Zkouška - PA156 Dialogove systemy

Budou tam 4 otázky na témata:

1. **počítačová akustika a zpracovávání signálu**
2. **digitalizace zvuku(a základy kolem), zpracování zvuku ve frekvenční/časové oblasti**
3. **syntéza rozpoznávání řeči**
4. **dialogové systémy**

1-V jakých mezích se pohybuje úspěšnost nástrojů pro rozpoznávání řeči a jakými prostředky ji lze zvýšit?

Rozpoznavaní řeci: převádí mluvené slovo na text
využívá se:

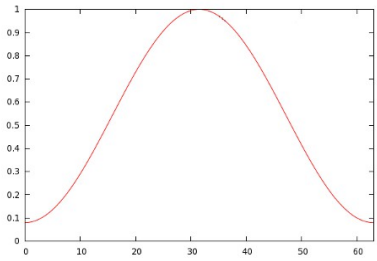
- rozpoznavan plynule reci
- rozpoznavan izolovanych slov

Úspěšnost rozpoznávání řečí se pohybuje cca **50% - 99%** v závislosti na úkolu,jakzyku,...
Lze zvýšit **omezením domény rozpoznávání**:

- rozpoznání tématu
- použití gramatik pro rozpoznání řeči

2-Co je Hammingovo okénko a kdy se používá?

Jedno z nejčastěji používaných váhových okének
Hammingovo okénko
Vychází s předpokladu, že čím je vzorek blíže ke kraji mikrosegmentu, tím více je ovlivněn okolím.
Váha vzorků uvnitř mikrosegmentu $w(sn) = 0.54 - 0.46*\cos((2*\pi*n)/(N-1))$
N - počet vzorků v mikrosegmentu.
Váha vzorků mimo mikrosegment $w(s) = 0$.



3-Jaká znáte váhová okénka a kdy se používají?

Krátkodobá analýza:
zpracování signálu na časovém intervalu, o němž se předpokládá, že na něm nedochází k výraznějším dynamickým změnám.
Tento interval se nazývá mikrosegment (někdy také stručneji segment) a jeho velikost se obvykle od 10 do 40 ms.

Tím že se rozhodneme pro určitou velikost mikrosegmentu, implicitně předpokládáme, že zvukový signál je v okolí okénka periodický s periodou okénka. Chyba, která vzniká nesouladem s tímto předpokladem, může být do jisté míry kompenzována použitím tzv. okénka. Okénko je posloupnost vah pro prvky mikrosegmentu.

Váhová okénka:
Hammingovo okénko: (viz výše)
Pravoúhlé okénko:
přiřadí každému prvku mikrosegmentu jednotkovou váhu, tj. je definováno vztahem

- $w(n) = 1$ pro $n = 0,...,N-1$
- $w(n) = 0$ pro ostatní n (mimo mikrosegment)

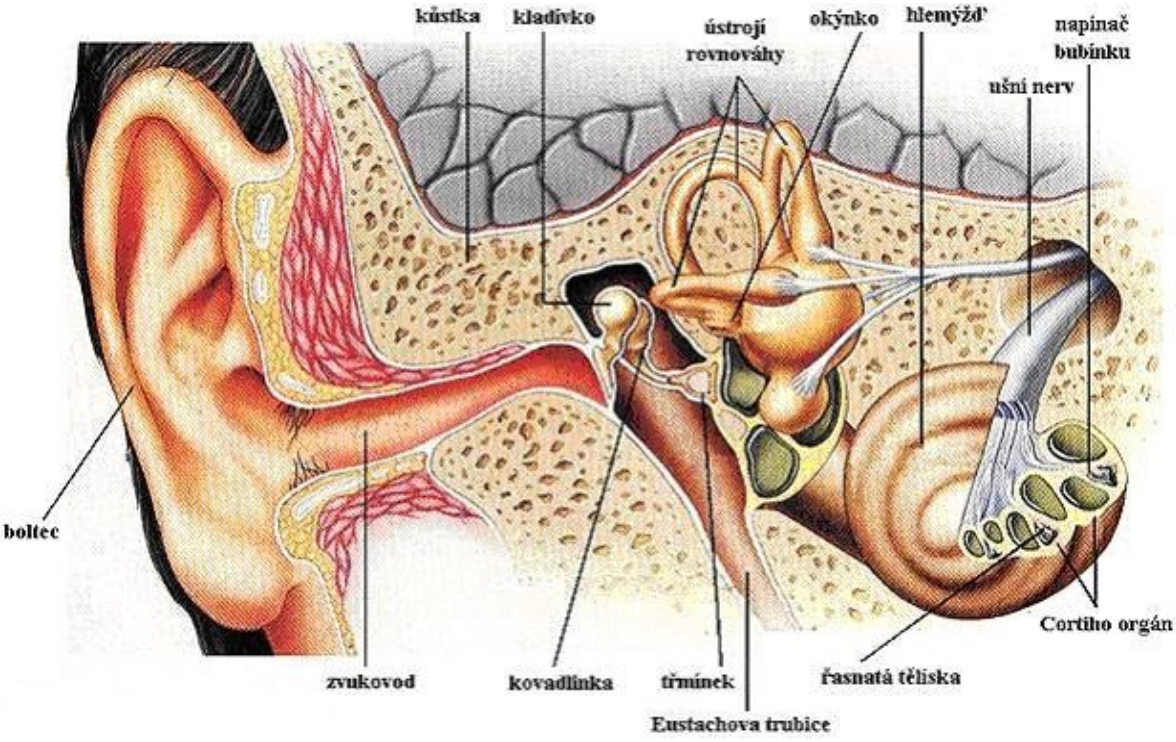
4-Popište mechanismus fyziologického vnímání zvuku.

Mechanizmus vnímání řeči

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 mechanizmy k vyrovnání rozdílů tlaku mezi vnějším prostředím a sluchovým orgánem
- vnitřní ucho - převádí zvukovou energii na vzruchy, které jsou vedeny dále do mozku.



Vnější ucho Obsahuje:

- Ušní boltec - soustřeďuje 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 přenese zvukovou energii na kůstky středního ucha.

Střední ucho Obsahuje:

- Kůstky středního ucha:
  - kladívko - přiléhá k bubínku
  - kovadlinka
  - 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.

Vnitřní ucho

- Hlemýžď (Cochlea):
  - Je naplněn vodnatým roztokem.
  - Ústrojí ve tvaru ulity hlemýždě, které obsahuje Cortiho Ústrojí.
  - **Cortiho Ústrojí** obsahuje zhruba 20000 vlákének s délkami 40 µm - 0,5 mm.
  - Vlákénka jsou jsou napojena na nervová zakončení, která vedou vzruchy do příslušného centra v mozku.
- Rovnovážný orgán.

5-Co mají společného emoce a dialogové systémy?

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 prozódií.

- TTS může modelovat emoce pomocí prozódie.
- Při rozpoznávání lze detekovat emoce pomocí prozódie.

(**Prozódie** - popisuje **zvukové vlastnosti jazyka (přízvuk**, **tón**, **intonace** (melodie), frázování)
Emotivní zabarvení hlasu - Projevuje se rychlymi změnami hlasitosti a zakladni frekvence. Často přesahuji hranici věty a jeho detekce u DS umožňuje zvolit vhodnou dialogovou strategii.
Emfatický přízvuk - Vytvařen emotivnim zbarvenim hlasu.Vyskyt např. ve větach v situacich s vyraznym emocionalnim kontextem: Boli to jak čert.
) (víc viz. Otázka 15)

Zjišťování emocí lze provádět pomocí:

- Změn galvanických vlastností kůže (změna odporu)
- Změn tlaku krve a pulsu
- Změn dýchání
- Změn elektrické aktivity mozku

K detekci emocí využívá:

- kameru
- emoční myš

6-Jak fungují skryté Markovovy modely a k čemu slouží?

Modelování řeči pomocí HMM vychází z následující představy o tvorbě řeči:

- Hlasové ústrojí se v krátkém čase nachází v jedné z konečně mnoha artikulačních konfigurací – generuje hlasový signál.
- Přejde do následující konfigurace.

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

Principy použití pro rozpoznávání

- 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 využívané levo-pravé Markovovy modely:
  - vhodné pro modelování procesů spjatých se vzrůstajícím časem.

Rozhodovací pravidlo při rozpoznávání izolovaného slova
Používá se princip maximální věrohodnosti.
1 Pro slovo O a všechna λ:

- 1 Spočítáme P(O|λ).[Určení pravděpodobnosti promluvy]

2 Jako výsledek vybereme třídu s maximální hodnotou P(O|λ).

Implementace

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

7-Popište PCM (/diferenční PCM/adaptivní PCM), jeho výhody a nevýhody.

WIKI:Pulzně kódová modulace (PCM) je modulační metoda převodu analogového zvukového signálu na signál digitální.Princip PCM spočívá v pravidelném odečítání hodnoty signálu pomocí A/D převodníku a jejím záznamu v binární podobě. Při nižší vzorkovací frekvenci je kvalita záznamu horší. Aby nedocházelo k aliasing, musí být podle Nyquistovy vzorkovací věty (též Shannon-Kotělnikův teorém) vzorkovací frekvence více než dvojnásobná oproti frekvenci zaznamenávaného signálu.

Pulsní kódová modulace:

- Přímo se ukládají získané hodnoty jednotlivých vzorků.
- Nevýhody:
- Relativně pomalé změny zvukového signálu ⇒ relativně malé rozdíly sousedních vzorků.⇒ velká redundance dat
- Řešení - diferenční PCM - ukládají se rozdíly mezi sousedními vzorky.
- V případě příliš velkých změny amplitudy signálu problém s nastavením kvantizačního kroku:
    - příliš velký krok - ztráta informace o částech signálu s malou amplitudou
    - příliš malý krok - přetečení hodnot v částech s velkou amplitudou.
- Řešení Adaptivní PCM - kvantizační krok se určí v závislosti na amplitudě signálu.
- Adaptivní diferenční PCM - ukládá rozdíly mezi sousedními vzorky a kvantizační krok se určuje v závislosti na velikosti změny.

8-Co je Fourierova transformace a jaké jsou podmínky jejího použití? Jaké jsou její modifikace pro použití v reálných dialogových systémech?

WIKI: Fourierova transformace je integrální transformace převádějící signál mezi časově a frekvenčně závislým vyjádřením pomocí harmonických signálů, tj. funkcí a , obecně tedy funkcí komplexní exponenciály. Slouží pro převod signálů z časové oblasti do oblasti frekvenční. Signál může být buď ve spojitém či diskrétním čase.

Podmínky

Získání spektra - Fourierova transformace:

- F(x) musí splňovat Dirichletovy podmínky
    - periodická funkce s periodou T
    - je na daném intervalu po částech spojitá (nejvýše konečný počet bodů nespojitosti 1. druhu)
    - má nejvýše konečný počet extrémů na daném intervalu
    - definována v krajních bodech daného intervalu

Analýza signálu ve frekvenční oblasti

Transformuje digitální řečový signál z časově oblasti do
frekvenční oblasti.
Využívá k tomu nejčastěji Fourierovu transformaci.
Nejčastěji používané druhy:

- krátkodobá Fourierova transformace
- krátkodobá diskrétní Fourierova transformace
- rychlá Fourierova transformace

Ve slajdech má napsané, že pro Fourierovu Transformaci musí mít funkce nejvýše konečně mnoho bodů nespojitosti. U zkoušky vám to ale neuzná, prý že jich musí mít nejvýše spočetně nekonečně mnoho (a pak v ústní části následuje vyptávání na mohutnosti číselných množin, kterých je víc, kterých je spočetně apod.).

9-Jmenujte alespoň 4 z pravidel pro vedení kooperativního dialogu.

Dialogovou komunikaci M = (S1; S2; E1; E2) nazveme:

Kooperativní <=> E1 = E2. Oba účastníci dialogu mají stejný cíl a snaží se spolupracovat.
Nekooperativní <=> E1 /= E2. Cíle obou účastníků dialogu se odlišují.
S nulovým součtem <=> E1 = -E2. Cíle obou účastníků dialogu jsou protichůdné.

Dialogový systém při komunikaci s uživatelem by měl brát ohled na následující aspekty:

- **aspekt informativnosti** – bud přiměřeně informativní (ne míň ani víc než je potřeba)
- **aspekt přesvědčivosti** – neuváděj lži nebo info co nejde dokázat/doložit
- **aspekt způsobu** – uváděj stručné,jednoznačné a explicitní info a udržuj v dialogu pořádek
- **aspekty zdvořilosti, empatie a etiky** – min. nároky na komun. ,max souhlas a max. empatii vůči partnerovi
- **aspekt asymetrie** – stručné info o možnostech systému, srozumitelné info o způsobu interakce se systémem
- **aspekt znalostí a schopností uživatele** – vem v úvahu znalosti uživatele a rozliš noob a zkušeného uživ.
- **aspekt vyjasňování a odstraňování chyb** – v případě selhání inic. Odstranění chyby nebo její vysvětlení

10-Popište jednotlivé fáze digitalizace akustického signálu.

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.

Vzorkování

Vzorkovací frekvence - mela by být minimálne dvojnásobkem nejvyšší frekvence, která je v signálu přítomna, aby bylo možné původní signál bez ztráty informace zrekonstruovat (Shannonův vzorkovací teorem).
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

Kvantizace - Metoda převodu spojitých hodnot na diskrétní.
Princip:

- Pokud hodnota signálu překročí n. násobek kvantizačního kroku je jí přiřazena hodnota n.
- kvantizační krok = rozsah hodnot měřené veličiny/počet diskrétních hodnot
- kvantizační chyba - zaokrouhlovací chyba způsobená velikostí kvantizačního kroku, přímo úmerná velikosti kvantizačního kroku.

Bežne Používané kvantizace:

- zpracování zvuku:
  - 2^8
  - 2^{16}
  - 2^{24}
- zpracování obrazu, . . . navíc
  - 2^{32}

Způsoby kódování průbehu vlny

Příme ukládání hodnot získaných kvantizací – kódování PCM (Pulse-Code Modulation).

- relativne pomale zmeny průbehu zvukoveho signálu – male rozdíly mezi sousedními vzorky.
- Velká redundance dat.
- Problem v případe příliš velkeho rozptylu amplitud v signálu (příliš velký kvantizační krok - příliš velká kvantizační chyba, příliš malý kvantizační krok – přetečení v okamžiku zvetšení amplitudy signálu).
- Diferenční PCM - ukládá se rozdíl mezi sousedními vzorky
- Adaptivní PCM | PCM s promenou velikostí kvantizačního kroku - kvantizační krok se uzpůsobí velikosti amplitudy signálu.

(víc k PCM viz otázka 7.)

11-Popište SIP.

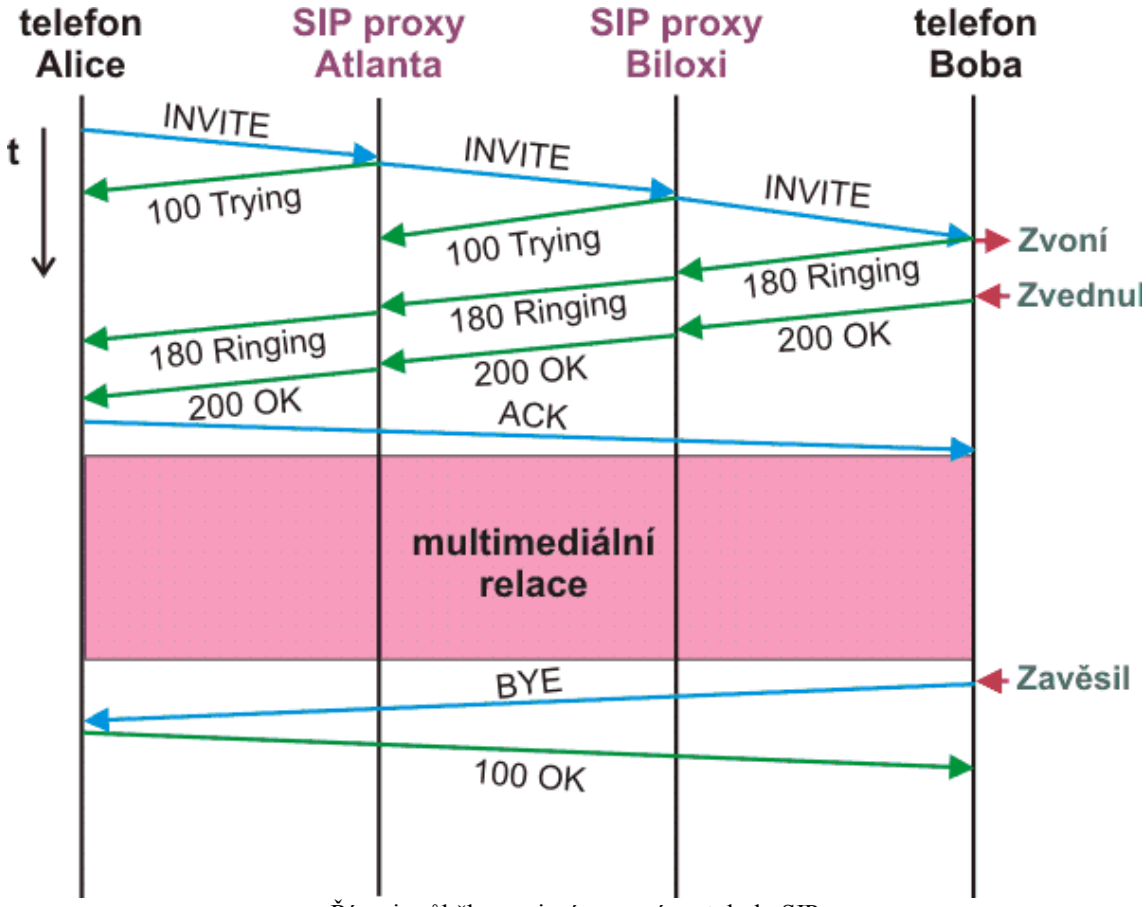
Session Initiation Protocol (SIP)

- 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).
  - . . .

- 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/videa, . . .
- Vlastní navázání spojení - využívá se protokol SDP
  - popisuje navazované spojení,
  - odkazuje na RTP/UDP datový tok, který je využit pro komunikaci účastníků.



Řízeni průběhu spojení pomocí protokolu SIP

12-Jmenujte a popište metody krátkodobé analýzy zvukového signálu ve frekvenční/časové oblasti.

Zpracování digitalizovaného signálu
Úvod
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, ktere jsou následně zpracovány.

Modelování funkce Cortiho ústrojí - pomocí diferenciálních rovnic se simuluje rezonance na určitých vlákenkách Cortiho ústrojí.

Váhove okénko
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 mely odpovídat tomu, jak je daný vzorek
ovlivnen okolím mikrosegmentu.
Nejčastěji Používané typy okének:

- pravoůhle okénko
- Hammingovo okénko

(detaily -> viz. otazka na okénka 2. a 3.)

Analýza digitalizovaného signálu v časové 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
- . . .

Funkce krátkodobé energie
Využívá funkci průměrné energie v rámci segmentu:
Výstupem je průmerná energie v daném okénku.
Použití:

- automaticke oddelení ticha řeči (signálu)
- příznaky v jednoduchých klasifikátorech slov
- oddelení znělých a neznelých částí promluvy.

Funkce krátkodobé intenzity
Funkce intenzity signálu v daném okénku.
Použití - stejne jako funkce krátkodobé energie.
Oproti krátkodobé energii nezvýrazňuje tolik dynamiku řečového signálu.

Krátkodobá funkce středního počtu průchodu nulou
Počítá zmeny znamenka digitalizovaného signálu.
Použití:

- detekce ticha
- detekce začátku a konce i zašumené promluvy
- přibližne určení základního hlasivkoveho tónu a formantů
- příznaky jednodušších klasifikátorů slov

Autokorelační funkce
Vrací podobnost úseků daného mikrosegmentu (čím vetší výsledná hodnota, tím podobnejší úseky posunute o m vzorků).
Použití:
- Používá se k zjišťování periodicity signálu základního tónu řeči.
- Základ pro výpočet koeficientů LPA.

Analýza 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
- . . .

(FT viz. otazka 8.)

Kepstrální analýza
Vychází z modelu činnosti hlasoveho ústrojí.
Kepstrální analýza umožňuje z řeči oddelit parametry buzení a parametry hlasoveho ústrojí.
Využití:
- ocenení foneticke struktury řeči - znelost, perioda
- základního tónu, formanty, . . .
- rozpoznávání slov
- verifikace a identifikace mluvčího
- . . .

Lineární prediktivní analýza
Jedna z nejefektivnejších metod analýzy akustickeho signálu - zajišťuje velmi přesne odhady parametrů při relativne male záteži.
Použití:
- určování spektrálních charakteristik modelu hlasoveho ústrojí
- z chyby predikce lze odvodit poznatky o znelosti a určit frekvenci základního hlasivkoveho tónu
- koeficienty a_i nesou informaci o spektrálních vlastnostech - lze je použít jako příznaky pro rozpoznávání řeči.

14-K čemu slouží dialogová strategie?

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.

13-K čemu slouží DTW? Jak funguje?

Dynamic Time Warping (DTW) - Metoda borcení časové osy

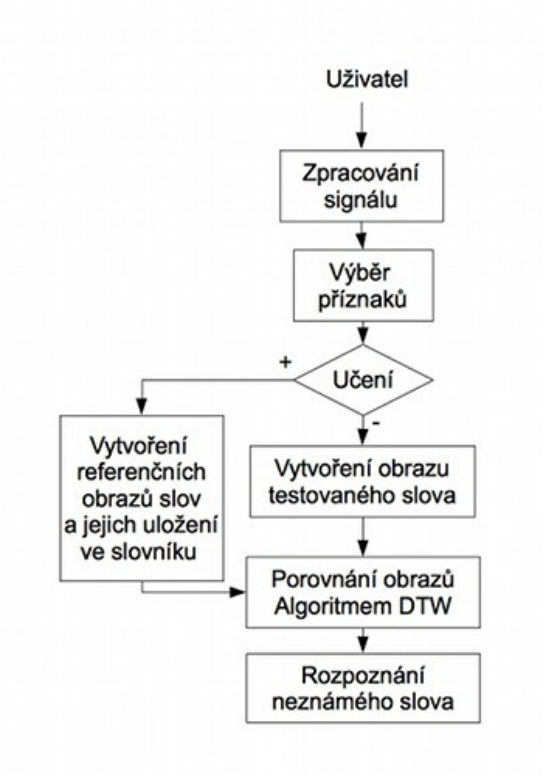
Používá se pro porovnání dvou číselných řad - dvou úseků promluv (dvou slov).
Vstup:

- posloupnost akustických vektorů získaných pomocí metod krátkodobé analýzy signálu
- databáze akustických vektorů rozpoznávaných slov.

Výstup - rozpoznané slovo resp. povel.

Základní postup

- Vytvoříme databázi 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.



Obrazek: Blokove schema klasikatoru slov

15-Co je Prozódie? Které jevy pod ni spadají?

Prozódie - popisuje **zvukové vlastnosti jazyka (výška řeči, h**lasitost, **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.

Prozódické vlastnosti
Intenzita (hlasitost)
***Doba trvání* -** Slabika může mít různou délku trvání v různém kontextu (typicky 50 - 200 milisekund)
***Kvalita hlasu* -**chvění hlasu ,zbarvení tonu, 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

Základní odvozené Prozódické vlastnosti
***Rytmus* -** Prozódický prvek odvozený z dob trvání slabik nebo pauz v daném časovém Úseku
Slovní přízvuk - je výrazně jazykově závislý
Větný přízvuk (intonační centrum) - zjednodušeně jde o Prozódické zvýraznění jádra výpovědi věty.
Intonace - nejobecněji - časový průběh časového spektra hlasu
Emotivní zabarvení hlasu - Projevuje se rychlými změnami hlasitosti a základní frekvence. Často přesahují hranici věty a jeho detekce u DS umožňuje zvolit vhodnou dialogovou strategii.
Emfatický přízvuk - Vytvářen emotivním zbarvením hlasu.Vyskyt např. ve větách v situacích s výrazným emocionálním kontextem: Bolí to jak čert.
Kontrastní přízvuk - snaha o zdůraznění slova nebo slabiky v kontrastu s jiným slovem nebo slabikou:
„řekl jsem do šakvic ne Rakvic.""
„Byte ne bit.""

Opakování - Prozódický atribut silně svázaný s mluvčím. 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, Mívá návaznost na další prozódické prvky: zaváhání, opakování, vyplněnou pauzu. . . .

16-Co značí v SRGS speciální pravidlo GARBAGE?

GARBAGE – značí část vstupu který lze považovat za NULL
Zvláštní pravidlo GARBAGE - Slouží k zadání libovolné nespecifikované promluvy

Informative example: given the definitions of US cities and states, a speech recognizer may implement the following rule definitions to match "Philadelphia in the great state of Pennsylvania" as well as simply "Philadelphia Pennsylvania".

```
$location = $city $GARBAGE $state;          // ABNF
```

```
<rule id="location">                    // XML
  <ruleref uri="#city"/>
  <ruleref special="GARBAGE"/>
  <ruleref uri="#state"/>
</rule>
```

17-Popište průběh FIA.

FIA - Form Interpretation Algorithn
FIA určuje pořadí provedení ve VoiceXML formuláři nebo menu. Cyklí přes všechna pole ve formuláři,žádá uživatele, aby zadal hodnoty pro každé nevyplněné pole,

Formuláře jsou interpretovány implicitním algoritmem pro interpretaci formulářů (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:

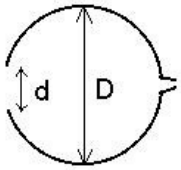
- pokud se má provést přesměrování hovoru (např. Element goto)
- pokud má dojít k předání dat dokumentovému serveru (element submit)
- pokud je explicitně požadováno ukončení (element exit).

Ukázka

```
<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>
```

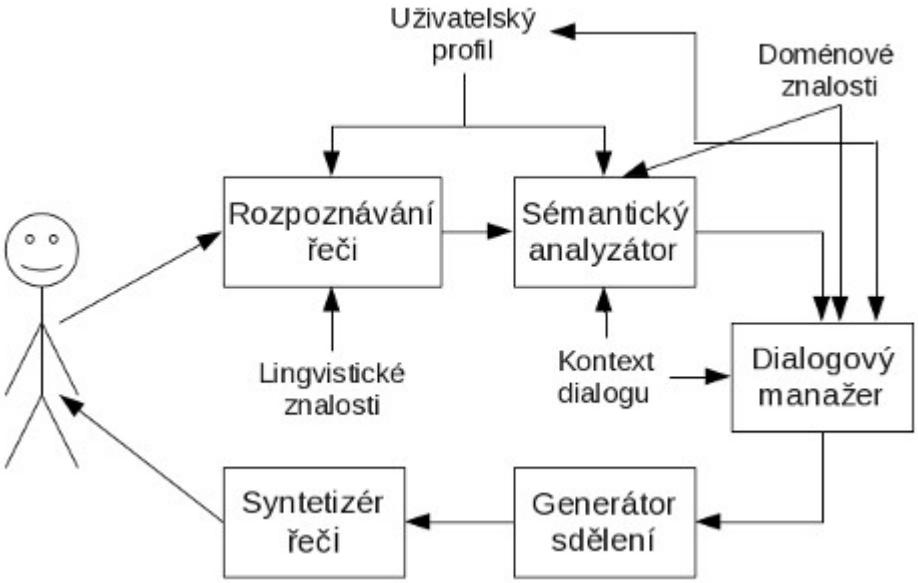
Otázky který si myslím že můžou být:

Helmholtzův rezonátor



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.

Struktura dialogového systému + popis



Uživatel - koncove zařízení, ktere uživateli umožňuje komunikovat s dialogovým systemem:

- telefon - komunikace prostřednictvím PSTN přes VoIP gateway - VoIP gateway převádí hlas na data a Zpět
- VoIP klient - komunikace prostřednictvím VoIP protokolu přímo s dialogovým systemem (SIP, H.323, Skype, . . .)
- textový klient - komunikace prostřednictvím protokolů DTMF+VoIP protokol, telnet, ssh, XMPP,. . .

Rozpoznávání řeči:

- převádí mluvene slovo na text
- využívá se:
  - rozpoznávání plynule řeči
  - rozpoznávání izolovaných slov
- pro zvýšení úspešnosti se používají gramatiky popisující množinu očekávaných vstupů.

Semantický analyzátor

- získává relevantní údaje z rozpoznaneho textu
- využívají se např. atributove gramatiky.

Dialogový manažer

- konečný automat
- na základe aktuálního stavu a vstupu od uživatele rozhoduje o dalším průbehu dialogu.

Generátor promluv - na základe údajů od dialogoveho manažeru generuje promluvy, ktere jsou následně syntetizovány.

Řečový syntetizer - převádí promluvy od generátoru promluv na mluvenou řeč, která je poslána uživateli.

Definuj zvuk, Druhy kmitání (zvuk) a co je akustika

Harmonické kmitání
- na teleso nepůsobí žádná vnejší 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 promenná síla

Zvuk - mechanické vlnění pružného prostředí (vzduch,
voda, kov, . . .)
- Problém - zvuk je periodický pouze na určitých
- intervalech.
  - analýza na krátkém intervalu, kde se předpokládá, že je periodický.

Akustika - věda studující zvuk
- Akustická intenzita - Vyjadřuje množství akusticke energie, ktere projde jednotkovou plochou za jednotku času.
  - Je přímoúmerná druhé mocnině akustickeho tlaku.
  - Orientační hodnoty akusticke intenzity
    - šepot - 10 - 20 dB
    - tlumený hovor - 35 - 45 dB
    - symfonický orchestr - 70 - 90 dB
    - rocková hudba - 110 - 130 dB.

Mechanismus vytváření řeči

Řeč vzniká pomocí hlasoveho ústrojí (umísteno v hrtanu).
Hlasivky vytváří úzkou hlasovou štěrbinu a jsou rozechvívány procházejícím vzduchem.
Frekvence jejich kmitání určuje základní hlasivkový tón - F0.
Zvuk, který vzniká v hrtanu pomocí hlasivek (samohlásky, znele souhlásky) je modifikován v rezonančních dutinách:
- hrtanove
- ústní
- nosohltanove.

Rezonanční dutiny fungují na stejnem principu jako Helmholtzův rezonátor(viz výše).

Fonetika

Zkoumá zvukovou stránku jazyka z různých aspektů.
Základní pojmy, ktere souvisejí se zpracováním řeči a dialogovými systemy:
- fonem
  - samohlásky - formanty
  - souhlásky - znelost/neznelost souhlásek
- koartikulace
- spodoba znelosti

Fonemy a fonetická transkripce
Fonem - elementární zvukový segment, který je vymezen na základe sve schopnosti diferencovat vyšší, znakove jednotky jazykoveho systemu (morfemy).
Fonetická transkripce (přepis) - převod psaneho textu do odpovídající foneticke podoby:
na shledanou -> na zhledanou | na schledanou
Fonetická abeceda - slouží k zápisu fonetickeho přepisu
- Mezinárodní fonetická abeceda (IPA) - součástí standardu UNICODE
- Fonetická abeceda pro metody zpracování řeči (Speech Assessment Methods Phonetic Alphabet - SAMPA) - sedmibitový přepis foneticke abecedy, využívá se při automatizovanem zpracování (např. řečový syntetizer MBrola, . . .)..

Samohlásky
Samohláska - samostatne tvoří slabiku
Obsahují:

- základní hlasivkový tón - frekvence kmitání hlasivek (100 - 400 Hz)
- formanty - frekvence vznikle a zesílene rezonancí v hlasových dutinách.

Formanty
Frekvence vznikle a zesílene rezonancí v hlasových dutinách

- F1 - vzniká rezonancí v dutine ústní.
- F2 - vzniká rezonancí v dutine hrdelní.

Existují i vyšší formanty (F3, . . .) - výskyt je často individuální.
Výskyt a intenzita formantů se může lišit v závislosti na:

- pohlaví - muž/žena
- veku - detství/dospívání/dospelost/seniorský vek
- zdravotním stavu - např. nachlazení, ochraptelost, němoci hlasivek a hrtanu, . . .
- . . .

Rozpoznávání plynule řeči

Hlavní rozdíly oproti rozpoznávání slov:

- nelze vytvořit databázi vzorů
- nutno brát zřetel na Prozódicke 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 pravdepodobnost např. Pro slova „máma" a „nána" - nejspíše se použije máma – je častejší.

(úspěšnost viz otázka 1.)

Gramatiky pro podporu rozpoznávání řeči
Java Speech Grammar Speci cation (JSGF)

- Textový zápis gramatiky nezávislý na platforme 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 interpretru VoiceGlue, . . .
- Později nahrazen SRGS

Ukázka JSGF
<koren> = Chci jet <cim> :j
Chci jet <cim> z <odkud> do <kam> :j
Chci jet <cim> z <odkud> do <kam> v <kdy> :;
<cim> = vlakem j autobusem;
<odkud> = <czMesto>;
<kam> = <czMesto>;
<kdy> = <czCas>;

W3C Speech Recognition Grammar Specification (SRGS)

- Standard W3C.
- Aktuální verze 1.0 (březen 2004).
- Definuje způsob zápisu pravidel a jejich odkazování.
- Dva způsoby zápisu:
  - XML
  - ABNF (Augmented BNF).

Ukázka W3C SRGS
#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>

Základní pojmy

Sémantika - je nauka o významu výrazů
Pragmatika – sleduje nějaký záměr slov
Syntax - zabývá se vztahy mezi slovy ve větě
Prozódie - popisuje zvukové vlastnosti jazyka (vyška řeči, hlasitost, doba trvani)

- Dialog - rozhovor dvou a více účastníku (sled promluv).
- Promluva - Souvisle sdelení, ktere učiní jeden účastník dialogu smerem k druhemu.
- Obrat - Promluva a reakce druheho účastníka na ni.
- Dialogová strategie
  - Postup, který k dané promluve přiřazuje následující promluvu.
  - Využívá znalost stavu dialogu:
    - zadané a požadovane informace
    - schopnosti účastníků dialogu
    - . . .
  - Je vlastností každého účastníka dialogu.

Dialogová komunikace
Hodnotící funkce:
funkce přiřazující každému dialogu reálne číslo.
Označuje se E(L), kde L je dialog.
Dialogová komunikace - Uspořádaná čtveřice
M = (S_1; S_2; E_1; E_2)
Si ; i € {1; 2} - dialogová strategie příslušneho účastníka.
Ei ; i € {1; 2} - hodnotící funkce příslušneho účastníka.

Kooperativita dialogu (viz otázka 9.)

Zpětná vazba v DS

Před tím, než system zpracuje Získané informace, je vhodne provest jejich verifikaci:

- oprava chyb rozpoznávání řeči
- oprava chyb uživatele
- . . .

Způsoby oveření získaných dat:

- **Sumarizující Zpětná vazba** - po zadání veškerých dat uživatelem je zopakuje a případne umožní jejich opravu.
- **Zpětná vazba „echo“** - po zadání každého údaje ho uživateli zopakuje, poskytne mu možnost případne opravy.
- **Implicitní Zpětná vazba** - posledne zadaná data jsou součástí dotazu na následující údaj.
- **Explicitní Zpětná vazba** - system validuje zadaná data pomocí explicitních dotazů na jejich hodnoty.

Příklady
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.
. . .

Implicitní Zpětná vazba:
System: Jmeno studenta.
Uživatel: Jan Novák.
System: Ve kterem meste se Jan Novák narodil.
. . .

Explicitní Zpětná vazba:
System: Zadejte jmeno studenta.
Uživatel: Jan Novák.
System: Student se jmenuje Jan Novák. Je to tak?
. . .

Pawlakův informační system

Pawlakův informační system formálne popisuje vztahy mezi objekty, jejich atributy a jejich hodnotami.
Souvislost s dialogovými systemy - hledání minimální množiny hodnot atributů, ktere nám určují jednotlive objekty.

Příklad

	Prvek1	Prvek2	Prvek3	Prvek4
Atribut1	1	1	0	0
Atribut2	0	1	1	1
Atribut3	1	1	1	0

Vyhledávací strom
Konstrukce vyhledávacího stromu pro Pawlakův IS:

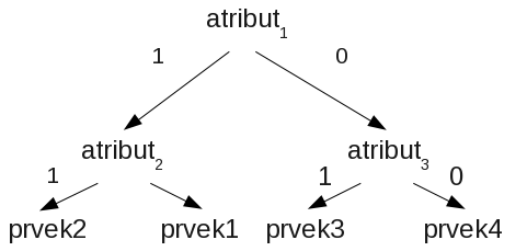
- 1 Postupne bereme jednotlive atributy obsažene v IS a ptáme se na jeho přítomnost (hodnotu).
- 2 Listy jsou jednotlive prvky, uložene v IS.

Souvislost s dialogovým rozhraním (s iniciativou systemu):

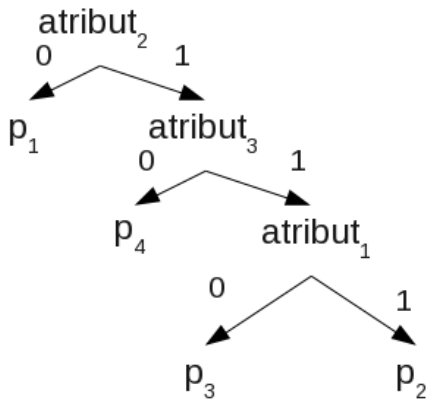
- Na každé úrovni stromu se ptáme na hodnotu/přítomnost odpovídajícího atributu.
- Uživatelova odpoved' určuje pokračování dialogu.

Lze použít i dialog se smíšenou iniciativou:

- 1 Uživatel zadá hodnoty libovolneho počtu atributů.
- 2 System odpoved' zpracuje a doptá se na chybející hodnoty.



Obrázek: Vyhledávací strom pro Pawlakův IS 4



Obrázek: Jiný vyhledávací strom pro Pawlakův IS 4

Dialog a strategicke hry

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

Strategicke hry
Vezňovo dilema
Strategická hra dvou hráčů.
Předpokládá, že každý hráč se stará především o svůj prospech.
Simuluje vyšetřování zločinu, ze ktereho jsou podezřelí dva lide.
Pravidla:

1. Pokud oba vezni 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).

	Bob mlčí.	Bob mluví.
Adam mlčí.	Oba odsoudí na 2 roky.	Adam dostane 10 let, Bob bude volný.
Adam mluví.	Adam bude volný, Bob dostane 10 let.	Oba odsoudí na 5 let.

Vezňovo dilema - jak se zachová part'ák?

Strategicke hry s podobnou výplatní funkcí
Válka pohlaví
Manžele preferují společne strávený čas. Co budeme delat dnes odpoledne?
Půjdeme (budeme se dívat) na módní přehlídku nebo na fotbal?
Hlava - Orel
Dva lide se sázejí, co padne na minci. Hlava nebo orel?

Iterovane strategicke hry

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

Prozódie - podrobně

WIKI:
Prozódie v lingvistice popisuje **zvukové vlastnosti jazyka**, které se uplatňují na úrovni **vyšší než** jednotlivý **foném** (**hláska**, **segment**). Souhrnně se hovoří o tzv. suprasegmentálních jevech, kterými jsou **slabika**, **přízvuk**, **tón**, **intonace** (melodie), frázování, … .

Prozódie
Úvod
Výstup syntézy je monotonní řeč bez intonace a přízvuku - zní nepřirozeně.
Náprava - doplnění Prozódie.
Základní Prozódické prvky:

- výška řeči
- hlasitost
- doba trvání.

Základním nositelem Prozódie v běžné řeči je slabika.
Prozódie závisí na typu věty:

- oznamovací, tázací zjišt'ovací, rozkazovací – klesající intonace
- otázka doplňovací (odpověď' ano/ne) - rostoucí intonace.

Modelování Prozódie - modulace F0.

Další Prozódické vlastnosti
Intenzita (hlasitost)

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 ,zbarvení tonu, 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 Prozódické vrstvě jazyka.

Základní odvozené Prozódické vlastnosti
Rytmus
- Prozódický prvek odvozený z dob trvání
    - slabik
    - pauz v daném časovém Úseku

Slovní přízvuk
je výrazně jazykově závislý:
- umístění přízvuku ve slově/přízvučné jednotce
- míra použití Prozódických prostředků k jeho vytváření - zejména použití hlasitosti oproti výšce.

Větný přízvuk (intonační centrum) - zjednodušeně jde o Prozódické zvýraznění jádra výpovědi věty.

Intonace
nejobecněji - časový průběh časového spektra hlasu

Emotivní zabarvení 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: Bolí to jak čert.

Kontrastní přízvuk - snaha o zdůraznění slova nebo slabiky v kontrastu s jiným slovem nebo slabikou:
„řekl jsem do šakvic ne Rakvic.""
„Byte ne bit.""

Opakován - Prozódický 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 n
Mívá návaznost na další Prozódické prvky: zaváhání, opakování, vyplněnou pauzu. . . .

Synteza řeči

Cíl - převod psaneho textu na mluvenou řeč. Výsledná řeč by mela znít co nejpřirozeneji.
Přirozená řeč by mela obsahovat:

- správnou intonaci
- správne umístení přízvuků
  - slovní
  - vetný
- korektní koartikulaci
- správný rytmus (časování)
- . . .

Druhy syntezy řeči

- Synteza ve frekvenční oblasti - simuluje chování řečového ústrojí.
- Synteza v časové oblasti - spojování řečových segmentů do vetších celků (veta, promluva, . . .)
- Korpusová - varianta syntezy v časové oblasti – jako databáze řečových segmentů slouží řečový korpus.
- Problemove orientovaná synteza:
  - varianta syntezy v časové oblasti
  - využívá vetší celky - vety, . . .
  - příklady:
    - hlášení nádražního rozhlasu
    - automatizovane linky telefonicke podpory
    - . . .

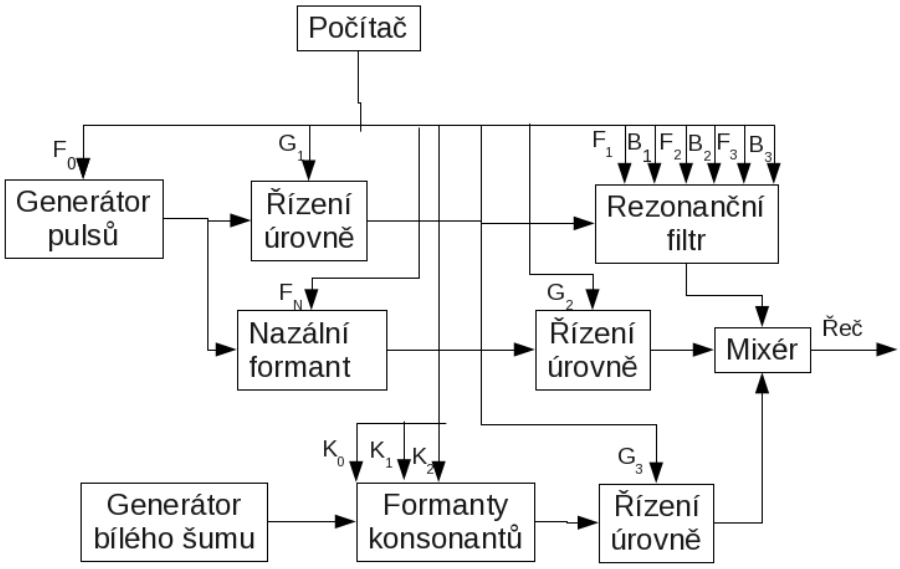
Fáze syntezy řeči

1. Fonetický přepis textu.
2. Synteza foneticky přepsaneho textu:
   - Synteza ve frekvenční oblasti - volba průbehu parametrů syntezy (F0/generátor šumu, vyšší harmonicke frekvence, jejich intenzita, . . .)
   - Synteza v časové oblasti - výber vhodných segmentů a jejich spojení.
3. Případný postprocessing:
   - doplnení intonace
   - doplnení přízvuků
   - . . .

Fonetický přepis

Slouží k přesnemu, jednoznačnemu zápisu mluvene 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 kra:sni: jazik
- . . .



Obrazek: Schema serioveho formantoveho syntetizeru

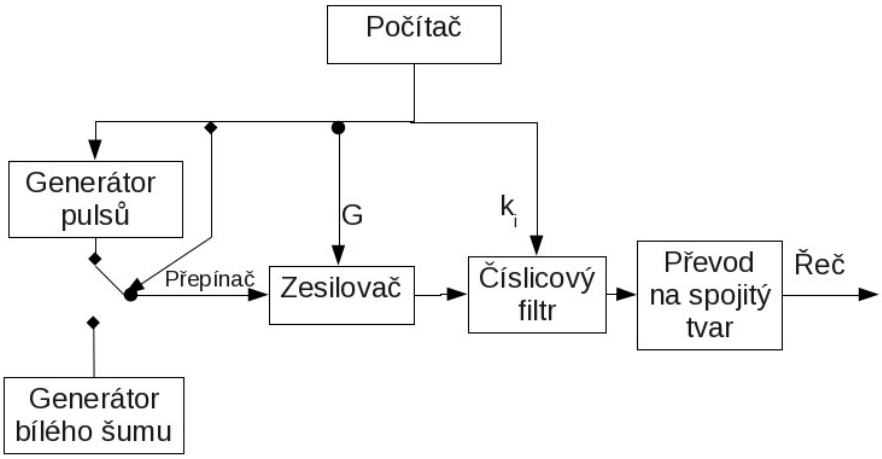
LPC syntetizer + schema

Charakteristiky pro LPC syntetizer:

- perioda základního hlasivkoveho tónu T0
- charakteristika hlásky - znelá/neznelá
- amplituda budícího signálu G
- koeficienty číslicoveho filtru.

Způsob získání koeficientu číslicoveho filtru:

- vrcholy v LPC spektrální obálce analyzovaneho mikrosegmentu
- kořeny charakteristicke rovnice zdrojoveho filtru
- reflexní koeficienty.



Synteza ve frekvenční oblasti

Shrnutí

Výhody a nevýhody syntezy ve frekvenční oblasti:

+ Male pamet'ove nároky - model použiteho mluvčího.
+ Syntezu lze realizovat hardwarove.
- Hlas bývá mene přirozený oproti synteze v časové oblasti.
  - Problem přesnosti matematickeho modelu.
- Softwarová synteza ve frekvenční oblasti bývá výpočetne náročnejší než synteza v časové oblasti.

Obvykle využití:

- doplnení syntezy v časové oblasti o:
  - vetnou intonaci
  - vetný a slovní přízvuk
  - další Prozódicke faktory.
- Občas pro syntezu na zařízeních, která nedisponují dostatečnou kapacitou pameti (mobilní telefony, PDA, . . .).
- Občas pro multiliguální syntezu.

Používané řečove segmenty

Alofóny:

- poziční varianty fonemů - obsahují
  - fonem
  - okolí ovlivnene koartikulací.
- počet alofónů - n3 (n - počet fonemů).

Difóny:

- začínají uprostřed jednoho fonemu a končí uprostřed následujícího fonemu
- počet difónů - n2
- často využívane pro syntezu i pro rozpoznávání (např. syntetizer MBrola)

Trifóny:

- Začínají uprostřed leveho sousedního fonemu a končí uprostřed praveho sousedního fonemu.
- Počet - n3.
- Často Používané pro rozpoznávání a syntezu řeči.

Slabične segmenty:

- Snaha, aby co nejvíce odpovídaly slabikám.

- délka - 1 | 3 fonemy.
- Využívá se např. v TTS systemu Demosthenes.

Standardy pro syntezu řeči

Snaha sjednotit jazyky pro popis promluvy pro řečove syntetizery.
Definují značkování postihující:

- prozódii - rychlost řeči, F0, zdůraznení části promluvy, pauzu, hlasitost, . . .
- mluvčího - pohlaví, vek, . . .

Používané standardy:

- SABLE
- SSML

SABLE

Otevřený standard pro Prozódicke značkování textu.
aplikace XML/SGML
snaha o zkombinování 3. značkovacích jazyků pro syntezu řeči:

- SSML - Speech Synthesis Markup Language
- STML - Spokén Text Markup Language
- JSML - Java Synthesis Markup Language

Základní značky (SABLE)

- SABLE - kořenová značka
- DIV
  - Slouží k členení dokumentu na odstavce a vety.
  - Typ části dokumentu určuje atribut type.
    <DIV TYPE="paragraph"> ... </DIV>
  - Prozódicke značky:
    - EMPH - zdůraznení části promluvy
    - PITCH - výška promluvy
    - VOLUME - úroveň hlasitosti
    - RATE - rychlost
    - BREAK - pauza

Popis mluvčího:

- element SPEAKER:
  - AGE - vek mluvčího (older, middle, younger, teen, child)
  - GENDER - pohlaví (male, female)
  - NAME - jmeno mluvčího, závisle na TTS - TTS musí daného mluvčího znát.

Foneticke:

- PRON - foneticky přepsaná promluva, lze použít IPA.
- SAYAS - způsob fonetickeho přepisu (datum, telefon, url, poštovní adresa,. . .)
- LANGUAGE - jazyk promluvy.

Ukázka

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

SSML

- Otevřený standard W3C
- Aplikace XML.
- Součást rodiny W3C Voice Browser Activity
- Aktuální verze 1.0 (září 2004)

Základní značky
kořenový element speak
strukturní elementy:

- p - odstavec
- s - veta

foneticke:

- **say-as** - způsob fonetickeho 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, vek, . . .)

Prozódicke značkování:

- emphasis - zdůraznení části promluvy
- break - pauza
- prosody - ovlivňuje základní Prozódicke jevy:
  - vlastnost dána atributem - pitch, rate, duration, volume

Ukázka

```
<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.</prosody>
  </p>
 </voice>
</speak>
```

W3C Voice Browser Activity

1999 - založena W3C Voice Browser Working Group.
Cíl - návrh standardů umožňujících přístup k Webu
pomocí hlasu a telefonu.Členove: HP,Motorola,ScanSoft,IBM,. . .

Standardy W3C Voice Browser Activity

- VoiceXML - jazyk pro popis dialogových strategií.
- Speech Recognition Grammar Specification - jazyk prozápis gramatik pro podporu rozpoznávání řeči.
- Semantic Interpretation for Speech Recognition – jazyk pro podporu semanticke interpretace.
- Speech Synthesis Markup Language - jazyk pro popis Prozódických charakteristik pro syntezu řeči.
- Pronunciation Lexicon Specification - popis výslovnostipro rozpoznávání a syntezu řeči.
- Call Control XML - jazyk pro popis řízení telefonního spojení uživatele a systemu.
- State Chart XML - jazyk pro popis obecne použitelných stavových automatů.

Zpracování
Standardy jsou značkovací jazyky - nutná interpretace
Existuje řada platforem:

- Volne dostupne desktopove- JvoiceXML, PublicVoiceXML, . . .
- Komerční desktopove - Optimtalk - dříve existovala volne dostupná verze; laboratoř LSD má zakoupenou licenci na laboratorní stroje.
- Volne dostupne on-online - Asterisk+VoiceGlue resp. OpenVXI, . . .
- komerční on-line - Voxeo Prophecy, Bevocal Cafe – lze vyzkoušet a omezene používat on-line (max. 2 paralelní hovory).

Semantic Interpretation for Speech Recognition

Sémantika - přiřazuje význam tvrzením.
Semantika v dialogových systemech:

- přiřazuje interpretaci promluvám a jejich částem
- umožňuje získání relevantních údajů.

SISR - standard z rodiny W3C Voice Browser Activity

- slouží k semanticke 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 vnitřne reprezentována pomocí objektů ve formátu JSON.

Přiřazení interpretace části promluvy

- Semantická interpreta bývá součástí pravidel SRGS.
- Přiřazení interpretace k pravidlu - pomocí „tagu":
    - 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: finterpretaceg
          $potvrzení = $souhlas -ano} | $nesouhlas -ne}

Odvozování interpretace na základe dílčích interpretací

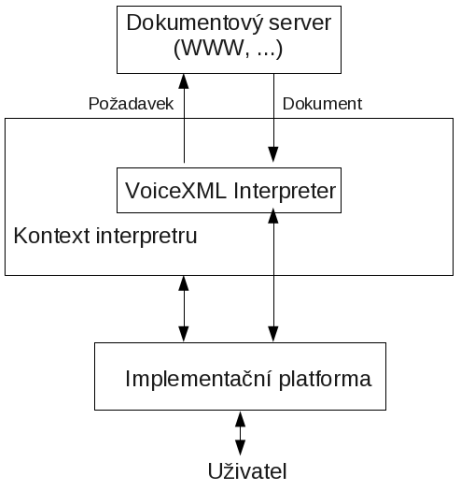
- Zápis odvození - pomocí výrazů v jazyce ECMAScript.
- Přiřazení pravidel pro odvození k pravidlům gramatiky - pomocí atributu/elementu tag.
- Výsledná interpretace reprezentována pomocí objektů ve formátu JSON.
- Vyhodnocování promluv:
    - přístup k dílčím interpretacím – interpretace neterminálních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

VoiceXML

Základní informace

- Jazyk pro popis dialogových strategií.
- Součást standardů W3C Voice Browser Activity.
- Cíl: přinest výhody weboveho vývoje a doručování obsahu do interaktivních hlasových aplikací.

Model komunikace



Obrázek : Model architektury aplikací postavených na VoiceXML

Struktura aplikací

- 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 opetovnemu provádení jiste části dialogu (např. zjištení e-mailove adresy, . . .).
  - Realizovány jako formuláře, kterým mohou být předány parametry, a ktere mohou vrace hodnotu (viz dále).
- Sezení:
  - Začíná v okamžiku zahájení komunikace s VoiceXML interpretrem.
  - 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, ktere sdílejí kořenový dokument.

Zápis dialogů pomocí VoiceXML

- vxml - kořenový element každého dokument.
- Musí obsahovat atributy:
  - version - použitá verze standardu VoiceXML
    - aktuální 2.1
    - hodnota závisí na použite platforme - OptimTalk 1.9 - 2.1, JVoiceXML - zatím neůplná podpora verze 2.1, VoiceGlue - podpora 2.0 + nektere možnosti z 2.1, . . .
  - xmlns - deklarace implicitního jmenneho prostoru. Hodnota musí být http://www.w3.org/2001/vxml.
  - xml:lang - hodnotou je kód jazyka, pro který je dialogove rozhraní navrženo.
- Element obsahuje:
  - jeden nebo více elementů form,
  - element menu,
  - . . .

Formulář

FIA - Algoritmus interpretace formulářů (viz. otazka 17.)

Jeden ze základních elementů VoiceXML dokumentů.
Ohraničen značkami < form > a < =form >.
Obsahuje:

- sadu vstupních polí
- deklarace promenných daného formuláře - element var
- definice gramatik platných v daném formuláři
- bloky výkonneho kódu.
- . . .

Atributy:

- id - povinný atribut:
  - slouží jako identifikátor daného formuláře
  - jeho hodnota musí být unikátní v daném dokumentu
  - lze použít k předávání řízení do daného formuláře.

Možný obsah

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

Vstupní pole a řídící struktury - ukázka užití (str.275)

```
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml"
xml:lang="cs-CZ">
<form id="hello">
<block name="hello">
<prompt>Welcome to the VoiceXML!.</prompt>
</block>
<field name="greating">
<prompt>Hello.</prompt>
<grammar src="greatings.grxml"/>
<noinput>
<prompt>Tell mi something nice, like hello, hi,
good day.</prompt>
</noinput>
<nomatch>
<prompt> I didn't understand you, but thanks anyway.
</prompt>
<exit/>
</nomatch>
<noinput count="2">
<prompt> When you don't want to speek to me good
bye.</prompt>
<exit/>
</noinput>
</field>
<filled>
<prompt> you said <value expr="greating"/></prompt>
<submit src="SomeURI" namelist="greating"/>
</filled>
</form>
```

Element field
Představuje vstup od uživatele. Může být zadán bud' hlasem nebo pomocí DTMF.
Atributy:

- name - jmeno pole. Používá se k přístupu k zadané hodnote (pomocí stínove promenne se shodným jmenem).
- expr - výraz v jazyce ECMAScript, který slouží k inicializaci hodnoty vstupního pole.
- cond - vstupní podmínka nutná pro zpracování vstupního pole.

Obsah elementu:

- Výzva s popisem požadovane hodnoty (element prompt).
- Gramatika (element grammar) - gramatika s popisem akceptovaných vstupů.
  - Typ gramatiky závisí na použite platforme (zabudovanem rozpoznávači řeči, např. Voxeo Prophecy, OptimTalk - SRGS, JVoiceXML - JSGF,. . .).
- Ošetření událostí:
  - noinput - nebyl zadán žádný vstup
  - nomatch - zadán neakceptovaný vstup (vstup neodpovídá gramatice)
  - filled - umožňuje zpracování vstupu po vyplnení vstupního pole
  - . . .

Ukázka užití

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.0" xmlns="...">
<form id="main">
<field name="name">
<prompt>Zadejte Vaše jmeno</prompt>
<grammar src="..." type="application/xml+srgs"/>
<noinput>Zadejte prosím Vaše křestní jmeno
</noinput>
<nomatch>Je mi líto, ale zadané jmeno není
v kalendáři</nomatch>
</field>
<filled>
<submit next="applicationURI" namelist="name"/>
</filled>
</form>
</vxml>
```

Element record
Umožňuje systemu nahrát zprávu od uživatele.
Lze využít např. pro tvorbu hlasoveho 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 výsledne nahrávky, musí být podporována VoiceXML platformou
- . . .

Obsah elementu:

- Výzva/výzvy s popisem požadovaneho požadovaneho vstupu.
- Ošetření událostí:
  - noinput - uživatel zprávu nezačal nahrávat.
  - connection.disconnect.hangup - uživatel zavesil.

Ukázka užití

```
<?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 zaznení
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 problem).
Jeden a tentýž subdialog se dá volat opakovane.
Vyvolání subdialogu:

- element subdialog - vlastní volání subdialogu.
- Obsahuje:
  - param - definice hodnoty parametru.
  - filled - kód, který se má provest po návratu z dílčího dialogu.
- Atributy:
  - name - jmeno volaneho subdialogu.
  - src - URI formuláře s kódem subdialogu.

Kód subdialogu:

- formulář
- ukončený elementem return.

Ukázka užití

```
<?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="greating" src="šay_hello">
<param name="param1" expr="'ahoj'"/>
<filled>
<prompt>Hodnota subdialogu je <value
expr="greating.great"/></prompt>
</filled>
</subdialog>
<filled>
<prompt>Řekl jste <value expr="greating.great"/>
</prompt>
</filled>
</form>
<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 neodpovedel.
Nashledanou.</prompt>
```

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

Element block
Obsahuje proveditelný obsah.

- atributy:
  - name - název bloku.
  - expr - iniciální hodnota promenne formuláře.
  - cond - podmínka omezující provádení bloku.
- struktura - shodná s obsahem elementu filled:
  - řídící struktury - elementy if, else, elseif
  - přiřazovací příkaz - element assign, clear, . . .
  - příkazy skoku - element goto, exit, return, . . .