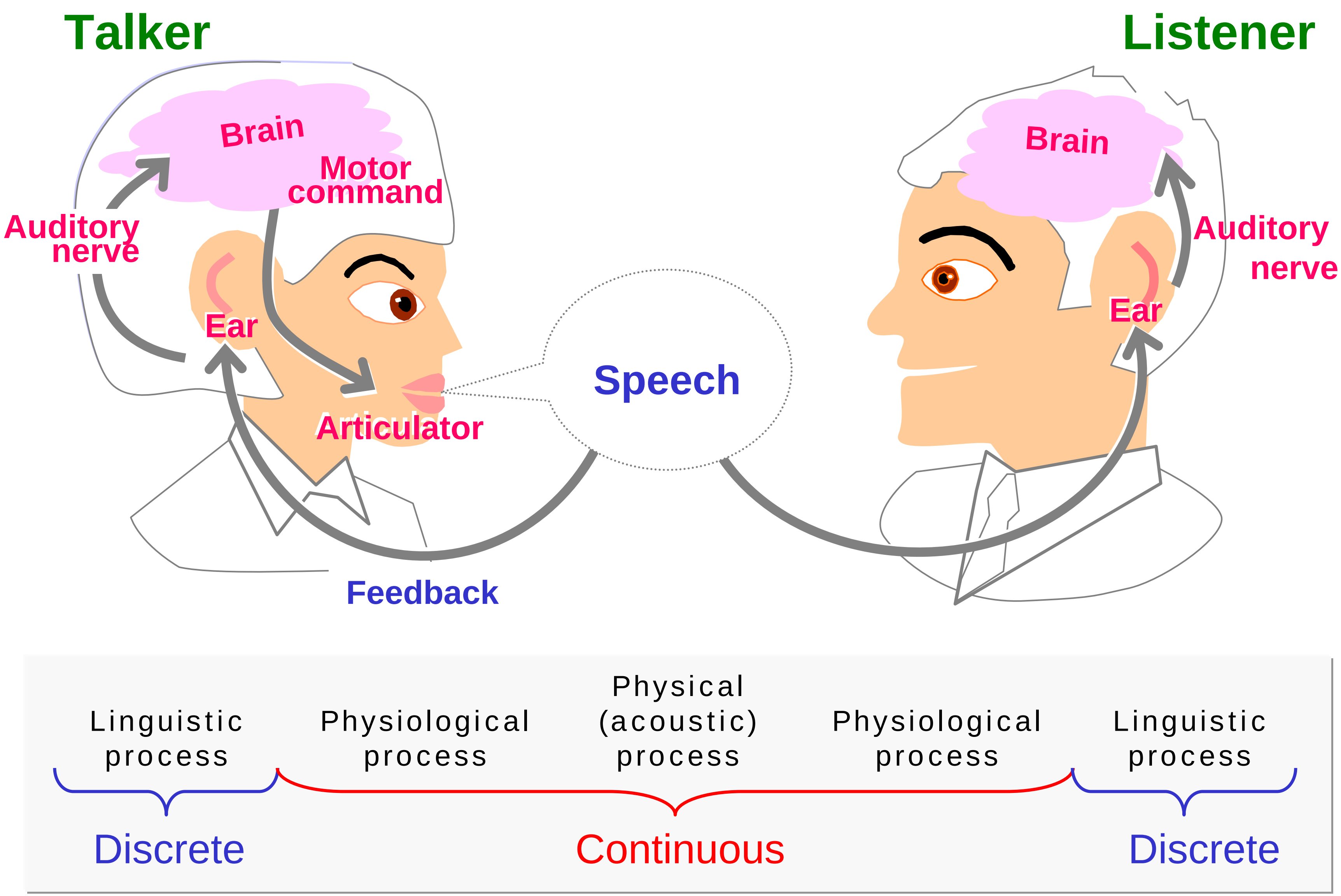
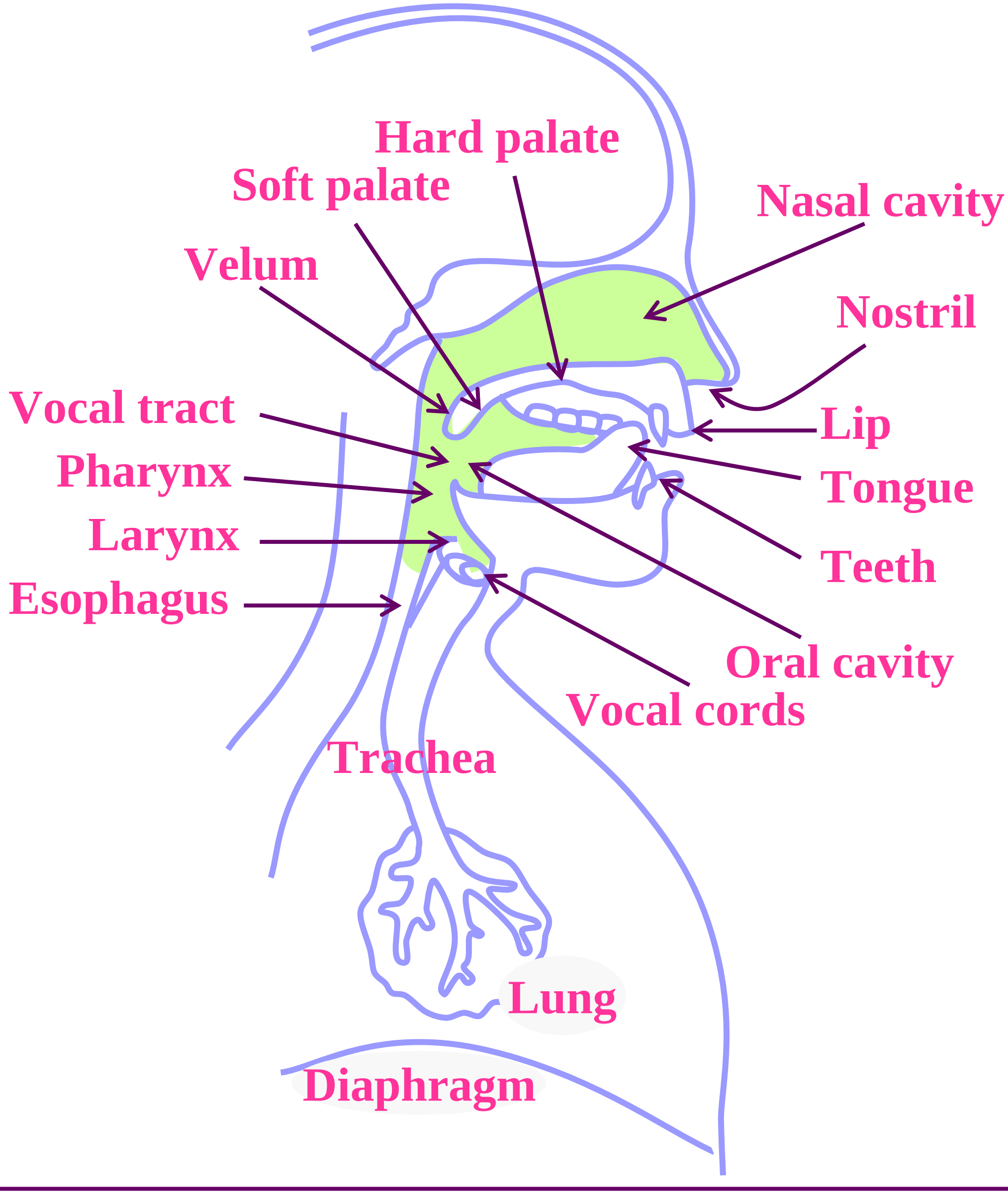
Speech Production

Sadaoki Furui
Tokyo Institute of Technology
Department of Computer Science
furui@cs.titech.ac.jp

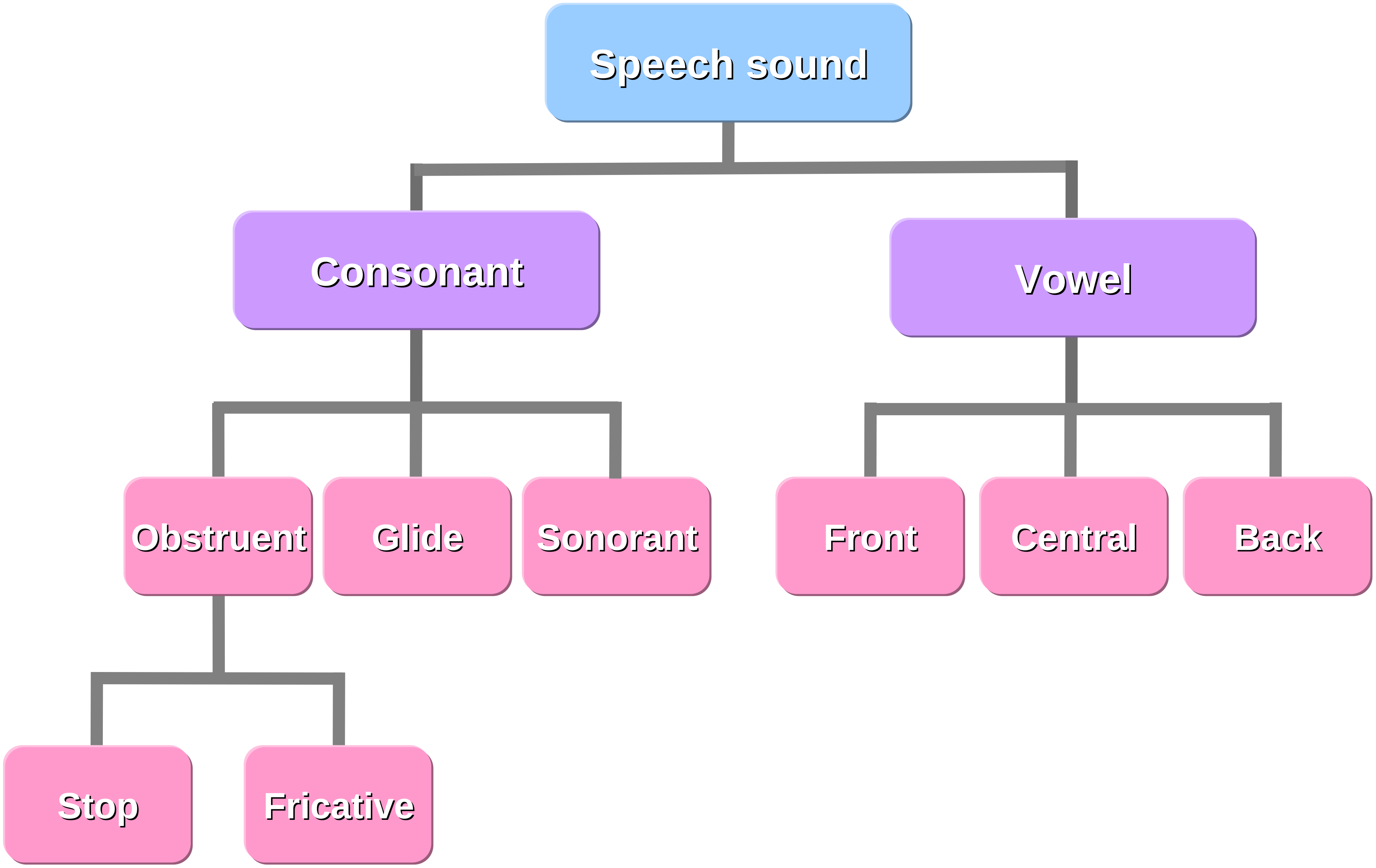
Speech chain



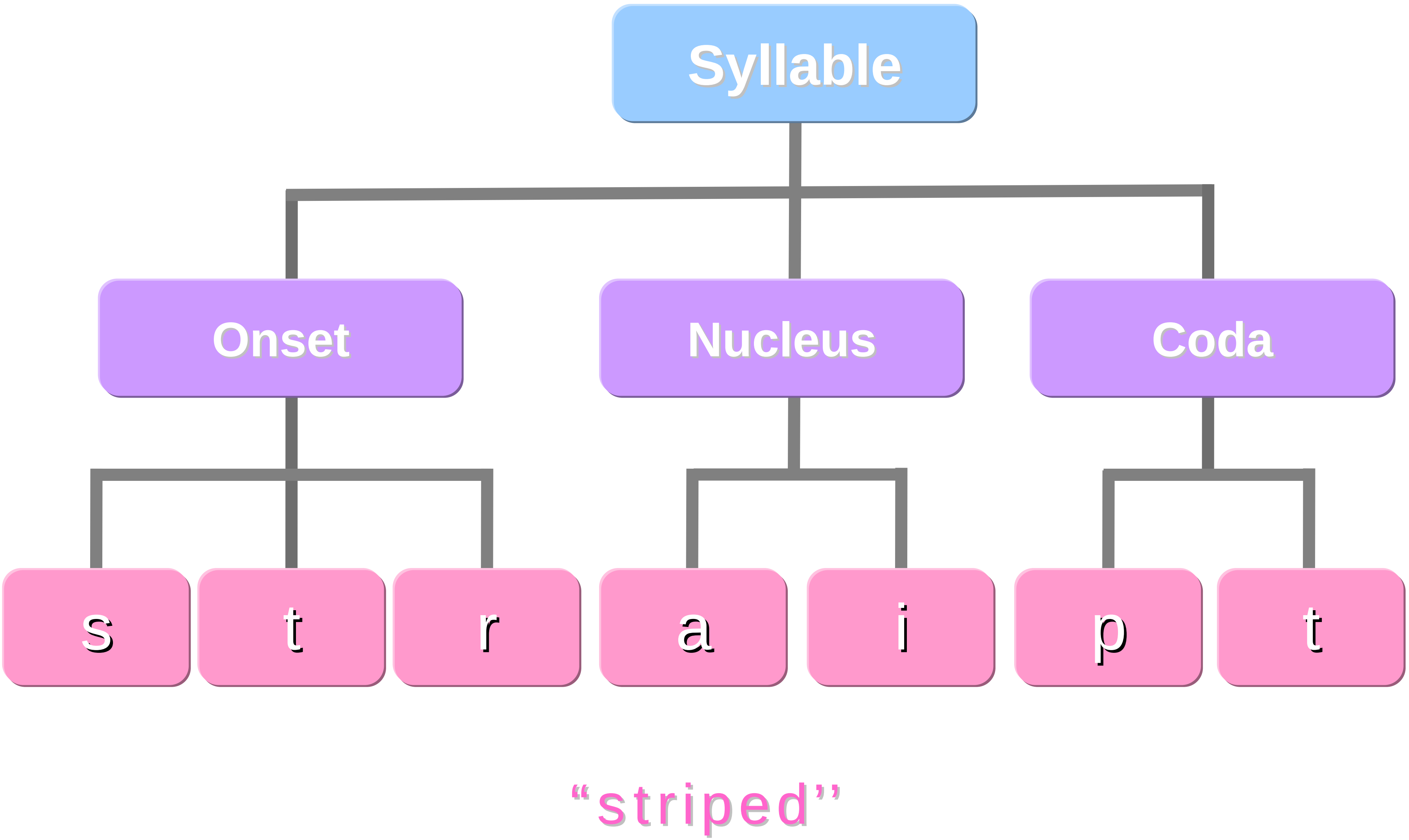
Schematic diagram of the human vocal mechanism



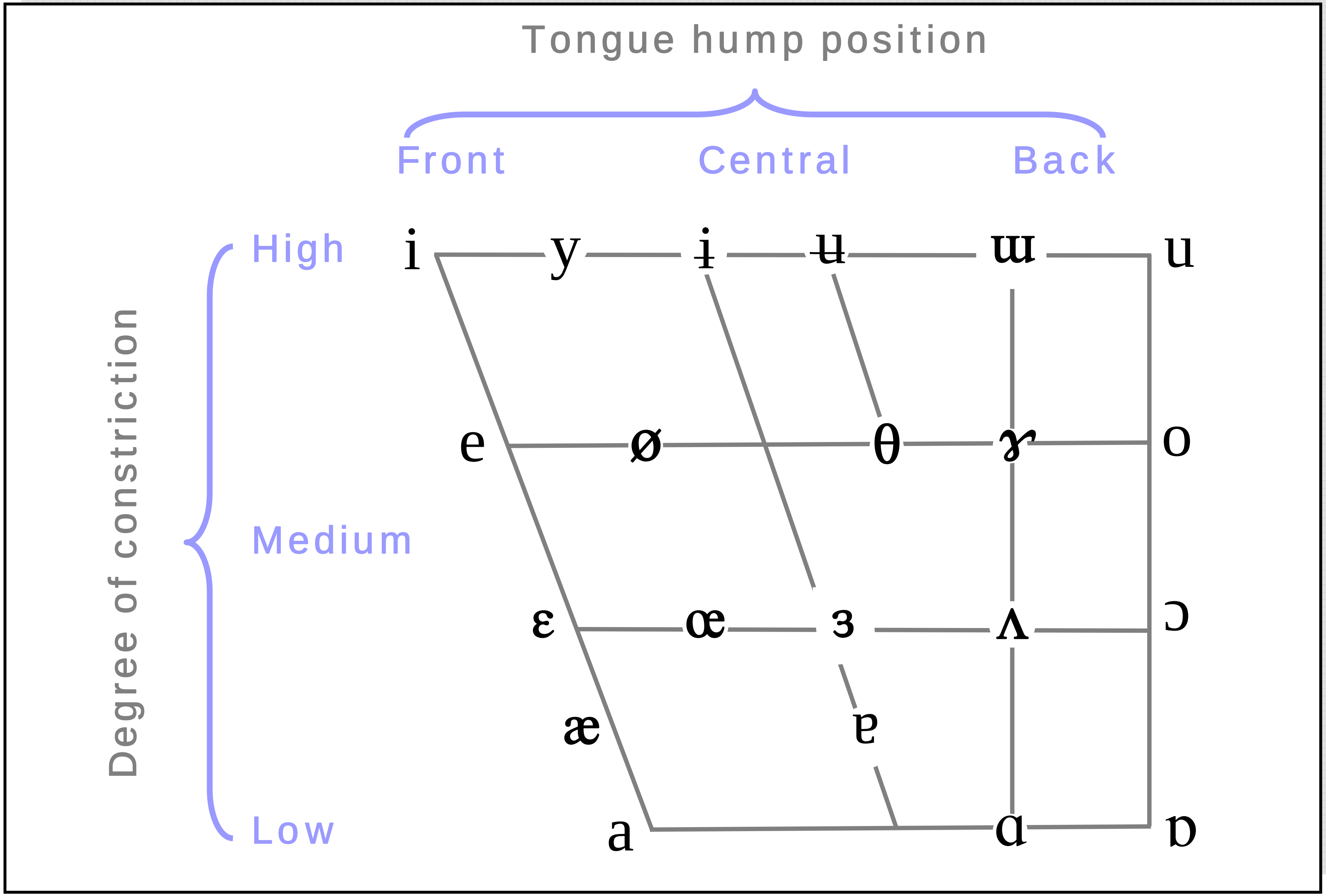
Example of a logical concept hierarchy of phonemes



Example of an ontological hierarchy



Vowel classification from approximate vocal organ representation



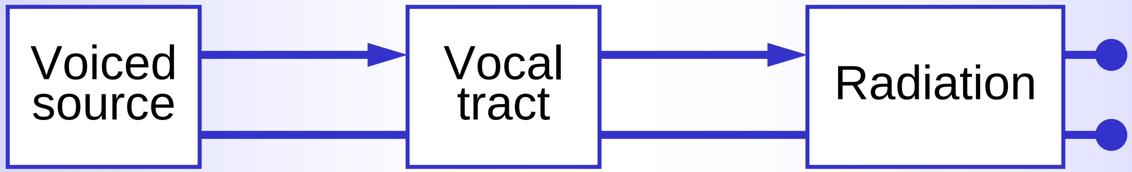
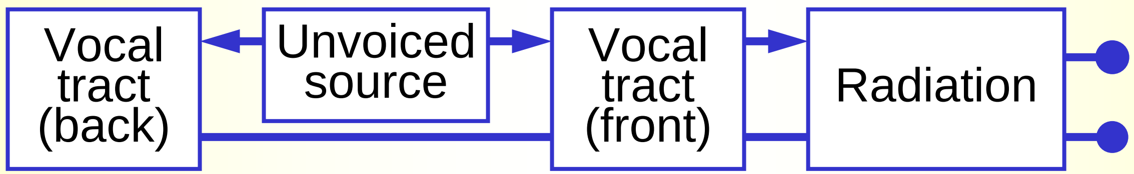
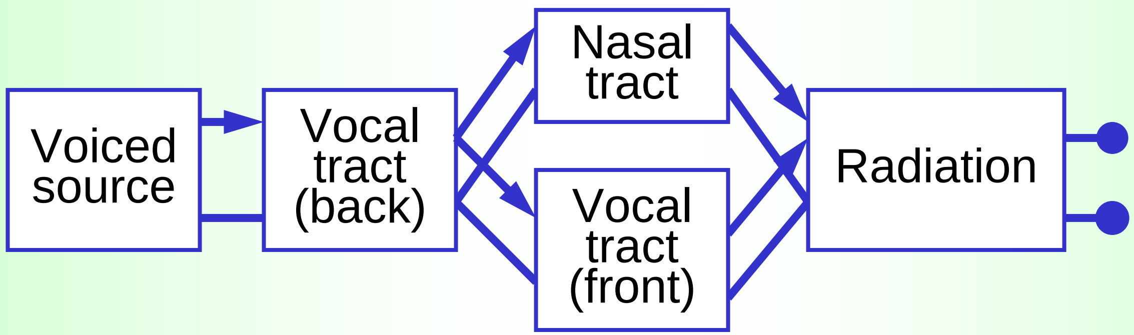
Consonants

Articulation place		Labial		Dental		Alveolar		Palatal		Glottal	
Source		V	UV	V	UV	V	UV	V	UV	V	UV
Articulation manner	Fricatives	v	f	ð	θ	z	s	ʒ	ʃ		h
	Affricates					dz	ts	dʒ	tʃ		
	Plosives	b	p			d	t	g	k		
	Semivowels	w				l		j, r			
	Nasals	m				n		ŋ			

V= voiced; UV= unvoiced

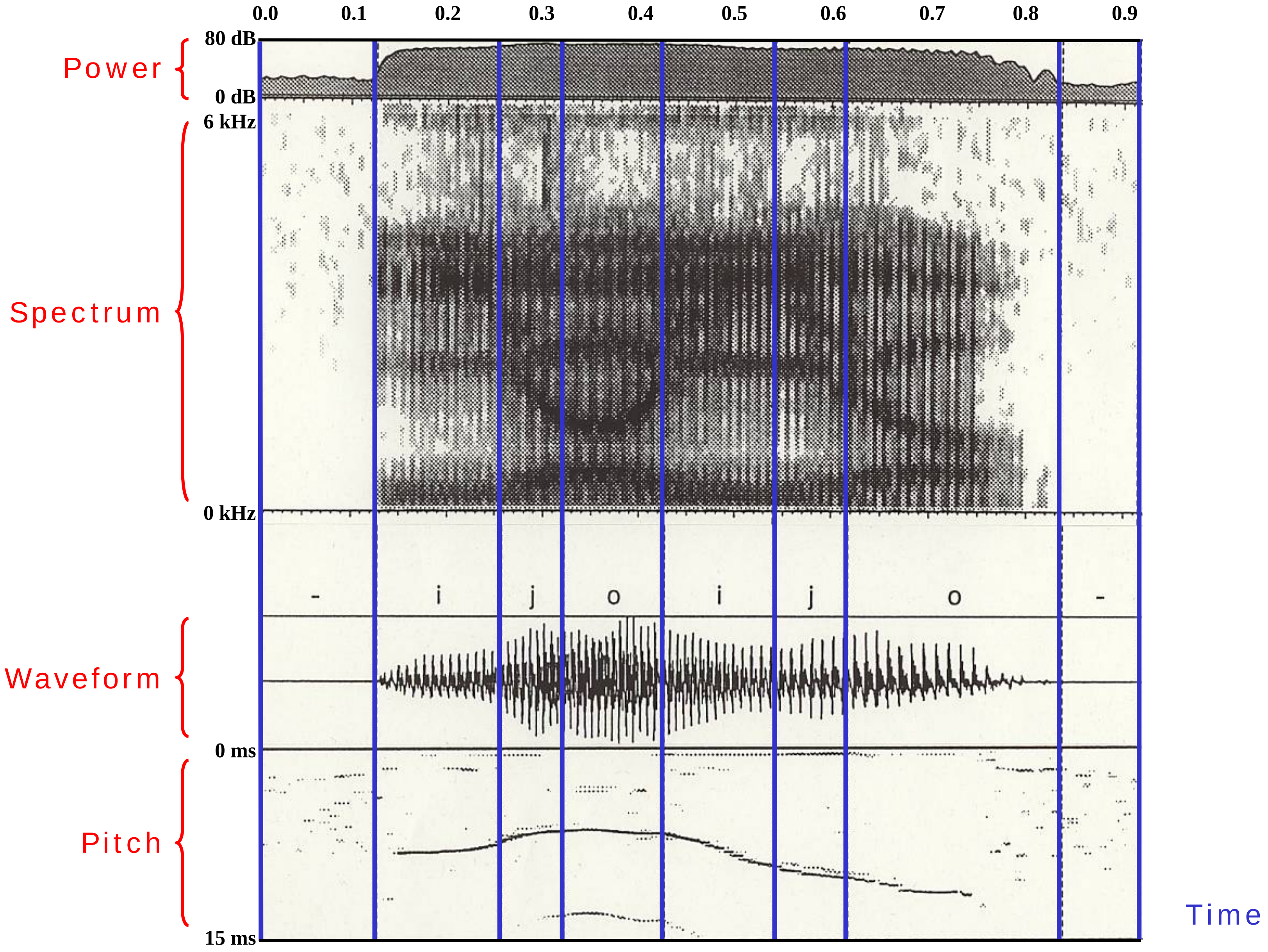
p.12, table 2.1

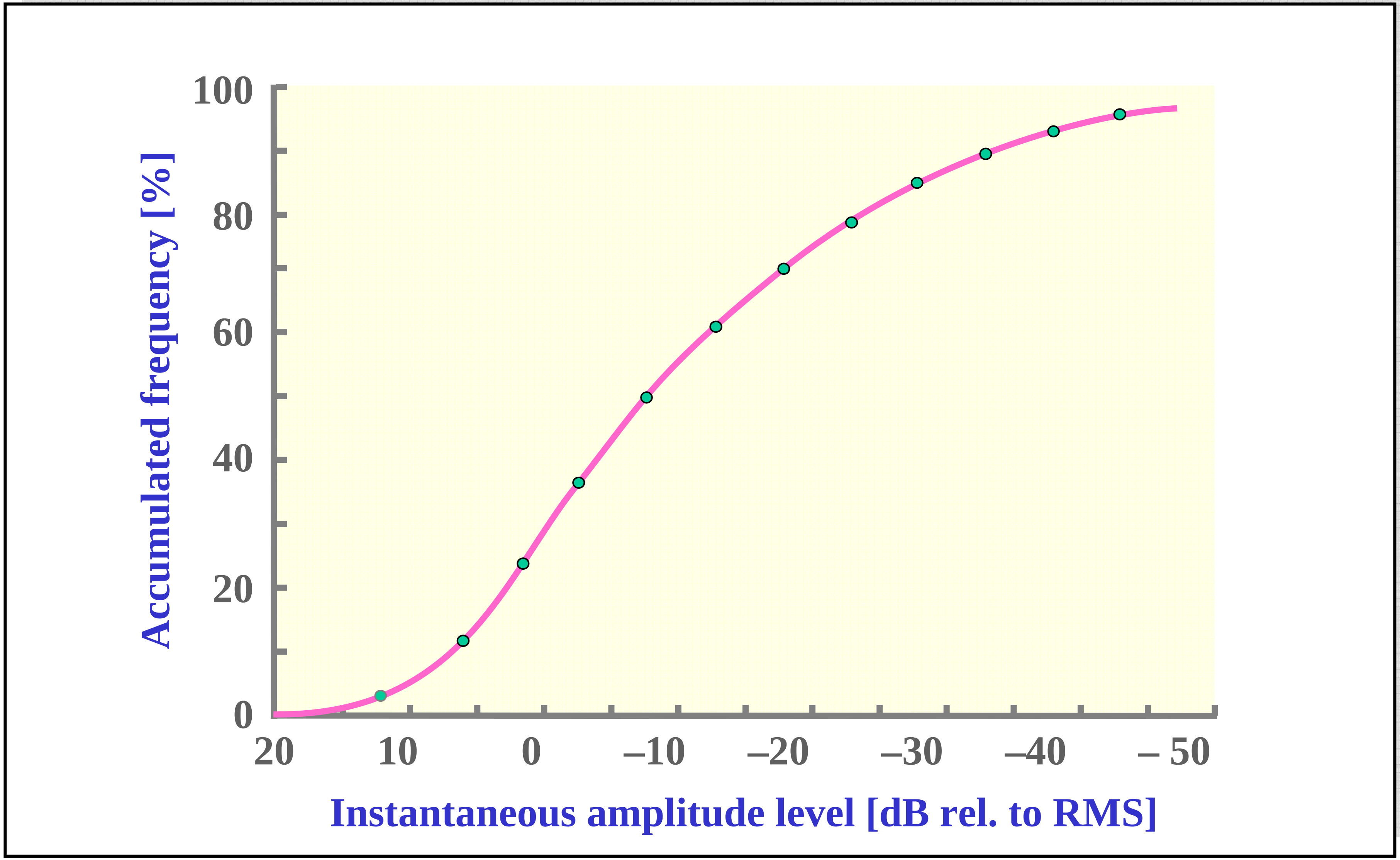
Speech production process models

Type	Speech production model	System function
Vowel type		**Resonance only (all-pole model)**
Consonant type		**Resonance and anti-resonance (pole-zero)**
Nasal type (nasal & nasalized vowel)		**Resonance and anti-resonance (pole-zero)**

(p.28, table3.1)

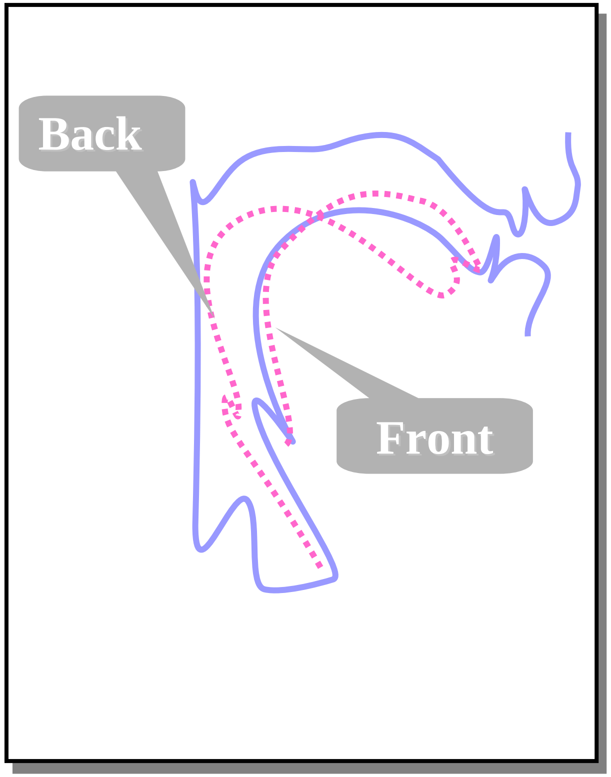
Speech energy, sound spectrogram, waveform and pitch





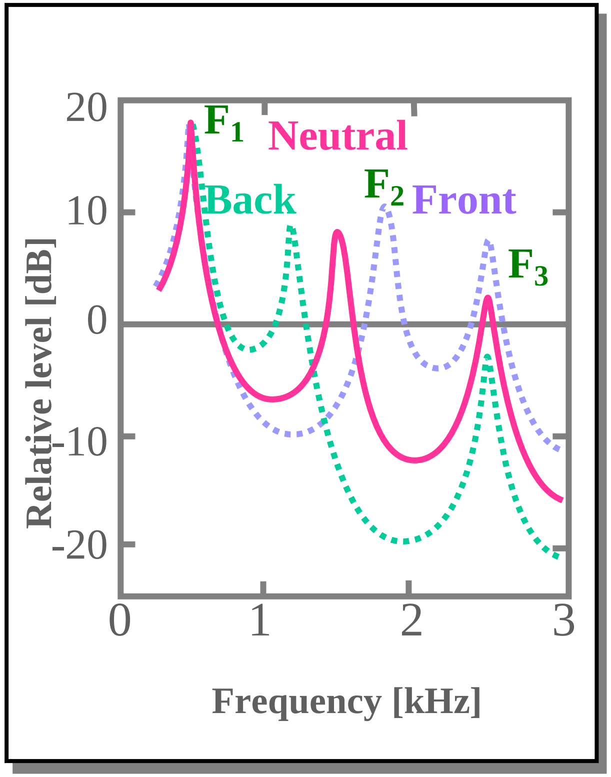
Accumulated distribution of speech amplitude level calculated for utterances made by 80 speakers having a duration of roughly 37 min.

Examples of the relationship between vocal tract shapes and vowel spectral envelopes



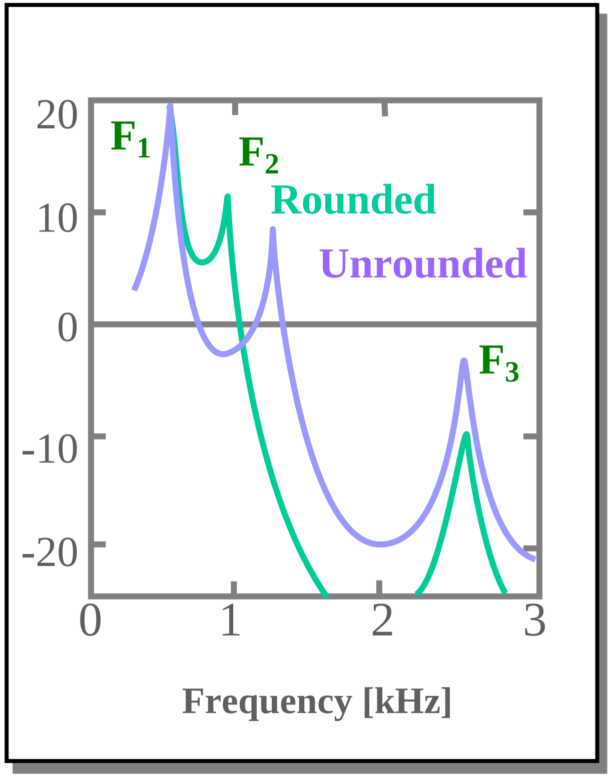
(a)

Schematization of mid-sagittal section of vocal tract for a neutral vowel (solid contour), and for back and front tongue-body positions.



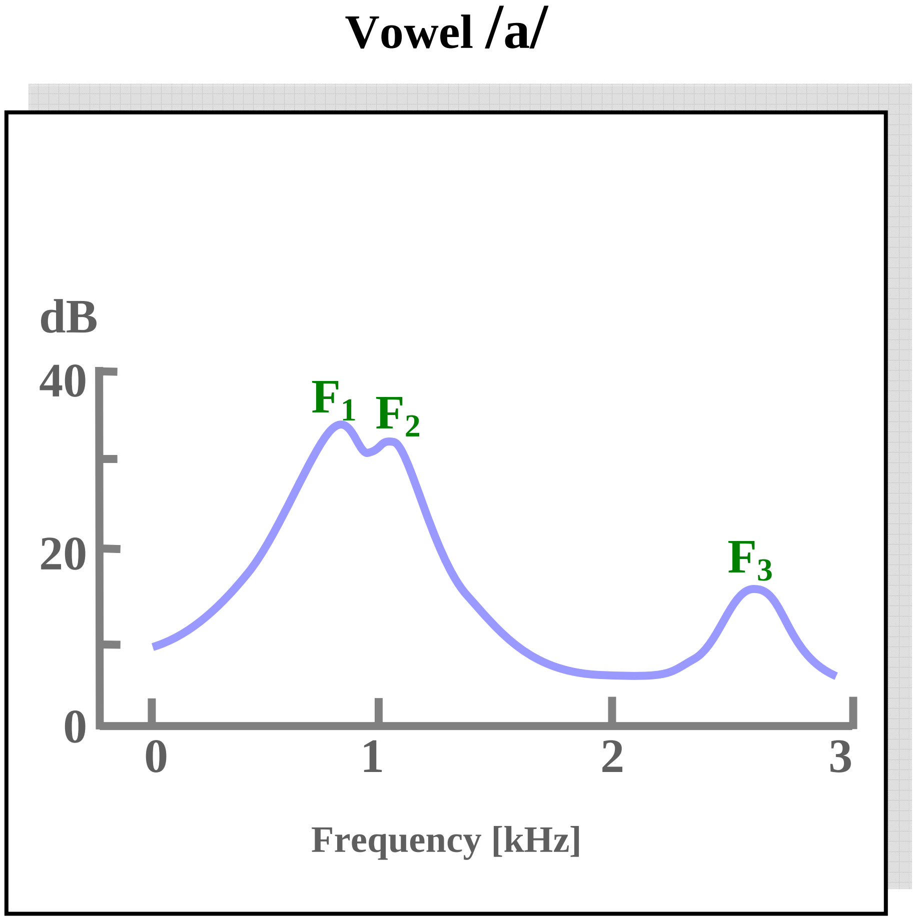
(b)

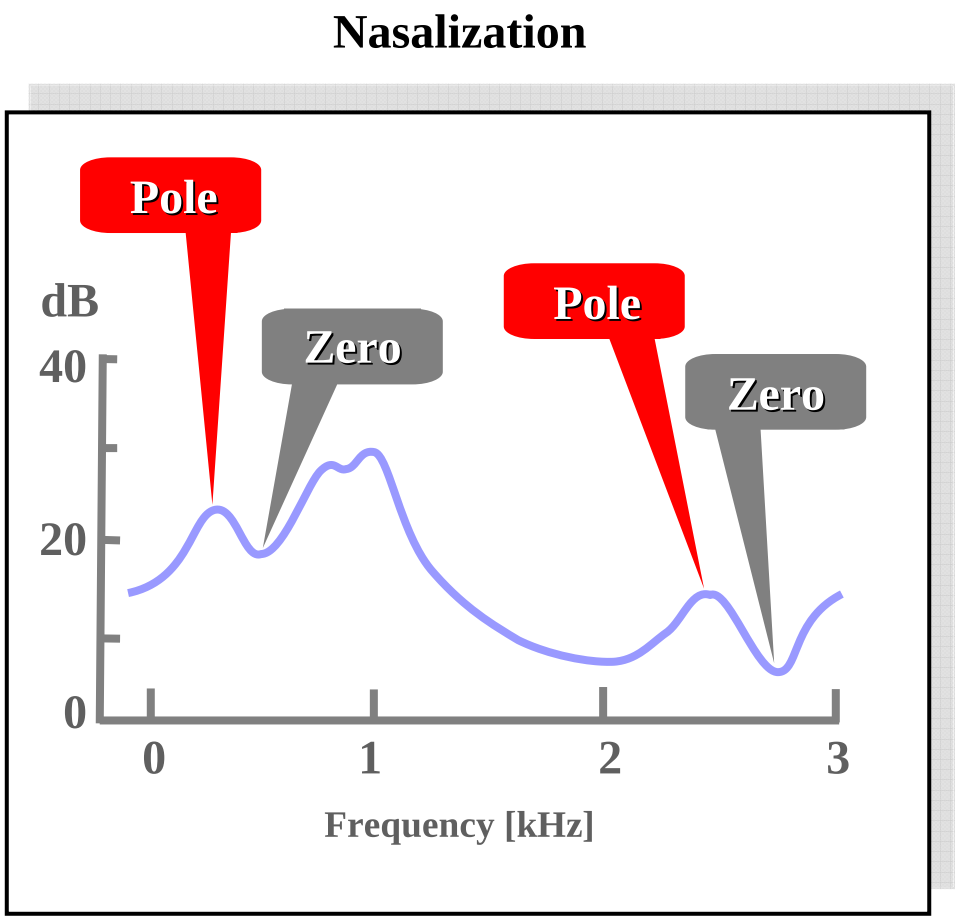
Idealized spectral envelopes corresponding to the three tongue-body configurations in (a).



(c)

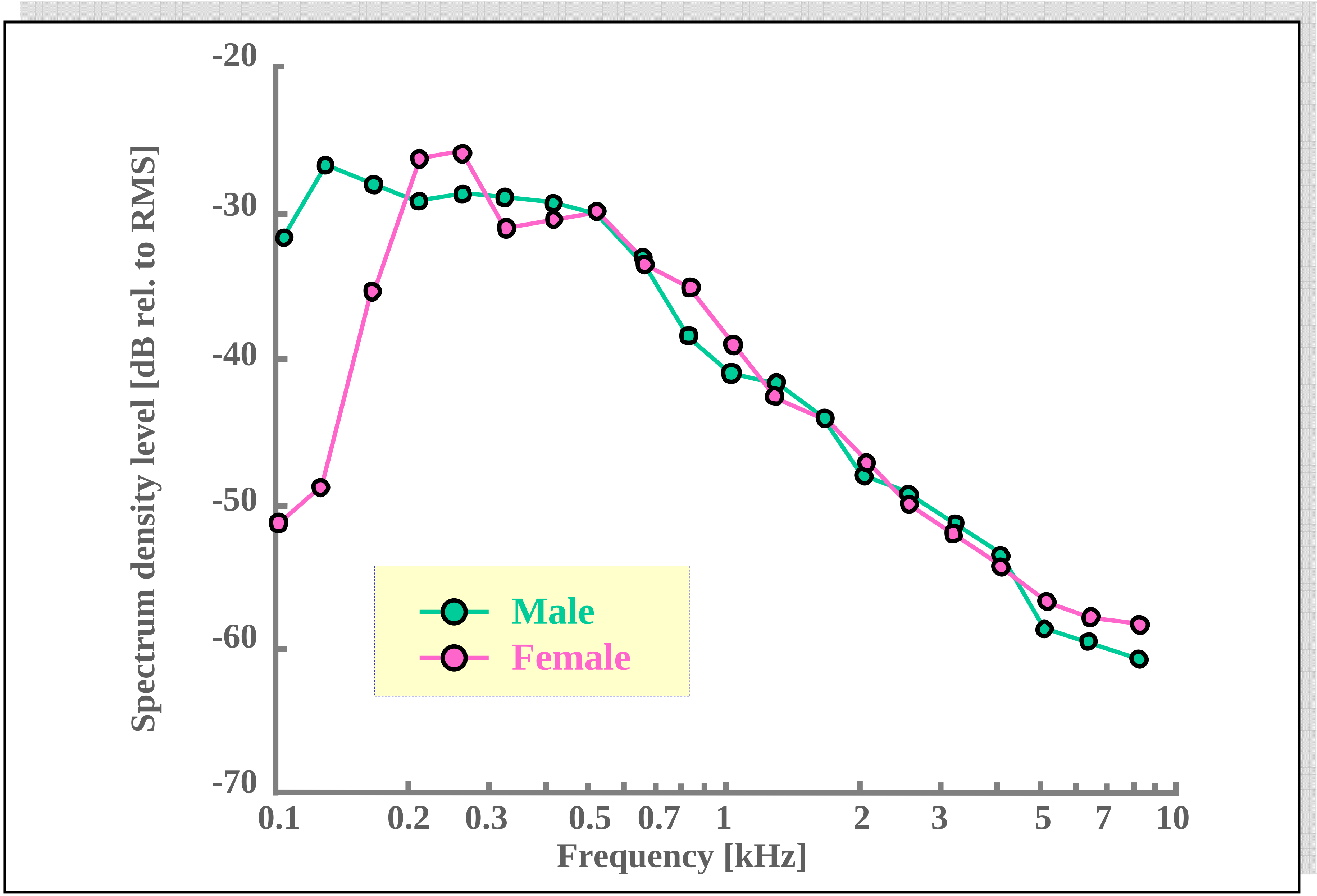
Approximate effect of lip rounding on the spectral envelope for a back vowel.



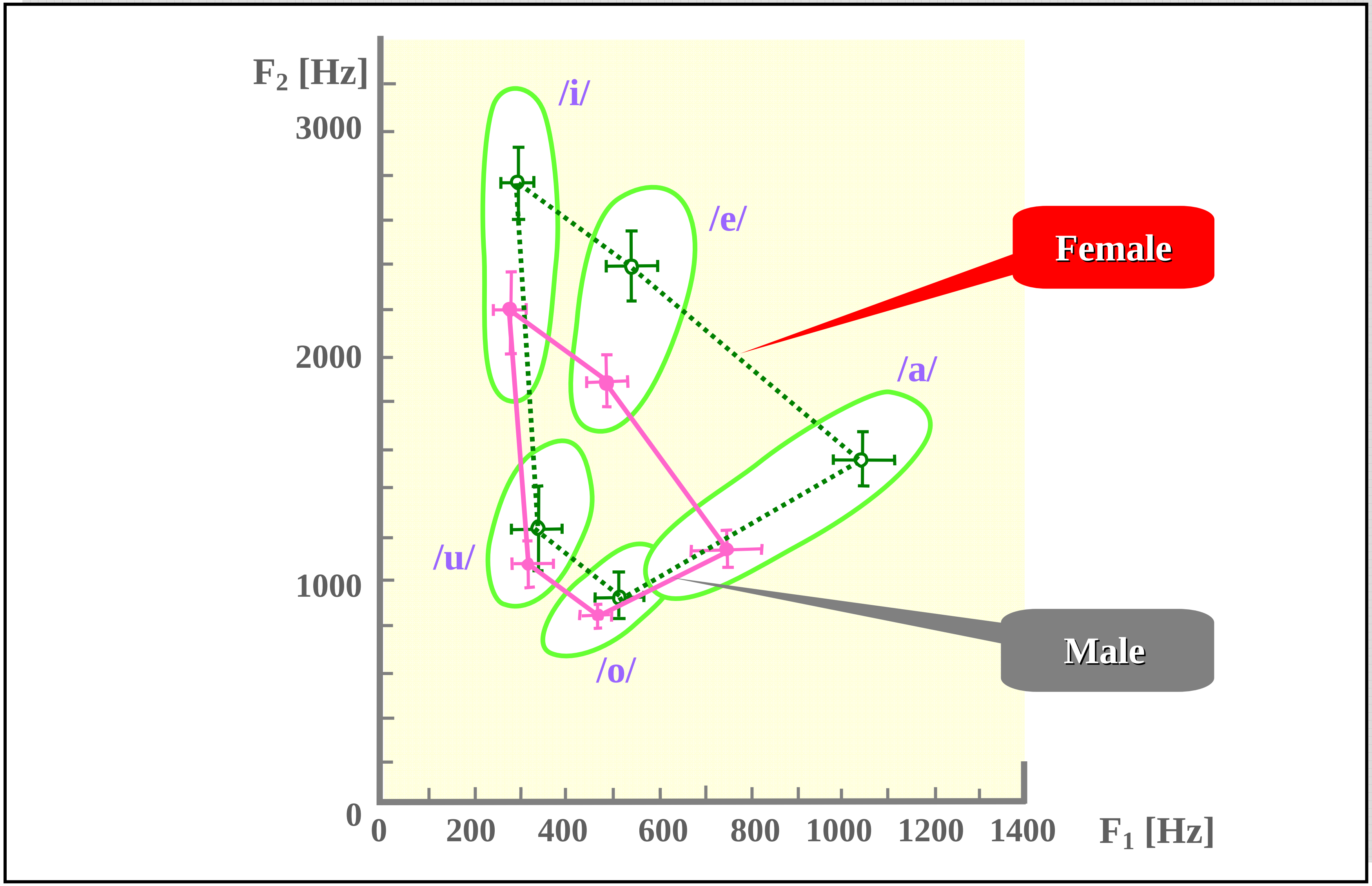


An example of spectral change caused by the nasalization for vowel /a/. It is characterized by pole-zero pairs at 300~400 Hz and at around 2500 Hz. F_1, F_2, F_3 are formants.

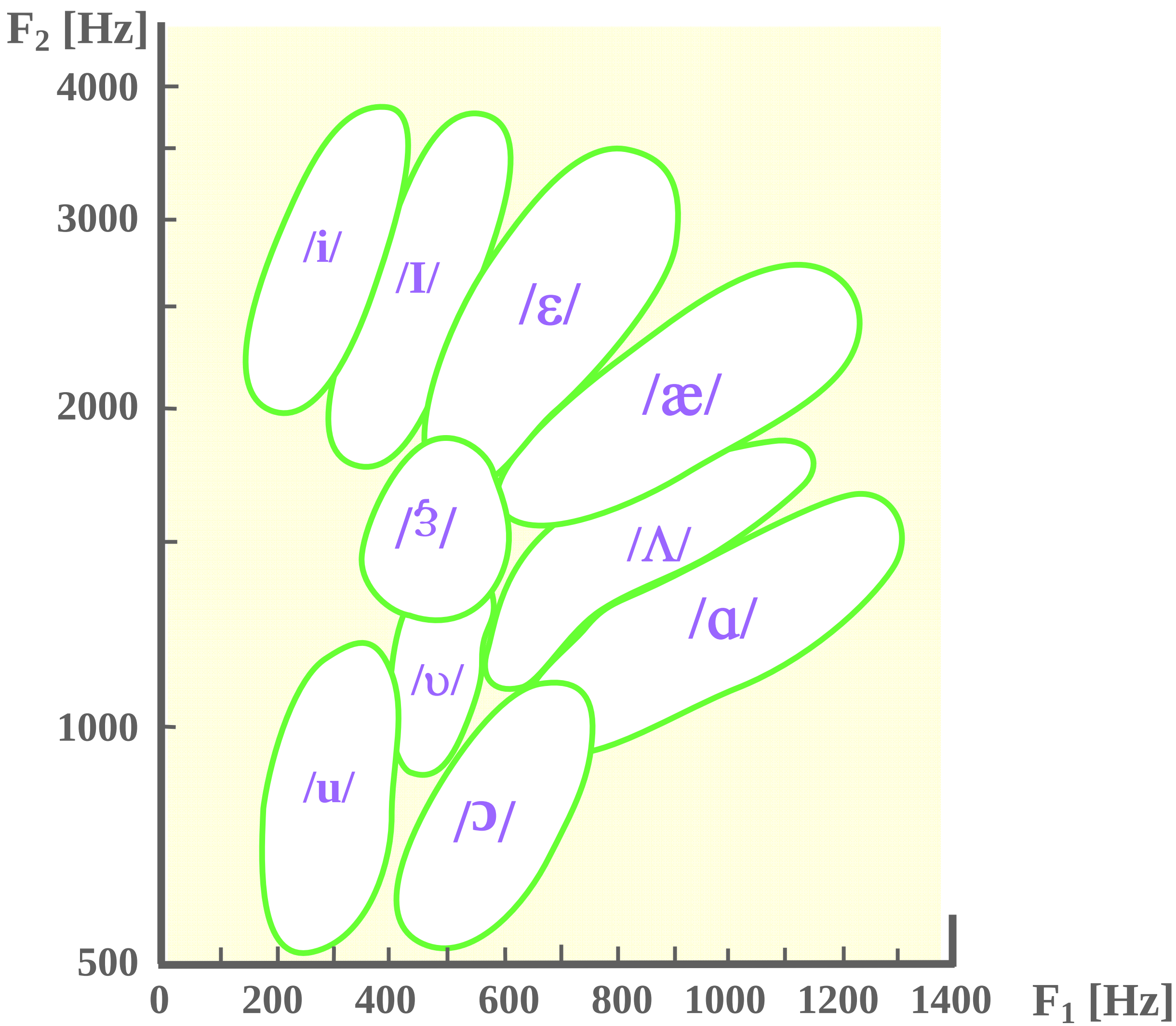
Long-time averaged speech spectrum calculated for utterances made by 80 speakers



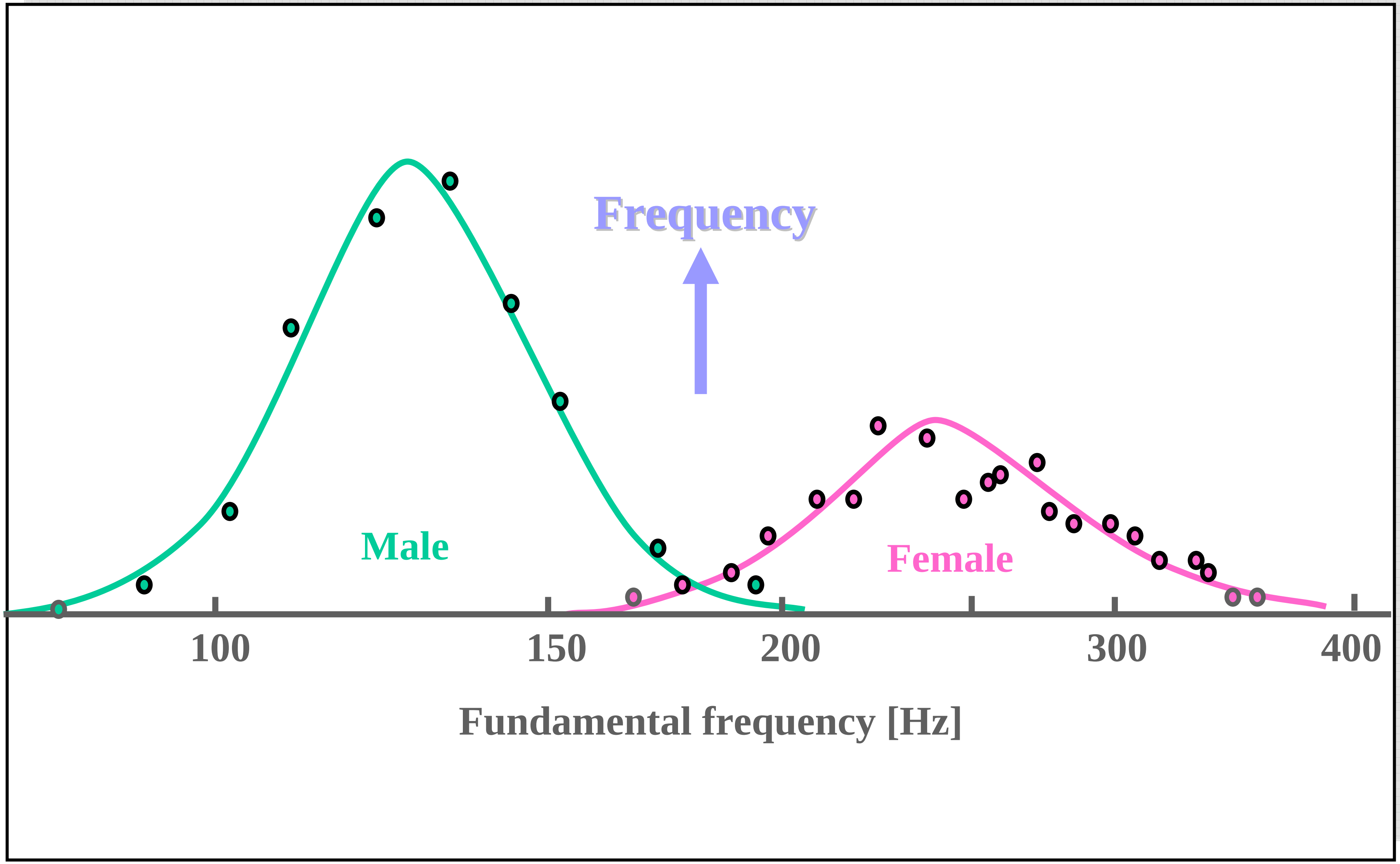
0105-06



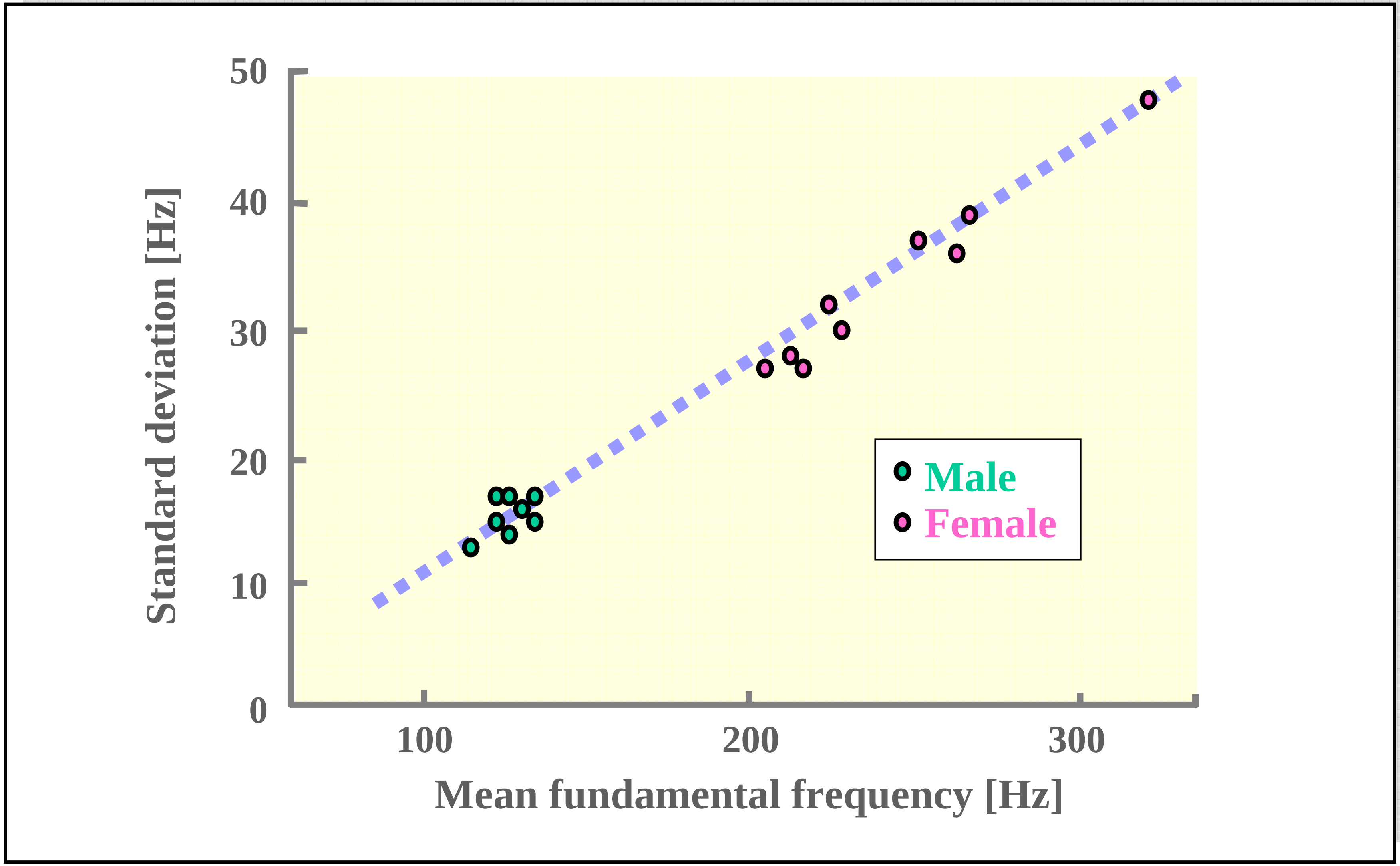
Scatter diagram of formant frequencies of five Japanese vowels uttered by 60 speakers (30 males and 30 females) in the F_1-F_2 plane



Scatter diagram of formant frequencies of 10 English vowels uttered by 76 speakers (33 adult males, 28 adult females, and 15 children) in the F_1~F_2 plane.

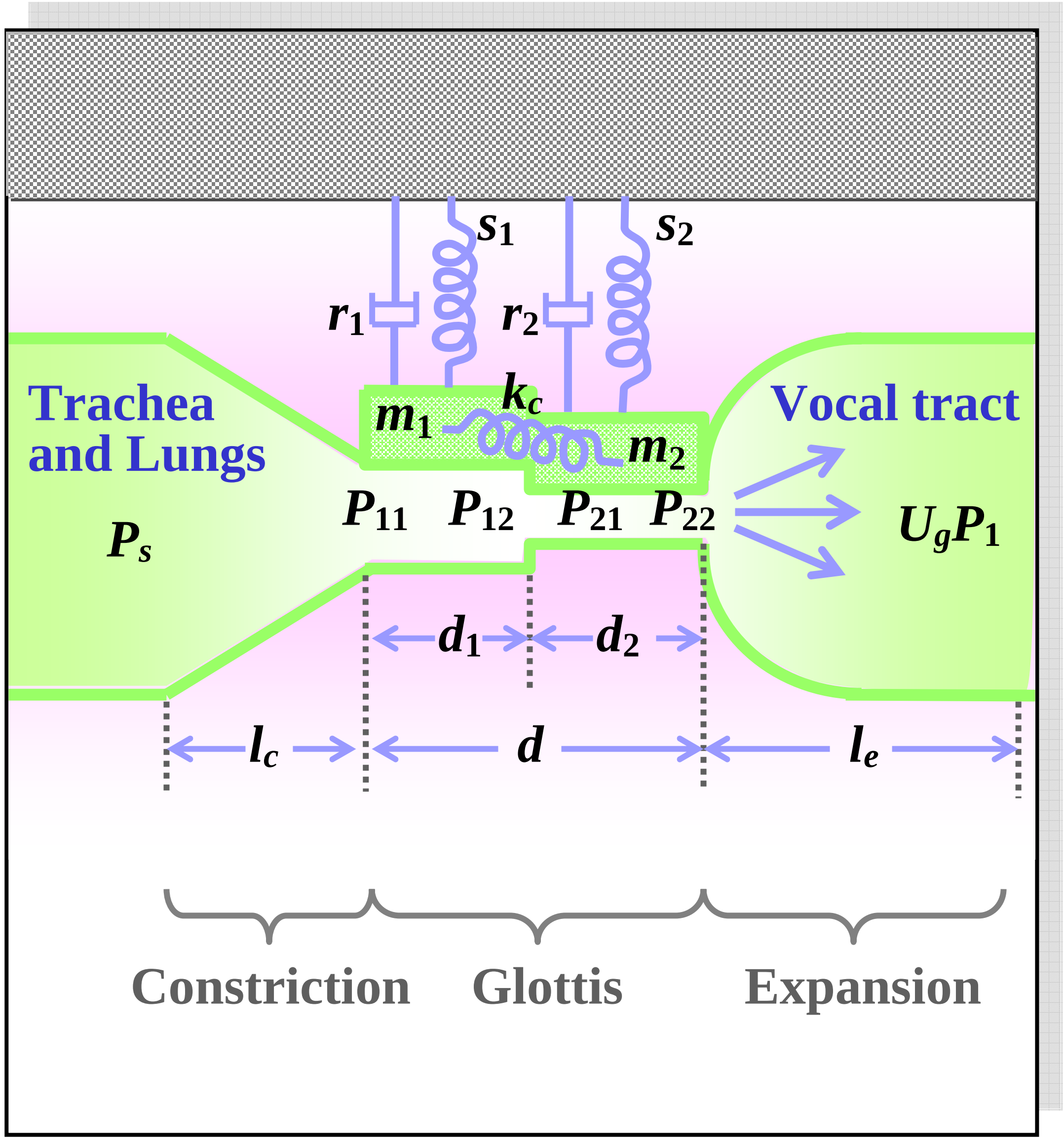


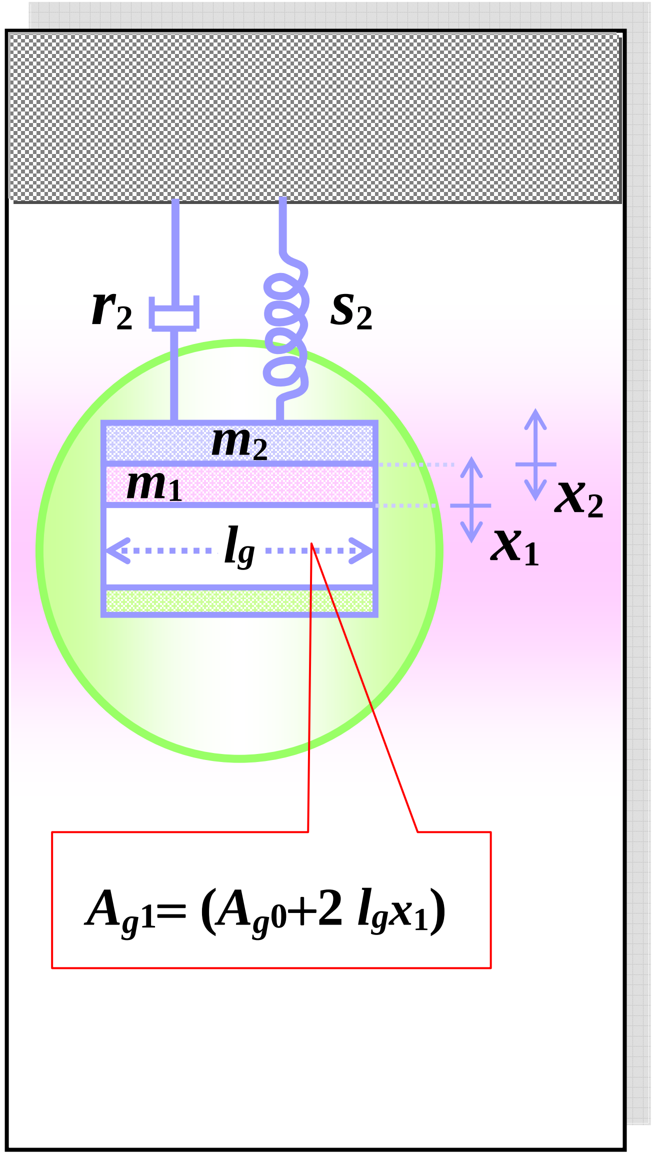
Fundamental frequency distribution over speakers

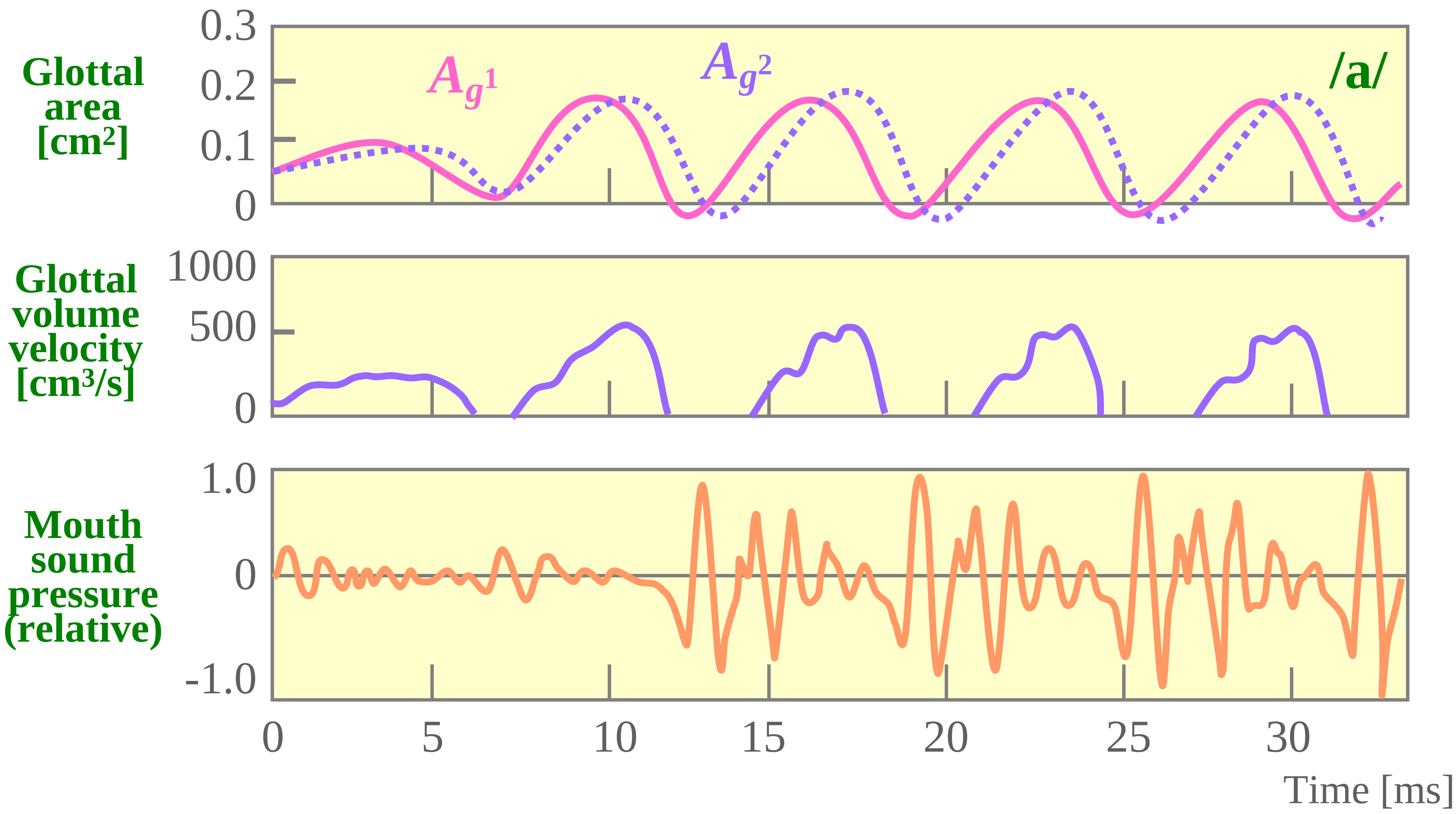


Mean and standard deviation of temporal variation in fundamental frequency during conversational speech for various speakers

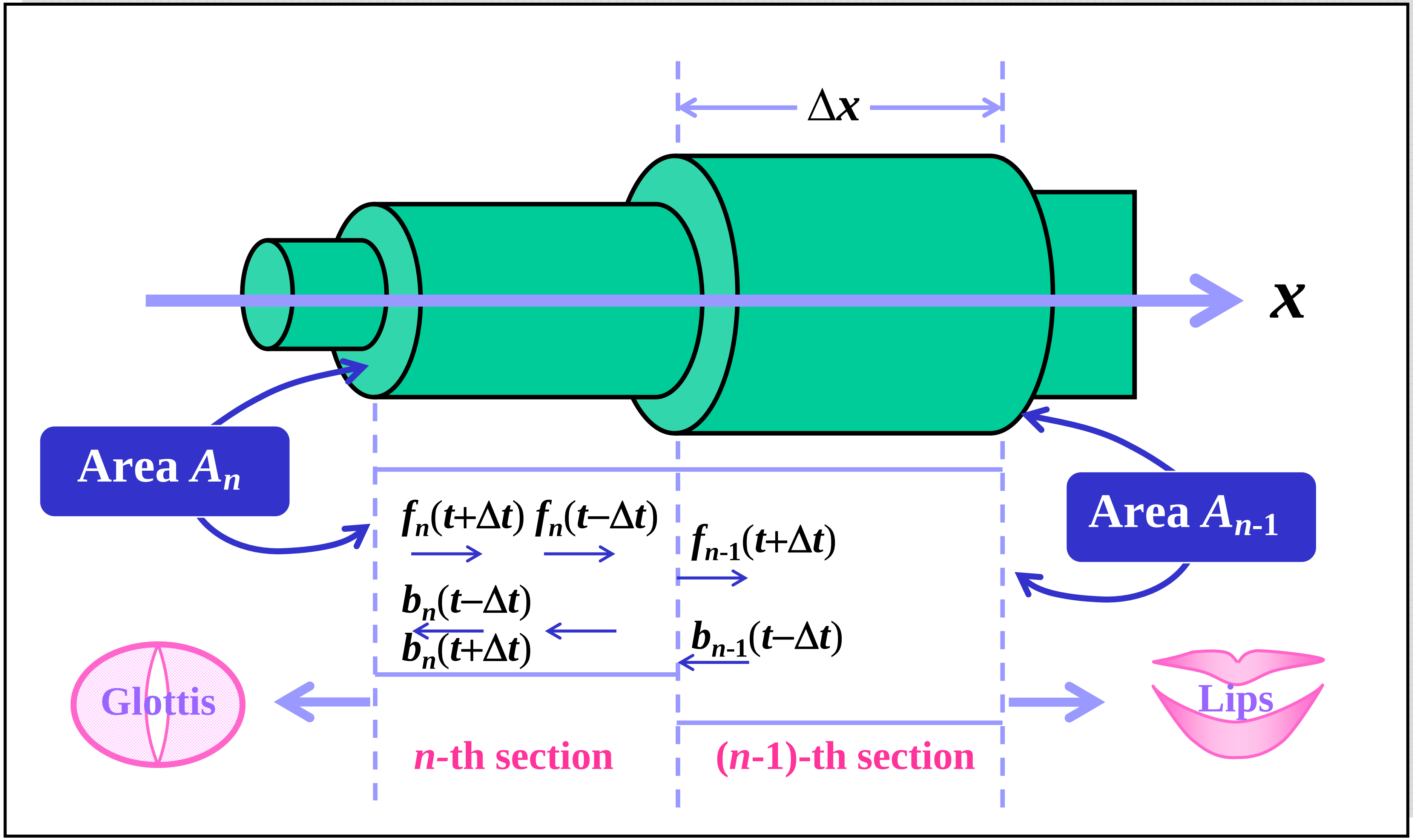
Configuration of two-mass model; cross section of glottis (A_{g1}=area at d_1 section; A_{g0}=area in the neutral state at d_1 section)





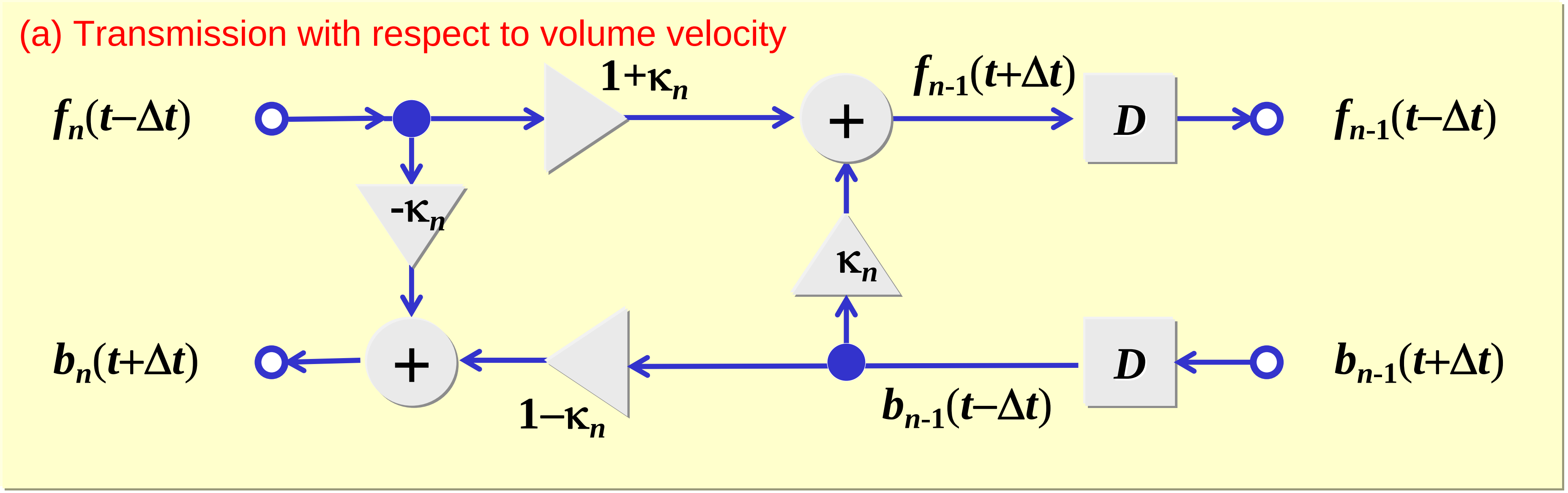


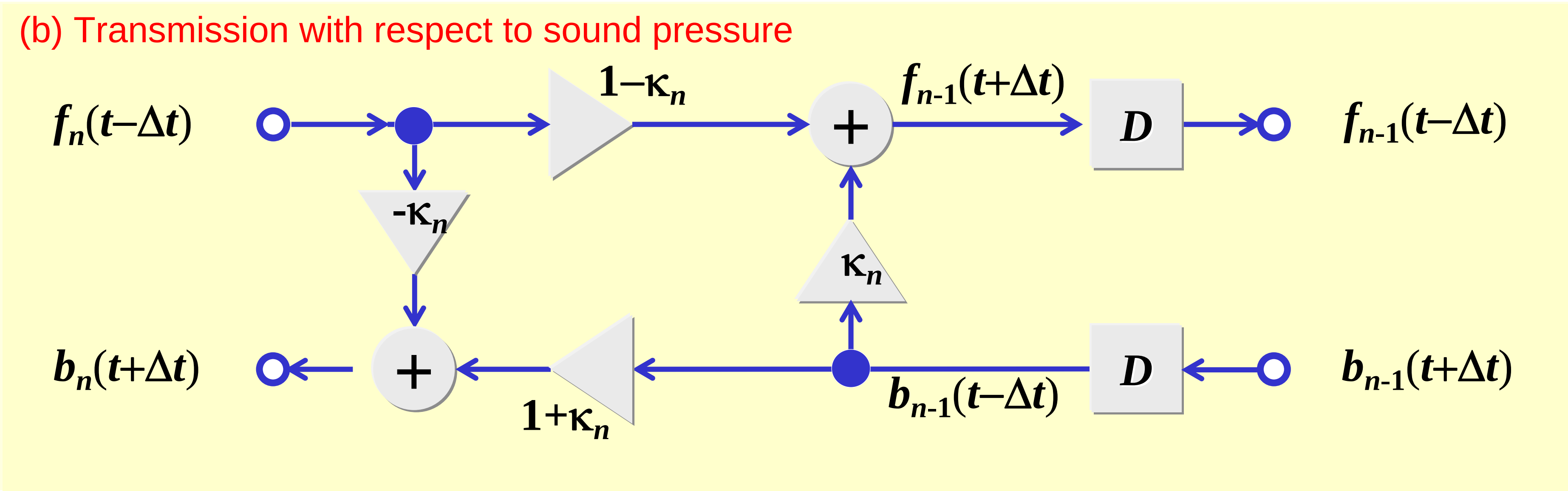
Simulation of speech production for vowel /a/ using the two-mass model



Definition of forward and backward waves with respect to volume velocity at the *n*th cross section, and continuity condition for the volume velocity at the boundary between the (*n*-1)th and *n*th sections

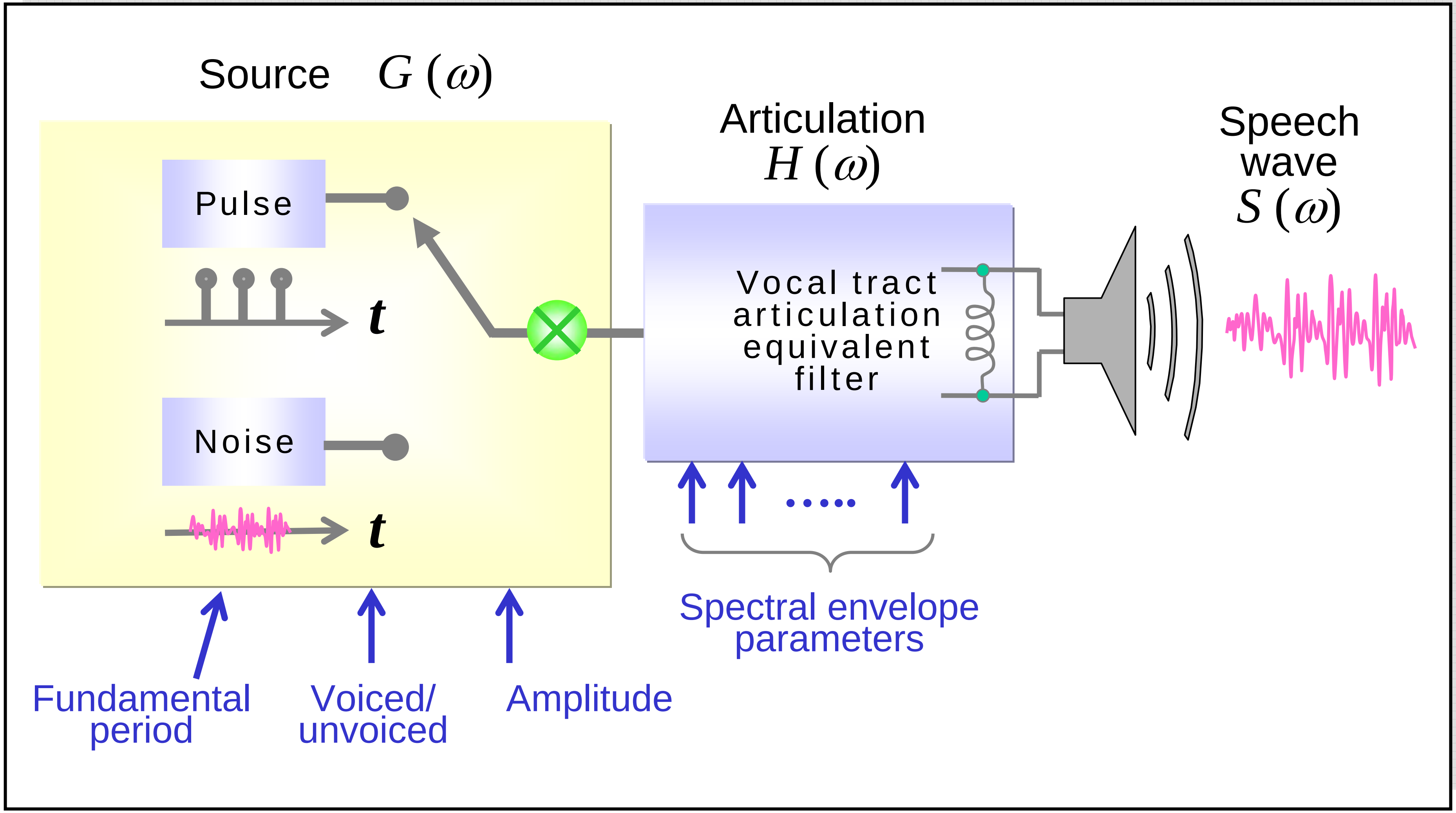
Transmission model of acoustic waves in the vocal tract





(D = time delay of $2\Delta t$)

Linear separable equivalent circuit model of the speech production mechanism



$$S(\omega) = G(\omega) \cdot H(\omega)$$