

## On the use of the Glottal Source for Expressive Speech Analysis

Thomas Drugman, Thomas Dubuisson, Thierry Dutoit

TCTS Lab, University of Mons, Belgium

This contribution summarizes our recent investigations in the use of the glottal source for characterizing expressive voice. It is organized in three main parts. First, we study which methods are the most suited for estimating the glottal flow directly from the speech signal. This is a particularly difficult task which is a typical case of blind separation, since neither the vocal tract nor the glottal components are observable. Secondly, we focus on the parameterization of the resulting glottal flow estimates, highlighting which features are the most appropriate to characterize it. Finally, we report our results of glottal analysis of expressive speech, revealing interesting modifications in the glottal behavior when producing Lombard speech, various voice qualities, or hypo/hyperarticulated speech.

As mentioned above, reliably and accurately estimating the glottal source from speech recordings is a complex issue. This usually requires to process speech frames synchronized on glottal closure instants and whose length is proportional to the pitch period. For this, three of the most efficient approaches are the following [1]. The Closed Phase Inverse Filtering (CPIF, [2]) method computes an estimation of the vocal tract response during the glottal closed phase, during which the effects of the subglottal cavities are minimized. The Iterative Adaptive Inverse Filtering (IAIF, [3]) technique is based on an iterative refinement of both the vocal tract and the glottal components in order to improve the quality of the estimates. Finally the Mixed-Phase Separation (MPS, [4]) approach is a non-parametric technique which relies on the causal/anticausal properties of speech. More precisely, it isolates the anticausal component of speech as it corresponds to the glottal open phase. In [1], we have shown that CPIF and MPS lead to the most efficient estimation of the glottal flow for clean recordings.

From the resulting estimates of the glottal flow, several features can be extracted [1]. In the time domain, we found out that the Normalized Amplitude Quotient (NAQ) and the Quasi Open Quotient (QOQ) are amongst the most suited glottal characteristics. In the spectral domain, the Harmonic Richness Factor (HRF) and the ratio between the two first harmonic amplitudes (H1-H2) provided an efficient description of the glottal source [1].

Based on the conclusions drawn here above, several types of expressive were analyzed.

**Lombard Speech:** The Lombard effect refers to the speech changes due to the immersion of the speaker in a noisy environment. In such a context, the speaker tends to modify its way of uttering so as to maximize the intelligibility of its message. We have shown in [5] that this is also reflected by a significant modification of the glottal behavior. Important variations of the glottal parameters were observed, depending on the level and type of the noise. For example in a factory noise of 84 dB, NAQ is decreased by 26.4%, QOQ by 12.6%, H1-H2 by 2.9dB and HRF is increased by 4.1dB.

**Voice Quality:** Our study was here led on a database where the same speaker produces modal, but also soft and loud voice. It was shown in [1] that when the vocal effort becomes stronger, NAQ and H1-H2 are significantly decreased while HRF is consistently increased.

**Hypo/Hyperarticulated Speech:** For hyperarticulated voice, speech clarity tends to be maximized, while hypoarticulation refers to speech produced with minimal efforts. We have shown in [6] that the stronger the degree of articulation, the higher the glottal formant frequency, the maximum voiced frequency and the fundamental frequency.

### References

- [1] T. Drugman, B. Bozkurt, T. Dutoit, *A Comparative Study of Glottal Source Estimation Techniques*, Computer Speech and Language, accepted for publication.
- [2] D. Wong, J. Markel, A. Gray, *Least squares glottal inverse filtering from the acoustic speech waveform*. IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 27, pp. 350-355, 1979.
- [3] P. Alku, J. Svec, E. Vilkmán, F. Sram, *Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering*, Speech Comm., vol. 11, pp. 109-118, 1992.
- [4] T. Drugman, B. Bozkurt, T. Dutoit, *Causal-Anticausal Decomposition of Speech using Complex Cepstrum for Glottal Source Estimation*, Speech Comm., accepted for publication.
- [5] T. Drugman, T. Dutoit, *Glottal-based Analysis of Lombard Speech*, Proc. Interspeech, pp. 2610-2613, 2010.
- [6] B. Picart, T. Drugman, T. Dutoit, *Analysis and Synthesis of Hypo and Hyperarticulated Speech*, Proc. SSW7, pp. 270-275, 2010.