Spracherkennung zur Transkription von Radiospots

Kristoffer Witt – 13.01.2009
Master of Science „Verteilte Systeme“
HAW HAMBURG, AW1

Gliederung

Wie ist der Vortrag unterteilt?

- Einleitung/ Szenario
- Problemstellung
  - Mögl. Verwendung
- Grundlagen
  - Segmentierung
  - Spracherkennung
- Probleme/Risiken
- Perspektive AW2 und Projekt
- Diskussion

Szenario

Was sind die Rahmenbedingungen?

- AdVision Digital GmbH
- Webportal „Ädzyklopädie“
- Zugriff auf Werbemaßnahmen für
  - TV
  - PRINT
  - PLAKAT
- Durchsuchbar nach bestimmten Kriterien

- Händisch vergeben:
  - 3 Stufige Hierarchie
    - Konzern, z.B. EADS
    - Marke, z.B. AIRBUS
    - Produkt, z.B. A380
- Zusätzlich weitere Informationen

- http://v2.adzyklopaedie.com

Szenario (2)

Wie funktioniert eine Recherche in der Adzyklopädie?

Welche Kriterien gibt es?

- Neues Medium:
  - Radio
- Beispielspot „Saitenbacher“ (http://www.saitenbacher.de)

- Transkription zwecks verbesserter Durchsuchbarkeit
- Versehen der Audiodaten mit Text

Problemstellung

Welches Problem gilt es zu lösen bzw. zu untersuchen?

- Transskript für Gehörlose
- Unterstützung bei Vergabe von manuellen Kriterien
- Fingerprinting
- Navigation innerhalb des Clips

Mögl. Verwendungen

Weitere Mögliche Anwendungen für automatisch erfassten Sprachinhalt

I. Audiodatendarstellung
II. Audiosegmentierung
III. Spracherkennung

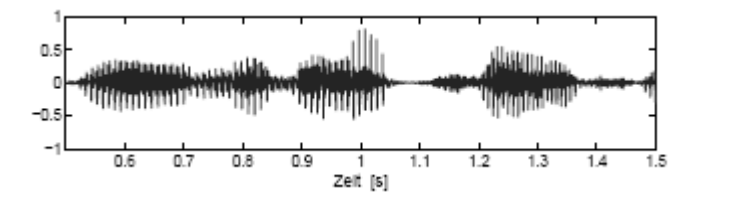
Grundlagen

Welches sind die grundlegenden Technologien?

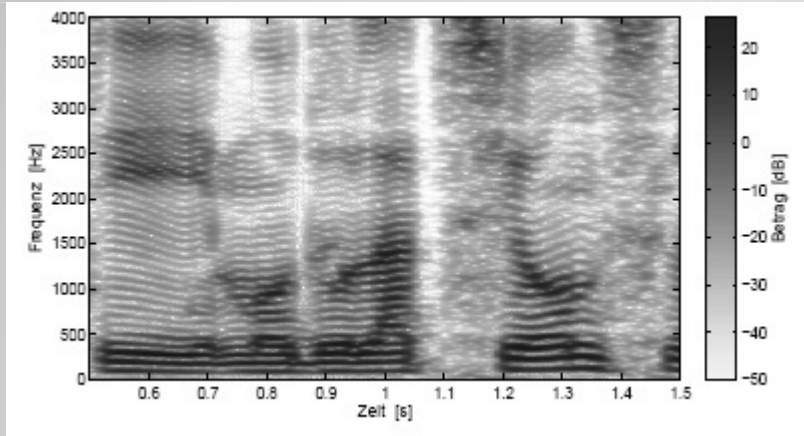
Exkurs „Audiosignaldarstellung“

Wie können Audiodaten dargestellt werden?

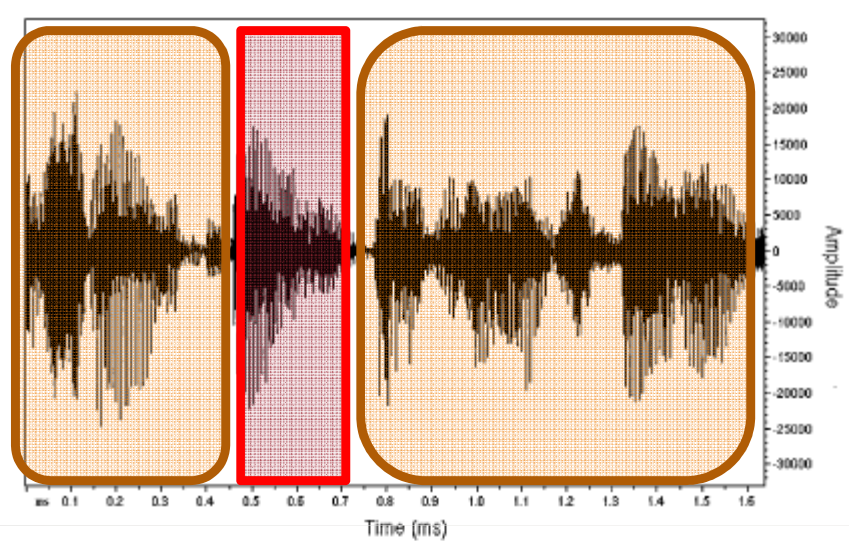
- Schwingungsverlauf



Aus [Pfister und Kaufmann 2008], Seite 49



Aus [Pfister und Kaufmann 2008], Seite 49



Segmentierung

Unterteilung der Audiodaten zur besseren Verarbeitung.

Erstmal nur in Musik und Sprachdaten

Kriterien	Sprache	Musik
Bandbreite	0-7 kHz	0-20 kHz
Stilleanteil	hoch	niedrig
„Beat“	Keiner	häufig

[Lu 2001]

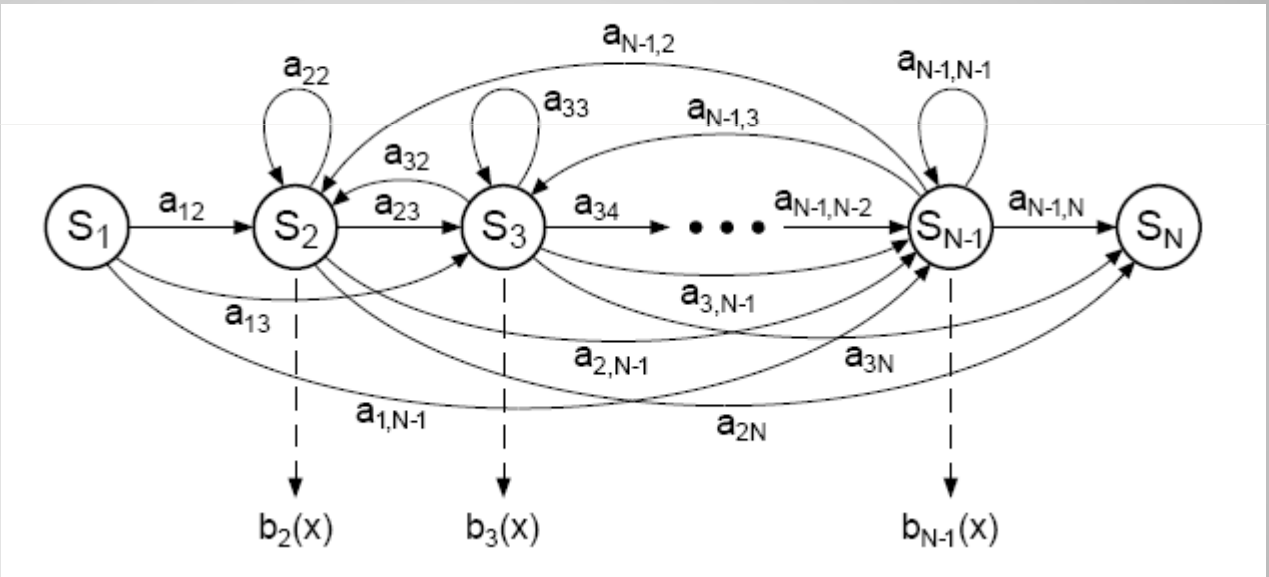
Spracherkennung

Wie wird aus Audiodaten Text gewonnen?

- Unterschiedliche Typen:
  - Keywordspotter =>Vergleich
  - Continous Word Sequence Recognizer =>Stochastisch Wortmodelle (HMM)
- Abbildung von Merkmalen auf Laute

Hidden Markov Modell

- Zustände(S_i) entsprechen einem Laut



- Heterogenität
  - Komplett Unterschiedliche Daten
- Nicht für Sprachererkennung optimiert
  - Nebengeräusche
  - Hintergrundmusik
  - Mehrere Sprecher (keine Profile)
  - Unbekannte Umgebungen
  - Prosodie

Probleme

Welche Schwierigkeiten treten bei der Erkennung auf?

Perspektive für AW2 und das Projekt

- Akquise von geeigneten Testdaten
- Entwurf und Umsetzung einer Testplattform
  - Unterschiedliche Erkennungssysteme
  - Verschiedene Segmentierungslösungen
  - Möglichst hohe Erkennungsrate

Vielen Dank für die Aufmerksamkeit

Gibt es noch Fragen?