Introducing COVAREP:

A collaborative voice analysis repository for speech technologies

John Kane

Wednesday November 27th, 2013
SIGMEDIA-group
TCD

Introduction

(a) Gilles Degottex

(b) Thomas Drugman

(c) Tuomo Raitio

(d) Stefan Scherer

Motivation


OCT 26, 2013

The Big Data Brain Drain: Why Science is in Trouble

"...open, well-documented, and well-tested scientific code is essential not only to reproducibility in modern scientific research, but to the very progression of research itself."

Related toolkits

openSMILE: The Munich Versatile and Fast Open-Source Audio Feature Extractor.

Authors: Florian Eyben, Felix Weninger, Martin Woellmer, Bjoern Schuller

KALDI - Speech recognition toolkit

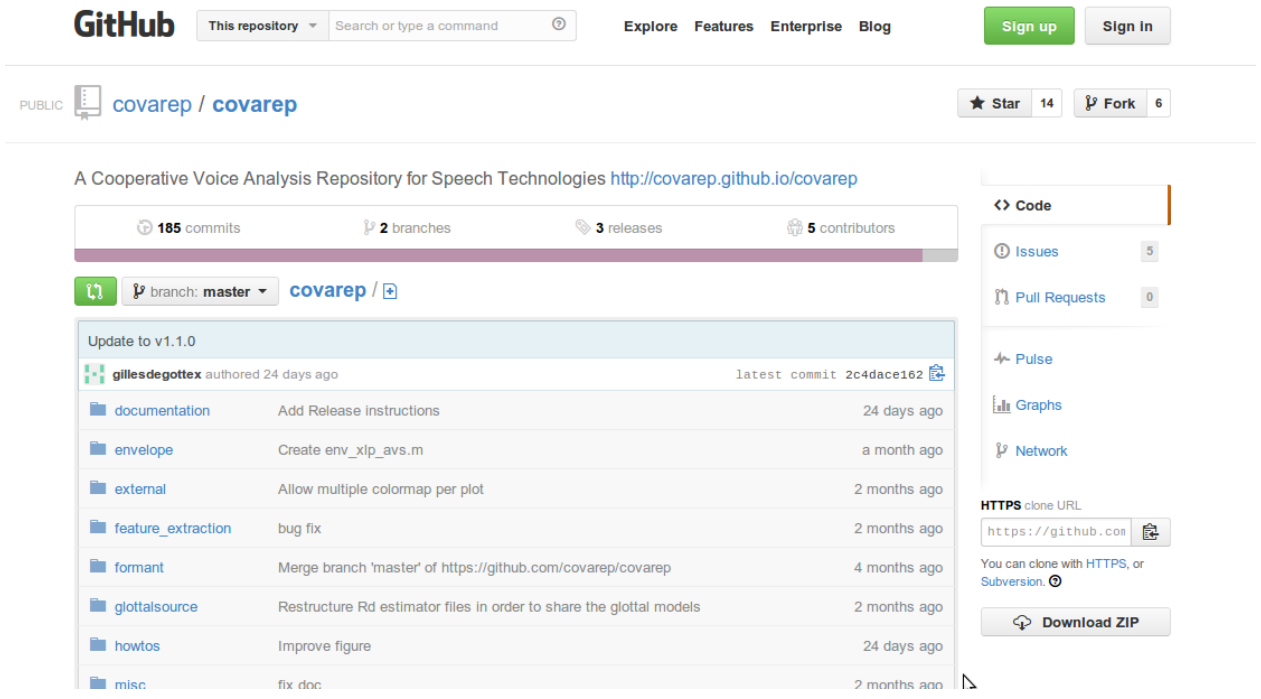
SPTK - Speech processing toolkit

VOICEBOX - Speech analysis toolkit

Solution?

Fast, effective results every time

COVAREP - Aims



Website: `http://covarep.github.io/covarep/index.html`
GitHub: `https://github.com/covarep/covarep`

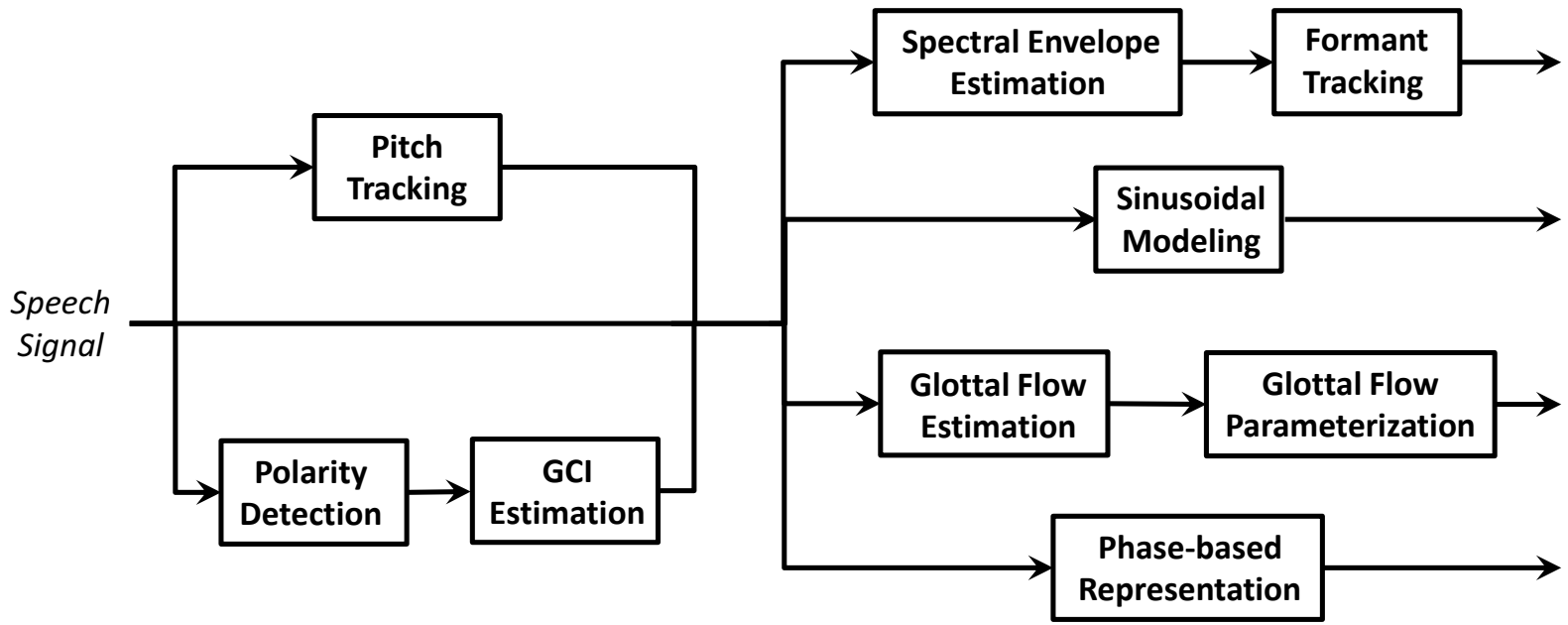
COVAREP - Aims

- More reproducible research
- Increase the availability and impact of speech processing algorithms
- Participation and feedback

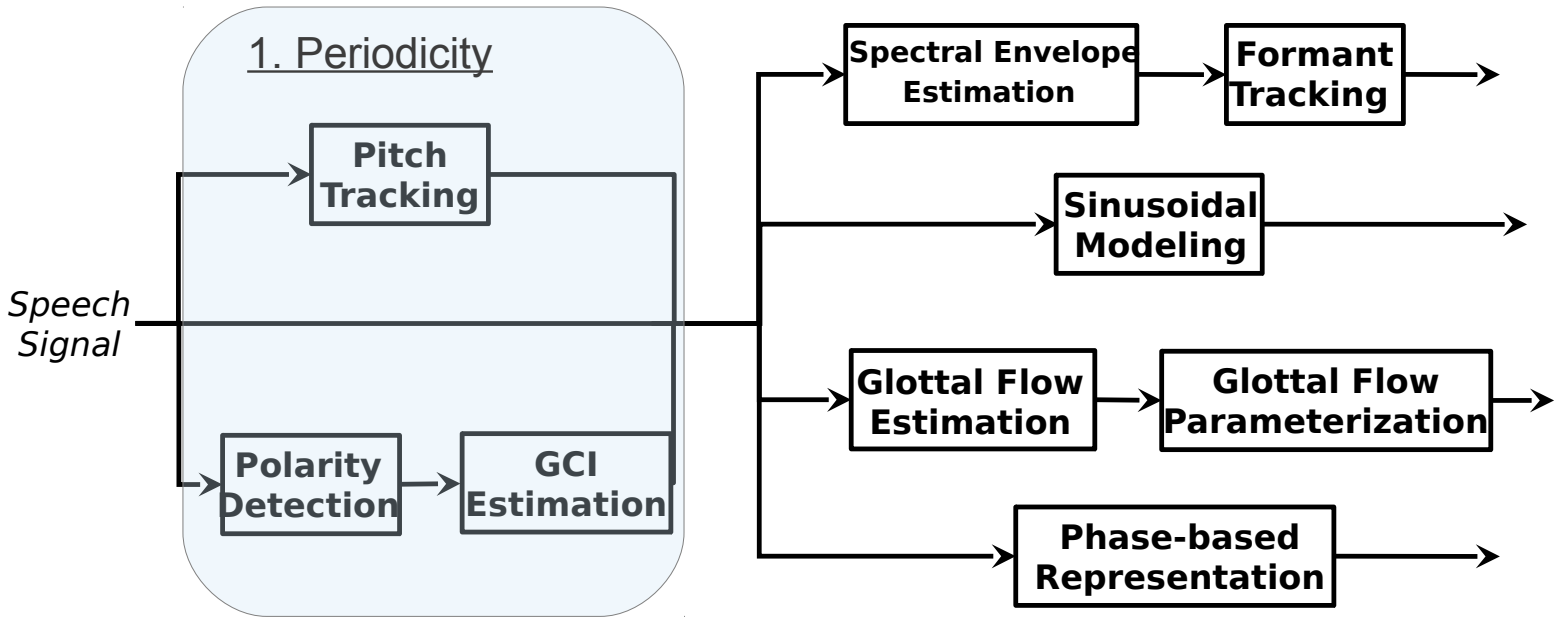
COVAREP - Scope

- Broad scope - any speech signal processing algorithms
- Speech analysis, synthesis, conversion, transformation, speech quality, enhancement, glottal source/voice quality analysis, etc.
- Use! Contribute!

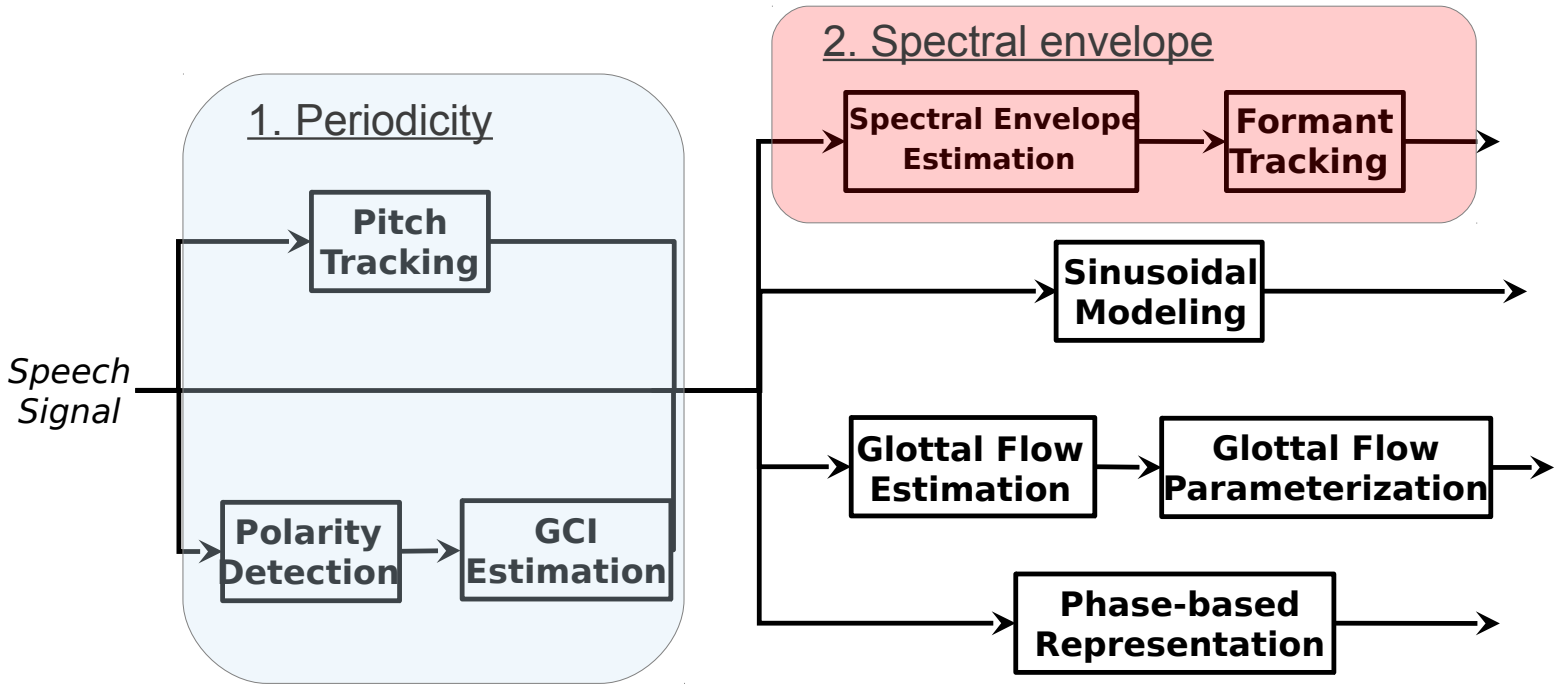
Overview of COVAREP



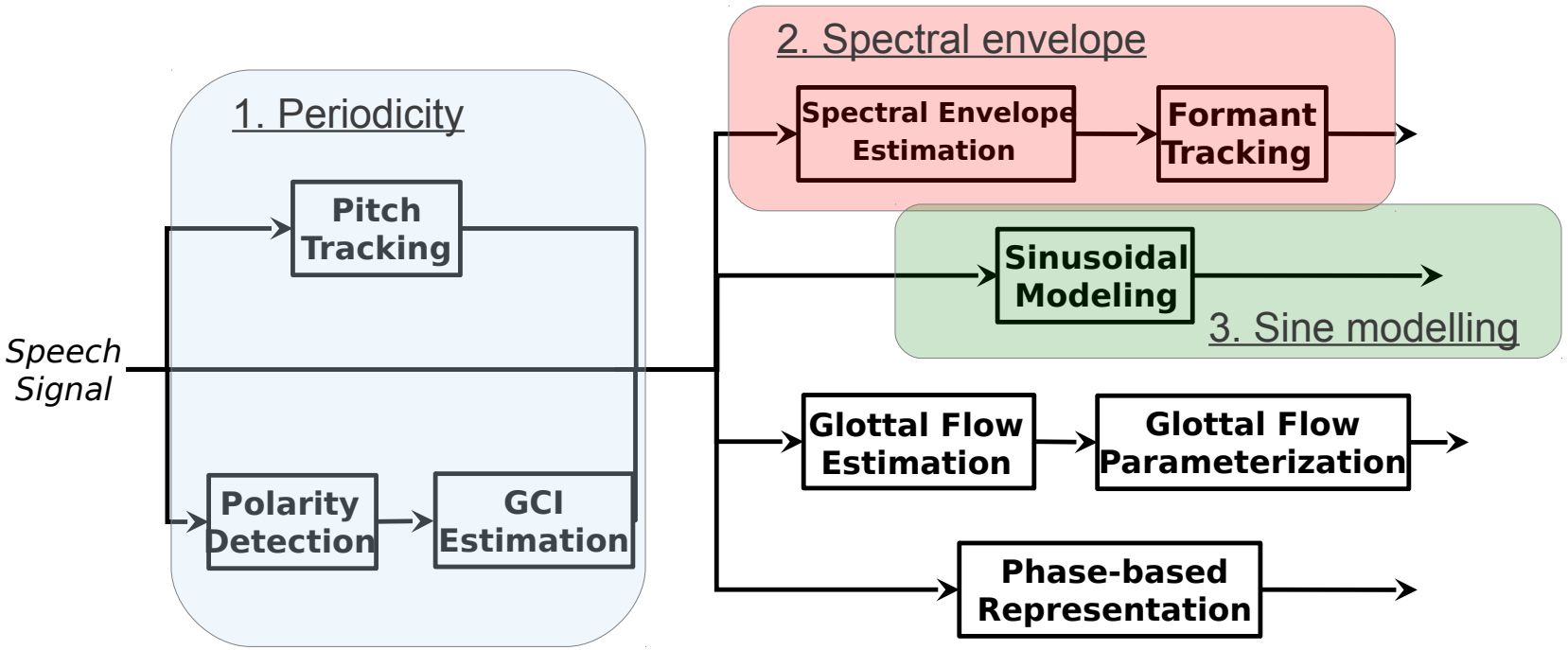
Overview of COVAREP



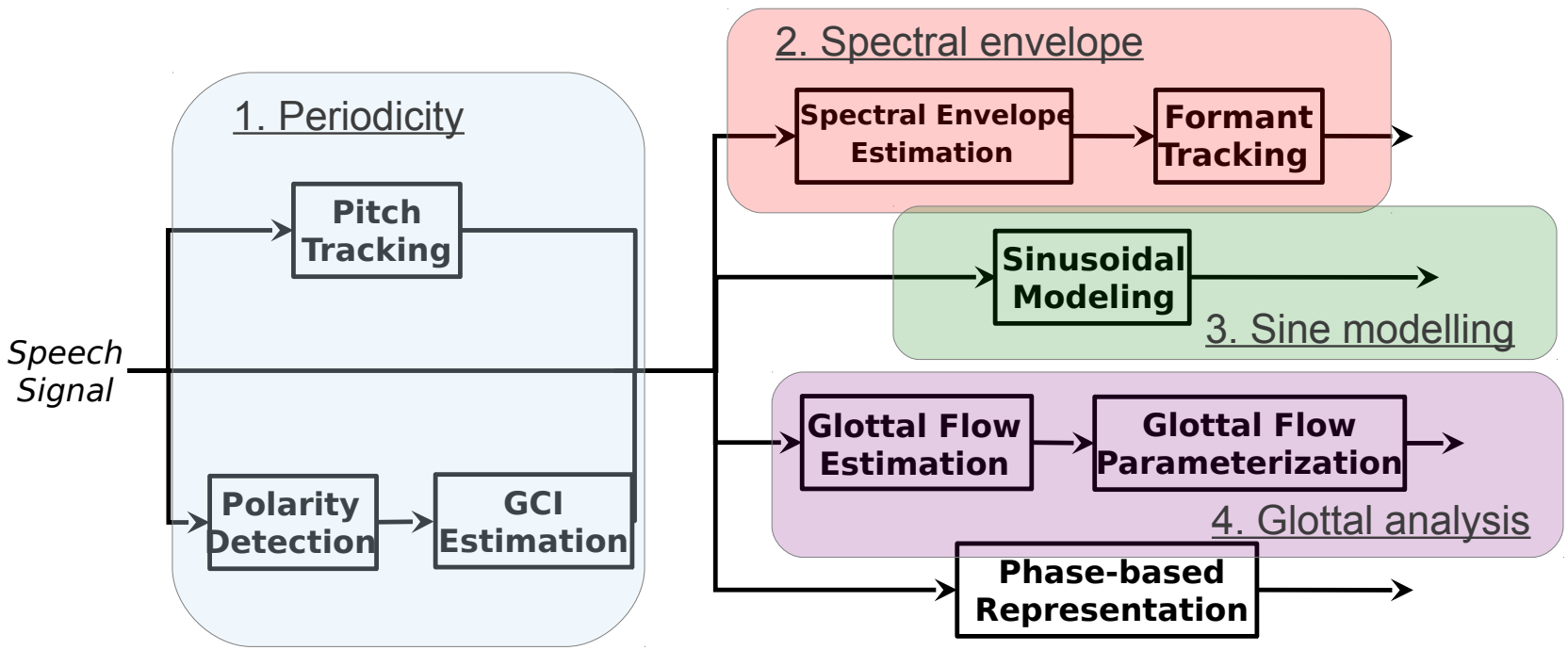
Overview of COVAREP



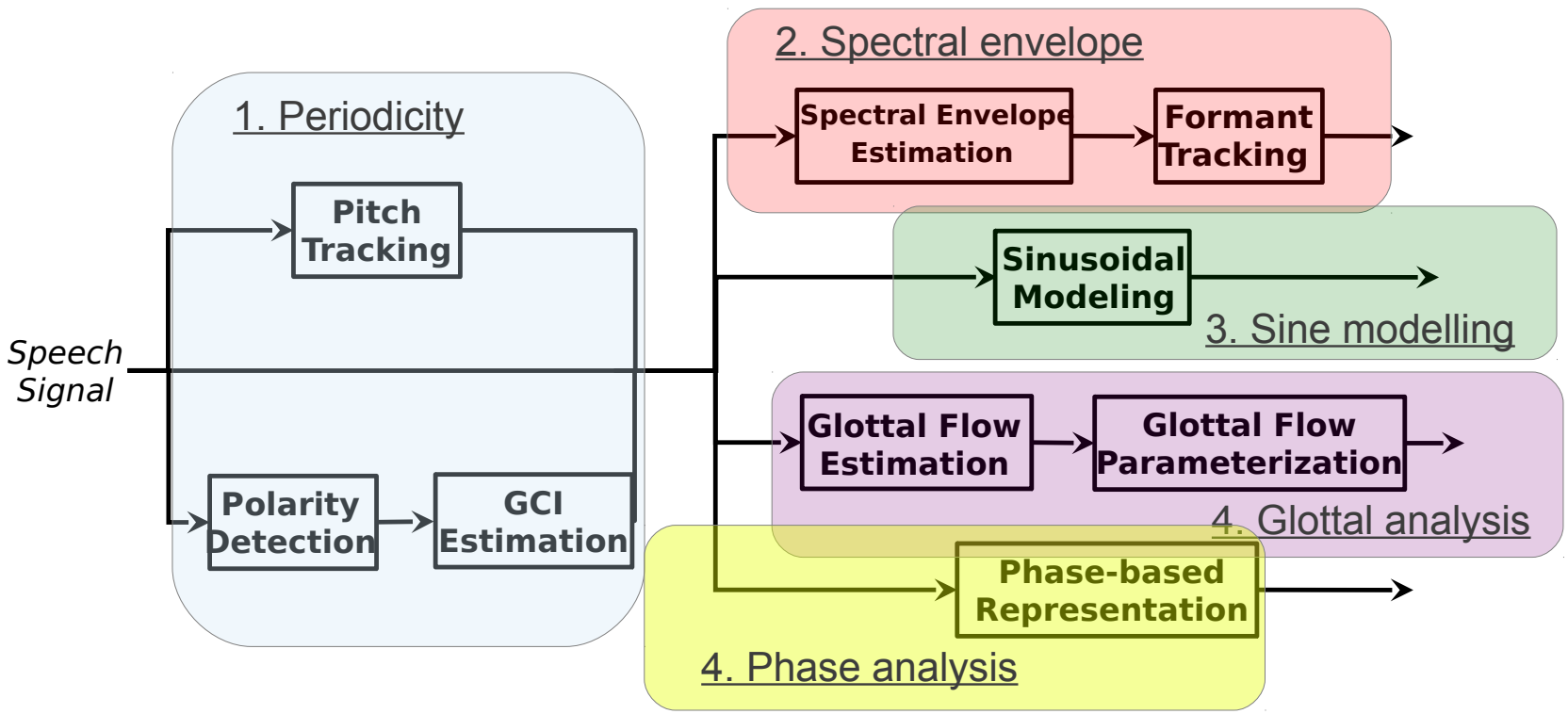
Overview of COVAREP



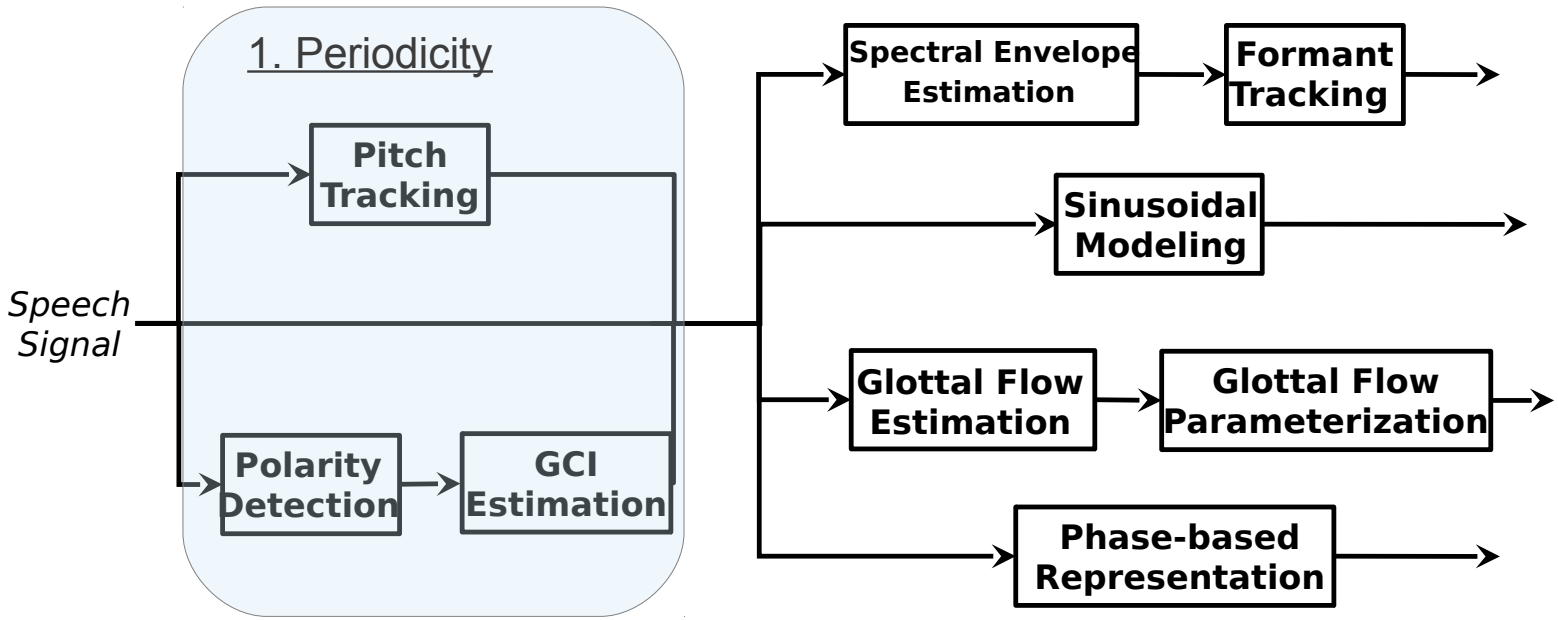
Overview of COVAREP



Overview of COVAREP



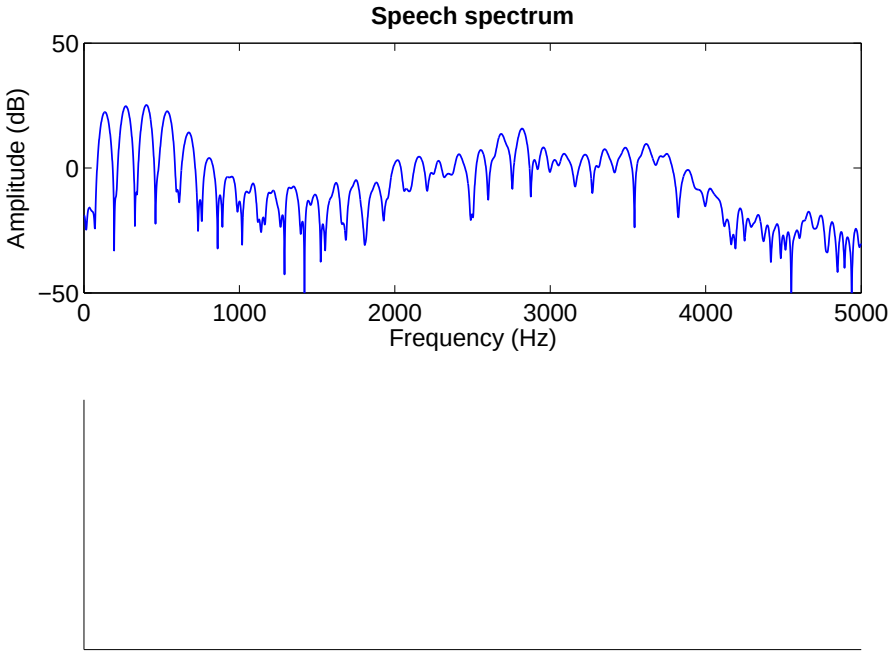
COVAREP - Periodicity & synchronicity



COVAREP - Periodicity & synchronicity

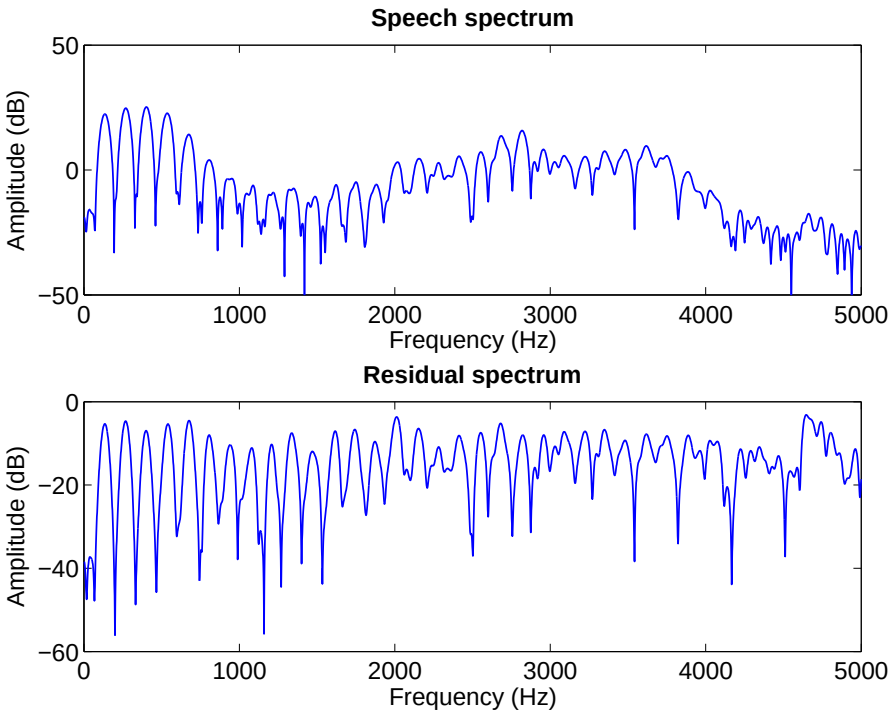
- Polarity detection
- f_0 and voicing decision extraction
- Detection of glottal closure instants

Periodicity & synchronicity - F0 extraction



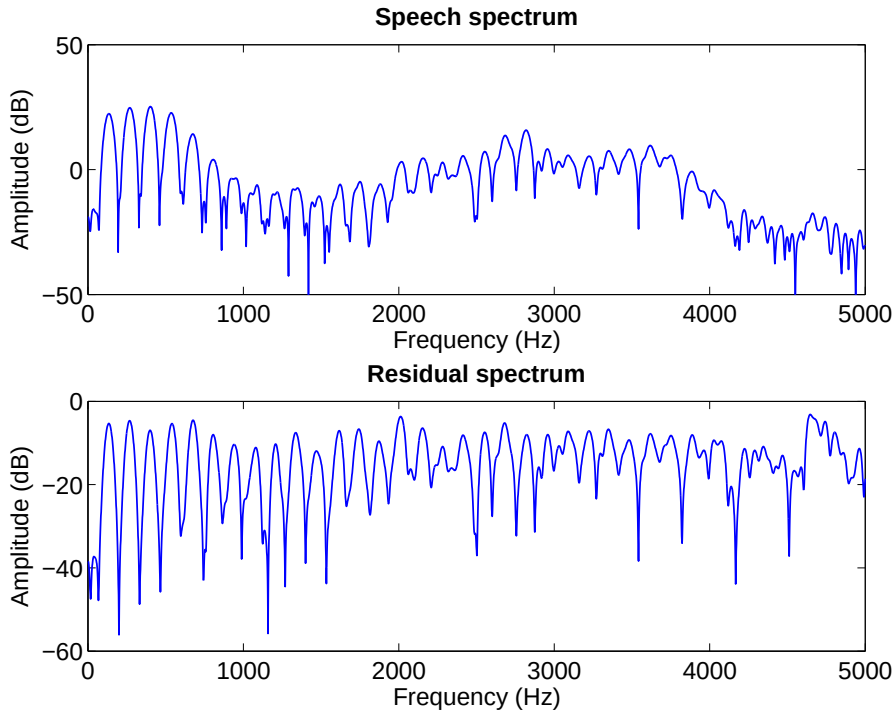
Speech amplitude spectrum

Periodicity & synchronicity - F0 extraction



Envelope-removed speech amplitude spectrum

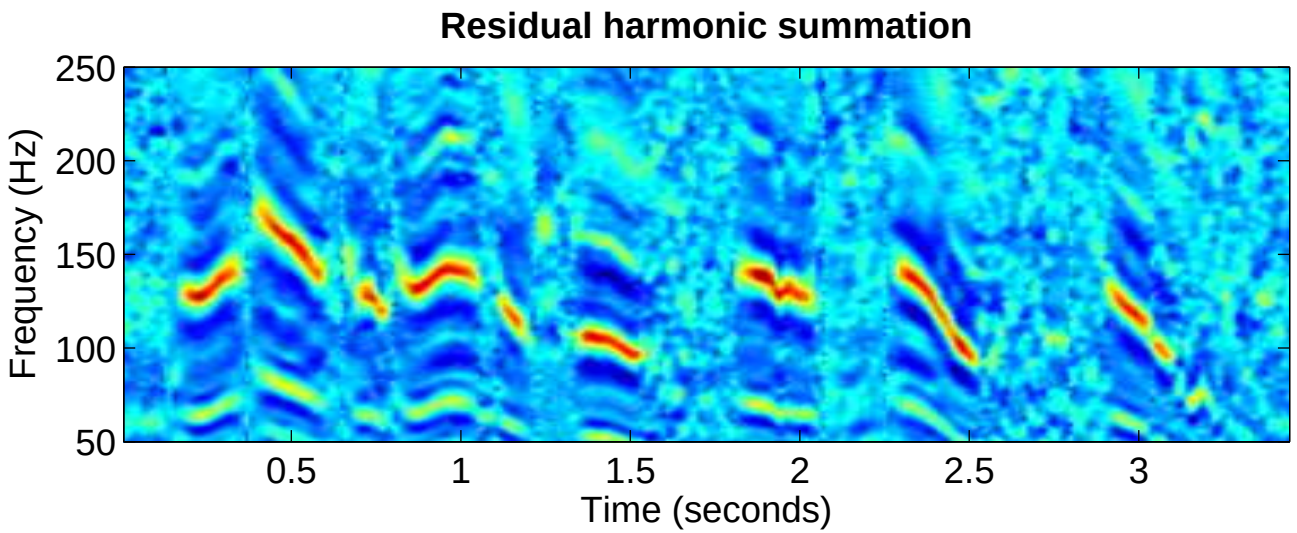
Periodicity & synchronicity - F0 extraction



$$\mathrm{SRH(f)} = E(f) + \sum_{k=2}^{N} [E(k \cdot f) - E((k-0.5) \cdot f)] \quad \text{for} \quad f \in [F0_{min}, F0_{max}]$$

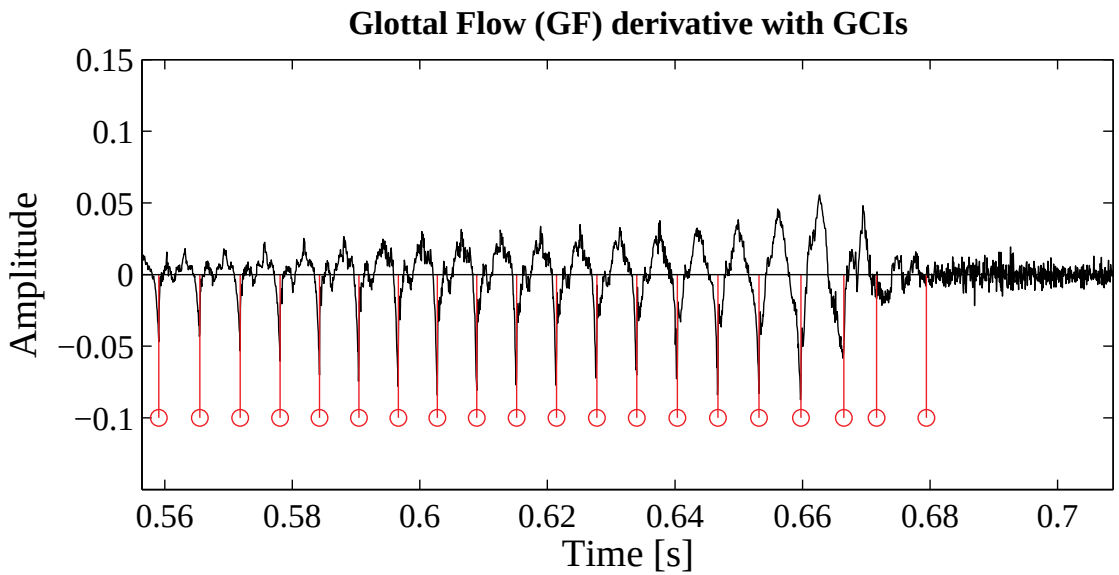
where E is the residual spectrum, f is frequency (Hz) and N is the number of harmonics considered

Periodicity & synchronicity - F0 extraction



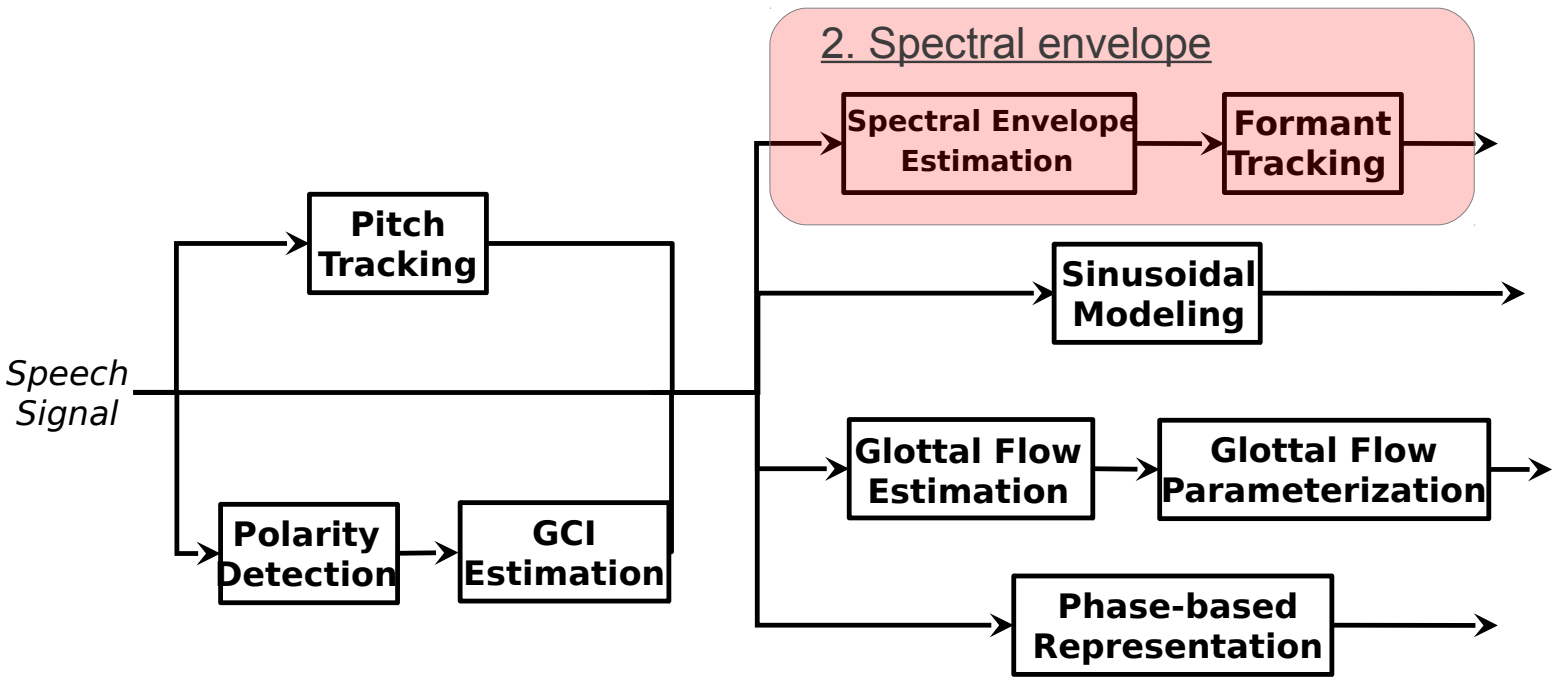
Residual harmonic summation over time

COVAREP - Periodicity & synchronicity



Detected glottal closure instants

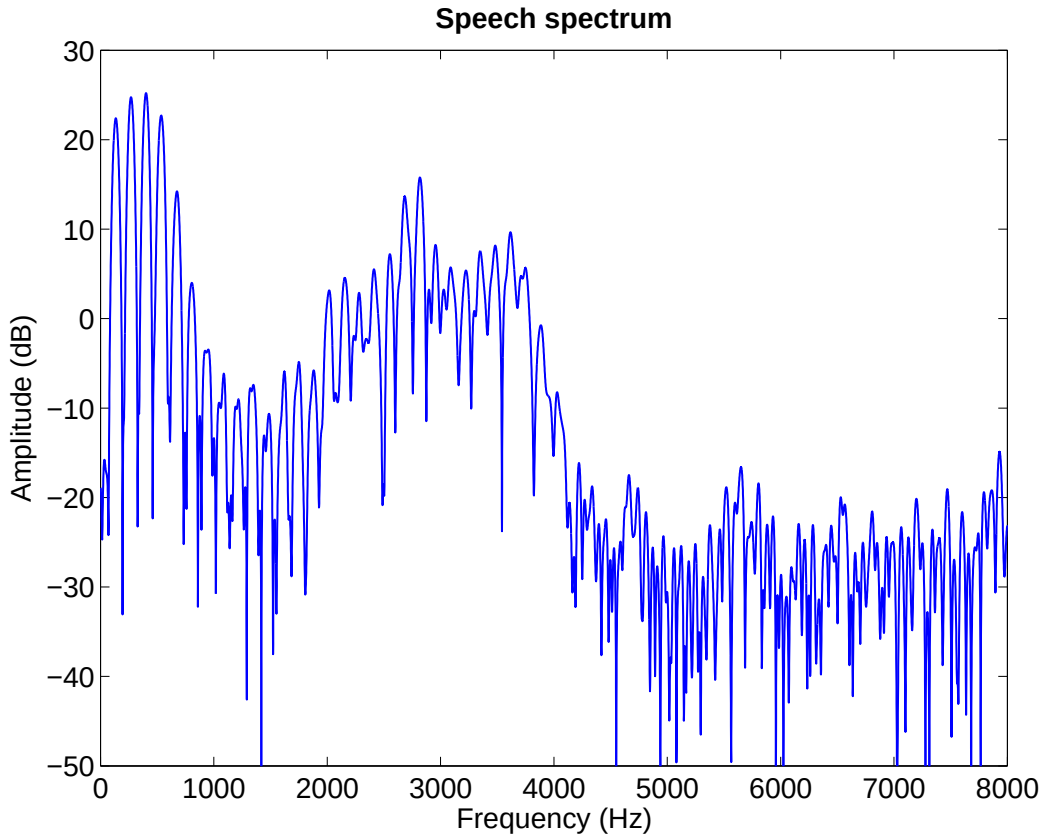
COVAREP - Spectral envelope estimation



COVAREP - Spectral envelope estimation

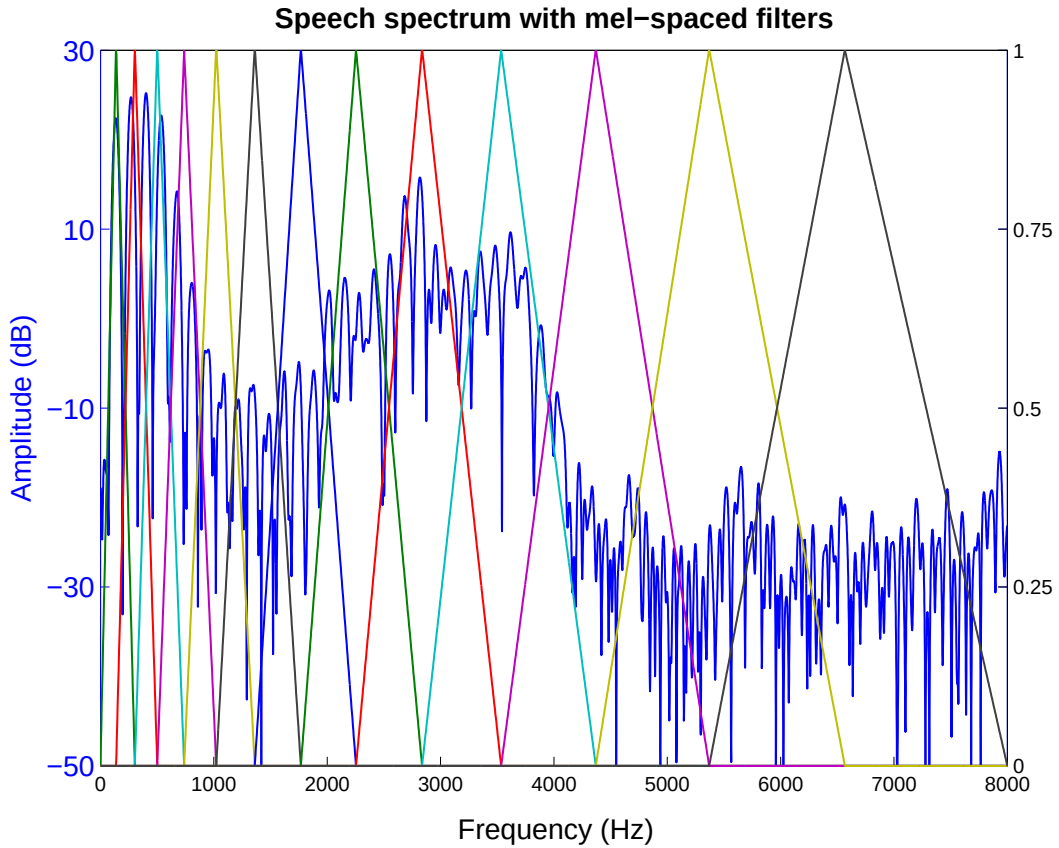
- Discrete all-pole (DAP) model
- "True envelope" (TE) - spectral envelope by iterative cepstral smoothing
- Weighted linear prediction
- Conversion from envelope to Mel-Frequency Cepstral Coefficients (MFCC)

COVAREP - Spectral envelope estimation



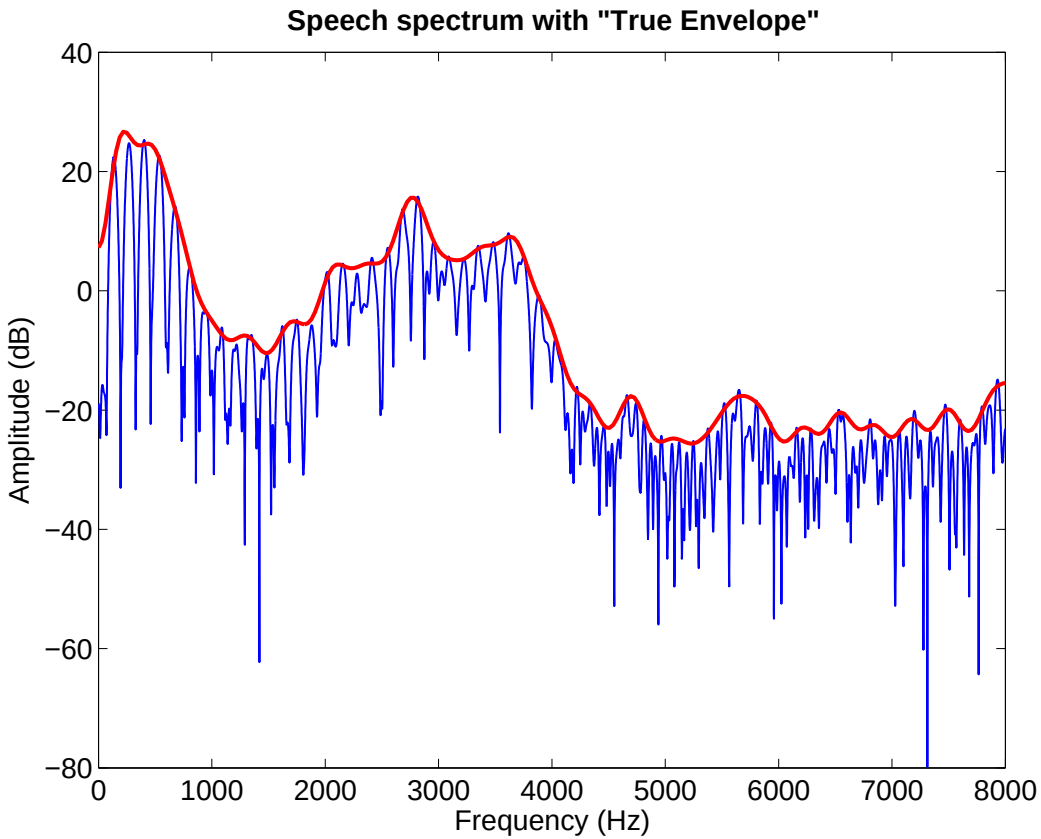
Speech amplitude spectrum

COVAREP - Spectral envelope estimation



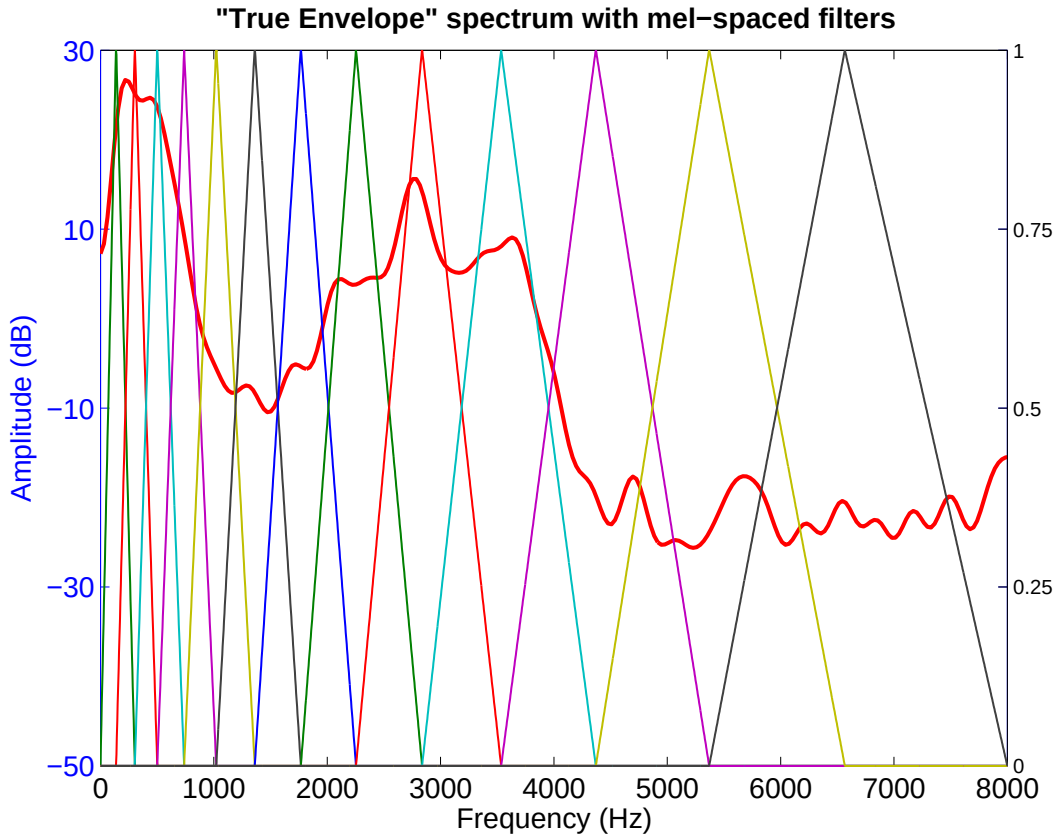
Speech spectrum with mel-spaced triangular filters

COVAREP - Spectral envelope estimation



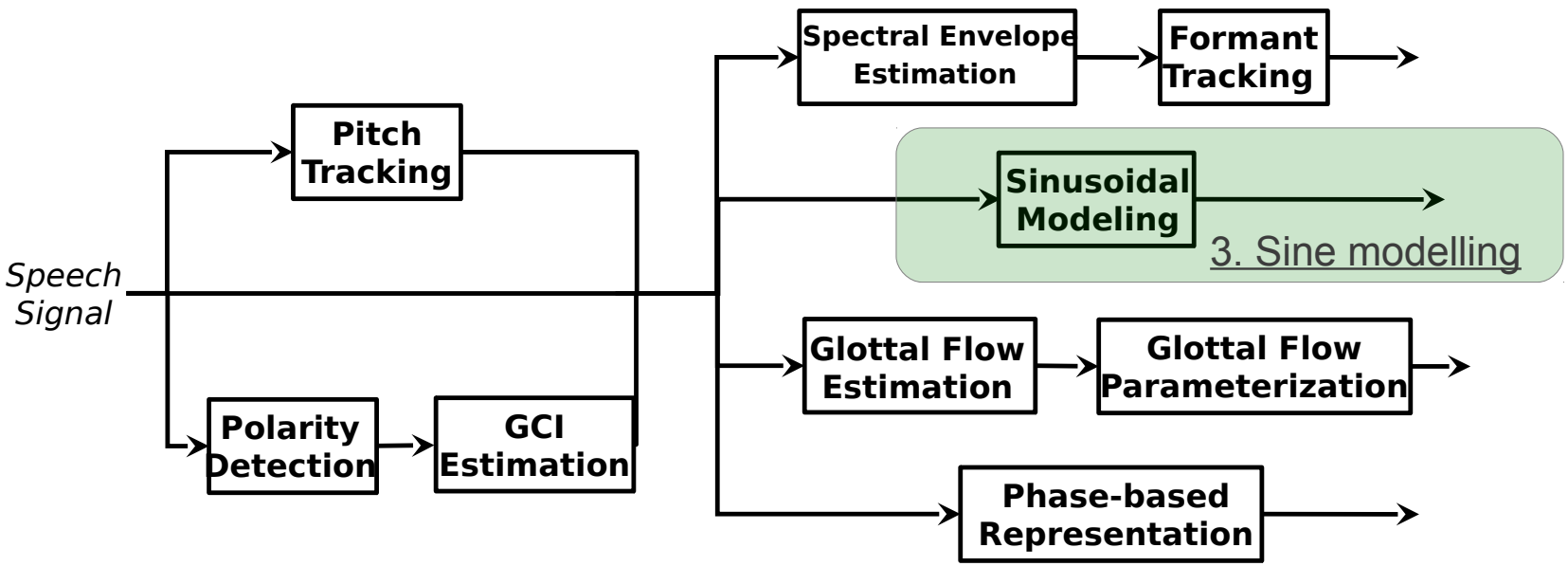
Speech spectrum with TE spectral envelope

COVAREP - Spectral envelope estimation



TE spectral envelope with mel-spaced triangular filters

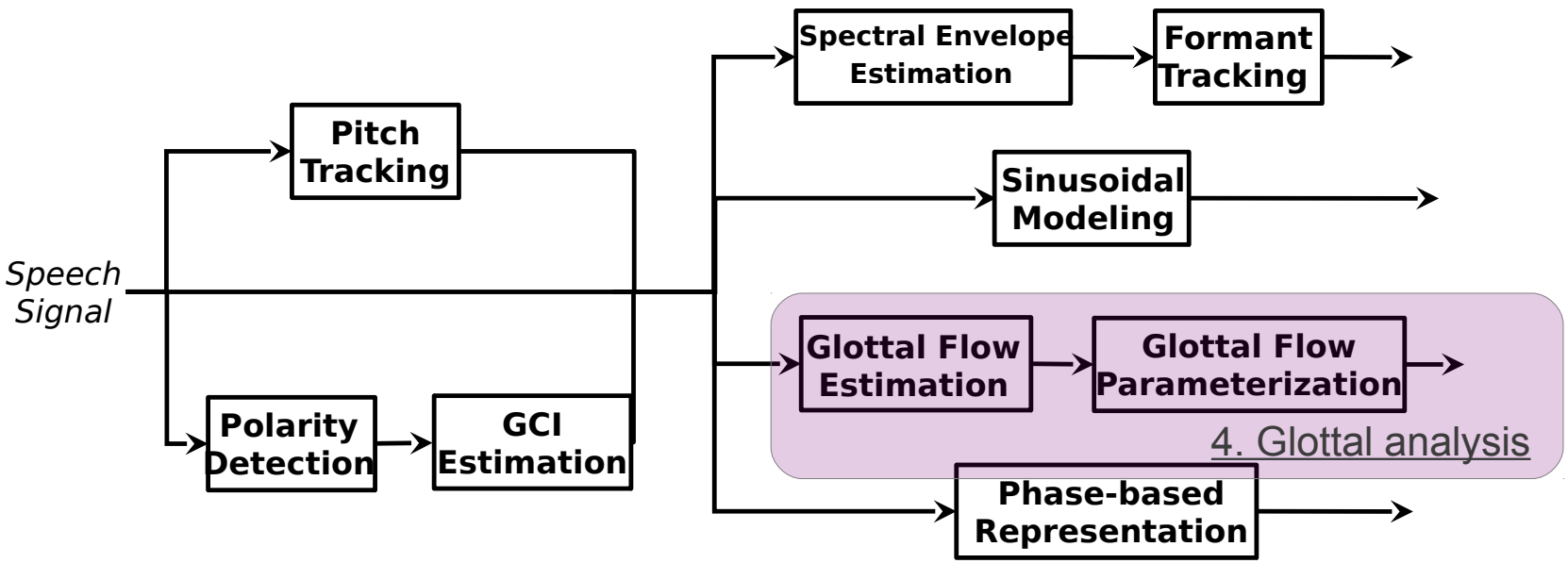
COVAREP - Sinusoidal modelling



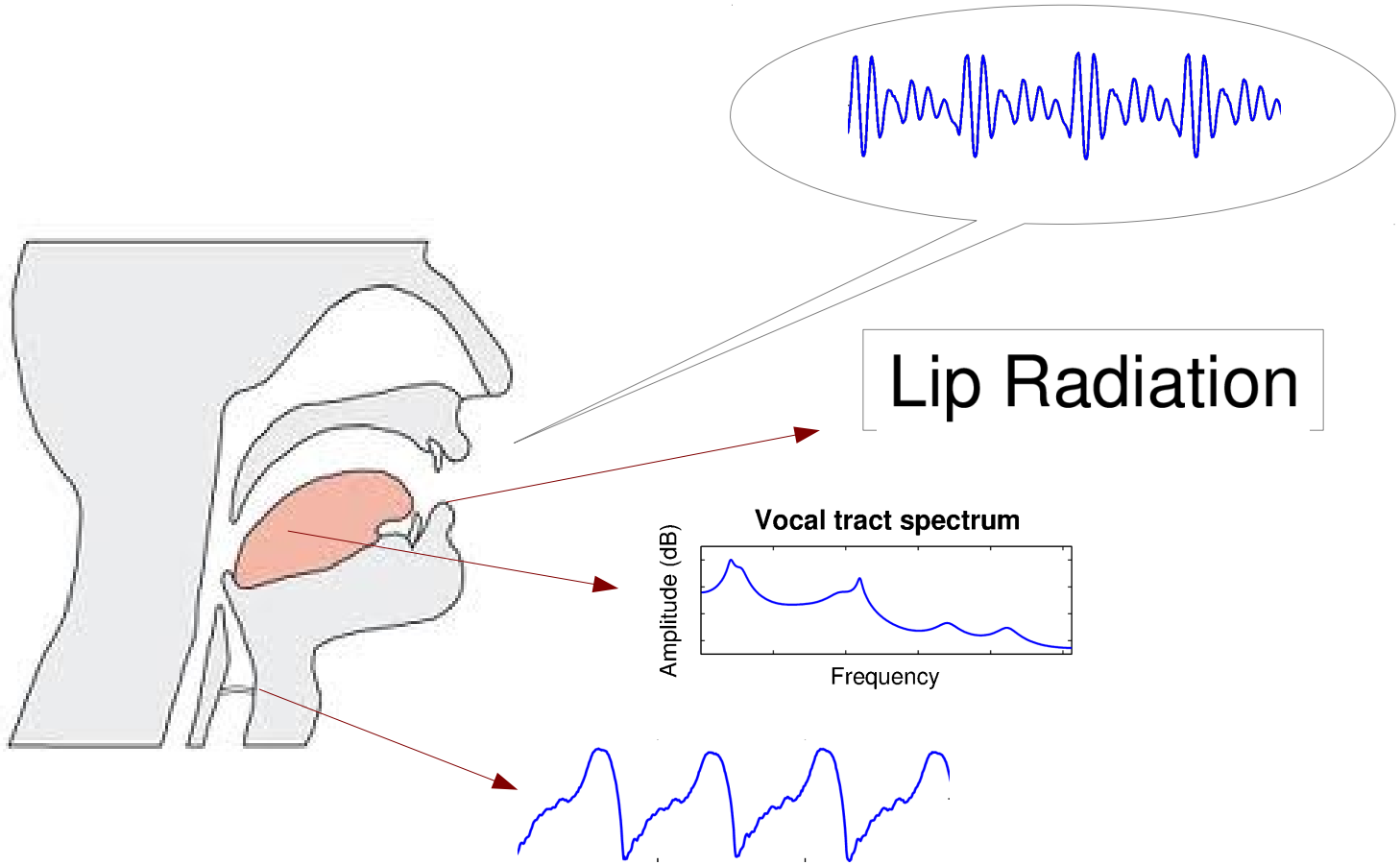
COVAREP - Sinusoidal modelling

- Harmonic model
- Quasi-Harmonic Model (QHM)
- Adaptive Harmonic Model (aHM)
- Harmonic synthesis

COVAREP - Glottal analysis



COVAREP - Glottal analysis



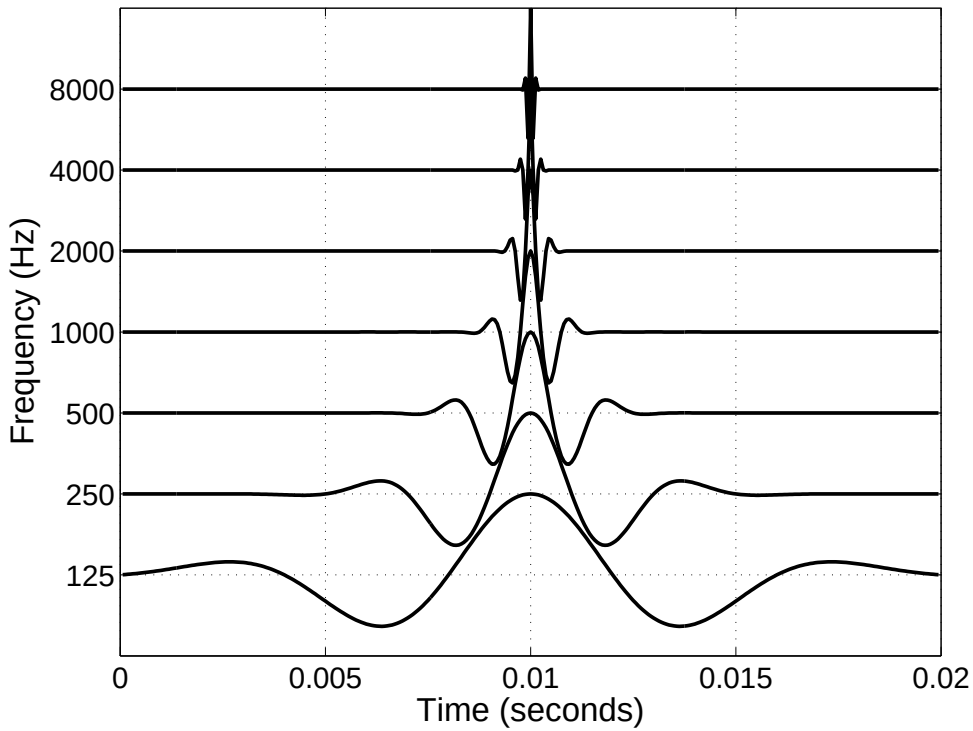
COVAREP - Glottal analysis

- Deconvolution of glottal source and vocal tract components
- Algorithms for parameterising the glottal source
- Detection of changes in tone-of-voice and voice quality

COVAREP - Glottal analysis

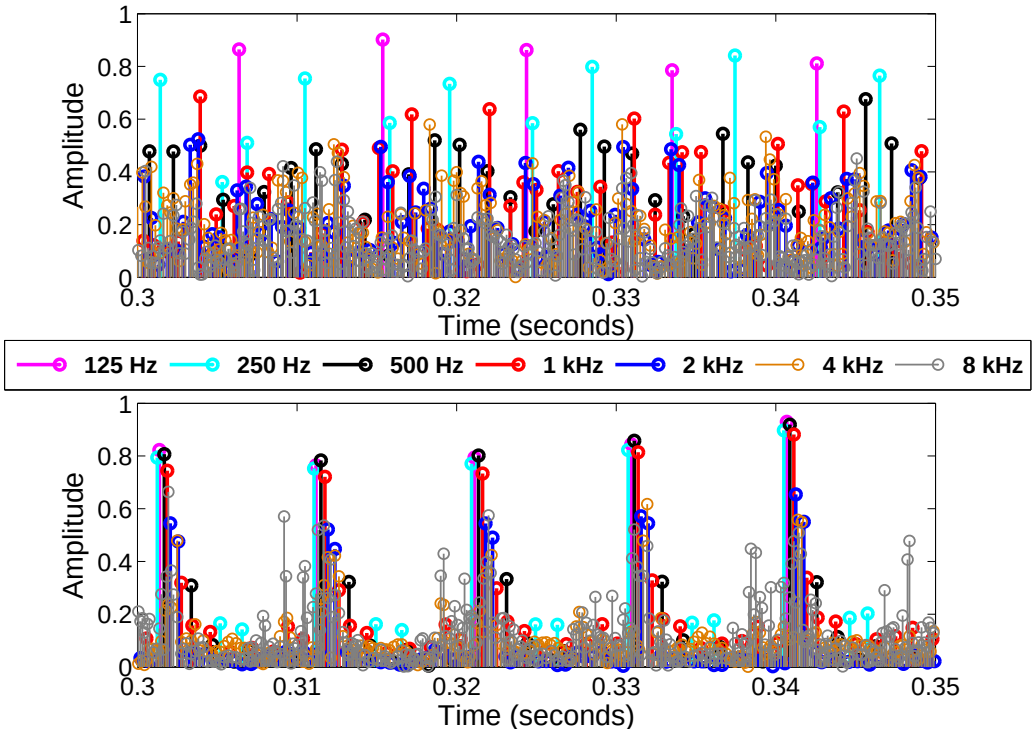
Vocal effort

COVAREP - Glottal analysis



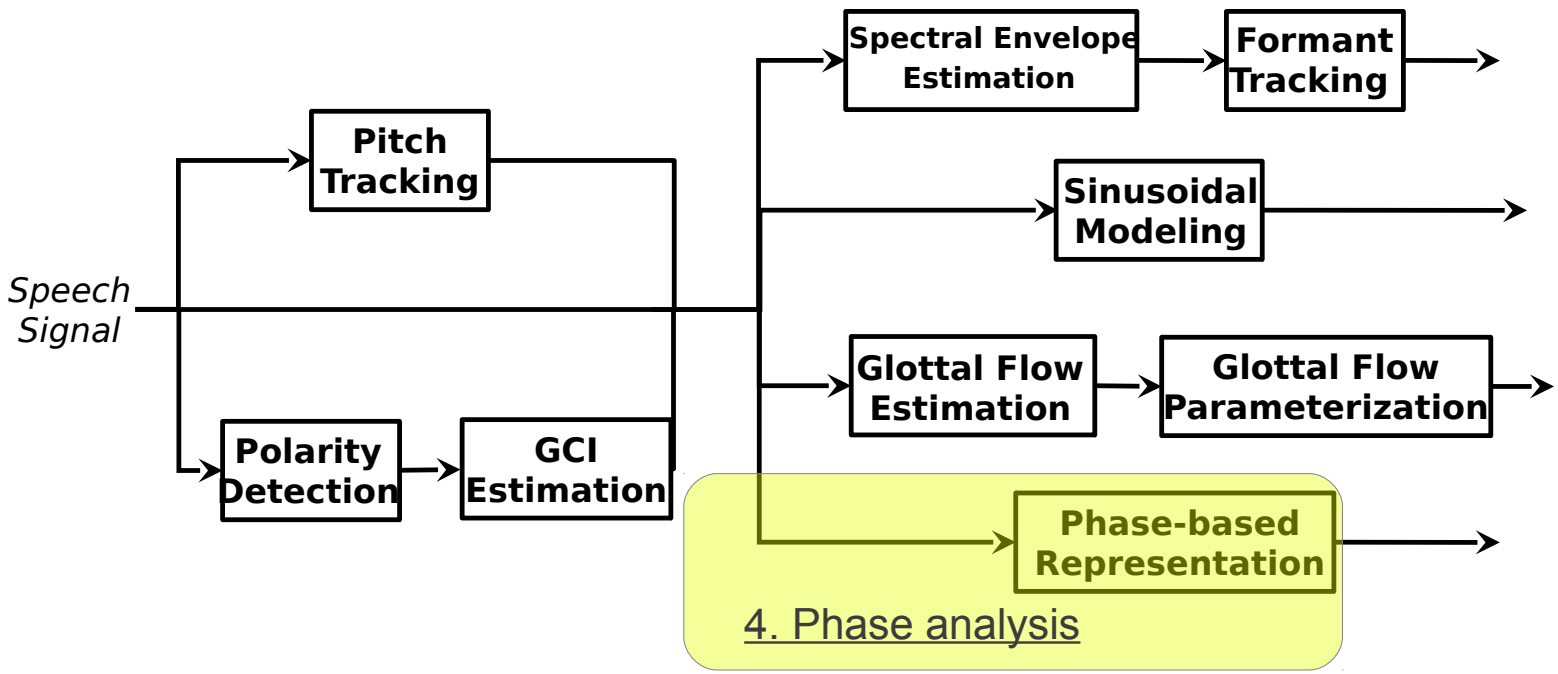
Wavelet decomposition of an impulse

COVAREP - Glottal analysis



All peaks across the different frequency bands
for breathy (top) and tense (bottom) speech samples

COVAREP - Phase processing



COVAREP - Phase processing

- Relative phase shift - speaker verification
- Phase distortion - emotional valence detection
- Chirp group delay represenation - detection of voice disorders

Emotion classification experiment

- **Speech data:** Berlin emotion database (10 speakers, 7 acted emotions, 500+ utterances)
- **Class labellng:** Emotion vs non-emotion (binary), Passive-neutral-active (3-class)
- **Feature extraction:** Using COVAREP v1.1.0
- **Classification:** Support vector machines (RBF kernel)
- **Validation:** Speaker independent, leave-one-speaker-out

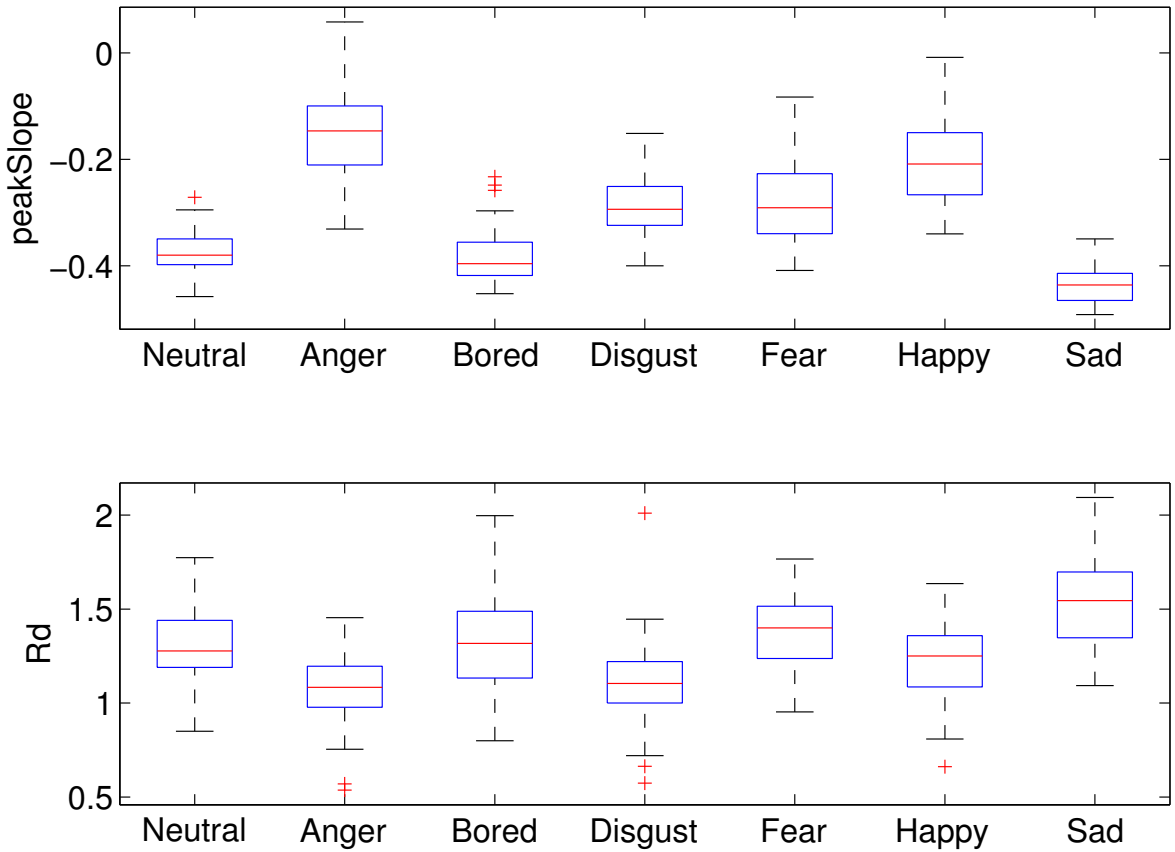
Emotion classification experiment

Feature sets

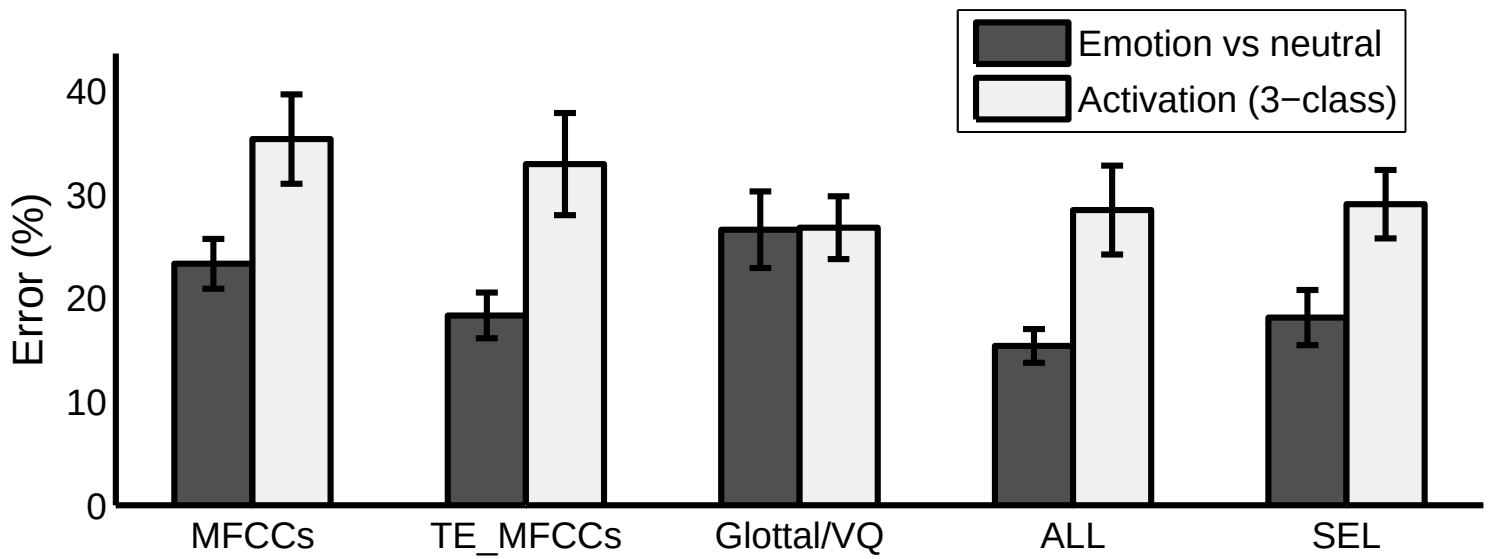
- **MFCC:** Standard Mel-frequency cepstral coefficients
- **TE-MFCC** MFCCs derived from True Envelope representation
- **Glottal/VQ:** Glottal and voice quality related features
- **ALL:** TE-MFCC and Glottal/VQ combined
- **SEL:** 10 most discriminative features

Speaker independent - Leave-one-speaker-out classification experiments

Emotion classification experiment - Results



Emotion classification experiment - Results



Emotion classification experiment - Results

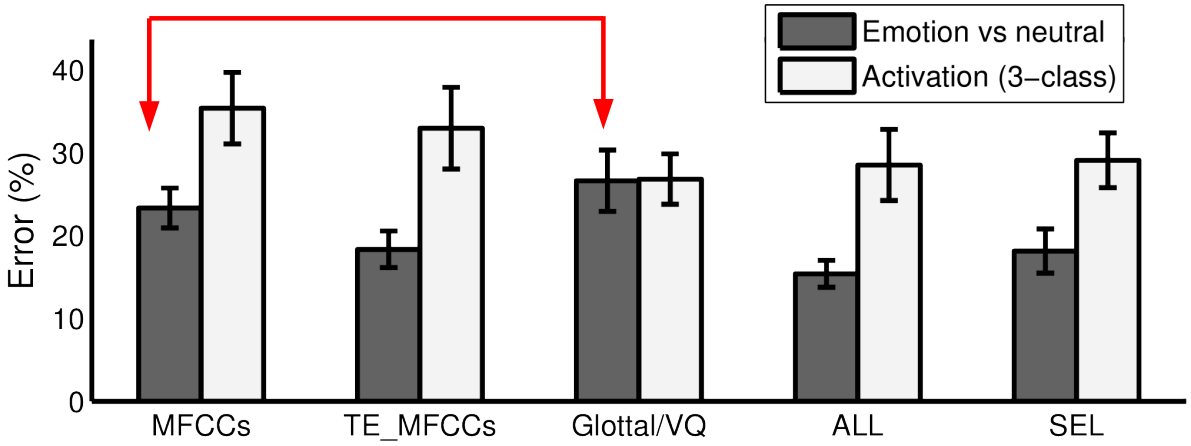
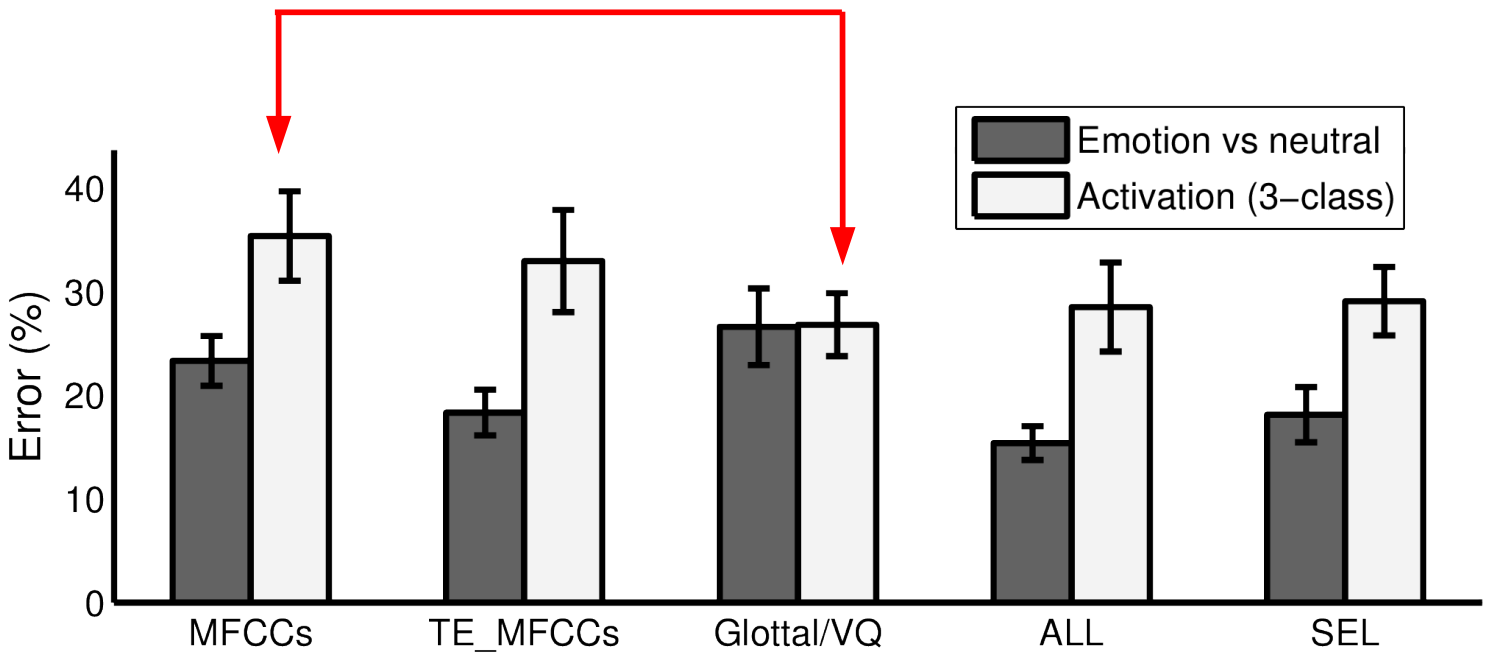


Table: Confusion matrix (%)

	MFCCs		Glottal/VQ	
	Neutral	**Emotion**	**Neutral**	**Emotion**
Neutral	**48**	52	**82**	18
Emotion	18	**82**	27	**73**

Emotion classification experiment - Results



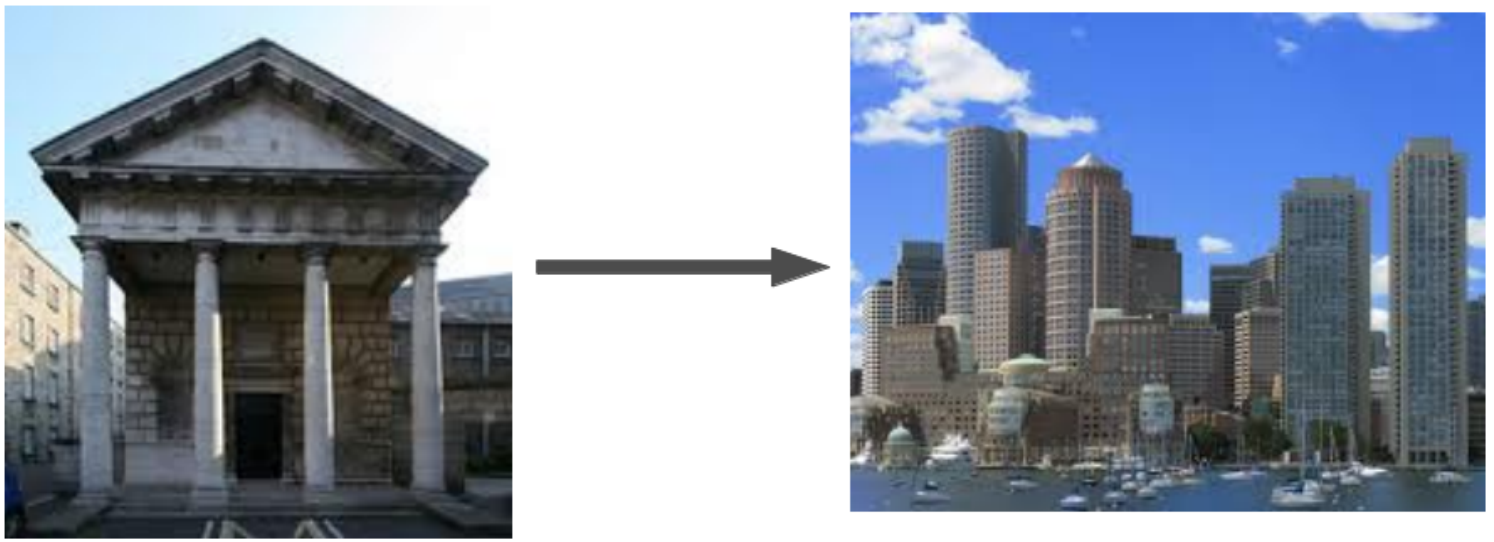
Potential applications for COVAREP algorithms

- Speech synthesis
- Speech recognition
- Modelling variation in speaking styles and affective states
- Speaker verification
- Voice pathology detection
- Lots of others!!

COVAREP summary

- Repository of open-source speech processing algorithms
- Cross-unversity/country effort
- Fast access to newly developed state-of-the-art algorithms
- Improve visability and impact
- More reproducible research

... and finally!



Thank you!

Resources:

Website: `http://covarep.github.io/covarep/`

GitHub: `https://github.com/covarep/covarep`

Paper: Degottex, G., Kane, J., Drugman, T., Raitio, T., "COVAREP - A collaborative voice analysis repository for speech technologies", Submitted to ICASSP 2014