PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Finite element simulation of /asa/ in a three-dimensional vocal tract using a simplified aeroacoustic source model

Marc Arnela(1) and Oriol Guasch(1)

(1)GTM Grup de recerca en Tecnologies Mèdia, La Salle - Universitat Ramon Llull, Barcelona, Catalonia, Spain,
marc.arnela@salle.url.edu, oriol.guasch@salle.url.edu

Abstract

The numerical simulation of fricative sounds in three-dimensional (3D) vocal tracts typically involves hybrid Computational Aeroacoustics (CAA) approaches. Unfortunately, those are very costly and require from supercomputer facilities to produce a mere few milliseconds of sound. The problem becomes prohibitive if one not only aims at generating a single fricative, but also sequences containing both, vowels and fricatives. It then seems wise to try to approximate somehow the flow noise sources, so that only an acoustic simulation becomes necessary. That avoids the very demanding computational fluid dynamics step in CAA. Aeroacoustic sources can be modeled to different levels of precision. In this work, it is suggested to follow a similar methodology to that in one-dimensional (1D) techniques, but applied to 3D dynamic vocal tracts. Vowel sounds are produced introducing glottal pulses at the vocal tract entrance (glottis), while monopole and dipole sources consisting of white noise are activated in the region where a fricative sound is generated. Acoustic wave propagation in a 3D dynamic vocal tract is simulated using a stabilized Finite Element Method (FEM) for the wave equation in mixed form, set in an Arbitrary Lagrangian-Eulerian (ALE) framework. The sequence /ɑsɑ/ is produced as an example.

Keywords: vocal tract acoustics, Finite Element Method, vowels, fricatives

1 INTRODUCTION

Intensive research on three-dimensional (3D) vocal tract acoustics has been carried out in recent years. Vowel sounds have extensively been studied with a large variety of models. For instance, in [1, 2] the finite element method (FEM) was used to respectively analyze the influence of the lips and vocal tract shape on the production of vowels, whereas in [3] finite differences were applied to study the 3D acoustics of side branches, such as the piriform fossae and valleculae. In [4], vowel sounds were synthesized using digital waveguide mesh (DWM) models and in [5] a multimodal theory was proposed to study the generation of higher order modes. Vowel-vowel sequences like diphthongs have also been simulated. To that goal, it does not suffice dealing with the propagation of sound waves within a static vocal tract, as for vowels, but one has to also consider that the vocal tract is moving. This problem was first solved using FEM and deforming simplified vocal tracts with circular cross-sections in [6]. It was then extended to complex vocal tract geometries based on Magnetic Resonance Imaging (MRI) in [7]. Besides, a DWM model was recently used in [8] to deal with dynamic MRI-based vocal tracts, but instead of distorting the computational domain accordingly, it was proposed to fix it and account for the vocal tract dynamics through impedance variations of the computational domain.

The generation of fricative sounds is much more intricate than the production of vowels because resorting to the linear acoustic wave equation is not enough. In the particular case of sibilant /s/, a channel constriction is formed between the tongue and the palate, which accelerates the airflow emanating from the glottis. A jet develops which impinges on the teeth, resulting in strong turbulent vortices that emit aerodynamic noise. Therefore, the production of a sibilant sound is essentially an aeroacoustic phenomenon which requires solving at the same time the flow dynamics and the sound that generates. To that purpose, one can resort to hybrid Computational Aeroacoustics (CAA) approaches in which the incompressible Navier-Stokes equations are first resolved and then an acoustic analogy is employed to get the sound field [9, 10, 11]. Unfortunately, that process is computationally very demanding and supercomputer facilities are needed to get only a few milliseconds of sound. The production of a syllable like /sɑ/, or the pseudoword /ɑsɑ/, are totally out of the range of current computers if one follows conventional CAA strategies.

In this paper, we propose an alternative for the 3D generation of /ɑsɑ/, inspired on 1D approaches (see e.g., [12, 13]). The key to being able to synthesize that pseudoword with a regular desktop computer consists in approximating the aeroacoustic source for /s/ and thus avoid solving the non-linear Navier-Stokes equations. Only the linear wave equation in mixed form in a moving framework becomes then necessary to generate /ɑsɑ/. There exist several options to simulate the aeroacoustic source for /s/. One could consist in building a quadrupole sound source distribution based on Kirchoff's spinning vortices that emulates Lighthill's source term [14]. An easier one, borrowed from 1D strategies and which we have followed herein, is that of employing monopole and dipole Gaussian noise sources. Those are introduced immediately downstream of the constriction and in the vicinity of the teeth. They respectively account for volume pulsations at the constriction exit and for force fluctuations at the teeth surface.

The sequence /ɑsɑ/ has been simulated in this work using simplified 3D vocal tract geometries made of circular cross-sections in a straightened midline. These geometries are constructed from the 1D area functions in [15], and deformed from /ɑ/ to /s/ and back to /ɑ/ by means of linear interpolation, as done in [6] and [16] for 3D and 2D simplified vocal tracts. It is worthwhile mentioning that the role of monopole sources in the 3D production of static /s/ was recently investigated in [17] using a multimodal approach.

The paper is structured as follows. Section 2 presents the methodology proposed to simulate /ɑsɑ/ with FEM, Section 3 shows the obtained results and, finally, Section 4 closes the paper with the conclusions.

2 METHODS

2.1 Problem statement

Acoustic waves propagating in a static 3D vocal tract can be simulated, in the corresponding computational domain Ω, by numerically solving the acoustic wave equation in mixed form, which is nothing but the linearized continuity and momentum equations for sound propagation in a stationary inviscid fluid,

$$\frac{1}{\rho_0 c_0^2}\partial_t p + \nabla\cdot\boldsymbol{u} = q \qquad \text{in } \Omega,\ t>0, \tag{1a}$$

$$\rho_0\partial_t\boldsymbol{u} + \nabla p = \boldsymbol{f} \qquad \text{in } \Omega,\ t>0. \tag{1b}$$

In Equation (1), $p(\boldsymbol{x},t)$ stands for the acoustic pressure, $\boldsymbol{u}(\boldsymbol{x},t)$ is the acoustic particle velocity, c_0 the speed of sound, ρ_0 the air density, and ∂_t the partial time derivative. $q(\boldsymbol{x},t)$ and $\boldsymbol{f}(\boldsymbol{x},t)$ respectively represent a volume source distribution and an external body force per unit volume. These two terms are usually set to zero in the simulation of vowel sounds (see e.g., [3, 6]), but in this work they will be used to generate the sibilant /s/. The irreducible wave equation is recovered as usual from (1) taking

$$\rho_0\partial_t(1a) - \nabla\cdot(1b), \tag{2}$$

which yields

$$\frac{1}{c_0^2}\partial_{tt}^2 p - \nabla^2 p = \rho_0\partial_t q - \nabla\cdot\boldsymbol{f}. \tag{3}$$

Observe that the first and second terms in the r.h.s. of Eq. (3) respectively describe the time derivative of a volume velocity and the divergence of a force, which are nothing else than standard monopole and dipole sound source distributions (see e.g., [18]). Although quadrupole in nature, a proper modeling of q and $\boldsymbol{f}$ will allow us to emulate turbulent sources within the vocal tract, and thus mimic the sound produced by a fricative.

The next step to generate a sequence like /ɑsɑ/ is that of considering dynamic vocal tracts. This means that the propagation of acoustic waves has to be simulated within a computational domain $\Omega(t)$ whose points move

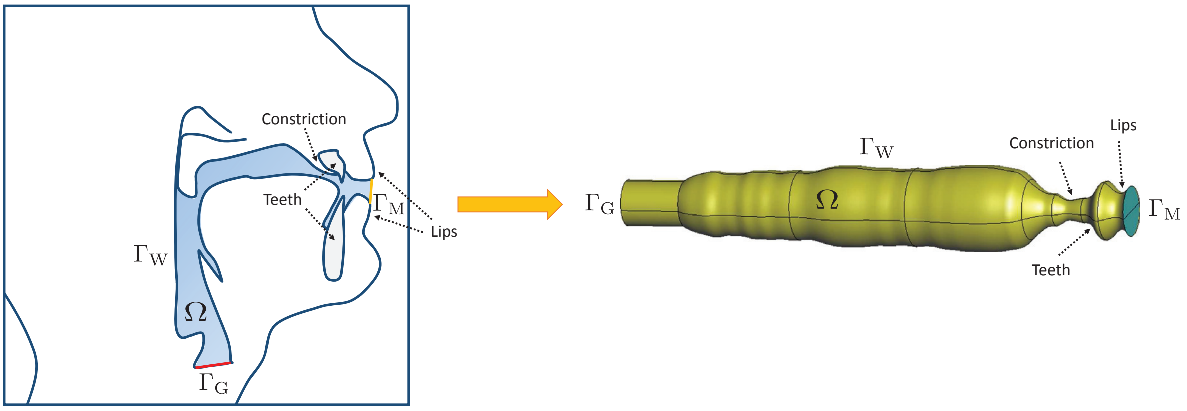


Figure 1. Sketch representing the articulation of fricative /s/ in the midsagittal plane (left) and the corresponding simplified 3D vocal tract Ω (right). Γ_G stands for the glottal cross-sectional boundary, Γ_W for the vocal tract walls and Γ_M for a fictitious boundary that closes the vocal tract at the mouth.

at a certain velocity $\boldsymbol{u}_\mathrm{dom}(\boldsymbol{x},t)$. In such a situation, the mixed wave equation (1) has to be expressed in an Arbitrary Lagrangian-Eulerian (ALE) frame of reference that moves with $\Omega(t)$, rather than in a static one. In particular, a quasi-Eulerian ALE approximation has been adopted in this work, in which the time derivative ∂_t is replaced by $\partial_t - \boldsymbol{u}_\mathrm{dom}\cdot\nabla$ in Eq. (1) to account for the domain motion. This results in the ALE mixed wave equation (see [6]),

$$\frac{1}{\rho_0 c_0^2}\partial_t p - \frac{1}{\rho_0 c_0^2}\,\boldsymbol{u}_\mathrm{dom}\cdot\nabla p + \nabla\cdot\boldsymbol{u} = q \qquad \text{in } \Omega(t),\ t>0, \tag{4a}$$

$$\rho_0\partial_t \boldsymbol{u} - \rho_0\,\boldsymbol{u}_\mathrm{dom}\cdot\nabla\boldsymbol{u} + \nabla p = \boldsymbol{f} \qquad \text{in } \Omega(t),\ t>0. \tag{4b}$$

Equation (4) must be supplemented with proper boundary and initial conditions for voice production. We have considered the following ones (see Figure 1)

$$\boldsymbol{u}\cdot\boldsymbol{n} = g(t) \qquad \text{on } \Gamma_\mathrm{G}(t),\ t>0, \tag{5a}$$

$$\boldsymbol{u}\cdot\boldsymbol{n} = p/Z_w \qquad \text{on } \Gamma_\mathrm{W}(t),\ t>0, \tag{5b}$$

$$p = 0 \qquad \text{on } \Gamma_\mathrm{M}(t),\ t>0, \tag{5c}$$

$$p = 0,\ \boldsymbol{u} = 0 \qquad \text{in } \Omega(t),\ t=0. \tag{5d}$$

Equation (5a) imposes an acoustic particle velocity $g(t)$ on the glottal cross-section $\Gamma_\mathrm{G}(t)$, so as to introduce the sound waves generated by the vocal folds within the vocal tract. Equation (5b) emulates losses in the vocal tract walls $\Gamma_\mathrm{W}(t)$ thought the impedance Z_w and Eq. (5c) considers an open-end boundary condition at the mouth termination $\Gamma_\mathrm{M}(t)$.

The finite element method (FEM) was used to numerically solve the ALE mixed wave equation (4) with boundary conditions (5). In particular, we followed the implementation in [6], which uses subgrid scales as a stabilization strategy for the numerical scheme.

2.2 Simplified aeroacoustic source model

Frication noise is generated when the airflow emerging from a supraglottal constriction forms a jet that impinges an obstacle. In the case of /s/, the obstacle are the teeth and the constriction channel is that formed between the tongue and the palate (see Figure 1). This physical phenomenon produces volume pulsations at the constriction exit and fluctuating forces in the vicinity of the teeth, which can respectively be related to monopole and dipole sound sources to some extent [12]. 1D approaches typically consider them as volume velocity and pressure sources represented by current and voltage sources in the case of transmission line circuit models [13].

In this work, these sound sources are directly considered in the acoustic wave equation. As shown in Section 2.1, monopoles and dipoles can easily be introduced within the vocal tract through $q(\boldsymbol{x},t)$ and $\boldsymbol{f}(\boldsymbol{x},t)$ in the r.h.s of the ALE mixed wave equation (4). A first approximation would be to treat them as point sources, that is to say

$$q(\boldsymbol{x},t) = w_q(t)\delta(\boldsymbol{x}), \tag{6}$$

$$\boldsymbol{f}(\boldsymbol{x},t) = \boldsymbol{w}_f(t)\delta(\boldsymbol{x}), \tag{7}$$

where $\delta(\boldsymbol{x})$ is the Dirac delta function used to locate a sound source in a particular point of the vocal tract. $w_q(t)$ and $\boldsymbol{w}_f(t)$ are two amplitude functions that mimic the turbulent noise generated by the two types of sound sources. In this work they are simply approximated by Gaussian white noise. A monopole is introduced at the constriction channel, whereas a dipole is placed at the teeth (see Figure 1).

On the other hand, note from (3) that whereas the monopole involves the scalar quantity $q(\boldsymbol{x},t)$, the dipole concerns $\boldsymbol{f}(\boldsymbol{x},t)$ which is a three-component vector describing the force variation per unit volume in a 3D domain. The three components of $\boldsymbol{f}(\boldsymbol{x},t)$ determine the dipole orientation. It is well-known from Curle's analogy that sound diffracted at a surface can be represented by means of a dipole distribution in its normal direction [18, 9]. Since the teeth are perpendicular to the streamwise jet flow, we have considered a dipole aligned in that direction, as a first rough approximation. Therefore, in a simplified straightened vocal tract, only the longitudinal component of $\boldsymbol{f}(\boldsymbol{x},t)$ differs from zero.

2.3 Simulation details

The first issue we shall address to generate /ɑsɑ/ is that of defining the motion of the vocal tract geometry. That will be based on linear interpolation between circular cross-sections set in a straight midline, whose areas have been selected to correspond to the 1D area functions in [15] for vowel /ɑ/ and sibilant /s/ (see Figure 1 for the geometry of /s/). For the FEM computations, an initial volumetric mesh is next generated. Rather than being that of an /ɑ/, the initial mesh corresponds to the vocal tract in a rest position and consists of structured tetrahedral elements of size ~ 0.2 cm, which discretize a circular tube of length 18 cm and radius 0.65 cm. Starting from such initial mesh prevents from severe element distortion during the articulation process. The initial mesh is deformed to reach the vocal tract shape of vowel /ɑ/, it then goes to /s/ and returns back to /ɑ/, to complete the sequence /ɑsɑ/. For such a process, wall node coordinates $\boldsymbol{x}^{n+1}_{\text{walls}}$ are directly obtained at a time t^{n+1} from the dynamic vocal tract representation, whereas inner node coordinates are computed through diffusion. In particular, at every time step the Laplacian equation for the node displacements $\boldsymbol{w}(\boldsymbol{x},t)$ is solved,

$$\nabla^2 \boldsymbol{w}^{n+1} = 0 \qquad \text{in } \Omega,\ t = t^{n+1}, \tag{8a}$$

with boundary conditions

$$\boldsymbol{w}^{n+1} = \boldsymbol{w}^{n+1}_{\text{walls}} = \boldsymbol{x}^{n+1}_{\text{walls}} - \boldsymbol{x}^{n}_{\text{walls}} \qquad \text{on } \Gamma_{\text{W}},\ t = t^{n+1}, \tag{8b}$$

$$\boldsymbol{w}^{n+1} \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\text{G}},\ t = t^{n+1}, \tag{8c}$$

$$\boldsymbol{w}^{n+1} \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\text{M}},\ t = t^{n+1}. \tag{8d}$$

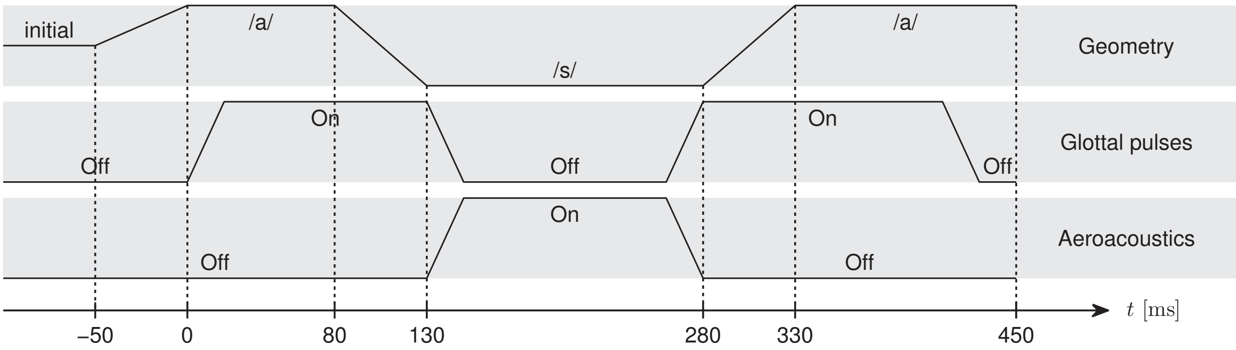


Figure 2. Time evolution scheme for the simulation of /ɑsɑ/ showing the transition times of the vocal tract geometries and the activation/deactivation of the glottal pulses and aeroacoustic models.

The second important point concerns the generation of the glottal pulses and aeroacoustic sound sources. An LF model [19] has been implemented for the former to obtain $g(t)$ in Eq. (5a). That is enhanced with a pitch curve and some shimmer and jitter, similar to what was done in [16]. Besides, and as said before, white noise has been used for the monopole and dipole sources $q(\boldsymbol{x},t)$ and $\boldsymbol{f}(\boldsymbol{x},t)$ in Eq. (4) to generate the fricative sound, as described in Section 2.2. A single dipole has been located at 1.15 cm from the mouth in the vicinity of the teeth, whereas the monopole has been placed just at the constriction exit, 1.85 cm from the mouth. The strength of the three sound sources (glottal pulses, monopoles and dipoles) has been adjusted empirically to resemble the relative levels in natural speech (see e.g., [20]).

The activation and deactivation of the glottal and aeroacoustic sources has been coordinated with the motion of the vocal tract geometry. Figure 2 summarizes the selected time instants at which sources are switched on and off. As observed and explained before, the initial geometry deforms to vowel /ɑ/, then to /s/ and back to /ɑ/, to produce the sequence /ɑsɑ/. Each geometry is respectively sustained for 80, 150 and 120 ms, with a transition time of 50 ms between them. One may think the same time intervals should work for the glottal pulses and aeroacoustic sources. However, this would lead to unconnected speech resulting in mute periods. To avoid that problem, we have followed the proposal in [20] and imposed a voicing maintenance in which glottal pulses are activated beyond the articulation of a vowel sound. This resulted in an overlapping of the glottal and aeroacoustic source models. Observe also in the figure that ascending and descending ramps of 20 ms are applied when switching on and off each source model. This allows one to better emulate the onset/offset of the glottal pulses and aeroacoustic sources. Note also that the former get deactivated 20 ms before finishing the simulation, which allows sound waves to vanish within the vocal tract.

The FEM simulations have been performed with a speed of sound of $c_0 = 350$ m/s and an air density of $\rho_0 = 1.14$ kg/m^3. The wall impedance Z_w is computed from the boundary admittance coefficient $\mu = \rho_0 c_0 / Z_w$, with $\mu = 0.005$. A sampling frequency of $f_s = 250$ kHz has been used for the time evolution of the ALE mixed wave equation, which ensures stability of the numerical schemes. In contrast, that rate is relaxed to 10 kHz for moving the vocal tract as this suffices for properly capturing its dynamics. A single simulation requires about 18 hours in a regular desktop computer (Intel(R) Core(TM) i7-6700 3.4 GHz).

3 RESULTS

Figure 3 depicts some snapshots from the FEM simulation of /ɑsɑ/, acquired at $t = 50, 100, 202, 300$ and 350 ms. Each snapshot shows the mesh and the acoustic pressure on the boundaries of the vocal tract, with the color limiting values, expressed in Pascals, adapted in each frame to enhance the visualization of acoustic waves. The figure also presents the waveform of the sequence with the red dots corresponding to the snapshot time instants. Observe in the figure that the vocal tract smoothly transitions from /ɑ/ to /s/ and back to /ɑ/, while acoustic

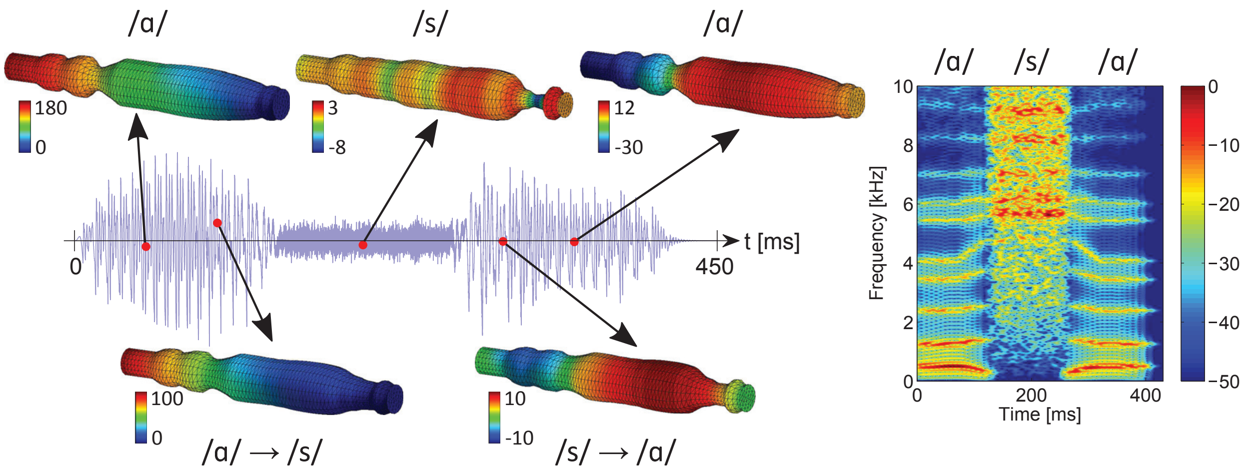
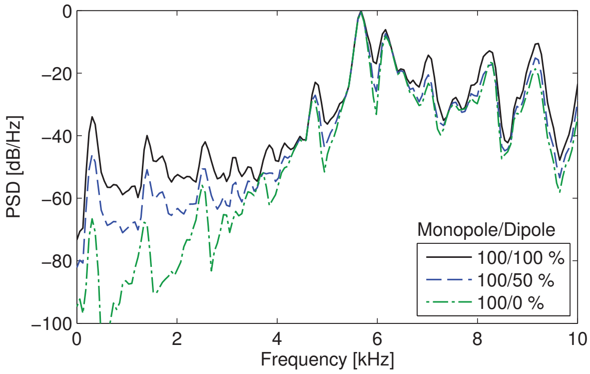


Figure 3. Evolution of the acoustic pressure for the FEM simulated sequence /ɑsɑ/ with some snapshots showing the acoustic pressure (in Pa) within the vocal tract. Each frame was acquired, from left to right, at $t = 50, 100, 202, 300, 350$ ms. Red dots are plotted on the waveform to illustrate these time instants. The spectrogram of /ɑsɑ/ is represented in the right panel (in dBs).

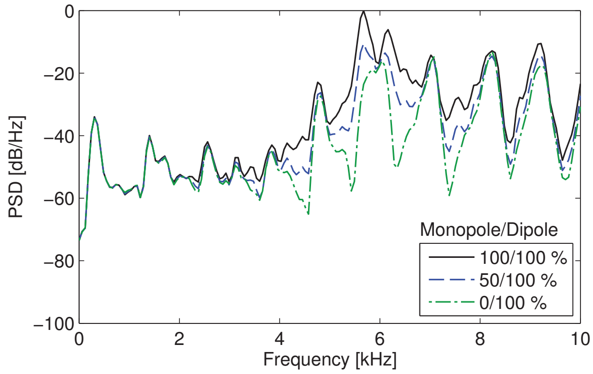
waves propagate inside it. This results in a smooth transition of the waveform between sounds. Moreover, the imposed glottal pulses for the phonation of vowels are naturally mixed with the Gaussian noise of the monopole and dipole used for /s/. In this way, we avoid performing two independent simulations, one for the glottal pulses and the other one for the aeroacoustic sources, and then superpose the resulting pressures as sometimes done in 1D approaches (see e.g., [12]).

On the other hand, the right panel of Figure 3 shows the spectrogram of the simulated /ɑsɑ/. This has been computed applying a preemphasis filter to enhance the higher frequencies, as typically done in speech analysis. Observe in the spectrogram that the typical formants of vowel /ɑ/ are generated in the first time instants, and that these start transitioning to those corresponding to the vocal tract of an /s/. However, such formants get smeared out when the glottal pulses are switched off and the monopole and dipole get activated in the articulation of /s/ (see Figure 2). This produces a significant increase of energy above 5 kHz, typically found in the production of sibilant /s/. Similar observations can be derived for the inverse transition from /s/ to /ɑ/.

Finally, it is also interesting to analyze the role played by the monopole and dipole sources in the generation of /s/. To that purpose, several /ɑsɑ/ sequences have been simulated varying the relative amplitudes of the monopole and the dipole. In particular, we have considered full, half and null activation (100, 50 and 0 %) and simulated all possible combinations that contain at least one fully operative sound source. Figure 4 presents the power spectral densities (PSD) of all the /s/ sounds extracted from various /ɑsɑ/ simulations. The black solid curve, to be taken as reference, represents a full activation of the monopole and dipole sources (this was the option adopted in this work). As seen, the energy in this configuration concentrates at the high frequency range, presenting the characteristic peak of an /s/ close to 5.6 kHz. Low frequency energy (<5 kHz) is, in average, about 25 dB lower than the higher one, which is also within typical values for the production of sibilant /s/ (see e.g., [17]). If we next check what occurs when we deactivate the dipole (see Figure 4a), we observe that it mainly contributes to the low frequency energy, although some small variations can also be observed for higher frequencies. As regards the monopole (see Figure 4b), it has the opposite behavior. The monopole essentially increases the PSD beyond 5 kHz leaving the low frequency range almost unaltered.



(a)



(b)

Figure 4. Power Spectral Density (PSD) of the /s/ considering a full (100 %), a half (50 %), and null (0 %) activation of the monopole and dipole sound sources. Values are normalized to the maximum value of the configuration with monopoles and dipoles fully activated (100/100%).

4 CONCLUSIONS

In this work, a 3D FEM approach has been presented to generate voice sequences containing vowels and fricatives. In particular, we have focused on the generation of the pseudoword /ɑsɑ/. This involves simulating 3D acoustic waves propagating in a dynamic vocal tract. To generate them, a glottal pulse model has been employed for the production of /ɑ/, whereas a simplified aeroacoustic source model based on monopoles and dipoles has been proposed for the generation of sibilant /s/. Results have shown that the spectrum of an /s/ can be properly reproduced with the proposed approach, and that the use of both, monopole and dipole sources, is essential to achieve this goal. The monopole mainly contributed to the high frequency range above 5 kHz, while the dipole governed the low frequencies. Above all, the proposed methodology has allowed for the 3D simulation of /ɑsɑ/ without needing supercomputer facilities. That would have proved impossible with conventional hybrid CAA. Future works will consider a more extensive analysis of the location, strength, and spectral content of the monopoles and dipoles, as well as the simulation of other sequences in simplified and realistic vocal tracts.

ACKNOWLEDGEMENTS

This research has been supported by the Agencia Estatal de Investigación (AEI) and FEDER, EU, through project GENIOVOX TEC2016-81107-P. The second author would also like to thank l'Obra Social de la Caixa and the Universitat Ramon Llull for their support under grant 2018-URL-IR2nQ-031.

REFERENCES

[1] Arnela, M., Blandin, R., Dabbaghchian, S., Guasch, O., Alías, F., Pelorson, X., Van Hirtum, A., and Engwall, O. Influence of lips on the production of vowels based on finite element simulations and experiments. J. Acoust. Soc. Am. 2016; 139(5):2852–2859.

[2] Arnela, M., Dabbaghchian, S., Blandin, R., Guasch, O., Engwall, O., Van Hirtum, A., and Pelorson, X. Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds. J. Acoust. Soc. Am. 2016; 140(3):1707–1718.

[3] Takemoto, H., Mokhtari, P., and Kitamura, T. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. J. Acoust. Soc. Am. 2010; 128(6):3724–3738.

[4] Speed, M., Murphy, D., and Howard, D. Modeling the vocal tract transfer function using a 3d digital waveguide mesh. IEEE/ACM Trans. Audio Speech Lang. Process. 2014; 22(2):453–464.

[5] Blandin, R., Arnela, M., Laboissière, R., Pelorson, X., Guasch, O., Van Hirtum, A., and Labal, X. Effects of higher order propagation modes in vocal tract like geometries. J. Acoust. Soc. Am. 2015; 137(2):832–843.

[6] Guasch, O., Arnela, M., Codina, R., and Espinoza, H. A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production. Acta Acust. united with Acustica 2016; 102(1):94–106.

[7] Arnela, M., Dabbaghchian, S., Guasch, O., and Engwall, O. Finite element generation of vowel sounds using dynamic complex three-dimensional vocal tracts. Proc. of 23th International Congress on Sound and Vibration (ICSV23); July 2016; Athens, Greece.

[8] Gully, A. J., Daffern, H., and Murphy, D. T. Diphthong synthesis using the dynamic 3d digital waveguide mesh. IEEE/ACM Trans. Audio Speech Lang. Process. 2018; 26(2):243–255.

[9] Guasch, O., Pont, A., Baiges, J., and Codina, R. Concurrent finite element simulation of quadrupolar and dipolar flow noise in low mach number aeroacoustics. Computers & Fluids 2016; 133:129–139.

[10] Yoshinaga, T., Nozaki, K., and Wada, S. Experimental and numerical investigation of the sound generation mechanisms of sibilant fricatives using a simplified vocal tract model. Phys. Fluids 2018; 30(3):035104.

[11] Pont, A., Guasch, O., Baiges, J., Codina, R., and Van Hirtum, A. Computational aeroacoustics to identify sound sources in the generation of sibilant /s/. Int. J. Numer. Method. Biomed. Eng. 2019; 35(1):e3153.

[12] Narayanan, S., and Alwan, A. Noise source models for fricative consonants. IEEE Trans. Audio Speech Lang. Process. 2000; 8(3):328–344.

[13] Birkholz, P., Jackèl, D., and Kröger, B. J. Simulation of losses due to turbulence in the time-varying vocal system. IEEE Trans. Audio Speech Lang. Process. 2007; 15(4):1218–1225.

[14] Pont, A., Guasch, O., and Arnela, M. Finite element generation of sibilants /s/ and /z/ using random distributions of Kirchhoff's vortices. Submitted 2019.

[15] Beautemps, D., Badin, P., and Laboissière, R. Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data. Speech Commun. 1995; 16(1):27–47.

[16] Arnela, M., and Guasch, O. Finite element synthesis of diphthongs using tuned two-dimensional vocal tracts. IEEE/ACM Trans. Audio Speech Lang. Process. 2017; 25(10):2013–2023.

[17] Yoshinaga, T., Van Hirtum, A., and Wada, S. Multimodal modeling and validation of simplified vocal tract acoustics for sibilant /s/. J. Sound Vib. 2017; 411:247–259.

[18] Howe, M. S. Theory of vortex sound, vol. 33. Cambridge University Press, 2003, p. 1–24.

[19] Fant, G., Liljencrants, J., and Lin, Q. G. A four-parameter model of glottal flow. STL-QPSR 1985; 77(4):1–13.

[20] Pape, D., L. M.T. Jesus, and Birkholz, P. Intervocalic fricative perception in european portuguese: an articulatory synthesis study. Speech Commun. 2015; 74:93–103.