PB095 - Úvod do počítačového zpracování řeči

Luděk Bártek[1]

[1]

Obsah

Cíle předmětu

Seznámení s oblastmi:

- digitálního zpracování zvuku
  - v časové oblasti
  - ve frekvenční oblasti
  - převod signálu z časové do frekvenční oblasti
- syntézy řeči
- rozpoznávání řeči
- dialogových systémů

Organizace předmětu

- Dvouhodinová přednáška
- Možnosti zakončení:
  - zkouška - písemka + ústní zkouška, termíny budou vypsány v IS MU během prosince
  - kolokvium - ústní rozprava na dané téma z oblasti zpracování zvuku
  - zápočet

Vyučující

- Luděk Bártek
- e-mail: bar@fi.muni.cz
- kancelář: B314
- konzultace viz osobní stránka v ISu (`https://is.muni.cz/auth/osoba/2154#vyuka`)

Doporučená literatura

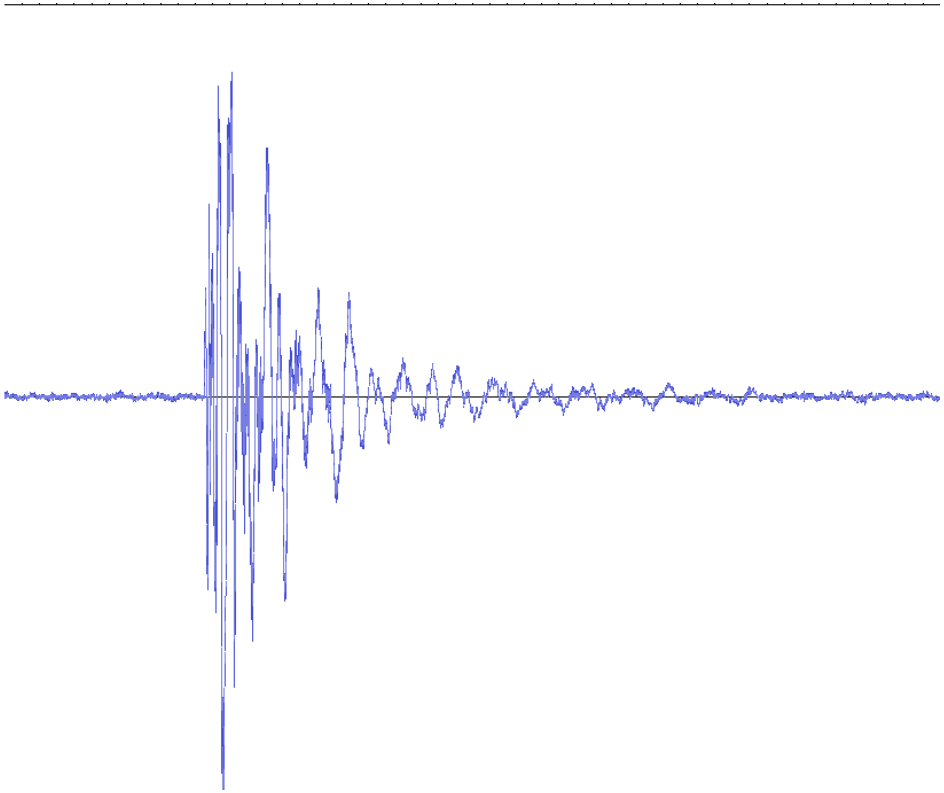
- J. Psutka et al, Mluvíme s počítačem česky, Academia 2006
- J. Psutka, Komunikace s počítačem mluvenou řečí, Academia, Praha, 1995
- Z. Kotek, V. Minařík, Metody rozpoznávání a jejich aplikace, Academia, Praha, 1993
- T. Dutoit, An introduction to Text-to-Speech Synthesis, Kluwer Acad. Publ., 1999
- M. R. Schroeder, Computer Speech, Springer 1999
- Původní stránky předmětu (`http://www.fi.muni.cz/~kopecek/upzr.htm`) doc. Kopečka
- Stránky Voice Browser Activity (???)

- Laboratoře LSD, NLP
- `http://lsd.fi.muni.cz/`
- `http://nlp.fi.muni.cz/`

Co je to zvuk?

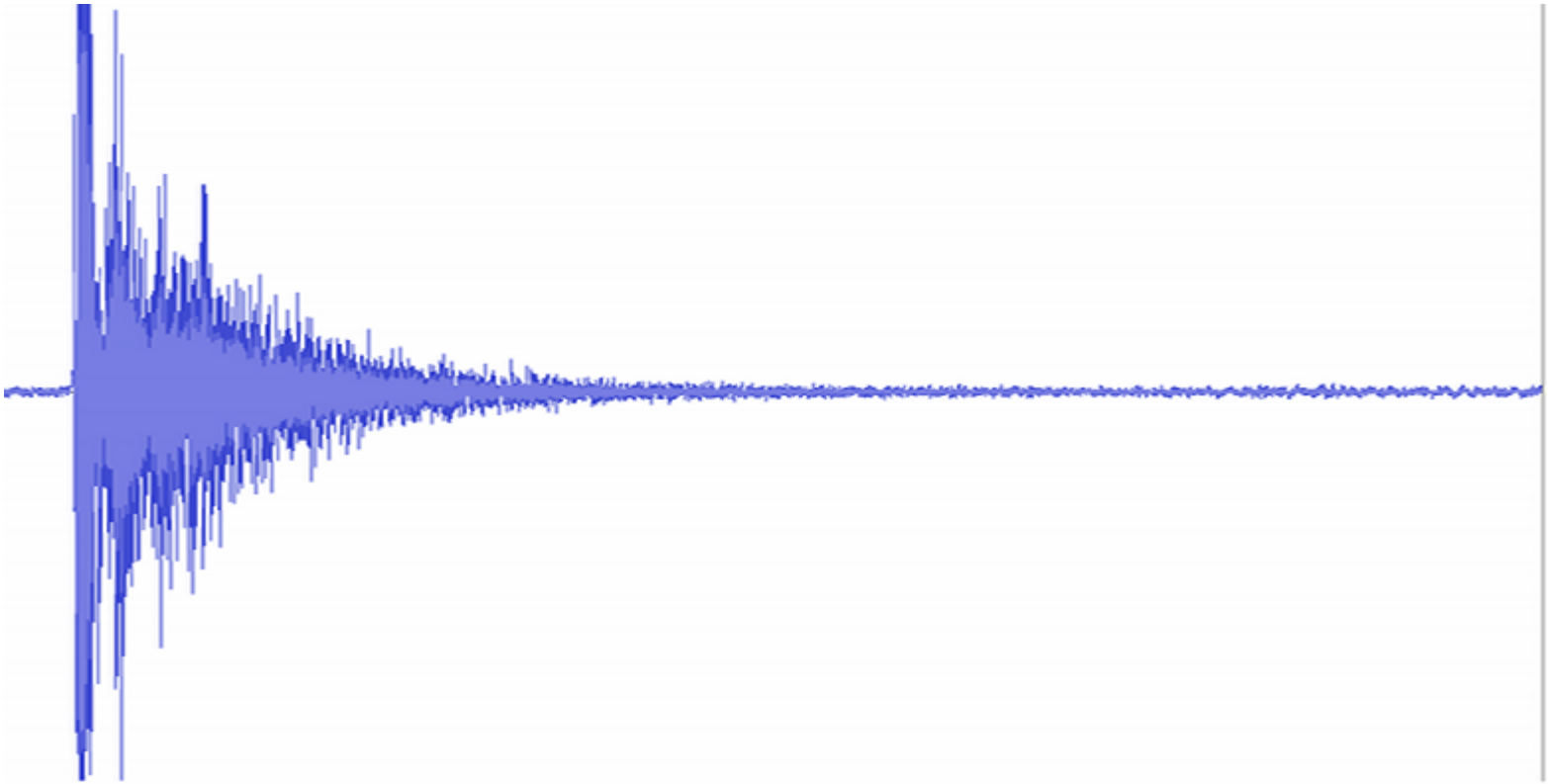
- Akustický signál.
- Jedná se o kmitavý pohyb molekul pružného prostředí.
  - vzduch
  - voda
  - kov
  - ...
- Vyvolán odporem prostředí - vede k opakovanému stlačování prostředí.
- Podrobněji v části fyzikální akustika.

Ukázka zvuku



Zvuk klepnutí na plastové tělo počítače

Ukázka zvuku

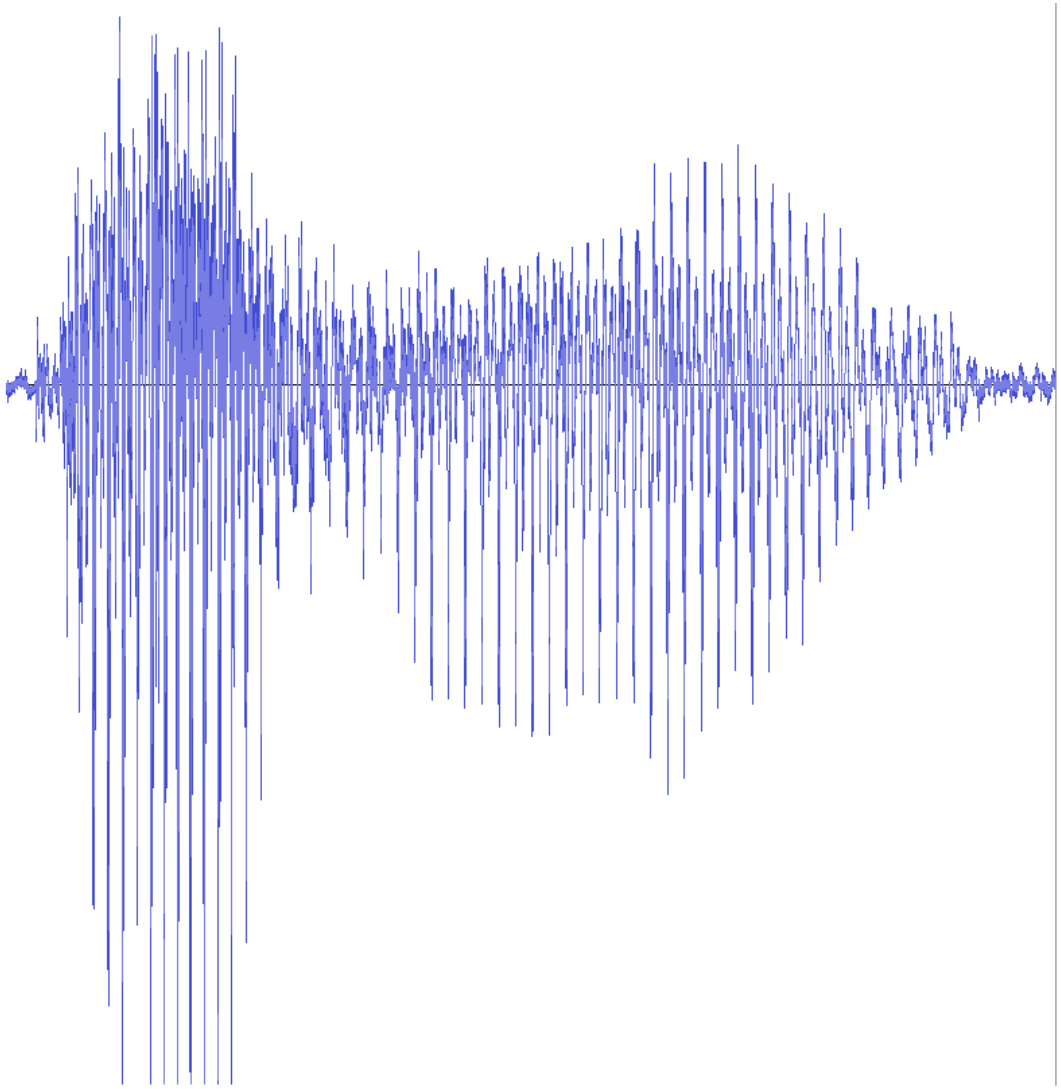


Zvuk tlesknutí

Co je to řeč?

- Akustický signál a gesta sloužící ke komunikaci.
- Obsahuje definované vzory (slova), která jsou dána jazykem.
- Velmi rozvinutý u člověka.
  - Příznaky schopnosti tvorby artikulované řeči již u Australopitéka (-3 milióny let).
  - Slouží ke sdělování: myšlenek, pocitů, emocí, ...
    - myšlenek - "Dnes budeme probírat láčkovce."
    - pocitů - "Je mi krásně.", "Radši se ke mně ani nepřibližuj!", ...
    - emocí - "Au!", "Jé!",
- Určité formy akustické komunikace (řeči) lze pozorovat i u dalších vyšších živočichů:
  - způsob zajištění kooperace při získávání obživy (delfín, vlk, ...)
  - vábení partnera (jelen, ...)
  - vyjádření emočních stavů (pes, opice, ...).
  - ...

Ukázka řeči



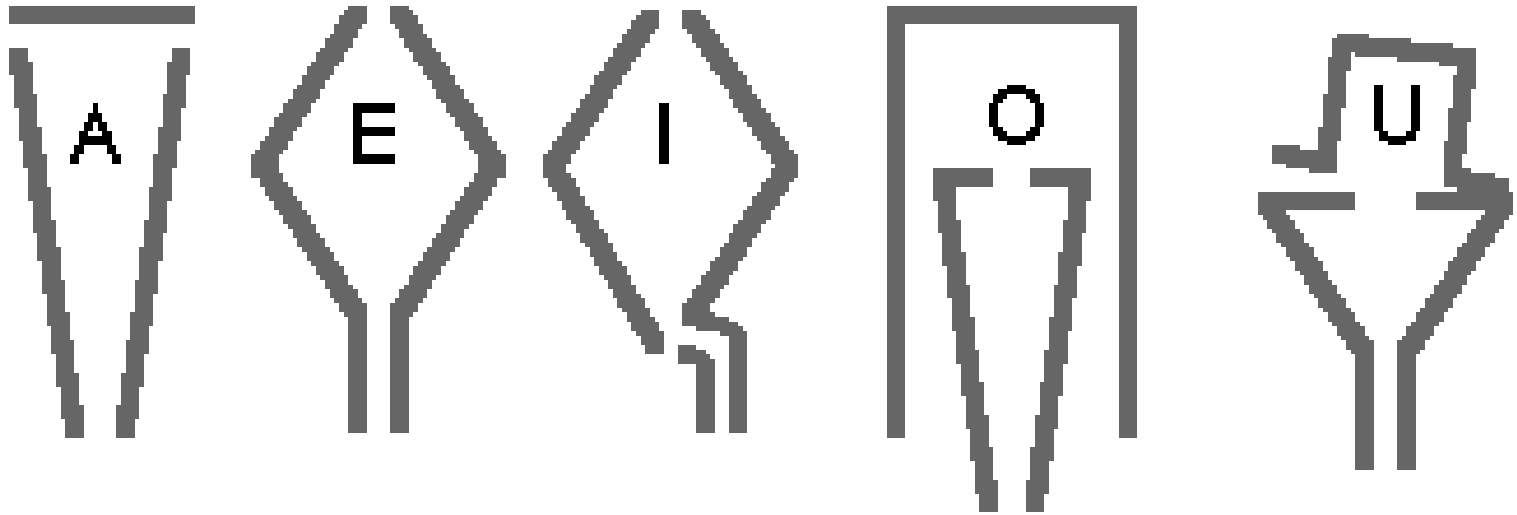
Zvukový záznam (`images/ahoj.wav`)

Studium řeči

- Fyzika - akustika.
- Biologie - medicína (fyziologie, fyziologická akustika).
- Jazykověda - fonetika.

Přehled historie zpracování a napodobování řeči

- Schopnost artikulované řeči - australopitekus - cca. -3 000 000 let
- Starověk - budování mluvících soch
- Galileo Galilei - souvislost mezi tónem a frekvencí
- 1779 - Christian Gottlieb Kratzenstein - systém rezonátorů pro samohlásky a, e, i, o ,u



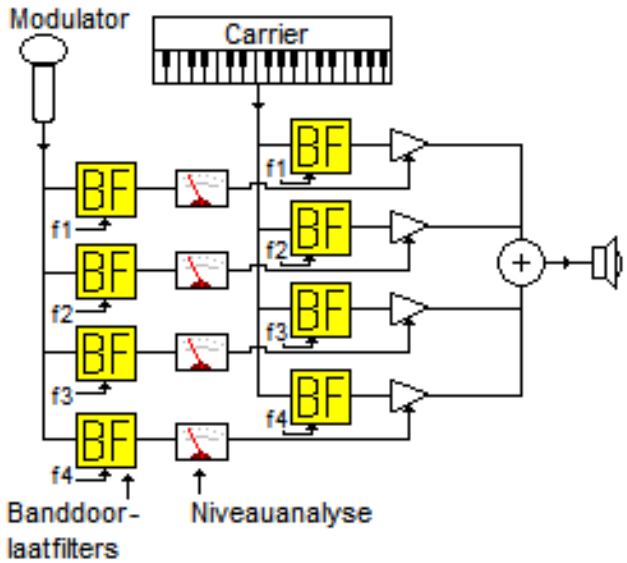
Přehled historie zpracování a napodobování řeči

- 1791 - (Johann) Wolfgang von Kempelen (de Pázmánd) - první mechanický řečový syntetizér

- 1835 - zrekonstruován a upraven Wheatonem - navíc pružná "ústní dutina".
- 1846 - J. Faber - mluvící stroj Euphonia

Přehled historie zpracování a napodobování řeči

- 1937 - R. R. Riesz - mechanický mluvící stroj
- 1939 - H. Dudley
  - VODER - elektromechanický řečový syntetizér
  - VOCODER - systém pro kódování a přenos řeči



- 50. léta 20. století - syntéza ve frekvenční oblasti
  - později v časové oblasti
- 70. léta 20. století - počítačové zpracování zvuku

Historie analýzy řeči

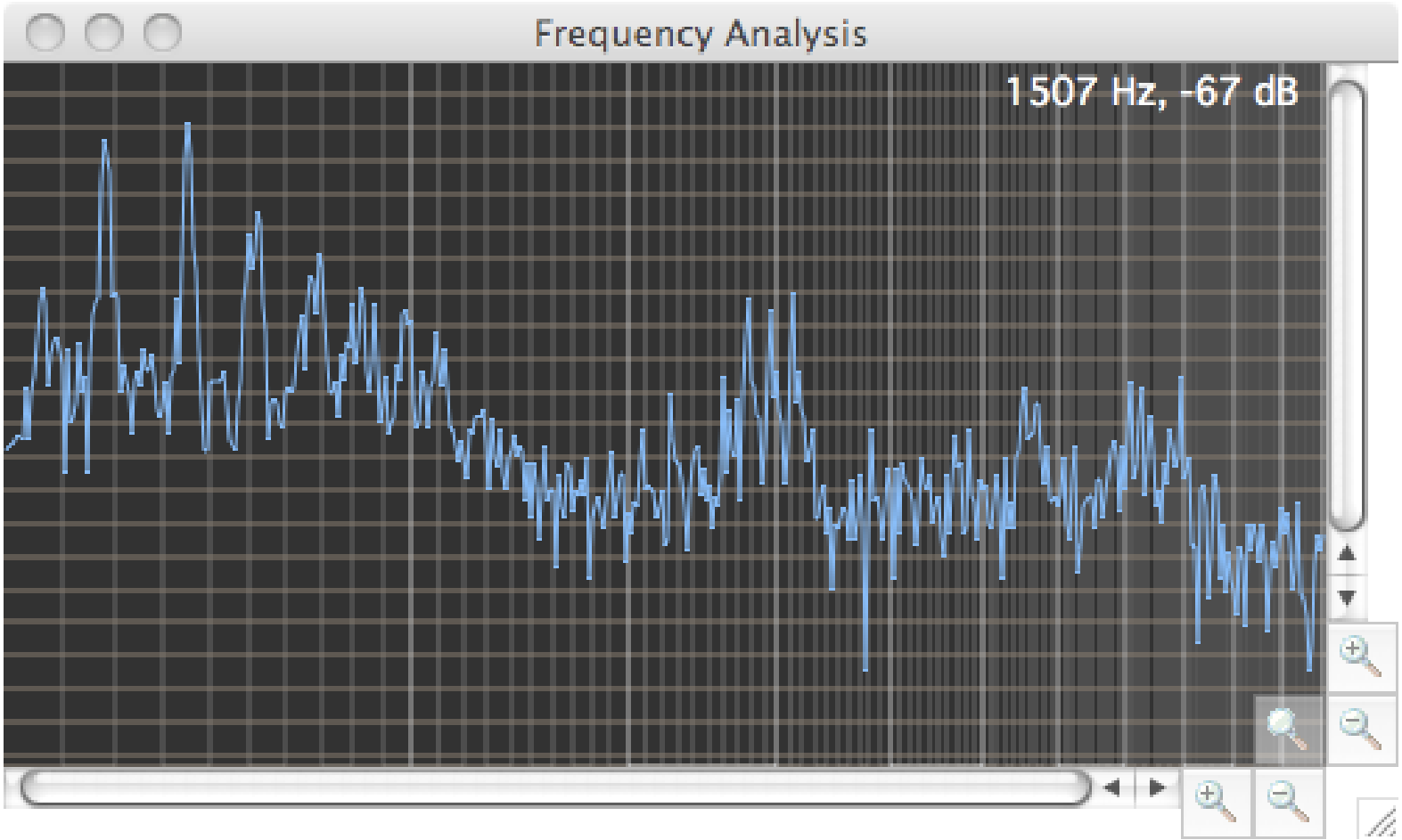
- 19. století porozumění principů tvorby a zpracování řeči (rezonanční teorie, základy fonetiky):
  - J. B. Fourier - Fourierova věta
    - principy spektrální analýzy zvuku
  - H. Helmholtz
    - fyziologie vnímání hudby
    - Helmholtzů rezonátor

  - J.R. Ewald - fyziologie sluchu.

Historie analýzy řeči

- Dvacáté století:
  - 1924 - spektrální analýza řeči na bázi formantové analýzy samohlásek
  - Vokodéry - komprese řečového záznamu
  - 1946 - 47 zařízení pro grafické zobrazení řeči
  - 2. polovina 20. století - intenzivní rozvoj teorie a počítačových aplikací

Spektrum zvuku



Textová data k obrázku. (`images/spektrum-a.txt`)

Současnost

- syntéza řeči:
  - komerční TTS:
    - AT&T Natural Voices
    - IBM Research TTS
    - Loquendo TTS
  - nekomerční TTS:
    - MBrola
    - Festival
    - Demosthenes

Současnost

- Rozpoznávání řeči:
  - izolovaných slov
  - souvislé promluvy
    - komerční: Dragon, ViaVoice Desktop Products
    - nekomerční: Sphinx4, ...
- Dialogové systémy
  - Infocity Liberec (TU Liberec, Prof. J. Nouza), 485353100
  - MIT Cambridge, Spoken Language System Group
    - Mercury - 001-877-648-8255
    - Jupiter - 001-888-573-8255

- Syntéza a rozpoznávání řeči
  - Demosthenes
  - NLP - čeština pro syntetizér MBrola - využit řečový korpus CLAP
- Asistivní technologie:
  - Audi-C - dialogové programování v C++
  - Audis - řečový hypertextový prohlížeč
  - ...
- Dialogové systémy
  - WebGen (`http://lsd.fi.muni.cz/webgen/`) - dialogové generování webových prezentací
  - GATE (`http://lsd.fi.muni.cz/GATE/`- dialogové kreslení obrázků, dialogové prohlížení obrázků, zvukové zobrazení obrázků
  - ...
- Spolupráce s laboratořemi NLP, VR, ...

Akustika

- Věda zabývající se zvukem.
- Z řeckého akustikos - vztahující se k slyšení.
- Akustika zkoumá zvuk z hlediska:
  - fyzikálního (fyzikální akustika) - zvuk jako fyzikální vlnění
    - rychlost šíření, vztah mezi různými fyzikálními veličinami zvuku, šíření zvuku, ...
  - fyziologického (fyziologická akustika) - tvorba a vnímání řeči u člověka
  - hudebního (hudební akustika) - zvuky a jejich kombinace s ohledem na potřeby hudby
    - jak lidem zní kombinace a sekvence zvuků a tónů, ...
  - molekulárního (molekulární) - vztah molekulární struktury a akustických vlastností
    - k měření se využívá hyperzvuk (¿100 MHz].
  - zpracování zvuku na počítači (počítačová akustika) - digitální zpracování zvuku.

- Během semestru se budeme zabývat:
  - fyzikální akustikou
  - fyziologické akustikou
  - počítačovou akustikou

Základy fyzikální akustiky

- Zvuk je fyzikální vlnění:
  - kmitavý pohyb molekul
  - mechanické vlnění látkového prostředí vyvolávácí sluchový vjem
  - charakterizován:
    - frekvencí
    - amplitudou

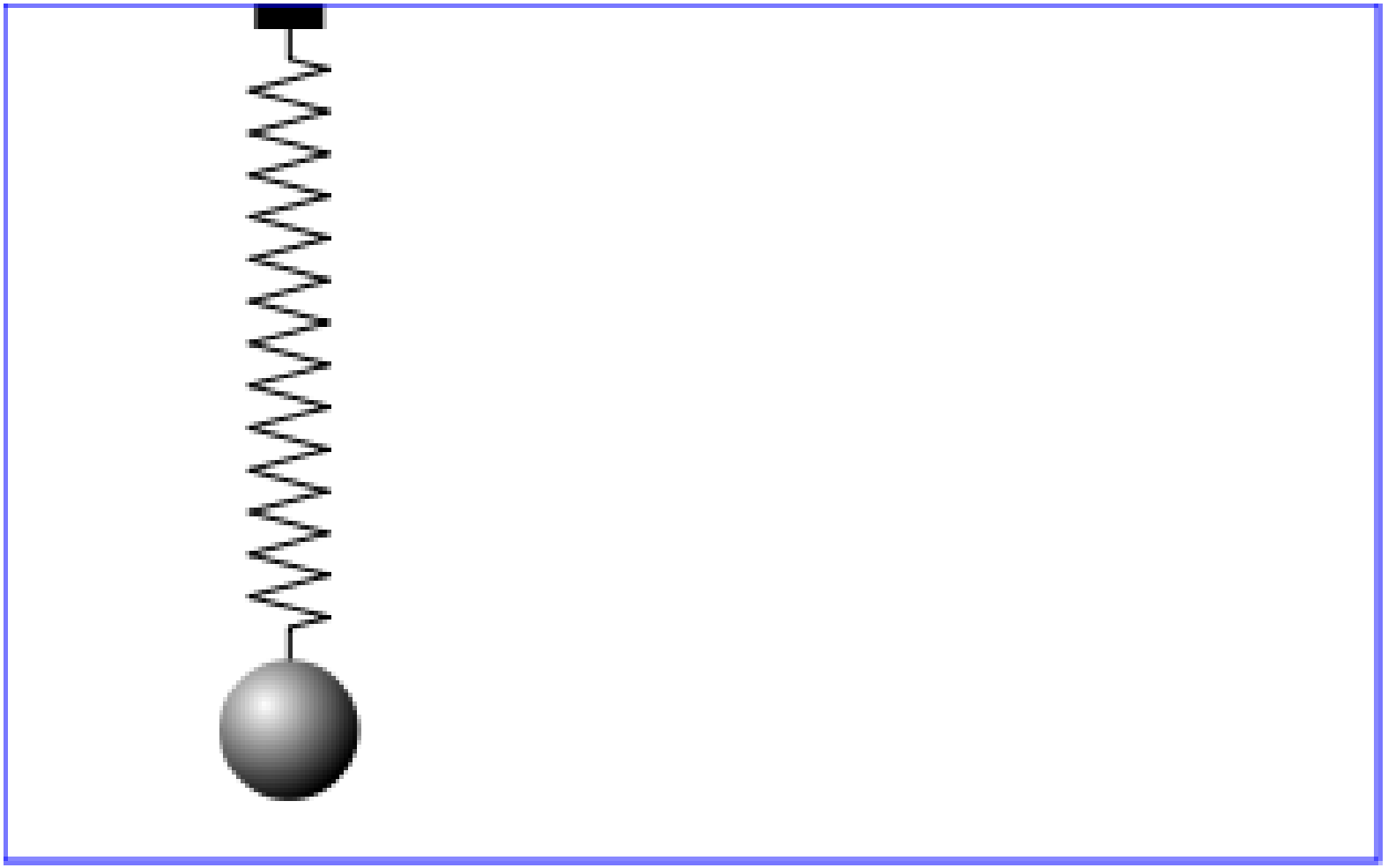
Frekvence

- perioda (T) - nejkratší doba, kterou tělesu trvá průchod stejnou fází pohobu.
- $f = 1/T$
- jednotka Hz
- 1 Hz = 1 perioda za sekundu

Rychlost zvuku

- závisí na prostředí a řadě dalších fyzikálních faktorů
  - teplota
  - tlak
  - ...
- rychlost zvuku v různých prostředích
  - vzduch (13,4 stupňů) - 340 m/s
  - voda (25 stupňů) - 1500 m/s
  - rtuť - 1400 m/s
  - beton - 1700 m/s
  - led - 3200 m/s
  - ocel - 5000 m/s
  - sklo - 5200 m/s

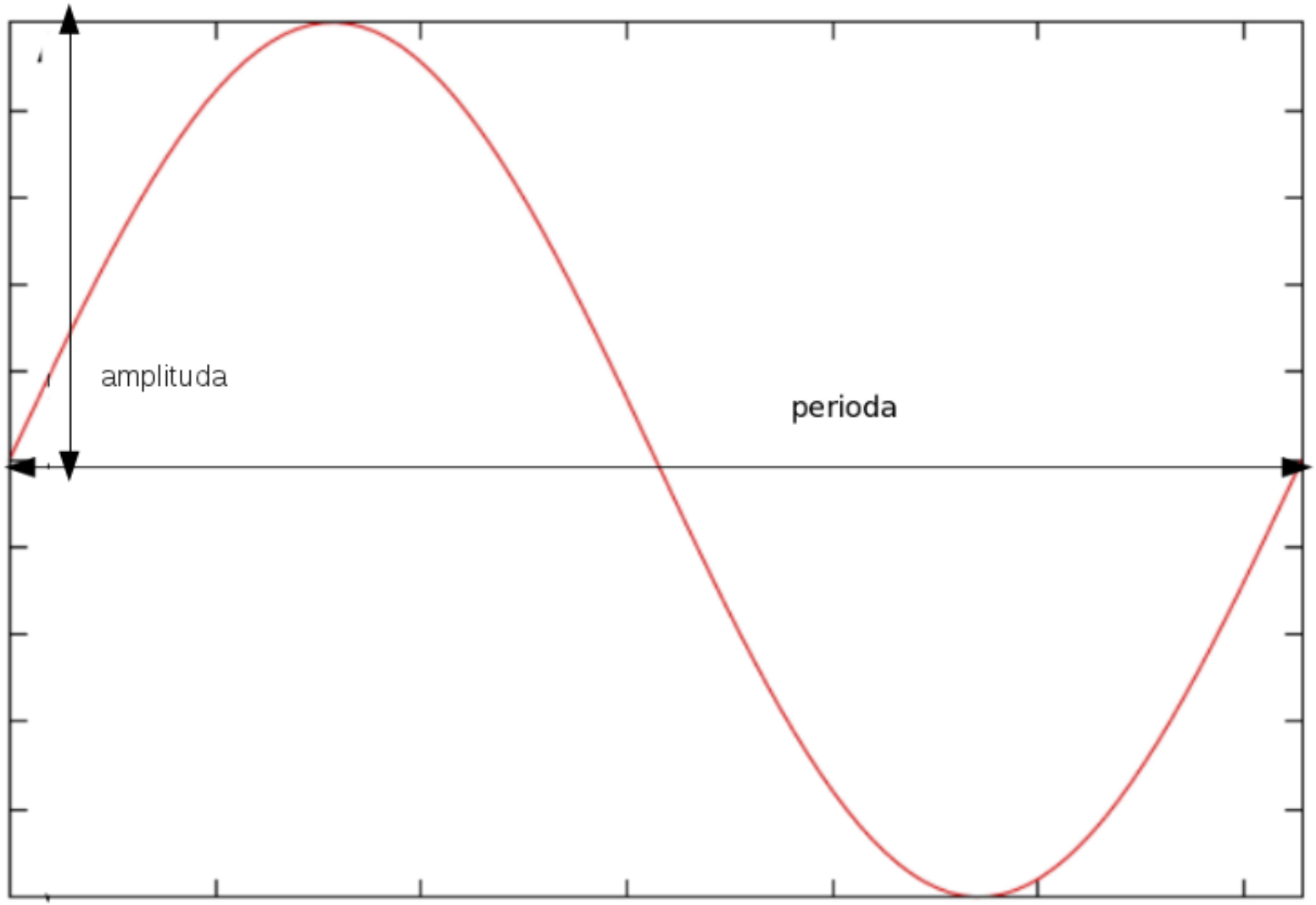
Hmotný bod na nehmotné pružině



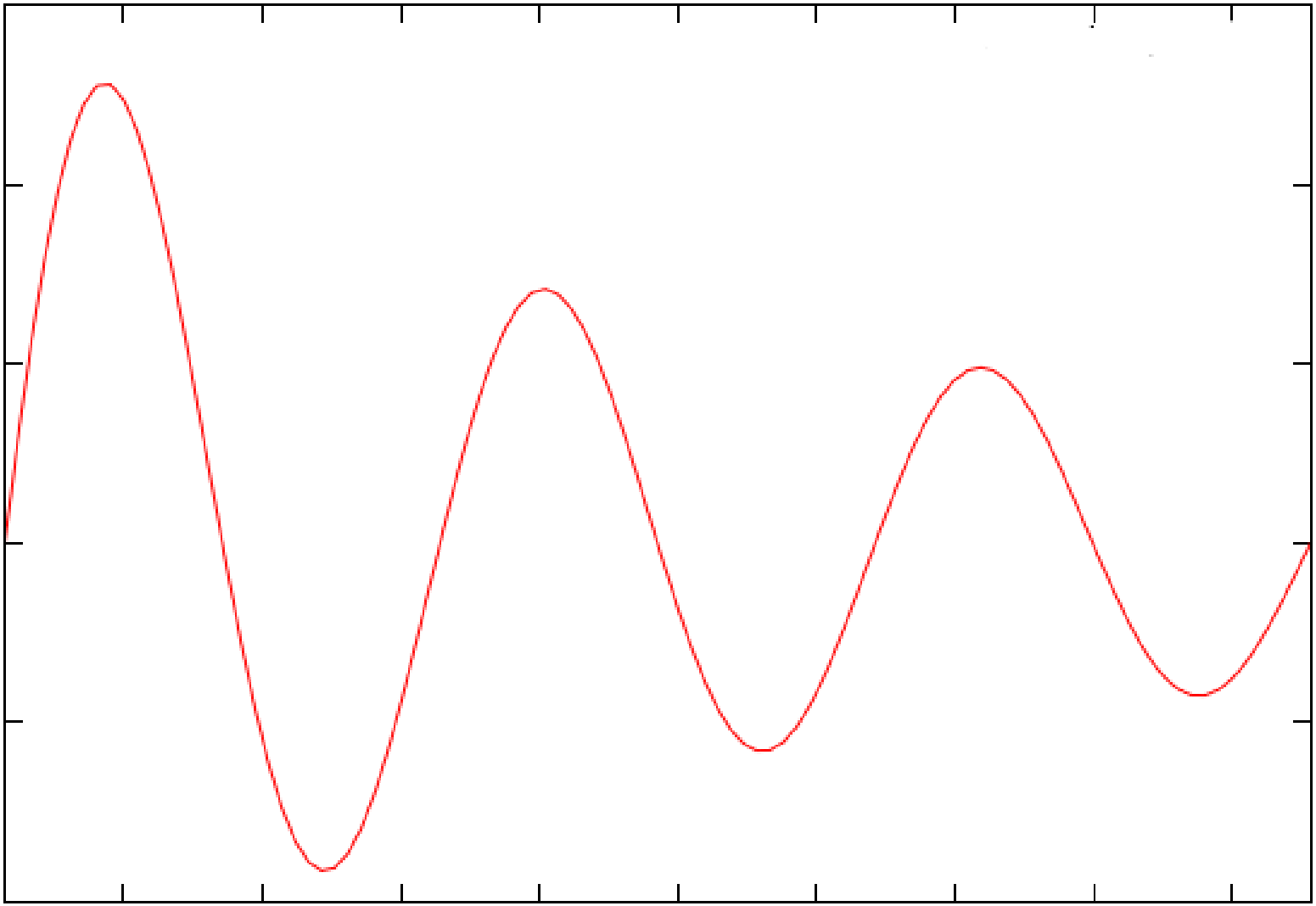
Hmotný bod na nehmotné pružině

- Zanedbáváme:
  - odpor prostředí
  - gravitaci
  - ...
- Základní veličiny:
  - amplituda - maximální hodnota výchylky dané periodické veličiny (ymax)
  - okamžitá výchylka - $y = y_{max} sin(\omega t)$
    - ω - úhlová rychlost periodického jevu $\omega = 2\pi / T = 2\pi F$ [rad/s]
    - t - čas
  - perioda (T) - doba trvání jednoho opakování daného jevu. Měří se v sekundách.
  - frekvence - F = 1/T [Hz]

Perioda a amplituda



Tlumené kmity



Vlastní a vynucené kmity, rezonance

- Vlastní kmity - jsou kmity soustav bez působení vnějších sil
- Vnější kmity - vynuceny vnějším prostředím systému (buzením)
- Rezonance - fyzikální jev, malá budící síla může způsobit značné změny kmitajícího systému

$$A_r = \frac{\frac{S}{m}}{2b\sqrt{\omega_0^2 - b^2}} = \frac{S}{2mb\omega}$$

- A_r - rezonanční amplituda
- S - amplituda budící síly
- m - hmotnost kmitajícího tělesa
- b - tlumení kmitající soustavy (řádově menší než omega)
- ω - úhlová rychlost tlumených kmitů

Akustický tlak a akustická intenzita

- Akustická intenzita:
  - množství energie, které projde jednotkovou plochou za jednotku času - jednotka Wm-2
    - P - tlak, S - plocha

$$I = \frac{P}{S}$$

- Akustický tlak - síla působíci na element plochy v prostředí vlnivého děje (jednotka Pascal [Pa])
  - má-li sinusový průběh:

$$p = p_0 sin(\omega t)$$

p0 - maximální akustický tlak v průběhu periody

- Akustická intenzita je úměrná druhé mocnině akustického tlaku.

Akustická intenzita (2.)

- práh citlivosti (slyšení) - $I_0 = 10\text{-}12\ W/m^2$ $\quad$ 20 μPa
- práh bolesti - $1 W/m^2$ $\quad$ 130 Pa
- Intenzita není vnímána lineárně (lineární nárust vnímané intenzity odpovídá geometrickému nárustu intenzity)
  - Weber-Fechnerův psychofyzikální zákon
- Hladina intenzity (hlasitost) zvuku L

$$L = 10log(\frac{p}{p_0})^2$$

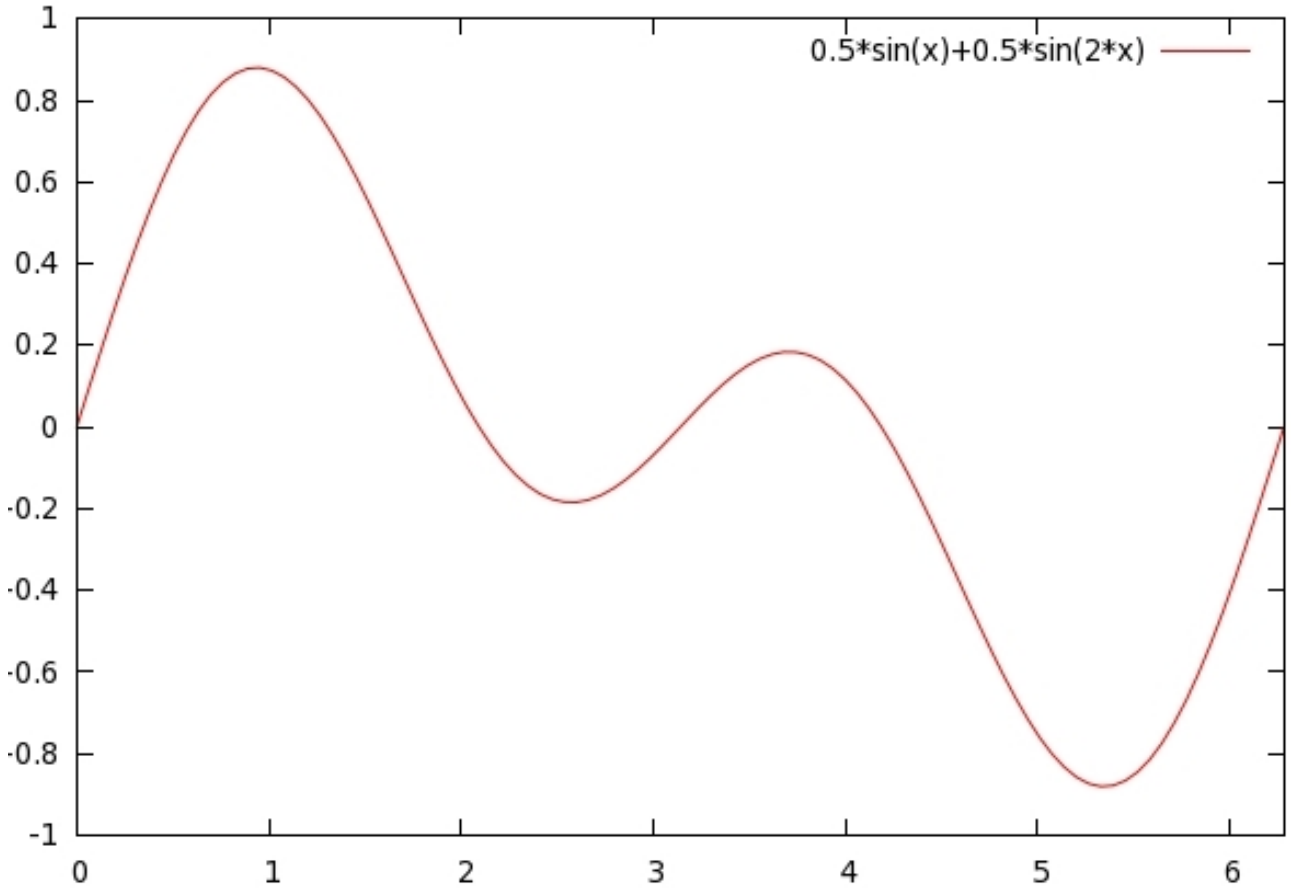
- Jednotka 1 Bel (1B) - rozsah hladin cca 13 Belů

Akustická intenzita - orientační hodnoty

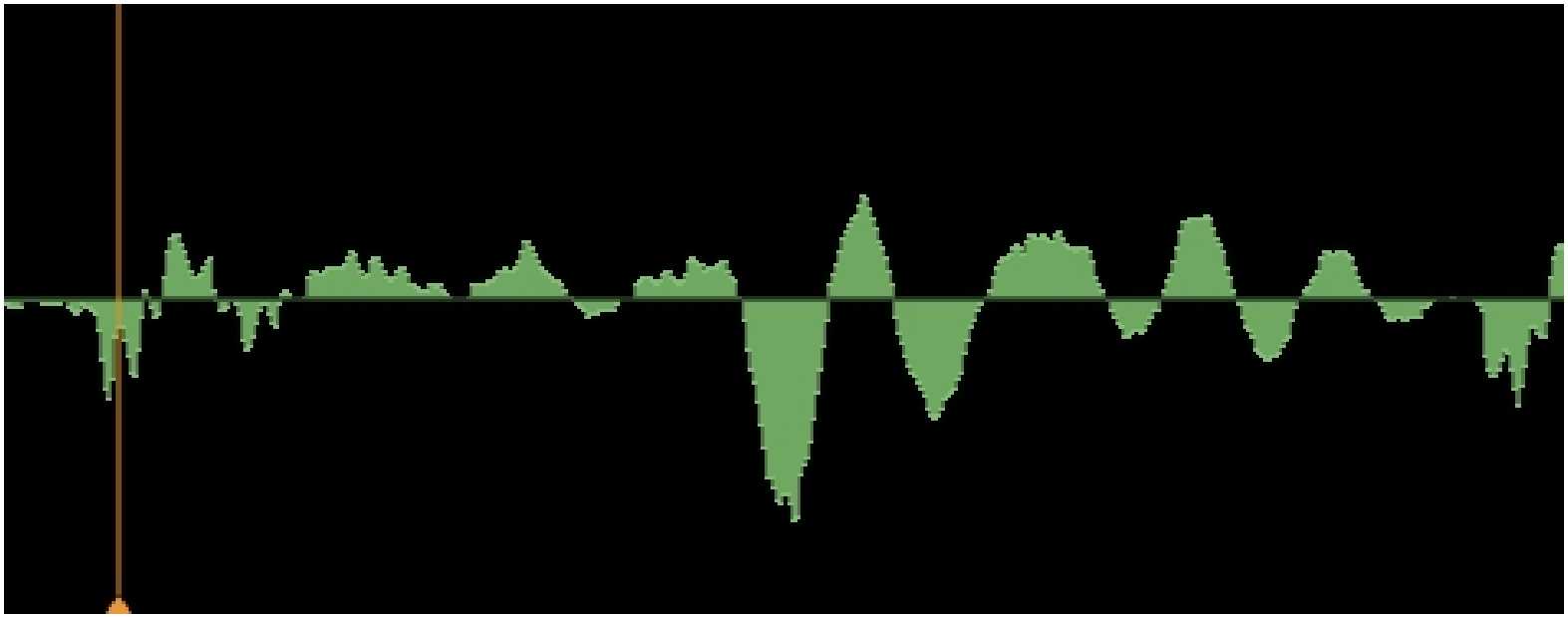
- šepot - 10 - 20 dB
- tlumený hovor - 35 - 45 dB
- hovor střední hlasitosti - 50 - 55 dB
- symfonický orchestr - 70 - 90 dB
- rocková hudba 110 - 130 dB
- vzlet proudového letadla 190 dB
- subjektivní vnímání zavisí na frekvenci

Základní a složený tón

- Základní tón - zvukovou intenzita v závislosti na čase odpovídá sinusoidě
- Složený tón - lineární kombinace základních tónů
  - vetšina zvuků

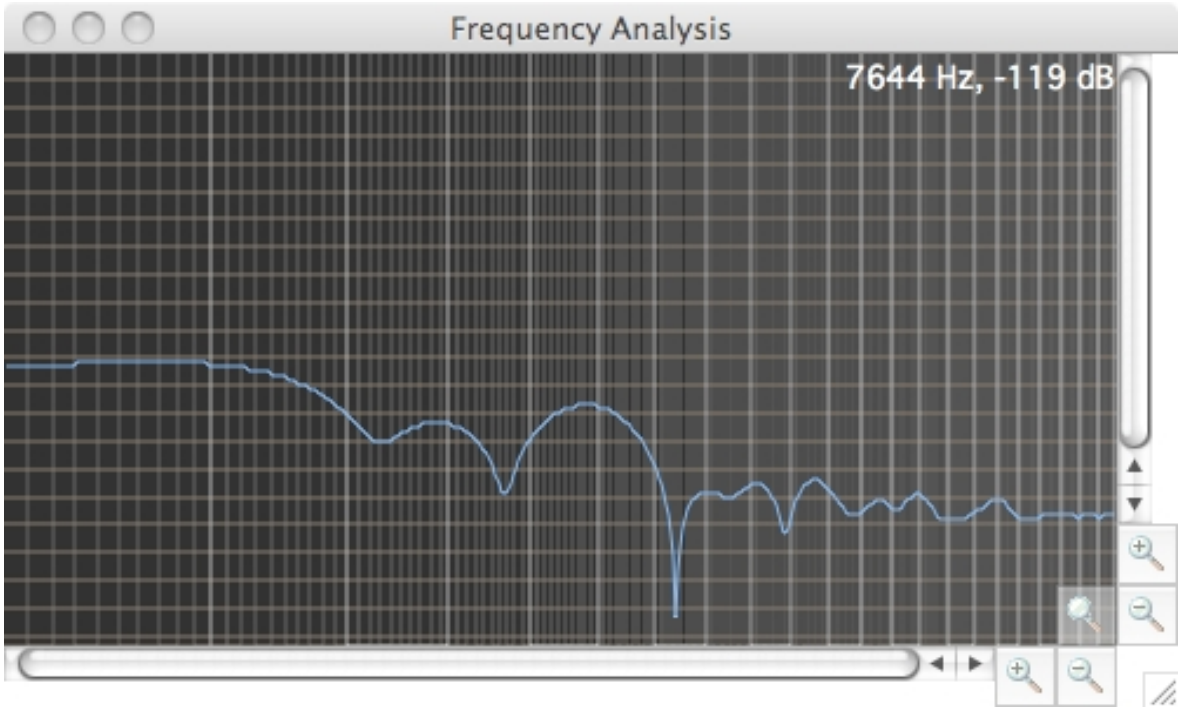


Reálný zvuk



Akustické spektrum zvuku

- Reálné zvuky:
  - Jedná se většinou složené tóny.
  - Složeny ze základních tónů.
  - Lze je rozložit na jednotlivé složky - akustické spektrum



  - K získání frekvenčních charakteristik lze využít např Fourierovu transformaci, lineární predikci, ...

Fourierovy řady

- f(x) - periodická funkce s periodou T
- $$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k cos kx + b_k sin kx$$
- nejlepší aproximace pro:
  - $\alpha, \alpha + T$ - interval periodicity funkce f(x)

$$a_k = \frac{2}{T} \int_{\alpha}^{\alpha+T} f(x) cos(k\omega x) dx$$

$$b_k = \frac{2}{T} \int_{\alpha}^{\alpha+T} f(x) sin(k\omega x) dx$$

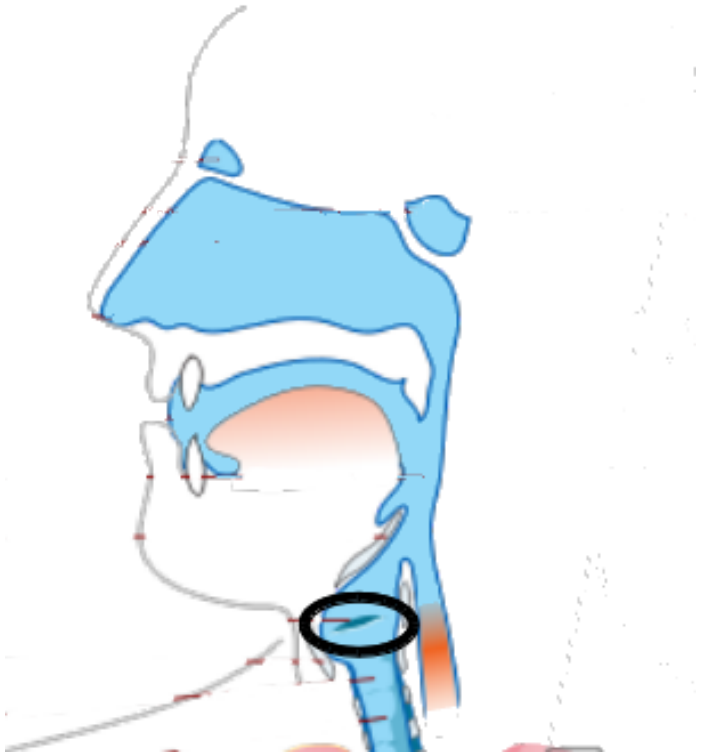
Základy fyziologické akustiky

- Mechanismus vytváření řeči
- Mechanismus vnímání řeči
- Helmholtzova rezonanční teorie
  - G Bekesy - Nobelova cena za fyziologii a medicínu za výzkum funkce cochlei(3. 6. 1899, Budapest - 13. 6. 1972, Honolulu)
- Helmholtzův rezonátor

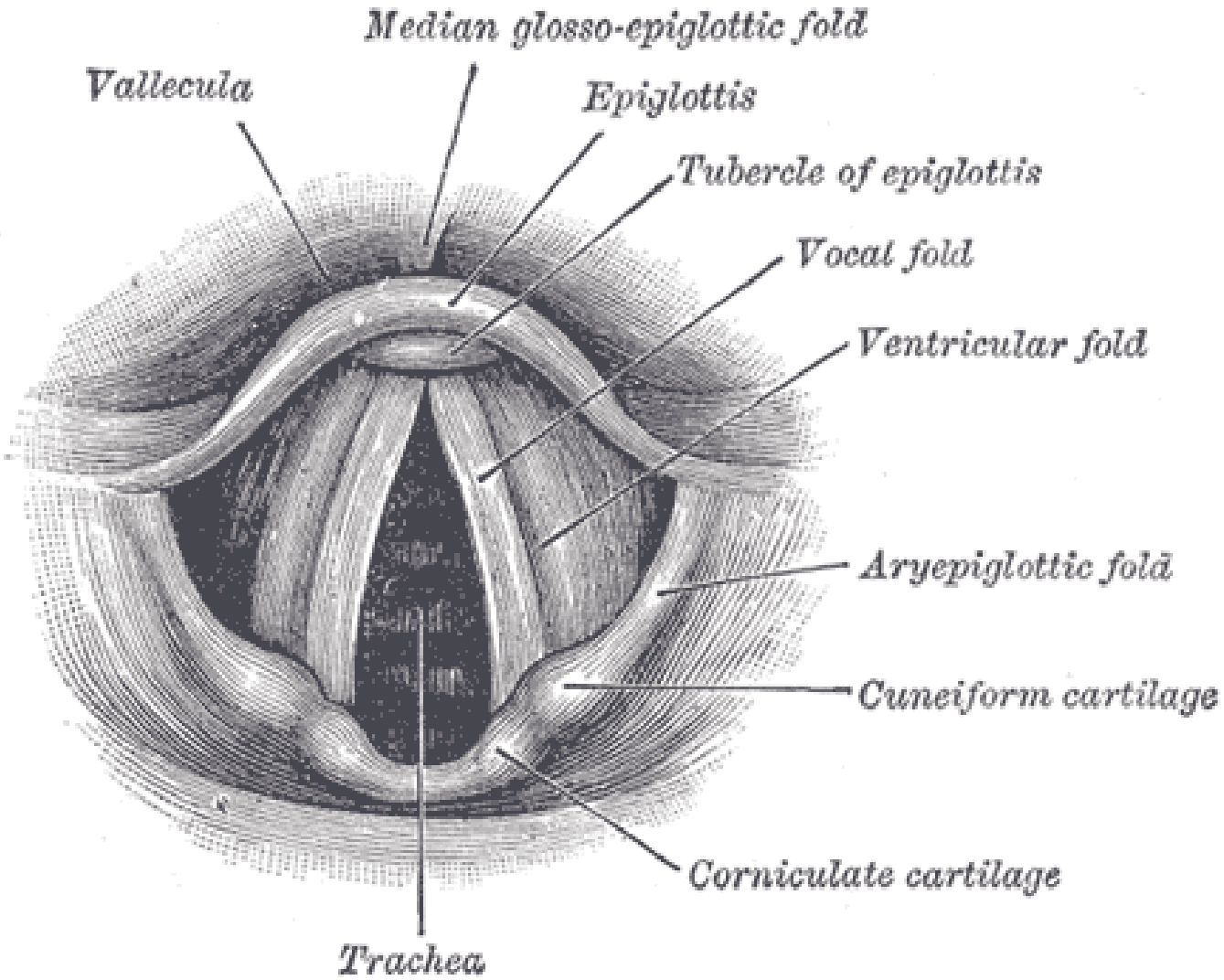
Mechanismus vytváření řeči

- Řeč vytváří hlasové ústrojí - hlasivky v hrtanu.
  - Hlasivky vytváří úzkou hlasovou štěrbinu - vzduch je při průchodu rozkmitá.
  - Frekvence kmitání hlasivek - základní hlasivkový tón.
  - Zvuk vzniklý v hrtanu (s výjimkou např. sykavek) je modifikován v rezonančních dutinách (obdoba Helmholtzova rezonátoru).
  - Rezonanční dutiny:
    - hrtanové
    - ústní
    - nosohltanové

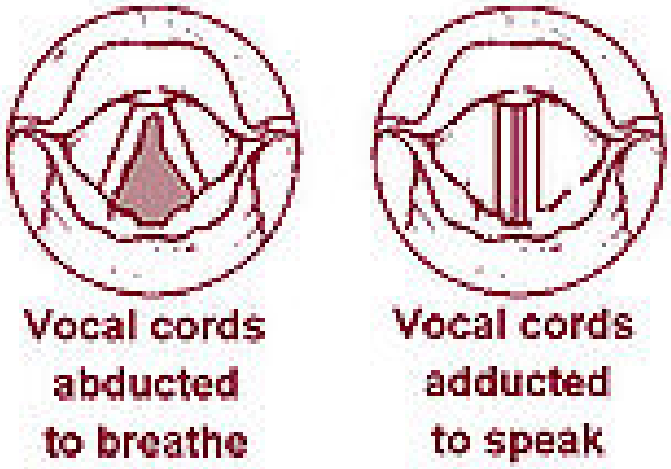
Umístění hlasivek



Hlasové ústrojí - schéma hlasivek



Hlasové ústrojí - činnost

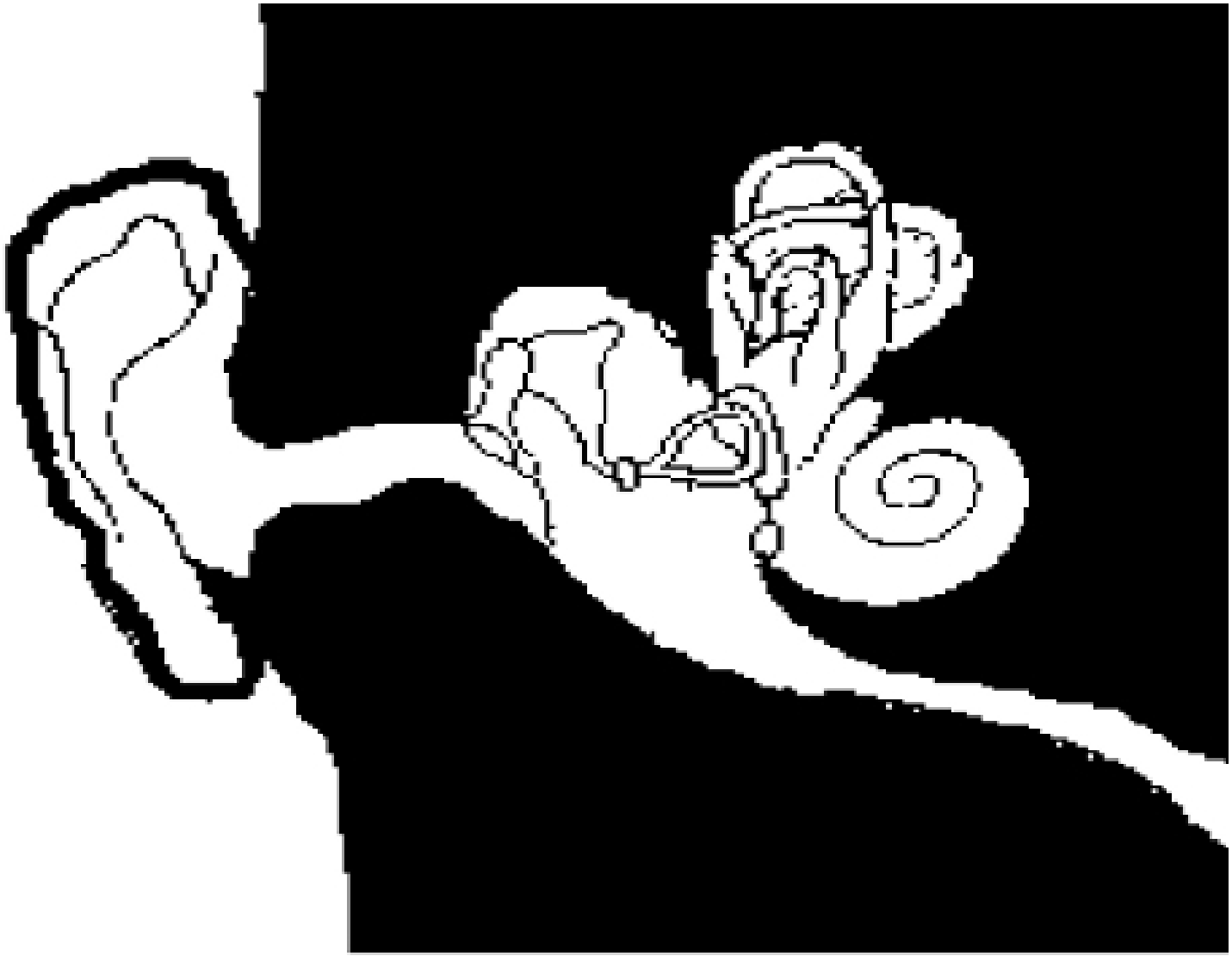


- pří dýchání jsou hlasivky rozevřeny
- při řeči se dutina zužuje a proudící vzduch je rozechvívá
- tím se vytváří základní hlasivkový tón
- ten je dále modifikován v hlasových dutinách:
  - hrtanové
  - nosohltanové
  - ústní

Vnímání zvuku

- sluchový orgán:
  - ušní boltec - zachycuje a koncentruje zvukovou energii
  - zvukovod - vede zachycenou energii k bubínku
  - ušní bubínek - rezonancí rozkmitán a přenáší vlnění na kůstky středního ucha:
    - kladívko
    - kovadlinka
    - třmínek
  - Eustachova trubice
    - vede ze středního ucha do dutiny ústní
    - slouží k vyrovnávání případných přetlaků (brání poškození středního a vnitřního ucha)
  - oválné okénko - jemná membrána tvořící rozhraní mezi středním a vnitřním uchem
  - hlemýžď (Cochlea)
    - součást vnitřního ucha
    - ústrojí ve tvaru ulity hlemýždě

Sluchový orgán (1.)

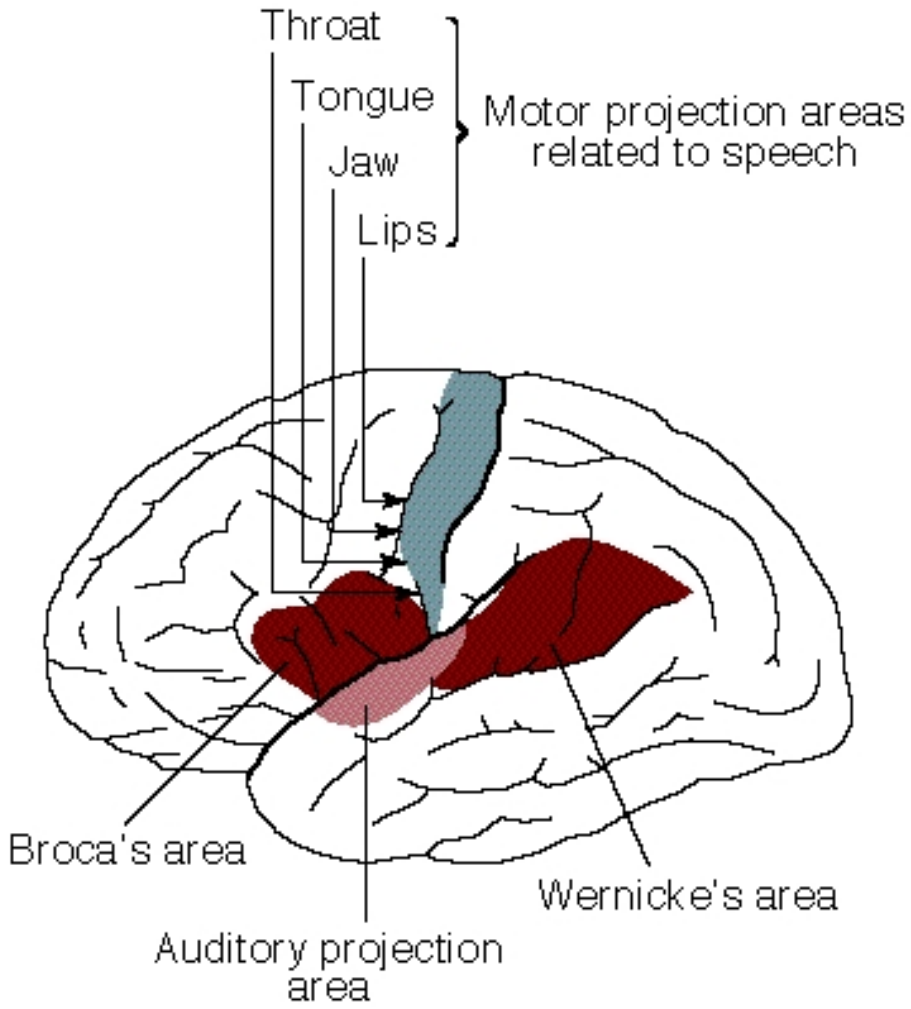


Vnímání zvuku (2.)

- Hlemýžď (Cochlea):
  - součást vnitřního ucha
  - ústrojí ve tvaru ulity hlemýždě
  - naplněno vodnatým mokem
  - obsahuje Cortiho ústrojí
    - obsahuje cca 20 000 vlákének
    - jejich délka od cca 40 μm do 0,5 mm
    - rezonují s jednotlivými tóny ve zvuku
    - vlákénka slouží k přenosu informací o jednotlivých složkách zvuku do mozku.

Zpracování zvuku v mozku

- Řečové centrum v mozku



Zpracování zvuku v mozku
pokračování

- Brocova oblast:
  - obsahuje artikulační vzorce - sekvence zapojení jednotlivých svalů potřebných k vyslovení slova
  - Brocova expresivní, motorická - afázie - rozumí řeči, má problémy s výslovností:
    - vynechávání slov
    - telegrafická kvalita řeči
    - řeč je kostrbatá
    - ...
- Wernickeho oblast
  - obsahuje sluchové vzorce a významy slov.

Fonetika

- Přírodní věda na pomezí lingvistiky, anatomie, fyziologie a fyziky (akustiky)
- Zkoumá zvukovou stánku jazyka z různých aspektů jazyka:.
  - fyziologickou činnost mluvidel
  - akustickou podstatu zvuků.
- Dělení fonetiky:
  - artikulační - tvorba fónů ve zvukovém ústrojí
  - akustická - přenos zvuků prostředím, jejich frekvence, ...
  - percepční - jak jsou zvuky přijímány,

Základy fonetiky

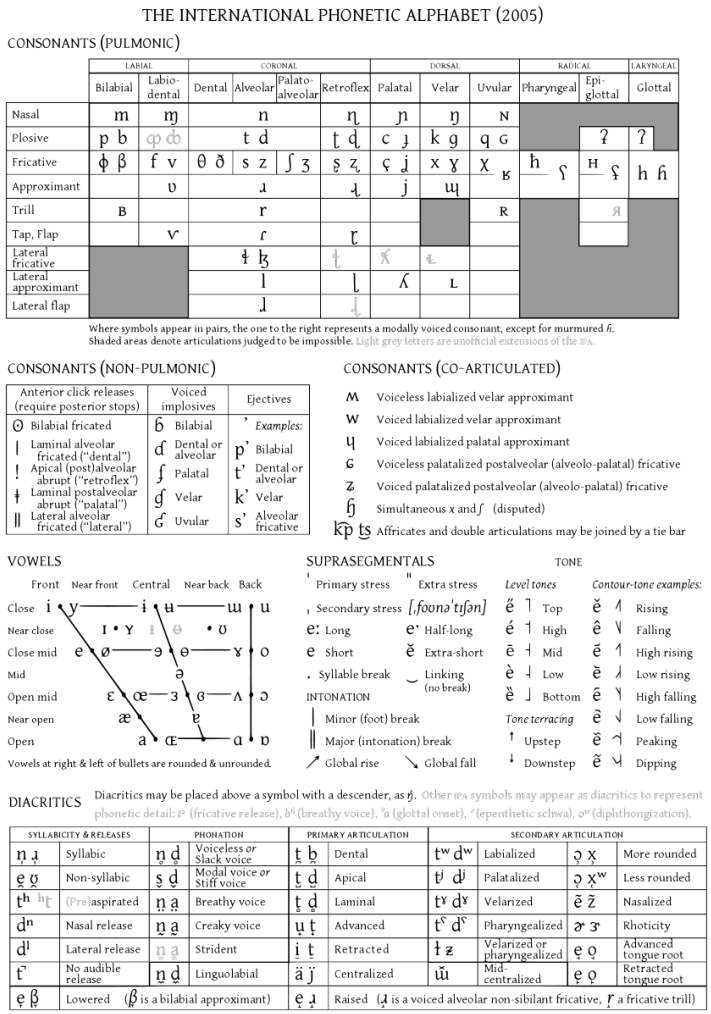
- Foném - elementární zvukový segment, který je schopný diferencovat vyšší znakové jednotky jazykového systému (morfémy).
- Fonémy:
  - samohlásky:
    - základní frekvence
    - formanty
  - samohlásky:
    - znělé - na vzniku se spolupodílí hlasivky
    - neznělé.
- Koartikulace - vzniká změnou parametrů řečového ústrojí při přechodu z jedné hlásky na druhou
  - další řečové jednotky:
    - alofón
    - difón
    - trifón

- Fonetický přepis - psaná a mluvená forma téže promluvy se mohou lišit.
  - Jednoznačný a přesný zápis mluvené řeči.

Fonetický přepis

- Přesný a jednozačný zápis mluvené řeči.
- IPA - International Phonetic Alphabet
  - součást standardu UNICODE.
- Národní fonetické abecedy.
- TTS - většinou využívají 7bitový ASCII přepis znaků z IPA (SAMPA - Speech Assessment Methods Phonetic Alphabet)např. DITe
- pravidla pro přepis:
  - změna znělosti na hranici znělá souhláska - neznělá souhláska
  - měkčení souhlásek, pokud následuje *i, ě*
  - ...
- Občas může být přepis regionálně závislý:
  - *sh*:
    - Čechy - sch
    - Morava - zh

IPA



THE INTERNATIONAL PHONETIC ALPHABET (2005)

CONSONANTS (PULMONIC)

	LABIAL		CORONAL				DORSAL			RADICAL		LARYNGEAL
	Bilabial	Labio-dental	Dental	Alveolar	Palato-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Epi-glottal	Glottal
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ			
Plosive	p b	ɓ		t d		ʈ ɖ	c ɟ	k g	q ɢ		ʡ	ʔ
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ʜ ʢ	h ɦ
Approximant		ʋ		ɹ		ɻ	j	ɰ				
Trill	ʙ			r					ʀ		ʀ̠	
Tap, Flap		ⱱ		ɾ		ɽ						
Lateral fricative				ɬ ɮ								
Lateral approximant				l		ɭ	ʎ	ʟ				
Lateral flap				ɺ								

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured ɦ.
Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

CONSONANTS (NON-PULMONIC)

Anterior click releases (require posterior stops)	Voiced implosives	Ejectives
ʘ Bilabial fricated	ɓ Bilabial	' *Examples:*
ǀ Laminal alveolar fricated ("dental")	ɗ Dental or alveolar	pʼ Bilabial
ǃ Apical (post)alveolar abrupt ("retroflex")	ʄ Palatal	tʼ Dental or alveolar
ǂ Laminal postalveolar abrupt ("palatal")	ɠ Velar	kʼ Velar
ǁ Lateral alveolar fricated ("lateral")	ʛ Uvular	sʼ Alveolar fricative

CONSONANTS (CO-ARTICULATED)

ʍ Voiceless labialized velar approximant
w Voiced labialized velar approximant
ɥ Voiced labialized palatal approximant
ɕ Voiceless palatalized postalveolar (alveolo-palatal) fricative
ʑ Voiced palatalized postalveolar (alveolo-palatal) fricative
ʩ Simultaneous x and ʃ (disputed)
k͡p t͡s Affricates and double articulations may be joined by a tie bar

VOWELS

	Front	Near front	Central	Near back	Back
Close	i y		ɨ ʉ		ɯ u
Near close		ɪ ʏ		ʊ	
Close mid	e ø		ɘ ɵ		ɤ o
Mid			ə		
Open mid	ɛ œ		ɜ ɞ	ʌ	ɔ
Near open	æ		ɐ		
Open	a	ɶ		ɑ	ɒ

Vowels at right & left of bullets are rounded & unrounded.

SUPRASEGMENTALS

ˈ Primary stress ˌ Extra stress [ˌfoʊnəˈtɪʃən]
ˌ Secondary stress
eː Long eˑ Half-long
e Short ĕ Extra-short
. Syllable break ‿ Linking (no break)

INTONATION

| Minor (foot) break
‖ Major (intonation) break
↗ Global rise ↘ Global fall

TONE

Level tones		*Contour-tone examples:*	
e̋ ˥	Top	ě ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e᷄ ˧˥	High rising
è ˨	Low	e᷅ ˩˧	Low rising
ȅ ˩	Bottom	e᷈ ˥˧	High falling
Tone terracing		e᷌ ˧˩	Low falling
↑	Upstep	e᷉ ˩˥˩	Peaking
↓	Downstep	e᷊ ˥˩˥	Dipping

DIACRITICS Diacritics may be placed above a symbol with a descender, as ŋ̊. Other IPA symbols may appear as diacritics to represent phonetic detail: tˢ (fricative release), bʱ (breathy voice), ʔa (glottal onset), ᵊ (epenthetic schwa), oᵘ (diphthongization).

SYLLABICITY & RELEASES		PHONATION		PRIMARY ARTICULATION		SECONDARY ARTICULATION			
n̩ ɹ̩	Syllabic	n̥ d̥	Voiceless or Slack voice	t̪ b̪	Dental	tʷ dʷ	Labialized	ɔ̹ χ̹	More rounded
e̯ o̯	Non-syllabic	s̬ d̬	Modal voice or Stiff voice	t̺ d̺	Apical	tʲ dʲ	Palatalized	ɔ̜ χ̜ʷ	Less rounded
tʰ ʰt	(Pre)aspirated	n̤ a̤	Breathy voice	t̻ d̻	Laminal	tˠ dˠ	Velarized	ẽ z̃	Nasalized
dⁿ	Nasal release	n̰ a̰	Creaky voice	u̟ t̟	Advanced	tˤ dˤ	Pharyngealized	ɚ ɝ	Rhoticity
dˡ	Lateral release	n̼ a̼	Strident	i̠ t̠	Retracted	ɬ̪ ʑ	Velarized or pharyngealized	e̘ o̘	Advanced tongue root
t̚	No audible release	n̪ d̪	Linguolabial	ä ï	Centralized	ů	Mid-centralized	e̙ o̙	Retracted tongue root
e̞ β̞	Lowered (β̞ is a bilabial approximant)			e̝ ɹ̝	Raised (ɹ̝ is a voiced alveolar non-sibilant fricative, r̝ a fricative trill)				

Podrobnosti viz `http://www.langsci.ucl.ac.uk/ipa/`

Česká fonetická abeceda

- Krátké samohlásky:
  - a pata, e led, i lid, o rod, u ruka
- Dlouhé samohlásky:
  - a:, á pátá; e:, é léto; i:, í lípa; o:, ó tón; u:, ú úkol
- Dvojhlásky:
  - au auto; eu euforie; ou houba

Česká fonetická abeceda

- Souhlásky:
  - m matka; μ tramvaj; n nos ň kůň nk banka
  - p pes; b babička; t táta; d dům; ť ťapka; ď ďábel; k kost; g gram
  - c co; dz; leckdo; č čáp; dž džem
  - f fuj; v voda; s sen; z zub; š šíp; ž žena; x, ch chléb; ych abych byl; h hra
  - r rak; ř řeka; ř rybář
  - j j já; l l les
  - r krk; l vlk; m Rožmberk

Ukázka textu v abecedě SAMPA spolu s výsledkem.

- Fonetický přepis věty "Čeština je krásná řeč"
  - tSeSTina je kra:sna: r/etS
  - Syntetizovaná věta "Čeština je krásná řeč." (`data/cestina.wav`)

Fonetika - samohlásky

- Krátké samohlásky - a, e, i, o, u
- Dlouhá samohlásky - á, é, í, (ó), ú
- Dvojhlásky - (eu), (au), ou
- Samohlásky:
  - základní hlasivkový tón - 100 - 400 Hz
  - formanty - rezonancí v dutinách hlasového traktu zesílené části akustického spektra

Formanty

- Určující pro rozpoznávání samohlásek
- Formant F1 vzniká rezonancí v dutině ústní
- Formant F2 vzniká rezonancí v dutině hrdelní
- Hlavní formanty - spektrální poloha a intenzita může být dána:
  - muž
  - žena
  - dítě
  - individuálně
- Vyšší formanty F3 -
  - výskyt bývá individuální

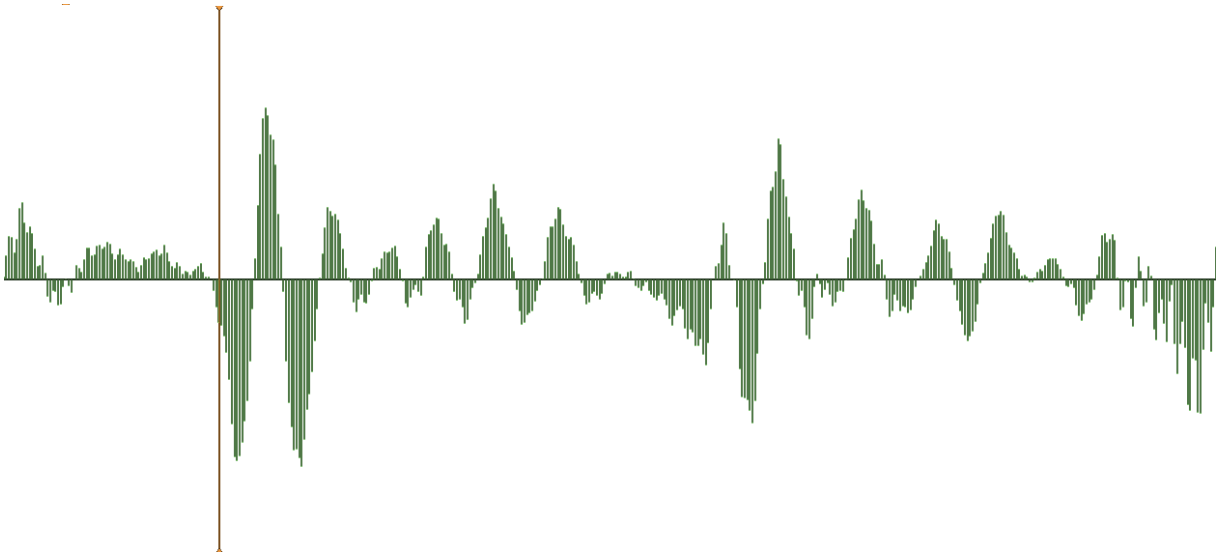
Formanty F1 a F2 pro české samohlásky

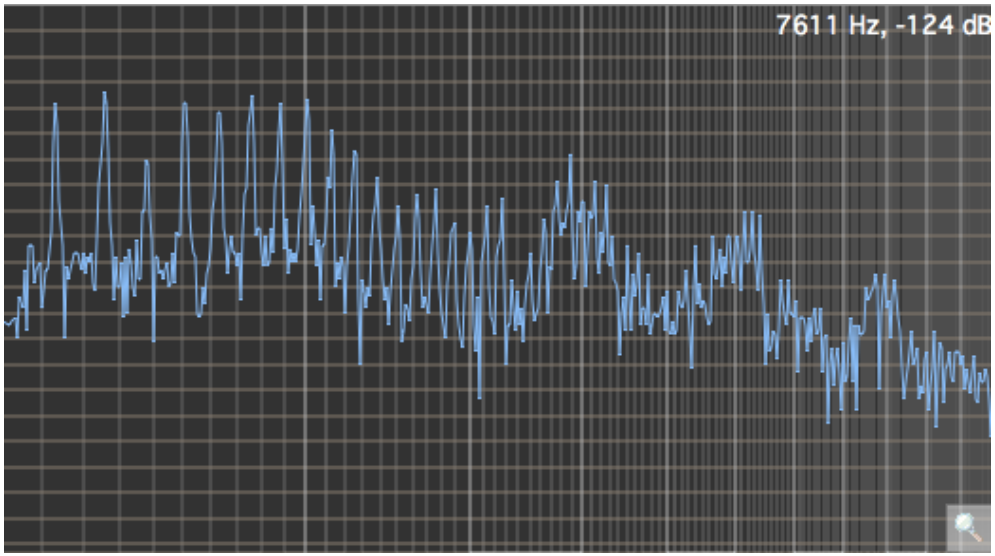
hláska	F1 [Hz]	F2 [Hz]
a	750 - 1100	1100 - 1500
e	500 - 700	1500 - 2000
i	300 - 500	2000 - 3000
o	500 - 700	900 - 1200
u	300 - 500	600 - 1000

Četnost výskytu samohlásek

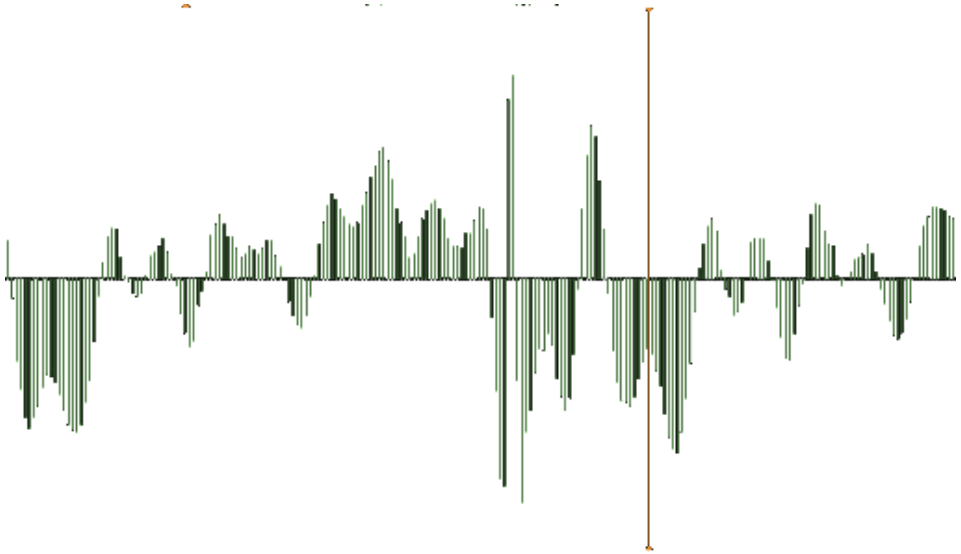
- e - 10
- a, o, i - 6 - 7
- í - 4
- další jen s nepatrnou frekvencí:
  - á, u, é, ou, ú
  - ó, au, eu

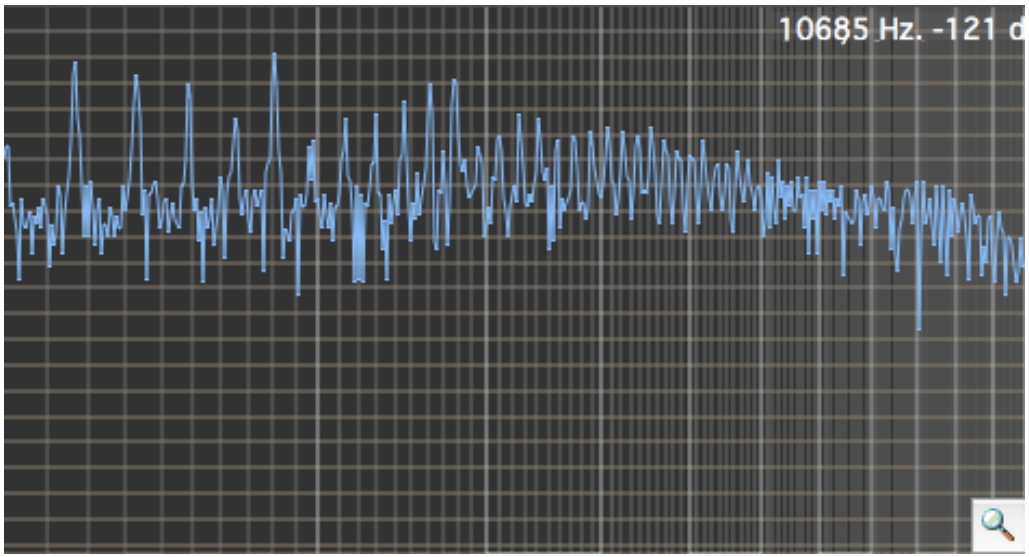
České samohlásky - a



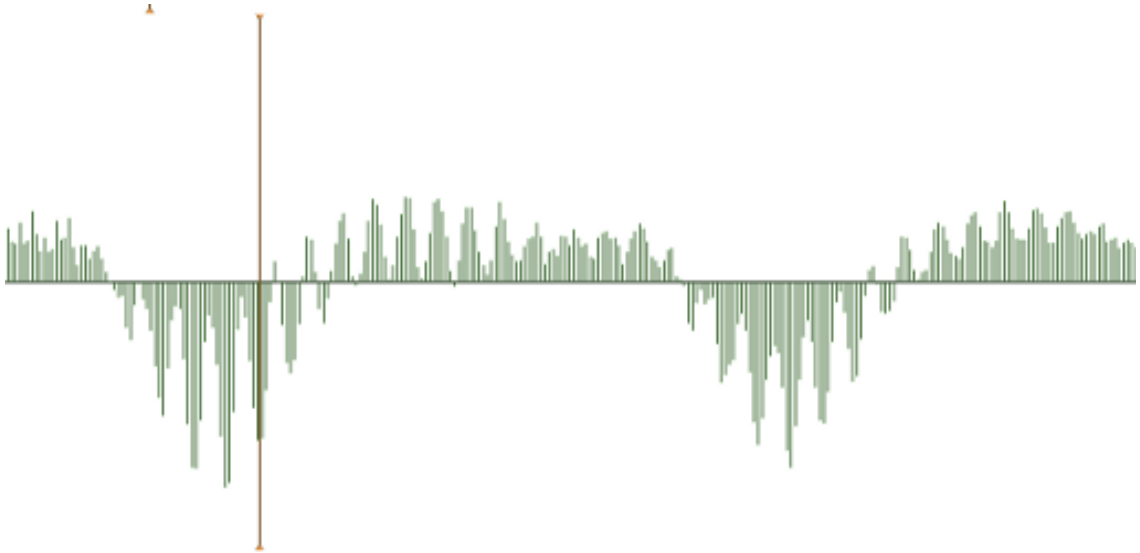


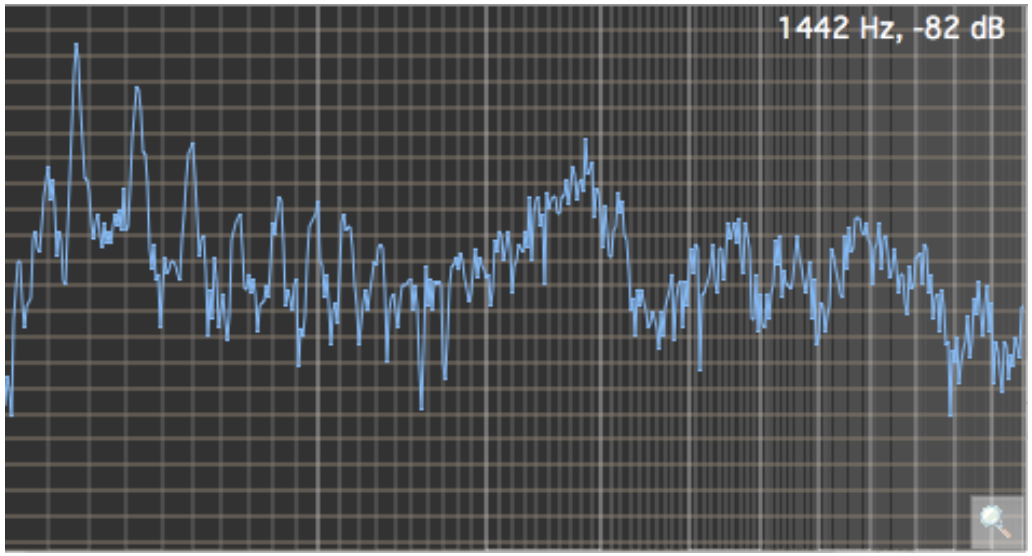
České samohlásky - e



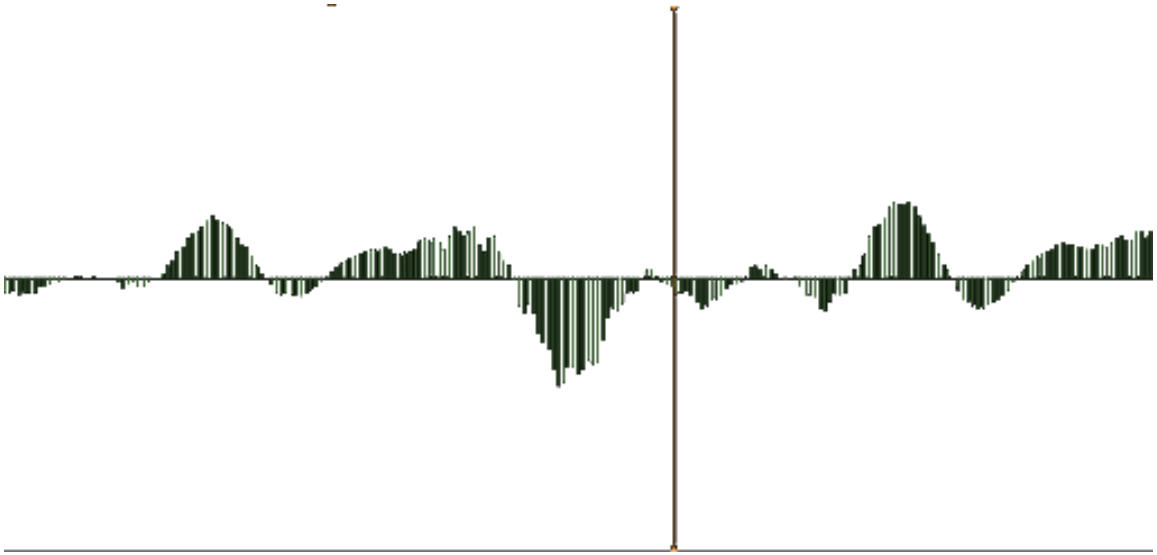


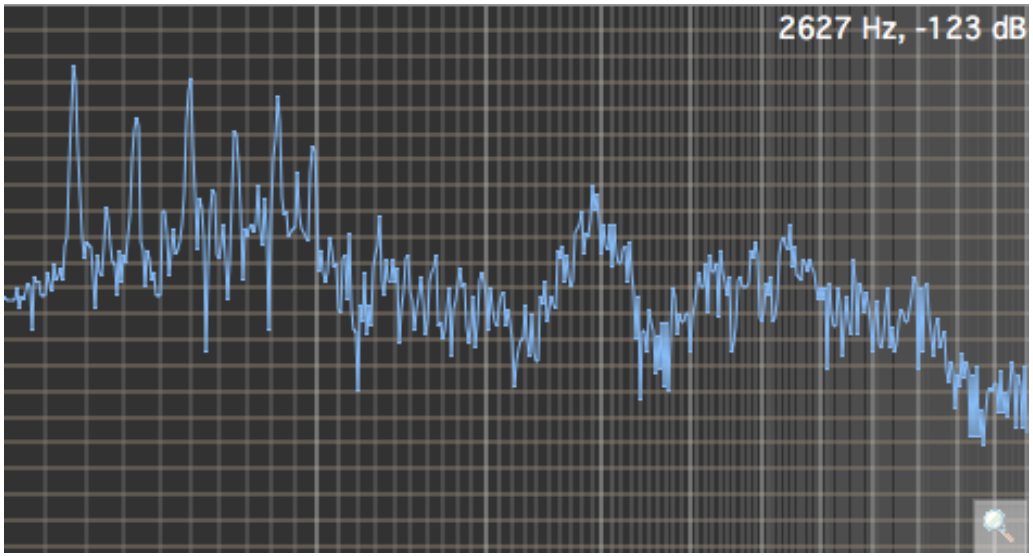
České samohlásky - i



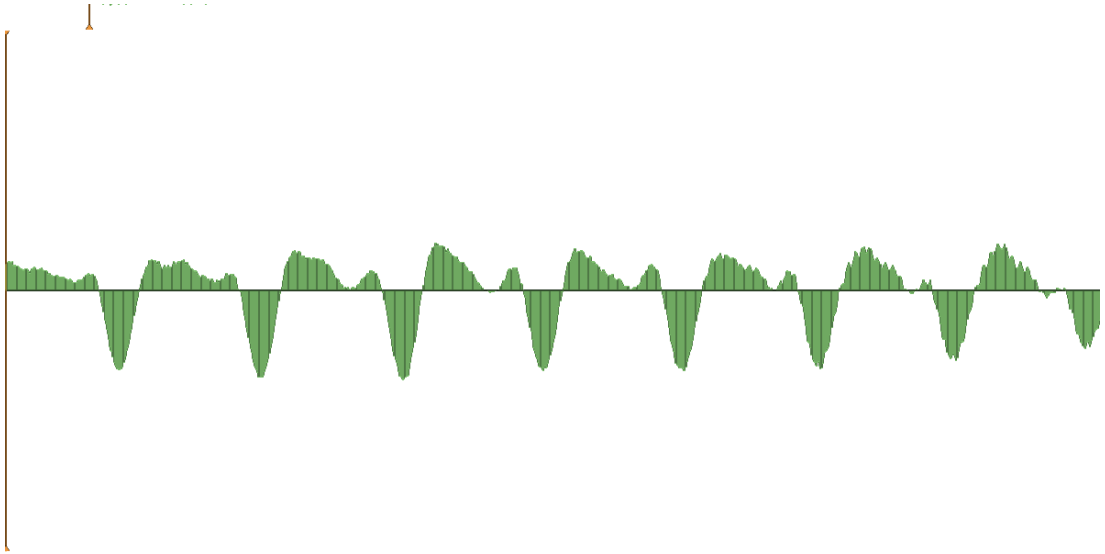


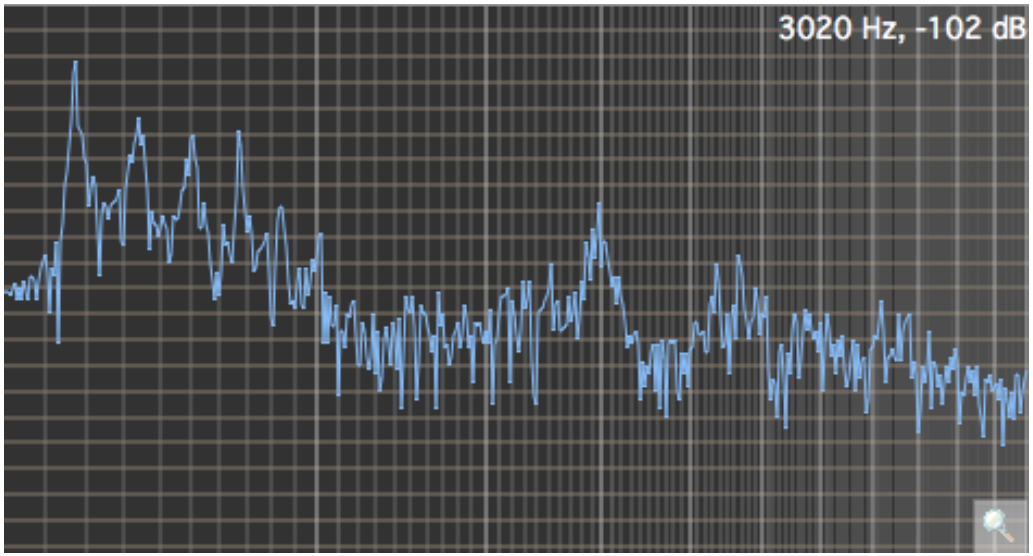
České samohlásky - o



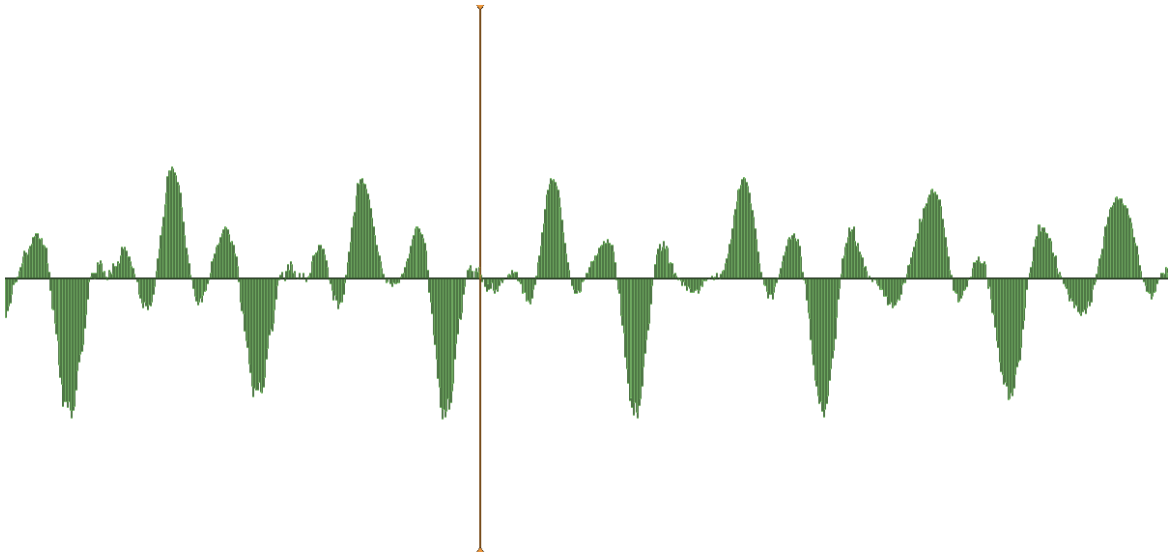


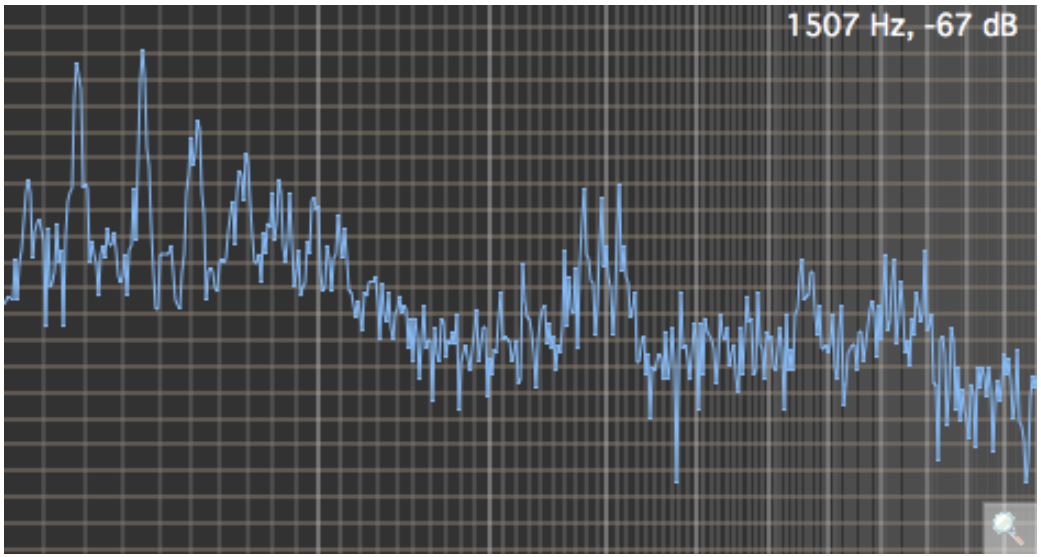
České samohlásky - u



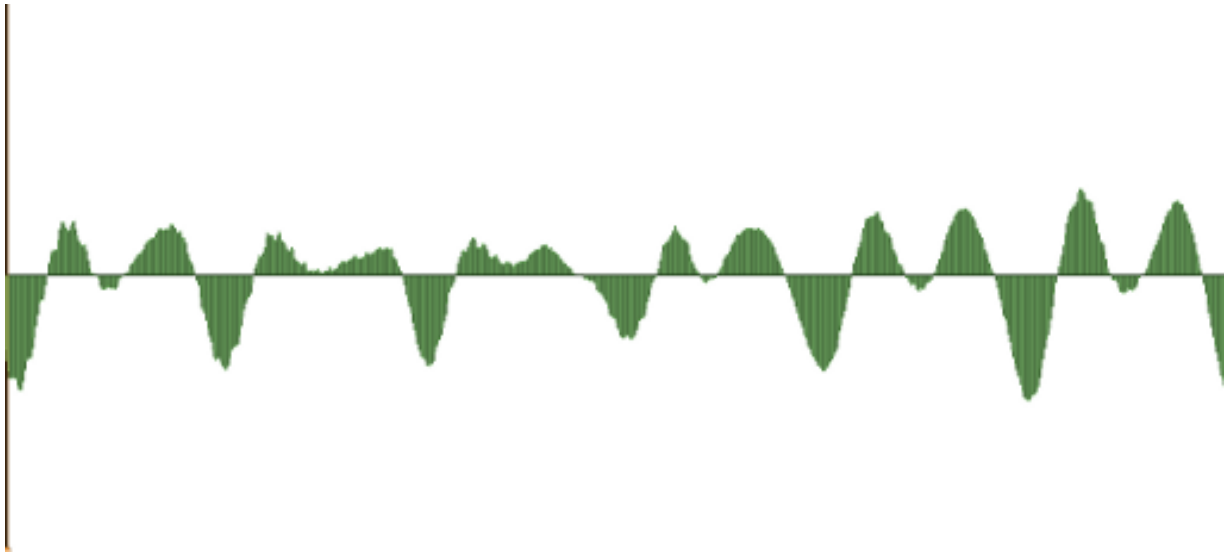


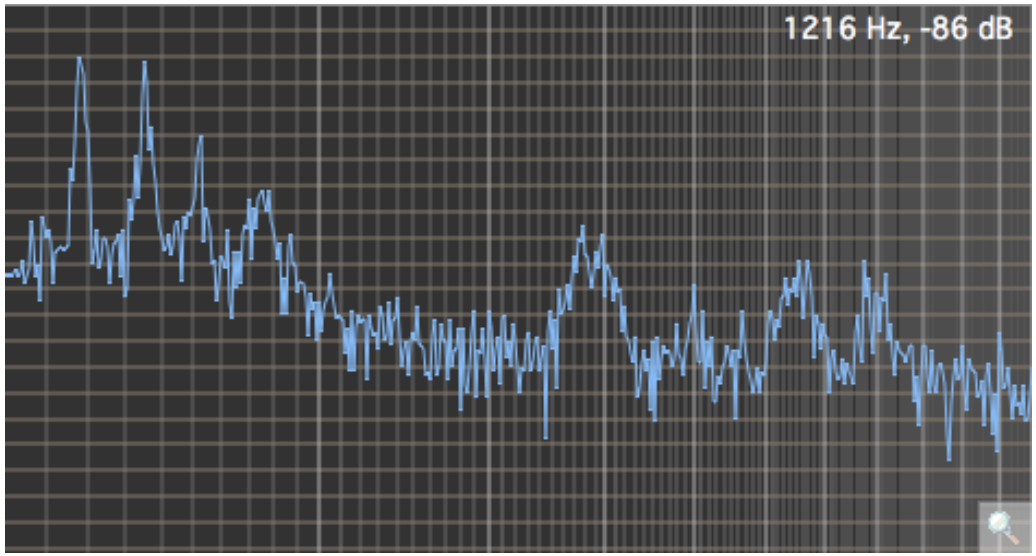
České dvojhlásky - au



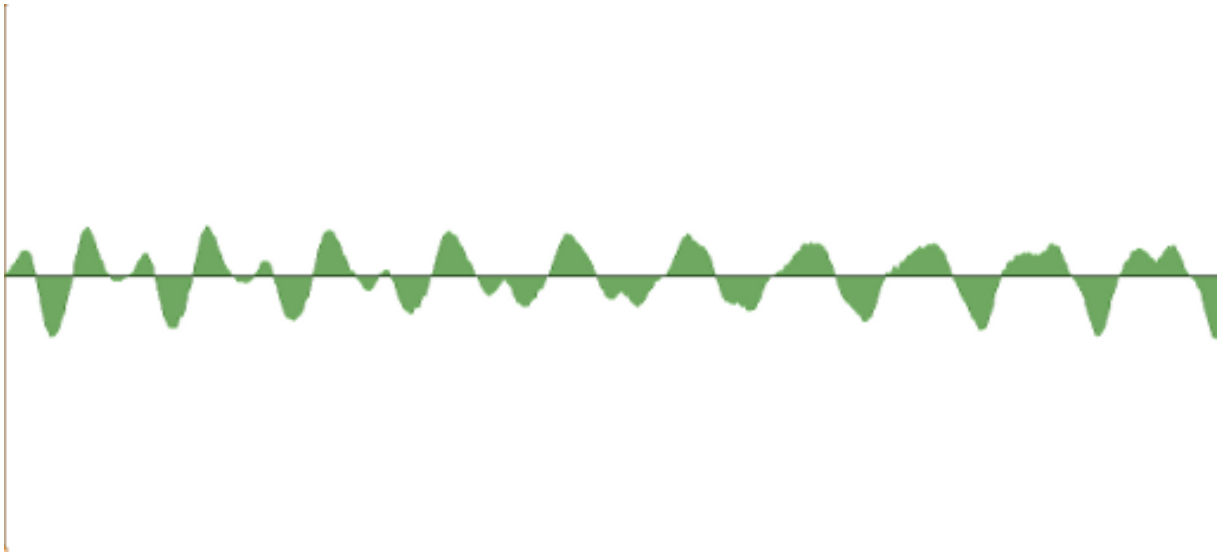


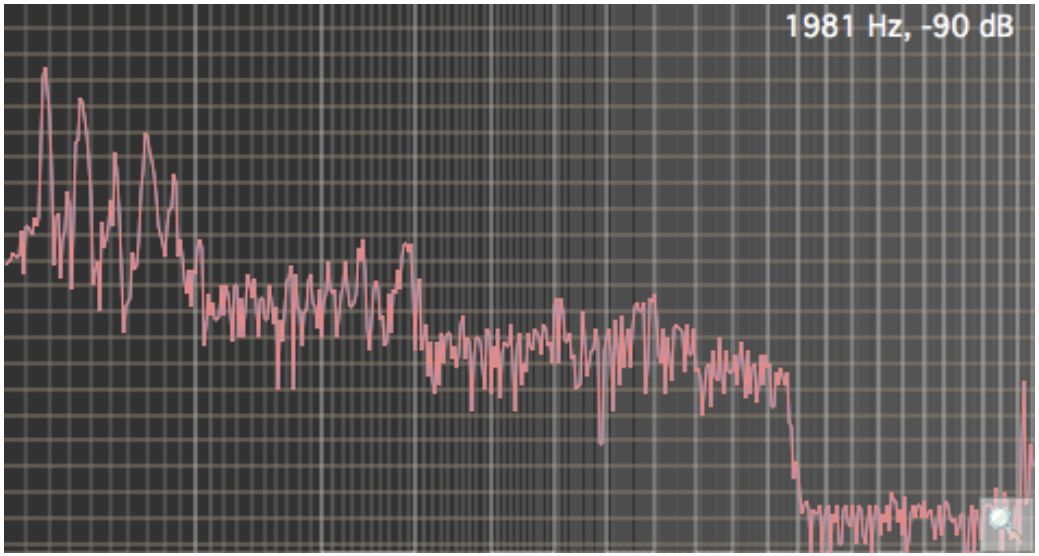
České dvojhlásky - ou





České dvojhlásky - eu





Souhlásky (konsonanty)

- Zvukově dynamické děje.
- Pojem formantu ztrácí význam
  - tónový charakter mají pouze části některých souhlásek.
- Klasifikace:
  - znělé (sonorní)
  - neznělé (šumové)
  - fonetikové dále podle místa a způsobu artikulace na:
    - retné - m, b, p, w, v, f
    - zubní - n, d, t, dh, th
    - dásňové - c, z, s, dz
    - patrové - ň, ď, ť, ž, š
    - závěrové (okluzívy, režené, explozívy) - b, d, ď, g, p, ť
    - úžinové - v, z, ž, f, th, s, š, ...

Znělé a neznělé souhlásky

- Znělé souhlásky
  - charakteristické přítomností základního tónu
  - na vytváření se aktivně podílejí hlasivky.
- Neznělé souhlásky
  - hlasivky jsou pasivní (otevřené)
- Párové
  - neliší se artikulací, pouze znělostí
  - např. b-p, d-t. z-s, ...

Porovnání párových souhlásek (waveform)
S – Z

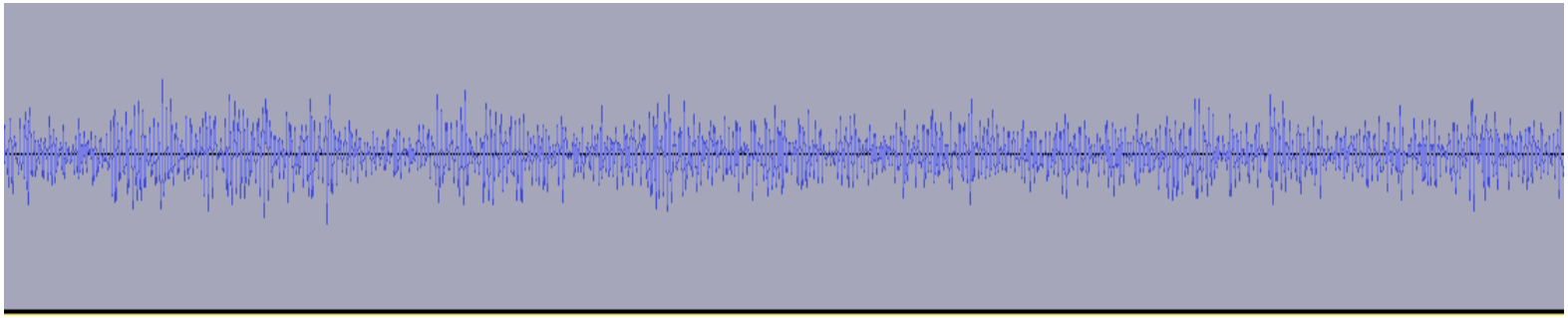


Figure: Průběh hlásky S

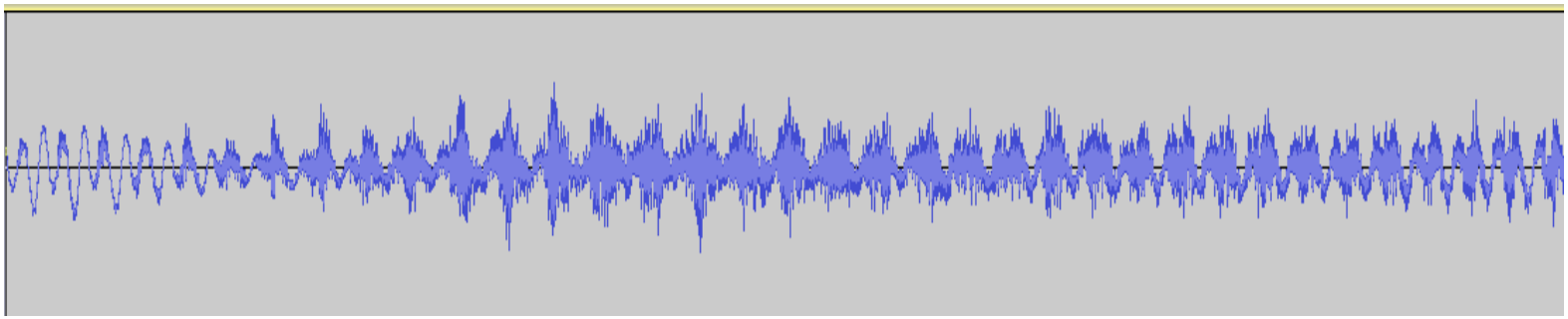


Figure: Průběh hlásky Z

Porovnání párových souhlásek (spektrum)
s – z

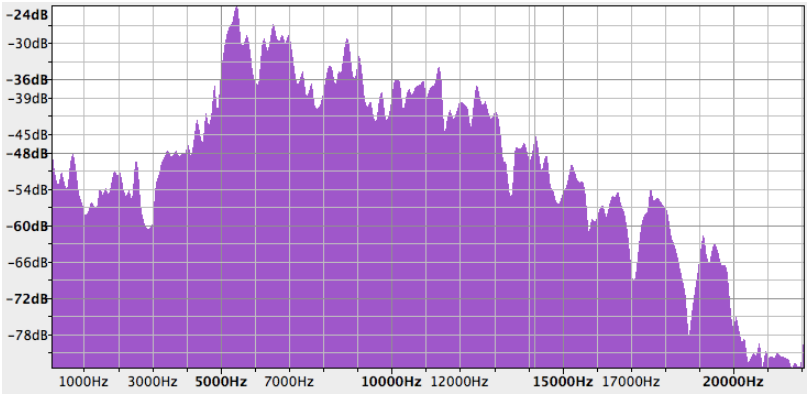


Figure: Spektrum hlásky s

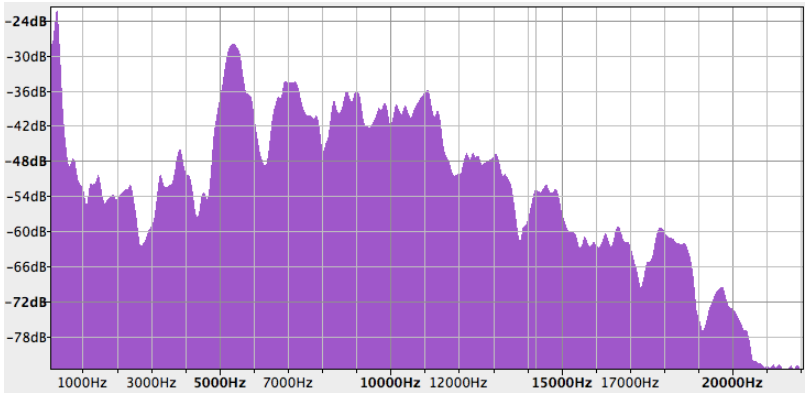


Figure: Spektrum hlásky z

Podle způsobu tvorby

- Okluzívy
  - závěrové souhlásky
  - vytvořena překážka výdechovému proudu vzduchu:
    - jazyk
    - zuby
    - rty
  - (p, b), (t, d), (ť, ď). (k, g), m, n, ň
- Frikativy
  - úžinové
  - zúžení výdechové cesty při artikulaci
  - (s, z), (š, ž), (f, v), (ch, h), l, j, r, ř
- Semiokluzívy
  - polouzávěrové
  - vytváří se jak překážkou, tak zúžením výdechové cesty
  - c, č

Ukázka okluzívy

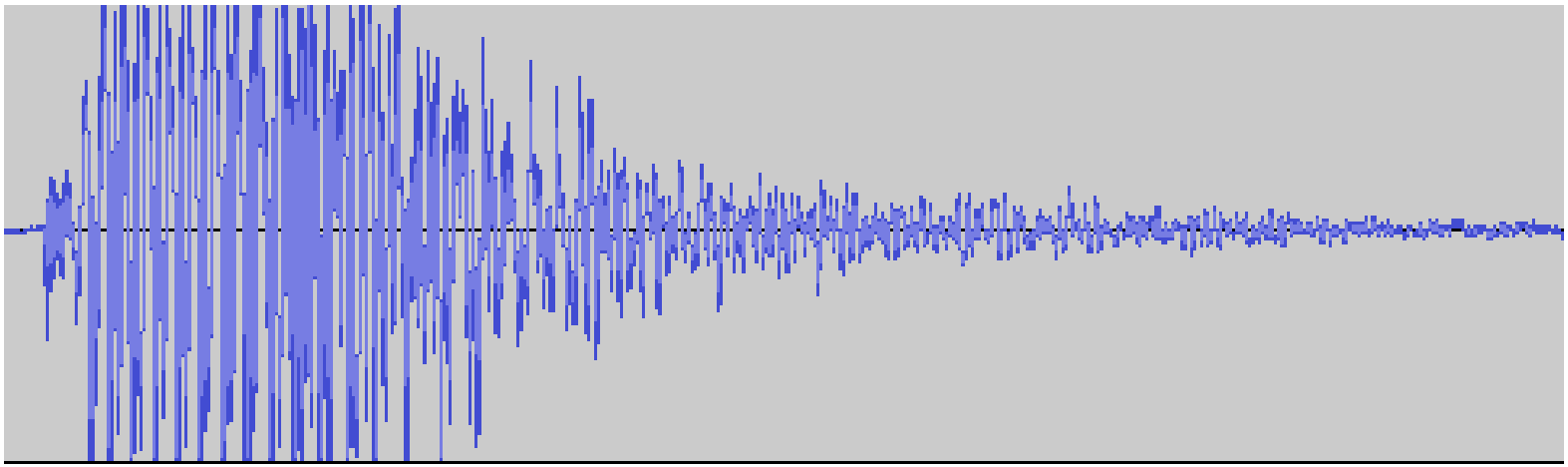


Figure: Průběh hlásky p

Ukázka frikativy

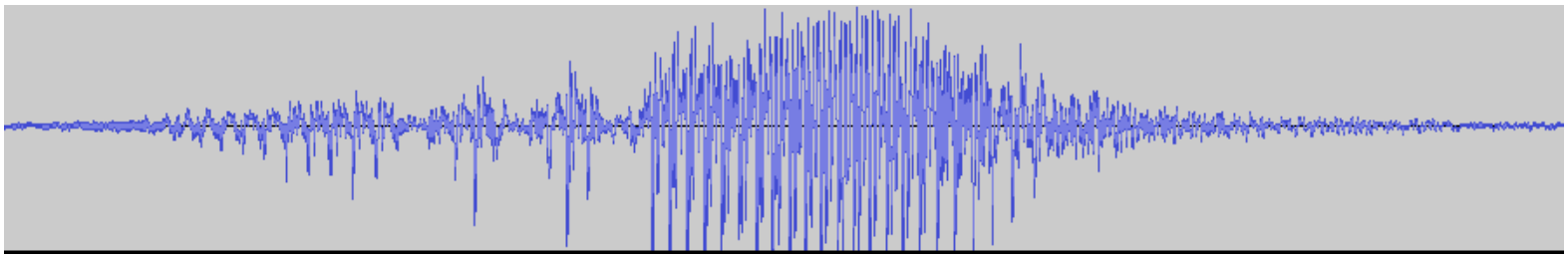


Figure: Průběh hlásky r

Ukázka semiokluzívy

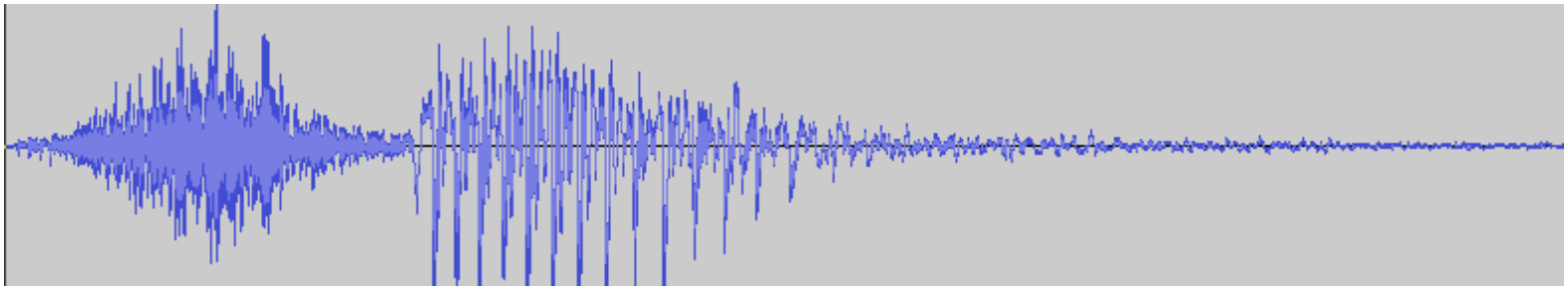


Figure: Průběh hlásky c

Koartikulace

- Modifikace fonému v řečovém kontextu.
- Nutnost přenastavit řečový trakt na další foném.
- Způsobuje problémy při:
  - syntéze řeči
  - rozpoznávání řeči.

Ukázka vlivu koartikulace - původní fonémy

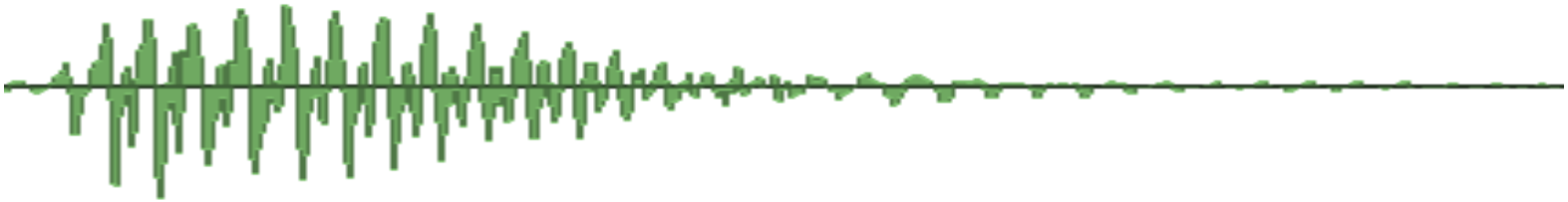


Figure: Samotné p

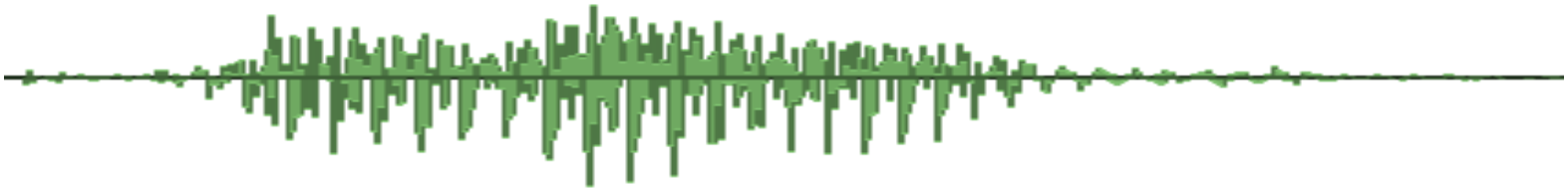


Figure: Samotné á

Ukázka vlivu koartikulace (2.)

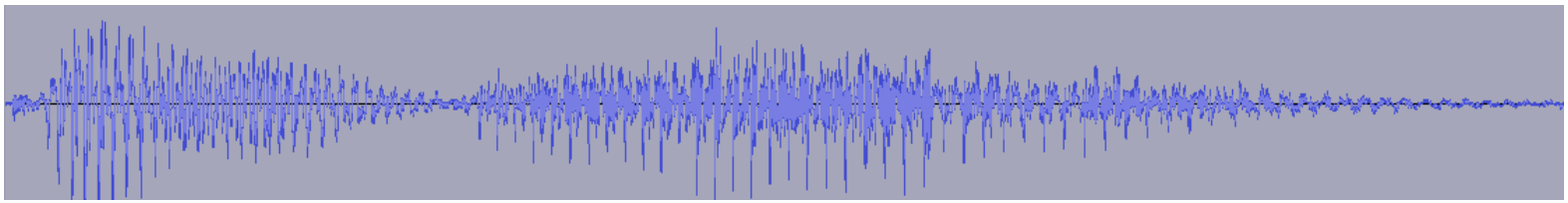


Figure: Spojení fonémů p a á

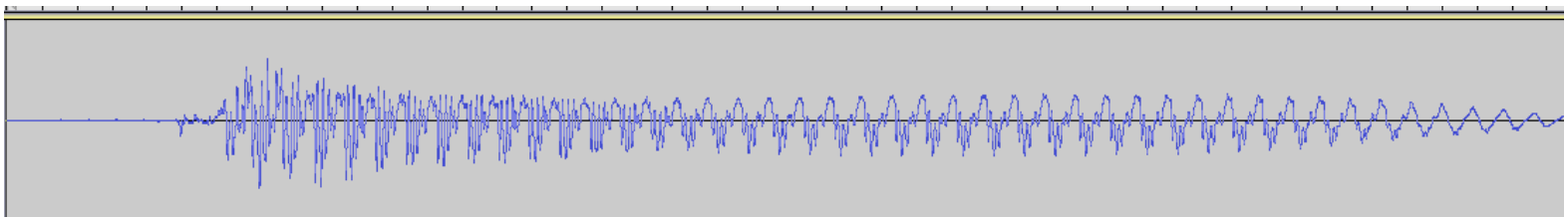


Figure: Vyslovená slabika pá

Digitalizace zvuku

- Cíl - převod spojitého signálu na posloupnost digitálních hodnot vhodných pro uchování v počítači.
- Postup digitalizace:
  1. Vzorkování - převod reálných vstupních hodnot na posloupnost diskrétních reálných čísel.
  2. Kvantizace - převod posloupnosti reálných čísel na posloupnost celých čísel.
  3. Kódování - způsob uložení a kódování posloupnosti celočíselných hodnot získaných v kroku 2.

Vzorkování

- Transformace spojitého časové závislého signálu s(t) na časově diskrétní posloupnost s(nT) = 0, 1, 2, ...
  - T - perioda vzorkování.
  - Pokud nemá dojít ke ztrátě informace, musí být vzorkovací frekvence aspoň dvojnásobkem nejvyšší frekvence, která je signálu obsažena.
- Po čase T je sejmuta a dána na výstup (ke kvantizaci) hodnota ze vstupního snímače.
  - většinou okamžitá úroveň napětí nebo proudu na vstupu.

- Digitální zpracování zvuku:
  - audio CD
  - mp3 - navíc použita ztrátová komprese
  - miniDisc - navíc použita ztrátová komprese ATRAC
  - DAT
  - ...
- Digitální zpracování signálu obecně:
  - digitalizace dat z různých analogových měřících zařízení
  - digitální zpracování obrazu
  - ...

Ukázka digitalizovaného signálu

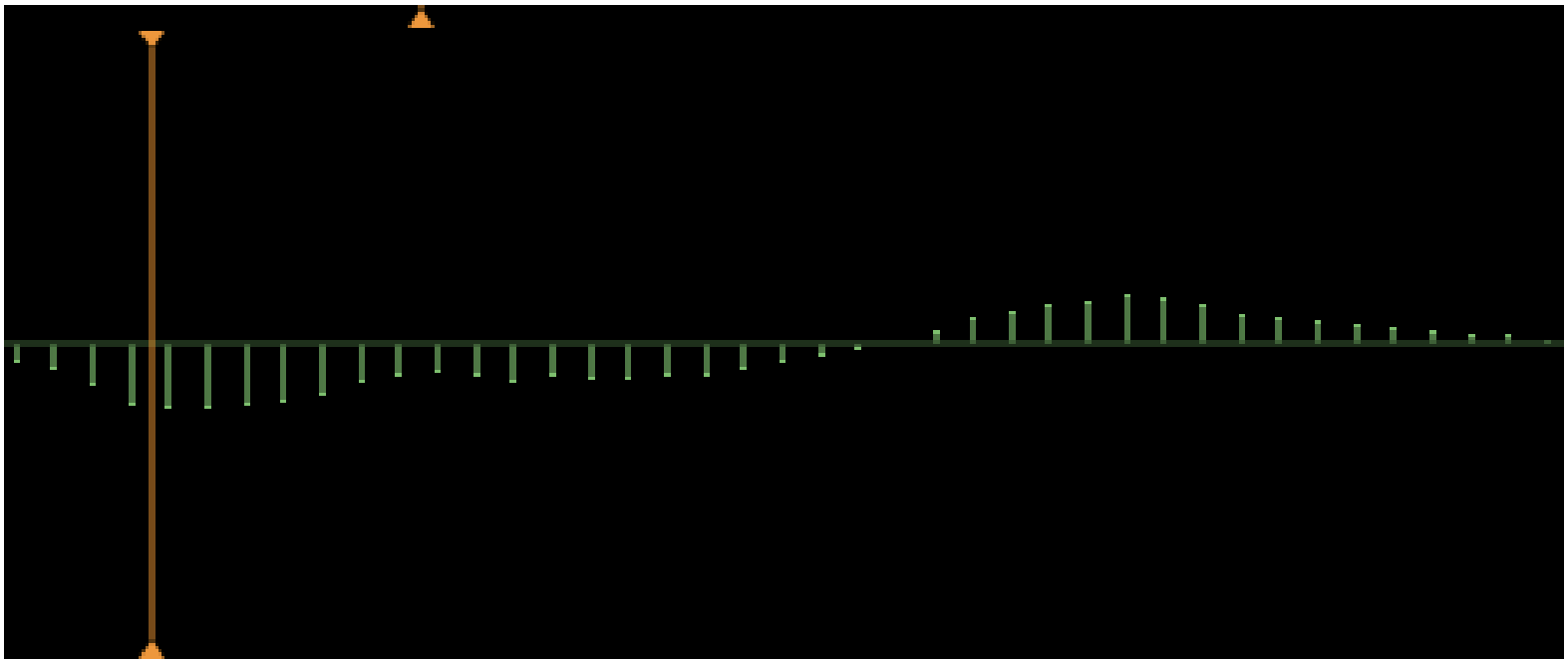


Figure: Ukázka digitalizovaného zvuku

Shanonův vzorkovací teorém

- Analogový signál s(t) lze rekonstruovat z hodnot vzorků s(nT) následovně:

$$s(t) = \sum_{n=-\infty}^{\infty} s(nT) \frac{sin(\pi(\frac{t}{T} - n))}{\pi(\frac{t}{T} - n)}$$

  právě tehdy když je vzorkovací frekvence alespoň dvojnásobkem nejvyšší frekvence obsažené ve vstupním signálu.
- Důsledky:
  - Vzorkovací frekvence by měla být alespoň dvojnásobkem nejvyšší frekvence vstupního signálu.
  - Je-li menší dochází ke zkreslení složek vyšších frekvencí.
  - Spor příznivců a odpůrců audio CD - je 44kHz dostačující vzorkovací frekvence pro hudbu?

Kvantizace

- Převod reálných navzorkovaných hodnot na celočíselné hodnoty.
- Počet celočíselných hodnot = počet úrovní kvantování
  - 256
  - 65 536
  - 16 777 216
- Kvantizační krok - reálný interval přiřazený kvantizované jednotce.
  - Na vstupu je signál s amplitudou 128 mA (-128 - 128 mA).
  - 8bitová kvantizace - 256 kvantizačních úrovní
  - kvantizační krok = 256 [mA]/256 [kvantizačních úrovní] = 1[mA].

Kvantizace
pokraačování

- Běžně používané kvantizace:
  - 8 bitů
  - 16 bitů
  - 24 bitů
- Realizováno pomocí A/D převodníků
  - součást zvukových karet
  - mobilních telefonů
  - ...

Běžně používané parametry digitalizace zvuku

- Vzorkovací frekvence:
  - 8 kHz - telefonní kvalita
  - 16 kHz - běžná řeč
  - 22 kHz - rozhlasová kvalita
  - 44 kHz - audio CD
  - 48 kHz - DVD
- kvantizace:
  - 8 bitů
  - 16 bitů
  - 24 bitů
- počet audio kanálů
  - 1
  - 2
  - 4
  - 6 (5.1, 5 směrových kanálů + basy)

Způsoby kódování signálu

- PCM - přímé ukládání hodnot získaných kvantizací.
- výhody:
  - jednoduché na zpracování
  - nedochází k další ztrátě informací
- nevýhody:
  - Často malé rozdíly mezi hodnotami sousedních vzorků - značná redundance dat.
  - Konstantní hodnota kvantizačního kroku (závisí na parametrech AD převodníku):,
    - V případě malé amplitudy vstupního signálu - ztráta informace (signál nepřekročí kvantizační krok).
    - V případě velké amplitudy - hodnota překročí rozsah - zkreslení signálu.
    - Oba případy brání kvalitní rekonstrukci původního signálu.

- Diferenční PCM - uchovávání rozdílů sousedních vzorků místo uchovávání jejich hodnot. Hodnota rozdílu bývá podstatně menší než hodnota vzorku - lze uchovat pomocí méně bitů.
- Adaptivní PCM - kvantizační krok se určuje na základě amplitudy vstupního signálu.

Metody krátkodobé analýzy

- Zvuk je "periodický" pouze na krátkém intervalu.
- Zpracování signálu na krátkém časovém intervalu, kde se nepředpokládají výraznější dynamické změny (*mikrosegment*).
  - velikost od 10 do 40 ms
- Metody krátkodobé analýzy:
  - v časové oblasti
  - ve frekvenční oblasti

Krátkodobá analýza

- Nevýhoda použití mikrosegmentu:
  - chyba způsobená předpokladem, že zvuk v okolí okénka zůstává periodický s periodou okénka
  - tuto chybu lze kompenzovat použitím okénka
- Okénko - posloupnost vah pro prvky mikrosegmentu
- Nejběžněji používané typy okének:
  - hammingovo
  - pravoúhlé

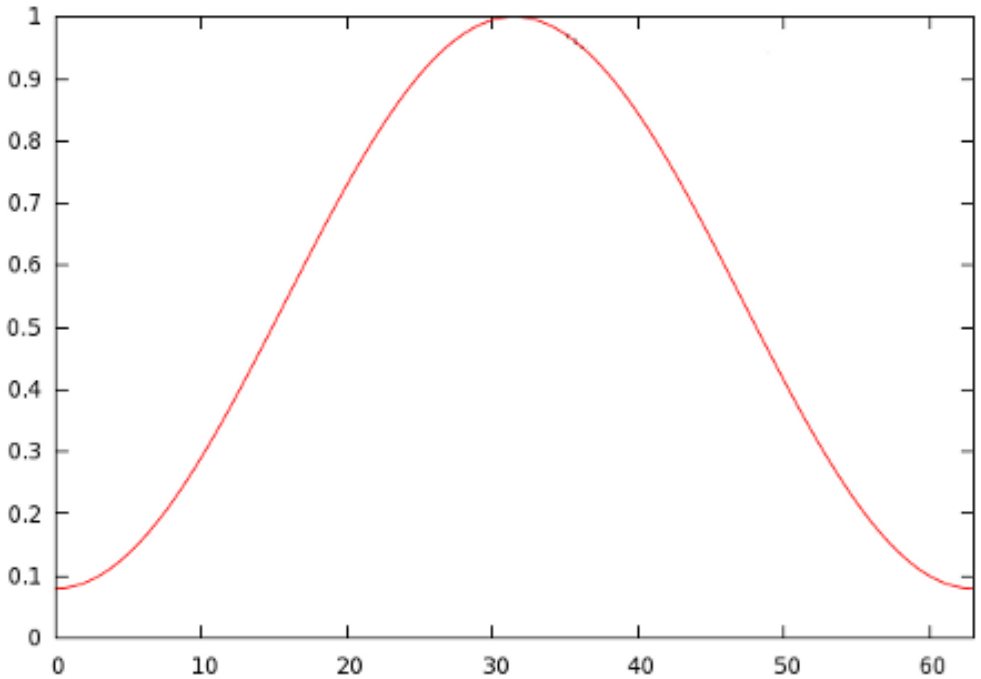
Hammingovo okénko

- Pro výpočet n-té váhy se využívá vztah

$$w(n) = \begin{cases} n = 0 \ldots N-1 & 0.54 - 0.46 cos(2\pi n/(N-1)) \\ n < 0 \vee n \geq N & 0 \end{cases}$$

  N - počet vzorků v mikrosegmentu

- Hammingovo okénko pro mikrosegment délky 64



Pravoúhlé okénko

- Přiřadí každému prvku mikrosegmentu váhu 1:

$$w(n) = \begin{cases} n = 0 \ldots N-1 & 1 \\ n < 0 \vee n \geq N & 0 \end{cases}$$

N - délka mikrosegmentu

Analýza digitalizovaného signálu v časové oblasti 1.

- Vychází se přímo z hodnot vzorků, nikoliv z hodnot spektra.
- Používá se funkce krátkodobé energie:

$$E(n) = \sum_{k=-\infty}^{\infty} (s(k)w(n-k))^2$$

  - Ukázka výpočtu funkce krátkodobé energie v Octave (`octave/ste.m`)
  - s(k) - vzorek v čase k, w(n-k) - váha odpovídajícího okénka pro čas k
  - Výstupem je průměrná energie vrámci segmentu.
  - Značně citlivá na velké změny úrovně signálu v rámci segmentu.
  - Druhá mocnina zvyšuje dynamiku zvukového signálu.

Analýza digitalizovaného signálu v časové oblasti 2.

- Funkce krátkodobé intenzity:

$$I(n) = \sum_{k=-\infty}^{\infty} |s(k)| w(n-k)$$

  - Používá se např. pro detekci ticha.

Ukázka průběhu funkce krátkodobé energie

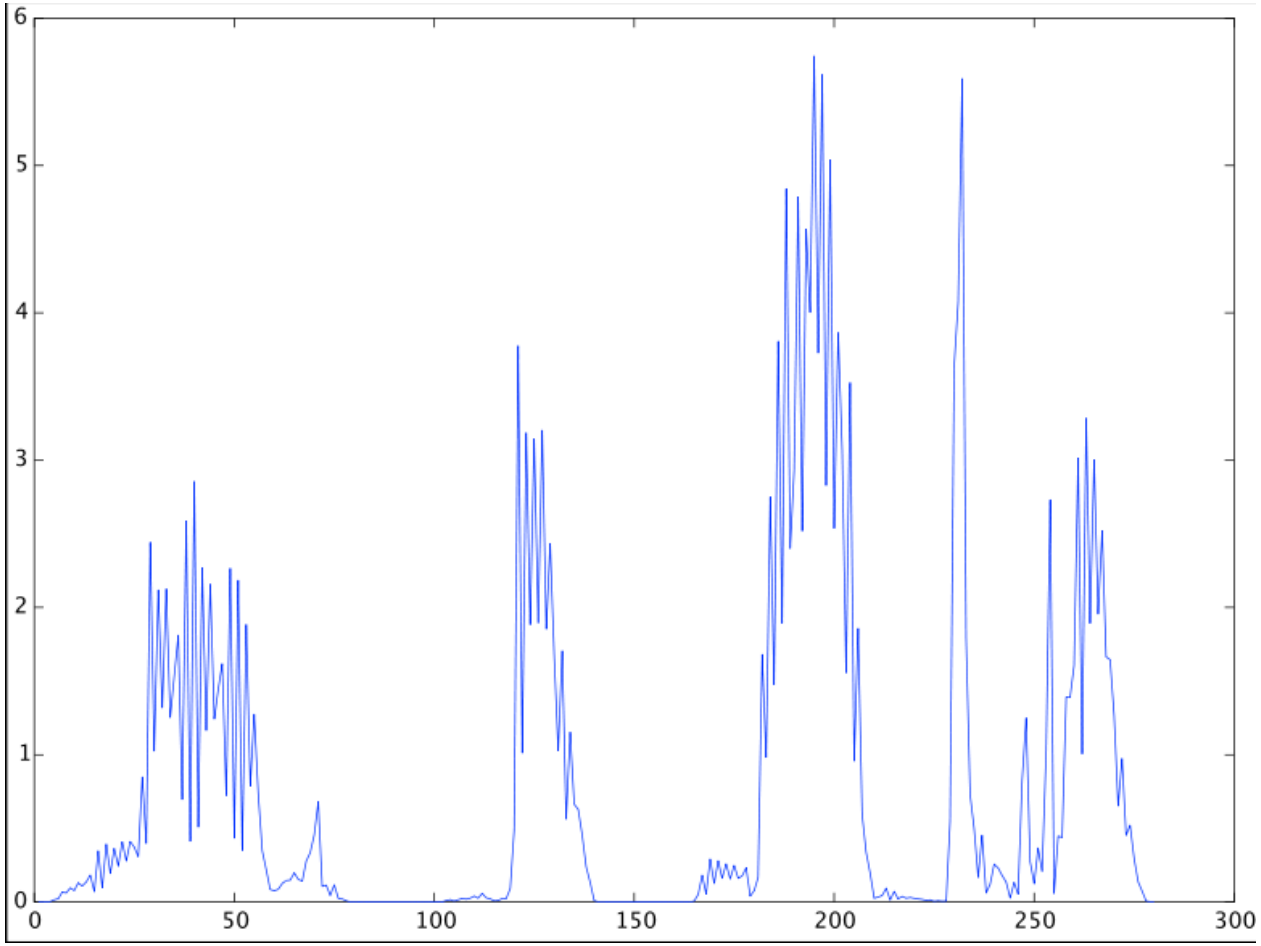


Figure: Ukázka průběhu funkce krátkodobé energie

Ukázka průběhu funkce krátkodobé intenzity

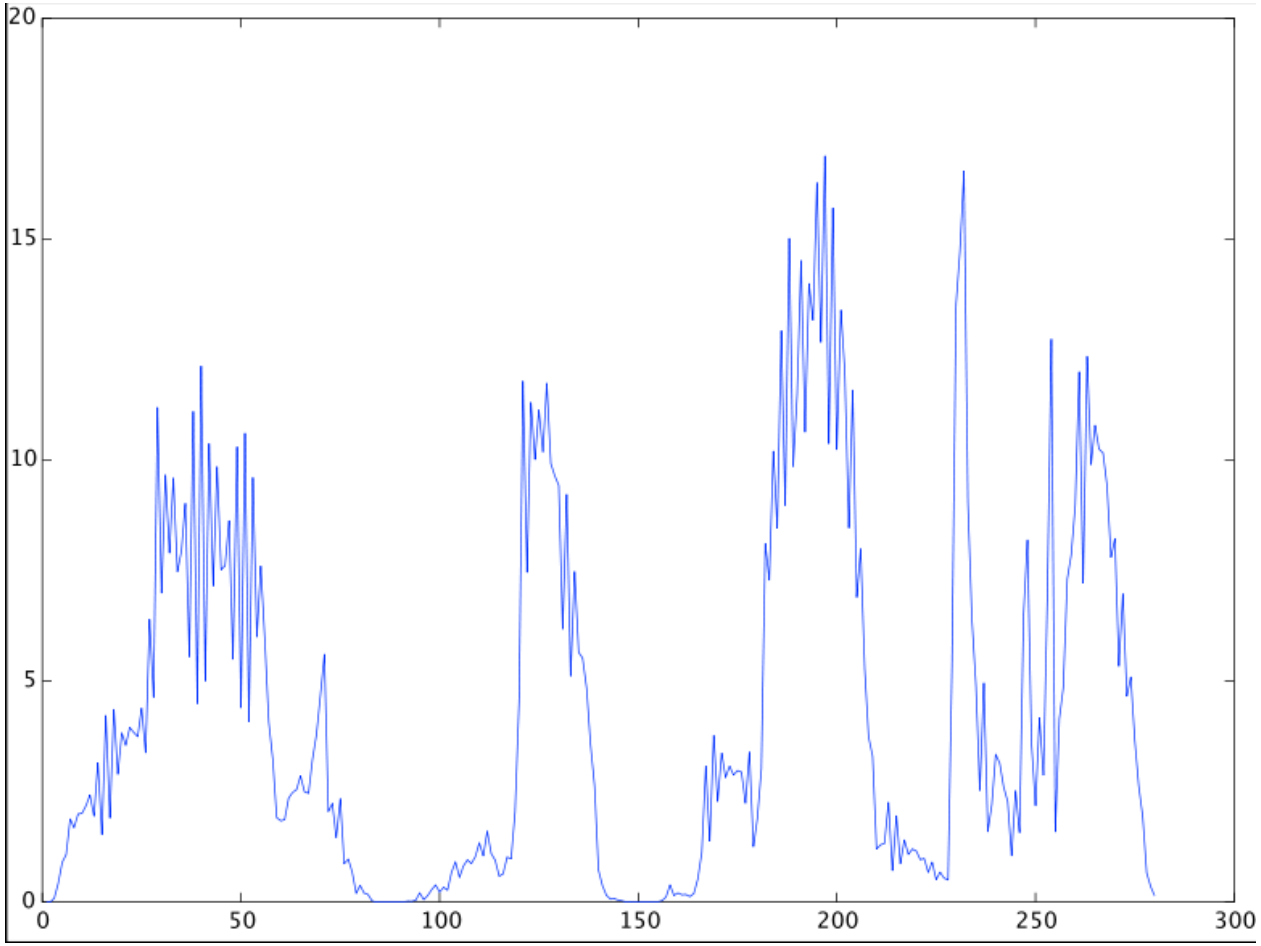


Figure: Ukázka průběhu funkce krátkodobé intenzity

Analýza digitalizovaného signálu v časové oblasti 2.

- Krátkodobá funkce středního počtu průchodu nulou:
  - součet všech průchodů signálu nulou

$$Z(n) = \sum_{k=-\infty}^{\infty} |sgn[s(k)] - sng[s(k-1)]| w(n-k)$$

  - varianta - počet lokálních extrémů
  - obě mohou být negativně ovlivněny šumem zvukového pozadí
- Diferenční klasifikátory:
  - diference prvního řádu

$$D_n = \sum_{k=-\infty}^{\infty} |s(k) - s(k-1)| w(n-k)$$

Ukázka průběhu funkce středního počtu průchodů nulou

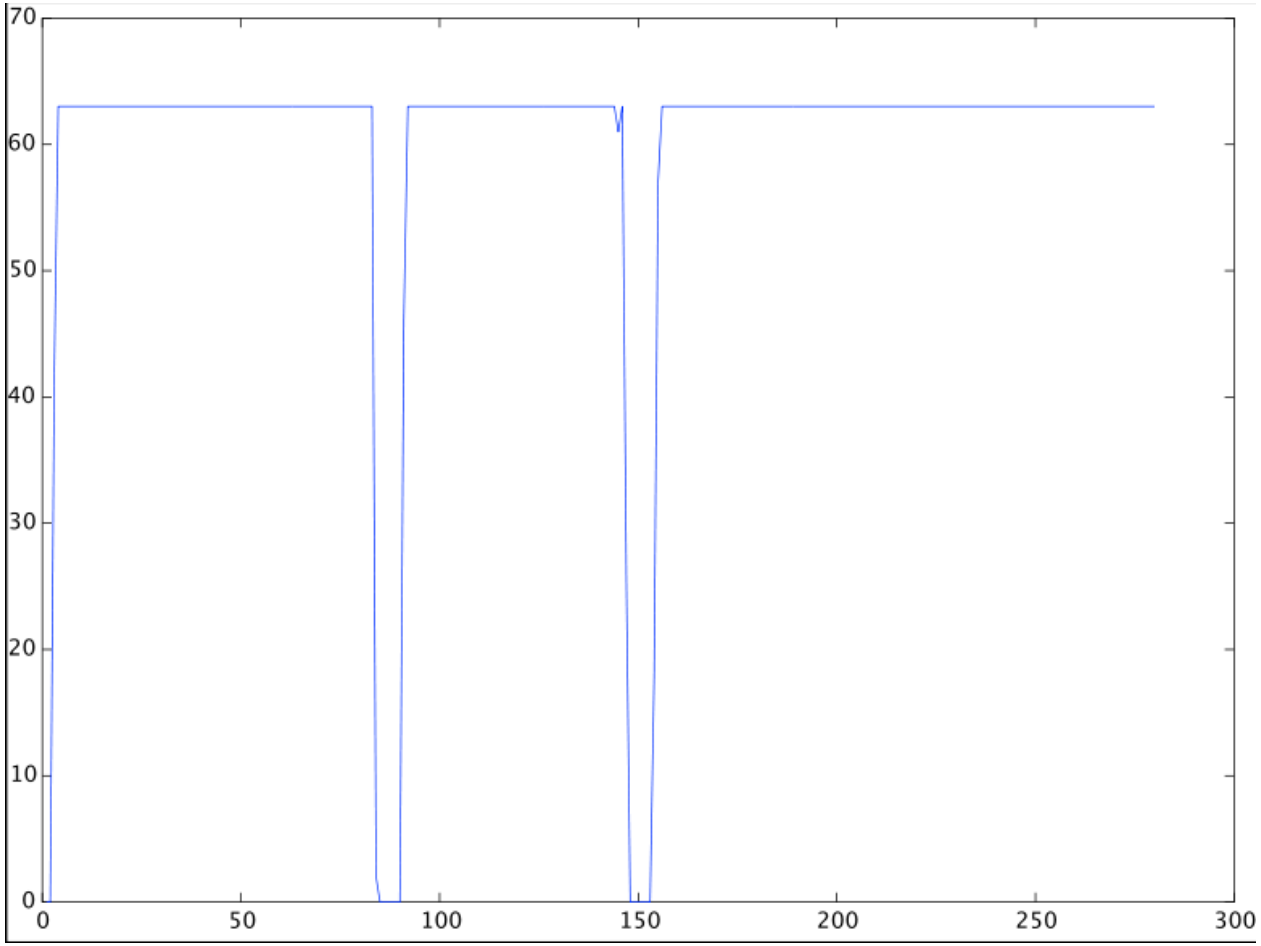


Figure: Ukázka průběhu funkce středního počtu průchodů nulou

Analýza digitalizovaného signálu v časové oblasti 3.

- Krátkodobá autokorelační funkce:

$$R(m,n) = \sum_{k=-\infty}^{\infty} (s(k)w(n-k))(s(k+m)w(n-k+m))$$

  - používá se při zjišťování periodicity signálu základního tónu řeči
  - je-li signál periodický s periodou P, R(m,n) nabývá maxima pro m=0, P, 2P, ...
  - předpokládá délku mikrosegmentu aspoň 2P

Zpracování signálu ve frekvenční oblasti

- Nejvíce používané:
  - krátkodobá Fourierova transformace
  - kepstrální analýza
  - lineární prediktivní analýza

Fourierova řady

- f(x) - periodická spojitá funkce s periodou T

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k cos(kx) + b_k sin(kx))$$

- Způsob výpočtu koeficientů a_i a b_i:
  - α, $\alpha + T$ - interval periodicity funkce f

$$a_k = \frac{2}{T} \int_{\alpha}^{\alpha+T} x cos(kx) dx$$

$$b_k = \frac{2}{T} \int_{\alpha}^{\alpha+T} f(x) sin(k\omega x) dx$$

- Nelze přímo použít - digitalizovaný zvuk není spojitý a je periodický pouze na omezených úsecích.

Diskrétní Fourierova Transformace (DFT)

- Používá se pro vyjádření spektrálních vlastností periodických posloupností s periodou N vzorků případně konečných posloupností délky N vzorků.
- Výpočet koeficientů X(k) DFT:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} x(n) W_N^{-kn}$$

  - —X(k)— - intenzita k. spektrálního koeficientu; frekvence závisí na velikosti mikrosegmentu N a vzorkovací frekvence T
  - x(n) - n. vzorek daného mikrosegmentu.
  - $W_n = e_j * 2\pi/N = cos(2\pi/N) + j * sin(2pi/N)$

Inverzní Diskrétní Fourierova Transformace (IDFT)

- Výpočet n. vzorku na základě hodnot X(k) - Inverzní diskrétní Fourierova transformace (IDFT):

$$x(n) = \frac{1}{N} \sum_{i=0}^{N-1} X(k) e^{j \frac{2\pi}{N} kn} = \frac{1}{N} \sum_{i=0}^{N-1} X(k) W_N^{kn}$$

Rychlá Fourierova transformace (FFT)

- Časová složitost výpočtu spektrálních koeficientů pomocí DFT - n^2 operací na komplexními čísly.
- Pomocí FFT - $N * log_2 \frac{N}{2}$ operací násobení.
- FFT požaduje, aby délka analyzovaného segmentu byla mocninou 2.

Kepstrální analýza

- Vychází z modelu činnosti hlasového ústrojí:
  - Řečové kmity lze modelovat jako odezvu lineárního systému na buzení sestávající ze sledu pulzů pro znělou hlásku a šumu pro neznělou.
- Kepstrum - $X(k) = IFFT(FFT(x(k)))$
- Kepstrální analýza umožňuje z řeči oddělit parametry buzení a parametry hlasového ústrojí.
- Využití:
  - ocenění fonetické struktury řeči
    - znělost
    - F0, F1, F2, ...
  - rozpoznávání slov
  - verifikace a identifikace mluvčího
  - ...

Lineární predikce

- Jedna z nejefektivnějších metod analýzy akustického signálu.
  - Zajišťuje velmi přesné odhady parametrů při relativně malé zátěži.
- Vychází z předpokladu, že s(k) lze popsat jako lineární kombinaci N předchozích vzorků a buzení u(k) s koeficientem zesílení G:

$$s(k) = -\sum_{i=1}^{N} a_i s(k-i) + Gu(k)$$

- Použití:
  - určování spektrálních charakteristik modelu hlasového ústrojí
  - z chyby predikce lze odvodit poznatky o znělosti a určit frekvenci základního tónu
  - koeficienty ai nesou informaci o spektrálních vlastnostech
    - lze je použít jako příznaky pro rozpoznávání řeči.

Software pro analýzu signálu

- HTK (`http://htk.eng.cam.ac.uk/`) - Hidden Markov Model Toolkit (Engineering Department of Cambridge University) - toolkit pro tvorbu rozpoznávačů řeči založených na skrytých Markovových modelech.
- ESPS toolkit (`http://www.speech.kth.se/software/#esps`)
- NICO toolkit (`http://nico.nikkostrom.com/`) - toolkit pro vytváření umělých neuronových sítí, využívá se např. pro rozpoznávání řeči.
- Matlab - knihovny pro analýzu řeči
  - labrosa.ee.columbia.edu/matlab/ (`http://labrosa.ee.columbia.edu/matlab/`)
  - Audio processing in Matlab (`http://www.umiacs.umd.edu/~ramani/cmsc828d_audio/Audio%20processing%20using%20Matlab.ppt`)
  - ...

Software pro analýzu signálu

- Octave (`http://www.gnu.org/software/octave/index.html`) - opensource alternativa Matlabu
  - Měly by jít použít tytéž knihovny.
- SMP Tool (`http://www.speech.kth.se/smptool/`)
- ...

Cíle rozpoznávání řeči

- Cíle rozpoznávání řeči:
  - interpretace příkazů uživatele – hlasové ovládání různých zařízení.
    - telefon
    - navigace
    - ...
  - převod mluveného slova na text – přepis mluveného slova
    - záznamy soudních přelíčení
    - ...
- Druhy rozpoznávání řeči:
  - rozpoznávání izolovaných slov
  - rozpoznávání plynulé promluvy.

Úvod

- Cíl – rozpoznání částí promluvy ohraničených z obou stran pauzou.
- Uživatel může zadávat pouze jednotlivé povely nebo musí po vyřčení slova udělat pauzu.
- Odpadá problém se stanovením rozhraní dvou slov/povelů.
  - Povel může být víceslovný, ale pro tyto účely představuje jedno slovo.
- Obvykle jde o systémy závislé na uživateli
  - nutnost tréninku.
- Mívají omezenou kapacitu slovníku
  - slovník – seznam rozpoznávaných slov.
- Používají obvykle vektor příznaků.
  - Vektor hodnot získaných analýzou signálu (spektrum, kepstrum, energie, intenzity, ...).
  - Získán některou z metod krátkodobé analýzy.

Vektory příznaků a jejich porovnávání

- Vektor příznaků
- Vektorový prostor nad tělesem F je množina V společně s dvěma operacemi sčítání vektorů a násobení skalárem, které splňují následující axiomy:
  - (V, +) je komutativní grupa
  - Násobení skalárem ($FxV - V$) je asociativní a(b$\mathbf{v}$) = ab($\mathbf{v}$)
  - 1$\mathbf{v}$=v, kde 1 je jednotkový prvek tělesa
  - a dále platí distributivní zákon:
    - a($\mathbf{v}$+$\mathbf{w}$) = a$\mathbf{v}$ + a$\mathbf{w}$
    - (a+b)$\mathbf{v}$ = a$\mathbf{v}$ +b$\mathbf{v}$
- Metrický prostor: Množina M se zobrazením d (metrikou), pro které platí:
  - $d(x, y) \geq 0$
  - $d(x, y) = 0 \Leftrightarrow x = y$
  - $d(x, y) = d(y, x)$
  - $d(x, z) \leq d(x, y) + d(y, z)$
- Příklad metriky je např. Euklidovská vzdálenost.

Klasifikátory

- Klasifikátory využívající porovnání slov metodou DTW (Dynamic Time Warping)
  - umožňují porovnání podobnosti dvou dynamických jevů, které probíhají různými rychlostmi
- Klasifikátory založené na statistických metodách
  - modelování pomocí skrytých Markovových modelů
- Hierarchické klasifikátory
  - Pracují hierarchicky:
    1. Akustická analýza sígnálu.
    2. Rozdělení signálu promluvy na segmenty.
    3. Fonetické dekódování jednotlivých segmentů.
    4. Rozpoznání slova (povelu) probíhá ve druhé vyšší úrovni na základě posloupnosti klasifikovaných segmentů.
  - Podobný princip se využívá pro rozpoznávání plynulé řeči.

Metoda DTW (Borcení časové osy)

- Používá se pro porovnání dvou úseků promluv (slov).
  - Úseky jsou vyjádřeny posloupností vektorů příznaků
    - úsek promluvy rozdělen do mikrosegmentů
    - klasifikovány souborem krátkodobých charakteristik
- Postup:
  1. Pro rozpoznávané posloupnosti vytvoříme soubor referenčních posloupností akustických vektorů.
  2. Vytvoříme posloupnost akustických vektorů pro rozpoznávané slovo.
  3. Metodou DTW porovnáme rozpoznávanou posloupnost s referenčními a vybereme tu, s největší shodou.

- Algoritmus hledá parametrizaci f,g takovou, že i=f(k),j=g(k), k=1, ..., K, minimalizuje výraz:

$$D(A,B) = \sum_{k=1}^{K} d(a(f(k)), b(g(k)))$$

  - d je vzdálenost mezi akustickými vektory (např. Euklidovská metrika)
  - Euklidovská metrika

$$d(a,b) = \sqrt{(\sum_{i=1}^{n}(a_i - b_i)^2)}$$

Metoda DTW (2.)

- Omezující podmínky:
  - f,g – neklesající funkce
  - omezení na lokální souvislost a strmost:
    - $0 \leq f(k) - f(k-1) \leq I^*$
    - $0 \leq g(k) - g(k-1) \leq J^*$
    - většinou platí I^*, $J^* = 1, 2, 3$
    - Z praktických testů vyplynulo, že při příliš strmém přírůstku může dojít např. k nevhodné korespondenci mezi příliš krátkým segmentem vzorku A a příliš dlouhým segmentem vzorku B.
  - Omezení na hraniční body:
    - $f(1) = 1, f(K) = I$
    - $g(1) = 1, g(K) = J$

Metoda DTW (3.)

- Omezující podmínky
  - Globální vymezení oblasti pohybu funkce DTW:
    - Omezení minimální a maximální přípustné směrnice přímky omezující přípustnou oblast, při splnění podmínky na hraniční body
      $$1 + \alpha[i(k) - 1] \leq j(k) \leq 1 + \beta[i(k) - 1]$$
    - α – minimální směrnice přímky omezující přípustnou oblast
    - β – maximální směrnice přímky omezující přípustnou oblast
- ...

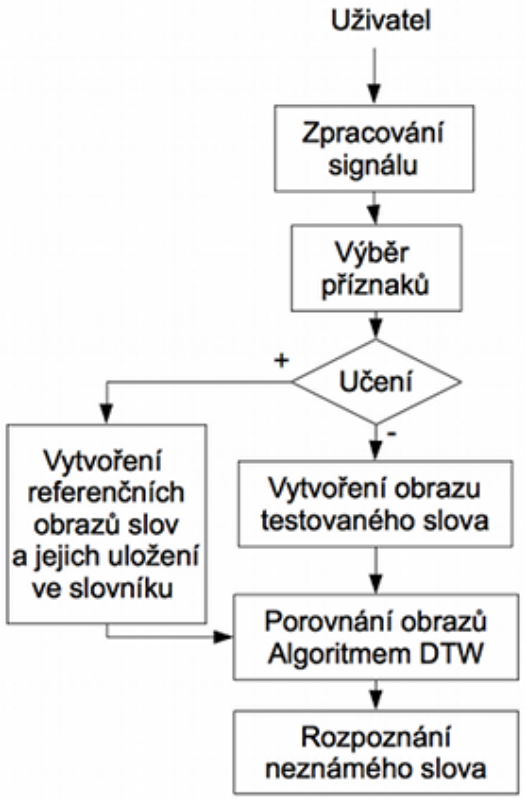


Figure: Blokové schéma algoritmu DTW

- Obecný postup:
  1. Řečník resp. skupina řečníků vysloví postupně každé trénované slovo požadovaného slovníku. Buď jednou nebo opakovaně
  2. Vstupní slova jsou zdigitalizována, nejčastěji do formátu PCM.
  3. Dále jsou převedena zvolenou metodou krátkodobé analýzy na posloupnost vektorů příznaků.
  4. Detekce hranic slov
     - může být náročné na provedení např. kvůli rušivému pozadí.
     - Nekorektní detekce hranic slov zhoršuje úspěšnost rozpoznávání
     - Metody odstraňující i jen částečně vliv pozadí zvyšují výpočetní náročnost.
  5. Vytvoření referenčních obrazů slov.

DTW – Metody vytváření referenčních obrazů slov

- Přímé využití obrazů trénovací množiny jako referenčních obrazů slov
  - namluvená slova od jednoho nebo více řečníků jsou použita jako referenční vzory
  - DTW nevyžaduje, aby obrazy téhož slova byly stejně dlouhé, ale z důvodu možnosti aplikace pomocných kritérií je vhodné provést časovou normalizaci každého obrazu.
- Vytváření průměrného vzorového obrazu pro každou třídu slov *w:*
  - používají se metody lineárního a nebo dynamického průměrování
  - lineární průměrování:
    - provedeme lineární časovou normalizaci všech akustických obrazů trénovací množiny
    - výsledné referenční složky obrazu určíme jako průměr odpovídajících složek obrazů pro dané slovo
  - dynamické průměrování:
    - vzorový obraz se vytváří použitím algoritmu DTW

DTW – Metody vytváření referenčních obrazů slov
pokračování

- Vytváření vzorových obrazů shlukováním
  - rozdělíme vzorové obrazy pro dané slovo do shluků tak, že obrazy uvnitř shluku jsou si "podobné" a obrazy z různých shluků jsou "nepodobné"
  - shlukování lze realizovat:
    - interaktivně (poloautomaticky) – metoda řetězové mapy, algoritmus ISODATA (viz Levinson, Rabiner , Sondhi – Interactive Clustering Techniques for Selecting Speaker-Independed Reference Templates for Isolated Word Recognition, IEEE Transactions on ASSP, 27, 1979, č 2)
    - automaticky – algoritmy založené na MacQueenově algoritmu (viz např. Komunikace s počítačem mluvenou řečí)

- Během klasifikace probíhá zpracování řečového signálu stejně jako při učení:
  - pokud jsou referenční obrazy normalizovány je nutné normalizovat i rozpoznávaná slova.
- Pravidla využívaná při klasifikaci:
  - minimální vzdálenost
  - varianty pravidla nejbližšího souseda

Redukce výpočetních a paměťových nároků při použití DTW

- Nevýhody DTW:
  - vysoké paměťové a výpočetní nároky
  - mohou znesnadňovat klasifikaci v reálném čase i při relativně malém slovníku
- Metody řešení:
  - hrubá síla – využití drahých paralelních procesorů případně zákaznických obvodů
  - vhodné zakódování parametrů jednotlivých mikrosegmentů referenčních i testovacích obrazů
  - redukce počtu mikrosegmentů akustického obrazu slova – využívají se oblasti spektrální stacionarity řečového signálu
  - snížení výpočetní náročnosti při hledání nejbližšího souseda ve slovníku
    - vhodná volba prohledávacích postupů

Redukce výpočetních a paměťových nároků při použití DTW 2.

- redukce oblasti prohledávání funkce DTW
  - pomocí heuristik do operací porovnávání obrazů
- vhodné zakódování parametrů mikrosegmentů
  - využívá vektorovou kvantizaci a kódovou knihu
  - kódová kniha
    - abeceda konečného počtu kvantizovaných vzorků:
    - každý vektor ve vzorku lze nahradit jeho pořadovým číslem
    - při předem definované kódové knize lze dopředu spočítat matici vzájemných vzdáleností mezi kvantizačními vzory
- využití oblastí spektrální stacionarity řečového signálu
  - využívá se přítomnost oblastí spektrální stacionarity
  - metoda spektrální stopy:
    - spektrální stopa – spojnice koncových bodů vektorů příznaků
    - aproximace – např. lineárními úseky.

Redukce výpočetních a paměťových nároků při použití DTW 3.

- Zavedení účinných způsobů vyhledávání nejbližšího souseda.
  - Viz metody prohledávání metrických prostorů.
    - Nutno ověřit, že vzdálenost použitá v algoritmu DTW je metrika.
- Redukce výpočetních nároků pomocí heuristik při porovnávání:
  - vícestupňový rozhodovací postup
    1. Porovnáváme promluvu proti celému slovníku pomocí pouze několika příznaků.
    2. Výstupem je soubor perspektivních kandidátů (řádově jednotky desítek), ve kterém se vyhledává pomocí klasického DTW.
  - práh zamítnutí
    - Po každém kroku porovnáváme spočítanou vzdálenost.
    - Překročíme-li experimentálně získanou hodnotu prahu obraz je zamítnut.

Metoda HMM (Hidden Markov Model)
Úvod

- Modelování řeči pomocí HMM vychází z následující představy o tvorbě řeči:
  - Hlasové ústrojí se v krátkém čase nachází v jedné z konečně mnoha artikulačních konfigurací – generuje hlasový signál.
  - Přejde do následující konfigurace.
- Tuto činnost lze chápat statisticky.
- Kvantizací akustických vektorů (vytvořením kódové knihy) lze dosáhnout konečnosti všech parametrů odpovídajícího modelu.

Princip použití HMM pro rozpoznávání

- Jsou generovány dvě vzájemně svázané časové posloupnosti náhodných proměnných:
  - Podpůrný Markovův řetězec – posloupnost konečného počtu stavů.
  - Řetězec konečného počtu spektrálních vzorů.
- Náhodné funkce ohodnocující pravděpodobnostmi vztah vzorů k jednotlivým stavům.
- Pro rozpoznávání řeči nejčastější využívané levo-pravé Markovovy modely.
  - Vhodné pro modelovaní procesů spjatých se vzrůstajícím časem.

Markovův proces

- Markovův proces G se skrytým Markovovým modelem je pětice $G = (Q, V, N, M, \pi)$
  - $Q = \{q_1, ...q_k\}$ – množina stavů
  - $V = \{v_1, ..., v_m\}$ – množina výstupních symbolů
  - $N = (n_{i,j})$ – matice přechodu
    - určuje pravděpodobnost přechodu ze stavu q_i v čase t do stavu q_j v čase t_1
  - $M = (m_{i,j})$ – matice přechodu, určující pravděpodobnost generování akustického vektoru v_j, v kterémkoliv čase ve stavu q_i
  - π – (π_i) – vektor pravděpodobností počátečního stavu (pravděpodobnost toho, že i. stav je počáteční)
- Trojice $\lambda = (N, M, \pi)$ – soubor parametrů HMM; vytváří model řečového segmentu (slova, ...)
  - např. Vintsjukův model pro slovo:
    - počet stavů 40 — 50 – odvozeno od průměrného počtu mikrosegmentů ve slově (délka mikrosegmentu 10 ms)

Určení pravděpodobnosti promluvy

- Značíme $P(O|\lambda)$.
- Promluva O standardně zpracována do posloupnosti $O = (o_1, \ldots O_T)$.
  - T – počet mikrosegmentů promluvy
  - o_i – odpovídají výstupním symbolům
- Určení $P(O|\lambda)$ – metoda využívající rekurzivní výpočet odpředu nebo odzadu generované posloupnosti (forward-backward algorithm).

Určení pravděpodobnosti
Pokračování

- Výpočet odpředu:
  - α_i – pravděpodobnost přechodu do stavu q_i při generování posloupnosti $\{o_1, ..., o_t\}$ ($\alpha_i = P(o_1, o_2..o_t, q_i(t)|\lambda)$
  - Rekurzivní výpočet:
    1. inicializace
       $$\alpha_1(i) = \pi_i m_i(o_1)$$
       pro $1 \leq i \leq N$
    2. rekurze pro t=1,2, ... T-1
       $$a_{t+1}(j) = [\sum_{i=1}^{N} a_t(i) n_{i,j}] m_j(o_{t+1})$$
       pro $1 \leq j \leq N$, $m(o_t)$ je ekvivalentní zápisu $m_i(l)$, pokud $o_t = v_l$
    3. Výsledná pravděpodobnost:
       $$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

Alternativní způsob výpočtu $P(O|\lambda)$

- Nevýhoda předchozího postupu:
  - ve výsledném vztahu jsou zahrnuty pravděpodobnosti všech možných posloupností stavů délky T
- Lze nahradit výpočtem maximálně pravděpodobné posloupnosti Q.
- Výpočet realizován pomocí Viterbiova algoritmu:
  - problém řešen rekurzivně s použitím techniky dynamického programování

Trénování parametrů modelu $\lambda = (N, M, \pi)$

- Nutno stanovit postup při trénování parametrů modelu
  - maximalizace pravděpodobnosti $P(O|\lambda)$
  - neexistuje analytická metoda k zajištění globálního maxima
  - používají se iterativní algoritmy zajišťující aspoň lokální maximalitu
- Nejpoužívanější postup – Bauman-Welchův algoritmus
- Problém při trénování modelu:
  - vliv konečné trénovací množiny – čím menší je trénovací množina a čím vetší matice M, tím vetší pravděpodobnost, že některé prvky matice budou nastaveny na 0 – problém chybějících (neadekvátních) dat

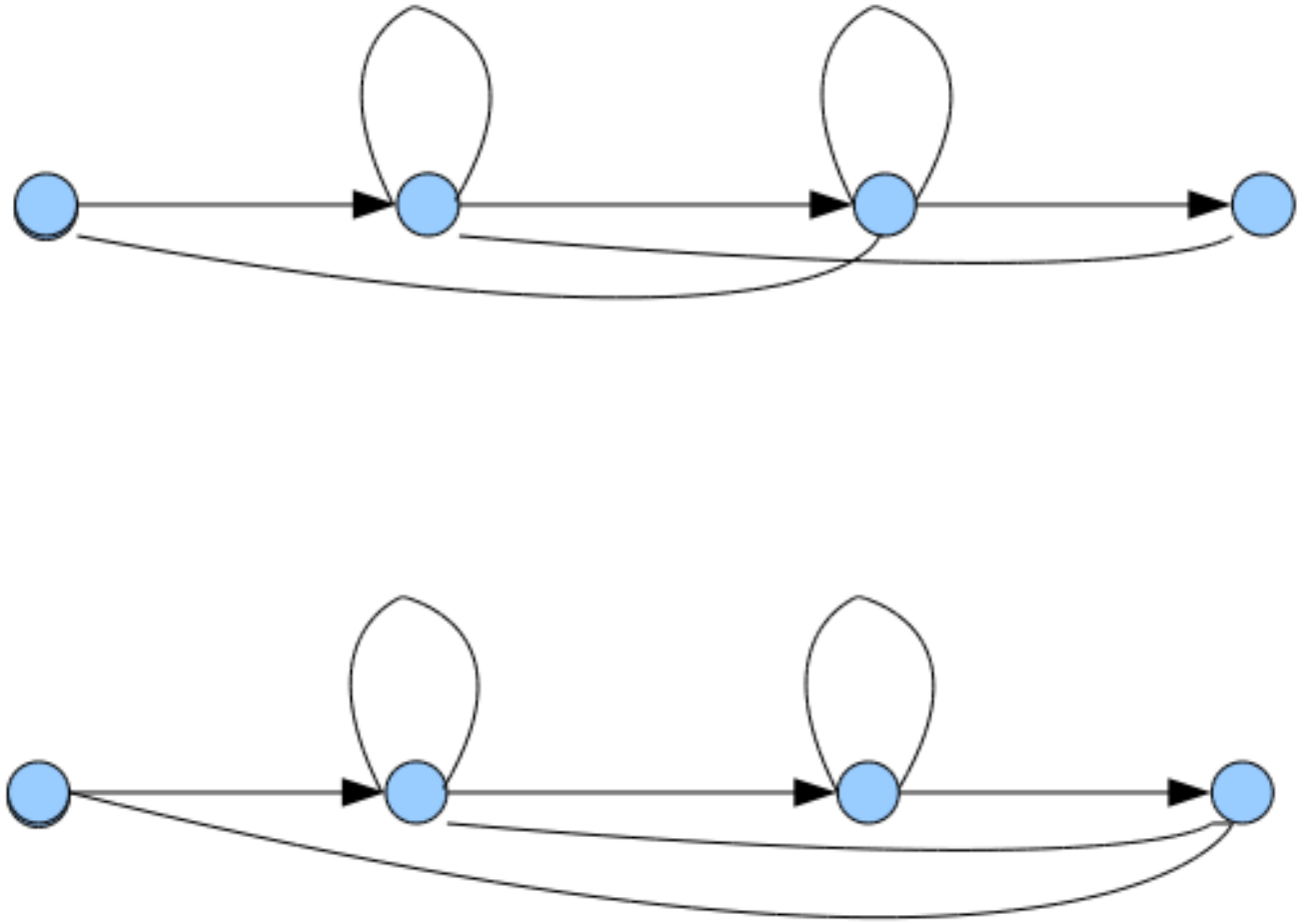
Rozhodovací pravidlo – rozpoznávání slova

- Princip maximální věrohodnosti:
  - Pro neznámé slovo O určíme hodnoty $P(O|\lambda)$ pro všechna modely λ.
  - Jako výsledek vybereme třídu s maximální hodnotou.

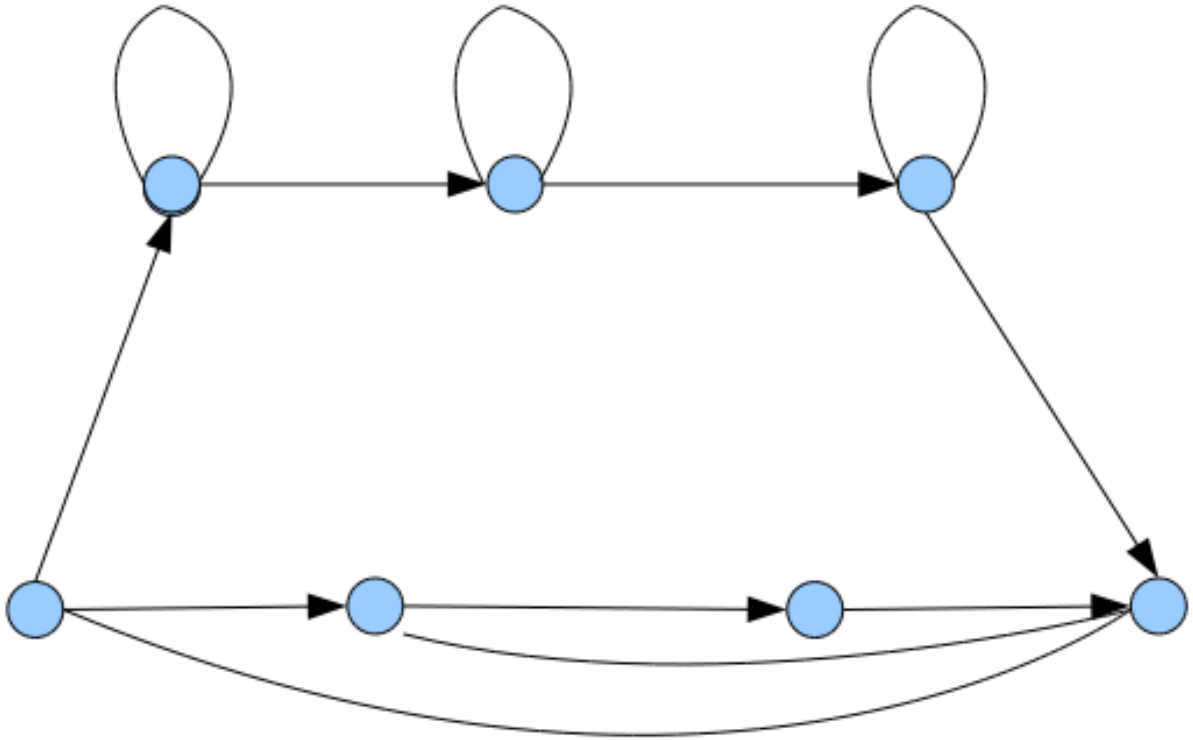
Implementace

- Modelování povelů:
  - nejčastější se používají modely se 4-7 stavy
  - lze využít SW nástroje pro tvorbu HMM:
    - HTK – Hidden Markov Model ToolKit (http://htk.eng.cam.ac.uk/)
- Modelování fonémů:
  - obvykle 4-7 stavů
  - model slova – zřetězení modelů fonémů
  - problémy s výpočtem v reálném čase
  - speciální algoritmy na vyhledávání

Příklady struktur HMM pro fonémy



Příklady struktur HMM pro fonémy



Obtíže při rozpoznávání izolovaných slov

- Určení začátku a konce promluvy:
  - šum kontra sykavky
  - detekce nahodilého zvukového vzruchu (klepnutí, ...) kontra okluzivy, které obsahují pauzy
  - možná přítomnost infrazvuků.

Ukázka plozívy

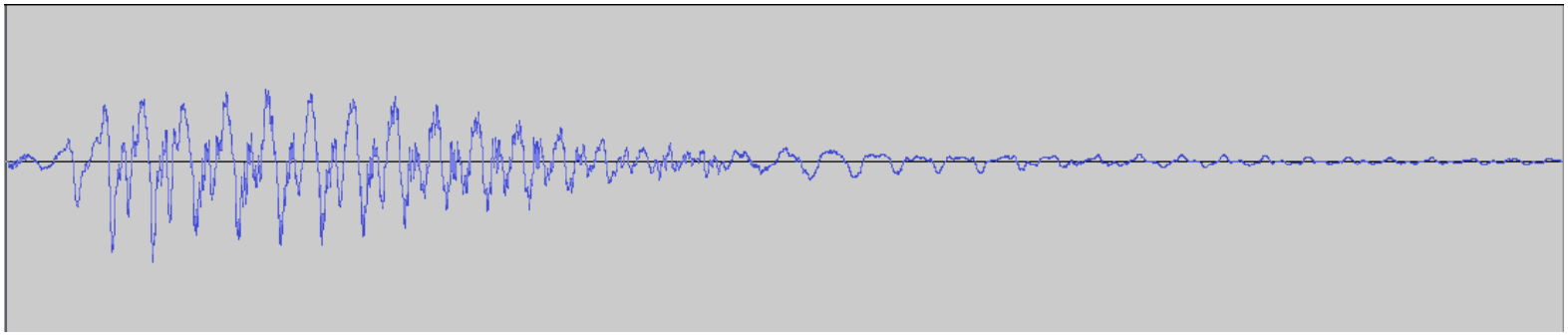


Figure: Hláska P

Úvod

- Hlavní rozdíly oproti rozpoznávání slov:
  - nelze vytvořit analogii databáze vzorů
  - prozodické faktory
  - nutnost určovat hranice mezi slovy
  - výplňkové zvuky a chyby řeči
- Řešení - statistický přístup
  - použití jazykových modelů
  - HMM vrátí stejnou pravděpodobnost např. pro slova "máma" a "nána"
    1. máma je častější - vhodné použít máma

Jazykové modely

- Posloupnost slov (promluva) W = (w(1)w(2)...w(n)).
- Posloupnost akustických vektorů - O = O(o(1)o(2)...o(t)).
- Chceme nalézt W* (množinu všech promluv) maximalizující P(W—O).
- Dle Bayesova pravidla platí:P(W*—O) = max P(W—O) = max P(W)*P(O—W)/P(O)
- Pro nalezení maxima potřebujeme znát:
  - model řečníka P(O—W)
  - jazykový model P(W)
- Model řečníka se nahrazuje pravděpodobností generování W odpovídajícím Markovovým modelem.
- Trigramový model:
  - Platí: $P(w(n)|w(1)..w(n-1)) \cong P(w(n)|w(n-2)w(n-1))$

Rozpoznávání tématu - topic recognition

- Úspěšnost rozpoznávání plynulé řeči 50-99
  - úkolu
  - jazyku
  - mluvčím
  - ...
- Úspěšnost rozpoznávání může zvýšit:
  - znalost tématu promluvy
  - použití gramatiky pro rozpoznávání řeči.
- Mění se stavový prostor a pravděpodobnosti trigramů
  - např. mějme burzovní zprávy - bylo rozpoznáno slovo honey nebo money?
- Známé téma - může být přesnější jazykový model.

Gramatiky pro podporu rozpoznávání řeči

- Umožňují omezit množinu rozpoznávaných promluv:
  - výhoda - vyšší úspěšnost rozpoznávání
  - nevýhoda - nižší volnost vyjadřování
- Používají se bezkontextové gramatiky.
- V praxi často používané formáty gramatik:
  - JSGF (`http://www.w3.org/TR/jsgf/`) - původně definována v Java Speech API (`http://java.sun.com/products/java-media/speech/`)
  - SRGS (`http://www.w3.org/TR/speech-grammar/`) - součást standardů W3C Voice Browser Activity (`http://www.w3.org/Voice`)
    - Určeny pro tvorbu dialogových a hlasových rozhraní.

Ukázka gramatiky ve formátu JSGF

```
#JSGF
 <koren> = Chci jet <cim>.|
           Chci jet <cim> z <odkud> do <kam>.|
           Chci jet <cim> z <odkud> do <kam> v <kdy>.;
 <cim> = vlakem | autobusem;
<odkud> = <czMesto>;
<kam> = <czMesto>;
<kdy> = <czCas>;
```

Ukázka odpovídající gramatiky v XML formátu SRGS

```
<grammar root="koren" version="1.0" xml:lang="cs-CZ">
 <rule id="koren">
  <one-of>
   <item>Chci jet <ruleref uri="\#cim"/>.</item>
   <item>Chci jet <ruleref uri="\#cim"/>
    z <ruleref uri="url db názvů stanic"/>
    do <ruleref uri="url db názvů stanic"/>
   </item>
   ...
  </one-of>
 </rule>
```

Ukázka odpovídající gramatiky v XML formátu SRGS

Pokračování

```
 <rule id="cim">
  <one-of>
   <item tag="vlak">vlakem</item>
   <item tag="autobus">autobusem</item>
    ...
  </one-of>
 </rule>
</grammar>
```

Ukázka gramatiky v ABNF formátu SRGS

```
root=$koren;
language = cs-CZ;
...
$koren = Chci jet $cim. |
         Chci jet $cim z $<url db stanic>
          do $<url db stanic>|
         ...
$cim = autobusem {$out=autobus} | vlakem {$out=vlak}
```

Úvod

- Úkol:
  - Převod psaného textu na mluvenou řeč.
  - Co nejpřirozenější řeč - ideálně k nerozeznání od člověka:
    - správná intonace
    - správné umístění přízvuků
    - správná koartikulace
    - správný rytmus
    - ...

Druhy syntézy řeči

- Druhy syntézy řeči
  - ve frekvenční oblasti
  - v časové oblasti
  - korpusová
  - problémově orientovaná syntéza:
    - hlášení nádražního rozhlasu
    - automatizované linky telefonické podpory

Fáze syntézy řeči

1. Fonetický přepis.
2. Syntéza fonetické transkripce
3. Případný postprocessing:
   - intonace
   - správné časování - modifikace délky fonémů, ...
   - větné přízvuky
   - ...

Fonetický přepis

- Slouží k přesnému, jednoznačnému zápisu mluvené řeči.
- Využívá fonetickou abecedu:
  - mezinárodní fonetická abeceda - IPA (součást standardu UNICODE)
  - 7bitový přepis IPA pomocí ASCII - SAMPAma:S se dobr'e / ma:S se dobRe
- Nelze si pamatovat fonetický přepis každé promluvy - nutno zabezpečit automatický přepis:
  - fonologická pravidla
- Při transkripci češtiny se některé české znaky nevyužívají:
  - ch - x
  - w - v
  - y/ý - i/í
  - q - kv
- Koartikulace

Pravidla fonetického přepisu češtiny

- ch → x
- ů → ú
- w → v
- q → kv
- y → i
- ý → í
- ě → je /po b,p,f,v
- dě, tě, ně, mě
  - dě → ďe
  - tě → ťe
  - ně → ňe
  - mě → mňě

Pravidla fonetického přepisu češtiny

- di, ti, ni
  - di $\rightarrow$ ďi
  - ti $\rightarrow$ ťi
  - ni $\rightarrow$ ňi
- X:
  - x $\rightarrow$ ks — začátek slova před samohláskou, mezi samohláskami nebo před neznělou souhláskou a nebo na konci slova, s výjimkou ex¡samohláska¿ $\rightarrow$ egz
  - x $\rightarrow$ gz — před znělou souhláskou

Změny na při spojování souhlásek

- Dochází k nim při spojování souhlásek.
- Způsobeny přenastavováním mluvidel.
- 2 druhy:
  - spodoba znělosti - změna znělosti párových souhlásek
    - ZPS → ZPS
    - NPS → NPS
    - dub → dup
    - zpěv → spjef
    - sběr → zbjer
    - když → gdiš
  - spodoba artikulační - při spojení dvou souhlásek s různou artikulací
    - banka, tango
    - tramvaj, nymfa
    - punťa, pindík
    - odpovědně, sto dní, vodní
    - ts → c, tš → č
    - ds → c, dš → č

On-line přístupné ukázky syntézy řeči

- AT&T Labs Natural Voices© Text-To-Speech (`http://www2.research.att.com/~ttsweb/tts/demo.php`)
- Free demo to create avatars using TTS by SitePal (`http://www.oddcast.com/home/demos/tts/tts_example.php`)
- Cepstral Text-to-Speech (`http://cepstral.com/demos/`)
- Festival Online Demo (`http://www.cstr.ed.ac.uk/projects/festival/onlinedemo.html`)
- Speechtech s.r.o. (`http://www.speechtech.cz/cs/produkty/demo.html#Iva210`)

Syntéza ve frekvenční oblasti

- Emulace funkce hlasového ústrojí pomocí FM syntezátoru.
- Nutno uchovávat:
  - frekvenční charakteristika použitého hlasu
  - parametry buzení.
- Využívá:
  - systém frekvenčních generátorů - simulují hlasivky
  - filtry a zesilovače - simulace rezonance v dutinách
  - Tyto komponenty ovládány parametry modelu.
- Nejběžněji použité způsoby kódování zdroje:
  - Řečová syntéza formantového typu - uchovávají se parametry průběhu jednotlivých formantů a buzení.
  - LPC řečová syntéza - uchovávají se F0, příznak znělosti, amplituda budícího signálu G a koeficienty LPC,

Syntéza ve frekvenční oblasti

- Výhody
  - menší paměťové nároky - uchovávají se pouze parametry modelu.
- Nevýhody:
  - oproti syntéze v časové oblasti může být výsledek méně přirozený - "robotické" hlasy
  - Softwarová - výpočetně relativně náročné - lze implementovat přímo na úrovni HW
    - skládání jednotlivých frekvencí, které tvoří příslušné fonémy
    - řešení koartikulace
    - ...
  - Neexistuje dostatečně přesný matematický model

Využití syntézy ve frekvenční oblasti

- Využití dříve:
  - malé paměťové nároky
  - domácí počítače (Amiga, Atari, ...)
  - syntéza realizována většinou hardwarově
- Dnes:
  - Syntéza na zařízeních s nedostatkem paměti.
  - Syntéza realizovaná hardwarově pomocí zákaznických obvodů.
- Doplnění syntézy v časové oblasti o prozodické jevy:
  - Větná intonace
  - ...
  - Realizováno programově pomocí modifikace F_0 a formantů.

Schéma syntetizéru formantového typu

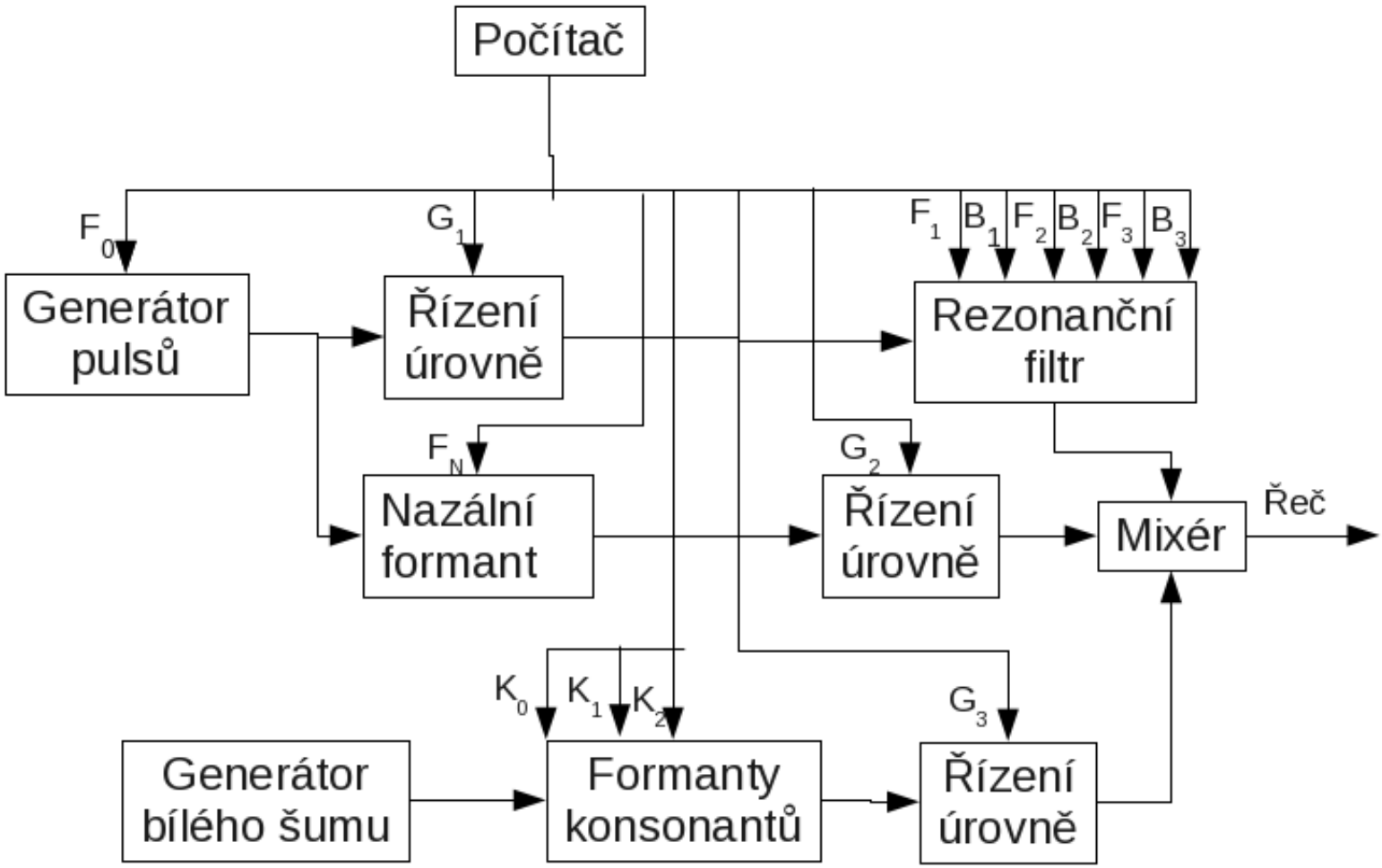
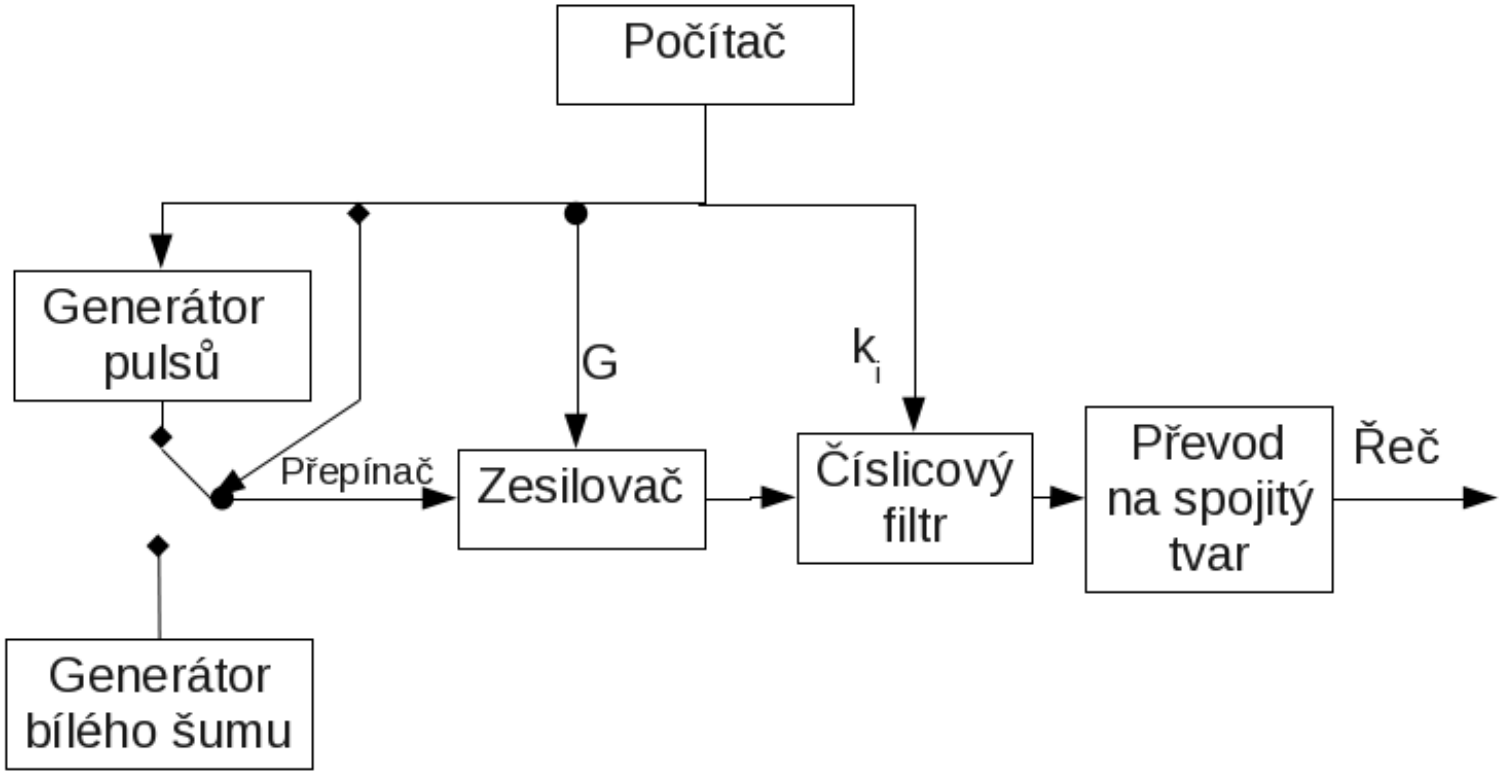


Schéma LPC syntetizéru



Syntéza v časové oblasti

- Princip
  - spojování navzorkovaných řečových segmentů uložených v databázi.
- Využívají se různé typy základních segmentů:
  - větší
    - lépe se modelují některé další charakteristiky jako intonace, přízvuky, ...
    - větší nároky na paměť - větší množství segmentů (potenciálně až 2n, kde n je délka segmentu)
    - příklady – slova, části vět, ...
  - menší
    - menší paměťové nároky - menší množství segmentů
    - horší možnost modelování větné intonace, přízvuků, ... (viz oblasti spektrální stacionarity řeči).

Používané řečové segmenty

- Alofóny
  - poziční varianty fonémů - obsahuje i části okolních fonémů
  - počet n^3 (n - počet fonémů)
- Difóny
  - začínají uprostřed jednoho fonému a končí uprostřed následujícího
  - počet n^2
  - často využívané pro syntézu i rozpoznávání:
    - MBrola (`http://tcts.fpms.ac.be/synthesis/mbrola.html`)

Používané řečové segmenty
Pokračování

- Trifóny
  - začínají uprostřed levého sousedního fonému a končí uprostřed pravého sousedního
  - počet n^3
  - často využívané pro rozpoznávání a syntézu
- Slabičné segmenty.
- Segmenty proměnné délky získané z korpusu.
- Rámce

Slabiky

- Slabika
  - Slabikovat se učí už děti v první třídě.
  - Nejmenší jednotka organizační jednotka řeči.
  - Nelze odvodit strukturu slabik - nejednoznačnost dělení některých slov na slabiky
    - funk-ční vs funkč-ní.
  - Počet slabik - uvádí se cca 10000.
  - Struktura slabiky
    - preatura (onset)
    - nukleus (vokalické jádro) - bývá to samohláska, příp. dvojhláska, sonora - např. k**r**k, frikativa - např. p**s**t, nazála - např. sed**m**
    - koda - nemusí se vyskytovat
    - nukleus + koda jsou považovány za základ slabiky
    - svahy – preatura a koda; jedná se většinou o jednu nebo více souhlásek.

Slabičné segmenty

- Definovány uměle
- Řešení nejednoznačnosti hranice slabiky.
- Frekventované slabičné typy:
  - V (samohláska/dvojhláska) - ú - kol
  - KV (souhláska - samohláska) - vo - da
  - KVK - jed-not-ka
  - KK - tr-sy
  - KKV - dna
  - KKVK - dmout
- Tvoří vice než 95 % slabik
- Umožňují automatickou segmentaci textu.
- Používají se např. v syntetizéru Demosthénes (doc. Kopeček LAF (LSD) FI)

Syntéza v časové oblasti

- Princip
  - spojování navzorkovaných řečových segmentů uložených v databázi.
- Využívají se různé typy základních segmentů:
  - větší
    - lépe se modelují některé další charakteristiky jako intonace, přízvuky
    - větší nároky na paměť – větší množství segmentů (potenciálně až 2n, kde n je délka segmentu)
    - příklady – slova, části vět
  - menší
    - menší paměťové nároky – menší množství segmentů
    - horší možnost modelování větné intonace, přízvuků, ... (viz oblasti spektrální stacionarity řeči).

Používané řečové segmenty

- Alofóny
  - poziční varianty fonémů – obsahuje i části okolních fonémů
  - počet n3 (n – počet fonémů)
- Difóny
  - začínají uprostřed jednoho fonému a končí uprostřed následujícího
  - počet n2
  - často využívané pro syntézu i rozpoznávání:
    - MBrola (`http://tcts.fpms.ac.be/synthesis/mbrola.html`)
- Trifóny
  - začínají uprostřed levého sousedního fonému a končí uprostřed pravého sousedního
  - počet n3
  - často využívané pro rozpoznávání a syntézu
- Slabičné segmenty.
- Segmenty proměnné délky získané z korpusu.
- Rámce

Slabiky

- Slabika
  - Slabikovat se učí už děti v první třídě.
  - Nejmenší jednotka organizační jednotka řeči.
  - Nelze odvodit strukturu slabik – nejednoznačnost dělení některých slov na slabiky
    - funk-ční vs funkč-ní.
  - Počet slabik – uvádí se cca 10000.
  - Struktura slabiky
    - preatura (onset)
    - nukleus (vokalické jádro) – bývá to samohláska, příp. dvojhláska, sonora – např. k**r**k, frikativa – např. p**s**t, nazála – např. sed**m**
    - koda – nemusí se vyskytovat
    - nukleus + koda jsou považovány za základ slabiky
    - svahy – preatura a koda;jedná se většinou o jednu nebo více souhlásek

Slabičné segmenty

- Definovány uměle
- Řešení nejednoznačnosti hranice slabiky.
- Frekventované slabičné typy:
  - V (samohláska/dvojhláska) – ú – kol
  - KV (souhláska – samohláska) – vo – da
  - KVK – jed-not-ka
  - KK – tr-sy
  - KKV – dna
  - KKVK – dmout
- Tvoří vice než 95
- Umožňují automatickou segmentaci textu.
- Používají se např. v syntetizéru Demosthénes (doc. Kopeček LAF (LSD) FI)

Vlastní syntéza

1. Fonetický přepis.
2. Segmentace dle použitých řečových segmentů.
3. Výběr odpovídajících akustických segmentů
   - databáze segmentů.
4. Spojení segmentů
   - nutné, aby odpovídala F_0 – jinak se vyskytují různé ruchy (lupnutí, ...)
     - vhodné řešit už při vytváření db segmentů.
5. Případný postprocessing

Korpusová syntéza

- Konkatenativní syntéza v časové oblasti.
- Jako db segmentů využívá řečový korpus.
- Nutno doplnit značky pro syntézu:
  - fonetický přepis
  - hranice řečových segmentů
  - průběh F_0
  - ...
- Umožňuje přesnější výběr segmentů
  - snižuje výpočetní složitost spojování a postprocessingu.
- Příklad – viz dizertační práce dr. Batůška v knihovně FI.

Syntéza na bázi rámců

- Většinou se jedná o problémově orientovanou syntézu.
- Syntéza se skládá z:
  - rámců – neměnící se části vět
  - slotů – měnící se části promluvy
- Výhoda:
  - rámce jsem dopředu namluveny a mohou obsahovat intonaci
  - syntetizuje se pouze obsah slotů
    - omezená množina
    - lze použít celá slova
- Příklady:
  - hlášení nádražního rozhlasu:
    - `Osobní vlak číslo <číslo_vlaku> ze směru <seznam_stanic> přijede k <číslo_nástupiště>. nástupišti v <čas>.`

Prozodie

- Výstupem syntézy je monotóní hlas bez intonace a přízvuku – zní nepřirozeně
- Doplnění prozodie
  - základní prozodické prvky:
    - výška
    - hlasitost
    - doba trvání
  - nositelem je slabika
  - Větná intonace (prozodie) – závisí na typu věty:
    - otázky zjišťovací (odpověď ano/ne) – rostoucí
    - oznamovací, tázací doplňovací, rozkazovací – klesající
    - řeší se modulací F_0
  - Doplnění přízvuku/důrazu
    - modifikace F_0 a intenzity
    - lokální modifikace větné melodie

Prozodie – ukázky větné intonace

- Originální promluva (`data/masse.wav`)
- Oznamovací věta (`data/masse-ozn.wav`)
- Otázka zjišťovací (`data/masse-dotaz.wav`)

Výška základního tónu

- Výška základního tónu odpovídá formantu F_0.
- Průběh F_0 na vokalickém jádru bývá nelineární.
- Změna intonace není pouhou změnou F_0
  - nutno modifikovat i vyšší formanty.
- Na základě důležitosti F_0 se jazyky dělí na:
  - tónové (čínština, vietnamština, ...)
    - čínské slovo -ma- v závislosti na průběhu F_0 může znamenat matka, konopí, kůň, nadávat
  - jazyky s melodickým přízvukem (srbština, slovinština, litevština, norština, svédština, ...)

Další prozodické vlastnosti

- Intenzita (hlasitost):
  - fyzikální pohled – intenzita signálu v daném časovém okamžiku
  - fyziologický pohled – reakce vnitřního ucha (cortiho ústrojí) na vnímaný zvuk.
  - Tato hlediska se různí.
    - Subjektivní vnímání zvuku neodpovídá ani v prvním přiblížení fyzikální intenzitě signálu.
- Doba trvání:
  - Slabika může mít různou dobu trvání v různém kontextu.
  - Drobné odchylky mohou být i ve stejném kontextu.
  - Typická doba trvání slabiky 50 — 200 milisekund.

Další prozodické vlastnosti

- Kvalita hlasu
  - chvění hlasu (jitter)
  - nepravidelné výchylky v amplitudě F_0 (shimmer)
  - zbarvení tónu
  - ochraptělost
  - níra znělosti
  - ...
- Rychlost řeči
  - Lze chápat jako převrácenou hodnotu průměrné délky slabiky
  - Lze měřit i jinými způsoby:
    - počtem vyslovených textových znaků za jednotku času (vyhodnocování syntetizérů řeči).

Další prozodické vlastnosti
Pokračování

- Pauza
  - tichá
  - vyplněná – obsahuje nějaký charakteristický zvuk (např. eeh)
    - ztížená detekce – hlavní format je blízký formantům samohlásek "a", "e".
- Zaváhání
  - Přímo vypovídá o pragmatice projevu.
  - Důležitý např. pro modifikaci dialogové strategie u dialogových systémů.
  - Typický případ informace obsažené zejména v prozodické vrstvě jazyka.

Základní odvozené prozodické vlastnosti

- Rytmus (časování):
  - Prozodický prvek odvozený z dob trvání
    - slabik
    - pauz v daném časovém úseku.
- Slovní přízvuk
  - Je odvozen ze všech základních atributů.
  - Je výrazně jazykově závislý:
    - umístění přízvuku ve slově/přízvučné jednotce
    - míra použití prozodických prostředků k jeho vyjádření zejména použití hlasitosti oproti výšce.
- Větný přístup (intonační centrum):
  - zjednudešeně jde o prozodické zvýraznění jádra výpovědi věty

Základní odvozené prozodické vlastnosti (2.)

- Intonace
  - nejobecnější – časový průběh zvukového spektra hlasu
  - za určující pro melodii se obvykle považuje základní hlasová frekvence – lze zobrazit grafem v závislosti na čase
    - časová závislost základní hlasové frekvence
  - související terminologie:
    - melodie
    - kadence
    - intonační kadence
    - melodém
    - průběh F0
- Emotivní zabarvení hlasu
  - projevuje se:
    - rychlými změnami hlasitosti a základní frekvence
  - Často přesahují hranici věty.
  - Detekce je důležitá např. pro dialogové systémy – umožňuje zvolit vhodnou dialogovou strategii.

Základní odvozené prozodické vlastnosti (3.)

- Emfatický přízvuk
  - Vytvářen emotivním zbarvením hlasu.
  - Vyskytuje se např. ve větách pronesených v situacích s výrazným emocionálním kontextem, např.
    - To je tedy opravdu **neslýchané.**
    - Bolí to jak **čert**.
- Kontrastní přízvuk
  - snaha o zdůraznění slova nebo slabiky v kontrastu s jiným slovem nebo slabikou během promluvy nebo dialogu:
    - "řekl jsem do **Šakvic** ne **Rakvic"**
    - "**byte** ne **bit"**

Základní odvozené prozodické vlastnosti (4.)

- Opakování
  - prozodický atribut silně svázaný s mluvčím.
  - Opakování bývá často variantou výplňkových částí promluvy – mluvčí si ji často ani neuvědomuje (nezaměňovat s koktáním – porucha řeči).
  - Může se jedna o formu zdůraznění – v krajním případ může být považováno za vadu řeči.
- Výplňkové části
  - kromě výplňkové funkce mohou charakterizovat
    - styl mluvčího: ,,Byl jsi včera na akci, **viď**?"
  - nářečí resp. slang: ,,**Vole**, ta včerejší spářka byla hustá, že **vole**?"

Základní odvozené prozodické vlastnosti (5.)

- Přerušení:
  - častý jev v mluvené řeči na úrovní:
    - vyšších celků (výpověď/promluva, věta, prozodická fráze, ...)
    - uvnitř slov.
  - Mívá návaznost na další prosodické prvky:
    - zaváhání
    - opakování
    - vyplňenou pauzu
    - ...
  - Zvyšuje obtížnost rozpoznávání mluvené řeči – nutno s ním počítat.
- Korekce částí promluvy:
  - Častý jev a to vzhledem k rozdílným částem.
  - Příčiny vzniku:
    - důsledek přeřeknutí,
    - upřesnění předchozí části promluvy,
    - oprava předchozí části promluvy.
  - Často následuje přerušení nebo další prozodické jevy.

Prozodické segmenty mluvené řeči

- Prozodické segmenty mluvené řeči:
  - Promluva.
  - Prozodická fráze
    - Skupina slov vytvářející jednotný intonační celek.
    - Představuje základní, z prozodického hlediska kompaktní strukturu.
    - Členění do prozodických frází ve velké míře souvisí se syntaktickou strukturou odpovídající věty.
  - Přízvukový takt
    - skupina slabik podřízená jednomu slovnímu přízvuku.
    - V češtině typicky slovo nebo slovo a jednoslabičné slovo.
  - Slabika

Standardy pro syntézu řeči

- Snaha sjednotit jazyky pro popis promluvy pro řečové syntetizéry.
- Definují značkování postihující:
  - prozódii
    - rychlost řeči
    - F_0
    - zdůraznění části promluvy
    - pauzu
    - hlasitost
    - ...
  - mluvčího
    - pohlaví
    - věk
    - ...
  - ...
- Používané standardy:
  - SABLE
  - SSML

SABLE

- Vývoj započat v 2. polovině 90. let
- aplikace XML/SGML
- snaha o zkombinování 3 značkovacích jazyků pro syntézu řeči:
  - SSML – Speech Synthesis Markup Language (`http://www.w3.org/TR/2010/REC-speech-synthesis11-20100907/`) (W3C, 1999)
  - STML – Spoken Text Markup Language (`http://www.bell-labs.com/project/tts/stml.html`) (CSTR Edinburgh University, Lucent Technologies, 1997)
  - JSML – Java Synthesis Markup Language (`http://www.w3.org/TR/jsml/`) (Sun Microsystems, 2000)
- `http://www.bell-labs.com/project/tts/sable.html`

- SABLE – kořenová značka
- div – slouží k logickému členění dokumentu (odstavec, věta)
- prozodické:
  - EMPH – zdůraznění části promluvy
  - PITCH – výška promluvy
  - VOLUME – úroveň hlasitosti
  - RATE – rychlost
  - BREAK – pauza
- popis hlasu:
  - SPEAKER – popisuje pohlaví a věk mluvčího
- fonetické
  - PRON – výslovnost – fonetický přepis
  - SAYAS – způsob fonetického přepisu (datum, telefon, url, poštovní adresa, ...)
  - LANGUAGE – jazyk promluvy

SABLE – ukázka

```
<SABLE>
 <DIV TYPE="paragraph">
  <VOLUME LEVEL="quiet">Šepot.</VOLUME>
  <VOLUME LEVEL="medium">
   <RATE SPEED="fast">Rychlá věta.</RATE>
   <PITCH BASE="+50%">Vysoko posazená věta</PITCH>
  </VOLUME>
 </DIV>
</SABLE>
```

SSML

- Vývoj započat v koncem 90. let
- součást W3C Voice Browser Activity (`http://www.w3.org/Voice`)
- Aktuální verze 1.0 (září 2004)

- kořenový element – speak
- strukturní elementy
  - p – odstavec
  - s – věta
- fonetické:
  - say-as – způsob fonetického přepisu (výslovnosti, datum, telefon, url, číslo, ...)
  - phoneme – fonetický přepis dané promluvy
  - sub – substituce (např. přepis zkratek, ...)
- popis hlasu:
  - voice – popis hlasu, kterým se má text přečíst (pohlaví, věk, ...)
- prozodie:
  - emphasis – zdůraznění částí promluvy
  - break – pauza
  - prosody – ovlivňuje prozodické jevy: výšku, průběh základní frekvence, rychlost, item délka trvání promluvy, hlasitost.

```
<?xml version="1.0"?>
<speak>
 <voice gender="female">Female voice.</voice>
 <voice gender="male">Male voice.</voice>
 <emphais level="soft">Soft emphasis</emphasis>
 <p>Speech with 5 seconds <break time="5s"/> break.</p>
 <prosody volume="+6dB">Speech at double volume.</prosody>
 <prosody volume="-6dB">Speech at half volume.</prosody>
</speak>
```

Motivace

- Dialogový systém - informační systém s dialogovým (hlasovým/textovým) rozhraním.
- Přirozenější způsob komunikace pro většinu uživatelů než GUI.
- Poskytují nové způsoby komunikace s aplikacemi:
  - telefon
  - hlasová komunikace prostřednictvím počítače s náhlavní soupravou (mikrofonem, reproduktory).
- Možnost komunikace bez použití končetin.
- Zlepšují přístupnost pro uživatele s různými druhy postižení:
  - zrakově postižení
  - imobilní uživatelé
  - ...
- Při dobře navrženém dialogovém rozhraní může být komunikace podobně efektivní jako GUI.
  - grafická komunikace - paralelní
  - hlasová komunikace - sekvenční

Historie dialogových systémů

- Eliza
  - počátek 60. let
  - počítačová simulace rozhovoru s psychoterapeutem
  - textové komunikace v přirozeném jazyce
- Parry
  - autor K. M. Colby (1963)
  - simulace paranoidního pacienta - reakce na Elizu
  - v řadě dialogů nebylo možné jednoznačně určit, zda se jedná o simulaci nebo reálného pacienta

- Pracují se znalostní databází vytvořenou experty v dané oblasti
- Znalostní databáze obsahují:
  - fakta
  - inferenční pravidla - pravidla pro odvozování závěrů na základě zjištěných faktů.
- DENDRAL - expertní systém z oblasti organické chemie
- INTERNIST I- expertní systém pro pomoc při diagnostice (1970, University of Pittsburgh Medicine School)
- MYCIN
  - Stanford University (70. léta)
  - navazuje na INTERNIST I (jeden z auto společného jednoho z
  - obsahoval i pokročilá odvozovací pravidla
  - diagnostika bakteriálních onemocnění
  - ve 3/4 případů shoda s lidským expertem

Aktuální stav oblasti dialogových systémů

FI MU

- Laboratoře:
  - LSD (`http://lsd.fi.muni.cz/`) – Laboratoř vyhledávání a dialogu
    - vedoucí – doc. Kopeček, prof. Zezula
    - zaměření – vyhledávání, dialogové systémy, zpracování zvuku, item asistivní technologie, ...
  - NLP (`http://nlp.fi.muni.cz`) – Laboratoř zpracování přirozeného jazyka
    - Vedoucí – doc. Pala
    - zaměření – textové korpusy, slovníky, morfologie, syntaktická analýza, sémantická analýza, ...

Aktuální stav oblasti dialogových systémů

ČR

- FIT VUT Brno:
  - analýza signálu
  - rozpoznávání řeči
  - systémy pro automatizovaný záznam a zpracování konferencí
  - ...
- ZČU Plzeň
  - rozpoznávání řeči
  - syntéza řeči
  - dialogové systémy
  - ...
- ČVUT Praha
  - syntéza řeči
  - počítačová lingvistika
  - ...

- World Wide Web Consortium Voice Browser Working Group (`http://www.w3.org/Voice/`)
  - vývoj a správa standardů pro tvorbu dialogových rozhraní
  - vývoj a správa standardů pro tvorbu multimodálních dialogových rozhraní
  - členové: IBM, Nuance Communication, Lucent Technologies, Motorola, ScanSoft, Tellme Networks, Vocalocity, ...
- MIT
- Carnegie Mellon University (CMU)
- OGI
- EPF Lausane
- ...

Současné oblasti použití

- Dialogové informační systémy o dopravních spojeních.
- Dialogové ovládání některých systémů v automobilech:
  - navigace
  - telefon
  - ...
- Dialogové systémy pro zdravotně postižené -
- Vojenské aplikace.
- ...

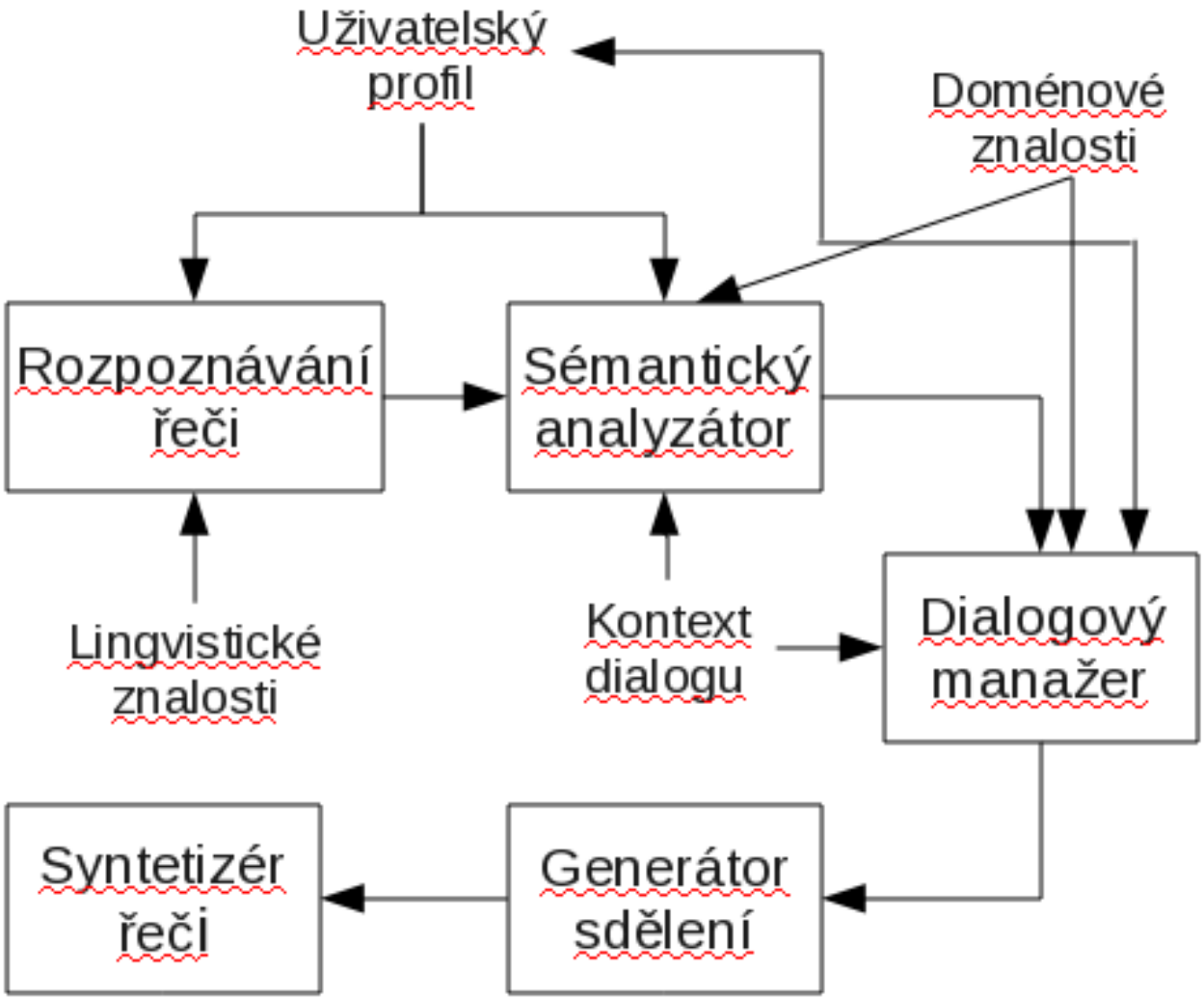
Základní pojmy

- Dialogový systém - informační systém disponující dialogovým rozhraním.
- Dialogové rozhraní - UI, které umožňuje uživateli komunikaci s aplikací prostřednictvím dialogu.
- Dialog - komunikace dvou účastníků (pro nás člověk - počítač).
- Promluva - souvislé sdělení jednoho účastníka dialogu.
- Obrat - promluva a reakce druhého účastníka na ni.
- Dialogová strategie - určuje ke každé promluvě následníka.
- Hodnotící funkce E(L) - přiřazuje dialogu reálné číslo (jeho ohodnoceni).
- Dialogová komunikace - uspořádaná čtveřice M=(S1, S2, E1, E2).

Charakteristické rysy dialogových systémů

- Převládající řečová komunikace.
  - Problémy s rozpoznáváním řeči - řešení:
    - omezení problémové domény
    - gramatiky pro rozpoznávání řeči.
  - Bývá doplněna o vstup pomocí:
    - DTMF - telefonní aplikace
    - klávesnice - webové aplikace komunikující dialogem v přirozené řeči.
- Tendence ke komunikaci přirozenou řečí s co nejmenšími omezeními.
  - Vede na rozpoznávání plynulé řeči.
  - Řešení např. omezením množiny akceptovaných promluv pomocí gramatik.
- Snaha o co největší efektivitu a optimalitu komunikace.
  - Změny dialogové strategie v závislosti na zkušenostech uživatele - lze odhadnou z průběhu komunikace:
    - zkušený uživatel - průběh dialogu určuje spíše uživatel
    - nezkušený uživatel - DS se snaží uživatele co nejvíce vést a poskytovat mu co nejvíce nápovědy.

Dialogový systém



Komponenty dialogového systému

- Rozpoznávání řeči - rozpoznání promluvy
  - rozpoznání promluvy
  - ke zvýšení úspěšnosti využívá
    - lingvistické znalosti
    - uživatelský profil
- Sémantický analyzátor
  - zjištění významu promluvy
  - využívá:
    - uživatelský profil
    - doménové znalosti
    - kontext dialogu
- Dialogový manažer
  - na základě známých faktů rozhoduje o dalším kroku dialogu ze strany systému
- Generátor sdělení
  - generuje sdělení podle požadavků dialogového manažeru.

Informace využívané dialogovým systémem

- Lingvistické znalosti - jazykový model, ...
- Uživatelský profil - model řečníka, emoční model, ...
- Doménové znalosti - informace o oblasti DS použitelné pro interpretaci rozpoznané promluvy, pro rozhodování dialogové strategie, ...
  - informace od oblasti dialogového systému
  - použitelné pro:
    - interpretaci rozpoznávané promluvy
    - rozhodování o dalším kroku dialogu (dialogovou strategii)
  - patří sem:
    - např. jaká data je zapotřebí zadat
- Kontext dialogu
  - uchovává aktuální stav dialogu
    - zadané údaje
    - informace o chybách (nerozpoznané promluvy, chyby v zadávaných údajích)
  - lze použít např. pro použití vhodné gramatiky pro sémantickou interpretaci rozpoznané promluvy

Dialogová strategie

- Zobrazení $UxQ \rightarrow UxQ$.
  - U - promluva
  - Q - stav
- Určuje následující krok dialogu v závislosti na stavu dialogu a vstupní promluvě.
- Hodnotící funkce dialogu E přiřazuje danému dialogu reálné číslo popisující úspěšnost dialogu z pohledu dané strany.
- Dělení dialogů z hlediska hodnotící funkce:
  - d je kooperativní dialog - $E_1(d) = E_2(d)$
  - d je nekooperativní dialog - $E_1(d) \neq E_2(d)$
  - d je dialog s nulovým součtem - $E_1(d) = -E_2(d)$

Pravidla pro vedení kooperativního dialogu (H. P. Grice)

- Autorem Herbert Paul Grice - anglický jazykovědec
- Aspekt informativnosti
  1. Buď přiměřeně informativní (ne méně než je potřeba, ne více než je potřeba).
- Aspekt přesvědčivosti
  1. Neuváděj nepravdivé informace.
  2. Neuváděj informace, které nelze dokázat nebo doložit.
- Aspekt způsobu
  1. Informace v replice by měla být co nejvíce explicitní
  2. Vyhýbejte se nejednoznačnostem.in
  3. Usilujte o stručnost.
  4. Buďte disciplinovaní, udržujte v dialogu pořádek.

Pravidla pro vedení kooperativního dialogu
Pokračování

- Aspekt zdvořilosti, empatie a etiky
  1. Minimalizujte nároky vůči komunikačnímu partnerovi, maximalizujte výhody pro něj.
  2. Minimalizujte nedostatky komunikačního partnera a maximalizujte jeho přednosti.
  3. Maximalizujte souhlas s partnerem a minimalizujte nesouhlas.
  4. Maximalizujte empatii vůči partnerovi.

Pravidla pro vedení kooperativního dialogu (H. P. Grice) - komunikace člověk počítač

- Aspekt asymetrie
  1. Informujte uživatele o všech důležitých charakteristikách, která vybočují z očekávaného normálního průběhu dialogu a která by měl vzít v úvahu k zajištění kooperativity.
  2. Zajistěte stručné avšak dostatečné informování uživatele o možnostech systému a jeho omezeních.
  3. Informujte srozumitelně a dostatečně uživatele o způsobu interakce ze systémem.

Pravidla pro vedení kooperativního dialogu (H. P. Grice) - komunikace člověk počítač

Pokračování

- Aspekt znalostí a schopností
  1. Vezměte v úvahu relevantní znalosti uživatele.
  2. Vezměte v úvahu možné uživatelovy chybné analogie.
  3. Rozlišujte mezi začínajícím a zkušeným uživatelem systému.
  4. Vezměte v úvahu legitimní představy uživatele o znalostech a schopnostech systému.
- Aspekt vyjasňování a odstraňování chyb
  1. V případě selhání komunikace iniciujte metakomunikaci zajišťující odstranění chyby nebo její vysvětlení.
  2. Zajistěte vysvětlující metakomunikaci v případě nekonsistentních nebo nejednoznačných uživatelových vstupních dat.

Aspekty komunikace kooperativního dialogového systému

- aspekt informativnosti
- aspekt přesvědčivosti
- aspekt způsobu
- aspekt zdvořilosti, empatie a etiky
- aspekt asymetrie
- aspekt znalostí a schopností uživatele
- aspekt vyjasňování a odstraňování chyb

Iniciativa v dialogu

- Další krok dialogu je vždy určen dialogovou strategií jedné z komunikujících stran - jedna strana klade dotazy, druhá na ně odpovídá.
- V případě komunikace člověk - počítač lze rozlišit
  - dialog s iniciativou uživatele
  - dialog s iniciativou systému
  - dialog se smíšenou iniciativou.
- V reálném nasazení se používají:
  - dialogy se smíšenou iniciativou
  - dialogy s iniciativou systému.

Iniciativa v dialogu - příklady

- Dialog s iniciativou systému:

```
Systém: Zadejte Vaše křestní jméno
Uživatel: Jan
Systém: Zadejte Vaše příjmení
Uživatel: Novák
```

- Dialog s iniciativou uživatele:

```
Uživatel: Chtěl bych bych si rezervovat knihu XY
Systém: Dobře.
Uživatel: A film UV.
Systém: Dobře.
Uživatel: To je vše.
Systém: Vaše rezervace knihy XY a fulmu UV byla
        přijata.
```

Iniciativa v dialogu - příklady

- Dialog se smíšenou iniciativou:

```
Uživatel: Chtěl bych si zaregistrovat předmět PB123.
Systém: S jakým zakončením?
Uživatel: Zkouška.
Systém: Registruji Vám předmět PB123 se zakončením
        zkouškou. Souhlasí?
Uživatel: Ano.
```

Zpětná vazba v dialogovém systému

- Před tím, něž systém předá získané informace k dalšímu zpracování je vhodné provést jejich verifikaci.
  - chyby rozpoznávání řeči
  - chyba uživatele
  - ...
- Způsoby ověření získaných dat
  - sumarizující zpětná vazba
  - zpětná vazba "echo"
  - implicitní zpětná vazba
  - explicitní zpětná vazba.
- V případě nesouhlasné reakce uživatele následuje opravný dialog.

Ukázky zpětné vazby

- Sumarizující zpětná vazba:

```
Uživatel: Jmenuji se Jan Novak a narodil jsem se
          8.8. 1988 v ABC.
Systém: Bylo zadáno jméno Jan Novák, datum narození
        8.8. 1988 a místo narození ABC.
```

- Zpětná vazba "echo":

```
Systém: Jak se jmenujete?
Uživatel: Jan Novák.
Systém: Jmenujete se Jan Novák. Kdy jste se narodil.
Uživatel: 8.8. 1988
Systém: Narodil jste se 8.8. 1988. Ve kterém městě?
Uživatel: ABC Systém: Narodil jste se v městě ABC.
```

Ukázky zpětné vazby
Dokončení

- Implicitní zpětná vazba:

```
Systém: Jméno studenta.
Uživatel: Jan Novák.
Systém: Ve kterém městě se Jan Novák narodil?
...
```

- Explicitní zpětná vazba:

```
Systém: Zadejte jméno studenta.
Uživatel: Jan Novák
Systém: Student se jmenuje Jan Novák, je to tak?
```

Ukázka opravného dialogu

```
Uživatel: Jmenuji se Jan Horák a narodil jsem se
          8.8. 1988 v ABC.
Systém: Bylo zadáno jméno Jan Novák, datum narození
        8.8. 1988 a místo narození ABC.
Uživatel: Ne. Jméno je Jan Horák.
Systém: Zadal jste jméno Jan Horňák, datum narození
        8.8. 1988 a místo ABC.
Uživatel: Ne, jméno má být Jan Novák.
Systém: Zadal jste jméno Jan Novák, datum narození
        8.8. 1988 a místo ABC.
```

Další aspekty dialogových strategií

- Možnost přerušit systém.
- Možnost iniciovat korekce a metakomunikaci uživatelem.Nerozumím, mohl byste to prosím zopakovat?
- Možnost získání nápovědy.Nevím, co po mě požadujete.
- Detekce znalostí a schopností uživatele.
  - Částečně lze odvodit z předchozího průběhu dialogu.
  - Uzpůsobení iniciativy dialogové strategie - iniciativa systém vs. smíšená iniciativa.

Další aspekty dialogových rozhraní

- Přizpůsobení dialogové strategie - např. uživatel ve spěchu, vystresovaný uživatel, ...
- Detekce emocí uživatele:
  - na základě neverbálních charakteristik hlasu
  - nutná podpora v modulu rozpoznávání řeči.
- Vícejazyčnost (multilingualita):
  - nutnost podpory na straně použité platformy:
    - rozpoznávání řeči
    - sémantická analýza řeči
    - syntéza řeči.
  - Možnost pokusit se o automatickou detekci použitého jazyka.
    - Paralelní rozpoznávání všemi dostupnými rozpoznávači - použije se výstup s maximální pravděpodobností.

Další aspekty dialogových rozhraní

- Multimodalita:
  - Umožňuje paralelní komunikaci více kanály - (obraz, zvuk, hmat).
    - Zlepšuje přístupnost
  - Příklady multimodálních rozhraní:
    - Rozhovor vede avatar (talking head) - vhodné pro uživatele s poruchou slyšení.
    - Ruce/avatar (celé tělo resp. horní polovina) - provádí tlumočení do znakové řeči.
  - Alternativní způsoby vstupu - klávesnice, kamera, snímače aktivity mozku, svalů (krk, obličej, ...), různé joysticky, ...
  - Nutnost synchronizace jednotlivých kanálů.

Další aspekty dialogových rozhraní
Pokračování

- Zdvořilost - viz pravidla vedení kooperativního dialogu
- Prozódie - určení sémantiky a pragmatiky promluvy:
  - určení druhu věty (tázací (`data/masse-dotaz.wav`)/oznamovací (`data/masse-ozn.wav`)), ...
  - detekce emocí
  - ...
- Učení se z chyb.
  - Zapamatování si nerozpoznané promluvy a pokud uspěje opravný dialog (zpětná vazba), pokus o analýzu původní promluvy a přidání typu promluvy do lingvistických znalostí.

Přístup k dialogovému rozhraní

- Telefonní:
  - PSTN (Public Switched Telephone Network) - veřejná telefonní sít.
  - Nutnost digitalizace uživatelského vstupu a připojení počítače k PSTN:
    - voice-modem
    - ISDN modem
    - výstup lze již přímo zpracovat pomocí ASR.
  - VoIP - protokol pro přenos hlasu přes IP.
    - Většinou jako rozšíření ústředny o dialogový manažer (platformu).
    - Např. Asterisk + VoiceGlue, ...
    - komunikace prostřednictvím VoIP protokolů
  - Možnost využití DTMF (Dual Tone Multi-Frequency).

Přístup k dialogovému rozhraní

- Textové:
  - Odpadá nutnost digitalizace uživatelského vstupu.
  - Velmi vhodný pro ladění rozhraní:
    - není nutno řešit chyby ASR, ...
  - Lze využít i IM
    - Asterisk + XMPP (Jabber).

Části dialogových systémů

- Rozpoznávání řeči:
  - ASR - nejlépe s podporou rozpoznávání plynulé řeči
  - Pro zvýšení úspěšnosti jsou použity bezkontextové gramatiky - občas slouží jako základ i pro sémantickou interpretaci.
  - Lze částečně nahradit pomocí DTMF.
- Sémantická interpretace:
  - Atributy se sémantickou interpretací u CFG.
  - Občas se používá keyword spotting.
- Dialogový manažer:
  - logické programování
  - různá řešení pomocí procedurální programovacích jazyků
  - proprietární řešení
  - otevřená řešeni.

Části dialogových systémů

- Generátor promluv:
  - pracuje s výstupem z dialogového manažeru
  - bývá součástí dialogového manažeru
  - generuje textovou verzi promluvy.
- Hlasový syntetizér:
  - TTS.

VoIP

- Rodina protokolů pro přenos hlasu přes internet (IP síť).
- Využívá se např. pro IP telefonii.
- Na transportní vrstvě - UDP.
- Na relační vrstvě - Real-Time Protocol (RTP).
- Řada implementací
  - Liší se použitými standardy:
    - H323 - na ústupu
    - SIP
    - firemní - Skinny (Cisco), HFA - Siemens.
  - Službami a signalizací:
    - podpora video hovorů, zasílání textových zpráv, ...
- Více viz `http://www.protocols.com/pbook/VoIP.htm`

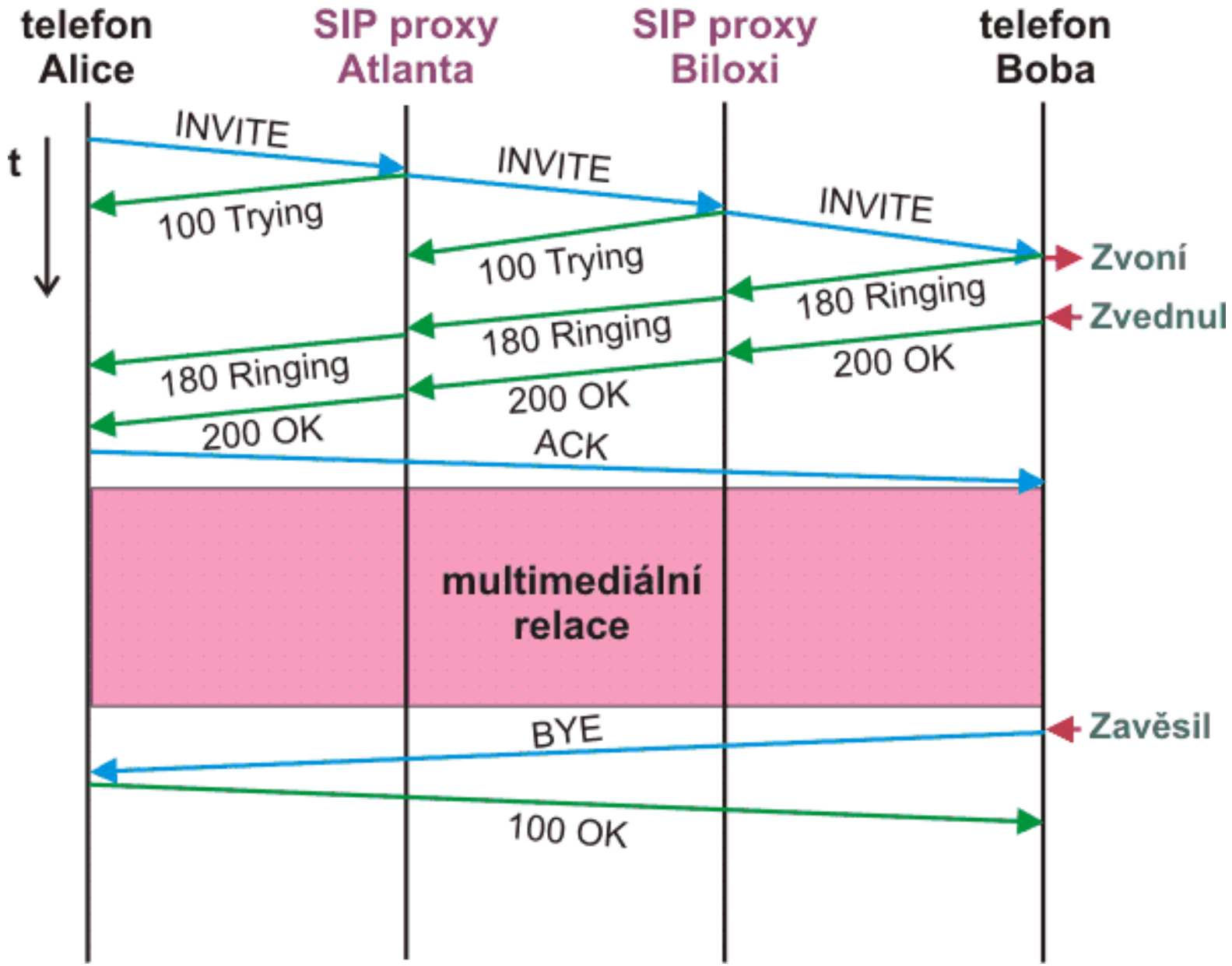
SIP

- Session Initiation Protocol (viz SIP RFCs and Drafts (`http://www.cs.columbia.edu/sip/drafts.html`), resp.IETF (`http://www.ietf.org/rfc/`)/sbírka Internetových standardů na FI (`http://ftp.fi.muni.cz/pub/internet-drafts/`), ...)
- Protokol na aplikační vrstvě.
- Protokol určený pro přenos signalizace v internetové telefonii.
- Na transportní vrstvě využívá UDP.
- Vznikl jako reakce na H.323 - příliš komplexní (obtížně implementovatelný).
  - Je jednodušší.

SIP

- Pro vytvoření a řízení multimediální relace zajišťuje:
  - Lokalizaci účastníka (jednoznačná identifikace - uživatel, adresa SIP serveru).
  - Zjištění stavu účastníka - dostupný, obsazený, přesměrovaný.
  - Zjištění možností účastníka - použitelný kodek, max. přenosová rychlost, použitelnost video hovorů, ...
  - Navázání spojení:
    - využívá se protokol SDP popisující navázané spojení (Session Description Protocol)
    - vlastní hovor využívá protokol RTP (Real-Time Protocol).
  - Řízení probíhajícího spojení:
    - změny parametrů spojení v jeho průběhu
    - ukončení spojení.

Průběh relace přes protokol SIP



Podpora rozpoznávání řeči

- Omezení domény možných vstupů.
- Bezkontextové gramatiky popisující množinu možných vstupů.
- Používané způsoby:
  - prostředky logického programování
  - různá proprietární řešení
  - otevřené standardy
    - JSGF
    - SRGS
    - ...

Java Speech Grammar Specification (JSGF)

- Textový zápis gramatiky nezávislý na platformě a prodejci, sloužící pro podporu rozpoznávání řeči.
- Určen pro použití při rozpoznávání řeči.
- Používá styl a konvence styl a konvence jazyka Java.
- Součást Java Speech API.
- Aktuální verze 1.0 (říjen 1998).
- Využit např. v rozpoznávači Sphinx4J (`http://cmusphinx.sourceforge.net/`), VoiceXML interpretru VoiceGlue, ...

JSGF - Syntaxe

- Gramatika se skládá z pravidel, které popisují co může být řečeno.
- Syntaxe je case-sensitive.
- Kódování znaků - Unicode.
- Formát hlavičky:

```
#JSGF version [char-encoding [locale]];
#JSGF V1.0;
#JSGF V1.0 ISO8859-2;
#JSGF V1.0 UTF-8 cs_CZ;
```

JSGF - Syntaxe

- Názvy pravidel nesmí obsahovat bílé znaky.
- Neterminální symboly:
  - <názevNeterminálu>, <mesto>, <anone>
- Terminální symboly
  - víceslovní terminály a zvláštní symboly mohou být uzavřeny do uvozovek:"Nové Město na Moravě" "+"
- Zvláštní pravidla
  - <NULL> - pravidlo, které je automaticky použito, aniž by uživatel cokoliv řekl
  - <VOID> - pravidlo, které nemůže být řečeno
- Deklarace gramatiky:
  - grammar *názevBalíku.názevGramatiky;*

JSGF - syntaxe (pokračování)

- Vkládání gramatik - umožňuje používat pravidla nebo gramatiky definované v jiném souboru.
  - import *fullyQualifiedRuleName*;
  - import *fullGrammarName*;import <com.sun.speech.app.numbers.one>; import <com.sun.speech.app.numbers.*>;
- Deklarace gramatiky
  - grammar *názevBalíku.názevGramatiky;*
- Tělo gramatiky
  - *neterminál* = *pravidlo*;

```
<jmeno> = Jan | Jana |...;
<jmeno> = <krestniJmeno> <prijmeni>;
```

- Více viz specifikace (`http://java.sun.com/products/java-media/speech/forDevelopers/JSGF/JSGF.html#16587`).

JSGF - ukázka

```
JSGF V1.0
#import cz.mesta.*;
#import cz.hodiny.*;
<koren> = Chci jet <cim>.| Chci jet <cim> z <odkud> do <kam
          Chci jet <cim> z <odkud> do <kam> v <kdy>;
<cim> = vlakem | autobusem;
<odkud> = <czMesto>;
<kam> = <czMesto>;
<kdy> = <czCas>;
```

SRGS

- Standard W3C Voice Browser Activity WG.
- Aktuální verze 1.0 (březen 2004).
- Definuje způsob zápisu pravidel a jejich odkazování.
- Dva způsoby zápisu
  - XML
  - ABNF (Augmented BNF).
- Více později při probírání VoiceXML

Sémantická interpretace promluvy

- Většinou řešeno pomocí atributů v gramatice pro rozpoznávání řeči.
- Slouží k určení umístění a hodnoty významných částí uživatelské promluvy.
- JSGF:
  - K pravidlu je přiřazena jeho sémantická interpretace.
  - Zapisuje se:

    ```
    {sémantická interpretace};
    ```

  - Příklad:

    ```
    <souhlas> = <ano> {ano} | <ne> {ne};
    <ano> = ano | jo | jasně;
     <ne> = ne | ani náhodou;
    ```

Sémantická interpretace promluvy
Pokračování

- SRGS:
  - K pravidlu je přiřazena jeho sémantická interpretace.
  - Používá se standard Semantic Interpretation for Speech Recognition (SISR (`http://www.w3.org/TR/semantic-interpretation/`))
  - Sémantická interpretace může obsahovat výrazy v jazyce ECMAScript (`http://www.ecma-international.org/publications/files/ECMA-ST/Ecma-262.pdf`).
    - Skriptovací jazyk standardizovaný organizací ECMA (`http://www.ecma-international.org/`) (European Computer Manufacturer Association).
    - Používá se pro skriptování na straně klienta ve webových stránkách.
    - Implementace - JavaScript, JScript, ActionScript.
    - Jednotlivé implementace přidávají nestandardní knihovny (práce s prohlížečem, ...)
  - Více později při probírání VoiceXML.

Popis dialogových rozhraní

- Popis ve vyšším programovacím jazyce:
  - prostředky logického programování (Prolog)
  - procedurální programovací jazyky (C/C++, Java, ...) - např. projekt AudiC (LSD FI),
- Proprietární řešení.
- Otevřené standardy:
  - VoiceML - předchůdce VoiceXML (2. polovina 90. let)
  - VoiceXML - součást W3C VoiceBrowser Activity (`http://www.w3.org/Voice/`)
  - CallXML - Voxeo Prophecy (`http://www.voxeo.com/`)

Syntéza řeči

- Tvorba promluvy:
  1. Dialogový manažer zvolí rámec pro požadovanou výstupní promluvu.
  2. Doplní se do ní hodnoty slotů.
  3. Předá se řečovému syntetizéru.
- Značkování prozodických jevů:
  - závislé na použitém TTS
  - Speech Synthesis Markup Language (SSML) - součást standardů W3C VoiceBrowser Activity
    - TTS musí obsahovat podporu pro tento standard.

Online nástroje pro tvorbu dialogových rozhraní

- Webové nástroje
  - Nuance (BeVocal) Café (???)
    - podporované standardy: VoiceXML 2.1, SRGS 1.0
    - obsahuje sadu nástrojů pro ladění dialogových rozhraní
  - Tellme Studio (`https://studio.tellme.com/`)
    - podporované standardy: VoiceXML 2.x, SRGS + SISR
  - Voxeo Prophecy (`http://www.voxeo.com`)
    - podporované standardy: CallXML, VoiceXML 2.0, SRGS, CCXML
  - další viz např. seznamy na W3C VoiceBrowser Activity (`http://www.w3.org/Voice/#implementations`) nebo na Wikipedii (`http://en.wikipedia.org/wiki/Dialogue_systems`)

Desktopové nástroje pro tvorbu dialogových rozhraní

- Trinidkit (`http://www.ling.gu.se/projekt/trindi//trindikit/`)
  - toolkit pro tvorbu dialogových rozhraní založený na logickém programovaní (Sicstus Prolog)
- CSLU Toolkit (`http://www.cslu.ogi.edu/toolkit/`)
  - vývojové prostředí pro tvorbu dialogových rozhraní
  - umožňuje snadnou tvorbu i multimodálních dialogových rozhraní
  - pro Win32
- Voxeo Prophecy (`http://www.voxeo.com/prophecy/home.jsp`)
  - on-line VoiceXML platforma - bezplatně max. 2 připojení k hostovanému rozhraní
  - možnost instalace serveru na vlastní počítač
  - podobné možnosti jako online verze

Desktopové nástroje pro tvorbu dialogových rozhraní

- Voxeo VoiceObjects (`http://www.voxeo.com/voiceobjects/`) - platforma pro tvorbu a testování hlasových aplikací.
- Textový editor, nejlépe se zvýrazněním syntaxe + (desktopová) VoiceXML platforma:
  - JVoiceXML (???)
  - publicVoiceXML (`http://publicvoicexml.sourceforge.net/`)
  - OptimTalk (`www.optimsys.cz/cs/optimtalk/optimtalk`)
  - ...
- Další viz např. seznamy na W3C VoiceBrowser Activity (`http://www.w3.org/Voice/#implementations`) nebo na Wikipedii. (`http://en.wikipedia.org/wiki/Dialogue_systems`)

Úvod

- 1876 - udělen patent na telefon A. G. Bellovi
- WWW
  - 1989 - článek HyperText and CERN (Tim Burns Lee) koloval po CERNu k připomínkám
  - Vánoce 1990 - demonstrován řádkový webový prohlížeč a editor
  - 1991 - všeobecná dostupnost WWW na počítačích v CERNu
  - 1994 - první setkání W3 konsorcia (`www.w3.org`)
- W3C VoiceBrowser Working Group (`http://www.w3.org/Voice`)
  - založena 1999
  - cíl - návrh standardů umožňujících přístup k WWW pomocí hlasu a telefonu
  - zastoupeny firmy jako:
    - HP
    - Nuance Communications
    - Lucent Technologies
    - Motorola
    - ScanSoft
    - IBM

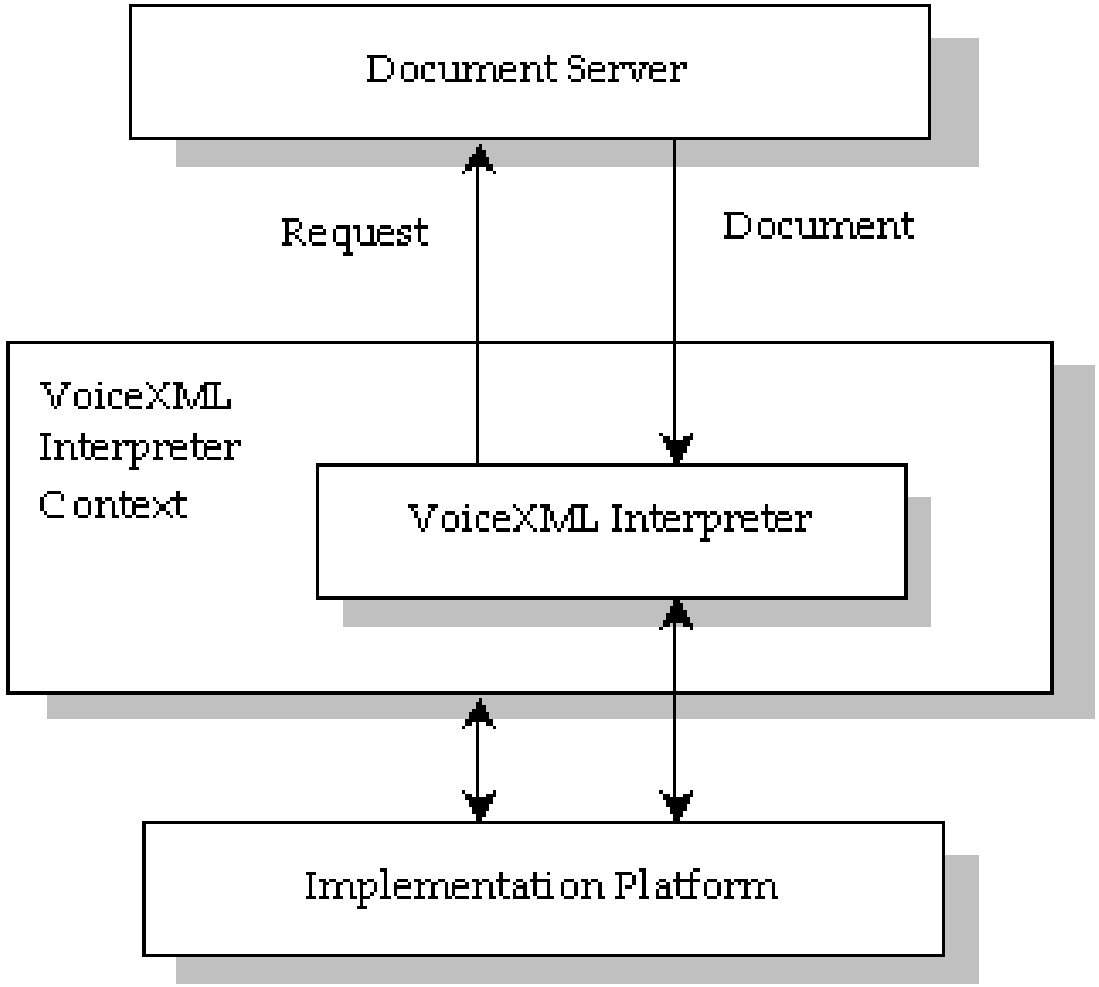
Standardy W3C Voice Browser Activity

- VoiceXML (`http://www.w3.org/TR/voicexm20/`)
- Speech Recognition Grammar Specification (SRGS (`http://www.w3.org/TR/speech-grammar/`))
- Speech Synthesis Markup Language (SSML (`http://www.w3.org/TR/speech-synthesis/`))
- Semantic Interpretation for Speech Recognition (SISR (`http://www.w3.org/TR/semantic-interpretation/`))
- Pronunciation Lexicon Specification (PLS (`http://www.w3.org/TR/pronunciation-lexicon/`))
  - slouží k popisu fonetických informací pro rozpoznávání a syntézu řeči
  - výslovnost zkratek, místních jmen, ...
- Call Control XML (CCXML (`http://www.w3.org/TR/ccxml/`))
- State Chart XML (SCXML (`http://www.w3.org/TR/scxml/`))

Základní informace

- Jazyk pro popis dialogových rozhraní
- Cíl - přinést výhody webového vývoje a doručování obsahu do interaktivních hlasových aplikací
- vývoj započat 1995 - AT&T Phone Markup Language
- 1998 - konference hostovaná W3C na téma hlasového procházení WWW - předvedeny jazyky PML, VoxML, SpeechML, TalkML, VoiceHTML, ...
- 1999 - založeno VoiceXML Forum - spojení sil při vývoji jazyka pro značkování dialogů
- 2000 - VoiceXML 1.0, krátce na to přijato jako standard W3C
- Aktuální verze:
  - doporučení 2.1 (červen 2007)
  - draft 3.0 (srpen 2010)

Architektura VoiceXML aplikací



Struktura VoiceXML aplikací

- VoiceXML dokument(y)
  - formuláře - konečně stavové automaty.
  - Uživatel se nachází v jednom z konverzačních stavů.
  - Přechody definovány pomocí URI - odkazují na další krok dialogu.
    - URI - Uniform Resource Identifier
    - jednoznačná identifikace zdroje (souboru, obrázku, ...) na Internetu
    - rozšíření URL (URL je odkaz na soubor, cíl URI nemusí existovat).
  - Dialog končí, pokud tento přechod není definován.
- Dva druhy dialogů:
  - formuláře - definuje proces pro získání hodnot sady položek
  - menu - poskytuje uživateli sadu možností a odkazů na pokračování dialogu

Struktura VoiceXML aplikací

- Subdialogy
  - obdoba funkcí
  - slouží k opětovnému provádění jisté části dialogu a vrácení získaných hodnot.
- Sezení - začíná v okamžiku, kdy uživatel zahájí interakci se VoiceXML interpretrem a končí, když je ukončena buď uživatelem, VoiceXML dokumentem nebo kontextem dialogu.
- Aplikace - sada dokumentů, které sdílejí kořenový dokument

VoiceXML formulář

- Základní komponenta VoiceXML dokumentů.
- Obsahuje:
  - sadu položek
  - deklarace proměnných nepatřících položkám
  - ošetření událostí.
- Základní atribut - id
  - název formuláře
  - lze se pomocí něj na formulář odkazovat
  - musí být unikátní.
- Zpracování formuláře - FIA
  1. Výběr a přebrání jedné nebo více výzev.
  2. Získání uživatelských odpovědí, které naplní jednu nebo více položek a nebo vyvolání události (žádost o nápovědu).
  3. Zpracování sekcí *filled* u všech zadaných položek.

Ukázkový VoiceXML formulář

```
<vxml version="2.0" ...>
<form id="hello">
 <block name="hello">
  <prompt>Welcome to the VoiceXML!.</prompt>
 </block>
 <field name="greating">
  <prompt>Hello.</prompt>
  <grammar root="greating" src="greating.grxml"/>
  <noinput>
   <prompt>Tell mi something nice, like hello, hi, good day
  </noinput>
  <nomatch>
   <prompt>I didn't understand you, but thanks anyway.</pro
   <exit/>
  </nomatch>
```

Ukázkový formulář
Pokračování

```
  <noinput count="2">
   <prompt> When you don't want to speek to me good bye.</p
   <exit/>
  </noinput>
 </field>
 <filled>
  <prompt> you said <value expr="greating"/></prompt>
 </filled>
 </form>
</vxml>
```

Položky formuláře

- Vstupní položky
  - field
  - record
  - transfer
  - object
  - subdialog.
- Vstupním položkám odpovídají proměnné s názvem, který odpovídá hodnotě atributu name, příslušné vstupní položky.
- Řídící položky
  - block
  - initial.
- Provádění lze omezit pomocí atributu cond.

Element field

- Představuje vstup od uživatele.
- Atributy:
  - name - jméno pole
    - přístup k výsledné hodnotě pomocí stínové proměnné s tímto jménem.
  - expr - případná počáteční hodnota, lze použít výrazy jazyka ECMAScript
  - cond - podmínka nutná pro zpracování vstupu
  - vice viz specifikace (`http://www.w3.org/TR/2004/REC-voicexml20-20040316/#dml2.3.1`).

Element field
Dokončení

- Obsah:
  - případná výzva s popisem vstupu (element prompt)
  - gramatika - popisuje množinu akceptovatelných vstupů
  - ošetření událostí
    - noinput
    - nomatch
    - filled
    - ...
  - ...

Ukázka použití elementu field

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml">
 <form id="main">
  <field name="name">
   <prompt>Zadejte Vaše jméno</prompt>
   <grammar src="..." type="application/xml+srgs"/>
   <noinput>Zadejte prosím Vaše křestní jméno</noinput>
   <nomatch>Je mi líto, ale zadané jméno není v kalendáři</
  </field>
  <filled>
   <submit next="http://some.uri.cz/aplikace" namelist="nam
  </filled>
 </form>
</vxml>
```

Element record

- Umožňuje systému nahrát zprávu.
- Lze využít např. pro dialogový záznamník.
- Atributy:
  - name
  - expr
  - cond
  - beep - má-li se před začátkem nahrávání přehrát zvukový signál
  - maxtime - maximální délka nahrávky
  - type - mime-type výsledné nahrávky; musí být podporována VoiceXML platformou
  - ...
- Obsah:
  - případná výzva s popisem vstupu
  - ošetření událostí
    - noinput
    - connection.disconnect.hangup (použití elementu catch).

Ukázka použití elementu record

```
<?xml version="1.0" encoding="utf-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml">
 <form id="zaznamnik">
  <record name="zaznam" beep="true" maxtime="30s"
          type="audio/x-wav">
   <prompt> Bohužel zde nikdo není. Po zaznění signálu
    můžete zanechat vzkaz.  </prompt>
   <noinput> Bohužel nic neslyším. Zkuste to znovu.
   </noinput>
   <catch event="connection.disconnect.hangup">
    <submit next="http://some.uri.cz/zaznamnik"/>
   </catch>
  </record>
 </form>
</vxml>
```

Element subdialog

- Slouží k vyvolání dialogu, řešícího dílčí problém.
- Element subdialog.
- Jeden a tentýž subdialog lze volat opakovaně.
- Elementy:
  - subdialog - volání dílčího dialogu
  - param - definice hodnoty parametru
  - filled - kód, který se má provést po návratu z dílčího dialogu.
- Atributy
  - name - jméno volaného dílčího dialogu
  - src - URI dokumentu, který obsahuje kód dialogu.
- Kód subdialogu - formulář, ukončený elementem return.

Ukázka subdialogu

```
<?xml version="1.0" encoding="utf-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml">
<form id="demo">
 <subdialog name="greating" src="\#say_hello">
  <param name="param1" expr="'ahoj'"/>
  <filled>
   <prompt> Hodnota subdialogu je
     <value expr="greating.great"/></prompt>
  </filled>
 </subdialog>
 <filled>
  <prompt>Řekl jste <value expr="greating.great"/>
  </prompt>
 </filled>
</form>
```

Ukázka subdialogu

```
<form id="say_hello">
 <var name="param1"/>
 <field name="great">
  <prompt><value expr="param1"/></prompt>
  <grammar root="pozdrav" src="pozdrav.grxml"/>
  <noinput count="2">
   <prompt>
    Na pozdrav jste mi neodpověděl. Nashledanou.
   </prompt>
  <return/>
  </noinput>
```

Ukázka subdialogu
Dokončení

```
    <nomatch>
     <prompt>
      Bohužel jsem Vám nerozuměl, ale stejně dekuji.
      Nashledanou.
     </prompt>
     <return/>
    </nomatch>
   </field>
   <filled>
    <return namelist="great"/>
   </filled>
  </form>
</vxml>
```

Element block

- Obsahuje proveditelný obsah.
- Provádí se pokud:
  - má hodnotu 'undefined' (nebyl dosud navštíven)
  - atribut cond se vyhodnotí jako true.
- Struktura - viz předchozí příklady.
- Atributy:
  - name - jméno bloku
  - expr - iniciální hodnota proměnné formuláře
  - cond - podmínka omezující provádění bloku.

Element initial

- Typické využití - dialogové strategie se smíšenou iniciativou.
- Umožňuje uživateli zadat více informací naráz.
- Na rozdíl od ostatních vstupních elementů nemůže obsahovat:
  - gramatiku - využívá se gramatika formuláře (viz ukázka na dalším slidu)
    - pokud je uživatelova odpověď gramatikou formuláře rozpoznána, je nutno nastavit hodnotu elementu initial - vetšinou se používá true
  - filled.
- Dceřiné elementy:
  - výzvy (prompt)
  - ošetření událostí (catch, nomatch, noinput).

Ukázka použití elementu initial

```
<?xml version="1.0" encoding="utf-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml">
 <form id="main">
  <grammar src="registrace.grxml" type="application/srgs+xm
  <block>
   <prompt>
    Vítejte v registraci předmětů na FI
   </prompt>
  </block>
  <initial name="mixed">
   <prompt>
    Zde můžete zadat, jaké předměty s jakým ukončením
    si chcete zaregistrovat
   </prompt>
```

Ukázka použití elementu initial

Pokračování

```
 <noinput>
  Řekněte něco jako Předmět PB095 na zkoušku
 </noinput>
 <noinput count="2">
  Dobře zkusíme to postupně.
  <assign name="mixed" expr="true"/>
  <reprompt/>
 </noinput>
<nomatch>
 Můžete si zaregistrovat předměty PB095, PB125,
 PB162 s hodnocením zkouškou nebo zápočtem.
</nomatch>
<nomatch count="2">
 Dobře zkusíme to postupně.
 <assign name="mixed" expr="true"/>
 <reprompt/>
```

Ukázka použití elementu initial
Pokračování

```
 <nomatch count="2">
  Dobře zkusíme to postupně.
  <assign name="mixed" expr="true"/>
  <reprompt/>
 </nomatch>
</initial>
<field name="predmet">
 <grammar src="registrace.grxml\#predmet"/>
 <prompt>Zadejte kód předmětu</prompt>
 <nomatch>
  Zatím si lze zaregistrovat předměty PB162, PB095, PB125.
 </nomatch>
```

Ukázka použití elementu initial
Pokračování

```
 <nomatch count="3">
  Bohužel se nám zadávání nedaří. Nashledanou.
  <exit />
 </nomatch>
 <noinput count="3">
  Jelikož jste nic nezadal, tak se s Vámi loučím.
  <exit />
 </noinput>
</field>
<field name="ukonceni">
 <grammar src="registrace.grxml\#ukonceni"/>
 <prompt>Zadejte požadované ukončení.</prompt>
 <nomatch>
  Předměty lze ukončit zkouškou nebo zápočtem.
 </nomatch>
```

Ukázka použití elementu initial
Pokračování

```
 <nomatch count="3">
  Bohužel se Vám zadávání nedaří, zkuste to klasicky na
  adrese is.muni.cz. Nashledanou.
 </nomatch>
 <noinput>
  Zadejte, zda chcet předmět ukončit zkouškou nebo zápočte
 </noinput>
 <noinput count="3"> ...  </noinput>
</field>
```

Ukázka použití elementu initial
Dokončení

```
 <filled>
  <prompt>
    Provádím registraci předmětu s kódem
    <value expr="predmet"/>
    a ukončením <value expr="ukonceni"/>.
  </prompt>
 </filled>
 </form>
</vxml>
```

Další informace o VoiceXML

- Popis na W3C VoiceBrowser Activity (`http://www.w3.org/TR/2004/REC-voicexml20-20040316/`)
- www.voicexml.org (`http://www.voicexml.org`)
- Šimek, Richard - Tutoriál jazyka VoiceXML (bakalářská práce FI), 2005
- ...

SRGS

- W3C specifikace jazyka pro zápis bezkontextových gramatik pro podporu rozpoznávání řeči.
- Aktuální verze 1.0 (březen 2004).
- Nahradil původně používaný standard JSGF.
- Dvě varianty zápisu gramatiky:
  - XML
  - Augmented Backus-Naur From (ABNF).
- Liší se pouze zápis nikoliv vyjadřovací síla.
- Možnost použití způsobu zápisu závisí na použité platformě.
  - Vetší podpora pro XML formát.

Gramatiky (Teorie Formálních Jazyků)

- Gramatika $G = (N, \Sigma, P, S)$
  - N - množina neterminálních symbolů
  - Σ - abeceda
  - P - množina pravidel
  - S - kořenový neterminál.
- Bezkontextová gramatika
  - gramatika $G = (N, \Sigma, P, S)$
  - pravidla ve tvaru: $N \rightarrow \{N \cup \Sigma\}^*$.

XML Formát Gramatiky

- XML prolog.
- Kořenový element - grammar.
- Atributy:
  - root - pravidlo odpovídající kořenovému neterminálnímu symbolu
  - xml:lang - jazyk gramatiky
  - version - použitá verze SRGS (aktuálně pouze 1.0)
  - mode
    - dtmf
    - voice - implicitní hodnota
  - ...
- Element grammar - obsahuje množinu pravidel (elementů rule).

ABNF Gramatika

- ABNF hlavička
  - #ABNF *verze* [*kódování*]

    ```
    #ABNF 1.0 ISO-8859-2
    ```

  - root $ *jméno pravidla*; - kořenový neterminální symbol
  - language *jazyk*;
  - mode voice|dtmf;

    ```
    #ABNF 1.0 UTF-8 root $pozdrav;
    language cs-CZ;
    mode voice;
    ```

Zápis pravidla

- Levá strana pravidla:
  - XML formát
  - element rule
    - atribut id - jednoznačný identifikátor pravidla
    - obsah - pravá strana pravidla

      ```
      <rule id="pozdrav"> ahoj </rule>
      ```

  - ABNF
    - *<id pravidla>*

      ```
      $pozdrav = ahoj;
      ```

Zápis pravidla

- Pravá strana pravidla
  - může obsahovat terminální a neterminální symboly:
    - sekvenci
    - varianty
  - XML formát
    - tělo elementu rule

```
<rule id="vstup">
  Proved <ruleref uri="#prikazy"/>
  s parametry <ruleref uri="#parametry"/>.
</rule>
```

  - ABNF
    - $<neterminál> = <pravá strana>

```
$vstup =
Proveď $<http://www.nekde.cz/grammar.gram#prikazy>
s parametry $<http://www.nekde.cz/grammar.gram#parametry>
```

nebo

```
$vstup = Proved $prikazy s parametry $parametry
```

Sekvence

- Posloupnost terminálních a neterminálních symbolů.
  - $X \rightarrow YZa$
- Lze ji rozdělit na logické části.
- XML zápis:
  - zapsat přímo

```
<rule id="spojeni">
 Chci jet z <ruleref uri="#misto"/>.
</rule>
```

  - dělení na logické části
  - využitelnost
    - počet opakování dané části (atribut repeat
    - sémantická interpretace

- XML Formát:

```
<rule id="spojeni">
 Chci jet <item>z <ruleref uri="#misto"/> </item>
 <item> do <ruleref uri="#misto"/> </item>
 <item> <ruleref uri="#druh"/></item>
 <item> <ruleref uri="#datum"/></item>
 <item> v <ruleref uri="#cas"/></item>
</rule>
```

- ABNF zápis:

```
$spojeni = Chci jet z $misto
```

Varianty

- umožňují uživateli zadat jeden z možných vstupů
  - $X \rightarrow Y|Z|a$
- XML zápis:

```
<rule id="barvy">
 <one-of>
  <item>cervena</item>
  <item>zelena</item>
  <item>modra</item>
 </one-of>
</rule>
```

- ABNF zápis

```
$barvy = (cervena|zelena|modra)
```

Opakování

- Umožňuje specifikaci:
  - nepovinných částí promluvy
  - opakovaných částí promluvy
- XML zápis
  - pomocí atributu repeat u elementu item

```
<rule id="adresa">
 www
 <item repeat="1-2">
  tečka <ruleref uri="#castAdresy"/>
 </item>
 tečka <ruleref uri="#tld"/>
</rule>
```

Opakování

- ABNF zápis
  - za prvek uvedeme počet opakování uzavřený do <>

```
$adresa = www $castAdresy <1-2> $tld
          $castAdresy = tecka $text
```

- počet opakování
  - číslo - *číslo* krát
  - číslo1- číslo2 - *číslo1* - *číslo2* krát
    - číslo- - *číslo* - ∞ krát

Zvláštní pravidla

- GARBAGE - odpovídá libovolné promluvě až po následující blíže specifikovanou část
- VOID - pravidlo, které nelze vyslovit (zakázaní určité promluvy)
- NULL - pravidlo, které je vždy rozpoznáno (může být i prázdné)
- XML formát:
  - <ruleref special=”*pravidlo*” / >

```
<rule id="spojeni">
 <ruleref special="GARBAGE"/>
 z <ruleref uri="#misto"/> do <ruleref uri="#misto"/>
 <ruleref uri="#prostredek"/>
</rule>
```

- ABNF
  - $*pravidlo*

```
$spojeni = $GARBAGE z $misto do $misto $prostredek
```

Literatura

- Specifikace W3C.
- Příklady použité na přednášce.

SISR

- Sémantika - přiřazuje význam slovům
- Sémantika v dialogových systémech
  - přiřazuje význam promluvám a jejich částem
- SISR - standard W3C pro zpracování sémantiky promluvy.
  - aktuální verze 1.0
  - publikován - duben 2007
  - úzce spjat se standardy
    - ECMAScript
    - SRGS
- Umožňuje přiřazení základních interpretací částem promluvy a vytváření odvozených interpretací pro nadřazená tvrzen
  - přiřazení interpretace částem promluvy
  - odvozování interpretace na základech dílčích interpretací
  - přiřazení interpretace vstupním polím dialogu

Přiřazení interpretace části promluvy

- Sémantická interpretace bývá součástí pravidla SRGS.
- K pravidlu přiřazena pomocí elementu/atributu tag.
- XML formát SRGS gramatiky:
  - element tag

```
<item>
 <ruleref uri="souhlas"/><tag>{out='ano'}</tag>
</item>
```

  - atribut tag elementu item

```
<item tag="ano">jo</item>
```

- ABNF tvar:
  - uveden za interpretovanou část promluvy
  - tvar: interpretace

```
$souhlas = jo {ano}
```

Odvozování interpretace na základě dílčích interpretací

- Zápis pomocí výrazů v jazyce ECMAScript.
- Přiřazeno k pravidlům pomocí elementu tag.
- Interpretace reprezentována pomocí objektů jazyka ECMAScript.
- Stínové proměnné:
  - pro pravidla - objekt rules
  - výstup - objekt out

Odvozování interpretace na základě dílčích interpretací

XML formát SRGS gramatiky

```
<rule id="vlastnictvi">
 <item>Mám
   <item repeat="0-1">
      <ruleref uri="#barva"/>
   </item>
   <ruleref uri="prostredek"/>
  <tag>{out = rules.barva + ';' + rules.prostredek;}</tag>
 </item>
</rule>
```

Odvozování interpretace na základě dílčích interpretací

ABNF gramatika

```
$vlastnictvi = mam $barva <0-1> $prostredek
 {out = rules.barva + ';' + rules.prostredek;};
$barva = (cervenou {cervena}
      |cervene{cervena}
      |zelenou{zelena}
      |zelene{zelena});
$prostredek = (auto{auto} | kolobezku{kolobezka});
```

Přiřazení interpretace vstupním polím VoiceXML formuláře

- Využívají se atributy stínového objektu out
- XML formát:

```
<rule id="vlastnictvi">
 <item>
    Mám <item repeat="0-1"><ruleref uri="#barva"/>
    <ruleref uri="#prostredek"/>
    <tag>
     {
      out.barva = rules.barva;
      out.prostredek = rules.prostredek;
     }
    </tag>
  </item>
<rule>
```

Přiřazení interpretace vstupním polím VoiceXML formuláře

- ABNF gramatika:

```
$vlastnictvi = mam $barva <0-1> $prostredek
{
 out.barva = rules.barva;
 out.prostredek = rules.prostredek;
};
$barva = (cervenou {cervena}|
          cervene{cervena}|
          zelenou{zelena}|
          zelene{zelena});
$prostredek = (auto{auto} | kolobezku{kolobezka});
```

Literatura

- Specifikace SISR
- ECMAScript
- Příklady použité na přednášce.

Další standardy W3C VoiceBrowser Activity

- SSML
- Pronunciation Lexicon Specification
- Call Control XML
- State Chart XML

Speech Synthesis Markup Language

- Značkovací jazyk pro podporu syntetizované řeči ve webových aplikacích.
- Standard W3C
- Aktuální verze 1.0 (září 2004)
- Vychází z JSGF/JSML (JSpeech Markup Language)
- Cíle:
  - musí umožňovat konzistentní ovládání hlasového výstupu řečovým syntetizérem.
  - musí dovolovat TTS pro co nejširší škálu aplikací a domén
  - musí být internacionalizovaný
  - musí být snadno použitelný pro psaní dokumentů
  - musí být implementovatelný pomocí stávajících technologií
  - JSML dokumenty musí být lidsky čitelné.
- Zbytek viz syntéza řeči.

Pronunciation Lexicon Specification (PLS)

- Standard W3C
- Aktuální verze 1.0 (říjen 2008)
- Definuje značkování pro specifikaci slovníků výslovnosti pro podporu syntézy a rozpoznávání řeči.
- Specifikace W3C

Call Control XML

- Standard W3C
- Aktuální verze 1.0 (červenec 2011)
- navržen pro ovládání telefonních hovorů z dialogových systémů
- Specifikace W3C
- Umožňuje:
  - sestavení a ovládání konferenčních hovorů
  - přesměrování hovoru
  - ...

Pronunciation Lexicon Specification (PLS)

- Standard W3C
- Aktuální verze 1.0 (říjen 2008)
- Definuje značkování pro specifikaci slovníků výslovnosti pro podporu syntézy a rozpoznávání řeči.
- Specifikace na stránkách W3C Specifikace na stránkách W3C

Základní elementy PLS

- Kořenový element - lexicon
  - atributy - xmlns - specifikace jmenného prostoru (http://www.w3.org/2005/01/pronunciation-lexicon)
  - xml:lang - jazyk dokumentu
  - version - verze dokument (1.0)
  - alphabet - abeceda použitá pro fonetický přepis
- lexeme - obsahuje popis pro jednu lexikální jednotku (slovo, zkratku,...)
  - musí obsahovat aspoň jeden dceřiný element grapheme
- phoneme - obsahuje fonetický přepis dané lexikální jednotky (většinou se používá IPA).

Ukázka lexikonu

```
<?xml version="1.0" encoding="utf-8"?>
 <lexicon
    version ="1.0"
    xmlns="..."
    alphabet="ipa"
    xml:lang="en-US">
 <lexeme>
  <grapheme>color</grapheme>
  <phoneme>kaler</phonem>
 </lexeme>
</lexicon>
```

Ukázka propojení lexikonu a gramatiky

- XML formát SRGS

```
<grammar xmlns="..." xml:lang="en" version="1.0">
 <lexicon
   uri="http://www.example.com/lexicon.file"/>
 <lexicon
   uri="http://www.example.com/strange-city-names.file"
   type="media-type"/>
 ...
</grammar>
```

Ukázka propojení lexikonu a gramatiky

- ABNF formát SRGS

```
#ABNF V1.0 ISO-8859-1;
language en-US;
lexicon <http://www.example.com/lexicon.file>;
lexicon <http://www.example.com/strange-city-names.file>
        <media-type>;
...
```

Ukázka použití lexikonu v SSML

```
<speak version="1.1" xmlns="..." xml:lang="en-US">
 <lexicon uri="lexicon.pls" xml:id="pls"/>
 <lexicon uri="strange-words.file" xml:id="sw"
          type="media-type"/>
 <lookup ref="pls"> tokens here are looked up in
  lexicon.pls
  <lookup ref="sw"> tokens here are looked up first in
     strange-words.file and then, if not found, in
     lexicon.pls
  </lookup>
  tokens here are looked up in lexicon.pls
 </lookup>
 tokens here are not looked up in lexicon documents ...
</speak>
```

Metoda Wizard of Oz

- Simulace dialogového rozhraní modelem člověk – člověk.
- Založena na principu popsaném v knize The Wonderful Wizard of Oz (Lyman Frank Baum)
- Princip:
  - Funkce dialogového rozhraní je (skrytě) simulována člověkem.
  - Průběh dialogu je protokolován.
  - Průběh se řídí navrženou dialogovou strategií.
    - Pokud je dostupný prototyp může Wizard pouze modifikovat a předávat komunikaci mezi uživatelem a systémem.
- Občas snaha navodit zdání, že uživatel komunikuje s dialogovým systémem – využívají se různé prostředky:
  - vzdálená komunikace kde osoba simulující dialogové rozhraní komunikuje prostřednictvím TTS
  - použití vocodérů, které změní hlas osoby, která provádí testování, aby zněl jako výstup TTS
  - ...

Tvorba dialogového rozhraní pomocí metody WoZ

- Z korpusu dialogů na dané téma (pro danou doménu) lze vygenerovat dialogové rozhraní následovně:
  1. Vytvoříme iniciální korpus metodou WoZ
     - Komunikace pouze čaroděj – uživatel.
  2. Odstraní se konflikty a na základě korpusu se vytvoří dialogové rozhraní.
  3. Kombinovaně vytvoříme nový korpus.
     - "Čaroděj" se snaží maximálně využívat navržené dialogové rozhraní..
  4. Odstranění konfliktů a vygenerování nové verze dialogového rozhraní.
  5. Pokud je rozhraní v pořádku, generování končí, jinak se pokračuje krokem 3.

Multimodální dialogová rozhraní

- Mimo mluvenou řeč umožňuje alternativní způsoby komunikace člověk – počítač:
  - textová komunikace
  - grafická komunikace
  - ...
- Výhoda - lepší přístupnost.
  - uživatelé s poruchami sluchu,
  - uživatelé s poruchami řeči,
  - ...

Multimodální komunikace počítač – člověk

- Textová:
  - Mimo hlasový výstup je navíc zobrazen i odpovídající textový výstup.
  - Lze využít prostředky pro IM, SMS, ...
- Grafická:
  - Talking Heads – mimo hlasový výstup je navíc zobrazena tvář (hlava, celý člověk, ...), jejíž pohyby, zejména úst, odpovídají mluvené řeči.
  - Komunikace znakovou řečí – mluvené slovo je překládáno na znakovou řeč (viz Guimeraes et al. – Structure of the Brazilian Sign Language (Libras) for Computational Tools: Citizenship and Social, in Organizational, Busines, and Technological Aspects of the Knowledge Society, CCIS vol. 112, Springer, Heidelberg, 2010, pp. 365 – 370.)
    - Znaková řeč prezentována pomocí rukou nebo avatara.

Multimodální komunikace člověk – počítač

- Široké spektrum možností zadávání vstupu uživatelem jinak než hlasem:
  - klávesnice (počítač, DTMF, SMS, ...)
  - rukou psaný vstup – dotyková obrazovka + pero, ...
  - ústy ovládaná zařízení
  - ovládání pomocí pohybů očí a víček
  - rozpoznávání řeči pomocí sond detekujících činnost svalů a mozku (viz Schultz, T. – Silent and Weak Speech Based on Elektromyography, in Proceedings of 12th Internation Conference ICCHP 2010 Part 1, Wien, Springer, Heidelberg, pp. 595 – 604, 2010.)
  - rozpoznávání znakové řeči
  - ...
- Často jako doplněk řečového vstupu.

Existující nástroje a standardy

- Proprietární řešení:
  - Součást CSLU Toolkitu.
  - Projekt August.
- Otevřená řešení:
  - Návrhy doporučení W3C týkající se multimodálního přístupu – zatím bez implementace.
    - Využívají a propojují i další standardy W3C (CCXML, XHTML, VoiceXML, SVG, SMIL, ...).
  - Výstup W3C Multimodal Interaction WG

Ukázka multimodální syntetizované řeči

Demonstrační video se syntetizovanou multimodální řečí.

Co jsou to emoce?

- "This is a very tough question, that has produced significant amounts of headaches to scientists in the past ...", "... many researchers have to opted to study systematically phenomena that most consider emotional." (Laval University of Quebec)
- "Only mathematics is certain, so all must be based on mathematics." (R. Descartes)
- Dělení emocí:
  - Primární (základní) – vyskytují se u všech lidí a u části vyšších živočichů.
  - Sekundární (vyšší) – mohou být intelektuální, morální a estetické. Mohou se lišit mezi jednotlivými kulturami.
- Velkých šest:
  - hněv
  - zklamání
  -

Základní emoce

- Velkých šest (R. Descartes):
  - hněv
  - zklamání
  - smutek
  - strach
  - překvapení
- Další autoři:
  - Arnold – hněv, averze, odvaha, sklíčenost, touha, zoufalství, strach, nenávist, láska, smutek.
  - Ekman, Friesen, Ellsworth – hněv, odpor strach, radost, smutek, překvapení.
  - Frijda — touha, štěstí, zájem, překvapení, údiv, zármutek,
  - ...

Detekce emocí

- Lze provádět pomocí detekce změn různých biometrických vlastností.
  - Změny galvanických vlastností kůže.

  - Změny tlaku krve a pulsu.

Detekce emocí

- Použitelné biometrické charakteristiky:
  - změny dýchání

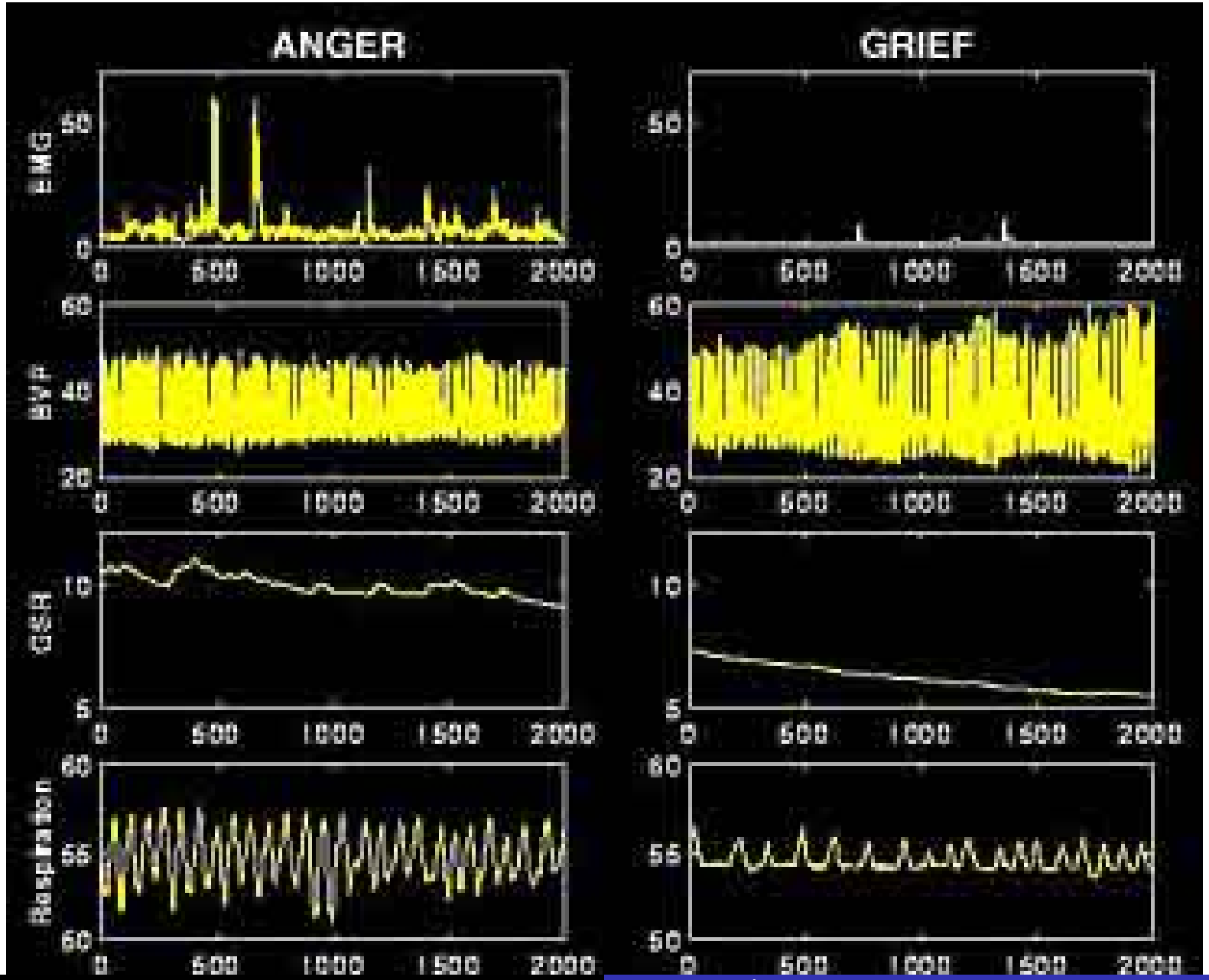
Ukázky z Yale Face Database

- Radost

Ukázky z Yale Face Database

- Ospalost

Ukázky průběhů charakteristik pro smutek a hněv



Využití emocí v dialogovém rozhraní

- Dialogová rozhraní informačních systémů
  - uzpůsobení dialogové strategie emočnímu stavu uživatele (klid, stres, hněv, ...)
  - přepojení uživatele na lidského operátora.
- Výukové DS:
  - uzpůsobení dialogové strategie koncentraci uživatele.
- ...