

Merkmalsextraktion

Stefan Scherer

Abteilung Neuroinformatik
Universität Ulm

5. Mai 2006

Inhalt

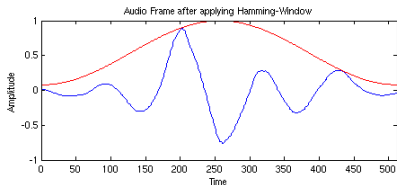
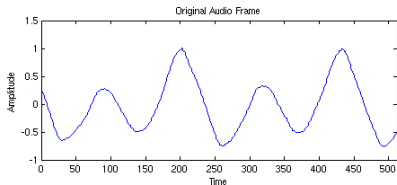
- 1 Generelle Vorgehensweise
- 2 LPC Merkmale
- 3 MFCC Merkmale
- 4 Pitch Merkmale

Generelle Vorgehensweise

- Aufteilen in Frames (32 ms)
- Hamming Fenster
- Features für einzelne Frames berechnen

Hamming Fenster

$$w(n) := 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad \forall n = 0, 1, \dots, N-1$$



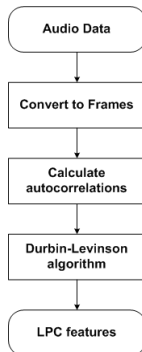
Gliederung

- 2 **LPC Merkmale**
 - Einführung
 - Mathematische Grundlagen
 - Durbin-Levinson Algorithmus

Linear Predictive Coding I

- *kommendes* Sprachsignal als Linearkombination der *Vergangenen*
- minimieren der Summe der quadrierten Differenzen zwischen dem eigentlichen Sprachsignal und den vorhergesehenen Koeffizienten
- Verwendungszwecke:
 - Datenkomprimierung
 - Sprechererkennung

Linear Predictive Coding II



Gliederung

- 2 **LPC Merkmale**
 - Einführung
 - **Mathematische Grundlagen**
 - Durbin-Levinson Algorithmus

Mathematische Grundlagen I

- $s_n = -\sum_{k=1}^p a_k s_{n-k} + Gu_n$
- $e_n = s_n - \hat{s}_n = s_n + \sum_{k=1}^p a_k s_{n-k}$
- $E = \sum_n e_n^2 = \left\{ \sum_n (s_n + \sum_{k=1}^p a_k s_{n-k}) \right\}^2$

Mathematische Grundlagen II

- $\sum_{k=1}^p a_k \sum_n s_{n-k} s_{n-i} = - \sum_n s_n s_{n-i} \quad \forall i = 1, \dots, p$
- $R_i = \sum_n s_n s_{n-i} \quad \forall i = 1, \dots, p$
- $\sum_{k=1}^p a_k R_{|i-k|} = R_i \quad 1 \leq i \leq p.$

Mathematische Grundlagen III

$$\begin{pmatrix} R_0 & R_1 & R_2 & \cdots & R_{p-1} \\ R_1 & R_0 & R_1 & \cdots & R_{p-2} \\ R_2 & R_1 & R_0 & \cdots & R_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{p-1} & R_{p-2} & R_{p-3} & \cdots & R_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} = - \begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_p \end{pmatrix}$$

Gliederung

- 2 LPC Merkmale
 - Einführung
 - Mathematische Grundlagen
 - Durbin-Levinson Algorithmus

Durbin-Levinson Algorithmus

- 1 $E(0) = R(0)$
- 2 $k_i = \frac{R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(|i-j|)}{E(i-1)}$
- 3 $\alpha_j^{(i)} = k_i$
- 4 $\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad \forall 1 \leq j \leq i-1$
- 5 $E(i) = (1 - k_i^2) E(i-1)$
- 6 Repeat steps 2 - 5 recursively $\forall i = 1, \dots, p$

Abbildung: The Durbin-Levinson algorithm.

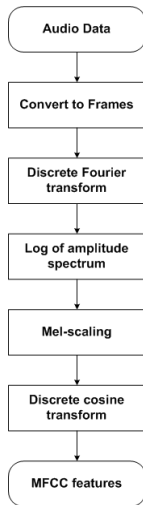
Gliederung

- 3 MFCC Merkmale
 - Grundlagen

Mel Frequency Cepstral Coefficients

- angelehnt an das menschliche Gehörsorgan
- Verdopplung des Mel Wertes entspricht Verdopplung der Wahrgenommenen Tonhöhe
- 0 - 500 Hz gleichzusetzen mit Mel, weiter nicht mehr linear
- Beispiele: 100 Hz = 100 Mel, 200 Hz = 200 Mel →
Verdopplung des Mel Wertes = Verdopplung der Tonhöhe;
1500 Hz = 1100 Mel, 10000 Hz = 2200 Mel
- Comirva Bibliothek

Mel Frequency Cepstral Coefficients



Gliederung

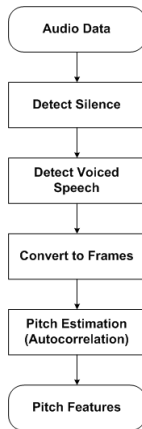
- 4 Pitch Merkmale
 - Grundlagen
 - Vorverarbeitung
 - Simple Inverse Filter Transform

Grundlagen I

- → Grundfrequenz
- Prosodische Information (Satzmelodie, Betonungen, Pausen etc. → sekundre Sprachmerkmale)
- ANSI Definition:

Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends primarily on the frequency of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus.

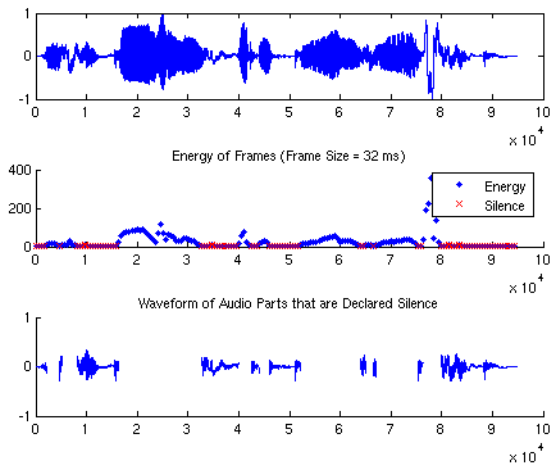
Grundlagen II



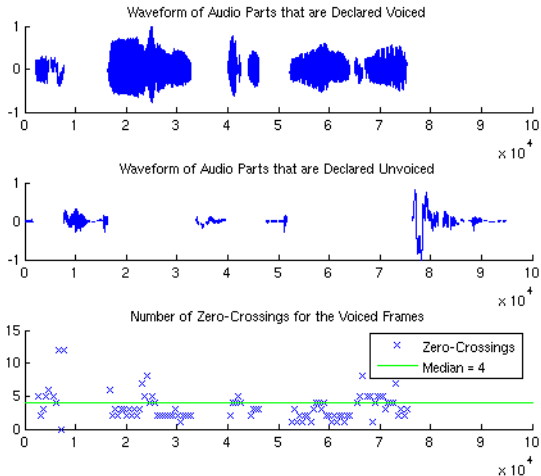
Gliederung

- 4 Pitch Merkmale
 - Grundlagen
 - **Vorverarbeitung**
 - Simple Inverse Filter Transform

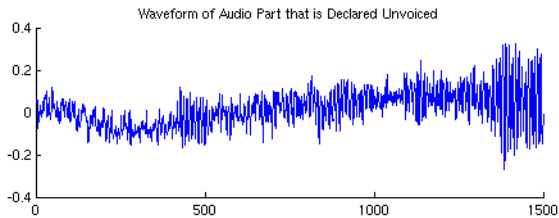
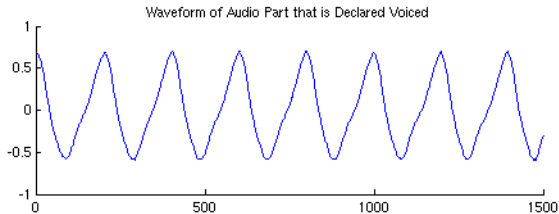
Vorverarbeitung I



Vorverarbeitung II



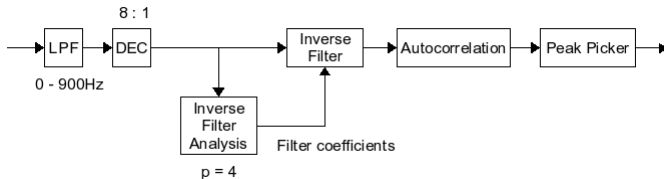
Vorverarbeitung III



Gliederung

- 4 Pitch Merkmale
 - Grundlagen
 - Vorverarbeitung
 - Simple Inverse Filter Transform

Simple Inverse Filter Transform



Autokorrelation

