

## Multimodálne aspekty komunikácie a uchovávania dát

Kurdel Peter · Informačné technológie

26.09.2011



Komunikácia medzi ľuďmi je založená na niekoľkých komunikačných modalitách, ako napríklad reč, gestá a podobne. Multimodalita znamená použitie viacerých komunikačných kanálov na prenos informácie. Unimodálne systémy prenášajú informáciu iba cez jeden komunikačný kanál, napríklad reč. Multimediálne systémy taktiež používajú na prenos informácie niekoľko komunikačných kanálov, napríklad zvuk a obraz. Multimodálne systémy spracovávajú informácie na vyššej abstraktnej úrovni, čím umožňujú získať, uchovávať a spracovávať význam dát, čo pre multimediálne systémy nie je potrebné.

### Úvod

Senzory pracujúce na základe rôznych princípov (modalít) generujú multimodálne signály. Patria sem optické, mechanické, chemické, termálne, elektrické, chromatografické, magnetické, biologické, ultrazvukové a iné snímače [6]. Hlavným cieľom v oblasti interakcie medzi človekom a počítačom (HCI - Human Computer Interaction) je dosiahnuť reakciu počítača v reálnom čase na multimodálny vstup od používateľa, ako aj prístup k uloženým multimodálnym dátam.

### Metadáta, anotácie a spravovanie multimodálnych dát

Pri analýze, spracovaní a zdieľaní vedeckých dát hrá mimoriadne dôležitú úlohu popis dát, teda metadáta. Termíny metadáta a anotácie sú používané s istými obmenami v závislosti na študovanej oblasti. V spracovaní obrazov a videí sa ako metadáta označujú charakteristiky získané zo signálov. V oblasti štúdia reči a jazykov sa obvykle hovorí o anotáciách, hoci prepis reči (manuálny alebo automatický) nie je vždy považovaný za anotáciu rečového signálu.

Technické metadáta popisujú fyzikálne charakteristiky digitálnych objektov. Riadiace (manažovacie) metadáta dokumentujú životný cyklus elektronických zdrojov, ako napríklad údaje o objednaní, získaní, zachovávaní, licenčné a vlastnícke práva, pôvod a podobne. Technické a manažovacie metadáta musia byť doplnené o vyhľadávacie metadáta, ktoré používateľom umožňujú identifikovať, lokalizovať a získať vhodné zdroje. Štandardný spôsob vyjadrenia metadát, predovšetkým o zdrojoch na webe, je RDF (Resource Description Framework). Je založený na trojiciach A(O,V), kde každá trojica vyjadruje skutočnosť, že objekt O má atribút A s hodnotou V.

Anotácia je priradenie významu k jednotlivým segmentom dát na popísanie ich obsahu. Môžeme ich rozdeliť do troch kategórií:

- štruktúrna anotácia: popisuje fyzické a logické vlastnosti obsahu (napr. formát videa)
- lingvistická anotácia: text určený na čítanie človekom (niekedy sú používané aj pomenovania textová anotácia alebo lexikálna anotácia)
- sémantická anotácia: priradenie sémantických dát (metadát) k obsahu.

Počítačom zrozumiteľná informácia musí byť sémanticky anotovaná, aby počítač vedel interpretovať význam dát. Kolekcia anotácií sa nazýva ontológia. Ontológia popisuje pojmy (koncepty) v určitej oblasti (doméne) a vzťahy medzi týmito konceptmi. Na manažovanie a anotovanie multimodálnych dát boli vyvinuté rôzne systémy, napríklad Annotation Graphs [3], Exmeralda [19], NITE XML Toolkit [4], Multitool [1], Anvil [8], Elan [21], SIDGrid [9] a iné.

### **Metódy analýzy a archivovania multimodálnych dát**

Multimodálna dátová integrácia bola pôvodne vyvinutá pre oblasť vojenských aplikácií a robotiky. Keďže táto interdisciplinárna technológia vznikala v rôznych oblastiach aplikačného výskumu, terminológia dodnes nie je jednotná. V rôznych publikáciách sú používané pojmy: sensorová fúzia, multimodálna dátová fúzia, multi-sensorová dátová fúzia, prípadne dátová alebo informačná fúzia. Dátovú fúziu možno klasifikovať na základe charakteru vstupných a výstupných informácií na konkurenčnú, komplementárnu a kooperatívnu dátovú fúziu. Existuje mnoho prístupov k multimodálnej dátovej fúzii, napríklad fuzzy logika.

Návrh multimodálneho systému závisí na charaktere dát a požiadavkách aplikácie. Zvolené modality by vo všeobecnosti mali poskytovať vyhovujúce klasifikačné vzory. Niektoré aplikácie môžu mať špecifické požiadavky, napr. v oblasti biometrických dát požiadavka akceptovateľnosti: multimodálny biometrický systém môže využívať biometrické rysy ako odtlačky prstov alebo očnú dúhovku na identifikáciu, ale pre pozorovacie aplikácie sú vhodnejšie tvár a chôdza. Iný príklad z oblasti spracovania reči: prozodické rysy sú považované za hlavný indikátor citov v reči a fonetické rysy sú používané na rozpoznávanie reči.

Dolovanie multimodálnych dát v databázach je rýchlo sa rozvíjajúca oblasť. Boli vytvorené rôzne databázy: MyIDEa je multimodálna biometrická databáza [13], multimodálna emočná databáza AvID [10], biometrická databáza BIOMET [2]. Multimodálny prístup do databázy pomocou zariadenia upevneného na hlave používateľa je popísaný v [12].

### **Príklad použitia: Multimodalita v databáze giest**

Reč tela je dôležitou súčasťou komunikácie, ktorá môže pri rozhovore sprostredkovať aj viac ako 50% informácií [20]. Tieto neverbálne informácie môžu byť niekedy dokonca v rozpore s hovorenými slovami. Signály môžu byť vysielané jednotlivými časťami tela (hlava – tvár, brada, ústa, pery, zuby, jazyk, nos, oči, obočie, čelo; rameno – lakeť, ruka, prsty; trup – krk, plecia, hrud', chrbát, brucho, boky; nohy – koleno, chodidlo) alebo niekoľkými partiami tela súčasne. Napríklad komunikácia dotykom (haptic communication) sa používa v rôznych súvislostiach, ale môže byť nebezpečná,

ak ju osoba s ktorou komunikujeme chápe ako útok.

Komunikácia pohybom tela (kinesic communication) je najznámejšia forma neverbálnej komunikácie. Postoj (držanie tela) pri rozhovore poskytuje veľa rôznych informácií. Napodobovanie postoja druhej osoby vyjadruje súhlas, dôveru a náklonnosť.

Komunikácia na základe relatívnej polohy tela (proxemic communication) rozlišuje 4 základné vzdialenosti (zóny): intímna, neformálna osobná, spoločenská, verejná. Tieto vzdialenosti sa môžu výrazne líšiť v závislosti od jazykového a kultúrneho prostredia a od povahového charakteru osoby. Záleží aj na vzájomnom uhle natočenia postáv. Priamy postoj tvárou v tvár môže byť niekedy konfrontačný. Naopak voľba postoja, keď sa diskutujúci pozerajú jedným smerom naznačuje, že sa postavili „tvárou k rovnakému problému“.

Gestá obvykle vyjadrujú neverbálnu komunikáciu pohybom tela a rúk. Pri rozhovore sa zvyčajne pozeráme do tváre partnera, keďže 90 rôznych svalov tváre vysiela mnoho informácií. Spôsob naklonenia hlavy tiež môže zmeniť význam hovorených slov. Oči sú obzvlášť dôležité a pri rozhovore sa snažíme nadviazať kontakt očami. Počas rozhovoru tento kontakt opakovane prerušujeme a obnovujeme. Obočie a čelo poskytujú ďalšie dôležité signály od prekvapenia po strach a hnev.

Jednoduché základné gestá sú reprezentované iba jednou časťou tela [17], napríklad oči, pery, pravá ruka a pod. Tieto gestá môžu byť spojené do jednoduchých kompozitných giest, ktoré obsahujú celú hlavu alebo ruky. Tieto gestá môžu byť ďalej kombinované do multi-modálnych giest. Zložené gestá možno vytvárať z jednoduchších použitím dvoch operátorov: paralelná kombinácia a sekvenčné zretáženie.

Hranicu medzi jazykom a gestom alebo medzi verbálnou a neverbálnou komunikáciou v niektorých prípadoch nie je jednoduché identifikovať. Psychológovia Paul Ekman a Wallace Friesen v roku 1969 [7] navrhli rozdeliť gestá do 5 kategórií:

- emblémy: vedome používané gestá, ktoré sú priamou náhradou slov (existuje preklad do verbálnej formy), ako napríklad zamávanie rukou pri rozlúčke; majú blízko k znakovkej reči; v rôznych jazykových a kultúrnych prostrediach môže byť ich význam chápaný odlišne
- ikonické gestá: ilustrujú, čo sa hovorí verbálne, ako napríklad otáčanie imaginárnym volantom pri hovore o riadení auta; úzko súvisia s rečou a ich načasovanie je synchronizované s rečou; väčšinou sú jazykovo a kultúrne univerzálne
- prejavy afektu: tlmočenie citov, ako napríklad úsmev; výrazy tváre – hnev, znechutenie, strach, radosť, smútok a prekvapenie sú univerzálne
- regulátory: slúžia na riadenie prúdu konverzácie, vzájomné pôsobenie diskutujúcich, napríklad osobe, ktorá dohovorela klesnú ruky, naopak osoba, ktorá chce prehovoriť môže zdvihnúť ruku
- adaptory: gestá, ktoré slúžia na uvoľnenie napätia hovoriaceho, ako napríklad podupávanie nohou.

Iná kategorizácia delí gestá na rečovo nezávislé gestá a gestá súvisiace s rečou. Prvá skupina giest má aj bez hovorenej informácie jednoznačnú interpretáciu akceptovanú v danom jazykovom a kultúrnom prostredí. Význam takéhoto gesta sa dá preložiť do

verbálnej informácie, napríklad zamávanie na pozdrav alebo znak mieru a pokoja. Druhú skupinu tvoria gestá súvisiace s rečou, ktoré sú používané paralelne s verbálnou komunikáciou. Táto forma neverbálnej komunikácie sa používa na zdôraznenie správy, ktorú komunikujeme a poskytuje dodatočnú informáciu k verbálnej správe, napríklad počas diskusie ukážeme na určitý objekt.

Gestá sú na rozdiel od iných foriem reči tela oveľa viac asociované s rečou a jazykom a môžu mať špecifický lingvistický obsah. Reč obsahuje aj neverbálne prvky (paralanguage, vocalics), tzv. neverbálne pokyny hlasu. Rôzne akustické vlastnosti reči, medzi ktoré patrí kvalita hlasu a štýl reči, ale aj prozodické črty: rytmus, intonácia a dôraz, môžu v niektorých prípadoch zmeniť význam hovorených slov. Lingvista George L. Trager navrhol nasledovný klasifikačný systém[14]:

- kontext hlasu: kontext, v ktorom hovoriaci hovorí, napríklad situácia, nálada, vek, pohlavie, kultúra
- kvalita hlasu: hlasitosť, tón, tempo, rytmus, artikulácia, rezonancia, nazalita (nosovosť) a akcent dávajú každému individu jedinečný „hlasový odtlačok“
- vokalizácia pozostáva z troch častí:
  - charakter (city prejavene počas reči, napríklad smiech, plač, zívanie)
  - kvalifikácia (štýl doručenia správy, napríklad zarevanie „Hej prestaň!“ a naopak zašepkanie „Hej prestaň!“)
  - oddeľovače (napríklad „hm-hm“ oznamuje hovoriacemu, že počúvajuci počúva.

Počas komunikácie sa verbálne a neverbálne správy navzájom ovplyvňujú: opakujú sa, sú v rozpore, dopĺňajú sa, nahrádzajú sa, regulujú sa navzájom a zdôrazňujú sa navzájom alebo sa zmiernujú.

Cieľom vytvorenia digitálneho slovníka giest bolo poskytnúť vedecko-výskumným pracovníkom možnosť porovnávať rôzne gestá, ich význam a rôzne atribúty v rôznych jazykových a kultúrnych prostrediach spolu so súvisiacimi verbálnymi a neverbálnymi prejavmi.

Základným materiálom pri tvorbe digitálneho slovníka giest bola kniha *Picture Dictionary of Gestures* autorky Evy Růžičkovej [18]. V [16] bol navrhnutý prístup k riešeniu problémov klasifikácie a anotácie zvukov reči, ktoré majú vlastný fonetický obsah, fonologickú funkciu a prozódium, ale nemajú adekvátnu lingvistickú (resp. textovú) reprezentáciu. Na ilustráciu sú na obr. 1 informácie, ktoré slovník poskytuje o geste, ktoré znázorňuje chlad.



Obr. 1 Slovník DiGest poskytuje rôzne informácie o geste

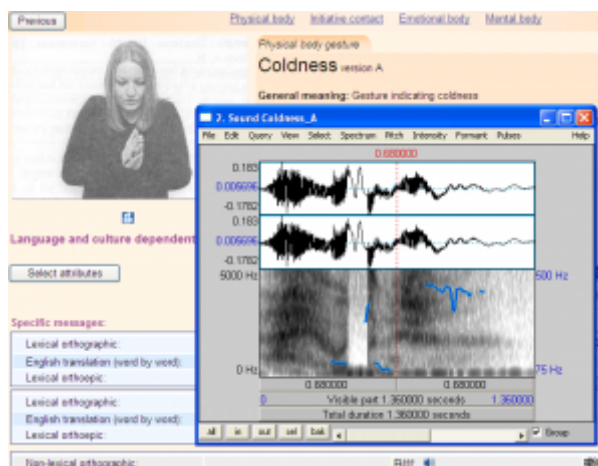
Slovník ponúka pre každé gesto všeobecné informácie v anglickom jazyku: názov, typ, význam a popis gesta. Informácie sú doplnené ilustračným obrázkom, prípadne aj videom a inými doplňujúcimi informáciami v podobe pripojených súborov rôznych formátov – zvukové, video, obrazové, textové, grafické. Ďalej slovník ponúka synonymá, teda gestá, ktorých vzhľad je rovnaký alebo veľmi podobný, ale ich význam je iný.

Jazykovo a kultúrne závislé informácie sú zobrazené v podobe tabuľky. Používateľ si môže zvoliť množinu atribútov, ktoré sa majú zobrazovať a má možnosť zvoliť jazyky, v ktorých sa budú zobrazovať jazykovo závislé informácie. Slovník obsahuje pre každé gesto nasledovné špecifické informácie:

- lexikálna správa:
  - orthografický zápis (môže byť doplnený videom a/alebo zvukovým súborom)
  - orthoepický prepis (použitím IPA International Phonetic Alphabet fonu)
- nelexikálna správa:
  - orthografický zápis (môže byť doplnený videom a/alebo zvukovým súborom)
  - orthoepický prepis (použitím IPA fonu)
  - pre správy v jazykoch, ktoré na zápis písomnou formou nepoužívajú latinku, napr. čínština, japončina, tradičná mongolčina a iné, je uvedený aj prepis latinkou
- doplňujúce informácie v súboroch pripojených k danému gestu vo zvolenej jazykovej mutácii, ktoré môžu bližšie špecifikovať používanie gesta v určitom jazykovom a kultúrnom prostredí.

V dolnej časti stránky je uvedený kontext, v ktorom sa dané gesto používa, špecifikovaný atribútmi: formálnosť, rod, vek a sociálne postavenie.

Zvukové súbory pripojené k informáciám o gestách v rôznych jazykových mutáciách umožňujú skúmať a porovnávať lingvistické atribúty verbálnych aj neverbálnych správ súvisiacich s gestami v rôznych jazykových a kultúrnych prostrediach. Používateľ môže otvoriť zvukové súbory napríklad pomocou prostriedku Praat [15], ktorý zobrazí príslušné grafy a prípadné anotácie, čo je ilustrované na obr.2.



Obr. 2 Zvukový súbor zo slovníka otvorený pomocou Praat

Dátové gridy sú určené pre digitálne úložiská a uchovávanie rozsiahlych dát [5]. Využívajú gridovú technológiu Storage Resource Broker (SRB), ktorá bola vyvinutá v Super-počítačovom centre v San Diegu [11]. Poskytujú jednotný prehľadateľný prístup k virtualizovaným distribuovaným zdrojom, takže údaj o umiestnení dát je izolovaný od fyzického miesta uloženia dát a je nezávislý od typu úložiska, čo umožňuje bezproblémový prechod na nový hardvér. Dátové gridy sú škálovateľné, s pribúdajúcim rozsahom digitálnej knižnice je možné k úložisku dát dynamicky pridávať ďalšie hardvérové prostriedky. Sú veľmi efektívne pre spravovanie úložiska, ale poskytujú iba relatívne obmedzenú podporu pre metadáta.

Digitálne knižnice typicky predpokladajú lokálne úložisko dát. Podporujú uchovávanie rozličných komplexne štruktúrovaných digitálnych prostriedkov a umožňujú flexibilné spravovanie metadát.

## Záver

V článku sme diskutovali o rôznych aspektoch multimodality v komunikácii a pri analýze, uchovávaní a dolovaní multimodálnych dát. Práca je riešená v rámci projektu „Priemyselný výskum v oblasti efektívnej práce s rozsiahlymi dátami v používateľsky orientovaných aplikáciách“, ktorý je spolufinancovaný zo zdrojov EÚ, ITMS kód projektu: 26240220029. Na ilustráciu bola použitá digitálna multimodálna databáza giest vytvorená v oddelení analýzy a syntézy reči Ústavu informatiky Slovenskej akadémie vied.

## Literatúra

1. Allwood J., Groenqvist L., Ahlsen E., and Gunnarsson M.: 2001. Annotations and tools for an activity based spoken language corpus. In Proceedings of the Second SIGdial Workshop on Discourse and Dialogue, pages 1-10.
2. BIOMET [http://biometrics.it-sudparis.eu/english/index.php?item=10&menu=projects]
3. Bird S., Kazuaki Maeda, Xiaoyi Ma, Haejoong Lee, Beth Randall, and Salim Zayat, TableTrans, MultiTrans, InterTrans and TreeTrans: Diverse Tools Built on the Annotation Graph Toolkit, April 2002, [http://arxiv.org/PS\\_cache/cs/pdf/0204/0204006v1.pdf](http://arxiv.org/PS_cache/cs/pdf/0204/0204006v1.pdf)
4. Carletta J., Evert S., Heid U., Kilgour J., Robertson J., and Voormann H.: 2003. The NITE XML Toolkit: flexible annotation for multi-modal language data. Behavior

- Research Methods, Instruments, and Computers, special issue on Measuring Behavior, 35(3):353-363.
5. Hedges M., Hasan A.: Digital Library Storage using iRODS Data Grids, <http://www.kcl.ac.uk/iss/cerch/about/staff/hedges.html>
  6. Chang E.: Distributed, Multimodal Sensor-data Fusion and Analysis, August 7th, 2004, <http://sigmm.utdallas.edu/>
  7. Kinesic communication: [http://changingminds.org/explanations/behaviors/body\\_language/kinesic.htm](http://changingminds.org/explanations/behaviors/body_language/kinesic.htm)
  8. Kipp M.: 2001. Anvil- a generic annotation tool for multimodal dialogue. In Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech), pages 1367-1370.
  9. Levow G. A., Waxmonsky S., Bertenthal B., McNeill D.: SIDGrid: A Framework for Distributed, Integrated Multimodal Annotation, Archiving, and Analysis, Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, pages 231-234, Antwerp, September 2007, <http://www.sigdial.org/workshops/workshop8/Proceedings/SIGdial41.pdf>
  10. mAvID [<http://www.thefreelibrary.com/Multi-modal+emotional+database+AvID>]
  11. Moore R. W.: Digital Libraries, Preservation Environments, and Data Grids, San Diego Supercomputer Center, <http://irods.sdsc.edu>
  12. Multimodal Database Access on Hand-held Devices [<http://www.aclweb.org/anthology/P/P04/P04-3029.pdf>]
  13. MyIDea [<http://diuf.unifr.ch/diva/biometrics/MyIdea/en/db.html>]
  14. Nonverbal communication: [http://en.wikipedia.org/wiki/Non-verbal\\_communication](http://en.wikipedia.org/wiki/Non-verbal_communication)
  15. Praat: doing phonetics by computer <http://www.fon.hum.uva.nl/praat/>
  16. Rusko M., Juhar J.: Towards annotation of nonverbal vocal gestures in Slovak. In Verbal and nonverbal features of human-human and human-machine interaction. Editor Anna Esposito, Nikolaos G. Bourbakis, Nikolaos Avouris, Ionnis Hatzilygeroudis. - Berlin: Springer, 2008. ISBN 978-3-540-70871-1. ISSN 0302-9743, p. 255-265.
  17. Ruttkay Z.: Presenting in Style by Virtual Humans. In Verbal and Nonverbal Communication Behaviours. Editor Anna Esposito, Marcos Faundez-Zanuy, Eric Keller, Maria Marinaro. Berlin: Springer 2007, p. 25-36.
  18. Ružičková, E.: Picture dictionary of gestures (American, Slovak, Japanese, and Chinese), Comenius University Publishing House, Bratislava (2001), ISBN 80-22-1675-X.
  19. Schmidt T.: Transcribing and annotating spoken language with EXMARaLDA. In Proceedings of the LREC-Workshop on XML-based richly annotated corpora, 2004.
  20. Using Body Language: [http://changingminds.org/techniques/body/body\\_language.htm](http://changingminds.org/techniques/body/body_language.htm)
  21. Wittenburg P., Brugman H., Russel A., Klassmann A., and Sloetjes H.: 2006. Elan: a professional framework for multimodality research. In Proceedings of Language Resources and Evaluation Conference (LREC) 2006.

---

Spoluautorom článku je Jolana Sebestyénová, Ústav informatiky Slovenskej akadémie vied, Bratislava

---

