

Rozpoznávání a syntéza řeči

PLIN059

Mgr. Dana Hlaváčková, Ph.D.

Rozpoznávání řeči

- Automatic **Speech Recognition** (ASR)
- Speech-To-Text (STT)
- odlišit od – **rozpoznávání hlasu** – identifikace mluvčího (verifikace, bezpečnostní systémy)
- převod zvukového signálu na text
 - bigramy, n-gramy, pravděpodobnost výskytu slov na základě předchozího kontextu
 - velikost slovníku, korpusy
- vliv intonace, výslovnosti a okolního šumu
- komplikovanější pro flektivní jazyky

Rozpoznávání řeči

- rozpoznávání **izolovaných slov**
 - hlasové povely – hlasové ovládání počítače, hry, automobilu
- rozpoznávání **spojité řeči**
- jeden mluvčí – Speaker Dependent
- více mluvčích – Speaker Independent (možnost adaptace)
- **meze:**
 - mikrofon + eliminace šumu
 - zřetelná výslovnost
 - doménově specializovaná oblast
 - omezená slovní zásoba (právo, medicína)
 - natrénováno na jednoho mluvčího

Rozpoznávání řeči

- ověřování hlasem
- diktovací systémy
- transkripce audio nahrávek
- přepisy záznamu televizních a rozhlasových pořadů
- titulkovací systémy
- pomoc handicapovaným osobám (hlasové ovládání)
- dialogové systémy, automatická spojovatelka
- mobilní aplikace

Pracoviště v ČR

- ZČU Plzeň – projekt **MUSSLAP** (*musslap.zcu.cz*)
 - akustické rozpoznávání řeči
 - rozpoznávání znakové řeči
 - audiovizuální rozpoznávání řeči
- ZČU Plzeň - spin-off **Speechtech** (*speechtech.cz*)
 - SpeechTech ASR
 - MegaWord – NovaVoice, se společností Consulting Company Novasoft (*ccnovasoft.cz*)

Pracoviště v ČR

- TU Liberec – Laboratoř počítačového zpracování řeči
SpeechLab (*ite.tul.cz/speechlab*)
 - diktovací systém NewtonDictate, s firmou Newton Technologies
- VUT Brno – Fakulta informačních technologií
 - Ústav počítačové grafiky a multimédií
 - **Speech@FIT** – výzkum zaměřený na verifikaci hlasu
 - spin-off Phonexia (<http://www.phonexia.com/>)

Ukázky

NewtonDictate

- <http://www.youtube.com/watch?v=W0BT6DKzF3s>
- a další videa na YouTube

- pro pobavení:
- Windows Vista
<http://www.youtube.com/watch?v=kX8oYoYy2Gc>

Syntéza řeči

- **Speech Synthesis**
- Text-To-Speech (TTS)
- subslovní řečové segmenty, konkatenace (řetězení)
- **difóny** (jednotka od poloviny jedné hlásky do poloviny druhé)
- **trifóny** (kontext hlásky)
- modelování prozodických charakteristik řeči (melodie, tempo, hlasitost)
- mluvené (řečové) korpusy
- kvalita posuzována podle podobnosti lidskému hlasu

Syntéza řeči

- služby telefonních systémů
- software pro zrakově postižené osoby – odečítače obrazovky (screenreader, *pristupnost.cz/screen-readery*)
- spojení s vizualizací (audiovizuální syntéza)
- SpeechTech
- MUSSLAP (mluvící hlava)
- Acapela (<https://www.acapela-group.com/>)
- Google WaveNet – neuronové sítě, <https://cloud.google.com/text-to-speech>

Ukázky

- iPhone 4S (Apple) – asistentka Siri, rozumí a odpovídá (vliv šumu)
- <http://www.youtube.com/watch?v=TLtPMggOzD0>
- <https://www.youtube.com/watch?v=XSp0jbaSBZs>