

# Folgen einer Gesangsstimme in Echtzeit

Kann man einem Computer beibringen, wie ein Musiker zu hören?

Filip Korzeniowski

Johannes Kepler Universität, Department of Computational Perception

22.12.2011

# Problemstellung

- ▶ Live-Darbietung eines Musikstückes (Probe, Konzert, ...)
- ▶ Stimme wird mit Mikrofon aufgenommen
- ▶ Computer soll errechnen, an welcher Stelle im Notentext sich die Sängerin / der Sänger im Augenblick befinden



# Wozu?

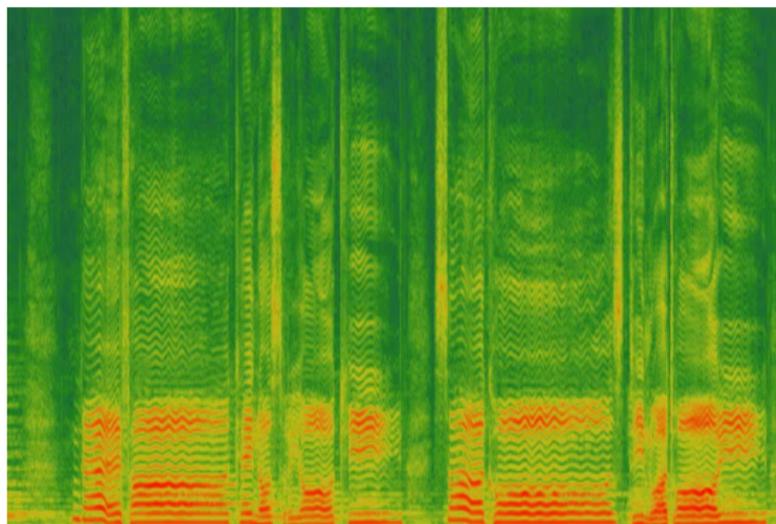
- ▶ Synchronisation für visualisierungen bei Live-Aufführungen
- ▶ Stilanalyse
- ▶ Virtuelles Orchester / virtuelle Band
- ▶ ...

Musik für menschliche Zuhörer:



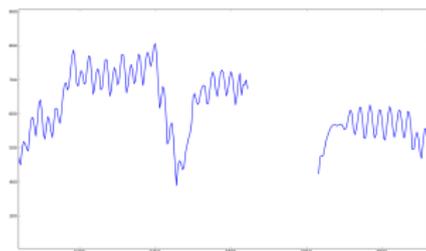
Musik für einen Computer:

-768, -674, -575, -460, -388, -300, -205, -122, -28, 76, 170, 308, 453, 601, 758, 885, 1003, 1107, ...

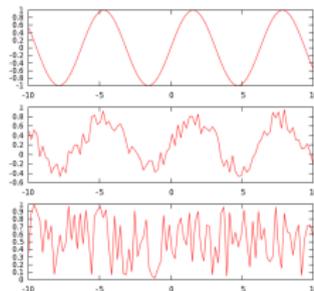


# Informationen aus Audio berechnen

## Tonhöhe



## Qualität



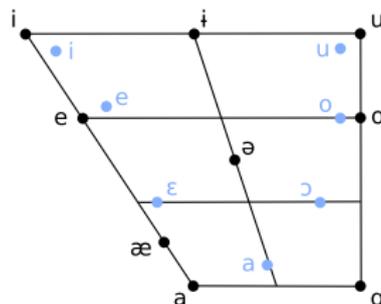
## Lautstärke

$$L_p = 20 \log_{10} \left( \frac{p_{rms}}{p_{ref}} \right)$$

$$p_{rms} = \sqrt{\frac{1}{N_{win}} \sum_{i=1}^{N_{win}} x_i^2}$$



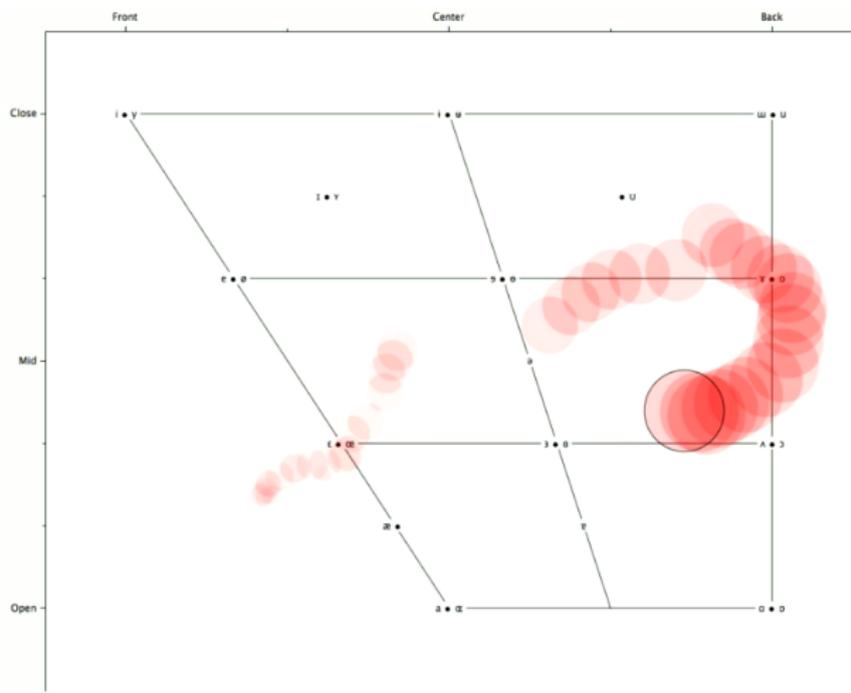
## Vokalinformation



# Ergebnis - Gesangsverfolgung

The image displays a musical score for a vocal line in G major, 2/4 time. The lyrics are: "O Lu na lu cen te, di Fe bo so rel la che can di da-e bel la ri splen di las su deh fa che-i no - stri oc - chi s'ac - co sti -no-ai tuo - i e soo - pri - ti a no - i che co - sa sei tu." The score is divided into five systems, with line numbers 11, 15, 19, 23, and 27. A blue highlight is placed under the word "che" in the second system. Below the score is a blue waveform visualization of the audio signal, with vertical red lines marking the time positions of the words: "bo so", "rel la", "che", "can", "di", "da-e", and "bel". The waveform shows the amplitude of the voice over time, with the highlighted word "che" corresponding to a specific peak in the signal.

# Ergebnis - Vokalerkennung





*That's all Folks!*