Semantic Analysis to Help Editing Recorded Speech

Jörn Loviscach
Music vs. speech

- Polyphony
- Rhythm
- Pitch
- Timbre
Applications

• Speech detection
• Speech recognition
  (a science in its own right)
• Speaker identification; diarization
• Emotion recognition
• Visualization for editing

music noise speech
Hello

happy neutral
no issues with robustness!
Examples

*Featuring:*
The MathWorks MATLAB®
and
Olivier Lartillot’s MIRtoolbox

(long URL; just search for it)

Example code online (URL at the end)
Vowel or consonant?

• Indicate the positions of sonorants
• Acoustic features:
  – Inharmonicity
  – Spectral rolloff
Vowels as colors

- Mapping formants to RGB colors
- MFCCs and PCA

Related to:
Detecting the annoying “um”

• MFCCs
• Template matching
Say again?

- Show several takes for a sentence
- MFCCs, dissimilarity matrix, Mahalanobis metric

Pitch

• Cepstrum

• What could that be good for?
Integration into an application

Loudness + timbre + speech recognition

Where to go from here

Models for: language, person, gender, accent, emotion, ...
Slides and code:

www.j3L7h.de/talks