Prosody: speech rhythms and melodies

3. Phonetic Basics

Dafydd Gibbon

Guangzhou Prosody Lectures, November 2016

Schedule

Week 1:

01 *Forms and functions of prosody: models and methods*
Nov. 2 (Wednesday) 2:30pm--4:30pm

02 *Forms and functions of prosody: prosodic semantics*
Nov. 4 (Friday) 10am--12am

Week 2:

03 ***Basics of digital phonetics***
Nov. 8 (Tuesday) 10am--12am

04 *Pitch Stylisation: tone and intonation*
Nov. 8 (Tuesday) 2:30pm--4:30pm

Week 3:

05 *Syllables and Prosody Modelling*
Nov.15 (Tuesday) 10am--12am

06 *SpeechTiming: durations and rhythm*
Nov.15 (Tuesday) 2:30pm--4:30pm

Contents

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

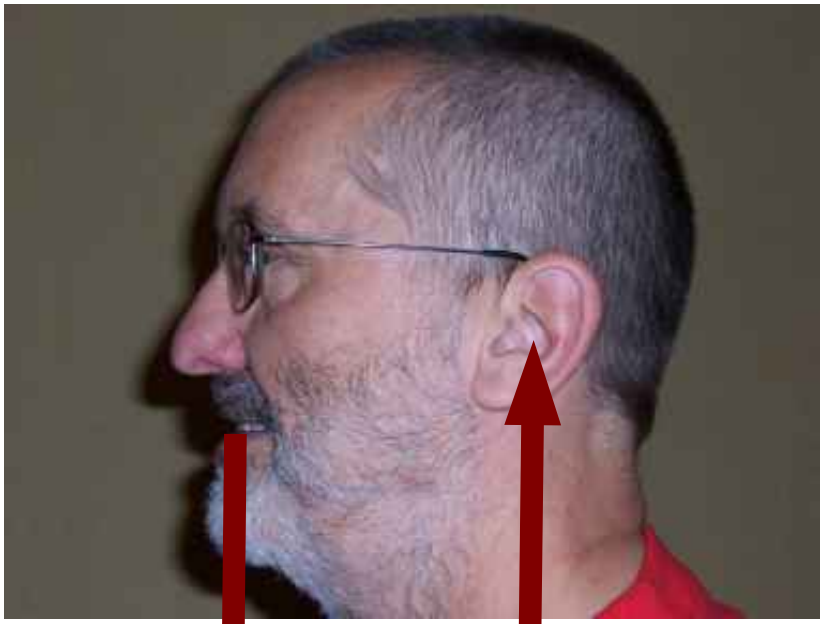
Domains and Methods of Phonetics

- 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 simple models such as articulatory phonetic categories
  - measurements based on simple models
    - of position and movement of articulatory organs
    - of the structure of speech signals
    - of the mechanisms of the ear and perception in hearing
  - statistical study of categories of observation and measurement
  - creation of more complex models:

    production, transmission, perception

Phonetic Domains

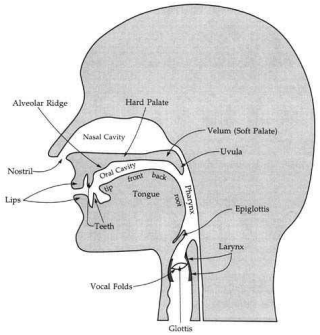
The Domains of Phonetics: the Phonetic Cycle



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

The Domains of Phonetics: the Phonetic Cycle

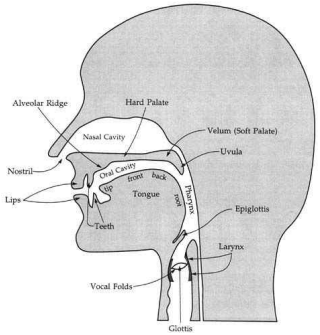
Sender: Articulatory Phonetics



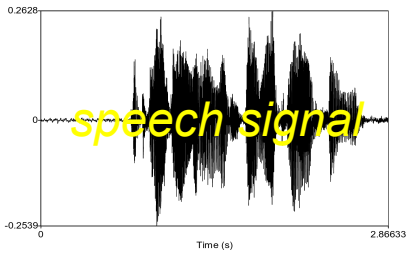
A tiger and a mouse were walking in a field...

The Domains of Phonetics: the Phonetic Cycle

Sender: Articulatory Phonetics



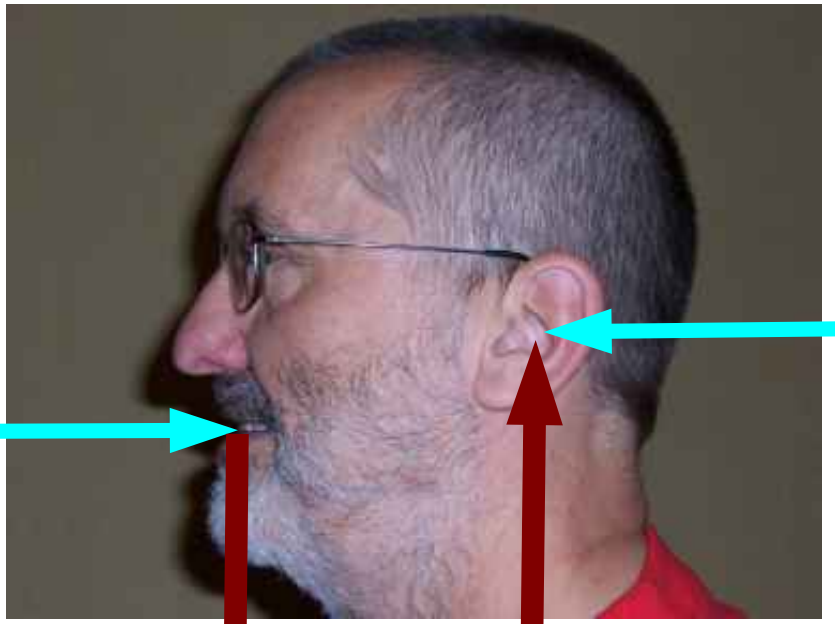
A tiger and a mouse were walking in a field...



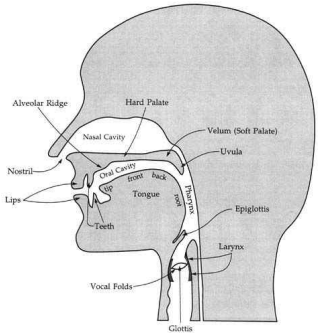
Channel: Acoustic Phonetics

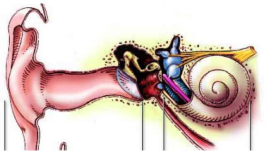
The Domains of Phonetics: the Phonetic Cycle

Sender: Articulatory Phonetics

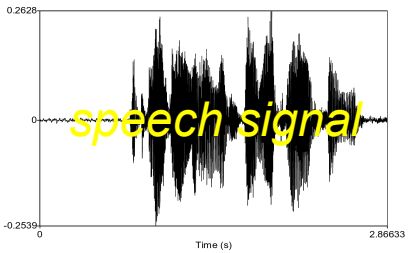


Receiver: Auditory Phonetics





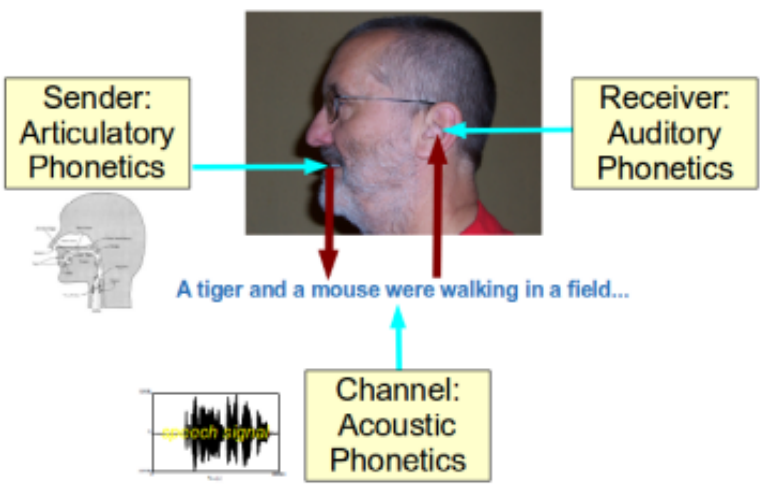
A tiger and a mouse were walking in a field...

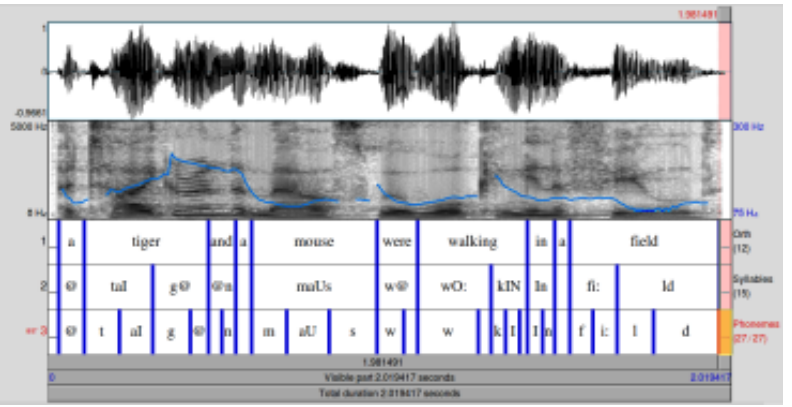


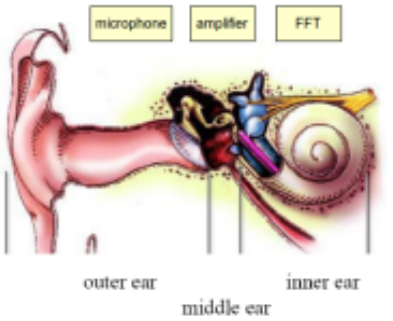
Channel: Acoustic Phonetics

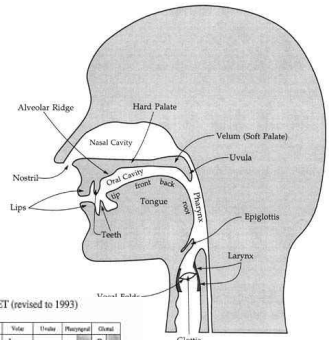
Phonetic Models

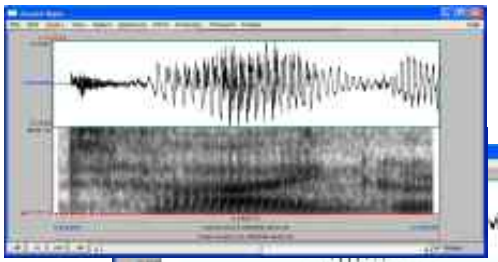
Phonetic Models

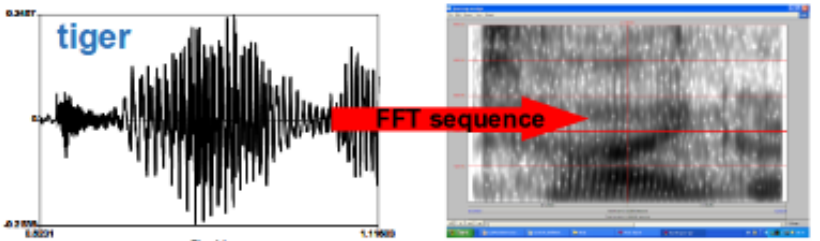


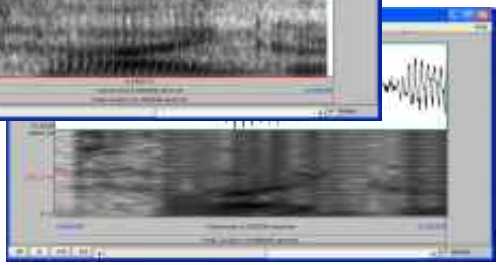


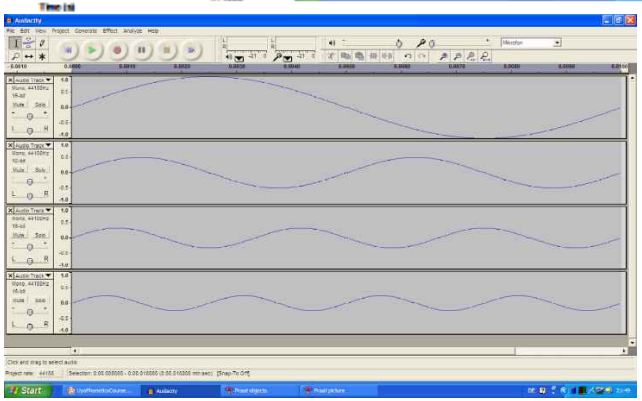


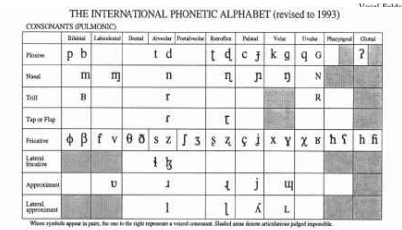


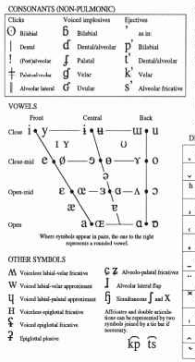


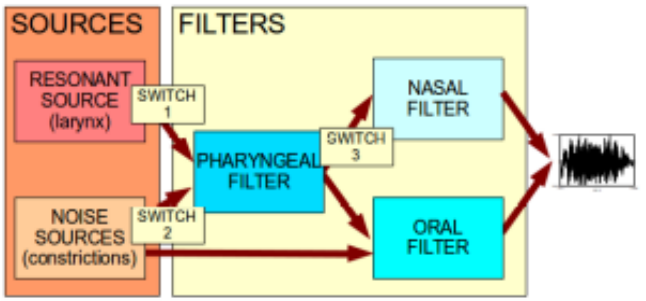


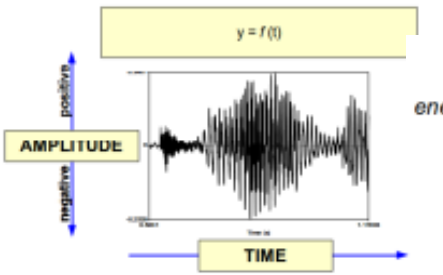


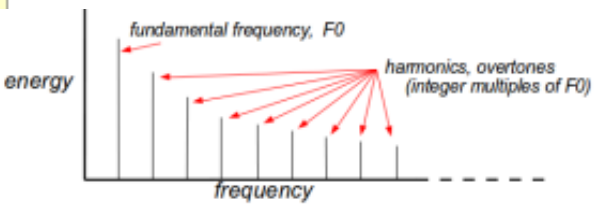










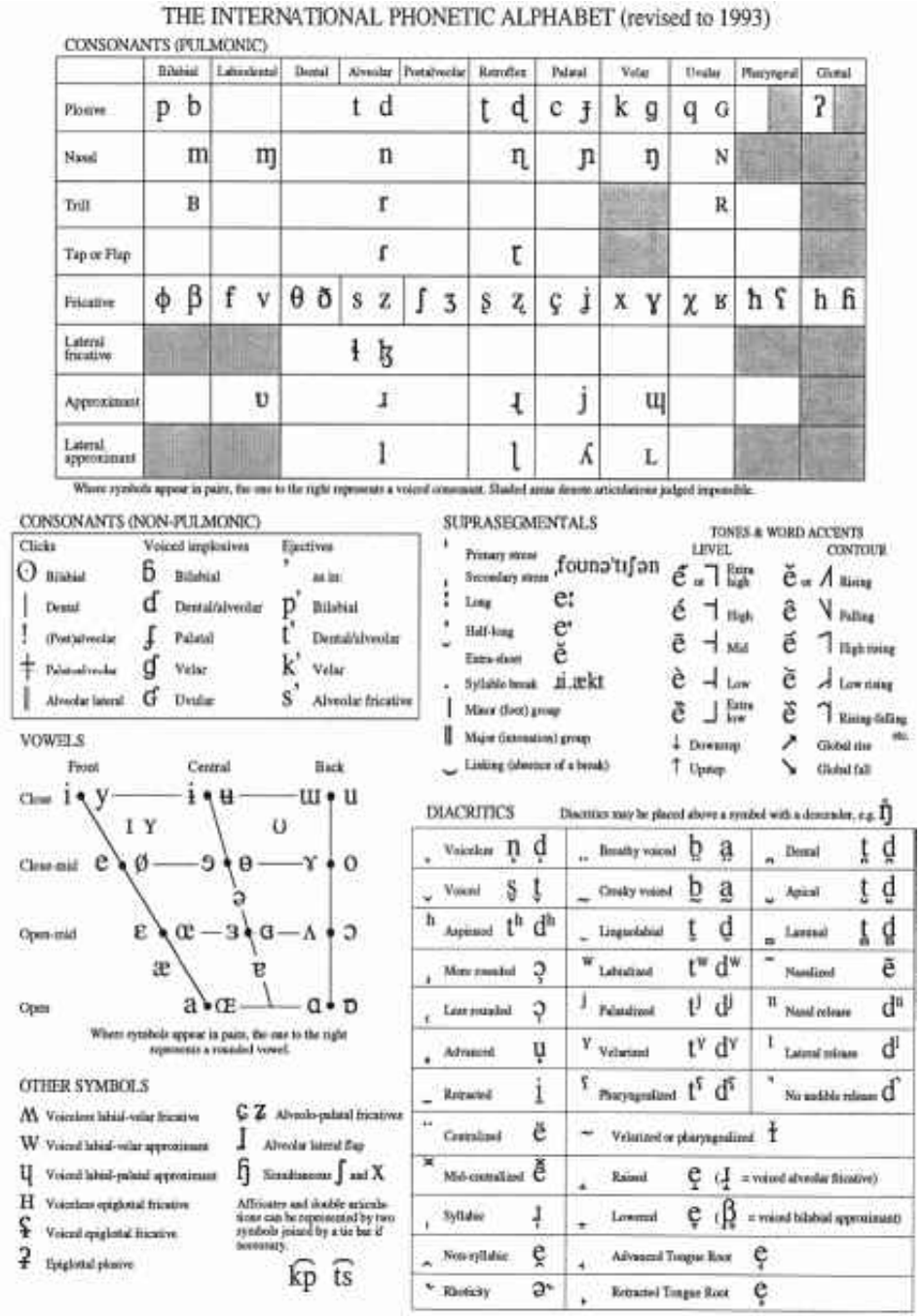


Quiz on the Phonetic Cycle

- 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)

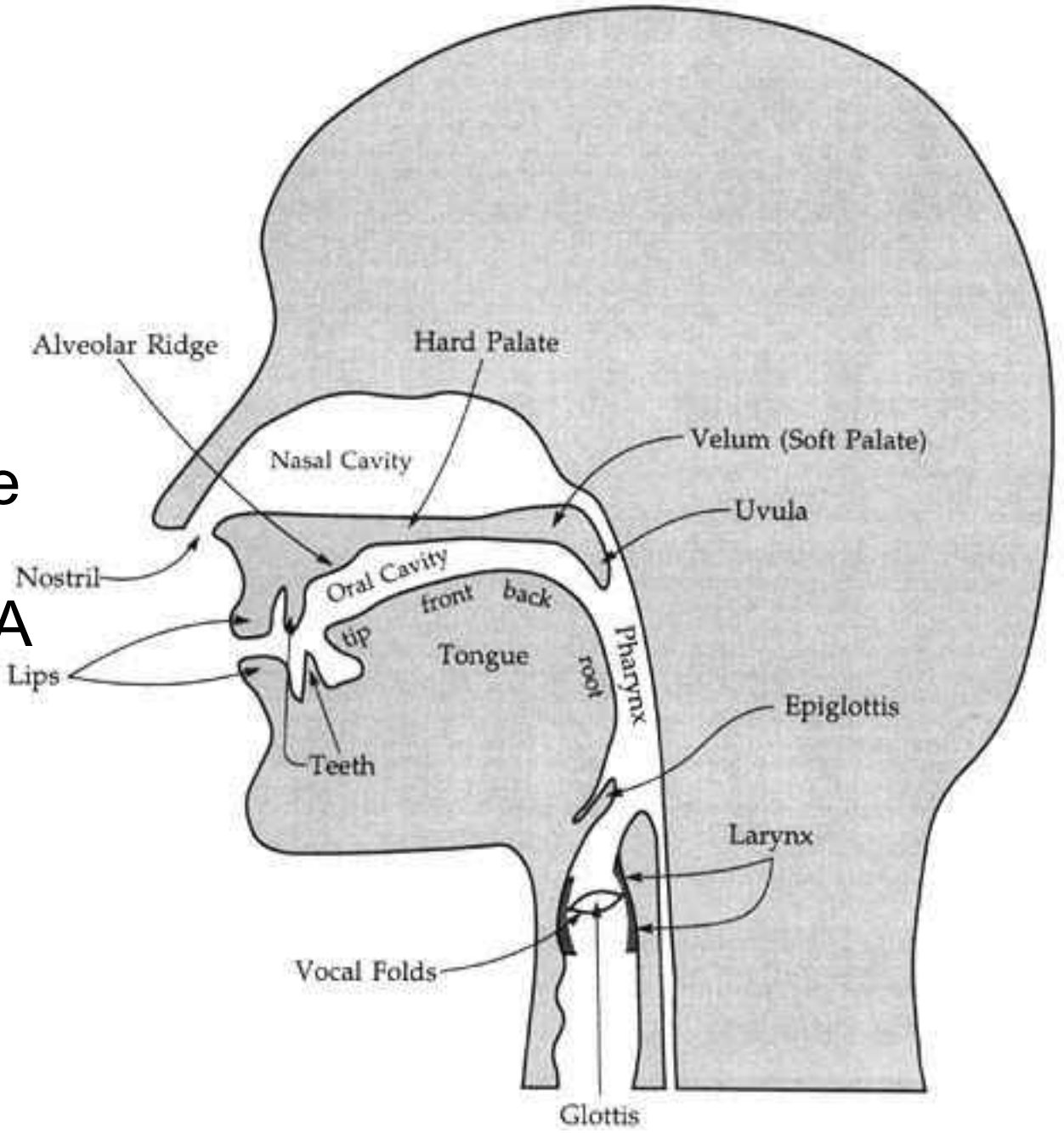
The IPA (A = Alphabet / Association)



- 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 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



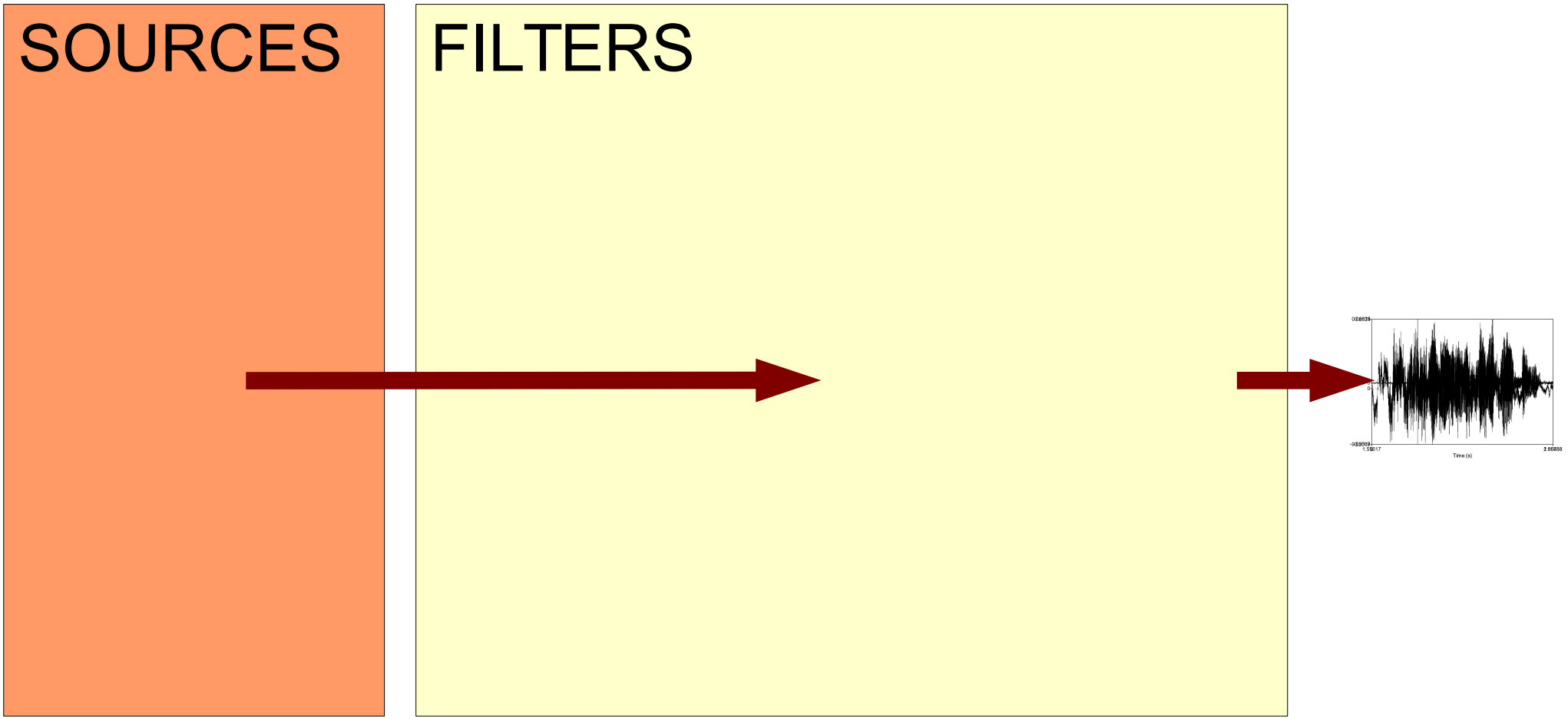
How does sound production work?

Sources, switches and filters

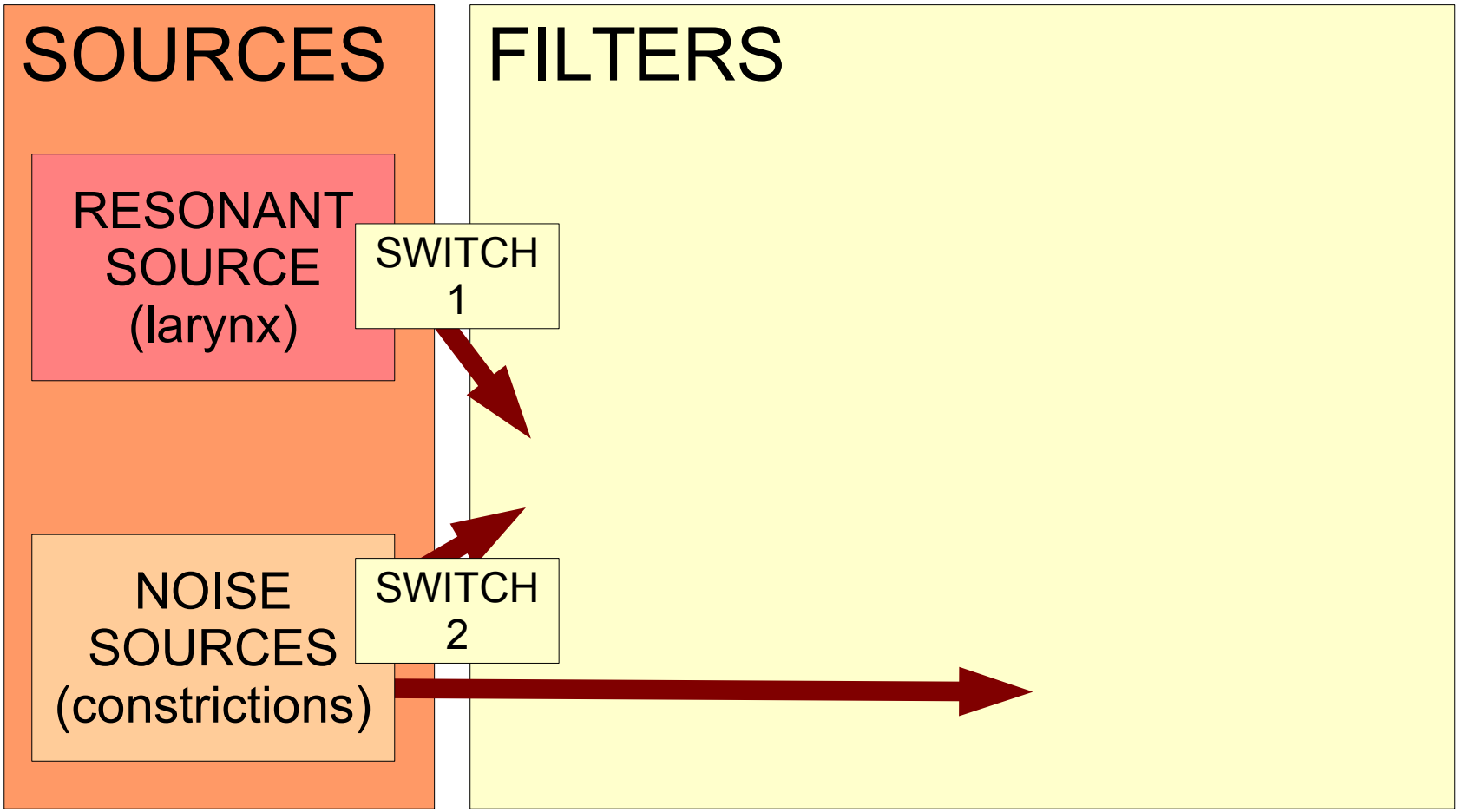
The Source-Filter Model of Speech Production

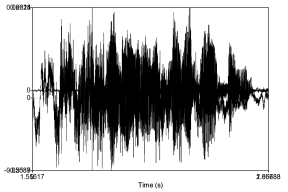
- A “model” is a simplified representation of relevant features of reality (but it also adds its own artefacts)
- In the Source-Filter Model of speech production, the sound is generated by the SOURCE and modified by the FILTER
- The Source-Filter Model represents the speech production process in two phases:
  - The SOURCE of the sound + SWITCHES:
    - 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

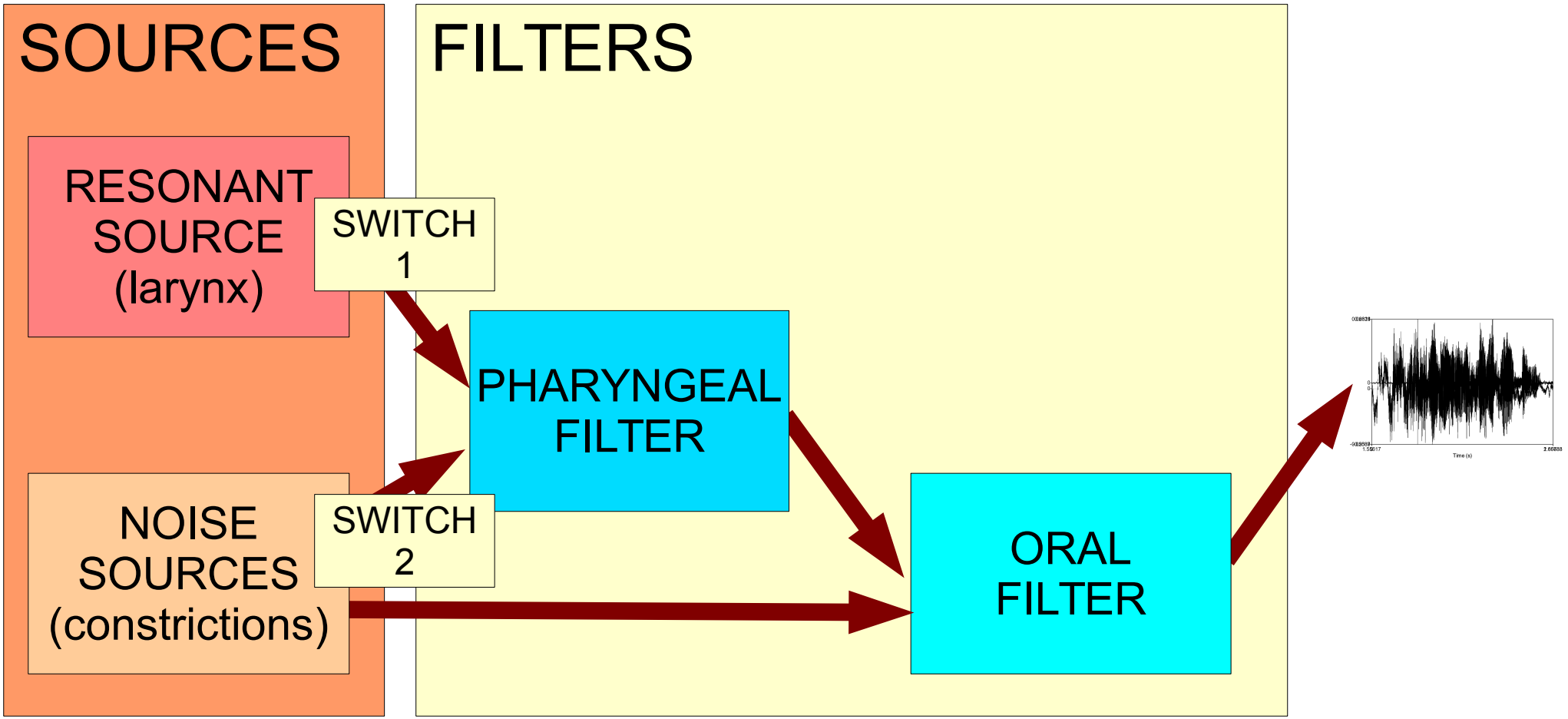


The Source-Filter Model of Speech Production

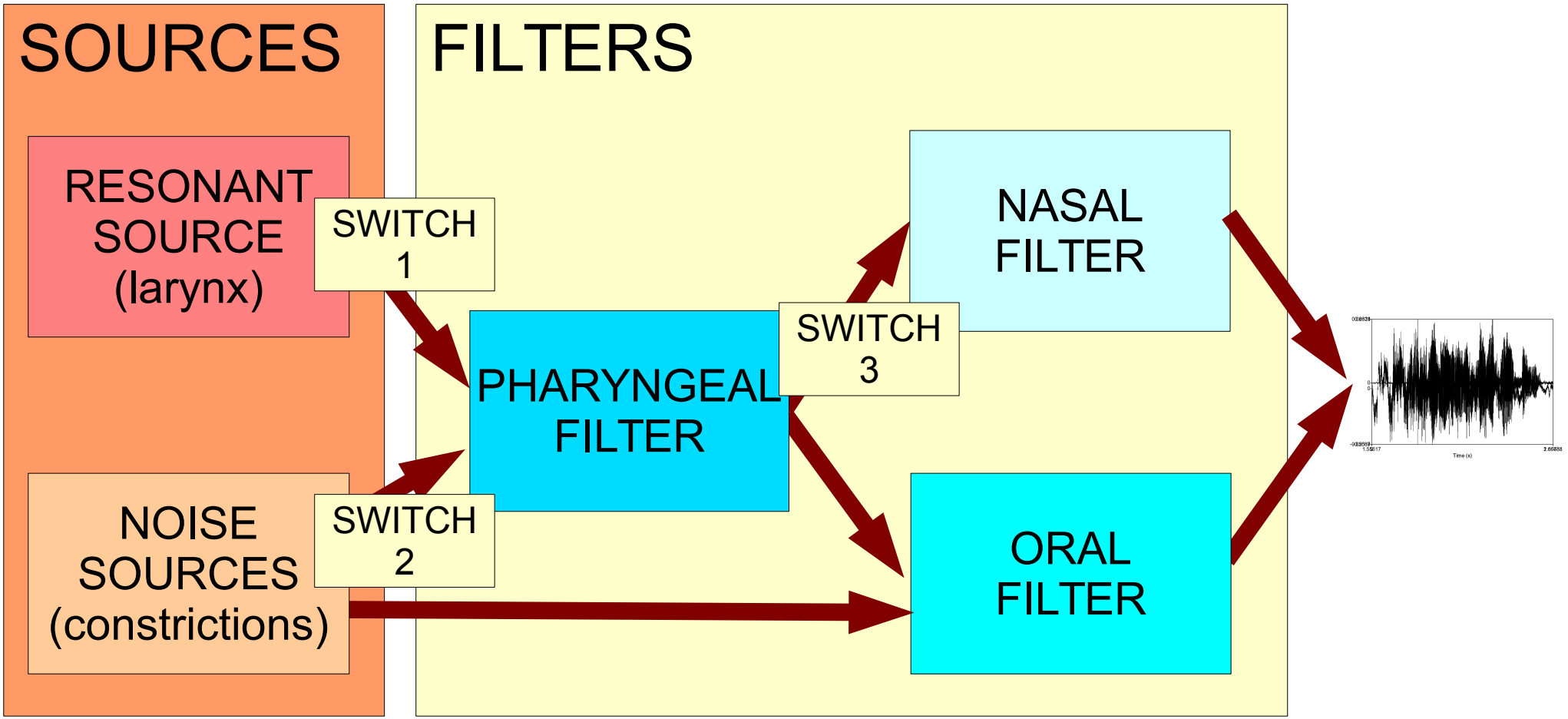




The Source-Filter Model of Speech Production

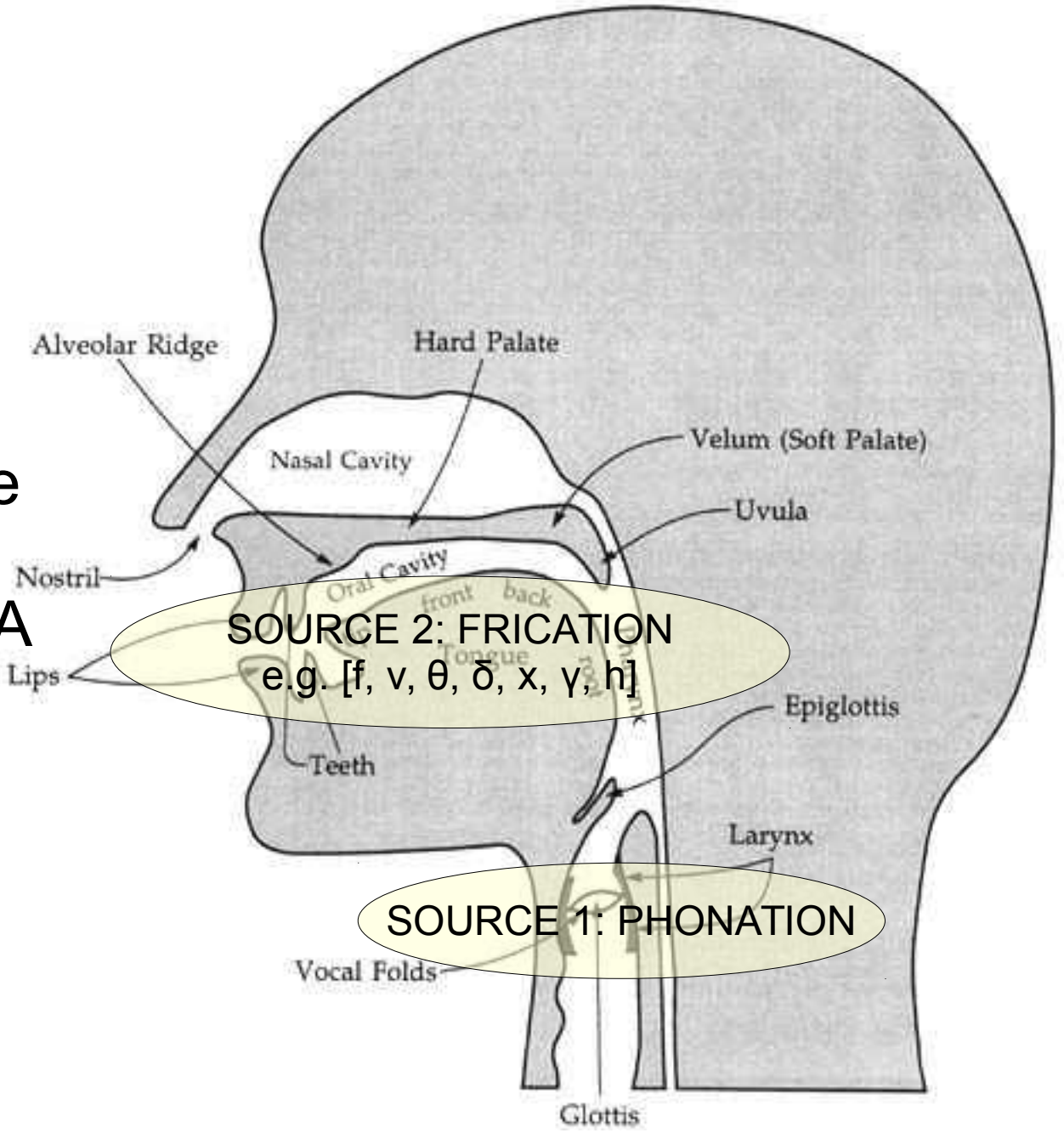


The Source-Filter Model of Speech Production



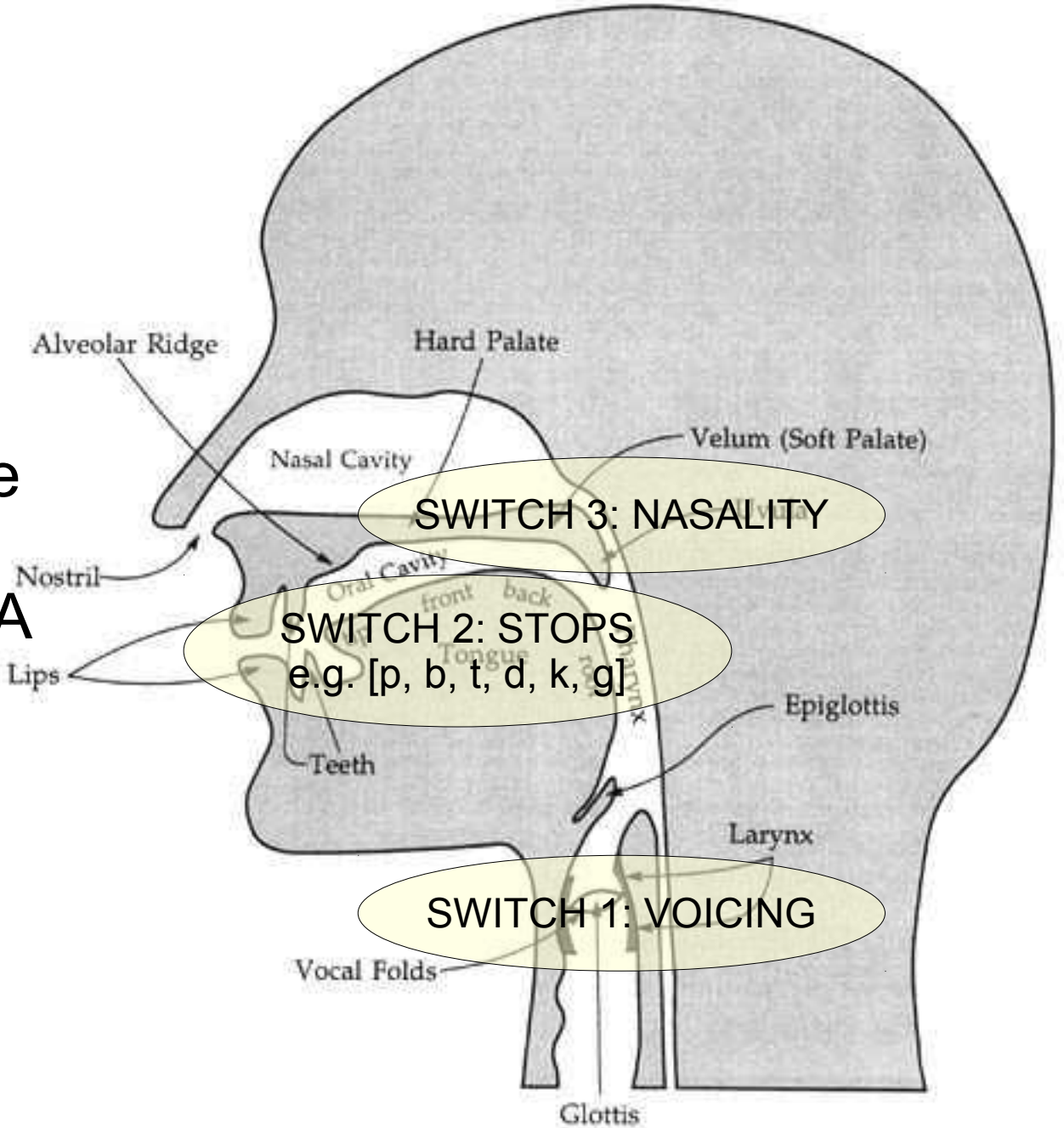
Articulatory domain: sound sources

- 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



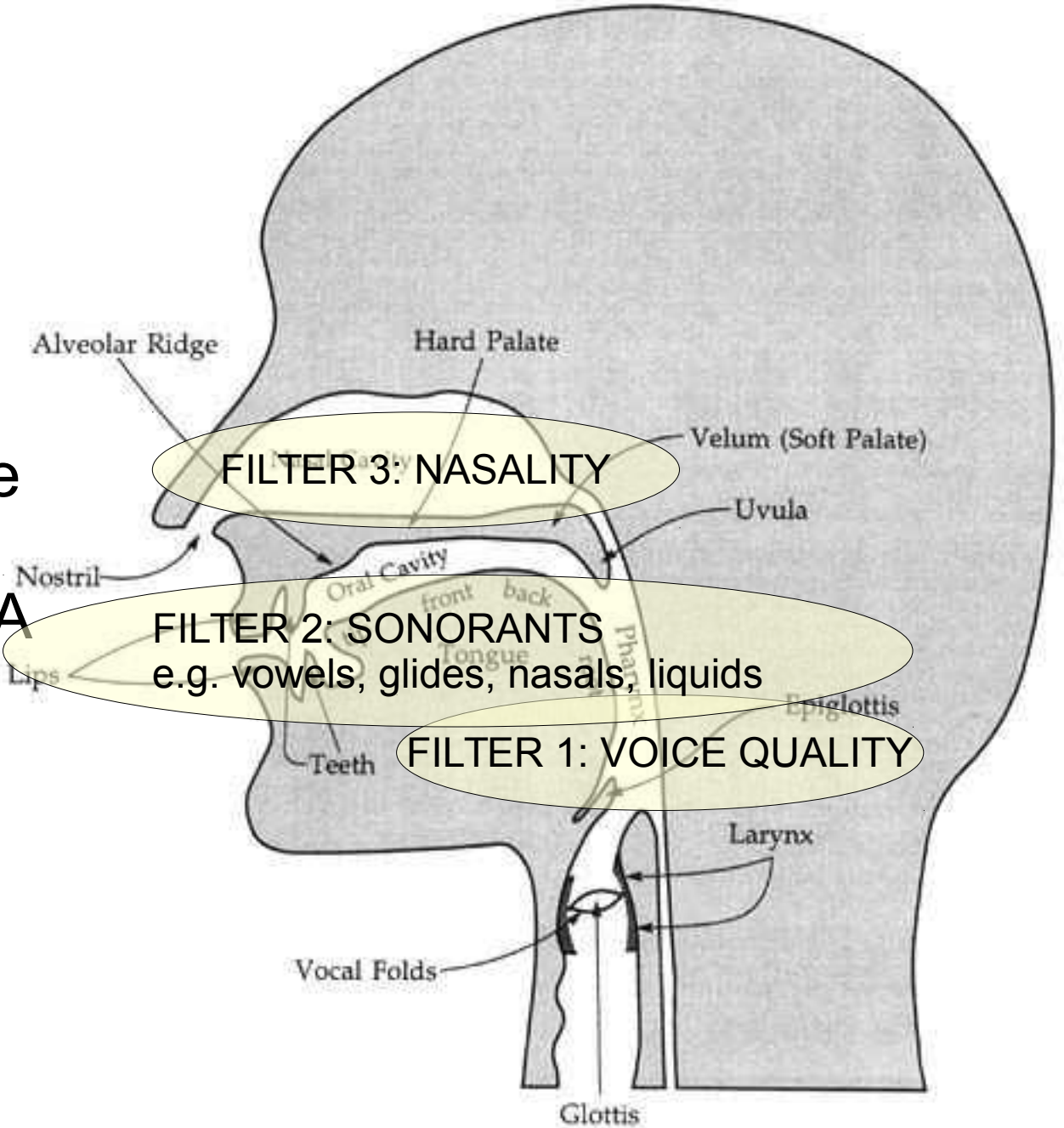
Articulatory domain: voice, stop, nasal switches

- 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

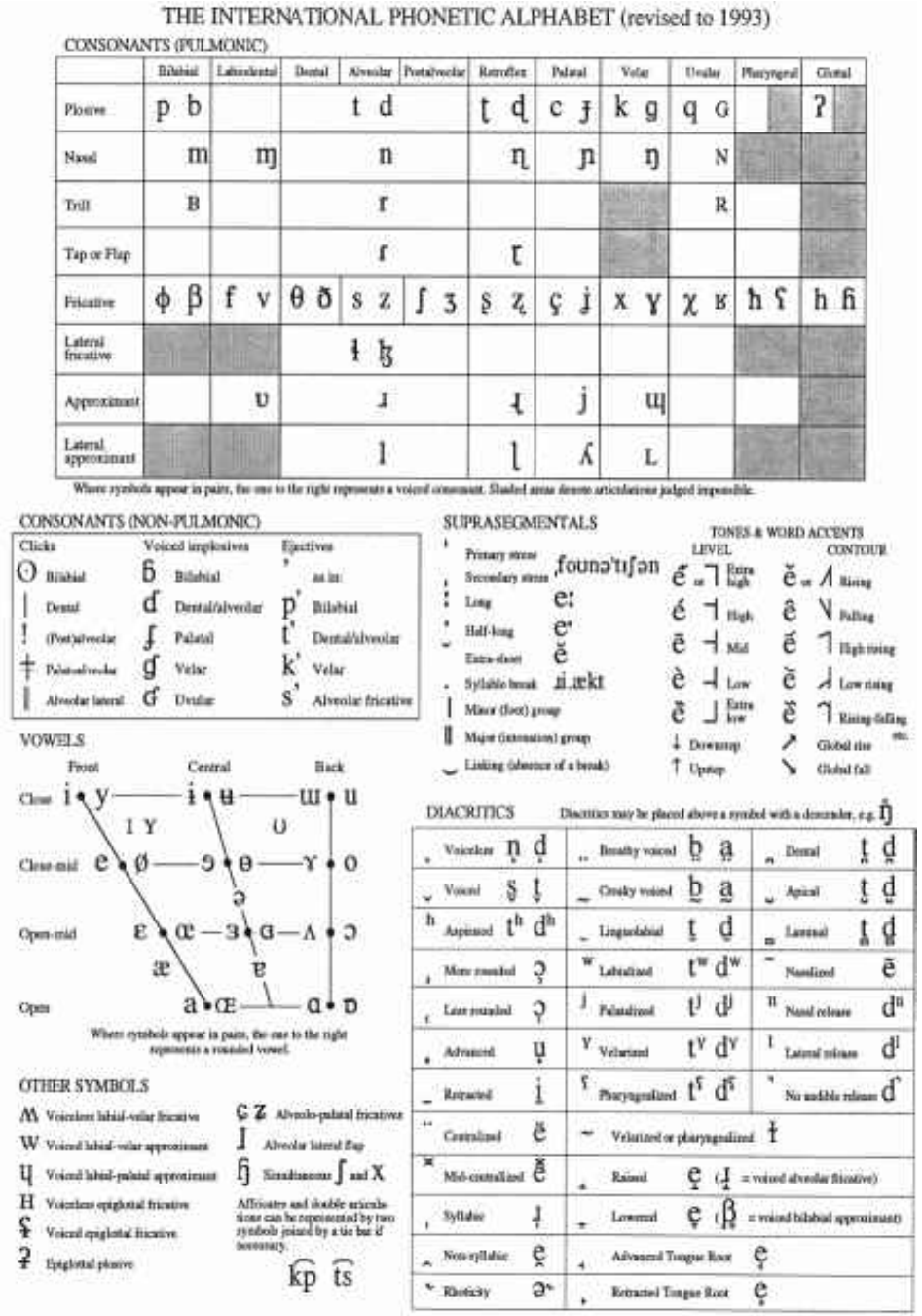


Articulatory domain: voice, stop, nasal switches

- 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)



- 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

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.

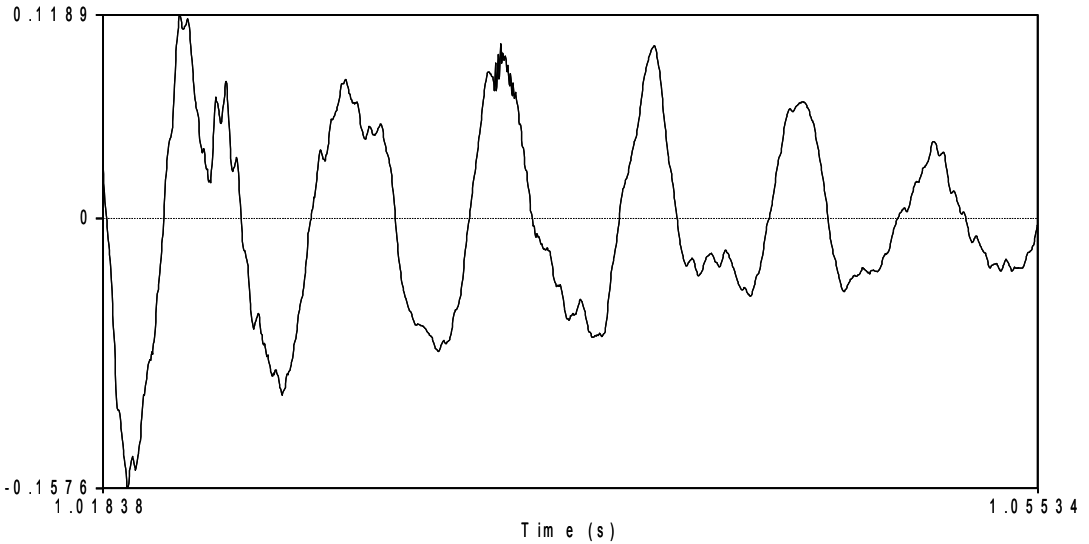
Acoustic Phonetics
(Speech Transmission)

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

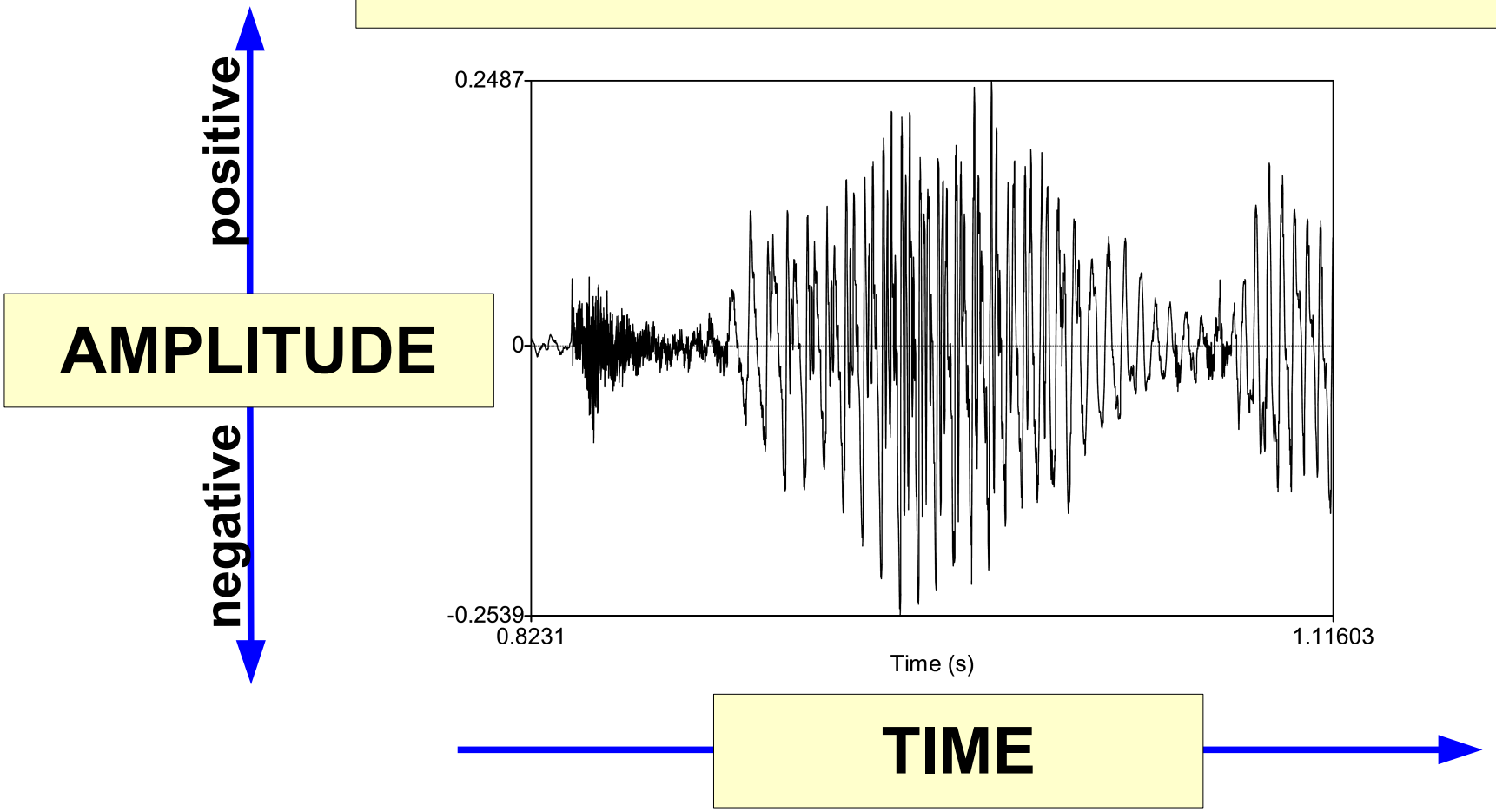
- Speech is transmitted through air (and other substances) as a regular wave of pressure changes:



- 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

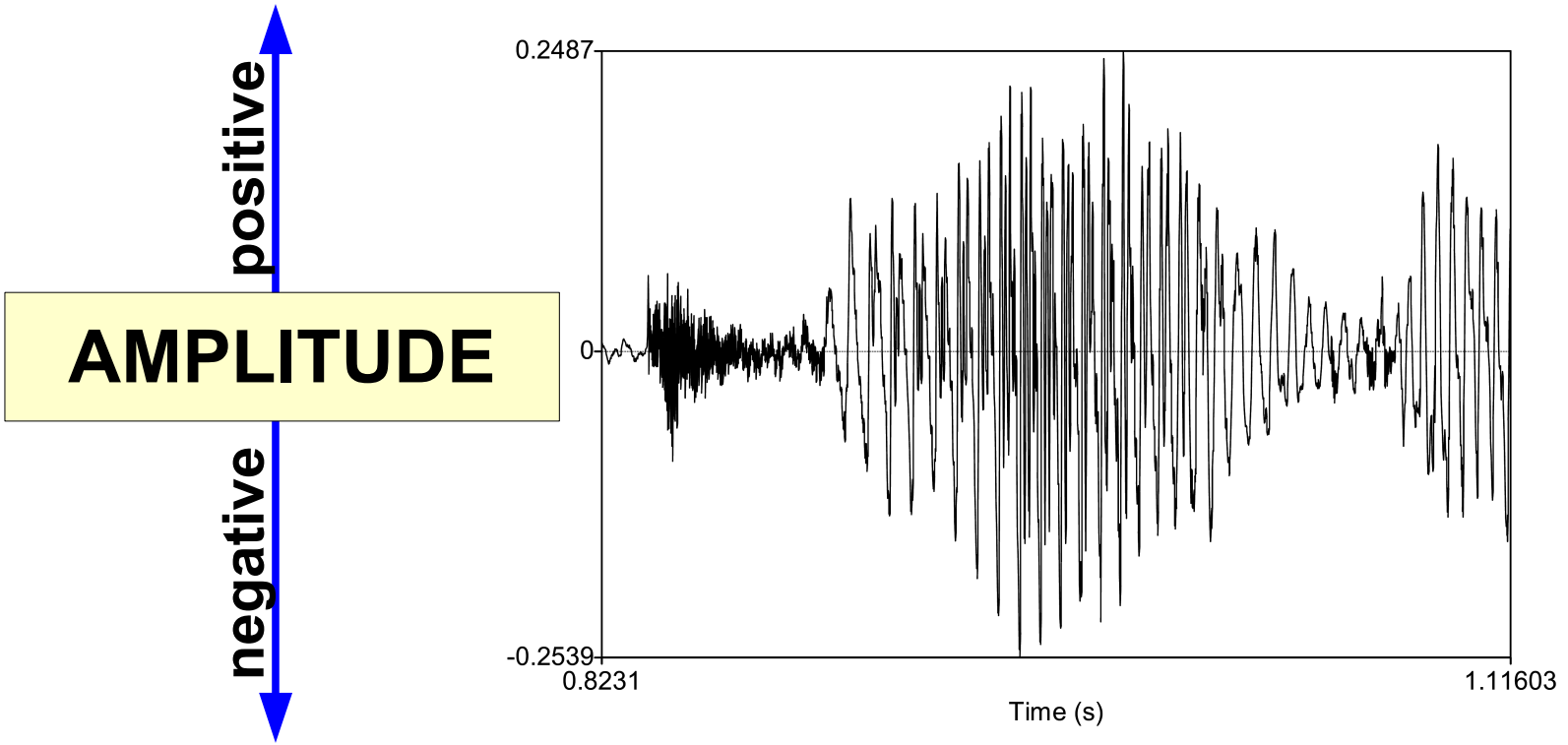
Basic Speech Signal Parameters

$y = f(t)$



The Time Domain: the Speech Wave-Form

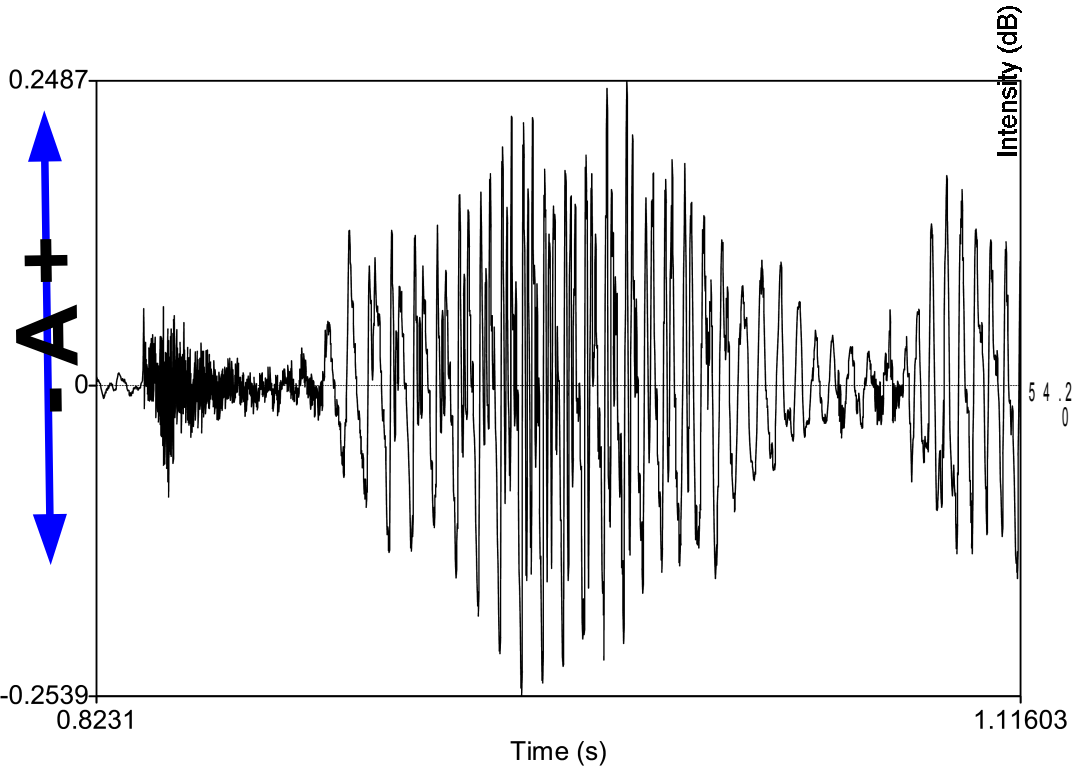
- The *positive* or *negative amplitude* **A** of the speech signal at any given point in time is the *distance* of the wave from zero at this point in time.

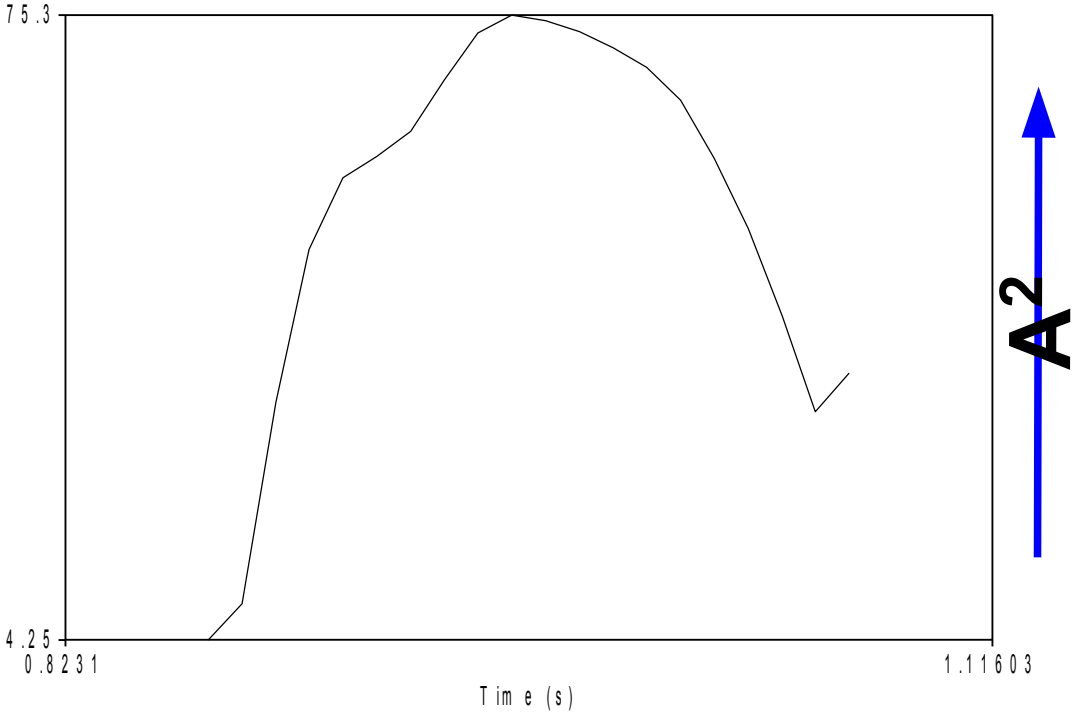


Derived parameter *INTENSITY*

- 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)

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

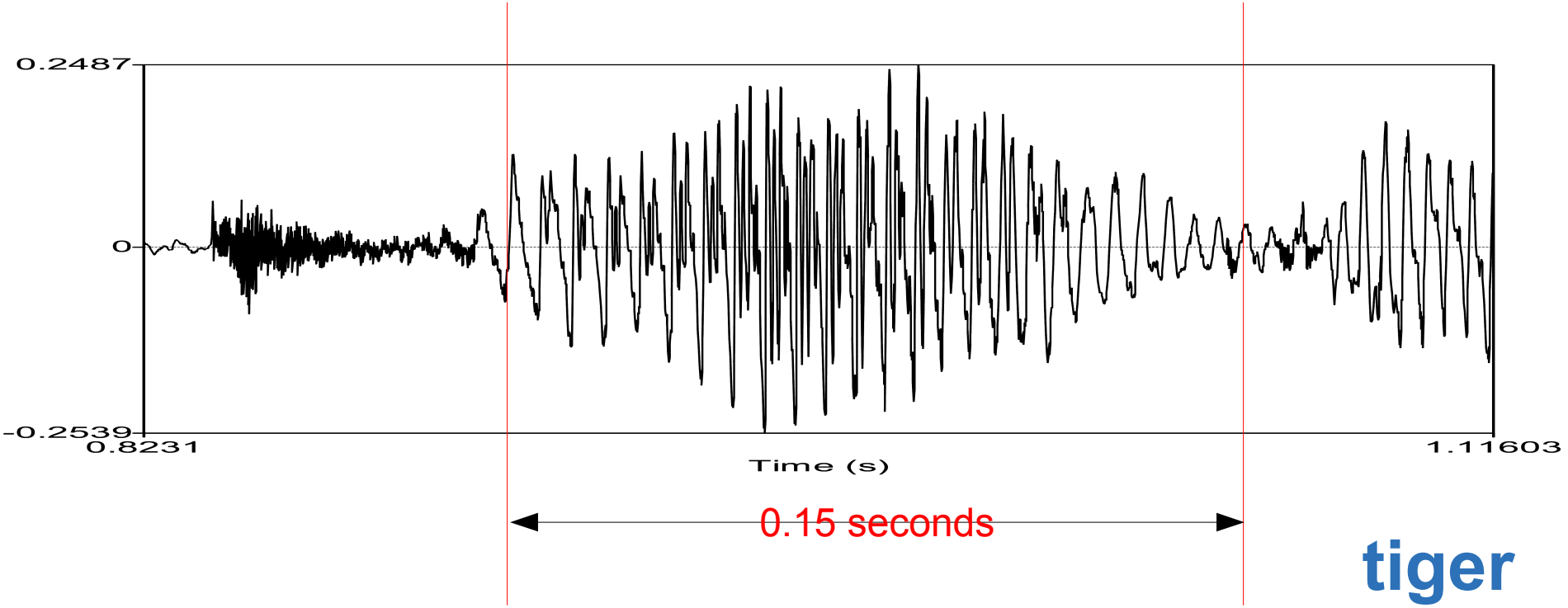
- Energy is therefore intensity averaged over time
  - In fact, intensity measurements are, in practise, energy measurements over very short periods of time
- Compare other measurement units per time unit:
  - miles per hour
  - kilowatts per hour

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

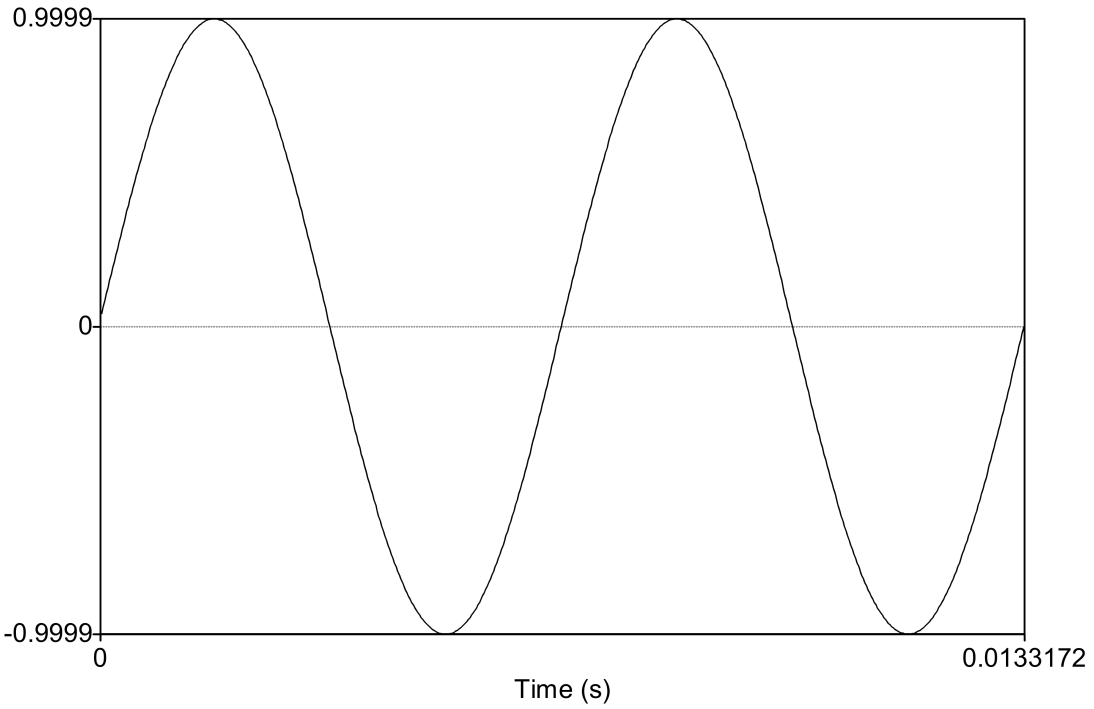
- The *frequency* of a speech signal is the number of waves (periods) per second in the waveform



- 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).



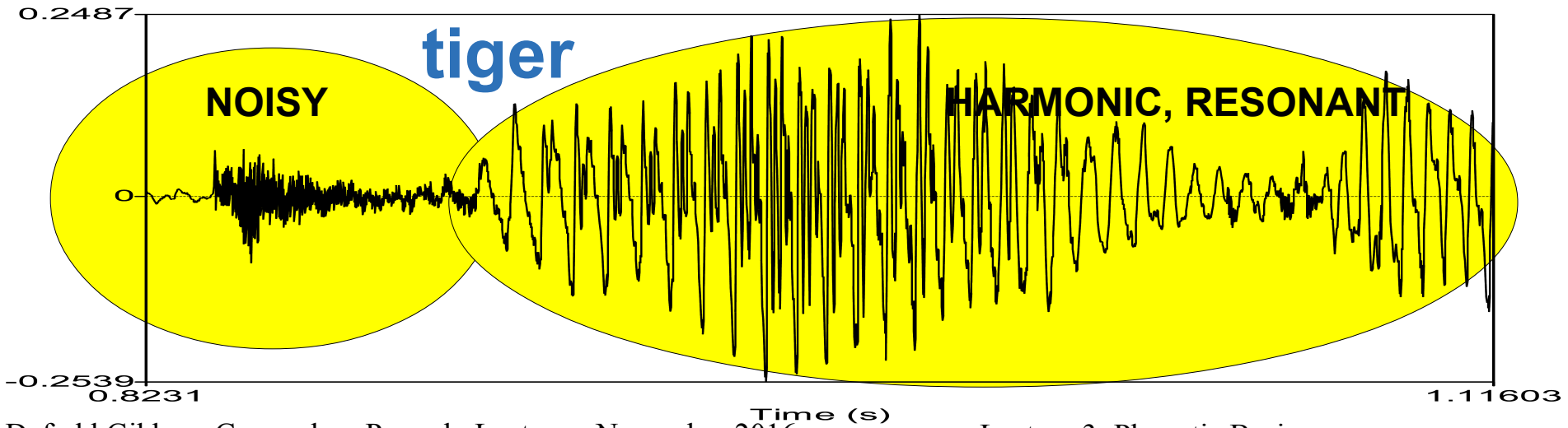
- The speech signal is not a simple sine wave, however, but a complex signal.

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

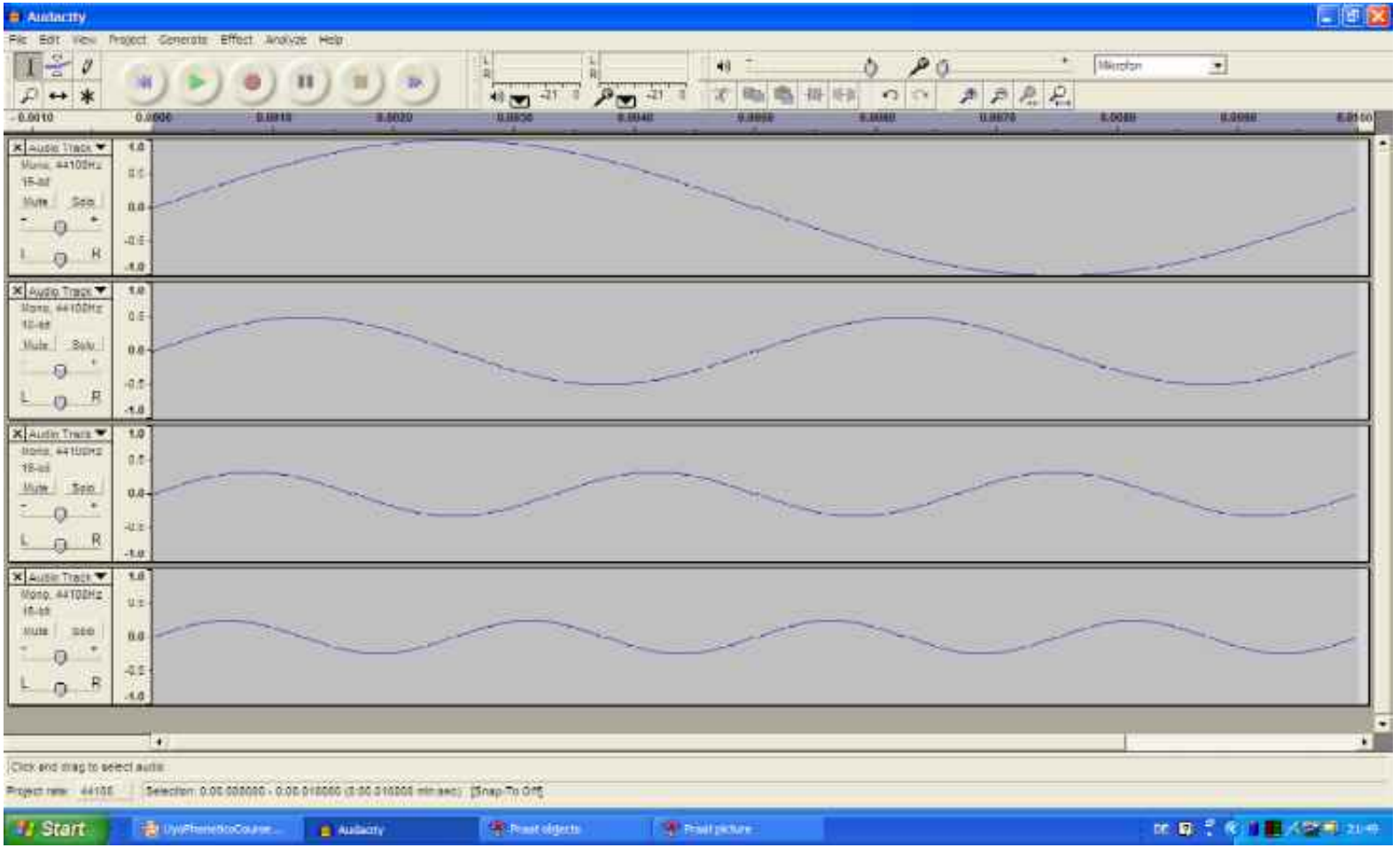
Complex Sources: noisy & harmonic signals

- 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



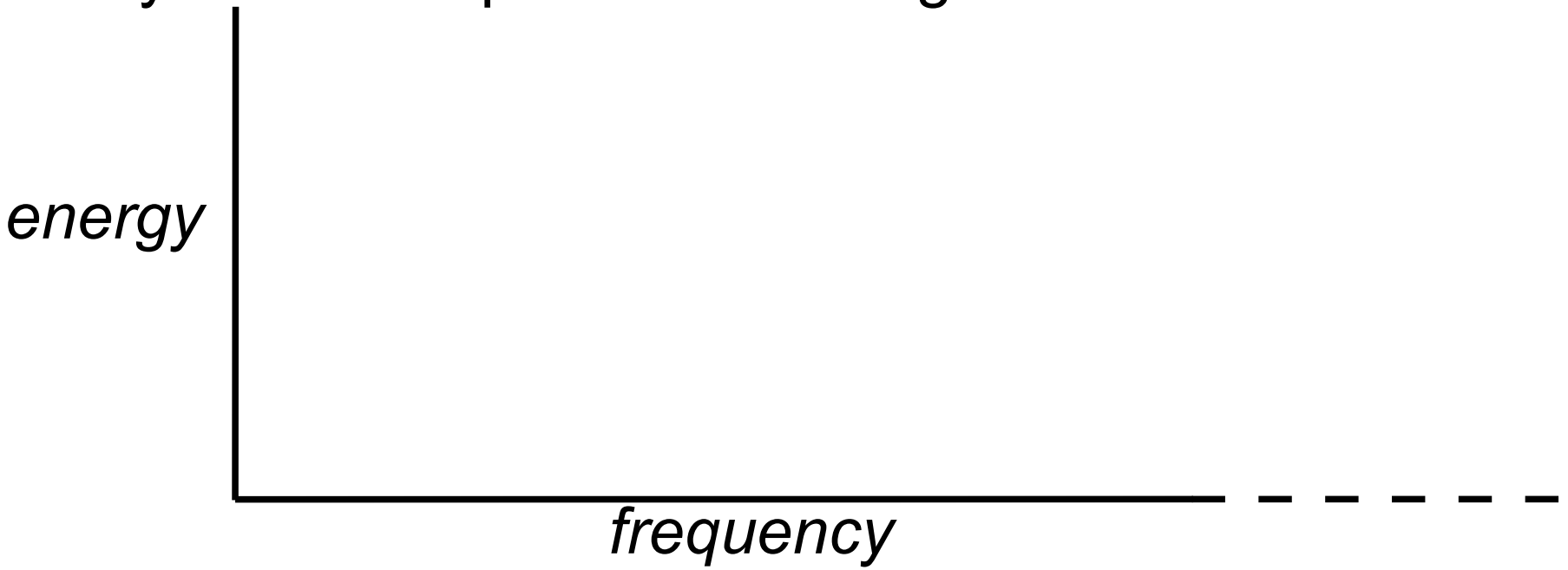
Sources with Integer Multiples of Sine Waves

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



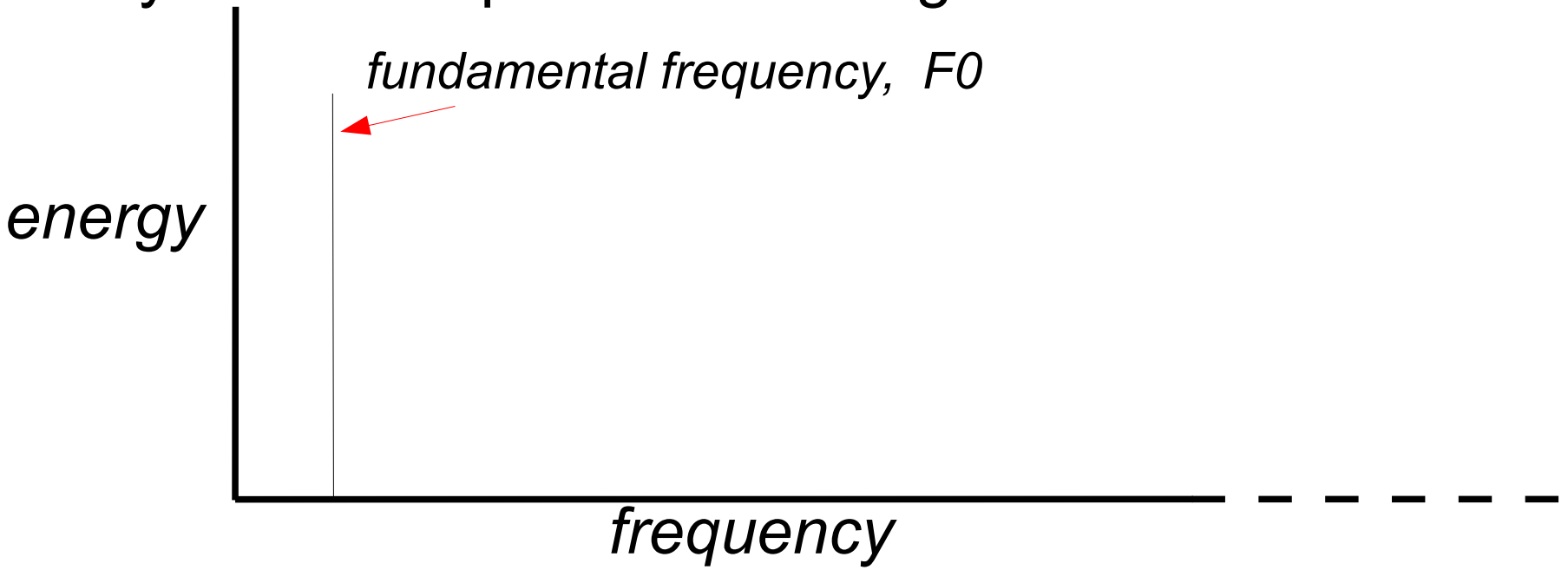
Harmonics / overtones in complex signals

- 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



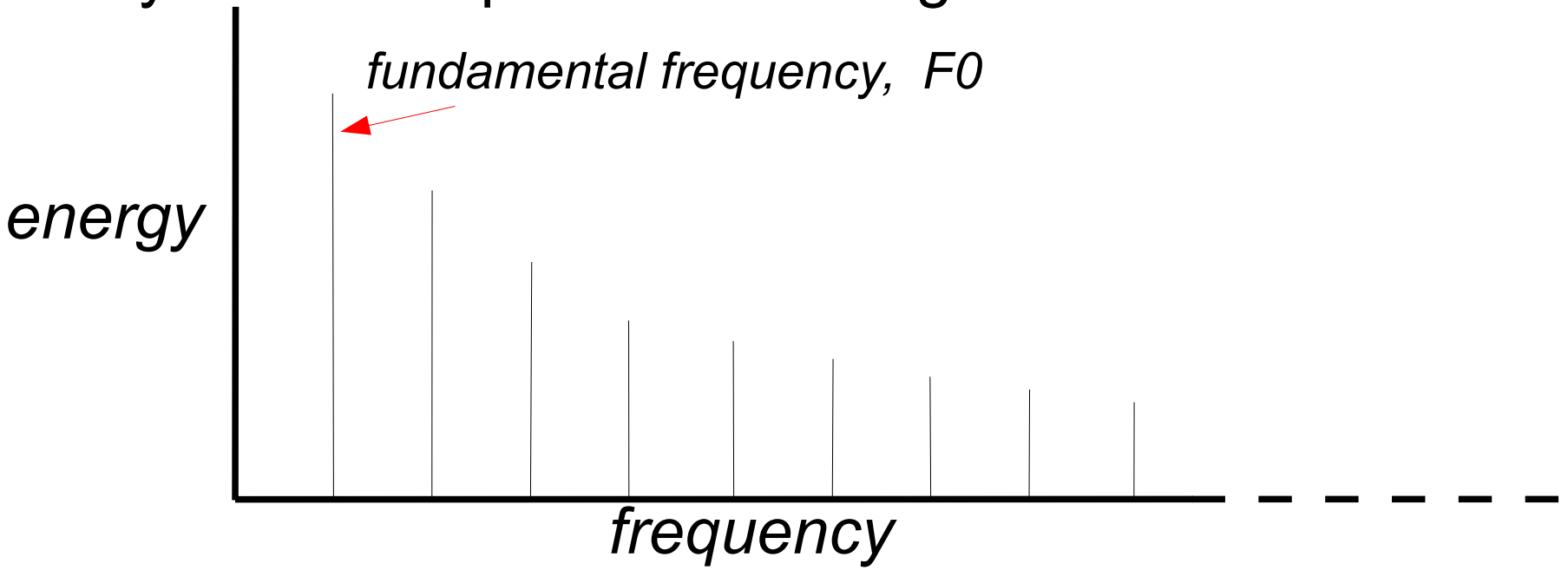
The Spectrum of Complex Signals

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



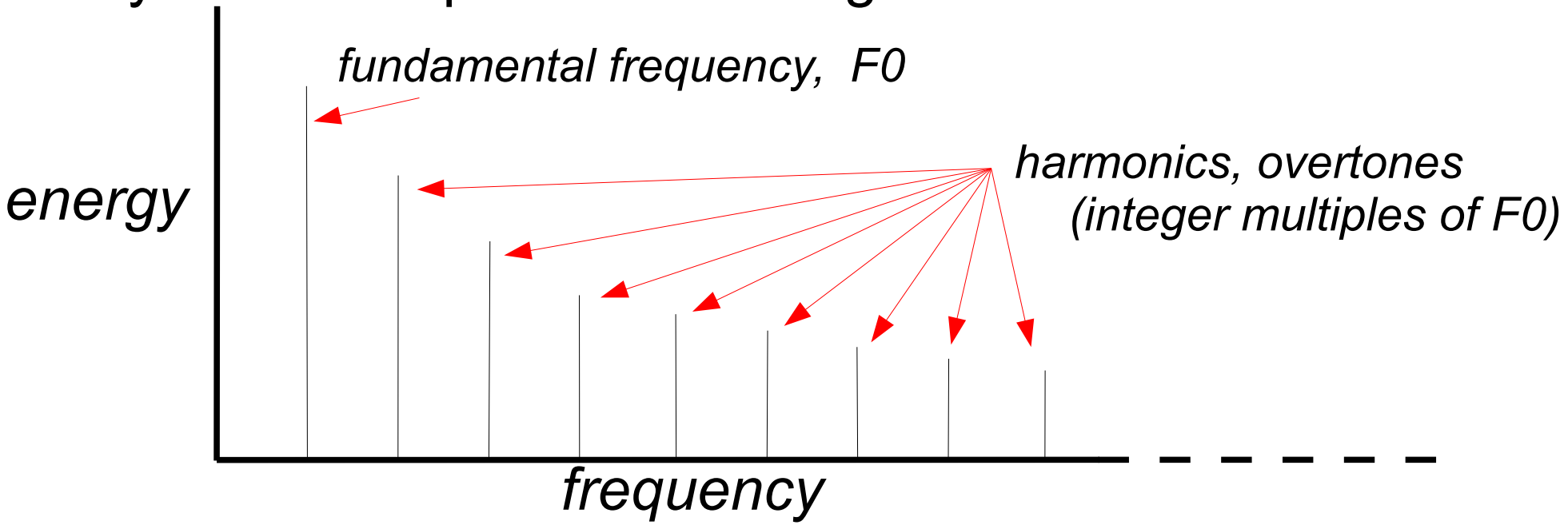
The Spectrum of Complex Signals

- 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



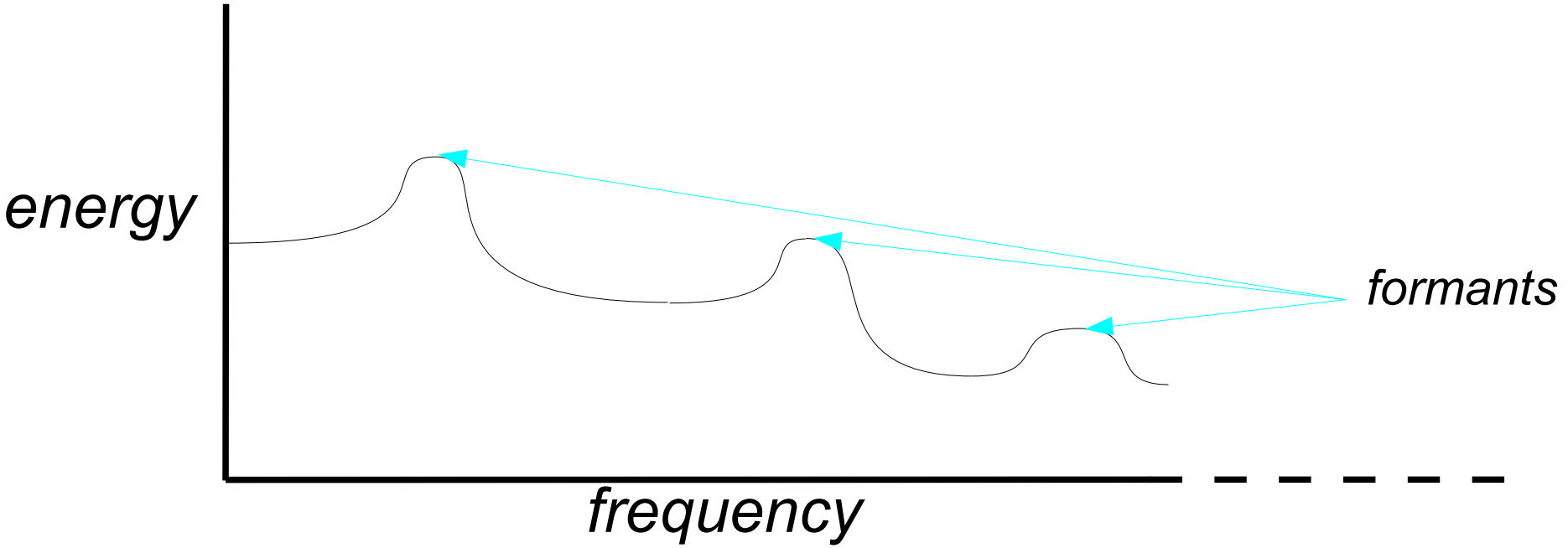
The Spectrum of Complex Signals

- 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



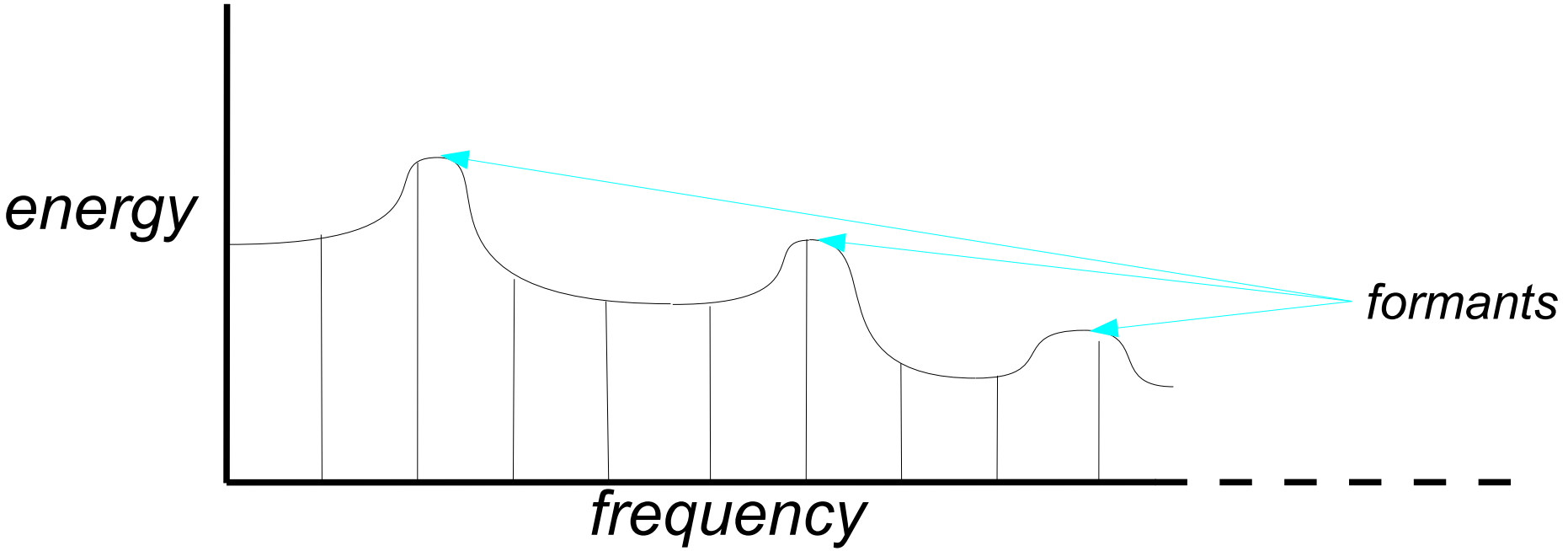
The Spectrum of Complex Signals

- 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*)



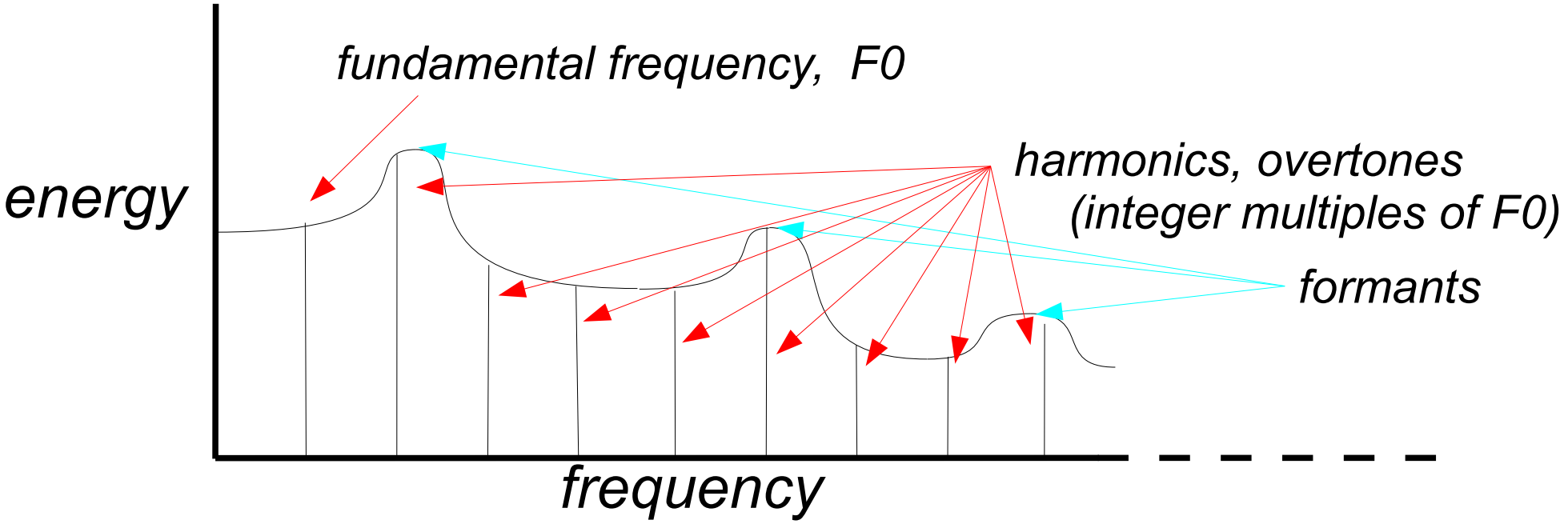
The Spectrum of Complex Signals

- 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



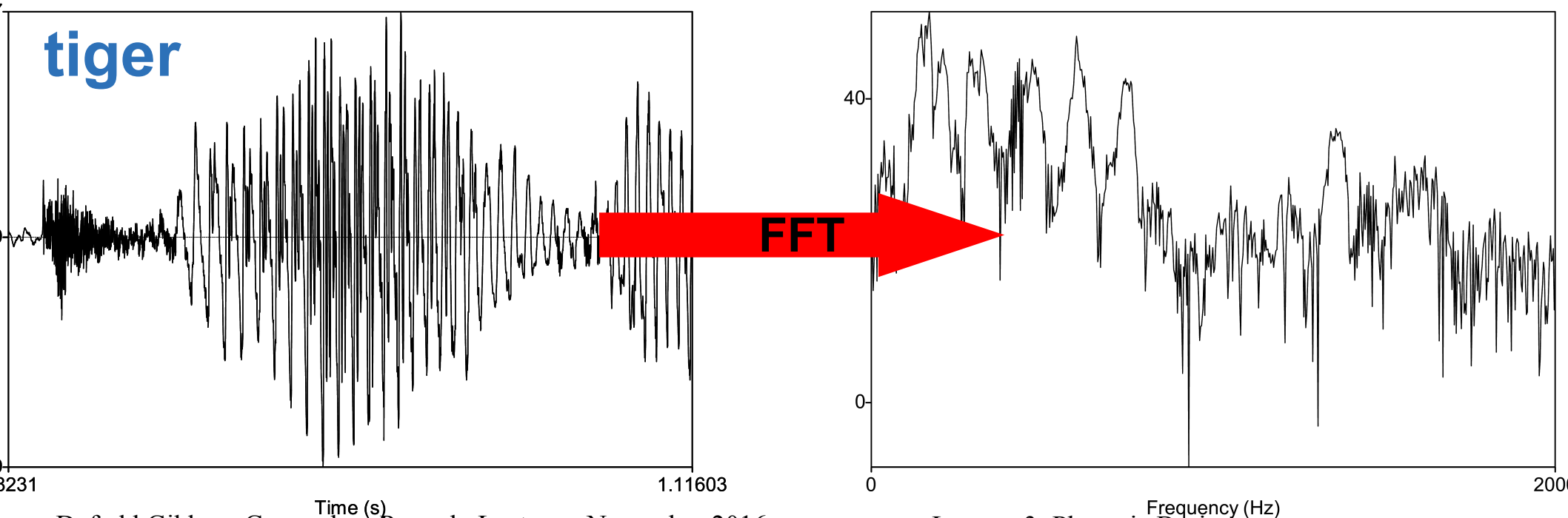
The Spectrum of Complex Signals

- 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:

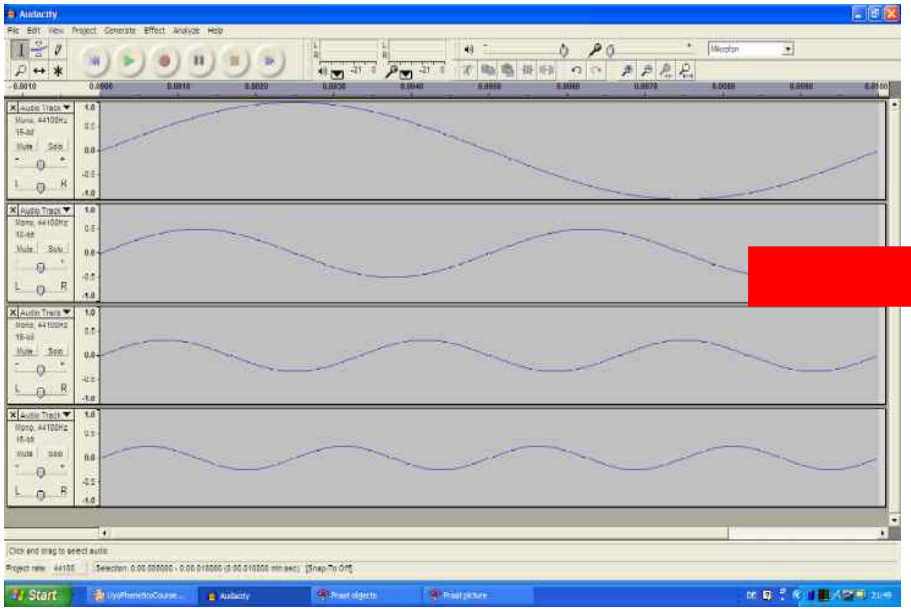


The Speech Sound Source: sawtooth waveforms

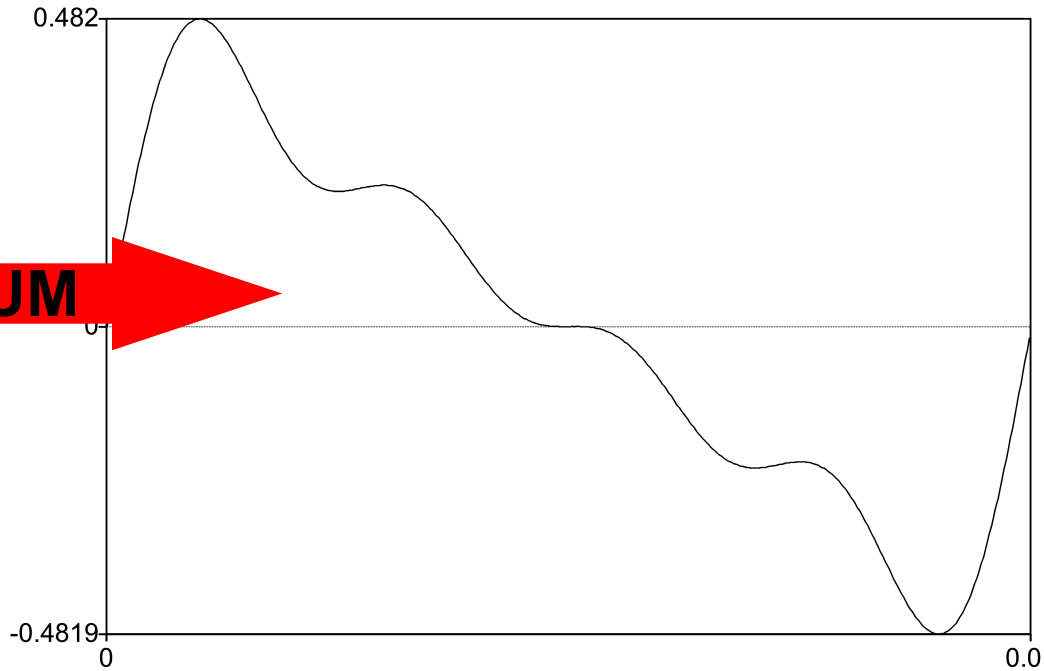
- The sum of harmonics which are integer multiples, with A inverse to F, creates a sawtooth waveform:

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

- This example illustrates the sum of four sine waves:
  100 Hz + 200 Hz + 300 Hz + 400 Hz

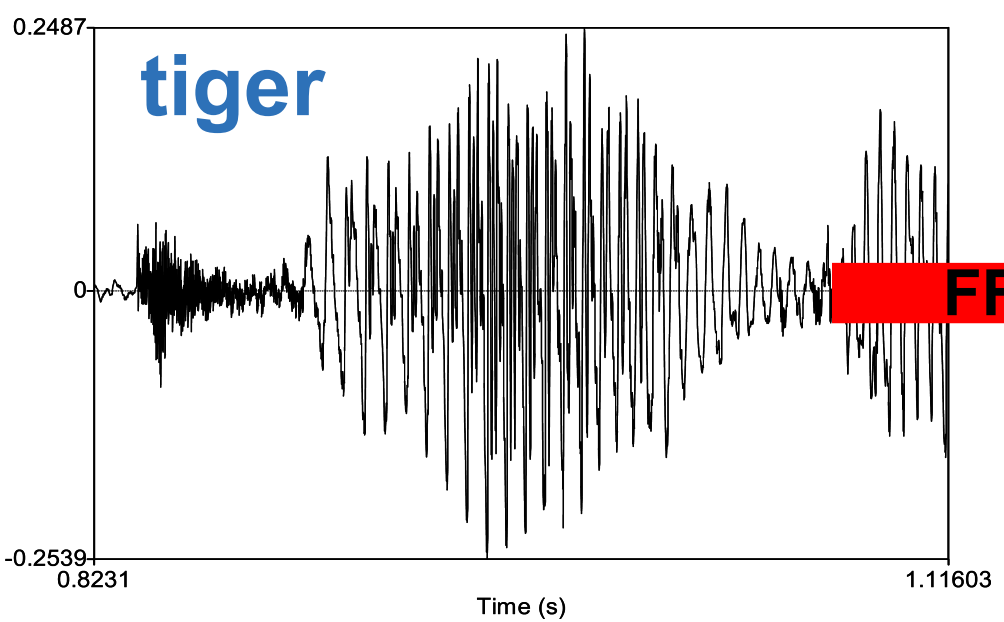


SUM

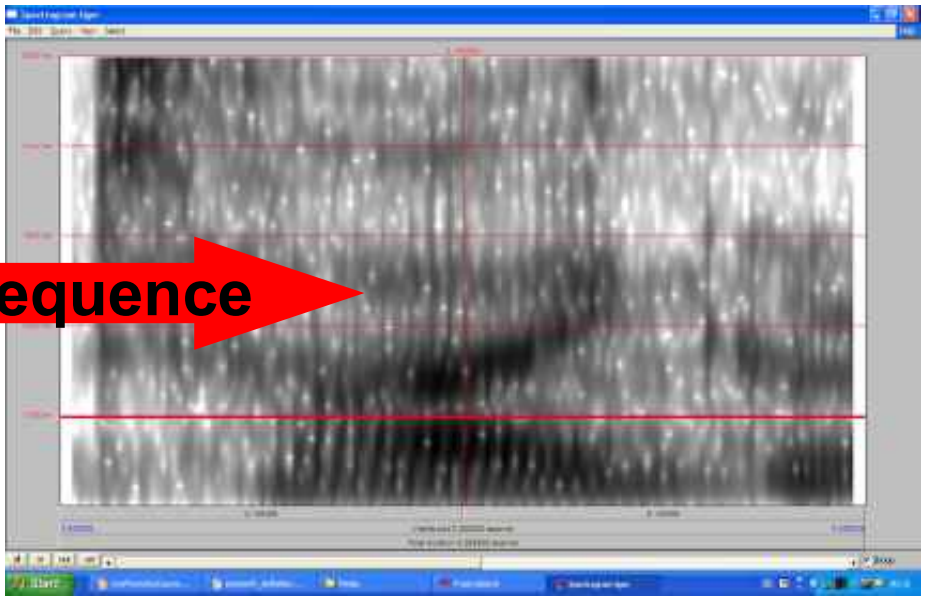


Fourier Analysis: the Spectrogram

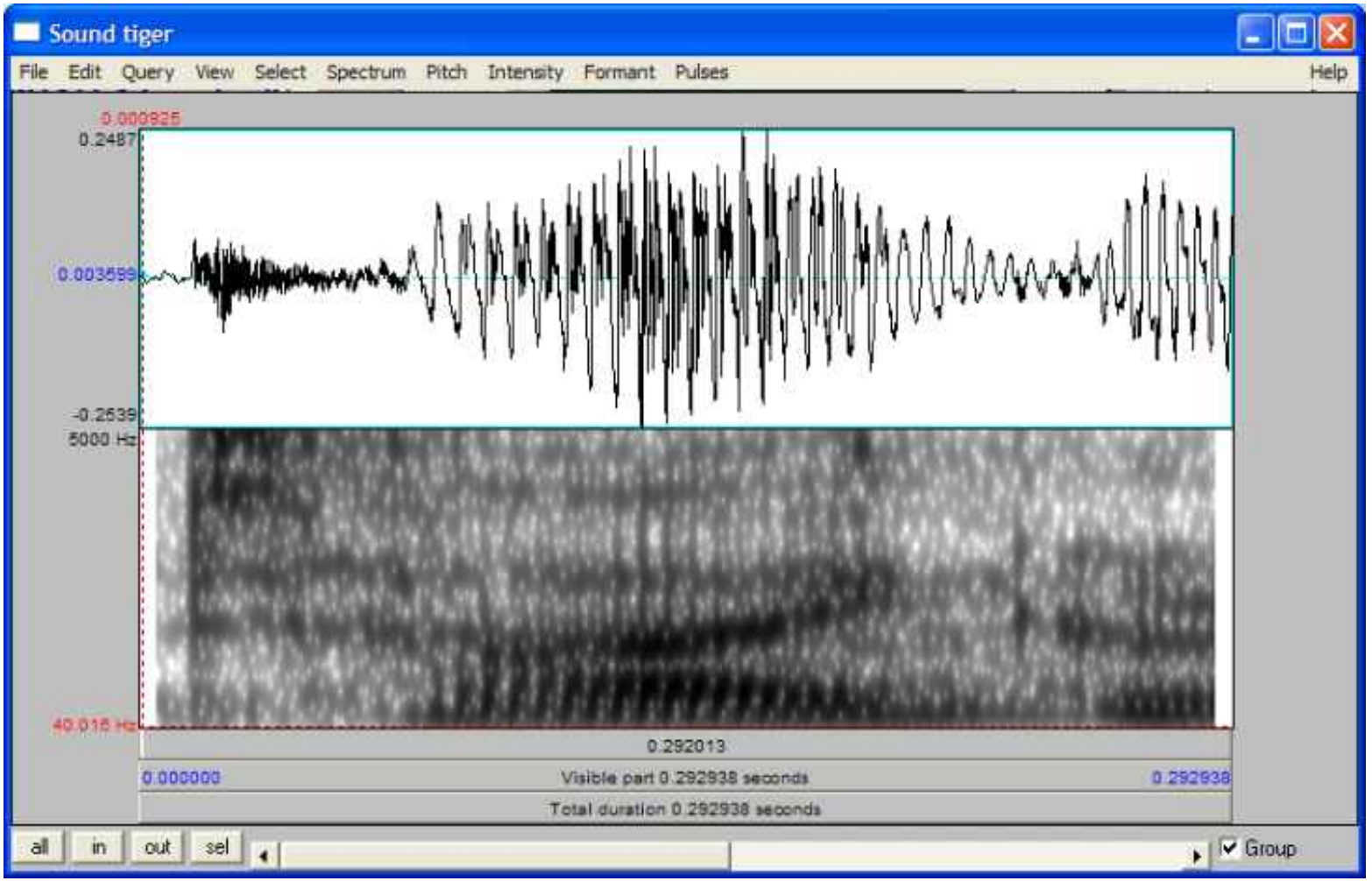
- A single spectral analysis of part of a speech signal, yielding a spectrum, requires a sequence of several periods:
  - In order to track the changing structure of a speech signal, a sequence of spectra is needed.
  - A representation of a sequence of spectra is called a spectrogram



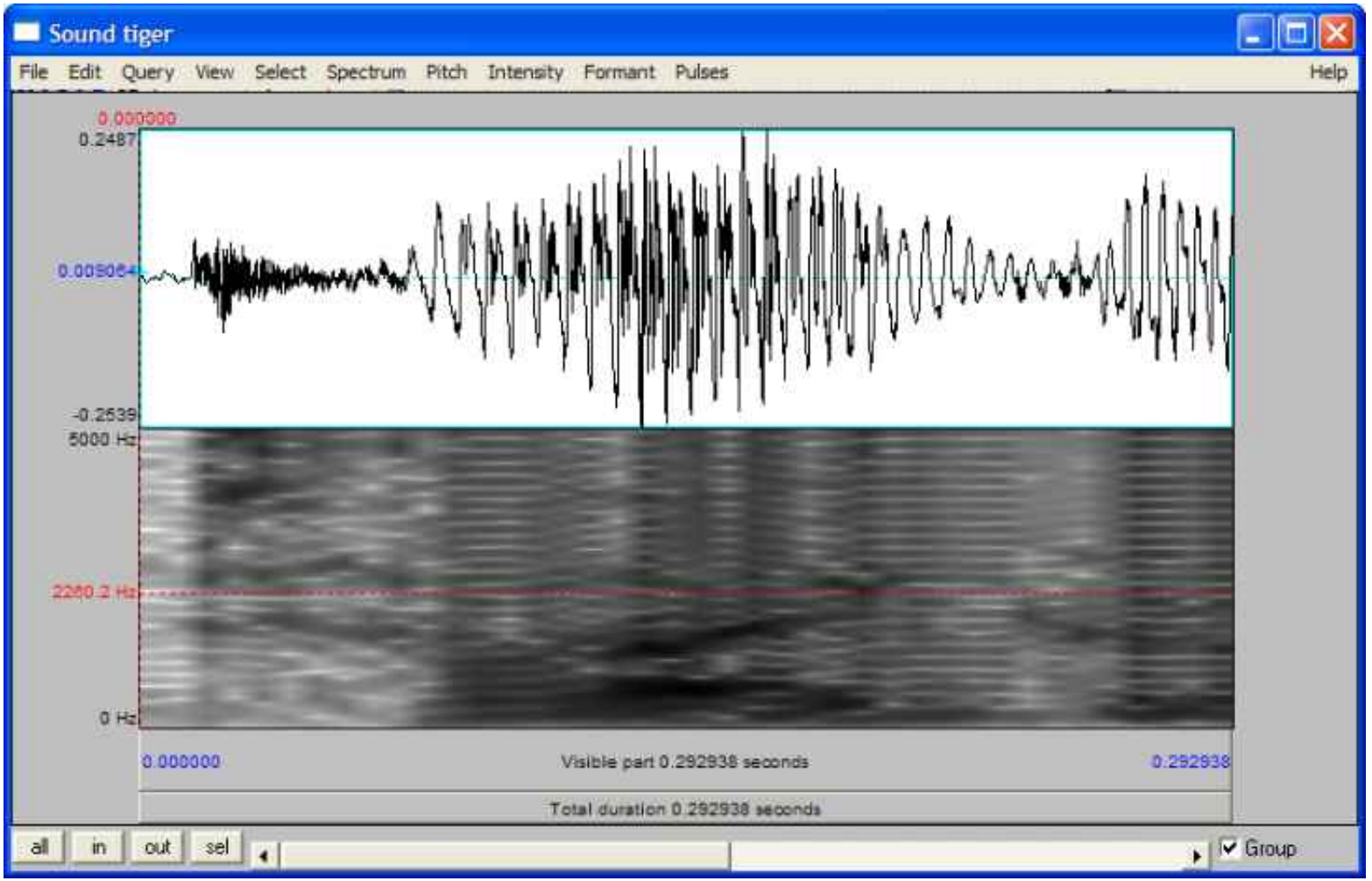
FFT sequence



Broad band spectrogram



Narrow band spectrogram

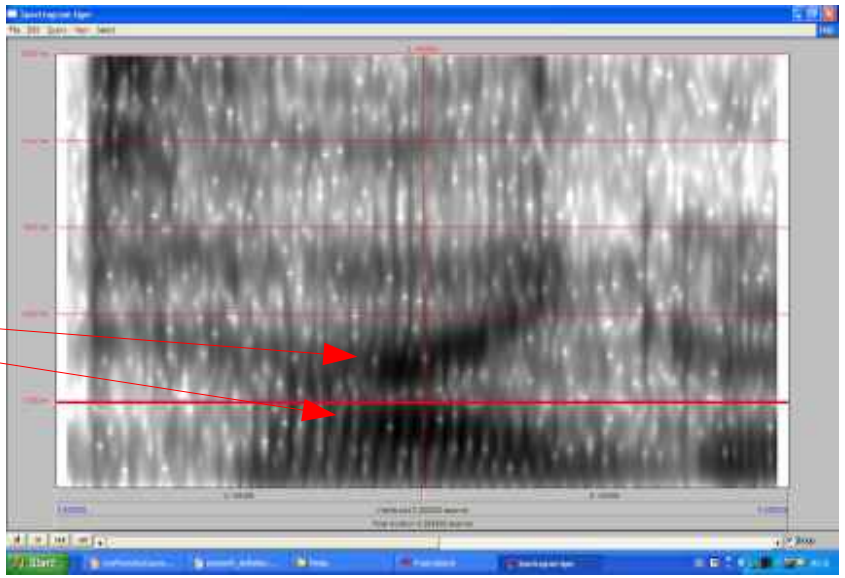


Spectrogram Filtering: Formants

- 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

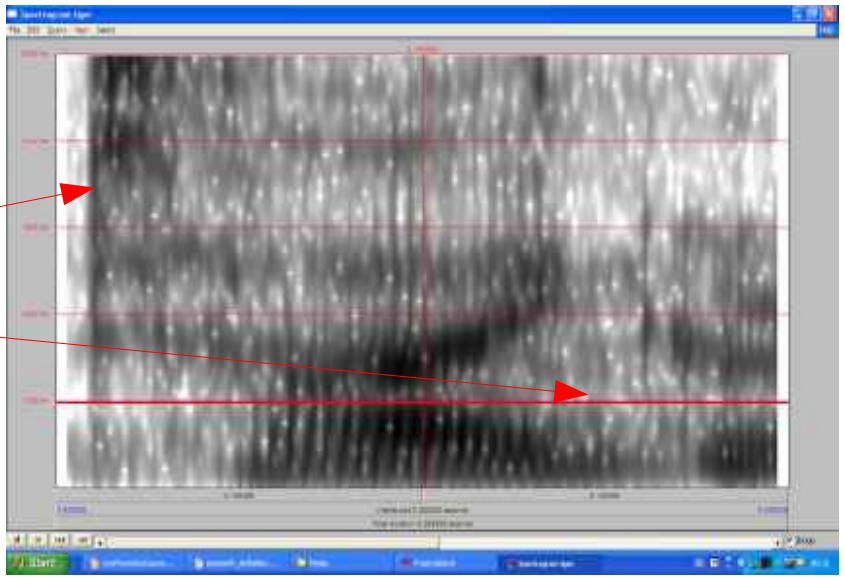


Spectrogram Filtering: Consonantal Noise

- 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

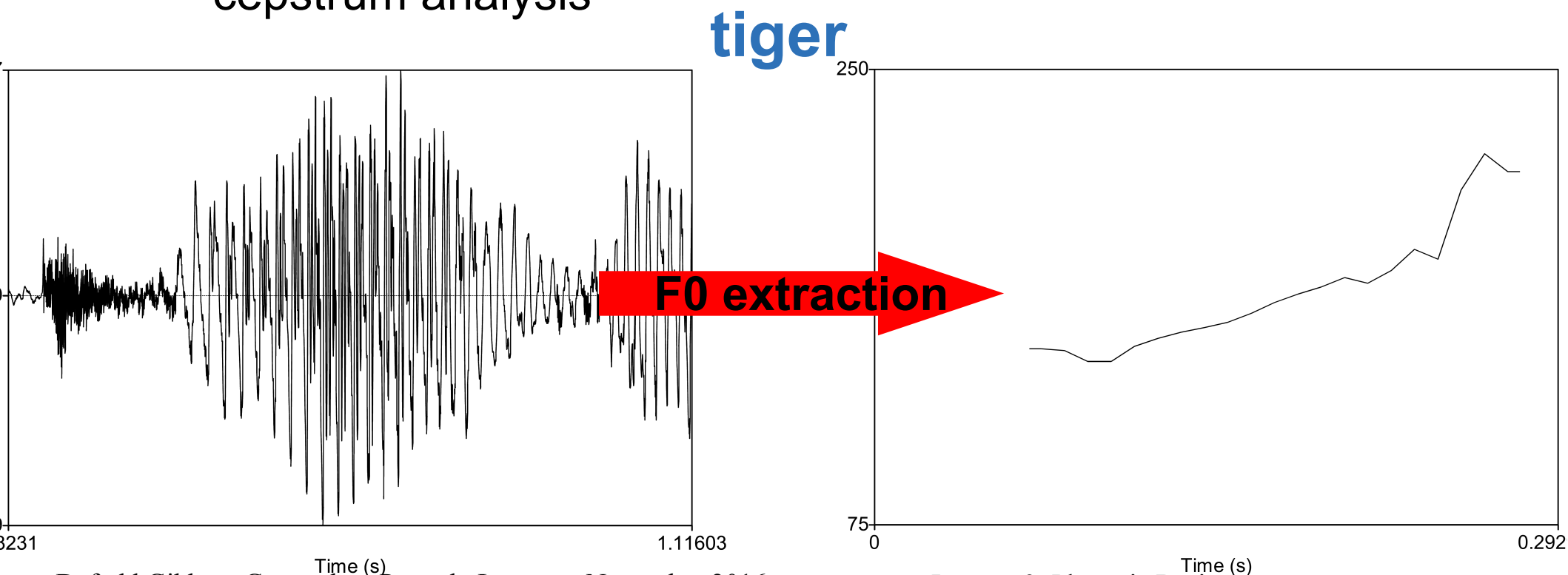
tiger

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

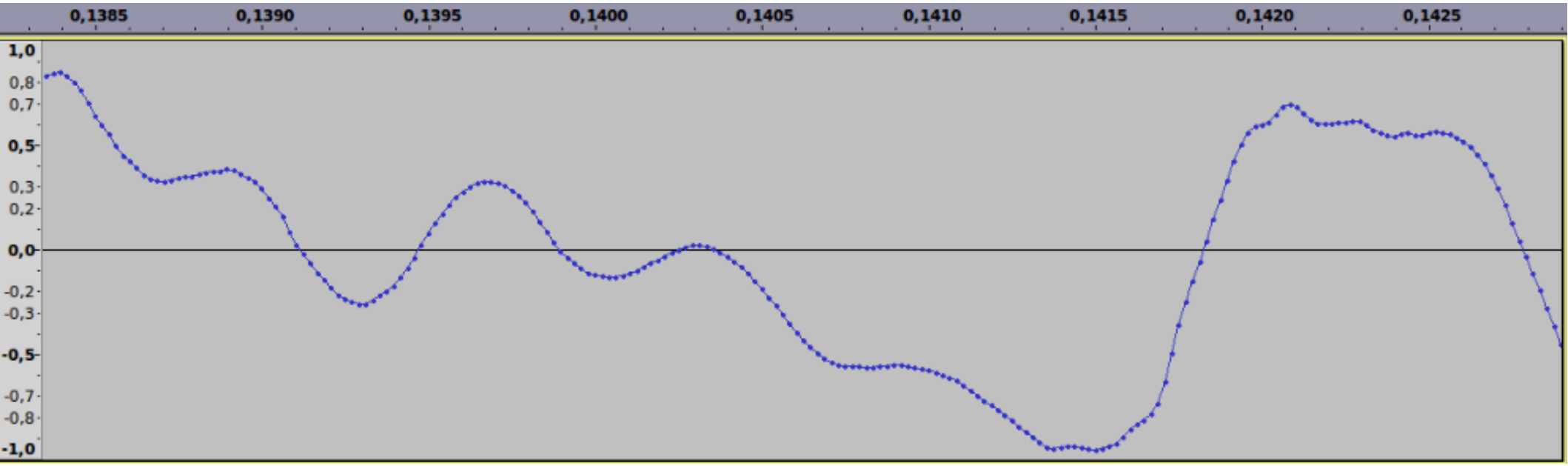


Analog-to-Digital (A/D) Conversion

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

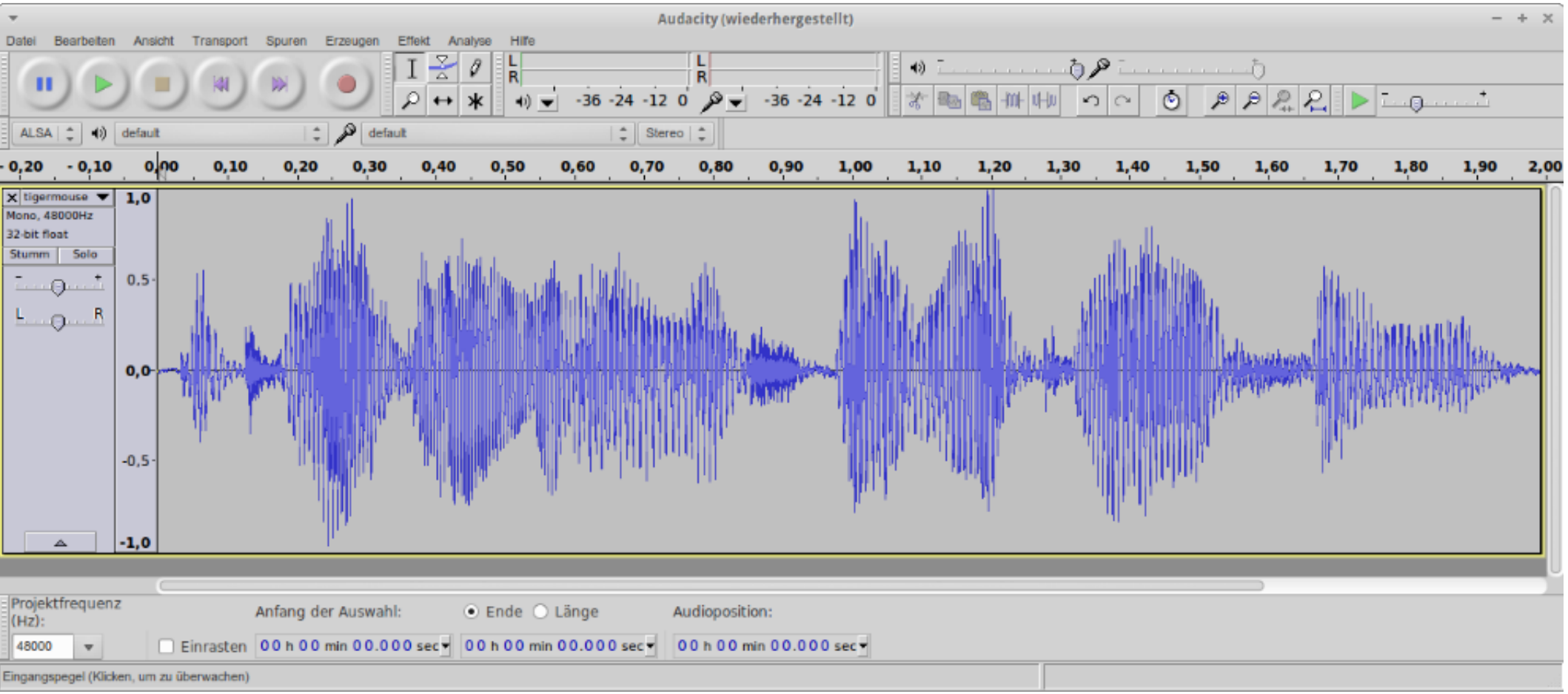


Quiz on Acoustic Phonetics

- 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

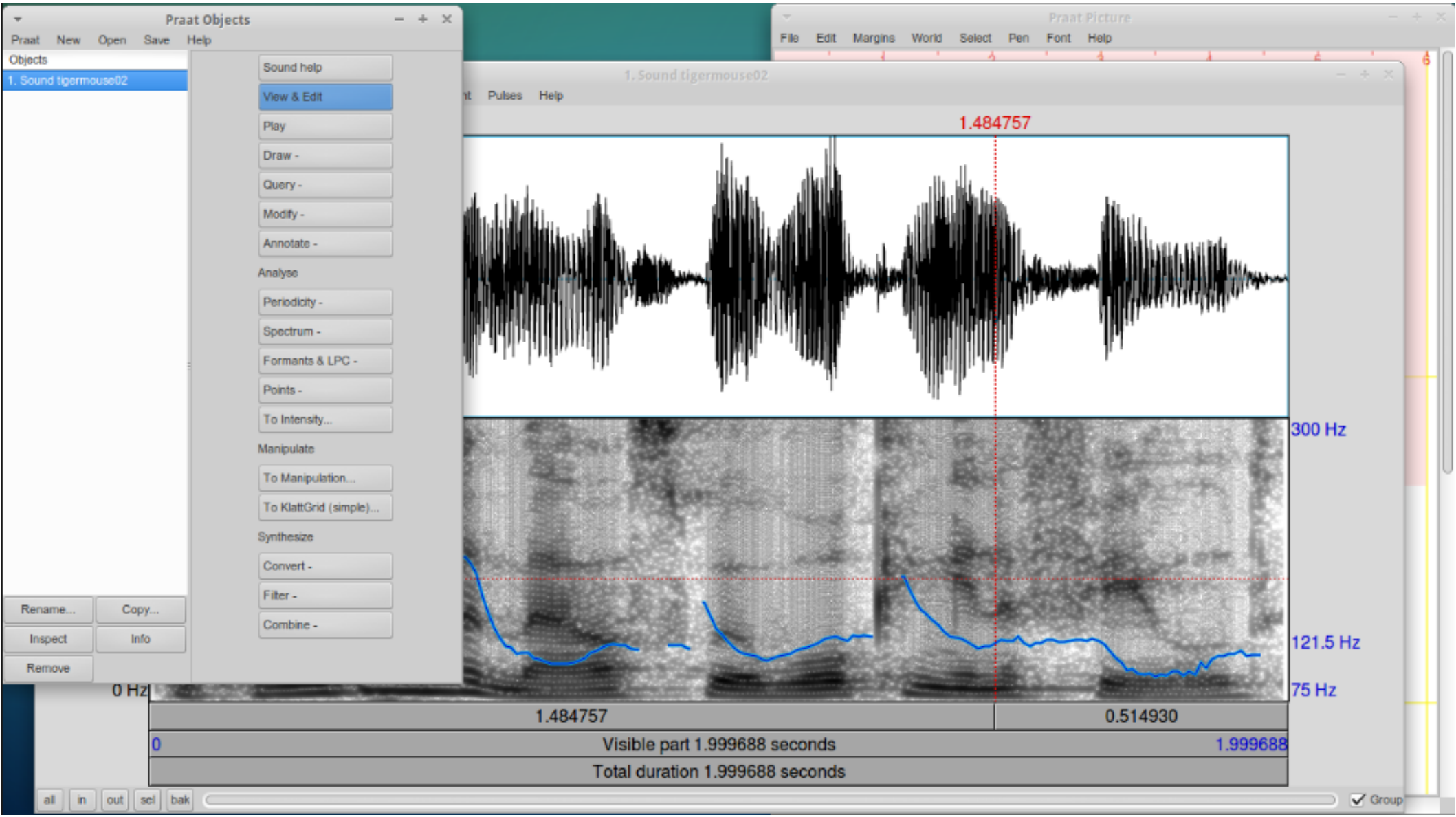
Digital Phonetics Software

Audacity: an audio signal editor for preprocessing (cutting, normalisation, filtering, ...)



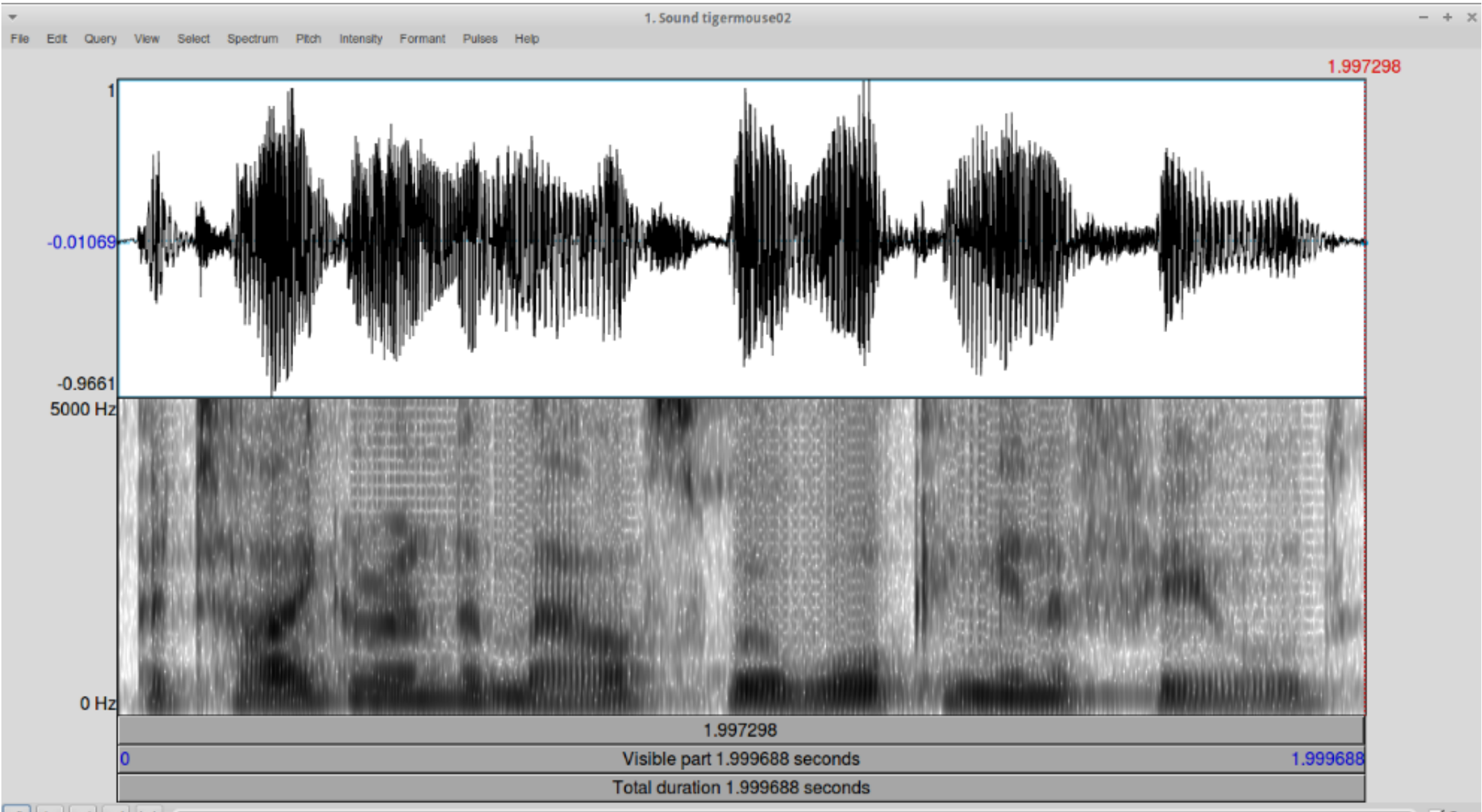
Digital Phonetics Software

Praat: a speech workbench for analysis (amplitude, spectrum, intensity, formants, ...)



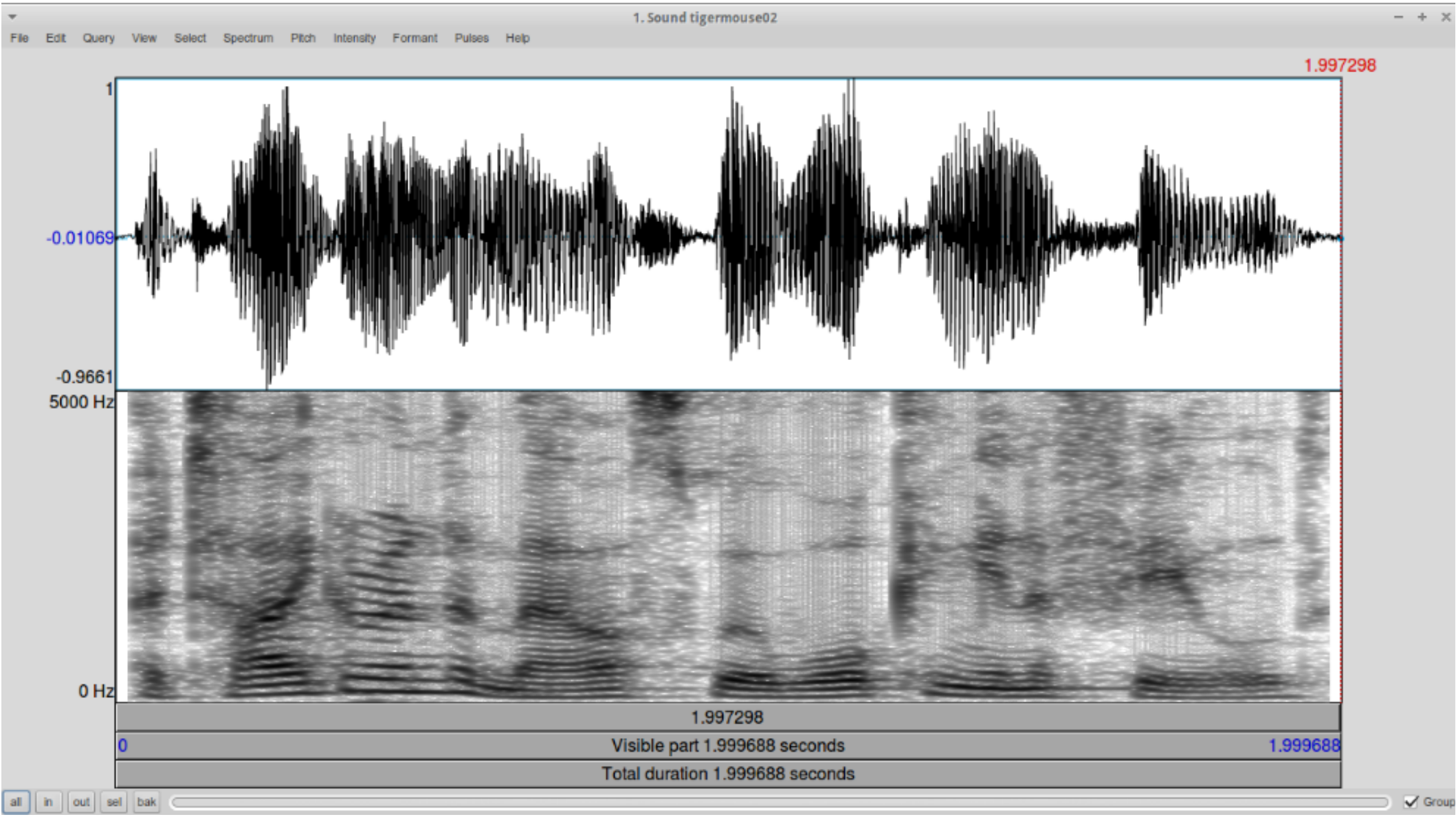
Digital Phonetics Software

Praat: broad band spectrogram



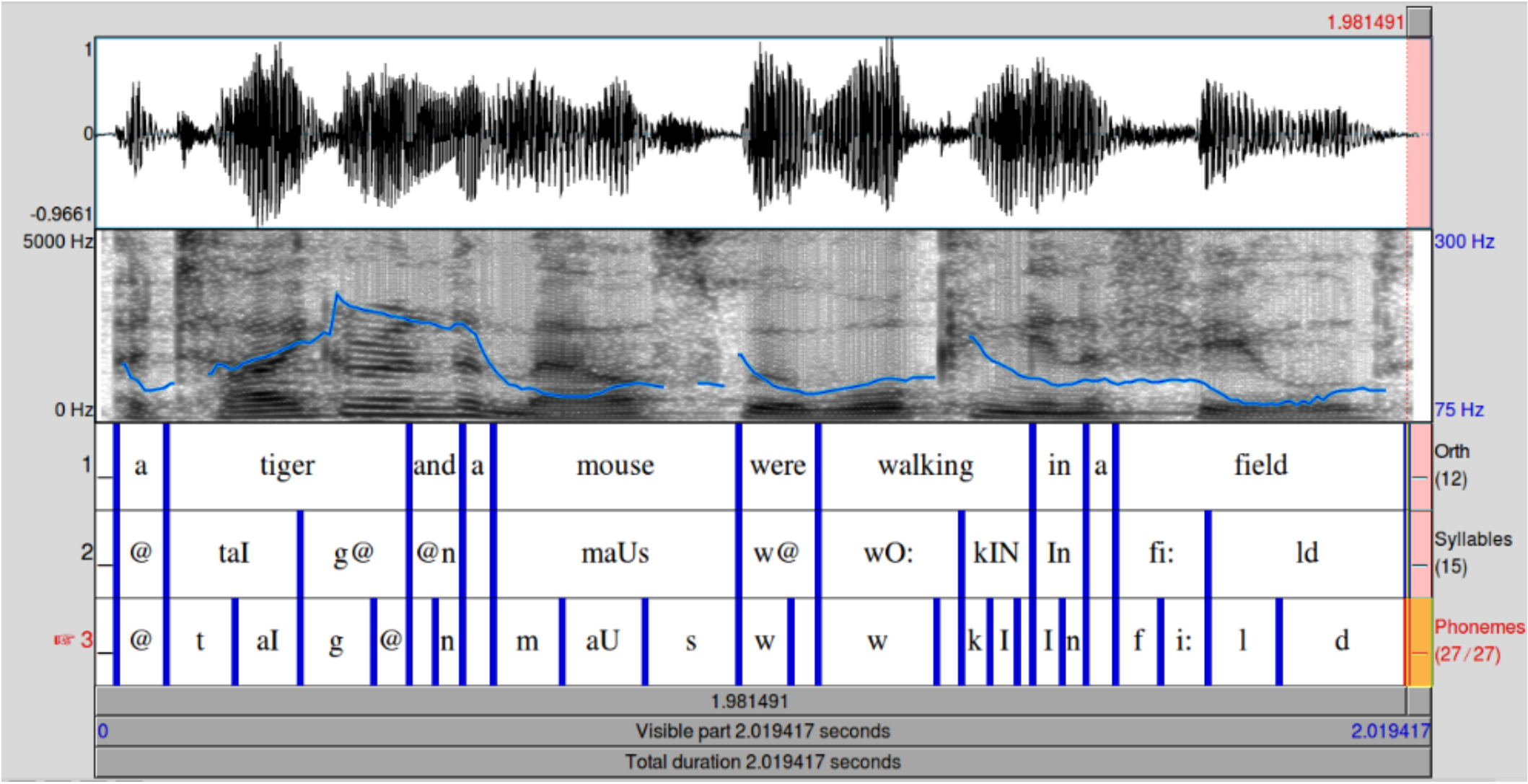
Digital Phonetics Software

Praat: narrow band spectrogram



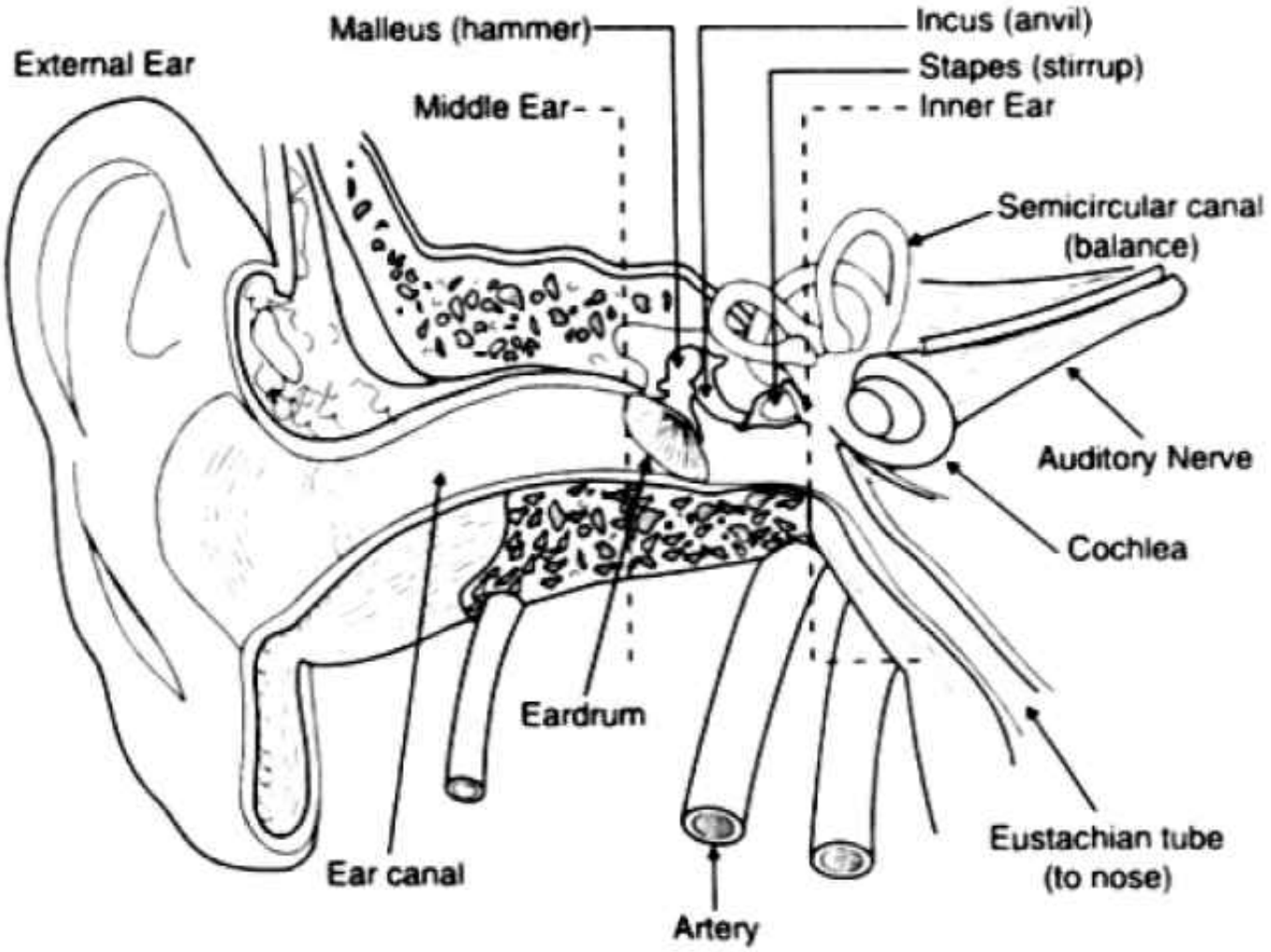
Digital Phonetics Software

Praat: spectrum and pitch annotation of the speech signal with transcriptions

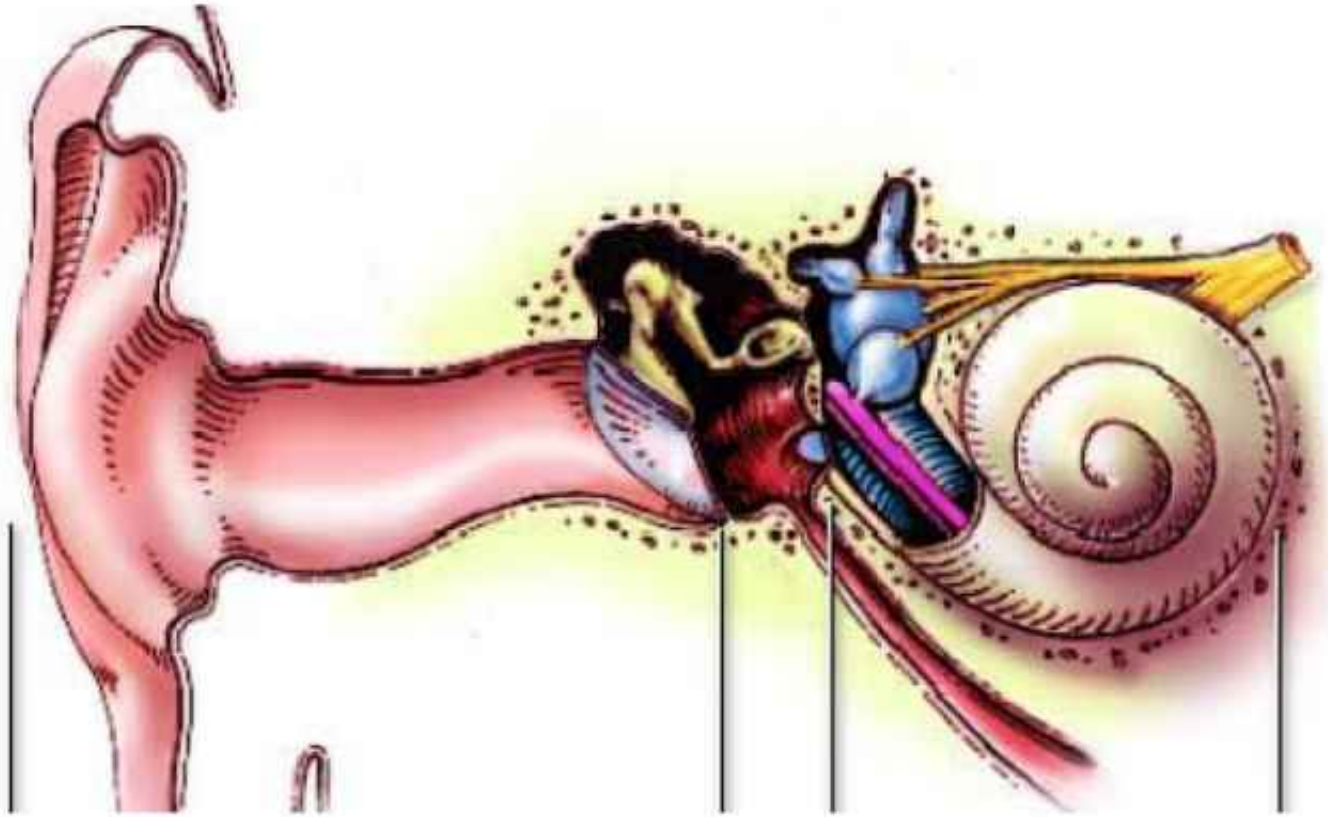


Auditory Phonetics
(Speech Perception)

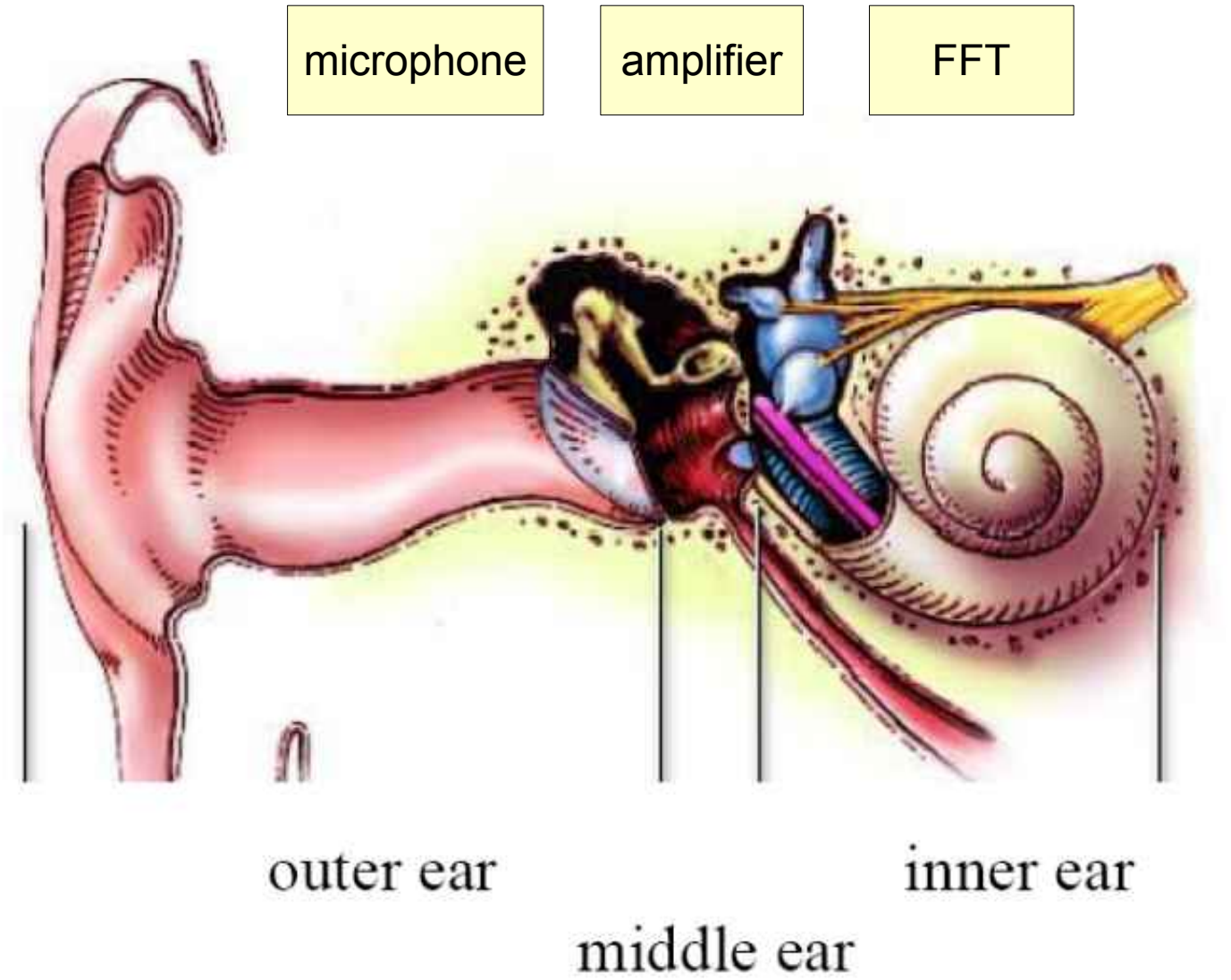
The Auditory Domain: Anatomy of the Ear



The Auditory Domain: Anatomy of the Ear



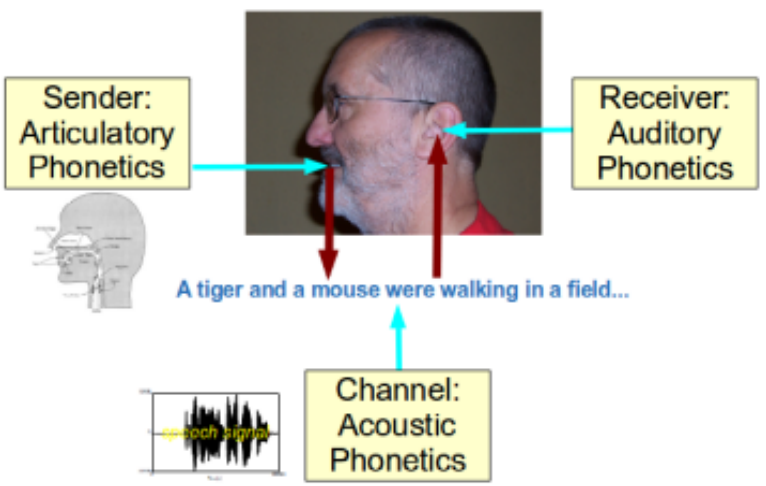
The Auditory Domain: Anatomy of the Ear

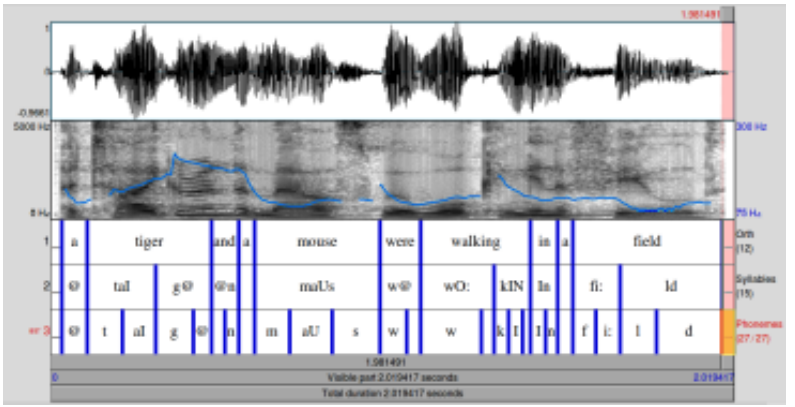


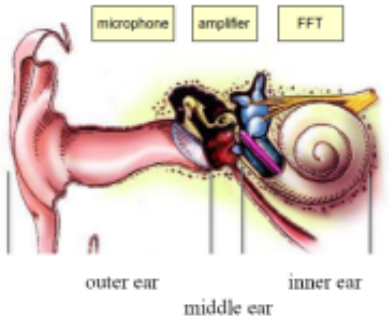
Quiz on Auditory Phonetics

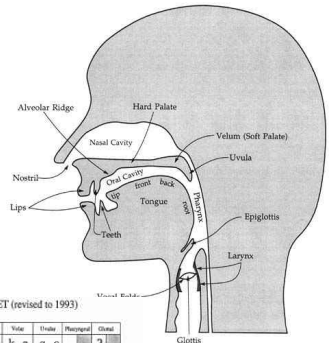
- What are the functions of
  - the outer ear?
  - the middle ear?
  - the inner ear?
- What are
  - the ossicles?
  - the oval window?
  - the cochlea?
  - the basilar membrane?

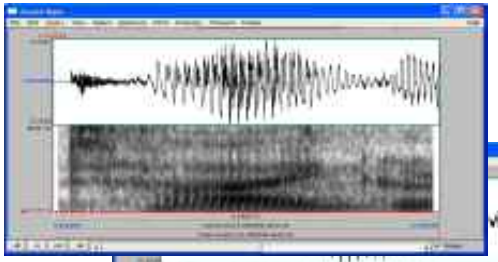
Summary: Phonetic Models

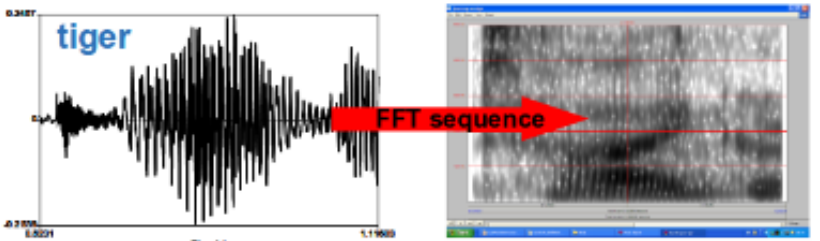


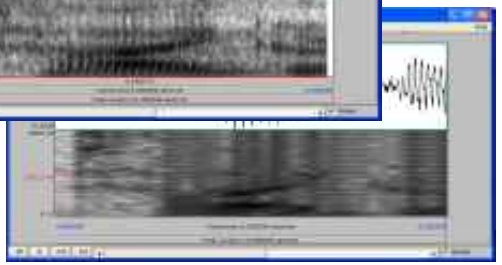


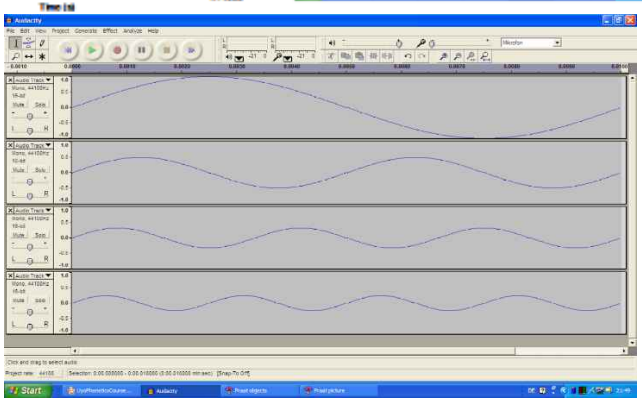


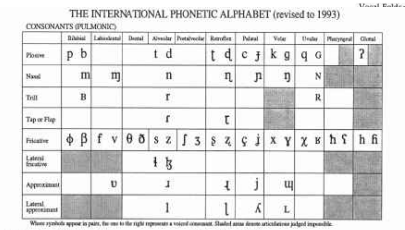


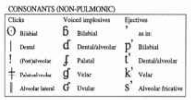


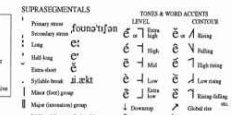


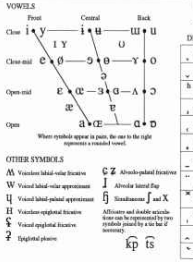


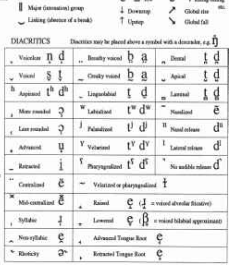


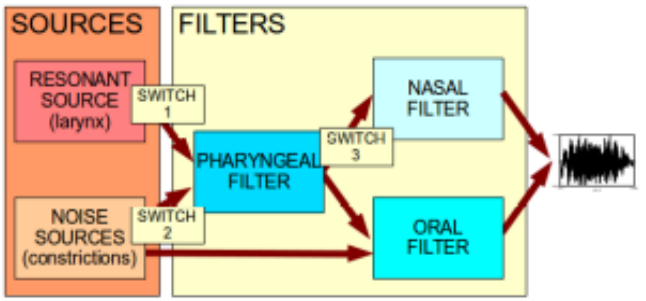


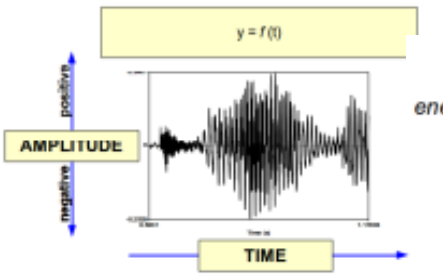


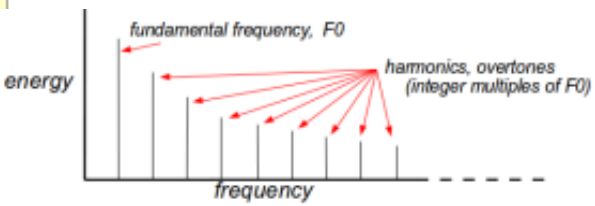












Summary Quiz

- With reference to the Source-Filter model of speech production,
  - How does the *source* relate to
    - the articulatory organs?
    - the acoustic signal?
    - the anatomy of the ear?
  - How does the *filter* relate to
    - the articulatory organs?
    - the acoustic signal?
    - the anatomy of the ear?
  - How do the *switch* relate to
    - the articulatory organs?
    - the acoustic signal?
    - the anatomy of the ear?