Aplikácie audio rozhrania

Juraj Kačur

UMIKT, FEI STU, Bratislava

Obsah

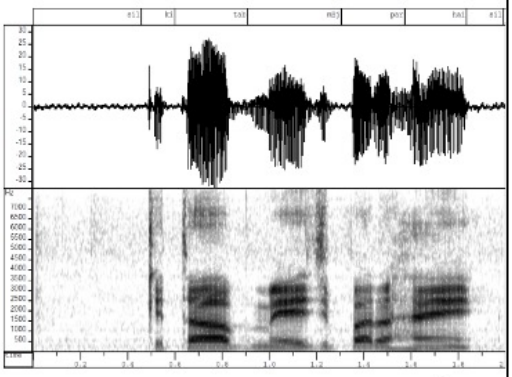
- Detekcia reči
- Potlačenie šumov
- Kompresia reči a audio signálov
- Identifikácia hovoriaceho
- Rozpoznávanie reči
- Syntéza reči
- Ďalšie použitia

2

Obsahom tejto state je vymenovanie základných aplikácii, ktoré môžu byt využite na strane stroja pri modalitie rozhrania využívajúcej zvukove signály, teda zmysel sluchu. Pre každú aplikáciu sú uvedene jej výhody , nevýhody, najčastejšie formy použitia, problémy s jej nasadením, rôzne druhy jej realizácie- typy postupov, atď..

Detekcia reči

- Úloha: zistiť či je v signály obsiahnutá reč a zdetegovať jej polohu
- Detekcia reči (VAD)
- Algoritmy využívajú známe vlastnosti reči
  - Nestacionárny signál
  - Frekv. Pásmo 300-4000 Hz
  - Periodické a neznelé časti
  - Dynamika reči 50dB
  - Rôzne spektrá
  - Plynú za sebou v čase
  - Dôležitá časová postupnosť



3

Najjednoduchšia a najzakladanejšia úloha je aby vôbec stroj vedel že na vstupe je rečový signál a odkiaľ po kial sa tento signál v case nachádza. Existujú rôzne algoritmy ktoré sa opierajú o základne vlastnosti rečových signálov.

Detekcia reči

- Využitie VAD
- Prenos rečových signálov
  - Systém prerušovaného vysielania (napr. GSM)
  - Systém variabilnej rýchlosti prenosu
- Systémy automatického rozpoznávania reči
  - Pre klasickú DTW je to nutnosť
  - Pre HMM VAD znamená: zrýchlenie, spresnenie a zjednodušenie celého procesu
- Zlepšenie kvality rečových signálov
  - Jednokanálový prístup: odhad šumu počas páuz a jeho následná filtrácia
- Analýza rečových signálov

4

Prehlaď kde sa takéto algoritmy najčastejšie nachádzajú

Odstránenie šumov

- Šumy
- Aditívne

  y=x+šum

  - Odhad spektra šumu z nerečových časti
  - Viacero prístupov
    - „Odčítanie“ priemerného šumu zo spektra zašumeného signálu
    - Periodický signál: priemerovanie periód
    - Existencia viacerých mikrofónov: zašumený a požadovaný(menej zašumený) signál, adaptívna filtrácia

5

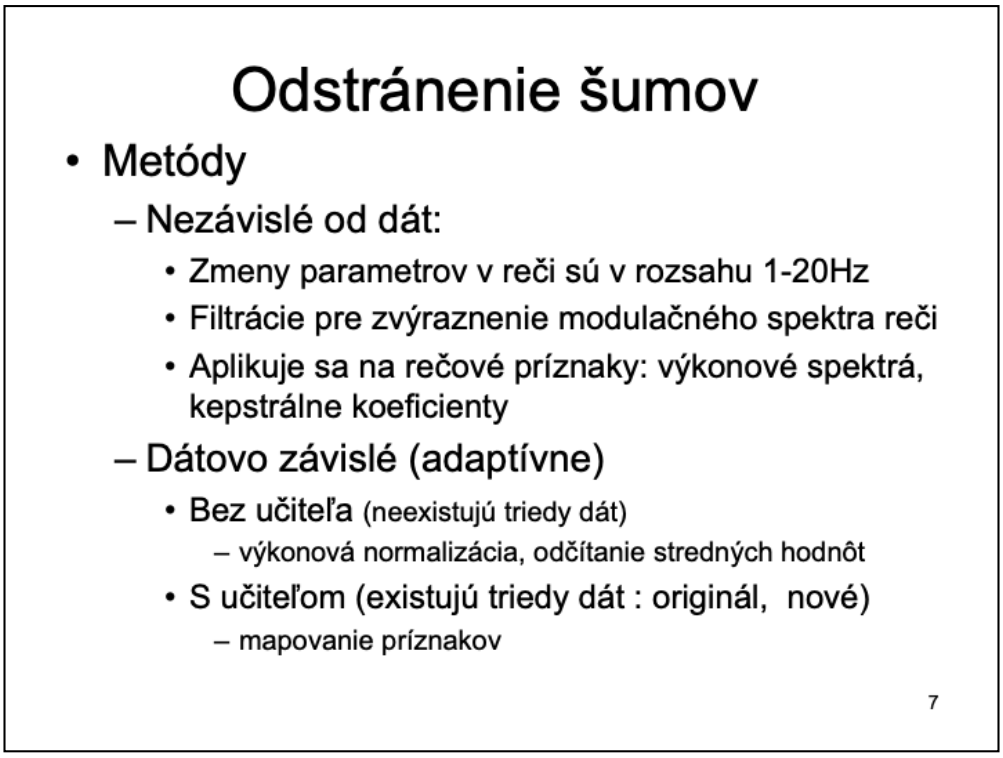
Ďalšou úlohou je potlačenie rôznych druhov šumov ktoré sa v reálnom prostredí môžu do užitočného audio signálu dostáť. Tu sú spomenuté 2 základné typy šumov- skreslení.

Odstránenie šumov

- Šumy
- Konvolútorne
  - Vplyv prostredia - „ozveny“

  Y=conv(x,šum)
    - Odhad impulzovej odpovede (filtra) miestnosti
    - Viacero prístupov
      - Inverzný filter
      - Podelenie spektra odhadom priemerného spektra signálu
      - Modifikácia príznakov „zvuku“ aby štatistické vlastnosti zodpovedali nejakému prostrediu, modelu

6

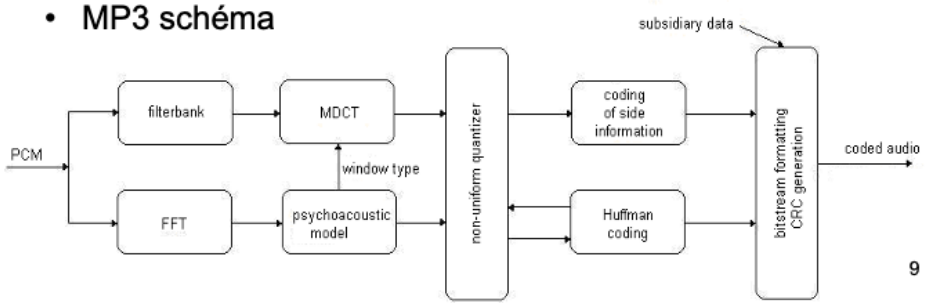


Na potlačenie týchto šumov sa používa veľa metód, ktoré sa podlá dostupných dát delia do rôznych tried.

Veľmi významnou skupinovo aplikácii pre rečové ako aj audio rozhranie sú algoritmy na kompresiu reči a kompresiu audio signálov. Tieto postupy značne (napr. Pre audio-hudbu až cca 12 krát) redukujú množstvo dátového priestoru na uchovanie a prenos týchto signálov. V nasledovnom sú spomenuté niektoré ich delenia a ich vlastnosti.

Kompresia audio signálov

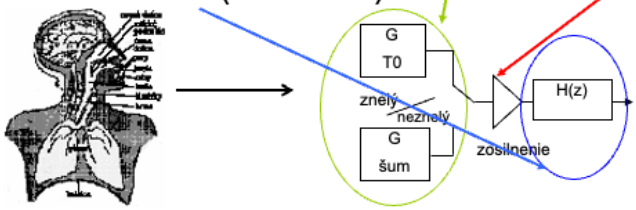
- Audio
- Je založená na využívaní psychoakustických princípov:
  - Absolútny prah počuteľnosti, frekvenčné maskovanie, časové maskovanie, nelineárné- psychoakustické frekvenčné mierky, nelineárne (log) vnímanie hlasitosti
- Najvýznamnejšie postupy: MP3, OGG, WMA; dosahujú cca 12 násobnú kompresiu
- MP3 schéma



Základom pre kompresiu audio signálov (hudba) je využitie nedokonalosti ucha teda odstránenie tých časti signálu ktoré ucho nevie vnímať. Vyžívajú sa tu poznatky z psychoakustiky najmä fenomén frekvenčného maskovania.

Kompresia audio signálov

- Reč
- Je založená na odhade parametrov lineárneho modelu produkcie reči
- Parametre: budenie (periodické/ “šum”), zosilnenie, parametre filtra (8-12 čísel)



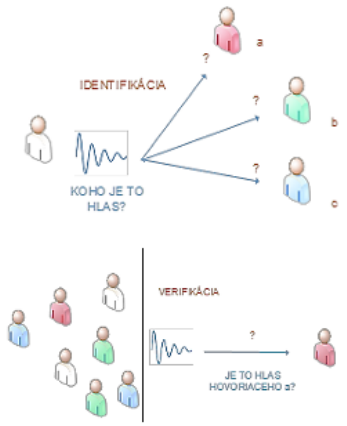
- Najvýznamnejšie postupy: CELP, ACELP,... Používajú sa vo VoIP a mobilných telefónoch
- Prenosová rýchlosť cca 9kbit/s až 800bit/s (hranica zrozumiteľnosti)

10

Pri kompresii rečových signálov, napr. Telefónia, sa naopak využíva známy model ako je reč produkovaná hlasovými organmi človeka. Potom parametre takéhoto modelu sú posielane, resp. uchovávane a nie samotný signál. Ten sa pomocou modelu a týchto parametrov vie jednoducho znova vygenerovať. Tým sa dosahuje ešte výraznejšia kompresia.

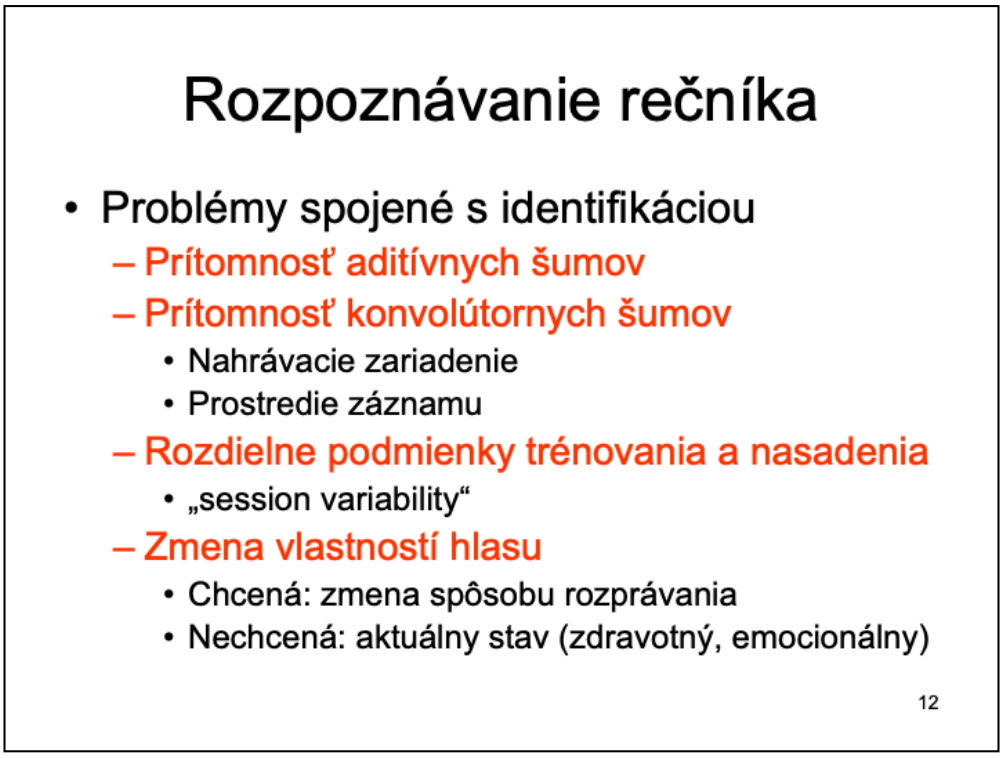
Rozpoznávanie rečníka

- Rozpoznávanie rečníka
  - Identifikácia rečníka
    - Kto zo skupiny hovorí?
    - Uzavretá skupina
    - Nekooperujúci používatelia
  - Verifikácia rečníka
    - Hovorí ten kto to tvrdí?
    - Otvorená množina
    - Kooperujúci používatelia



11

Často krát je potrebne na rečovom rozhraní riešiť problém identity hovoriacich. To riešia aplikácie z oblasti rozpoznávanie rečníka. Tie sa ďalej delia na problém identifikácie recnika a autentifikácie. Pri identifikácie je potrebne určiť kto zo skupiny rozpráva a pri autentifikácie je potrebne potvrdiť autentitu daného jedinca na základe jeho hlasu. To je možne preto lebo rečový signál nesie okrem lexikálnej informácie aj biometrickú informáciu (zdedenú aj naučenú)



V ďalšom sú rozobrane problémy, výhody a nevýhody využívania aplikácií rozpoznávania rečníka, ako aj rôzne spôsoby ich realizácie.

Rozpoznávanie rečníka

- Použitie systémov rozpoznania rečníka
  - Identifikácia rečníka
    - Kriminalistika
    - Automatické indexovanie rečových databáz (tzv. meta dáta)
    - Personifikácia služieb, nastavení, atď.
  - Verifikácia rečníka
    - Prístup k zabezpečeným údajom pomocou jednoduchej a neinvazivnej biometrie

13

Rozpoznávanie rečníka

- Podľa požadovaného vstupu
  - Textovo závislé
    - Vyžadujú presný text - frázu
    - Najpresnejšie (až 99% úspešnosť)
    - Menšie použitie
    - Využívajú systémy rozpoznania reči (HMM, DTW)
  - Textovo nezávisle
    - Neočakávajú konkrétny text
    - Úspešnosti 70-95%
    - Širšie a jednoduchšie použitie

14

Rozpoznávanie rečníka

- Výhody
  - Ľahko získaný biometrický signál
  - Neinvazívna metóda
  - Obsahuje viacero špecifických informácii o užívateľovi (fyzické, naučené)
- Nevýhody
  - Veľká variabilita rečníka
  - Nestálosť parametrov
  - Prítomnosť šumov (rozdiel medzi trénovaním a nasadením)
  - Ľahká reprodukcia (nahrávacie zariadenia)

15

Rozpoznávanie rečníka

- Existuje veľa príznakov rôznych vlastností
- Delenie:
  - Akustické
  - Prozodické
  - Príznaky vyššej úrovne

16

Rozpoznávanie rečníka

- Akustické príznaky (1. úroveň)
  - Odrážajú fyzické vlastnosti hlasových orgánov
  - Ľahko sa počítajú na úsekoch cca 10-30ms
  - Sú náchylné na zmeny prostredia a hovoriaceho (kompenzácia prostredia)
  - Najčastejšie
    - Obálka spektra

17

Rozpoznávanie rečníka

- Prozodické príznaky (2. úroveň)
  - Charakterové vlastnosti (naučené)
    - Dynamika, prízvuk, intonácia, rýchlosť, pauzy v reči, atď.
  - Ťažšie sa extrahujú a kvantifikujú
  - Počítajú sa z dlhších časových intervalov
  - Najúspešnejšie sú založené na vyhodnocovaní priebehu hlasivkovej periódy

18

Rozpoznávanie rečníka

- Príznaky vyššej úrovne (3. úroveň)
  - Jazyk
  - Dialekt
  - Slovná zásoba
  - Využívajú metódy rozpoznávania reči
  - Nie sú tak náchylné na vplyv prostredia
  - Málo diskriminujúce

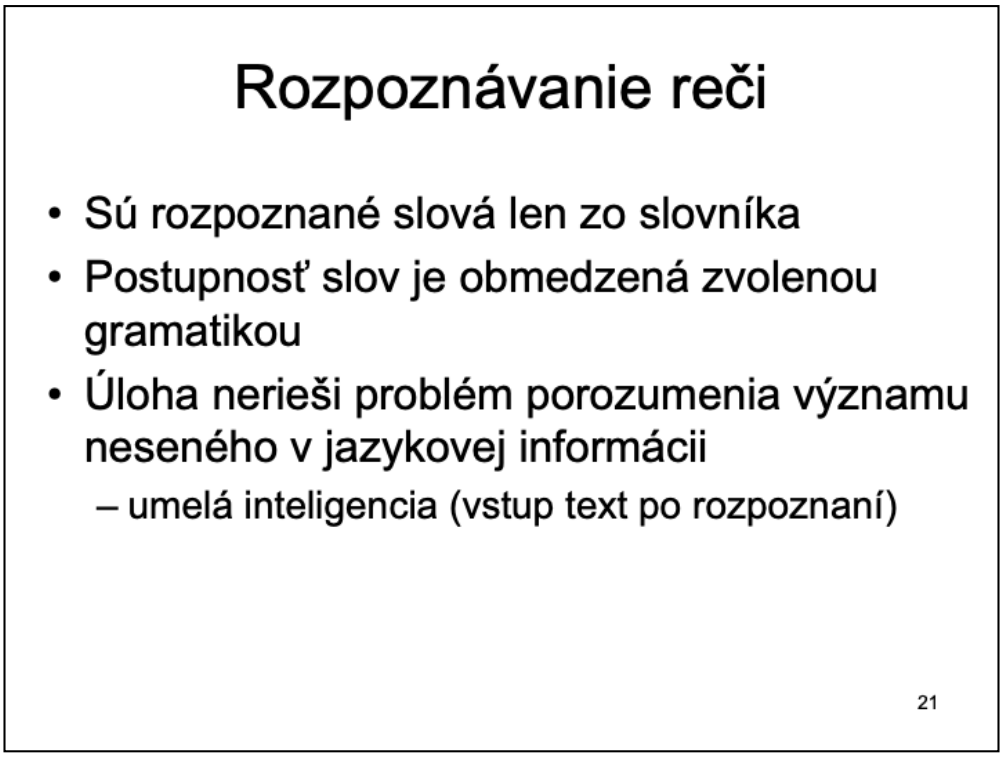
19

Rozpoznávanie reči

- Automatický prepis rečového signálu do textu
- Vstup rečový signál (PCM vzorky)
- Výstup: prepis čo bolo povedané, text

20

Asi najzložitejšou a najpraktickejšiu úlohov pre spracovanie reči je automatické rozpoznanie reči, teda úloha aby poctiac vedel prepísať (do formy textu) čo bolo povedane. Pozor to ešte nie je úloha aby významu aj porozumel a na to aj vhodne reagoval. Na to sú iné aplikácie ktoré spadajú do oblasti umelej inteligencie.



Zvyčajne takéto systémy majú rôzne očakávania čo sa týka vstupnej reči a obmedzenia na spôsob činnosti. Môžu preto existovať veľmi jednoduché systémy až po veľmi komplexne riešenia využívajúce veľké množstvo výpočtových a pamäťových prostriedkov. Preto sú j tie najzakladanejšie aspekty ďalej spomenuté ako aj spôsoby delenia systémov automatického rozpoznávania reči.

Rozpoznávanie reči

- Existuje veľa aplikácii, systémov a postupov riešenia
- Každé so svojimi predpokladmi, obmedzeniami na činnosť, formou vstupu a výstupu
- Majú rôzne zložitosti: systémové, časové, finančné,...

22

Rozpoznávanie reči

- Na ich korektné porovnanie je potrebná ich kategorizácia, zavedenie tried-kategórii
- Rôzne kritéria na ich delenie:
  - veľkosť slovníka
  - typ vstupnej reči
  - čas odpovedi systému
  - atď.

23

Rozpoznávanie reči

- Veľkosť slovníka
  - Veľmi dôležité vzhľadom na:
    - úspešnosť
    - čas výpočtu
    - zdroje systému
  - Rozlišuje sa:
    - Malý ~ (desiatky až stovky) slov
    - Stredný ~ (stovky až tisíce) slov
    - Veľký ~ sto tisíce slov

Pozn. počty sa menia vzhľadom na pokroky v technológiách

24

Rozpoznávanie reči

- Typ vstupnej reči
  - Rozpoznanie izolovaných slov
    - Na vstupe je len jedno slovo
  - Diktované prehovorenie
    - Slova idú za sebou, ale je medzi nimi dostatočne dlhá oddeľujúca pauza
  - Kontinuálna reč
    - Slová idú plynulo za sebou, predpokladá sa však obmedzenie na prostredie a typ reči
  - Prirodzená reč
    - Bežná konverzácia: šumy, chyby výslovnosti, artefakty rečníka, opakovanie slov, atď.

Úrovne zložitosti

- ľahká
- Stredná
- Najťažšia

25

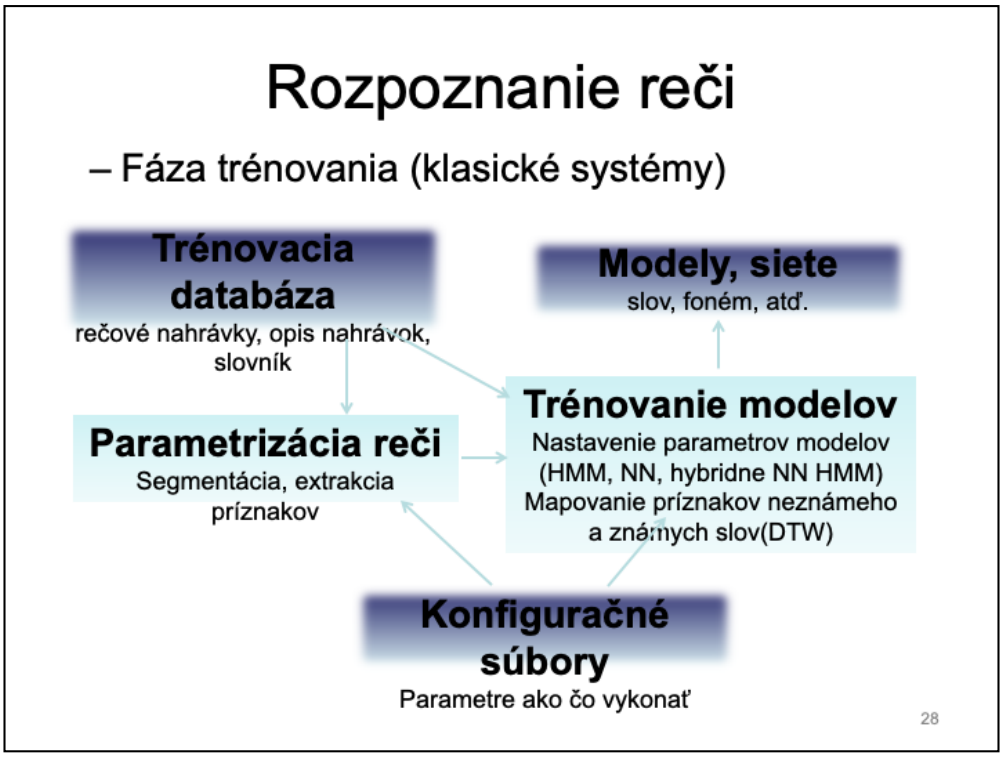
Rozpoznávanie reči

- Závislosť na hovoriacom
  - Závislé na hovoriacom
    - Vyššia úspešnosť
    - Menšia robustnosť pri zmene hlasu
    - Vhodný kompromis: adaptácia všeobecných modelov (natrénované na veľkom množstve hovoriacich) na konkrétneho užívateľa (existuje malé množstvo trénovacích dát)
  - Nezávislé na hovoriacom
    - Potenciálne menej presné
    - Majú väčšiu robustnosť

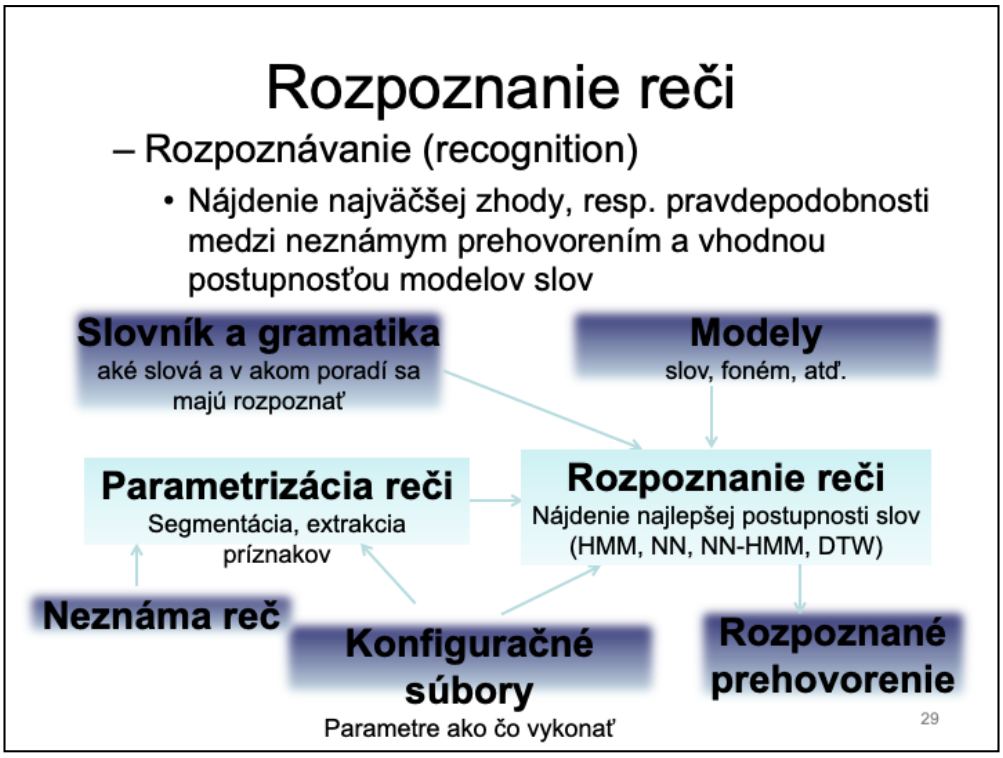
26

Rozpoznávanie reči

- Odozva systému
  - V reálnom čase
    - Tolerované oneskorenie cca sekunda/y
    - Je len jeden prechod cez všetky hypotézy
    - Zobrazujú sa postupne len čiastočné hypotézy; ešte pred skončením prehovorenia
    - Neoptimálne riešenie, menšie úspešnosti
  - „Off line“
    - Nemajú stanovenú časovú odozvu
    - Môžu čakať na skončenie prehovorenia
    - Môžu brať do úvahy všetky (viacero) hypotézy
    - potenciálne presnejšie
    - Môžu vykonať viacero prechodov cez sieť hypotéz



Pred samotným nasadením systému rozpoznávania reči je ho najprv potrebne natrénovať na danú reč, pripadne prostredie a rečníka. Preto existujú dve fázy a to fáza trénovcia a fáza nasedenia (rozpoznavania) systém. V závislosti od komplexnosti systému fáza trénovacia predstavuje najzložitejšiu úlohu tak z pohľadu teoretického, t.j. nájdenie vhodneho modelu reči a jeho následne nastavenie na reálnych dátach, tak z pohľadu praktickej realizácie, množstvo zložitých výpočtov, dát, parametrov, nastavení. Existuje viacero fundamentálnych metód rozpoznávania reči, napr. jednoduché porovnávanie vzorov (DTW), štatistické modelovanie reči (HMM) s modelovaním jazyka-gramatiky, metódy založene na rôznych typoch neurónových sieti az po tzv. end 2 end systemy. Každá z nich vyžaduje iné vstupne dáta, nastavenia a spôsoby trénovania. Všeobecná schéma trénovania je zobrazená na nasledovnom grafe.



Pri fáze rozpoznávania sa už použije natrénovaný model a na základe neho sa zo vstupného signálu vyberie najpravdepodobnejšia hypotéza, ktorá zodpovedá danému prehovoreniu. To ako sa to konkretne robí závisí od metódy. Všeobecná schéma fázy rozpoznávania je zobrazená na nasledovnom obrázku.

Rozpoznávanie reči

- DTW (dynamic time warping)
  - Rozpoznávanie izolovaných slov
  - Porovnáva slovo zo slovami zo slovníka
- HMM – Skryté Markovové modely
  - Spojitá reč
  - Štatistická technika modelovania dynamických systémov
  - Každá reči je modelovaná “HMM” modelmi

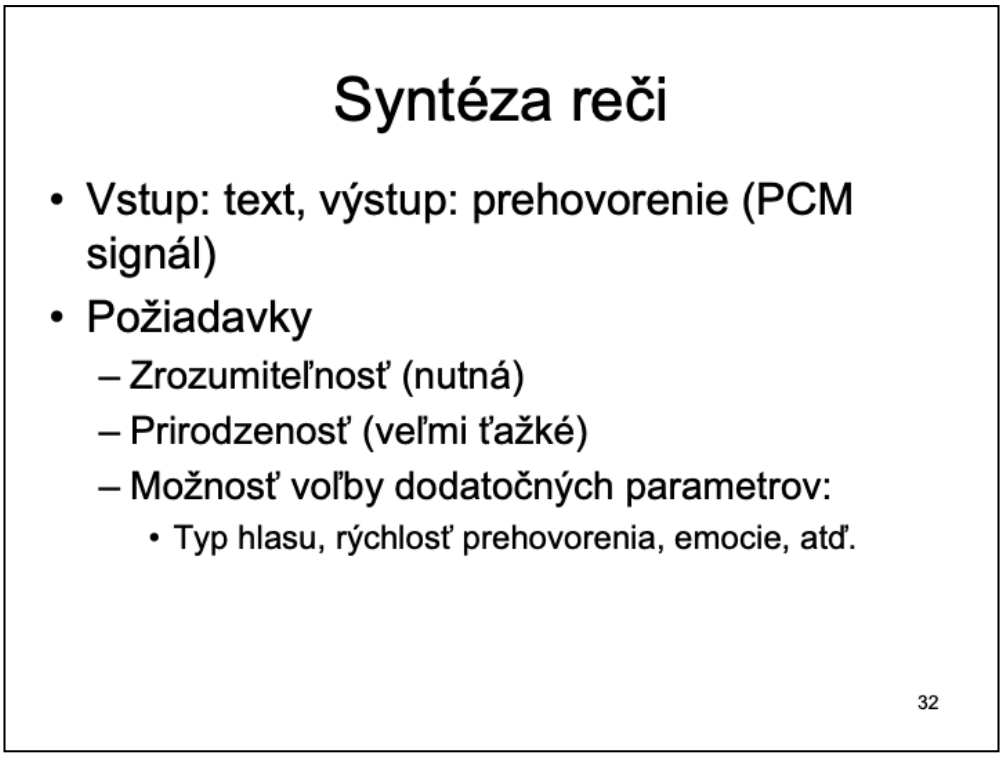
30

V ďalšom sú v stručnosti spomenuté najznámejšie používané postupy; niektoré sú už skôr historické.

Rozpoznávanie reči

- Hybridné neurónové siete a HMM
  - NN akustické modelovanie, HMM časove
  - najbežnejšie a najlepšie výsledky
  - Komplikované- znalosti z viacerých oblasti
- Neurónové siete
  - Najnovšie, vela dát, dlhé trénovanie
  - Jednoduchšie použitie bez hlbších znalosti (jazyk, fonológia, modely- jazyka, gramatiky)
  - End 2 end systémy: vstup signál alebo príznaky a výstup text bez ďalších medzistupňov, napr. Transformery, conformery, attention-encoder decoder, rekurentné siete

-31



Ďalšou veľkou oblasťou na audio rozhraní je syntéza reči. Teda umele generovanie prehovorenia (akustická forma) počítacom. Využiva sa vtedy keď poctiac prezentuje užívateľovi informáciu formou reči. V ďalšom sú spomenuté základné požiadavky na syntézu, jej vlastnosti, rôzne spôsoby realizacie so svojimi výhodami a nevýhodami.

Syntéza reči

- Existuje viacero prístupov
  - Syntéza spájaním jednotiek
    - Nájdenie a spojenie vhodných časti reči z nahovorenej databázy
    - najviac prirodzene znejúca syntetizovaná reč
    - Typy
      - Výber jednotiek: fonémy, slabiky, slová, vety. Spájanie jednotiek z databázy tak, aby plynulo nadväzovali, veľká databáza
      - Difónová: len dvojice foném, prechod medzi nimi je v strede segmentu, malá databáza
      - Syntéza na špecifickej oblasti: z nahratých slov a fráz vytvára reč. Používa sa v aplikáciách s limitovaný rečovým výstupom na určitú oblasť

33

Syntéza reči

– Formantová syntéza
  - Nepoužíva vzorky ľudskej reči
  - Reč vytvorená zo zvukových modelov
  - Parametre: základná hlasivková frekvencia, počet tvar a poloha rezonančných frekvencií (vrcholov spektra), časy trvania
  - Dobrá kontrola nad vlastnosťami reči
  - Problémy aby to znelo prirodzene

34

Syntéza reči

– Artikulačná syntéza
  - Je založená na modeli ľudského hlasového traktu a artikulácie

– HMM syntéza
  - Parametre ako frekvenčné spektrum, základná hlasivková frekvencia, prozódia reči sú modelované pomocou HMM
  - Priebeh reči je generovaný pomocou HMM a maximálnej pravdepodobnosti generovania parametrov za predpokladu textu

35

Okrem týchto základný aplikácii existuje viacero menej častých resp. novo skúmaných a nasadzovaných aplikácii. Tieto budú jednotne spomenuté v tomto odseku. Zaujímavá a aj potrebná úloha je detegovať emócie z reči. Jej výhody, nevýhody, a problémy s ňou spojene sú uvedené v ďalšom texte.

Ostatné aplikácie

- Problémy:
  - viacero emócii naraz
  - Rôzny prejav emócii- kultúrny, individuálny
  - Predstieranie, resp. hranie emócii: trénovacie databázy
  - Rozlišujeme cca 5 až 7 základných emócii: neutrálna, hnev, prekvapenie, radosť, odpor, nuda
  - Niekedy sa presne nemeria konkrétna emócia ale skôr trieda emócii
- Rozpoznávanie udalosti v audio signáloch
  - Rozbite sklo, výstrel, krik, plač
  - Automatické bezpečnostne dohľadové systémy (kamery všetko nevidia, zvuky sa šíria ďalej aj do nepokrytých miest)

37

Ďalšou dôležitou úlohou ktorá sa bude postupnejšie masívnejšie nasadzovať je detekcia resp. rozpoznanie rôznych udalosti zo zvukových signáloch. Jej využitie bude najmä v monitorovacích a bezpečnostných systémoch.

Ostatné aplikácie

- Monitorovacie systémy:
  - Živé, neživé a umelo vytvorené procesy
  - Či výrobné procesy prebiehajú tak ako majú
  - výskyt prírodných dejov, zvierat, atď..
  - Využívajú sa aj dlhšie časove intervaly: trvanie daného deju
  - Širšie frekvenčné pásma ako len reč, napr. 20- 20kHz, niekedy aj viac
  - Vzhľadom na rôznosť dejov sa môžu sledovať rôzne parametre signálu
  - Momentálne najlepšie systémy využívajú NN

38

Ostatné aplikácie

- Rozpoznávanie prostredia z audio signálu
  - Z nahrávky zistiť prostredie kde to bolo nahraté
  - Istým spôsobom je to pod úloha detekcie udalosti
- Detekcia príbuzenstva
  - Určiť zo zvukových – rečových nahrávok jedincov ktorí z nich sú v príbuzenskom vzťahu
  - Využívajú sa nie len biologické- fyzické podobnosti ale aj naučené správanie, ako sa ľudia v spoločnom vzťahu podobne vyjadrujú, atď..
  - Kvôli komplikovanosti úlohy, kde nie je úplne zrejme čo skúmať a vyhodnocovať sa s výhodou používajú NN prístupy

39

Ďalšie aplikácie môžu byt rozpoznanie prostredia, určenie príbuzenstva užívateľov na základe reči, vkladanie skrytej informácie do rečových resp. audio signálov (watermarking), atď.

Ďakujem za pozornosť

juraj.kacur@stuba.sk