Aalto University
School of Electrical
Engineering

Communication acoustics
Ch 5: Human voice

Ville Pulkki and Matti Karjalainen

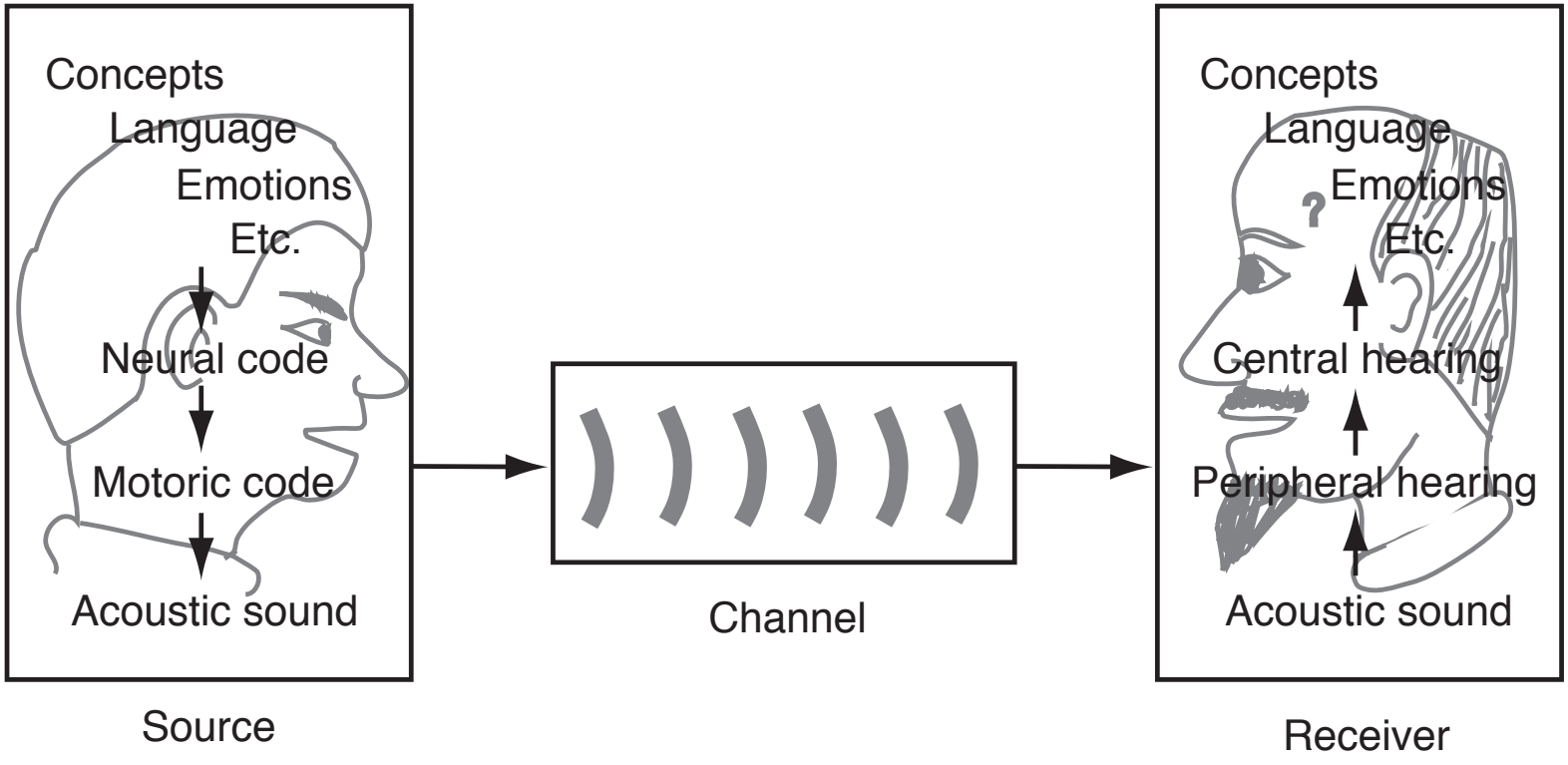
Department of Signal Processing and Acoustics
Aalto University, Finland

October 3, 2022

This chapter

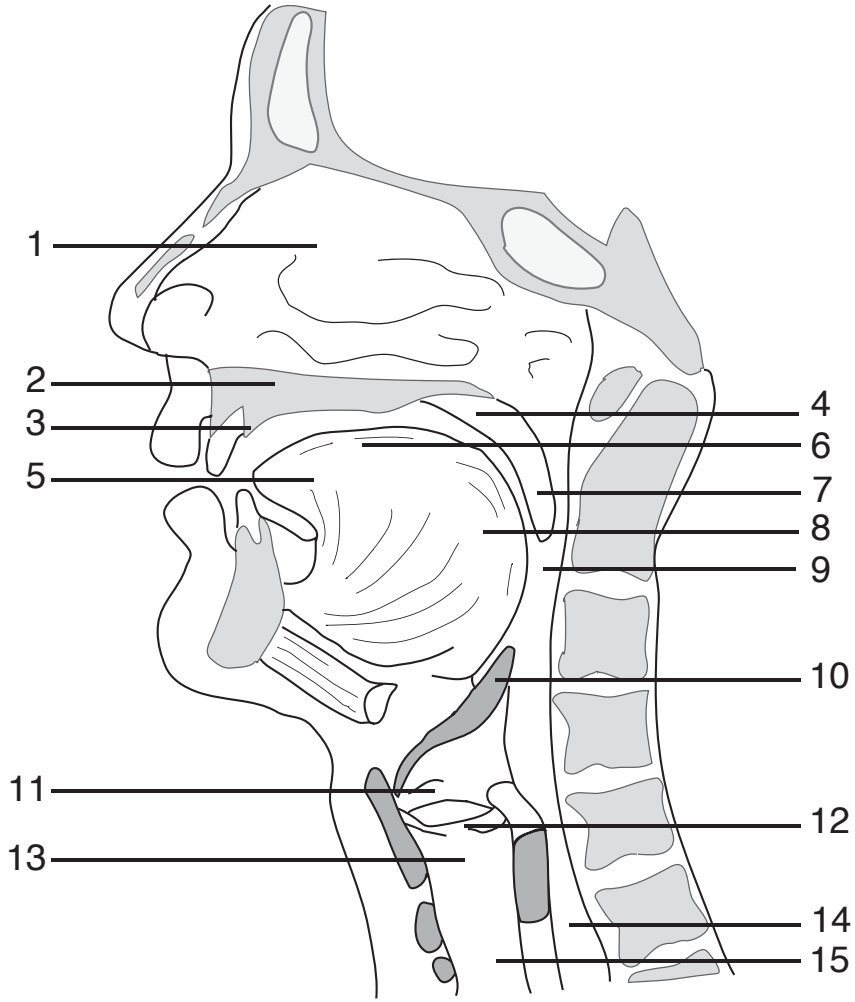
- Speech communication
- Speech production:
  - Speech production mechanism
  - Vocal folds - phonation
  - Vocal and nasal tract - articulation
  - Units and notation of speech: vowels, consonants
  - Prosody of speech
- Modeling of voice production
- Singing voice

Speech communication chain



Speech production mechanism

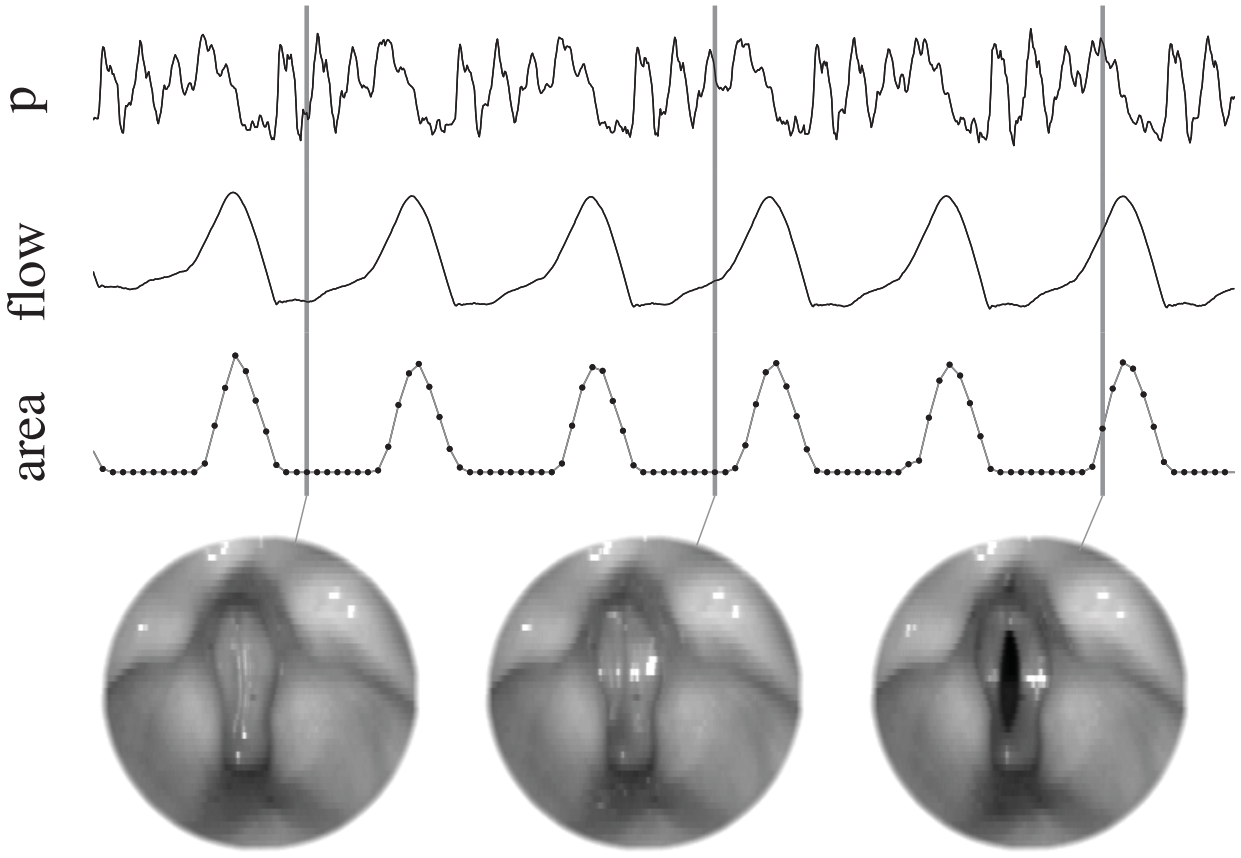
1. Nasal cavity
2. Hard palate
3. Alveolar ridge
4. Soft palate (velum)
5. Tongue tip
6. Dorsum
7. Velum
8. Radix
9. Pharynx
10. Epiglottis
11. False vocal folds
12. Vocal folds
13. Larynx
14. Esophagus
15. Trachea



Phonation and articulation

- Vocal fold – **phonation**
  - Generation and controlling of voiced sound in glottis
- Vocal tract and nasal tract – **articulation**
  - Controlling of voice features by articulation organs
- Concepts
  - Glottis (vocal fold opening)
  - Voiced / unvoiced / combined
  - Constriction
  - Formant
  - Vowel / consonant
  - Prosodic features

Vocal folds – phonation



▸ Link to vocal fold video

Vocal tract and nasal tract – articulation

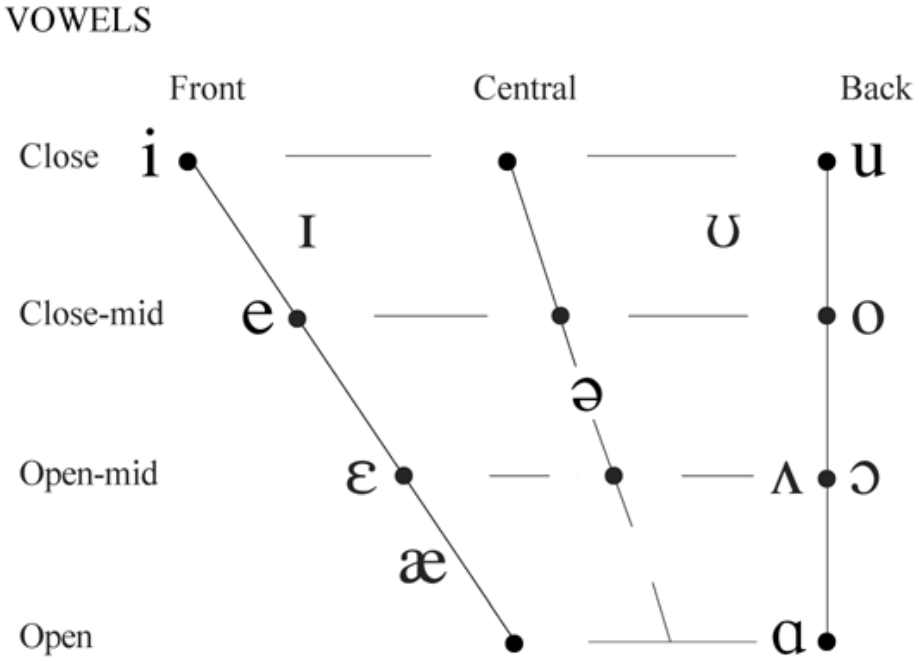
▸ Link to fMRI videos ▸ Singing in the MRI

Units and notation of speech – Phonetics

- Phonetics: study and description of spoken language
- Languages and language families
  - Indo-European, Dravidian, Uralic, ...
- Phonetic alphabet:
  - IPA (International Phonetic Alphabet)
  - Computerized: SAMPA, Speech synthesis markup language, ...
- Units of spoken language:
  - Phoneme (smallest linguistic unit), abstract unit class
  - Phone, a concrete unit of speech, including details of producing the sound
  - Allophone (alternative pronounciations of a phoneme)
  - Diphone (from mid phone via transition to the mid of next one
  - Triphone (similar combination of three successive phones)
  - Speech segment (typically subunit of a phone)

Vowels (English)

- **Front–back** position of tongue
- **Open-closed** position of tongue
- **Rounded–unrounded** opening of mouth



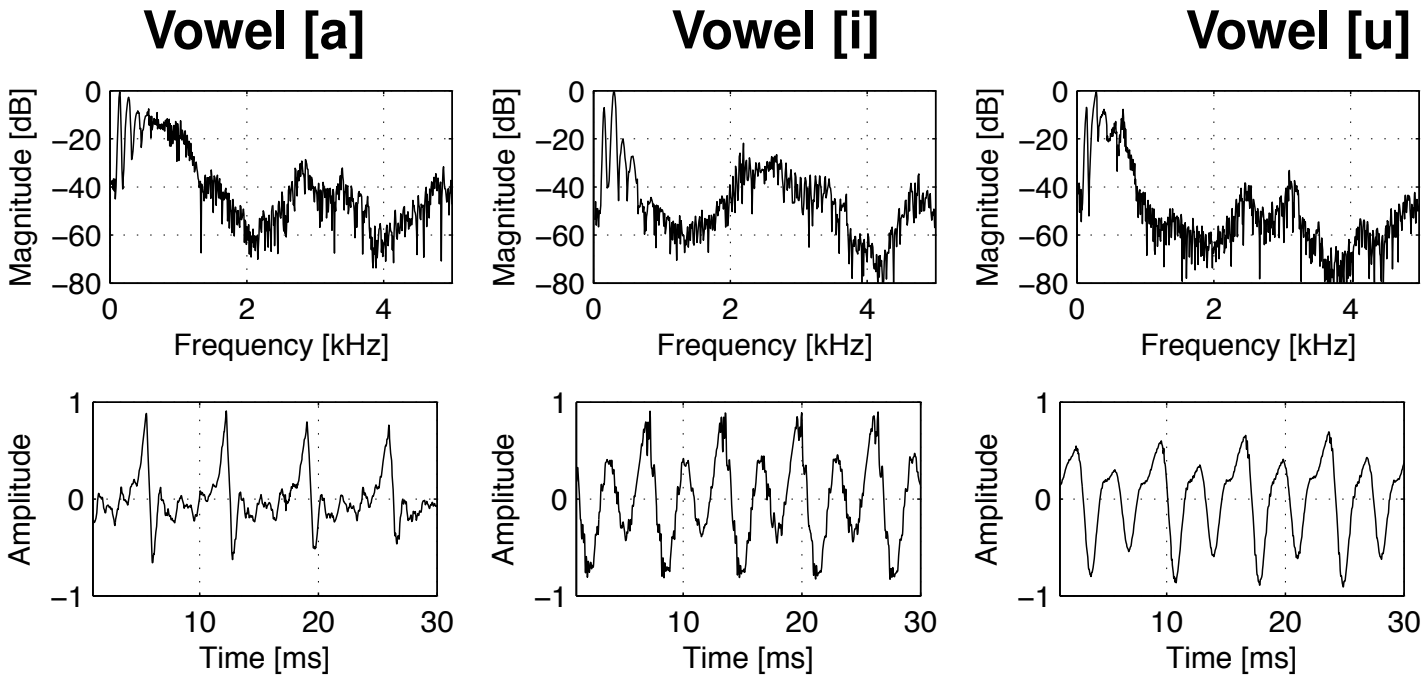
The symbols on the left of the vertical line have the mouth in an unrounded opening of the lips and those on the right in a rounded one. 2005 International Phonetic Association.

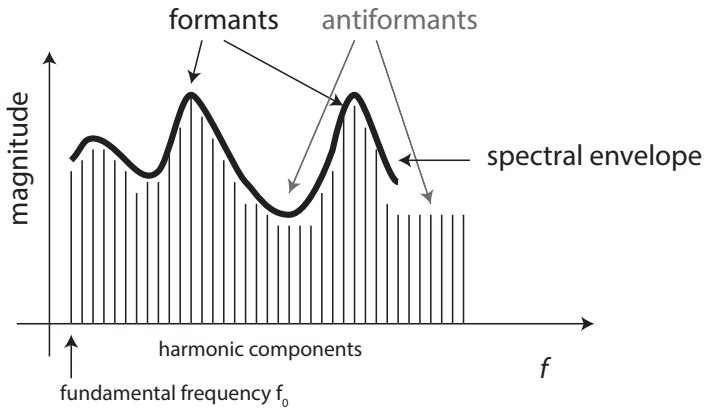
Vowels (English)

IPA symbol	examples	IPA symbol	examples
i	beat	ɪ	bit, (busy)
e	bait	ɛ	bet
æ	bat	ɑ	cot
ɔ	caught	o	coat
ʊ	book	u	boot
ʌ	but	ə	*a*bout

▸ Link to IPA page with sound examples of vowels and consonants

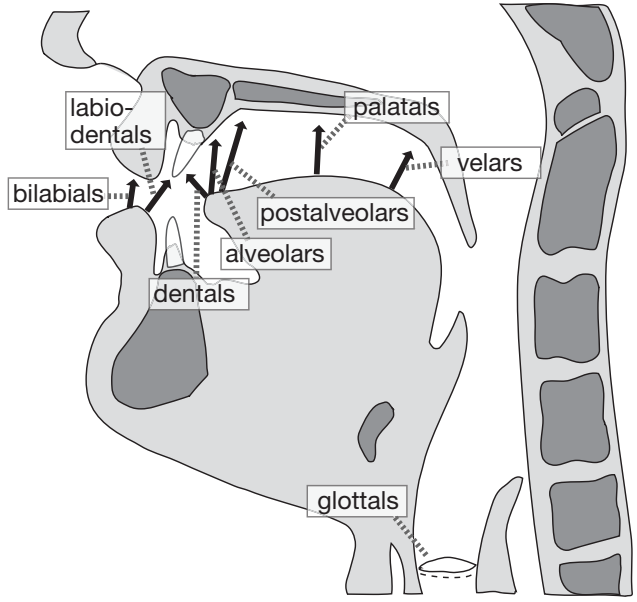
Vowel spectra





Consonants (English)

- Articulation place
  - Labial, dental, palatal, velar, laryngeal
- Articulation manner
  - Stop consonant, fricative, nasal, trill, lateral, approximant



Consonants (English)

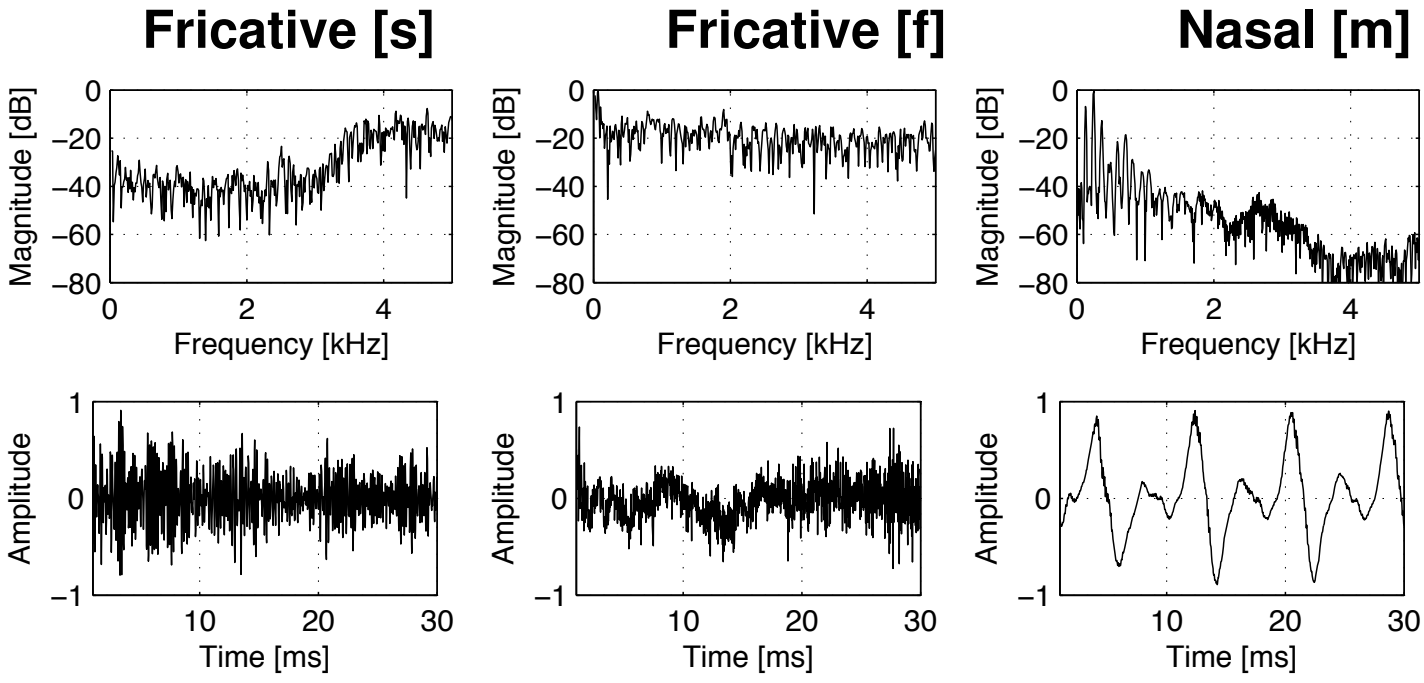
IPA symbol	example	manner	voiced	position
j	you	approximant	yes	palatal
w	wow	approximant	yes	labial-velar
ɹ	red (American dialect)	approximant	yes	alveolar
l	lull	approximant	yes	lateral
r	roar	trill	yes	alveolar
m	my	nasal	yes	bilabial
n	none	nasal	yes	alveolar
ŋ	hang	nasal	yes	velar
f	fine	fricative	no	labiodental
v	valve	fricative	yes	labiodental
θ	thigh	fricative	no	dental
ð	though	fricative	yes	dental
s	say	fricative	no	alveolar

Consonants (English)

z	zoo	fricative	yes	alveolar
ʃ	show	fricative	no	postalveolar
ʒ	measure	fricative	yes	postalveolar
h	how	fricative	no	glottal
p	pot	plosive	no	labial
b	bib	plosive	yes	labial
t	tot	plosive	no	alveolar
d	did	plosive	yes	alveolar
k	kick	plosive	no	velar
g	gig	plosive	yes	velar
tʃ	church	affricate	no	alveopalatal
dʃ	judge	affricate	yes	alveopalatal

Table: Table of common American English consonants

Consonants signals and spectra

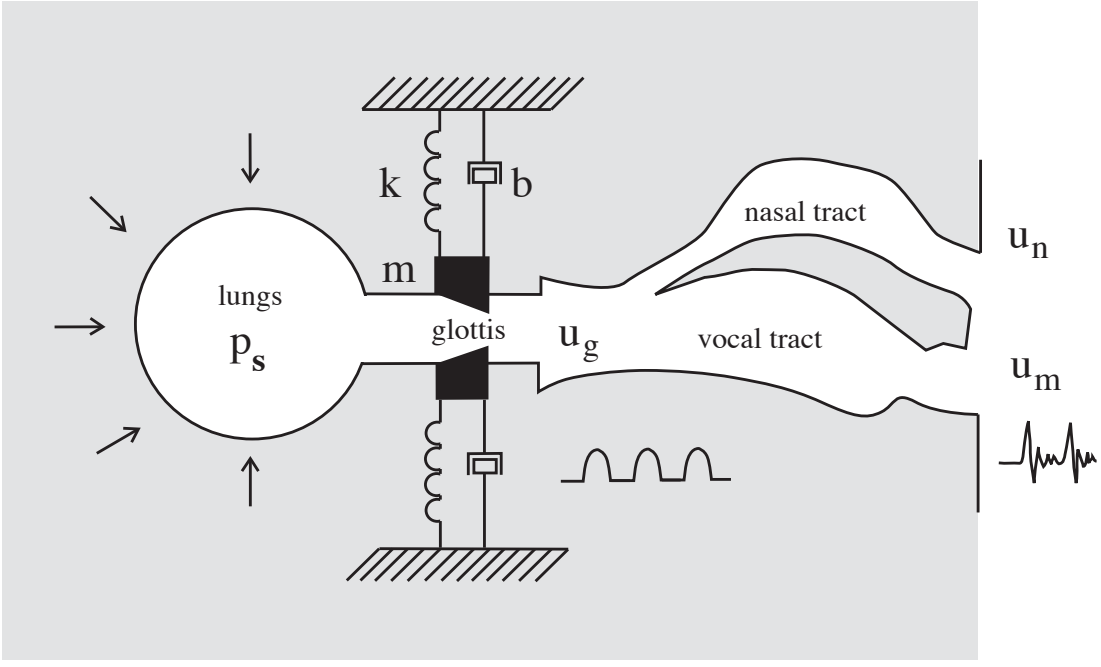


Prosody

- Intonation
  - Variations of f_0
- Stress
  - Pitch accent, dynamic accent, qualitative accent, quantitative accent
- Rhythm and timing
  - Division of words and phonemes in time

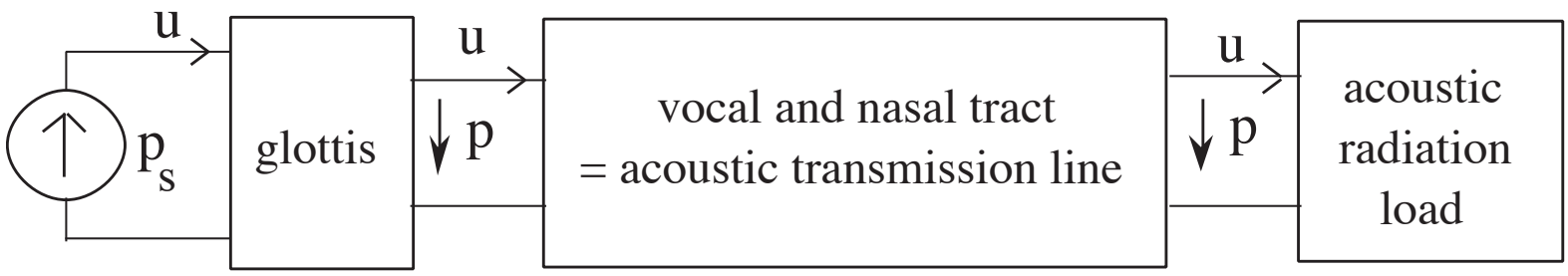
Modeling of speech production

- Simplification of the speech production mechanism, acoustic model



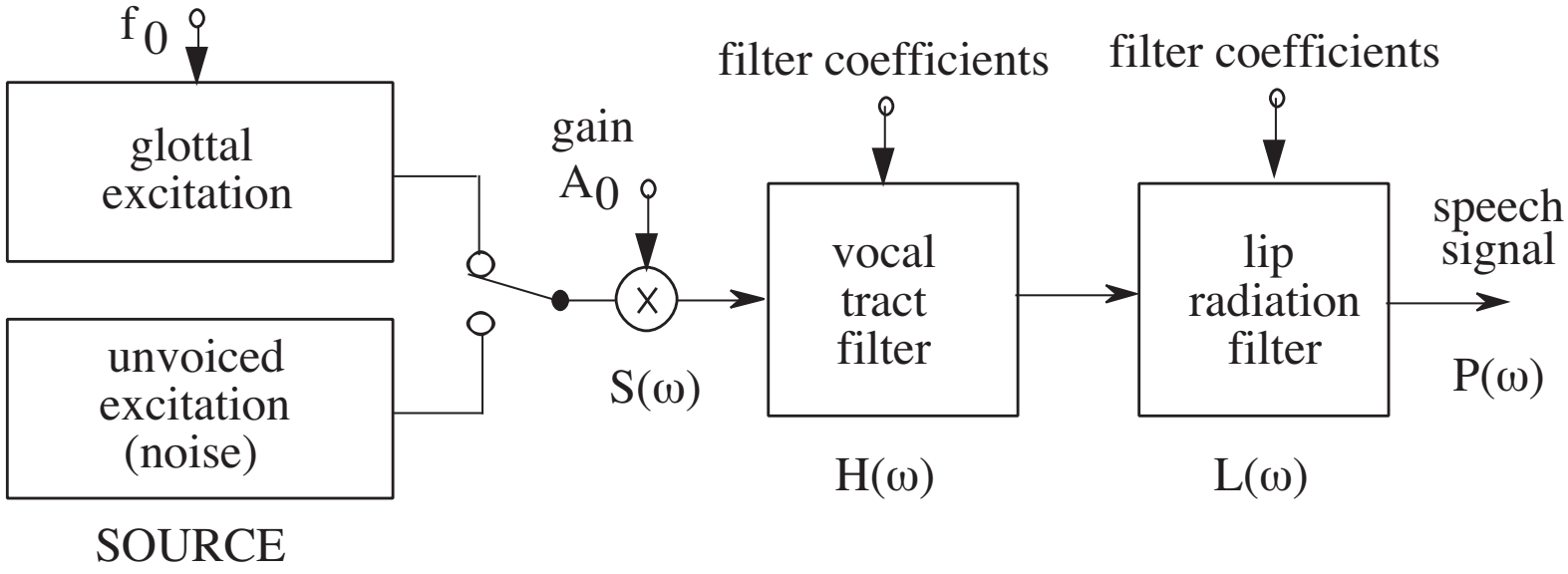
Transmission-line model of speech production

- Glottal oscillator
- Vocal tract as a transmission line
- Lip radiation as acoustic load

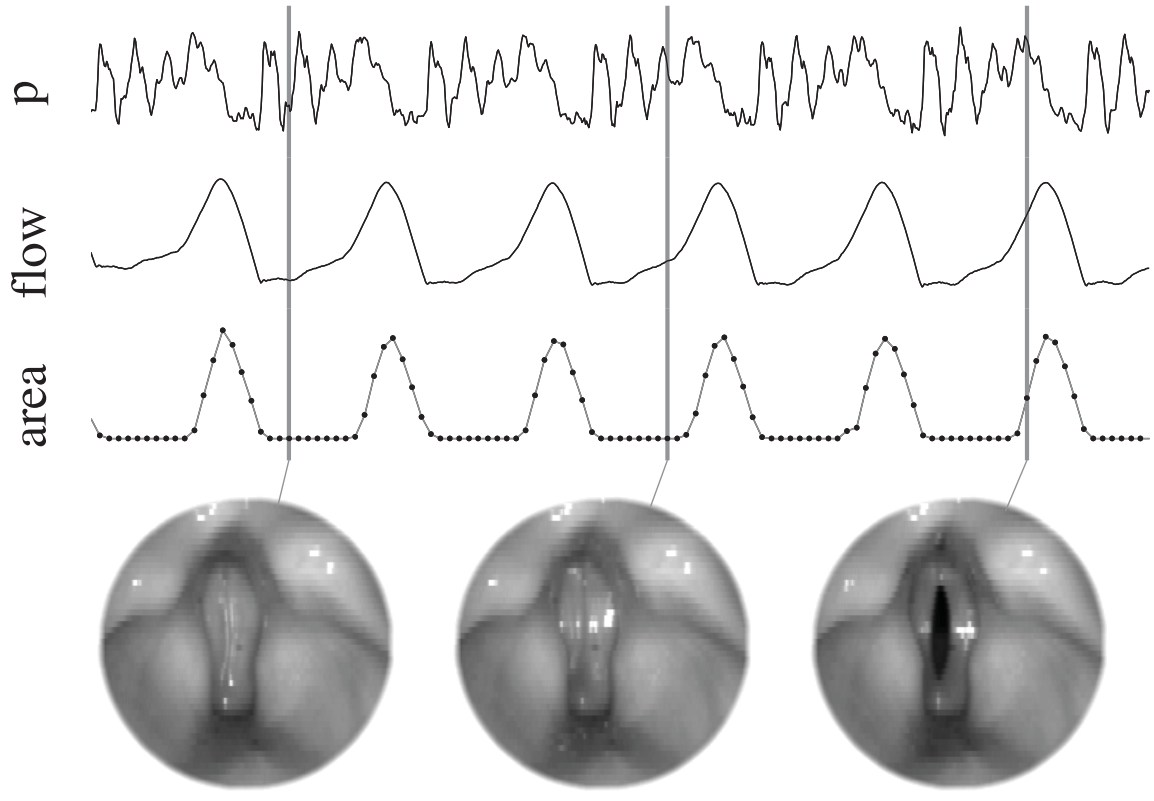


Signal model of speech production

- Source-filter model
- Excitation (source)
  - (a) voiced = quasiperiodic
  - (b) unvoiced = noiselike excitation
- Filter = response of vocal and nasal tract

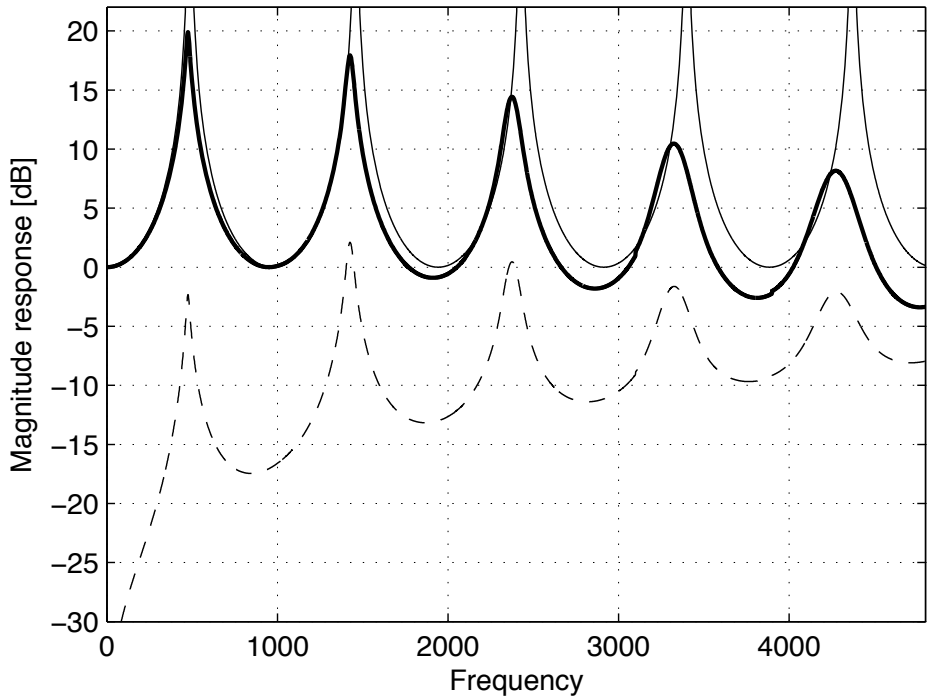


Glottal oscillation



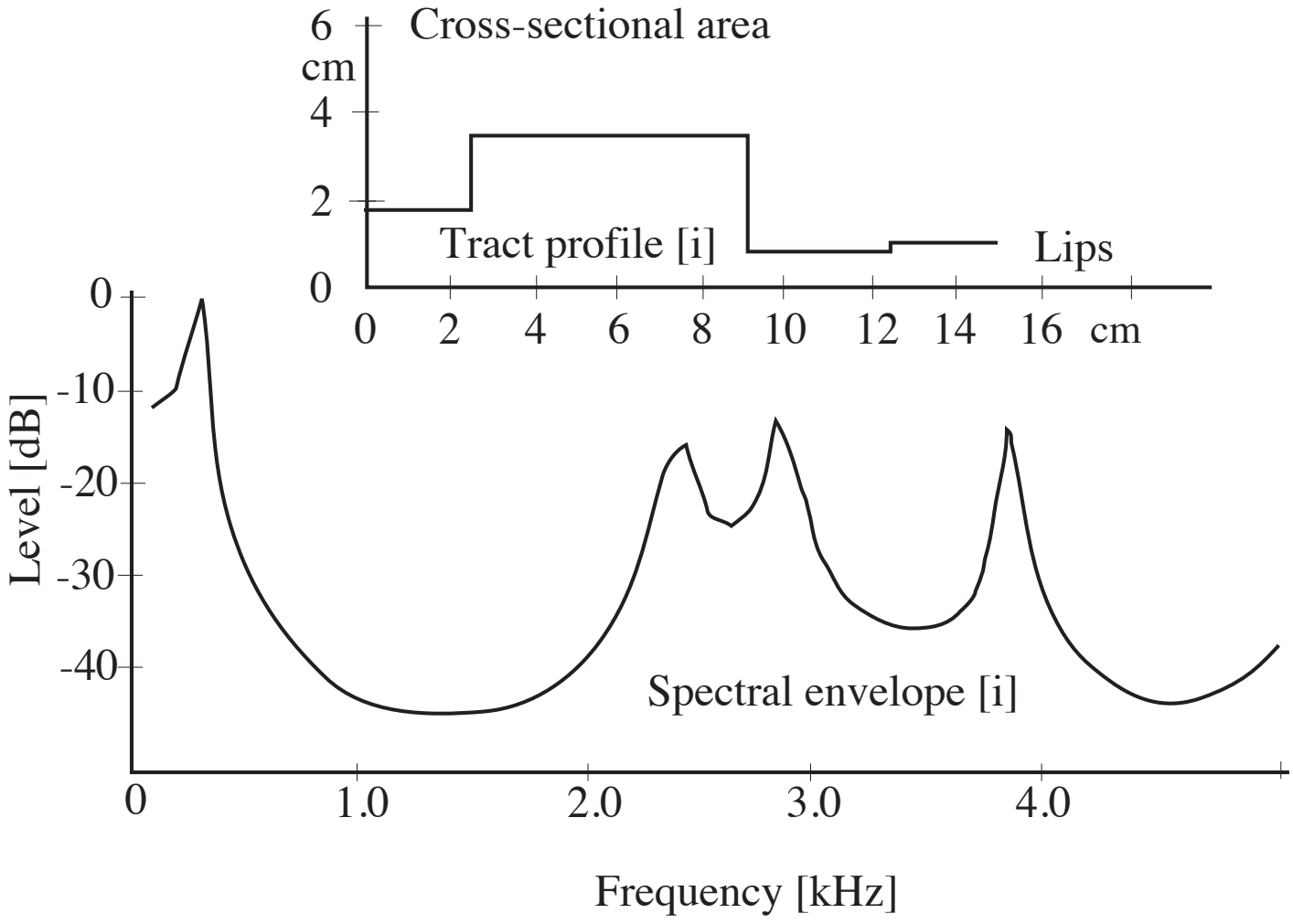
Courtesy of Hannu Pulakka, Aalto University.

Resonances of a tube

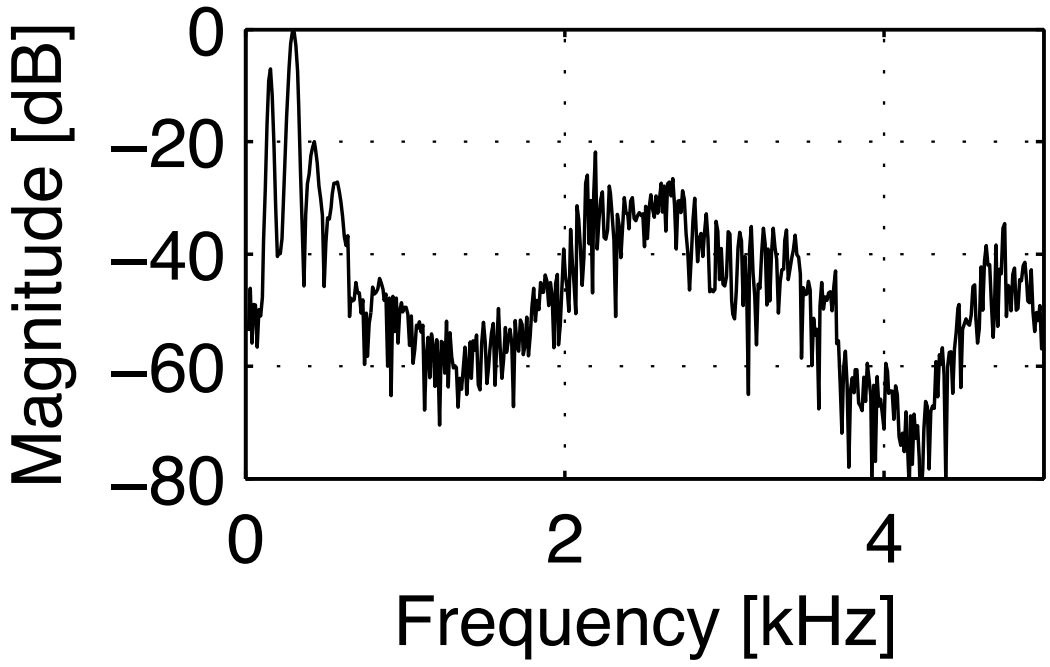


Thin line an idealized, lossless case. **Thick line** simulation with losses due to yielding walls, friction, and thermal loss. **Dashed line** shows the transfer function from the glottal velocity to pressure after lip radiation.

Vocal tract model resonances

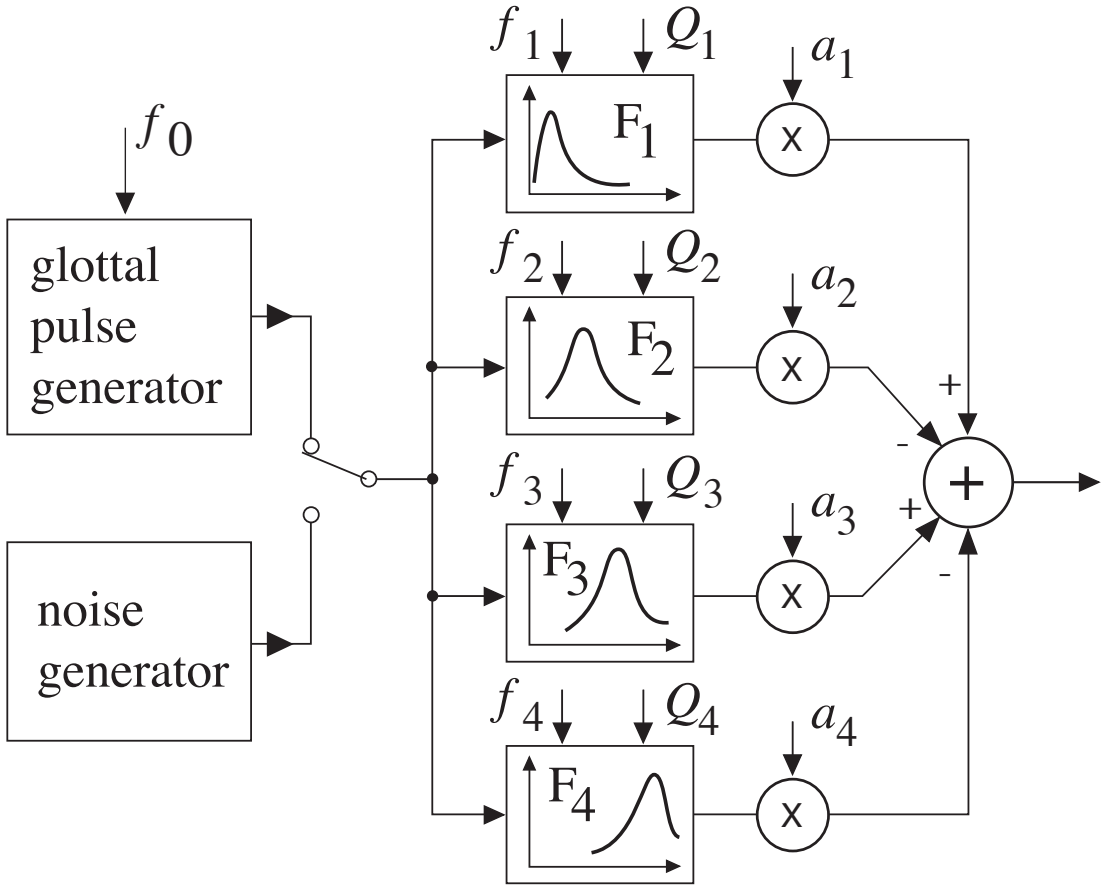


Vocal /i/ spectrum



compare with vocal tract model output one previous page

Formant synthesis model



Used to in certain speech synthesis methods

Singing voice

- Clear pitch, avoid glissandos, use stable f_0 within each note
- Spectrum is different from normal speech
- Vibrato, pitch and/or level is modulated with about 7 Hz
- Wider range of f_0 than in speech
- Classical singing style
  - Target is to be heard over a symphony orchestra without amplification
  - 'Singers formant' around 3 kHz makes voice more audible
  - In soprano singing the singer adjusts the frequency of formant(s) to match the frequency(ies) of partial(s) $\rightarrow$ high amplification

References

These slides follow corresponding chapter in: Pulkki, V. and Karjalainen, M. Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics. John Wiley & Sons, 2015, *where also a more complete list of references can be found*.