

Berechnung von digitalen Signalen

Jonathan Harrington

Analog Signale



1. Digitalisierung: Abtasten, Quantisierung

Praat

Digitale Zeitsignale



2. Anwendung von einem Fenster

EMU-tkassp

Zeitsignal-Aufteilung



3. Anwendung von einem Parameter

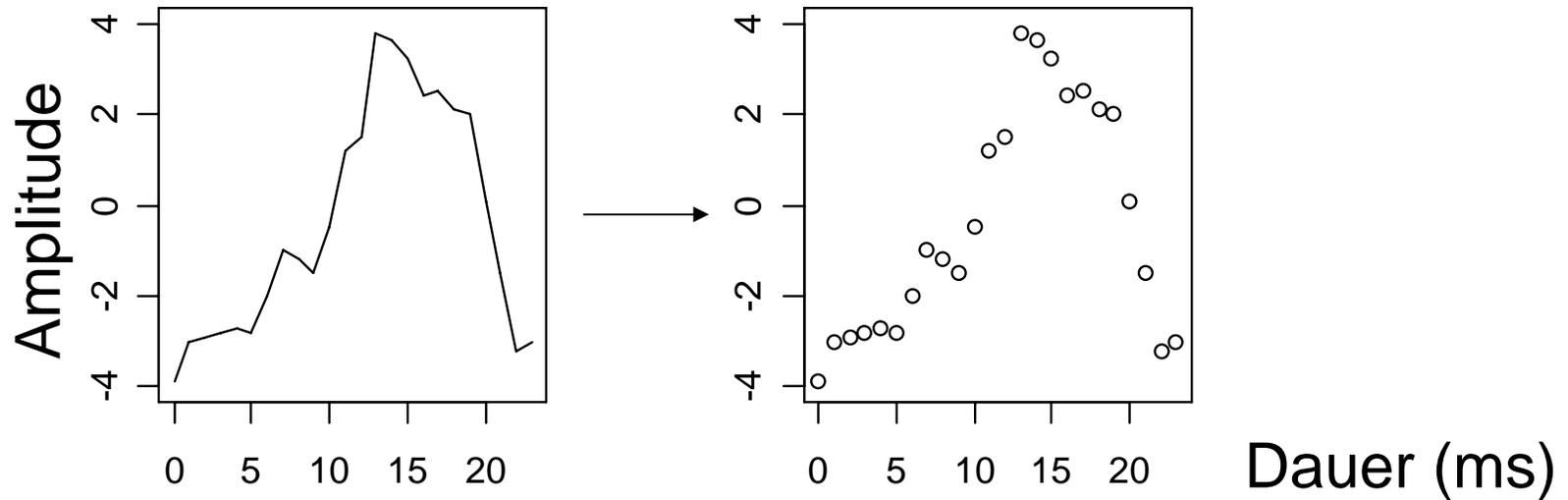
EMU-tkassp



Abgeleitete oder parametrisierte Signale (z.B. Formanten, Grundfrequenz)

1. Digitalisierung

Abtasten = die digitale Aufteilung der Zeitachse



Die Abtastperiode (T , in Sekunden): die (konstante) Dauer zwischen den digitalen Werten. Hier gibt es ein Punkt pro ms, die Abtastperiode = 0.001 s

Die Abtastfrequenz (f_s , Hz)

= $1/T$ (T in Sekunden)

= $1/0.001 = 1000$ Hz

Digitalisierung (fortgesetzt)

Quantisierung = Die digitale Aufteilung der Amplituden-Achse

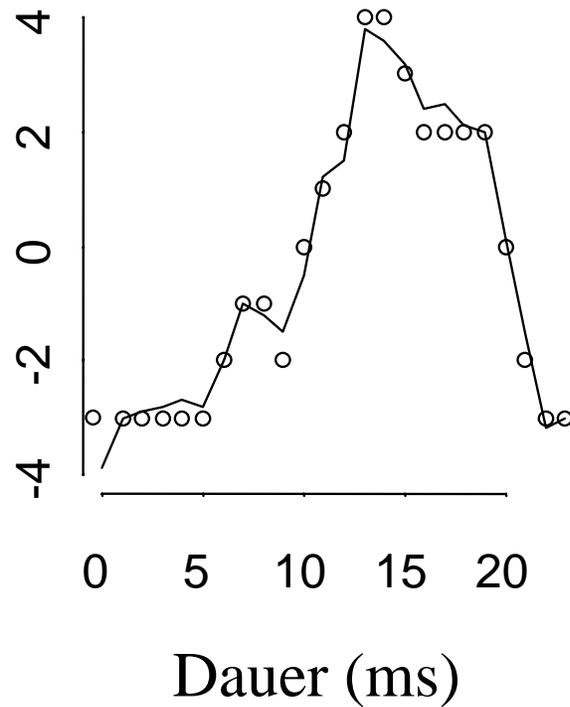
***n*-bit** Quantisierung bedeutet, dass die Amplitudenachse in 2^n Stufen aufgeteilt wird

| bit | Stufen | mögliche Werte der Amplitudenachse |
|-----|--------|------------------------------------|
| 1 | 2 | 0 1 |
| 2 | 4 | -1 0 1 2 |
| 3 | 8 | -3 -2 -1 0 1 2 3 4 |

Für die gesprochene Sprache: typischerweise mindestens 12-bit Quantisierung (-2047 bis 2048)

Quantisierung

3-bit Quantisierung, $f_s = 1000$ Hz



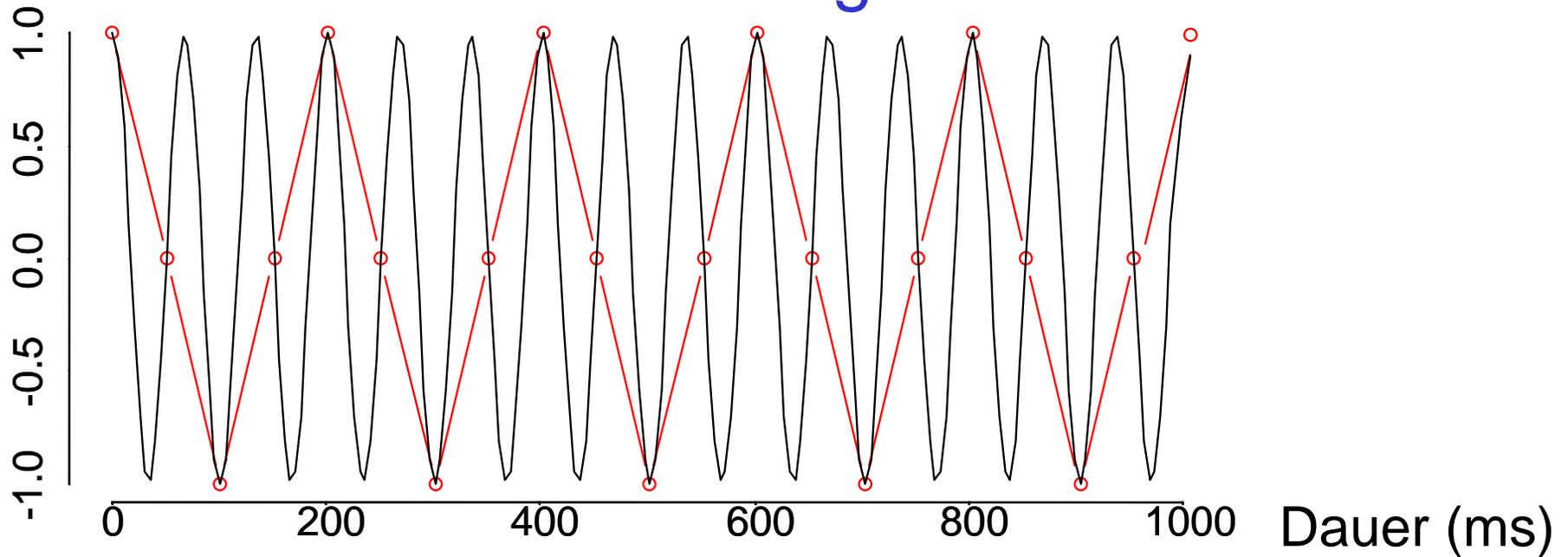
Digitalisierung: Nyquists Theorie (1928)

Im digitalen Signal erscheinen nur alle Frequenzen im Analogsignal bis zur Hälfte der Abtastfrequenz

z.B. benötigen wir eine Abtastrate von mindestens 30 Hz, um ein Analog-Signal von 15 Hz zu digitalisieren.

Eine **Faltung** entsteht, wenn f_s kleiner als 2 x die höchste Frequenz im Analog-Signal ist:

Faltung



Frequenz (Hz) des Analog-Sinusoids = 15 Hz

Abtastperiode, T , in Sekunden = $1/20$ s

Abtastrate, $f_s = 1/T$ Hz = 20 Hz

Frequenz des entsprechenden digitalen Sinusoids = 5 Hz

(Bei $f_s = 20$ Hz wird das 15 Hz Sinusoid auf ein 5 Hz Sinusoid gefaltet)

2. Anwendung von einem Fenster

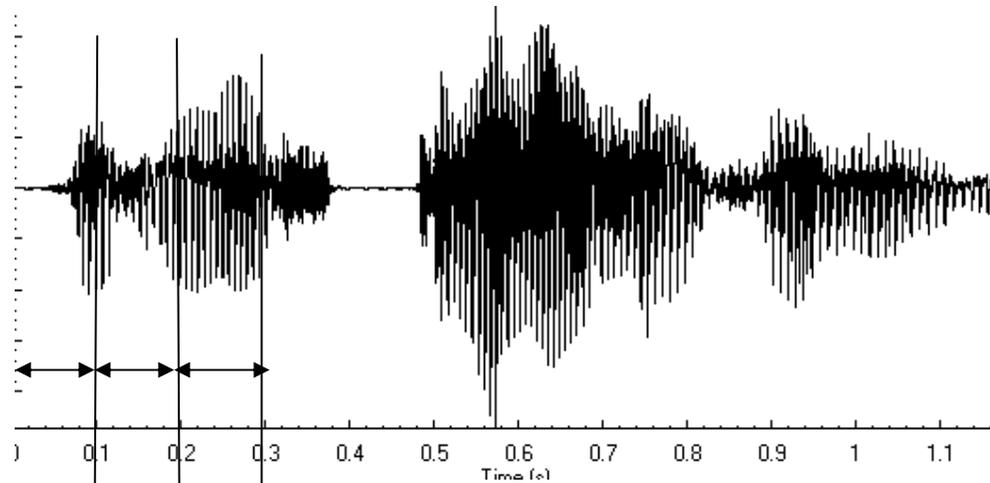
Ein Parameter (Formanten, F0, RMS) wird auf eine gewisse Dauer oder **Fenster** von Punkten angewendet

- **Fensterlänge** = die Dauer des Fensters

Für Sprachsignale, meistens 10 bis 30 ms = die Mindestdauer von phonetischen Ereignissen wie Plosiv-Verschlüssen, oder -Lösungen
- **Fensterverschiebung** = wie oft soll der Parameter berechnet werden? Meistens alle 5 oder 10 ms.
- Rechteckiges- oder Cosinus-Fenster?

Fensterlänge = 100 ms

Fensterverschiebung = 100 ms



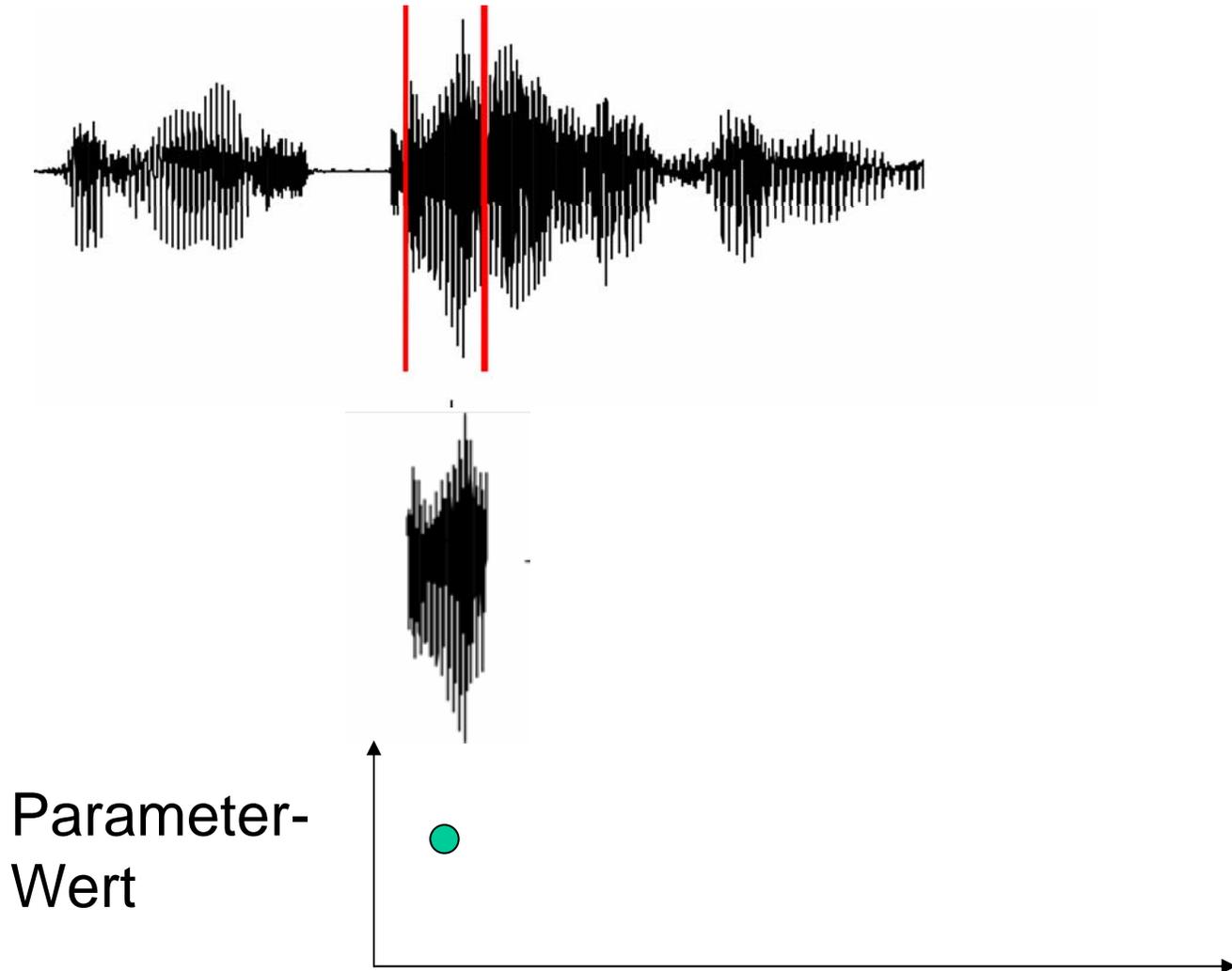
Parameter (Formanten, F0, RMS usw)

Parameter-
Wert



Rechteckiges (rectangular) Fenster

Das aufgeteilte Signal wird vor der Parameter-Berechnung **nicht** geändert

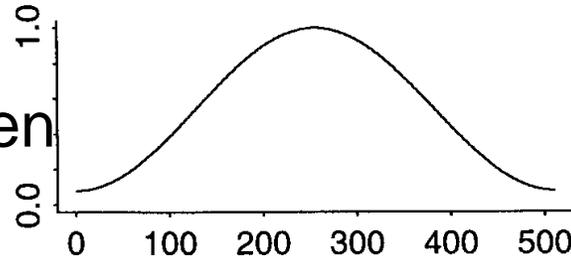


Ein rechteckiges Fenster hat 'scharfe' Kanten (das Signal beginnt und endet plötzlich), die ein Parameter manchmal beeinflussen.

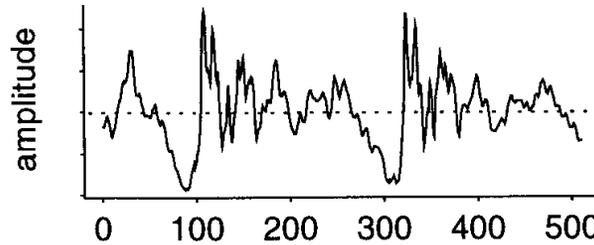
z.B. bei spektralen Berechnungen verursachen diese scharfen Kanten **hohe Frequenzen**, die mit dem Sprachsignal nichts zu tun haben.

Daher werden vor der Berechnung mancher Parameter die Kanten **geglättet** – durch die Multiplikation mit einem **Cosinus-Fenster**:

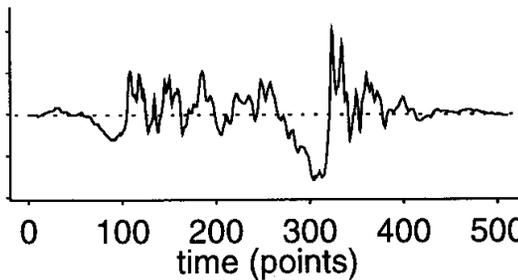
Cosinus-Fenster
(Werte liegen zwischen
0 und 1)



Signal



Signal mit dem
Cosinus-Fenster
multipliziert



Parameter-
Wert



Ein Cosinus-Fenster wird im Allgemeinen bei **Frequenz**berechnungen angewendet

Es gibt mehrere Sorten von Cosinus-Fenstern (Hamming, Hanning, Blackman...), die fast die selbe Wirkung haben

3. Parameter

(werden mit EMU-tkassp berechnet)

The image shows a screenshot of a software interface with a list of parameters. Each parameter has a checkbox to its right. Red arrows point from German labels on the left to the corresponding checkboxes. The parameters listed are:

- Use acfana
- Use f0ana
- Use forest
- Use rfcana
- Use rmsana
- Use spectrum
- Use zcrana

The German labels and their corresponding parameters are:

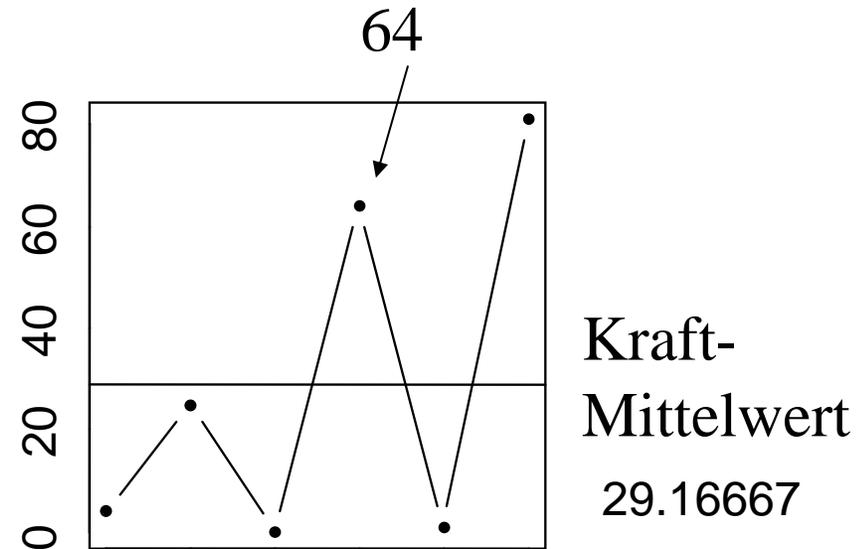
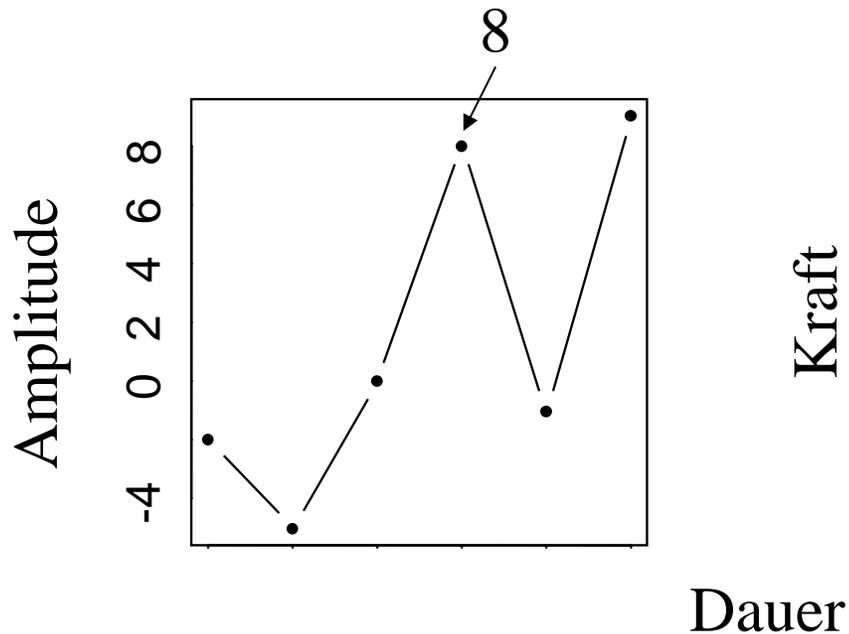
- Grundfrequenz (points to Use f0ana)
- Formanten (points to Use forest)
- dB-RMS (Lautstärke) (points to Use rmsana)
- ZCR (zero-crossing-rate) (points to Use zcrana)

(unterscheidet Frikative von Sonoranten)

dB-RMS (Lautstärke)

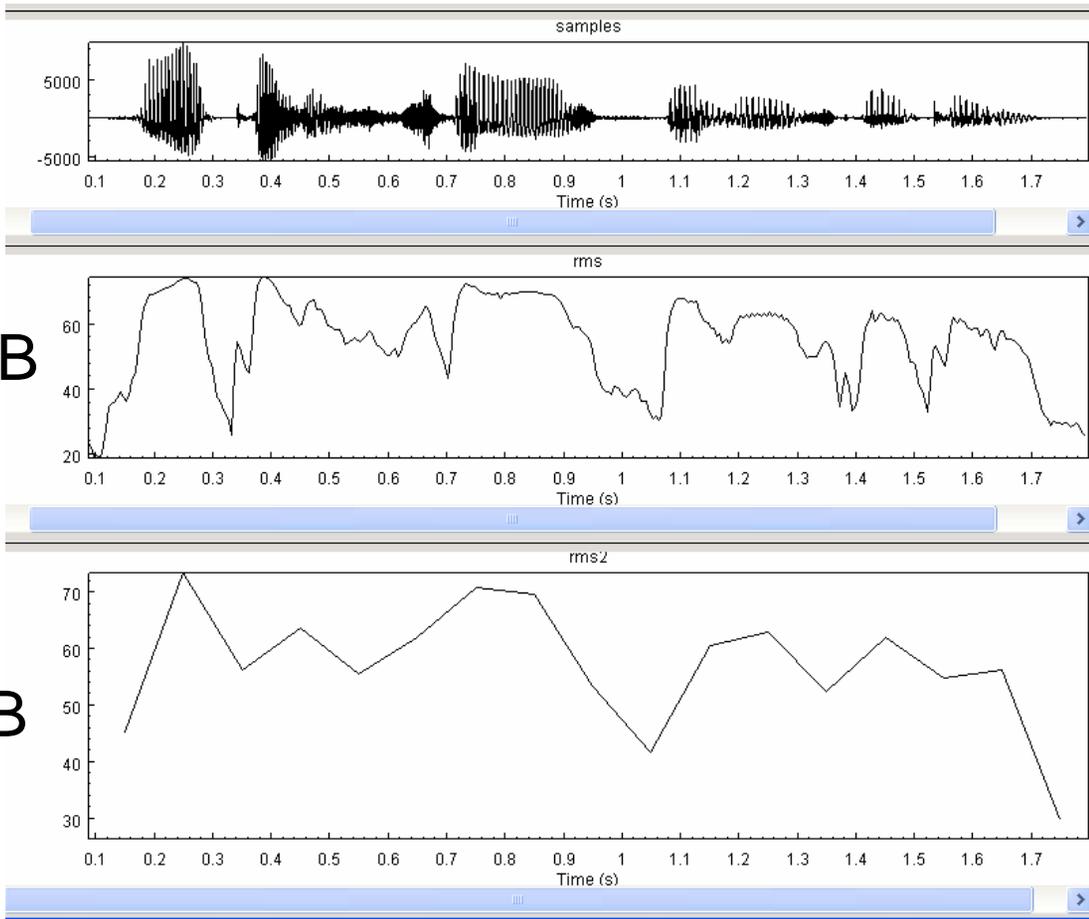
Lautstärke ist ca. im Verhältnis zum **Logarithmus vom Kraft-Mittelwert**

$$\text{Kraft} = \text{Amplitude}^2$$



$$\text{Lautstärke (dB-RMS)} = \log_{10}(29.16667) = 1.46 \text{ Bel} = 14.6 \text{ dB}$$

Parameter: Lautstärke (dB-RMS)



Fensterlänge 25 ms
-Verschiebung 5 ms

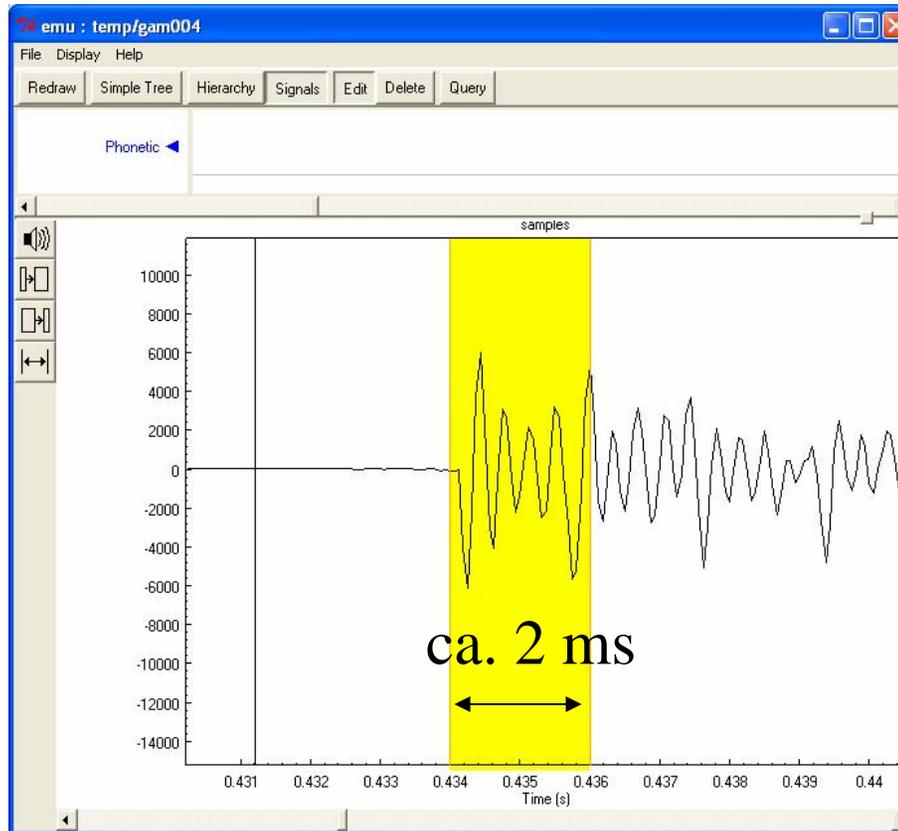
Fensterlänge 100 ms
-Verschiebung 100 ms

ZCR (zero-crossing-rate) oder Nulldurchgangsrate

ZCR (in Hz)

- die Frequenz, mit der das Signal die x-Achse (die Zeit-Achse) schneidet pro Sekunde, dividiert durch 2.
- Hohe ZCR-Werte: Laute mit Turbulenz (Frikative, Verschlusslösungen)

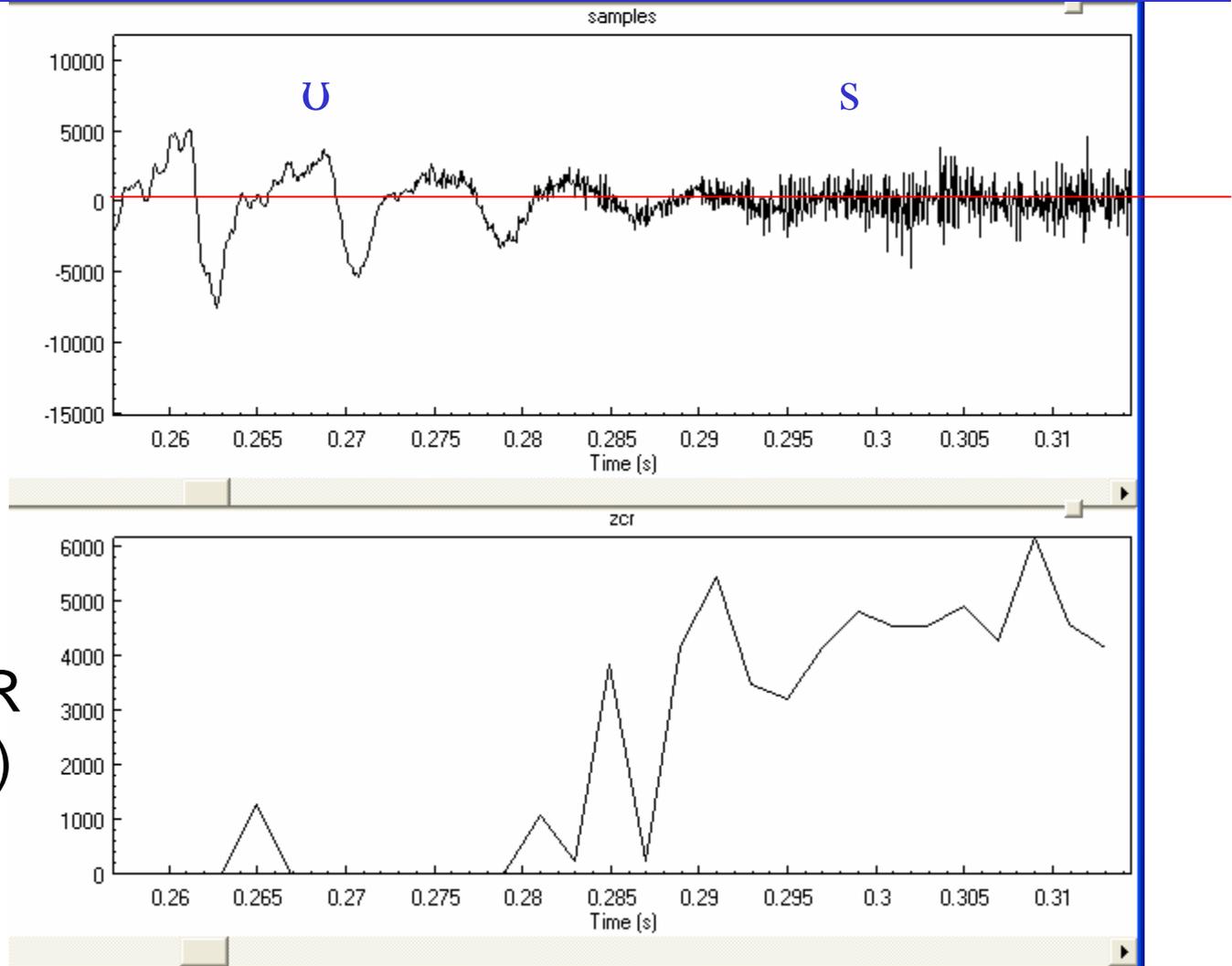
ZCR (fortgesetzt)



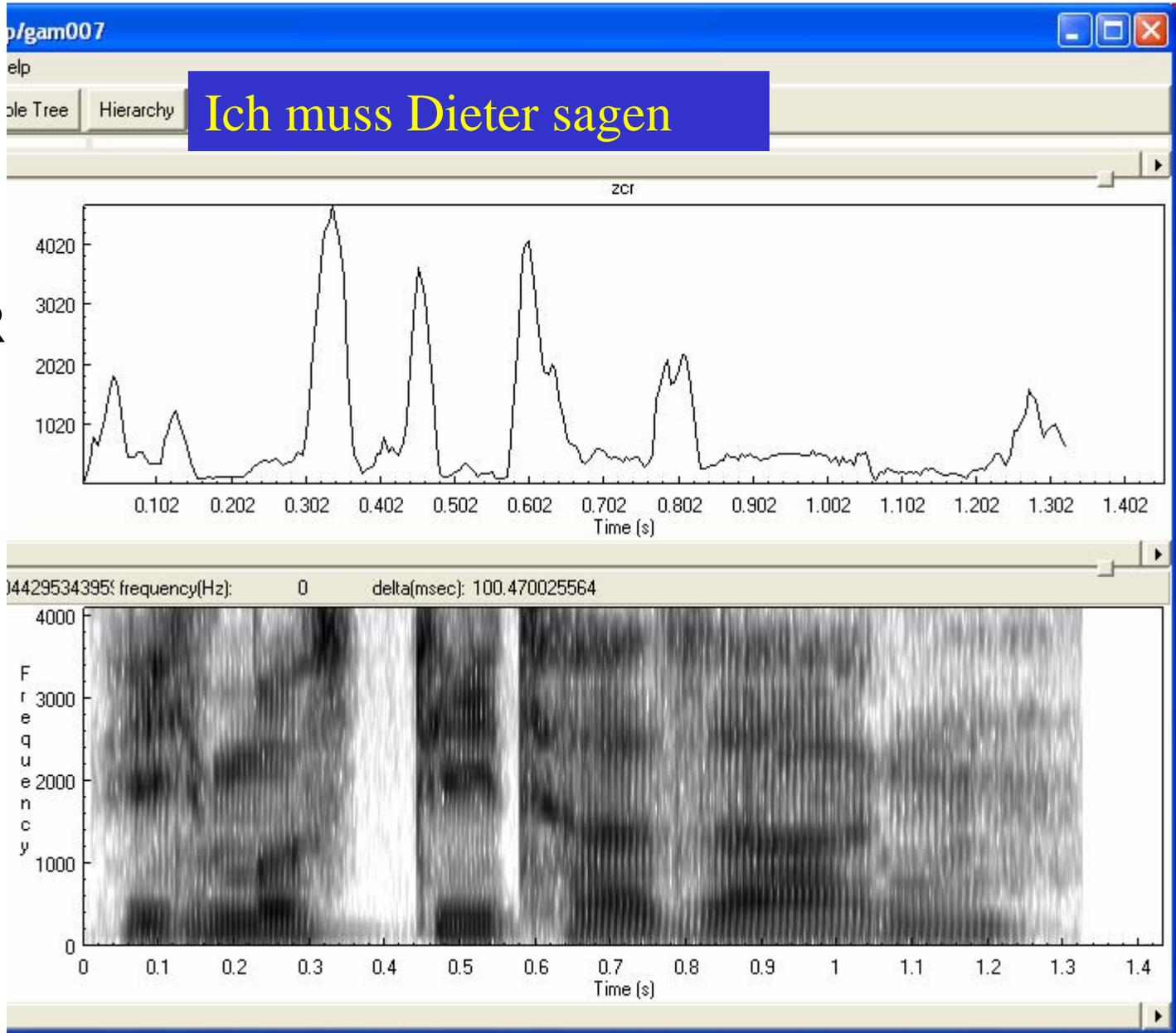
Das Signal schneidet die x-Achse ca. 9 Mal in 2 ms
= $9/2$ Mal pro ms
= $9000/2 = 4500$ Mal pro Sekunde

ZCR ist daher ca. $4500/2$ ca. 2250 Hz

Signale mit hoher Frequenz (wie Frikative und die Lösungen von Plosiven) schneiden die x-Achse mit hoher Frequenz und haben daher hohe ZCR-Werte

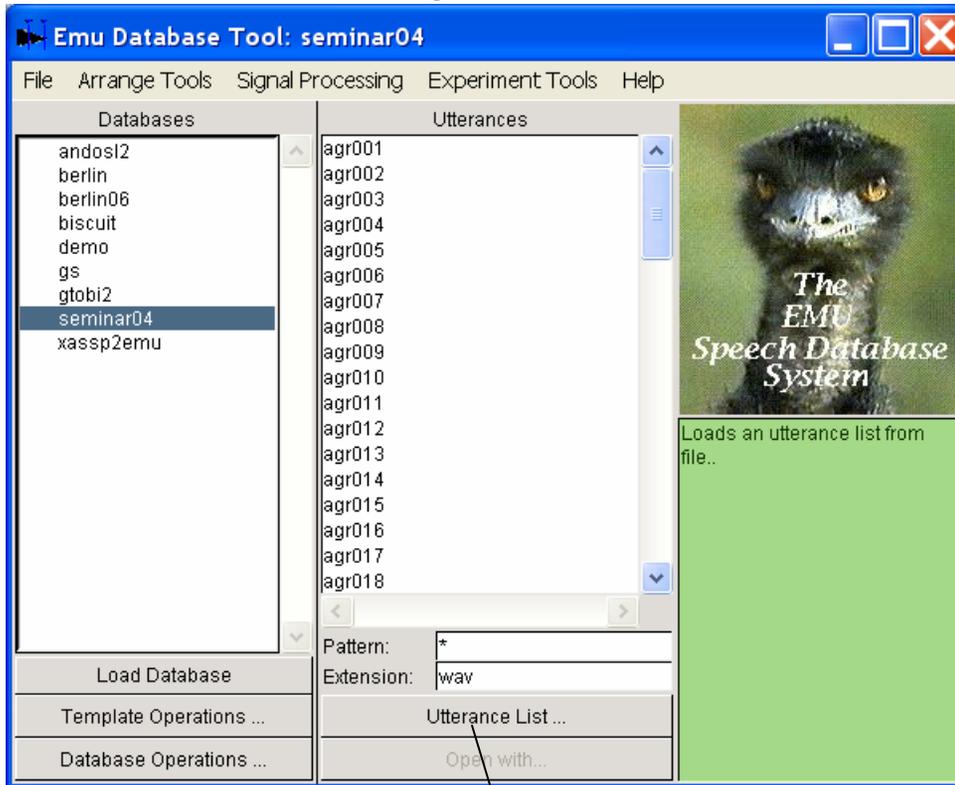


ZCR
(Hz)



Formanten und f_0 berechnen in den ersten 10
Äußerungen der **seminar04c** Sprachdatenbank.

Formanten und f0 berechnen in den ersten 10 Äußerungen der **seminar04c** Sprachdatenbank.



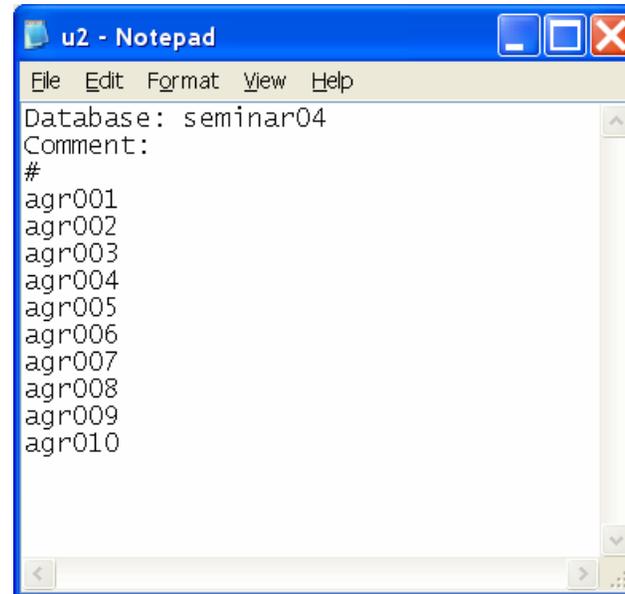
1. Utterance list als eine Textdatei (Extension .txt) speichern.

Save Utterance list

2. Die Textdatei manipulieren, um die Äußerungen auszusuchen, auf die wir die Signalverarbeitung anwenden wollen.

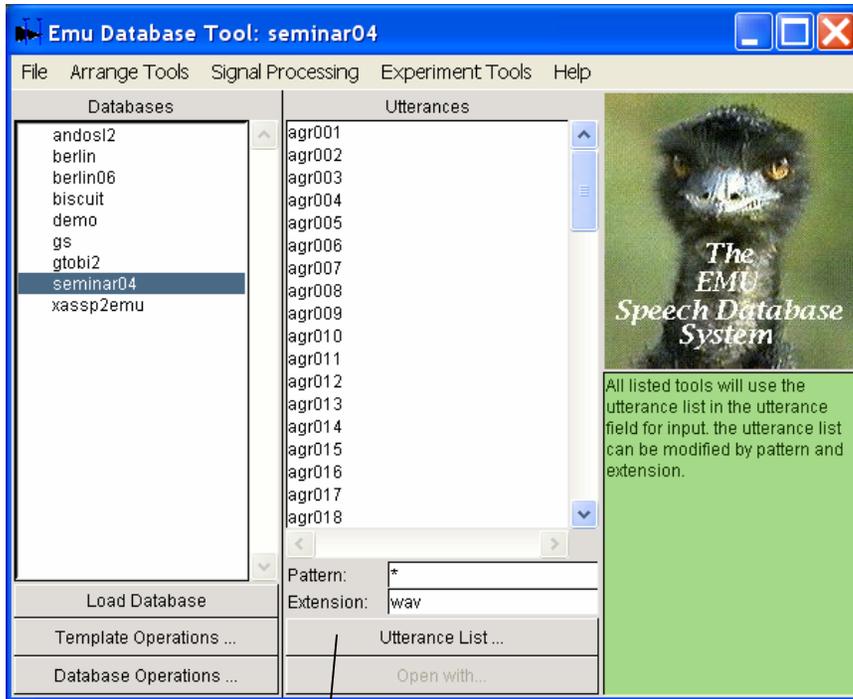


```
u3 - Notepad
File Edit Format View Help
Database: seminar04
Comment:
#
agr001
agr002
agr003
agr004
agr005
agr006
agr007
agr008
agr009
agr010
agr011
agr012
agr013
agr014
agr015
agr016
agr017
agr018
agr019
agr020
agr021
agr022
agr023
agr024
agr025
```

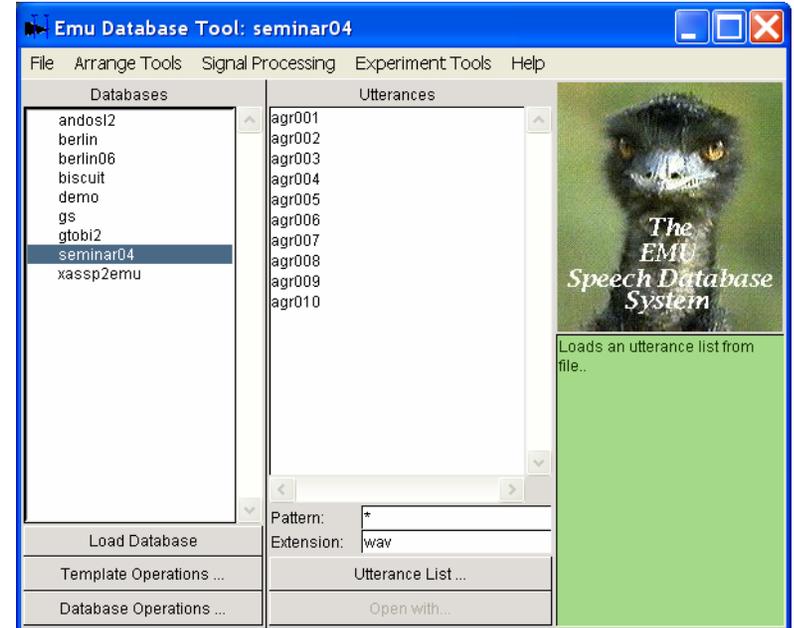


```
u2 - Notepad
File Edit Format View Help
Database: seminar04
Comment:
#
agr001
agr002
agr003
agr004
agr005
agr006
agr007
agr008
agr009
agr010
```

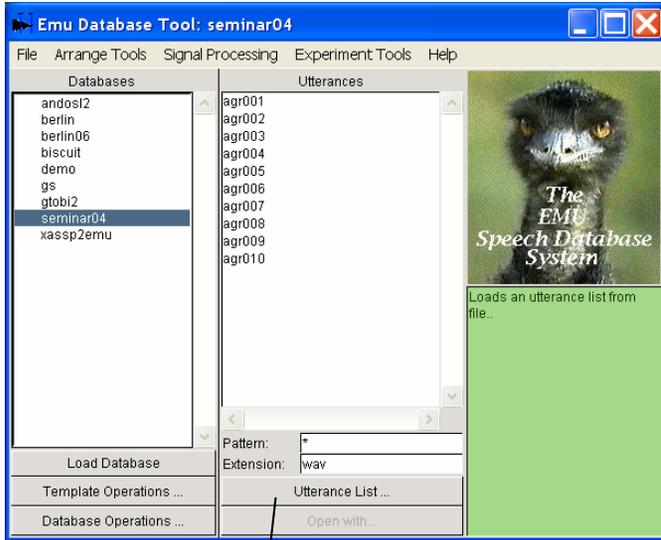
3. Die manipulierte Textdatei einlesen



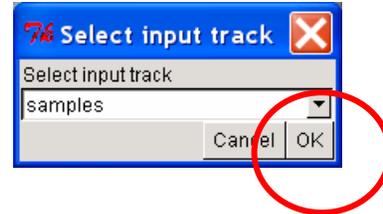
Load utterance list



4. tkassp starten

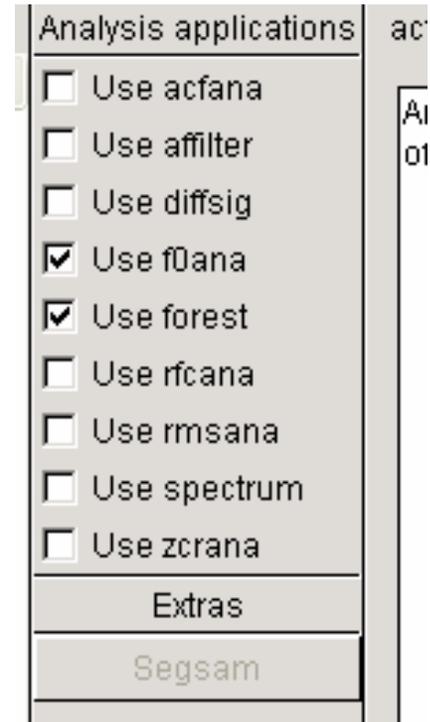
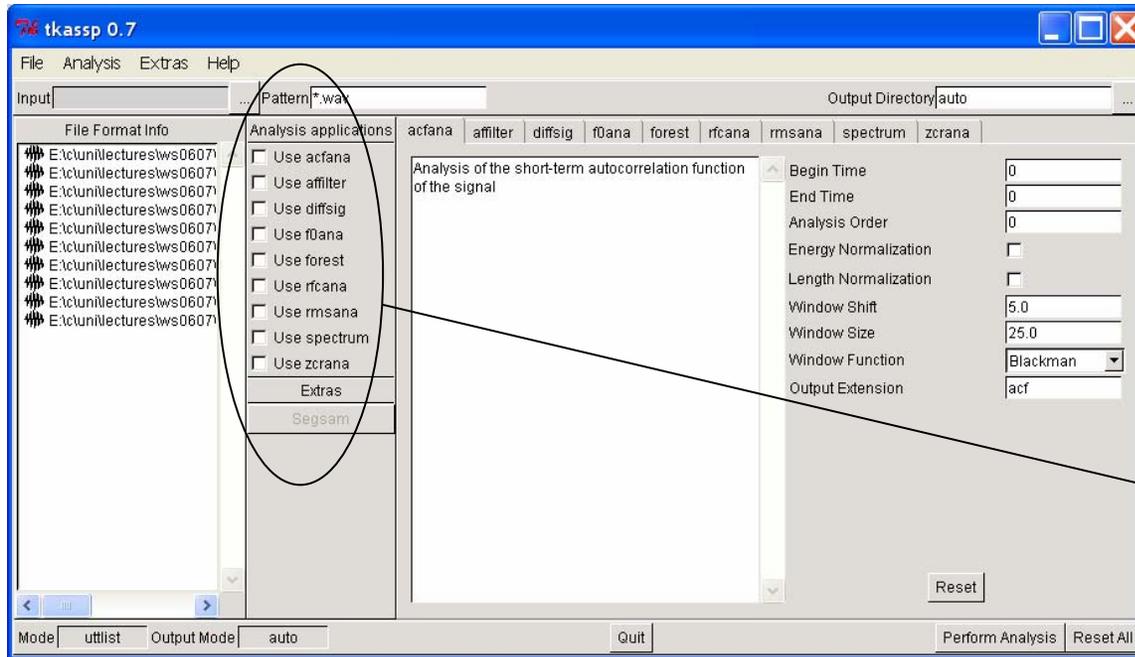


To tkassp



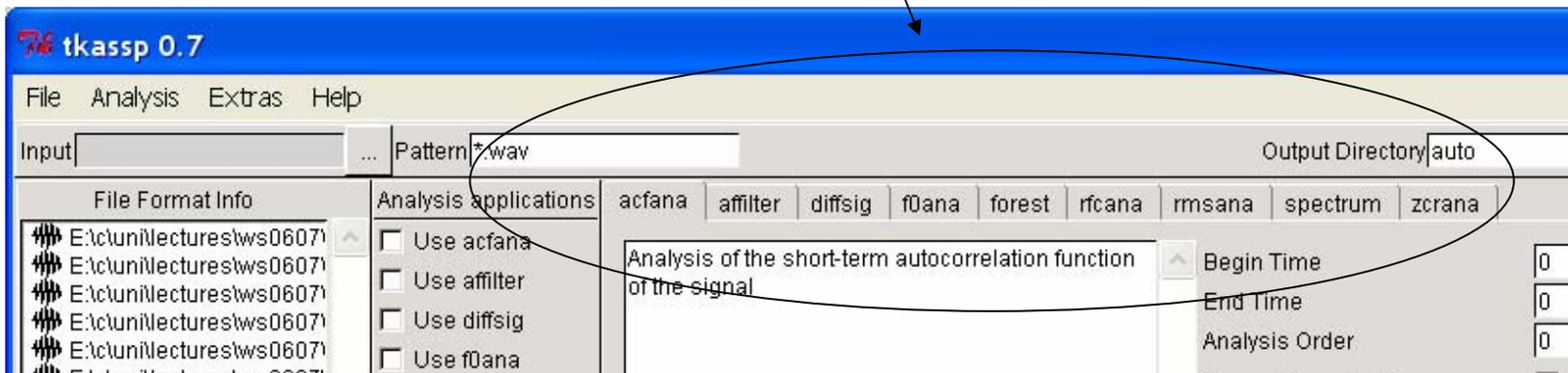
5. Parameter auswählen

(Für Formanten und f0).

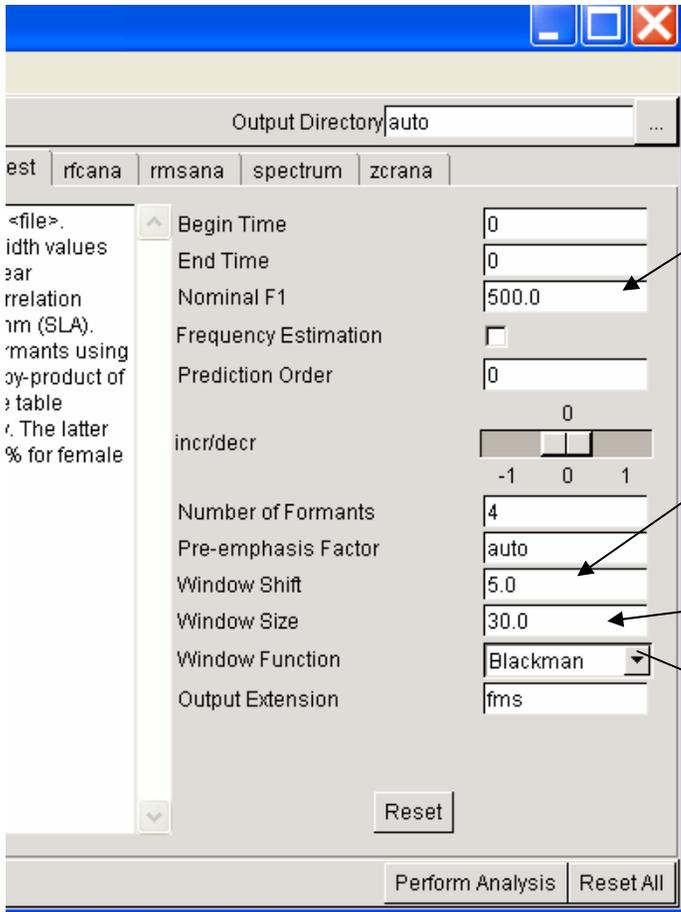


6. Parameterwerte festlegen

Mit dieser Registerkarte lassen sich die Parameter der jeweiligen Signalberechnung ändern



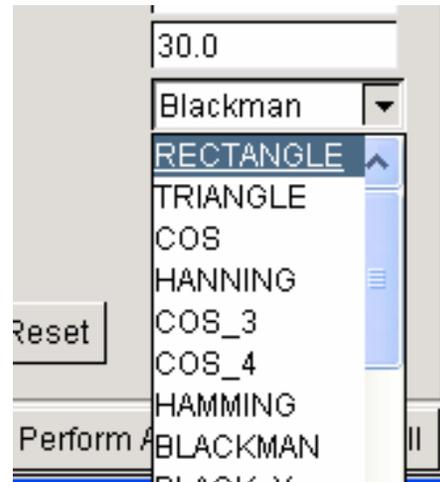
zB forest (Formant-Berechnungen)



Durchschnittlicher F1 (bei weiblichen Stimmen auf 600 – 700 Hz setzen)

Fensterverschiebung

Fensterlänge



Fenstersorte

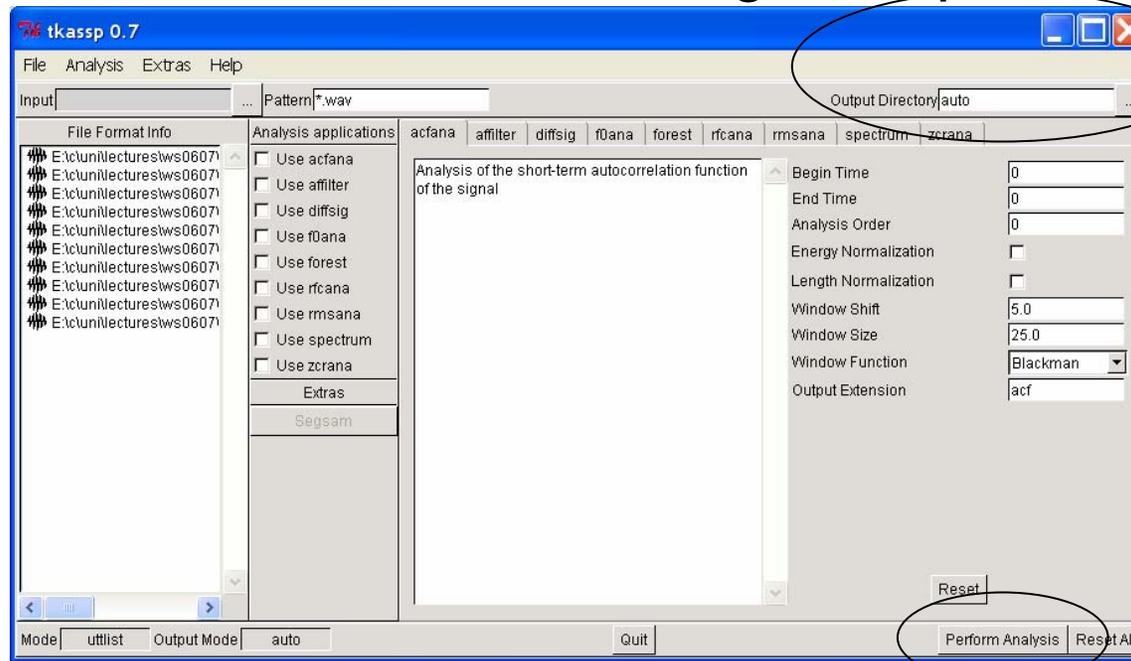
| | |
|---------------------|----------|
| Number of Formants | 4 |
| Pre-emphasis Factor | auto |
| Window Shift | 5.0 |
| Window Size | 30.0 |
| Window Function | Blackman |
| Output Extension | fms |

Extension

Für jede wav Datei, auf die die Formantberechnungen angewendet wird, wird eine entsprechende Datei mit Extension **fms** erzeugt, die die Formanten enthält (zB agr001.fms für agr001.wav)

7. Verzeichnis für die erzeugten Signale wählen

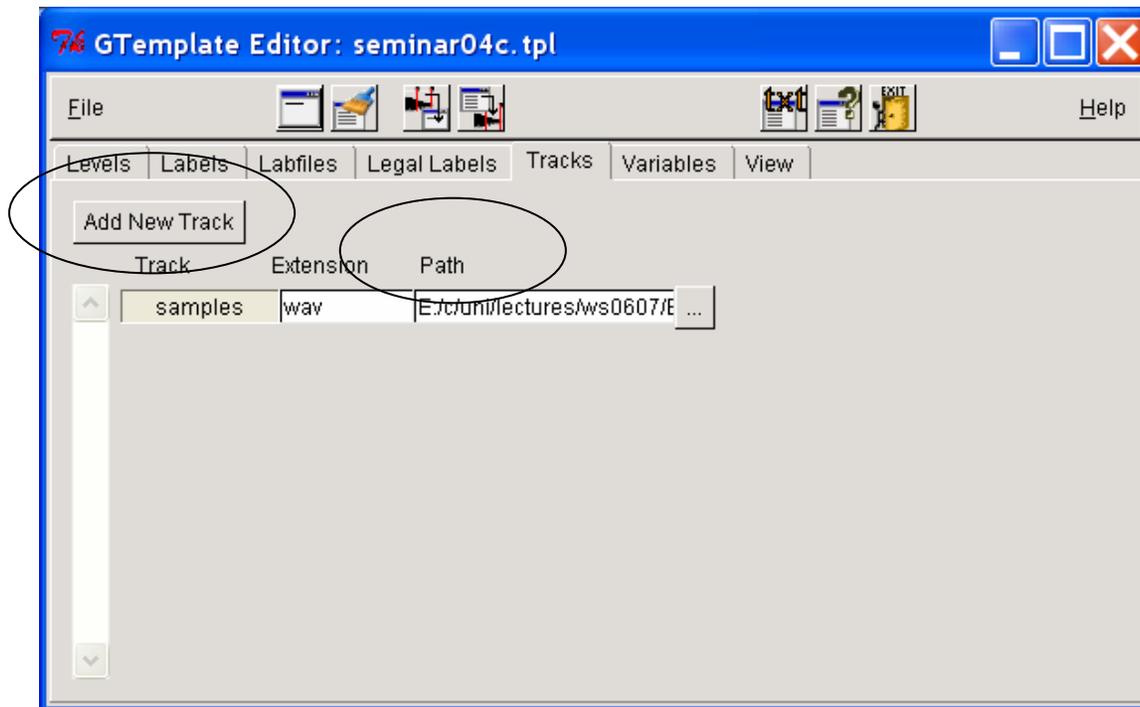
Verzeichnis wählen: wo möchte ich die Signale speichern?



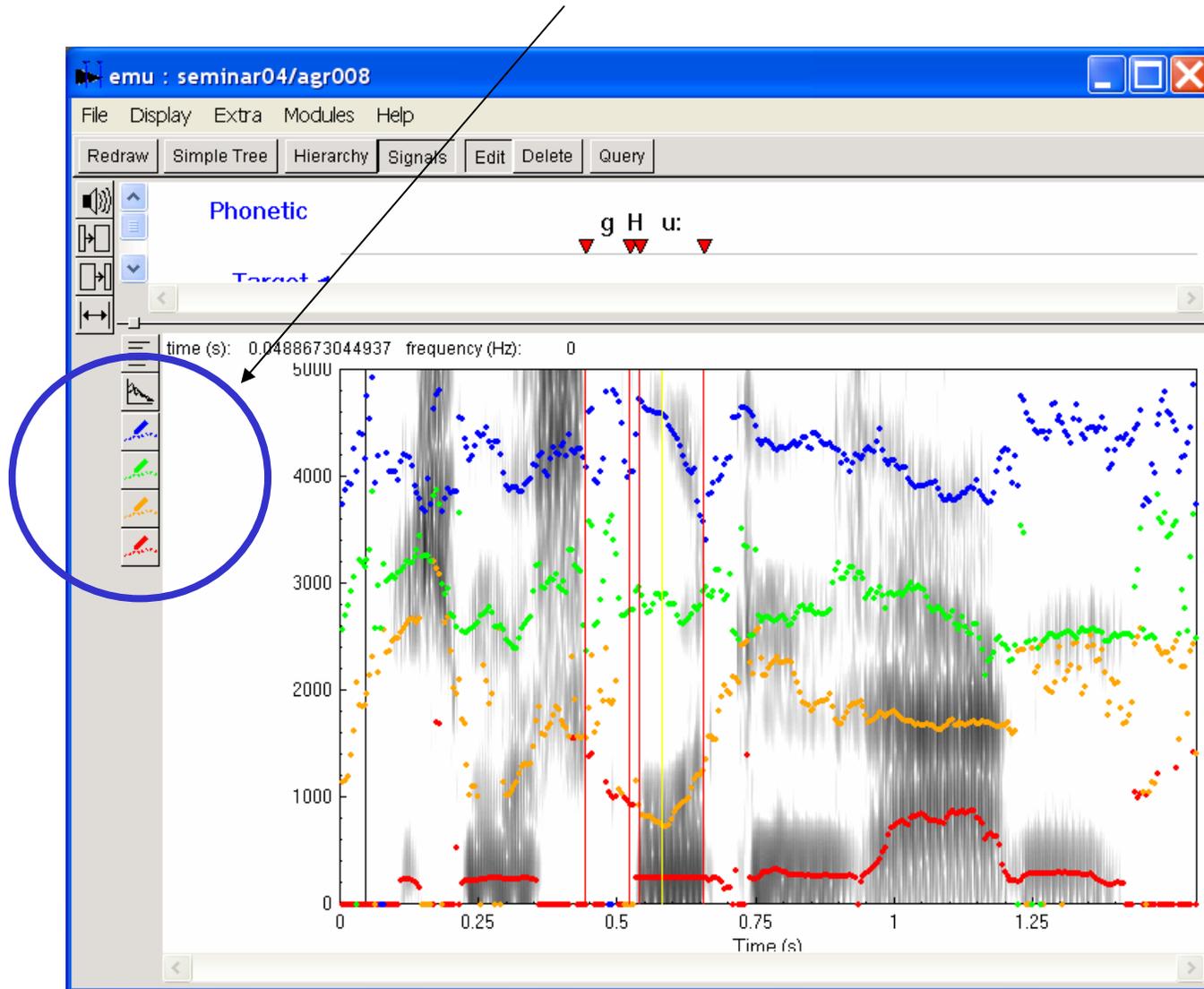
8. Signalberechnungen starten

9. Template Datei modifizieren

(Damit Emu weiß, wo sich die neuen Signale befinden)



10. EMU aufrufen, Formanten ggf. korrigieren.



Bitte jetzt zur Signalverarbeitung Übung in

<http://www.phonetik.uni-muenchen.de/~jmh/lehre/emur.htm>