Semantic Analysis to Help Editing Recorded Speech

Jörn Loviscach

FH Bielefeld
University of
Applied Sciences

Music vs. speech

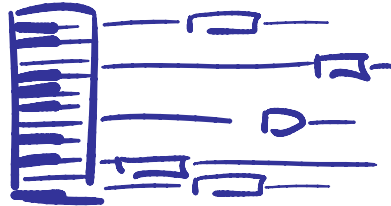- Polyphony

- Rhythm

- Pitch

- Timbre

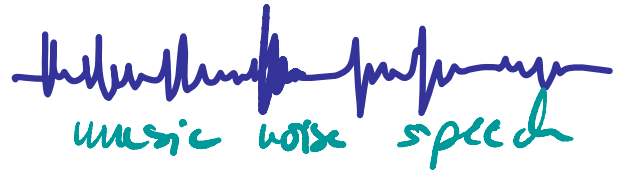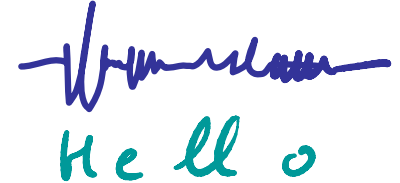# Applications

- Speech detection

- Speech recognition (a science in its own right)

- Speaker identification; diarization

- Emotion recognition

- Visualization for editing

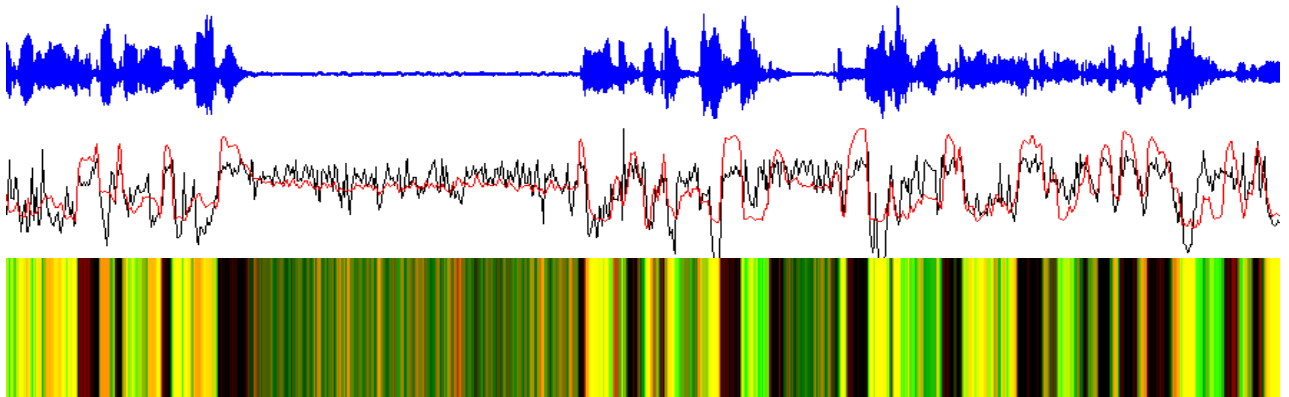# **Examples**

Featuring:

The MathWorks MATLAB®

and

Olivier Lartillot's MIRtoolbox

(long URL; just search for it)

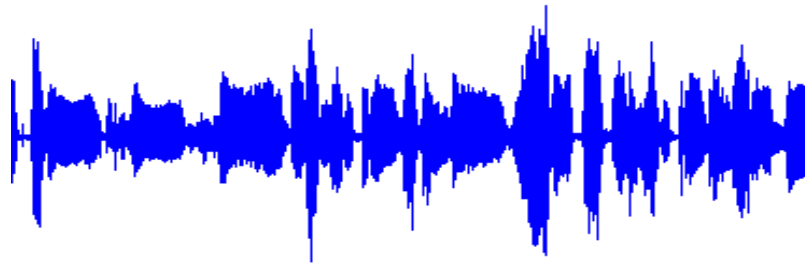Example code online (URL at the end)

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:

- "TimbreGrams" P. R. Cook and G. Tzanetakis.
  Audio information retrieval (AIR) tools. Proc. ISMIR 2000.
- "Comparisonics" S. V. Rice. Frequency-based coloring
  of the waveform display [...]. 119th AES Conv., 2005. Paper 6530.
- A. Mason, M. Evans, A. Sheikh. Music information retrieval in broad-
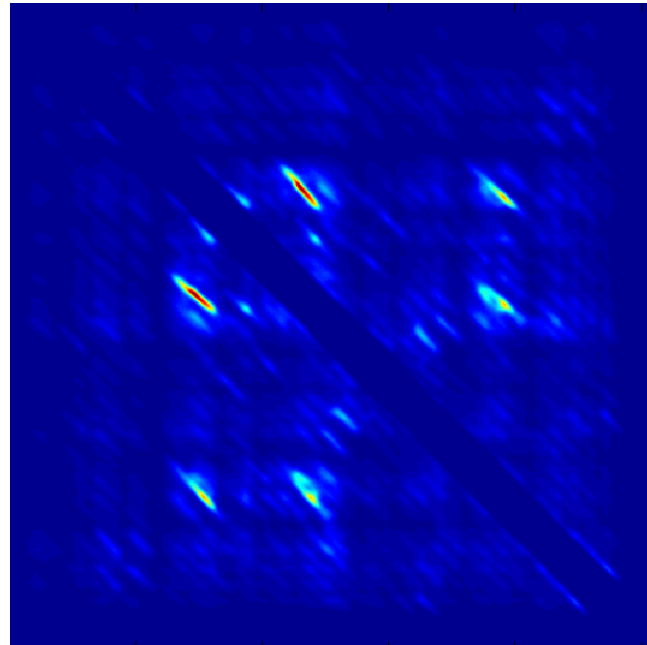 casting: Some visual applications. 123rd AES Conv., 2007. Paper 7238.

Detecting the annoying "um"
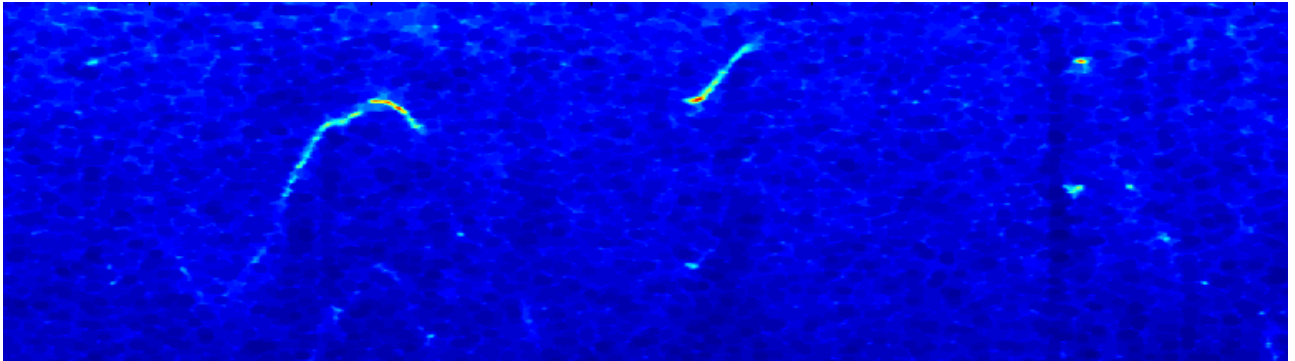
- MFCCs

- Template matching
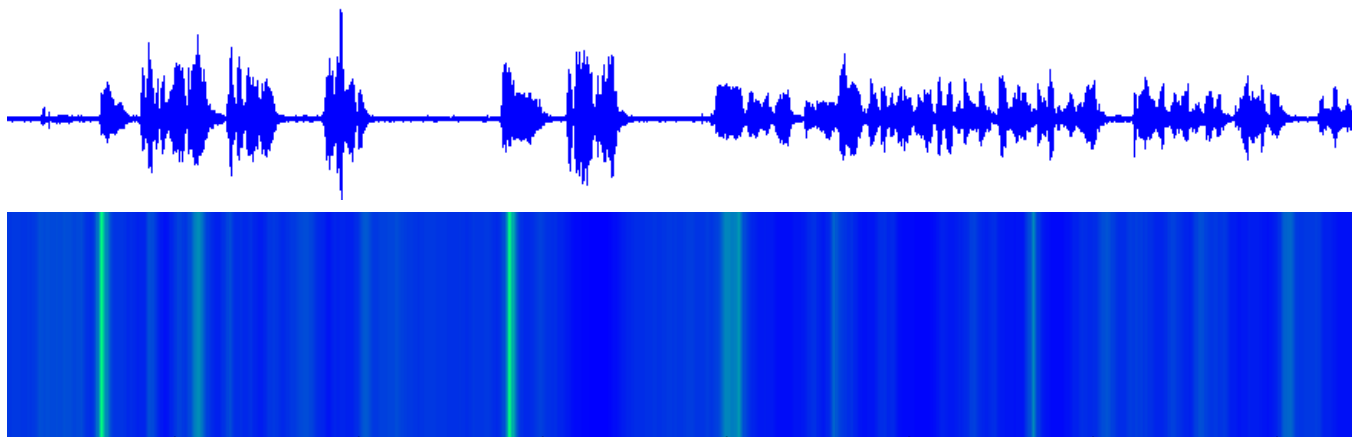
Say again?

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

Related to L. Couvreur et al. Audio Thumbnailing. Numediart QPSR 1(2), 2008.
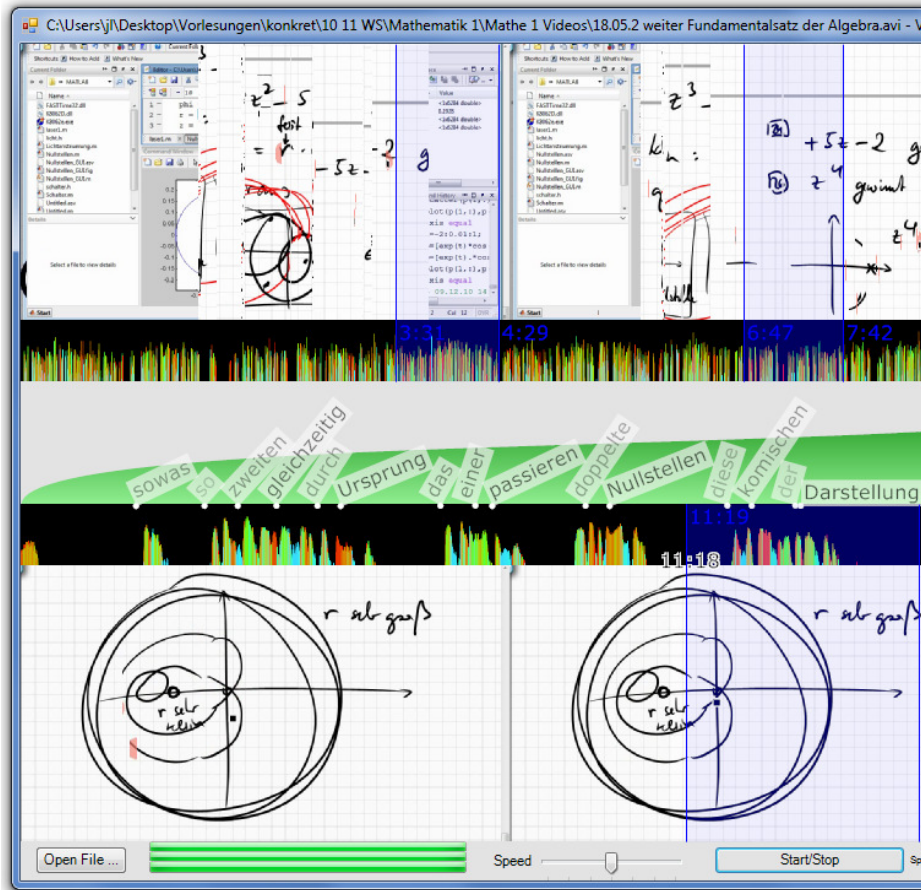
Pitch

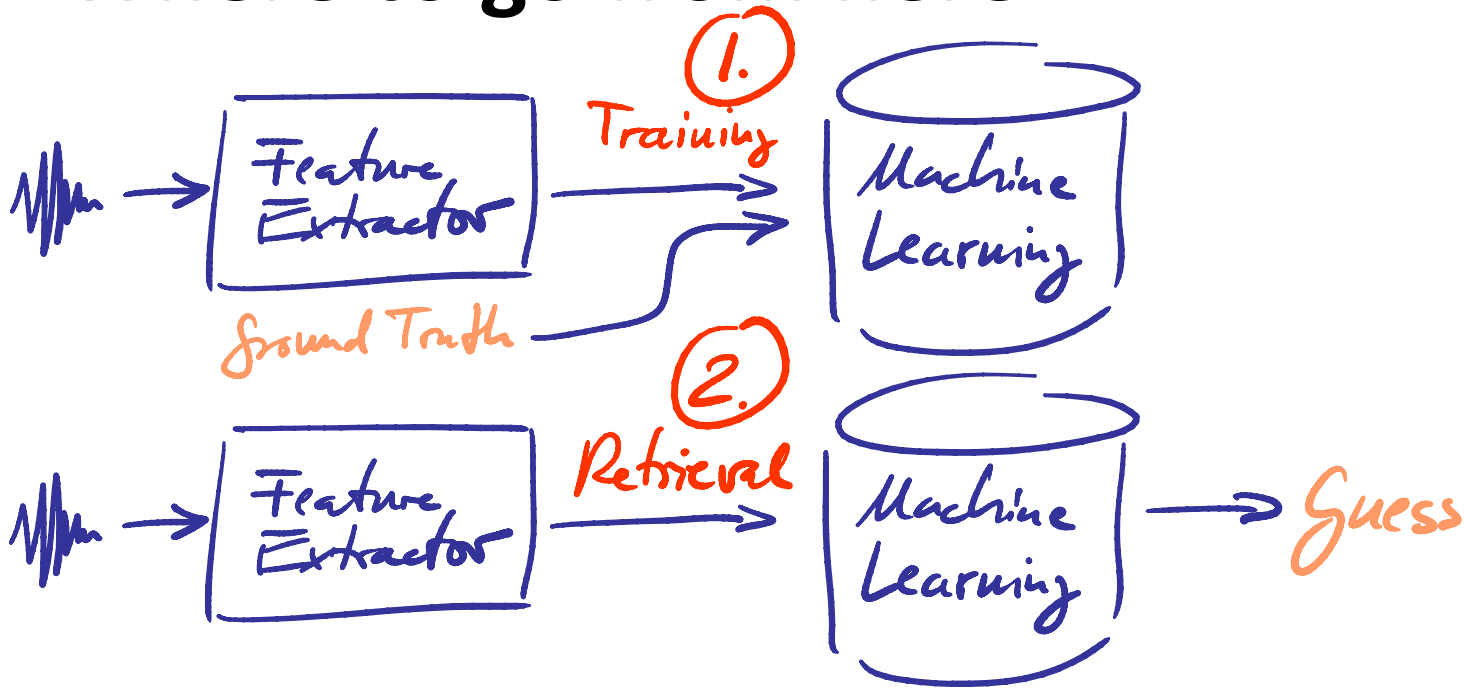- Cepstrum
- What could that be good for?

Integration into an application

Loudness
+ timbre
+ speech
  recognition

J. Loviscach. A nimble video
editor that puts audio first.
131st AES Conv., 2011.
Paper 8497.

Where to go from here

Models for: language, person, gender, accent, emotion, …

Slides and code:

www.j3L7h.de/talks