

The Music of Speech

Acoustic Phonetics

Dafydd Gibbon

Mannheim Summer School, July 2019

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

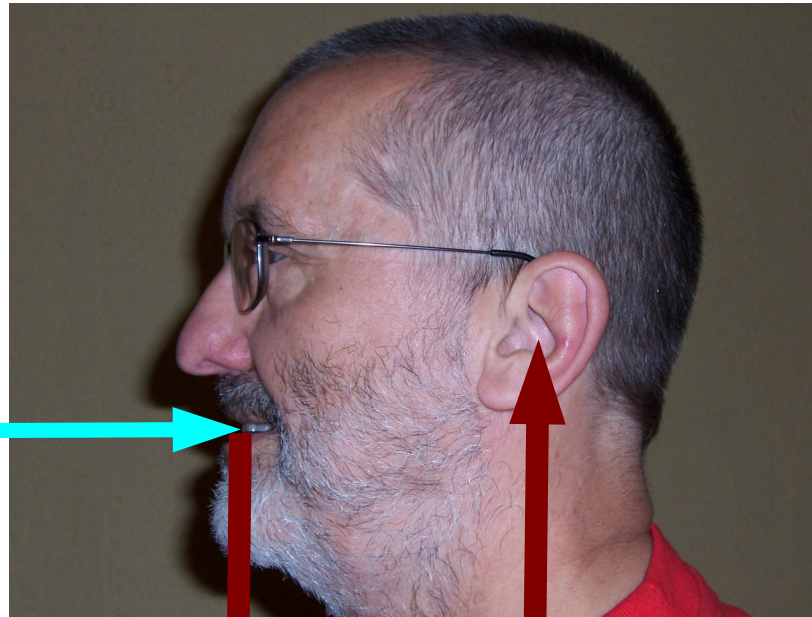
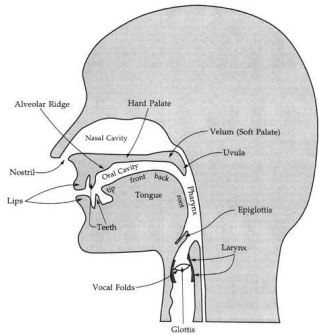


A tiger and a mouse were walking in a field

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

The Domains of Phonetics: the Phonetic Cycle

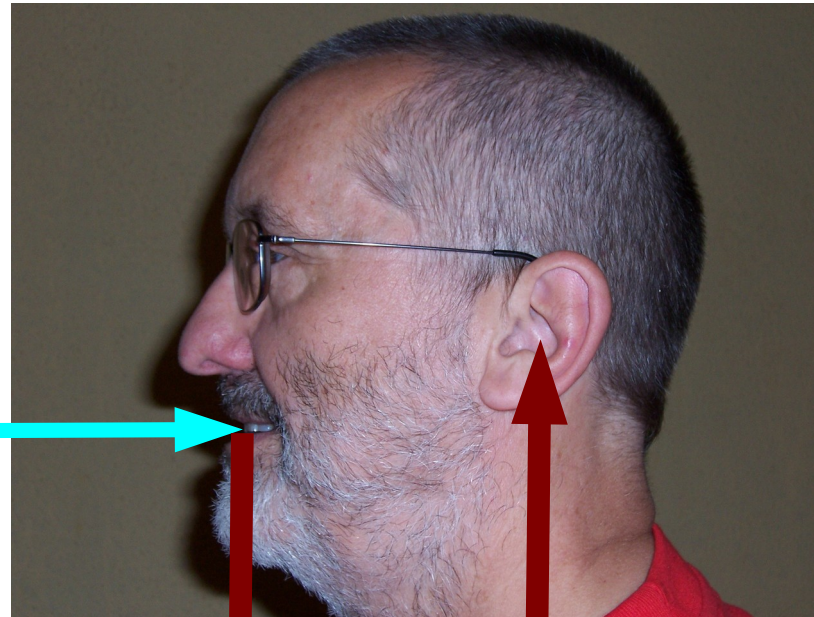
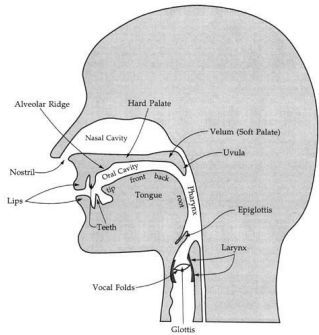
Articulatory
Phonetics



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

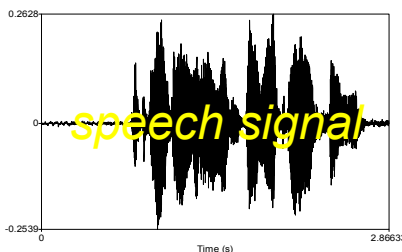
The Domains of Phonetics: the Phonetic Cycle

Articulatory
Phonetics



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

Acoustic
Phonetics



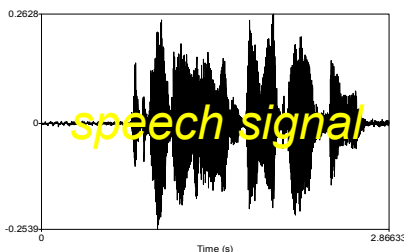
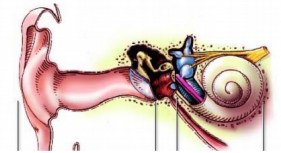
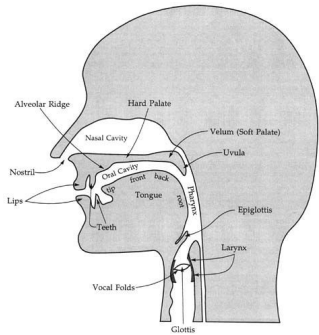
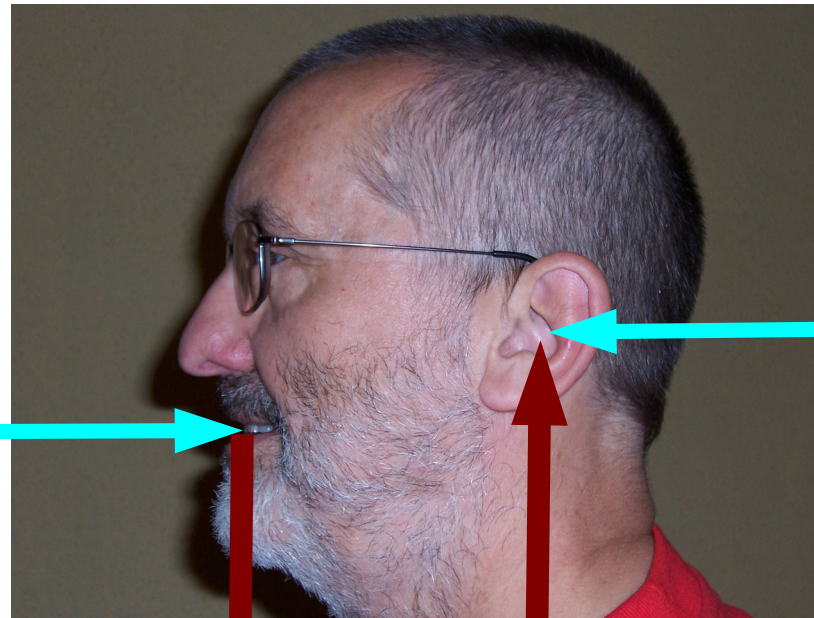
The Domains of Phonetics: the Phonetic Cycle

Articulatory Phonetics

Auditory Phonetics

Acoustic Phonetics

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



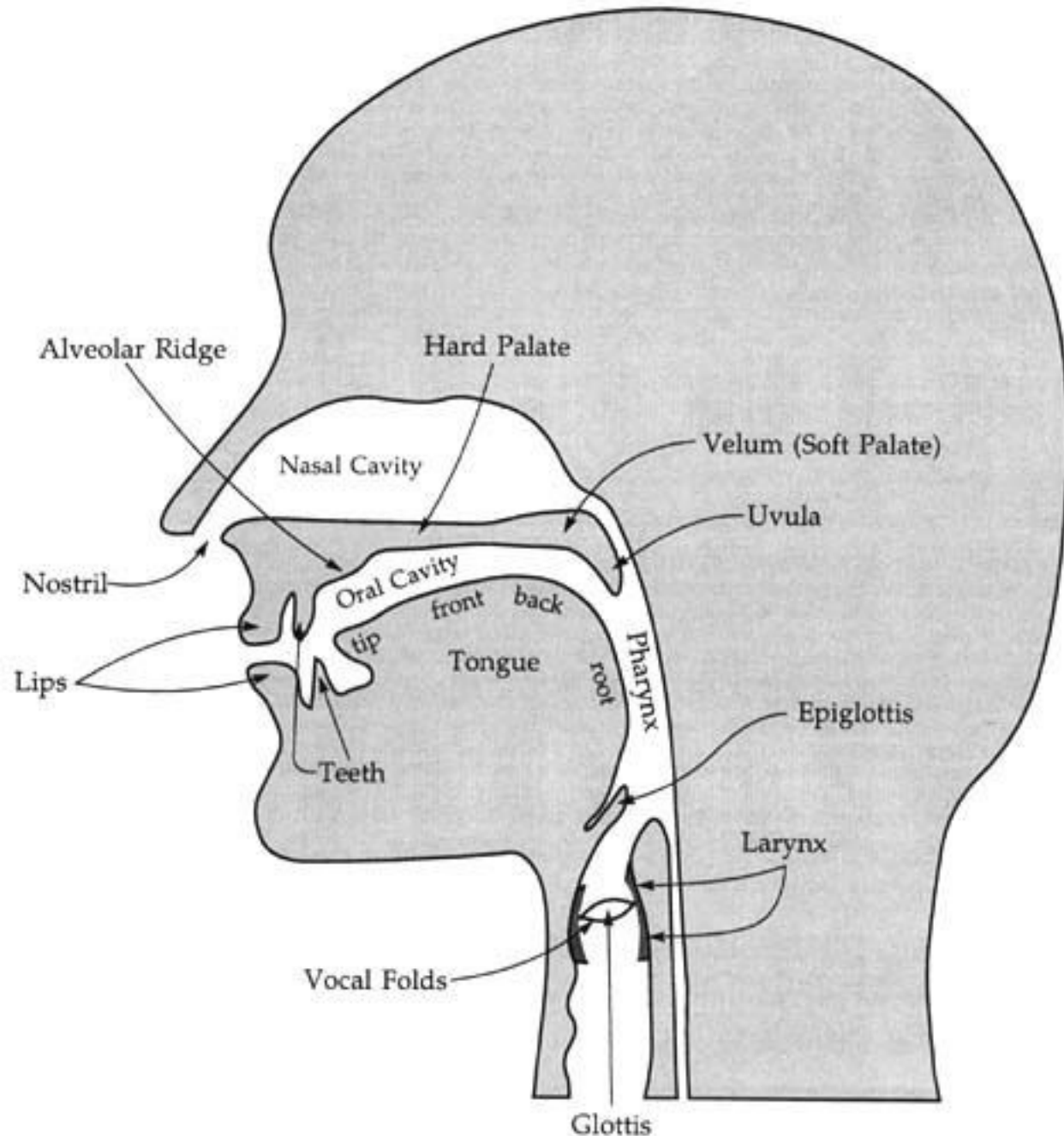
Quiz on the Phonetic Cycle

- 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				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.

- **IPA: 1897**
 - Latest revision: 2005
 - articulatory categories

- **transcription of**
 - the phonemes
 - of all languages of the world

- **phoneme:**
 - a vowel or consonant which distinguishes words in at least one language

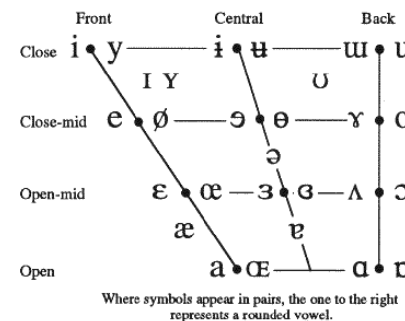
CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
◌ Bilabial	ɓ Bilabial	ʼ as in:
◌ Dental	ɗ Dental/alveolar	ɓ' Bilabial
◌ (Post)alveolar	ɟ Palatal	ɗ' Dental/alveolar
◌ Palatoalveolar	ɠ Velar	ɟ' Velar
◌ Alveolar lateral	ɣ Uvular	ɠ' Alveolar fricative

SUPRASEGMENTALS

	TONES & WORD ACCENTS
	LEVEL CONTOUR
ˈ Primary stress	ˈ or ˌ Rising
ˌ Secondary stress	ˌ or ˈ Falling
ː Long	ː High rising
ˑ Half-long	ˑ Mid rising
ˑ Extra-short	ˑ Low rising
◌ Syllable break	◌ Extra low rising-falling etc.
◌ Minor (foot) group	◌ Downstep
◌ Major (intonation) group	◌ Upstep
◌ Linking (absence of a break)	◌ Global fall

VOWELS



OTHER SYMBOLS

ɱ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
ɰ	Voiced labial-velar approximant	ɭ	Alveolar lateral flap
ɱ	Voiced labial-palatal approximant	ɥ	Simultaneous ʃ and X
ħ	Voiceless epiglottal fricative		Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
ʕ	Voiced epiglottal fricative		
ʔ	Epiglottal plosive		

kp ts

DIACRITICS

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

◌ Voiceless	◌ Breathy voiced	◌ Dental
◌ Voiced	◌ Creaky voiced	◌ Apical
◌ Aspirated	◌ Linguolabial	◌ Laminal
◌ More rounded	◌ Labialized	◌ Nasalized
◌ Less rounded	◌ Palatalized	◌ Nasal release
◌ Advanced	◌ Velarized	◌ Lateral release
◌ Retracted	◌ Pharyngealized	◌ No audible release
◌ Centralized	◌ Velarized or pharyngealized	
◌ Mid-centralized	◌ Raised	
◌ Syllabic	◌ Lowered	
◌ Non-syllabic	◌ Advanced Tongue Root	
◌ Rhoticity	◌ Retracted Tongue Root	

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		ɳ̠	ɲ	ŋ	ɴ		
Trill				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.

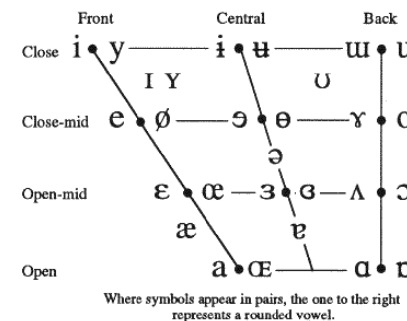
CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
◌ Bilabial	ɓ Bilabial	ʼ as in:
◌ Dental	ɗ Dental/alveolar	ɓ' Bilabial
◌ (Post)alveolar	ɗ' Palatal	ɗ' Dental/alveolar
◌ Palatoalveolar	ɠ Velar	ɠ' Velar
◌ Alveolar lateral	ɠ' Uvular	ɠ' Alveolar fricative

SUPRASEGMENTALS

	TONES & WORD ACCENTS
	LEVEL CONTOUR
ˈ Primary stress	ˈ founəˈtɪʃən
ˌ Secondary stress	ˌ
ː Long	ː eɪ
ˑ Half-long	ˑ eɪ
ˑ Extra-short	ˑ eɪ
· Syllable break	· iː.ækt
ˌ Minor (foot) group	ˌ eɪ
ˎ Major (intonation) group	ˎ
◌ Linking (absence of a break)	◌
	↑ Downstep
	↗ Upstep
	↘ Global fall
	↗ Rising
	↘ Falling
	↗ High rising
	↘ Low rising
	↗ Rising-falling etc.

VOWELS



OTHER SYMBOLS

ɸ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
ʋ	Voiced labial-velar approximant	ɭ	Alveolar lateral flap
ɰ	Voiced labial-palatal approximant	ɥ	Simultaneous ʃ and X
ħ	Voiceless epiglottal fricative		Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
ʕ	Voiced epiglottal fricative		
ʔ	Epiglottal plosive		

kp ts

- transcription of
 - the phonemes
 - of all languages of the world
- phoneme:

– a vowel or consonant which distinguishes words in at least one language

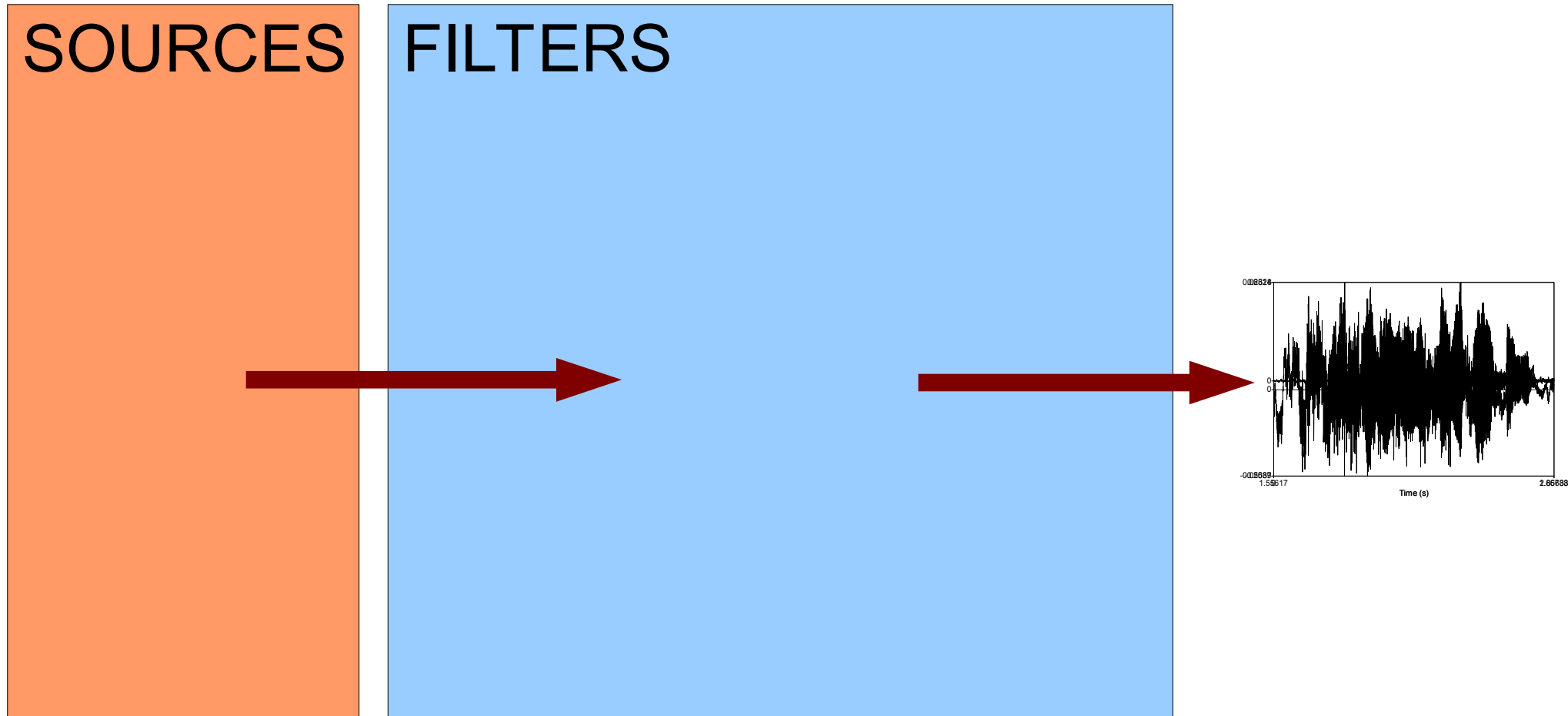
- /kit/ /kæɪt/ /kɒt/ /kʌt/ /ku:t/
- /kɜ:t/ /kɑ:t/ /kɔ:t/ /kəʊt/ kɔɪt/

- vowels are the main factors in rhythm and melody

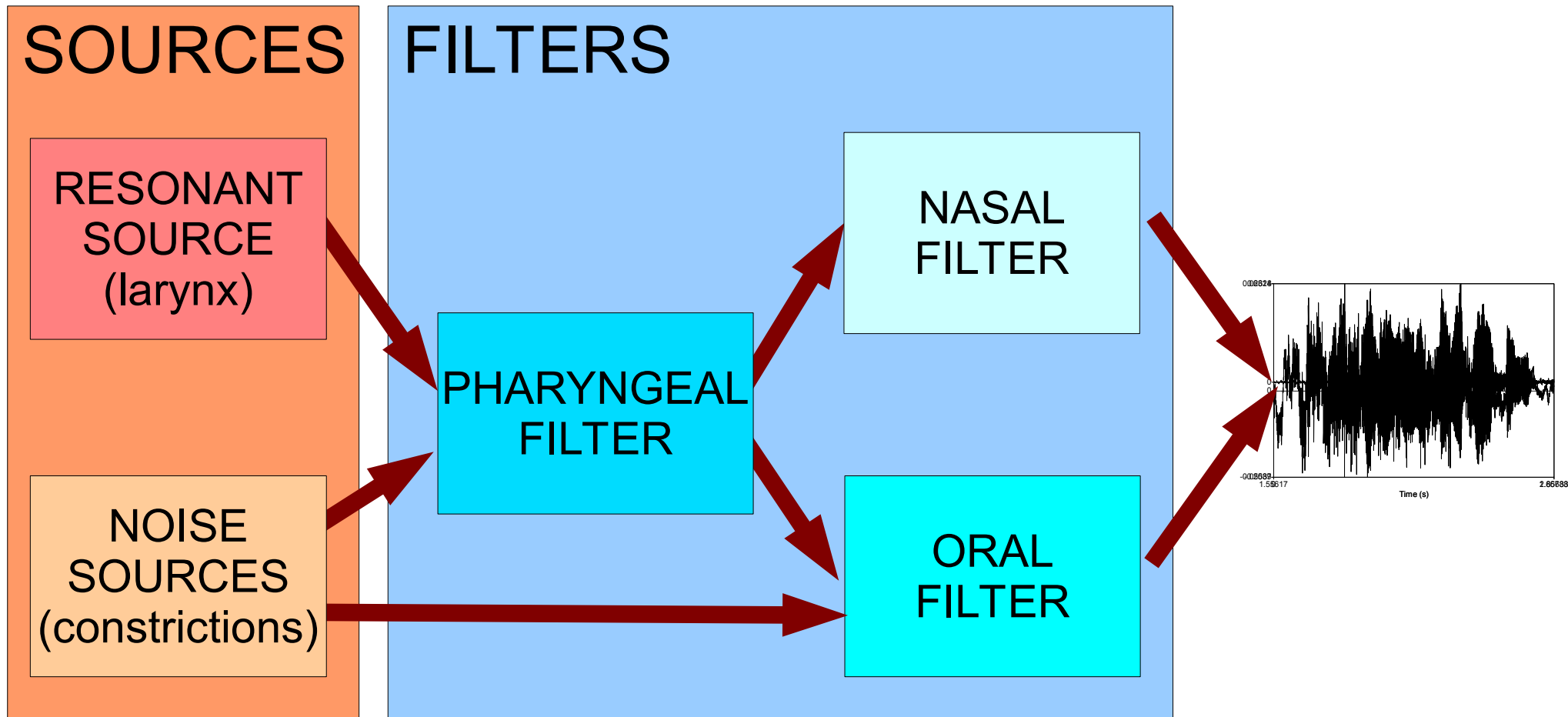
The Source-Filter Model of Speech Production

- A “model” is a simplified representation of relevant features of reality (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**
- Two types of **source**:
 - Larynx: melody (tone, intonation)
 - Narrowing / closing of the mouth (noisy consonants)
- Three types of **filter**:
 - the PHARYNGEAL CAVITY (throat)
 - the ORAL CAVITY (mouth)
 - the NASAL CAVITY (nose)

The Source-Filter Model of Speech Production



The Source-Filter Model of Speech Production



Quiz on Articulatory Phonetics

- Which are the main articulators involved in
 - vowel production?
 - consonant production?
 - tone production?
- Produce these consonants, followed by the vowel [a]:
 - voiceless bilabial fricative
 - 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.

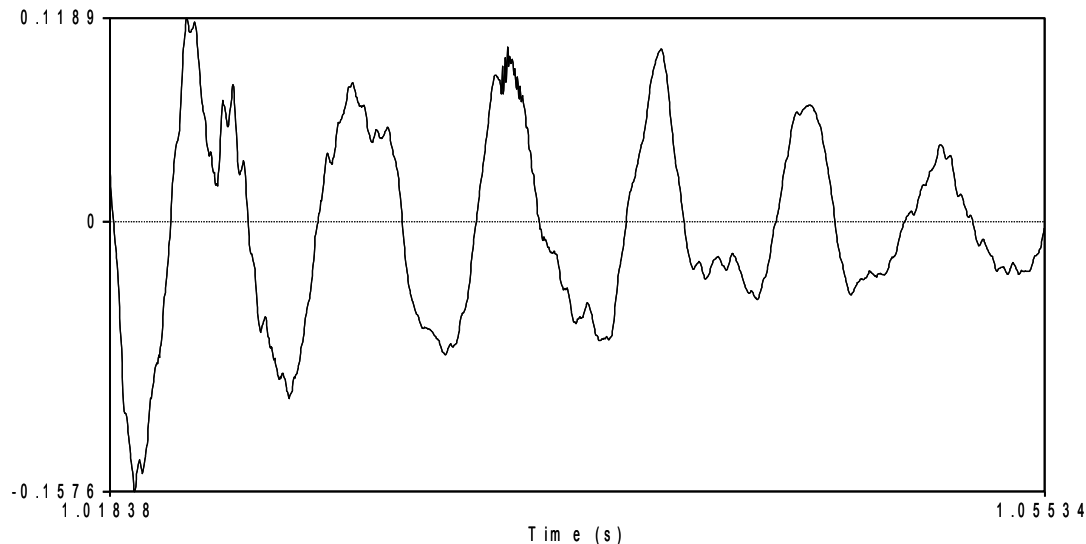
Acoustic Phonetics (Speech Transmission)

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** → **energy, intensity, loudness**
 - **Frequency** → **melody, pitch: tone, intonation**
 - **Time** → **duration, rhythm**
- The **methods** used to analyse speech signals are:
 - observation, transcription, description by a trained phonetician
 - analog-to-digital (A/D) conversion
 - mathematical definitions of filters and transformations

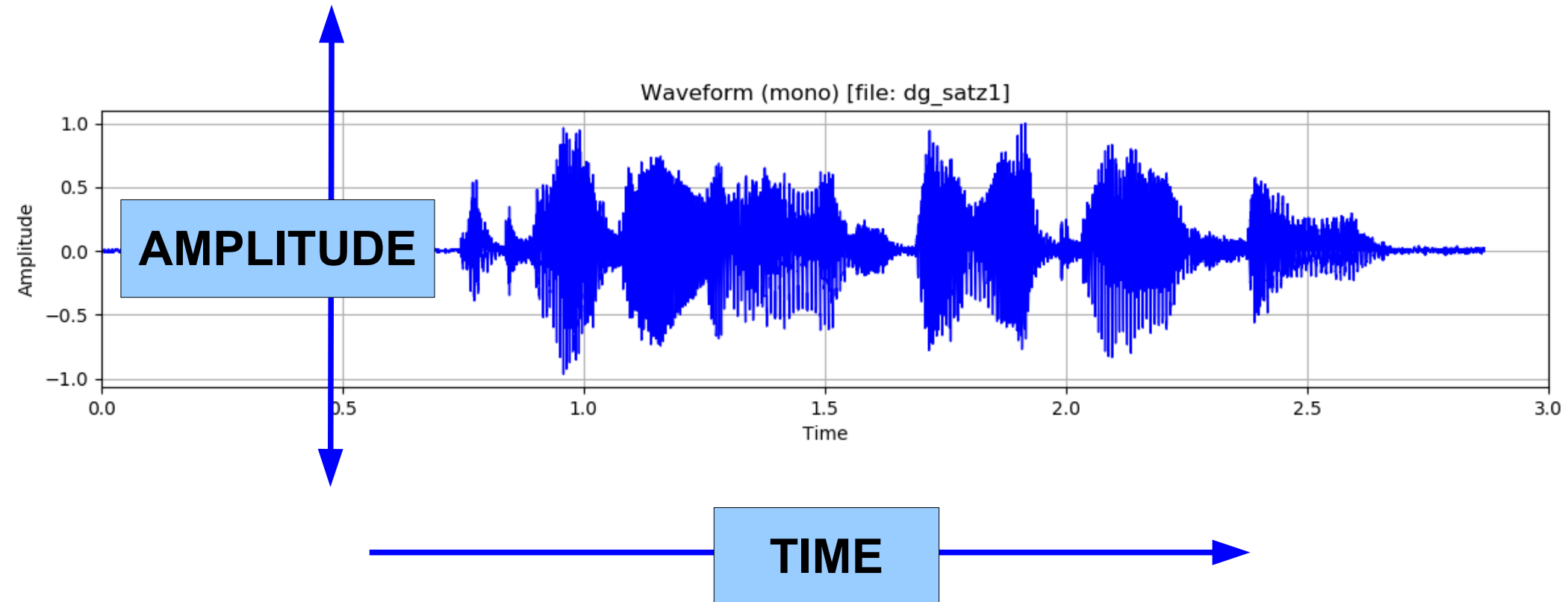
The Speech Wave-Form

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



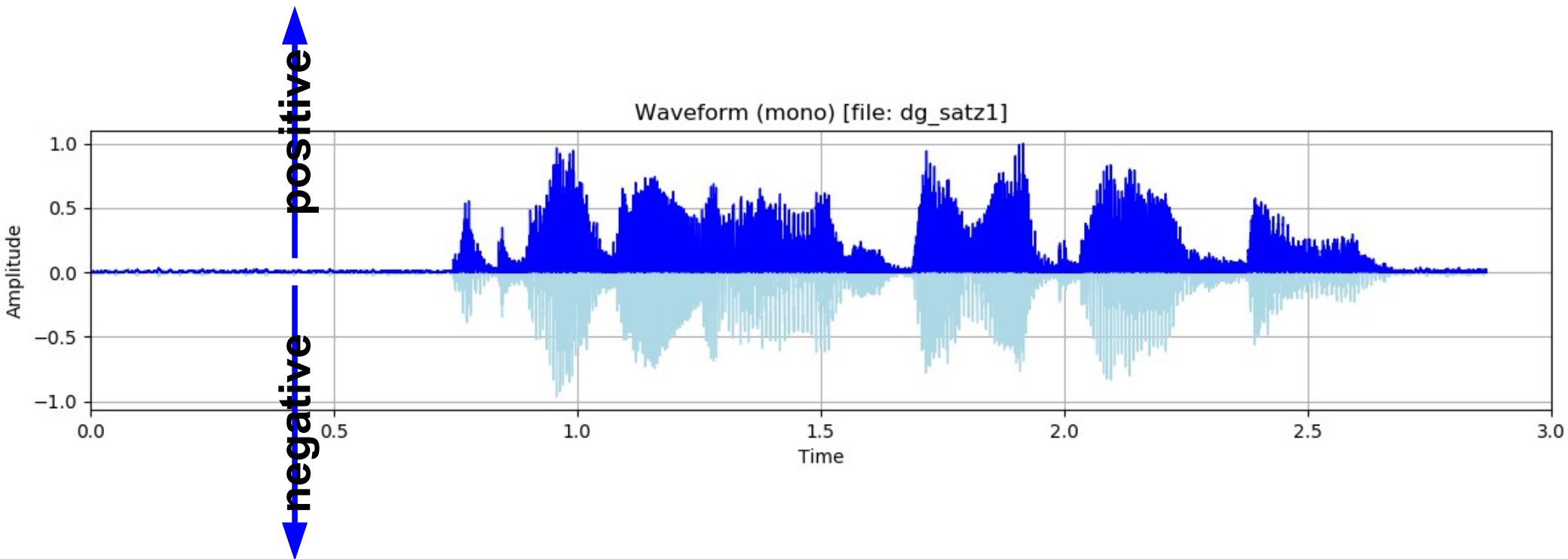
- The changes in air pressure
 - can be heard
 - can be measured (like the waves on the ocean)
 - the measurements can be visualised and used for calculating statistical models of the structure of speech

Basic Speech Signal Parameters



The Time Domain: the Speech Wave-Form

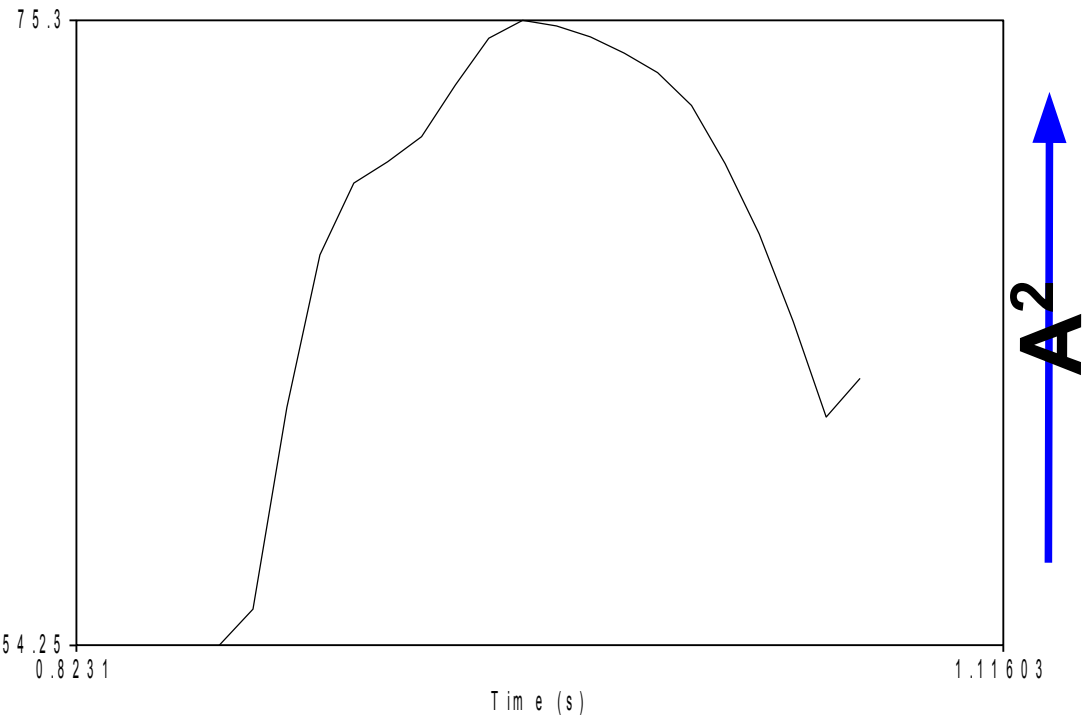
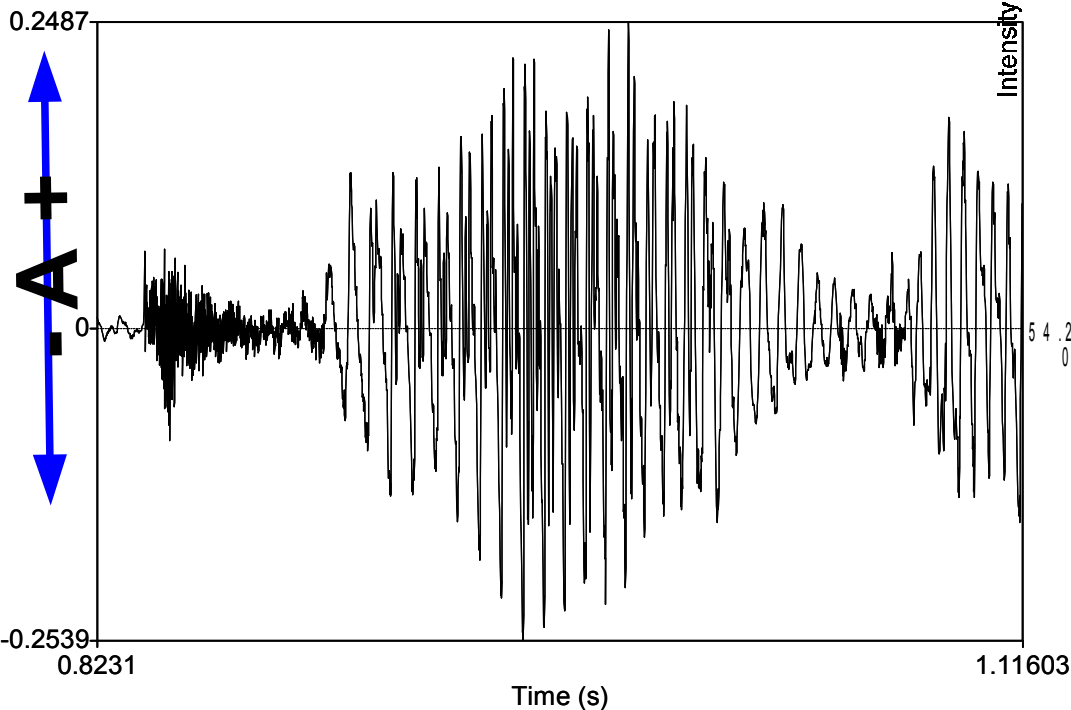
Positive or negative amplitude A of the speech signal:



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, \dots, I_n (remember: intensity is amplitude squared)

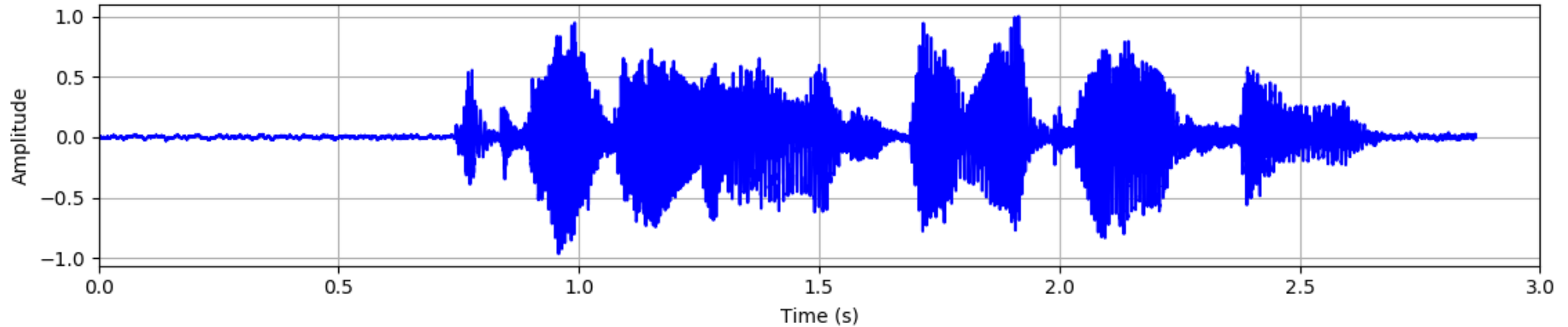
$$E = \sqrt{\frac{\sum_{i=1 \dots n} A(x_i)^2}{n}}$$

- Energy is 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

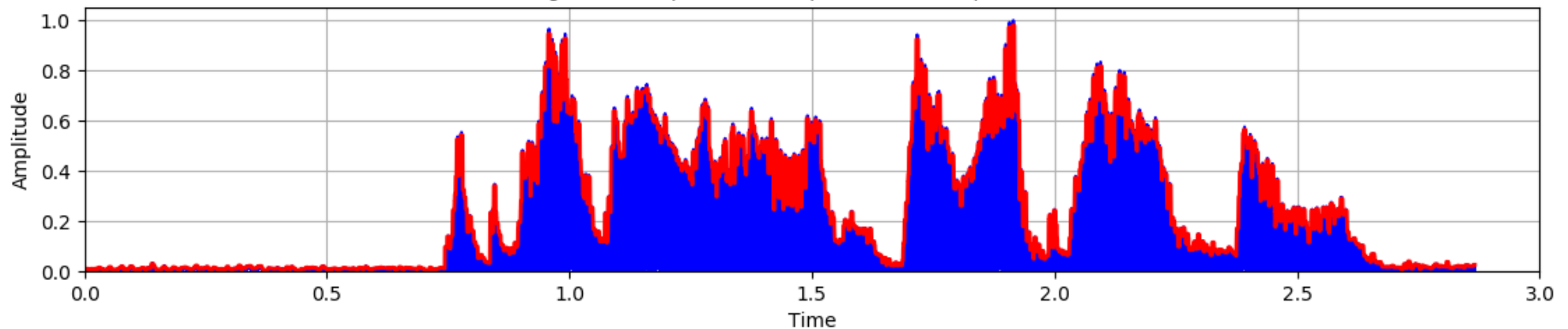
The Time Domain: the Speech Wave-Form

Envelope extraction [file: dg_satz1]

Waveform oscillogram (mono)



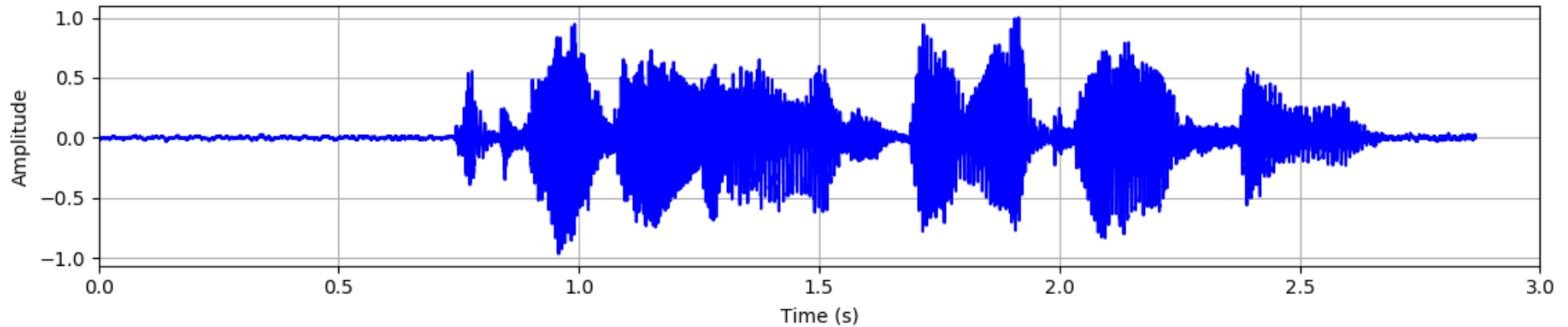
Rectified (absolute) signal and Amplitude Envelope (Peak-Picker: peakwin=0.0040, medfiltwin=0)



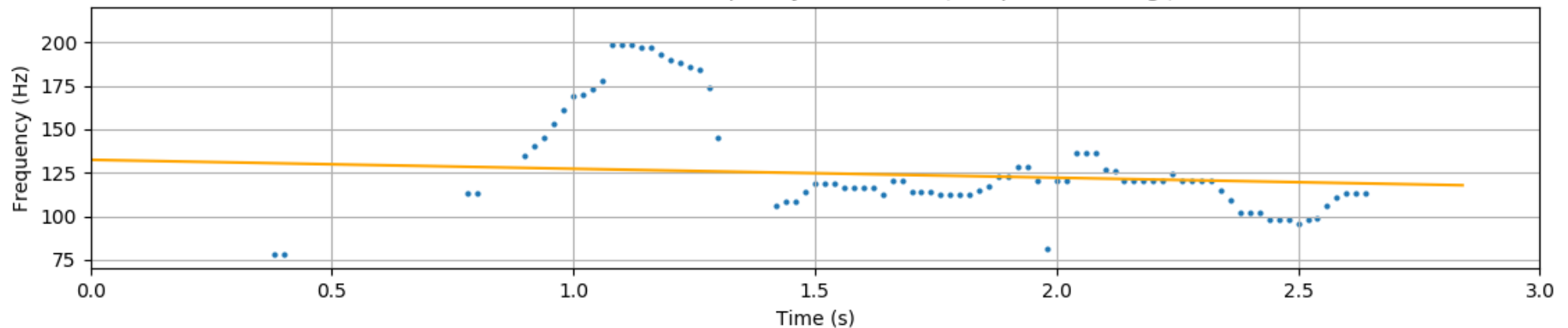
The Time Domain: the Fundamental Frequency

F0 estimation [file: dg_satz1]

Waveform oscillogram (mono)



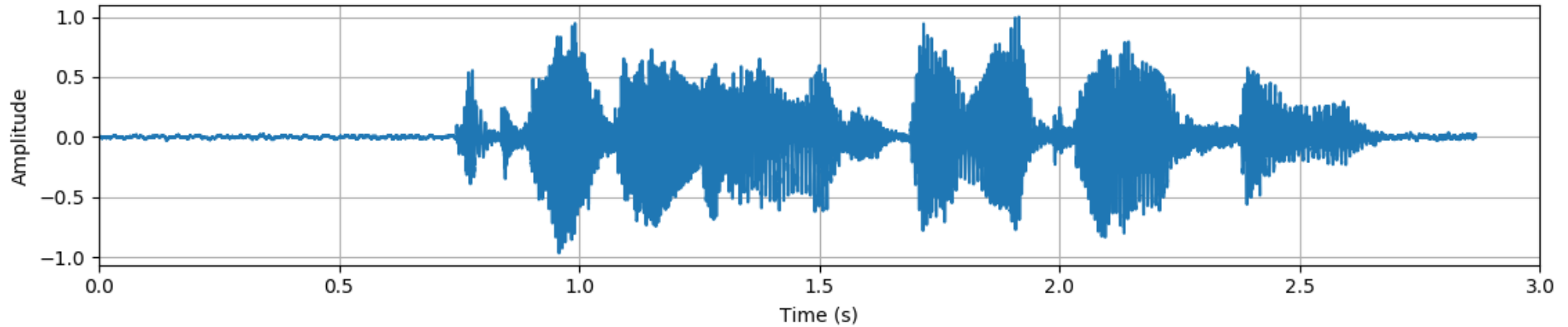
AMDF Fundamental frequency estimation (F0, 'pitch tracking')



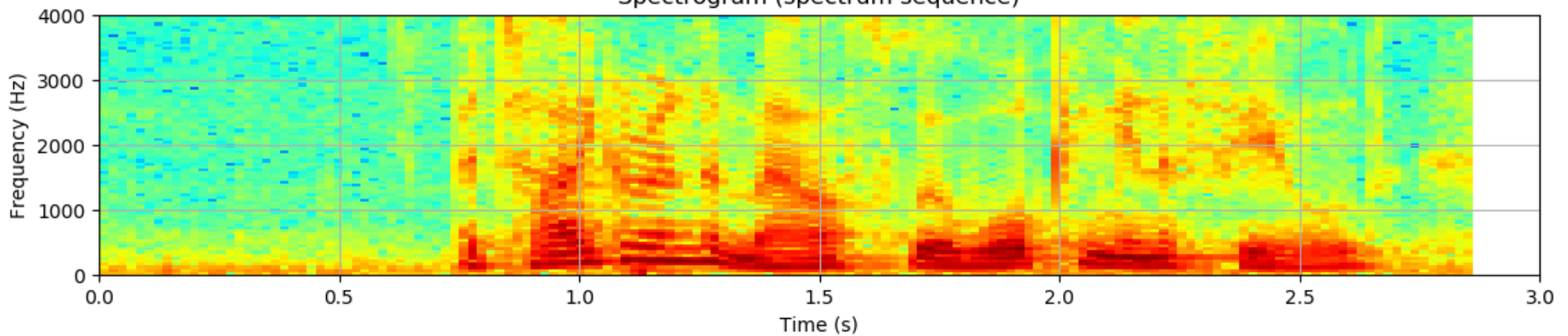
Time and Frequency Domains: the Spectrogram

Spectrogram [file: dg_satz1]

Waveform oscillogram (mono)



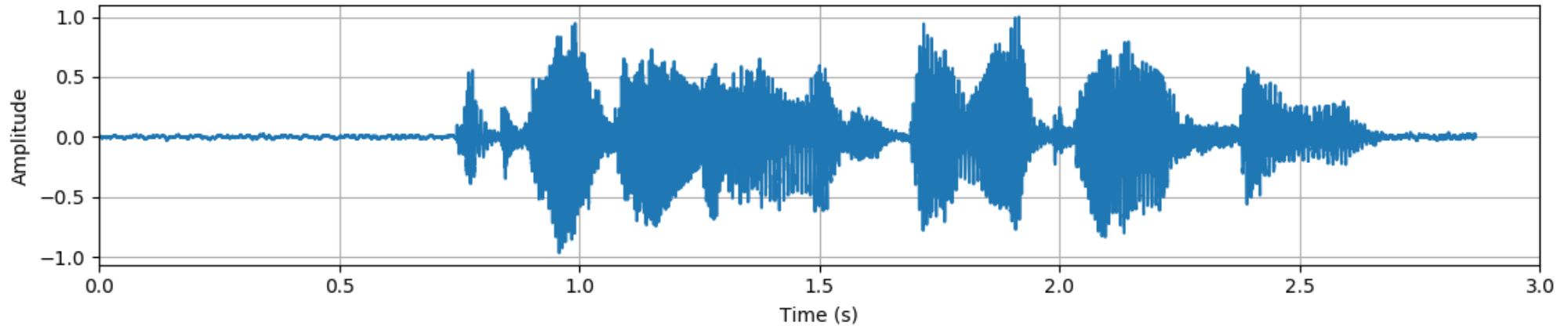
Spectrogram (spectrum sequence)



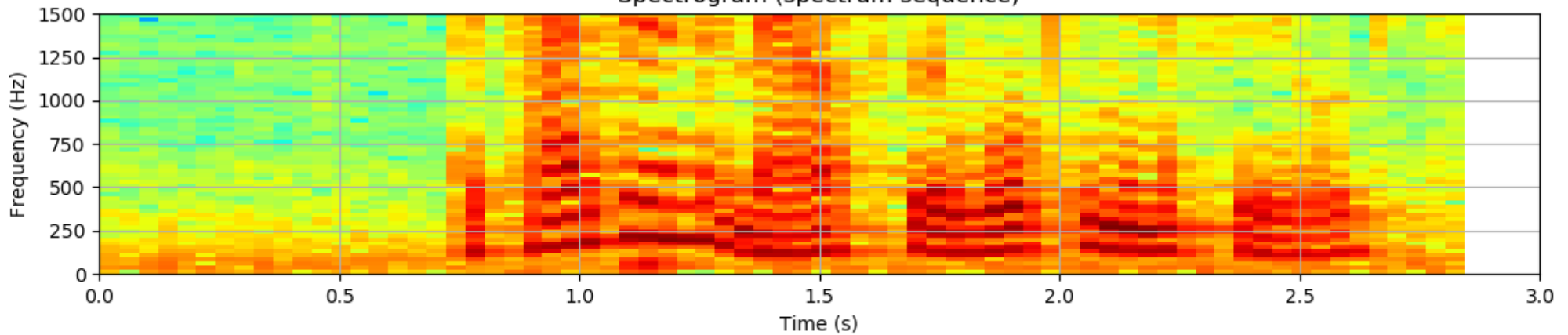
Time and Frequency Domains: the Spectrogram

Spectrogram [file: dg_satz1]

Waveform oscillogram (mono)

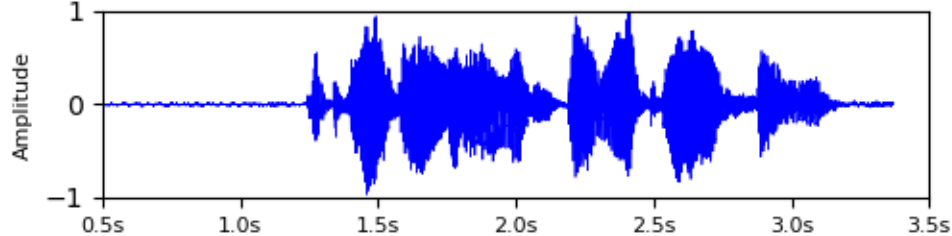


Spectrogram (spectrum sequence)

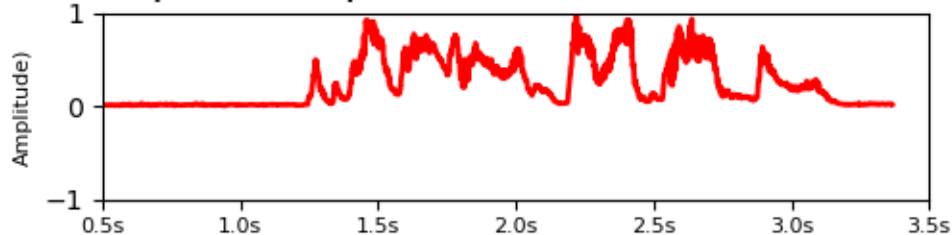


Four Acoustic Phonetic Visualisations

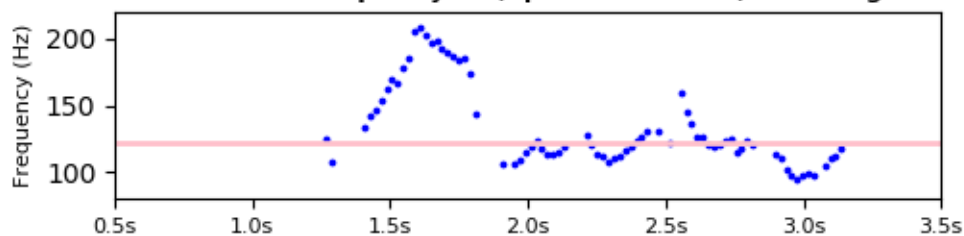
1. Waveform (mono)



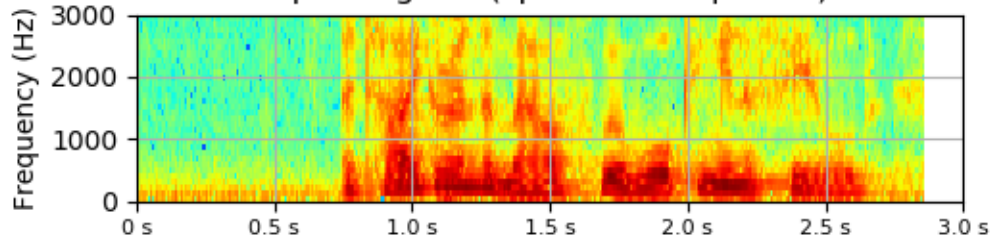
2. Amplitude Envelope (Emulation of Absolute Hilbert Transform)



3. Fundamental Frequency (F0, 'pitch') estimate, AMDF algorithm



Spectrogram (spectrum sequence)



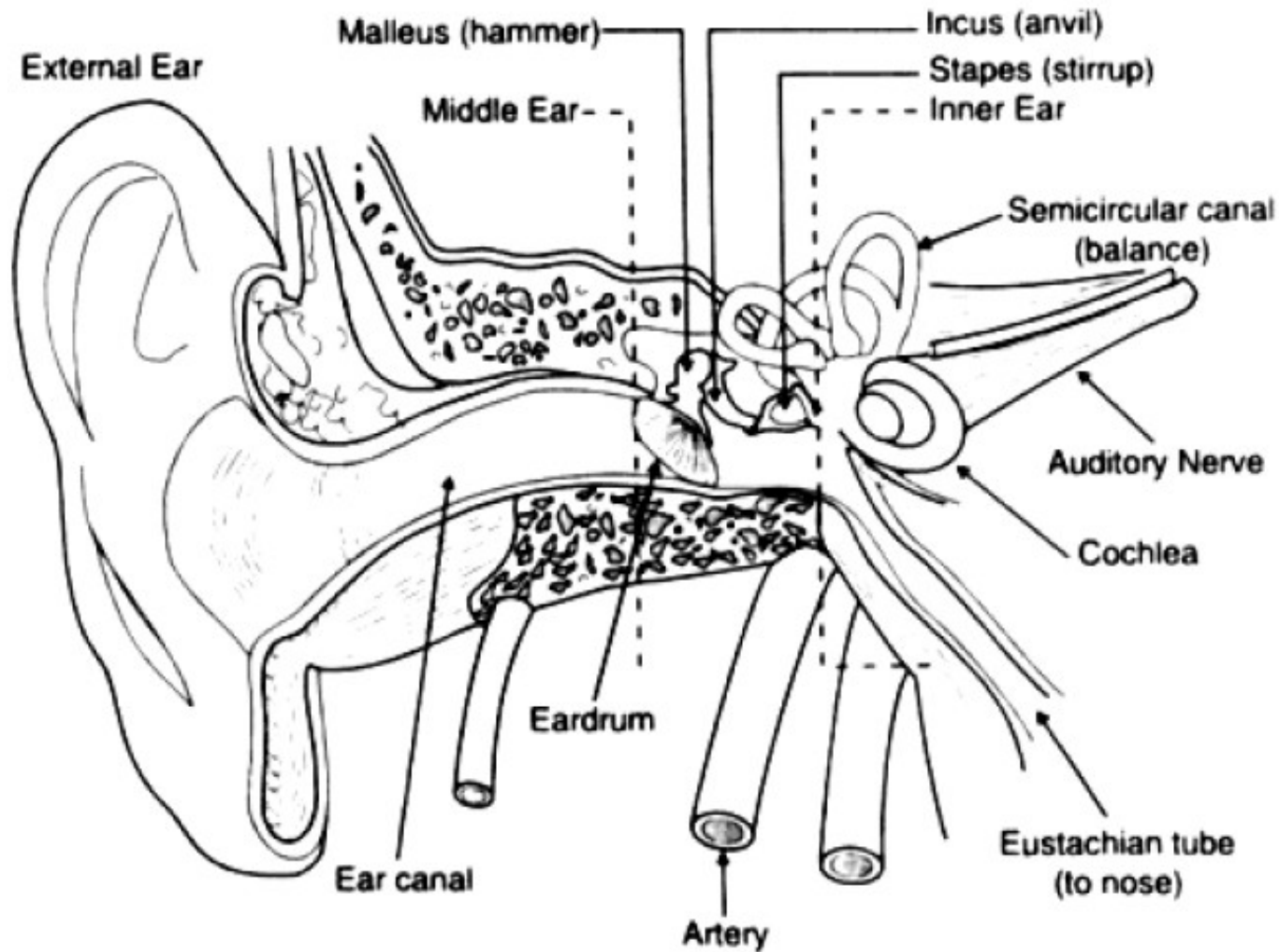
... almost all you need to know about the rhythms and melodies of speech ...

Quiz on Acoustic Phonetics

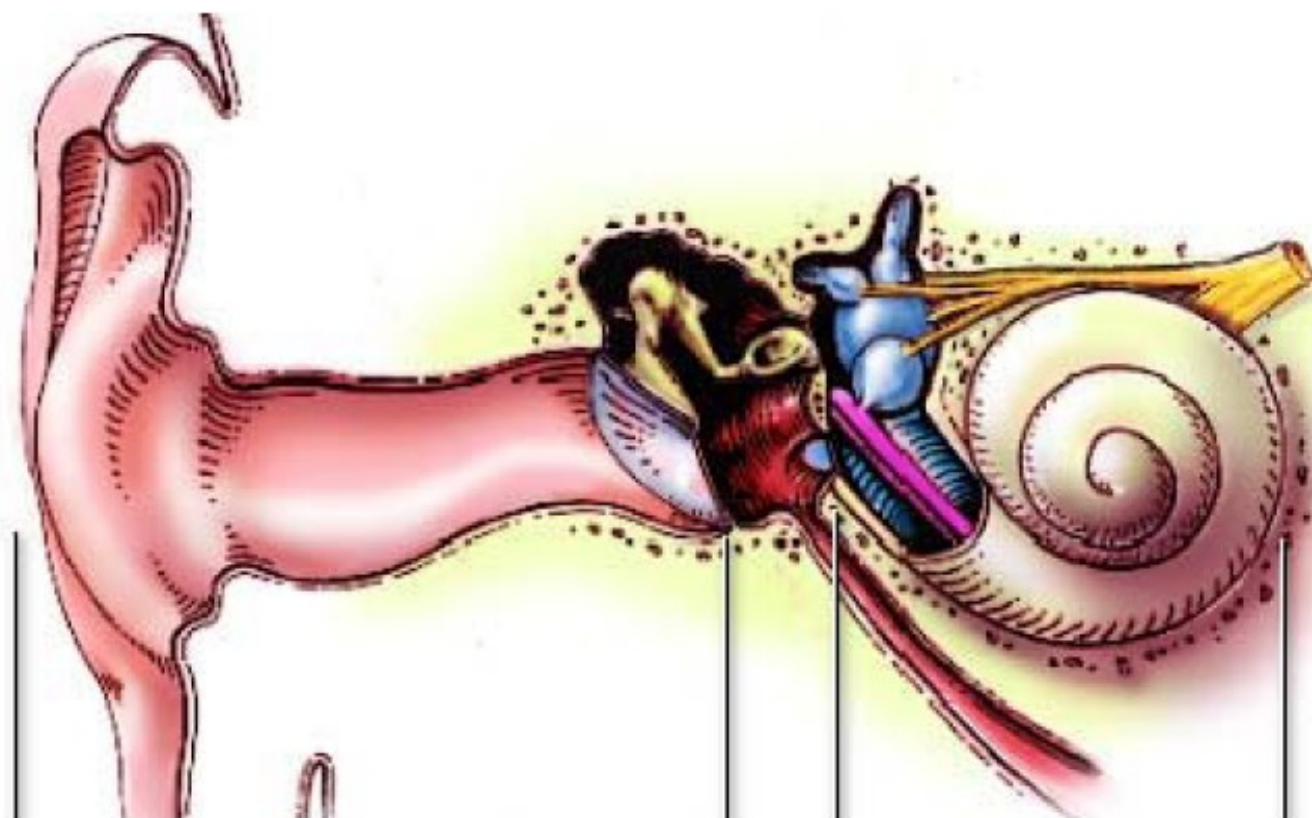
- What are the basic parameters of the speech signal?
- Define the following terms:
 - amplitude
 - intensity
 - energy
- How are time-domain representations of speech signal converted to frequency domain representations?
- Define the following terms:
 - Spectrum, spectrogram
 - fundamental frequency, 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

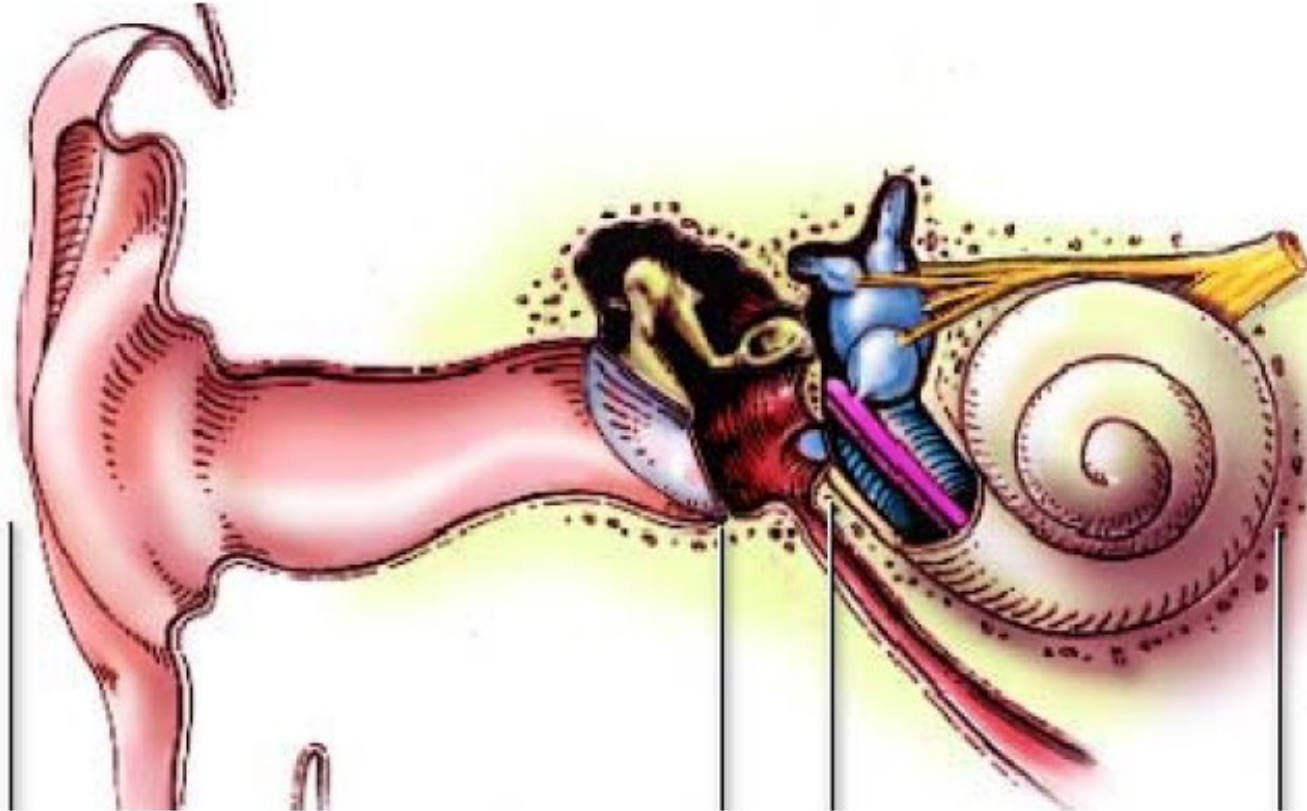
middle ear

The Auditory Domain: Anatomy of the Ear

microphone

amplifier

Fourier transform



outer ear

inner ear

middle 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 this unit

- you should know the basic physical foundations
- on which rhythms and melodies are based