Analysis of voiced fricative production using Videoendoscopy: *towards a model of the voicing offset mechanism*

Cátia M. R. Pinho[1], Luis M.T. Jesus[2], Anna Barney[3]

[1] IEETA, Universidade de Aveiro, Portugal; [2] IEETA and ESSUA, Universidade de Aveiro, Portugal; [3] ISVR, University of Southampton, UK
catiap@ua.pt, lmtj@ua.pt, ab3@soton.ac.uk

ieeta instituto de engenharia electrónica e telemática de aveiro

UNIVERSITY OF Southampton Institute of Sound and Vibration Research

CPMSP2 - 2010

Cognitive and Physical Models of Speech Production, Speech Perception and Production-Perception Interaction

Berlin 27th September – 1st October, 2010

Introduction

- Voicing is often maintained over only part of a voiced fricative [1]. This makes these sounds ideal for the study of onset and maintenance of voicing, a subject not investigated in great detail to date.
- During voiced speech the source of sound production arises from the vibrations of the vocal folds within the larynx. In order to terminate these vibrations it is necessary to change the mechanical properties of the folds so that the transglottal pressure drop is no longer sufficient to sustain voicing, or to reduce the pressure drop across them to below a threshold level [2].
- Videoendoscopic digital images acquired during speech production may help us extract some key features that can track the mechanical properties of the folds during VFV sequences.

Method

- Analysis of videoendoscopy images during the production of voiced fricatives /v, z, Z/ in middle word position.
- Extraction of key features and parameters related with vocal fold behaviour during the production of voiced fricatives, in a way that allows us to address some VT modelling questions:
  - How does voicing stop?
  - What is/are the articulatory mechanism(s) for voicing offset (in running speech);
  - Can we estimate unilateral vocal fold paralysis (UVFP) glottal source function to help with speech therapy?
- Currently developing an adequate theoretical modelling framework and testing new algorithms.
- The time resolution of 25 fps (a frame/image every 40 ms, i.e., one image contains 40 ms of vibratory information of the vocal folds) was not high enough to analyse with detail the behaviour of the vocal folds during the open-close cycles of speech production. We can see in Figure 1 that each frame corresponds to ≈ 5 open-close cycles of speech production.
- A summary of the methodology used to analyse and segment the key features of the digital images is described in the block diagram presented in Figure 2.

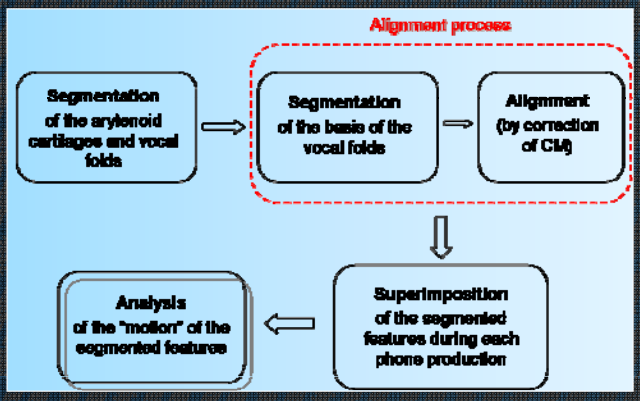

Fig. 2: Block diagram that describes the main steps used in the segmentation and analysis of the digital images during the VFV production.

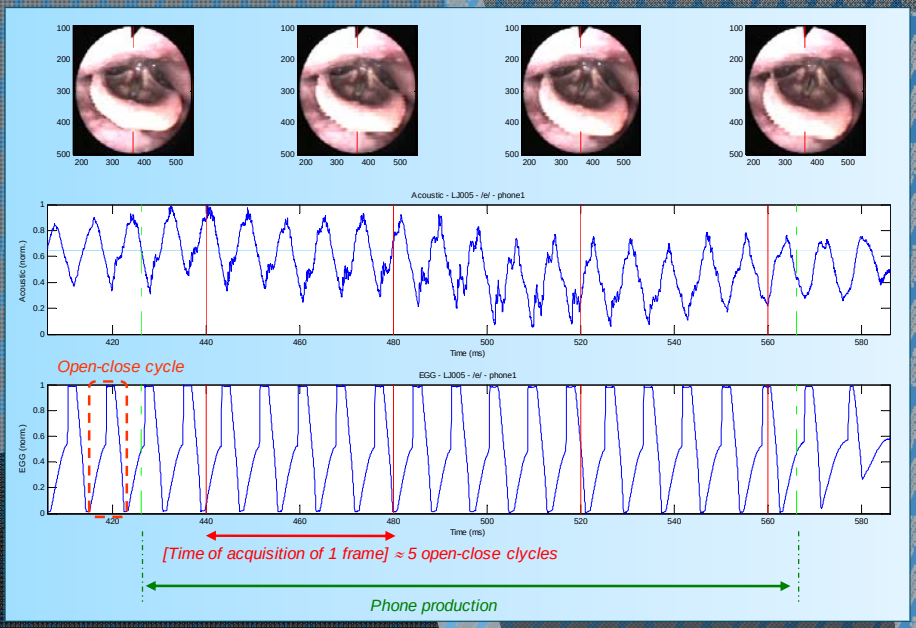

Fig. 1: From top to bottom: Digital video images of the vocal folds + audio signal + EGG signal of the production of the vowel /e/ (phone1) preceding the fricative /z/, speaker LJ – Corpus file 008. *(Although the EGG signal is clipped, this does not hinder our analysis, because the important parameter to extract is the width of the open-close cycle and not the height).*

Results

- An initial study using videoendoscopy to image the vocal folds during the production of voiced fricatives, a class of phoneme prone to devoicing, embedded in carrier phrases has given us a limited view of the voicing offset mechanism.
- The two main ideas that we are working on are:
  - When the arytenoids are rotating and the glottal width stays more or less the same, we believe that the stiffness in the vocal folds is increased.
  - When the arytenoids are opening together with the vocal folds it is likely that change to the transglottal pressure is the primary mechanism of voicing offset.
- However, to investigate these phenomena in more depth we require a greater acquisition frame rate to allow us to visualise the behaviour at several instants within the same oscillation cycle and therefore track the evolution of the articulatory gestures.
- An imaging technique is required that allows high speed data capture, producing images of the relevant articulatory structures (in particular the arytenoid cartilages and vocal folds) during running speech.
- We are interested to determine whether video kymography allows us to acquire the data we need (or if possible use high-speed videoendoscopy).

Future work

- Simulations of temporal patterns of oral airflow using a two-mass model (2MM) of the vocal folds under dynamic control [3] and modelling the different cases using an adapted version of 2MM of [4] allows dynamic change of pressure and/or mechanical properties of the folds.
- We also plan to apply the features and parameters extracted from the videoendoscopy high speed images in the two-mass model (2MM) [4], in a way to better mimic the signal produced during different VFV productions.

Acknowledgments

This work was supported by Fundação para a Ciência e a Tecnologia, Portugal (Research and Development Project PTDC/SAU-BEB/67384/2006 FCOMP-01-9124-FEDER-00747) and Association Francophone de la Communication Parlée (AFCP).

References

[1] L. M. T. Jesus and C. H. Shadle, "A Parametric Study of the Spectral Characteristics of European Portuguese Fricatives," Journal of Phonetics, vol. 30, pp. 437-464, Jully 2002.

[2] A. Barney, et al., "Investigation of the mechanisms of voicing onset," Journal of the Acoustical Society of America, vol. 123(5), p. 3576, 2008.

[3] J. C. Lucero and L. L. Koenig, "Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control," Journal of the Acoustical Society of America, vol. 117, pp. 1362-1372, March 2005.

[4] N. J. C. Lous, et al., "A symmetrical two-mass vocal-fold model coupled to vocal tract andtrachea, with application to prosthesis design," Acta Acustica united with Acustica, vol. 84, pp. 1135-1150, November/December 1998.