Interakce člověk–počítač v přirozeném jazyce (ICP)

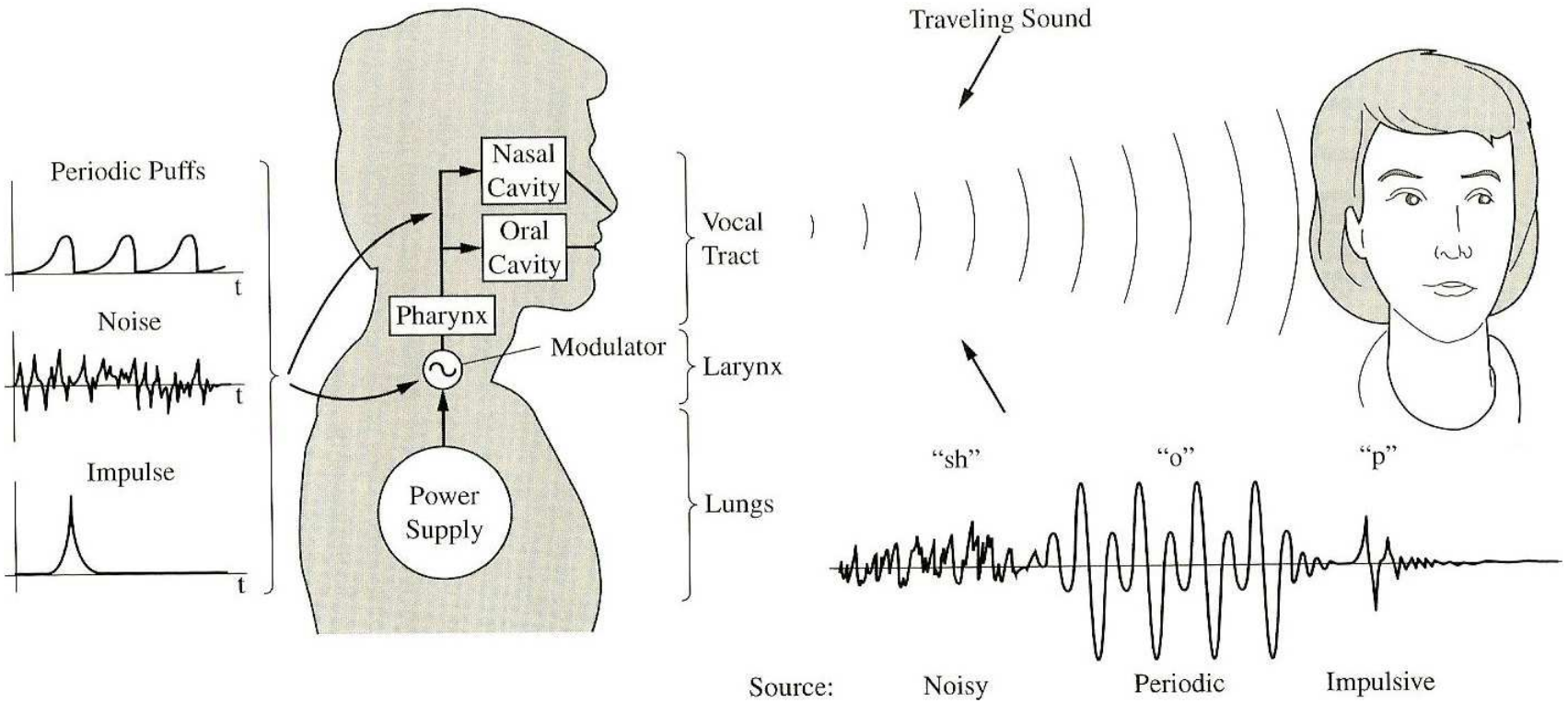
LS 2013 — Fonetické základy

Tino Haderlein, Elmar Nöth

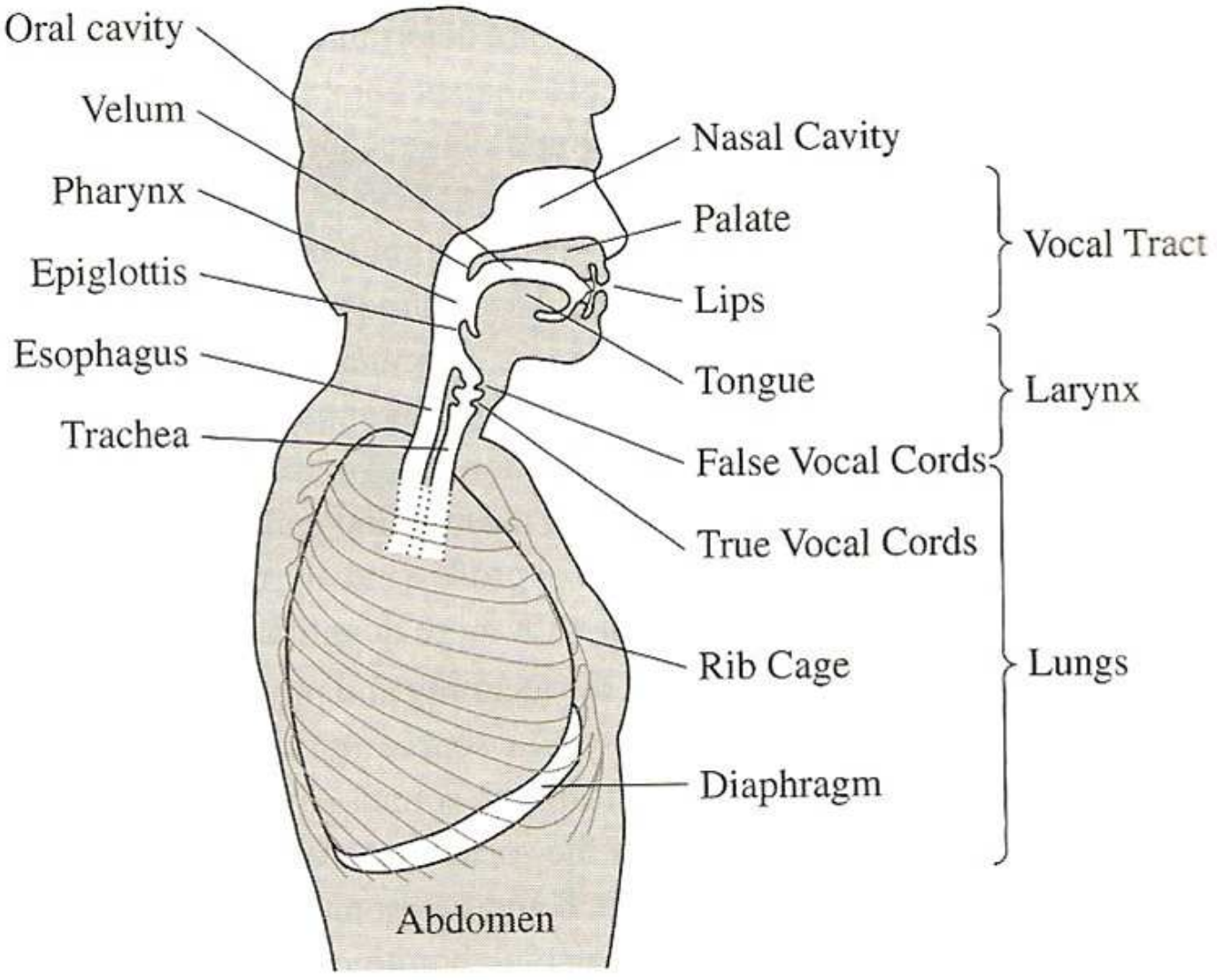
Katedra informatiky a výpočetní techniky (KIV)
Západočeská univerzita v Plzni

Lehrstuhl für Mustererkennung (LME)
Friedrich-Alexander-Universität Erlangen-Nürnberg

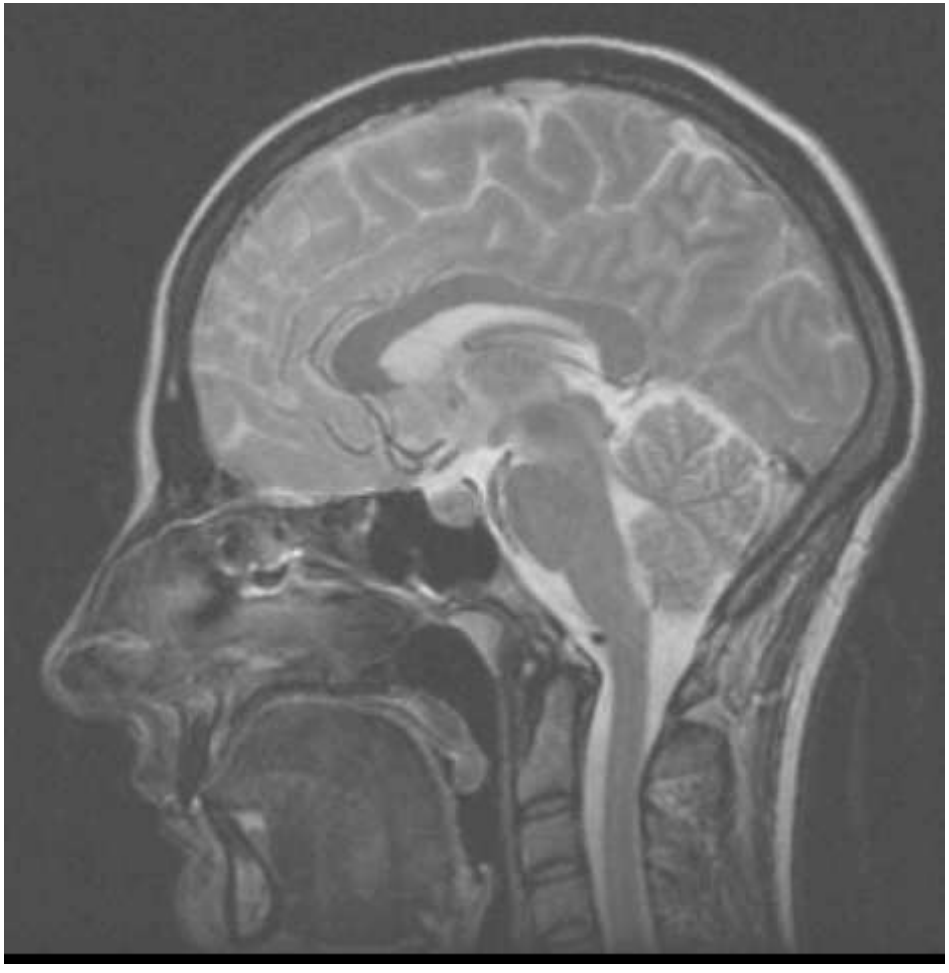
Speech Production



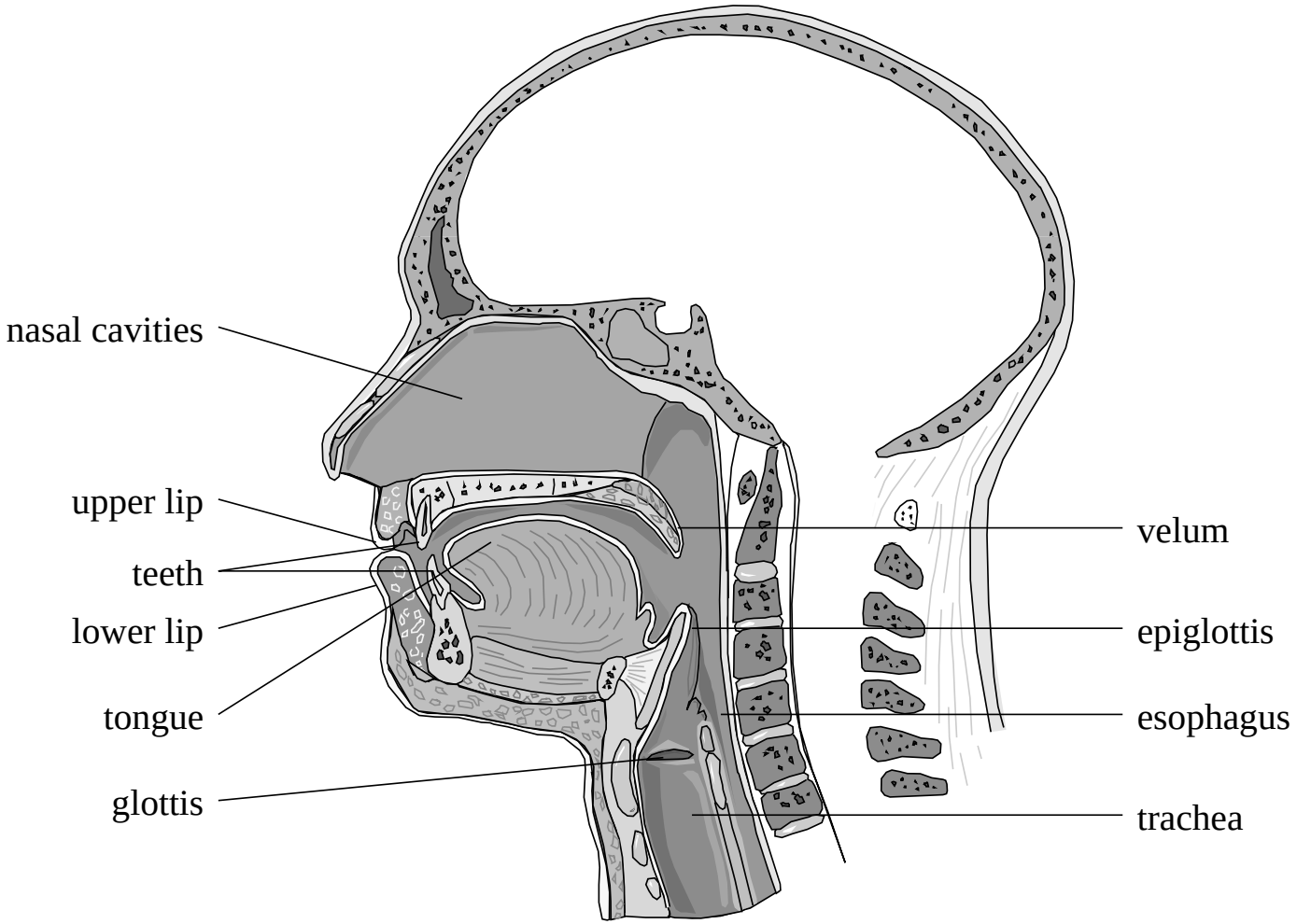
Speech Production



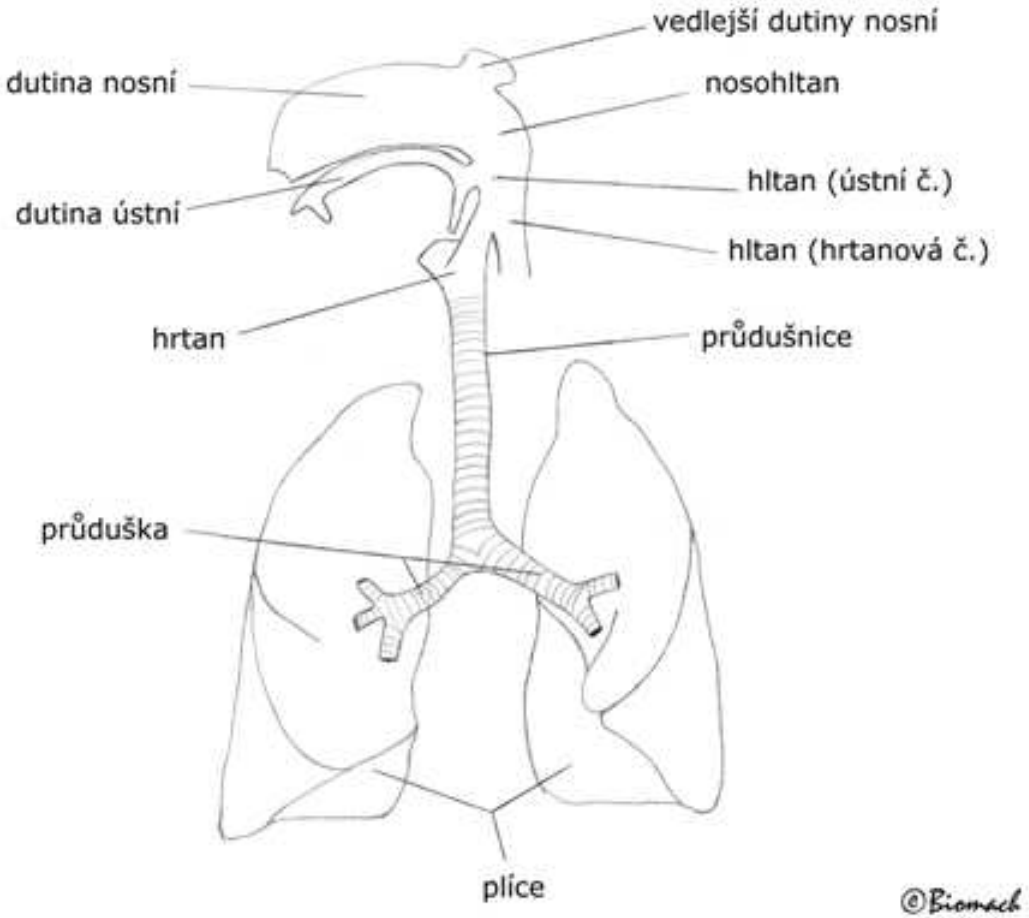
Speech Production



Speech Production

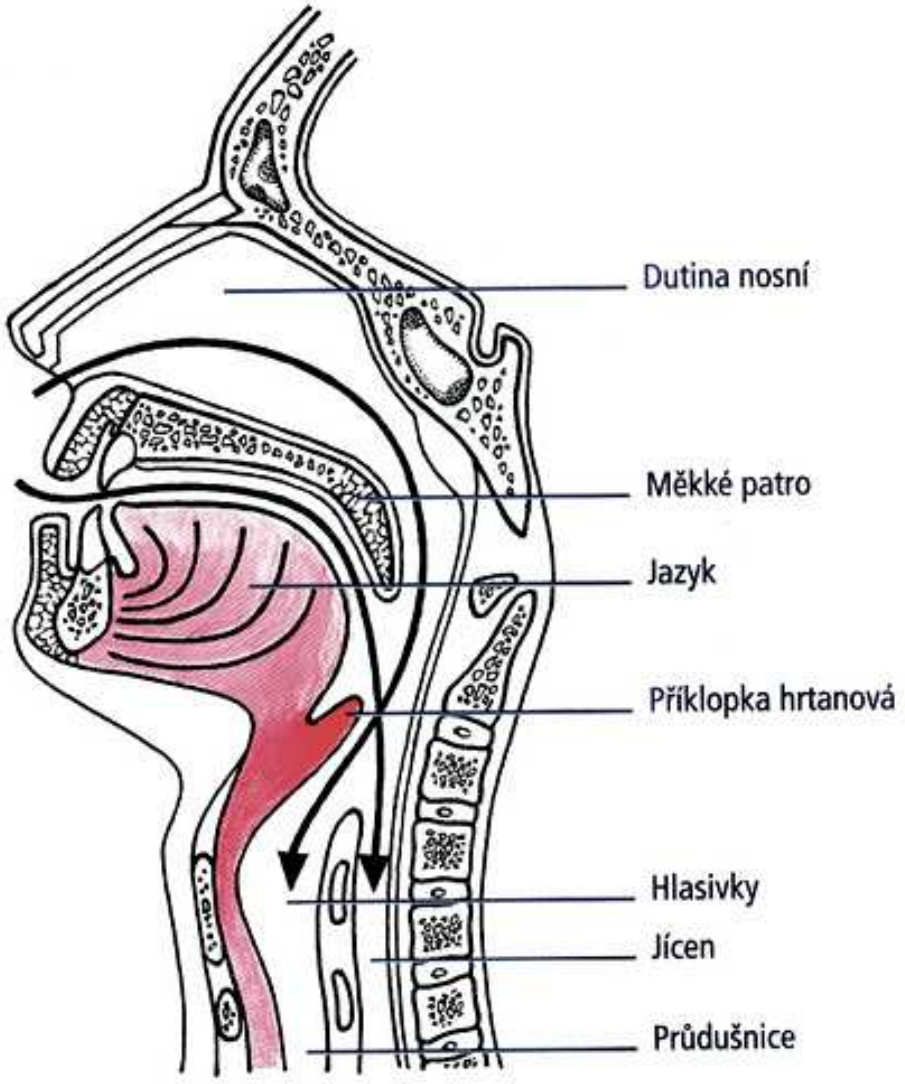


Produkce řeči



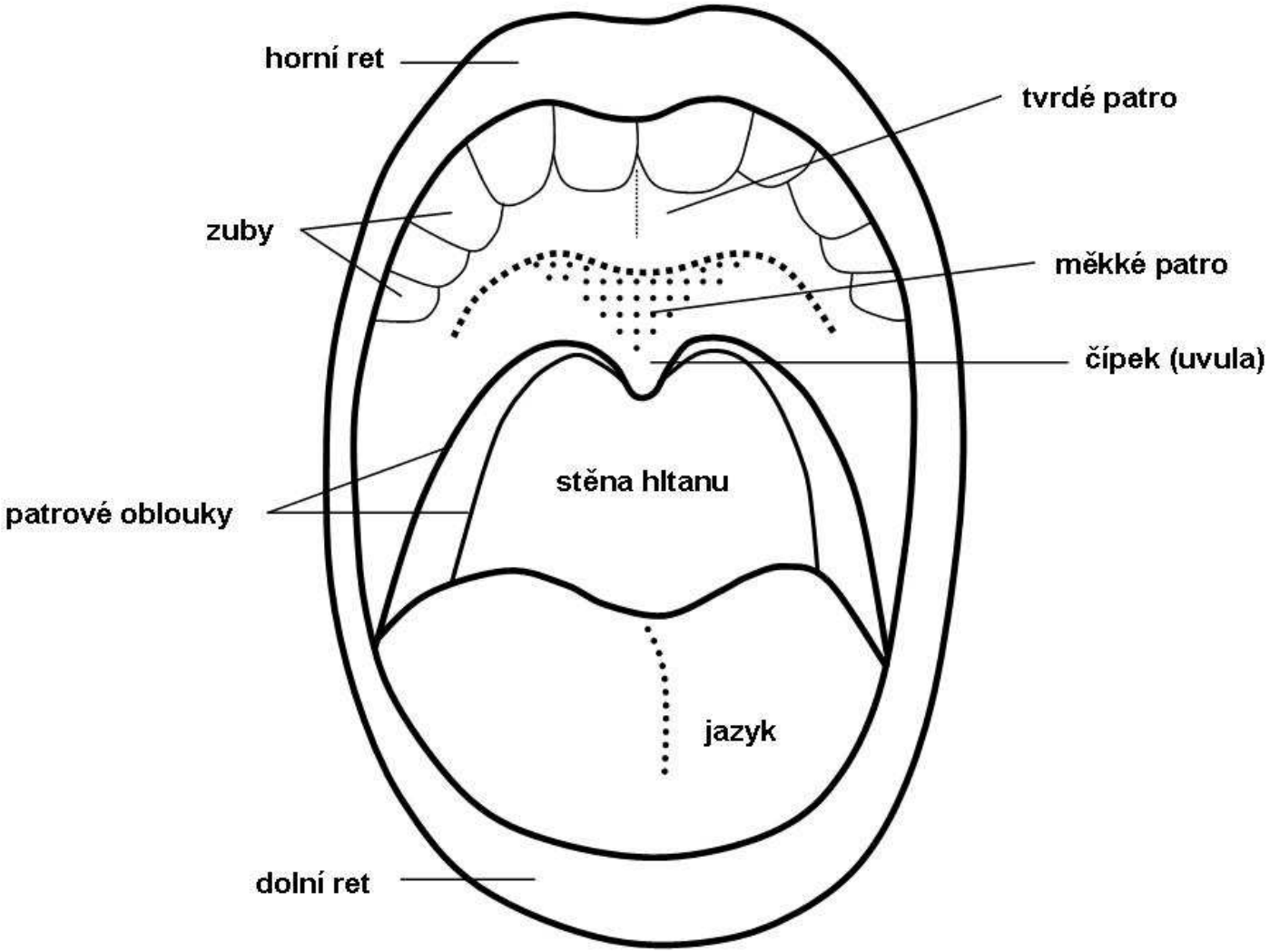
http://www.biomach.cz/_/rsrc/1323465036905/biologie-cloveka/dychaci-system/dychani1F.png

Produkce řeči



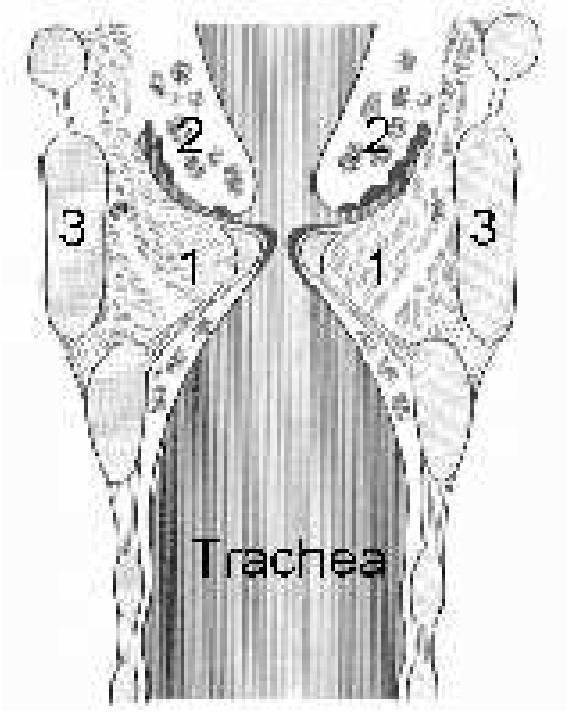
http://www.chatar-chalupar.cz/wp-content/uploads/2010/12/zdraví-23.jpg

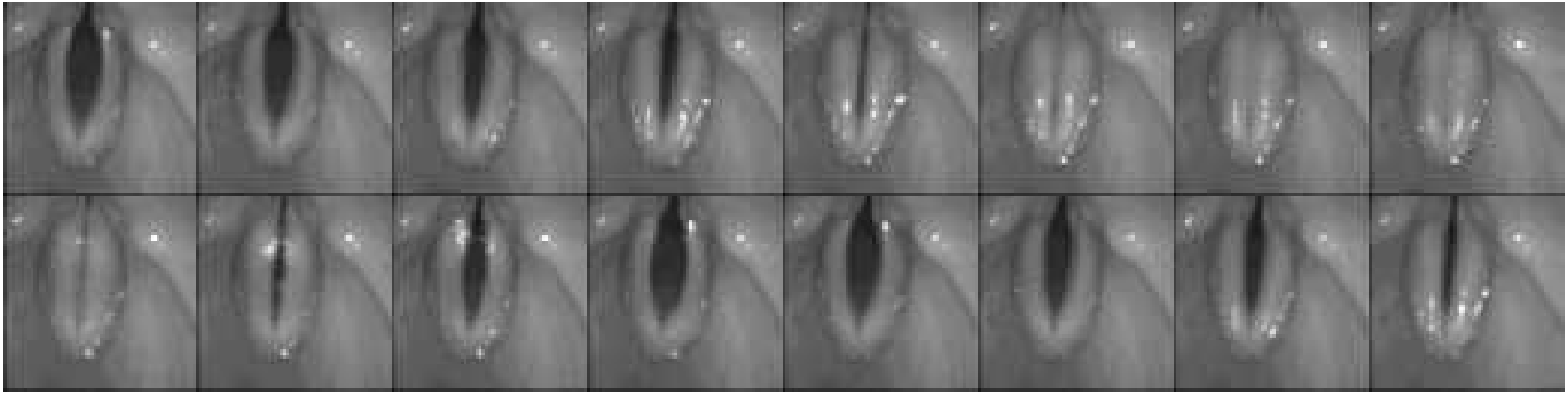
Produkce řeči



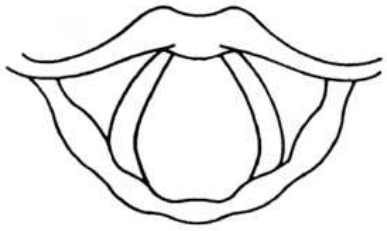
http://pfyziollfup.upol.cz/castwiki/wp-content/uploads/2012/11/Obr19.jpg

In the Beginning there was the Source

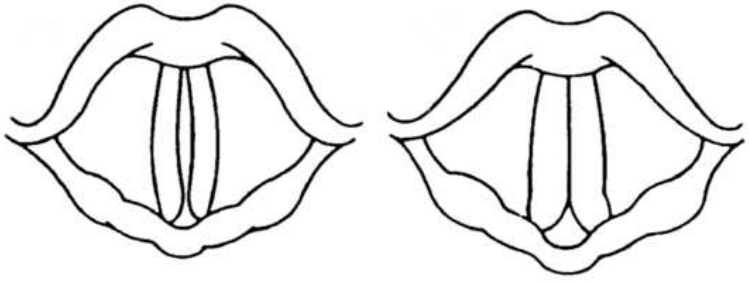




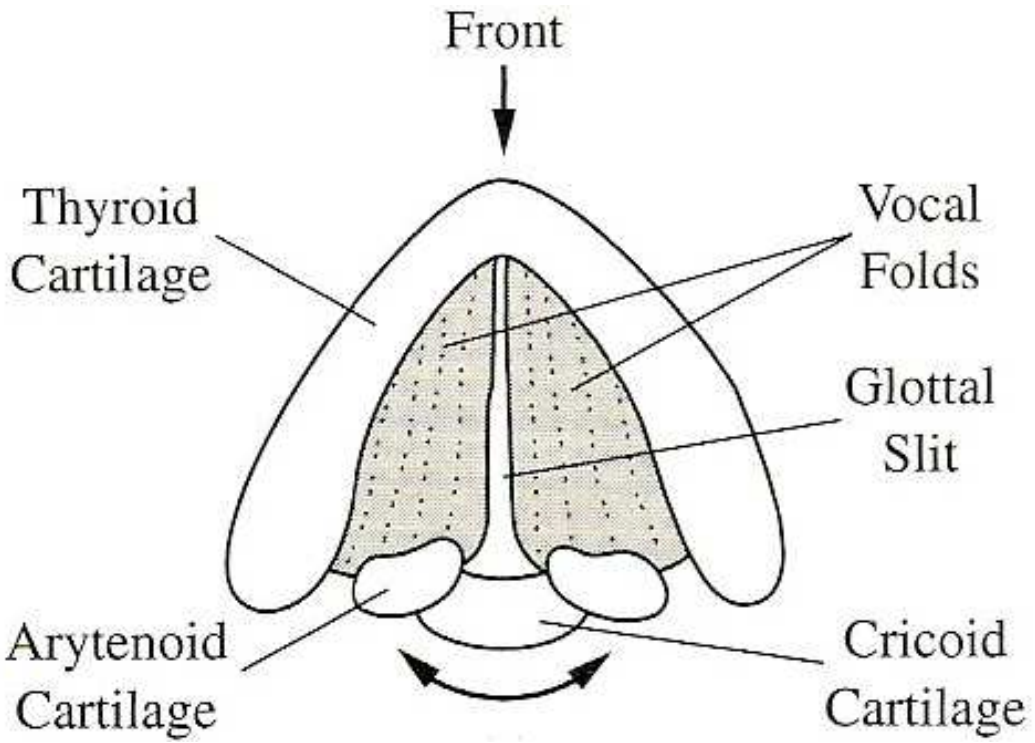
In the Beginning there was the Source

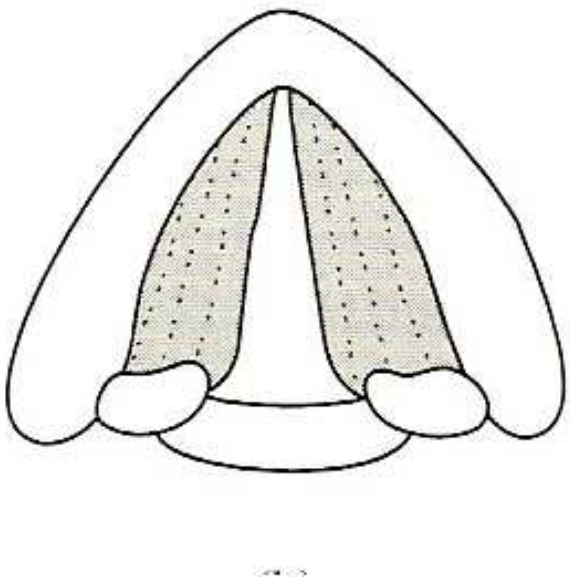


Phonation and Whispering

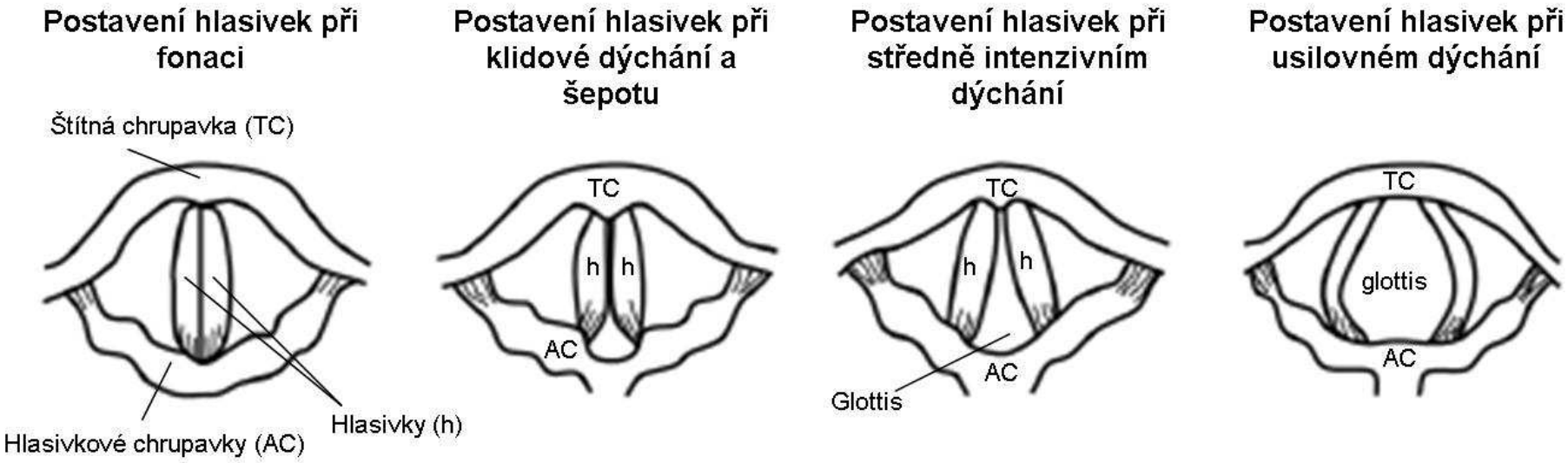


The Source – A Closer Look



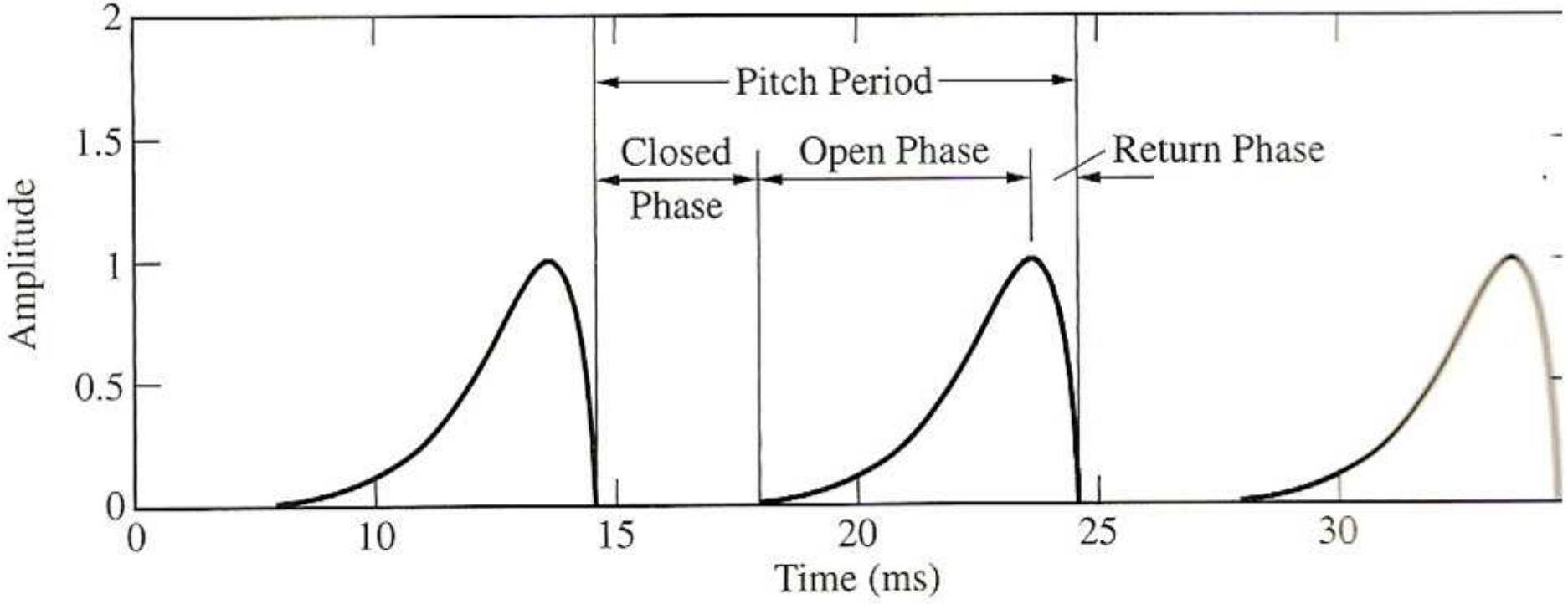


Zdroj hlasu

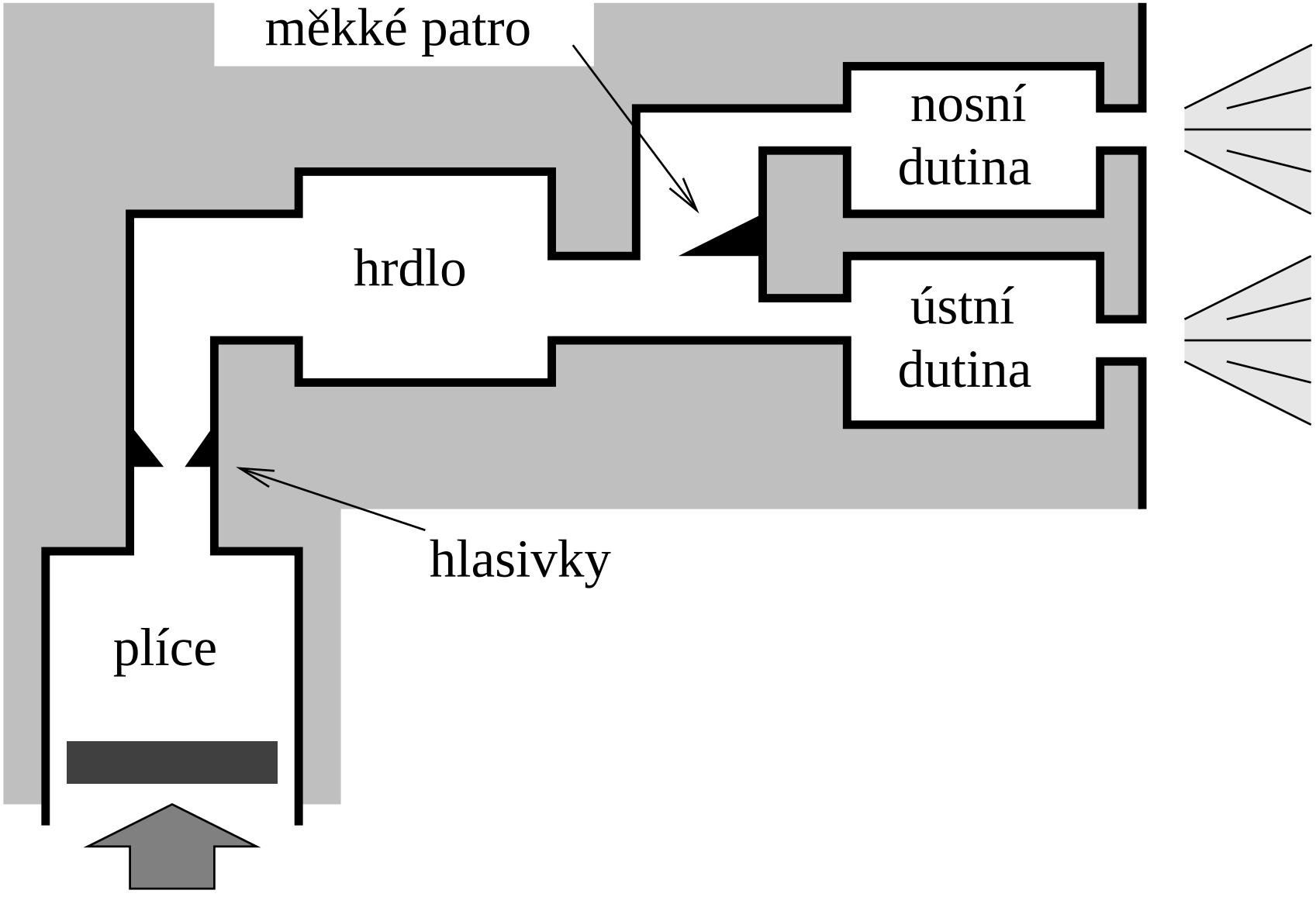


http://pfyziollfup.upol.cz/castwiki/wp-content/uploads/2012/11/Obr15.jpg

Glottal Airflow Velocity



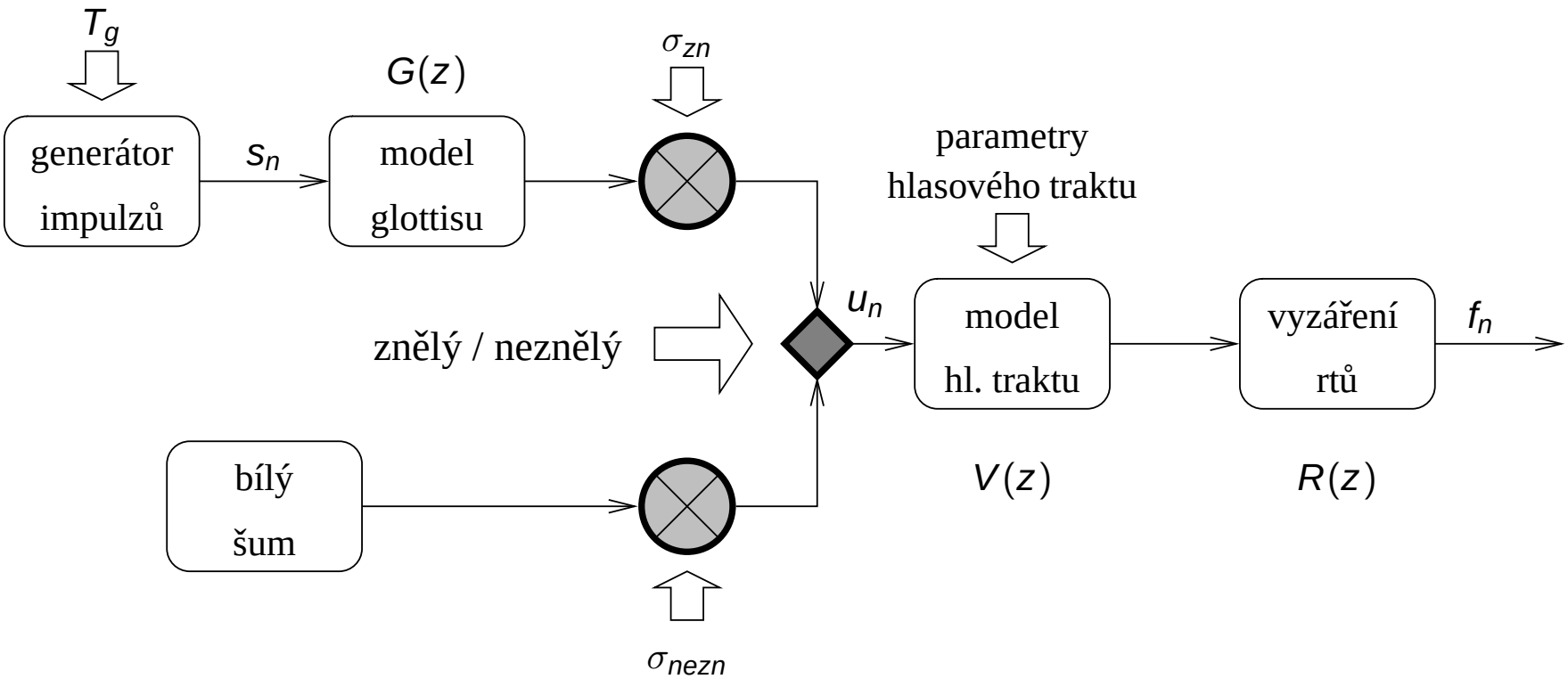
Mechanický model artikulace



Artikulace: Teorie zdroje a filtru

G. Fant (1960): Source–Filter Model

- Signál vznikne sériovým zapojením dvou popř. tří lineárních časově invariantních systémů.
- Artikulátory jsou reprezentovány jako filtry
  $\longrightarrow$ diskrétní signál jako konvoluce $f_n = u_n \star v_n \star r_n$
  z-transformace: $F(z) = U(z) \cdot V(z) \cdot R(z)$



Model výstupného signálu hlasivkové štěrbiny (glottis)

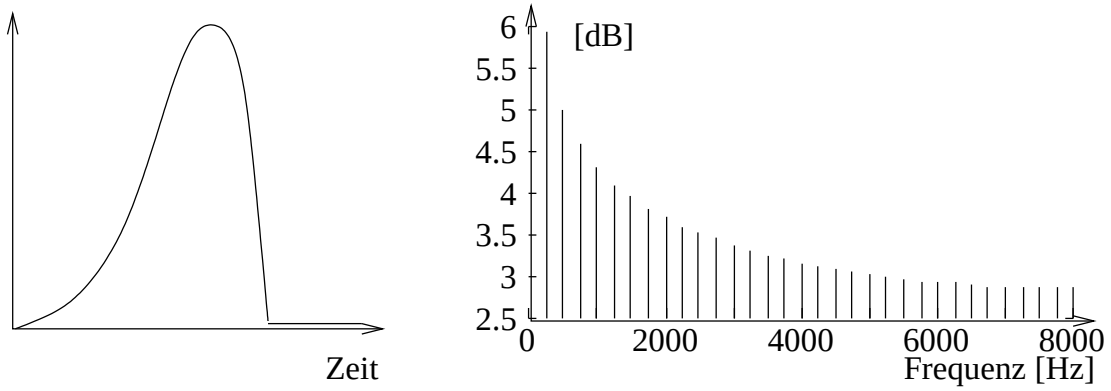
Neznělé hlásky: šumový signál s plochým spektrem

Znělé hlásky: periodický signál 50 až 400 Hz, modelován konvolucí sekvence s_n impulzů s libovolnou dobou kmitu T a průběhem signálu g_n hlasivek

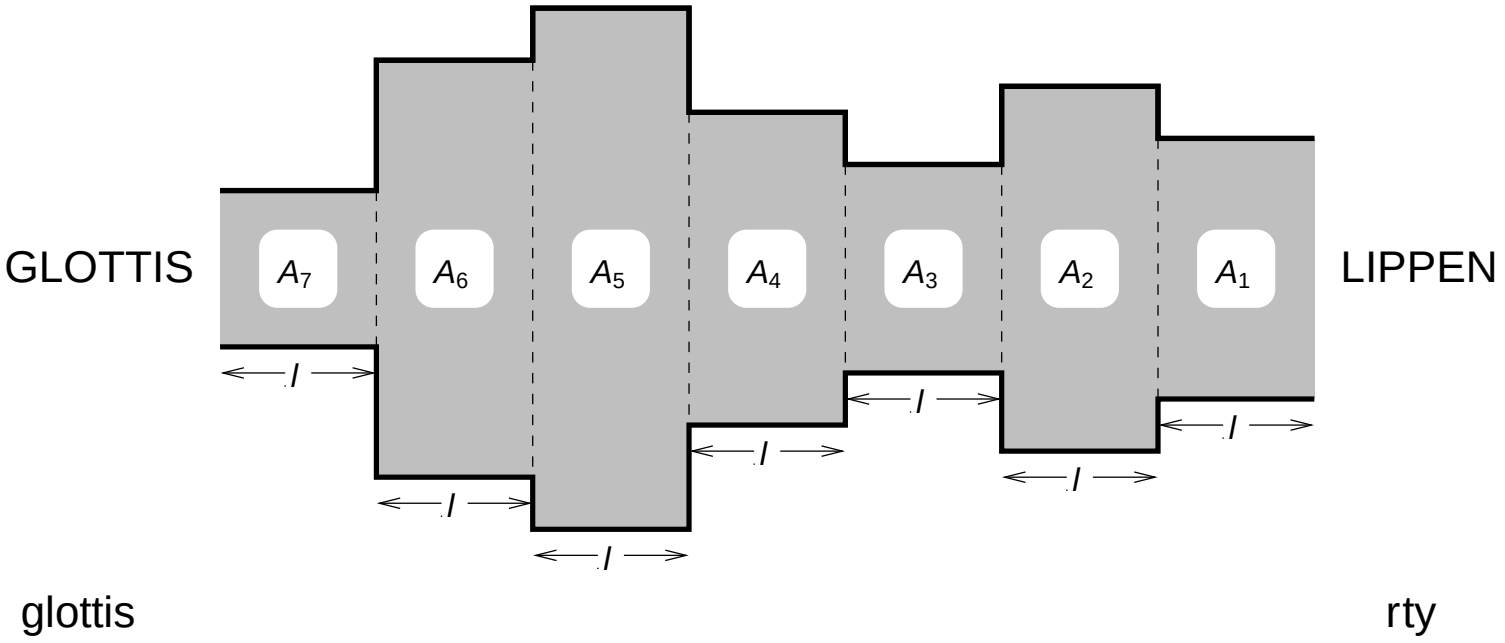
z-rovina: $G(z) = \frac{1}{(1-e^{-cT}z^{-1})^2}$ ($c \ll 1/T$ neměnný)

Cosinus Model A.E. Rosenberga.

Časový signál (vlevo) a frekvenční odpověď (vpravo):



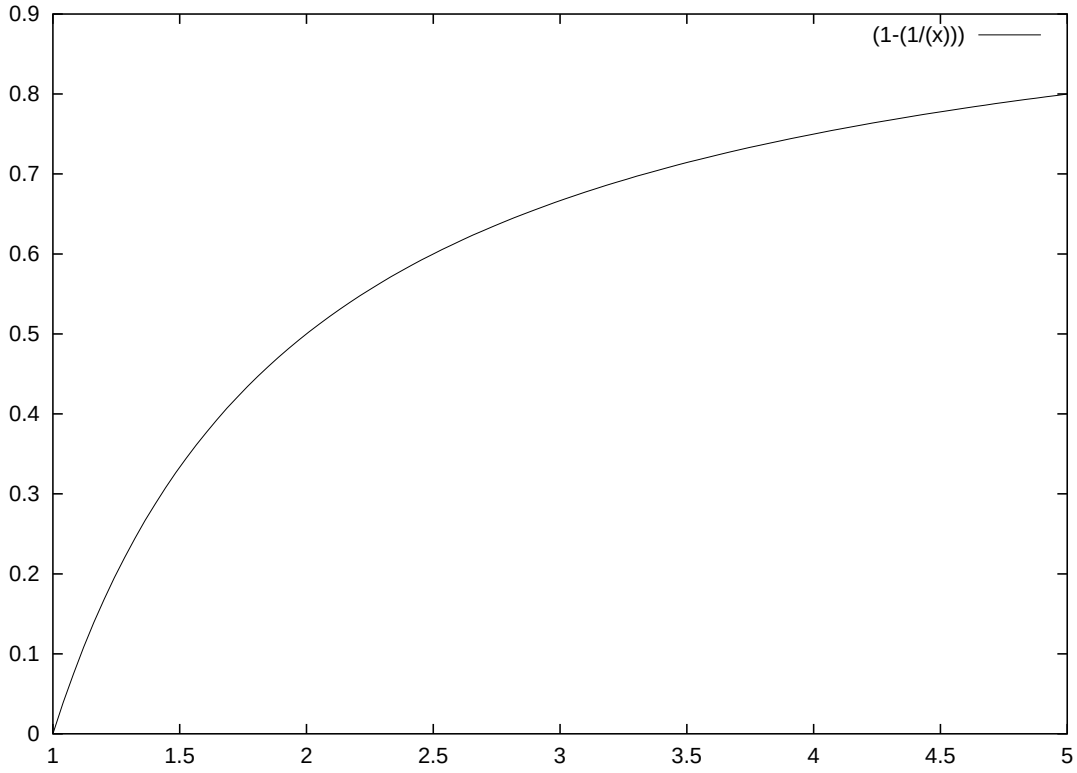
Hlasový trakt



- bezztrátová akustická trubice složená válcovitými úseky stejné délky
- ztráty na kraji hlasového traktu a v nosní dutině se zanedbávají
- **nástavná trubka** (ústní dutina a hrdlo) modelována jako **akustická trubice** délky L (typicky 170 mm) z M **válcovitých úseků** stejné délky s průřezovou plochou A_i, $i = 1, \ldots, M$
- resonance hlasového traktu $=$ **formant**

Rty

Výstup tlakové vlny přes malý otvor ve velmi velké ozvučnici idealizované: horní propust $R(z) = R_0(1 - z^{-1})$



Autoregresivní model

celková přenosní funkce pro znělé hlásky:

$$H(z) = \sigma \cdot G(z) \cdot V(z) \cdot R(z)$$

zjednodušení:

$$H(z) \simeq \sigma / A(z)$$

kde $A(z)$ je polynom $1 - \sum_{j=1}^{p} a_i z^{-i}$ proměnné z^{-1}
a σ je faktor zesílení
allpole systems, popř. **autoregresivní systémy** (kvůli vlastnostem v časové oblasti; jen závislé na předchozím outputu, ne na inputu)
$\rightarrow$ model tvoření hlasu: $F(z) = S(z) \cdot H(z) = S(z) \cdot \sigma / A(z)$
$A(z)$ známé $\rightarrow$ prvotní signál se může získat zpět **inverzní filtrací**
$S(z) = F(z) \cdot A(z)/\sigma$ řečového signálu.

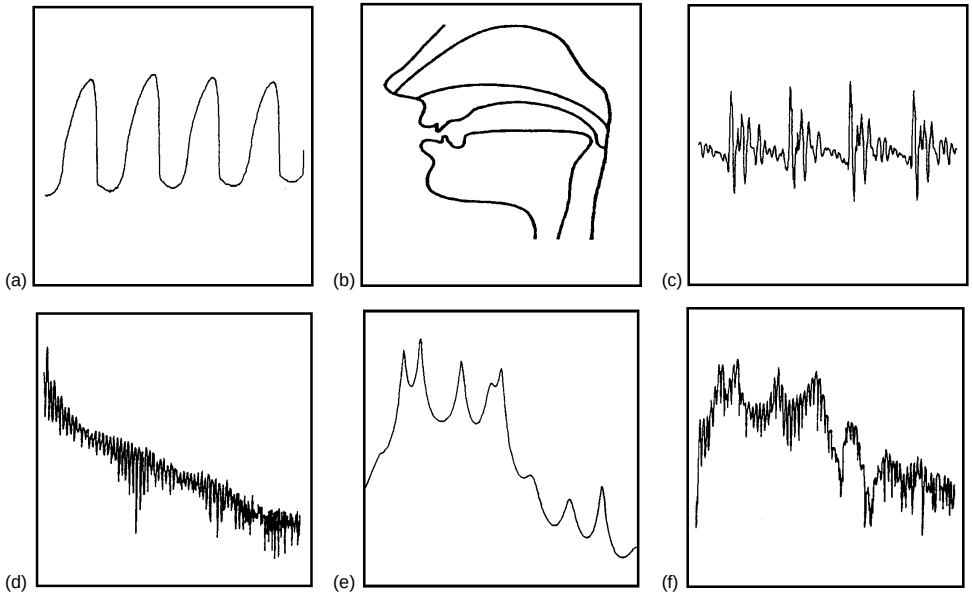
Autoregresivní model: znázornění

Vztahy mezi komponenty modelu na tvoření hlásek:
(a) měřený prvotní signál
(d) jeho výkonové spektrum (logaritmováno)

(b) schematizovaný tvar hlasového traktu
(e) frekvenční přenos autoregresivního modelu pro přenosní funkci hlasového traktu (logaritmován)

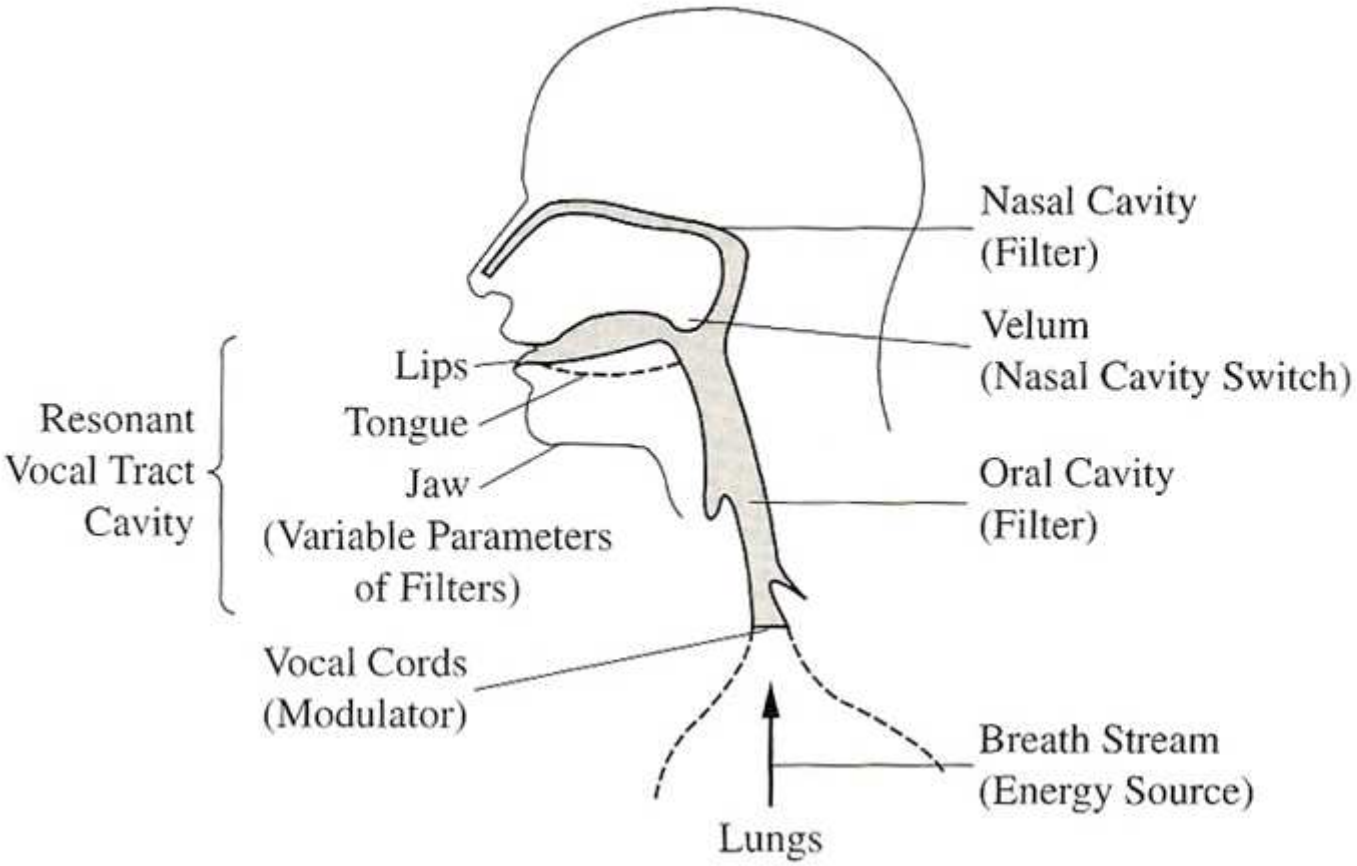
(c) časový signál vytvořené zvukové vlny

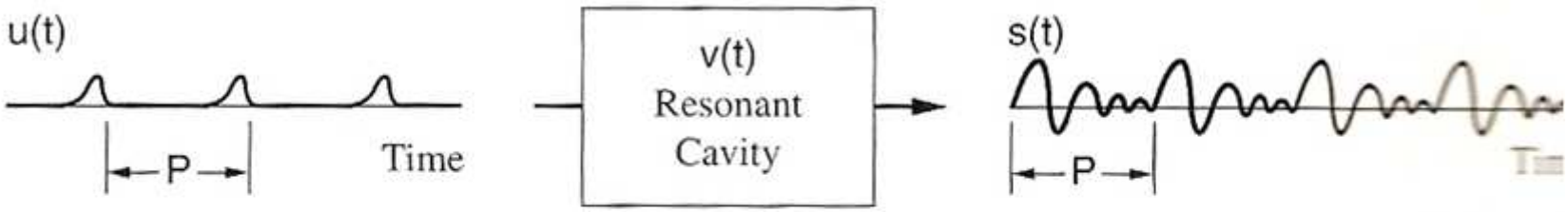
(f) krátkodobé spektrum (logaritmováno) vznikne z (d)+(e)



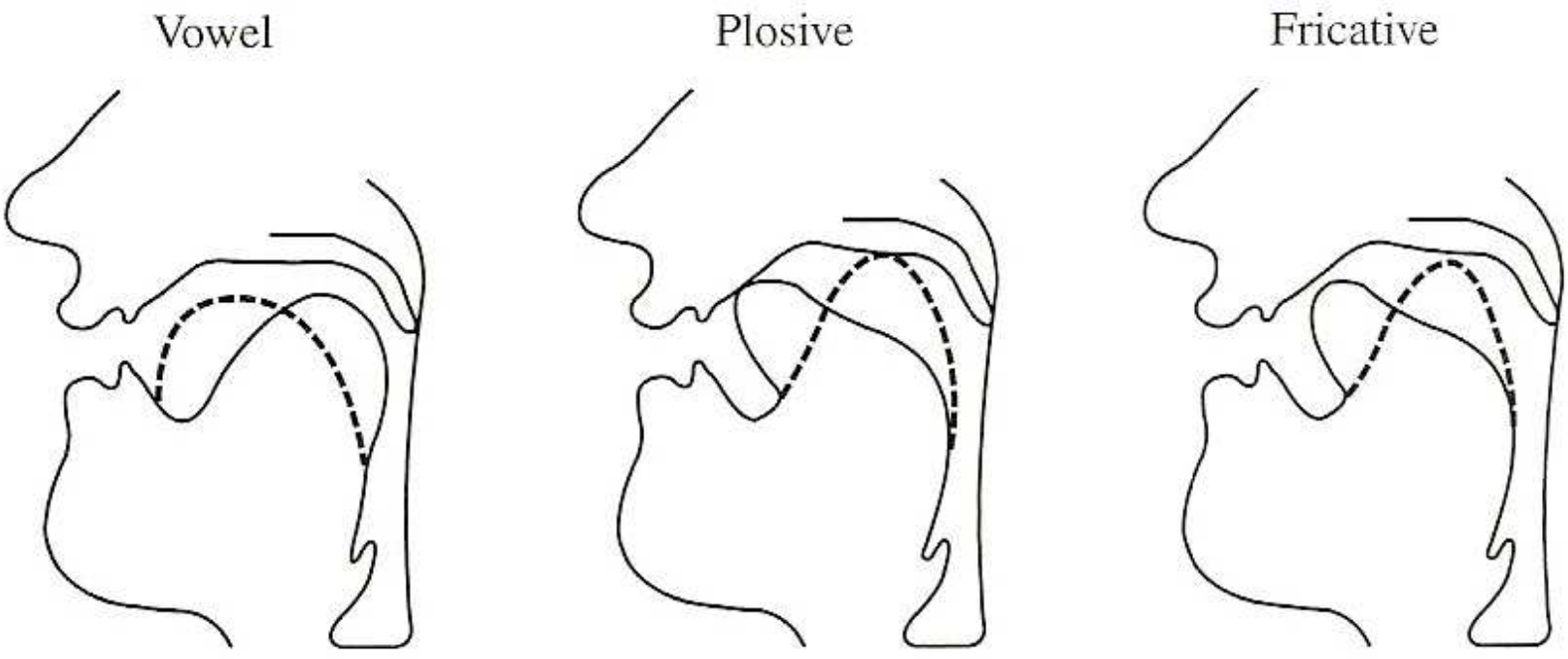
produkce hlásky [a]

Shaping of the Spectrum by a Resonator

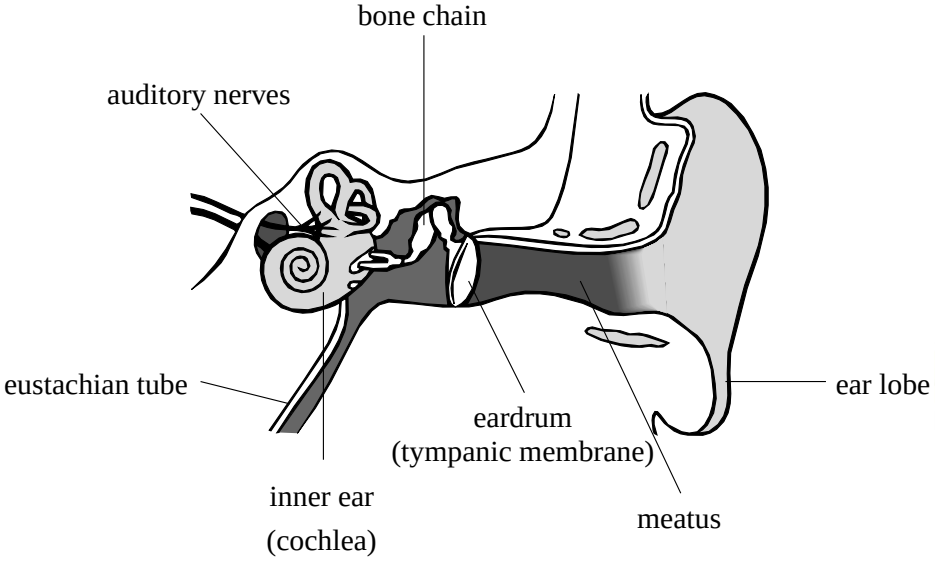


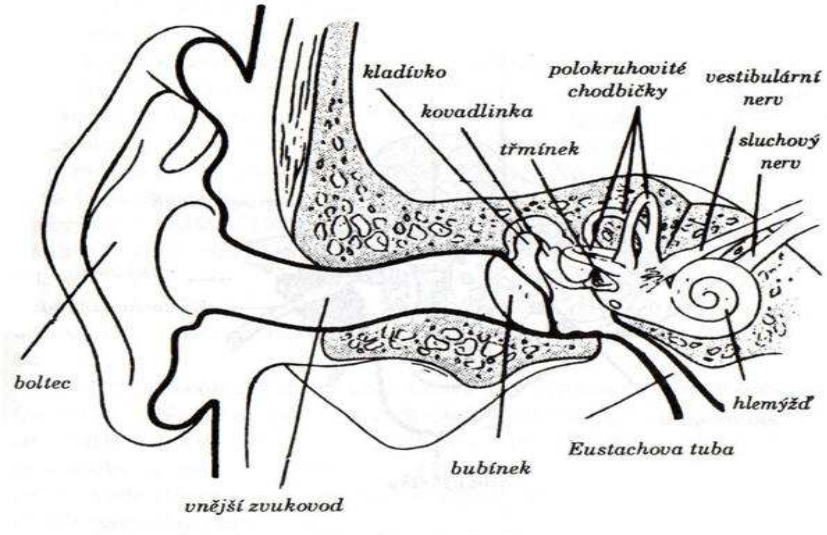


Shaping of the Spectrum by a Resonator



Percepce

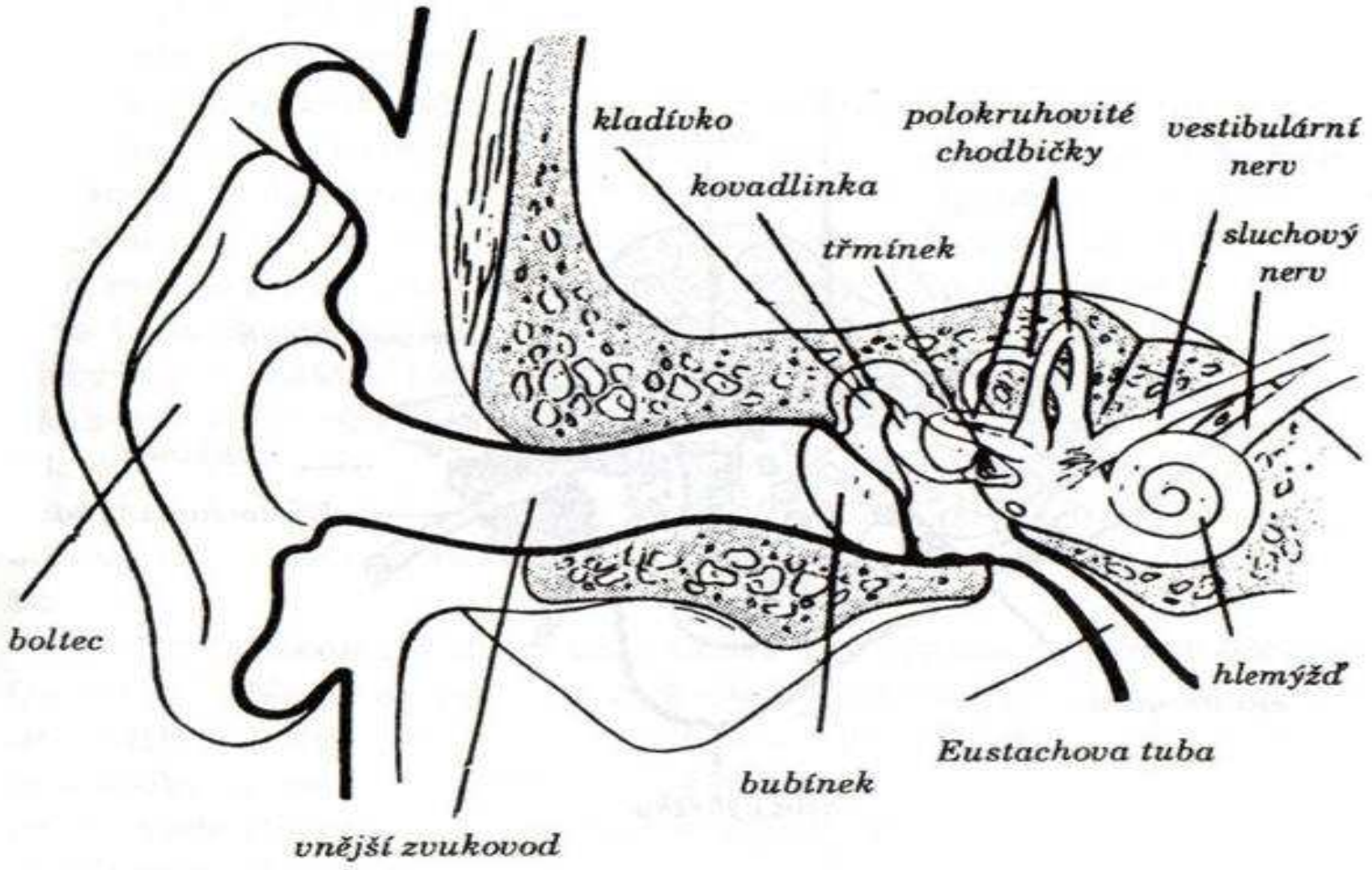




http://tichysvet.wz.cz/images/vnitrn1.jpg

- sluchový rozsah: 16 Hz – 20 kHz, nejvyšší citlivost na zvuk kolem 3 – 4 kHz
- ušní boltec zachytí zvuk – přes vnější zvukovod k bubínku přes kůstky na oválné okno vnitřního ucha (plně nestlačitelné lymfy)
- **hlemýžď** (**cochlea**, nejdůležitější část vnitřního ucha) vede od oválného okna ke sluchovému nervu: spirálovitá, zúžující se trubice, kolem 30 mm dlouhá, 2 1/2 závidů; dvě dělicí příčky **Reissnerova membrána** a **basilární membrána**

Percepce

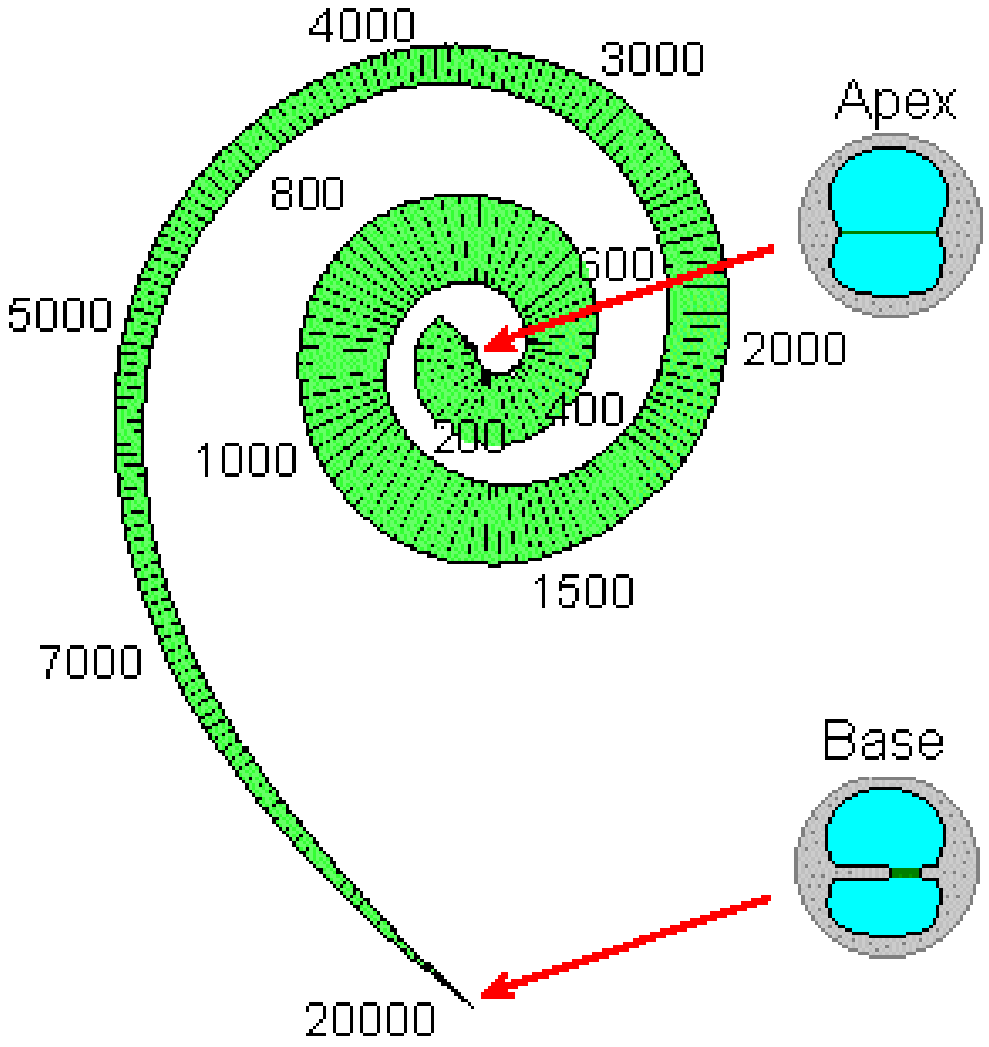


http://tichysvet.wz.cz/images/vnitrn1.jpg

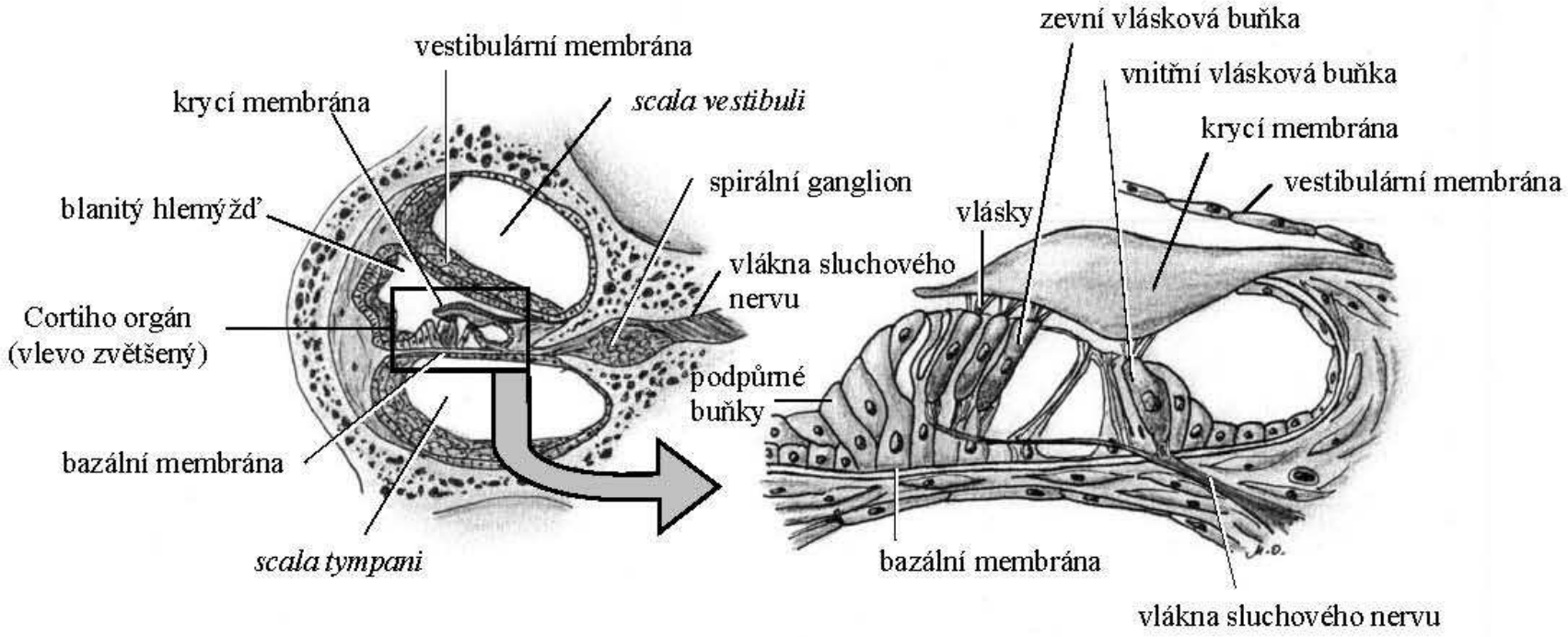
Hlemýžď (Cochlea)

- Kmitání basilární membrány kmitáním oválného okénka, amplituda je závislá na frekvenci, místo nejsilnější výchylky je závislé na frekvenčních komponentech.
- Místa na membráně jsou na různé frekvence různě citlivá.
- Citlivost na vysoké zvuky klesá rychleji než na hluboké zvuky.
- mechanické kmitání $\Rightarrow$ neuronální výboj v Cortiho orgánu
- Tam začínají vlákna sluchového nervu.
- Kódování fáze zvuku, popř. “phase locking”: Buňky nervu se synchronizují na sinusovou periodu prvotního signálu (až kolem 5 kHz).

Hlemýžď (Cochlea)



Cortiho orgán



Blanitý hlemýžď s Cortiho orgánem

http://pfyziollfup.upol.cz/castwiki/wp-content/uploads/2011/08/obr-08.jpg

Vnímání hlasitosti

Fyzikální síla zvukového signálu vyjádřena

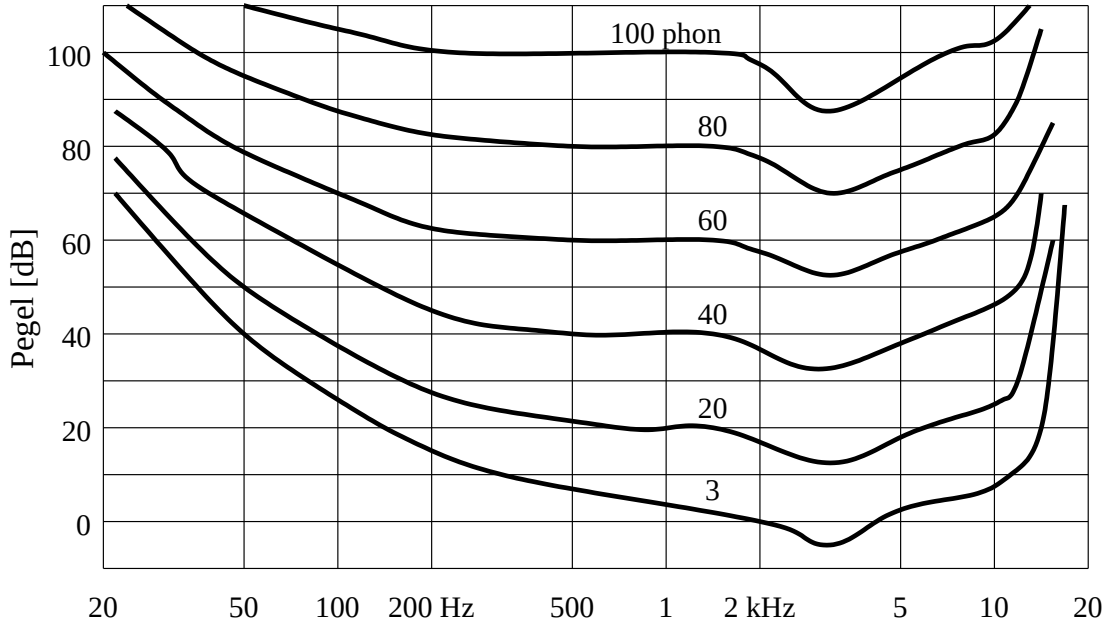
- jako akustický tlak p_s v pascalech (Pa)
- nebo jako intenzita I_s v $\mathrm{N/m^2}$

- I_s je proporcionální k p_s^2
- referenční veličiny: $p_0 = 2 \cdot 10^{-5}\,\mathrm{Pa};\ I_0 = 10^{-12}\,\mathrm{N/m^2}$
- definice: hladina akustického tlaku [dB] $= 20 \cdot \log \frac{p_s}{p_0} = 10 \cdot \log \frac{I_s}{I_0}$
- zdvojení intenzity $\rightarrow$ narůst hladiny o 3 dB

Vnímání hlasitosti

čáry stejné hlasitosti v sluchovém poli = izofony

- Práh slyšení je závislý na frekvenci.
- Zvukové vlny se stejnou hladinou, ale s různými frekvencemi **ne**vnímáme jako stejně hlasité.
- psychoakustická míra lidského vnímání hlasitosti: **phon**
- Hodnota míry odpovídá hladině v dB zvuku frekvence 1 kHz, který vnímáme jako stejně hlasitý (osa y dole: hladina).



Subjektivně vnímaná hlasitost

- psychoakustická míra
- zohlednění vnímaného poměru hlasitostí dvou zvuků
- vnímaná hlasitost zvuku s 1 kHz a 40 phon $=$ 1 son(e)
- L-krát tak nahlas vnímaný zvuk $\rightarrow$ L sone
- vnímaná hlasitost proporcionalní k $I^{0,3}$
- rozdíl hlasitosti dvou zvuků je pro hlasité zvuky vnímán snadněji: nutný rozdíl při 40 phon je 0,7 dB, při 80 phon 0,3 dB

Vnímání frekvencí

- Frekvence jsou na basilární membráně kódované místem stimulace.
- těsně sousedící frekvence $\rightarrow$ křížící se oblasti na membráně
- Příklad: Zvuk je skládán dvěma tóny s frekvencí F_a a F_b: střední frekvence $F_g = \frac{F_a + F_b}{2}$
- Kritická šířka pásma (critical bandwidth) je stanovitelná pro každou střední frekvenci F_g.
- Když $|F_a - F_b| <$ kritická šířka pásma, F_a a F_b jsou ve stejné frekvenční skupině.
- Vnímání skladaných zvuků je zavislé na tom, zda jsou komponenty ve stejné nebo v různých frekvenčních skupinách.
- Kritická šířka pásma je většinou nezávislá na hlasitosti a odpovídá pro všechny F_g kolem 1,2 mm na basilární membráně (kolem 1300 receptorových buněk).

Frekvenční skupiny

Psychoakustické fenomény uvnitř jedné frekvenční skupině:

- Čisté tóny jsou maskované úzkopásmovým šumem.
- citlivost k přesunu fáze
- integrace zvukové intenzity zvuků nebo úzkopásmový šum

- Kritická šířka pásma (critical bandwidth) Δf_{CB} frekvenční skupiny je funkce střední frekvence F_g.
- Δf_{CB} je kolem 100 Hz, když $F_g \leq 1000$ Hz; pro vyšší frekvence kolem $0{,}15 \cdot F_g$.
- Pásmo vnímatelných frekvencí se dá rozdělit do 24 nekřížících se frekvenčních skupin s kritickou šířkou pásma
  $\Rightarrow$ Barkova stupnice, definovaná podle vlastností vnímání frekvencí.

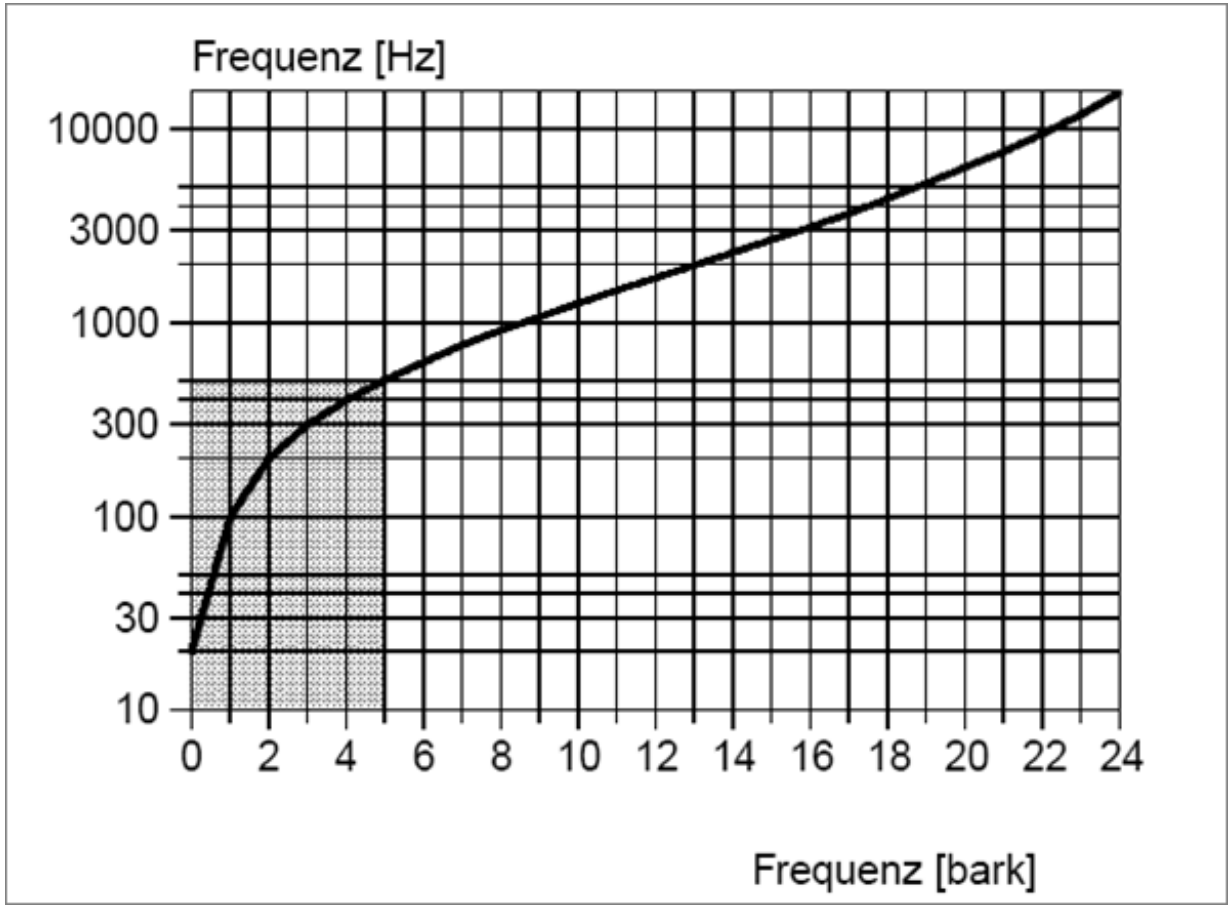
Barkova stupnice 1

Každá střední frekvence má specifickou kritickou šířku pásma.

Každá frekvence patří přesně k jedné frekvenční skupině.

Center frequ. Hz	Crit. bandwidth Hz	Frequency Hz	CB-rate bark
50	80	20	0
150	100	100	1
250	100	200	2
350	100	300	3
450	110	400	4
570	120	510	5
700	140	630	6
840	150	770	7
1000	160	920	8
1170	190	1080	9
1370	210	1270	10
1600	240	1480	11
1850	280	1720	12
2150	320	2000	13
2500	380	2320	14
2900	450	2700	15
3400	550	3150	16
4000	700	3700	17
4800	900	4400	18
5800	1100	5300	19
7000	1300	6400	20
8500	1800	7700	21
10500	2500	9500	22
13500	3500	12000	23
		15500	24

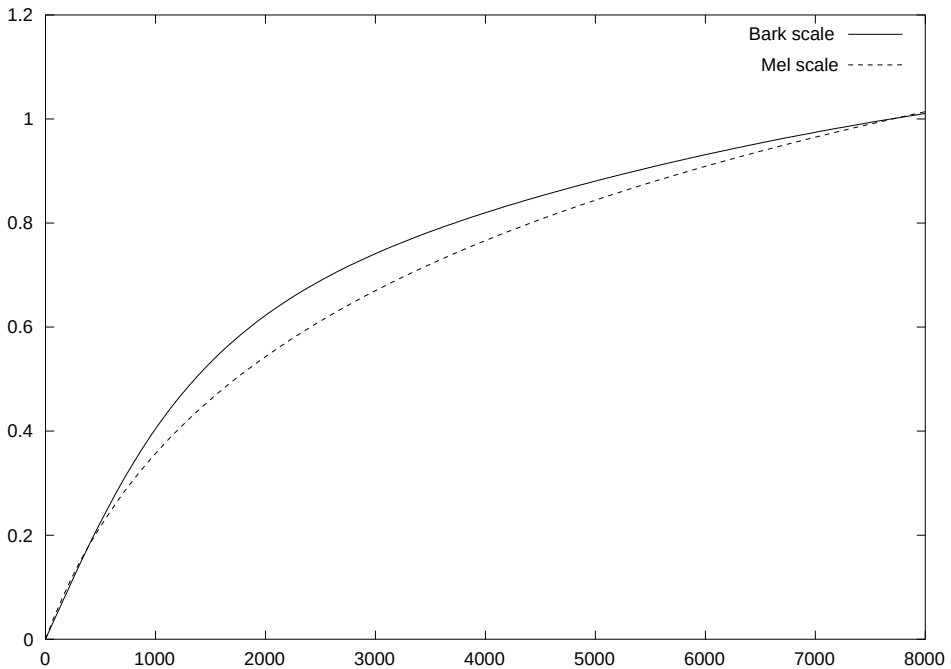
Barkova stupnice 2



Frekvenční stupnice, jednotka: *bark*.
Kritická šířka pásma každé střední frekvence je 1 *bark*.
$b(f) = 13 \arctan(0.00076f) + 3.5 \arctan(f/7500)^2$ [bark]

Vnímaná frekvence a mel(ody)-stupnice

- vnímaná frekvence má jednotku **mel**
- psychoakustická míra vnímání výšky tónu
- definice: tón s 131 Hz, tj. základní tón c_0 (C), dostane vnímanou frekvenci 131 mel
- tón s jinou frekvencí dostane dvojitou hodnotu na mel-stupnici jako tón, který je vnímán o polovinu nižší.
  $B(f) = 1125 \ln(1 + f/700)$ [mel]



Rozlišování výšek tónů

- čistý tón $+$ dlouhá délka $\Rightarrow$ rozdíl výšky tónů o 0,1% – 0,3% vnímatelný
  např. tón s 1 kHz: rozdíl o 3 Hz slyšitelný
- Rozlišování je možné kvůli kódování místa (cochlea) a phase–locking mechanismu.
- Rozlišování je mnohem horší, kde phase–locking mechanismus je slabý (velmi vysoké frekvence) nebo když je tón velmi krátký.
- Rozlišování je velmi špatné, když je zvuk skládán: kolem 2 kHz musí dva tóny mít rozdíl frekvencí kolem 200 Hz, aby byly rozlišitelné.

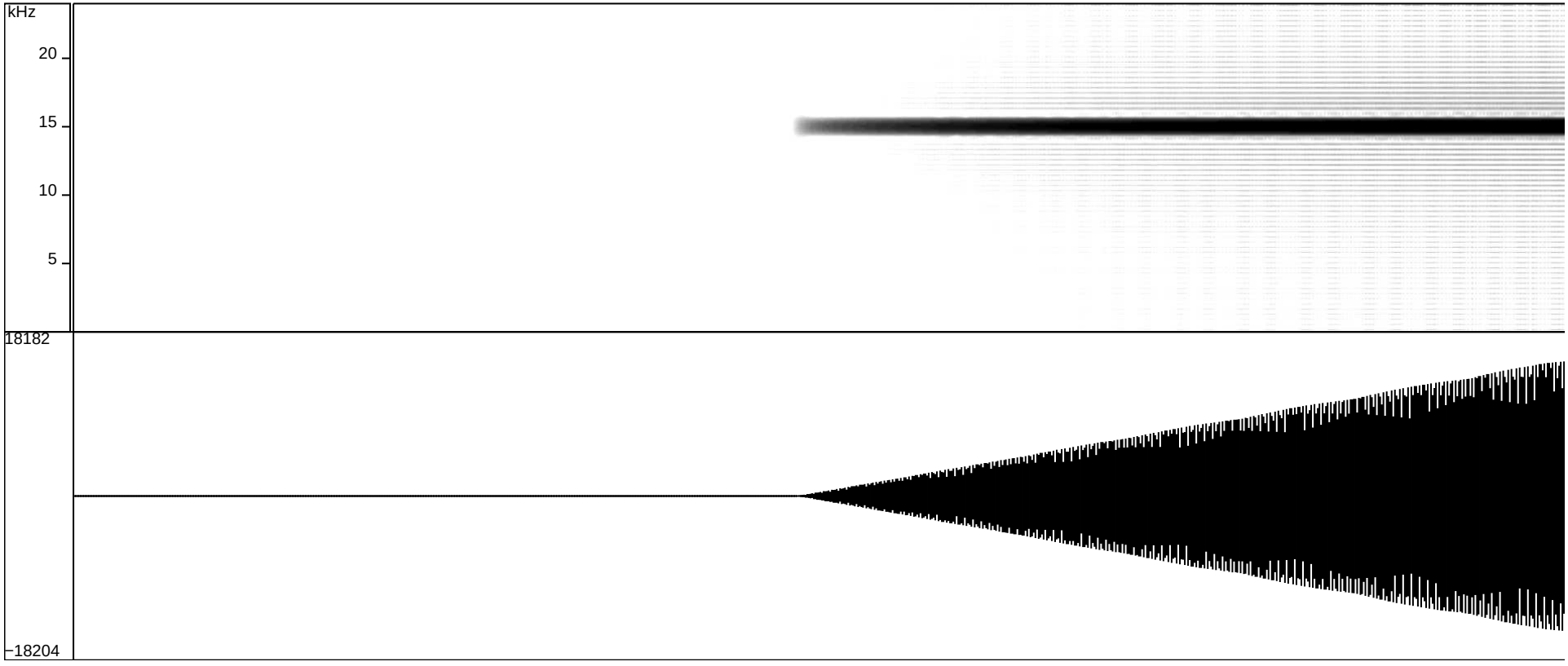
Sluch ve vyšším věku

Kdo neslyší nic?

Sluch ve vyšším věku

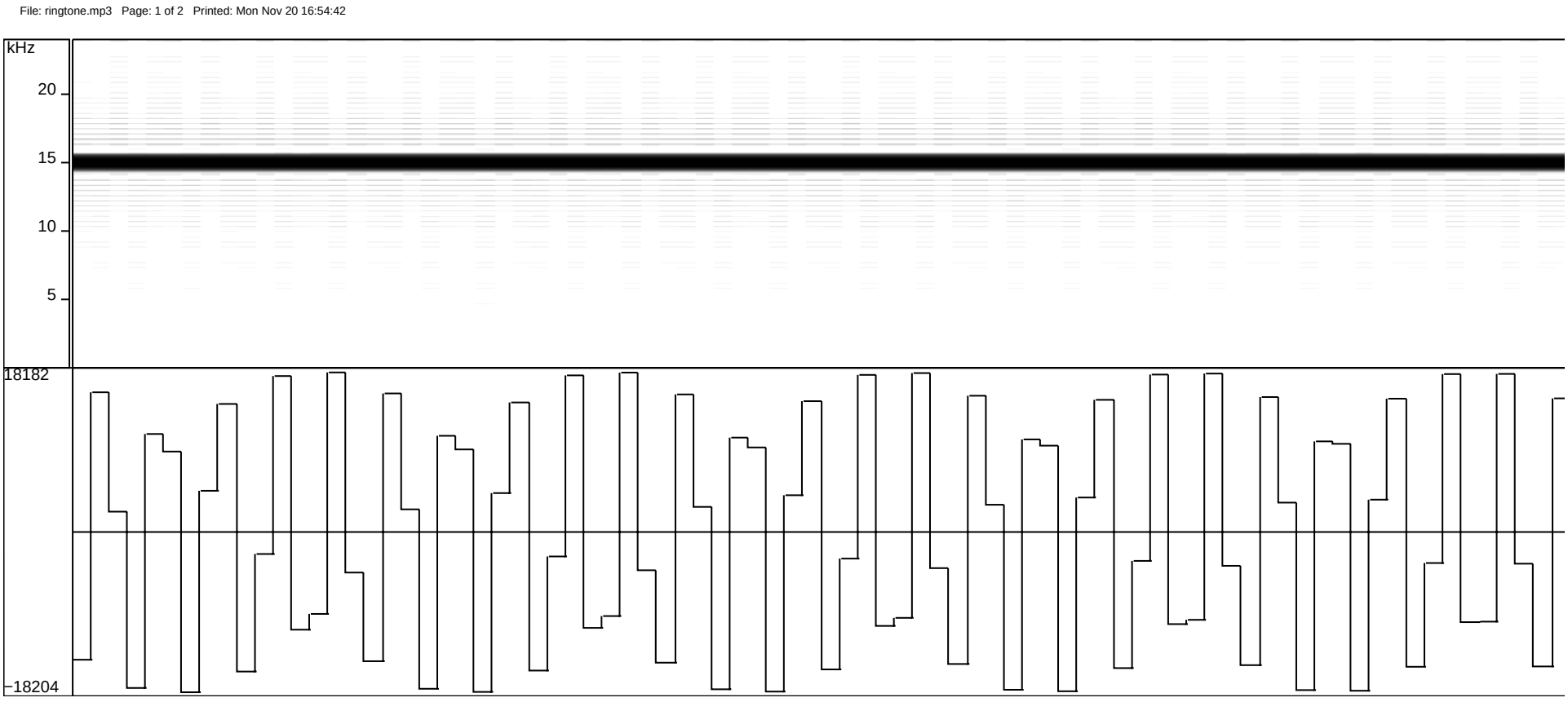
Kdo neslyší nic?

File: ringtone.mp3 Page: 1 of 2 Printed: Mon Nov 20 16:48:43



Sluch ve vyšším věku

Kdo neslyší nic?



Fonace

1. proud vzduchu: plíce $\rightarrow$ průdušnice $\rightarrow$ hrtan (larynx)
2. aktivace na hrtanové záklopce (glottis)
   - otevřená záklopka $\Rightarrow$ neznělá hláska, např. [t] [s]
   - zúžená záklopka, jaksi periodické otevírání a zavírání $\Rightarrow$ znělá hláska, např. z.B. [n] [z]
   - zavřená záklopka s explosivním otevíráním, např. před [a] (anglicky “glottal stop”)
3. kmity probíhají hlasový trakt
   charakteristické rezonanční vlastnosti
   formou/zúžením/uzavřením na určitých místech

Symbolický popis hlásek

řečové hlásky rozlišitelné podle své artikulace
souhlásky – dva důležité příznaky:

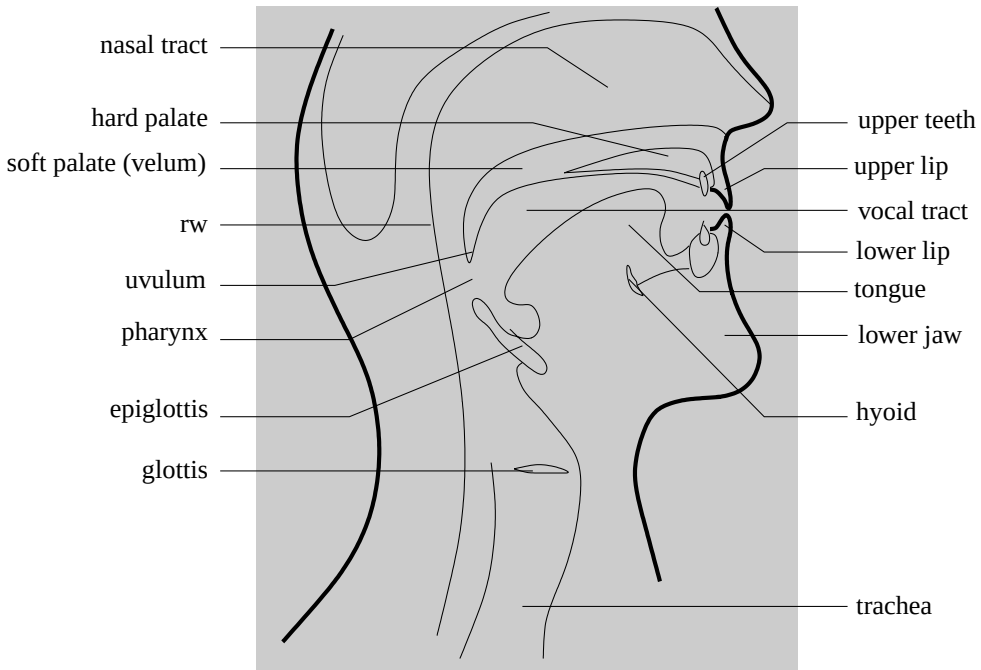
- způsob artikulace
- místo artikulace

samohlásky – artikulovány jako kontinuum; čtyři příznaky:

- svislá poloha jazyka
- vodorovná poloha jazyka
- pozice rtů
- délka samohlásky

Místo artikulace

- pasivní ústrojí artikulační
  - horní ret
  - horní zuby
  - patro:
    tvrdé patro (palatum),
    měkké patro (velum)
  - čípek (uvula)
- aktivní ústrojí artikulační
  - spodní ret
  - hrot jazyka
  - hřbet jazyka



Způsob artikulace: otvor mezi artikulačními ústrojími

- závěrové (okluzivy); výbuchové, ražové hlásky (plozivy), např. [p] [t] [g] – znělé, neznělé nebo aspirované
- nazály, např. [n] [m]
- konstriktivy, frikativy, např. [f] [s] [ʃ]
- aproximanty, např. [j] [l]
- vibranty, např. [r]

Artikulace souhlásek v češtině: přehled

Přehled a rozdělení českých souhlásek

Místo tvoření →					Bilabiály	Labiodentály	Alveoláry	Postalveoláry	Palatály	Veláry	Glotály
Způsob tvoření		Sluchový dojem	Akustická stavba								
Nazální okluzivy	Centrální	Nazální explozivy	Nekontinuální	Matné	m	(ɱ)	n		ɲ	(ŋ)	
Orální okluzivy		Orální explozivy			p b		t d		c ɟ	k g	(ʔ)
Semiokluzivy		Afrikáty		Drsné			t͡s d͡z	t͡ʃ d͡ʒ			
Konstriktivy		Frikativy	Kontinuální			f v	s z	ʃ ʒ		x (ɣ)	ɦ
Vibranty		Vibranty		Matné			(r̥) r				
							r				
Aproximanty		Aproximanty			(ʊ̯)				j		
	Laterální						l				
Hlasnost (fonace)		Znělost (sonorita)			- +	- +	- +	- +	- +	- +	- +
Artikulační orgán					labiální (rty)		lingvální (jazyk)				glotální (hlasivky)
Akustická stavba					gravisové		akutové			gravisové	
					nekompaktní				kompaktní		

http://upload.wikimedia.org/wikipedia/commons/2/29/Souhlasky_CS.PNG

Artikulace souhlásek v němčině: přehled

	bilabial	labiodental	alveolar	palatal	velar	uvular	glottal
Plosive	p b		t d		k g		ʔ
Nasale	m	ɱ	n		ŋ		
Frikative	ɸ β	f v	s z / ʃ ʒ	ç j	x ɣ	χ ʁ	h
laterale Frikative			ɬ				
Laterale			l				
Vibranten			r			ř	
Anschläge			ɾ			ř	
Halbvokale		ʋ		j	ɣ	ʁ	

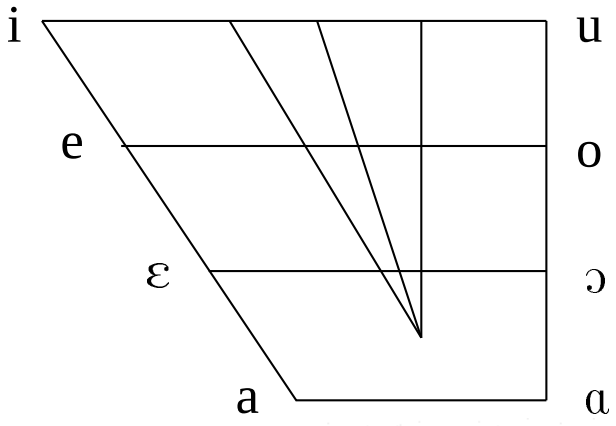
Kategorizace samohlásek

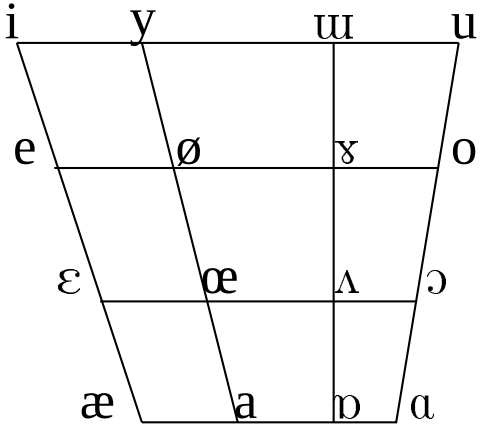
znaky popisu:

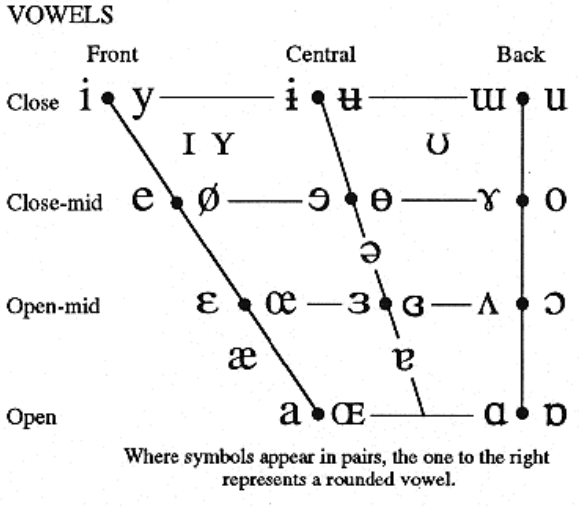
1. svislá poloha jazyka:
   nejvyšší bod jazyka svisle: vysoký, polovysoký, polonízký, nízký
   odpovídá stupni otevřenosti
   zavřené [i:]
   polozavřené [e:] (např. v němčině)
   polootevřené [ɛ]
   otevřené [a]
2. vodorovná poloha jazyka:
   nejvyšší bod jazyka vodorovně:
   přední ([i], [e]), zadní ([u], [o]), střední,
   neutrální samohláska *schwa* ([ə], [ɐ])
3. postavení rtů:
   zaokrouhlené ([o], [u]) vs. nezaokrouhlené ([i], [e])
4. délka samohlásky:
   např. [i] v „byt“ a [i:] v „být“

Trojúhelník samohlásek

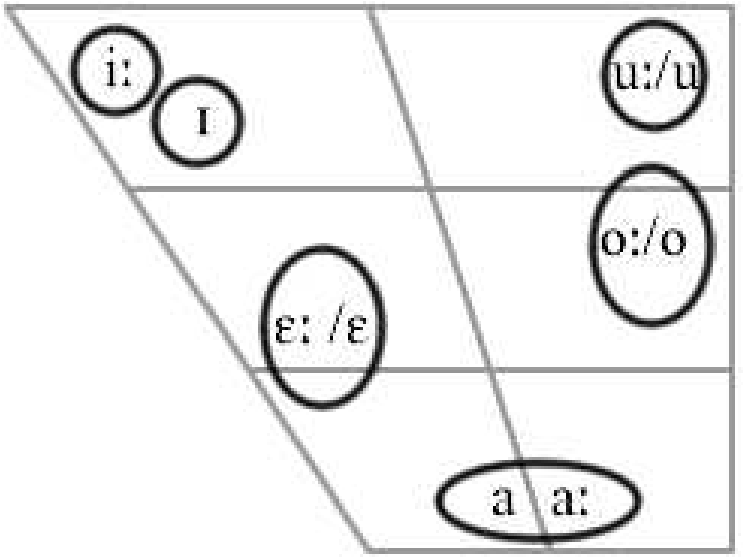
Trojúhelník (někdy nazýván jako čtyřúhelník) samohlásek jsou samohlásky doplněné podle (pří)znaků otevřenosti, popř. svislé polohy jazyka (vertikálně) a vodorovné polohy jazyka (horizontálně) – různá zobrazení:





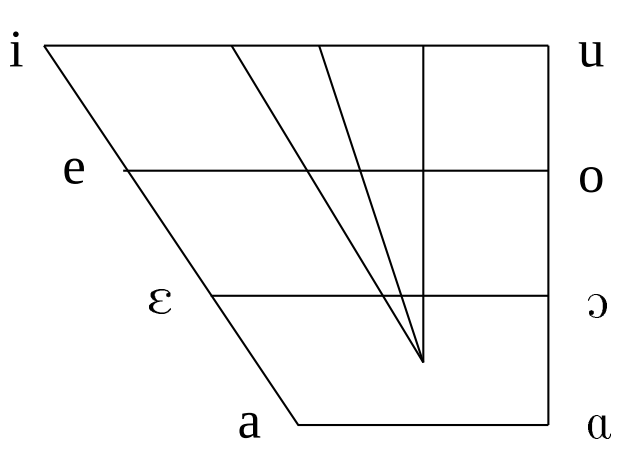


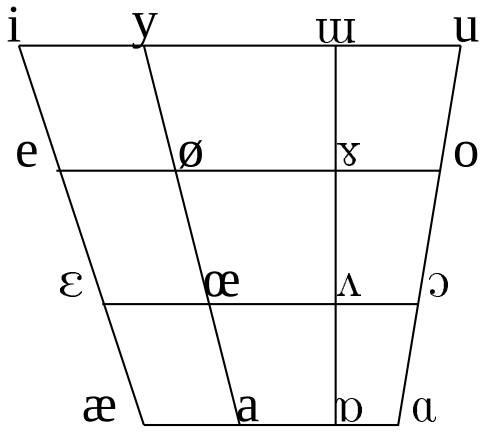
Trojúhelník samohlásek češtiny

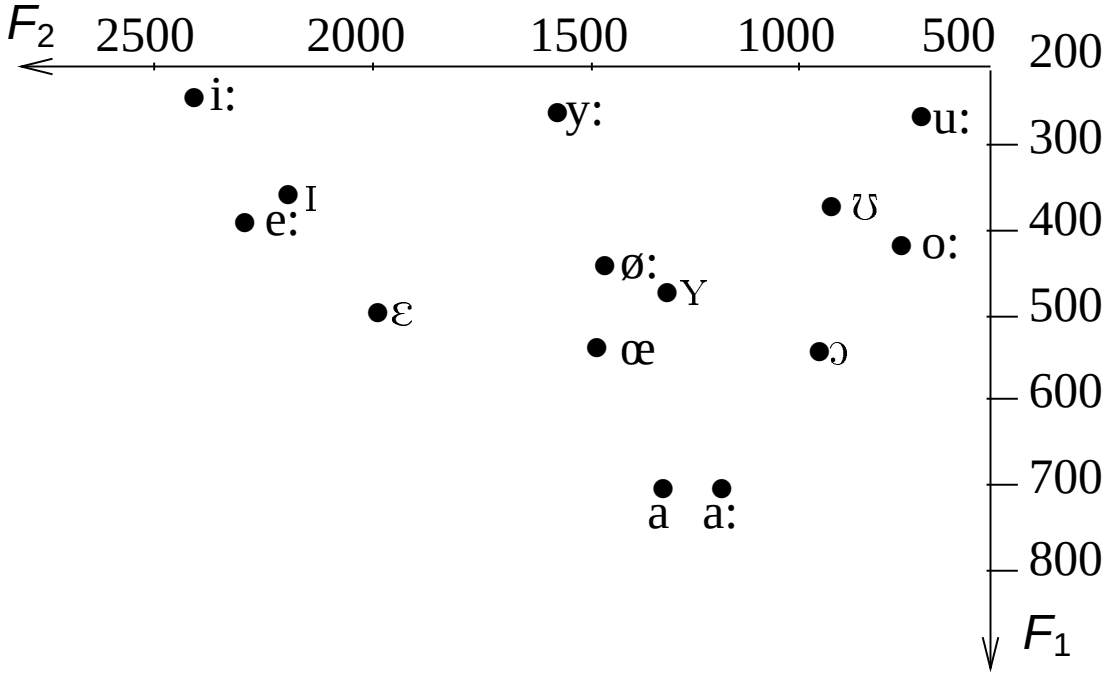


http://upload.wikimedia.org/wikipedia/commons/4/47/Czech_vowel_chart.png

Čtyřúhelník samohlásek (vyhlídka: počítání příznaků)







Fonologická kategorizace hlásek

Vlastnosti zvuku a artikulační pohyby zanedbáváme; bereme v úvahu pouze funkci hlásek k přenosu informací.
→ **foném**: nejmenší zvuková jednotka, která umožňuje rozlišovat významy

- **alofon** je konkrétní realizace fonému; např. jsou [n] a [ŋ] alofony psaného [n] – va**n**a [**n**] vs. ba**n**ka [ŋ]
- hláska/fón: zvuková jednotka, která ještě není klasifikována jako reprezentant fonému

Identifikace fonémů jazyku

- pražský (evropský) strukturalismus nebo funkcionalismus
  - Jazykový expert získá jazykový korpus introspekcí (sebepozorování).
  - segmentace
  - klasifikace analýzou minimálních párů: Dvě slova tvoří minimální pár, když se odliší jen v jediném hláskovém segmentu („let“ a „lep“). Fóny, které tvoří rozlišovací znak, patří k různým fonémům; tady reprezentují [t] a [p] fonémy /t/ a /p/.
- americký strukturalismus nebo distribucionalismus (Bloomfield)
  - řečové promluvy od jiných osob $\rightarrow$ reprezentativní korpus
  - segmentace
  - charakteristika klasifikace: dvě hlásky jsou v komplementární distribuci, když se neobjeví ve stejném prostředí (např. nahoře [n] a [ŋ]) $\rightarrow$ jeden foném
    Klasifikace se provádí pomocí substituce: Nahraď jeden prvek druhým a zeptej se jiné osoby, jestli se mění význam.

Fonémy v češtině

- **souhlásky (Palková 1994, str. 242):**
  p, b, m, t, d, n, ť, ď, ň, k, g, c, č, f, v, s, z, š, ž, x, h, ř, r, l, j
- **samohlásky a dvojhlásky (diftongy; Palková 1994, str. 193):**
  a, á, e, é, i, í, o, ó, u, ú, ou, au, eu

Zkreslení výslovnosti

výslovnost ovlivněná **tempem** a **stylem** mluvení
$\Rightarrow$ **redukce** standardní výslovnosti

- intonace: přízvuk redukován
- fonetické segmenty: **asimilace** + **elize**
- zkreslení hlásek

zkreslení hlásek koartikulací:

- přesnost pohybování
- nižší náročnost pohybování
- týká se místo a způsob artikulace a hlasové polohy
- vyšší tempo mluvení $\Rightarrow$ silnější zkreslení
- zejména silné pro bezpřízvučná funkční slova

Zkreslení výslovnosti, asimilace, elize atd.

Palková 1994, str. 144:

- intervokalické oslabování: lavice → [laice], [lajce]
- změna neznělé na znělou souhlásku: pět osob → [pjed osop]
- asimilace znělosti: s babičkou → [z babičkou]
- asimilace artikulační se změnou místa tvoření: [sčítat] → [ščítat], [banka] → [baŋka]
- asimilace artikulační se změnou způsobu tvoření: [slovenská] → [slovencká]
- redukce samohlásek: materiál → [matrijál]
- metateze (přeskupení hlásek nebo slabik): zvláštní→ [vzláštňí]
- elize: kostka → [koska], vždycky → [dicki]
- proteze (na začátku slova před samohláskou): okno → [vokno]
- epenteze (na kterémkoli jiném místě): osm → [osum]

Slabiky

různé definice **fonetické slabiky**

- **teorie o tvoření slabik podle tlaku:**
  - jeden výdechový impulz $\Rightarrow$ jedna slabika
- **teorie o tvoření slabik podle sonornosti:**
  - nejsonornější hláska tvoří jádro slabiky
  - třídy fónů (v němčině; klesající hodnoty):
    1. otevřené samohlásky
    2. zavřené samohlásky
    3. vibrant [r]
    4. laterální aproximant [l]
    5. nazály
    6. znělé frikativy
    7. neznělé frikativy
    8. plosivy
  - růst sonornosti v slabice až k nositeli maximální hodnoty (*sonor*) a potom pokles
  - hranice slabiky u minimální hodnoty

Intonační/prozodická informace

přízvuk

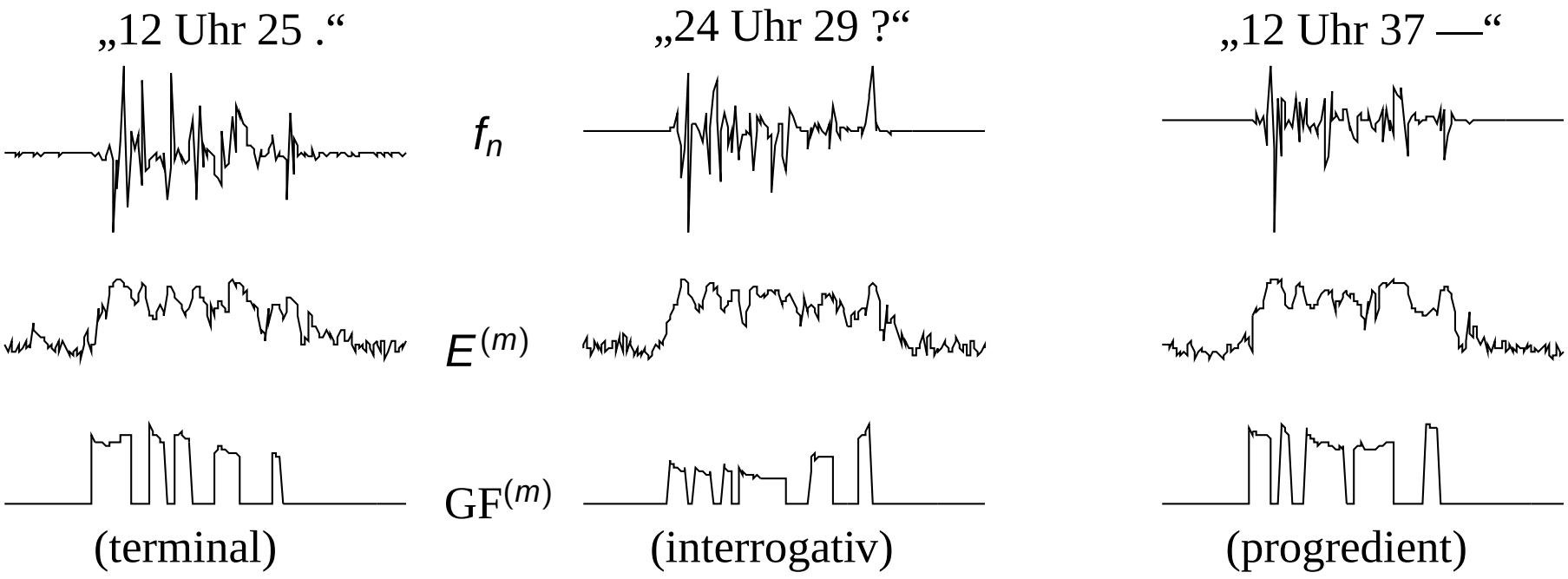
- rozlišování slov: (německy) *um’fahren* vs. *’umfahren*
- emfaticky: *to je ’ab’so’lut’ně špatně*
- zdůraznění: *Slo’vinsko, ne Slo’vensko*
- všeobecně zvýrazňování části promluvy

melodémy: oznamovací věta vs. otázka vs. progredující výpověď

intonační prostředky:

- umístění pauz pro lepší srozumitelnost
- výběr hlasové polohy
- melodie řeči (časový průběh)
- artikulační tlak, popř. zvuková intenzita
- časová struktura

Melodémy: příklad



melodém: ukončující klesavý — ukončující stoupavý — neukončující