

PB095 - Úvod do počítačového zpracování řeči

Luděk Bártek

Fakulta Informatiky
Masarykova Univerzita
Brno

podzim 2021

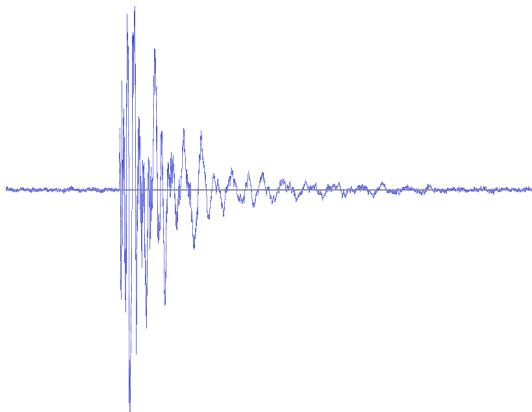
Obsah

- 1 Stručná historie zpracování zvuku a řeči

Co je to zvuk?

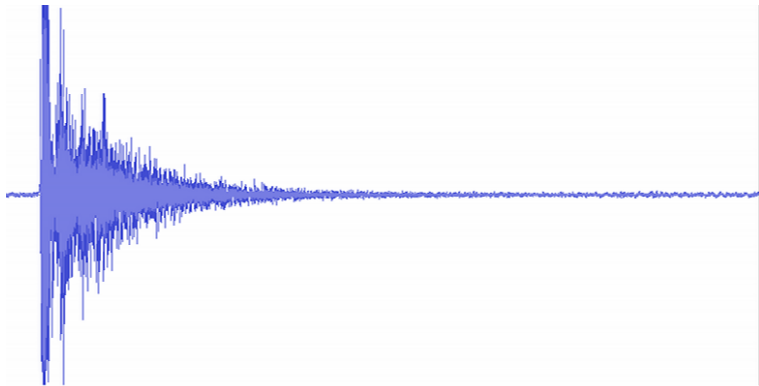
- Akustický signál.
- Jedná se o kmitavý pohyb molekul pružného prostředí.
 - vzduch
 - voda
 - kov
 - ...
- Vyvolán odporem prostředí - vede k opakovanému stlačování prostředí.
- Podrobněji v části fyzikální akustika.

Ukázka zvuku



Zvuk klepnutí na plastové tělo počítače
(images/klepnuti-plast.wav)

Ukázka zvuku

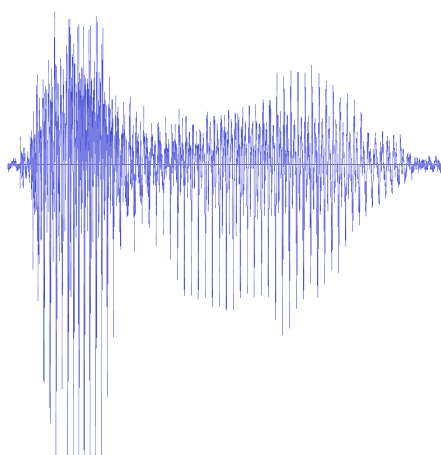


Zvuk tlesknutí (images/tlesknuti.wav)

Co je to řeč?

- Akustický signál a gesta sloužící ke komunikaci.
- Obsahuje definované vzory (slova), která jsou dána jazykem.
- Velmi rozvinutý u člověka.
 - Příznaky schopnosti tvorby artikulované řeči již u Australopitéka (-3 milióny let).
 - Slouží ke sdělování: myšlenek, pocitů, emocí, ...
 - myšlenek - "Dnes budeme probírat láčkovce."
 - pocitů - "Je mi krásně.", "Radši se ke mně ani nepřibližuj!", "Au!"
 - emocí - "Jé!", "Ach jo!", "Hurá!", ...
- Určité formy akustické komunikace (řeči) lze pozorovat i u dalších vyšších živočichů:
 - způsob zajištění kooperace při získávání obživy (kytovci, vlk, ...)
 - vábení partnera (jelen, ...)
 - vyjádření emočních stavů (pes, opice, ...).
 - ...

Ukázka řeči



Zvukový záznam (images/ahoj.wav)

Studium řeči

- fyzika - akustika
- biologie - medicína (fyziologie, fyziologická akustika)
- jazykověda - fonetika

Přehled historie zpracování a napodobování řeči

- Schopnost artikulované řeči - australopitekus - cca. -3 000 000 let
- Starověk - budování mluvících soch
- Galileo Galilei - souvislost mezi tónem a frekvencí
- 1779 - Christian Gottlieb Kratzenstein - systém rezonátorů pro samohlásky a, e, i, o, u



Přehled historie zpracování a napodobování řeči

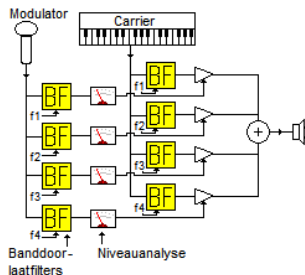
- 1791 - (Johann) Wolfgang von Kempelen (de Pázmánd) - první mechanický řečový syntetizér



- 1835 - zrekonstruován a upraven Wheatonem - navíc pružná "ústní dutina".
- 1846 - J. Faber - mluvící stroj Euphonia

Přehled historie zpracování a napodobování řeči

- 1937 - R. R. Riesz - mechanický mluvící stroj
- 1939 - H. Dudley
 - VODER - elektromechanický řečový syntetizér
 - VOCODER - systém pro kódování a přenos řeči



- 50. léta 20. století - syntéza ve frekvenční oblasti
 - později v časové oblasti
- 70. léta 20. století - počítačové zpracování zvuku

Historie analýzy řeči

- 19. století porozumění principů tvorby a zpracování řeči (rezonanční teorie, základy fonetiky):
 - J. B. Fourier - Fourierova věta
 - principy spektrální analýzy zvuku
 - H. Helmholtz
 - fyziologie vnímání hudby
 - Helmholtzův rezonátor

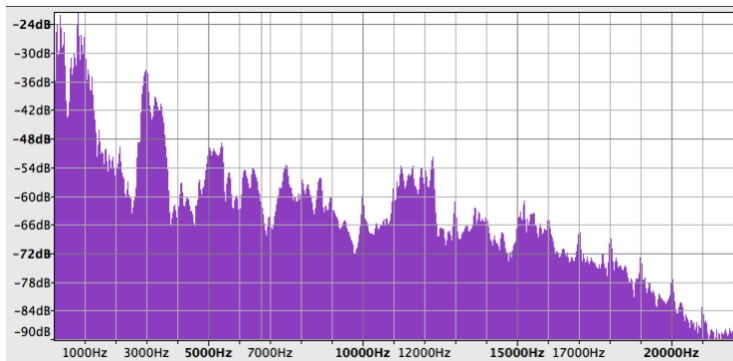


- J.R. Ewald - fyziologie sluchu.

Historie analýzy řeči

- Dvacáté století:
 - 1924 - spektrální analýza řeči na bázi formantové analýzy samohlásek
 - Vokodéry - komprese řečového záznamu
 - 1946 - 47 zařízení pro grafické zobrazení řeči
 - 2. polovina 20. století - intenzivní rozvoj teorie a počítačových aplikací

Spektrum zvuku



Textová data k obrázku. (images/spektrum-a.txt)

Současnost

- Syntéza řeči:
 - komerční TTS:
 - AT&T Natural Voices
 - IBM Research TTS
 - Loquendo TTS
 - nekomerční TTS:
 - MBrola
 - Festival
 - Demosthenes

Současnost

- Rozpoznávání řeči:
 - izolovaných slov
 - souvislé promluvy
 - komerční: Dragon, ViaVoice Desktop Products
 - nekomerční: Sphinx4, ...
- Dialogové systémy
 - Infocity Liberec (TU Liberec, Prof. J. Nouza), v letech 1998 — 2001 na dostupne na tel. 485353100
 - MIT Cambridge, Spoken Language System Group
 - Aktuální seznam veřejně dostupných projektů viz <http://web.sls.csail.mit.edu>
 - ...

- Syntéza a rozpoznávání řeči
 - Demosthenes
 - NLP - čeština pro syntetizér MBrola - využit řečový korpus CLAP
- Asistivní technologie:
 - Audi-C - dialogové programování v C++
 - Audis - řečový hypertextový prohlížeč
 - ...
- Dialogové systémy
 - WebGen – dialogové generování webových prezentací
 - GATE – dialogové kreslení obrázků, dialogové prohlížení obrázků, zvukové zobrazení obrázků
 - Dialogové rozhraní pro IDS JMK
 - ...
- Spolupráce s laboratořemi NLP, VR, ...