pb095 last day

fyz. akustika - zakladne pojmy , rychlost na com zavisu vo vzduchu vode , hmot. bod, amplituda atd., o kmitoch ..rezonance. akusticka intenzita/ tlak, jednotky .. akusticke spektrum (ziskam hodnoty frekvenci a intenzit zastupeni) .. furierove rady sa neda na digitalnom ne to nespojite

zaklad spracovania zvuku fiziolog.

Fonem (samohlasky (frekvence, formanty) spoluhlasky), koartikulace, .. zakladny hlasiv. ton (100-400hz)

cez spektrum urc frekvence audacity

metody analyzy kratkodoba

ste, sti, zcr

frekvencna anal. - fourier , linear. predikcia

DTW princip

increase accuracy: gramatiky (jazykovy a model recnika)

skryte markovove modely princip

alofon - fonem + okolie ovolyv. koartikulaci,

difon polovica fonemu jedneho a druheho

vokal. jadro : samohlaska/rl v slabike , preatura prve koda posledne (may not be)

Dve otazky,25 minut(lze libovolne prodlouzit)

Nejdriv pisemka na papir, pak inhed zkouska s opravou pisemky:

1.Akusticke spektrum
2.Skryte markovovske modely

PB095 - Uvod do pocitacoveho zpracovani reci

Zkouska - 26. 1. 2005
2 otazky, 20 minut (lze prodlouzit, pak ihned ustni "oprava" pisemky)

1. Konkatenativni synteza
2. Vektorova kvantizace, skryte Markovovy modely (HMM)

18. stol.
Kratzenstein – rezonátory
Wolfgang von Kempelen – první mechanický řečový syntetizér, 1791
Wheatstone – rekonstrukce Kempelenova syntetozéru, 1800
Nepomuk Bolzano – Bolzanova věta + funkce

19.stol
Fourier
Helmholtzův rezonátor

20. stol
- 1924 -spektrální analýza řeči na bázi formantové analýzy samohlásek.
- Vokodéry, komprese řečového záznamu
- Syntéza řeči
- Rozpoznávání řeči
- Dialogové systémy

1939 – elektronický sysntetizér VODER, analogický výstup, velká složitost zacházení

Akustika
- věda zkoumající zvuk
- infrazvuk < 16 Hz, 16 000 Hz > ultrazvuk
- zvuk: mechanické vlnění schopné vyvolat v lidském uchu sluchový vjem
- kmitání hmotného bodu: pohyb mezi amplitudami, perioda a frekvence kmitavého pohybu
- frekvence: 1 Hz = 1 kmit za sekundu

Rezonance

- označuje jev, který lze pozorovat při nuceném kmitání, kdy vhodně působicí malá budící síla může způsobit velké změny v kmitajícím systému
- těleso na pružině, při příliš velkých nebo nízkých freknvencích bude amplituda kmitání malá, naopak při vhodné frekvenci bude amplituda velmi velká → dojde k rezonanci
- frekvence ovlivňující objekt je podobná s vnitřní rezonanční frekvencí objektu

Akustická intenzita

- množství akustické energie působící na plochu za ejdnotku času
- = (akustický tlak (síla působící na plochu))^2
- rozsah AI ve kterém vnímáme tón se nazývá práh citlivosti
- **Weberův-Fechnerův zákon**
    - člověkem subjektivně vnímaná hlasitost roste při geometrickém nárůstu intenzity přibližně lineárně
- hladina akustické intenzity se proto vyjadřuje logaritmicky, používanou logaritmickou jednotkou je **1 bel** (bezrozměrná míra); šepot 10 dB, startující letadlo 120 dB

Tón

- základní: lze popsat sinusoidou v závislosti na čase
- složený tón: kombinace základních tónů
- Helmholtzův rezonátor – rozklad zvuku do základních tónů

Fyziologická akustika

Mechanismus vytváření řeči

- hlasové ústrojí v hrtanu
- mezi hlasivkami je *hlasová štěrbina*, která se mění podle kmitání hlasivek; podle frekvence kmitání vzniká hlasivkový tón
- tento zvuk se modifikuje v rezonačních dutinách (hrtanová, ústní a nosohltanový dutina)
- vnímání zvuků sluchovým orgánem
    - ušní boltec – zachycuje zvukovou energii
    - zvukovod – vede energii k bubínku
    - ušní bubínek – předává kmity třmínku
        - Eustachova trubice spojuje střední ucho a ústní dutinu, vyrovnává přetlaky
- hlemýžď (Cochlea) ve středním uchu, obsahuje Cortiho ústrojí
    - v něm se mechanická energie mění na nervové vzruchy pokračující do mozku

Helmholtzova rezonanční teorie

- vlákénka v Cortiho ústrojí představují rezonanční soustavu odpovídající strunám různých délek a rezonující proto pro různé tóny

Fonetika

- Foném: základní zvukový segment, který má rozlišovací funkci v systému konkrétního jazyka, každý jazyk má odlišnou sadu fonémů
- IPA (International Phonetic Alphabet): systém k fonetickému zápisu různých jazyků
- vokály (samohlásky), konsonanty (souhlásky)
- zvukové spektrum: kombinace všech frekvencí vlnění, ze kterých se zvuk skládá; základní (nejnižší) frekvence F0 a formanty

formanty
- vznikají rezonancí (F1) a v dutině hrdelní (F2)
- zesílené části generovaného zvukového spektra
- nejnižší frekvence (F0) udává výšku tónu, vyšší frekvence (formanty, F1, F2, F3 ...) udávají výsledný sluchový dojem
- určující pro rozpoznávání vokálů
- vlastnosti formantů jsou individuální; např.: a – F1: 750 – 1100 Hz, F2: 1100 – 1500

souhlásky
- podstatně více závislé na kontextu než samohlásky
- význam formantu pro ně nemá význam (mají tónový charaketr jen málo)
- znělé vs. neznělé, podle účasti hlasivek na vytváření souhlásky
- závěrové souhlásky (okluzívy) – výdechovému proudu vzduchu se vytváří překážka
- úžinové souhlásky (frikativy) – zúžení výdechové cesty
- polozávěrové souhlásky (semiokluzívy) (c, č)

Koartikulace
- modifikace fonému v řečovém kontextu

Digitalizace akustického signálu
- vzorkování: transformace spojitého vstupního signálu na diskrétní posloupnost
- Shannon: na bezeztrátovou transformaci je třeba, aby vzorkovací frekvence byla dvojnásobná oproti nejvyšší frekvenci vstupního signálu
    - jinak docházé ke zkreslení
- kvantizace: převod nevzorkovaných reálných hodnot na celočíselné; počet prvků intervalu → počet úrovní kvantování, většinou 8/16 bitů
- digitalizace řečového signálu: vzorkovací frekvence 16-20 kHz, 16ti bitové kódování

Zpracování digitalizovaného signálu

Krátkodobá analýza
- zpracování na časovém intervalu, na němž nepředpokládáme dynamické změny, tzv. **mikrosegmentu**

    - předpokládáme opakování tohoto segmentu v určitém intervalu; neshoda signálu v okolí vybraného „okénka“ může mít za následek chybu → použití „váhového oknénka“, Hammingova okénka nebo pravoúhlého okénka

Analýza v časové oblasti

- vychází z hodnot vzorku, nikoliv z hodnot spektra
- krátkodobá intenzita – detekce ticha
- střední počet průchodů nulou, počet lokálních extrémů
- krátkodobá autokorelační funkce: zjišťování periodicity signálu a základního tónu řeči

Analýza ve frekvenční oblasti

- krátkodobá Fourierova transformace
    - algoritmus FFT (Fast Fourier Transform)
    - časová reprezentace signálu <=> frekvenční reprezentace signálu
- Kepstrální analýza
- Lineární predikce: analýza na základě předchozích vzorků

Rozpoznávání izolovaných slov

- vyřčení oddělených povelů, odpadá problém stanovení rozhraní dvou slov
- akustický vektor, vektor příznaků: vztahující se k mikrosegmentu
- klasifikátory:
    - DTW (Dynamic Time Warping)
    - statistické metody
    - dvouúrovňové – segmentace a fonetické dekódování, rozpoznávání slova

Dynamic Time Warping

- porovnání dvou úseků promluv vzniklých rozdělením do mikrosegmentů
- pro množinu rozpoznávaných slov vytvoříme soubor referenčních posloupností akustických vektorů, porovnáváme posloupnost akustických vektorů slova s množinou posloupností referenčních akustických vektorů → bereme největší shodu
- algoritmus pro poměřování podobností dvou sekvencí v čase nebo rychlosti
- urychlení DTW vektorovou vkantizací

Kódová kniha

- vektorový prostor X rozdělíme na disjunktní podmnožiny X(i) a v každé zvolíme reprezentanta v(i)
- vektorový kvantizér přiřazuje vektoru x z X(i) vektor v(i)
- množina vektorů v(i) tvoří kódovou knihu
- vyhledávání v kódové knize: shluky, subshluky, prohledávací stromy

Skryté Markovovy modely (HMM)

- hlasové ústrojí je v jistém časovém intervalu v jednom z konečně mnoha stavů atrikulačních konfigurací, než přejde do stavu následujícího
- kvantizace akustických vektorů (vytvoření kódové knihy) → dosažení konečnosti všech parametrů modelu

- pětice (Q,V,N,M,π), kde Q jsou stavy, V je abeceda výstupních symbolů
    - N je matice přechodu (s jakou pravděpodobností se přejde v čase *t* do dalšího stavu v čase *t+1*)
    - M je matice přechodu (pravděpodobnost, že ve stavu *q* je generován akustický vektor)
    - π vektor pravděpodobností, že i-tý stav bude počáteční
- soubor trojice parametrů (N,M,π) vytváří model řečového segmentu (např. slova)
- **určení pravděpodobnosti promluvy**

Kódování

Rozpoznávání

Trénování

Modely

Řečový signál
Trénovací databáze
Trénování
Trénování

kódování signálu (segmentace, výpočet spektra, reprezentace např. kepstrální) → **trénovací část** (natrénování modelů jednotlivých promluv) → **rozpoznávací část** (vyhodnocení pravděpodobnosti, že daná sekvence vektorů byla generována daným modelem; model, který generuje danou promluvu, ji s nejvyšší pravděpodobností reprezentuje)

- modely slov vytvářeny zřetězením modelů fonémů
- obtíže:
    - šum pozadí vs. sykavky
    - přítomnost zvuků mimo oblast slyšitelnosti

Rozpoznávání souvislé řeči

- rozdíly oproti izolovaným slovům
    - nelze vytvořit databázi vzorů
    - prozodické faktory (prozodie = vlastnosti jazyka na úrovni vyšší než je úroveň jednoho fonému)
    - určování hranice mezi slovy, chyby řeči
- statistický přístup (máma vs. nána)

Jazykový model

- cíl je určit pravděpodobnost nějaké promluvy v daném jazyce
- přiřazuje pravděpodobnost vyslovení slova posloupnosti slov
- W = (w(1)w(2)…w(n)) - posloupnost slov
- O = (o(1)o(2)…o(t)) - posl. akust. vektorů
- jedná se n-gramové modely, kde pravděpodobnost n-tého slova se předvídá z n-1 slov
- nejčastější jsou trigramy - **P(w(n)|w(1)…w(n-1))**

Topic Recognition

- mění se stavový prostor a pravděpodobnost (př.: vím, že se jedná o burzu – honey → money)
- syntaktická struktura
- korpus: soubor textů opatřený metajazykovými značkami

Syntéza řeči

- ve frekvenční oblasti
  - syntetizér emuluje funkci hlasového ústrojí
  - formantová syntéza - lidské artikulační ústrojí jako skupina kaskádně řazených filtrů
  - LPC (Linear Prediction Coding) – metody lineární predikce
- v časové oblasti (**konkatenativní**)
  - nejjednoduší, nejčstější
  - rámce + sloty: doplnění potřebných slov do slotu (př.: jlášení odjezdů vlaků)
  - vytvoření segmentů řečové abecedy a následné skládání → je třeba zahrnout kontext
    - fonémy se nehodí kvůli koartikulaci
    - jednotlivá slova jako základní segment nelze použít kvůli velkému počtu
    - zákadní používané segmenty jsou
      - alofóny – u fonému si zapamatuju i jeho kontext (okolní dvě písmena) --> vyžaduje n^3 zapamatovaných možností (pro n fonémů)
      - difóny – ze středu jednoho fonému do středu druhého → n^2 možností
      - trifóny – ze středu levého sousedního do středu pravého sousedního fonému → n^3
      - slabičné segmenty – jsou přirozené, kontextu se vzájemně moc neovlivňují; vymezení obtížné - cca 10 000 slabik

Jazyky poměřované slabikou/přízvukem

- syllable timed (čeština)
- stress timed (angličtina)

Fonetická transkripce

- text → fonetická abeceda, co nejpřesnější záznam výslovnosti

- fonologická pravidla: znělost - (dub->dup, sjezd->sjest),

Prozodie

- zvukové vlastnosti mluvené řeči
- informační vrstva pro zvýšení srozumitelnosti, která obsahuje informace meobsažené na úrovni nižších celků
- základní jednotkou je **slabika**
    - atributy slabiky
        - výška: výška frekvence základního tónu (F0)
        - hlasitost (intenzita) – fyzikální (intenzita signálu v časovém okamžiku) a fyziologické hledisko (reakce Cortiho ústrojí na vnímaný zvuk)
        - trvání: podle kontextu
- další prozodické atributy
    - kvalita hlasu: jitter (chvění), nepravidelné výchylky v F0, zabarvení tónu apod.
    - rychlost řeči: trvání slabiky nebo např. počet vyslovených znaků za jednotku času
    - pauza: tichá pouza, vyplněná pauza („eeh“) – obtížná detekce
    - zaváhání: aspekt ne úplně zasahující do oblasti syntaxe a sémantiky; informace, ná základě které můžeme např. uzpůsobit dialogovu strategii
- odvozené atributy
    - rytmus: doby trvání pauz/slabik
    - slovní přízvuk: velmi jazykově závislý
    - větný přízvuk: prozodické zdůraznění jádra výpovědi věty (např. zdůraznění některých slov)
    - intonace: časový průběh zvukového spektra během výpovědi
    - emotivní zabarvení hlasu: kontext přesahující jedinou větu, důležitá např. pro dialogové systémy (volba dial. strategie)
    - emfatický (důrazný) přízvuk: "To je tedy opravdu neslýchané!“
    - kontrastní přízvuk: zdůraznění slova nebo slabiky v kontrastu s jiným slovem nebo slabikou
    - opakování: často jako výplňková část promluvy (kterou si mluvčí ani neuvědomuje)
    - výplňkové části: slova (subjektivně podle mluvčího) užívaná bez důležitosti neseného sdělení jako výplň: ("Máme zkoušku z matiky, víš co.", "Vole vykašli se na to vole.")
    - přerušení: větších celků nebo jednotlivých slov, např. v návaznosti na zaváhání, opakování apod.
    - korekce: přeřeknutí, upřesnění nebo upravení předchozí části promluvy

- Prozodické segmenty mluvené řeči
    - prozodické fráze: skupina intonačně jednotných slov (namísto věty)
        - koresponduje se syntaktickou strukturou věty
    - přízvukový takt: dělení prozodické fráze

        - skupina slabik podřízená jednomu slovnímu přízvuku (v češtině typicky slovo nebo slovo + jednoslabičné slovo)

Standardy pro syssntézu řeči:

- SABLE
    - snaha o zkombinování 3 značkovacích jazyků: SSML, STML, JSML
- SSML
    - součástí W3C, vývoj koncem 90. let
    -

Dialogové systémy

- Dialogovy system - informaˇcn´ı system s dialogovym (hlasovym/textovym) rozhraním.
- přirozenější než GUI
- Nové způsoby komunikace s applikacemi
- Pro lidi bez končetin

Historie:

- Eliza - 60. léta
- Parry

Základní pojmy:

- Dialogovy system - informaˇcn´ı system disponuj´ıc´ı dialogoo´ym rozhiran´ım.
- Dialogov´e rozhraní - UI, ktere umožňuje uzivateli komunikaci s aplikací prostrednictvım dialogu.
- Dialog - komunikace dvou ucastnıku (pro nas clovek ↔ pocıtac).
- Promluva - souvisle sdelenı jednoho úcastnıka dialogu.
- Obrat - promluva a reakce druhého ´uˇcastn´ıka na ni.
- Dialogova strategie - urcuje ke kazde promluve naslednıka.

Komponenty dialogového systému

- rozpoznávání řeči
- systematický analyzátor - zjištuje význam promluvy
- dialogový manažer - na základě faktů rozhoduje o dalším kroku
- generátor sdělení

Info využívané dialogovým systémem:

- lingvistické znalosti
- uživatelský profil
- Doménové znalosti
- kontext dialogu

DIALOGOVÁ KOMUNIKACE

Dialogova komunikace - Uspořádáná čtveřice M = (S1, S2, E1, E2).

- Určuje následující krok dialogu v závistolsti na stavu dialogu a vstupní promluvě
- Hodnotící funkcec dialogu E přiřazuje danému dialogu reálné číslo popisující úspěšnost dialogu z pohledu dané strany

Pravidla pro vedení kooperativního dialogu:

- Aspekty:
    - informovanost
    - přesevědčivost
    - způsobu - jendoznačná
    - zdvořilost
    - Asymetrie - informovat uživatele pokud něco vybočuje
    - Znalostí a schopností - jak moc je ten uživatel zkušený/vzdělaný
    - vyjasňování a odstra'nování chyb

Iniciativa v dialogu:

- podle toho kdo klade otázky a kdo odpovídá
    - dialog s iniciativou živatele - reálně se moc nepoužívá
    - dialog s iniciativou systému
    - dialog se smíšenou iniciativou
- Před tím než systém předá záskané ifnormace je dobré provést verifikaci a recap

Další aspekty dialogových strategií:

- možnost přerušit sysstém
- korekce chyb, opakování (uživatel je dement a nerozumí tomu)
- Přizpůsobení dialogové stragegie - užřivatel ve spěchu
- vícejazyčnost
- dettekce emocí
- multimodalita - umožnuje paralelní kuminkaci více kanály - (obraz, zvuk,...)
- učení se z chyb

Dialogové systémy:

VoIP:

- rodina protokolů pro přenos hlasu přes internet

SIP:

- Protokol na aplikační vrstvě
- pro přenos signalizace v internetové telefonii
- zjišťuje:
  - lokalizaci
  - stav - obsazený,...
  - možností - přenosová rychost
  - navázání spojení
  - řízení probíhajícího spojení

Podpora rozpoznávání řeči:

- omezení domény možných vstupů

Java Speech Grammar Specification (JSGF)

- Textový zápis gramatiky nezávislý na platformě a prodejci, pro podporu rozpoznávání řeič
- Gramtika se skládá z pravidel, které popisují co může být řečeno

SRGS:

- Standart W4C
- Definuje způsob zápisu a pravidel a jejich odkazování
- Dva zůsoby zápisu (XML, ABNF)

Sémantická interpretace promluvy:

- Většinou řešeno pomocí atributů v gramatice pro rozpoznávání řeči
- Slouží k urční umístění a hodnotvy významných částí uživatelsé promluvy

Popis dialogových rozhraní:

- Ve vyšším programovacím jazyce
- Proprietární řeení
- otevřené standardy: (VoxML, VoiceXML, CallXML)

Online nástorje pro implementaci dialovýcých rozhraní:

- Nuance Café,
- Tellme Studio
- Voxeo Prophecy

Desktopové:

- Trindikit
- CSLU toolkit
- Aspect Prophecy

Standardy W3C Voice Browser Activity

Historie:

- World Wide Web Consortium (W3C) je mezinárodní konsorcium, jehož členové společně s veřejností vyvíjejí webové standardy pro World Wide Web.
- Zloženo 1994
- VOice browser activity 1999

VoiceXML, SRGS, SSML, SISR, PLS,...

Základní info:

- jazyk pro popis dialogových rozhraní
- Cíl = výhody webového vývoje a doručení obsahu interaktivních hlasových aplikací
- 2000 - VoiceXML 1.0, kratce na to přijato jako standard W3C
-

Struktura VoiceXML applikací:

- VoiceXML dokument(y):
    - formluláře - končené stavoévé automaty
    - uživatel se nachází v jednom z konverzačních stavů
- Dva druhy dialogů:
    - formuláře - proces získání hodnot
    - menu
- Subdialogy - obdoba funkcí, (vrací hodnoty)
- Sezení - session
- Aplikace - sada dokumentů

VoiceXML formulář:

- základní komponenta dokumentů
- základní atribut = id

Položky formuláře:

- Vstupní položky - field, record, trnasfer, object, subdialog
- Vstupním položkám odpovídají proměnné
- řídící položky - block, initial

Element field
- představuje vstup od uživatele
- atributy má name, expr (hodnota), cond

Element record
- umožnuje systému nahrát zprávu, např dialogový záznamník

Element subdialog
- slouží k vyvolání dialogu, řešícího dílší problém (funkce)
- lze volat opakovaně
- kód subdialogu - formulář ukončený elementem return

Element block
- Obsahuje providitelný obsah

Element initial
- umožnuje uživateli zadat více informací naráz

SRGS

- nahrazuje JSGF
- Specifikace W3C
- Liší se pouze zápis nikoliv vyjadřovací síla

XML formát gramatiky:
- XML prolog.
- Kořenový element - grammar
- Atributy:
  - root - pravidlo
  - xml:lang - jazyk gramatiky
  - version
  - mode
  - …
- Element grammar - obsahuje množinu pravidel (elementů rule)

Sekvence:
- posloupnost terminálních a neterminácIních symbolů
- lze ji rozdělit na logické části

SISR

- Sémantika - přiřazuje význam slovům a promluvám
- standard W4C pro zpracování sémantiky promluvy
- umožnuje přiřazení základních interpretací částem promluvy a vytváření odvozených interpreteací pro nadřazená tvrtzení
    - přiřazení interpretace částem promluvy
    - odvozování
    - přiřazení vstupním poím dialogu

Odvození interpretace na základě dílčích interpreteací:

- Zápis pomocí ECMAScript
- Přiřazeno k pravidlům pomocí elementu tag

PLS

- Definuje značkování pro sepcifikaci slovníků výslovnosti pro podporu syntézy a rozpoznávání řeči
- Samozřejmě taky standard W3C

Základní elementy:

- Kořenový elemnt - lexikon
- lexeme - obsahuje popis pro jednu lexiální jednotku
- phoneme - obsahuje fonetický přepis dané lexikální jednotky

CCXML

- sloučí k ovládání řízení telefonních hovorů v průběhu interaktivních hlasových služeb

SCXML

- Souží k specifikaci konečných automatů

WIZARD OF OZ

- simulace dialogového rozhraní modelem člověk - člověk

Princip:

- Funkce dialogového rozhrají je (skrytě) simulována člověkem
- průběh dialogu je protokolován
- průběh se řídí navržečnou dialogovou strategií

Občas snaha navodit zdání že uživatel komunikuje s dialogovým systémem

MULTIMODÁLNÍ DIALOGOVÁ ROZHRANÍ

- mimo mluvenou řeč umožňuje i další způsoby komunikace člověk - počítač (textová, grafická, emoce,..)
- Výhody:
    - lepší přístupnost (pro neslyšící/nevidoucí)
    - lepší pochopení pragmatiky projevu

textová:

- prostě je zobrazen i text

Grafická:

- Talking heads
- komunikace znakovou řečí

- Široké spektrum možností zadávání vstupu uživatelem jinak než hlasem

Emoce:

- primární:
    - klasické i u živočichů
- sekundární
    - intelektuální, mordální, estetické
- Velkých šest:
    - hněv, zklamání, smutek, strach, překvapení
      // nechápu proč je jich jenom pět v prezentaci
- Detekce emocí:
    - biometrické vlastnosti
    - tlak, puls
    - dýchání elektciké aktivity moSku

Artificial Inteligence Markup Language (AIML)

- jazyk na bázi XML
- Popisuje znalostní bázi pro dotazovací systémy

Základní jednotky znalostí databáze:

- popisuje třídy objektů dat a částečně popisuje chování programů, které je zpracovávají
- objekty dat se skládají z jednotek zvaných témata a kategorie (struktorovaná nebo nestruktorovaná data)
- Používají se klíčová slova