

# 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, neuronové sítě
  - velikost slovníku, korpusy
- vliv intonace, výslovnosti, přízvuku 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 – **Katedra kybernetiky** Fakulty aplikovaných věd, Oddělení umělé inteligence, <http://www.kky.zcu.cz/cs>
  - akustické rozpoznávání řeči
  - audiovizuální rozpoznávání řeči
  - rozpoznávání znakové řeči
  - titulkovací systémy
- původně spin-off **SpeechTech** (<https://www.speechtech.cz>)
  - diktovací systém NovaVoice (technologie MegaWord), se společností Consulting Company Novasoft (<https://ccnovasoft.cz>)

# Pracoviště v ČR

- TU Liberec – Ústav informačních technologií a elektroniky  
Fakulty mechatroniky, informatiky a mezioborových studií
- **Laboratoř počítačového zpracování řeči SpeechLab**  
(<http://speechlab.tul.cz/index.html>)
  - diktovací systém NewtonDictate, s firmou Newton Technologies (<https://www.newtontech.net/cs/>)
- VUT Brno – Fakulta informačních technologií
  - **Ústav počítačové grafiky a multimédií**
  - **Speech@FIT** – výzkum zaměřený na verifikaci hlasu
  - původně spin-off Phonexia (<http://www.phonexia.com/>)

# Ukázky

Katedra kybernetiky

<https://www.youtube.com/channel/UCgLPJ6VT9rmwQ3GuZBajZ2w>

NewtonDictate

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

pro pobavení:

- Windows Vista Speech Recognition  
<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*)
- ZČU <http://www.kky.zcu.cz/cs>
- spojení s vizualizací (audiovizuální syntéza řeči)
- SpeechTech <https://www.speechtech.cz>
- Acapela (<https://www.acapela-group.com/>)
- Google WaveNet – neuronové sítě, <https://cloud.google.com/text-to-speech>

# Dialogové systémy

- Google Assistant <https://assistant.google.com/>
- Amazon Alexa [https://cs.wikipedia.org/wiki/Amazon\\_Alexa](https://cs.wikipedia.org/wiki/Amazon_Alexa)
- ELIZA computer therapist <https://cs.wikipedia.org/wiki/ELIZA>
- Talk to Poppy – terapie <https://www.talktopoppy.com/>
- <https://promethist.ai/>
- iPhone – asistentka Siri, rozumí a odpovídá  
<http://www.youtube.com/watch?v=TLtPMggOzD0>