

Automatische Spracherkennung

Dozent: Florian Schiel

1 Einführung

Dieser Kurs ist eine grundlegende Einführung in die maschinelle Erkennung gesprochener Sprache (automatische Spracherkennung). Es sind keine Voraussetzungen notwendig, insbesondere keine mathematischen oder Programmierkenntnisse.

Sind alle Teilnehmer in LSF zugelassen? Muss jemand nacherfasst werden?

In diesem Semester (WS 18/19) gibt 14 Vorlesungstermine; Termin am 5.2.17 wird ausfallen, weil an Tage danach die Modulprüfung gemeinsam mit dem Kurs P6.2 stattfindet.

1.1 Prüfungsleistung

Für diesen Kurs wird auf Wunsch ein Testat erteilt (Proseminar-Schein); Voraussetzung ist die erfolgreiche Teilnahme bei der Abschlussprüfung (per LSF zugelassene Studenten benötigen keinen Schein, die Note wird in LSF eingetragen). Eine Seminararbeit muss nicht erstellt werden.

Das aktuelle Skript können Sie online unter

<http://www.phonetik.uni-muenchen.de/>

im Menu

'Studium / Lehrmaterialien / P6.1 Einführung in die automatische Spracherkennung'

abrufen.

Wenn Sie Fragen zu den Inhalten des Kurses haben, können Sie mich jederzeit in meinem Büro (Schelling 3 Vordergebäude, Zi 219) ansprechen (wenn die Türe offen steht) oder per email.

1.2 Übung

Zur Vorlesung findet keine Übung statt.

1.3 Aufgabenstellung

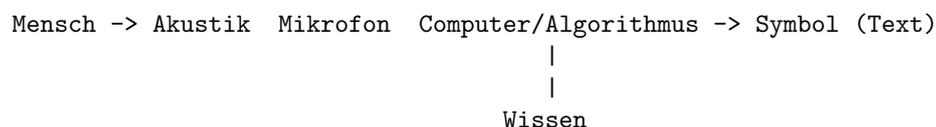
Automatische Spracherkennung =

'Die zuverlässige Erkennung einer lautsprachlichen Äußerung nur anhand des akustischen Sprachsignals mit endlichen Mitteln, in endlicher Zeit und ohne menschliche Hilfe.'

Hier:

Nur der 'akustische Teil' der Spracherkennung. D.h. keine höheren Verarbeitungsstufen wie Syntax, Semantik, Dialog- Modellierung, etc.

Skizze:



Es handelt sich also um ein Problem aus dem Bereich der Mustererkennung.

In Skizze einfügen: Störparameter (andere Farbe)

Mensch: Stimmung, Geschwindigkeit, gesundh. Zustand, Erschöpfung, kooperativ/n. kooperativ, Lombard-Effekt.

Akustik: Wechselnde Umgebung (Echo/Hall), Hintergrund- geräusche, Abstand zum Mikrofon.

Mikrofon: Versch. Charakteristik, Empfindlichkeit, Übertragungskanal, Rauschen, Störgeräusche.

Algorithmen: Modellannahmen, Rechenungenauigkeit.

1.4 Übersicht

1 Übersicht: Kursaufbau, Aufgabenstellung, Literatur

2 Grundlagen und Begriffe:

Sprachsignal

Abtasttheorem

Merkmale

Spektrum

Lineare Praediktion

Mel-Cepstrum

Vektorquantisierung

Mustererkennung

Bedingte Wahrscheinlichkeiten

Bayesscher Ansatz der Spracherkennung

Sprachmodellierung

Sucheproblem

Evaluation

3 Vertiefung:

3 wichtige Methoden/Algorithmen:

Einfacher Pattern-Match mit DTW
Statistisches Verfahren mit HMM
Neuronale Netze, Deep Learning (Hybride Verfahren)

4 Wissenschaftliches Arbeiten:
Evaluation von ASR
Typisches Experiment

5 Typische Fehlerquellen

1.5 Literatur

Gute Einführungen:

Euler S (2006): Grundkurs Spracherkennung. Vieweg Wiesbaden.

Pfister B, Kaufmann T (2008): Sprachverarbeitung - Grundlagen und Methoden der Sprachsynthese und Spracherkennung. Springer-Verlag Berlin Heidelberg.

Die Verwendung von ANN in der ASR:

Bouclard H, Morgan N (1994): Connectionist Speech Recognition - A Hybrid Approach Kluwer Academic Publishers, Engineering and Computer Science.

Beispiel für eine spezielle Merkmalsextraktion:

Hermansky H, Morgan N, Bayya A, Kohn P (1991): Compensation for the Effect of the Communication Channel in Auditory-Like Analysis of Speech (RASTA-PLP), EUROSPEECH 1991, Genova, p. 1367 - 1370.

Grundlegender Artikel über das Viterbi-Training von HMM:

Juang B H, Rabiner L R (1990): The segmental k-means algorithm for estimating parameters of Hidden Markov Models. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 38, No. 9, Sep 1990, S. 1639 - 1641.

Einführung in ANN:

Lippmann R S (1987): An Introduction to Computing with Neural Nets, IEEE, ASSP Magazine, April 1987, pp. 4 - 22.

Einführung über Dynamic Time Warping als Mustererkennung:

Ney H, Ortmanns S (1999): Dynamic Programming Search for Continuous Speech Recognition. in: IEEE Signal Processing Magazine, Sept 1999, pp. 64-83.

Gute Einführung in HMM in der ASR:

Picone J (1990): Continuous Speech Recognition Using Hidden Markov Models. IEEE ASSP Magazine, Jul 1990, S. 26 - 41.

Grundlegender Artikel über HMM:

Rabiner L R, Juang B H (1996): An Introduction to Hidden Markov Models IEEE ASSP Magazine, Jan 1986, p. 4.

Einführendes Buch über Akustische Phonetik; Kapitel 2.4 - 2.8 sind sehr gute Erklärungen zum digitalen Signal und Spektrum:

Reetz H (2003): Artikulatorische und akustische Phonetik, Wissenschaftlicher Verlag Trier.

Gutes Lehrbuch über alle Arten von Sprachverarbeitung:

Jurafsky D, Martin J H (2000): Speech and Language Processing. Prentice Hall, Kap I.7.

Mathematische Theorie der Spracherkennung; ziemlich schwierig; interessant (auch ohne viel Mathe-Kenntnisse): Kapitel 7:

Levinson S (2005): *Mathematical Models for Speech Technology*, John Wiley & Sons, UK.

Wissenschaftliches Arbeiten:

Bördlein S (2002): *Das Sockenfressende Monster in der Waschmaschine*. Alibri Verlag Gunnar Schedel.

Vorlesung "Wissenschaftliches Denken" von Chr. Roderer (dieses Semester ausnahmsweise auch für Bachelor Studenten), Termin: Do 8:30 S 226 (Bibliothek)