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

“...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

COVAREP - Open-source speech processing repository
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

1. Periodicity
   - Pitch Tracking
   - Polarity Detection
   - GCI Estimation

Speech Signal

2. Spectral envelope
   - Spectral Envelope Estimation
   - Formant Tracking
   - Sinusoidal Modeling
   - Glottal Flow Estimation
   - Glottal Flow Parameterization
   - Phase-based Representation

COVAREP - Open-source speech processing repository
Overview of COVAREP

1. Periodicity
   - Pitch Tracking
   - Polarity Detection
   - GCI Estimation

2. Spectral envelope
   - Spectral Envelope Estimation
   - Formant Tracking

3. Sine modelling
   - Sinusoidal Modeling
   - Glottal Flow Estimation
   - Glottal Flow Parameterization
   - Phase-based Representation

Speech Signal
Overview of COVAREP

1. Periodicity
   - Pitch Tracking
   - Polarity Detection
   - GCI Estimation

2. Spectral envelope
   - Spectral Envelope Estimation
   - Formant Tracking

3. Sine modelling
   - Sinusoidal Modeling
   - Glottal Flow Estimation
   - Glottal Flow Parameterization

4. Glottal analysis
   - Phase-based Representation

Speech Signal
Overview of COVAREP

1. Periodicity
   - Pitch Tracking

2. Spectral envelope
   - Spectral Envelope Estimation
   - Formant Tracking

3. Sine modelling
   - Sinusoidal Modeling

4. Glottal analysis
   - Glottal Flow Estimation
   - Glottal Flow Parameterization

4. Phase analysis
   - Phase-based Representation

Speech Signal

COVAREP - Open-source speech processing repository
COVAREP - Periodicity & synchronicity

1. Periodicity

- Pitch Tracking
- Polarity Detection
- GCI Estimation

Speech Signal

- Spectral Envelope Estimation
- Formant Tracking
- Sinusoidal Modeling
- Glottal Flow Estimation
- Glottal Flow Parameterization
- Phase-based Representation
COVAREP - Periodicity & synchronicity

- Polarity detection
- $f_0$ and voicing decision extraction
- Detection of glottal closure instants
Periodicity & synchronicity - F0 extraction

Speech spectrum

Speech amplitude spectrum
Periodicity & synchronicity - F0 extraction

Speech spectrum

Residual spectrum

Envelope-removed speech amplitude spectrum
Periodicity & synchronicity - F0 extraction

\[
\text{SRH}(f) = E(f) + \sum_{k=2}^{N} [E(k \cdot f) - E((k-0.5) \cdot f)] \quad \text{for} \quad f \in [F_{0\text{min}}, F_{0\text{max}}]
\]

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

Residual harmonic summation over time
COVAREP - Periodicity & synchronicity

Glottal Flow (GF) derivative with GCIs

Detected glottal closure instants
COVAREP - Spectral envelope estimation

2. Spectral envelope

- Spectral Envelope Estimation
- Formant Tracking
- Sinusoidal Modeling
- Glottal Flow Estimation
- Glottal Flow Parameterization
- Phase-based Representation

Speech Signal

- Pitch Tracking
- Polarity Detection
- GCI 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 filters

Speech spectrum with mel-spaced triangular filters
COVAREP - Spectral envelope estimation

Speech spectrum with "True Envelope"

Speech spectrum with TE spectral envelope
COVAREP - Spectral envelope estimation

"True Envelope" spectrum with mel-spaced filters

Frequency (Hz)
Amplitude (dB)

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

Speech Signal

- Pitch Tracking
- Polarity Detection
- GCI Estimation

4. Glottal analysis

- Spectral Envelope Estimation
- Formant Tracking
- Sinusoidal Modeling
- Glottal Flow Estimation
- Glottal Flow Parameterization
- Phase-based Representation
COVAREP - Glottal analysis

Lip Radiation

Vocal tract spectrum

Amplitude (dB)

Frequency
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

4. Phase analysis

Speech Signal

- Pitch Tracking
- Polarity Detection
- GCI Estimation
- Spectral Envelope Estimation
- Sinusoidal Modeling
- Glottal Flow Estimation
- Glottal Flow Parameterization
- Formant Tracking

Phase-based Representation
COVAREP - Phase processing

- Relative phase shift - speaker verification
- Phase distortion - emotional valence detection
- Chirp group delay representation - detection of voice disorders
Emotion classification experiment

- **Speech data**: Berlin emotion database (10 speakers, 7 acted emotions, 500+ utterances)

- **Class labelling**: 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

![Graph showing emotion classification results for different emotions: Neutral, Anger, Bored, Disgust, Fear, Happy, Sad. The graphs display box plots for peakSlope and Rd metrics across the emotions, with statistical comparisons indicated by symbols.]

COVAREP - Open-source speech processing repository
Emotion classification experiment - Results

The diagram shows the error (%) for different features: MFCCs, TE_MFCCs, Glottal/VQ, ALL, and SEL. The error is measured for two scenarios: Emotion vs neutral and Activation (3-class). The bars indicate the average error with error bars representing the variability.
Emotion classification experiment - Results

Table: Confusion matrix (%)

<table>
<thead>
<tr>
<th></th>
<th>MFCCs</th>
<th>Glottal/VQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neutral</td>
<td>Emotion</td>
</tr>
<tr>
<td>Neutral</td>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>Emotion</td>
<td>18</td>
<td>82</td>
</tr>
</tbody>
</table>

COVAREP - Open-source speech processing repository
Emotion classification experiment - Results

Error (%)

- MFCCs
- TE_MFCCs
- Glottal/VQ
- ALL
- SEL

- Emotion vs neutral
- Activation (3-class)
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-university/country effort
- Fast access to newly developed state-of-the-art algorithms
- Improve visibility 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