Proceedings of ISSP 2017, October 16-19, 2017, Tianjin, China, pp 87-89, http://www.issp2017.org.cn/index.html

Some current challenges in unified numerical simulations of voice production: from biomechanics to the emitted sound

Oriol Guasch[1] and Sten Ternström[2]

[1]GTM-Grup de Recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull
C/ Quatre Camins 30, 08022 Barcelona, Catalonia, Spain
e-mail: oguasch@salle.url.edu
[2]Department of Speech, Music and Hearing, School of Computer Science and Communication
KTH Royal Institute of Technology SE-100 44, Stockholm, Sweden
e-mail: stern@kth.se

Voice production all the way from muscle activation to sound - are we there yet? Three-dimensional (3D) numerical simulations of the entire process of voice generation appear to be very challenging. Muscle activations position the articulators, which define a vocal tract geometry and posture the vocal folds. Air emanating from the lungs induces self-oscillations of the vocal folds, which result in aeroacoustic sources and the subsequent propagation of acoustic waves inside the vocal tract (VT). There, many things could happen. For instance, the air could resonate to generate vowels, or, at constrictions, airflow may be accelerated to create turbulent sounds such as fricatives. The vocal tract walls are flexible and react to the inner acoustic pressure. Also, articulators can change the vocal tract geometry to generate vowel-vowel utterances or syllables. Sound is finally radiated from the mouth.

Attempting unified 3D numerical simulations of all the above processes, which involve coupling of a biomechanical model and the mechanical, fluid and acoustic fields, may seem unwise. Most research to date has addressed a few selected aspects of voice production. Unified approaches have been shunned for their daunting complexity and high-performance parallel computation requirements. This situation now seems to be changing. In this paper, we briefly review recent approaches towards 3D realistic voice simulation that unify, at least to some extent, some of the involved physical fields. Remaining challenges will be highlighted. We will focus on those works which end with the production of a given sound, thus leaving aside the huge amount of literature solely devoted to the complex simulation of phonation.

Probably, the easiest sound to generate is a vowel. If one has for example, a detailed MRI geometry of its corresponding VT, the finite element method (FEM) can be used to solve the wave equation and produce such a vowel, by prescribing a glottal train model at the glottis section. From a computational point of view this poses no excessive difficulties. The Laplacian operator is well behaved, so all that remains is to use an accurate scheme for the time discretization to cover all the audible frequency range. An associated problem is that of preventing waves radiated outside the mouth from reflecting at the boundaries of the computational domain. This can be achieved by imposing non-reflecting boundary conditions or by making use of perfectly matched layers. Yet, the production of a vowel in such a way involves only the acoustic pressure field. Recently, a one-way coupling with biomechanics has been proposed. Instead of using an MRI geometry for the VT, an ArtiSynth biomechanical model is employed to position the articulators to generate the vowel. Then, an automatic cavity extraction method is applied to obtain the VT geometry, which is used as the computational domain in a FEM software to produce the corresponding vowel sound.

One could proceed similarly for vowel-vowel (VV) utterances. However, in this case it is no longer possible to resort to the irreducible wave equation. Wave propagation takes place in a moving VT that

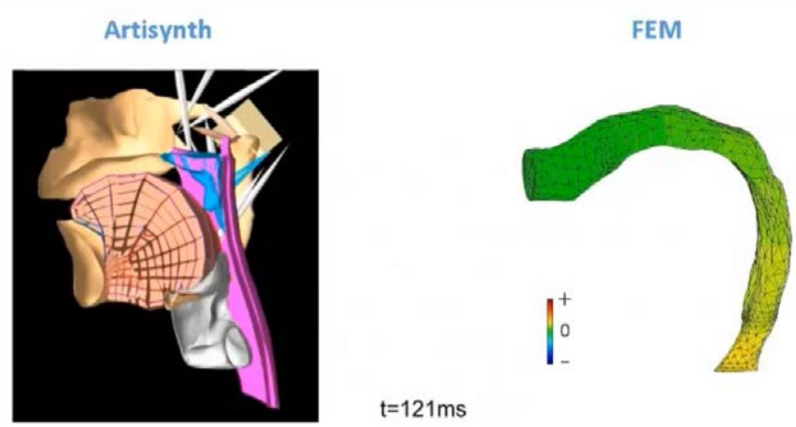

Figure 1: Snapshot of the generation of the VV utterance [ai]. On the left, the biomechanical model in Artisynth employed to define the geometry of the dynamic vocal tract cavity. On the right, the finite element model used to generate the sound [ai].

(Video available online at: www.youtube.com/user/eunisonFet).

evolves from the geometry of the initial vowel to that of the final one. Therefore, the linearized momentum and continuity equations that drive acoustic waves (sometimes referred to as the mixed wave equation), need to be expressed in an arbitrary Lagrangian-Eulerian (ALE) frame of reference, moving with the VT computational mesh. Solving the mixed wave equation poses more difficulties than solving its irreducible counterpart. The inf-sup condition that guarantees its well-posedness (in the Hadamard sense) does not transmit from the continuous weak form to the discrete one. Numerical stabilization techniques become necessary to use equal polynomial interpolations for the pressure and velocity fields. When simulating a VV utterance, one might simply interpolate between the geometries of the initial and final vowels to obtain the evolving VT. A recent improvement, however, is the achievement of a link with biomechanics, such that the trajectory between vowels is again driven by a biomechanical model in ArtiSynth (see Fig. 1).

Other links between physical fields in voice production have been established. The production of a sound, like [s], is essentially an aeroacoustic phenomenon which links the fluid and acoustic fields. The sibilant fricative [s] is generated when the turbulent jet passing between the tongue blade and the hard palate is deflected downwards through the incisors' gap, and impinges the inter-space between the lower incisors and the lower lips. The eddies in that region become an aeroacoustic source of sound, which in turn is diffracted by the upper incisors and radiated outwards. Ideally, one could recover all the physics from a numerical simulation of the compressible Navier-Stokes equations. Yet in practice, solving the latter at low Mach numbers is a very stiff problem. Consequently, one usually resorts to hybrid approaches in which the incompressible Navier-Stokes equations are first solved to compute the acoustic source terms, and those become input in a wave operator that yields the acoustic field. The price to be paid is that the coupling between the flow and the acoustics is no longer bidirectional, because there is no feedback from the acoustics to the flow dynamics. The Navier-Stokes equations are non-linear, time dependent and an example of a mixed problem requiring stabilized numerical strategies. At present, solving them in 3D inevitably requires large supercomputer facilities.

Finally, coupling of the mechanical, fluid and acoustic fields has also been attempted. Large and complex simulations of the fluid-structure problem of vocal fold self-oscillations, including contact and wave propagation inside the VT, have been recently reported by some authors. In general, some simplifications are assumed, for instance by prescribing the vocal fold movement in the acoustics, or by simplifying the vocal fold constitutive equations, or the contact model. Promisingly, it has already been possible to reproduce correctly the first formant positions of vowels in this way.

To our knowledge, a full unified simulation of a voice sound, comprising from the biomechanical model, to the vocal fold self-oscillations with contact, to the acoustic field, has not yet been carried out. However, this feat might be well within reach in the near future.