
Verarbeitung gesprochener Sprache

Proseminar

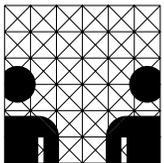
Wolfgang Menzel

`menzel@informatik.uni-hamburg.de`

Arbeitsbereich Natürlichsprachliche Systeme

Fachbereich Informatik

Universität Hamburg

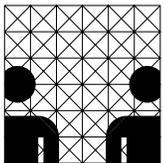


Proseminar, wozu?

“Lehrveranstaltungsarten sind ...

- ...
- Proseminare zur selbständigen Erarbeitung von Wissen und dessen Vermittlung,
- ...”

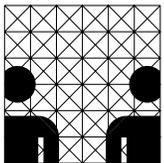
Prüfungsordnung Bachelor of Science der MIN-Fakultät



Proseminar, wozu?

- Schlüsselqualifikationen erwerben
 - Recherchieren
 - Strukturieren
 - Präsentieren
 - Moderieren
- Wechselwirkungen zwischen Informatiksystemen und ihrem Einsatzfeld
 - Verknüpfung von fachinhaltlichen Schwerpunkten mit Wirkungs- und gesellschaftlichen Fragestellungen
 - Technische Systeme im Kontext betrachten

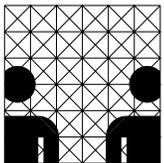
Fachspezifische Bestimmungen BSc Informatik



Proseminar, wozu?

primäre Fragestellungen:

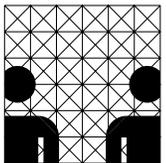
- Themenanalyse
- Themenaufbereitung, Systematisierung
- Ideenfindung
- Literaturarbeit: Recherche, Selektion, Auswertung
- Systemevaluation
- Präsentation
- Diskussion



Proseminar, wozu?

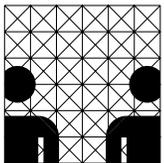
sekundäre Fragestellungen:

- Problemlösung
- Entwurf
- Realisierung



Proseminar, wie?

- Vortrag
- Diskussion
- Feedback
- schriftliche Ausarbeitung
- Erfahrungsbericht oder Kurzrezension
- Das Thema in der öffentlichen Wahrnehmung



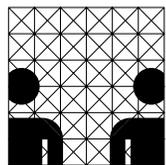
Verarbeitung gesprochener Sprache

A. Sprachverarbeitung, wozu?

- Analyse des Phänomenbereichs
- Brainstorming: Potenzielle Anwendungsfelder
- Analyse realer Anwendungsfälle
- Diskussion

B. Sprachverarbeitung, wie?

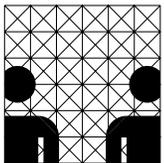
- Seminarvorträge zu Einzelthemen



A. Sprachverarbeitung, wozu?

1. Gesprochene Sprache

- Welche Informationen enthält ein Sprachsignal?
- Wie werden diese Informationen im Signal kodiert?
- Welche Bestandteile sind für technische Anwendungen interessant?

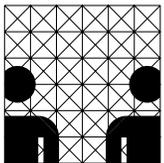


A. Sprachverarbeitung, wozu?

2. Sprachverarbeitung: Einsatzmöglichkeiten und Leistungsparameter

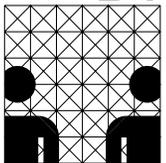
- Brainstorming: Anwendungsszenarien
- Was sind Störfaktoren? Wie lässt sich ihr Einfluss verringern?

Literatur: Tageszeitungen, populärwissenschaftliche Zeitschriften und Bücher



A. Sprachverarbeitung, wozu?

3. Anwendungsbereich I: Diktiersysteme
4. Anwendungsbereich II: Dialogsysteme
5. Anwendungsbereich III: Sprachlehrsysteme
6. Anwendungsbereich IV: Sprachassistenten
 - Evaluation
 - Erfahrungsbericht
 - Bewertung
7. Gesprochene Sprache, die alternative Mensch-Maschine-Schnittstelle?
 - Diskussion: Pro und Kontra



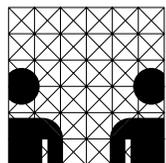
B. Sprachverarbeitung, wie?

7. symbolische Signalbeschreibung

- Beschreibungsebenen für gesprochene Sprache: Signal, Phonetik, Phonologie, Morphologie, Prosodie, ...
- Verhältnis zur geschriebenen Sprache
- Beschreibungseinheiten: akustische / artikulatorische Merkmale, Phone, Allophone, Phoneme, Silben, Morphe, Wörter, ...
- Visualisierung von Sprachsignalen

Literatur: Kohler (1977), Fellbaum (1984), Holmes (1991), jedes beliebige Handbuch der Sprachwissenschaft

Schwierigkeitsgrad: leicht



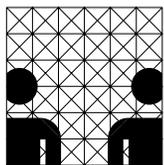
B. Sprachverarbeitung, wie?

8. parametrische Signalbeschreibung

- Maximalwertdetektion, Energie
- Autokorrelation, Nulldurchgangsdichte
- Spektrum, Fourier-Transformation, Spektrogramm, Formant
- Filter, Filterbänke
- Cepstrum
- Visualisierung von Signalparametern

Literatur: Vary et al. (1998), Schafer und Rabiner (1975), Eppinger und Herter (1993), Euler (2006)

Schwierigkeitsgrad: schwer



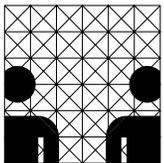
B. Sprachverarbeitung, wie?

9. Sprachsynthese

- Historische Vorläufer
- Formantsynthese
- Diphonsynthese
- PSOLA
- Korpus-basierte Synthese

Literatur: Eppinger und Herter (1993), Vary et al. (1998), Hess (1996), Portele (1996), Balestri et al. (1999), Stöber et al. (1999) Taylor (2009)

Schwierigkeitsgrad: mittel



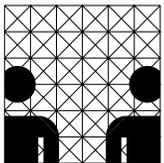
B. Sprachverarbeitung, wie?

10. Worterkennung I: Linearer Mustervergleich

- Abstandsmessung
- Dynamic Time Warping

Literatur: Vaissière (1985), Silverman und Morgan (1990), Furui (1989), Holmes (1991), Eppinger und Herter (1993), Ruske (1994), Euler (2006)

Schwierigkeitsgrad: schwer



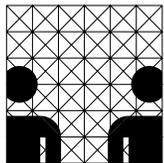
B. Sprachverarbeitung, wie?

11. Worterkennung II: Hidden-Markov-Modelle

- Markov-Modelle
- Hidden-Markov-Modelle
- Vektorquantisierung
- Training

Literatur: Rabiner und Juang (1986), Eppinger und Herter (1993), Picone (1990), Wendemuth (2004), Fink (2003) Euler (2006), Jurafsky und Martin (2008)

Schwierigkeitsgrad: sehr schwer



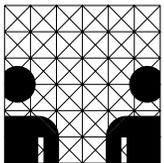
B. Sprachverarbeitung, wie?

12. Sprachmodellierung

- reguläre Grammatiken
- Bigramme, Trigramme
- stochastische Grammatiken
- Perplexität

Literatur: Jelinek (1976), Jelinek (1990), Ney (1992), Lee (1999), Jelinek (1998), Charniak (1993), Euler (2006), Jurafsky und Martin (2008)

Schwierigkeitsgrad: schwer



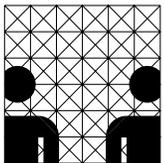
B. Sprachverarbeitung, wie?

13. Prosodie in der Spracherkennung

- Betonung
- Phrasengrenzen
- Satzmodus

Literatur: Vaissière (1988), Nöth (1989), Nöth et al. (1997), Kompe (1997)

Schwierigkeitsgrad: mittel

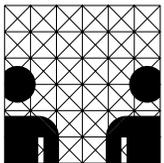


B. Sprachverarbeitung, wie?

14. Dialogmodellierung

- Anwendungen
- Dialogtypen
- Formalismen

Literatur: Bernsen et al. (1998), Harris (2005), Dahl (2004), McTear (2004), Chen (2006), Jokinen und McTear (2010) Lewis (2011) **Schwierigkeitsgrad:** mittel



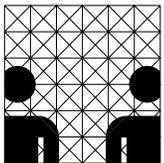
B. Sprachverarbeitung, wie?

15. Inkrementelle Dialog Systeme

- Warum inkrementelle Dialogsysteme?
- Inkrementelle Spracherkennung
- Hypothesenverwaltung, Hypothesenrevision
- Qualitätsmaße

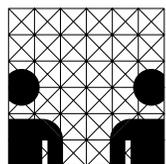
Literatur: Schlangen und Skantze (2009), Baumann et al. (2009), Comaneci et al. (2012)

Schwierigkeitsgrad: schwer



Ablaufplan

16.10.2012	Einführung	4.12.2012	Vorträge 3
23.10.2012	Diskussion: Aufgabenanalyse	11.12.2012	Vorträge 4
30.10.2010	Brainstorming: Anwendungen	18.12.2012	Vorträge 5
6.11.2012	Erfahrungsberichte 1	8.1.2013	Vorträge 6
13.11.2012	Erfahrungsberichte 2	15.1.2013	Vorträge 7
20.11.2012	Vorträge 1	22.1.2013	Vorträge 8
27.11.2012	Vorträge 2	29.1.2013	Abschlussdiskussion



Literatur

- Balestri, Marcello, Pacchiotti, Alberto, Quazza, Silvia, Salza, Pierluigi und Sandri, Stefano, 1999. Choose the best to modify the least: A new generation concatenative synthesis system. In *6th European Conference on Speech Communication and Technology*, S. 2291–2294. Budapest.
- Baumann, Timo, Atterer, Michaela und Schlangen, David, 2009. Assessing and improving the performance of speech recognition for incremental systems. In *Proc. NAACL/HLT 2011*. Boulder, CO.
- Bernsen, Niels Ole, Dybkjaer, Hans und Dybkjaer, Laila, 1998. *Designing interactive speech systems : from first ideas to user testing*. Springer, London.
- Charniak, Eugene, 1993. *Statistical Language Learning*. MIT Press, Cambridge, MA.
- Chen, Fang, 2006. *Designing human interface in speech technology*. Springer, New York.
- Comaneci, Radu, Paetzel, Maike, Pricop, Mircea und Schlesinger, Philipp, 2012. Projektbericht spoken dialog systems.
- Dahl, Deborah (Hg.) 2004. *Practical spoken dialog systems*. Kluwer Academic Publ., Dordrecht.
- Eppinger, B. und Herter, E., 1993. *Sprachverarbeitung*. Hanser, München.
- Euler, Stephen, 2006. *Grundkurs Spracherkennung : vom Sprachsignal zum Dialog ; Grundlagen und Anwendung verstehen*. Vieweg, Wiesbaden.
- Fellbaum, Klaus, 1984. *Sprachverarbeitung und Sprachübertragung*, Bd. 12 von *Nachrichtentechnik*. Springer, Berlin.
- Fink, Gernot A., 2003. *Mustererkennung mit Markov-Modellen : Theorie, Praxis, Anwendungsgebiete*. Teubner, Stuttgart etc.
- Furui, Sadaaki, 1989. *Digital Speech Processing, Synthesis, and Recognition*. Marcel Dekker, New York.
- Harris, Randy Allen, 2005. *Voice interaction design : crafting the new conversational speech systems*. Elsevier/Morgan Kaufmann, Amsterdam.
- Hess, Wolfgang, 1996. Neuere entwicklungen in der sprachsynthese. *Spektrum der Wissenschaft*, (12), 100–.
- Holmes, John N., 1991. *Sprachsynthese und Spracherkennung*. Oldenbourg, München.
- Jelinek, F., 1976. Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, Bd. 64(4), 532–556.
- Jelinek, F., 1990. Self-organized language modeling for speech recognition. In Alex Waibel und Kai-Fu Lee (Hg.) *Readings in Speech Recognition*, S. 450–506. Morgan Kaufmann, San Mateo.

- Jelinek, Frederick, 1998. *Statistical methods for speech recognition*. MIT Press, Cambridge, Mass.
- Jokinen, Kristiina und McTear, Michael, 2010. *Spoken Dialogue Systems*. Morgan & Claypool, San Rafael, CA.
- Jurafsky, Daniel und Martin, James H., 2008. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall, Upper Saddle River, NJ.
- Kohler, Klaus J., 1977. *Einführung in die Phonetik des Deutschen*. Schmidt, Berlin.
- Kompe, Ralf, 1997. *Prosody in speech understanding systems*, Bd. 1307 von *LNAI*. Springer-Verlag, Berlin.
- Lee, K. F., 1999. *Automatic Speech Recognition: The Development of the Sphinx System. 4th edition*. Kluwer, Boston.
- Lewis, James R., 2011. *Practical speech user interface design*. CRC Press, Boca Raton etc.
- McTear, Michael F., 2004. *Spoken dialogue technology : toward the conversational user interface*. Springer, London.
- Ney, Hermann, 1992. Stochastic grammars and pattern recognition. In Pietro Laface und Renato De Mori (Hg.) *Speech Recognition and Understanding: Recent Advances, Trends and Applications*, S. 319–344. Springer, Berlin.
- Nöth, E., Batliner, A., Kießling, A., Kompe, R., Gallwith, F., Warnke, V. und Niemann, H., 1997. Spracherkennung und prosodie. *KI*, (4), 14–19.
- Nöth, Elmar, 1989. *Prosodische Information in der automatischen Spracherkennung — Berechnung und Anwendung*. Dissertation, Universität Erlangen–Nürnberg.
- Picone, Joseph, 1990. Continuous speech recognition using hidden markov models. *IEEE Acoustics, Speech, and Signal Processing Magazine*, Bd. 7(3), 26–41.
- Portele, Thomas, 1996. Künstliche sprache aus dem baukasten: datenbasierte sprachsynthese. *Spektrum der Wissenschaft*, (12), 106–.
- Rabiner, L. R. und Juang, B. H., 1986. An introduction to hidden markov models. *IEEE Acoustics, Speech, and Signal Processing Magazine*, Bd. 3(1), 4–16.
- Ruske, Günther, 1994. *Automatische Spracherkennung: Methoden der Klassifikation und Merkmalsextraktion*. Oldenbourg, München.
- Schafer, Ronald W. und Rabiner, Lawrence R., 1975. Digital representations of speech signals. *Proceedings of the IEEE*, Bd. 63(4), 662–667.
- Schlangen, David und Skantze, Gabriel, 2009. A general, abstract model of incremental dialogue processing. In *Proc. EACL 2009*. Athens, Greece.

- Silverman, Harvey F. und Morgan, David P., 1990. The application of dynamic programming to connected speech recognition. *IEEE Acoustics, Speech, and Signal Processing Magazine*, Bd. 7(3), 6–25.
- Stöber, Karlheinz, Portele, Thomas, Wagner, Petra und Hess, Wolfgang, 1999. Synthesis by word concatenation. In *6th European Conference on Speech Communication and Technology*, S. 619–622. Budapest.
- Taylor, Paul, 2009. *Text-to-speech synthesis*. Cambridge University Press, Cambridge.
- Vaissière, J., 1985. Speech recognition: A tutorial. In F. Fallside und W. A. Woods (Hg.) *Computer Speech Processing*, S. 191–242. Prentice-Hall International, London.
- Vaissière, J., 1988. The use of prosodic parameters in automatic speech recognition. In H. Niemann (Hg.) *Recent Advances in Speech Understanding and Dialog Systems*, NATO ASI. Springer.
- Vary, Peter, Heute, Ulrich und Hess, Wolfgang, 1998. *Digitale Sprachsignalverarbeitung*. B. G. Teubner, Stuttgart.
- Wendemuth, Andreas, 2004. *Grundlagen der stochastischen Sprachverarbeitung*. Oldenbourg, München.