

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

# Cíl a náplň předmětu

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Cíl – seznámení s oblastí dialogových systémů.
- Obsah kurzu:
  - Úvod do dialogových systémů, historie
  - Základní technologie:
    - přenos hlasu prostřednictvím počítačové sítě
    - digitální zpracování zvuku
    - rozpoznávání řeči
    - syntéza řeči

# Cíl a náplň předmětu

## pokračování

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Obsah kurzu:
  - Dialogové systémy:
    - formální modely dialogu
    - analýza dialogu, kooperativní a nekooperativní dialog
    - dialogové strategie
    - information retrieval DS
    - simulace DS
    - multimodalita
    - nástroje pro tvorbu dialogových systémů – W3C VoiceBrowser Activity
    - aplikace.

# Ukončení předmětu

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Možná ukončení:
  - zkouška
  - kolokvium
  - zápočet
- Požadavky:
  - zkouška + kolokvium – dobrá orientace v probírané problematice
    - ústní forma
    - nástin řešení problémové úlohy a detailnější rozbor některé z použitých technologií.
  - zápočet – schopnost návrhu řešení daného problému z oblasti dialogových systémů.

# Doporučená literatura

## Knihy

### Dialogové systémy

Luděk Bártek

### O předmětu

Úvod do dialogových systémů

Stručná historie zpracování zvuku

Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči  
Vnímání zvuku

- J. Psutka, Komunikace s počítačem mluvenou řečí, Academia, Praha, 1995
- Z. Kotek, V. Mařík, Metody rozpoznávání a jejich aplikace, Academia, Praha, 1993
- T. Dutoit, An Introduction to Text-to-Speech Synthesis, Kluwer Academic Publishing, 1996
- A. Kobsa, W. Wahlster, User Models in Dialog System, Springer 1989
- D. B. Roe, J. G. Wilpon (editors), Voice Communication Between Humans and Machines, National Academy Press, Washington D.C., 1994
- F. Jelinek, Statistical Methods for Speech Recognition, MIT Press 1997

- Via Voice
- Odkazy na dialogové systémy (DS)
- Různé projekty z oblasti počítačové sémantiky na Stanford University
- Stránky W3C VoiceBrowser Activity
- ...

# Co je dialogový systém?

## Dialogové systémy

Luděk Bártek

O předmětu

Úvod do dialogových systémů

Stručná historie zpracování zvuku

Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči

Vnímání zvuku

- Dialogový systém – systém komunikující s uživatelem pomocí dialogu v přirozeném jazyce
  - většinou se jedná o dialogové rozhraní ke klasickému IS.
- Častá komunikace přirozenou řečí.
- Alternativně:
  - komunikace pomocí DTMF
  - textová komunikace přirozenou řečí
  - multimodální komunikace:
    - řeč + obraz (simulace lidské tváře, titulky ve znakové řeči, ...)
    - řeč + text
    - ...

# Výhody a nevýhody dialogových systémů

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

## ■ Výhody:

- + Přirozenější způsob komunikace.
- + Přístupnost:
  - zrakově a motoricky postižení uživatelé
  - další uživatelé, kterým činí problémy práce ovládání počítače
  - možnost podrobnějšího vedení uživatele krok za krokem celým procesem
  - dalším krokem k lepší přístupnosti – multimodální rozhraní
  - ...



# Výhody a nevýhody dialogových systémů

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

## ■ Výhody:

+ Větší množství potenciálních uživatelů:

- počet uživatelů počítačů a Internetu vs počet uživatelů telefonu.

+ ...

## ■ Nevýhody:

- rychlost komunikace

- sekvenční vnímání zvuku vs. paralelní vnímání obrazu
- lze částečně eliminovat pomocí vhodné dialogové strategie

- Laboratoře:
  - LSD – doc. Kopeček, prof. Zezula
  - zaměřuje se na:
    - vyhledávání
    - dialogové systémy a zpracování zvuku
  - NLP – doc. Pala
  - zaměřuje se na:
    - korpusy
    - slovníky
    - morfologii
    - syntaktickou analýzu
    - sémantiku

# Aktuální práce v oblastech souvisejících s dialogovými systémy

Výzkum Česká republika

Dialogové systémy

Luděk Bártek

O předmětu

Úvod do dialogových systémů

Stručná historie zpracování zvuku

Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči  
Vnímání zvuku

- FIT VUT Brno
  - analýza signálu
  - rozpoznávání řeči
  - systém pro automatizované zpracování konferencí
  - ...
- ZČU v Plzni
  - rozpoznávání řeči
  - dialogové systémy
  - ...
- ČVUT – syntéza řeči

# Aktuální práce v oblastech souvisejících s dialogovými systémy

Komerční sféra - Česká republika

Dialogové systémy

Luděk Bártek

O předmětu

Úvod do dialogových systémů

Stručná historie zpracování zvuku

Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči  
Vnímání zvuku

- FROG Systems s.r.o. – CS-voice 97
- OptimSys s.r.o – VoiceXML platforma OptimTalk
- ...

# Aktuální práce v oblasti dialogových systémů

## Výzkum a práce ve světě

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- W3C VoiceBrowser Working Group
  - IBM
  - Nuance Communication
  - Lucent Technologies
  - Motorola
  - ScanSoft
  - Tellme Networks
  - Vocalocity
  - ...
- MIT
- OGI
- EPF Lausane
- ...

# European Masters in Language and Speech

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Evropské navazující studium v oblastech zpracování řeči a přirozeného jazyka (na FI během magisterského studia).
- Zapojeny VŠ např. v Dánsku, Řecku, Španělsku, Belgii, Německu, Velké Británii, Nizozemí, ...
- Více informací:
  - Stránka o EuroMasters na FI
  - doc. Pala, doc. Kopeček.

# Struktura dialogového systému

Dialogové systémy

Luděk Bártek

O předmětu

Úvod do dialogových systémů

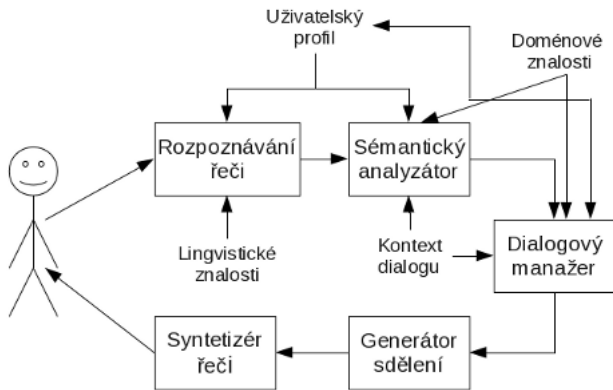
Stručná historie zpracování zvuku

Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči  
Vnímání zvuku



# Komponenty dialogového systému

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Uživatel – koncové zařízení, které uživateli umožňuje komunikovat s dialogovým systémem:
  - telefon – komunikace prostřednictvím PSTN přes VoIP gateway – VoIP gateway převádí hlas na data a zpět
  - VoIP klient – komunikace prostřednictvím VoIP protokolu přímo s dialogovým systémem (SIP, H323, Skype, ...)
  - textový klient – komunikace prostřednictvím protokolů DTMF+VoIP protokol, telnet, ssh, XMPP, ...
- Rozpoznávání řeči:
  - převádí mluvené slovo na text
  - využívá se:
    - rozpoznávání plynulé řeči
    - rozpoznávání izolovaných slov
  - pro zvýšení úspěšnosti se používají gramatiky popisující množinu očekávaných vstupů.



# Komponenty dialogového systému

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Sémantický analyzátor
  - získává relevantní údaje z rozpoznaného textu
  - využívají se např. atributové gramatiky.
- Dialogový manažer
  - konečný automat
  - na základě aktuálního stavu a vstupu od uživatele rozhoduje o dalším průběhu dialogu.
- Generátor promluv – na základě údajů od dialogového manažeru generuje promluvy, které jsou následně syntetizovány.
- Řečový syntetizér – převádí promluvy od generátoru promluv na mluvenou řeč, která je poslána uživateli.

# Údaje používané dialogovým systémem

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Lingvistické znalosti – údaje o jazyce, které využívá rozpoznávač řeči pro zvýšení úspěšnosti (pravděpodobnosti výskytů jednotlivých sekvencí řečových segmentů, gramatika, ...).
- Uživatelský profil – informace o uživatelích (charakteristiky hlasu, vyjadřování, používané fráze, ...).
- Doménové znalosti – informace odvoditelné z oblasti dialogového systému (gramatika, ...).
- Kontext dialogu – informace o aktuálním stavu dialogu (krok dialogu, uživatelský vstup, chybovost uživatele, ...).

# Historie zpracování a napodobování řeči

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- před 3 milióny let – Australopitekus – schopnost artikulované řeči
- starověk – budování mluvících soch bůžků
- 1779 – Kratzenstein – systém rezonátorů pro napodobení samohlásek a, e, i, o, u.



# Historie zpracování a napodobování řeči

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- 1791 – Wolfgang von Kempelen – mechanický mluvící stroj



- 1835 – zrekonstruován a upraven Wheatstonem v Dublinu – měl navíc pružnou „ústní dutinu“
- 1846 – J. Faber – mluvící stroj Euphonia

# Historie zpracování a napodobování řeči

Dialogové systémy

Luděk Bártek

O předmětu

Úvod do dialogových systémů

Stručná historie zpracování zvuku

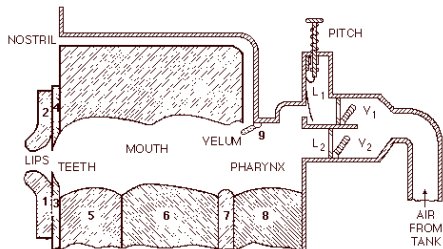
Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči  
Vnímání zvuku

- 1937 – R. R. Riesz – mechanický mluvící stroj napodobující lidské řečové ústrojí



- 1939 – H. Dudley – VODER (elektromechanický řečový syntetizér), VOCODER (elektrické zařízení kódování a přenos řeči)
- 50. léta 20. století – syntéza ve frekvenční oblasti, později syntéza v časové oblasti
- cca 1970 – počítače
- 1966 – J. Weizenbaum – Eliza (Communications of the

# Základy moderní analýzy řeči

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- 19. století
  - J.B. Fourier – Fourierova věta – využívá se při spektrální analýze zvuku
  - H. Helmholtz – zabýval se fyziologií vnímání hudby, Helmholtzův rezonátor
  - J. R. Ewald – fyziologie sluchu
- 1924 – spektrální analýza řeči na bázi formantové analýzy samohlásek
- 1939 – vokodér – zařízení pro kompresi řeči pro účely přenosu hlasu rádiem a transkontinentálním kabelem
- 1946 - 1947 – zařízení pro grafický záznam řeči
- 2. polovina 20. století – intenzivní vývoj jak teorie, tak počítačových aplikací.

# Základní řečové technologie

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- syntéza řeči
- rozpoznávání řeči
- související oblasti – zpracování jak v časové tak frekvenční oblasti
- rozpoznávání řečníka
- detekce emocí
- word spotting
- ...

- Zvuk
  - kmitavý pohyb molekul prostředí (vzduchu)
  - vyvoláván pružným odporem prostředí
- Kmit hmotného bodu
  - pohyb bodu z rovnovážné polohy do místa s maximální výchylkou (amplitudou), odtud do protilehlého místa s maximální výchylkou zpět do rovnovážného bodu.



# Kmity

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

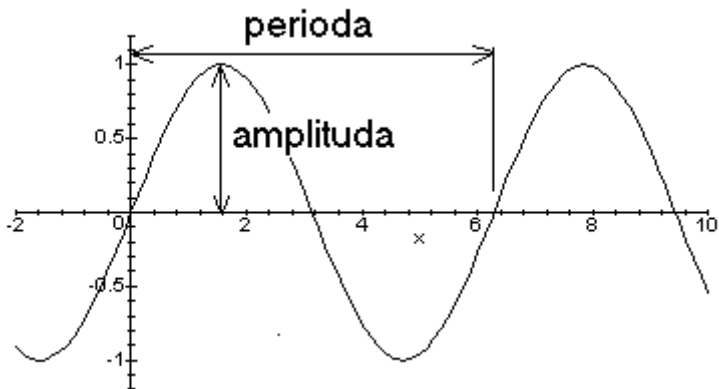
Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku



- Amplituda – maximální výchylka kmitavého pohybu.
- Perioda ( $T$ )
  - doba jednoho opakování periodického děje.
  - jednotka – 1 s (sekunda).
- Frekvence ( $f$ )
  - počet opakování periodického děje za jednotku času.
  - platí  $f = \frac{1}{T}$
  - jednotka 1 Hz (Hertz).

- Síla působící na kmitající bod:
  - $F = -ks$ ,  $k$  – tuhost pružiny,  $s$  – aktuální výchylka pružiny
  - $ma = -ks$ ,  $m$  – hmotnost tělesa,  $a$  – zrychlení
  - $a + \omega^2 s = 0$  ( $\omega^2 = \frac{k}{m}$ ,  $\omega$  – úhlová rychlost kmitavého pohybu:  $\omega = \frac{2\pi}{T}$ )
- fáze kmitavého pohybu:  $\psi = \omega t$
- okamžitá výchylka:  $y = y_m \sin \omega t = y_m \sin \psi$
- okamžitá rychlost:  $v = \omega y_m \sin \omega t = y_m \sin \psi$
- okamžité zrychlení:  $a = -\omega y_m \sin \omega t = y_m \sin \psi$

# Harmonické versus tlumené versus vynucené kmitání

Dialogové systémy

Luděk Bártek

O předmětu

Úvod do dialogových systémů

Stručná historie zpracování zvuku

Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči  
Vnímání zvuku

- Harmonické kmitání
  - na těleso nepůsobí žádná vnější síla
  - v praxi se s ním téměř nesetkáme (odpor vzduchu, ...).
- Tlumené kmitání
  - proti pohybu působí odpor prostředí
  - amplituda s časem (vzdáleností od zdroje) klesá
- Vynucené kmitání, rezonance
  - na hmotný bod působí navíc periodicky proměnné síla  $G = \sin\alpha t$ 
    - $F = ma = -ky + \sin\alpha t \Rightarrow a + \omega^2 y = \sin\alpha t$
    - partikulární řešení:  $\frac{\sin\alpha t}{\omega^2 - \alpha^2}$

- Zvuk – mechanické vlnění pružného prostředí (vzduch, voda, kov, ...)
- Akustika – věda studující zvuk (z řeckého akustikos – vztahující se k slyšení):
  - fyzikální – zvuk jako fyzikální vlnění
  - fyziologická akustika – vzniká a vnímání zvuku člověkem
  - hudební – zvuky z pohledu hudby
  - molekulární – vztah akustických vlastností a molekulární struktury.
- Rozdělení zvuku:
  - infrazvuk – frekvence  $< 16$  Hz
  - slyšitelný zvuk – 16 Hz – 16kHz
  - ultrazvuk –  $> 16$  kHz
  - hyperzvuk – až  $10^8$  Hz – využíván např. molekulární akustikou.

# Jednoduchý vs. složený tón

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

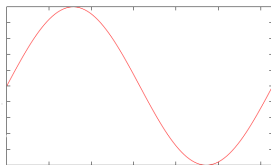
Fyzikální  
akustika

Fyziologická  
akustika

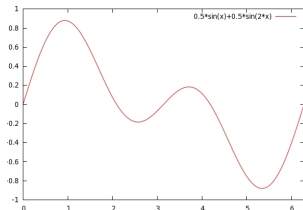
Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Základní tón – průběh intenzity v čase lze popsat jednoduchou sinusoidou.



- Složený tón – lineární kombinace jednoduchých tónů.



# Akustické spektrum zvuku

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Akustické spektrum – množina základních tónů, ze kterých je zvuk složen.

- Získání spektra – Fourierova transformace:

- $F(x)$  – periodická funkce s periodou  $T$ :

$$s_n(x) = \frac{a_0}{2} + \sum_{i=1}^n a_i \cos(i\omega x) + b_i \sin(i\omega x)$$

- $\omega = \frac{2\pi}{T}$

- aproximace  $F(x)$  je nejlepší při použití hodnot koeficientů  $a$  a  $b$ :

$$a_k = \frac{2}{T} \int_{-\infty}^{\infty} f(x) \cos(kx) dx$$

$$b_k = \frac{2}{T} \int_{-\infty}^{\infty} f(x) \sin(kx) dx$$

# Akustické spektrum zvuku

## pokračování

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- hodnoty spektra – zjistit v Psutkovi!!!
- Problém – zvuk je periodický pouze na určitých intervalech.
  - analýza na krátkém intervalu, kde se předpokládá, že je periodický.
- Z hlediska fyziologické akustiky – spektrum odpovídá rezonanci odpovídajících vláček Cortiho ústrojí, resp. odpovídající reakci neuronů.



## ■ Akustický tlak

- Odpovídá síle působící na element plochy v prostředí akustického vlnění.
- Pro sinusovou vlnu platí:

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

- $p_0$  – maximální akustický tlak v průběhu periody
- $\omega$  – úhlová rychlost
- $t$  – čas.

# Akustická intenzita a akustický tlak

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

## ■ Akustická intenzita

- Vyjadřuje množství akustické energie, které projde jednotkovou plochou za jednotku času.
- Je přímo úměrná druhé mocnině akustickému tlaku.
- Rozsah intenzity zvuku - dán rozsahem minimální ( $I_0$ ) a maximální ( $I_1$ ) akustické intenzity, kdy jsme schopni vnímat tón o frekvenci 1 kHz.
- Práh citlivosti –  $p_0 = 2 * 10^{-2} Nm^{-2}$ .
- Práh bolestivosti –  $p_1 = 10^2 Nm^{-2}$ .
- Rozsah –  $2,5 * 10^{13} Nm^{-2}$ .

- Weber-Fechnerův psychofyzikální zákon
  - Člověkem subjektivně vnímaná hlasitost roste při geometrickém nárůstu intenzity přibližně lineárně.
  - Pro stanovení hladiny intenzity zvuku ( $L$ ) volíme

$$L = 10 * \log \frac{I}{I_0}$$

- jednotka – 1 bel (originál bell) [B]
- Prakticky se využívá odvozená jednotka decibel [dB] ( $10^{-1}B$ ).

# Orientační hodnoty akustické intenzity

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- šepot – 10 - 20 dB
- tlumený hovor – 35 - 45 dB
- symfonický orchestr – 70 - 90 dB
- rocková hudba – 110 - 130 dB.

# Základy fyziologické akustiky

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Fyziologická akustika se zabývá:
  - mechanismem vytváření řeči
  - mechanismem vnímání řeči.
- Využívá Helmholtzovu rezonanční teorii.

# Helmholtzův rezonátor

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

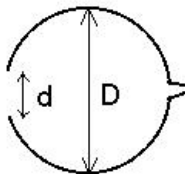
Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku



## ■ Princip činnosti:

- Přivedením vzduchu do rezonátoru v něm vznikne přetlak.
- Ten vytlačuje přebytečný vzduch ven a následně vzniká podtlak, který způsobí nasávání vzduchu z okolí.
- Takto vzniká periodický děj:

$$f = \frac{75,3}{D} \sqrt{\frac{d}{D}} [\text{Hz}]$$

# Mechanismus vytváření řeči

## Dialogové systémy

Luděk Bártek

O předmětu

Úvod do dialogových systémů

Stručná historie zpracování zvuku

Fyzikální akustika

Fyziologická akustika

Základy fyziologické akustiky

Mechanismus vytváření řeči

Vnímání zvuku

- Řeč vzniká pomocí *hlasového ústrojí* (umístěno v *hrtanu*).
- Hlasivky vytváří úzkou hlasovou štěrbinu a jsou rozechvívány procházejícím vzduchem.
- Frekvence jejich kmitání určuje *základní hlasivkový tón*.
- Zvuk, který vzniká v hrtanu pomocí hlasivek (samohlásky, znělé souhlásky) je modifikován v *rezonančních dutinách*:
  - hrtanové
  - ústní
  - nosohltanové.
- Rezonanční dutiny fungují na stejném principu jako Helmholtzův rezonátor).

# Hlasivky a schéma lidského hlasového ústrojí

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

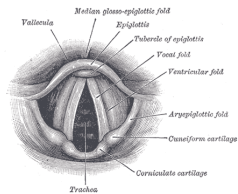
Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

## ■ Hlasivky



## ■ Jejich umístění





# Mechanismus vnímání řeči

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
Vnímání zvuku

- Zvuk vnímáme sluchovým orgánem.
- Sluchový orgán:
  - vnější ucho – zachycuje, soustřeďuje a přivádí zvukové vlny ke střednímu uchu
  - střední ucho
    - mechanickou cestou přenáší zvukovou energii mezi vnějším a vnitřním uchem
    - obsahuje mechanismy k vyrovnání rozdílů tlaku mezi vnějším prostředím a sluchovým orgánem
  - vnitřní ucho – převádí zvukovou energii na vzruchy, které jsou vedeny dále do mozku.

# Schéma sluchového orgánu

Dialogové  
systémy

Luděk Bártek

O předmětu

Úvod do  
dialogových  
systémů

Stručná  
historie  
zpracování  
zvuku

Fyzikální  
akustika

Fyziologická  
akustika

Základy  
fyziologické  
akustiky

Mechanismus  
vytváření řeči  
**Vnímání zvuku**



Obrázek: Schéma sluchového orgánu

- **Obsahuje:**
  - Ušní boltec – soustředí zvukové vlny do zvukovodu.
  - Zvukovod – vede zachycenou zvukovou energii (vlny) k bubínku.
  - Bubínek:
    - Tenká blána na konci zvukovodu – síla cca 0.1 mm.
    - Zesílí a přeneše zvukovou energii na kůstky středního ucha.

- **Obsahuje:**
  - **Kůstky středního ucha:**
    - kladívko – přiléhá k bubínku
    - kovádlinka
    - třmínek – přiléhá k oválnému okénku, kterým se zvuková energie předává do vnitřního ucha.
  - **Oválné okénko** – tvoří přístup k vnitřnímu uchu.
  - **Eustachova trubice:**
    - Vede ze středního ucha do nosohltanu.
    - Slouží k vyrovnání rozdílu tlaku mezi vnějším prostředím a středním uchem, aby nedošlo poškození sluchu.

- Hlemýžď' (Cochlea):
  - Je naplněn vodnatým roztokem.
  - Ústrojí ve tvaru ulity hlemýždě, které obsahuje Cortiho ústrojí.
  - Cortiho ústrojí obsahuje zhruba 20000 vláček s délkami  $40 \mu\text{m}$  — 0,5 mm.
  - Vláčekna jsou jsou napojena na nervová zakončení, která vedou vzruchy do příslušného centra v mozku.
- Rovnovážný orgán.

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

- Zkoumá zvukovou stránku jazyka z různých aspektů.
- Základní pojmy, které souvisejí s dialogovými systémy:
  - foném
    - samohlásky – formanty
    - souhlásky – znělost/neznělost souhlásek
  - koartikulace
  - spodoba znělosti

# Fonémy a fonetická transkripce

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Foném – elementární zvukový segment, který je vymezen na základě své schopnosti diferencovat vyšší, znakové jednotky jazykového systému (morfémy).
- Fonetická transkripce (přepis) – převod psaného textu do odpovídající fonetické podoby:

na shledanou → na zhledanou | na schledanou

- Fonetická abeceda – slouží k zápisu fonetického přepisu
  - Mezinárodní fonetická abeceda (IPA) – součástí standardu UNICODE
  - Řečové vyhodnocení metod fonetické abecedy (SAMPa) – sedmibitový přepis fonetické abecedy, využívá se při automatizovaném zpracování (např. řečový syntetizér MBrola, ... )..



- Samohláska – samostatně tvoří slabiku
- Rozdělení samohlásek:
  - krátké: a, e, i, o, u
  - dlouhé: á, é, í, ó, ú
  - dvojhlásky: eu, au, ou
- Obsahují:
  - základní hlasivkový tón – frekvence kmitání hlasivek (100 — 400 Hz)
  - formanty – frekvence vzniklé a zesílené rezonancí v hlasových dutinách.

- Frekvence vzniklé a zesílené rezonancí v hlasových dutinách
  - F1 – vzniká rezonancí v dutině ústní.
  - F2 – vzniká rezonancí v dutině hrdelní.
- Existují i vyšší formanty (F3, ...) – výskyt je často individuální.
- Výskyt a intenzita formantů se může lišit v závislosti na:
  - pohlaví – muž/žena
  - věku – dětství/dospívání/dospělost/seniorský věk
  - zdravotním stavu – např. nachlazení, ochraptělost, nemoci hlasivek a hrtanu, ...
  - ...

# Formanty F1 a F2 pro české samohlásky

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

Samohláska	Formant F1	Formant F2
a	700 — 1100 Hz	1100 — 1500 Hz
e	500 — 700 Hz	1500 — 2000 Hz
i	300 — 500 Hz	2000 — 3000 Hz
o	500 — 700 Hz	900 — 1200 Hz
u	300 — 500 Hz	600 — 1000 Hz

Tabulka: Formanty F1 a F2 u samohlásek

# Četnost výskytu samohlásek

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

Samohláska(y)	Relativní četnost
[e]	10 %
[a], [o], [i]	6 — 7 %
[í]	4 %
[á], [u], [é], [ou], [ú]	< 4 %
[ó], [au], [eu]	pouze nepatrná frekvence

- Na rozdíl od samohlásek jsou souhlásky dynamické děje.
- Silně závisí na kontextu, ve kterém se nacházejí.
- Tónový charakter mají pouze části některých souhlásek:
- Dělí se podle:
  - znělé – vznikají v hltanu, obsahují základní hlasivkový tón.
  - neznělé – vznikají v řečových dutinách (nosohltanové, ústní, ...), mohou mít charakter šumu (např. sykavky):
    - problematická detekce začátku promluvy při zašuměném zdroji.
  - Znělé a neznělé samohlásky se mohou vyskytovat v párech (párové souhlásky) např.:
    - r/l
    - b/p
    - d/t
    - ...

- Kroky digitalizace zvuku:
  - 1 vzorkování – snímání aktuální hodnoty signálu s danou frekvencí (vzorkovací frekvence)
  - 2 kvantizace – převod reálných hodnot na celočíselné
  - 3 kódování průběhu vlny – způsob ukládání informací o průběhu zvuku.

- Snímání aktuální hodnoty signálu s danou frekvencí – vzorkovací frekvence.
- Vzorkovací frekvence – měla by být minimálně dvojnásobkem nejvyšší frekvence, která je v signálu přítomna, aby bylo možné původní signál bez ztráty informace zrekonstruovat (Shannonův vzorkovací teorém).
- Získané hodnoty musí být následně kvantizovány a vhodným způsobem uloženy.
- Nejpoužívanější vzorkovací frekvence:
  - 8 kHz – telefonní kvalita
  - 16 kHz
  - 22050 Hz – rozhlasová kvalita
  - 44100 Hz – CD kvalita
  - 48 kHz – DVD kvalita

- Metoda převodu spojitych hodnot na diskretní.
- Princip:
  - Pokud hodnota signálu překročí  $n$ . násobek kvantizačního kroku je jí přiřazena hodnota  $n$ .
  - kvantizační krok = rozsah hodnot měřené veličiny/počet diskretních hodnot
  - kvantizační chyba – zaokrouhlovací chyba způsobená velikostí kvantizačního kroku, přímo úměrná velikosti kvantizačního kroku.
- Běžně používané kvantizace:
  - zpracování zvuku:
    - $2^8$
    - $2^{16}$
    - $2^{24}$
  - zpracování obrazu, ... navíc
    - $2^{32}$



# Způsoby kódování průběhu vlny

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Přímé ukládání hodnot získaných kvantizací – kódování PCM (Pulse-Code Modulation).
  - relativně pomalé změny průběhu zvukového signálu – malé rozdíly mezi sousedními vzorky.
  - Velká redundance dat.
  - Problém v případě příliš velkého rozptylu amplitud v signálu – příliš velký kvantizační krok – příliš velká kvantizační chyba, příliš malý kvantizační krok – přetečení v okamžiku zvětšení amplitudy signálu.
- Diferenční PCM – ukládá se rozdíl mezi sousedními vzorky
- Adaptivní diferenční PCM — diferenční PCM s proměnou velikostí kvantizačního kroku.

# Diferenční pulsní kódová modulace

Dialogové systémy

Luděk Bártek

Základy fonetiky

Úvod do počítačového zpracování zvuku

Komunikace uživatel – dialogový systém

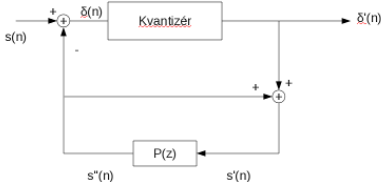
VoIP  
SIP

Zpracování digitalizovaného signálu

Analýza digitalizovaného signálu v časové oblasti

Analýza digitalizovaného signálu ve frekvenční oblasti

- Vychází z předpokladů:
  - Rozdíl dvou po sobě jdoucích vzorků je podstatně menší hodnota než hodnota vzorku.
  - Následující vzorek lze poměrně přesně odhadnout jako lineární kombinaci předchozích vzorků.
- Blokové schéma kódování signálu pomocí DPCM



- $s''(n)$  – odhad hodnoty řečového vzorku
- $s'(n)$  – rekonstruovaný signál, získaný jako součet kvantizovaného signálu  $\delta'(n)$  a  $s''(n)$
- $\delta(n) = s(n) - s''(n)$

# Adaptivní pulsní kódová modulace

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Možné velké změny amplitudy signálu:
  - Nepřesné zachycení slabého signálu – amplituda je příliš malá, srovnatelná s kvantizačním krokem (příliš velký kvantizační krok).
  - Zkreslení (ořezání) silného signálu – dojde k přetečení rozsahu hodnot určených pro zakódování signálu (příliš malý kvantizační krok).
- Řešení: přizpůsobení kvantizačního kroku amplitudě signálu.

# Způsoby komunikace uživatele s dialogovým systémem

Dialogové systémy

Luděk Bártek

Základy fonetiky

Úvod do počítačového zpracování zvuku

Komunikace uživatel – dialogový systém

VoIP  
SIP

Zpracování digitalizovaného signálu

Analýza digitalizovaného signálu v časové oblasti

Analýza digitalizovaného signálu ve frekvenční oblasti

- Hlasová:
  - komunikace většinou prostřednictvím telefonní sítě (PSTN, VoIP).
  - Digitalizace hlasu probíhá:
    - Na straně uživatele – komunikace pomocí VoIP.
    - Na straně telefonní ústředny – DS používá VoIP, uživatel používá PSTN.
    - Na straně DS – uživatel i DS používají PSTN.
  - Rozpoznávání řeči probíhá většinou na straně DS.

# Způsoby komunikace uživatele s dialogovým systémem

Dialogové systémy

Luděk Bártek

Základy fonetiky

Úvod do počítačového zpracování zvuku

Komunikace uživatel – dialogový systém

VoIP  
SIP

Zpracování digitalizovaného signálu

Analýza digitalizovaného signálu v časové oblasti

Analýza digitalizovaného signálu ve frekvenční oblasti

- textová:
  - uživatel komunikuje s DS buď pomocí specializovaného klienta nebo pomocí běžných protokolů z rodiny TCP/IP.
  - Odpadá nutnost rozpoznávání řeči.
  - Využívá se hlavně pro vývoj a ladění.
- hlasová+textová:
  - komunikace s DS buď VoIP nebo specializovaný klient.
  - V případě VoIP text buď pomocí DTMF nebo simulace SMS.

- VoIP – rodina protokolů pro řízení průběhu hlasové komunikace a přenos hlasu přes internet (sít' na bázi IP).
- Využívá se pro IP telefonii.
- Využívá protokoly:
  - UDP (transportní vrstva):
    - Stará se o přenos paketů přes počítačovou sít' mezi dvěma body.
    - Není zajištěno doručení paketů ani jejich pořadí.
    - Výhoda – nízká rezie přenosu dat.
    - Nevýhody – možná ztráta dat a možnost velkých rozdílů v rychlosti doručení jednotlivých paketů
  - RTP (relační vrstva):
    - Využívá se pro přenos multimediálních dat.
    - Zajišťuje doručení paketů.
    - Umožňuje řízení parametrů přenosu – zajistí malé rozdíly v rychlosti doručení paketů.

- VoIP – řada implementací
- liší se
  - použitými standardy – H.323 (na ústupu, standard ITU, komplexní, relativně komplikovaný), SIP (jednodušší náhrada H.323, v současnosti velmi rozšířený), firemní – Skinny (Cisco), HFA (Siemens), ...
  - službami – telefonie, TV (DVB), fax, zasílání zpráv, ...
  - signalizací – závisí na zvoleném standardu a použitých protokolech.
  - ...

# Session Initiation Protocol (SIP)

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- protokol pro řízení signalizace pro VoIP na aplikační vrstvě OSI modelu
- textový protokol pracující v režimu klient–server, poskytující mechanismy pro:
  - přesměrování hovoru
  - číselnou identifikaci volajícího a volaného
  - osobní mobilitu
  - autentizaci volajícího a volaného
  - podporu konferenčních hovorů prostřednictvím vícesměrového zasílání dat (multicast).
  - ...



# SIP – pokračování

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Identifikace účastníka – URI ve tvaru *sip:číslo@adresa\_počítače*
  - číslo – číslo přidělené uživateli na daném stroji (VoIP ústředně)
  - adresa počítače – adresa (FQDN/IP) ústředny, na které je uživatel registrován.
- SIP relace může být:
  - přímá – navázána přímo komunikujícími stranami
  - s použitím SIP proxy serveru/ů – tyto slouží jako registrátoři účastníků.
- Činnosti protokolu SIP:
  - Lokalizace účastníka – pomocí identifikace
  - Zjištění stavu účastníka – připravenost k přijetí hovoru vs. obsazeno/přesměrováno
  - Zjištění možností účastníka – dostupné kodeky, dostupná šířka pásma, podpora audia/video, ...
  - Vlastní navázání spojení – využívá se protokol SDP
    - popisuje navazované spojení,
    - odkazuje na RTP datový tok, který je využit pro

# Řízení průběhu spojení pomocí protokolu SIP

Dialogové systémy

Luděk Bártek

Základy fonetiky

Úvod do počítačového zpracování zvuku

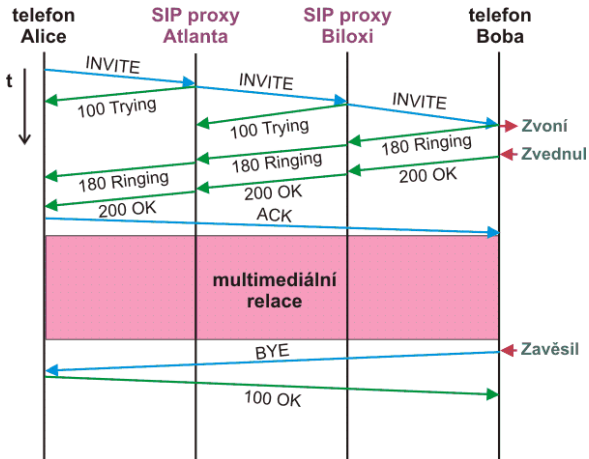
Komunikace uživatel – dialogový systém

VoIP  
SIP

Zpracování digitalizovaného signálu

Analýza digitalizovaného signálu v časové oblasti

Analýza digitalizovaného signálu ve frekvenční oblasti



# Zpracování digitalizovaného signálu

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Zvuk je neměnný pouze na krátkých časových úsecích – metody krátkodobé analýzy.
- Tento interval se nazývá mikrosegment – velikost 10 — 40 ms.
- Metody krátkodobé analýzy:
  - V časové oblasti – zpracovávají se přímo hodnoty jednotlivých vzorků.
  - Ve frekvenční oblasti – ze vzorků se získávají frekvenční charakteristiky, které jsou následně zpracovány.
- Modelování funkce Cortiho ústrojí – pomocí diferenciálních rovnic se simuluje rezonance na určitých vláčkách Cortiho ústrojí.

# Váhové okénko

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Při krátkodobé analýze předpokládáme, že signál je v okolí mikrosegmentu periodický se stejnou periodou jako uvnitř.
- Vzniklá chyba se kompenzuje použitím „okénka”.
- Okénko – posloupnost vah pro vzorky v mikrosegmentu.
- Tyto váhy by měly odpovídat tomu, jak je daný vzorek ovlivněn okolím mikrosegmentu.
- Nejčastěji používané typy okének:
  - pravoúhlé okénko
  - Hammingovo okénko

# Hammingovo okénko

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

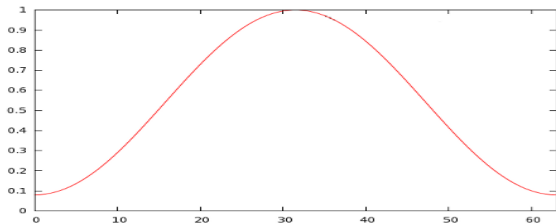
Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Vychází z předpokladu, že čím jsou vzorky blíže středu mikrosegmentu, tím méně jsou ovlivněny okolím.
- Pro výpočet vah se používá vzorec:

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

- Průběh vah okénka na mikrosegmentu:



# Pravoúhlé okénko

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Vychází se z předpokladu:
  - 1 vzorky mikrosegmentu nejsou pro naše potřeby ovlivněny okolím mikrosegmentu
  - 2 všechny vzorky mikrosegmentu jsou ovlivněny stejně.
- Všechny vzorky mikrosegmentu mají shodnou váhu.

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

# Analýza digitalizovaného signálu v časové oblasti

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Vychází přímo z hodnot vzorků, nikoliv z hodnot spektra.
- Používané metody:
  - funkce krátkodobé energie
  - funkce krátkodobé intenzity
  - funkce středního počtu průchodů nulou
  - diference 1. řádu
  - autokorelační funkce
  - ...

# Analýza v časové oblasti

## Funkce krátkodobé energie

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Využívá funkci průměrné energie v rámci segmentu:

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

- $s(k)$  – vzorek v čase  $k$
- $\omega(n-k)$  – váha odpovídajícího okénka pro čas  $k$
- Výstupem je průměrná energie v daném okénku.
- Druhá mocnina zvyšuje dynamiku zvukového signálu.
- Použití:
  - automatické oddělení ticha řeči (signálu)
  - příznaky v jednoduchých klasifikátorech slov
  - oddělení znělých a neznělých částí promluvy.



# Analýza v časové oblasti

## Funkce krátkodobé intenzity

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Funkce intenzity signálu v daném okénku.

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

- $|s(k)|$  – absolutní hodnota vzorku v čase  $k$
- $\omega(n-k)$  – váha odpovídajícího okénka pro čas  $k$
- Použití – stejné jako funkce krátkodobé energie.
- Oproti krátkodobé energii nezvýrazňuje tolik dynamiku řečového signálu.

# Analýza v časové oblasti

Krátkodobá funkce středního počtu průchodu nulou

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Počítá změny znaménka digitalizovaného signálu.

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

- Varianta – počet lokálních extrémů.
- Obě metody mohou být negativně zatíženy šumem zvukového pozadí.
- Použití:
  - detekce ticha
  - detekce začátku a konce i zašuměné promluvy
  - přibližné určení základního hlasivkového tónu a formantů
  - příznaky jednodušších klasifikátorů slov

# Analýza v časové oblasti

## Autokorelační funkce

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Vrací podobnost úseků daného mikrosegmentu (čím větší výsledná hodnota, tím podobnější úseky posunuté o  $m$  vzorků).

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

- Je-li signál periodický s periodou  $P$ ,  $R(m, n)$  nabývá maxima pro  $m=0, P, 2P, \dots$
- Předpokládá délku mikrosegmentu aspoň  $2P$ .
- Použití:
  - Používá se k zjišťování periodicity signálu základního tónu řeči.
  - Základ pro výpočet koeficientů LPA

# Analýza signálu ve frekvenční oblasti

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Transformuje digitální řečový signál z časové oblasti do frekvenční oblasti.
- Využívá k tomu nejčastěji Fourierovu transformaci.
- Nejčastěji používané druhy analýzy ve frekvenční oblasti:
  - krátkodobá Fourierova transformace
  - krátkodobá diskrétní Fourierova transformace
  - rychlá Fourierova transformace
  - kepstrální analýza
  - lineární predikce
  - ...

# Analýza signálu ve frekvenční oblasti

## Krátkodobá Fourierova transformace

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém  
VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Vychází z Fourierovy transformace:

$$S(\omega, t) = \sum_{k=-\infty}^{\infty} s(k)h(n-k)e^{-j\omega k}$$

- Obyčejnou Fourierovu transformaci získáme fixací času  $t$ .
  - $|S(\omega, t)|$  – amplituda složky akustického spektra odpovídající frekvenci  $\omega$  v čase  $t$ .
  - $h(n)$  – váhová funkce okénka.
- Předpokládá na vstupu periodickou funkci – zvuk je periodický na krátkých časových úsecích.
  - Při jejím použití se předpokládá, že zpracováváný mikrosegment se periodicky opakuje.

# Analýza signálu ve frekvenční oblasti

## Diskrétní Fourierova transformace

Dialogové systémy

Luděk Bártek

Základy fonetiky

Úvod do počítačového zpracování zvuku

Komunikace uživatel – dialogový systém

VoIP  
SIP

Zpracování digitalizovaného signálu

Analýza digitalizovaného signálu v časové oblasti

Analýza digitalizovaného signálu ve frekvenční oblasti

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

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

- $|X(k)|$  – intenzita k. spektrálního koeficientu, frekvence závisí na velikosti mikrosegmentu  $N$  a vzorkovací frekvenci.
- $x(n)$  –  $n$ . vzorek daného mikrosegmentu
- $W_n = \exp(j * 2\pi / N) = \cos(2\pi / N) + j * \sin(2\pi / N)$ .
- Výpočet  $n$ . vzorku na základě hodnot  $X(k)$  – IDFT:

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

# Analýza signálu ve frekvenční oblasti

Rychlá diskrétní Fourierova transformace

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Výpočet spektrálních koeficientů pomocí DFT –  $n^2$  operací nad komplexními čísly.
- Pomocí FFT –  $N * \log_2 N / 2$  operací násobení.
- FFT požaduje, aby délka analyzovaného segmentu byla mocninou 2.

# Analýza signálu ve frekvenční oblasti

## Kepstrální analýza

Dialogové  
systémy

Luděk Bártek

Základy  
fonetiky

Úvod do  
počítačového  
zpracování  
zvuku

Komunikace  
uživatel –  
dialogový  
systém

VoIP  
SIP

Zpracování di-  
gitalizovaného  
signálu

Analýza  
digitalizovaného  
signálu v časové  
oblasti

Analýza  
digitalizovaného  
signálu ve  
frekvenční  
oblasti

- Vychází z modelu činnosti hlasového ústrojí.
- Řečové kmity lze modelovat jako odezvu lineárního systému na buzení sestávající ze sledu pulzů pro znělou řeč 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 perioda základního tónu, formanty, ...
  - rozpoznávání slov
  - verifikace a identifikace mluvčího
  - ...



# Analýza signálu ve frekvenční oblasti

## Lineární prediktivní analýza

Dialogové systémy

Luděk Bártek

Základy fonetiky

Úvod do počítačového zpracování zvuku

Komunikace uživatel – dialogový systém

VoIP  
SIP

Zpracování digitalizovaného signálu

Analýza digitalizovaného signálu v časové oblasti

Analýza digitalizovaného signálu ve frekvenční oblasti

- 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  $u(k)$ :

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

kde  $G$  je koeficient zesílení a  $N$  řád modelu.

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

# Dialogové systémy

## Rozpoznávání řeči

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

- Rozpoznávání plynulé řeči – převádí souvislou promluvu na psaný text.
- Rozpoznávání izolovaných slov/příkazů.
- Princip rozpoznávání:
  - 1 Získání vektoru příznaků pomocí metod krátkodobé analýzy signálu.
  - 2 Klasifikace na základě vektoru příznaku získaného v předchozím kroku.

# Rozpoznávání izolovaných slov

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Slouží k rozpoznání povelů nebo slov (příkazů) zřetelně oddělených na začátku a konci mezerou.
- Odpadá problém stanovení začátku a konce slova v souvislé promluvě.
- Obvykle systémy závislé na uživateli:
  - nutnost natrénování
  - omezená kapacita slovníku.
- Obtíže při rozpoznávání izolovaných slov:
  - určení začátku a konce promluvy:
    - Odlišení šumu od sykavek.
    - Detekce nahodilého zvukového vzruchu (klepnutí, ...)  
kontra okluzívy, které obsahují pauzy.
    - Možná přítomnost infrazvuků.
    - ...

# Rozpoznávání izolovaných slov

## Typy klasifikátorů

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
roznávání  
řeči

- Klasifikátory využívající porovnání slov metodou DTW
  - Snaží se nalézt co největší shodu mezi rozpoznávaným slovem a slovy v databázi.
- Klasifikátory založené na statistických metodách – modelování pomocí skrytých Markovových modelů:
  - simulace procesu tvorby řeči.
- Klasifikátory pracující na dvou úrovních:
  - 1 Segmentace a fonetické dekódování jednotlivých segmentů.
  - 2 Rozpoznání slova na základě dekódovaných segmentů.

# Dynamic Time Warping (DTW)

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW

HMM  
Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Metoda borcení časové osy.
- Používá se pro porovnání dvou číselných řad – dvou úseků promluv (dvou slov).
- Vstup:
  - posloupnost akustických vektorů získaných pomocí metod krátkodobé analýzy signálu
  - databáze akustických vektorů rozpoznávaných slov.
- Výstup – rozpoznané slovo resp. povel.

- Vytvoříme db rozpoznávaných slov (referenční posloupnosti akustických vektorů).
  - Obvykle několik posloupností pro každé slovo, které odpovídají několika způsobům vyslovení příkazu.
- Rozpoznávané slovo převedeme na odpovídající posloupnost akustických vektorů.
- Metodou DTW nalezneme referenční posloupnost akustických vektorů s maximální shodou.

- Algoritmus DTW hledá parametrizaci  $f, g$ :

$$f, g : i = f(k), j = g(k), k \in \langle 1, K \rangle$$

minimalizující výraz

$$D(A, B) = \sum_{i=1}^K d(a_{f(i)}, b_{g(i)})$$

- $d$  – vzdálenost akustických vektorů (např. Euklidovská metrika)



- $f, g$  – 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 k nevhodné korespondenci mezi příliš krátkým segmentem vzorku  $a$  a příliš dlouhým segmentem vzorku  $b$ .
- Omezení na hraniční body:
  - $f(1) = 1, f(K) = I$ , kde  $I$  je počet vzorků slova  $a$ .
  - $g(1) = 1, g(K) = J$ , kde  $J$  je počet vzorků slova  $b$ .

- Globální vymezení oblasti pohybu funkce DTW:
  - omezení minimální a maximální přípustné směrnice přímky vymežující přípustnou oblast pohybu funkce DTW, při splnění podmínky na hraniční body:

$$1 + \alpha[i(k) - 1] \leq 1 + \beta[i(k) - 1]$$

- $\alpha$  – minimální směrnice přímky omezující přípustnou oblast
- $\beta$  – maximální směrnice přímky omezující přípustnou oblast.

# DTW – Praktická realizace klasifikátoru slov

## Blokové schéma

Dialogové  
systémy

Luděk Bártek

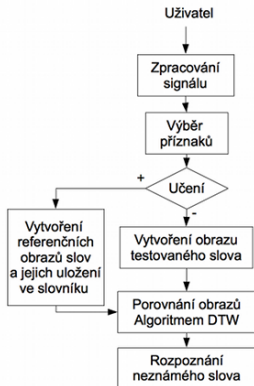
Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči



Obrázek: Blokové schéma klasifikátoru slov

### ■ Obecný postup:

- 1 Řečník resp. skupina řečníků vysloví postupně každé trénované slovo požadovaného slovníku, buď jednou nebo opakovaně.
- 2 Vstupní slova jsou zdigitalizována a následně převedena zvolenou metodou krátkodobé analýzy na posloupnost vektorů příznaků.
- 3 Detekce hranic (počátku a konce) slov:
  - Může být náročné na provedení, např. kvůli rušivému pozadí.
  - Nekorektní detekce hranic slov zhoršuje úspěšnost rozpoznávání.
  - Metody odstraňující i jen částečně vliv akustického pozadí zvyšují výpočetní náročnost.
- 4 Vytvoření referenčních obrazů slov.

- Přímé použití obrazů trénovací množiny jako referenčních obrazů slov – 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 dynamického průměrování.
- Vytváření vzorových obrazů shlukováním.
  - Vzorové obrazy pro dané slovo se rozdělí 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), automaticky (algoritmy založené na MacQueenově algoritmu).

- Nevýhody DTW – vysoké paměťové a výpočetní nároky mohou znesnadňovat klasifikaci v reálném čase i při relativně malém slovníku.
- Metody řešení:
  - Hrubá síla – využití paralelních procesorů popř. zákaznických obvodů – může být drahé.
  - Vhodné zakódování parametrů jednotlivých mikrosegmentů referenčních i testovacích obrazů. Využívá se:
    - vektorová kvantizace – počet různých vzorků je konečný – uloží se do kódové knihy a místo hodnoty vzorku se pracuje s jejich indexy v kódové knize.
    - kódová kniha – abeceda všech hodnot, které se vyskytly v signálu (lze kódovat úsporněji než při použití standardního PCM).

- Využití oblastí spektrální stacionarity – metoda segmentace spektální stopy.
  - spektrální stopa spojnice koncových bodů vektorů příznaků
  - lze ji aproximovat – např. lineárními úseky.
- Zavedení účinných způsobů vyhledávání nejbližšího souseda
  - metody prohledávání metrických prostorů
  - nutno ověřit, že vzdálenost použitá v DTW je metrika.
- Redukce výpočetních nároků pomocí heuristik při porovnávání:
  - vícestupňový rozhodovací postup:
    - 1 porovnání promluvy proti celému slovníku pomocí omezené množiny příznaků
    - 2 dohledání výsledku kroku 1. pomocí klasického DTW.
  - Práh zamítnutí:
    - 1 po každém kroku spočítáme vzdálenost slova a obrazu
    - 2 pokud překročí experimentálně stanovený práh, obraz je zamítnut.

# Skryté Markovovské Modely – HMM

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov

DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
roznávání  
řeči

- 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.
- Totu činnost lze modelovat statisticky.
- Kvantizací akustických vektorů lze dosáhnout konečnosti všech parametrů odpovídajícího modelu.



- Jsou generovány dvě vzájemně svázané časové posloupnosti náhodných proměnných:
  - podpůrný Markovův řetězec – posloupnost konečného počtu stavů
  - řetězec konečného počtu spektrálních vzorů
- Náhodná funkce ohodnocující pravděpodobnostmi vztah vzorů k jednotlivým stavům.
- Pro rozpoznávání řeči jsou nejčastěji využívány levo-pravé Markovovy modely:
  - vhodné pro modelování procesů spjatých se vzrůstajícím časem.

- Markovův proces  $G$  se skrytým Markovovým modelem je pětice  $G = (Q, V, N, M, n)$ 
  - $Q = q_1, \dots, q_k$  – množina stavů
  - $V = v_1, \dots, v_k$  – množina výstupních symbolů
  - $N = (n_{i,j})$  – matice přechodu. Určuje pravděpodobnost přechodu ze stavu  $q_i$  v čase  $t_1$  do stavu  $q_j$  v čase  $t_2$ .
  - $M = (m_{i,j})$  – matice přechodu, určující pravděpodobnost generování akustického vektoru  $v_j$ , v kterémkoliv čase ve stavu  $q_i$ .
  - $n = (n_i)$  – vektor pravděpodobností počátečního stavu (pravděpodobnost toho, že stav  $i$  je počáteční).
- Trojice  $\lambda = (N, M, n)$  – vytváří model řečového segmentu.
  - např. Vintsjukův model pro slovo – počet stavů 40 — 50 (odvozeno od průměrného počtu mikrosegmentů ve slově; délka mikrosegmentu 10 ms).

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

### ■ Výpočet odpředu:

- $\alpha_i$  – pravděpodobnost přechodu do stavu  $q_i$  při generování posloupnosti  $\{o_1, \dots, o_t\}$  ( $\alpha_i = P(o_1 \dots o_t, q(t)_i | \lambda)$ )
- Rekurzivní výpočet:

1 inicializace:  $\alpha_1(i) = \pi_i m_i(o_1), i \in \langle 1, N \rangle$

2 Rekurzivní krok pro  $t=1, \dots, T-1$ :

$$\alpha_{i+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) n_{i,j} \right] m_j(o_{i+1})$$

pro  $j \in \langle 1, N \rangle$ ,  $m(o_t)$  je ekvivalentní zápisu  $m_i(l)$ ,  
pokud  $o_t = v_l$ .

3 Výsledná pravděpodobnost:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$$

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

# HMM

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

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Nutno stanovit postup při trénování parametrů modelu.
- Cíl trénování:
  - maximalizace pravděpodobnosti  $P(O|\lambda)$
- Problém:
  - neexistuje analytická metoda ke zjištění globálního maxima funkce n proměnných.
- Řešení:
  - lze použít iterativní algoritmy zajišťující aspoň lokální maximalitu.
- Nejpoužívanější postup – Bauman-Welchův algoritmus.
- Další problém při trénování modelu:
  - vliv konečné trénovací množiny:
    - čím menší trénovací množina a čím větší matice M, tím větší pravděpodobnost, že některé prvky zůstanou nastaveny na 0 (problém chybějících/neadekvátních dat).

# HMM

## Rozhodovací pravidlo při rozpoznávání izolovaného slova

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Používá se princip maximální věrohodnosti.
  - 1 Pro slovo  $O$  a všechna  $\lambda$ :
    - 1 Spočítáme  $P(O|\lambda)$ .
  - 2 Jako výsledek vybereme třídu s maximální hodnotou  $P(O|\lambda)$ .

- Modelování povelů:
  - nejčastěji se používají modely se 4 — 7 stavy.
  - Pro modelování lze využít nástroje pro tvorbu HMM
    - HTK – Hidden Markov Model Toolkit.
- Modelování fonémů:
  - obvykle 4 — 7 stavů
  - model slova – zřetěžení modelů fonémů
  - problémy s výpočtem v reálném čase
    - lze řešit pomocí speciálních algoritmů pro hledání maxima  $P(O|\lambda)$ .



# Příklady struktur pro fonémy

Dialogové  
systémy

Luděk Bártek

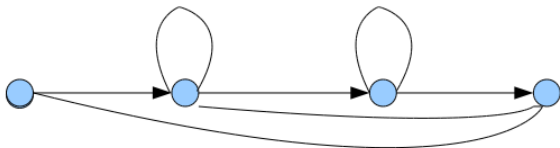
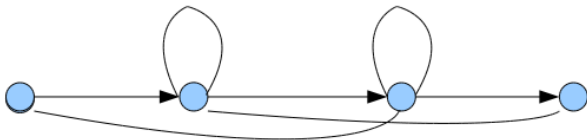
Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči



# Příklady struktur pro fonémy

Dialogové  
systémy

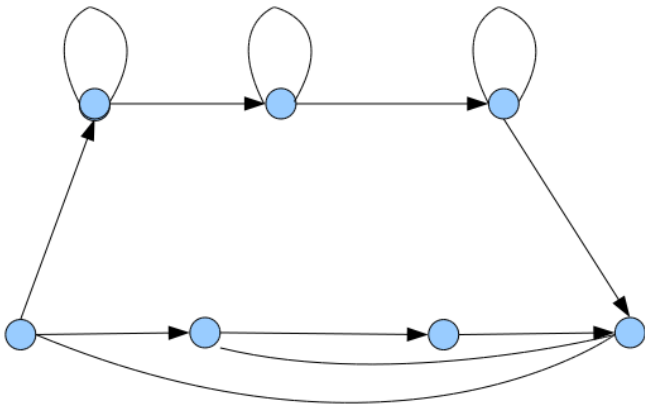
Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči



# Rozpoznávání plynulé řeči

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Hlavní rozdíly oproti rozpoznávání slov:
  - nelze vytvořit databázi vzorů
  - nutno brát zřetel na prozodické faktory
  - nutno určovat hranice mezi slovy
  - vypořádání se s výplňkovými zvuky a chybami řeči.
- Řešení – statistický přístup:
  - jazykový model
  - model uživatele.
- Příklad: HMM vrátí stejnou pravděpodobnost např. pro slova „máma“ a „nána“ – nejspíše se použije máma – je častější.

- Máme:
  - posloupnost slov (promluva)  $W = (w_1, \dots, w_n)$
  - posloupnost akustických vektorů  $O = (o_1, \dots, o_t)$ .
- Chceme nalézt  $W^*$  (množinu všech promluv), která maximalizuje  $P(W|O)$ .
- Dle Bayesova pravidla platí:

$$P(W^*|O) = \max P(W|O) = \max \frac{P(W) * P(O|W)}{P(O)}$$

# Rozpoznávání plynulé řeči

Jazykové modely – pokračování

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
roznávání  
řeči

- Pro nalezení maxima potřebujeme znát:
  - model řečníka –  $P(O|W)$
  - jazykový model –  $P(W)$ .
- Model řečníka lze nahradit pravděpodobností generování  $W$  odpovídajícím Markovovým modelem.
- Trigramový model:
  - Experimentálně ověřeno, že platí:

$$P(w_n | w_1 \dots w_{n-1}) \cong P(w_n | w_{n-2} w_{n-1})$$

# Rozpoznávání plynulé řeči

## Rozpoznávání tématu

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Úspěšnost rozpoznávání řeči se pohybuje cca 50 % — 99 % v závislosti na úkolu, jazyku, . . .
- Úspěšnost rozpoznávání lze zvýšit omezením domény rozpoznávání:
  - rozpoznání tématu
  - použitím gramatik pro rozpoznávání řeči.
- Známé téma:
  - změna stavového prostoru a pravděpodobnosti trigramů:
    - např. burzovní zprávy – rozpoznáno „honey“ nebo „money“?
  - možnost vytvoření přesnějšího jazykového modelu.

# Gramatiky pro podporu rozpoznávání řeči

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov

DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Úspěšnost obecného rozpoznávání plynulé řeči může klesnout až na cca 50 %.
- Zvýšení lze dosáhnout omezením domény – např. specifikováním přípustných vstupů.
- Lze použít gramatiky pro podporu rozpoznávání řeči:
  - bezkontextové gramatiky
- Způsoby zápisů gramatik:
  - prostředky logického programování
  - proprietární řešení
  - otevřené standardy – JSGF, W3C SRGS, ...

# Gramatiky pro podporu rozpoznávání řeči

## Java Speech Grammar Specification (JSGF)

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Textový zápis gramatiky nezávislý na platformě a prodejci.
- Určen pro použití při rozpoznávání řeči.
- Součást Java Speech API.
- Používá styl a konvence jazyka Java.
- Aktuální verze 1.0 (říjen 1998).
- Použit např. v rozpoznávači Sphinx-4, VoiceXML intepreteru VoiceGlue, ...
- Podrobněji v 2. polovině semestru při probírání tvorby dialogových rozhraní.



# Gramatiky pro podporu rozpoznávání řeči

## Ukázka JSGF

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov

DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

### #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>;

# Gramatiky pro podporu rozpoznávání řeči

## W3C Speech Recognition Grammar Specification (SRGS)

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW

HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

- Standard W3C.
- Aktuální verze 1.0 (březen 2004).
- Definuje způsob zápisu pravidel a jejich odkazování.
- Dva způsoby zápisu:
  - XML
  - ABNF (Augmented BNF).
- Podrobněji v 2. polovině semestru při probírání tvorby dialogových rozhraní.

# Ukázka W3C SRGS

Dialogové  
systémy

Luděk Bártek

Rozpoznávání  
řeči

Rozpoznávání  
izolovaných slov  
DTW  
HMM

Rozpoznávání  
plynulé řeči

Gramatiky pro  
podporu  
rozpoznávání  
řeči

```
#ABNF 1.0 UTF-8
```

```
root $pozdrav;
```

```
language cs-CZ;
```

```
mode voice;
```

```
$pozdrav = ahoj
```

```
<?xml version="1.0" encoding="utf-8"? >
```

```
<grammar root="pozdrav" xml:lang="cs-CZ"  
version="1.0" >
```

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

```
ahoj
```

```
< /rule>
```

```
< /grammar>
```

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

# Sémantická interpretace promluvy

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Cíl – počítači srozumitelná interpretace informací zadaných uživatelem.
- Příklad:  
Chtěl bych si koupit Zkrocení zlé ženy od Shakespeara.
  - akce = nákup
  - titul = Zkrocení zlé ženy
  - autor = Shakespeare
- Reprezentace – dvojice (atribut, hodnota).
- Implementace:
  - pravidlům gramatiky pro rozpoznávání řeči přidáme atributy, do kterých ukládáme jejich sémantickou reprezentaci.
  - Na attributech lze provádět operace sloužící k sestavení sémantické interpretace celé promluvy z interpretací jednotlivých pravidel.

# Popis Sémantické Interpretace

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

## ■ JSGF:

- přiřazena pomocí značek (tags)
- zápis – {sémantická interpretace}

< *sentence* > = < *intro* > < *titul* > od < *autor* >

< *titul* > = Pejska a kočičku

{Povídání o pejskovi a kočičce}|

(Zlou ženu|Zkrocení zlé ženy) {Zkrocení zlé ženy}|...

## ■ SRGS – standard SISR :

- standard W3C Voice Browser Activity.
- Je postaven na jazyce ECMAScript.
- K pravidlům se přidává pomocí značky nebo atributu *tag*.
- Do dialogu je interpretace vracena ve formátu JSON.

■ ...

# Základní pojmy

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Dialog – rozhovor dvou a více účastníků (sled promluv).
- Promluva – Souvislé sdělení, které učiní jeden účastník dialogu směrem k druhému.
- Obrat – Promluva a reakce druhého účastníka na ni.
- Dialogová strategie
  - Postup, který k dané promluvě přiřazuje následující promluvu.
  - Využívá znalost stavu dialogu:
    - zadané a požadované informace
    - schopnosti účastníků dialogu
    - ...
  - Je vlastností každého účastníka dialogu.

- Hodnotící funkce:
  - funkce přiřazující každému dialogu reálné číslo.
  - Označuje se  $E(L)$ , kde  $L$  je dialog.
- Dialogová komunikace – Uspořádaná čtveřice

$$M = (S_1, S_2, E_1, E_2)$$

- $S_i, i \in \{1, 2\}$  – dialogová strategie příslušného účastníka.
- $E_i, i \in \{1, 2\}$  – hodnotící funkce příslušného účastníka.



# Základní pojmy

## Kooperativita dialogu

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Dialogovou komunikaci  $M = (S_1, S_2, E_1, E_2)$  nazveme:
  - Kooperativní  $\Leftrightarrow E_1 = E_2$ . Oba účastníci dialogu mají stejný cíl a snaží se spolupracovat.
  - Nekooperativní  $\Leftrightarrow E_1 \neq E_2$ . Cíle obou účastníků dialogu se odlišují.
  - S nulovým součtem  $\Leftrightarrow E_1 = -E_2$ . Cíle obou účastníků dialogu jsou protichůdné.
- Toto hodnocení vychází z teorie her:
  - na dialog lze pohlížet jako na hru dvou účastníků.

# Pravidla pro vedení kooperativního dialogu

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Dialogový systém by se měl snažit o kooperativní dialog.
- Autor Herbert Paul Grice – anglický jazykovědec.
- Aspekt informativnosti:
  - 1 Bud' přiměřeně informativní – ne méně než je potřeba, ale ani 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 nejednoznačností.
  - 3 Usilujte o stručnost.
  - 4 Buďte disciplinovaní, udržujte v dialogu pořádek.

# Pravidla pro vedení kooperativního dialogu

## pokračování

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

### ■ 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 jeho nesouhlas.
- 4 Maximalizujte empatii vůči partnerovi.

### ■ Aspekt asymetrie:

- 1 Informujte uživatele o všech důležitých charakteristikách, které 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 o způsobu interakce se systémem.

## ■ Aspekt znalostí a schopností:

- 1 Vezměte v úvahu relevantní znalosti uživatele.
- 2 Vezměte v úvahu možné uživatelské 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 meta komunikaci zajišťující odstranění chyby nebo její vysvětlení.
- 2 Zajistěte vysvětlující meta komunikaci v případě nekonsistentních nebo nejednoznačných uživatelských vstupních dat.

# Aspekty komunikace kooperativního dialogového systému

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Dialogový systém při komunikaci s uživatelem by měl brát ohled na následující aspekty:
  - aspekt informativnosti
  - aspekt přesvědčivosti
  - aspekt způsobu
  - aspekty zdvořilosti, empatie a etiky
  - aspekt asymetrie
  - aspekt znalostí a schopností uživatele
  - aspekt vyjasňování a odstraňování chyb.

# Iniciativa v dialogu

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- 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.
- Reálné systémy používají:
  - dialogy se smíšenou iniciativou
  - dialogy s iniciativou systému.

# Iniciativa v dialogu

## Příklady

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

### ■ Dialog s iniciativou systému:

System: Zadejte Vaše uživatelské jméno

Uživatel: xyz

System: Aby Vás bylo možné ověřit řekněte větu:  
„Můj hlas je můj pas. Ověř si mě.“

Uživatel: Můj hlas je můj pas. Ověř si mě.

...

### ■ Dialog se smíšenou iniciativou:

Uživatel: Chtěl bych je dnes vlakem z Adamova  
do Kerkyry.

System: Chcete nalézt přímý spoj nebo spojení  
s přestupy?

Uživatel: Chtěl bych jet bez přestupů.

System: Je mi líto, ale přímý vlak z Adamova do Kerkyry  
nejezdí.

...

# Zpětná vazba v dialogovém systému

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Před tím, než systém zpracuje získané informace, je vhodné provést jejich verifikaci:
  - oprava chyb rozpoznávání řeči
  - oprava chyb uživatele
  - ...
- Způsoby ověření získaných dat:
  - Sumarizující zpětná vazba – po zadání veškerých dat uživatelem je zopakuje a případně umožní jejich opravu.
  - Zpětná vazba „echo“ – po zadání každého údaje ho uživateli zopakuje, poskytne mu možnost případné opravy.
  - Implicitní zpětná vazba – posledně zadaná data jsou součástí dotazu na následující údaj.
  - Explicitní zpětná vazba – systém validuje zadaná data pomocí explicitních dotazů na jejich hodnoty.



# Zpětná vazba v dialogovém systému

## Příklady

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Sumarizující zpětná vazba:
  - Uživatel: Chci jet vlakem z Adamova do Kerkyry.
  - System: Hledaný druh spojení Vlak. Odjezd Adamov, cílová stanice Kerkyra.
- Zpětná vazba „echo“:
  - System: Čím chcete jet?
  - Uživatel: Vlakem.
  - System: Chcete jet vlakem. Odkud chcete jet?
  - Uživatel: Z Adamova.
  - System: Chcete jet z Adamova. Kam chcete jet?
  - Uživatel: Do Kerkyry.
  - ...

# Zpětná vazba v dialogovém systému

## Příklady

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

### ■ 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

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

**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

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Možnost přerušit systém.
- Možnost iniciovat korekce a metakomunikaci uživatelem:  
Uživatel: Mohl, byste prosím zopakovat poslední dotaz?
- Možnost získání nápovědy:  
Uživatel: Mohl byste mi upřesnit požadované údaje?
- Detekce znalostí a schopností uživatele:
  - Částečně lze odvodit z předchozího průběhu dialogu:
    - počet chyb uživatele
    - počet žádostí o nápovědu
    - ...
  - Uzpůsobení iniciativy dialogového systému.
- Přizpůsobení dialogové strategie – např. uživatel ve spěchu, vystresovaný uživatel, ...

# Další aspekty dialogových strategií

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Detekce emocí uživatele:
  - na základě neverbálních charakteristik hlasu – nutná podpora v modulu rozpoznávání řeči
  - pomocí dalších čidel – EEG, EKG, ...
- vícejazyčnost (multilingualita).
- multimodalita
  - rozhovor vede avatar (talking head) – vhodné např. pro uživatele s poruchou slyšení.
  - alternativní způsob vstupu:
    - klávesnice
    - snímání obličeje
    - ...
- zdvořilost
- prozódie
- učení se z chyb.

# Typické chyby systému

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Chyby komunikačního rozhraní:
  - chybné rozpoznání řeči
  - chybná sémantická interpretace rozpoznané promluvy
- Technické chyby:
  - problémy spojené s komunikací pomocí VoIP
  - chybné parametry HW a periférií
- Chyby v dialogové strategii.
- Chybná funkce aplikace (řešiče problémů).
- Chybné funkce dalších modulů systému.

# Typické chyby uživatele

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Současná odpověď na více otázek, než je položeno:  
System: Který den chcete cestovat?  
Uživatel: Zítra dopoledne nebo o den později odpoledne rychlíkem z Brna do Vsetína.
  - Nedostatek lze eliminovat použitím různých způsobů analýzy promluvy podle úrovně znalostí uživatele.
- Odpověď na jinou otázku, než která je položena:  
System: Který den chcete cestovat?  
Uživatel: Chtěl bych jet někdy odpoledne z Brna do Vídně.
  - Následovat by měl dotaz s nabídkou řešení pro nejbližší dny.

# Typické chyby uživatele

## Pokračování

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Nedostatečná odpověď na otázku:

Systém: Odkud kam chcete cestovat?

Uživatel: Chci jet do Vídně.

- Systém, pokud je schopen lokalizovat uživatele, by mohl uživateli nabídnout spojení z jeho aktuální lokality pokud ne, je nutno se uživatele doptat na místo odkud chce cestovat.

- Odpověď na jinou otázku, než je položena:

Systém: Který den chcete cestovat?

Uživatel: Chtěl bych jet někdy odpoledne z Brna do Svitav.

- Systém může uživateli nabídnout různé alternativy odpoledních spojení a nebo si vyžádat upřesnění dne odjezdu.



# Typické chyby uživatele

## Pokračování

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

### ■ Ignorování informace:

Systém: V sobotu dopoledne není žádné spojení do Svitav. Zadejte jiný čas.

Uživatel: Něco tam jet musí.

### ■ Odbočení z dialogu:

Systém: Vítejte v automatickém systému pro získání informací o dopravních spojeních. . . .  
Kdy chcete cestovat?

Uživatel: A to se mnou opravdu nemluví člověk, ale počítač?

# Nástroje pro popis dialogových strategií

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Proprietární řešení:
  - Pro zápis dialogové strategie lze využít běžné programovací jazyky a jejich knihovny.
    - Nástroje logického programování – Prolog
    - Procedurální programovací jazyky – C/C++, ...
  - Specifické značkovací jazyky.
  - ...
- Otevřené standardy:
  - Standardy W3C Voice Browser Activity – VoiceXML.
  - CallXML – standard využívaný např. v platformě Voxeo Prophecy.
  - ...

# Proč simulovat dialog?

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Testování návrhu DS:
  - nelze provádět až po implementaci DS:
    - časová náročnost nové implementace
    - velká pravděpodobnost zanesení nových chyb.
  - Testování v době návrhu umožňuje včasné podchycení a odstranění slabých míst v komunikačním rozhraní (viz předchozí kapitola).

# Způsoby simulace dialogových systémů

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Simulace modelem člověk — člověk:
  - simulaci provádí dva lidé — snaží se napodobit chování dialogového systému
  - průběh dialogu je protokolován.
- Simulace v textové formě:
  - psané záznamy mluvených dialogů
  - textové simulace mluvených dialogů
  - dialogy prostřednictvím Internetu:
    - IRC
    - IM
    - Usenet News
    - ...

# Způsoby simulace dialogových systémů

Dialogové  
systémy

Luděk Bártek

Sémantická  
interpretace  
promluvy

Řízení průběhu  
dialogu

Popis  
dialogových  
rozhraní

Simulace  
dialogových  
systémů

- Simulace metodou Wizard of Oz (WOZ):
  - Funkce dialogového systému skrytě simulována člověkem.
  - Simulace napodobuje testované dialogové rozhraní – drží se testované dialogové strategie.
    - Pokud je dostupný prototyp, může jej použít – Wizard modifikuje výstupy systému, případně uživatelské vstupy.
  - Uživatel by se neměl dozvědět, že komunikuje s člověkem.
    - Používají se prostředky, které modifikují hlas operátora, aby získal „počítačové“ zbarvení, ...

- Podrobnou a funkční dialogovou strategii DS a požadavky pro uživatele, jaké akce má s dialogovým systémem provádět.
- Rozhodnutí, jaké moduly budou simulovány a jakým způsobem:
  - simulace různých chyb (např. rozpoznávání řeči)
  - simulace sémantického analyzátoru
  - ...
- Rozhodnutí, zda a do jaké míry se bude uživateli předstírat komunikace s reálným systémem.
- Reprezentativní výběr uživatelů:
  - výběr by měl rozložením odpovídat rozložení kategorií v cílové skupině uživatelů.

- Volba strategie simulace s ohledem na cíl:
  - testování dialogových strategií
  - sběr dat do dialogového korpusu
  - testování modulu pro ošetření chybových stavů
  - ...
- Vhodný výběr operátora – měl by:
  - dobře znát testovaný systém a jeho případná rozhraní.
  - dobře znát omezení jednotlivých komponent testovaného systému.
  - ...
- Způsob simulace komponent dialogového systému – aplikační logiky DS, ...

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011



# Dotazovací systémy pro relační databáze

Dialogové systémy

Luděk Bártek

Dotazovací systémy pro relační databáze

Databáze, atributy a vyhledávací stromy

Dialogové systémy a emoce

- Struktura databáze je určena relačním schématem (systémem atributů).
- Systém se snaží na základě uživatelem zadaných hodnot atributů (ne nezbytně všech) nalézt požadovanou odpověď.

# Pawlakův informační systém

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

- Autor — Zdzislaw I. Pawlak (1926 — 2006), polský matematik, člen polské Akademie věd.
- Pawlakův informační systém - čtveřice  $S = (U, T, V, f)$ :
  - $U$  – množina objektů
  - $T$  – množina atributů
  - $V$  – množina hodnot atributů
  - $f : U \times T \rightarrow V$
- Pawlakův informační systém formálně popisuje vztahy mezi objekty, jejich atributy a jejich hodnotami.
- Souvislost s dialogovými systémy – hledání minimální množiny hodnot atributů, které nám určují jednotlivé objekty.

# Pawlakův informační systém

## Příklad

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

	<i>Prvek<sub>1</sub></i>	<i>Prvek<sub>2</sub></i>	<i>Prvek<sub>3</sub></i>	<i>Prvek<sub>4</sub></i>
<i>Atribut<sub>1</sub></i>	1	1	0	0
<i>Atribut<sub>2</sub></i>	0	1	1	1
<i>Atribut<sub>3</sub></i>	1	1	1	0

- Konstrukce vyhledávacího stromu pro Pawlakův IS:
  - 1 Postupně bereme jednotlivé atributy obsažené v IS a ptáme se na jeho přítomnost (hodnotu).
  - 2 Listy jsou jednotlivé prvky, uložené v IS.
- Souvislost s dialogovým rozhraním (s iniciativou systému):
  - Na každé úrovni stromu se ptáme na hodnotu/přítomnost odpovídajícího atributu.
  - Uživatelská odpověď určuje pokračování dialogu.
- Lze použít i dialog se smíšenou iniciativou:
  - 1 Uživatel zadá hodnoty libovolného počtu atributů.
  - 2 Systém odpověď zpracuje a doptá se na chybějící hodnoty.

# Pawlakův informační systém

## Příklady vyhledávacích stromů

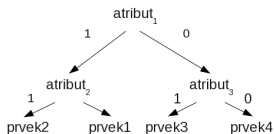
Dialogové systémy

Luděk Bártek

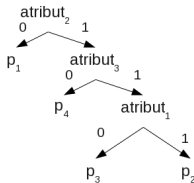
Dotazovací systémy pro relační databáze

Databáze, atributy a vyhledávací stromy

Dialogové systémy a emoce



Obrázek: Vyhledávací strom pro Pawlakův IS 2



Obrázek: Jiný vyhledávací strom pro Pawlakův IS 2

- Z daného systému atributů vybrat minimální systém jednoznačně rozlišující jednotlivé prvky.

	<i>Prvek<sub>1</sub></i>	<i>Prvek<sub>2</sub></i>	<i>Prvek<sub>3</sub></i>	<i>Prvek<sub>4</sub></i>
<i>atribut<sub>1</sub></i>	1	1	1	0
<i>atribut<sub>2</sub></i>	1	1	0	1
<i>atribut<sub>3</sub></i>	1	0	1	1
<i>atribut<sub>4</sub></i>	1	1	0	0
<i>atribut<sub>5</sub></i>	1	0	1	0

- Bylo dokázáno, že tento problém je NP-úplný.
- Na základě atributů vybrat optimální vyhledávací strom.
  - Kritérium – např. výška stromu – problém je opět NP-úplný.
- Existují aproximativní algoritmy.

# Dialogové systémy a emoce

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

- Počítače pracující s emocemi – počítače, které mají schopnost vyjádřit emoce, rozpoznat emoce a měnit své chování podle emocí uživatele.
- Schopnost určit emocionální stav uživatele – přizpůsobení dialogové strategie:
  - klidný uživatel vs. spěchající uživatel
  - klidný uživatel vs. rozčilený uživatel
  - rostoucí napětí uživatele
  - ...
- Emocionální stav má souvislost s prozodíí.
  - TTS může modelovat emoce pomocí prozodie.
  - Při rozpoznávání lze detekovat emoce pomocí prozodie.

# Dialogové systémy a emoce

## Oblasti využití

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

- Dialogová rozhraní informačních systémů – spokojenost/nespokojenost, spěch, ...
- DS pro výuku – uživatel se nudí, je napjatý, unavený, ...
- Umělá (počítačová) empatie.
- Automobilové systémy.
- Systémy bojových letadel, vrtulníků, simulace vojenských operací.
- ...



# Co jsou to emoce?

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

- This is a very tough question, that has produced significant amounts of headaches to scientists in the past ...
- ... many researchers have opted to study systematically phenomena that most consider emotional. (Laval University Quebec)
- René Descartes: „Only mathematics is certain, so all must be based on mathematics“
  - 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, štěstí, smutek, strach, překvapení.

- Arnold – hněv, averze, odvaha, sklíčenost, touha, zoufalství, strach, nenávisť, naděje, láska, smutek.
- Ekman, Friesen, and Ellsworth – hněv, odpor, strach, radost, smutek, překvapení.
- Frijda – touha, štěstí, zájem, překvapení, údiv, zármutek.
- Gray – vztek a hrůza, touha, radost.
- Izard – hněv, pohrdání, odpor, rozrušení, strach, vina, zájem, potěšení, hanba, překvapení.
- James – obava, žal, láska, vztek.
- Panksepp – očekávání, obava, vzteck, panika.

- University of Geneva (Klaus Scherer)
- Laval University, Quebec (Arvid Kappas)
- Queen's University, Belfast (R. Cowie)
- MIT Media Laboratory (R. W. Picard)
- IBM (Almanden Labs)
- University of California, Berkeley.

# Zjišťování emocí

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

- Lze provádět pomocí:
  - Změn galvanických vlastností kůže (změna odporu):



- Změn tlaku krve a pulsu:



# Zjišťování emocí

## pokračování

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

- Lze provádět pomocí:
  - Změn dýchání:



- Změn elektrické aktivity mozku:



# Zjišťování emocí

## Rozdíly charakteristik Hněv/Smutek

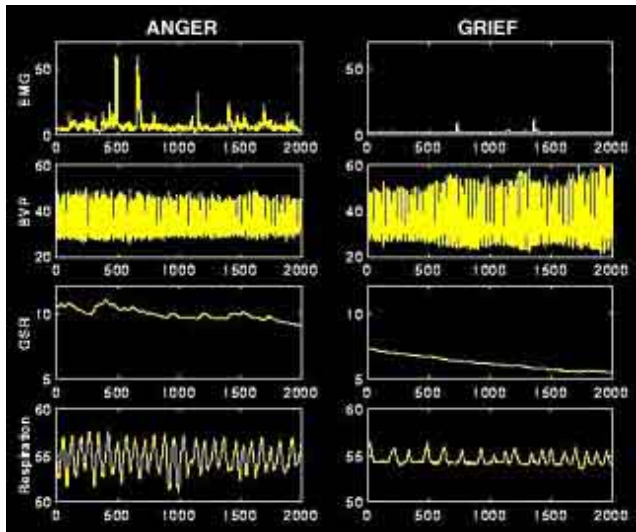
Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce



# Zjišťování emocí

## IBM Blue Eyes Project

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

- K detekci emocí využívá:
  - kameru:



- emoční myš:



# Výrazy tváře - Yale Face Database

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

## ■ Radost:



## ■ Smutek:





# Výrazy tváře - Yale Face Database

Dialogové  
systémy

Luděk Bártek

Dotazovací  
systémy pro  
relační  
databáze

Databáze,  
atributy a  
vyhledávací  
stromy

Dialogové  
systémy a  
emoce

## ■ Ospalost:



## ■ Překvapení:



# Dialogové systémy

Luděk Bártek

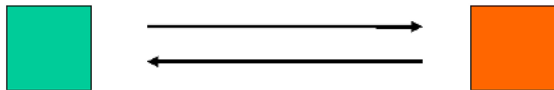
Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

- Sjednocení přístupu k dialogovým strategiím a uživatelským modelům.
- Affective computing – práce s emocemi.
- Kompatibilita s dalšími disciplínami:
  - teorie automatů a formálních jazyků
  - teorie her
  - univerzální algebra
  - ...
- Predikce kroků dialogu.
- Kompatibilita s VoiceXML.

# Matematický model dialogu

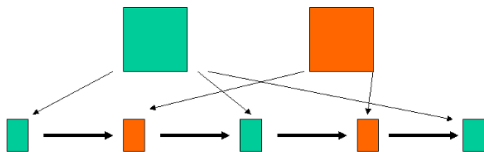
- Skládá se z:
  - dvou diskutujících stran.
  - jejich proslovů, které se střídají.



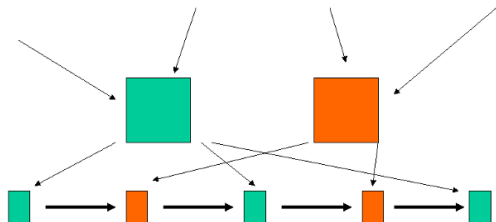
# Pragmatika dialogu

- Analyzujeme záznam dialogu z pohledu:
  - teorie řečových aktů
  - lexikální sémantiky a pragmatiky
  - dynamické sémantiky
  - dynamické epistemické logiky
  - teorie konverzace.
- Pragmatika může dále zahrnovat:
  - model uživatele/ů
  - model prostředí

- Pragmatika obsahující modely uživatelů:



- Pragmatika obsahující modely uživatelů a prostředí:



# Dialog jako Pawlakův informační systém

- Pawlakův IS  $S = (X, T, V, f)$ :
  - $X$  – množina objektů IS.
  - $T$  – množina jejich atributů.
  - $V$  – množina hodnot atributů.

$$f : X \times T \rightarrow V$$

- Příklad – dialog na trhu:
  - Atributy:
    - navrhovaná cena: -, 1 — 1000
    - nonverbální postoj: :-), :-, :-(, :->
  - Možný průběh dialogu: (1000, :-), (500, :-), (900, :-), (600, :-), (800, :-), (-, :->), (700, :-), (700, :-)

- Mealyho automat  $M$  – konečný automat s výstupem.
- $M = (Q, X, Y, \lambda, s_0)$ 
  - $Q$  – neprázdná množina stavů
  - $X$  – konečná vstupní abeceda
  - $Y$  – konečná výstupní abeceda
  - $\lambda$  – přechodová funkce:

$$\lambda : Q \times X \rightarrow Q \times Y$$



- Karetní hra, kdy hráči sázejí na sílu karet, které drží v ruce.
- Základní pravidla:
  - 1 Bankéř (rozdávající), rozdává každému hráči 5 karet.
  - 2 Hráči na základě karet, které jim byly rozdány mohou:
    - vsadit na karty určitou částku s případnou výměnou karet, které se jim nehodí a buď dorovnat poslední sázku nebo ji zvýšit.
    - Složit karty.
  - 3 Jakmile všichni pouze dorovnejí poslední navýšenou sázku, popřípadě složí všichni hráči, až na jednoho, hra končí a v případě, že zůstalo více jak jeden hráč, potom ten, který má vyšší hodnotu karet v ruce, vyhrává veškeré vsazené peníze. Pokud nesložil pouze jeden hráč, je tento vítězem dané hry.

- Tato hra má řadu variant.
- Při použití dobré strategie lze vyhrát i s velmi nízkými kartami:
  - minimalizace neverbálních projevů
  - vhodně zvolená strategie přiřazování, která budí zdání, že karty mají velmi vysokou hodnotu
  - vyhnutí se podobné strategii v následujících hrách – nemělo by jít vysledovat společné rysy chování pro danou situaci.

- Atributy stavu hry (množiny  $A_s$ ):
  - $AS_1$  – moje karty
  - $AS_2$  – celková výše sázky
  - $AS_3$  – důvěra ve vlastní karty
  - $AS_4$  – protihráčovi karty
  - $AS_5$  – hráčova strategie

# Poker

## Hodnoty atributů stavu hry

- moje karty –  $AS_1 \in \{bad, medium, good\}$
- celková výše sázky –  $AS_2 \in \{high, medium, low\}$
- důvěra ve vlastní karty –  $AS_3 \in \{high, medium, low\}$
- protihráčovi karty –  $AS_4 \in \{bad, medium, good\}$
- hráčova strategie –  $AS_5 \in \{careful, risky, bluffing\}$ .

# Poker

Atributy dialogových proslovů (množiny  $A_x$ )

Dialogové  
systémy

Luděk Bártek

- volba –  $AX_1 \in \{call, rise, fold\}$
- sebedůvěra proslovu –  $AX_2 \in \{high, medium, low\}$

# Poker

## Fragment vnitřních stavů

Dialogové  
systémy

Luděk Bártek

Stav	$AS_1$	$AS_2$	$AS_3$	$AS_4$	$AS_5$
$S_1$	medium	high	medium	medium	careful
$S_2$	medium	high	high	medium	risky
$S_3$	high	high	medium	medium	risky
$S_4$	high	high	medium	medium	bluffing
$S_5$	bad	low	medium	low	risky
$S_6$	bad	low	low	medium	careful
$S_7$	good	medium	medium	medium	risky
$S_8$	good	medium	low	good	careful

# Poker

## Fragment promluv dialogu

Dialogové  
systémy

Luděk Bártek

Promluva	$AX_1$	$AX_2$	$AX_3$	$AX_4$
$x_1$	raise	high	low	high
$x_2$	raise	high	high	low
$x_3$	call	low	high	medium
$x_4$	raise	high	medium	low
$x_5$	fold	low	high	low
$x_6$	call	low	high	medium

- Zkušený hráč:

$$\delta(S_1, X_1) = S_2, \lambda(S_1, X_1) = X_2$$

$$\delta(S_3, X_1) = S_4, \lambda(S_3, X_1) = X_3$$

- Začátečník:

$$\delta(S_5, X_4) = S_2, \lambda(S_5, X_4) = X_5$$

$$\delta(S_7, X_4) = S_8, \lambda(S_7, X_4) = X_6$$



# Dialog a strategické hry

- Dialog lze považovat za strategickou hrou.
- Strategická hra obsahuje množinu hráčů.
- Každý hráč má množinu akcí (strategií).
- Každý hráč má preferenční relaci (výplatní funkci (payoff function))

# Strategické hry

## Vězňovo dilema

- Strategická hra dvou hráčů.
- Předpokládá, že každý hráč se stará především o svůj prospěch.
- Simuluje vyšetřování zločinu, ze kterého jsou podezřelí dva lidé.
- Pravidla:
  - 1 Pokud oba vězni mlčí, jsou oba odsouzeni, za jiný, menší zločin, ke kratšímu trestu (např. 2 roky).
  - 2 Pokud jeden mlčí a druhý se přizná, ten který se přiznal, je osvobozen a ten, který mlčel je odsouzen k maximálnímu trestu (10 let)..
  - 3 Pokud se oba přiznají, oba jsou odsouzeni k polovičnímu trestu (5 let).
- Vězňovo dilema – jak se zachová parták?

# Přístup jestřáb-holubice

Dialogové  
systémy

Luděk Bártek

	holubice	jestřáb
holubice	5, 5	1, 10
jestřáb	10, 1	0, 0

# Strategické hry s podobnou výplatní funkcí

- Válka pohlaví
  - Manželé preferují společně strávený čas. Co budeme dělat dnes odpoledne? Půjdeme (budeme se dívat) na módní přehlídku nebo na fotbal?
- Hlava – Orel
  - Dva lidé se sázejí, co padne na minci. Hlava nebo orel?

# Iterované strategické hry

- Hráče necháme opakovaně hrát strategickou hru.
- Vzniká extenzivní hra s dokonalou informací.
  - Extenzivní hra – opakující se hra.
  - Dokonalá informace – znáte předchozí tahy všech hráčů.
- Příklady:
  - iterované vězňovo dilema
  - iterovaná hra "Válka pohlaví"

# Iterované věžňovo dilema

## Strategie

- Vždy spolupracovat – vždy se přiznat bez ohledu na to, jak se dříve zachoval spoluhráč (naivní mírotvorce).
- Vždy zradit – vždy zapírat.
- Tit-for-tat – spolupracovat nebo nespolupracovat, podle toho zda spoluobviněný minule spolupracoval/nespolupracoval.
- Zlomyslná – spolupracuje, dokud spoluobviněný nezradí. Potom vždy zradí.
- Mistrust – nejdřív zradí, a potom vždy opakuje tah spoluobviněného.
- Pavlov – spolupracuje pouze tehdy, pokud si spoluobviněný zvolil stejnou možnost v předchozím kole.
- Hard Tit-for-tat – spolupracuje, pokud spoluobviněný nezradil v žádném z dvou předchozích kol.
- Náhodná – spolupracuje s pravděpodobností 0.5.

# Iterované prostorové věžňovo dilema

- Objekty (buňky, jednotlivci, hráči) hrají iterované věžňovo dilema se svými sousedy.
- Po skončení hry objekty v závislosti na výsledku mění strategii pro další kolo.
- Chování kolonie strategií je nerozhodnutelný problém.
  - Důkaz – P. Grim 1994.

# Aplikace strategických her

Dialogové  
systémy

Luděk Bártek

- Ekonomie – chování trhu, zákazníků, ...
- Sociologie
- Psychologie
- Politika
  - strategická rozhodnutí v zájmu státu.
- Ekologie
- ...



# Generování dialogových rozhraní z korpusů

- Na základě korpusu dialogů lze automaticky generovat dialogový korpus, který danému korpusu „vyhovuje“.
- Postup:
  - 1 Vytvoříme korpus metodou Wizard of Oz (pracuje pouze „čaroděj“).
  - 2 Odstraníme konflikty a převedeme korpus na dialogové rozhraní.
  - 3 Kombinovaně vytvoříme nový korpus („čaroděj“ se snaží maximálně využívat navržené dialogové rozhraní).
  - 4 Odstraníme konflikty a vygenerujeme z korpusu nové dialogové rozhraní.
  - 5 Pokud je rozhraní v pořádku, končíme, jinak pokračujeme krokem 3.

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

# Generování promluv dialogovým systémem

- Generátor promluv získá od dialogového manažeru informace, které mají být sděleny uživateli.
- Generátor promluv musí vytvoří korektní větu v jazyce, který je použit pro komunikaci s uživatelem:
  - doplnění prezentovaných informací do předem připravených rámcových promluv
  - vygenerováním promluv ze sémantické reprezentace promluvy.

- Cíl – převod psaného textu na mluvenou řeč.
  - Výsledná řeč by měla znít co nejpřirozeněji.
- Přirozená řeč by měla obsahovat:
  - správnou intonaci
  - správné umístění přízvuků
    - slovní
    - větný
  - korektní koartikulaci
  - správný rytmus (časování)
  - ...

- Syntéza ve frekvenční oblasti – simuluje chování řečového ústrojí.
- Syntéza v časové oblasti – spojování řečových segmentů do větších celků (věta, promluva, ...)
- Korpusová – varianta syntézy v časové oblasti – jako databáze řečových segmentů slouží řečový korpus.
- Problémově orientovaná syntéza:
  - varianta syntézy v časové oblasti
  - využívá větší celky – věty, ...
  - příklady:
    - hlášení nádražního rozhlasu
    - automatizované linky telefonické podpory
    - ...

- 1 Fonetický přepis textu.
- 2 Syntéza foneticky přepsaného textu:
  - Syntéza ve frekvenční oblasti – volba průběhu parametrů syntézy (F0/generátor šumu, vyšší harmonické frekvence, jejich intenzita, ...)
  - Syntéza v časové oblasti – výběr vhodných segmentů a jejich spojení.
- 3 Případný postprocessing:
  - doplnění intonace
  - doplnění přízvuků
  - ...

- Slouží k přesnému, jednoznačnému zápisu mluvené řeči.
- Využívá fonetickou abecedu:
  - mezinárodní fonetická abeceda (IPA) – součást standardu UNICODE
  - SAMPA (Speech Assessment Method Phonetic Alphabet)
    - sedmibitový přepis IPA
    - navržena v 80. letech
    - používá se v různých TTS
    - příklad:

tSeSTina je krAsnl jazyk

- ...

# IPA

## Ukázka

Dialogové  
systémy

Luděk Bártek

### 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	ɸ β	t d			ʈ ɖ	c ɟ	k ɡ	q ɢ			
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	ħ ʕ	h ɦ
Approximant		ʋ	ɹ			ɻ	j	ɰ				
Trill	ʙ		r						ʀ		ʀ	
Tap, Flap		ⱱ	ɾ			ɽ						
Lateral fricative			ɬ ɮ		ɮ	ɬ	ɬ	ɮ				
Lateral approximant			l		ɭ	ɭ	ɭ	ɭ				
Lateral flap			ɺ		ɻ	ɻ	ɻ					

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.



- Nelze si pamatovat fonetické přepisy všech promluv:
  - Nutno zabezpečit automatický přepis.
- Pravidla fonetického přepisu:
  - mohou mít regionální charakter.
  - Příklad – výslovnost na shledanou v ČR:
    - Čechy – naschledanou
    - Morava – nazhledanou.
    - Obě varianty jsou spisovné.
  - Obecně přepis nemusí používat všechny znaky dané abecedy (i/y = i, c = ts, ...)
- Zohledňuje koartikulaci (spodobu znělosti).

# Pravidla fonetického přepisu češtiny

- ch → x , ů → ú, w → v, q → kv, y → i, ý → í
- ě:
  - bě → bje, pě → pje, fě → fje, vě → vje
  - dě → d'e, tě → t'e, ně → ňe, mě → m'ě
- i/í:
  - di/í → d'i/í, ti/í → t'i/í, ni/í → ňi/í
- X:
  - x → ks — začátek slova před samohláskou, mezi samohláskami nebo před neznělou souhláskou a nebo na konci slova.
  - x → gz:
    - *exsamohláska*
    - před znělou souhláskou

# Změny při spojování souhlásek

- Dochází k nim při spojování souhlásek.
- Způsobeny přenastavováním mluvidel.
- Dva druhy:
  - spodoba znělosti – změna znělosti párových souhlásek:
    - ZPS → ↯ ZPS: dub → dup, zpěv → spjev
    - NPS → ↯ NPS: sběr → zbjer, když → gdiš
  - spodoba artikulační – při spojení dvou souhlásek s různou artikulací:
    - nk/ng – banka, tango
    - mv/mf – tramvaj, nymfa
    - nt'/nd – punťa, pindík
    - dň – odpovědně, sto dní, vodní
    - ts → c
    - tš → č
    - ds → c
    - dš → č

# Syntéza řeči ve frekvenční oblasti

- Simuluje tvorbu hlasu v řečových orgánech.
- Uchovává se:
  - frekvenční charakteristika hlasu použitého pro syntézu
  - parametry buzení
- Princip:
  - Emulace hlasových orgánů s využitím:
    - frekvenčních generátorů
    - filtrů
    - zesilovače (zesilovačů).
  - Tyto komponenty jsou ovládány parametry modelu.
- Využívají se dva způsoby kódování zdroje:
  - řečová syntéza formantového typu
  - LPC řečová syntéza.

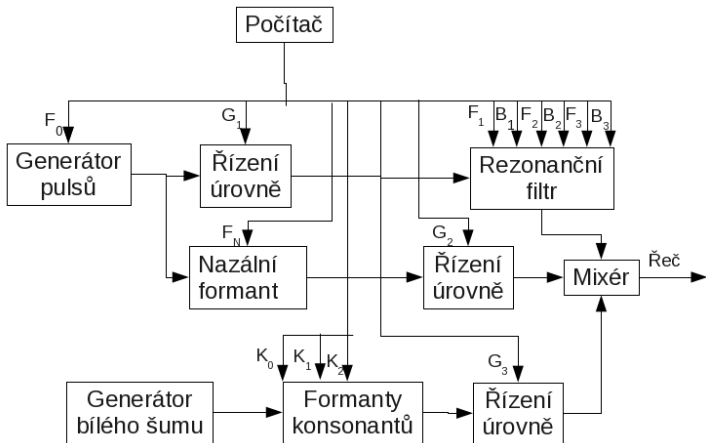
# Řečová syntéza formantového typu

- Rekonstruuje formanty hlasového traktu pomocí sériových a paralelních spojení několika rezonančních obvodů.
- Jejich frekvence a šířky pásma jsou ovládány elektronicky.
- Parametry syntetizéru:
  - $F_0$  – základní frekvence
  - $F_i$  – formanty
  - $F_N$  – nazální formant
  - $B_i$  – pásmové filtry pro  $F_i$
  - $G_i$  – parametry řízení zisku/zesílení
  - $K_i$  – formanty pro konsonanty

# Schéma sériového formantového syntetizéru

Dialogové  
systémy

Luděk Bártek



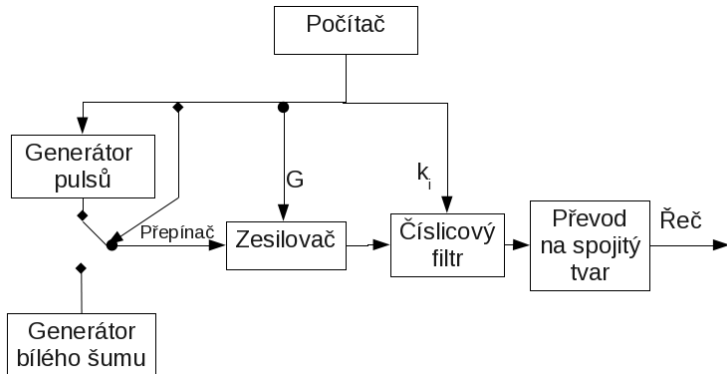
Obrázek: Blokové schéma sériového formantového syntetizéru

- Charakteristiky pro LPC syntetizér:
  - perioda základního hlasivkového tónu  $T_0$
  - charakteristika hlásky – znělá/neznělá
  - amplituda budícího signálu  $G$
  - koeficienty číslicového filtru.
- Způsob získání koeficientu číslicového filtru:
  - vrcholy v LPC spektrální obálce analyzovaného mikrosegmentu
  - kořeny charakteristické rovnice zdrojového filtru
  - reflexní koeficienty.

# Schéma LPC syntetizéru

Dialogové  
systémy

Luděk Bártek



Obrázek: Blokové schéma LPC syntetizéru



- Výhody a nevýhody syntézy ve frekvenční oblasti:
  - + Malé paměťové nároky – model použitého mluvčího.
  - + Syntézu lze realizovat hardwarově.
    - Hlas bývá méně přirozený oproti syntéze v časové oblasti.
      - Problém přesnosti matematického modelu.
    - Softwarová syntéza ve frekvenční oblasti bývá výpočetně náročnější než syntéza v časové oblasti.
- Obvyklé využití:
  - doplnění syntézy v časové oblasti o:
    - větnou intonaci
    - větný a slovní přízvuk
    - další prozodické faktory.
  - Občas pro syntézu na zařízeních, která nedisponují dostatečnou kapacitou paměti (mobilní telefony, PDA, ...).
- Více viz např. J. Psutka – Komunikace s počítačem mluvenou řečí.

# Syntéza v časové oblasti

- Cíl – převod obecného textu na řeč.
- Postavena na spojování segmentů řeči.
- Využívají se různé délky základních segmentů:
  - Větší:
    - lze lépe modelovat prozodické charakteristiky řeči
    - větší paměťové nároky – větší množství segmentů (potenciálně až  $2^n$ , kde  $n$  je délka segmentu).
    - příklady segmentů – slova, části vět, věty, ...
  - Menší:
    - horší možnost modelování prozodických jevů (větná intonace, přízvuky, ...)
    - menší paměťové nároky – menší množství menších segmentů.

# Používané řečové segmenty

- Alofóny:
  - poziční varianty fonémů – obsahují
    - foném
    - okolí ovlivněné koartikulací.
  - počet alofónů –  $n^3$  ( $n$  – počet fonémů).
- Difóny:
  - začínají uprostřed jednoho fonému a končí uprostřed následujícího fonému
  - počet difónů –  $n^2$
  - často využívané pro syntézu i pro rozpoznávání (např. syntetizér MBrola)

# Používané řečové segmenty

- Trifóny:
  - Začínají uprostřed levého sousedního fonému a končí uprostřed pravého sousedního fonému.
  - Počet –  $n^3$ .
  - Často používané pro rozpoznávání a syntézu řeči.
- Slabičné segmenty:
  - Snaha, aby co nejvíce odpovídaly slabikám.
  - Délka – 1 — 3 fonémy.
  - Využívá se např. v TTS systému Demosthénés.

- Slabika:
  - Slabikovat se učí už děti v první třídě.
  - Nejmenší organizační jednotka řeči.
  - Nelze odvodit strukturu slabik – nejednoznačnost dělení některých slov na slabiky:
    - funk-ční vs. funkč-ní
  - Počet slabik – uvádí se cca 10 000.

- Struktura slabiky:
  - preatura (onset)
  - nukleus (vokalické jádro slabiky) – bývá to:
    - samohláska resp. dvojhláska
    - sonora – např. *krk*
    - frikativa – např. *pst*
    - nasála – např. *sedm*
  - koda – nemusí se vyskytovat
  - nukleus + koda jsou považovány za základ slabiky
  - svahy:
    - preatura a koda
    - jedná se většinou o jednu nebo více souhlásek.

# Syntéza v časovém pásmu

## 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 – tma
  - KKVK – dmout
- Tyto segmenty tvoří více než 95 % slabik.
- Umožňují automatickou segmentaci textu.
- Používají se např. v syntetizéru Demosthénes (doc. Kopeček, LSD FI)

- 1 Fonetický přepis.
- 2 Segmentace textu dle použitých řečových segmentů.
- 3 Výběr odpovídajících akustických segmentů z db segmentů.
- 4 Spojení segmentů
  - Nutné, aby bylo možné spojitě hladké navázání segmentů:
    - shodné nebo velmi blízké hodnoty konce a začátku po sobě jdoucích segmentů
    - shodné nebo velmi blízké hodnoty 1. derivace navazujících konců obou segmentů.
- 5 Případný postprocessing
  - doplnění prozodie.
  - ...



- Konkatenativní syntéza v časové oblasti.
- Jako databází segmentů využívá řečový korpus.
  - Obsahuje označovanou mluvenou řeč.
  - Značkování obsahuje:
    - fonetický přepis dané řeči
    - hranice řečových segmentů
    - průběh  $F_0$  a případně i dalších formantů.
  - Umožňuje přesnější výběr řečových segmentů:
    - snižuje výpočetní složitost spojování a postprocessingu.
- Algoritmus výběru segmentů:
  - 1 Výběr odpovídajícího segmentu podle požadovaného obsahu.
  - 2 Pokud je segmentů více zvolí se z nich ten, který nejlépe navazuje.

# Syntéza v časové oblasti

## Syntéza na bázi rámců

- Většinou se jedná o problémově orientovanou syntézu.
- Syntéza se skládá z:
  - rámců – neměnicí se části vět
  - slotů – měnicí se částí promluvy.

- Výhody:

- Rámce jsou dopředu namluveny a mohou obsahovat intonaci.
- Syntetizuje se pouze obsah slotů:
  - velmi dobře specifikovaná množina slov
  - lze použít celá slova.

- Příklad:

- hlášení nádražního rozhlasu:

Osobní vlak číslo *číslo vlaku* ze směru *stanice* přijede  
k *číslo nástupiště* nástupišti v *čas*.

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

- Cíle dodatečného zpracování – obohatit syntetizovanou řeč o:
  - intonaci
  - přízvuky (větný, slovní)
  - důrazy
  - přestávky.
- Prostředky – modifikace:
  - F0, případně dalších formantů
  - lokální modifikace větné melodie.
  - intenzity – amplitudy.

- Výstup syntézy je monotóní řeč bez intonace a přízvuku – zní nepřirozeně.
- Náprava – doplnění prozódie.
- Základní prozodické prvky:
  - výška řeči
  - hlasitost
  - doba trvání.
- Základním nositelem prozódie v běžné řeči je slabika.
- Prozódie závisí na typu věty:
  - oznamovací, tázací zjišťovací, rozkazovací – klesající intonace
  - otázka doplňovací (odpověď ano/ne) – rostoucí intonace.
- Modelování prozódie – modulace F0.

# Ukázky větné intonace

Dialogové  
systémy

Luděk Bártek

Postprocessing  
Prozódie

Standards pro  
syntézu řeči

SABLE  
SSML

- Originální promluva bez intonace
- Oznamovací věta
- Otázka zjišťovací

# Výška základního tónu

Dialogové  
systémy

Luděk Bártek

Postprocessing  
Prozodie

Standards pro  
syntézu řeči

SABLE  
SSML

- Výška základního hlasivkového tónu odpovídá formantu F0.
- Průběh F0 na vokalickém jádru bývá nelieární.
- Změna intonace není pouhou změnou F0 – nutno modifikovat i vyšší formanty.
- Na základě důležitosti F0 se jazyky dělí na:
  - tónové (číntina, vietnamština, ...) – čínské slovo –ma– v závislosti na průběhu F0 může znamenat:
    - matka
    - konopí
    - kůň
    - nadávat
  - jazyky s melodickým přízvukem (srbština, slovinština, litevština, norština, švédština, ...)

- **Intenzita (hlasitost):**
  - fyzikální pohled – intenzita signálu v daném časovém okamžiku
  - fyziologický pohled – reakce vnitřního ucha (coortiho 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 délku trvání v různém kontextu.
  - Drobné odchylky mohou být i ve stejném kontextu.
  - Typická doba trvání slabiky – 50 — 200 milisekund.



### ■ Kvalita hlasu

- chvění hlasu (jitter)
- nepravidelné výchylky v amplitudě F0 (shimmer)
- zbarvení tónu
- ochraptělost
- míra 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).

### ■ Pauza

- tichá
- vyplněná – obsahuje nějaký charakteristický zvuk:
  - eeh
  - áá
  - éé
  - ...

### ■ Zaváhání

- Přímo vypovídá o pragmatice projevu.
- Důležitý např. pro modifikaci dialogové strategie u dialogových systémů.
- Typický případ informace obsažené zejména v prozodické vrstvě jazyka.

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

### ■ Intonace

- nejobecněji – časový průběh časového spektra hlasu
- za určující pro melodii se považuje základní hlasová frekvence
  - časová závislost základní hlasové frekvence
  - lze zobrazit grafem v závislosti na čase
- související terminologie:
  - melodie – průběh F0
  - kadence – určena např. důrazem, ...
  - intonační kadence
  - melodém – základní melodického průběhu určený na základě jeho gramatické funkce.
  - průběh F0

- Emotivní zbarvení hlasu
  - Projevuje se rychlými změnami hlasitosti a základní frekvence.
  - Často přesahují hranici věty.
  - Jeho detekce u DS umožňuje zvolit vhodnou dialogovou strategii.
- Emfatický přízvuk
  - Vytvářen emotivním zbarvením hlasu.
  - Vyskytuje se např. ve větách pronesených v situacích s výrazným emocionálním kontextem:

To je tedy opravdu *neslýchané*.  
Bolí to jak čert.

- Kontrastní přízvuk – snaha o zdůraznění slova nebo slabiky v kontrastu s jiným slovem nebo slabikou:

„Řekl jsem do *Šakvic* ne *Rakvice*.“

„*Byte* ne *bit*.“

### ■ 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.

### ■ Výplňkové části

- Kromě výplňkové funkce mohou charakterizovat:
  - styl mluvčího:

„Byl jsi včera na akci, *vid’?*”

- nářečí resp. slang:

„*Vole*, ta včerejší spářka byla ale hustá, co *vole?*”

### ■ Přerušení

#### ■ častý jev v mluvené řeči na úrovni:

- vyšších celků (výpoěd'/promluva, věta, prozodická fráze, ...)
- uvnitř slov.

#### ■ Mívá návaznost na další prozodické prvky:

- zaváhání
- opakování
- vyplněnou pauzu
- ...



- Korekce částí promluvy
  - častý jev a to vzhledem k různým částem.
  - Příčiny vzniku:
    - důsledek přeroknutí
    - upřesnění části promluvy
    - oprava předchozí části promluvy.
  - Často následuje přerušení nebo další prozodické jevy.

- Promluva.
- Prozodická fráze
  - Skupina slov vytvářející jednotný intonační celek.
  - Představuje základní, z prozodického hlediska kompaktní, strukturu.
  - Členění do prozodických frází souvisí ve velké míře 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

Dialogové  
systémy

Luděk Bártek

Postprocessing  
Prozodie

Standardy pro  
syntézu řeči

SABLE  
SSML

- 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

- Otevřený standard pro prozodické značkování textu.
- Vývoj započat v 2. polovině 90. let
- aplikace XML/SGML
- snaha o zkombinování 3. značkovacích jazyků pro syntézu řeči:
  - SSML – Speech Synthesis Markup Language (W3C, 1999).
  - STML – Spoken Text Markup Language (CSTR Edinburgh University, Lucent Technologies, 1997)
  - JSML – Java Synthesis Markup Language (Sun Microsystems, 2000)

- SABLE – kořenová značka
- DIV
  - Slouží k členění dokumentu na odstavce a věty.
  - Typ části dokumentu určuje atribut type.

```
<DIV TYPE="paragraph" > ... </DIV>
```

- prozodické značky:
  - EMPH – zdůraznění části promluvy
  - PITCH – výška promluvy
  - VOLUME – úroveň hlasitosti
  - RATE – rychlost
  - BREAK – pauza

- Popis mluvčího:
  - element SPEAKER:
    - AGE – věk mluvčího (older, middle, younger, teen, child)
    - GENDER – pohlaví (male, female)
    - NAME – jméno mluvčího, závislé na TTS – TTS musí daného mluvčího znát.
- Fonetické:
  - PRON – foneticky přepsaná promluva, lze použít IPA.
  - SAYAS – způsob fonetického přepisu (datum, telefon, url, poštovní adresa, . . .)
  - LANGUAGE – jazyk promluvy.

# SABLE

## Ukázka

Dialogové  
systémy

Luděk Bártek

Postprocessing  
Prozodie

Standardy pro  
syntézu řeči

SABLE  
SSML

```
<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>
```

- Otevřený standard W3C
- Vývoj započat koncem 90. let.
- Aplikace XML.
- Součást rodiny W3C Voice Browser Activity
- Aktuální verze 1.0 (září 2004)



- kořenový element speak
- strukturní elementy:
  - p – odstavec
  - s – věta
- fonetické:
  - say-as – způsob fonetického přepisu.
    - typ textu (telefon, URI, číslo, ...)
  - phoneme – fonetický přepis dané promluvy
  - sub – substituce – např. přepis zkratek, ...
- popis hlasu:
  - voice – popis hlasu, kterým se má text přečíst (pohlaví, věk, ...)

- Prozodické značkování:
  - emphasis – zdůraznění části promluvy
  - break – pauza
  - prosody – ovlivňuje základní prozodické jevy:
    - vlastnost dána atributem – pitch, rate, duration, volume
- Další viz specifikace

```
<speak version="1.0"
  xmlns="http://www.w3.org/2001/10/synthesis"
  xml:lang="en-US">
  <voice gender="male" age="18">
    <p>
      <prosody rate="1">I don't</prosody>
      <break time="1s"/>
      <prosody rate="0" pitch="x-low">speak Japanese.
    </p>
  </voice>
</speak>
```

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

- 1876 – udělen patent na telefon A. G. Bellovi
- WWW
  - 1989 – po CERN koluje článek HyperText and Cern (Tim Burnes Lee) –
  - Vánoce 1990 – demonstrován řádkový webový prohlížeč a editor.
  - 1991 – všeobecná dostupnost WWW na počítačích v CERN.
  - 1994 – první setkání W3 konsorcia.

# W3C Voice Browser Activity

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity

SRGS

XML Formát  
SRGS

ABNF formát  
SISR

- 1999 – založena W3C Voice Browser Working Group.
- Cíl – návrh standardů umožňujících přístup k Webu pomocí hlasu a telefonu.
- Členové:
  - HP
  - Nuance Communications
  - Lucent Technologies
  - Motorola
  - ScanSoft
  - IBM
  - Tellme Networks
  - Vocalocity
  - ...

# Standardy W3C Voice Browser Activity

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity

SRGS

XML Formát  
SRGS

ABNF formát  
SISR

- VoiceXML – jazyk pro popis dialogových strategií.
- Speech Recognition Grammar Specification – jazyk pro zápis gramatik pro podporu rozpoznávání řeči.
- Semantic Interpretation for Speech Recognition – jazyk pro podporu sémantické interpretace.
- Speech Synthesis Markup Language – jazyk pro popis prozodických charakteristik pro syntézu řeči.

# Standardy W3C Voice Browser Activity

pokračování

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity

SRGS

XML Formát  
SRGS

ABNF formát  
SISR

- Pronunciation Lexicon Specification – popis výslovnosti pro rozpoznávání a syntézu řeči.
- Call Control XML – jazyk pro popis řízení telefonního spojení uživatele a systému.
- State Chart XML – jazyk pro popis obecně použitelných stavových automatů.



# Standardy W3C Voice Browser Activity

## Zpracování

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity

SRGS

XML Formát  
SRGS

ABNF formát  
SISR

- Standardy jsou značkovací jazyky – nutná interpretace
- Existuje řada platforem:
  - volně dostupné desktopové– JVoiceXML, PublicVoiceXML, ...
  - komerční desktopové – Optimtalk – dříve existovala volně dostupná verze; laboratoř LSD má zakoupenou licenci na laboratorní stroje.
  - volně dostupné on-online - Asterisk+VoiceGlue resp. OpenVXI, ...
  - komerční on-line – Voxeo Prophecy, Bevocal Cafe – lze vyzkoušet a omezeně používat on-line (max. 2 paralelní hovory).

# Speech Recognition Grammar Specification

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

- W3C specifikace jazyka pro zápis bezkontextových gramatik pro podporu rozpoznávání řeči.
- Aktuální verze 1.0.
- Nahradil původně používaný standard JSGF
- Existují dvě varianty zápisu:
  - XML
  - Augmented Backus-Naur Form (ABNF).
- Liší se pouze způsob zápisu nikoliv vyjadřovací síla.
- Možnost použitého zápisu závisí na použité platformě.
  - širší podpora pro XML formát

- Gramatika  $G = (N, \Sigma, P, S)$ 
  - $N$  – konečná množina neterminálních symbolů
  - $\Sigma$  – konečná množina terminálních symbolů (abeceda jazyka)
  - $P$  – množina pravidel
  - $S$  – kořenový neterminální symbol
- Bezkontextová gramatika:
  - gramatika  $G = (N, \Sigma, P, S)$
  - pravidla ve tvaru  $N \rightarrow \{N \cup \Sigma\}^*$

- Začíná XML prologem
  - `<?xml version="1.0" encoding="..."?>`.
- Kořenový element – *grammar*; obsahuje množinu pravidel (elementů *rule*).
- Atributy:
  - *version* – použitá verze standardu SRGS (aktuálně 1.0).
  - *xml:lang* – kód jazyka gramatiky.
  - *root* – id pravidla odpovídajícího kořenovému neterminálu.
  - *mode* – pro jaký způsob komunikace je gramatika určena:
    - *dtmf* – pomocí DTMF kódů
    - *voice* – hlasově; implicitní hodnota.

- Element *rule*:
  - atributy:
    - *id* – identifikátor pravidla (odpovídá neterminálnímu symbolu na levé straně pravidla).
  - Obsah – pravá strana pravidla:
    - textový obsah – posloupnost terminálních symbolů
    - element *ruleref* – neterminální symbol; odkazovaný pomocí atributu *uri*.
    - element *one-of* – varianty (operátor |).
    - element *item* – logické členění sekvence; umožňuje např. uvést počet opakování dané části promluvy.

- Posloupnost terminálních a neterminálních symbolů.

*SAMPLE* → Mám rád *TYP* formát SRGS.

```
<rule id="sample">
```

```
  Mám rád <ruleref uri="#typ"/> formát SRGS.
```

```
</rule>
```

- Lze ji rozdělit na logické části:

```
<rule id="address">
```

```
  <item repeat="0-1">
```

```
    <ruleref uri="#server"/>
```

```
  </item>
```

```
  <item repeat="1-5">
```

```
    tečka <ruleref uri="#domena"/>
```

```
  </item>
```

```
</rule>
```

# Varianty

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

- Element *one-of*.
- Umožňuje specifikovat různé varianty očekávaných vstupů.
- Jednotlivé varianty jsou ohraničeny elementem *item*.
- Příklad:

```
<rule id="barvy">  
  <one-of>  
    <item>červená</item>  
    <item>zelená</item>  
    <item>modrá</item>  
  </one-of>  
</rule>
```

# Opakování

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity

SRGS  
XML Formát  
SRGS  
ABNF formát  
SISR

- Umožňuje specifikaci:
  - nepovinných částí promluvy
  - opakujících se částí promluvy
- Zápis – pomocí atributu *repeat* u elementu *item*.
- Možnosti počtů opakování:

- $n$  krát –  $n$ :

```
<item repeat="2">opakování</item>
```

- $\langle m, n \rangle$  krát –  $m-n$

```
<item repeat="0-1">  
  Chtěl bych  
</item>
```

- $\langle n, \infty \rangle$  krát –  $m-$

```
<item repeat="1-">Ahoj</item>
```



# Zvláštní pravidla

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát  
SRGS

ABNF formát  
SISR

- Slouží k zadání:
  - libovolné nespécifikované promluvy – *GARBAGE*
  - nevyslovitelného pravidla (zakázání určité promluvy) – *VOID*
  - vždy platného pravidla (i prázdného) – *NULL*
- Používají se jako zvláštní neterminální symboly:

```
<ruleref special="GARBAGE"/>
```

- Příklad použití:

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

# ABNF formát SRGS

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát  
SRGS

ABNF formát  
SISR

- Čistě textový formát gramatiky vycházející z tradičního formátu BNF.

```
<spojeni> ::= Chci jet <kdy> <cim>  
              z <odkud> do <kam>".
```

```
<kdy> ::= <den> <cas>
```

```
<cim> ::= vlakem | autobusem
```

```
...
```

- BNF podobný formát využívá dále např. JSGF

# Struktura ABNF zápisu SRGS

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

- Hlavička gramatiky – může obsahovat:
  - specifikaci jazyka gramatiky
  - režim gramatiky – voice/dtmf
  - kořenový neterminál
  - ...
- Pravidla gramatiky
  - formát –  $\$neterminál = (neterminál|terminál)^*$
  - *neterminál* = identifikátor pravidla u XML formátu.

# Struktura hlavičky ABNF zápisu SRGS

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity  
SRGS  
XML Formát  
SRGS  
ABNF formát  
SISR

- Začíná identifikací typu dokumentu.
  - #ABNF verze\_SRGS kódování\_gramatiky  
#ABNF 1.0 ISO-8859-2
- Následuje:
  - specifikace kořenového neterminálu – root \$*neterminál*;
  - jazyk gramatiky – language *kód jazyka*;  
language en-US;
  - režim použitelnosti gramatiky – mode (voice|dtmf);

# ABNF zápis pravidel gramatiky

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

- Sekvence – sekvence terminálních a neterminálních symbolů oddělených mezerou:

```
$pozdrav = dobrý den;  
$datum = $den $mesic $rok;
```

- Varianty – příslušné sekvence terminálních a neterminálních symbolů oddělené symbolem '|':

```
$dopravni_prostredok = autobus | vlak;
```

- Opakování:

- volitelné části – uzavřeny do '[' ]'
- m—n – <m-n>

# Ukázka SRGS gramatiky v ABNF notaci

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

```
#ABNF 1.0 UTF-8;
```

```
root $url;
```

```
language cs-CZ;
```

```
mode voice;
```

```
$url = [<protokol>][server] tečka
```

```
($domena tečka)<1-3>
```

```
tečka $tld[$cesta];
```

```
$protokol = http | ftp | telnet | gopher | ...;
```

```
$cesta = / $adresar <1-> /[$soubor];
```

```
...
```

# Semantic Interpretation for Speech Recognition

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity  
SRGS  
XML Formát  
SRGS  
ABNF formát  
SISR

- Sémantika – přiřazuje význam tvrzením.
- Sémantika v dialogových systémech:
  - přiřazuje interpretaci promluvám a jejich částem
  - umožňuje získání relevantních údajů.
- SISR – standard z rodiny W3C Voice Browser Activity
  - slouží k sémantické interpretaci promluv
  - publikován v dubnu 2007
  - aktuální verze 1.0.
  - Je úzce spjat se standardy:
    - ECMA Script – vyhodnocování interpretace používá výrazy jazyka ECMA Script
    - SRGS – vyhodnocování je pomocí atributů přiřazeno gramatice pro rozpoznávání promluvy.
    - JSON – interpretace je vnitřně reprezentována pomocí objektů ve formátu JSON.

# Přiřazení interpretace části promluvy

- Sémantická interpretace bývá součástí pravidel SRGS.
- Přiřazení interpretace k pravidlu – pomocí „tag“:
  - XML formát SRGS:

- element tag:

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

- atribut tag:

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

- ABNF formát SRGS:

- interpretace uvedena za interpretovanou částí promluvy.
- tvar:  $\{interpretace\}$

```
$potvrzení = $souhlas {ano} | $nesouhlas {ne}
```



# Odvozování interpretace na základě dílčích interpretací

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

- Zápis odvození – pomocí výrazů v jazyce ECMAScript.
- Přiřazení pravidel pro odvození k pravidlům gramatiky – pomocí atributu/elementu tag.
- Výsledná interpretace reprezentována pomocí objektů ve formátu JSON.
- Vyhodnocování promluv:
  - přístup k dílčím interpretacím – interpretace neterminálních symbolů na pravé straně:
    - atributy stínové proměnné rules
    - neterminálu  $N$  odpovídá atribut  $N$ .
  - vrácení výsledné interpretace z pravidla do nadřazeného pravidla – objekt out.
  - vrácení interpretace do dialogu:
    - atributy objektu out
    - vstupnímu poli  $N$  odpovídá atribut  $N$ .

# Vyhodnocování promluv

## XML formát

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

```
<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>
...
```

# Vyhodnocování promluv

## ABNF Formát

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity

SRGS  
XML Formát  
SRGS  
ABNF formát  
SISR

```
$vlastnictvi = Mám $barva <0-1> $prostredek  
{  
    out = rules.barva + ";" + rules.prostredek;  
};
```

# Přiřazení interpretace vstupním polím

## XML Formát

Dialogové  
systémy

Luděk Bártek

W3C Voice

Browser

Activity

SRGS

XML Formát

SRGS

ABNF formát

SISR

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

# Přiřazení interpretace vstupním polím

## ABNF formát

Dialogové  
systémy

Luděk Bártek

W3C Voice  
Browser  
Activity

SRGS  
XML Formát  
SRGS  
ABNF formát  
SISR

```
$vlastnictvi = mam $barva <0-1> $prostredek  
{  
  out.barva = rules.barva;  
  out.prostredek = rules.prostredek;  
};
```

- Specifikace SRGS
- Specifikace SISR
- Specifikace ECMAScript
- Specifikace JSON

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

- Jazyk pro popis dialogových rozhraní.
- Součást standardů W3C Voice Browser Activity.
- Cíl:
  - přinést výhody webového vývoje a doručování obsahu do interaktivních hlasových aplikací.
- Historie:
  - 1995 – započat vývoj značkovacího jazyka AT&T Phone Markup Language.
  - 1998 – konference pořádaná W3C na téma hlasového procházení webu:
    - předvedeny jazyky PML, VoxML, SpeechML, TalkML, VoiceHTML, ...



- Historie (pokračování):
  - 1999 – založeno VoiceXML Forum – cílem je spojení sil při vývoji jazyka pro značkování dialogů.
  - 2000:
    - vydána specifikace VoiceXML 1.0
    - specifikace VoiceXML 1.0 přijata jako standard W3C.
- Aktuální stav:
  - doporučení – VoiceXML 2.1. (červen 2007)
  - draft – VoiceXML 3.0 (srpen 2010)

# VoiceXML

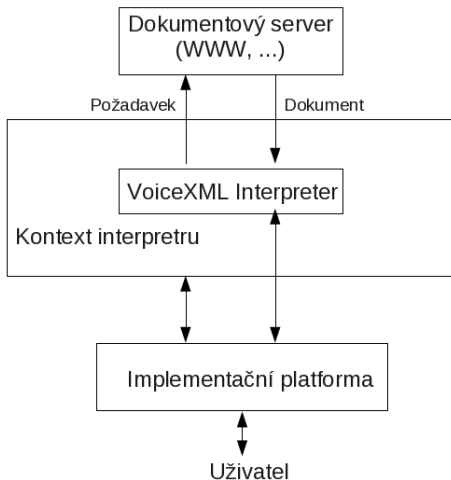
## Model komunikace

Dialogové  
systémy

Luděk Bártek

VoiceXML

Zápis dialogů  
pomocí  
VoiceXML



- VoiceXML dokument(y):
  - Skládají se z formulářů.
  - Uživatel se v daném okamžiku nachází v jednom z konverzačních stavů.
  - Přechody mezi stavy definovány pomocí URI – odkazují na další krok dialogu.
  - Dialog končí, pokud tento přechod není definován.
- VoiceXML definuje dva druhy dialogů:
  - Formuláře – definuje proces nutný pro získání hodnot sady položek.
  - Menu – poskytuje uživateli sadu možností a odkazů na pokračování dialogu.

- **Subdialogy:**
  - Obdoba funkcí v procedurálním programu.
  - Slouží k opětovnému provádění jisté části dialogu (např. zjištění e-mailové adresy).
  - Realizovány jako formuláře, kterým mohou být předány parametry, a které mohou vrace hodnotu (viz dále).
- **Sezení:**
  - Začíná v okamžiku zahájení komunikace s VoiceXML interpreterem.
  - Končí:
    - na přání uživatele (např. ukončení spojení, žádost o ukončení interpretace, ...)
    - VoiceXML dokumentem – není definován další přechod, předání dat k dalšímu zpracování, ...
- **Aplikace – sada dokumentů, které sdílejí kořenový dokument.**

- *voicexml* – kořenový element každého dokument.
- Musí obsahovat atributy:
  - *version* – použitá verze standardu VoiceXML
    - aktuální 2.1
    - hodnota závisí na použité platformě – OptimTalk 1.9 – 2.1, JVoiceXML – zatím neúplná podpora verze 2.1, VoiceGlue – podpora 2.0 + některé možnosti z 2.1, ...
  - *xmlns* – deklarace implicitního jmeného prostoru. Hodnota musí být <http://www.w3.org/2001/vxml>.
  - *xml:lang* – hodnotou je kód jazyka, pro který je dialogové rozhraní navrženo.
- Element obsahuje:
  - jeden nebo více elementů form,
  - element menu,
  - ...

- Jeden ze základních elementů VoiceXML dokumentů.
- Ohraničen značkami `< form >` a `< /form >`.
- Obsahuje:
  - sadu vstupních polí
  - deklarace proměnných daného formuláře – element `var`
  - definice gramatik platných v daném formuláři
  - bloky výkoného kódu:
  - ...
- Atributy:
  - `id` – povinný atribut:
    - slouží jako název daného formuláře
    - jeho hodnota musí být unikátní v daném dokumentu
    - lze použít k předávání řízení do daného formuláře.

- Formuláře jsou interpretovány implicitním algoritmem pro interpretaci formulářů (Form Interpretation Algorithm, FIA):
  - 1 Přehraj všechny výzvy, které jsou potomky tohoto elementu form.
  - 2 Dokud existuje vstupní pole formuláře s nedefinovanou hodnotou:
    - 1 Vyber 1. vhodný nezadaný vstup.
    - 2 Přehraj všechny výzvy, které se váží k danému poli.
    - 3 Získej hodnotu vstupu daného vstupního pole nebo zpracuj vyvolanou událost (help, nomatch, ...)
    - 4 Zpracuj část filled daného vstupního pole.

- FIA může dále skončit pokud:
  - pokud se má provést přesměrování hovoru (např. element goto)
  - pokud má dojít k předání dat dokumentovému serveru (element submit)
  - pokud je explicitně požadováno ukončení (element exit).



# Formulář

## Ukázka

Dialogové  
systémy

Luděk Bártek

VoiceXML  
Zápis dialogů  
pomocí  
VoiceXML

```
<vxml version="2.0"
      xmlns="http://www.w3.org/2001/vxml"
      xml:lang="en-US">
  <form id="hello">
    <prompt>
      Hello world!
      This is our first VoiceXML form.
    </prompt>
  </form>
</vxml>
```

- Vstupní pole - odpovídají různým možnostem zadání vstupních položek formuláře:
  - field – vstup od uživatele, možnost zadání hlasem nebo pomocí DTMF.
  - record – slouží k nahrání zprávy od uživatele.
  - subdialog – slouží k vyvolání dialogu řešícího dílčí problém, např. zadání adresy, ...
- Řídící položky:
  - block – příkazový blok, lze využít např. k různým výstupům pro uživatele, vyhodocování vstupních dat, ...
  - initial – iniciální část formuláře. Využívá se hlavně v dialogových rozhraních se smíšenou strategií.
  - transfer – přesměrování uživatele na novou lokaci (aplikaci, telefonního operátora, ...)
  - object – slouží ke zpřístupnění funkcionality, která může být závislá na platformě (dll, JSP+ servlet, ...)

# Formulář

Vstupní pole a řídicí struktury – ukázka užití

Dialogové  
systémy

Luděk Bártek

VoiceXML  
Zápis dialogů  
pomocí  
VoiceXML

```
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxm
xml:lang="cs-CZ">
<form id="hello">
  <block name="hello">
    <prompt>Welcome to the VoiceXML!.</prompt>
  </block>
  <field name="greeting">
    <prompt>Hello.</prompt>
    <grammar src="greetings.grxml"/>
    <noinput>
      <prompt>Tell mi something nice, like hello, hi,
        good day.</prompt>
    </noinput>
```

# Formulář

Vstupní pole a řídicí struktury – ukázka užití – pokračování

Dialogové  
systémy

Luděk Bártek

VoiceXML  
Zápis dialogů  
pomocí  
VoiceXML

```
<nomatch>
  <prompt> I didn't understand you, but thanks anyw
</prompt>
  <exit/>
</nomatch>
<noinput count="2">
  <prompt> When you don't want to speak to me good
  bye.</prompt>
  <exit/>
</noinput>
</field>
<filled>
  <prompt> you said <value expr="greeting"/></prompt>
  <submit src="SomeURI" namelist="greeting"/>
</filled>
</form>
```

- Představuje vstup od uživatele. Může být zadán buď hlasem nebo pomocí DTMF.
- Atributy:
  - name – jméno pole. Používá se k přístupu k zadané hodnotě (pomocí stínové proměnné se shodným jménem).
  - expr – výraz v jazyce ECMAScript, který slouží k inicializaci hodnoty vstupního pole.
  - cond – vstupní podmínka nutná pro zpracování vstupního pole.
  - Více viz specifikace.

### ■ Obsah elementu:

- Výzva s popisem požadované hodnoty (element prompt).
- Gramatika – SRGS gramatika s popisem akceptovaných vstupů.
- Ošetření událostí:
  - noinput – nebyl zadán žádný vstup
  - nomatch – zadán neakceptovaný vstup (vstup neodpovídá gramatice)
  - filled – umožňuje zpracování vstupu po vyplnění vstupního pole
  - ...

# Element field

Ukázka užití

Dialogové  
systémy

Luděk Bártek

VoiceXML

Zápis dialogů  
pomocí  
VoiceXML

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.0" xmlns="...">
  <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</nomatch>
    </field>
    <filled>
      <submit next="applicationURI" namelist="name"/>
    </filled>
  </form>
</vxml>
```

# Element record

- Umožňuje systému nahrát zprávu od uživatele.
- Lze využít např. pro tvorbu hlasového záznamníku.
- Atributy:
  - name – název vstupního pole
  - expr – viz field
  - cond – viz field
  - beep – má-li být začátek nahrávání být signalizován zvukovým signálem
  - maxtime – maximální délka nahrávky
  - type – mime-type výsledné nahrávky, musí být podporována VoiceXML platformou
  - ...



- Obsah elementu:
  - Výzva/výzvy s popisem požadovaného požadovaného vstupu.
  - Ošetření událostí:
    - noinput – uživatel zprávu nezačal nahrávat.
    - connection.disconnect.hangup – uživatel zavěsil.

# Element record

Ukázka užití

Dialogové  
systémy

Luděk Bártek

VoiceXML  
Zápis dialogů  
pomocí  
VoiceXML

```
<?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 dílčího dialogu (dialogu řešícího dílčí problém).
- Jeden a tentýž subdialog se dá volat opakovaně.
- Vyvolání subdialogu:
  - element subdialog – vlastní volání subdialogu.
  - Obsahuje:
    - param – definice hodnoty parametru.
    - filled – kód, který se má provést po návratu z dílčího dialogu.
  - Atributy:
    - name – jméno volaného subdialogu.
    - src – URI dokumentu s kódem subdialogu.
- Kód subdialogu:
  - formulář
  - ukončený elementem return.

# Element subdialog

Ukázka užití

Dialogové  
systémy

Luděk Bártek

VoiceXML  
Zápis dialogů  
pomocí  
VoiceXML

```
<?xml version="1.0" encoding="utf-8"?>
<vxml version="2.0" xmlns="..." xml:lang="cs-CZ">
  <form id="demo">
    <block>
      <prompt>Ukázka použití subdialogu ve VoiceXML
    </prompt>
    </block>
    <subdialog name="greeting" src="#say_hello">
      <param name="param1" expr="'ahoj'"/>
      <filled>
        <prompt>Hodnota subdialogu je <value
expr="greeting.great"/></prompt>
      </filled>
    </subdialog>
```

# Element subdialog

Ukázka užití – pokračování

Dialogové  
systémy

Luděk Bártek

VoiceXML  
Zápis dialogů  
pomocí  
VoiceXML

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

# Element subdialog

Ukázka užití – pokračování

Dialogové  
systémy

Luděk Bártek

VoiceXML  
Zápis dialogů  
pomocí  
VoiceXML

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

- Obsahuje proveditelný obsah.
  - atributy:
    - name – název bloku.
    - expr – iniciální hodnota proměnné formuláře.
    - cond – podmínka omezující provádění bloku.
  - struktura – shodná s obsahem elementu filled:
    - řídicí struktury – elementy if, else, elseif
    - přiřazovací příkaz – element assign, clear, ...
    - příkazy skoku – element goto, exit, return, ...

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011



# Řízení toku dialogu

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup  
Procedurální  
přístup  
Nastavení  
vlastnosti  
interpretace

- VoiceXML poskytuje dva způsoby řízení průchodu dialogem:
  - Funkcionální:
    - Založen na FIA.
    - Omezení provádění daného bloku/pole – atribut cond, přiřazení hodnoty danému vstupnímu poli, . . .
    - Použití elementu initial – iniciální část dialogu, většinou je iniciativa na straně uživatele.
  - Procedurální:
    - Využívají se klasické řídicí struktury – if/then/else/elseif, goto.

# Funkcionální přístup řízení průběhu dialogu

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Průběh dialogu řízen FIA – opakovaně se vybírá další vhodný nevyplněný element.
- Ke změně pořadí vyhodnocování vstupních polí a bloků lze využít:
  - změny hodnot vstupních polí:
    - přiřazení – (nepřejeme si toto pole provádět
    - smazání – dané pole bude znovu vyplněno; lze využít pro opakované zadávání nějaké hodnoty (např. koníček, kód předmětu při zápisu, ...)
  - atribut cond:
    - Obsahuje podmínku, která musí být splněna, aby se vstup/blok provedl.
    - Podmínka zapsána v jazyce ECMAScript.

# Přiřazení hodnoty

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Element *assign*
  - Má atributy:
    - name – název proměnné/vstupního pole
    - expr – obsahuje výraz (může být i konstantní), jehož hodnota se má proměnné přiřadit.
  - Proměnná musí být před použitím definována – stínová proměnná vstupního pole; pomocí elementu var.

```
<var name="uri" expr=""/>
```

```
...
```

```
<assign name="uri" expr="protokol +':://' +  
server +'/ ' + cesta"/>
```

- V gramatice – pomocí atributů objektu out, více viz přednáška o SISR.

# Přiřazení hodnoty

## Dokončení

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastnosti  
interpretace

- Pomocí kódu v jazyce ECMAScript.
- Smazání hodnoty proměnné – element *clear*:
  - seznam proměnných uveden v atributu *namelist*
  - názvy jsou odděleny mezerou.

```
<clear namelist="protokol server cesta"/>
```

# Atribut *cond*

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Obsahuje podmínku, která musí být splněna, aby se daný element zpracoval.
- Podmínka zapsána jako logický výraz v jazyce ECMAScript:
  - Syntakticky hodné resp. velmi blízké podmínkám v jazycích C/Java/JavaScript.
  - Operandy relačních operátorů jsou:
    - konstanty/konstantní výrazy
    - proměnné/stínové proměnné z dokumentu.

```
<field name="zakonceni" cond="predmet!=''">  
  <prompt> Zadejte zakončení předmětu  
    <value expr="predmet"/>.  
  </prompt>  
  <grammar src="zakonceni.grxml"/>  
  ...  
</field>
```

# Element *initial*

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastnosti  
interpretace

- Typicky se využívá u dialogů se smíšenou iniciativou.
- Slouží k zadání informací platných v rozsahu celého dialogu.
- Využívá gramatiku formuláře – obsažena jako dceřiný uzel elementu form – většinou jako 1.

# Element *initial*

Ukázka použití

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

```
<?xml version="1.0" encoding="utf-8"?>
<vxml version="2.0"
      xmlns="http://www.w3.org/2001/vxml">
  <form id="registrace">
    <grammar src="registration.grxml"
            type="application/srgs+xml"/>
    <initial name="regfull">
      <prompt>
        Zadejte, který předmět si přejete
        zaregistrovat a s jakým ukončením.
      </prompt>
```

# Element *initial*

Ukázka použití

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

```
<nomatch count="2">
  Bohužel Vám nerozumím. Zkusíme to jinak.
  <assign name="regfull" expr="true"/>
  <reprompt/>
</nomatch>
</initial>
<field name="predmet">
  <prompt>Název nebo kód registrovaného
    předmětu</prompt>
  <grammar src="registration.grxml#predmet"
    type="application/xml+srgs"/>
</field>
```



# Element *initial*

## Ukázka použití

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

```
<field name="ukonceni">
  <prompt>Způsob ukončení předmětu.</prompt>
  <help>Jedna z možností zkouška, kolokvium,
    zápočet.</help>
  <grammar src="registration.grxml#ukonceni"
    type="application/xml+srgs"/>
</field>
<field name="potvrzeni">
  <prompt>Přejete si zaregistrovat
    <value expr="predmet"/> s ukončením
    <value expr="ukonceni"/>?
  </prompt>
  <grammar src="yesno.grxml"
    type="application/xml+srgs"/>
</field>
```

# Element *initial*

Ukázka použití

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

```
<filled>
  <block cond="potvrzeni=='y'">
    <submit
      next="https://voiceis.muni.cz/auth/reg/"
      namelist="predmet ukonceni"/>
  </block>
  <block cond="potvrzeni=='n'">
    <clear
      namelist="predmet ukonceni potvrzeni"/>
  </block>
</filled>
</form>
</vxml>
```

# Řízení průběhu dialogu

## Procedurální přístup

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Umožňuje psát dialogová rozhraní způsobem, který je obvyklý u strukturovaného programování.
- Řídící struktury:
  - podmíněné provádění
  - skoky.
- Podmíněný příkaz:
  - elementy if, elseif, else
- Příkazy skoku:
  - elementy goto, submit, exit, return, subdialog.

# Podmíněný příkaz

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Umožňuje vykonání určité části dialogu za předem stanovených podmínek.
- Realizován pomocí elementů:
  - if
  - else
  - elseif
- Elementy if a elseif mají atribut cond – obsah shodný s atributem cond u funkcionálního přístupu.
- Elementy else a elseif musí být potomky elementu if.
- Element if může být na místech, kde je povolen proveditelný obsah:
  - element block
  - element filled

# Podmíněný příkaz

## Příklad

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

```
<if cond="color=='red'">  
  <assign name="barva" expr="'r'"/>  
<elseif cond="color=='green'">  
  <assign name="barva" expr="'g'"/>  
<elseif cond="color=='blue'">  
  <assign name="barva" expr="'b'"/>  
<else/>  
  <assign name="barva" expr="'u'"/>  
</if>
```

# Příkazy skoku

goto

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Příkaz slouží k přechodu na:
  - jinou položku daného formuláře
  - jiný formulář (dialog) v daném dokumentu
  - jiný dokument v dané aplikaci.
- Element – *goto*
- Atributy:
  - *nextitem* – umožňuje předat řízení jinému vstupnímu poli daného formuláře. Hodnotou je řetězcová konstanta s URI vstupního pole.
  - *expritem* – umožňuje předat řízení jinému vstupnímu poli daného formuláře. Hodnotou může být libovolný výraz v jazyce ECMAScript, jehož výsledkem je URI vstupního pole.
  - *next* – URI formuláře, kterým se má pokračovat.
  - *expr* – výraz v jazyce ECMAScript, který se vyhodnotí jako URI formuláře, kterým se má pokračovat.

# Příkazy skoku

## submit

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Příkaz slouží k odeslání získaných hodnot ke zpracování serverové části aplikace a v pokračování dokumentem, který je vrácen jako odpověď.
- Element – *submit*
  - atributy *next*, *expr* – URI aplikace, která má data zpracovat, a která vrací pokračování dialogu.
  - atribut *method* – obsahuje jednu z hodnot *get* a *post*. Značí zda se má se serverem komunikovat pomocí HTTP metody *post* a nebo *get*.

# Příkazy skoku

exit, return, subdialog

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Element *exit*
  - Slouží k ukončení aplikace.
  - může obsahovat atributy:
    - *expr* – návratová hodnota dialogu.
    - *namelist* – seznam proměnných, které se mají vrátit interpretu.
- Element *subdialog* – slouží k vyvolání dílčího dialogu (více viz minulá přednáška).
- Element *return* – slouží k návratu ze subdialogu (více viz minulá přednáška).



# Nastavování vlastností interpretace properties

Dialogové  
systémy

Luděk Bártek

Řízení toku  
dialogu

Funkcionální  
přístup

Procedurální  
přístup

Nastavení  
vlastností  
interpretace

- Nastavování parametrů interpretace – pomocí vlastností (properties).
- Element *property*:
  - Slouží k nastavování vlastností interpretace.
  - Atributy:
    - name – jméno nastavované vlastnosti
    - value – hodnota, která se má nastavit.
- Typy vlastností:
  - platformně závislé vlastnosti
  - obecné vlastnosti rozpoznávání řeči
  - obecné vlastnosti DTMF
  - vlastnosti výzev
  - vlastnosti komunikace po síti
  - ostatní vlastnosti.

- Platformě závislé vlastnosti:
  - Vnášejí do aplikací nekompatibilitu – dané vlastnosti jsou dostupné pouze na některé platformě.
  - Platformě závislé vlastnosti by měly používat reverzní název domény autora.
- Obecné vlastnosti rozpoznávání řeči:
  - Vlastnosti jsou převážně přebrány z Java Speech API.
  - Patří sem např. `confidencelevel` – nastavuje minimální hodnotu důvěry ve výsledek vrácený rozpoznávačem, aby byl vstup akceptován.
- Obecné vlastnosti DTMF:
  - *interdigittimeout* – maximální délka prodlevy při zadávání série číslic pomocí DTMF.
  - *termtimeout* – maximální prodleva, po které je ukončeno zadávání vstupu pomocí DTMF.
  - ...

- Vlastnosti výzev a vstupů:
  - *timeout* – po jak dlouhé době se má generovat událost noinput.
  - *bargein* – jestli má skončit předčítání promptu v okamžiku, kdy je detekován vstup od uživatele.
  - *bergeintype* – jaký typ vstupu od uživatele má přerušit předčítání výzvy (speech, dtmf).
- Vlastnosti načítání obsahu:
  - Slouží k nastavování různých vlastností, které se vztahují k načítání obsahu (dokument, audio, gramatika, skript, ...).
  - Patří sem např. *fetchtimeout* – doba, po které skončí načítání chybou.
- a další.

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

# Speech Synthesis Markup Language

Dialogové  
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- Značkovací jazyk, určený pro zvýšení kvality syntézy řeči.
- Standard W3C.
- Aktuální verze 1.1 (září 2010).
- Vychází ze specifikací JSGF a JSML (Sun Microsystems).
- Vychází z něj jazyk SABLE.

- Vytvořit standard pro značkování prozodických jevů mluvené řeči.
- Jazyk by měl být podporován různými TTS.
- Zvýšení kvality syntézy řeči pomocí ovládnání:
  - výslovnosti
  - hlasitosti
  - průběhu základního hlasivkového tónu
  - rychlosti
  - ...

- Kořenový element – *speak*.
- Atributy:
  - *version* – použitá verze SSML (aktálně 1.0, 1.1)
  - *xml:lang* – přirozený jazyk použitý obsahem tohoto elementu.
- Může obsahovat elementy:
  - výslovnost – *lexicon*, *phoneme*, *say-as*
  - struktura – *p*, *s*
  - prozodie – *emphasis*, *prosody*, *voice*, *break*
  - ostatní – *audio*, *meta*, *metadata*, ...

- Element *p*:
  - Ohraničuje odstavec.
  - Atribut – *xml:lang* – přirozený jazyk tohoto odstavce.
  - Může obsahovat elementy:
    - audio, break, emphasis, mark, phoneme, prosody, say-as, sub, s, voice.
- Element *s*:
  - Ohraničuje větu.
  - Atribut – *xml:lang*.
  - Může obsahovat elementy:
    - audio, break, emphasis, mark, phoneme, prosody, say-as, sub, voice.



- Element *lexicon*:
  - Vkládá odkaz na lexikon výslovnosti (více viz ).
  - Atributy:
    - *uri* – URI odkazující na soubor s lexikonem výslovnosti.
    - *type* – mime typ odpovídající typu lexikonu.
- Element *phoneme*:
  - Obsahuje fonetický přepis textu.
  - Atributy:
    - *alphabet* – použitá fonetická abeceda (ipa, případně ještě x-JEITA, x-JEITA-2000 – japonské fonetické abecedy, většinou znaková využívá znakovou sadu UNICODE).
    - *ph* – fonetický přepis textu uzavřeného do tohoto elementu.

- Element *say-as*
  - Popisuje jakým způsobem se má daný text vyslovovat (datum, množství peněz, ...).
  - Atributy:
    - *interpret-as* – o jaký typ dat se jedná (currency, date, ...)
- Element *sub*:
  - Umožňuje definovat aliasy pro daný text (např. přepis zkratek, ...).
  - Atributy:
    - *alias* – alias pro text, který je obsahem daného elementu.

- Umožňuje popsat prozodické vlastnosti promluvy počítače.
- Do jaké míry budou obsaženy ve výsledné řeči závisí na podpoře v konkrétním TTS.
- *voice* – umožňuje ovlivňovat některé charakteristiky použitého hlasu:
  - pohlaví – atribut *gender* – povolené hodnoty male, female, neutral
  - věk – atribut *age* – kladné celé číslo udávající věk mluvčího.
  - variantu – atribut *variant* – kladné celé číslo, které značí která varianta daného hlasu se má použít – musí být podpora v TTS
  - jazyk – atribut *xml:lang* – pokud je dostupný použije se tento jazyk, jinak by se měl použít jiný, co nejbližší jazyk.

- Element *emphasis*
  - daný text by se měl říct s důrazem – pomocí přízvuku, hlasitosti, . . .
  - míra důrazu popsána atributem *level* – hodnoty jsou none, reduced, moderate, strong.
- Element *break*
  - výsledkem by měla být pauza v řeči
  - její síla (výraznost) je ovlivněna atributem *strength* – jedna z hodnot none, x-weak, weak, medium, strong, x-strong
  - doba trvání atributem *length* – čas ve formátu shodným s formátem použitým ve specifikaci CSS2.

- Element *prosody* – umožňuje ovlivňovat prozodické charakteristiky promluvy, která je jeho obsahem. Je nutná podpora na straně TTS::
  - $F_0$  (atribut *pitch*) – hodnota může udávat výšku v Hz, relativní změnu a nebo některou z hodnot x-low, low, medium, high, x-high a nebo default.
  - Průběh  $F_0$  (atribut *contour*) – hodnotou jsou mezerou oddělené uspořádané dvojice (time, pitch), kde time je vyjádřen pomocí procentuálně a výška stejným způsobem jako u atributu pitch.
  - Rozsah  $F_0$  na daném úseku (atribut *range*) – hodnota buď rozsah v Hz, nebo relativní rozsah a nebo jedna z hodnot x-low, low, medium, high, x-high a default.
  - Doba trvání (atribut *duration*) – jak dlouho se má daný text číst (ms resp. s).
  - Hlasitost (atribut *volume*) – hlasitost proslovu – hodnoty v intervalu 0.0 – 100.0 nebo jedna z silent (=0.0), x-soft, soft, medium, loud, x-loud a nebo default (=100.0).

# Pronunciation Lexicon Specification

Dialogové  
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- Standard W3C VoiceBrowser Activity.
- Aktuální verze 1.0 (říjen 2008).
- Popisuje jazyk pro tvorbu lexikonů výslovnosti použitelných při syntéze a rozpoznávání řeči.
  - výslovnost cizích slov
  - výslovnost zkratk
  - ...

- Kořenový element *lexicon*:
  - Atributy:
    - *xml:lang* – přirozený jazyk dokumentu
    - *version* – aktuální verze 1.0
    - *xmlns* – musí být propojen se jmenným prostorem <http://www.w3.org/2005/01/pronunciation-lexicon>
    - *alphabet* – použitá fonetická abeceda.
  - Obsah:
    - Element *metadata* – informace o dokumentu.
    - Element(y) *lexeme* – jednotlivé položky slovníku.
- Element *lexeme*
  - Atribut *role* – popisuje mluvnické kategorie slova, tak aby bylo možné zvolit nejvhodnější výslovnost (např. sloveso vs podstatné jméno – red vs. red)
  - Obsah:
    - Element(y) *grapheme* – psaná podoba slova.
    - Element(y) *phonemes* – výslovnost(i) slova.
    - Element(y) *alias* – v případě, že *grapheme* obsahuje zkratku, tak její plný tvar (např. ČR – Česká republika).

- Element *phoneme*
  - Atribut *preferred* – pokud je u pojmu uvedeno více různých výslovností, tato je preferovaná.
  - Obsah – fonetický zápis výslovnosti pojmu.
- Element *alias*
  - Atribut *preferred* – pokud je u pojmu uvedeno více různých výkladů, toto je preferovaný.
  - Obsah – plný zápis zkratky.
- Více viz specifikace.



# State Chart XML

Dialogové  
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- Návrh standardu W3C (poslední varianta duben 20110)
- Značovací jazyk pro popis konečných automatů používaných v dialogových rozhraních.
- Kandidát na řídicí jazyk v:
  - VoiceXML 3.0 (aktuálně ve vývoji)
  - budoucích verzích CCXML
  - jazyce pro popis multimodálních rozhraní.

- Konečný automat  $(S, \Sigma, \phi, q_0, Q)$ :
  - $S$  – konečná neprázdná množina stavů
  - $\Sigma$  – vstupní abeceda
  - $\phi$  – přechodová funkce  $S \times \Sigma \rightarrow S$
  - $q_0$  – počáteční stav
  - $Q$  – množina koncových stavů.
- Zápis pomocí SCXML:
  - stav – element *state*:
    - povinný atribut *id* – název stavu
    - počáteční stav – obsahuje dceřinný element *initial*
    - koncový stav – obsahuje dceřinný element *final*
  - přechod(y) – pomocí elementu/ů *transition*:
    - atribut *event* – událost, která vyvolá přechod (nepovinný)
    - atribut *target* – identifikátor cílového stavu
- Příklady a podrobnosti viz specifikace.

# Call Control XML

Dialogové  
systémy

Luděk Bártek

SSML

PLS

SCXML

CCXML

- CCXML je navrženo, aby umožnilo ovládat telefonní hovory z dialogových rozhraní popsaných např. pomocí VoiceXML, . . . .
- Umožňuje ovládat hovory na úrovni, která je mimo možnosti VoiceXML:
  - konferenční hovory
  - přiřadit každému hovoru vlastní VoiceXML interpret
  - ovládání odchozích hovorů
  - . . .
- Aktuálně kandidát na doporučení ve verzi 1.0 (duben 2010)

# Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2011

# Multimodální dialogová rozhraní

Dialogové  
systémy

Luděk Bártek

Multimodální  
dialogová  
rozhraní

- Multimodální dialogové rozhraní – mimo mluvenou řeč umožňuje alternativní způsoby komunikace člověk ↔ počítač:
  - textová komunikace
  - grafická komunikace
  - ...
- Výhody:
  - lepší přístupnost – např. pro uživatele s poruchami sluchu/řeči.

# Multimodální komunikace počítač → člověk

Dialogové  
systémy

Luděk Bártek

Multimodální  
dialogová  
rozhraní

- **Textová:**
  - Mimo hlasový výstup je navíc zobrazen odpovídající textový výstup.
  - Lze využít např. prostředky pro IM, SMS, . . . .
- **Grafická:**
  - Talking Heads – mimo hlasový výstup je navíc zobrazena tvář, jejíž pohyby, zejména úst odpovídají řeči (ukázky[1]).
  - Komunikace znakovou řečí
    - Zobrazeny ruce, které se pohybují synchroně s řečí.[2]
    - Často jako doplněk mluvící hlavy nebo formou avatara.

# Multimodální komunikace člověk → počítač

Dialogové  
systémy

Luděk Bártek

Multimodální  
dialogová  
rozhraní

- Velmi široké spektrum možností zadávání vstupu jinak než hlasem:
  - klávesnice (počítač, DTMF, SMS, ...)
  - rukou psaný vstup – dotyková obrazovka + pero
  - různá ústy ovládaná zařízení
  - ovládání pomocí pohybů očí a víček
  - rozpoznávání řeči pomocí sond detekujících činnost svalů a mozku [4] – pomocí EEG, EMG, ...
  - ...
- Často jako doplněk řečového vstupu.

- Proprietární řešení:
  - dříve např. součást CSLU Toolkitu [3]
  - projekt August – viz [5]
- Otevřená řešení:
  - Doporučení W3C týkající se multimodálního přístupu – zatím ve stádiu návrhů, bez implementace.
  - Využívají a propojují i další standardy W3C (CCXML, XHTML,
  - Výstup W3C Multimodal Interaction Working Group (VoiceXML, SVG, SMIL, ...).



# Návrhy standardů Multimodal Interaction WG

Dialogové  
systémy

Luděk Bártek

Multimodální  
dialogová  
rozhraní

- Multimodal Architecture and Interfaces – popis architektury a podporovaných rozhraní.
- EMMA (Extensible MultiModal Annotation Markup Language) – standard pro výměnu údajů mezi jednotlivými komponentami multimodálního dialogového systému.
- InkML – standard pro data zadaná pomocí dotykových obrazovek a elektronického pera.
- EmotionML – slouží k anotování dat z pohledu emocí.

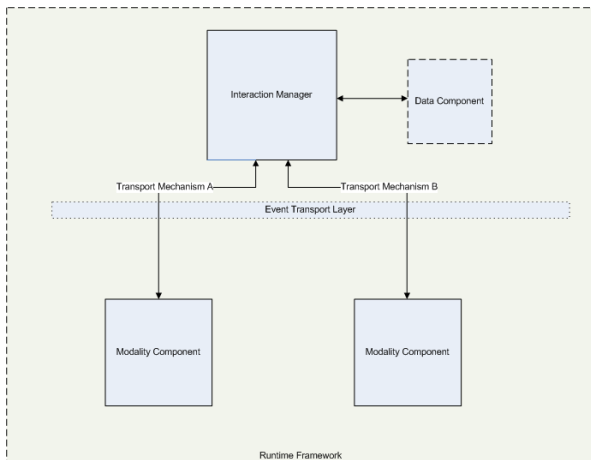
# Multimodal Initiative WG

## Návrh architektury systému

Dialogové  
systémy

Luděk Bártek

Multimodální  
dialogová  
rozhraní



**Obrázek:** Návrh architektury multimodálního systému (převzato ze specifikace W3C)



Massaro, D., Cohen, M. M. – Demos From The Perceptual Science Lab, dostupné na adrese <http://mambo.ucsc.edu/demos.html> (květen 2011).



Guimeraes, K., Antunes, D. R., Guilhermino, de F. Guilhermino, D., Lopes da Silva, R. A., Garcia, L. S. – Structure of the Brazilian Sign Language (Libras) for Computational Tools: Citizenship and Social, in Organizational, Business, and Technological Aspects of the Knowledge Society, CCIS vol. 112, Springer, Heidelberg, 2010, pp. 365 – 370.



Barnard, E. et al – CSLU Toolkit, dostupné na adrese <http://www.cslu.ogi.edu/toolkit/index.html> (květen 2011).



Schultz, T. – Silent and Weak Speech Based on Elektromyography, in Proceedings of 12th International

Conference ICCHP 2010 Part 1, Wien, Springer,  
Heidelberg, pp. 595 – 604, 2010.



the August Home Page,  
<http://www.speech.kth.se/august/> (květen 2011).