

Glottal area waveforms study from high speed video-endoscopic recordings and voice production model with aeroacoustic coupling driven by a forced glottal folds model

Thomas Hézard, Thomas Hélie, Boris Doval, René Caussé, Gilles Degottex

▶ **To cite this version:**

Thomas Hézard, Thomas Hélie, Boris Doval, René Caussé, Gilles Degottex. Glottal area waveforms study from high speed video-endoscopic recordings and voice production model with aeroacoustic coupling driven by a forced glottal folds model. 9th Pan European Voice Conference (PEVOC 9), Aug 2011, Marseille, France. hal-01134014

HAL Id: hal-01134014
https://hal.science/hal-01134014

Submitted on 21 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Glottal area waveforms study from high speed video-endoscopic recordings and voice production model with aeroacoustic coupling driven by a forced glottal folds model

Thomas Hézard[(1)], Thomas Hélie[(1)], Boris Doval[(2)], René Caussé[(1)], Gilles Degottex[(3)]

[(1)] Institut de Recherche et de Création Acoustique-Musique (IRCAM), Paris, France
[(2)] Lutherie-Acoustique-Musique (LAM) - Institut Jean Le Rond d'Alembert (IJLRA), Paris, France
[(3)] University of Crete - Computer Science Department, Heraklion, Crete, Greece
thomas.hezard@ircam.fr, thomas.helie@ircam.fr, boris.doval@upmc.fr, rene.causse@ircam.fr, degottex@csd.uoc.gr

Source-filter paradigm is one of the most common approaches used by the scientific community to establish models for analysis and synthesis of voice sounds. But, if the source-filter models are easy to compute, they ignore some physical phenomena which are important for the naturalness of the sounds. In this presentation, we investigate how we can modify the usual source-filter model in order to take into account some simplified aeroacoustic coupling between vocal folds and vocal tract while keeping low-cost computation and efficient analysis methods. After presenting a glottal area waveform model, we establish a complete voice model based on the preceding glottal area waveform model and evaluate its relevance by comparing analysis and synthesis methods based on this model and on a usual source-filter model.

1. A glottal area waveform model based on high speed video-endoscopic recordings

The most used source in the source-filter models of voice production is obviously the Liljencrants-Fant model, or LF model [1]. As the LF model is originally a glottal flow model, this waveform model appears to be also relevant to model glottal area waveforms. A study on a database of glottal area waveforms extracted from high speed video-endoscopic recordings highlighted that most of them can be accurately approximated with an LF model. In our lecture, we present these results with some examples from the database using the one-parameter LF-Rd model [2]. As the LF-Rd model has only one parameter, it can't represent the diversity of waveforms we observe in phonation and we'll introduce some other interesting glottal area waveform models.

2. Voice production model

In this second part, we use the previous observations to establish a simplified aeroacoustic vocal folds model. The vocal folds geometry is reduced to an open area whose temporal dynamics is driven by an LF-Rd model. The aeroacoustic coupling between the vocal folds and the vocal tract is ensured by a simplified jet model based on standard Bernoulli equations under quasi-stationarity hypothesis and the continuity of the acoustic pressure and flow at the interface between the jet and the vocal tract, a basic model commonly used for voice and wind instruments modeling [3]. This model, driven by a subglottal pressure p_{sub}, is described by

$$\left\{ \begin{array}{rcl} p_{sub} & = & p_G(t) + \frac{1}{2}\rho {v_G}^2(t) \\ u_G(t) & = & A_G(t) v_G(t) \end{array} \right. \tag{1}$$

$$\left\{ \begin{array}{rcl} p_G(t) & = & p(0,t) \\ u_G(t) & = & u(0,t) \end{array} \right. \tag{2}$$

where $p_G(t)$, $u_G(t)$ and $v_G(t)$ are the pressure, flow and speed inside the glottis, ρ is the air density, $A_G(t)$ is the forced glottal area and $p(0,t)$ and $u(0,t)$ are the pressure and the flow at the inner end of the vocal tract. Finally, the vocal tract is modeled with an usual auto-regressive (AR) filter [4].

3. Signal analysis, model inversion and validationl

An analysis method for source-filter model with LF-Rd source and AR filter has been developed at IRCAM [5]. As we use a similar source-filter model in which we added a jet model, we adapt this analysis method to our model. We can then evaluate our model and the improvements brought by the aeroacoustic coupling by comparing the analysis and re-synthesis results of our model and the LF-Rd / AR model.

References

[1] Fant G, Liljencrants J and Lin Q. *A four-parameter model of glottal flow*. STL-QPSR 4 : 1-13, 1985.

[2] G. Fant. *The LF-model revisited. Transformations and frequency domain analysis*. STL-QPSR, 36 : 119–156, 1995.

[3] Vilain C. *Contribution à la synthèse de parole par modèle physique. Application à l'étude des voix pathologiques*. PhD Thesis, Institut National Polytechnique de Grenoble, 2002.

[4] J.D. Markel and A.H. Gray. *Linear Prediction of Speech*. Springer Verlag, 1976.

[5] G. Degottex. *Glottal source and vocal-tract separation. Estimation of glottal parameters, voice transformation and synthesis using a glottal model*. PhD Thesis, Universite Paris VI - Pierre et Marie Curie, 2010.