

Phonetik und Sprachtechnologie

Sprachtechnologie umfasst die Bereiche Spracherkennung und -synthese, Sprechererkennung und Dialogsysteme und wurde traditionell eher den Ingenieurwissenschaften und der Informatik zugeordnet. Zunehmend wird jedoch deutlich, dass Grundlagenwissen über die menschliche Kommunikation dabei nicht zu vernachlässigen ist. Hier ist die Phonetik gefragt mit ihren Untersuchungen zur Produktion, zur akustischen Übertragung und zur Wahrnehmung gesprochener Sprache. Passend hierzu hat sich die klassische Phonetik mittlerweile zunehmend naturwissenschaftlich orientiert und Forschung im Bereich der automatischen Sprachverarbeitung mit phonetischem Grundlagenwissen verbunden. Die Grundlagenforschung profitiert im Gegenzug von sprachtechnologischen Entwicklungen, indem sich nun empirisches Arbeiten mit Sprachdaten zu einem großen Teil automatisieren lässt.

Auch der Student profitiert natürlich durch ein vielseitiges Studium, in dem sowohl grundlagen- als auch anwendungsorientiertes Forschen vermittelt wird, und das ihn mit guten Aussichten in den Arbeitsmarkt entlässt. Auch als Nebenfach in Studiengängen wie der Informatik oder der Computerlinguistik ist die Phonetik eine sinnvolle Wahl.

Das **Institut für Phonetik und Sprachverarbeitung (IPS)** bietet für technisch interessierte Studenten einige anregende Studienfelder, darunter die folgenden:

Spracherkennung Spracherkennung bedeutet die Umwandlung eines akustischen Sprachsignals in den zugrundeliegenden Text, also eine Umwandlung von Signalen in Symbole (Lautsymbole, Buchstaben, Wörter). Hierzu werden Sprachsignale in kleine Abschnitte unterteilt (Fensterung) und in jedem dieser Fenster akustische Merkmale extrahiert, anhand de-

rer die Zuordnung von Signalabschnitten zu Lauten möglich wird. Hierbei kommen statistische Methoden (u.a. sog. Hidden Markov Modelle) zum Einsatz. Bei der akustischen Repräsentation des Sprachsignals wird die menschlichen Wahrnehmung nachempfunden, indem beispielsweise die Frequenzskala durch die perzeptiv fundierte mel-Skala ersetzt wird. Als weitere Wissensquellen werden statistische Sprachmodelle für Laut- und Wortfolgewahrscheinlichkeiten genutzt. Zu den aktuellen Forschungsgebieten zählt hier die Spracherkennung bei Hintergrundlärm (z.B. im Auto).

Sprachsynthese Im Gegensatz zur Spracherkennung bedeutet Sprachsynthese die Umwandlung von Text in Sprachschall, also die Umwandlung von Symbolen in Signale. Simple Module dieser Art sind beispielsweise in Navigationssystemen zu finden. Der Text wird hierbei zunächst in eine Folge von Lautsymbolen umgewandelt. Anhand höherer linguistischer Analysen wird bestimmt, welche Wörter der Äußerung betont werden sollen und wie die Äußerung zu phrasieren ist. Hierdurch ergeben sich die prosodischen Aspekte der Äußerung (Sprechmelodie, Dauer der Sprachlaute, Lautstärkeverhältnisse der Silben). Mit den aus diesen Verarbeitungsschritten gewonnenen Informationen wird je nach Synthesystem das akustische Signal direkt generiert (Formantsynthese), ein artikulatorisches Modell angesteuert, das die Äußerung dann spricht (artikulatorische Synthese), oder aus einer Datenbank mit kleinen Sprachsignalstücken die geeigneten Stücke selektiert und verkettet (konkatenative Synthese). Aktuelle Forschungsgebiete sind hier unter anderem Verbesserung der Sprechmelodie und Modellierung von Emotionen in der Stimme.

Sprachdialogsysteme Sprachdialogsysteme erlauben es einem menschlichen Benutzer, mit einer Maschine mittels natürlichsprachlicher Ein- und Ausgabe zu kommunizieren (Mensch-Maschine-Kommunikation). Die Forschung für die Entwicklung von Dialogsystemen findet in interdisziplinärer Zusammenarbeit mit verschiedensten Fächern wie Sprachwissenschaften, Computerlinguistik, Informatik, Psychologie statt. Zu den Komponenten dieser Systeme gehören: Spracherkennung und -verstehen, Dialogsteuerung, Abruf von Datenbanken, Antwortgenerierung und Sprachsynthese. Heutige Sprachdialogsysteme zeichnen sich u.a. dadurch aus, dass sie komplexe Anfragen verarbeiten können, eine natürliche Kommunikation ermöglichen, jederzeit unterbrochen werden können (Barge-In) und sich an den jeweiligen Benutzer adaptieren.

Sprachdatenbanken Sprachdatenbanken sind notwendig zum Trainieren und Evaluieren der oben beschriebenen Systeme. Im Rahmen des am IPS beheimateten BAS (Bayerisches Archiv für Sprachsignale) werden solche Datenbanken erstellt. Da auch die Wirtschaft sehr an solchen Daten interessiert ist, ergeben sich gerade hier viele Industriekontakte und auch Drittmittel. Die Erstellung von Datenbanken erfordert ihrerseits diverse Methoden aus Informatik und Ingenieurwissenschaften, beispielsweise zur Alinierung von Signal und Text, also der Abbildung der Lautsymbole und Textpassagen auf die entsprechenden Stellen im Sprachsignal.

Darüberhinaus lässt sich anhand dieser Daten unter Verwendung einfacher Script-Sprachen statistisch fundierte empirisch-linguistische Grundlagenforschung betreiben, die sich noch nicht überall in den Sprachwissenschaften mangels ausreichender Datenmengen und Auswertungsmöglichkeiten durchgesetzt hat.

Zusammenfassung Eine technisch ausgerichtete Phonetik leistet also erstens durch die hier erworbenen Kenntnisse über das Funktionieren lautsprachlicher Kommunikation einen wichtigen Beitrag für ingenieurwissenschaftliche Fachbereiche. Zweitens kann auch die linguistische Forschung im Allgemeinen

von den empirischen Methoden und den technischen Hilfsmitteln der Phonetik profitieren. Für Studenten ergibt sich hier ein interessantes Studiengbiet mit der Möglichkeit, als studentische Hilfskraft in diversen Drittmittelprojekten praktische Erfahrung zu sammeln und damit Qualifikationen für den Arbeitsmarkt.

Literatur

- Harrington, J., Cassidy, S. (1999). *Techniques in Speech Acoustics*. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Jurafski, D., Martin J.H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice-Hall, Upper Saddle River, New Jersey.
- Terhardt, E. (1998). *Akustische Kommunikation*. Springer, Berlin/Heidelberg.
- Van Santen, J., Sproat, R.W., Olive, J.P., Hirschberg, J. (Hrsg.) (1996). *Progress in Speech Synthesis*. Springer, New York

Institutsadresse

Institut für Phonetik und Sprachverarbeitung
LMU München
Schellingstr. 3/II
80799 München
Tel.: +49 (0)89 2180-2758
Homepage: <http://www.phonetik.uni-muenchen.de>

Weitere Informationen

Uwe Reichel
Schellingstr. 3, Raum 249 VG
Tel.: +49 (0)89 2180-2812
Mail: studienberatung@phonetik.uni-muenchen.de