

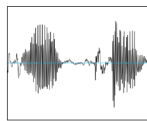
WebMAUS: automatisches Segmentieren und Etikettieren von Sprachsignalen über das Web

Thomas Kisler

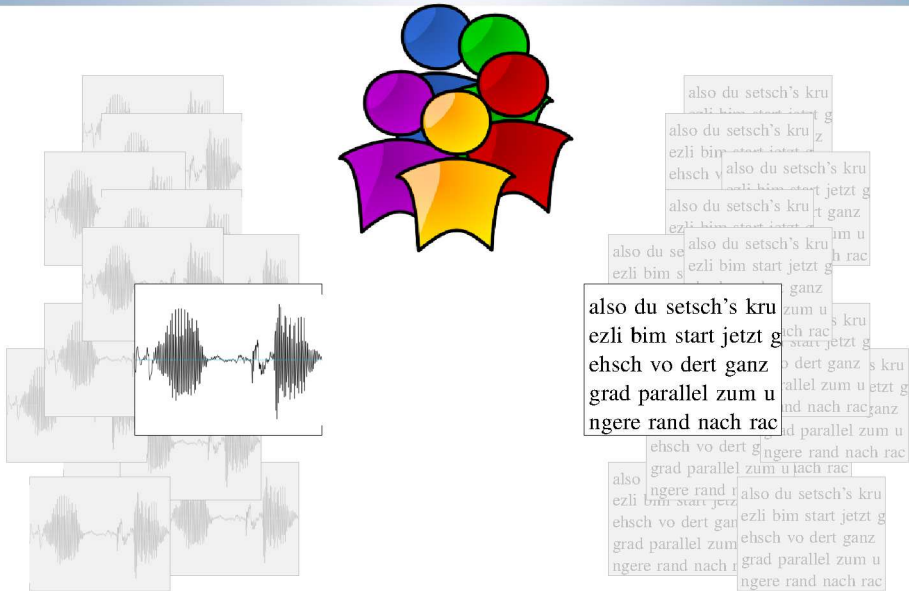
Bayerisches Archiv für Sprachsignale, Institut für Phonetik und Sprachverarbeitung

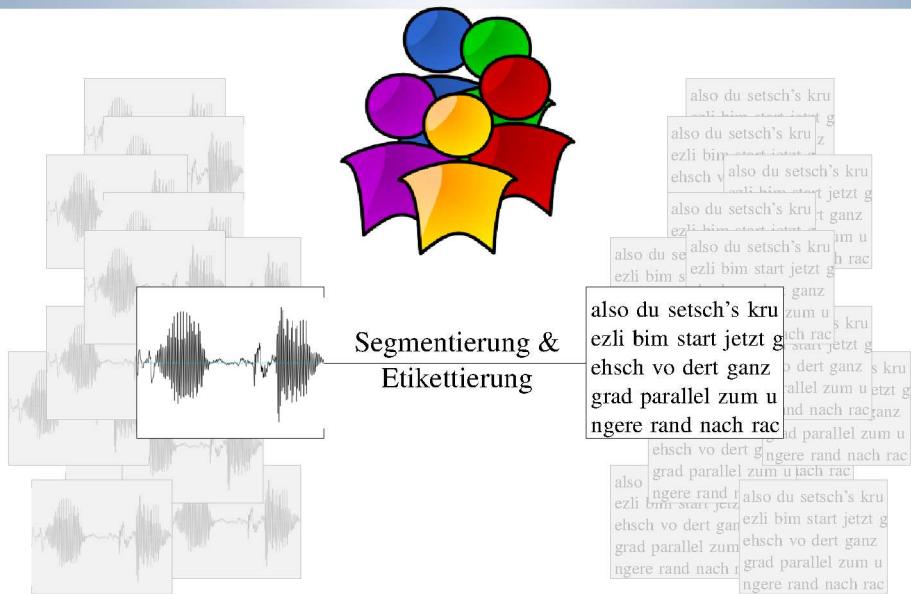
October 15, 2014





also du setsch's kru
ezli bim start jetzt g
ehsch vo dert ganz
grad parallel zum u
ngere rand nach rac





Segmentieren und Etikettieren



Manuelles Segmentieren & Etikettieren (S&E) von Audio-Signalen

- Sehr zeitaufwändig
- Phonetiker / ausgebildetes Personal
- Fehleranfällig
- Interlabeller agreement
- Mühsame Arbeit

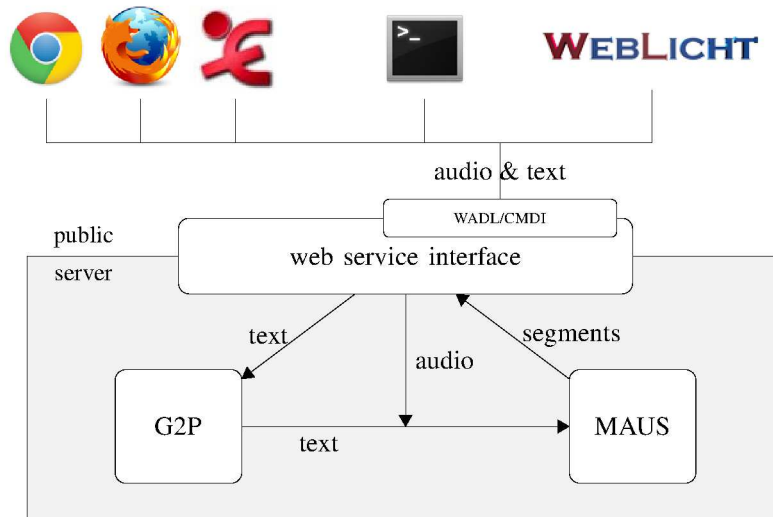
Segmentieren und Etikettieren

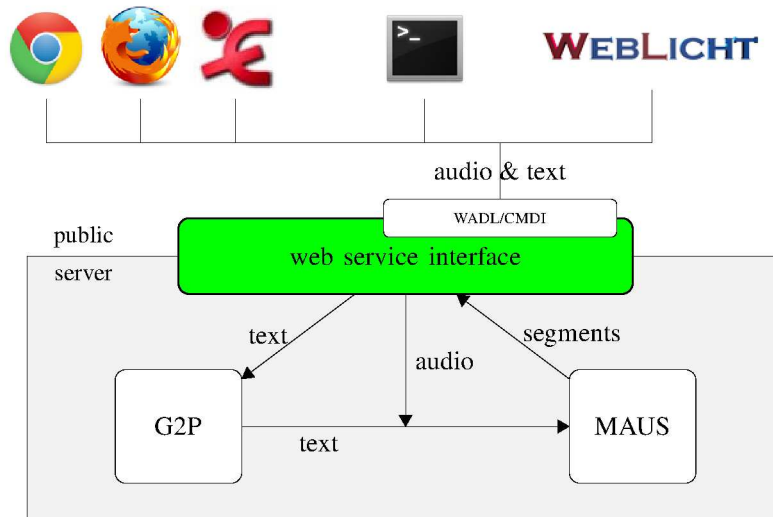


Manuelles Segmentieren & Etikettieren (S&E) von Audio-Signalen

- Sehr zeitaufwändig
- Phonetiker / ausgebildetes Personal
- Fehleranfällig
- Interlabeller agreement
- Mühsame Arbeit

Lösung ⇒ Automatisierung



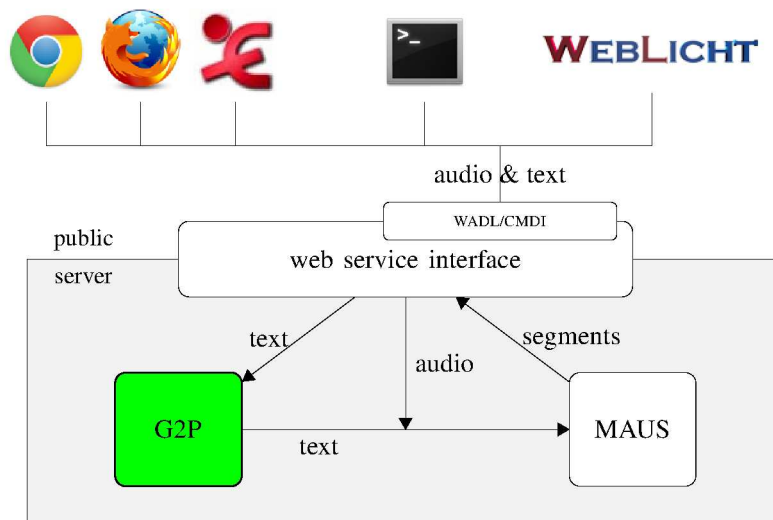


- Web Services (beschrieben durch WADL & CMDI)
 - einfach durch Skripte aufrufbar (z.B. mit curl)
 - automatische Service Aufrufe (z.B. Hilfe mit Beispielen ist automatisch generiert)
- Integrierbarkeit
- Web 2.0 Schnittstelle
 - einfach zu benutzen
 - hübsch(er)
 - keine Installation notwendig
 - immer aktuellste Version

WebMAUS - unter der Haube



- WebMAUS
 - Java Wrapper um G2P & MAUS
 - RESTful Schnittstelle mit Jersey (JAX-RS)
 - Web Schnittstelle Typescript, angularjs, HTML5
- G2P → Perl Programm (Uwe Reichel)
- MAUS → C++ Framework um HTK (Florian Schiel)
Warum MAUS? Munich AUtomatic Segmentation



- Input: orthographische Verschriftung, z.B. “schönes Frühlingswetter”
- Output: kanonisches Transkript, z.B. “dax” in SAM-PA¹
- Methode 1: Nachschlagen im Lexikon - wenn Wort existiert

Beispiel:

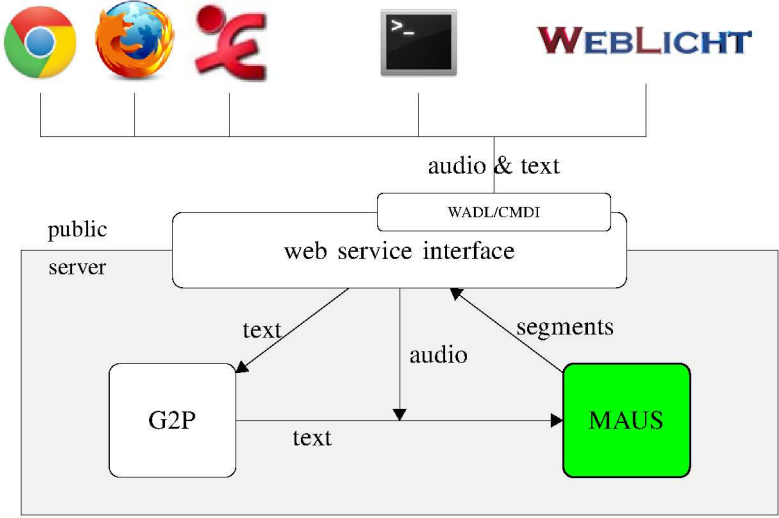
- Input: “Zug”
- Output: “tsu:k”

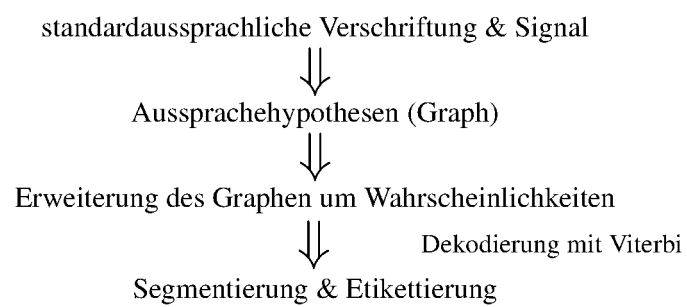
¹SAM phonetisches Alphabet

- Methode 2 (falls nicht 1): Konvertierung basierend auf Entscheidungsbaum in sequentiell kleiner werdender Phonem-Nachbarschaft

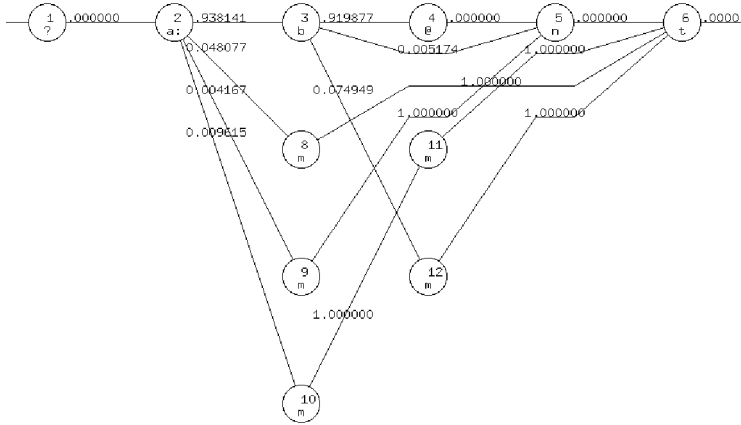
Beispiel:

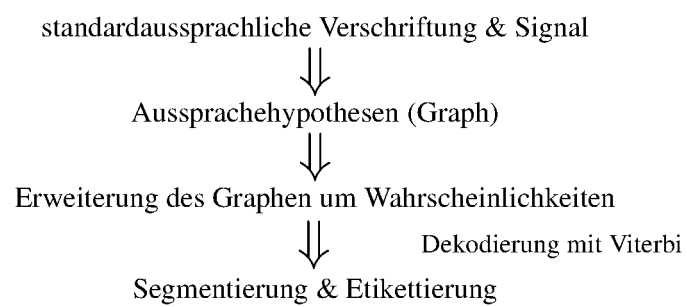
- Input: "puckedidudel"
- Output (deutsches Modell): "pUk@didu:d@l"
- Output (englisches Modell): "pVkdIdju:dI"





MAUS - Probability Space





MAUS - Result

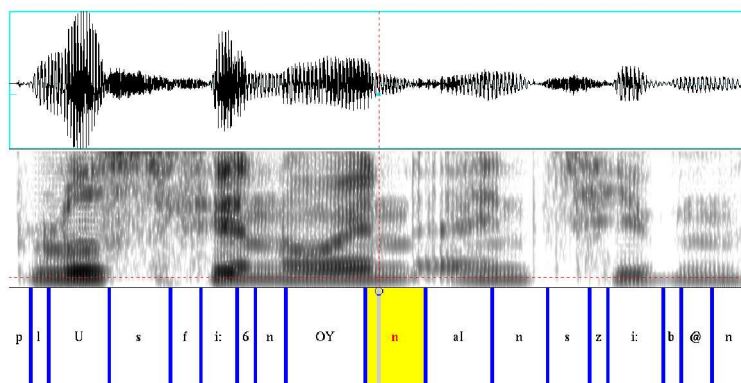


Abbildung: Beispiel einer von MAUS erzeugten Segmentierung und Etikettierung

MAUS - Vorteile



- Produziert gute und objektive/reproduzierbare Grenzen
- Besonders bei großen Korpora enorme Zeitersparnis
- Grenzen bedürfen evtl. Korrektur, aber Labels und Grenzen existieren schon
- Einfach erweiterbar
 - phonetisch reiches Material
 - 15-20 Sprecher
 - Sprecher äußern unterschiedliche Sätze
 - idealerweise mit Aussprachelexikon (G2P)

WebMAUS - Features



- Web-Schnittstelle & -Services öffentlich verfügbar
- Signal Dateien mit bis zu 200MB pro Datei
- Signal Datei: ca. 50 Sekunden Sprache, ca. 600 Segmente → ca. 10 Sekunden
- WebMAUS Multiple: erfolgreich getestet mit 4000+ Dateien und bis zu 2000MB pro Anfrage
- WebMAUS Multiple: verarbeitet 10 Dateien gleichzeitig
- 8 Sprachen mit Graphem-zu-Phonem Konvertierung (eng-aus, deu, eng, ita, hun, nld, nze, pol, fre, ge)
- 12 Sprachen forced-alignmenet + !sprachunabhängiges Modell!
(deu-de, eng-GB, eng-AUS, eng-NZ, eng-US, ita-IT, fin-FI, hun-HI, nld-NL, pol-PL, ekk-EE, por-PT, spa-ES, kat-GE, sampa)

*unterstrichene Sprachen wahlweise mit oder ohne adaptivem forced-alignment

Demo WebMAUS



`http://clarin.phonetik.uni-muenchen.de/
BASWebServices`

Zukunfftige Pläne



- Mehr Sprachen
- Sprachvarianten (z.B. Schweizer Deutsch)
- Weitere Web Services:
 - ASSP (Phonetische Signal Verarbeitung)
 - Konvertierungstools
 - Text-Alignment
 - Silbifizierung