GFM-Voc: a tool for analysis and modification of the glottis signal

Olivier Perrotin[1]*, Ian V. McLoughlin[2]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, Grenoble, France
[2]University of Science and Technology of China, Hefei, P.R. China

Keywords: glottal flow model; source-filter decomposition; voice quality

Introduction

Estimation of the glottal source signal from recorded voice has been a long-sought goal in voice research [1]. Most recent methods for glottal inverse filtering have addressed the issue of separating the first vocal tract (VT) resonance from the glottal formant, i.e., the low-frequency resonance that describes the open phase of the vocal fold vibration. However, the glottis signal is broadband and few methods are capable of estimating the high-frequency glottis spectral tilt, correlated to the closing phase of the vocal fold vibration, and crucial to vocal force perception [2]. Moreover, such methods often estimate the glottis signal in the time domain [3]. We present here the Glottal Flow Model-based vocoder (GFM-Voc) [4] that allows the real-time extraction and modification of both VT and glottis contributions to the voice signal in the frequency domain, as compact sets of filter parameters. In particular, the glottis filter has a wide-band frequency response, including both glottal formant and spectral tilt characteristics that are closely related to vocal force perception [2]. GFM-Voc thus provides a straightforward way to analyse voice quality, and allows the real-time modification of VT and glottis filters before re-synthesis. Examples of applications include expressive speech synthesis ; auditory feedback perturbation ; and speech therapy.

Methods

A key ingredient of GFM-Voc is the analysis part performed by an improved Iterative Adaptive Inverse Filtering (IAIF) method [5] based on a Glottal Flow Model, which we call GFM-IAIF [6]. Basic IAIF separates a speech frame into 4 components: a flat-spectral-envelope excitation, a high-order linear prediction (LP) VT filter, a low-order LP glottis filter, and a first order high-pass lip radiation filter. In particular, glottis and lip radiation are combined into a first order filter, that reduces the glottis filter to the glottal formant contribution [5]. To include the spectral tilt in the glottis filter, a recent method called Iterative Optimal Pre-emphasis (IOP)-IAIF [7] uses an unconstrained order to encompass all the slope of the speech frame spectrum in the glottis filter. While the IOP-IAIF improvements are merited, we believe that a high filter order risks endowing the glottal model with too much complexity. Our GFM-IAIF method is built on the assumption that a third order filter is enough to model both glottal formant (with a complex conjugate pole pair) and spectral tilt (with one real pole), based on the findings of [2]. A Matlab implementation of GFM-IAIF is available in [8]. GFM-Voc then implements a full analysis-resynthesis pipeline [4]. GFM-IAIF extracts the glottis and VT filters, that are then modified by changing the position of their poles. Changing the glottis filter poles is equivalent to varying the position and bandwidth of the glottal formant, and position of spectral tilt, resulting in modification of voice quality (vocal force and tension). Changing the VT poles allows the shifting of formants that correlates with changes of jaw, tongue, and lip positions. Finally, the speech signal is reconstructed by filtering the unchanged excitation with the modified filters.

Results

GFM-IAIF has been evaluated against IAIF and IOP-IAIF on both synthetic and natural speech, various vowels and voice qualities (from soft to loud voice), and male and female speakers [6]. Overall, we observed similar performances for estimation of glottal formant parameters. However, GFM-IAIF provided spectral tilt parameters that were closest to ground truth and more discriminative for voice quality than other methods. Specifically, IOP-IAIF and IAIF tend to attribute too much and not enough spectral tilt to the glottis, respectively. Finally, a demonstration of the full GFM-Voc framework for real-time voice modification can be found in [9].

Discussion

GFM-Voc is the first framework allowing high-quality and real-time modification of vocalic formants and voice quality. It relies on GFM-IAIF that extracts vocal tract and glottis contributions as an intuitive set of filter parameters. This method is frequency-domain based as it describes the glottis with spectral parameters, and was evaluated on those. It is not yet investigated how well GFM-IAIF can extract a glottal waveform that is faithful to the temporal vibration of vocal folds, and this is let for future work.

References

[1] Drugman *et al*, Computer Speech & Language, 28(5): 1117-1138, 2014.
[2] Doval *et al*, Acta Acustica, 92(6): 1026-1046, 2006.
[3] Degottex *et al,* Speech Communication, 55: 278-294, 2013.
[4] Perrotin *et al*, Proc. of Interspeech, 3685-3686, 2019.
[5] Alku *et al*, Speech Communication, 11(2-3): 109-118, 1992.
[6] Perrotin *et al*, ICASSP, 7160-7164, 2019.
[7] Mokhtari *et al*, Speech Communication, 104: 24-38,2018.
[8] *http://github.com/operrotin/GFM-IAIF*, (online).
[9] *http://youtu.be/9djf7ljkrsY,* (online).

*corresponding author e-mail

*corresponding author e-mail