

PB095 – Úvod do počítačového zpracování řeči

Luděk Bártek

Fakulta Informatiky
Masarykova univerzita

podzim 2021

Obsah

- 1 Multimodální dialogové systémy
- 2 Dialogové systémy a emoce

Multimodální dialogová rozhraní

- Mimo mluvenou řeč umožňuje alternativní způsoby komunikace člověk – počítač:
 - textová komunikace
 - grafická komunikace
 - ...
- Výhoda - lepší přístupnost.
 - uživatelé s poruchami sluchu,
 - uživatelé s poruchami řeči,
 - ...

Multimodální komunikace počítač – člověk

- Textová:
 - Mimo hlasový výstup je navíc zobrazen i odpovídající textový výstup.
 - Lze využít prostředky pro IM, SMS, ...
- Grafická:
 - Talking Heads – mimo hlasový výstup je navíc zobrazena tvář (hlava, celý člověk, ...), jejíž pohyby, zejména úst, odpovídají mluvené řeči.
 - Komunikace znakovou řečí – mluvené slovo je překládáno na znakovou řeč (viz Guimeraes, K., Antunes, D. R., Guilhermino, de F. Guilhermino, D., Lopes da Silva, R. A., Garcia, L. S – Structure of the Brazilian Sign Language (Libras) for Computational Tools: Citizenship and Social, in Organizational, Business, and Technological Aspects of the Knowledge Society, CCIS vol. 112, Springer, Heidelberg, 2010, pp. 365 – 370.)
 - Znaková řeč prezentována pomocí rukou nebo avatara.

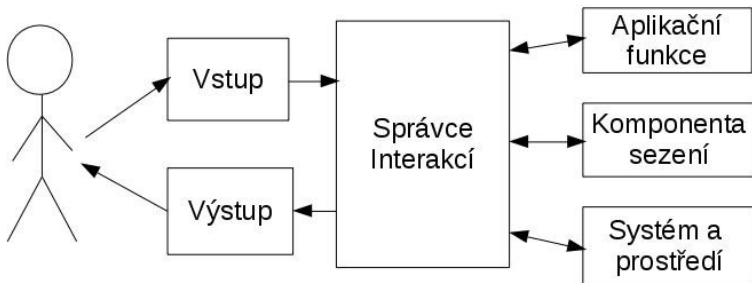
Multimodální komunikace člověk – počítač

- Široké spektrum možností zadávání vstupu uživatelem jinak než hlasem:
 - klávesnice (počítač, DTMF, SMS, ...)
 - rukou psaný vstup – dotyková obrazovka + pero, ...
 - ústy ovládaná zařízení
 - ovládání pomocí pohybů očí a víček
 - rozpoznávání řeči pomocí sond detekujících činnost svalů a mozku (viz Schultz, T. – Silent and Weak Speech Based on Elektromyography, in Proceedings of 12th International Conference ICCHP 2010 Part 1, Wien, Springer, Heidelberg, pp. 595 – 604, 2010.)
 - rozpoznávání znakové řeči
 - ...
- Často jako doplněk řečového vstupu.

Existující nástroje a standardy

- Proprietární řešení:
 - Součást CSLU Toolkitu. (ukázka multimodálního dialogového rozhraní (demos/support.rad)pomocí CSLU Toolkitu)
 - Projekt August
- Otevřená řešení:
 - Návrhy a doporučení W3C týkající se multimodálního přístupu.
 - Využívají a propojují i další standardy W3C (CCXML, XHTML), VoiceXML, SVG), SMIL,...
 - Výstup W3C Multimodal Interaction WG

Architektura multimodálních aplikací



Komponenty multimodálních aplikací

- vstup/výstup – různé vstupní komponenty (viz 4).
- Správce interakcí – logická komponenta, která řídí a synchronizuje komunikaci prostřednictvím různých vstupních a výstupních kanálů.
- Aplikační funkce – funkce poskytované aplikační logikou systému.
- Komponenta sezení – poskytuje správci interakcí rozhraní pro správu stavu spojení a aplikace.
- Komponenta systému a prostředí – poskytuje správci interakcí způsob jak zjistit a reagovat na změny v prostředí, v němž aplikace běží (dostupná I/O zařízení a jejich možnosti, uživatelské preference, ...)

Ukázka multimodální syntetizované řeči

data/avatar-demo.mpg

Figure: Ukázka multimodálního grafického výstupu

Standardy W3C MMI WG

- Emotion Markup Language (EmotionML) 1.0 – slouží ke značkování (automatickému/ručnímu) emocí v dalších dokumentech (viz demos/emotional.ssm1)
- Multimodal Architecture and Interfaces – popisuje architekturu rámce pro MMI a rozhraní mezi jeho složkami (např. zpracování CCXML – implementační platforma, SCXML – dokument popisující řízení interakce, HTML – modalita, VoiceXML – modalita, SVG – modalita)
- Ink Markup Language (InkML) – popisuje vstupy od elektronických per a stylusů (zadávání psaného textu, obrázků, ...)
- EMMA: Extensible MultiModal Annotation markup language – reprezentace informací získaných od uživatele libovolným kanálem.

Standardy W3C MMI WG

Dokončení

- EmotionML – jazyk pro reprezentaci emocí a odpovídajících stavů pro aplikace. Vyžití:
 - manuální anotace dat – ruční doplnění průběhu emocí u komunikace
 - automatické rozpoznávání emočního stavu z chování uživatele (očekávané přechody lze modelovat pomocí SCXML)
 - generování emocí na straně systému
 - popis slovníků pro použití s EmotionML – definice emocí (viz příklad na stránkách W3C MMI WG)

Standardy W3C MMI WG

Implementace

- MMI Architecture – multimodální uživatelské rozhraní pro budoucí telekomunikační služby od Deutsche Telekom, multimodální uživatelské rozhraní pro mobilní zařízení od France Telecom, JVoiceXML (využití různých komponent pro modality (text, hlas)), . . . , více viz zpráva o implementaci
- EMMA – prototyp multimodálního DS od společnosti Avaya, NLWorkbench (EMMA producent) a prohlížeč EMMA dokumentů od Conversational Technologies, řada multimodálních aplikací od Deutsche Telekom, SmartWeb (multimodální přístup k sémantickému webu) a OMDIP (Interaction of Dialog Management Components) od DFKI, . . . , více viz zpráva o implementaci
- InkML – interakce s OS od firmy Microsoft, InkChat od University of Western Ontario, Cue-me od Openstream Inc, . . . , více viz zpráva o implementaci

Standardy W3C MMI WG

Implementace – dokončení

- Implementace EmotionML:
 - knihovna LORIA/SYNALP z University of Lorraine
 - knihovna pro C# z Chemnitz University of Technology,
 - podpora pro EmotionML v Mary TTS,
 - ... (více viz zpráva o implementaci)

Další standardy použitelné pro tvorbu multimodálních rozhraní

- Standardy použitelné pro popis jednotlivých kanálů:
 - GUI – HTML
 - dialogové rozhraní – VoiceXML + další standardy W3C Voice Browser WG
 - Grafika/video – SVG, SVG + SMIL, SVG + JavaScript, ...

Co jsou to emoce?

- "This is a very tough question, that has produced significant amounts of headaches to scientists in the past ...", "... many researchers have to opted to study systematically phenomena that most consider emotional." (Laval University of Quebec)
- "Only mathematics is certain, so all must be based on mathematics." (R. Descartes)
- Dělení emocí:
 - Primární (základní) – vyskytují se u všech lidí a u části vyšších živočichů.
 - Sekundární (vyšší) – mohou být intelektuální, morální a estetické. Mohou se lišit mezi jednotlivými kulturami.
- Velkých šest:
 - hněv
 - zklamání
 -

Základní emoce

- Velkých šest (R. Descartes):
 - hněv
 - zklamání
 - smutek
 - strach
 - překvapení
- Další autoři:
 - Arnold – hněv, averze, odvaha, sklíčenost, touha, zoufalství, strach, nenávisť, láska, smutek.
 - Ekman, Friesen, Ellsworth – hněv, odpor strach, radost, smutek, překvapení.
 - Frijda — touha, štěstí, zájem, překvapení, údiv, zármutek,
 - ...

Detekce emocí

- Lze provádět pomocí detekce změn různých biometrických vlastností.
 - Změny galvanických vlastností kůže.



- Změny tlaku krve a pulsu.



Detekce emocí

- Použitelné biometrické charakteristiky:
 - změny dýchání



- změny elektrické aktivity mozku



- změny charakteristik řeči
- výraz tváře (Yale Face Database).
- ...

Ukázky z Yale Face Database

- Radost



- Smutek



Ukázky z Yale Face Database

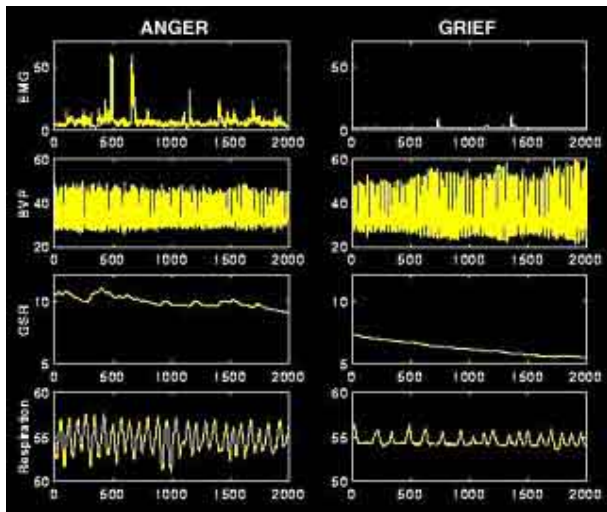
- Ospalost



- Překvapení



Ukázky průběhů charakteristik pro smutek a hněv



Využití emocí v dialogovém rozhraní

- Dialogová rozhraní informačních systémů
 - uzpůsobení dialogové strategie emočnímu stavu uživatele (klid, stres, hněv, ...)
 - přepojení uživatele na lidského operátora.
 - emoce lze použít ve vnitřních stavech Mealyho automatu pro ovlivnění chování přechodové funkce.
- Výukové DS:
 - uzpůsobení dialogové strategie koncentraci uživatele.
- ...