

# Spraakherkenning en -synthese, lecture 1

Paul Boersma, September 2019

# Contents of lecture 1

- Organization
- The sound chain

# Organization

- 14 lectures:
  - Tuesdays 13:00–15:00
  - Wednesdays 17:00–19:00 (except lecture 12: Thursday 10 Oct 13:00–15:00)
- 14 laptop classes:
  - Tuesdays 15:00–19:00
  - Thursdays 15:00–19:00
- 3 small tests (in laptop classes 4, 8 and 11)
- 1 final exam in week 8

# Organization: homework

- Every second hour of a lecture leads to homework.
- Homework consists of reading and doing assignments.
- Homework is largely done during laptop class, where you can interact about it with your fellow students and your teachers.
- If you hand in your homework (on Canvas) at least 5 hours before the next lecture, you are guaranteed to receive some personal feedback.
- The personal feedback is usually given (on Canvas) before the next lecture.
- The homework is discussed in the first hour of the next lecture.

# Organization: homework for odd lectures

- Subject 3 presented at Tuesday lecture (14:00–15:00, i.e. second hour)
- Homework 3 started during Tuesday laptop class (15:00–19:00)
  - reading on subject 3; doing assignments 3
- Homework 3 (perhaps) continued at home
- Hand in assignments 3:
  - before Wednesday 12:00 if you want to receive feedback
  - deadline Wednesday 17:00
- Assignments 3 discussed at Wednesday lecture (17:00–18:00)

# Organization: homework for even lectures

- Subject 4 presented at Wednesday lecture (18:00–19:00, i.e. 2<sup>nd</sup> hour)
- Homework 4 started at home
  - reading subject 4; doing assignments 4
- Homework 4 continued during Thursday laptop class: 15:00–19:00
- Homework 4 perhaps continued at home
- Hand in assignments 4:
  - before next Tuesday 8:00 if you want to receive feedback
  - deadline next Tuesday 13:00
- Assignments 4 discussed at next Tuesday lecture (13:00–14:00)

# Course contents: four modules

1. Sound & Pitch
  2. Spectrum (requires knowledge of Sound & Pitch)
  3. Speech Technology (requires knowledge of Spectrum)
  4. Deep Speech (requires knowledge of Speech Technology)
- The first three modules end in a short test (15% each).
  - The final exam (55%) tests all four modules.

# Do look on Canvas

- “Modules”:
  - see all 28 classes
- “Home”:
  - “Ten ways to learn”
  - “Teachers”
  - “General material”
  - “Grading”
    - handing in your homework assignments is not optional!

# Teachers

- prof.dr. Paul Boersma, Phonetic Sciences, UvA
  - paul.boersma@uva.nl
- Bence Halpern MSc, NKI (not available in weeks 1 and 3)
  - b.halpern@nki.nl

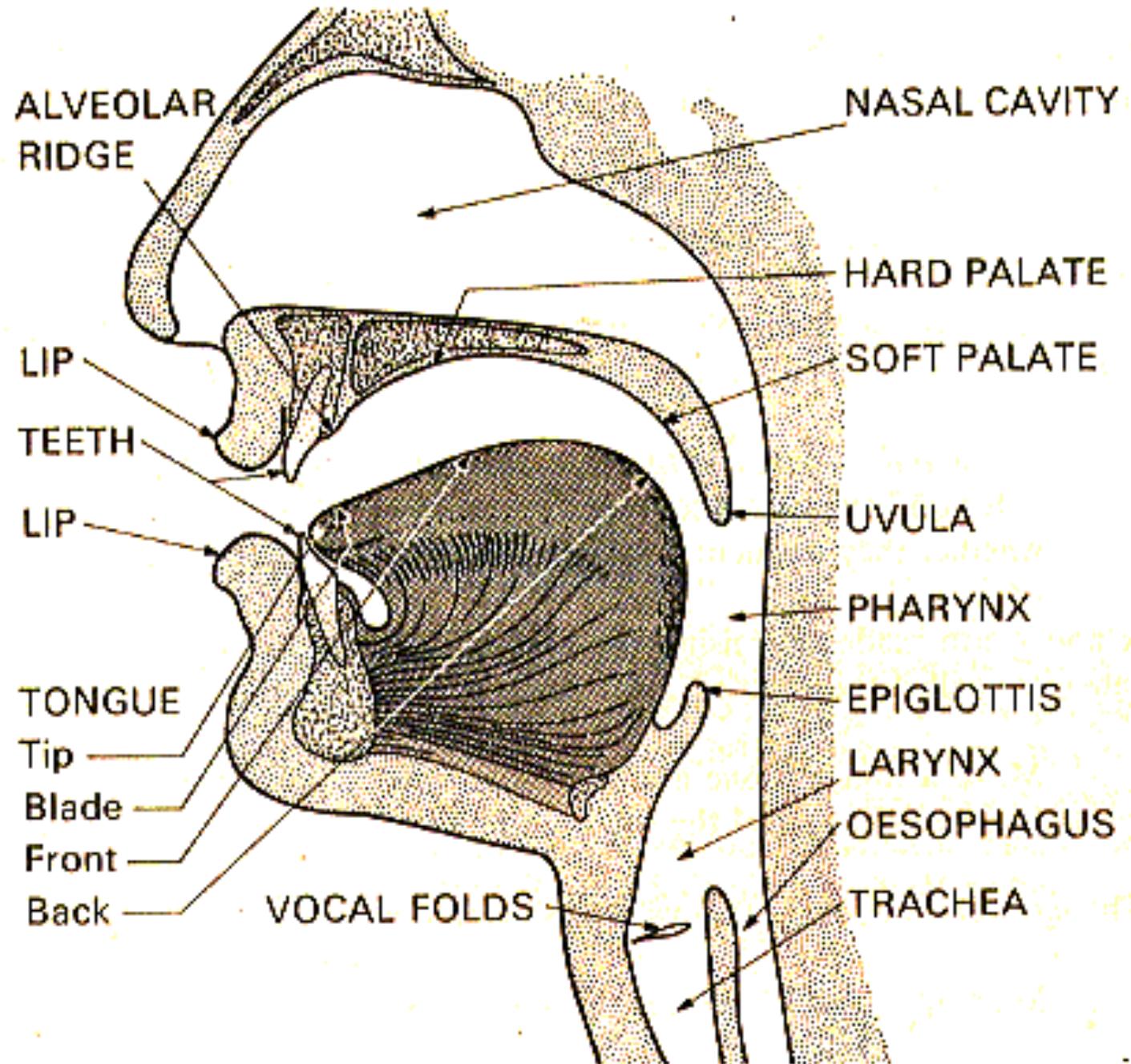
# The speech chain

- Intention of the speaker (higher centres of the brain)
- Language (semantics, syntax, morphology, lexicon, phonology)
- Phonetics (auditory targets, motor program)
- Muscle movement (diaphragm, larynx, tongue, jaw, lips,...)
- Sound wave in air (vibrations radiate through air)
- Hearing (eardrum, ossicles, basilar membrane, auditory nerve)
- Auditory processing (brainstem, thalamus, auditory cortex)
- Language (phonology, lexicon, morphology, syntax, semantics)
- Understanding by the listener (higher centres of the brain)

# Phonetics and phonology

- Phonetics
  - The physics of speaking, sound and hearing
  - Continuous events: movements, gestures, spectra, pitch
  - Processes: coarticulation, overlap
- Phonology
  - The organization of speech in the mind
  - Discrete elements: phrases, segments, phonemes, features
  - Processes: assimilation, deletion
- Phonetics–phonology interface (bidirectional)
  - Auditory cues to phonological elements (high-frequency noise  $\sim$  /s/)
  - Production and perception mechanisms

# Vocal tract



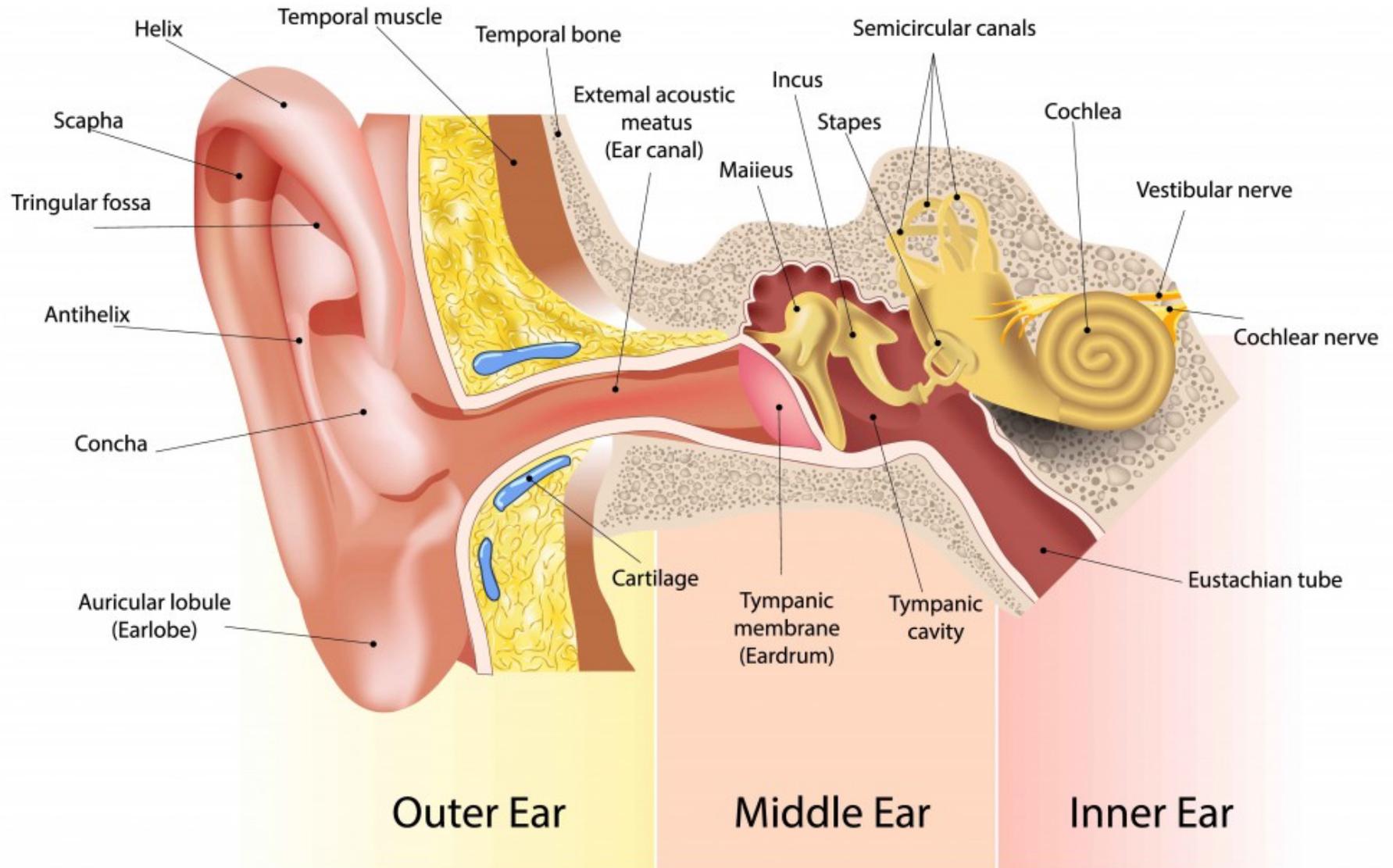
(from Wells & Colson 1971)

# Human voiced sound production

- Air pressure from lungs (diaphragm, intercostal muscles) creates steady airflow through trachea (windpipe), larynx (voice box), and pharynx (throat).
- Vocal folds are brought together, *so that* (Bernoulli effect) they start to vibrate, generating air pressure fluctuations known as sound waves.
- Resonances in vocal tract modify these waves according to position and shape of lips, jaw, tongue, soft palate, and other speech organs, creating multiple prominent regions in the spectrum (*formants*).
- Mouth, nose and neck radiate sound waves into the environment.

# Ear

## Anatomy of the Ear



*(malleus)*

# Source–filter model of speech production

- — an approximation of speech as a combination of a sound *source* and a linear acoustic *filter*, which are largely independent.
- Source:
  - Glottal vibrations (periodically closing constriction)
  - Noisy airflow through a fixed constriction
  - Jets (nonlocal noise following a fixed constriction)
  - Trills (periodically closing constriction downstream from the glottis)
- Filter:
  - Oral, nasal and pharyngeal cavities, mostly downstream from the constriction.

# The International Phonetic Alphabet

- <http://www.internationalphoneticassociation.org/content/ipa-chart>
- full 2015 version on Canvas at lecture 1 (license: CC-BY-SA)

# Phoneme inventories

- Speech is made up of segments.
- The number of possible segments is finite.
  - Possible segments are *phonemes*: vowels and consonants.
- The Dutch word *pop* consist of three segments: /pɔp/.
- In *pop*, the phoneme (consonant) /p/ occurs twice, and the phoneme (vowel) /ɔ/ occurs once.
- In *bij* /bɛi/, we may have three segments, because /ɛi/ is considered a single phoneme.
- Perhaps *boei* /buj/ may consist of three segments, because /u/ and /j/ are considered separate phonemes.

# The Dutch vowel inventory

- Closed (or high) vowels are /i/ *piet*, /y/ *puut*, /u/ *poet*
- Perhaps long ones in /i:/ *prestige*, /y:/ *centrifuge*, /u:/ *rouge*
- Short half-closed (or mid-high) vowels are /ɪ/ *pit*, /ʏ/ *put*
- Long half-closed (or mid-high) vowels are /e:/ *peet*, /ø:/ *peut*, /o:/ *poot*
- Short half-open (or mid-low) vowels are /ɛ/ *pet*, /ɔ/ *pot*
- Perhaps long ones in /ɛ:/ *beige*, /œ:/ *freule*, /ɔ:/ *zone*
- The short open (or low) vowel is /a/ *pat*
- The long open (or low) vowel is /a:/ *paat*
- The diphthongs are /ɛi/ *pijt*, /œy/ *puit*, /au/ *pout*

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): .....
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth):
  - Non-sibilants (generation of noise at the constriction):
- Nasals:
- Approximants:
- Trills:
- Taps:
- Affricates (plosive–fricative sequence):

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): .....
  - Non-sibilants (generation of noise at the constriction):
- Nasals:
- Approximants:
- Trills:
- Taps:
- Affricates (plosive–fricative sequence):

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): s, z, ʃ(ç) *sjaal*, ʒ(z) *rouge*
  - Non-sibilants (generation of noise at the constriction): .....
- Nasals:
- Approximants:
- Trills:
- Taps:
- Affricates (plosive–fricative sequence):

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): s, z, ʃ(ç) *sjaal*, ʒ(z) *rouge*
  - Non-sibilants (generation of noise at constriction): f, v, x/χ *lach*, ɣ *negen*, h
- Nasals: .....
- Approximants:
- Trills:
- Taps:
- Affricates (plosive–fricative sequence):

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): s, z, ʃ(ç) *sjaal*, ʒ(z) *rouge*
  - Non-sibilants (generation of noise at constriction): f, v, x/χ *lach*, ɣ *negen*, h
- Nasals: m, n, ŋ *bang*
- Approximants: .....
- Trills: .....
- Taps: .....
- Affricates (plosive–fricative sequence):

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): s, z, ʃ(ç) *sjaal*, ʒ(z) *rouge*
  - Non-sibilants (generation of noise at constriction): f, v, x/χ *lach*, ɣ *negen*, ɦ
- Nasals: m, n, ŋ *bang*
- Approximants: v *wie*, w *eew*, j, l, ɹ *boer*
- Trills: .....
- Taps: .....
- Affricates (plosive–fricative sequence):

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): s, z, ʃ(ç) *sjaal*, ʒ(z) *rouge*
  - Non-sibilants (generation of noise at constriction): f, v, x/χ *lach*, ɣ *negen*, h
- Nasals: m, n, ŋ *bang*
- Approximants: v *wie*, w *eeuw*, j, l, ɹ *boer*
- Trills: r/R *rood*
- Taps: .....
- Affricates (plosive–fricative sequence):

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): s, z, ʃ(ç) *sjaal*, ʒ(z) *rouge*
  - Non-sibilants (generation of noise at constriction): f, v, x/χ *lach*, ɣ *negen*, h
- Nasals: m, n, ŋ *bang*
- Approximants: v *wie*, w *eeuw*, j, l, ɹ *boer*
- Trills: r/R *rood*
- Taps: (ɾ *groot*)
- Affricates (plosive–fricative sequence): .....

# The Dutch consonant inventory by manner

- Plosives (closure plus burst): p b t d k g (*goal*)
- Fricatives (stationary noise)
  - Sibilants (directing a jet of air into the teeth): s, z, ʃ(ç) *sjaal*, ʒ(z) *rouge*
  - Non-sibilants (generation of noise at constriction): f, v, x/χ *lach*, ɣ *negen*, h
- Nasals: m, n, ŋ *bang*
- Approximants: v *wie*, w *eew*, j, l, ɹ *boer*
- Trills: r/R *rood*
- Taps: (ɾ *groot*)
- Affricates (plosive–fricative sequence): tʃ?

# The Dutch consonant inventory by articulator

- Labial (lower lip):
  - Bilabial:
  - Labiodental:
- Coronal (tongue tip or blade):
  - Alveolar:
  - Palatoalveolar:
- Dorsal (tongue body):
  - Palatal:
  - Velar:
  - Uvular:
- Glottal:

# How to work with Praat

- Recording sounds
- Viewing sounds
- Playing sounds
- Cutting and pasting sounds