

Einführung in die Computerlinguistik

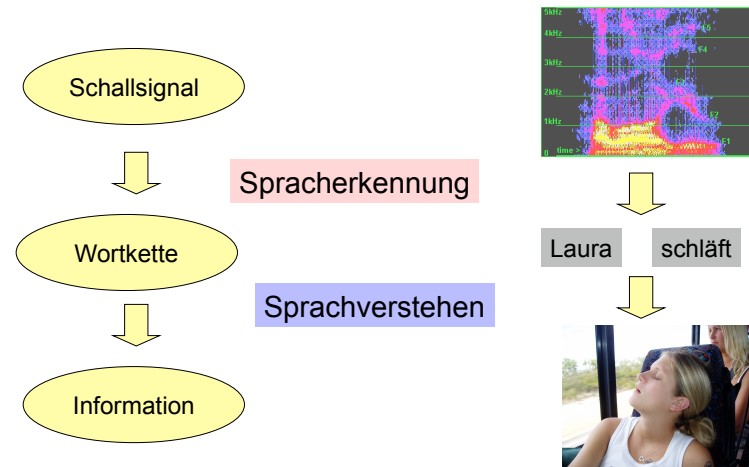
Verarbeitung gesprochener Sprache

WS 2011/2012

Manfred Pinkal

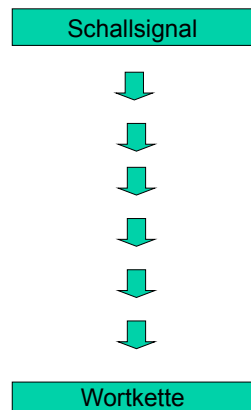
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Sprachverarbeitung



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

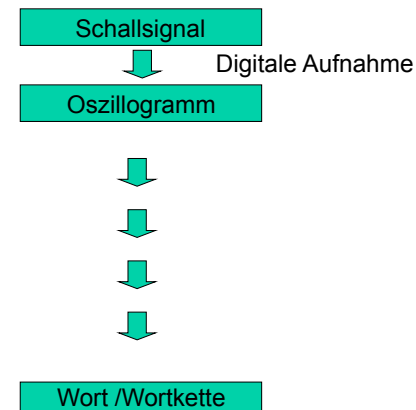
Spracherkennung



- Die Grundaufgabe der Spracherkennung: Gegeben ist ein kontinuierliches Schallsignal. Welche Kette von Wörtern wurde vom Sprecher geäußert?

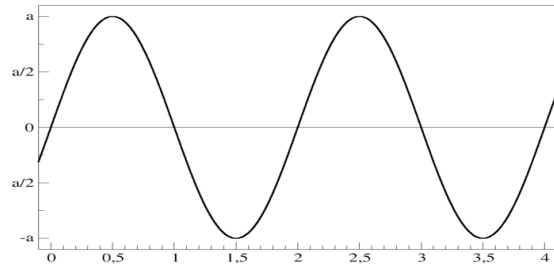
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spracherkennung



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

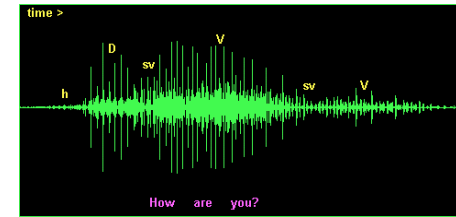
Reine Schwingung



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

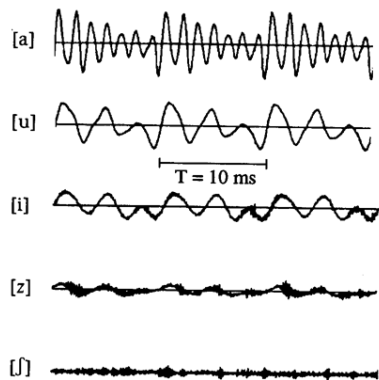
Ein Oszillogramm

- Das Oszillogramm für eine Äußerung des englischen Satzes „How are you“



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

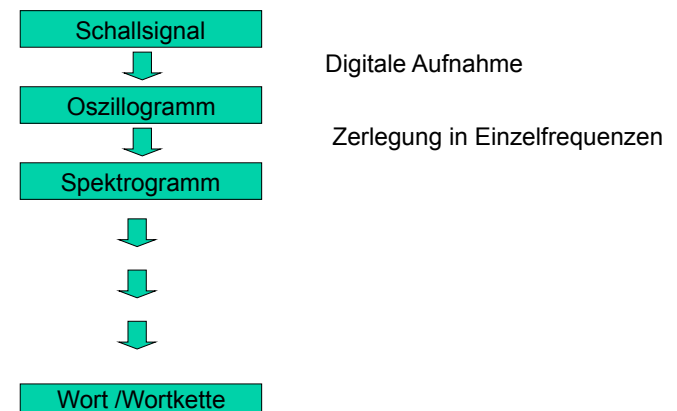
Einzelne Laute als Oszillogramme



- Laute werden charakterisiert durch Kombination von Schwingungen verschiedener Frequenzen
- Im Oszillogramm **schwer erkennbar** (Überlagerung)
- Daher: Geschicktere Repräsentation durch Komponentenanalyse (Fourier-Transformation)
- Ergebnis: Zeit-Frequenz-Diagramm (**Spektrogramm**)

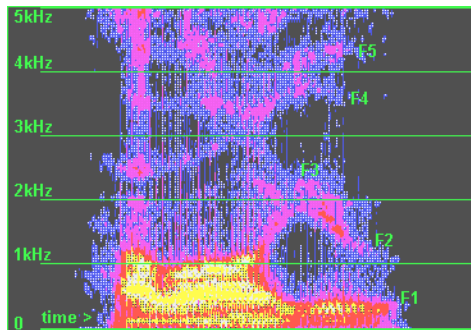
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spracherkennung: (Vereinfachtes) Schema



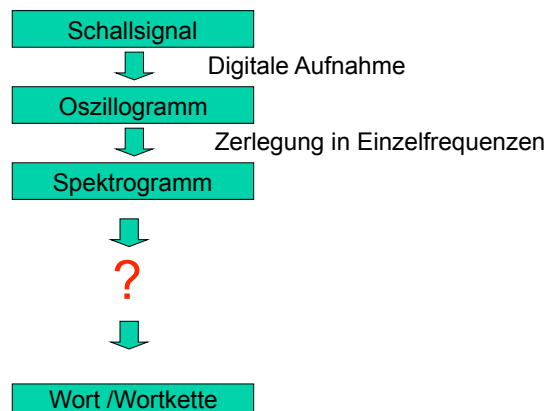
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Ein Spektrogramm



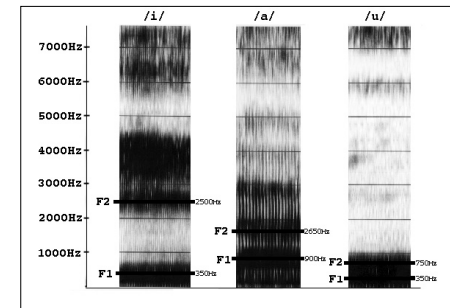
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spracherkennung: (Vereinfachtes) Schema



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spektrogramm für die Vokale i,a,u



- Dunkle Färbung: große Schallenergie in einem bestimmten Frequenzbereich.
- Die **Formanten** (Obertöne) F1 und F2 sind für die charakteristische Vokalqualität verantwortlich.
- Der Verlauf des **Basisformanten** F0 (hier nicht sichtbar) gibt die Intonation der Äußerung wieder.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spracherkennung: Versuch 1

- Identifikation von Lautgrenzen im Spektrogramm (Segmentierung) + Abgleich der Spektrogramm-segmente mit einer Datenbank "idealer" Laute (Identifikation); Verknüpfung der identifizierten Laute zu Wörtern und Sätzen.
- **Funktioniert nicht, wegen der Varianz des Signals.**

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Problem 1: Varianz des Signals

- Gleicher Laut/ gleiches Wort wird nicht immer gleich ausgesprochen
 - Verschiedene Dialekte
 - Verschiedene Sprecher
 - Unterschiedliche Sprechgeschwindigkeit
 - Physischer und emotionaler Zustand des Sprechers
 - Abhängig von Tonhöhe und Akzent
- Sprachexterne Einflüsse verändern das Signal
 - Raumakustik, Hall, Entfernung
 - Medium: direkte Kommunikation, Telefon, Handy
 - Mikrofonqualität und -charakteristik
 - Hintergrundgeräusche

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spracherkennung: Versuch 2

- Identifikation von Lautgrenzen im Spektrogramm (Segmentierung)
- Erstellung eines Trainingskorpus mit Lautannotationen (alignierte phonetische Annotation)
- Bestimmung von Merkmalsmustern für die Spektrogrammsegmente
- Training eines statistischen Laut-Klassifikators

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Problem 2: Kontinuität des Signals

- Die **Laute** eines Wortes lassen sich schwer gegeneinander abgrenzen
 - Wo hört Laut 1 auf, wo fängt Laut 2 an?
 - Dazu kommt das Phänomen der **Koartikulation**: Laute beeinflussen sich gegenseitig.
 - In Lautfolgen wie [am], [um], [an] kann man nicht den Vokal vom Nasal trennen: Vokal hat Nasal-Qualität und umgekehrt.
 - /k/ wird verschieden realisiert in Koffer, Kind, Kabel
- **Wörter** sind nur in der Orthografie sauber getrennt.
 - In der gesprochenen Sprache gibt es zwischen Wörtern meistens keine Pause
 - Pausen kommen in spontaner Sprache auch innerhalb von Wörtern vor

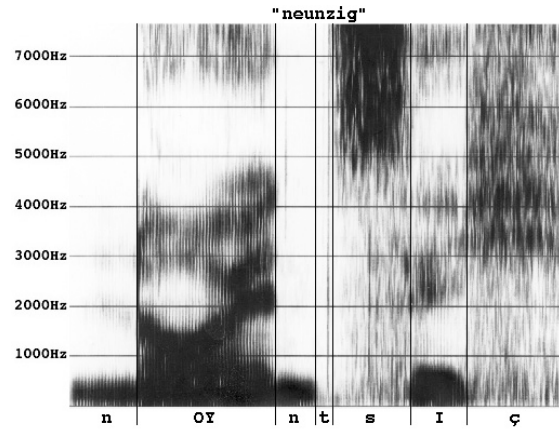
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Statistische Modellierung

- Ermittlung der **wahrscheinlichsten** Wortkette
 - $W = w_1 w_2 \dots w_n$, die einem beobachteten akustischen Signal entspricht.
- Die akustische Information, die durch die Lautspektrographie gemessen wird, ist zu differenziert für statistische Berechnungen.
- Wir erzeugen eine handhabbare Charakterisierung der akustischen Information durch **Merkmalsextraktion**.

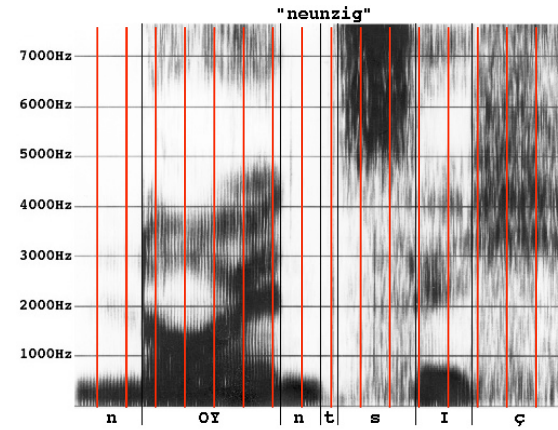
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spektrogramm für ein deutsches Wort



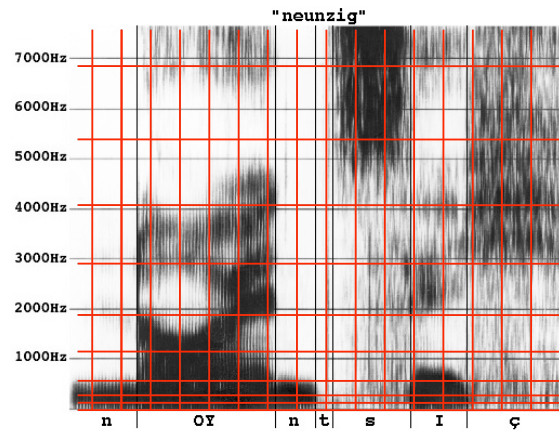
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spektrogramm für ein deutsches Wort



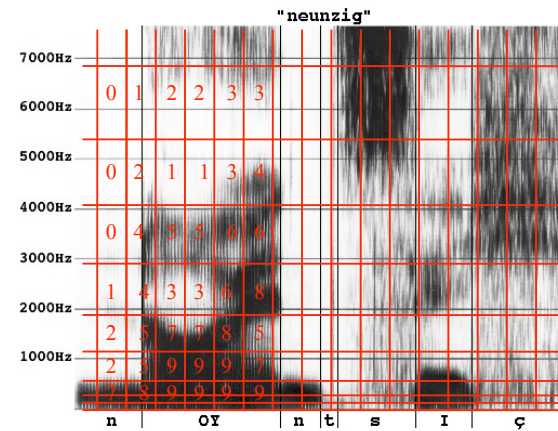
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spektrogramm für ein deutsches Wort



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Spektrogramm für ein deutsches Wort



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Erinnerung: Bayes-Theorem für WSD

- Merkmalsmuster v : Symptom
- Wortsinn s : Ursache
- Mit Bayes-Regel :
$$P(s|v) = \frac{P(v|s) \cdot P(s)}{P(v)}$$
- Der wahrscheinlichste Wortsinn:
$$\max_s P(s|v) = \max_s \frac{P(v|s) \cdot P(s)}{P(v)}$$
$$= \max_s P(v|s) \cdot P(s)$$
- $P(s)$ ist die globale, "a priori"-Wahrscheinlichkeit des Wortsinns s .
- $P(v)$, die Wahrscheinlichkeit des Merkmalsmusters, wird nicht mehr benötigt.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Akustisches Modell und Sprachmodell

$$\max_w P(W|O) = \max_w P(O|W) \cdot P(W)$$

- $P(O|W)$ ist die Wahrscheinlichkeit, dass eine Wortfolge in einer bestimmten (durch den Merkmalsvektor bezeichneten) Weise ausgesprochen wird: **Akustisches Modell**
- $P(W)$ ist die Wahrscheinlichkeit, dass eine bestimmte Wortfolge geäußert wird: „**Sprachmodell**“

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Wie bestimmen wir $P(W|O)$?

- Sparse-Data-Problem!
- Akustisches Merkmalsmuster O : Symptom
- Tatsächlich geäußerte Wortkette W : Ursache
- Mit **Bayes-Regel** :
$$P(W|O) = \frac{P(O|W) \cdot P(W)}{P(O)}$$
- Die wahrscheinlichste Wortkette:
$$\max_w P(W|O) = \max_w \frac{P(O|W) \cdot P(W)}{P(O)}$$
$$= \max_w P(O|W) \cdot P(W)$$
- $P(W)$ ist die globale, "a priori"-Wahrscheinlichkeit der Wortkette W .
- $P(O)$, die Wahrscheinlichkeit des Merkmalsmusters, wird nicht mehr benötigt.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Sprachmodelle

$$\max_w P(W|O) = \max_w P(O|W) \cdot P(W)$$

- Wie berechnen wir $P(W) = P(w_1 w_2 \dots w_n)$?
- Grundlage ist die Frequenz von Wortfolgen in Korpora.
- Sparse-Data-Problem: Ganze Sätze kommen viel zu selten vor.
- **Kettenregel** erlaubt die Reduktion von $P(w_1 w_2 \dots w_n)$ auf bedingte Wahrscheinlichkeiten:

$$P(w_1 w_2 \dots w_n)$$
$$= P(w_1) * P(w_2|w_1) * P(w_3|w_1 w_2) * \dots * P(w_n|w_1 w_2 \dots w_{n-1})$$

aber:

- $P(w_n|w_1 w_2 \dots w_{n-1})$: Sparse-Data-Problem ist nicht beseitigt!

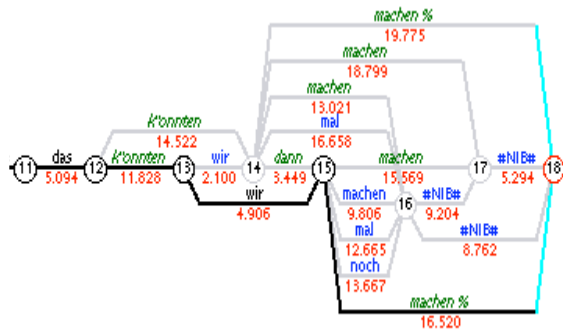
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

n-Gramme

- n-Gramm-Methode:
 - Wir approximieren die Wahrscheinlichkeit, dass ein Wort w im Kontext einer beliebig langen Wortfolge auftritt, durch die relative Häufigkeit, mit der es in einem auf n Wörter begrenzten Kontext auftritt ("Markov-Annahme")
 - Dabei wird das Wort selbst mitgezählt. N-Gramm-Wahrscheinlichkeit berücksichtigt also einen Vorkontext von $n-1$ Wörtern.
- Meistens wird mit Bigrammen und Trigrammen gearbeitet.
- Beispiel Bigramm-Approximation:
 - $P(w_n | w_1 w_2 \dots w_{n-1}) \approx P(w_n | w_{n-1})$
 - $P(w_1 w_2 \dots w_n) \approx P(w_1) * P(w_2 | w_1) * P(w_3 | w_2) * \dots * P(w_n | w_{n-1})$

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal Uds Computerlinguistik

Worthypothesengraph



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal Uds Computerlinguistik

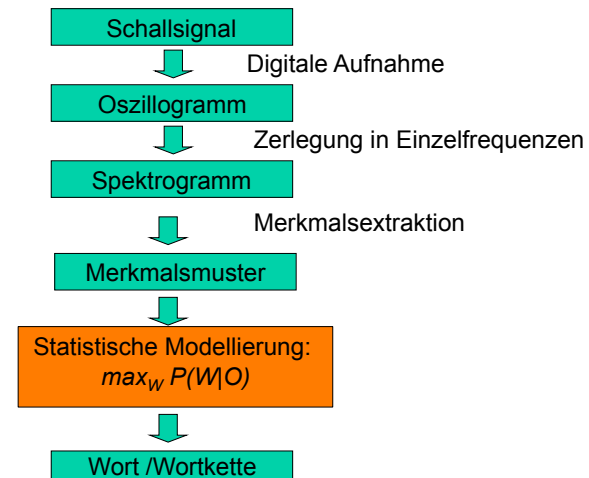
Akustische Modelle

$$\max_w P(W | O) = \max_w P(O | W) \cdot P(W)$$

- Aussprachewörterbuch, das für jedes Wort die phonetische Umschrift enthält
 - Genauer: Die Umschrift für alternative Aussprachen, die in einem gewichteten endlichen Automaten kodiert sind.
- Training von „Lautmodellen“ auf Datensammlungen für gesprochene Sprache: Aufnahmen von Sprachlauten mit ihrer phonetischen Kategorie/ Umschrift: Liefert die Wahrscheinlichkeit, mit der bestimmte Laute durch Merkmalsmuster realisiert werden.
- Für die statistische Zuordnung von Merkmalsmustern und Wörtern wird die HMM-Methode („Hidden Markov Models“) verwendet.

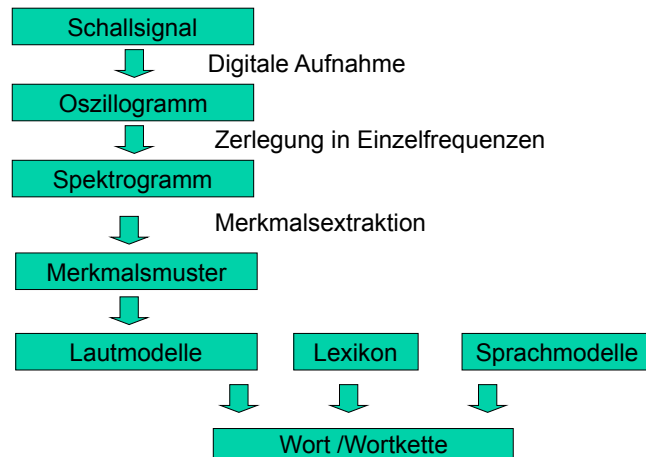
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal Uds Computerlinguistik

Spracherkennung: (Vereinfachtes) Schema



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal Uds Computerlinguistik

Spracherkennung: Schema



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Erkennerperformanz ist abhängig von:

- Sprechmodus: Einzelwort, kontinuierlich, spontan
- Sprecherbindung: abhängig, unabhängig, adaptiv
- Größe des Lexikons:
 - Einfache Sprachsteuerungssysteme: 100-200 Wortformen
 - Dialogsysteme: 500-1000 Wortformen (+ spezieller Wortschatz)
 - Diktiersysteme: ab 50000 Wortformen
- **Perplexität**: Maß für die Uniformität der Eingabe
beschränkte Domäne, gesteuerter Dialog: niedrige Perplexität
keine Domänenbeschränkung, freie Rede: hohe Perplexität
- Eingabequalität
- Verarbeitungszeit

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Stand der Spracherkennungstechnik

- Maß für die Erkennerperformanz: **Wortfehlerrate**
(wieviele Wörter der „besten Kette“ wurden falsch verstanden/gar nicht verstanden/hinzuphantasiert?)
- Wortfehlerrate hängt von der verfügbaren Verarbeitungszeit und verschiedenen externen Faktoren ab.
- Gängige Systeme analysieren in Echtzeit (Verarbeitungszeit \leq Sprechzeit) und sind in der Wortfehlerrate in einem akzeptablen Bereich .

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik