COVAREP

A collaborative voice analysis repository for speech technologies

Authors/Maintainers

Gilles Degottex <degottex@csd.uoc.gr>
Computer Science Department, University of Crete, Heraklion, Greece

John Kane <kanejo@tcd.ie>
Phonetics and Speech Laboratory, Trinity College Dublin, Ireland

Thomas Drugman <thomas.drugman@umons.ac.be>
TCTS Lab - University of Mons, Belgium

Tuomo Raitio <tuomo.raitio@aalto.fi>
Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland

Stefan Scherer <scherer@ict.usc.edu>
Institute for Creative Technologies, University of Southern California, United States

In Short

COVAREP is an open-source repository of advanced speech processing algorithms and stored in a **GitHub.com** project where researchers in speech processing can store implementations of published algorithms.

Project page: `http://covarep.github.io/covarep`

Contact us:

ADDRESSED ISSUE

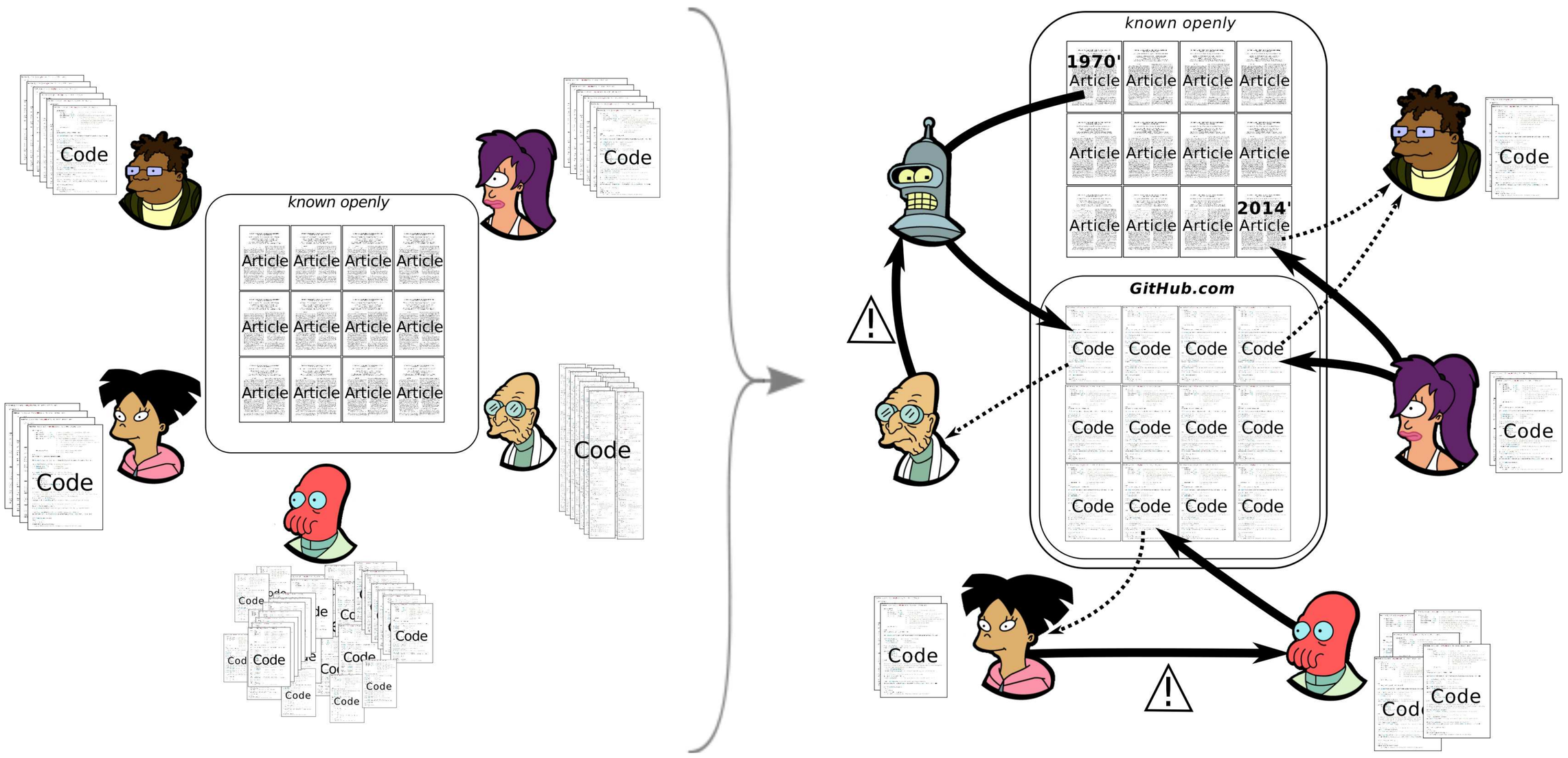
Algorithms can have some degree of complexity and, hence, can be difficult to accurately re-implement based on article descriptions.
Consequences:

– Re-implementations frequently have **significant differences** from the original ones.
– Many promising developments are **under-exploited or discarded**.
– Researchers tend to **stick to conventional** algorithms.

SUGGESTED SOLUTION

Encourage authors to include original implementations in a common repository.

+ **Single de facto version** for the speech community to refer to.
+ **Reproducible research**: Fairer comparisons of algorithms in articles.
+ **Encouraged usage**: The free availability encourages interdisciplinary works AND innerdisciplinary cooperations.
+ The GitHub platform allows **feedback** from the whole community.



REPOSITORY CHARACTERISTICS

Scope

- Speech analysis
- Glottal source/voice quality analysis
- Speech synthesis, conversion, transformation, enhancement
- etc. (basically not limited)

To maintain high standard: **only published works** in well-known conferences and journals.

The *GitHub.com* platform

- Report bugs and submit bug corrections
- Discuss existing implementations
- Suggest new implementations
- etc.

Legal issues and Intellectual Property (IP)

Because of multiple contributing institutions, homogeneous IP policy is impossible. Thus:

- **Repository** and not a toolbox (each method has its own license)
- Need compatible open-source licenses (e.g. GPL/LGPL, Apache, MIT).

Application Programming Interface (API)

- In the repository, each method is present only once.
- Dependent on *Voicebox* (`www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html`)
- A coding convention for:
  - Argument and return values, using the International System of Units (SI)
  - The documentation (e.g. arguments and return values descriptions)
  - Nothing else (e.g. no requirement to format the inner part of the implementations)

Language

- Initially written in **Matlab**® (`www.mathworks.com`)
- Strongly encourage to make the code compatible with **GNU Octave** (`octave.org`)
- The code has to be **platform independent** (Linux, Mac OS, Windows®)

We will consider implementations in Python or C/C++ in the future.

Current Content

- **Spectral envelope**
  - Discrete All-Pole (DAP)
  - "True"-Envelope (TE)
  - Stabilized weighted linear prediction using short-time-energy (SWLP), Short-time-energy-weighted linear prediction (WLP), Extended weighted linear prediction using Absolute-Value-Sum weighting (AVS-XLP)
  - Compression using Frequency Warped CEPpstrum (FWCEP) (similar to MCEP)
  - Relative Phase Shift (RPS), Phase Distortion (PD)
- **Sinusoidal modeling**
  - Peak Picking
  - Harmonic Model (HM)
  - Quasi-Harmonic Model (QHM)
  - Adaptive Harmonic Model and Adaptive Iterative Refinement (aHM-AIR)
  - Harmonic synthesis, Sinusoidal synthesis using OverLap-Add (OLA)
- **Glottal source analysis**
  - Polarity detection (RESKEW)
  - Pitch tracker using Summation of the Residual Harmonics (SRH)
  - Speech Event Detection based on the Residual Excitation And Mean-based Signal (SEDREAMS) for Glottal Closure Instant (GCI) determination, SEDREAMS-based GCI method optimised for non-modal voice qualities (SE-VQ)
  - Detection of creaky voice (vocal/glottal fry)
- **Glottal source parameters**
  - Spectral tilt correlate (PEAKSLOPE)
  - Maxima Dispersion Quotient (MDQ)
  - Normalised amplitude quotient (NAQ)
  - Quasi-open quotient (QOQ)
  - Difference in amplitude of the first two harmonics of the differentiated glottal source spectrum (H1-H2)
  - Harmonic Richness Factor (HRF)
  - Parabolic Spectral Parameter (PSP)
- **Glottal flow estimation**
  - Iterative Adaptive Inverse Filtering (IAIF) Glottal Inverse Filtering
  - Glottal flow derivative estimation based on the complex cesptrum
  - Model-based glottal analysis: Spectrum of the Liljencrants-Fant (LF) glottal flow derivative model, Rd parameter estimation of the LF model based on Mean Squared Phase and 2nd order Difference (MSPD2)
- **Electroglottographic (EGG) analysis**: Open quotient measurement with number of peaks at glottal opening and closing (DECOM)
- **Formant tracking**: Chirp Group Delay formant tracking (CGD)
- **Spectral analysis** (Fan-Chirp Transform (FChT)
- Automatic **feature extraction** available in COVAREP