

Rozpoznávání řeči



Úvod

- Automatické rozpoznávání řeči - ASR (Automatic Speech Recognition)
- Snaží pracovat podobně jako lidský mozek
- Dnes již umí řeč převést do textu, ale stále nerozumí obsahu
- Nejpoužívanější - statistický přístup



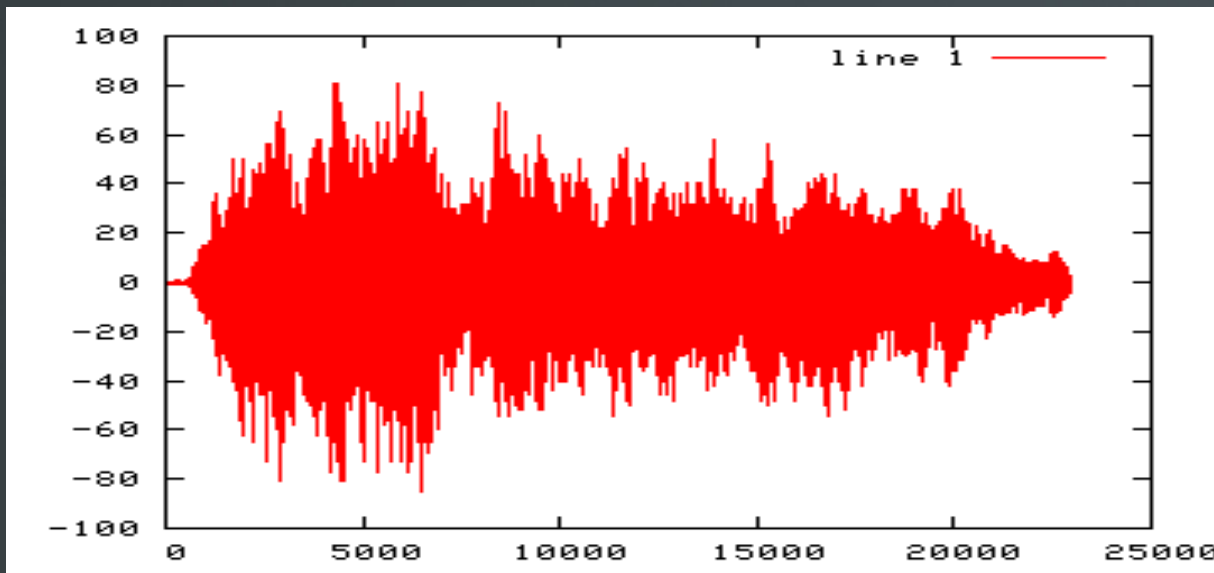
Historie

- ASR existují již od počátků počítačů
- První systémy – jen několik slov, porovnávaly proti nahrávkám, závislé na řečníkovi (nižší úspěšnost)
- S růstem paměťové a výkonové kapacity rostla i míra schopnosti rozpoznávat řeč
- Dnešní ASR systémy rozpoznají až milióny slov



Zvuk a řeč

- Řeč je ve své fyzikální podobě spojitá funkce změny akustického tlaku
- Člověk, ale zvuk slyší jako tóny – slyší jejich frekvenci



Fonémy

- Nejmenší lingvistická jednotka schopná rozlišovat významové jednotky
- Princip minimálních párových slov (liší se jedním fonémem)
 - [a] – [á] hrábě - hrabě, [t] – [d] ten – den
 - Ale [n] a [ŋ] nemění význam – nejsou 2 odlišné fonémy




Princip rozpoznávání

- Akustický model
- Jazykový model
- Výslovnostní slovník



Akustický model

- Amplituda akustického vlnění je snímána v pravidelných intervalech a uložena ve formě celého čísla (digitalizace a vzorkování signálu) – příznaky pro každý interval – vstupem akustického modelu
 - Anotovaná řečová data - jak který foném zní. (jak vypadá jeho sekvence příznaků)
 - Sada možných fonémů
 - Vytvořen statistický model – ukazuje, jak vypadá matice příznaků jednotlivých fonémů
- 

Jazykový model

- Každý jazyk má svá pravidla – gramatiku
- Posbírají se statistiky výskytů dvojic nebo trojic slov → vytvoří se jazykový model, který modeluje pravděpodobnost výskytu posloupnosti slov.



Výslovnostní slovník

- Přiřadí každému slovu výslovnost (posloupnost fonémů) a propojí tak jazykový model s akustickým modelem.
- Z akustického, jazykového modelu a slovníku se vytvoří rozpoznávací síť na jejímž základě je řeč rozpoznávána



Zajímavosti

- Anglické rozpoznávače obsahují kolem 50 000 slov
- České mají 500 000 – 1 000 000 slov, klidně i 4 000 000



Ovládání hlasem

- Ovládání počítače hlasem – hlavní operační systémy již začaly tyto funkce podporovat
- Dragon Naturally Speaking, Simon, MyVoice (Liberec)
- Ovládání mobilů – Apple, Android...
- Ovládání inteligentích systémů domu
- Bezdrátové hlasové ovládání domácích zařízení – SmartRoom (Liberec)



Ovládání hlasem

- Ovládání invalidního vozíku hlasem
- Ovládání nastavení postele u lidí trvale připoutaných na lůžko



Diktování, řeč na text

- Přímé zapisování mluvené řeči – odpadá nutnost přepisování/zapisování – soudy, kanceláře, medicínské využití, zápis přednášek na univerzitách
- Psaní (diktování) textů na počítači – MS Word
- Sphinx4, Voice to Text (FIT VUT Brno), MyDictate, Dragon Naturally Speaking



Další využití

- Textové vyhledávání ve videu (VUT), automatické otitulkování videa (youtube.com)
- Kombinace s dialogovým systémem, kdy vstupem je naše řeč
- Rozpoznávání mluvčího – kdo, mluví
- Jedna z možností autentizace – voice recognition



Závěr

- Uspěšnost současných ASR se pohybuje nad kolem 94%
- Už si lépe poradí s šumem a ruchem, nezávislé na mluvčím
- Vědci odhadují, že srovnatelné s člověkem budou nejdříve za 20 let
- Přesto jsou dnes velmi užitečnými pomocníky a hendikepovaným umožňují vykonávat činnosti, které by dříve dělat nemohli



Videoukázky

- FIT VUT – vyhledávání v řeči
<http://www.prednasky.com/>
- Dragon Naturally Speaking – ovládání počítače hlasem a diktování -
<http://www.youtube.com/watch?v=ZZ0N3WpmZho>
- MyVoice, MyDictate – TU Liberec
<http://www.fugasoft.cz/index.php?cont=myvoice&su>
- Pro pobavení – prezentace MS VISTA -
<http://www.youtube.com/watch?v=kX8oYoYy2Gc>



Zdroje

- PSUTKA, Josef. Mluvíme s počítačem česky. Vyd. 1. Praha : Academia, 2006. 746 s. ISBN 8020013091.
- <http://www.osel.cz/index.php?clanek=5152>
- <http://www.fugasoft.cz/index.php>



Děkuji za pozornost.

