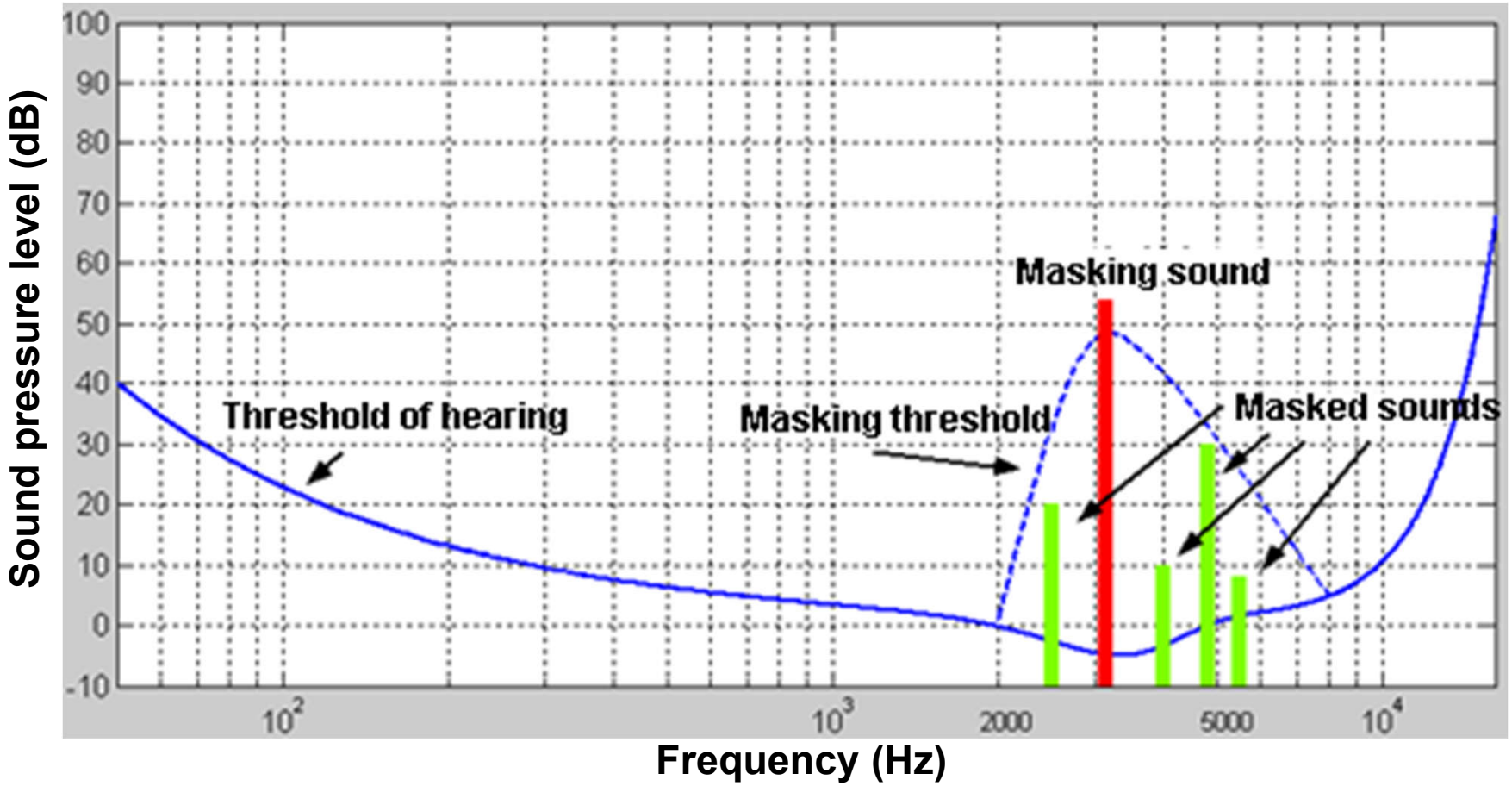
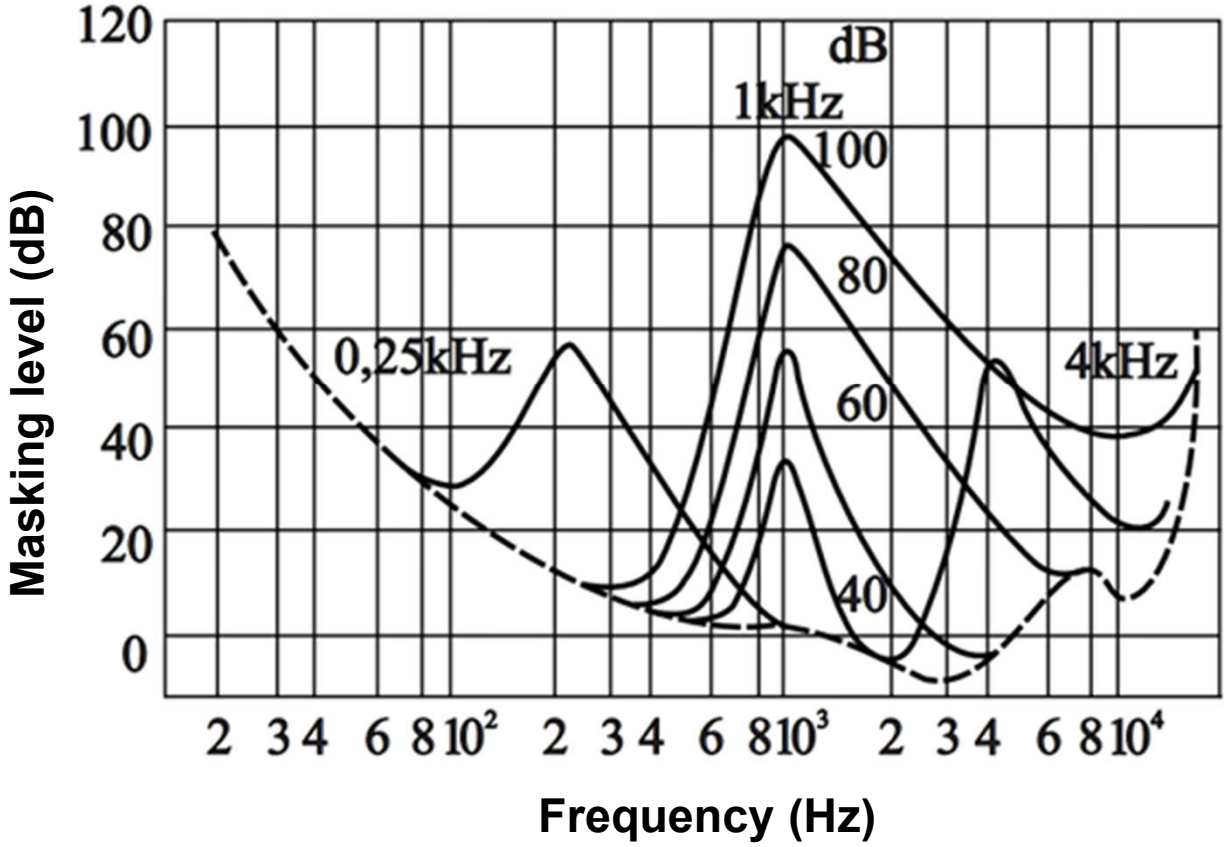
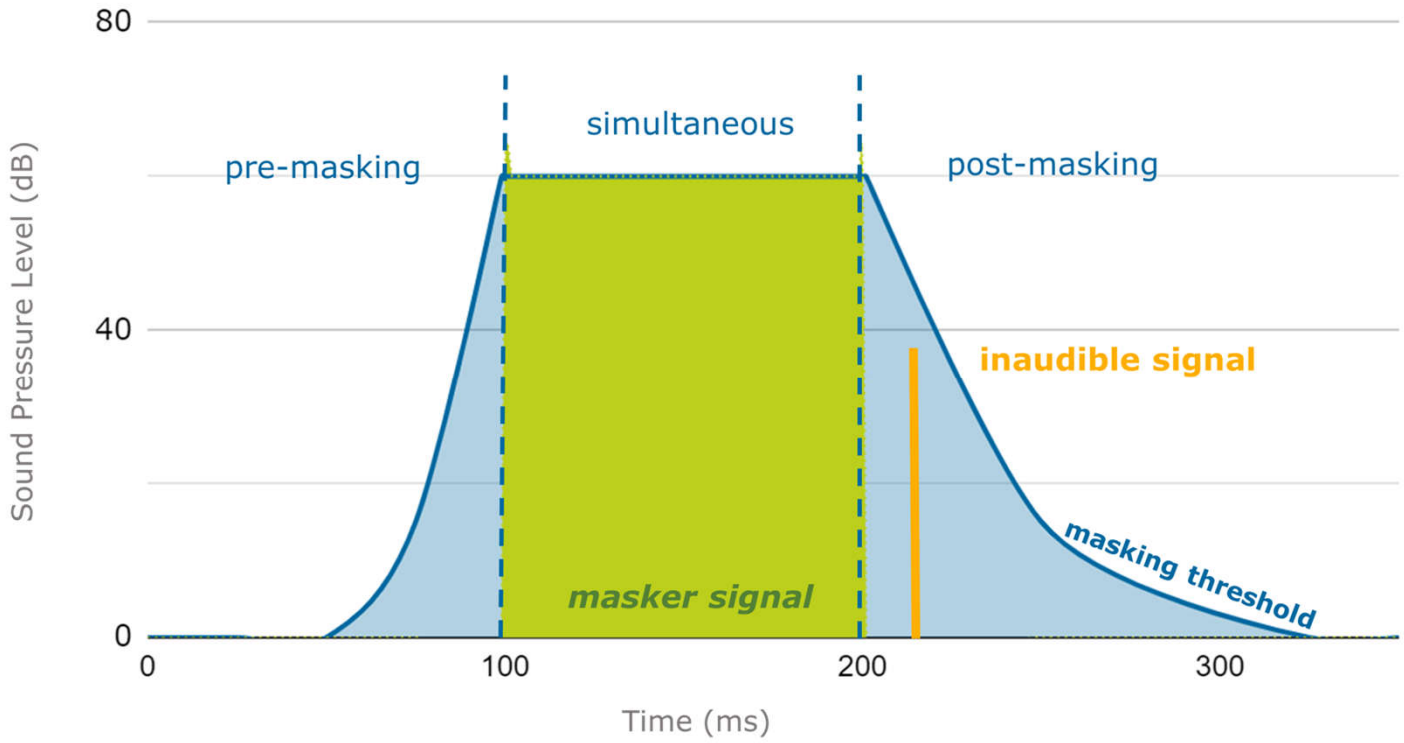
Frequency domain level masking



Level-dependent masking of narrowband noise



Time domain masking



Parameters influencing masking phenomena

- Sensing person(s)
- The test signal
  - Type
  - Level
  - Duration
- The impact of the environment

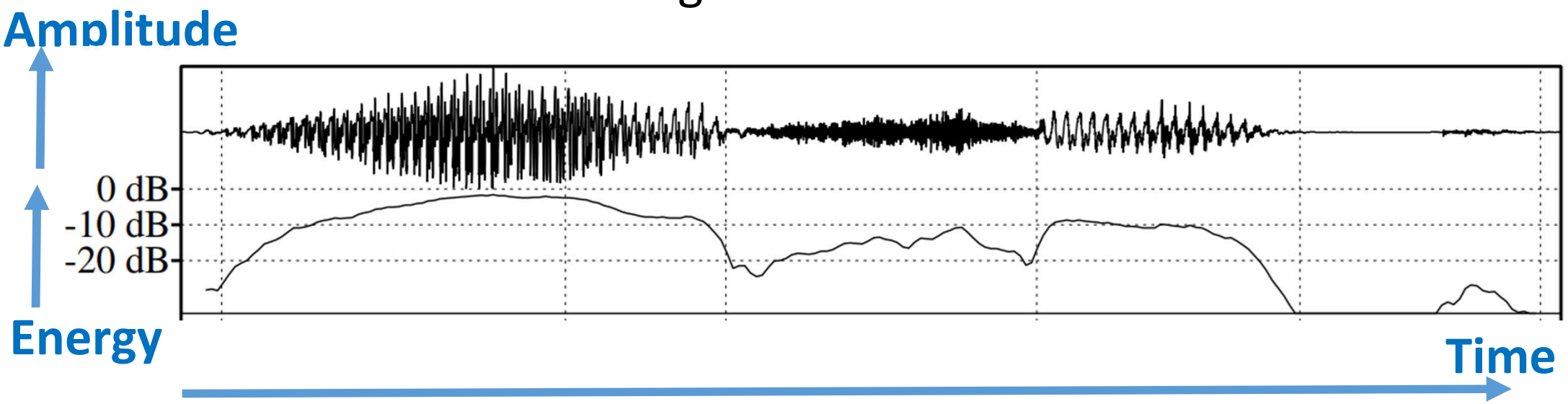
Human modality processing capacity

- Hearing ~ 50 bit/s (speech coding 1kbit/s-64kbit/s)
- Vision ~ 6-10 bits/s
- Probing ~ 1-3 bits/s
- All others ~0.5 bit/s

4. The linguistic structure of speech

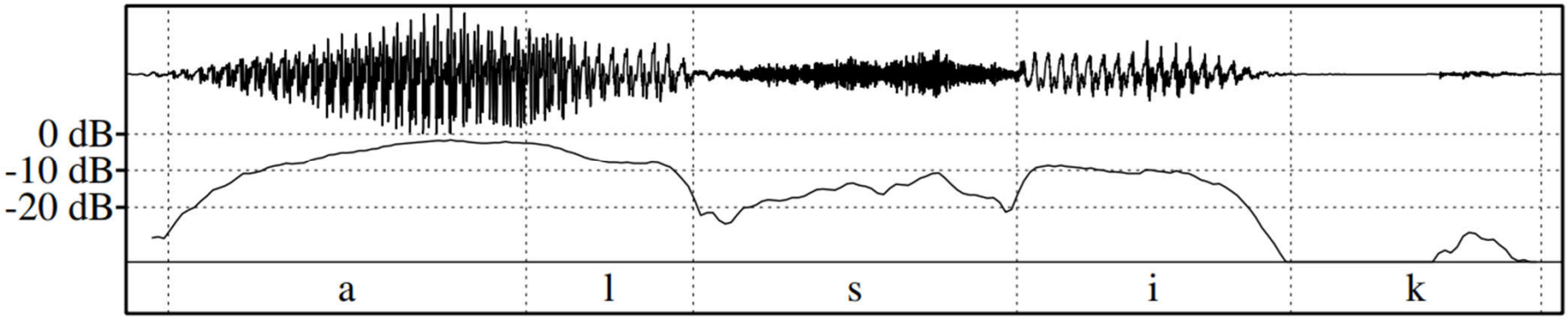
Acoustic level

- Continuous signal



We will not discuss the visual modality now.

Brain level



s1 s2 s3 s4 s5

Time

A series of discrete symbols

Mapping alternatives between levels

- Many options
  - Sound
  - Syllable
  - Word
  - Prosodic unit / phrase (between two breaths)
  - Sentence
  - Unity of thought
  - …..

Definition of phoneme

- Phonemes are the minimal set of linguistic elements from which any communication of a given language can be reconstructed at the brain level with correct meaning, but in only one way.
- Speech sounds are the physical realizations of phonemes
- Allophones are different realization solutions of the same phoneme (they do not have a meaning-distinguishing role)

Variation of speech

- Every announcement is a unique and unrepeatable event.
- Intraindividual variation
  - We can't say the same thing exactly the same way twice.
  - The speaker's speech is also constantly changing.
    - During the day (morning, noon, evening)
    - From health status
    - From mood
    - Depending on the communication context (with whom, when, where, about what, …)
- Interindividual variation
  - Every person has a characteristic voice character.

Learned automated processes related to speech

- The brain's storage of the sounds of a given language is the basis of articulation and perception.
- Articulatory base: elementary cognitive and physiological processes related to speech production (~min. 8 months – 3 years)
- Perceptual basis: elementary cognitive and physiological processes related to speech perception (can be developed up to the age of ~13-15)

Levels of language proficiency

- The lexical knowledge base refers to the knowledge of words.
- The set of syntactic knowledge determines how words can be linked together to form meaningful sentences.
- The semantic level helps in interpreting the meaning.
  For example: *Laci and Mari's friends showed up at the wedding.*
- The subject of pragmatics is how context (the actors in the dialogue, their assumed intentions, the location of the conversation, knowledge of the social environment, etc.) contributes to meaning.
  For example: “ *Where is the Eiffel Tower?”*
  The answer might be: *If you go out from behind the block, you will see it.* , or *in Paris.*

The relationship between writing and speech

Written form	Name
A, a, B, b, 1, 2, @, #, €	Characters (elements of written communication)
A, a, B, b, one, two, worm,	Orthographic signs
|e| | gy |	Grapheme (pronounced together)
ɛ ɟ / ɟ̬	IPA phonetic transcription https://scripts.sil.org/cms/scripts/page.php?item_id=IPAhome
This is it.	Phonetic transcription with ASCII characters (SAMPA) https://www.phon.ucl.ac.uk/home/sampa/hungaria.htm

Hungarian speech sound symbols

Letter	IPA	SAMPA	E1	E2	Letter	IPA	SAMPA	E1	E2
á	[a:]	A:	A:	a1	m,mm	[m], [m:]	m, m:	m, m:	m, m:
a	[ɔ]	O	a	a	m*	[ɱ]	M	-	-
o	[o]	o	o	o	n, nn	[n], [n:]	n, n:	n, n:	n, n:
ó	[o:]	o:	o:	o1	n*	[ŋ]	N	-	-
u	[u]	u	u	u	ny, nny	[ɲ], [ɲ:]	J, J:	N, N:	ny, ny:
ú	[u:]	u:	u:	u1	j, ly, jj, lly	[j], [j:]	j, j:	j, j:	j, j:
ü	[y]	y	U	u2	j*	[ç]	X)	-	-
ű	[y:]	y:	U:	u3	h, hh	[h], [h:]	h, h:	h, h:	h
i	[i]	i	i	i	h*	[ɦ]	h	-	-
í	[i:]	i:	i:	i1	h*, ch*	[x]	x	-	-
é	[e:]	e:	E:	e1	h*	[ɣ]		-	-
ö	[ø]	2	O	o2	v, vv	[v], [v:]	v, v:	v, v:	v, v:
ő	[ø:]	2:	O:	o3	f, ff	[f], [f:]	f, f:	f, f:	f, f:
e	[ɛ]	E	e	e	z, zz	[z], [z:]	z, z:	z, z:	z, z:
- *(svá)*	[ə]				sz, ssz	[s], [s:]	s, s:	s, s:	s, s:
b, bb	[b], [b:]	b, b:	b, b:	b	zs, zzs	[ʒ], [ʒ:]	Z, Z:	Z, Z:	Z, Z:
p, pp	[p], [p:]	p, p:	p, p:	p	s, ss	[ʃ], [ʃ:]	S, S:	S, S:	S, S:
d, dd	[d], [d:]	d, d:	d, d:	d	dz	[d͡z], [d͡z:]	dz, dz:	dz, dz	dz
t, tt	[t], [t:]	t, t:	t, t:	t, t:	dzs	[d͡ʒ], [d͡ʒ:]	dZ, dZ:	dZ, dZ	dZ
gy, ggy	[ɟ], [ɟ:]	d', d':	G, G:	gy, gy:	c, cc	[t͡s], [t͡s:]	ts, ts:	c, c:	c, c:
ty, tty	[c], [c:]	t', t':	T, T:	ty, ty:	cs, ccs	[t͡ʃ], [t͡ʃ:]	tS, tS	C, C:	C, C:
g, gg	[g], [g:]	g, g:	g, g:	g, g:	l, ll	[l], [l:]	l, l:	l, l:	l, l:
k,kk	[k], [k:]	k, k:	k, k:	k, k:	r, rr	[r], [r:]	r, r:	r, r:	r, r:

Phonemic chart of British English

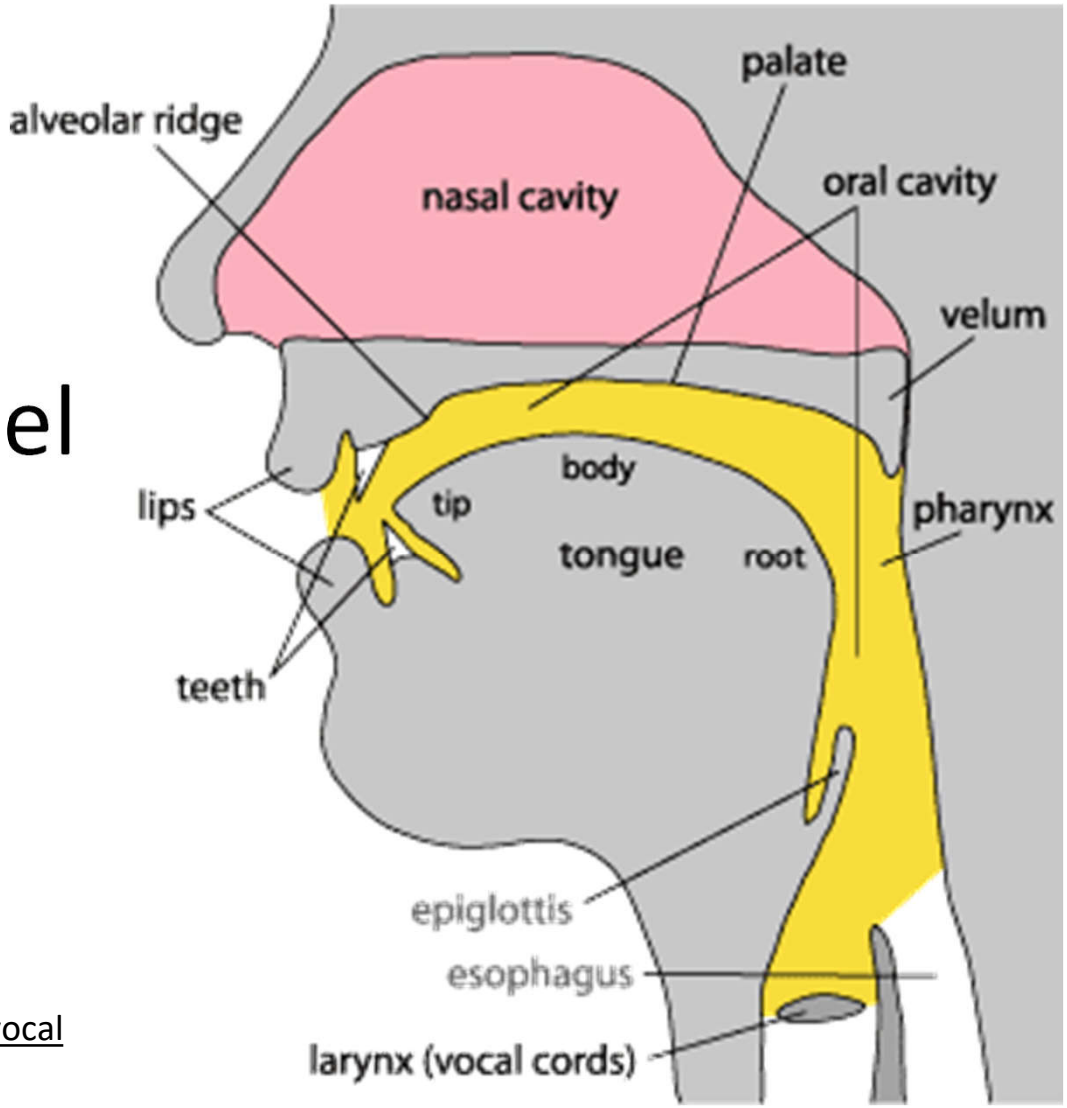
	monophthongs				diphthongs			
VOWELS	i: sheep	ɪ ship	ʊ good	u: shoot	ɪə here	eɪ wait		
	e bed	ə teacher	ɜ: bird	ɔ: door	ʊə tourist	ɔɪ boy	əʊ show	
	æ cat	ʌ up	ɑ: far	ɒ on	eə hair	aɪ my	aʊ cow	
CONSONANTS	p pea	b boat	t tea	d dog	ʧ cheese	ʤ June	k car	g go
	f fly	v video	θ think	ð this	s see	z zoo	ʃ shall	ʒ television
	m man	n now	ŋ sing	h hat	l love	r red	w wet	j yes

Phonemic Chart
voiced
unvoiced

Source: https://www.englishclub.com/pronunciation/phonemic-chart.php

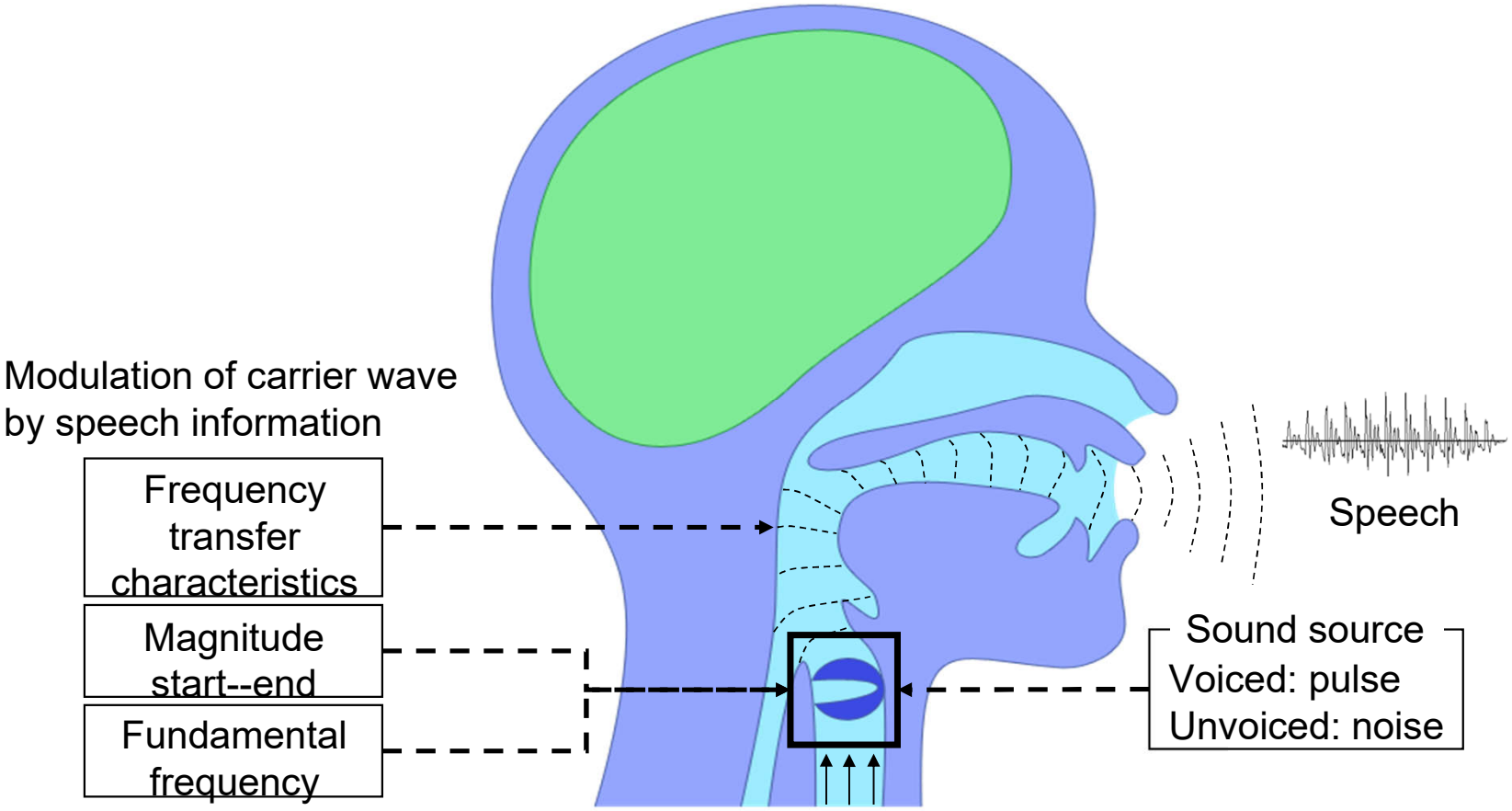
5. Voice formation, physiological foundations

The articulatory channel (vocal tract)



Source:
http://www.englishbaby.com/lessons/3201/member_submitted/vocal_organs_of_speech

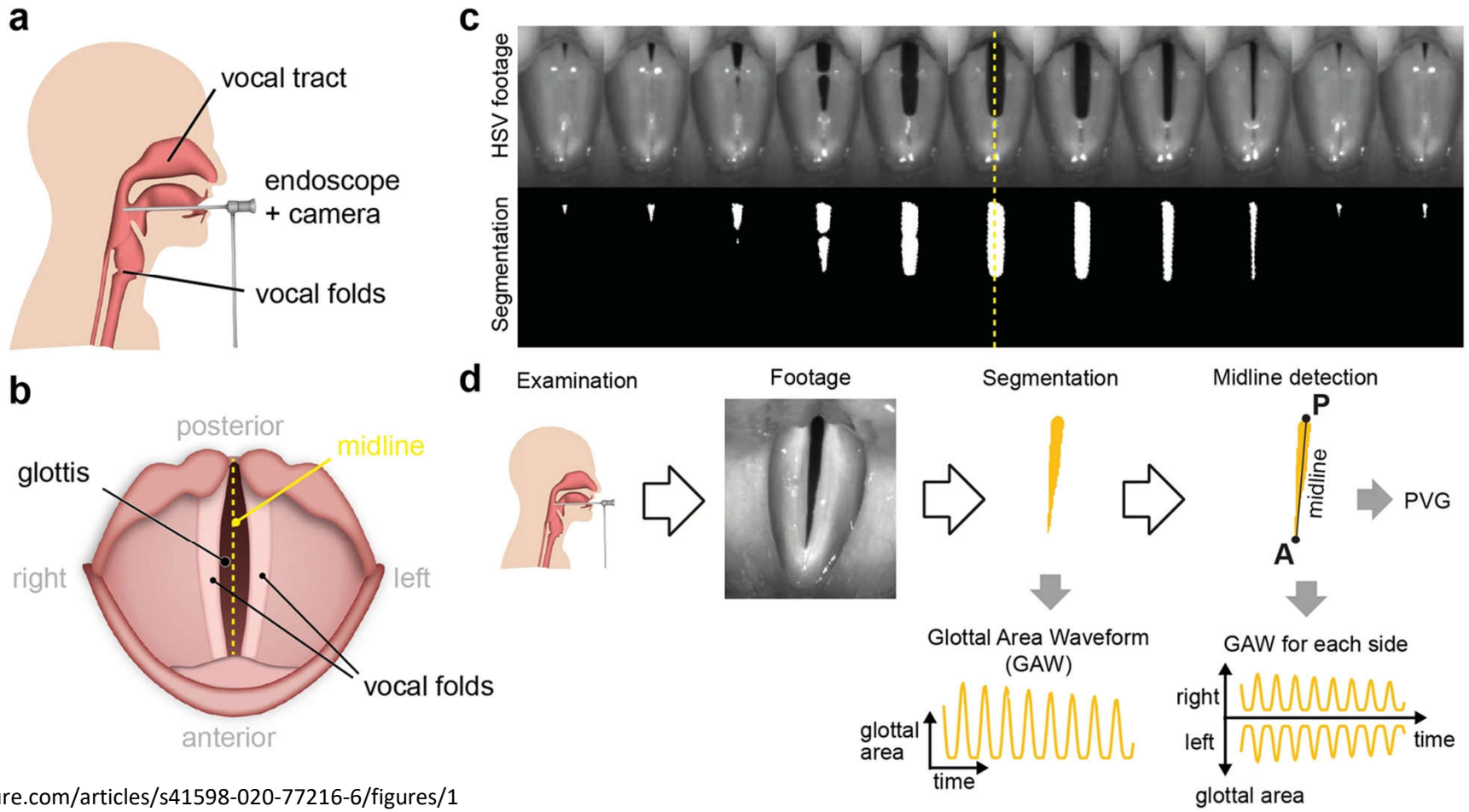
Speech production mechanism



Sound generation mechanisms (in Hungarian)

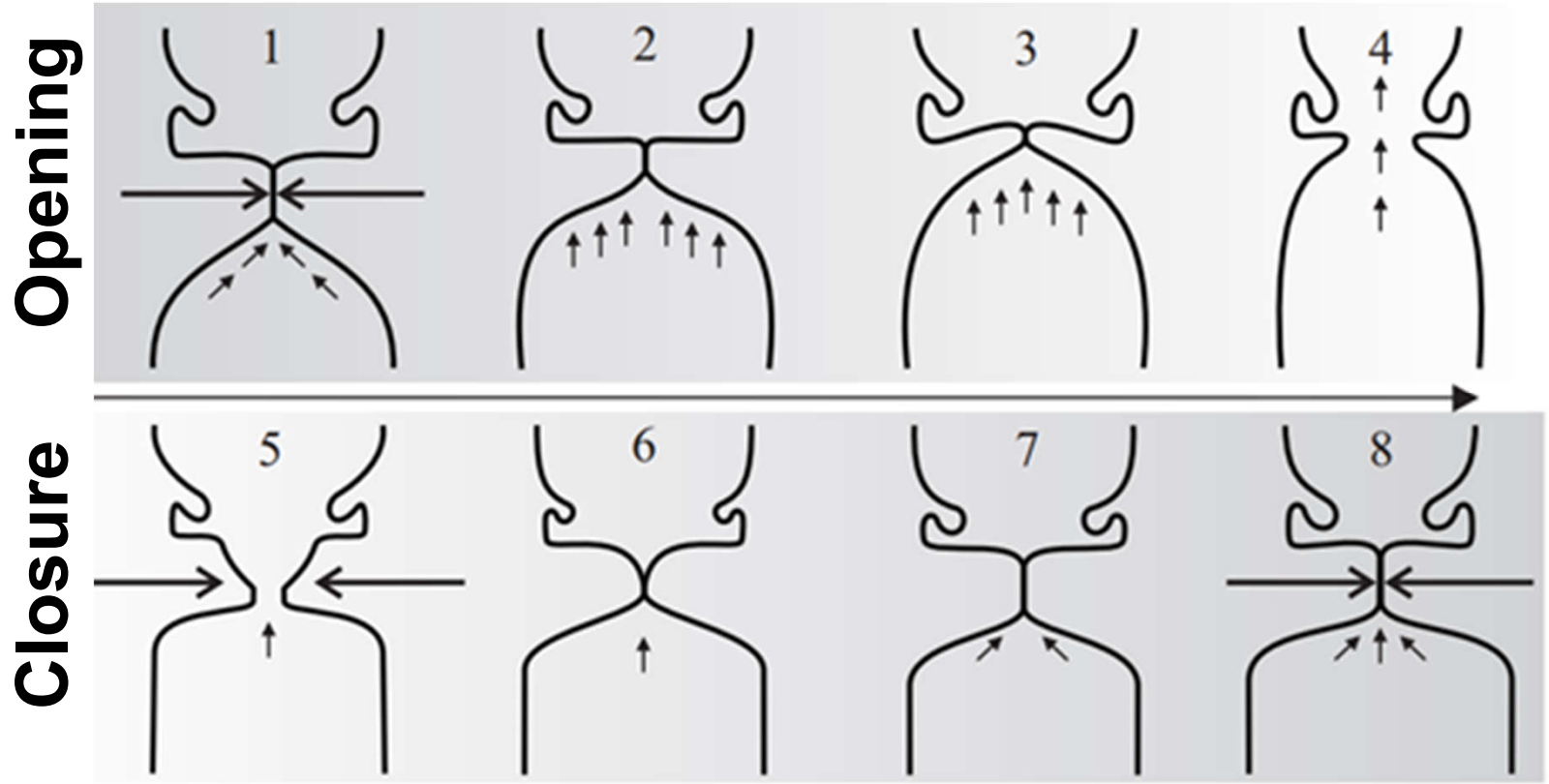
- Sound (quasi) periodic
  - The vocal cords vibrate
  - Vowels and voiced consonants
- Turbulent flow
  - The vocal cords are open.
  - e.g. s, sz , f
- Plosive sounds
  - Closure + closure release (burst)
  - Vocal cords continuously closed, then suddenly open, e.g. p, t, k
- Mixed excitation

Typical positions of the vocal cords (glottis)

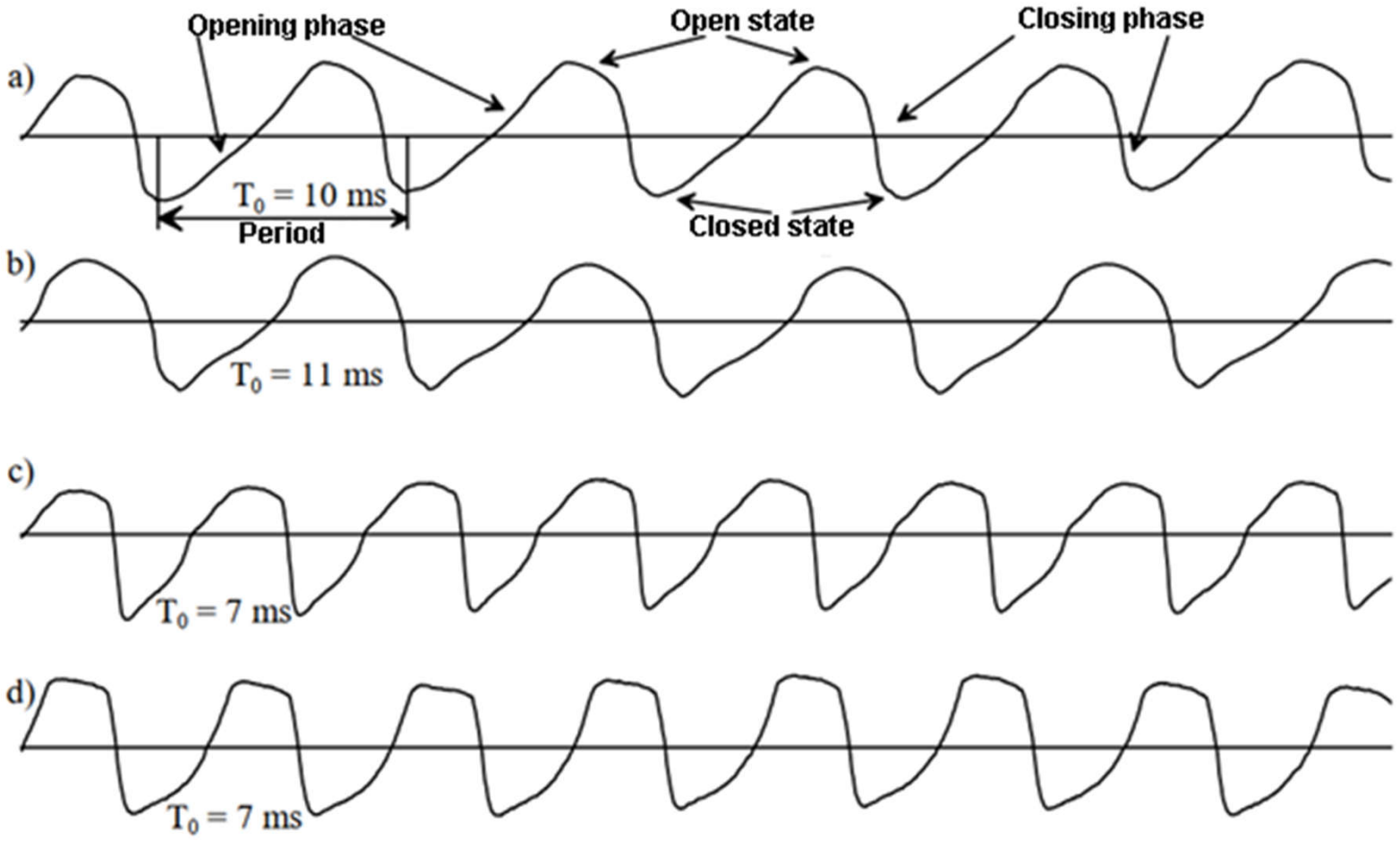


Source: https://www.nature.com/articles/s41598-020-77216-6/figures/1

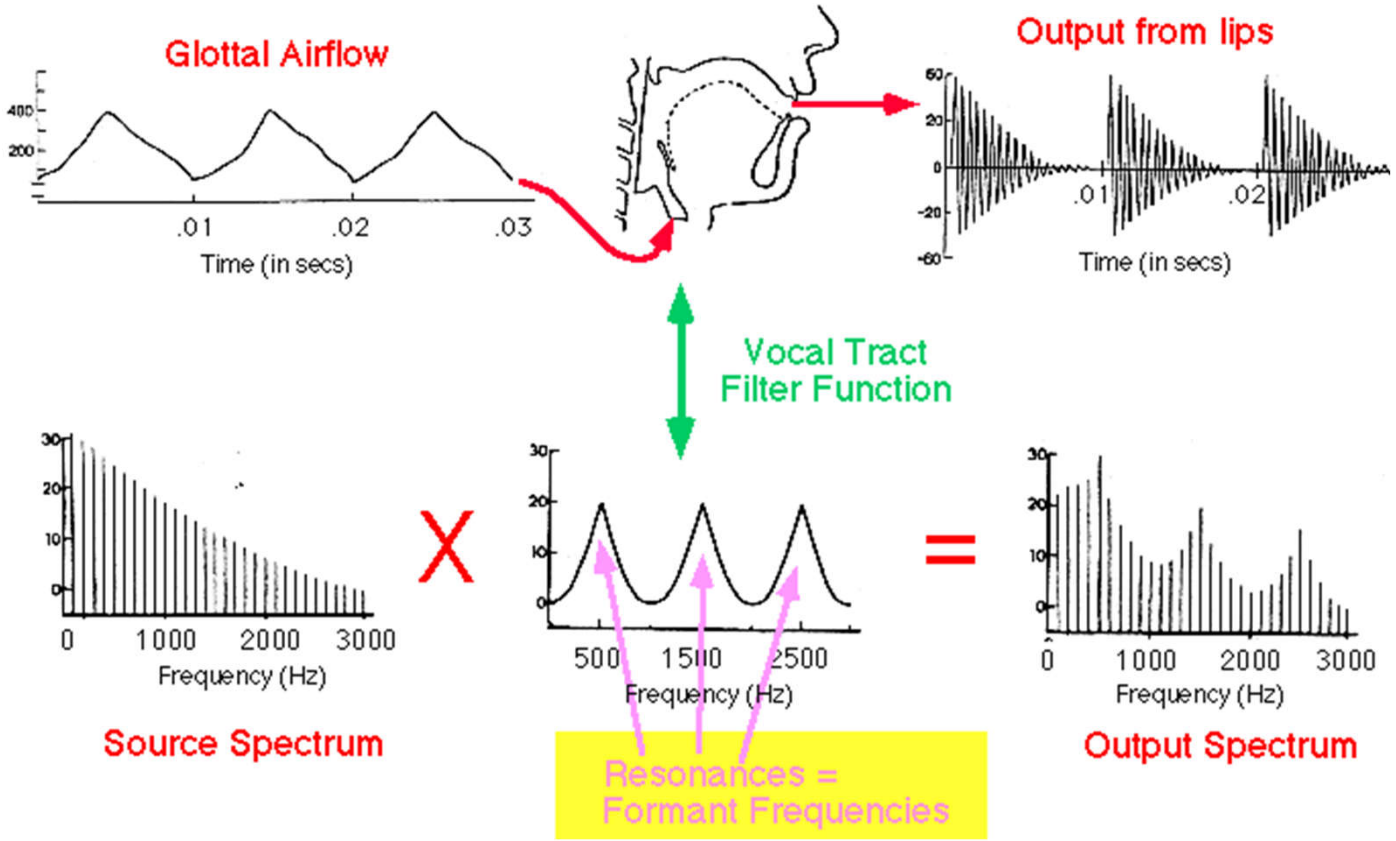
Movement of the vocal cords during a vocal period



Male and female vocal waveform (a and i sound)



Formation of voiced sounds at physiological and model levels



Discrete-time model of speech generation