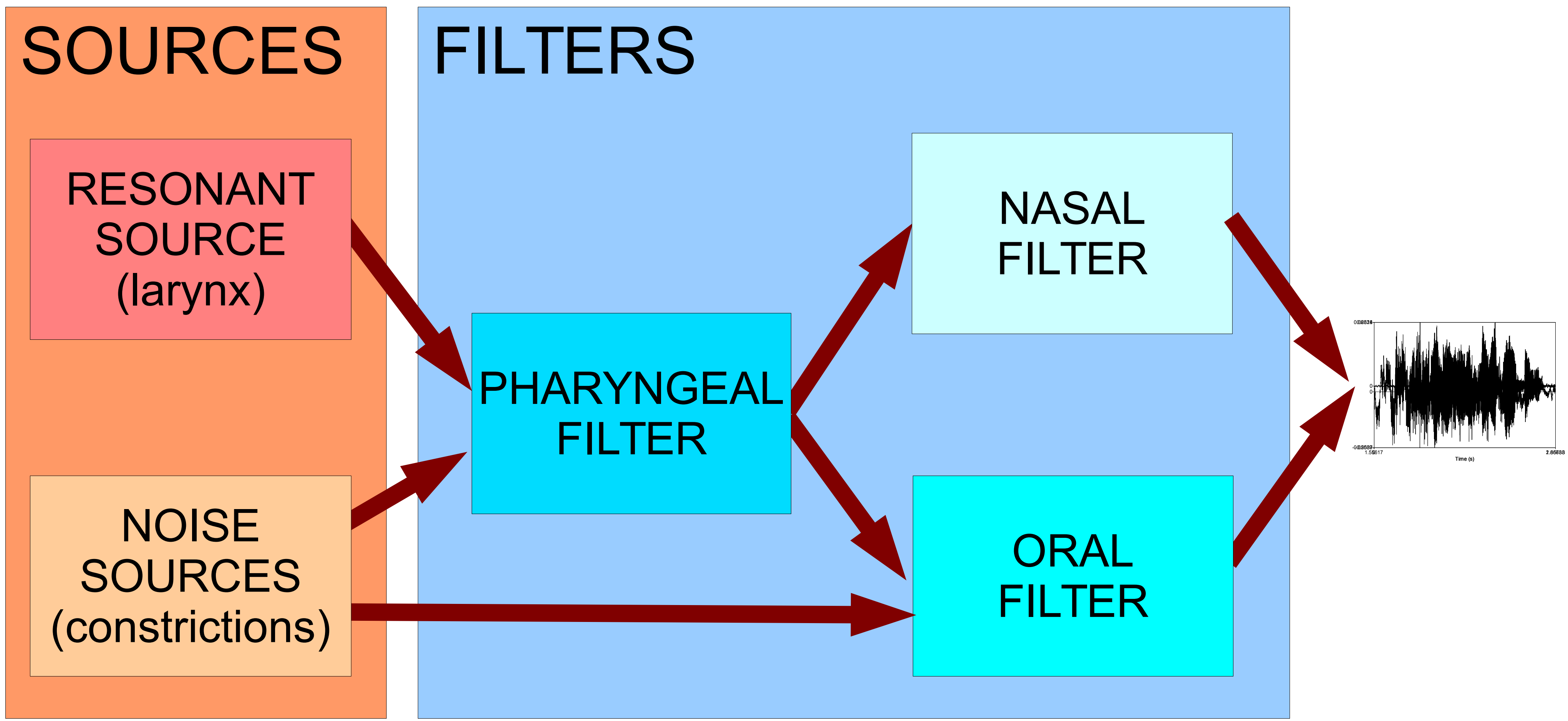
The Source-Filter Model of Speech Production

- 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.

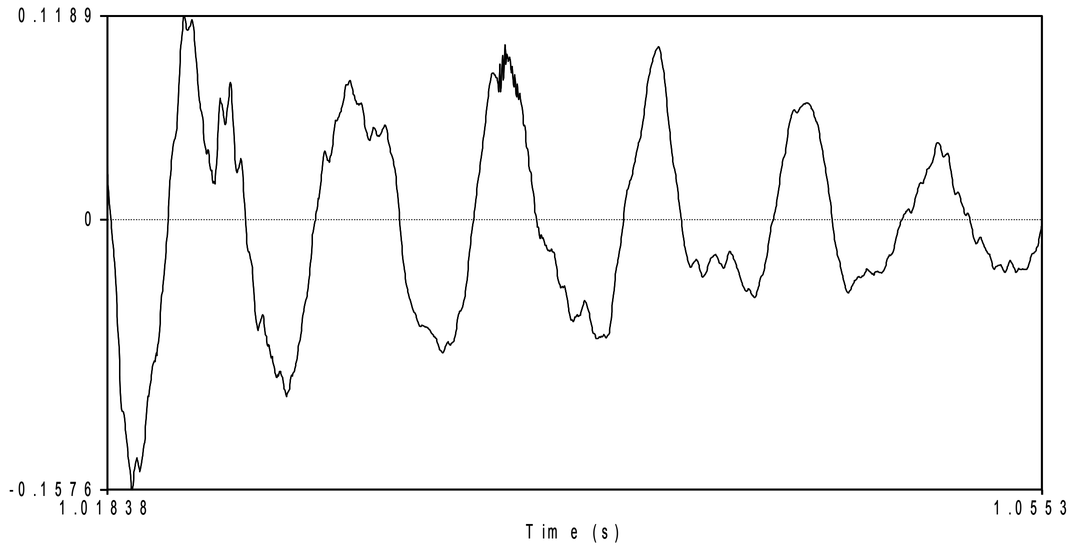
Acoustic Phonetics (Speech Transmission)

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

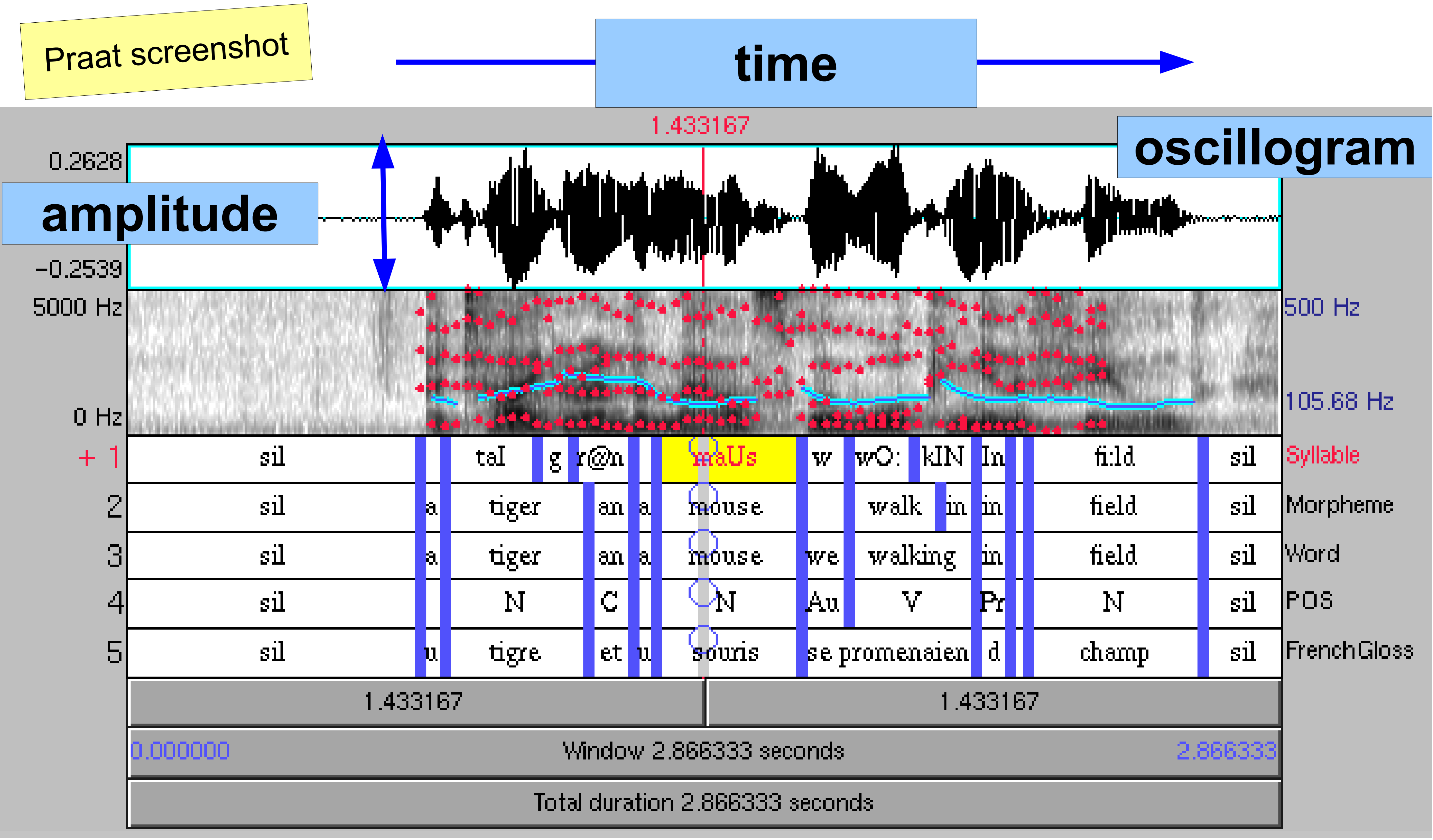
The Speech Wave-Form

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



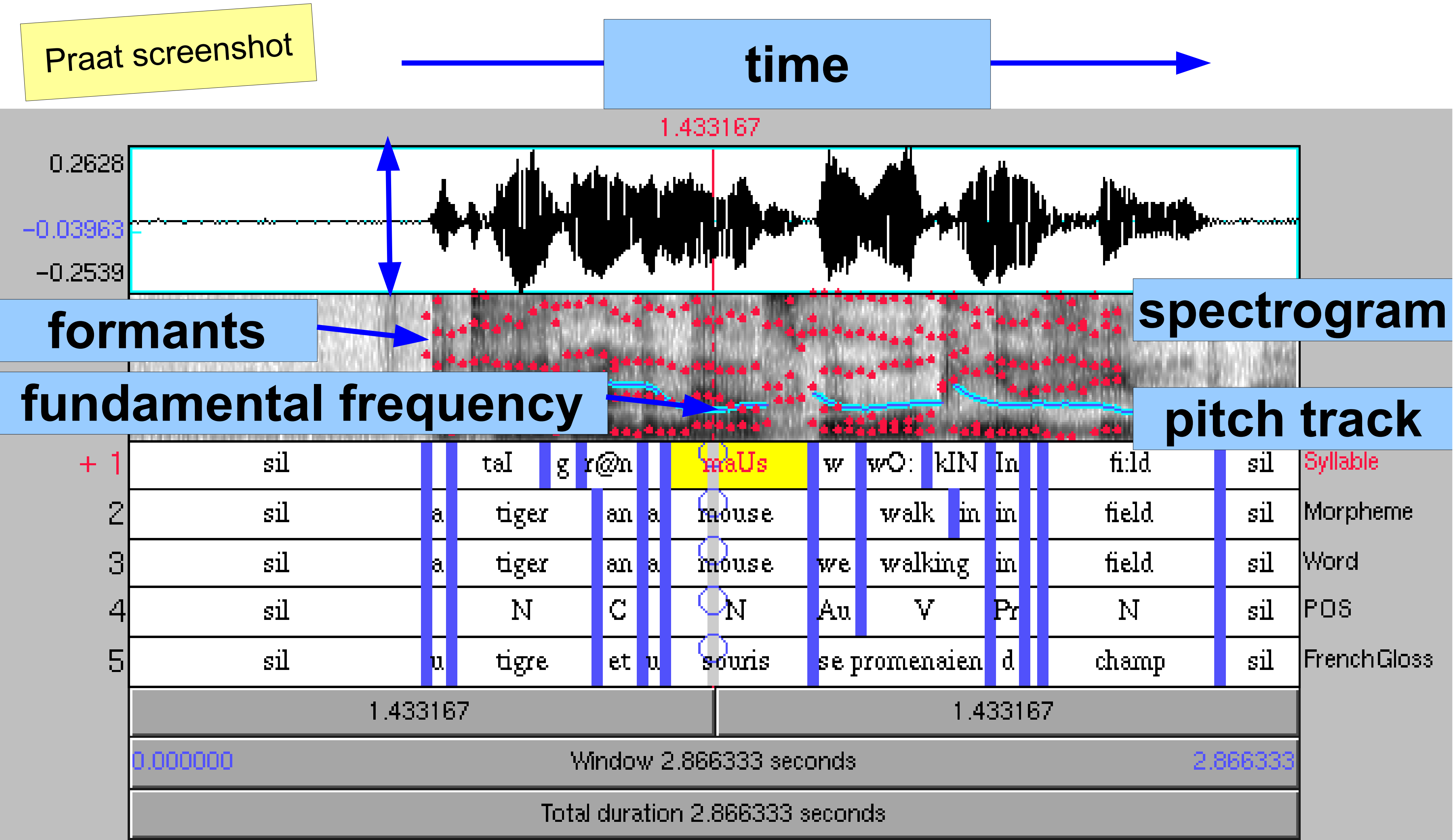
- The changes in air pressure
  - but can be heard
  - and cannot be seen (unlike the waves on the ocean)
  - but can be measured (like the waves on the ocean)
  - and the measurements can be visualised and used for calculating statistical models of the structure of speech

Visualisation of Speech Signal Parameters



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

Visualisation of Speech Signal Parameters

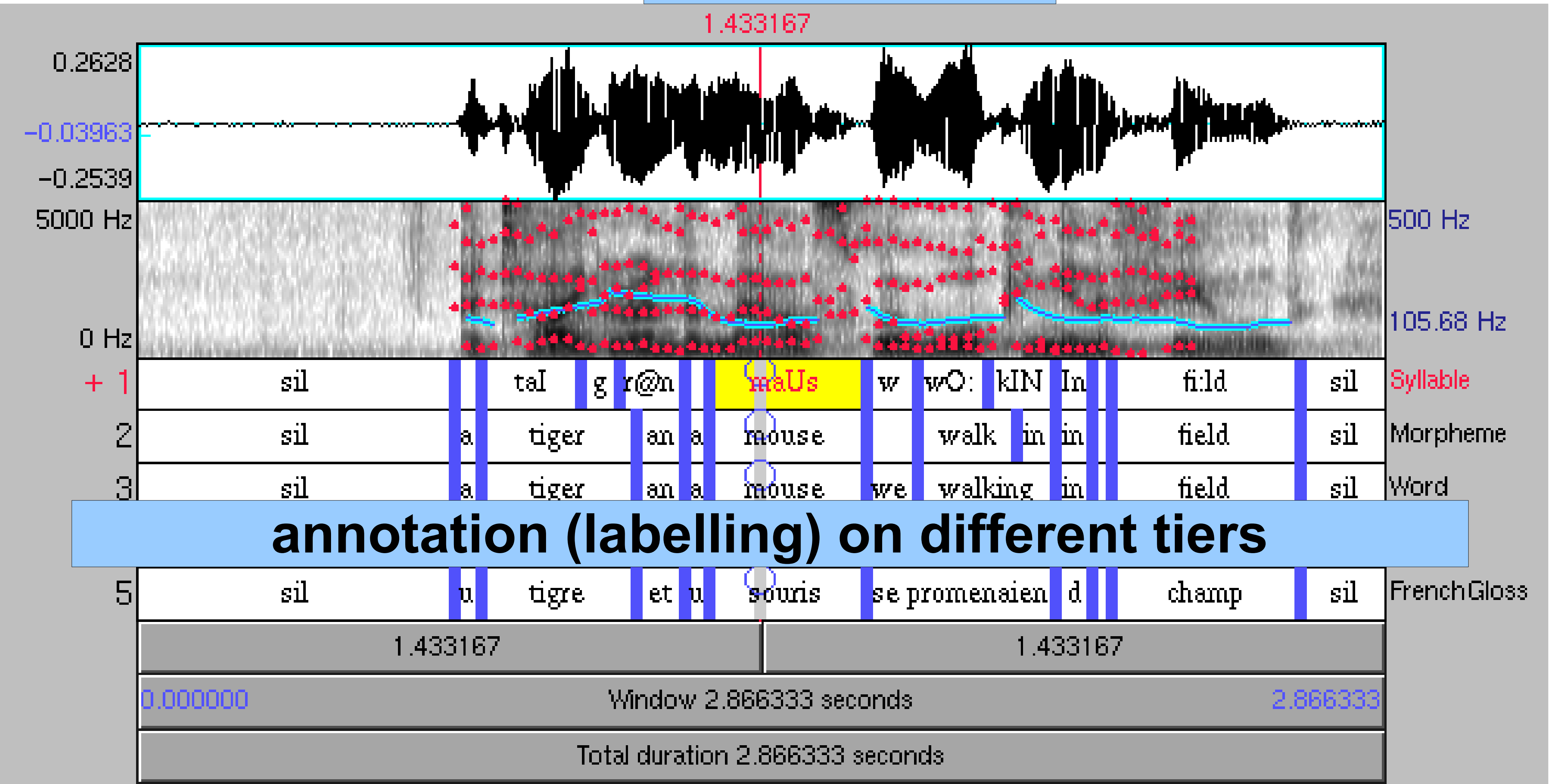


A tiger and a mouse were walking in a field...

Visualisation of Speech Signal Parameters

Praat screenshot

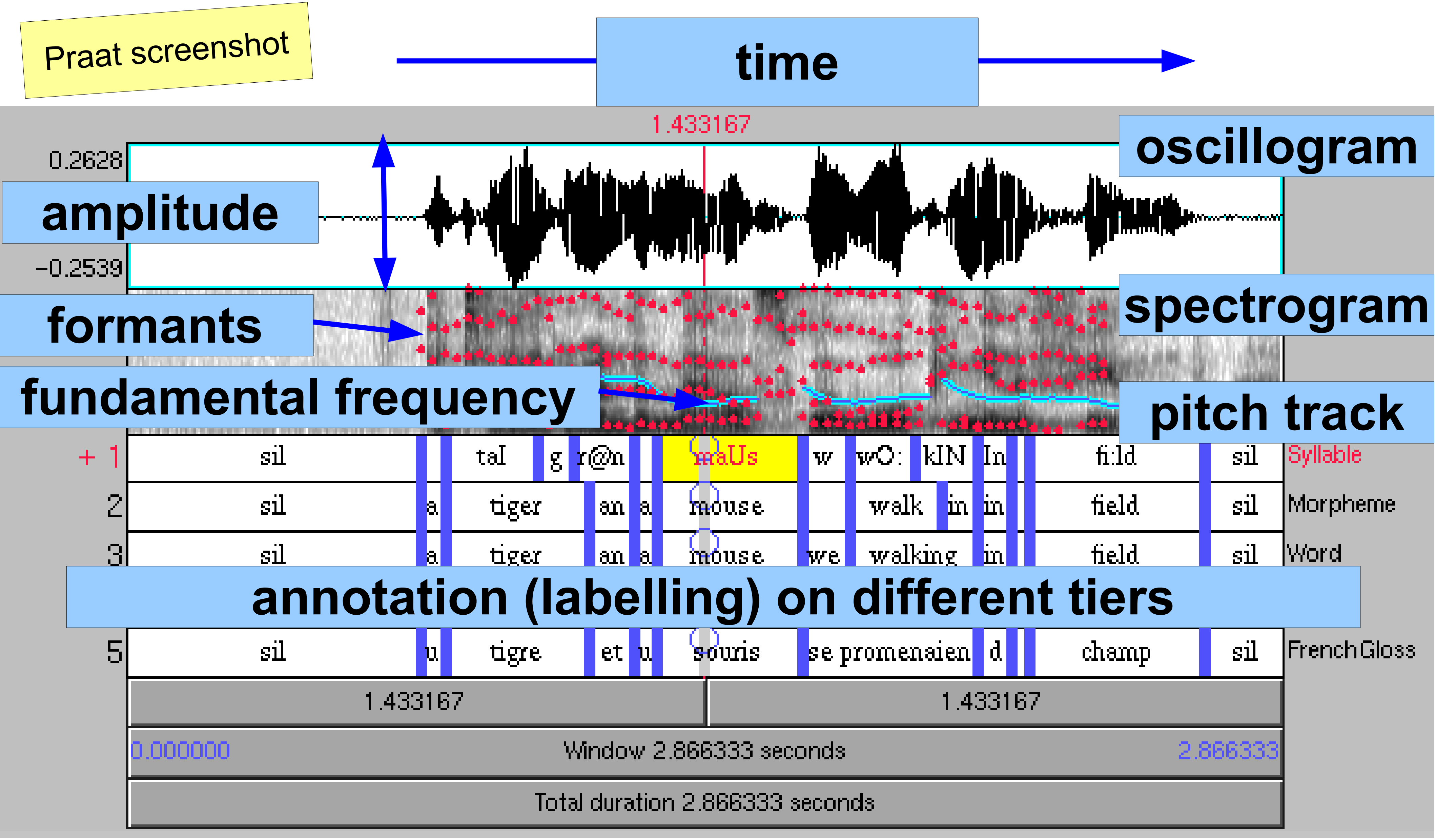
time



annotation (labelling) on different tiers

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

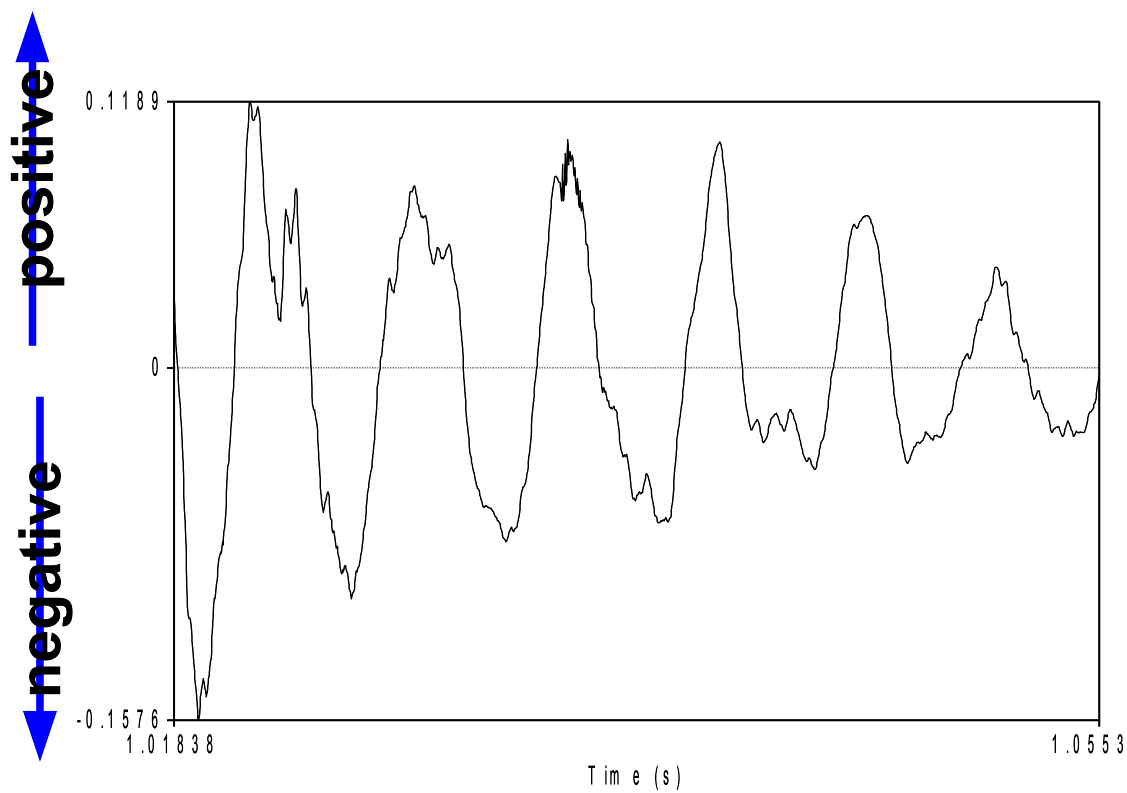
Visualisation of Speech Signal Parameters



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

The Time Domain: the Speech Wave-Form

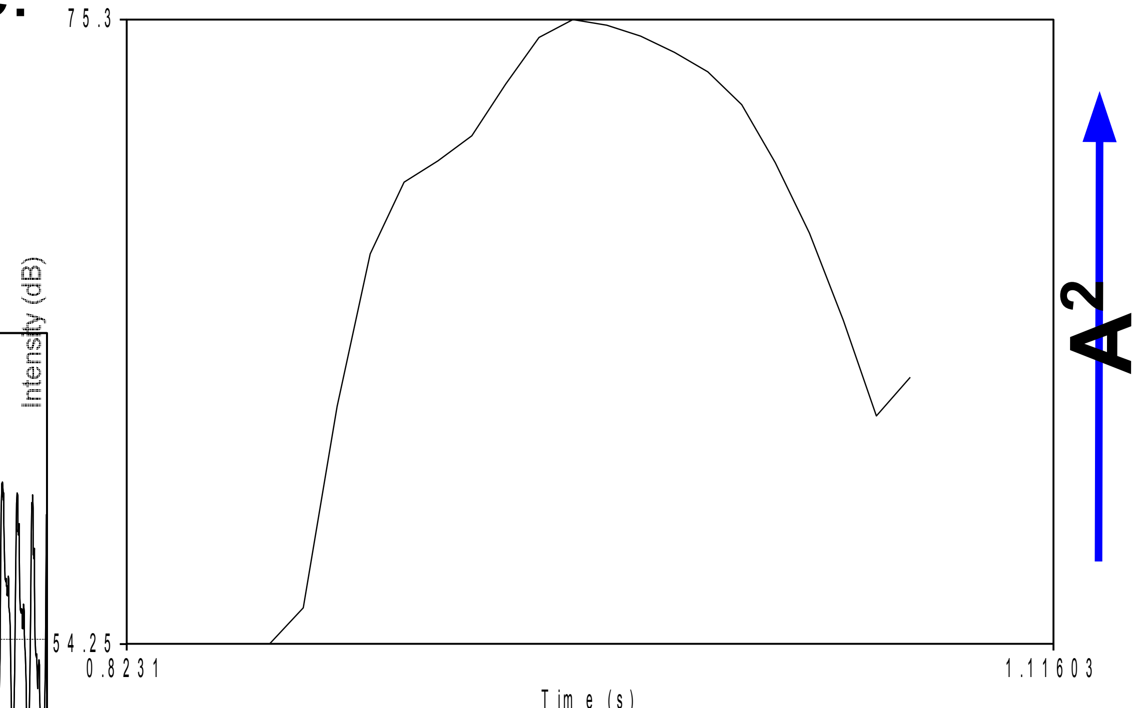
- The *positive* or *negative amplitude* **A** of the speech signal at any given point in time is the *distance* of the wave from zero at this point in time.

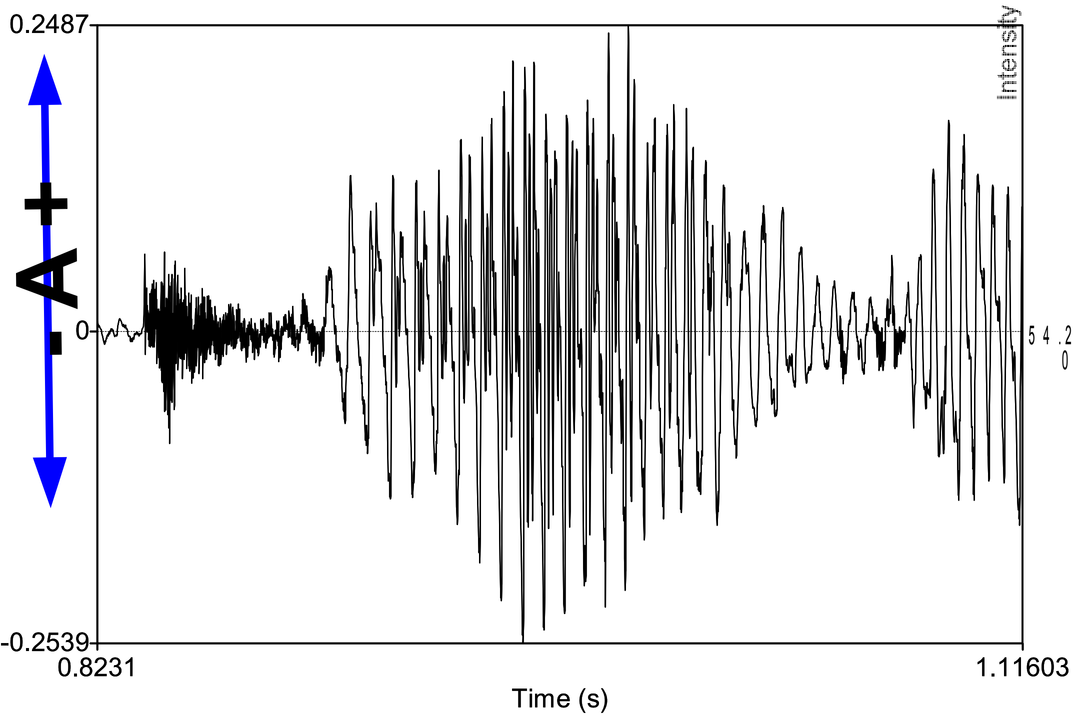


Derived parameter *INTENSITY*

- 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)

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

- Energy is therefore intensity averaged over time
  - In fact, intensity measurements are, in practice, 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 signal is *resonant* if its periods are regular in duration.
  - A signal is *noisy* if its periods are irregular in duration

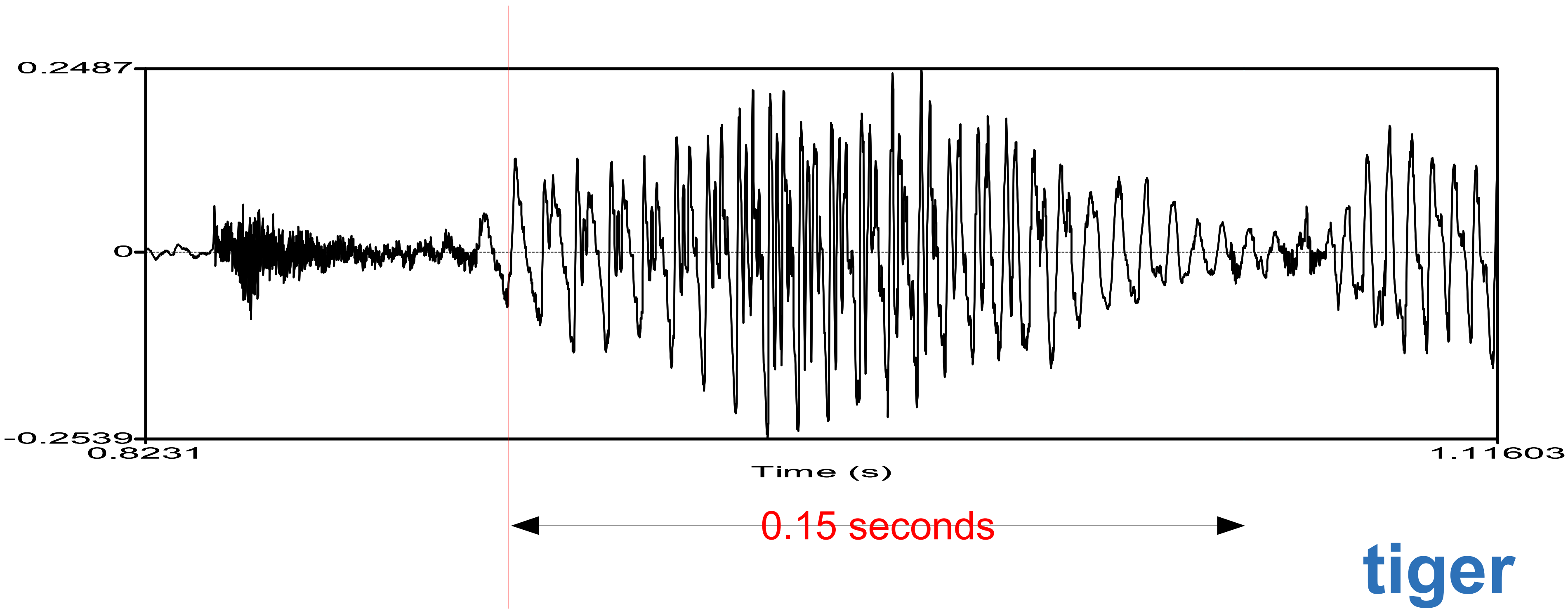
- The *wavelength* λ (lambda) in metres of a speech signal is the speed of sound in m/sec divided by the number of periods per second.

A task:

- *What is the speed of sound?*
- *What is the wavelength of a sound with 100 periods per second?*

The Frequency Domain: simple & complex signals

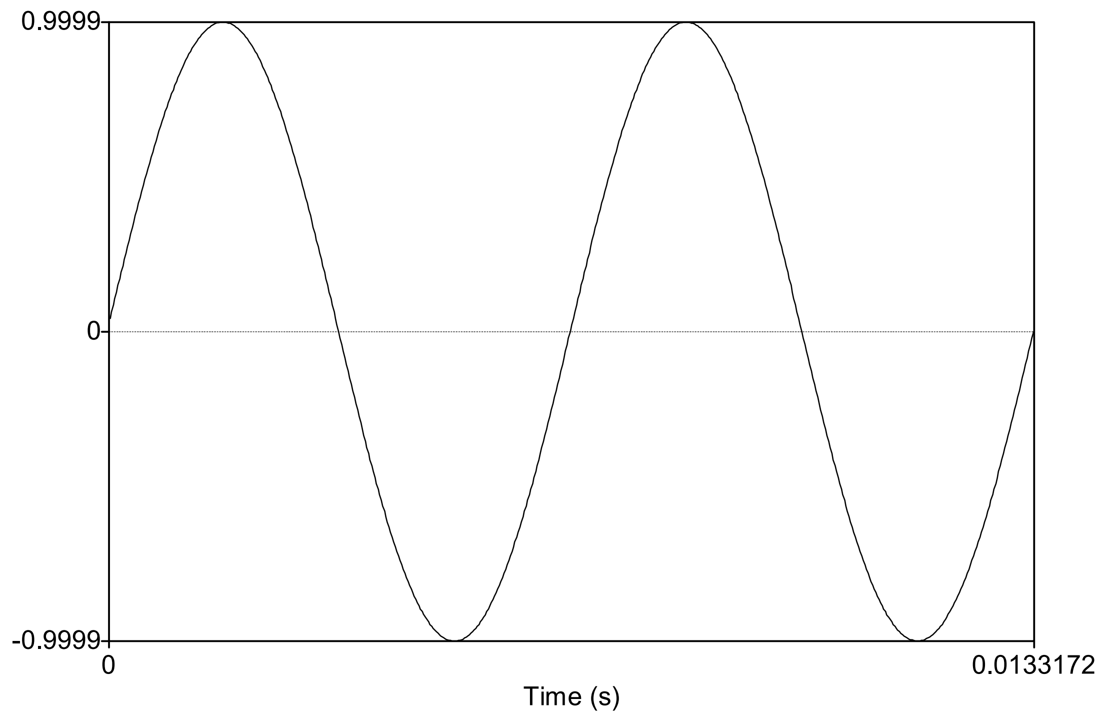
- The *frequency* of a speech signal is the number of waves (periods) per second in the waveform



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

The Simplest Sources produce Sine Waves

- A sine wave with frequency F is produced by an evenly swinging pendulum – a rather slow sine wave!



- The speech signal is not a simple sine wave but a complex signal composed of many sine waves of different frequencies.

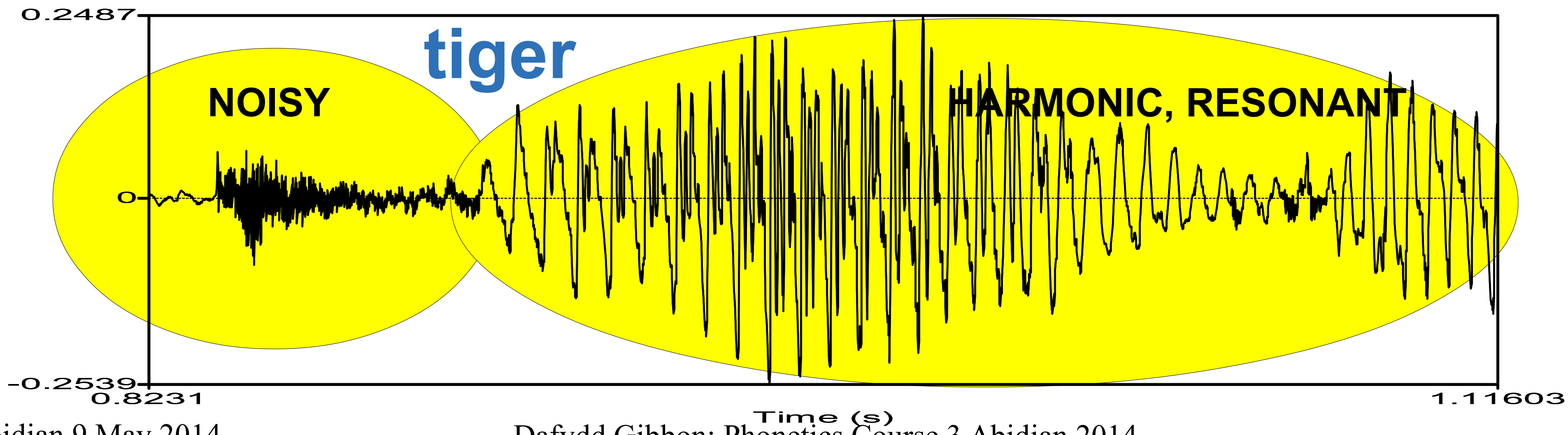
The Frequency Structure of Speech

- The SOURCE
  - for harmonic, voiced sounds
  - is the larynx (‘voicebox’, ‘Adam’s apple’)

- The larynx produces:
  - an approximately triangular complex waveform, consisting of
    - a fundamental frequency
      - about 80 Hz - 150 Hz for men (greater range possible)
      - about 160 Hz - 300 Hz for women (greater range possible)
    - many overtones, which are audible up to about 20 kHz
    - different intensities of overtones, relative to each other, which determines the overall waveform, and therefore the timbre or quality of the sound which the source produces
    - during voicing, the larynx generates a waveform which is rather like a “sawtooth” sequence

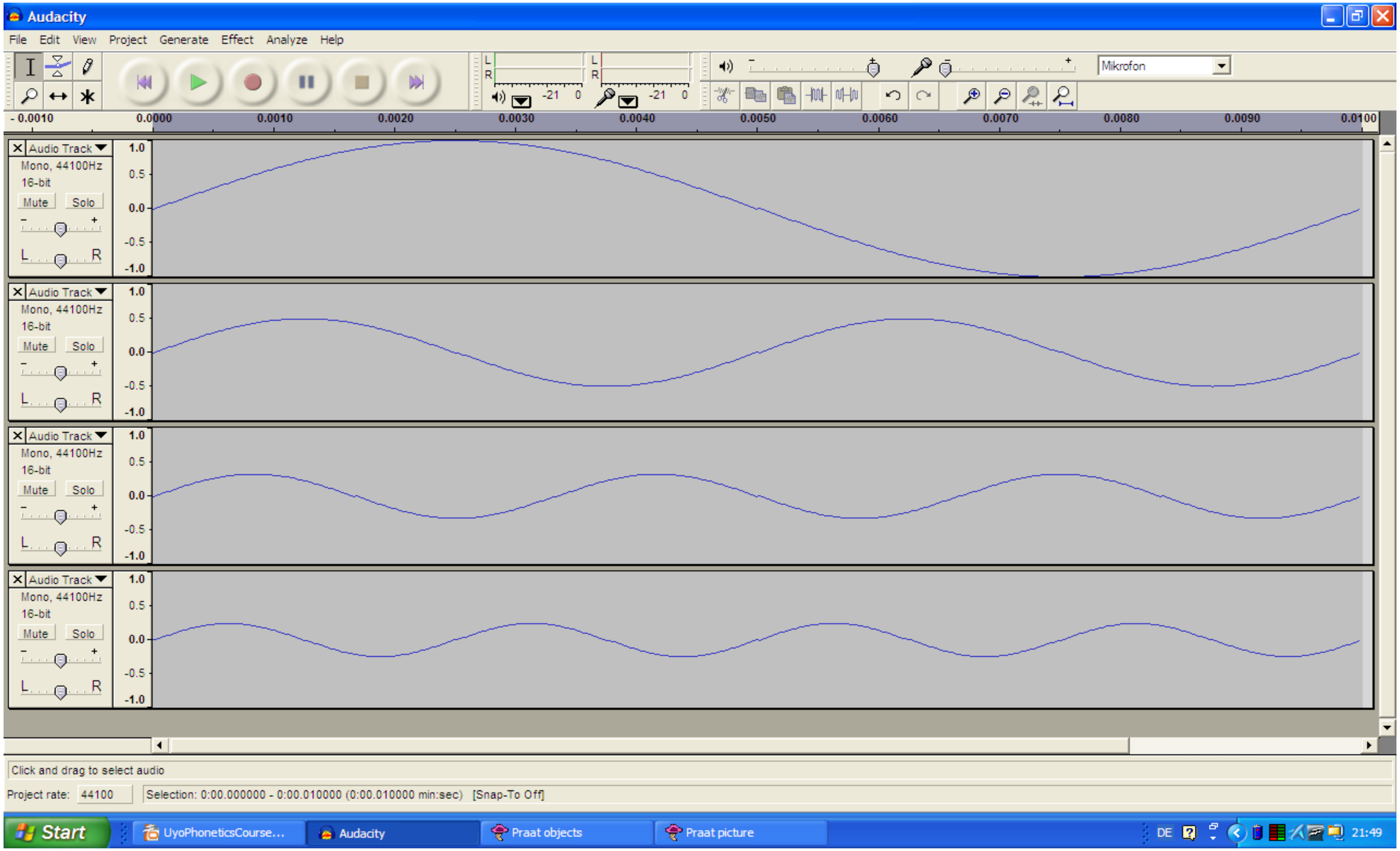
Complex Sources: noisy & harmonic signals

- If many sine waves of arbitrary frequencies occur together, the result is NOISE.
- If many sine waves occur together, 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



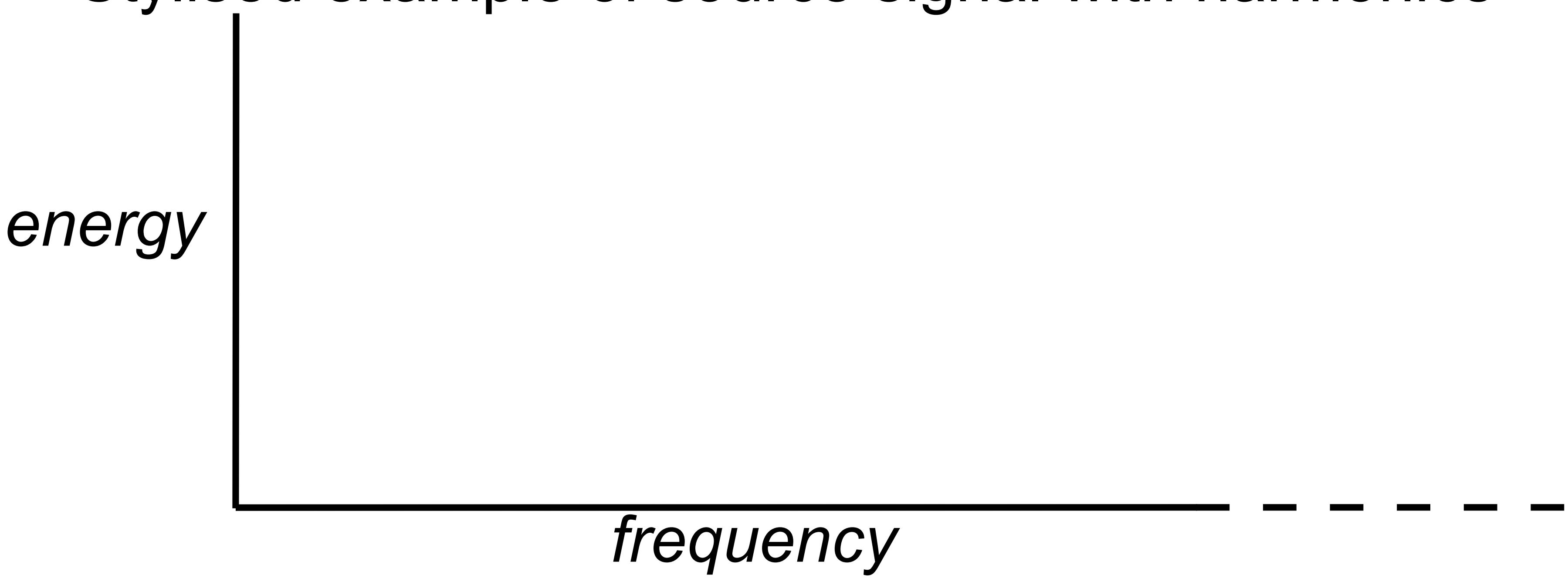
Sources with Integer Multiples of Sine Waves

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



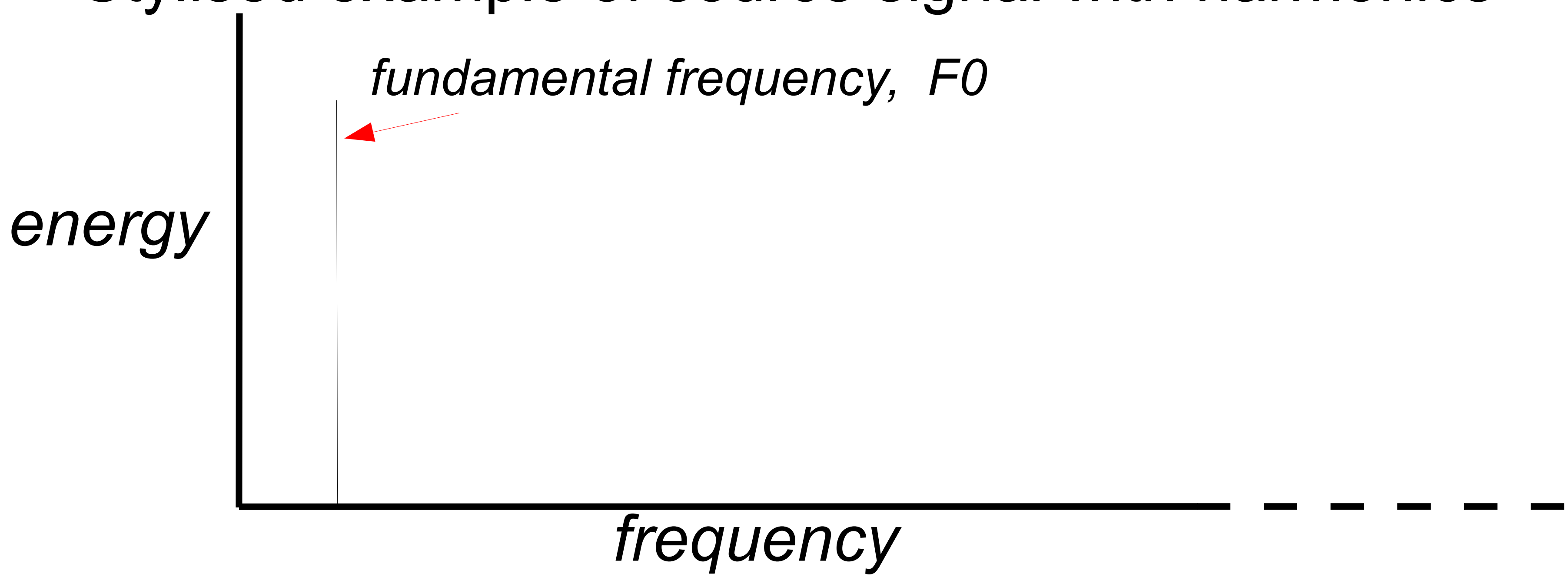
Harmonics / overtones in complex signals

- 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



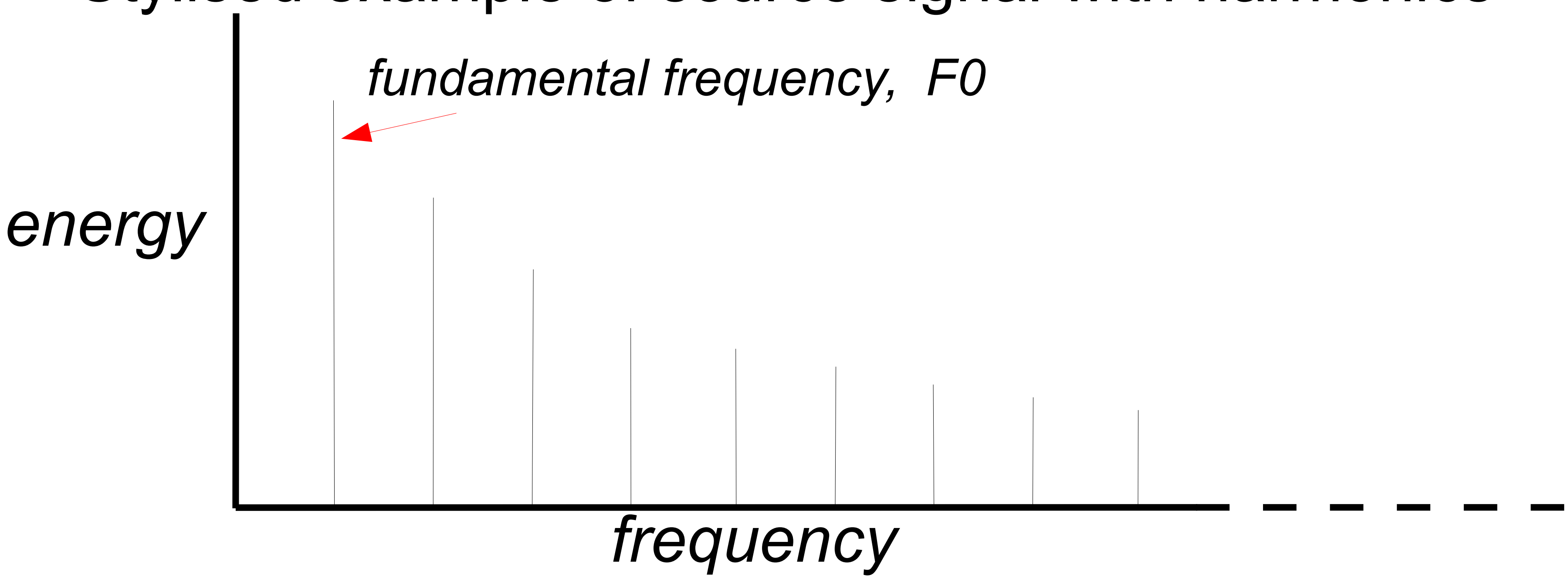
The Spectrum of Complex Signals

- 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



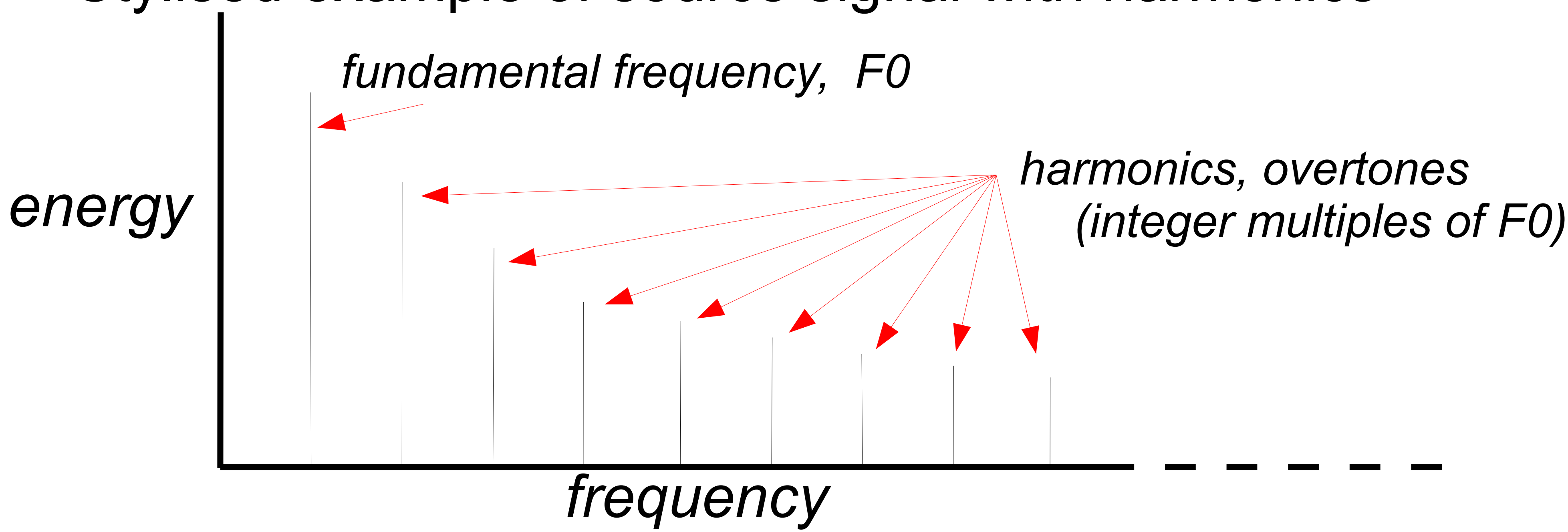
The Spectrum of Complex Signals

- If a complex signal consists of
  - a series of sine waves with frequencies of *f*, *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



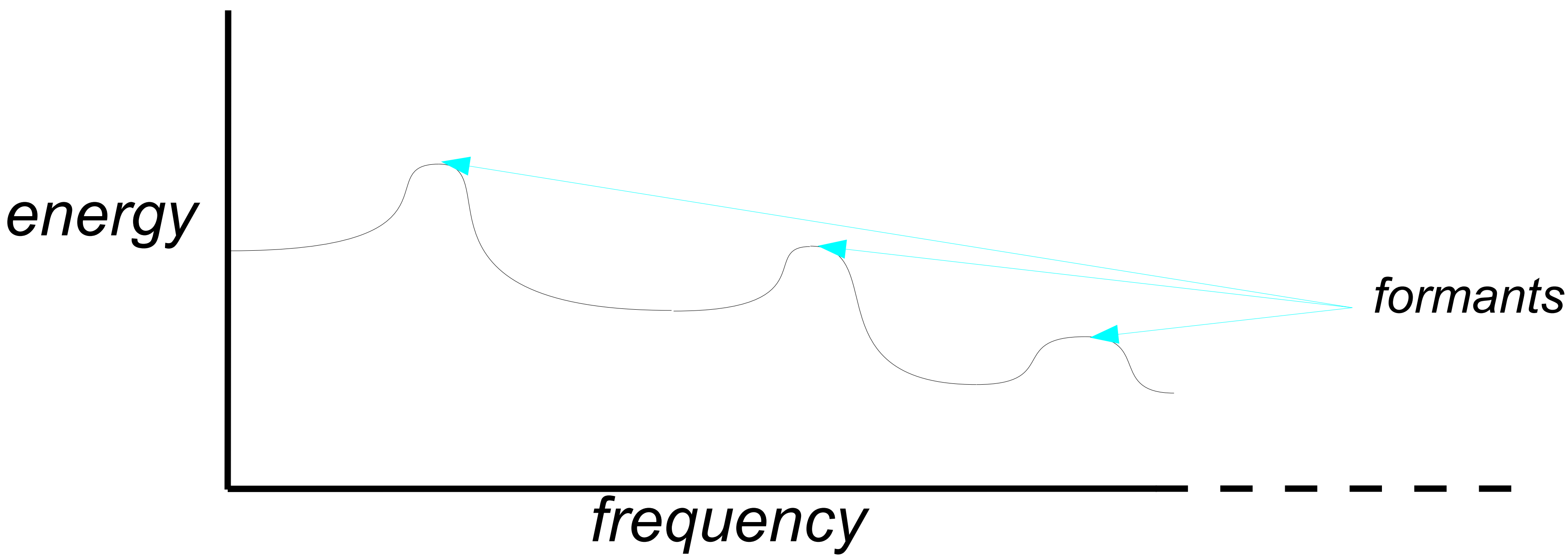
The Spectrum of Complex Signals

- 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



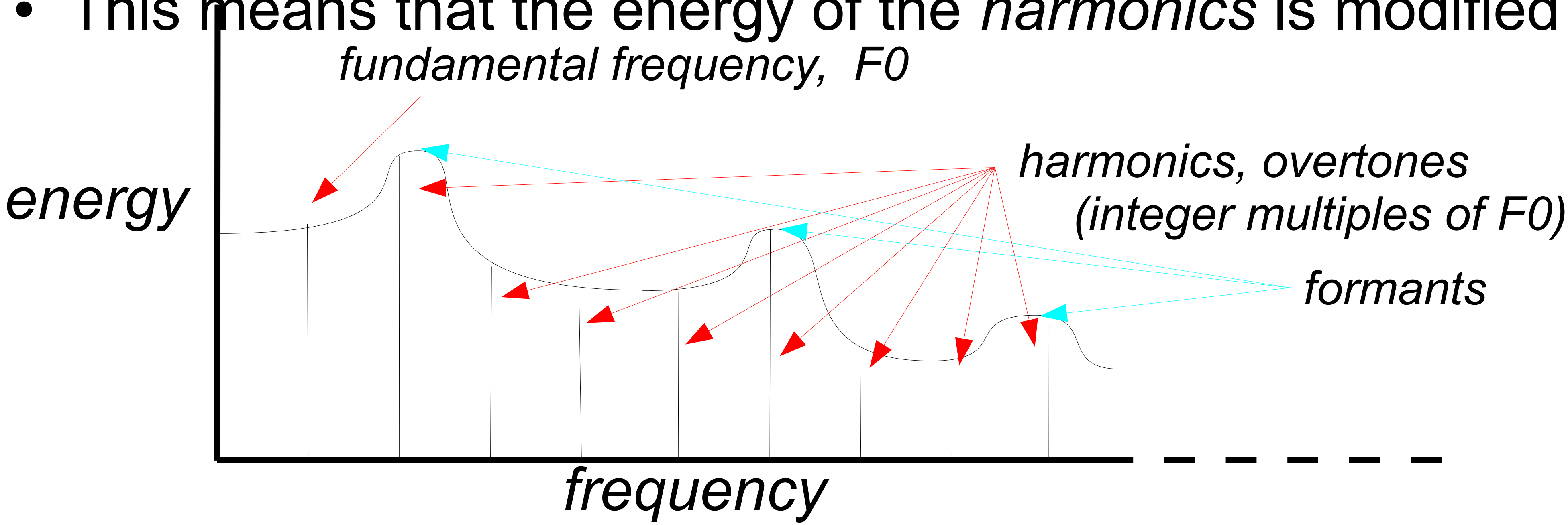
The Spectrum of Complex Signals

- The filter system consists of pharyngeal, nasal, oral cavities, with resonant frequencies which amplify or damp the overtones with these 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*)



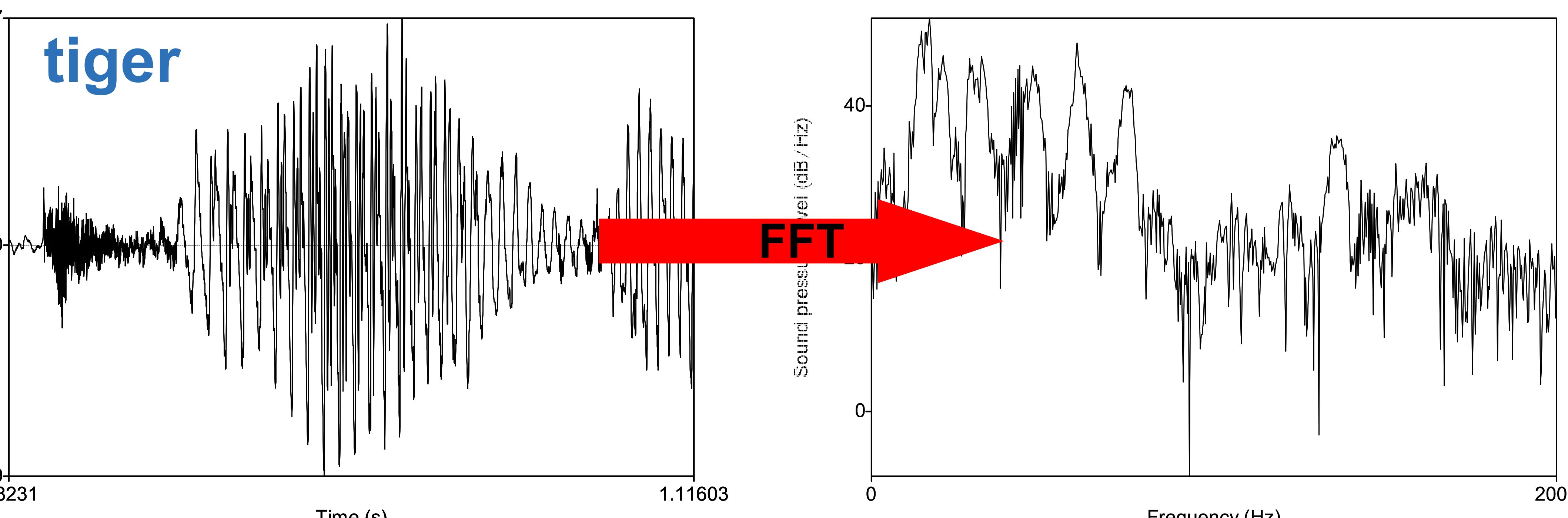
The Spectrum of Complex Signals

- The filter system consists of pharyngeal, nasal, oral cavities, with resonant frequencies which amplify or damp the overtones with these 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*)
- 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:



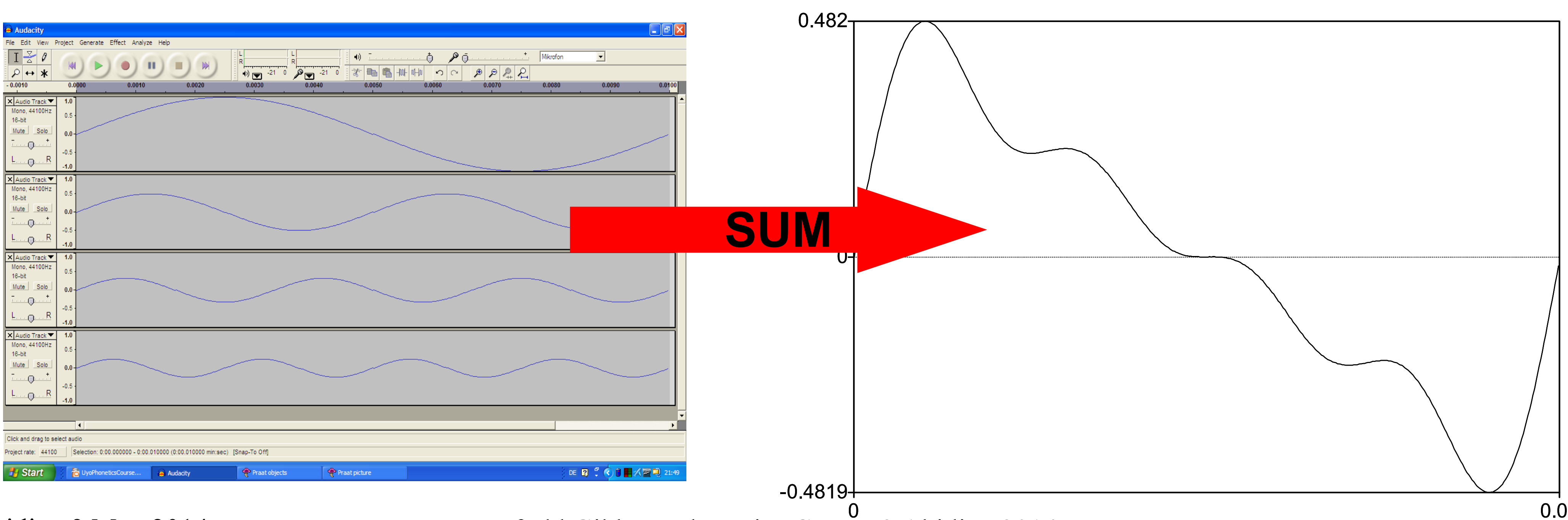
The Speech Sound Source: sawtooth waveforms

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

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

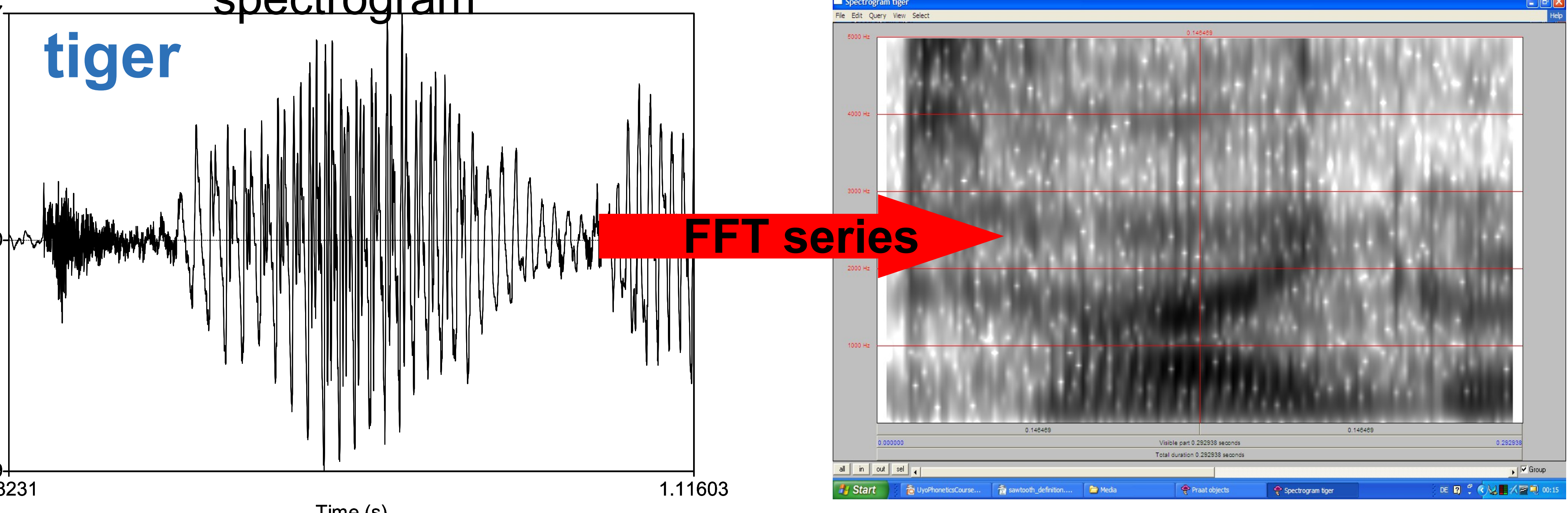
- This example illustrates the sum of four sine waves:

  100 Hz + 200 Hz + 300 Hz + 400 Hz

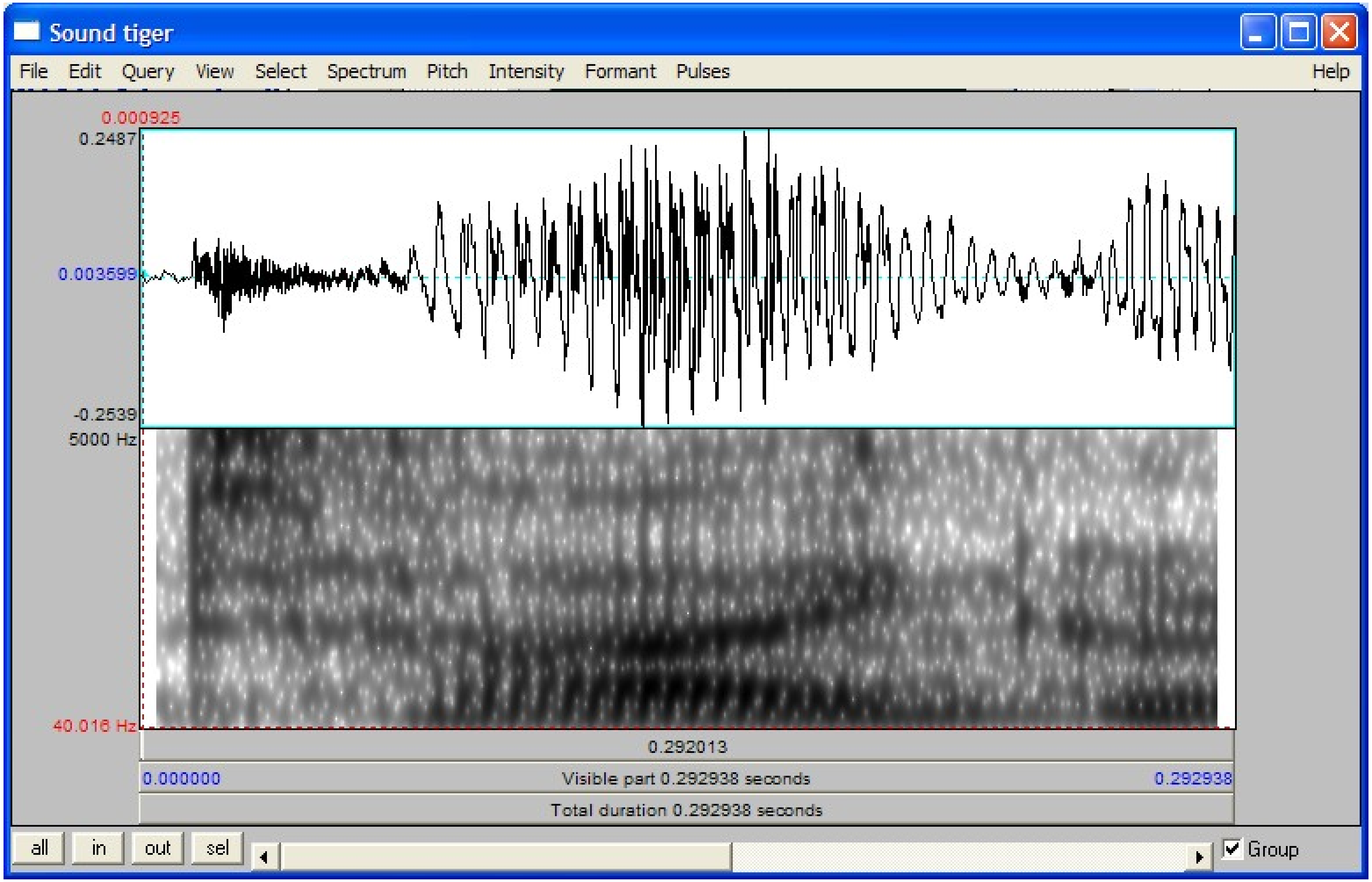


Fourier Analysis: the Spectrogram

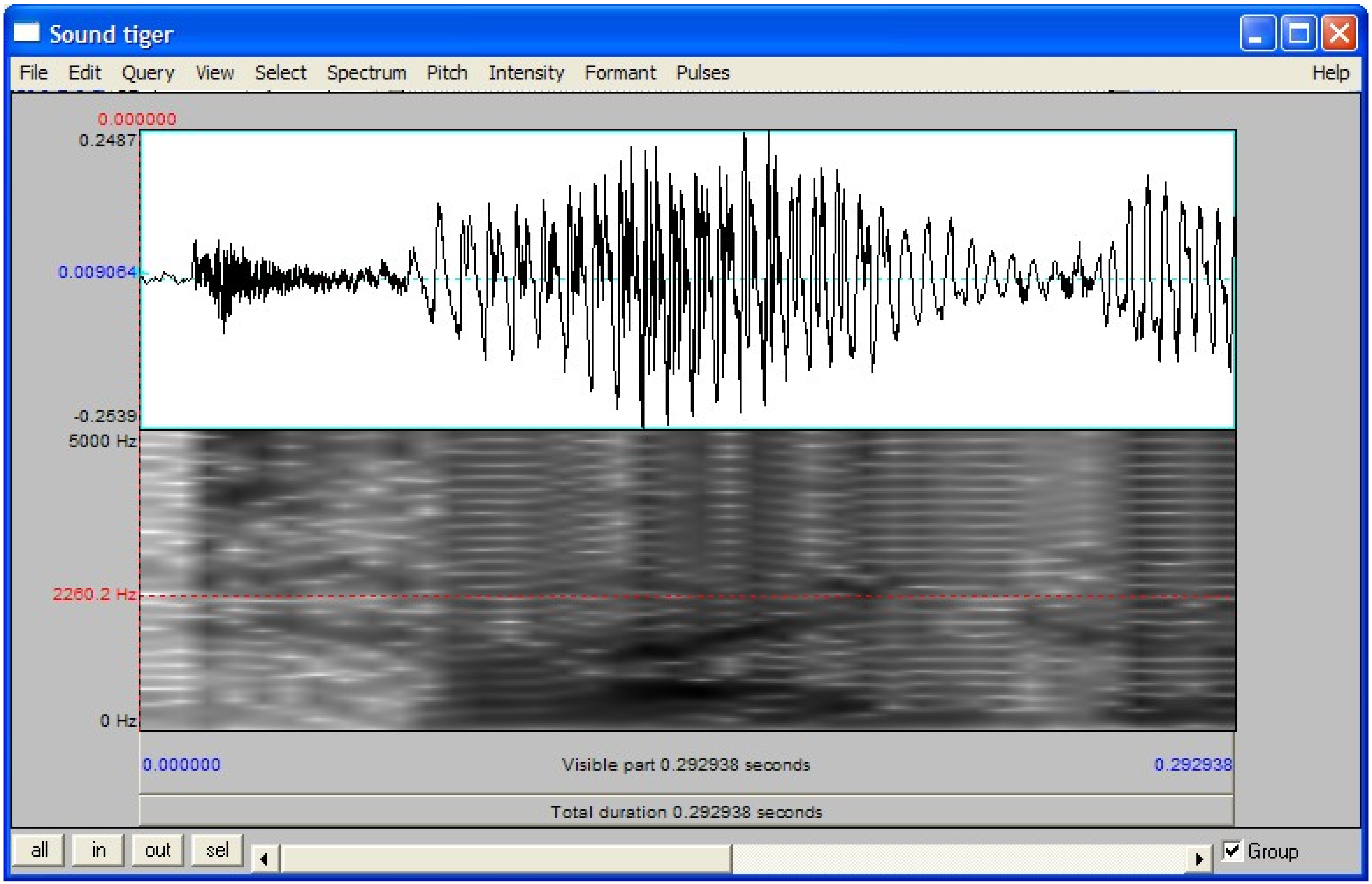
- A single spectral analysis of an interval in a speech signal, yields a spectrum and requires a at leat one period:
- In order to track the changing structure of a speech signal, a sequence of spectra is needed.
  - A representation of a sequence of spectra is called a spectrogram



Broad band spectrogram



Narrow band spectrogram

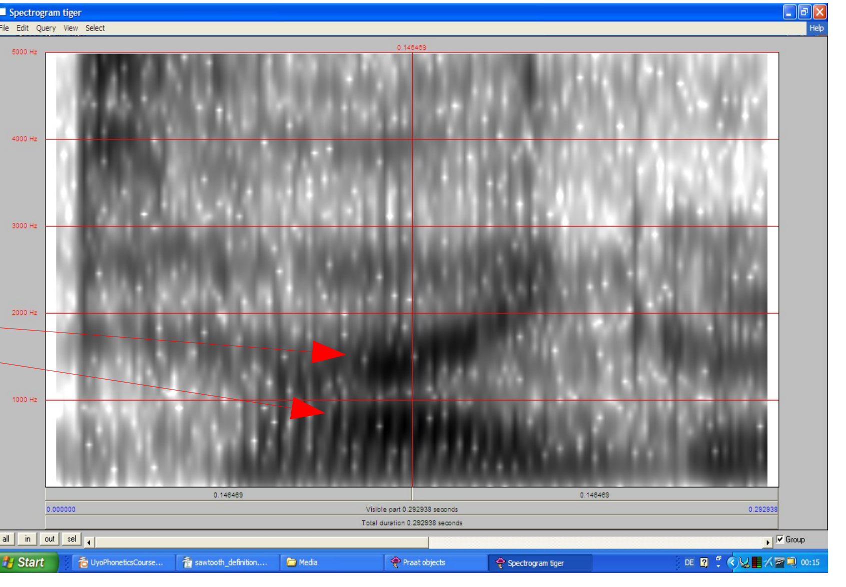


Spectrogram Filtering: Formants

- 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

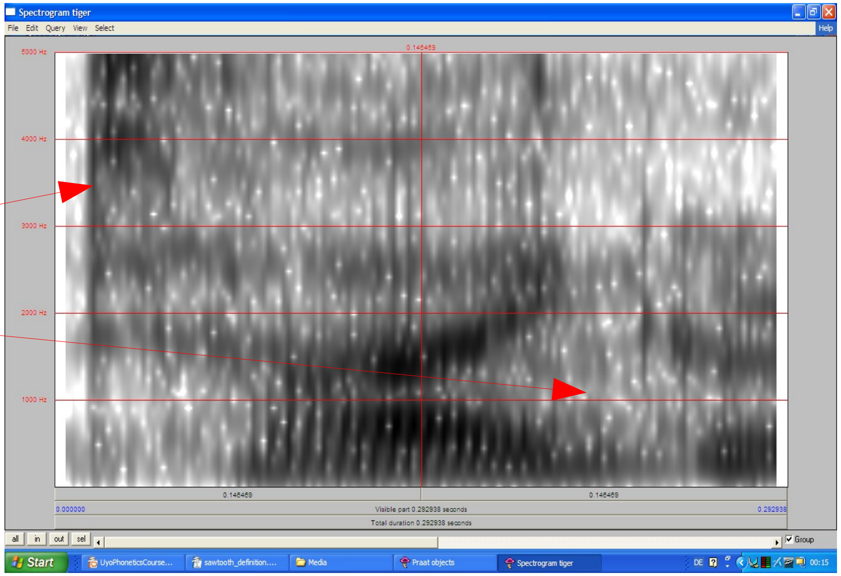


Spectrogram Filtering: Consonantal Noise

- 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

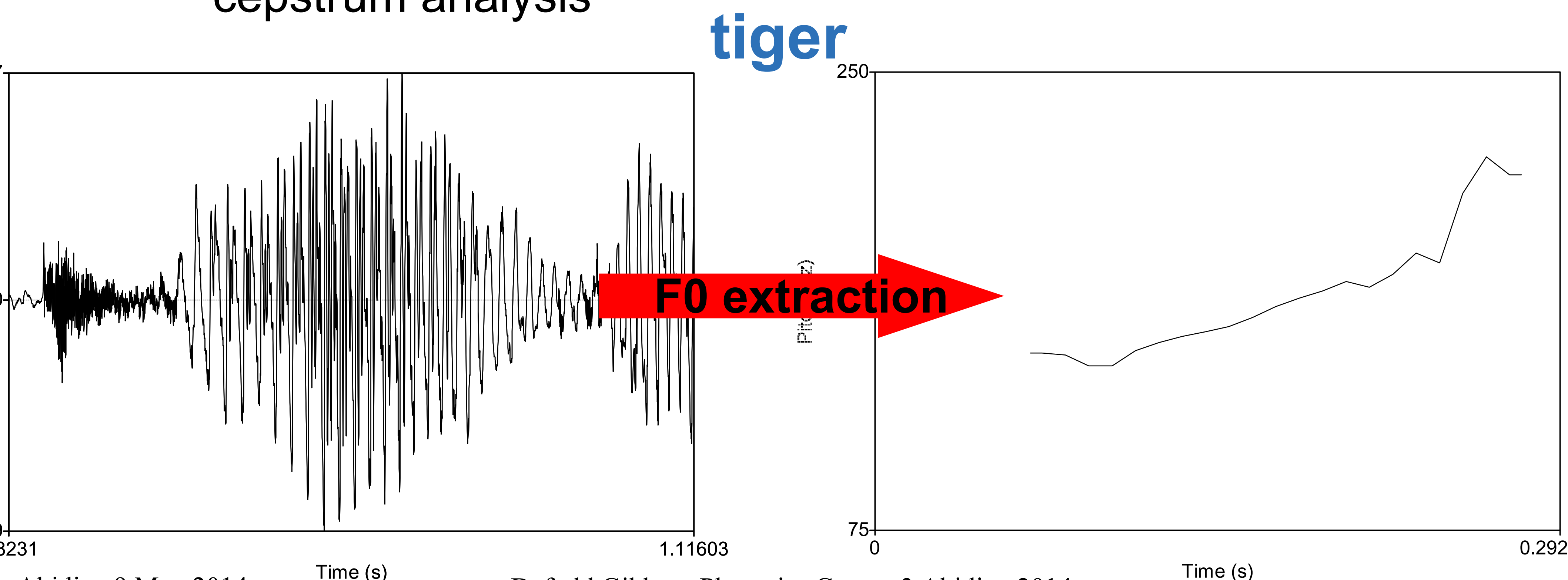
tiger

CONSONANTS



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

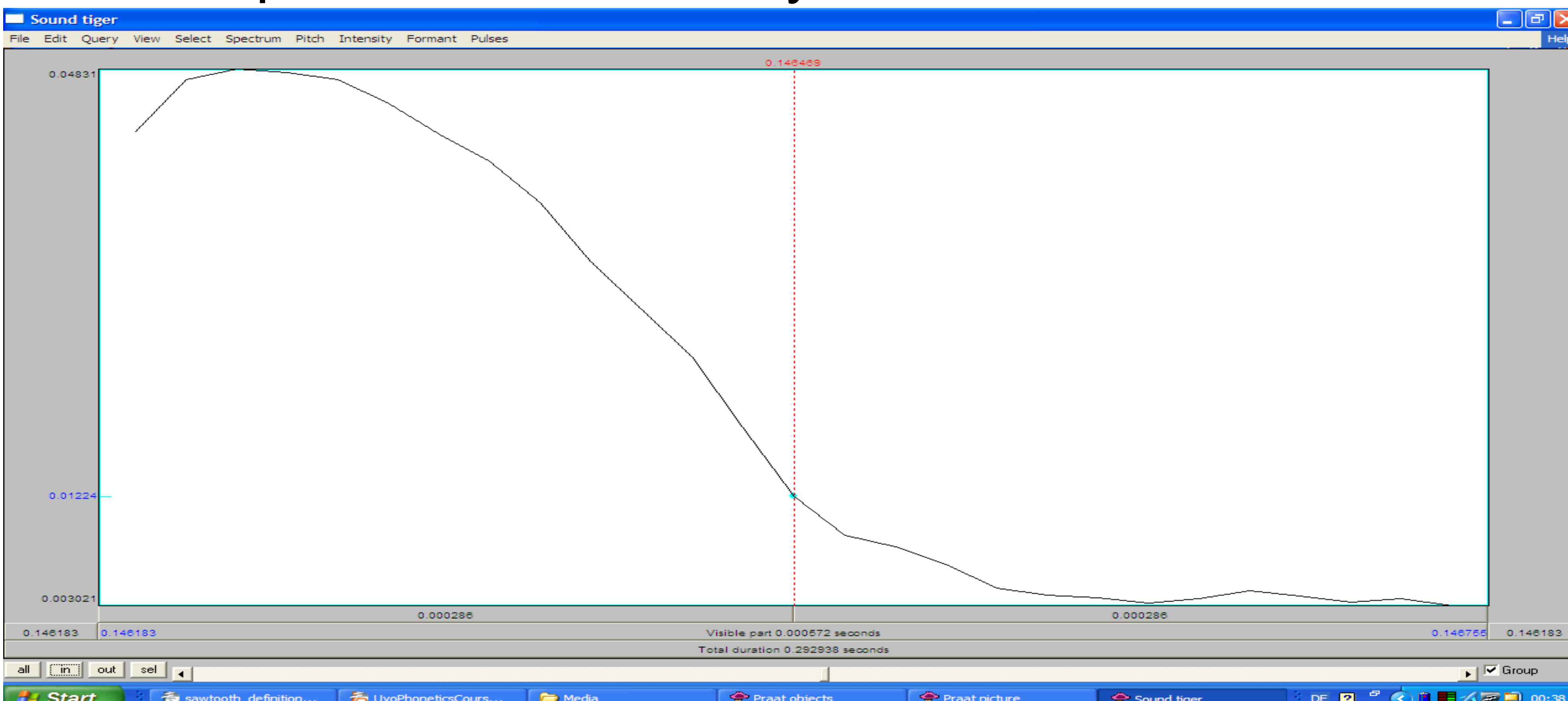


Analog-to-Digital (A/D) Conversion

- 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 (prime numbers)
    - 48 kHz (DAT tapes)
    - 22.05 kHz (laboratory recordings)
    - .... (other sampling rates, e.g. 16 Hz, 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

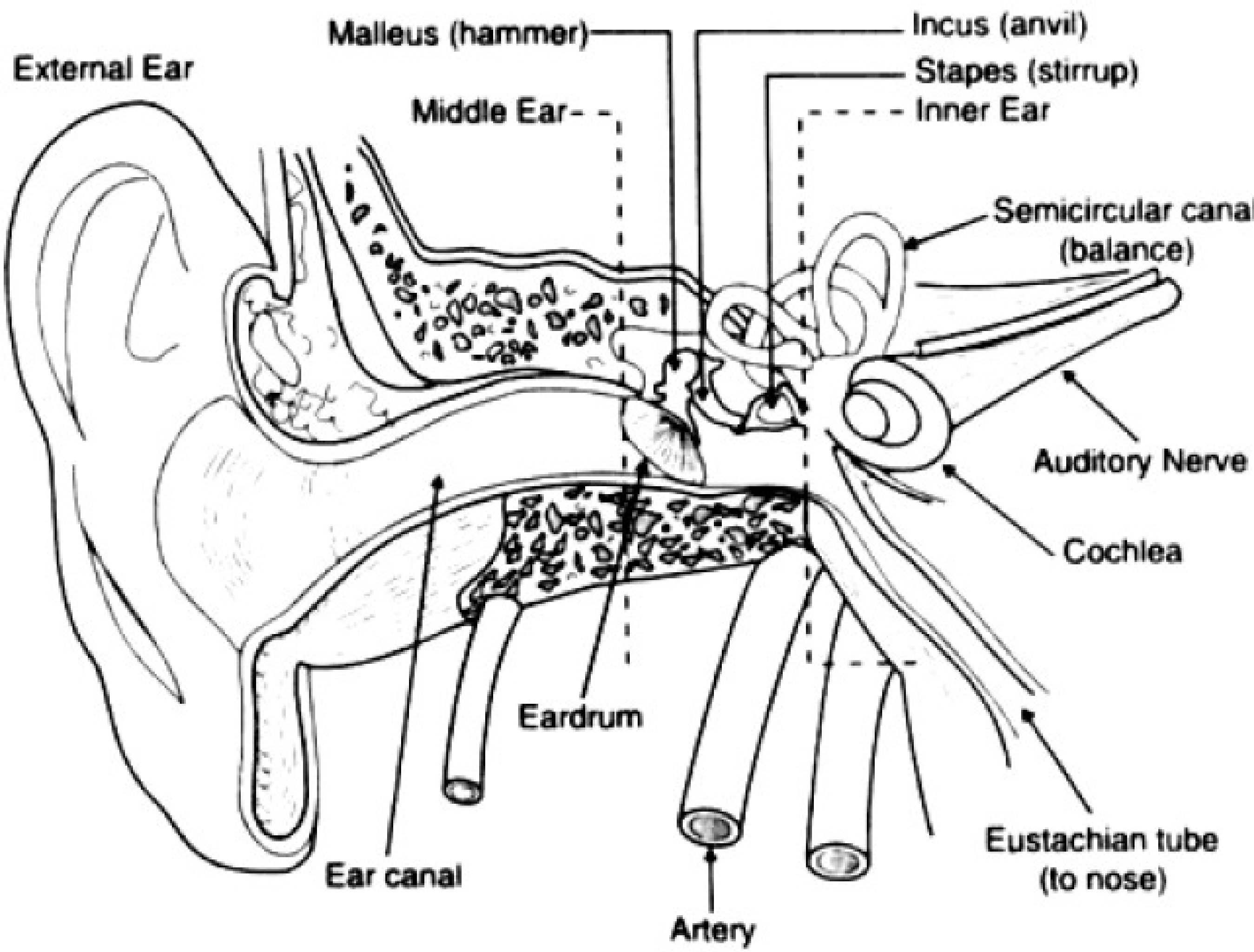


Quiz on Acoustic Phonetics

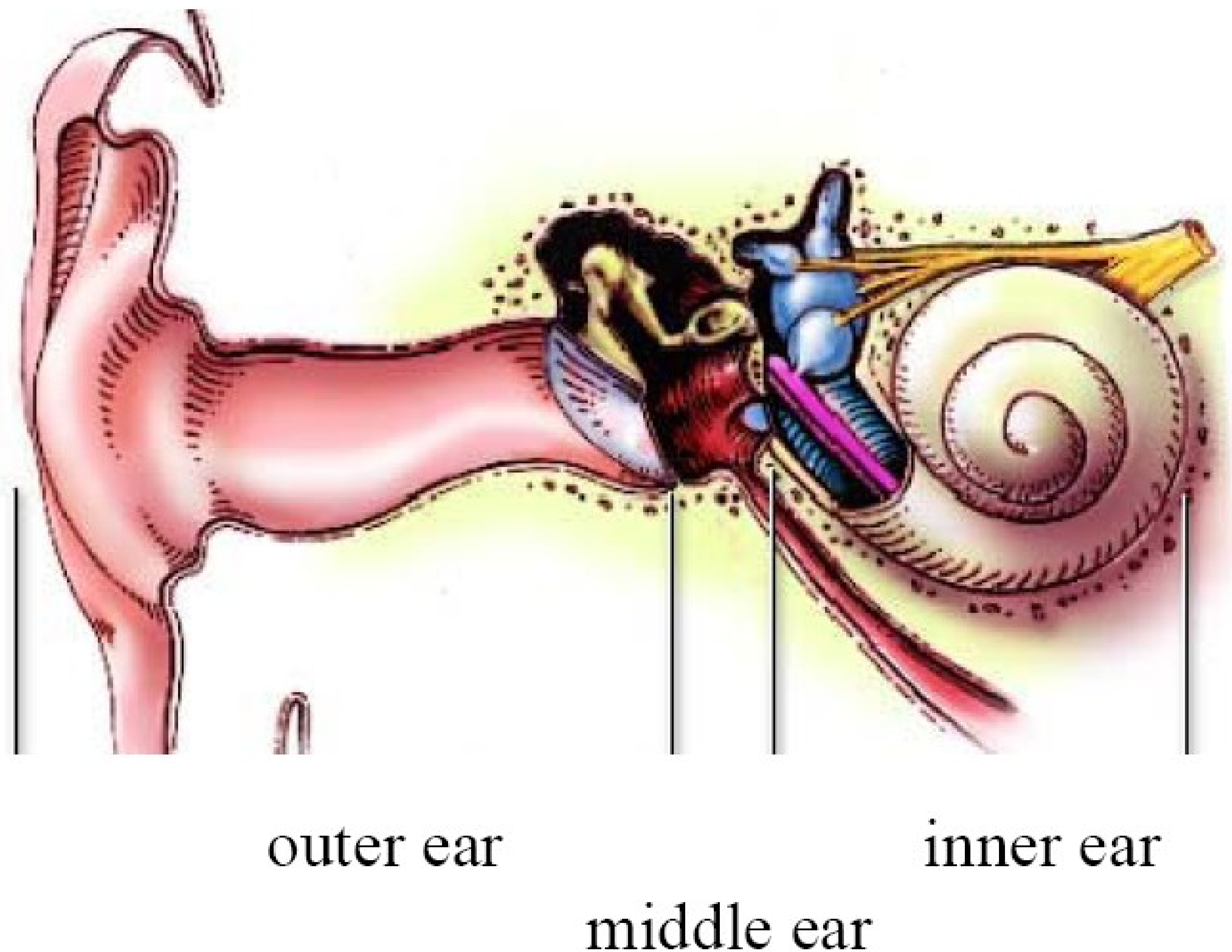
- What are the basic parameters of the speech signal?
- Define the following terms:
  - amplitude
  - intensity
  - energy
- How are time-domain representations of speech signal converted to frequency domain representations?
- Define the following terms:
  - spectrum
  - spectrogram
  - fundamental frequency,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

Auditory Phonetics (Speech Perception)

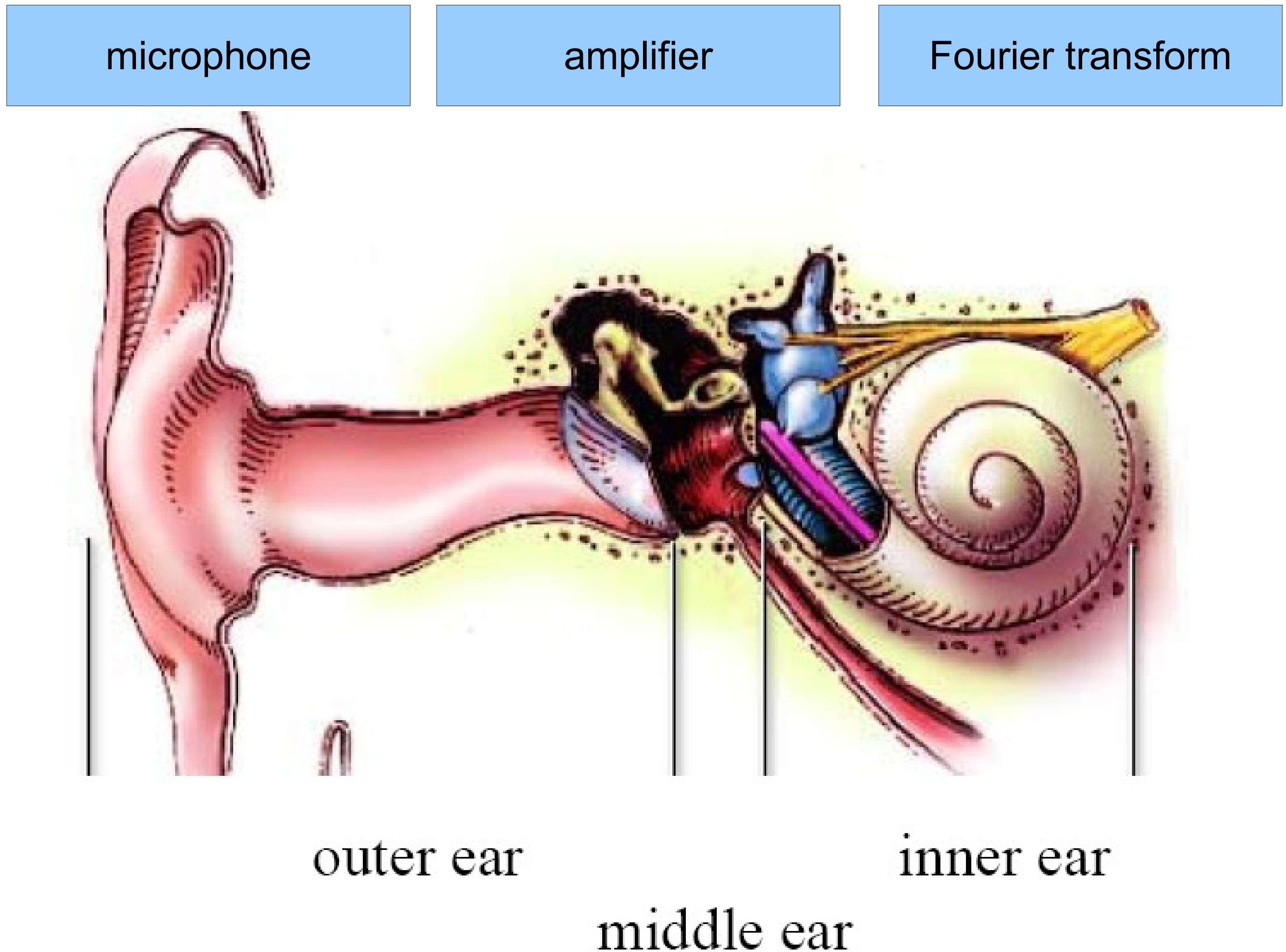
The Auditory Domain: Anatomy of the Ear



The Auditory Domain: Anatomy of the Ear



The Auditory Domain: Anatomy of the 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