

Rozpoznávání a syntéza řeči

Rozpoznávání řeči

- Automatic **Speech Recognition** (ASR)
- Speech-To-Text (STT)
- odlišit od – **rozpoznávání hlasu** – identifikace mluvčího
- 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, automobily
- 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

- diktovací systémy
- transkripce audio nahrávek
- přepisy záznamu televizních a rozhlasových pořadů
- titulkovací systémy
- pomoc handicapovaným osobám (hlasové ovládání)
- dialogové systémy, automatická spojovatelka
- mobilní aplikace
- systémy orientované na kvalitu řečového signálu

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)
- TU Liberec – Laboratoř počítačového zpracování řeči **SpeechLab** (ite.tul.cz/speechlab)
 - diktovací systém NewtonDictate, s firmou Newton Technologies

Ukázky

- NovaVoice <http://www.youtube.com/watch?v=CK9dxFJh47Q>
- NewtonDictate
<http://www.youtube.com/watch?v=W0BT6DKzF3s>
- Windows Vista
<http://www.youtube.com/watch?v=kX8oYoYy2Gc>
- Windows 7 Dragon NaturallySpeaking Version 11.5
<http://www.youtube.com/watch?v=VJ0i1dHZH48>

Syntéza řeči

- **Speech Synthesis**
- Text-To-Speech (TTS)
- subslovní řečové segmenty, konkatenace
- **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 (*acapela-group.com*)

Ukázky

- iPhone 4S (Apple) – asistentka Siri, rozumí a odpovídá (vliv šumu)
- Siri <http://www.youtube.com/watch?v=6FBOUY7NPIA>
- <http://www.youtube.com/watch?v=TLtPMggOzD0>
- Trulyhandsfree Voice Control 2.0 (Sensory) – aktivace hlasem, dosah 6 m, potlačení šumu