

Rozpoznávání a syntéza řeči

PLIN059

Mgr. Dana Hlaváčková, Ph.D.

Rozpoznávání a syntéza řeči

- **další z úkolů NLP**
- **analýza textu**
 - znaky a řetězce znaků
 - rozpoznání i produkce (strojový překlad, automatické generování textů)
- **analýza řeči**
 - zvukový signál
 - převod signálu na text (znaky)
 - převod textu na zvukový signál (co nejvíce podobný lidské řeči)

Rozpoznávání řeči

- **Speech-To-Text (STT), Automatic Speech Recognition (ASR)**
- **rozpoznávání hlasu**
 - identifikace mluvčího (Speaker Identification)
 - rozpoznání věku a pohlaví
 - odlišení mluvčího od ostatních hlasů (Speaker Diarization)
 - verifikace, bezpečnostní systémy, Voice Biometrics Technologies, např. [Phonexia](#)
- **hlasové povely**
 - hlasové ovládání počítače, hry, mobilu, domácnosti, automobilu...
 - Voice Assistant – Google, Siri, Alexa, chatbots – Gemini
 - rozpoznávání izolovaných slov nebo spojitě řeči

Rozpoznávání řeči

- **převod zvukového signálu na text**
- **pravidlové systémy – HMM (Hidden Markov Model)**
 - stavy a přechody mezi stavy
 - bigramy, n-gramy, pravděpodobnost výskytu slov na základě předchozího kontextu
 - velikost slovníku, vliv intonace, výslovnosti, přízvuku a okolního šumu, komplikovanější pro flektivní jazyky
- **neuronové sítě – transformery**
 - úspěšnější v rozpoznávání i syntéze
 - náročné na množství dat, dlouhý čas trénování, black box

Rozpoznávání řeči

- jeden mluvčí – **Speaker Dependent (SD)**
- více mluvčích – **Speaker Independent (SI)**
 - možnost adaptace
- **meze:**
 - mikrofon + eliminace šumu, zřetelná výslovnost
 - doménově specializovaná oblast
 - omezená slovní zásoba (právo, medicína)
- **možnosti**
 - běžně mluvený jazyk, dialekty, přízvuky
 - rozpoznávání a syntéza emocí
 - komplexní systémy pro více jazyků

Rozpoznávání řeči – uplatnění

- ověřování hlasem, hlasové povely
 - pomoc handicapovaným osobám (hlasové ovládání)
- diktovací systémy
- transkripce audio nahrávek
- přepisy záznamu televizních a rozhlasových pořadů
- titulkovací systémy, automatické titulky (v reálném čase)
- analýza sentimentu
- překlad v reálném čase
- hlasoví asistenti

Syntéza řeči

- **Text-To-Speech (TTS), Automatic Speech Synthesis (ASS)**
- modelování zvukového signálu do podoby lidské řeči
- **konkatenáční syntéza** (řetězení)
 - subslovní řečové segmenty, **difóny** (jednotka od poloviny jedné hlásky do poloviny druhé)
 - **trifóny** (kontext hlásky)
 - modelování prozodických charakteristik řeči (melodie, tempo, hlasitost)
- **neuronové sítě**, Deep Learning Synthesis, Deep Neural Networks (DNN)
 - přirozeně znějící hlasy
 - syntéza přízvuků, emocí, stylu mluveného projevu

Syntéza řeči – uplatnění

- služby telefonních systémů
- hlášení na nádraží, v hasičském sboru apod.
- software pro zrakově postižené osoby – odečítače obrazovky
- spojení s vizualizací (audiovizuální syntéza řeči)
- předabování videí
- využití hlasů u podcastů, dokumentů, zpráv, sociálních sítích apod.
- audio deepfakes

Pracoviště v ČR

- **ZČU Plzeň** – Katedra kybernetiky Fakulty aplikovaných věd, **Výzkumné centrum NTIS**, <http://www.kky.zcu.cz/cs>
- firma **SpeechTech** (<https://www.speechtech.cz>)
 - diktovací systém *MegaWord*, syntéza řeči
- **TU Liberec** – Ústav informačních technologií a elektroniky Fakulty mechatroniky, informatiky a mezioborových studií
- **Laboratoř umělé inteligence** (<https://ailab.ite.tul.cz/>)
 - diktovací systém *NewtonDictate*, s firmou **Newton Technologies** (<https://www.newtontech.net/cs/>)
 - *Beey* – platforma pro titulkování v 18 jazycích

Pracoviště v ČR

- VUT Brno – Fakulta informačních technologií
 - **Ústav počítačové grafiky a multimédií**
(<https://speech.fit.vut.cz/>)
 - **Speech@FIT** – výzkum zaměřený na verifikaci hlasu a identifikaci mluvčího
 - **Phonexia** (<http://www.phonexia.com/>)

Ukázky

- Využití v psychoterapii
- **ELIZA** computer therapist <https://cs.wikipedia.org/wiki/ELIZA>
- **elysai** – Promethist <https://promethist.ai/>

- **Google Cloud** – <https://cloud.google.com>
- <https://revoicer.com/>
- <https://elevenlabs.io/>