Unified numerical simulation of the physics of voice. The EUNISON project.

Oriol Guasch[1], *Sten Ternström*[2], *Marc Arnela*[1] *and Francesc Alías*[1]

[1]GTM Grup de recerca en Tecnologies Mèdia,
La Salle, Universitat Ramon Llull, Barcelona, Catalonia, Spain
[2]Department of Speech, Music and Hearing, Kungliga Tekniska Hgskolan, Stockholm, Sweden

oguasch@salleurl.edu, stern@csc.kth.se, marnela@salleurl.edu, falias@salleurl.edu

Abstract

In this demo we will briefly outline the scope of the european EUNISON project, which aims at a unified numerical simulation of the physics of voice by resorting to supercomputer facilities, and present some of its preliminary results obtained to date.

Index Terms: Voice production, finite element methods, physics of voice

1. Demo description

Several strategies have been followed in the past decades to simulate the human voice. These usually focus on the final goal of generating a realistic acoustic signal of voice, making whatever simplifications may be necessary for it. For instance, current commercial speech synthesis systems are based on concatenation of pre-recorded audio segments, whereas many articulatory physics-based approaches rely on one-dimensional formulations that need to incorporate several tricks, so as to emulate three dimensional (3D) voice effects.

However, the amazingly growing capacity of computers combined with intense research on numerical mathematics, has opened the door to go one step beyond. It is at the core of the European FET project EUNISON (Extensive UNIfied-domain SimulatiON of the human voice, http://www.eunison.eu), to build a new voice simulator which, from given inputs representing topology or muscle activations or phonemes, it will render the complex 3D physics of the voice, including of course its acoustic output. This will give important insights into how the voice works, and how it fails.

The physics of voice is extremely intricate. It involves complex interactions between laminar and turbulent airflow; vibrating, deforming, colliding elastic solids; and sound waves resonating in a contorting, dynamic duct. The simulation of these phenomena demands tailored numerical strategies which will be accounted for within EUNISON by adopting and developing new stabilized finite element methods (FEM) to be implemented in parallel platforms. Yet, no good progress can be made without testing the numerical simulations against experimental data. Another crucial aspect of the project will be then to perform

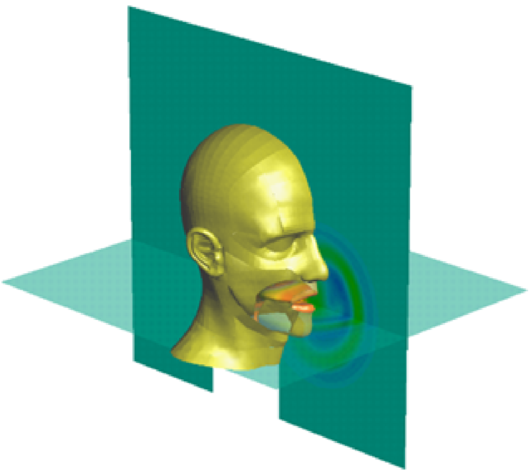

Figure 1: *Finite Element simulation for vowel /a/. Courtesy of La Salle R&D, Universitat Ramon Llull, Barcelona, Catalonia, Spain.*

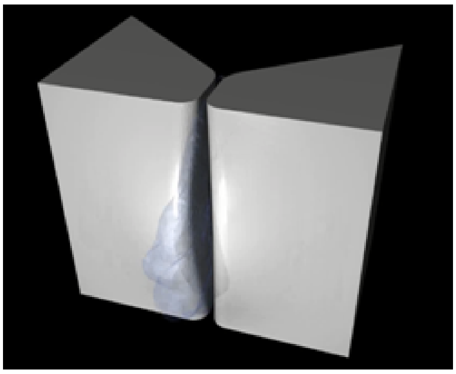

Figure 2: *Vocal folds numerical simulation. Courtesy of the Computational Technology Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden.*

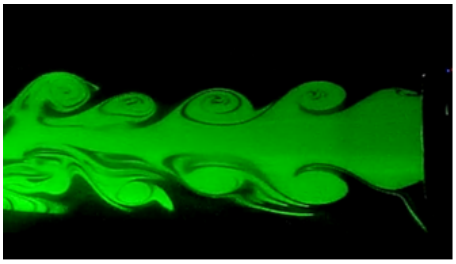

Figure 3: *Experimental flow for a fricative sound. Courtesy of the GIPSA-lab, Centre National de la Recherche Scientifique (CNRS), Grenoble, France.*

experimental investigation, which will not only serve for testing purposes, but also to better understand several mechanisms of voice production related e.g., to the generation of fricatives, plosives and syllables.

In this demo we will present some preliminary results attained at the initial steps of EUNISON (2013-2016). These will consist of several animations and figures corresponding to simulations of some particular phenomena that ideally, by the end of the project, will be integrated in a single framework. In Fig.1 we can observe a snapshot of acoustic waves emanating from a realistic real head when pronouncing vowel /a/. Fig.2 presents a snapshot of the self-oscillation of vocals folds driven by air emanating from lungs, which already includes the possibility of vocal folds contact. In Fig.3, we show the experimental visualization of flow corresponding to a fricative consonant.

The final goal of EUNISON is not developing a text-to-speech synthesis system, but rather a voice simulation engine, with many applications; given the right controls and enough computer time, it could be made to speak in any language, or sing in any style. The model will be operable on-line, as a reference and a platform for others to exploit in further studies. The long-term prospects include more natural speech synthesis, improved clinical procedures, greater public awareness of voice, better voice pedagogy and new forms of cultural expression.

2. Acknowledgements

This research is supported by EU-FET grant EUNISON 308874.