THE SPEECH

Study aid for learning of Communications Acoustics

VIHIM 000

Prof. Fülöp Augusztinovicz

BME Dept. of Networked Systems and Sefvices

fulop@hit.bme.hu

2018. október 16.,
Budapest

Introduction

- The language:
  - Communication between the individuals of the human society
  - Main tool of the human thinking

- The speech:
  - Primary manifestitation of the language
  - Most widespread form of human communication (but by far not the only one)
  - The very first, from the chronological point of view: well before writing

Suggested literature

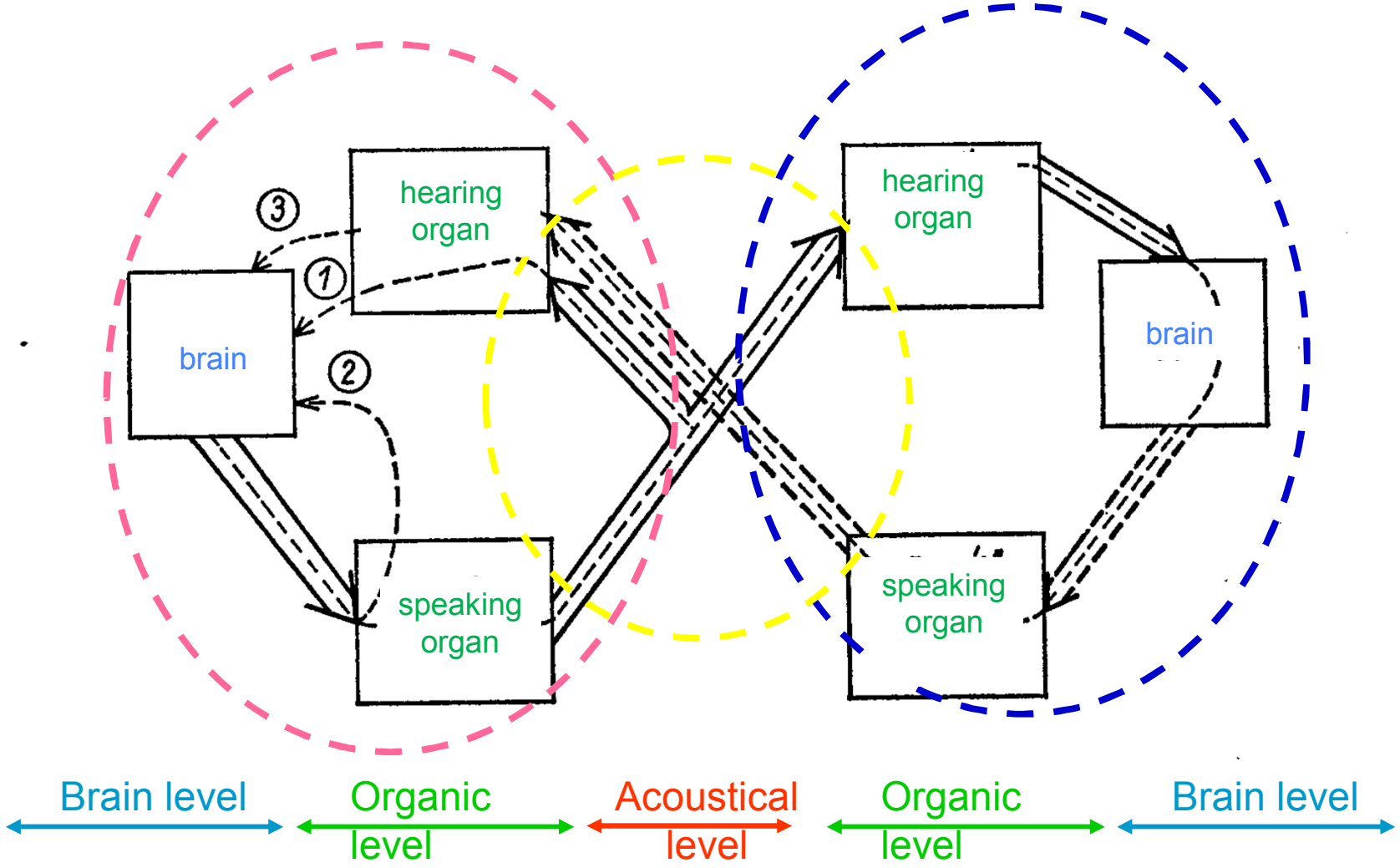
- P. Ladefoged & I. Maddieson: **The sounds of the world's languages** Blackwell Publishers, 1985

- E. Keller (Ed.), **Fundamentals of speech synthesis and speech recognition.** Basic concepts, state-of-the-art and furure challenges. John Wiley, 1994.

- Wendy J. Holmes: **Speech synthesis and recognition.** 2nd Ed. Taylor & Francis, 2001 (in form of an e-book: 2003)

- Manfred R. Schroeder, **Computer speech: Recognition, compression, synthesis.** Springer, 2004

The natural chain of speech



The structure of speech

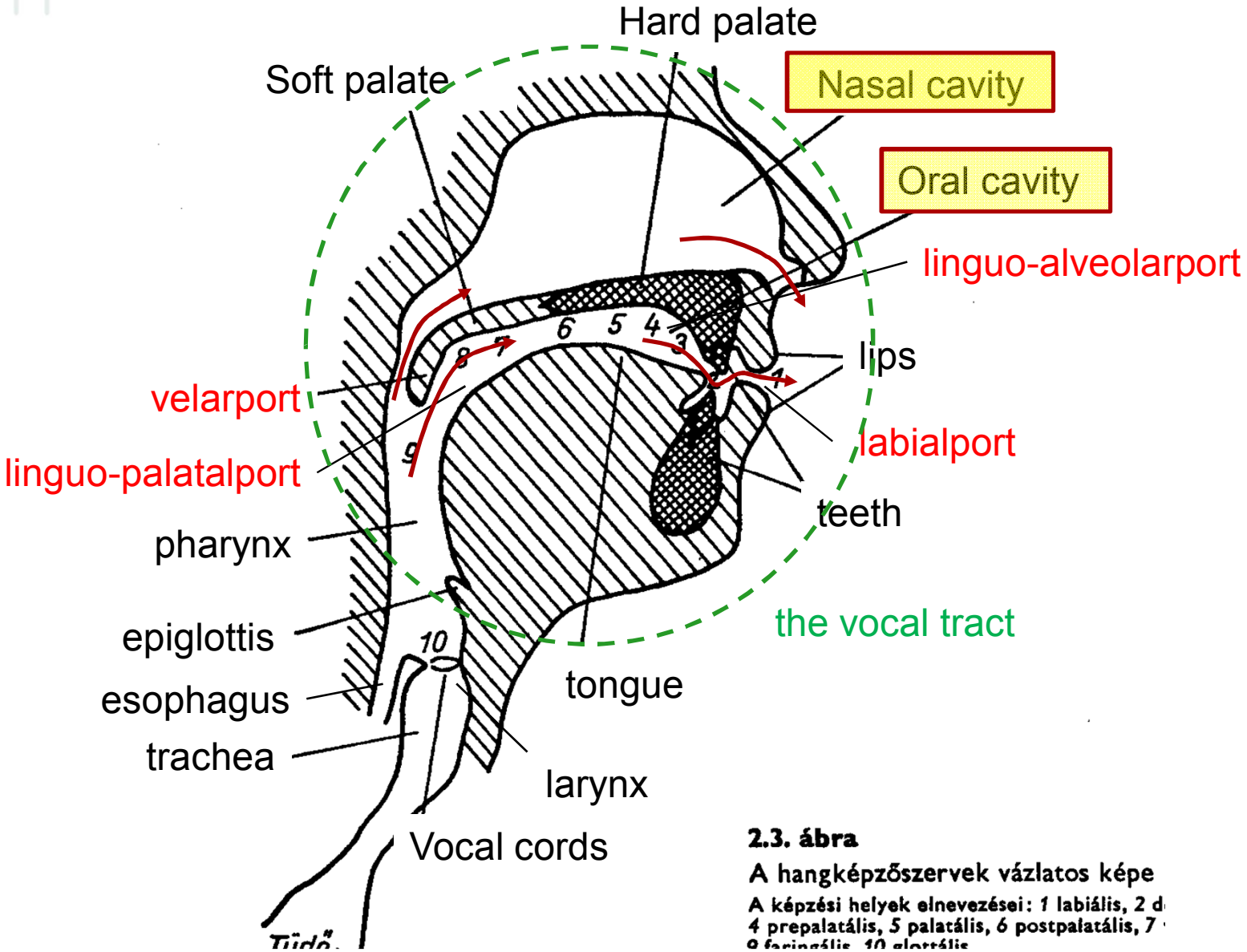
- Sentence > word > syllable > **phonema** > speech sound

  - Notion of phonemae:

The smallest contrastive linguistic unit which may bring about a change of meaning

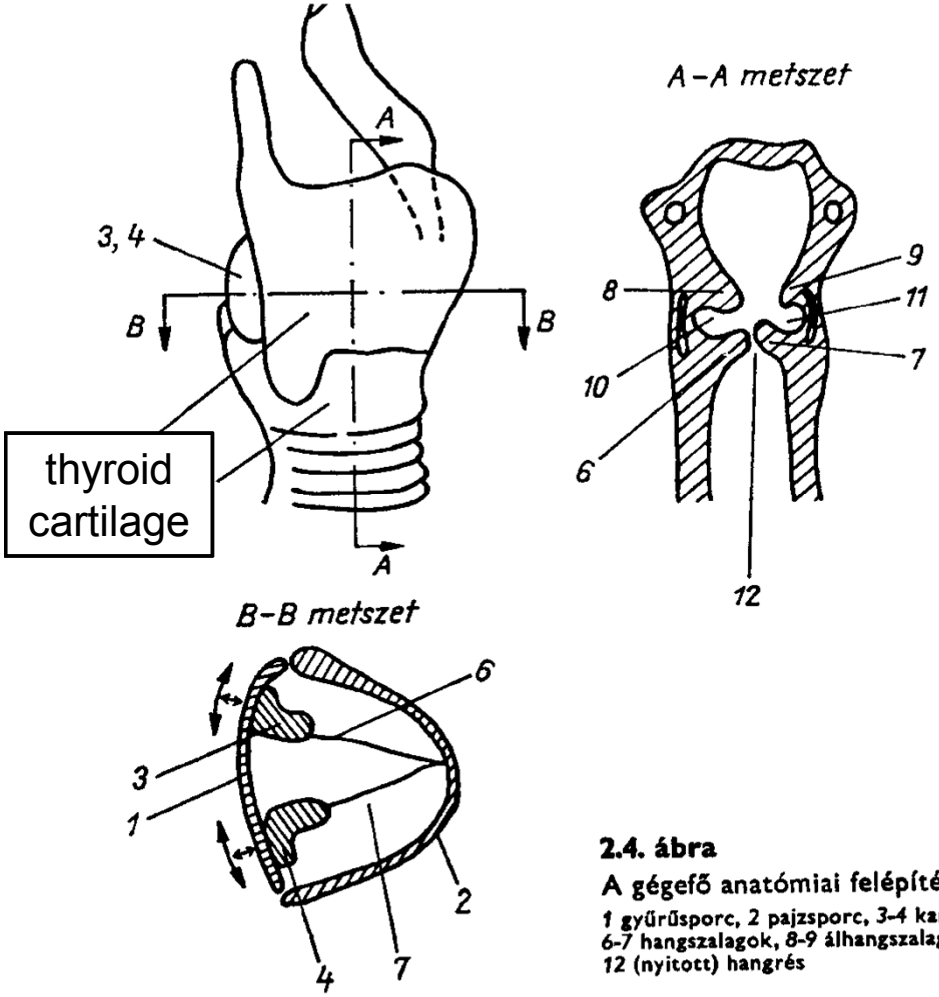
Fonéma x_i	Kulcsszó	$P(x_i)$ %	$I(x_i)$	Fonéma x_i	Kulcsszó	$P(x_i)$ %	$I(x_i)$
(ɛ)	eke	11,22	0,3541	(l)	lap	5,26	0,2234
(e:)	élet	3,26	0,1610	(t:)	mellé	0,32	0,0265
(i)	igen	4,69	0,2070	(r)	ara	3,57	0,1616
(i:)	így	0,48	0,0370	(r:)	arra	0,01	0,0013
(ø)	öröm	0,84	0,0590	(n)	nap	6,88	0,2660
(ø:)	őrült	0,94	0,0633	(n:)	annál	0,12	0,0116
(y)	üres	0,39	0,0312	(p)	anya	0,43	0,0338
(y:)	űz	0,18	0,0264	(p:)	fonnyad	0,08	0,0082
(a)	akar	10,09	0,3340	(m)	malom	4,9	0,2132
(a:)	ádám	2,98	0,1510	(m:)	zümmög	0,05	0,0055
(o)	ott	4,82	0,2109	(ts)	cápa	0,18	0,0164
(o:)	óra	0,95	0,0228		játszik		
(u)	ugat	0,96	0,0644	(ts:)	viccel	0,01	0,0013
(u:)	úr	0,26	0,0223	(dz)	brindza	0,002	0,0007
(p)	pipa	0,7	0,0501	(dz:)	edzz!	?	?
(p:)	nappal	0,05	0,0055	(dʒ)	lándzsa	?	?
(b)	baka	1,33	0,0829	(dʒ:)	briddzsel	?	?
(b:)	abba	0,3	0,0251	(tʃ)	csap	0,44	0,0344
(t)	túró	6,5	0,2563	(tʃ:)	fröccsen	?	?
(t:)	attól	0,64	0,0466	(c)	atya	0,01	0,0013
(d)	meder	2,37	0,1280	(c:)	hattyú	?	?
(d:)	addig	0,02	0,0024	(J)	agy	2,25	0,1232
(k)	eke	4,24	0,1933	(J:)	buggyan	0,001	0,0001
(k:)	akkor	0,07	0,0073	(v)	java	1,92	0,1095
(g)	egér	2,18	0,1203	(v:)	evvel	?	?
(g:)	reggel	0,03	0,0035	(s)	szó	1,56	0,0936
(f)	fa	0,77	0,0541	(s:)	lasszó	0,09	0,0091
(f:)	puffan	0,01	0,0013	(z)	záp	3,19	0,1586
(j)	java folyó	1,68	0,0990	(z:)	bízz!	0,03	0,0035
				(ʃ)	sava	3,12	0,1561
(j:)	bajjal	0,001	0,0003	(ʃ:)	kossal	0,04	0,0045
(h)	hó	2,54	0,1346	(ʒ)	zsák	0,02	0,0024
(h:)	ahhoz	0,002	0,0005	(ʒ:)	rúzzsal	?	?

The speaking organ / 1



The speaking organ / 2

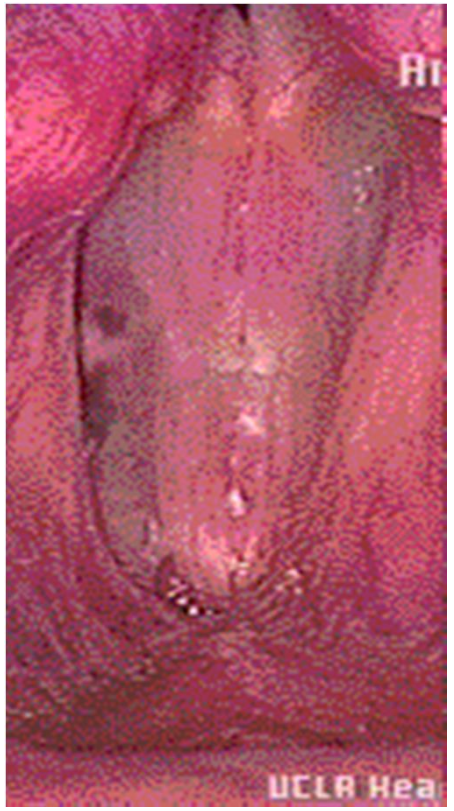
- Larynx and the vocal cords



2.4. ábra

A gégefő anatómiai felépítése

1 gyűrűsporc, 2 pajzsporc, 3-4 kannaporcok, *5* gégefedő porc, 6-7 hangszalagok, 8-9 álhangszalagok, *10-11* Morgani-féle üregek, *12* (nyitott) hangrés



The human singing

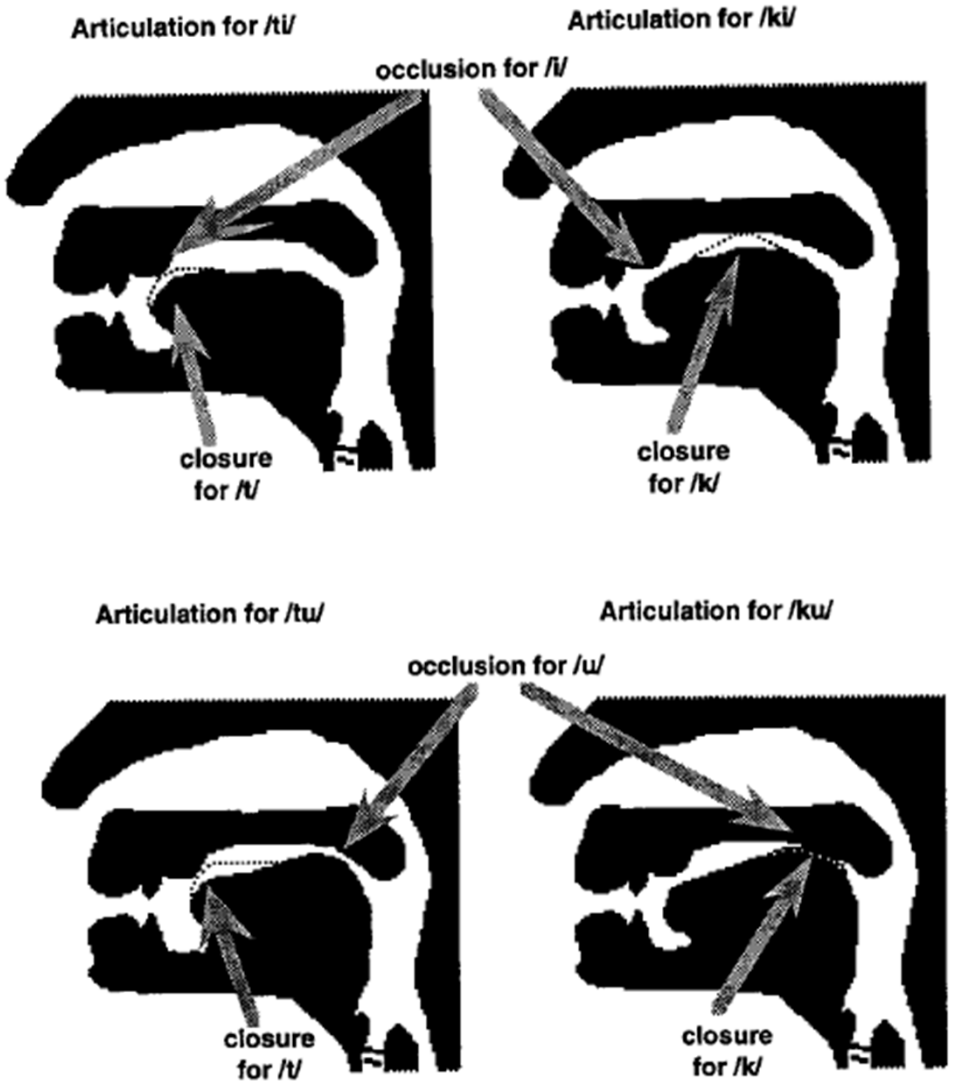
- Laryngoscope:

- http://mentalfloss.com/article/56045/listen-quartet-sing-while-you-watch-close-their-vocal-cords

The speech production process / 1

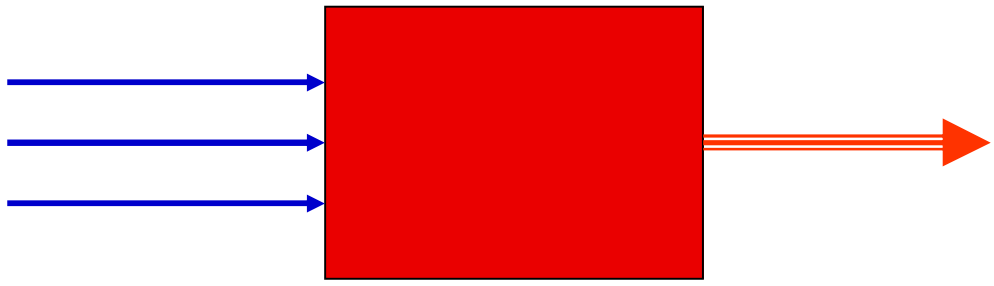
- release of air from the lung
- the air at the larynx either
  - passes through: **unvoiced** sound (consonants like ***s***, ***t*** or ***f***)
  - Is set into vibration by the vocal cords: **voiced**
    (vowels or voiced consonants like ***m*** or ***z***)
- operation of ports (or valves)->production of various sounds:
  - if velar port is open: nasal consonants or vowels
    (***m***, ***n***, of fr***a***nc)
  - if linguo-palatal port is closed: ***k***, ***g*** or ***ng***
  - in the linguo-alveolar port: mostly consonants
    e.g. explosives ***(t, d)*** and fricatives ***(s, z, ʃ, ʒ)***
  - in labial port: ***p, b, t, v***

Example: articulation of voiceless plosives



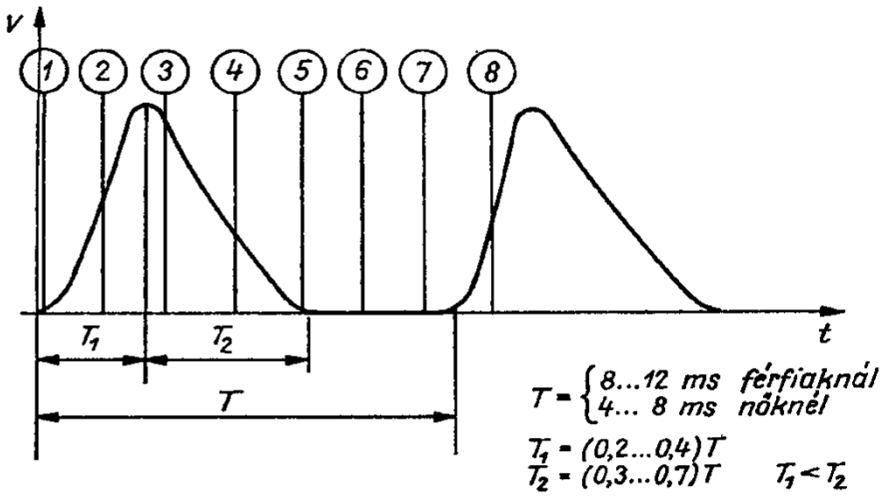
The speech production process / 2

- The transmission system
  - system of resonators (aka vocal tract, articulation channel)
- Excitations
  - voicing
  - turbulence
  - shock wave



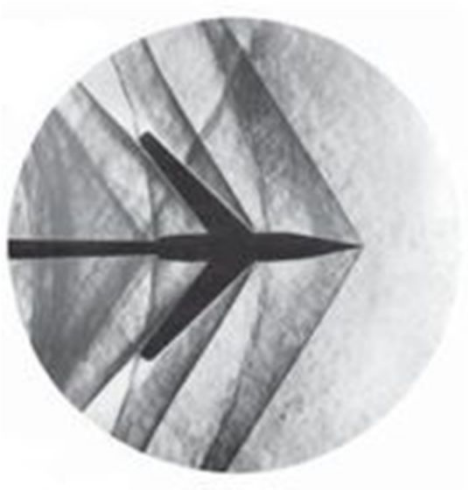
Excitation mechanisms

voice



turbulence

shock wave



Classification of speech sounds

- Vowels (V) **i, é, ü, ö, e á, a, o, u**
- Consonants (C)
  - stops
    - nasals **m, n, ny, ng**
    - plosives
      - voiced **b, d, g, gy**
      - voiceless **p, t, k, ty**
    - rhotic / trill **r**
  - lateral fricatives **l, j**
  - fricatives
    - voiced **v, z, zs**
    - voiceless **f, s, sz, h**
  - stop-fricatives (affricate)
    - voiced **dz, dzs**
    - voiceless **c, cs**

The international phonetic alphabet

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		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.

https://en.wikipedia.org/wiki/International_Phonetic_Alphabet
http://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/

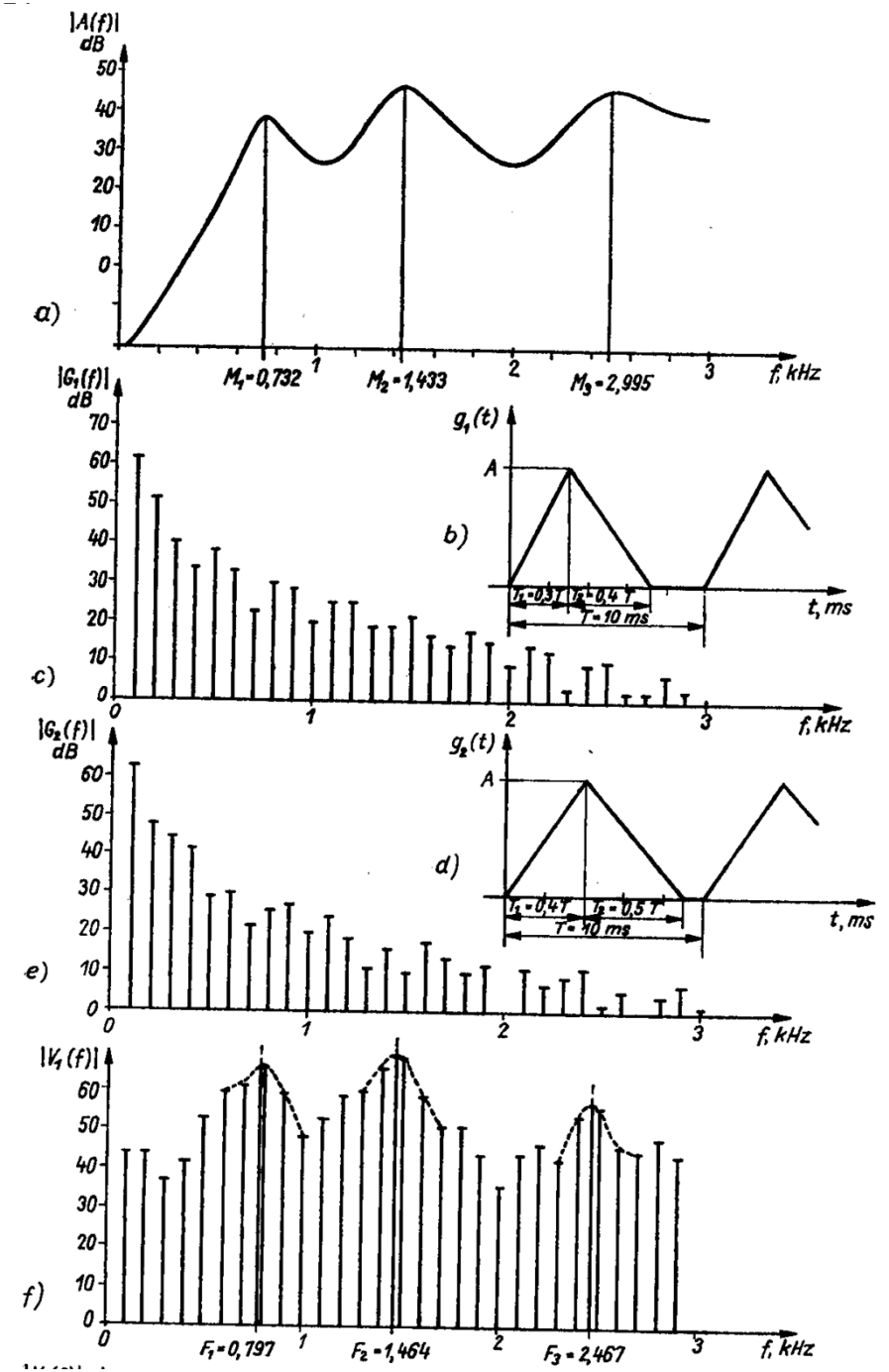
Characteristics of vowels

FRF of the vocal tract

line spectrum of the voice, female

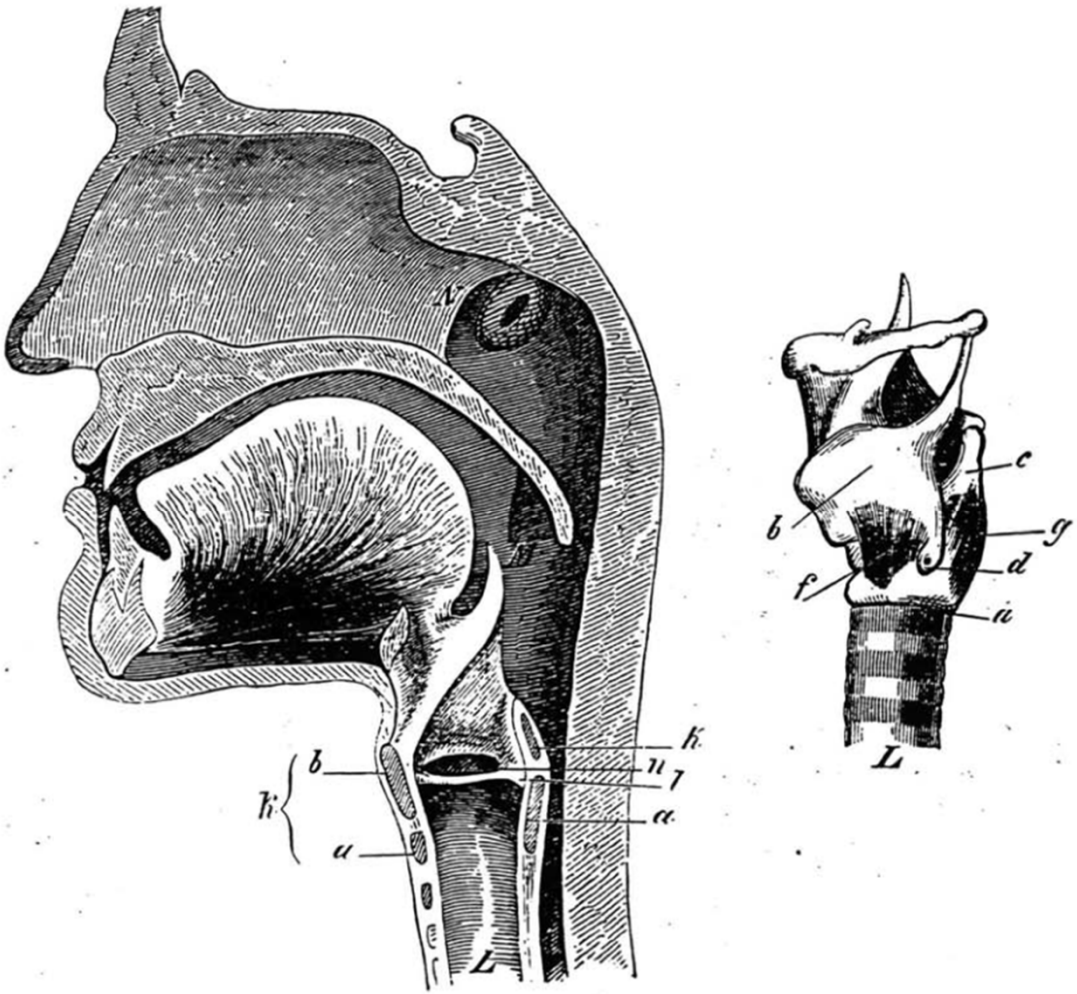
line spectrum of the voice, male

final spectrum, modified by the transmission of the vocal tract

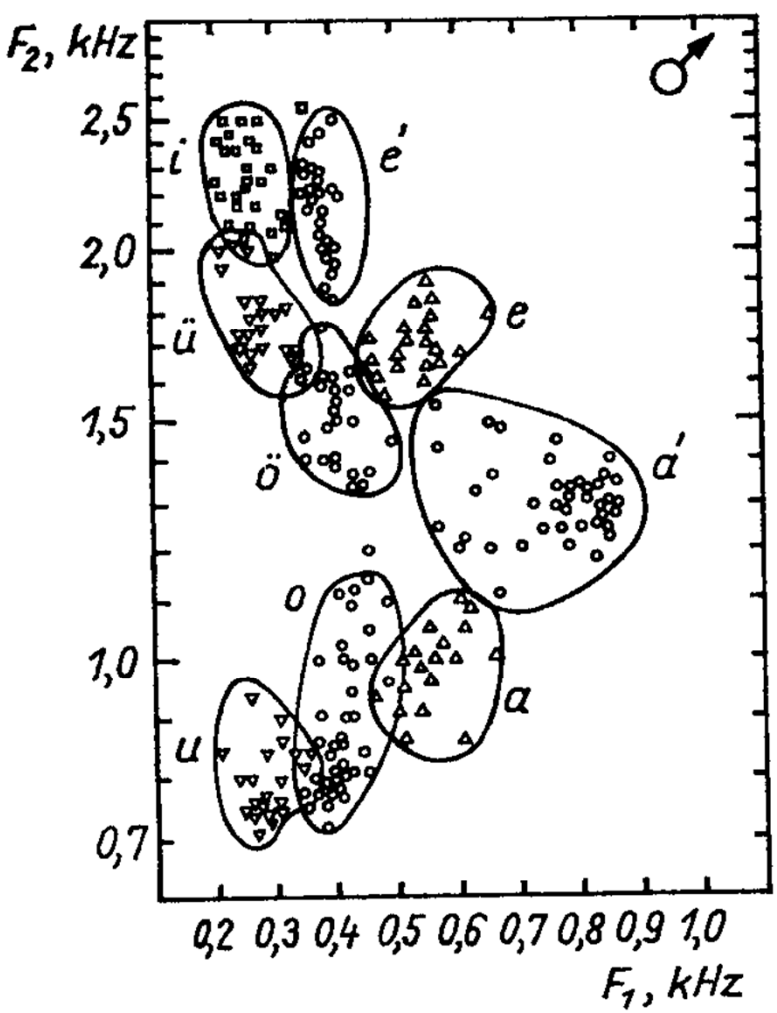


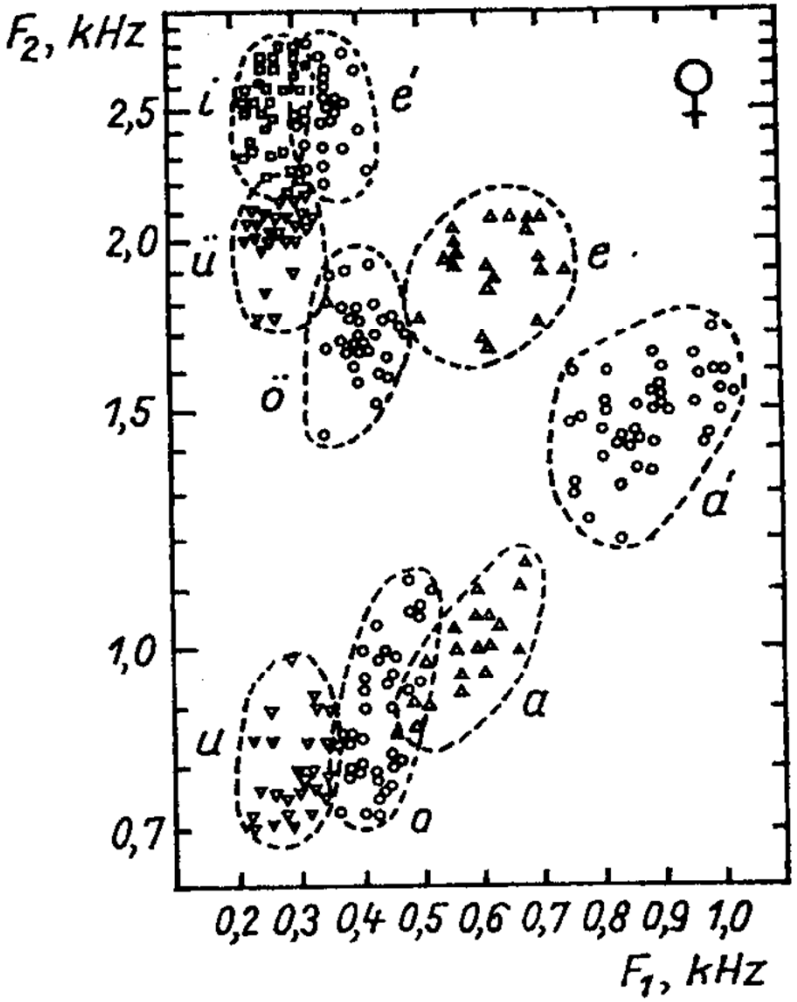
The resonant system

- The **nasal** and the **oral** cavities



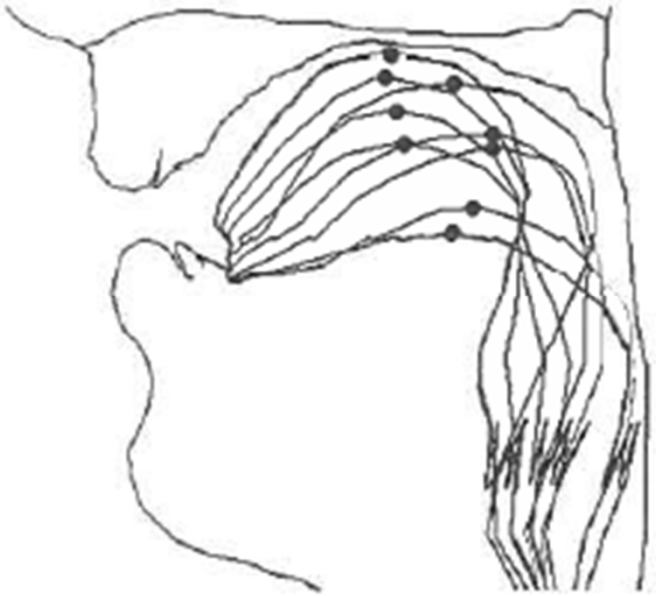
Formants of vowels

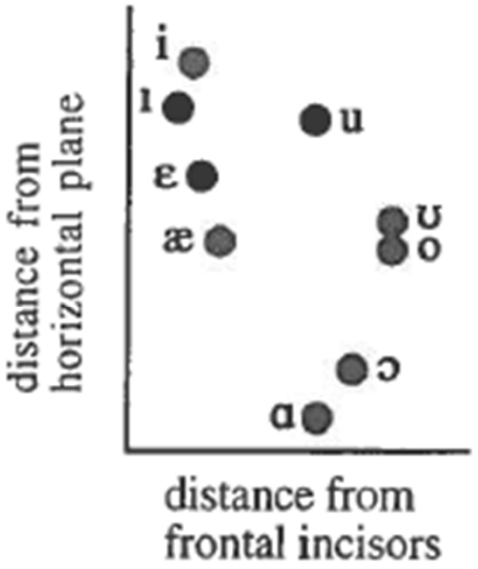




The oral cavity as a tuned resonator

- Tuning element: the tongue



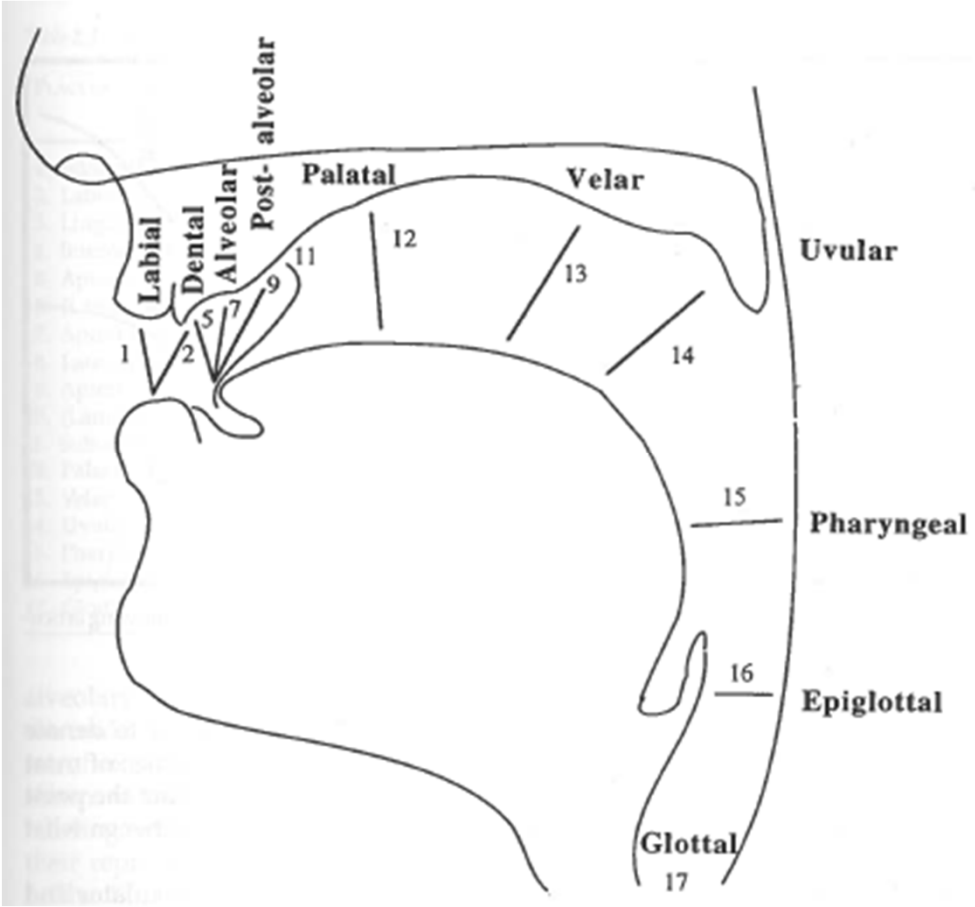


Characteristics of consonants

- Fricatives
  - turbulent excitation > resonance of the vocal tract
- Plosives
  - cannot be held for long time
  - no structure of formants
  - shock wave > time out of vocal cords > stabilised vowel

Articulation places of consonants

Palatal	Velar	Uvular	Pharyngeal	Glottal
c ɟ	k g	q G		ʔ
ɲ	ŋ	N		
		R		



Daabase of Hungarian speech sounds

http://magyarbeszed.tmit.bme.hu/index.php?p=interaktiv

Music and singing: form of joy

https://www.youtube.com/watch?v=5hksKXZQ3us

https://www.youtube.com/watch?v=Tz7RQ-C8muk

THE SPEECH INTELLIGIBILITY

Influencing factors of intelligibility

1. articulation capability of the speaker
2. listening capability of the listener
3. characteristics of the sound field (room acoustical parameters)
4. level of background noise

Parameter 3 and 4 are expressed in form of ratios:

3. direct to reverberant ratio

and

4. signal to noise ratio

Measurement of intelligibility

- Subjective tests
  - test with meaningless syllables
    - invoking many test persons
    - result is the ratio of correctly understood syllables to the total number of syllables, in %
  - test with full sentences
- Objective tests
  - articulation index (**AI**): only background noise is considered
  - Speech Intelligibility Index (SII): reverberation is also taken into account
  - Speech Transmission Index (**STI**): subjective quantity is approached by objective measurements (international standard)
    - the speech consists of components of various frequencies, which are amplitude modulated by a low frequency envelope
    - in order to remain intelligible, this modulation should be retained
    - the transmission of this modulation without distortion is deteriorated both by multiple-path transmission (reverberation) and by background noise

Effect of noise: Articulation index (AI)

$$AI = \sum g \cdot \Delta L, \qquad (2.4)$$

ahol g egy az oktávsávtól függő sulyozó tényező és ΔL a hallgató helyén mért beszéd-hangnyomásszint csucsok és a beszédzavaró zaj oktávsávban mért hangnyomásszintjei között mért különbség a 250, 500, 1000, 2000 és 400 Hz középfrekvenciáju oktávsávokban. A g sulyzó értékek:

$f_{közép}$	250	500	1000	2000	4000
g	0,0018	0,0050	0,0075	0,0107	0,0083

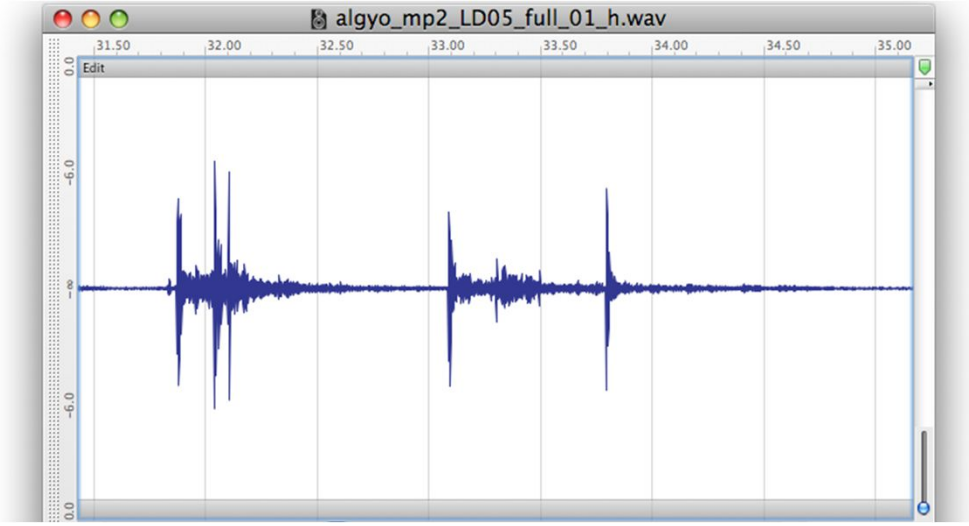
A beszéd érthetősége és az AI artikulációs index közötti összefüggés a következő:

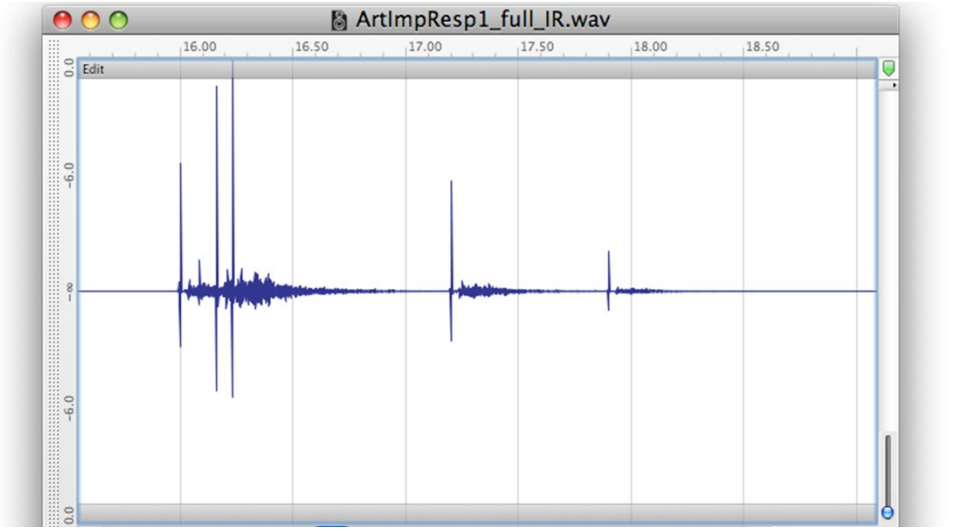
AI	Beszédérthetőség
0,1	igen rossz
0,3	nem megfelelő
0,3 ÷ 0,5	megfelelő
0,5 ÷ 0,7	jó
> 0,7	igen jó

Replacement rule of thumb: the signal to ratio should be as high as possible, but min. 5-10 dBA

Effect of reverberation on the intelligibility

- Original sound sample
- In-situ recording
- Measured impulse response
- Artificially reverberated sound





Arfiticial reverberation / 1

Cave theatre in the
Fertőrákos quarrier

Arfiticial reverberation / 2

Underground concert hall in the
Baradla cave in NW-Hungary

Arfiticial reverberation / 3

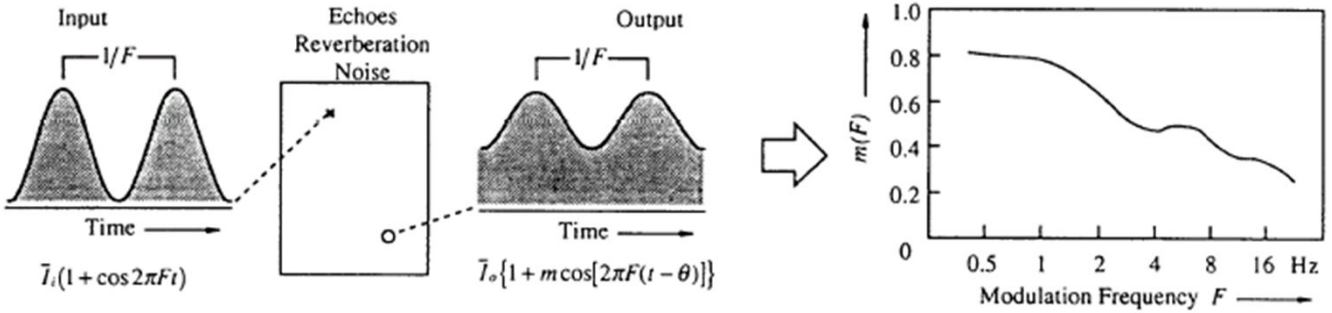
Esztergom
Basilica

Consequence: difficulties with PA systems

- Very difficult to develop/install good quality PA systems in reverberant spaces
- Example: the aula of this university (building K)

Concept of the calculation of STI

- Test signal: modulated band-limited noise
  - nowadays: by means of impulse reponse measurement



$$SNR_{kf} = 10\log_{10}\frac{m_{kf}}{1-m_{kf}}$$

- Calculation of apparent S2N ratio:
- Frequency weighting:

$$STI = \sum_{k=1}^{7} W_k MTI_k$$

w_k	0.13	0.14	0.11	0.12	0.19	0.17	0.14

Evaluation of STI values

STI value	Quality according to IEC 60268-16	Intelligibility of syllables in %	Intelligibility of words in %	Intelligibility of sentences in %
0 – 0.3	bad	0 – 34	0 – 67	0 – 89
0.3 – 0.45	poor	34 – 48	67 – 78	89 – 92
0.45 – 0.6	fair	48 – 67	78 – 87	92 – 95
0.6 – 0.75	good	67 – 90	87 – 94	95 – 96
0.75 – 1	excellent	90 – 96	94 – 96	96 – 100

Many figures of merit of intelligibility can be compared by using the **CIS** (Combined Intelligibility Scale) value:

$$CIS = 1 + \lg(STI).$$

Characteristics of the STI

- It depends on
  - signal to noise ratio
  - reverberation
  - bandwidth
  - nonlinear distortions
  - etc.

Sávszélesség	**CIS**
350-3000 Hz	0,62
350-5700 Hz	0,77
180-3000 Hz (nagyjából ez felel meg a végpont elektroakusztikai elemei sávszélességének)	0,75
180-5700 Hz	0,86
85-11000 Hz	1

Relationship of measurable SI descriptors