

Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,
Brno

jaro 2014

Základy fyziologické akustiky

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém
VoIP
SIP

- Fyziologická akustika se zabývá:
 - mechanismem vytváření řeči
 - mechanismem vnímání řeči.
- Využívá Helmholtzovu rezonanční teorii.

Helmholtzův rezonátor

Dialogové
systémy

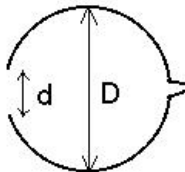
Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém
VoIP
SIP



■ Princip činnosti:

- Přivedením vzduchu do rezonátoru v něm vznikne přetlak.
- Ten vytlačuje přebytečný vzduch ven a následně vzniká podtlak, který způsobí nasávání vzduchu z okolí.
- Takto vzniká periodický děj:

$$f = \frac{75,3}{D} \sqrt{\frac{d}{D}} [\text{Hz}]$$

Mechanismus vytváření řeči

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém
VoIP
SIP

- Řeč vzniká v *hlasovém ústrojí* (umístěno v *hrtanu*).
- Hlasivky vytváří úzkou hlasovou štěrbinu a jsou rozechvívány procházejícím vzduchem.
- Frekvence jejich kmitání určuje *základní hlasivkový tón* – F_0 .
- Zvuk, který vzniká v hrtanu pomocí hlasivek (samohlásky, znělé souhlásky) je modifikován v *rezonančních dutinách*:
 - hrtanové
 - ústní
 - nosohltanové.
- Rezonanční dutiny fungují na stejném principu jako Helmholtzův rezonátor).

Mechanismus vytváření řeči

Neznělé souhlásky

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- Na vzniku se nepodílejí hlasivky.
- Vznikají tak, že vydechovanému proudu vzduchu je do cesty vložena překážka.
 - jazyk
 - zuby
 - rty.

Hlasivky a schéma lidského hlasového ústrojí

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

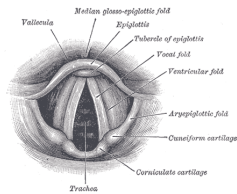
Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

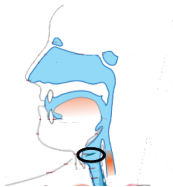
Komunikace
uživatel –
dialogový
systém

VoIP
SIP

■ Hlasivky



■ Jejich umístění



Mechanismus vnímání řeči

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- Zvuk vnímáme sluchovým orgánem.
- Sluchový orgán:
 - vnější ucho – zachycuje, soustřeďuje a přivádí zvukové vlny ke střednímu uchu
 - střední ucho
 - mechanickou cestou přenáší zvukovou energii mezi vnějším a vnitřním uchem
 - obsahuje mechanismy k vyrovnání rozdílů tlaku mezi vnějším prostředím a sluchovým orgánem
 - vnitřní ucho – převádí zvukovou energii na vzruchy, které jsou vedeny dále do mozku.

Schéma sluchového orgánu

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP



Obrázek: Schéma sluchového orgánu

- **Obsahuje:**
 - Ušní boltec – soustředí vlny do zvukovodu.
 - Zvukodod – vede zachycenou zvukovou energii (vlny) k bubínku.
 - **Bubínek:**
 - Tenká blána na konci zvukovodu – síla cca 0.1 mm.
 - Zesílí a přeneše zvukovou energii na kůstky středního ucha.

- **Obsahuje:**
 - **Kůstky středního ucha:**
 - kladívko – přiléhá k bubínku
 - kovadlinka
 - třmínek – přiléhá k oválnému okénku, kterým se zvuková energie předává do vnitřního ucha.
 - **Oválné okénko – tvoří přístup k vnitřnímu uchu.**
 - **Eustachova trubice:**
 - Vede ze středního ucha do nosohltanu.
 - Slouží k vyrovnání rozdílu tlaku mezi vnějším prostředím a středním uchem, aby nedošlo poškození sluchu.

- Hlemýžď' (Cochlea):
 - Je naplněn vodnatým roztokem.
 - Ústrojí ve tvaru ulity hlemýždě, které obsahuje Cortiho ústrojí.
 - Cortiho ústrojí obsahuje zhruba 20000 vláček s délkami $40 \mu\text{m}$ — 0,5 mm.
 - Vlákénka jsou napojena na nervová zakončení, která vedou vzruchy do příslušného centra v mozku.
- Rovnovážný orgán.

Základy fonetiky

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- Zkoumá zvukovou stránku jazyka z různých aspektů.
- Základní pojmy, které souvisejí se zpracováním řeči a dialogovými systémy:
 - foném
 - samohlásky – formanty
 - souhlásky – znělost/neznělost souhlásek
 - koartikulace
 - spodoba znělosti

Fonémy a fonetická transkripce

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- Foném – elementární zvukový segment, který je vymezen na základě své schopnosti diferencovat vyšší, znakové jednotky jazykového systému (morfémy).
- Fonetická transkripce (přepis) – převod psaného textu do odpovídající fonetické podoby:
 - na shledanou → na zhledanou | na schledanou
- Fonetická abeceda – slouží k zápisu fonetického přepisu
 - Mezinárodní fonetická abeceda (IPA) – součástí standardu UNICODE
 - Fonetická abeceda pro metody zpracování řeči (Speech Assessment Methods Phonetic Alphabet – SAMPA) – sedmibitový přepis fonetické abecedy, využívá se při automatizovaném zpracování (např. řečový syntetizér MBrola, ...)..

- Samohláska – samostatně tvoří slabiku
- Rozdělení samohlásek:
 - krátké: a, e, i, o, u
 - dlouhé: á, é, í, ó, ú
 - dvojhlásky: eu, au, ou
- Obsahují:
 - základní hlasivkový tón – frekvence kmitání hlasivek (100 — 400 Hz)
 - formanty – frekvence vzniklé a zesílené rezonancí v hlasových dutinách.

- Frekvence vzniklé a zesílené rezonancí v hlasových dutinách
 - F_1 – vzniká rezonancí v dutině ústní.
 - F_2 – vzniká rezonancí v dutině hrdelní.
- Existují i vyšší formanty (F_3, \dots) – výskyt je často individuální.
- Výskyt a intenzita formantů se může lišit v závislosti na:
 - pohlaví – muž/žena
 - věku – dětství/dospívání/dospělost/seniorský věk
 - zdravotním stavu – např. nachlazení, ochraptělost, nemoci hlasivek a hrtanu, ...
 - ...

Formanty F_1 a F_2 pro české samohlásky

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

Samohláska	Formant F_1	Formant F_2
a	700 — 1100 Hz	1100 — 1500 Hz
e	500 — 700 Hz	1500 — 2000 Hz
i	300 — 500 Hz	2000 — 3000 Hz
o	500 — 700 Hz	900 — 1200 Hz
u	300 — 500 Hz	600 — 1000 Hz

Tabulka: Formanty F_1 a F_2 u samohlásek

Četnost výskytu samohlásek

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

Samohláska(y)	Relativní četnost
[e]	10 %
[a], [o], [i]	6 — 7 %
[í]	4 %
[á], [u], [é], [ou], [ú]	< 4 %
[ó], [au], [eu]	pouze nepatrná frekvence

- Na rozdíl od samohlásek jsou souhlásky dynamické děje.
- Silně závisí na kontextu, ve kterém se nacházejí.
- Tónový charakter mají pouze části některých souhlásek:
- Dělí se podle:
 - znělé – vznikají v hltanu, obsahují základní hlasivkový tón.
 - neznělé – vznikají v řečových dutinách (nosohltanové, ústní, ...), mohou mít charakter šumu (např. sykavky):
 - problematická detekce začátku promluvy při zašuměném zdroji.
 - Znělé a neznělé samohlásky se mohou vyskytovat v párech (párové souhlásky) např.:
 - r/l
 - b/p
 - d/t
 - ...

■ Kroky digitalizace zvuku:

- 1 vzorkování – snímání aktuální hodnoty signálu s danou frekvencí (vzorkovací frekvence)
- 2 kvantizace – převod reálných hodnot na celočíselné
- 3 kódování průběhu vlny – způsob ukládání informací o průběhu zvuku.

- Snímání aktuální hodnoty signálu – snímání se opakuje s určitou frekvencí (vzorkovací frekvence).
- Vzorkovací frekvence – měla by být minimálně dvojnásobkem nejvyšší frekvence, která je v signálu přítomna, aby bylo možné původní signál bez ztráty informace zrekonstruovat (Shannonův vzorkovací teorém).
- Získané hodnoty musí být následně kvantizovány a vhodným způsobem uloženy.
- Nejpoužívanější vzorkovací frekvence:
 - 8 kHz – telefonní kvalita
 - 16 kHz
 - 22050 Hz – rozhlasová kvalita
 - 44100 Hz – CD kvalita
 - 48 kHz – DVD kvalita

- Metoda převodu spojitych hodnot na diskretní.
- Princip:
 - Pokud hodnota signálu překročí n . násobek kvantizačního kroku je jí přiřazena hodnota n .
 - kvantizační krok = rozsah hodnot měřené veličiny/počet diskretních hodnot
 - kvantizační chyba – zaokrouhlovací chyba způsobená velikostí kvantizačního kroku, přímo úměrná velikosti kvantizačního kroku.
- Běžně používané kvantizace:
 - zpracování zvuku:
 - 2^8
 - 2^{16}
 - 2^{24}
 - zpracování obrazu, ... navíc
 - 2^{32}

Způsoby kódování průběhu vlny

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém
VoIP
SIP

- Přímé ukládání hodnot získaných kvantizací – kódování PCM (Pulse-Code Modulation).
 - relativně pomalé změny průběhu zvukového signálu – malé rozdíly mezi sousedními vzorky.
 - Velká redundance dat.
 - Problém v případě příliš velkého rozptylu amplitud v signálu (příliš velký kvantizační krok – příliš velká kvantizační chyba, příliš malý kvantizační krok – přetečení v okamžiku zvětšení amplitudy signálu).
- Diferenční PCM – ukládá se rozdíl mezi sousedními vzorky
- Adaptivní PCM — PCM s proměnou velikostí kvantizačního kroku – kvantizační krok se uzpůsobí velikosti amplitudy signálu.

Diferenční pulsní kódová modulace

Dialogové systémy

Luděk Bártek

Fyziologická akustika

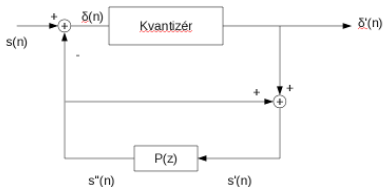
Základy fonetiky

Úvod do počítačového zpracování zvuku

Komunikace uživatel – dialogový systém

VoIP
SIP

- Vychází z předpokladů:
 - Rozdíl dvou po sobě jdoucích vzorků je podstatně menší hodnota než hodnota vzorku.
 - Následující vzorek lze poměrně přesně odhadnout jako lineární kombinaci předchozích vzorků.
- Blokové schéma kódování signálu pomocí DPCM



- $s''(n)$ – odhad hodnoty řečového vzorku
- $s'(n)$ – rekonstruovaný signál, získaný jako součet kvantizovaného signálu $\delta'(n)$ a $s''(n)$
- $\delta(n) = s(n) - s''(n)$

Adaptivní pulsní kódová modulace

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- Možné velké změny amplitudy signálu:
 - Nepřesné zachycení slabého signálu – amplituda je příliš malá, srovnatelná s kvantizačním krokem (příliš velký kvantizační krok).
 - Zkreslení (ořezání) silného signálu – dojde k přetečení rozsahu hodnot určených pro zakódování signálu (příliš malý kvantizační krok).
- Řešení: přizpůsobení kvantizačního kroku amplitudě signálu.

- Hlasová:
 - komunikace většinou prostřednictvím telefonní sítě (PSTN, VoIP).
 - Digitalizace hlasu probíhá:
 - Na straně uživatele – komunikace pomocí VoIP.
 - Na straně telefonní ústředny – DS používá VoIP, uživatel používá PSTN.
 - Na straně DS – uživatel i DS používají PSTN.
 - Rozpoznávání řeči probíhá většinou na straně DS.
 - Kdy je vhodné rozpoznávání řeči na straně klienta?
 - Jaké mohou být výhody rozpoznávání řeči na straně klienta?

Způsoby komunikace uživatele s dialogovým systémem

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- textová:
 - uživatel komunikuje s DS buď pomocí specializovaného klienta nebo pomocí běžných protokolů z rodiny TCP/IP.
 - Odpadá nutnost rozpoznávání řeči.
 - Využívá se hlavně pro vývoj a ladění.
- hlasová+textová:
 - komunikace s DS
 - VoIP – text pomocí DTMF (alá SMS).
 - specializovaný klient.

- VoIP – rodina protokolů pro řízení průběhu hlasové komunikace a přenos hlasu přes internet (sít' na bázi IP).
- Využívá se pro IP telefonii.
- Využívá protokoly:
 - UDP (transportní vrstva):
 - Stará se o přenos paketů přes počítačovou sít' mezi dvěma body.
 - Není zajištěno doručení paketů ani jejich pořadí.
 - Výhoda – nízká režie přenosu dat.
 - Nevýhody – možná ztráta dat a možnost velkých rozdílů v rychlosti doručení jednotlivých paketů
 - RTP (relační vrstva):
 - Využívá se pro přenos multimediálních dat.
 - Zajišťuje doručení paketů.
 - Umožňuje řízení parametrů přenosu – zajistí malé rozdíly v rychlosti doručení paketů.

- VoIP – řada implementací.
- Liší se:
 - použitými standardy
 - H.323 (na ústupu, standard ITU, komplexní, relativně komplikovaný)
 - SIP (jednodušší náhrada H.323, v současnosti velmi rozšířený)
 - firemní – Skinny (Cisco), HFA (Siemens), ...
 - službami – telefonie, TV (DVB), fax, zasílání zpráv, ...
 - signalizací – závisí na zvoleném standardu a použitých protokolech.
 - ...

Session Initiation Protocol (SIP)

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- Protokol pro řízení signalizace pro VoIP na aplikační vrstvě OSI modelu.
- Textový protokol pracující v režimu klient–server, poskytující mechanismy pro:
 - přesměrování hovoru
 - číselnou identifikaci volajícího a volaného
 - osobní mobilitu
 - autentizaci volajícího a volaného
 - podporu konferenčních hovorů prostřednictvím vícesměrového zasílání dat (multicast).
 - ...

- Identifikace účastníka – URI ve tvaru *sip:číslo@adresa_počítače*
 - číslo – číslo přidělené uživateli na daném stroji (VoIP ústředně)
 - adresa počítače – adresa (FQDN/IP) ústředny, na které je uživatel registrován.
- SIP relace může být:
 - přímá – navázána přímo komunikujícími stranami
 - s použitím SIP proxy serveru/ů – tyto slouží jako registrátoři účastníků.

Činnosti protokolu SIP

Dialogové
systémy

Luděk Bártek

Fyziologická
akustika

Základy
fonetiky

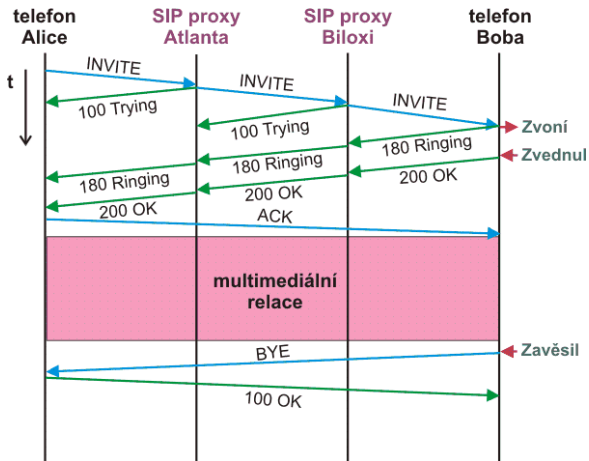
Úvod do
počítačového
zpracování
zvuku

Komunikace
uživatel –
dialogový
systém

VoIP
SIP

- Lokalizace účastníka – pomocí identifikace
- Zjištění stavu účastníka – připravenost k přijetí hovoru vs. obsazeno/přesměrováno
- Zjištění možností účastníka – dostupné kodeky, dostupná šířka pásma, podpora audia/video, ...
- Vlastní navázání spojení – využívá se protokol SDP
 - popisuje navazované spojení,
 - odkazuje na RTP/UDP datový tok, který je využit pro komunikaci účastníků.

Řízení průběhu spojení pomocí protokolu SIP



Obrázek: Obrázek převzat z Wikipedie

Dialogové systémy

Luděk Bártek

Fyziologická akustika

Základy fonetiky

Úvod do počítačového zpracování zvuku

Komunikace uživatel – dialogový systém

VoIP
SIP