

## Bachelorarbeit / Masterarbeit

# “Unüberwachtes Lernen von akustischen Einheiten zum maschinellen Spracherwerb”

## Themengebiet

Ziel ist es, Algorithmen zu entwickeln die mit minimalem Vorwissen, also unüberwacht, Struktur in Audioaufzeichnungen lernen. Traditionelle Spracherkennungsalgorithmen werden überwacht trainiert. Zu erkennende Muster, also Worte und Phoneme (die akustischen Einheiten der Worte) sind vorab bekannt. Statistische Modelle werden anhand von transkribierten und gelabelten Trainingsdaten gelernt. Das Transkribieren und Labeln von Trainingsdaten ist allerdings aufwendig und teuer, daher ist es das Ziel, Algorithmen zu entwickeln die Phoneme und Phonemfolgen, also Worte, selbstständig lernen können. Als Analogie könnte man sagen: Wir möchten, dass der Algorithmus eine Sprache wie ein Kind erlernt.

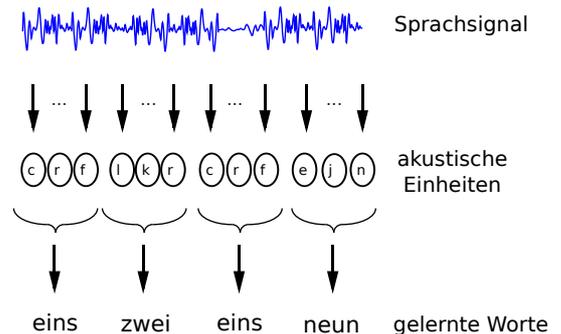


Abbildung 1: Vom Signal zum Wort

## Aufgabenstellung

Im Rahmen der studentischen Arbeit soll ein Algorithmus zum Lernen von akustischen Einheiten untersucht werden. Ziel ist hierbei ein System zum unüberwachten Lernen von Worten zu erweitern. Hierzu ist ein Algorithmus zum Lernen von akustischen Einheiten zu implementieren. Anschließend ist eine Bewertung der Qualität der gefundenen Einheiten durchzuführen. Des Weiteren kann eine Integration in den bestehenden Algorithmus zum Lernen von Worten durchgeführt werden und eine Rückkopplung der gelernten Worte in den Lernalgorithmus zur Verbesserung der akustischen Einheiten untersucht werden. Die Aufgabenstellung umfasst die folgenden Punkte:

- Einarbeitung in die Literatur zum unüberwachten Lernen und maschinellen Spracherwerb
- Implementierung eines Algorithmus zum Lernen von akustischen Einheiten
- Evaluierung des Algorithmus und Einbindung in den Algorithmus zum Lernen von Worten
- Dokumentation und Präsentation der Arbeit

- **Betreuer:** Oliver Walter
  - ⇒ E-Mail: walter@nt.uni-paderborn.de
  - ⇒ Telefon: 05251 - 60 / 3624
  - ⇒ Raum: P7.2.05.1

Wir haben ihr Interesse geweckt? Dann besuchen Sie uns in unseren Büros! Auch weitere Arbeiten möglich!