Influence of lips on the production of vowels based on finite element simulations and experiments

Marc Arnela[a)]
GTM–Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, C/Quatre Camins 30, Barcelona, E-08022, Catalonia, Spain

Rémi Blandin
GIPSA-Lab, Unité Mixte de Recherche au Centre National de la Recherche Scientifique 5216, Grenoble Campus, St. Martin d'Heres, F-38402, France

Saeed Dabbaghchian
Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden

Oriol Guasch and Francesc Alías
GTM–Grup de recerca en Tecnologies Mèdia, La Salle, Universitat Ramon Llull, C/Quatre Camins 30, Barcelona, E-08022, Catalonia, Spain

Xavier Pelorson and Annemie Van Hirtum
GIPSA-Lab, Unité Mixte de Recherche au Centre National de la Recherche Scientifique 5216, Grenoble Campus, St. Martin d'Heres, F-38402, France

Olov Engwall
Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden

(Received 14 October 2015; revised 29 March 2016; accepted 4 May 2016; published online 19 May 2016)

Three-dimensional (3-D) numerical approaches for voice production are currently being investigated and developed. Radiation losses produced when sound waves emanate from the mouth aperture are one of the key aspects to be modeled. When doing so, the lips are usually removed from the vocal tract geometry in order to impose a radiation impedance on a closed cross-section, which speeds up the numerical simulations compared to free-field radiation solutions. However, lips may play a significant role. In this work, the lips' effects on vowel sounds are investigated by using 3-D vocal tract geometries generated from magnetic resonance imaging. To this aim, two configurations for the vocal tract exit are considered: with lips and without lips. The acoustic behavior of each is analyzed and compared by means of time-domain finite element simulations that allow free-field wave propagation and experiments performed using 3-D-printed mechanical replicas. The results show that the lips should be included in order to correctly model vocal tract acoustics not only at high frequencies, as commonly accepted, but also in the low frequency range below 4 kHz, where plane wave propagation occurs. © *2016 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4950698]

[JL] Pages: 2852–2859

I. INTRODUCTION

Radiation losses take place when sound waves propagating through the vocal tract emanate from the mouth aperture towards infinity. The most natural way to model this loss mechanism in voice production simulation is to use free-field radiation solutions, which consider a computational domain containing not only the vocal tract airway but also a finite portion of the free-field space that surrounds the human head. Perfectly matched layers (PMLs) (e.g., Takemoto *et al.*, 2010; Arnela and Guasch, 2013), infinite elements (e.g., Švancara and Horáček, 2006; Vampola *et al.*, 2013) or non-reflecting boundary conditions (e.g., Kako and Touda, 2006; Espinoza *et al.*, 2014) are then usually applied to absorb the sound waves reaching the outer boundary of the computational domain so as to emulate free-field propagation. However, the price to be paid for such accurate approaches is a high computational cost. Alternatively, the computational domain can be truncated at the mouth exit and the lips removed to impose a radiation impedance (e.g., Vampola *et al.*, 2008), or simply a zero pressure release condition (e.g., Hannukainen *et al.*, 2007; Aalto *et al.*, 2014), on a closed cross-section. This allows one to reduce the computational domain and hence the time and memory requirements of the simulations. However, analytical expressions for the radiation impedance, including the lips, are not available in the literature. Theoretical models that neglect their influence such as the circular piston set either in a

[a)]Electronic mail: marnela@salleurl.edu

 0001-4966/2016/139(5)/2852/8/$30.00

spherical or in an infinite baffle (see, e.g., Morse and Ingard, 1968; Kinsler *et al.*, 2000) are often used instead. In this work, we will analyze some of the effects of removing the lips in the production of vowels.

In Arnela *et al.* (2013), the radiation effects of human head geometry simplifications during vowel production were studied. It was concluded that the human head could be well approximated by a spherical baffle without any significant perceptual effect, but that lips should be taken into account for frequencies higher than 5 kHz. Below 5 kHz, lips could be discarded and the piston set in a sphere load impedance model should match the free-field radiation solution. However, it is to be noted that in that work, simplified vocal tract geometries were constructed using standard area functions available in the literature (Story, 2008), which were attached to a realistic human head with lips to define a reference case. The aim of this work is to extend the study in Arnela *et al.* (2013) and analyze the radiation effects of lips using three-dimensional (3-D) vocal tract geometries generated from magnetic resonance imaging (MRI).

Vowels /ɑ/, /i/, and /u/ are considered for this purpose. The MRI-based vocal tract geometries have been obtained from a freely available three-dimensional database (Aalto *et al.*, 2014). The lips have been removed from the vocal tract up to the last closed cross-section, where an impedance load can be imposed. Results for these configurations will be compared with those including the lips. Time-domain finite element simulations have been carried out for each case and experiments have been performed using equivalent 3-D-printed replicas for the vowel /ɑ/.

This paper is organized as follows. The methodology for the analysis of the lip radiation effects is detailed in Sec. II. The results are then presented in Sec. III and discussed in Sec. IV. Conclusions close the work in Sec. V.

II. METHODOLOGY

A. Vocal tract models

The MRI-based 3-D vocal tracts generated by Aalto *et al.* (2014) were adapted for the current work. Some additional elements not necessary for this study were removed, such as the subglottal tube and part of the face. The resulting vocal tract geometries were then set in a rigid flat baffle located at the mouth termination plane, which is defined as the last front-plane that produces a closed outline when it intersects with the vocal tract (see Dabbaghchian *et al.*, 2015). It has to be noted that a flat baffle was used in order to facilitate the construction of the mechanical replicas and the latter comparison between experiments and simulations, although any other configuration (e.g., spherical baffle) could be used instead.

Vowels /ɑ/, /i/, and /u/ were selected, since /ɑ/ and /u/ correspond to limiting cases for radiation losses (large and small mouth aperture) and vowel /i/ to an intermediate situation. Two configurations were considered for each vowel vocal tract: with lips and without lips. To construct the latter, the lips were simply removed so that the last closed cross-section of the vocal tract coincided with the flat baffle. The resulting vocal tract geometries can be observed in Fig. 1 for each vowel and vocal tract ending configuration.

B. Finite element numerical simulations

The finite element method (FEM) was applied to solve the time-domain wave equation

$$(\partial_{tt}^2 - c_0^2\nabla^2)p = 0, \tag{1}$$

combined with a PML to account for free-field propagation (Arnela and Guasch, 2013). In Eq. (1) $p(\boldsymbol{x}, t)$ stands for the acoustic pressure, c_0 for the speed of sound, and ∂_{tt}^2 for the second order time derivative. In what concerns boundary conditions, wall losses were introduced at the vocal tract using a constant boundary admittance coefficient value of 0.005. This value corresponds to the impedance of the vocal tract tissue $Z_w = 83\,666\,\mathrm{kg/m^2s}$ (see Švancara and Horáček, 2006). A wideband impulse was imposed at the glottal cross-section representing an input volume velocity. It consists of a Gaussian pulse of the type

$$gp(n) = e^{[(\Delta t\, n - T_{gp})0.29T_{gp}]^2}\,[\mathrm{m^3/s}], \tag{2}$$

with $T_{gp} = 0.646/f_0$ and $f_0 = 10$ kHz. This pulse was low-pass filtered at the maximum frequency of analysis (10 kHz) to avoid spurious errors in the simulations. The rest of the boundaries in the computational domain were considered acoustically rigid.

Each of the vocal tract models described in Sec. II A was set in a rigid flat baffle of dimensions 0.3×0.3 m (see Fig. 2). This baffle constitutes one side of a volume of $0.3 \times 0.3 \times 0.2$ m, which was used to allow sound waves to emanate from the vocal tract exit and naturally account for radiation losses. In turn, this volume was surrounded with a PML, 0.1 m in width, to emulate free-field propagation. The PML was configured to have a relative reflection coefficient of 10^{-4}. The resulting computational domains were meshed using linear tetrahedra with element sizes comprising from $h = 0.001$ m within the vocal tract, $h = 0.0025/0.005$ m in the free-field volume to $h = 0.0075$ m in the PML region. This gave a total number of elements ranging between

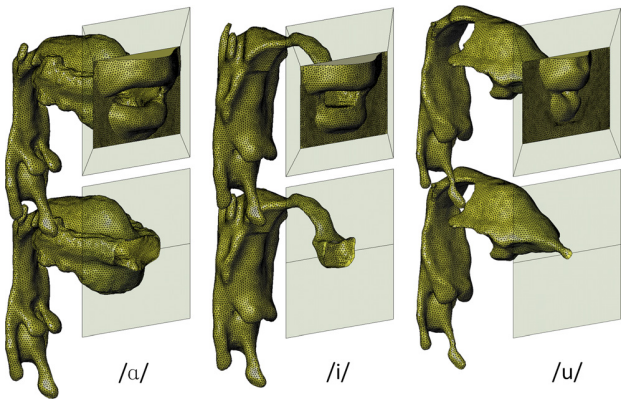


FIG. 1. (Color online) Vocal tract models for vowels /ɑ/, /i/, and /u/ with lips (top) and without lips (bottom).

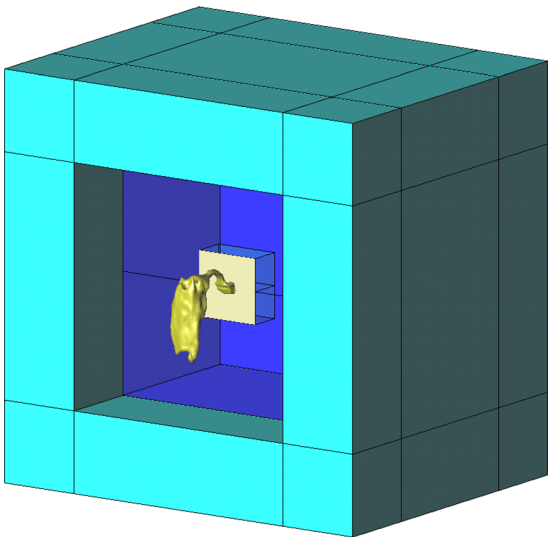


FIG. 2. (Color online) Computational domain used for the finite element simulation of vowel /i/, consisting of the corresponding vocal tract geometry, a free-field volume, and a PML. The most outer volumes correspond to the PML, which are surrounding the free-field radiation space. The vocal tract is set in a rigid flat baffle, which constitutes one side of the free-field volume.

3.5 and 4×10^6, depending on the vocal tract model configuration.

An FEM numerical simulation was then conducted using a sampling frequency of $f_s = 8000$ kHz, which fulfills a restrictive stability condition of the Courant–Friedrich–Levy-type required by explicit numerical schemes. Time events of 20 ms were simulated using a speed of sound of $c_0 = 350$ m/s and an air density of $\rho_0 = 1.14\,\mathrm{kg/m^3}$. These values were modified in the comparisons with experiments to match the corresponding temperature measured inside the insulated room ($c_0 = 343.19$ m/s and $\rho_0 = 1.2057\,\mathrm{kg/m^3}$ for the configuration with lips, and $c_0 = 343.86$ m/s and $\rho_0 = 1.2011\,\mathrm{kg/m^3}$ for the configuration without lips, corresponding, respectively, to a temperature of 19.65°C and 20.8°C). Simulation times ranged between 70 and 80 hours in a serial computing system with processor Intel® Core™ i5 2.8 GHz.

C. Experimental setup

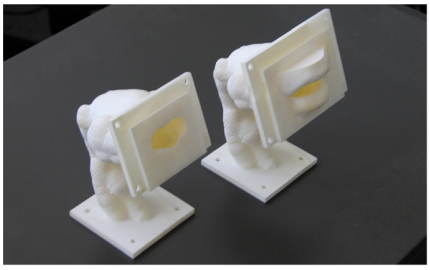
Mechanical replicas of the geometry corresponding to vowel /ɑ/ (see Fig. 1) were 3-D-printed with a ProJet 3510 SD 3-D printer using a UV curable plastic visijet M3X material [see Fig. 3(a)]. A rigid flat rectangular Plexiglas screen (365 × 360 mm) was set on the exit of the replicas.

An experimental setup based on the work of Motoki *et al.* (1992) was used to measure the acoustic pressure inside and outside these replicas. As shown in Fig. 3(b), it consists of a sound source (PSD:2002S-8) which allows one to generate sounds in the frequency range from 2 to 10 kHz. The experimental analysis will then be limited to this frequency range (see Blandin *et al.*, 2015, for previous comparisons between experiments and FEM in the frequency range below 2 kHz). The sound source is connected to the replica with an adaptation part featuring a 2 mm diameter communication hole. A probe microphone (B&K 4182) equipped with a 1 mm in diameter and 200 mm long probe allows one to measure the acoustic pressure. The microphone is displaced by a 3-D positioning system (OWIS PS35) with an accuracy of ±0.1 mm so that it is possible to measure the acoustic pressure on a fine grid of points. After being amplified by a microphone conditioner B&K 5935L, the signal of the microphone is transmitted to a data acquisition card (NI PCI-MIO 16 XE) and sampled at a rate of 44150 Hz. The same card was used to generate the excitation signal which is amplified by an Onkyo a-807 amplifier and transmitted to the sound source. The input electric current of the sound source was recorded as a reference to be compared to the microphone signal. The whole measurement process is computer controlled (NI LABVIEW).

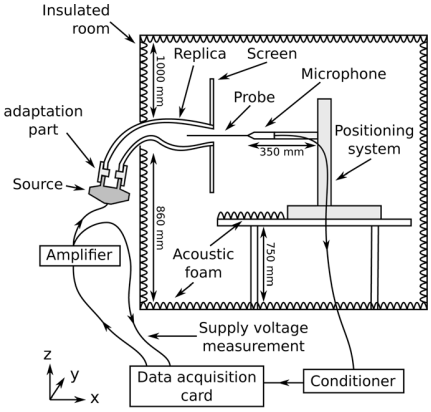
In order to limit the disturbance due to the ambient noise and the reflections on the wall of the room, the exit of the replica, the microphone and its positioning system were placed in an insulated room (1.92 m × 1.95 m × 1.99 m, Vol = 7.45 $\mathrm{m^3}$, Van Hirtum and Fujiso, 2012). The sound source was placed outside this room to avoid interference with the sound radiated by the replica.

D. Calculation of transfer functions and pressure maps

First, vocal tract transfer functions were computed from the finite element simulations. The Gaussian pulse in Eq. (2) was imposed as an input volume velocity $q_i(t)$ at the glottal cross-section and the acoustic pressure $p_o(t)$ was collected over time at a node located in front of the vocal tract exit,



(a)



(b)

FIG. 3. (Color online) (a) Mechanical replicas for vowel /ɑ/ with (right) and without (left) lips and (b) experimental setup.

0.04 m from the mouth center. A vocal tract transfer function was then calculated as

$$H(f) = \frac{P_o(f)}{Q_i(f)}, \tag{3}$$

where $P_o(f)$ and $Q_i(f)$, respectively, denote the Fourier transform of $p_o(t)$ and $q_i(t)$.

Second, pressure–pressure transfer functions between two arbitrary points i and j were obtained for both FEM simulations and experiments. These can be computed as

$$H_{ij}(f) = \frac{P_j(f)}{P_i(f)}, \tag{4}$$

with $P_i(f)$ and $P_j(f)$ standing for the Fourier transform of the acoustic pressure collected at points i and j, respectively. It is to be noted that the picks that could appear in a pressure–pressure transfer function do not correspond to the vocal tract resonances or formants typically observed in vocal tract transfer functions (see, e.g., Guasch and Magrans, 2004, for an explanation in mechanical transmissibility functions). For instance, if $P_i(f)$ and $P_j(f)$ are, respectively, obtained at the mouth exit and within the vocal tract, $P_i(f)$ will contain all the formants that appear in a vocal tract transfer function, while some of them will also be present in $P_j(f)$, depending on their pressure distribution. Therefore, note that shared resonances will be canceled once $H_{ij}(f)$ is computed. However, this magnitude is of special interest for the comparison between simulations and experiments (Blandin *et al.*, 2015), since the microphone and source calibration can be avoided, a single microphone can be used for the measurements and the influence of the experimental setup (amplifier, wires, and microphone conditioner) on them can be neglected.

Finally, the acoustic pressure distribution in the near field and in a region within the oral cavity was also obtained for FEM simulations and experiments. To do so, the vocal tract was excited with a broadband signal (a Gaussian pulse for simulations and a sweep tone for experiments) at the glottal cross-sectional area. Then, the acoustic pressure was collected in a structured grid of points separated a distance of 2.5 mm apart (some of the FE mesh nodes were forced to match this grid of points, so value interpolation was avoided in the simulations). This grid was located in a transverse plane that contains the center point of the mouth aperture. An example of this grid for vowel /ɑ/ considering the lips can be found in Fig. 4. The points were selected so that the microphone probe could reach any position (the probe cannot bend). The time evolution of the acoustic pressure at each point of the grid was then Fourier transformed. The pressure map was finally obtained after selecting the desired frequency (e.g., that of a vocal tract resonance). All pressure maps have been scaled with respect to the maximum value for comparison purposes between FEM simulations and experiments.

III. RESULTS

A. Comparison between simulations and experiments

First of all, FEM simulations and experiments have been cross-validated. Vowel /ɑ/ was selected for that purpose since measurements are easier to perform for it. This is so because, in contrast to other vowels, /ɑ/ presents a large mouth aperture and a wide oral cavity, which allow the microphone probe to cover a large area within the vocal tract before it bends. Moreover, vowel /ɑ/ may present noticeable radiation effects due to its large mouth aperture. The influence of lips for this vowel sound has been analyzed by comparing the results obtained with the two exit configurations (with lips and without lips) shown in Fig. 1 for the FEM simulations and in Fig. 3a for the experiments.

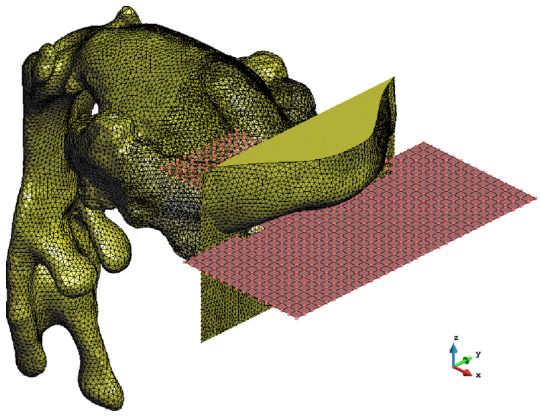


FIG. 4. (Color online) Grid of tracking points for the vocal tract geometry of vowel /ɑ/ with lips used in FEM and experiments to compute pressure maps. The grid covers the near-field and a portion within the oral cavity accessible by the microphone probe.

Pressure–pressure transfer function between two points within the vocal tract have been first computed [see Eq. (4) in Sec. II D for definition]. In Fig. 5, an example of transfer function between point #1 and point #3 is shown for the case with lips (top) and without lips (bottom). These points were, respectively, located at $(-0.0325, 0, 0)$ m and (0, 0, 0) m, with the origin of coordinates placed at the center of the mouth exit (center of the last closed cross-section in Fig. 4). As can be observed in Fig. 5, experiments and simulations match to a large extent for both configurations. However, some small peaks and dips not present in the FEM results can be observed in the experimental ones. These can be attributed to small irregularities of the mechanical replica.

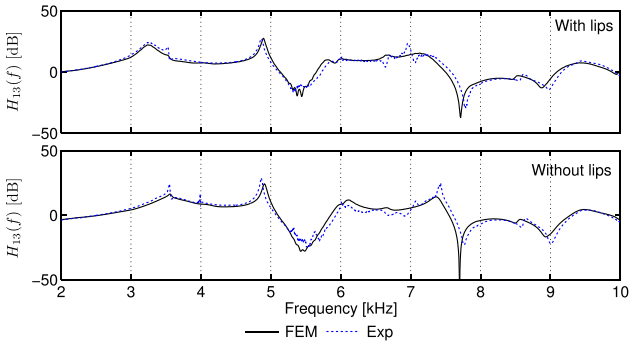


FIG. 5. (Color online) Pressure−pressure transfer function $H_{13}(f) = P_3(f)/P_1(f)$ for vowel /ɑ/ with lips (top) and without lips (bottom) obtained by finite element simulations (FEM) and experiments (Exp). $P_1(f)$ and $P_3(f)$ stand for the Fourier transform of the acoustic pressure collected at point #1 and point #3, which are, respectively, located within the oral cavity and at the mouth exit.

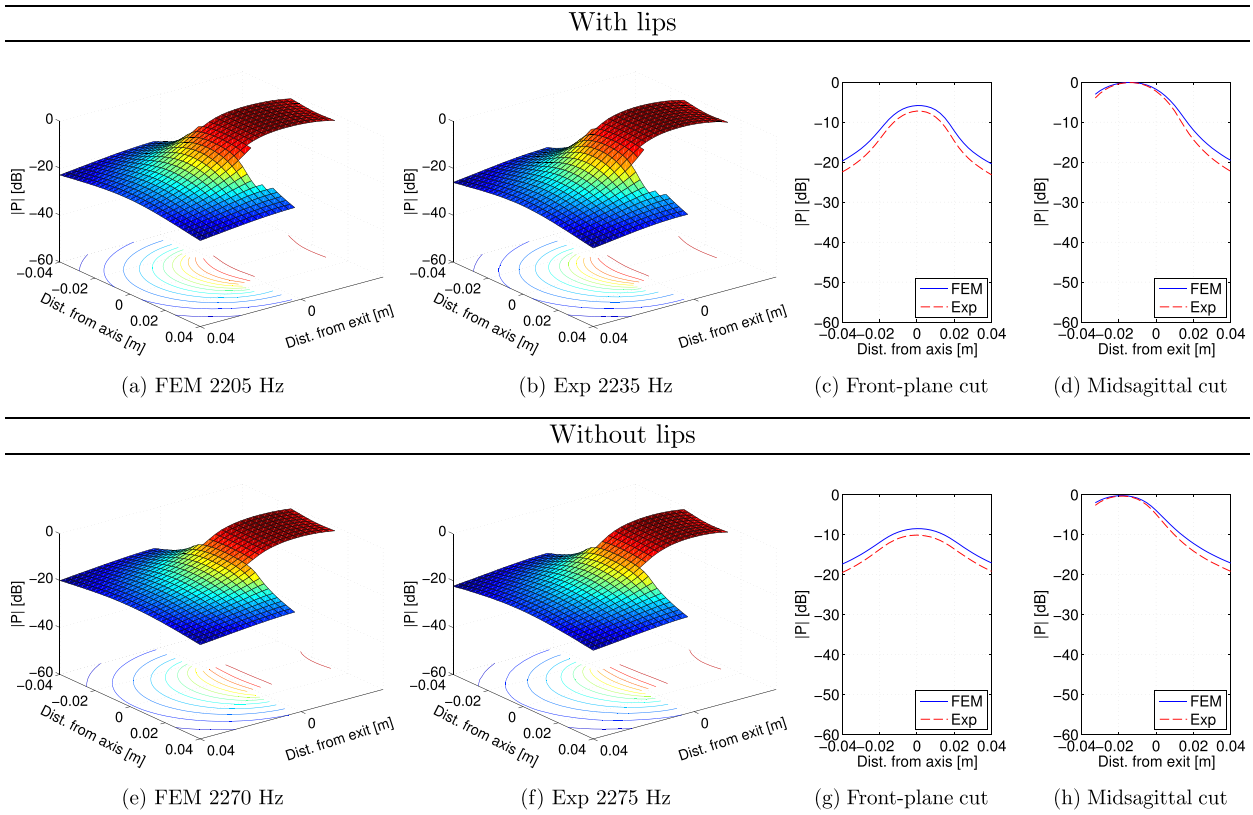
On the other hand, some resonance and antiresonance amplitudes and bandwidths are slightly different since FEM is only considering constant-frequency wall losses.

Pressure maps were next obtained from FEM simulations and experiments for both configurations (the procedure is described in Sec. II D). The third formant and the last formant just before 10 kHz were selected, the former being located in the low frequency region where the plane wave assumption holds and the latter belonging to the high frequency range. The resulting formant pressure maps are presented in Fig. 6 and Fig. 7, respectively. In addition, cuts of these pressure maps in a frontal plane located 1 cm in front of the mouth exit and in the midsagittal plane are also shown in the figures to facilitate comparison between FEM simulations and experiments. Let us first focus on the pressure map for the low frequency formant (see Fig. 6). Note in the figure that for the two cases planar waves are produced within the oral cavity, which exhibit a spherical radiation pattern when exiting the mouth. No significant differences can be observed between the case with lips and without lips and neither between FEM simulations and experiments. In what concerns the pressure maps for the high frequency formant in Fig. 7, it can be observed that the wavefronts within the oral cavity are no longer plane and that the radiation pattern is not symmetric. Note that in contrast to what occurs in the low frequency range (Fig. 6), the radiation pattern presents a strong dip, its location depending on the exit configuration. Besides, it should be remarked that numerical and experimental results closely match for the lip and no-lip cases.

B. Vocal tract transfer functions

Once the FEM approach has been validated against experiments, it will be used to analyze further radiation effects of lips. In particular we will resort to computed vocal tract transfer functions to do so [see Eq. (3) in Sec. II D for definition]. Figure 8 shows them for the vocal tracts of vowels /ɑ/, /i/, and /u/ with and without lips (see Fig. 1). For completeness, the first formant locations and −3 dB bandwidths are listed in Table I for each vowel and exit configuration.

The results for vowel /ɑ/ in Fig. 8 show that the inclusion of lips induces a significant shift of the formants to lower frequencies. This is clearly appreciated below 5 kHz, formant deviations rising up to 12.09% (see Table I). The shifting of the formants was an expected effect since for low frequencies lips mainly produce an increase of the radiation reactance (imaginary part of the radiation impedance) so that, depending on the formant-cavity affiliation phenomenon (see, e.g., Fant, 1970), formant positions are shifted down (see, e.g., Arnela *et al.*, 2013). On the other hand, lips have also affected some dips below 5 kHz, mainly originated by transverse modes or side branches such as the piriform sinuses and valleculae (see, e.g., Vampola *et al.*, 2015; Takemoto *et al.*, 2013; Takemoto *et al.*, 2010, for a detailed study of their effects). Note for instance that a small antiresonance between 3 and 4 kHz has appeared in the vocal tract transfer function when lips are considered (see Fig. 8). As the formant bandwidths are concerned (see Table I), the



(a) FEM 2205 Hz (b) Exp 2235 Hz (c) Front-plane cut (d) Midsagittal cut

(e) FEM 2270 Hz (f) Exp 2275 Hz (g) Front-plane cut (h) Midsagittal cut

FIG. 6. (Color online) Pressure amplitude maps for the third formant of vowel /ɑ/ obtained by finite element simulations (FEM) and experiments (Exp) with lips (top) and without lips (bottom). The acoustic pressure distribution within the oral cavity and at an area close to the vocal tract exit (mouth) is presented. Moreover, cuts of the pressure amplitude maps in a front-plane located 1 cm in front of the mouth exit and in the midsagittal plane are also presented.

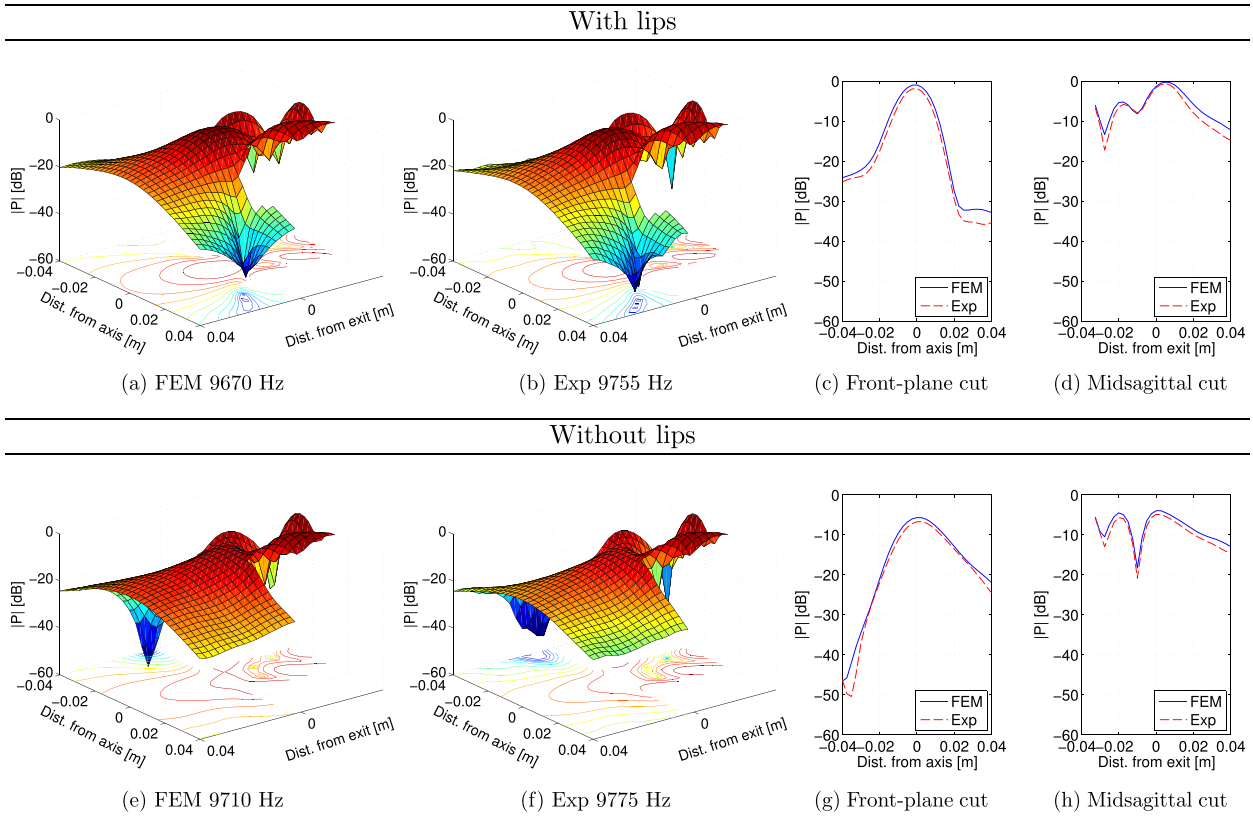


FIG. 7. (Color online) Pressure amplitude maps obtained for the last formant before 10 kHz of vowel /ɑ/ by finite element simulations (FEM) and experiments (Exp) with lips (top) and without lips (bottom). The acoustic pressure distribution within the oral cavity and at an area close to the vocal tract exit (mouth) is presented. Moreover, cuts of the pressure amplitude maps in a front-plane located 1 cm in front of the mouth exit and in the midsagittal plane are also presented.

inclusion of the lips has significantly reduced their values. This is a direct consequence of the shifting down of the formant locations. The radiation resistance (real part of the radiation impedance) diminishes when lowering the frequency, which results in smaller formant bandwidths. The significance of this effect will depend again on the cavity affiliation of each formant. As regards the high frequency range (above 5 kHz), a large variation of the amplitude levels takes place, which can be as high as 7 dB (once the levels have been normalized to that of the first formant for every function in the figure).

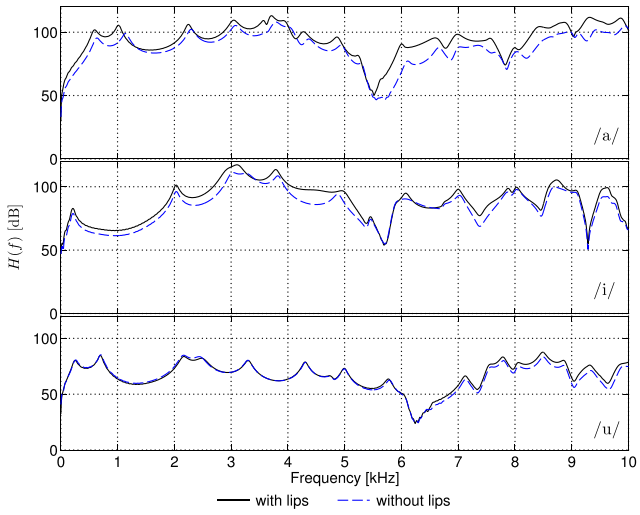


FIG. 8. (Color online) Vocal tract transfer functions $H(f)$ computed from finite element simulations for vowels /ɑ/, /i/, and /u/, with and without lips.

In what concerns vowel /u/, as one could expect, the radiation effects of lips are less important than for vowel /ɑ/, since /u/ entails a smaller mouth aperture. Below 5 kHz no significant differences are observed in the vocal tract transfer functions between the two cases (see the bottom of Fig. 8). Indeed, the obtained formant deviations are less than about 3.5% and the reduction in bandwidth smaller than −3%. For high frequencies a small increment of the amplitude levels (up to about 4–5 dB) can be appreciated when the lips are considered (see Fig. 8).

The case of vowel /i/ corresponds to an intermediate situation between /ɑ/ and /u/ in terms of radiation effects. Below 5 kHz, the observed differences for the first two formants are not really important (below 1.5%, see Table I). However, significant changes are produced for the third and fourth formants. When lips are concerned, the location of the formants not only shifts down but also gets closer to each other compared to the case without lips. In addition, the pressure level for the fourth formant becomes higher than for the third one. The proximity of both formants together with this amplitude modification has resulted in some difficulties to measure their location and bandwidths (denoted with a dash in Table I). On the other hand, it can be observed that the

TABLE I. First formant locations F_i (Hz) and bandwidths BW_i (Hz) for the cases with lips and without lips for vowels /ɑ/, /i/, and u/. Deviations in % ($\Delta F_i/F_i$) with respect to the case with lips are also reported.

		Formant location						Formant bandwidth					
Vowel	Case	F_1	F_2	F_3	F_4	F_5	F_6	BW_1	BW_2	BW_3	BW_4	BW_5	BW_6
/ɑ/	Lips	598	1017	2249	3049	3712	-	119	97	105	158	132	-
	No lips	638	1140	2297	3080	3792	-	123	115	123	162	197	-
	Δ(%)	6.69	12.09	2.13	1.02	2.16	-	3.36	18.56	17.14	2.53	49.24	-
/i/	Lips	214	2033	-	3104	3787	4944	67	91	-	-	130	-
	No lips	216	2035	3036	3255	3823	4879	67	90	-	-	131	145
	Δ(%)	0.93	0.1	-	4.87	0.95	−1.31	0	−1.1	-	-	0.76	-
/u/	Lips	254	697	2170	2480	3301	4305	90	79	-	-	118	98
	No lips	263	704	2165	2439	3296	4310	89	79	-	-	115	97
	Δ(%)	3.54	1	−0.23	−1.65	−0.15	0.12	−1.11	0	-	-	−2.54	−1.02

sixth formant in Fig. 8 (∼5 kHz) moves to higher frequencies and becomes much more attenuated when lips are considered (see also Table I). The shift of this formant to higher frequencies was unexpected, since as mentioned before, lips produce an increase of the reactance which should shift down the formant locations. This all seems to indicate that this formant does not correspond to a purely longitudinal mode but rather to a higher order mode, which strongly depends on the shape of the vocal tract. In what concerns the high frequency range, the shape of the vocal tract transfer function is not significantly altered, yet an increment of the amplitude levels, not higher than 5 dB, can be appreciated.

IV. DISCUSSION

The results obtained are mainly in line with those previously reported in Arnela *et al.* (2013), where the effects of lips were also studied among other human head geometry simplifications. In this work, it has also been observed that lips can produce an increase of the energy for high frequencies together with a shift down of the formant locations, depending on the produced vowel sound and the formant-cavity affiliation. However, some significant differences have been found between both works, especially for large mouth apertures.

Regarding vowel /ɑ/, the formant deviations reported in Arnela *et al.* (2013) for the lower frequencies (<5 kHz) were below 3% − 5% so they were not considered relevant from a perceptual point of view (Flanagan, 2008). The inclusion of the lips was then only recommended to account for the level increments in the high frequency range, which were up to about 6 dB. As reported in Monson *et al.* (2011), minimum level changes between 1 dB and 5 dB are perceptually detectable in the 8 kHz band. Therefore, the 6 dB amplitude variations could be perceptually relevant and modify the naturalness of the produced sound. In the current work, significant level increments for high frequencies have also been observed for vowel /ɑ/ (up to 7 dB), so lips can again be considered relevant in this frequency range. However, and in contrast to that previous study, deviations higher than 3% − 5% have been reported in this work for the first two formants (up to 12.09%), which may be perceptually important. Therefore, lips should be considered in this occasion to correctly model the radiation effects not only for higher frequencies, but also below 5 kHz.

In what concerns vowel /u/, no significant differences have been observed in the vocal tract transfer functions with lips and without lips. The formant deviations are below 3.5% and the level increments less than 5 dB, which can be considered not relevant from a perceptual point of view. Therefore, the lips can be removed for vowel /u/ without involving a significant perceptual modification, which agrees with the conclusions reached in Arnela *et al.* (2013) for this vowel sound.

Vowel /i/ was not addressed in Arnela *et al.* (2013). However, some considerations can be made for it. In this work the formant deviations, bandwidths and levels differences are quite small. Yet, significant deviations have been observed in the vocal tract transfer functions even for low frequencies, although they are difficult to quantify objectively. A subjective evaluation would be valuable in order to validate the importance of lips for this vowel from a perceptual point of view. However, such a study lies out of the scope of this paper.

Finally, it is to be pointed out that in Arnela *et al.* (2013) simplified vocal tract geometries with elliptical cross-sections were constructed by using area functions available in the literature (Story, 2008), which were in turn set in a realistic human head with lips. In contrast, this work considers full MRI-based vocal tract geometries already containing the lips. Therefore, the possible error derivable from the coupling of the simplified vocal tract with the realistic face with lips is avoided. In addition, a more complex acoustic field is generated for mid and high frequencies due to the intricate shape of the MRI-based vocal tracts and the presence of the side branches (e.g., piriform sinuses and valleculae).

All in all, it seems clear that the lips can also significantly influence the low frequency range of vowel sounds, especially for large mouth apertures. However, a large number of vocal tract geometries from different individuals should be analyzed to generalize the herein exposed conclusions with statistical confidence.

V. CONCLUSIONS

In this work, time-domain finite element simulations and experiments were conducted to analyze the effects of lips on the production of vowels. MRI-based vocal tract geometries with and without lips for vowels /ɑ/, /i/, and /u/ were selected. Results were contrasted for each vowel sound and vocal tract exit configuration. Pressure–pressure transfer functions and pressure distribution patterns for vowel /ɑ/ were used to cross-validate FEM simulations with measurements on 3-D-printed mechanical replicas. A fairly good match was achieved. At high frequencies, non-symmetrical radiation patterns were appreciated, exhibiting also a large dependence with the presence of lips. FEM was also used to compute the vocal tract transfer functions of vowels /ɑ/, /i/, and /u/. The inclusion of lips had a clear influence for vowel /ɑ/ at all frequencies. It produced a significant shift down of the formant locations, a reduction of their bandwidths, and a level increase that could be perceptually relevant. These effects were much weaker for vowel /i/ and almost negligible for vowel /u/.

ACKNOWLEDGMENTS

This research has been supported by EU-FET grant EUNISON 308874.

Aalto, D., Aaltonen, O., Happonen, R.-P., Jääsaari, P., Kivelä, A., Kuortti, J., Luukinen, J.-M., Malinen, J., Murtola, T., Parkkola, R., Saunavaara, J., Soukka, T., and Vainio, M. (**2014**). "Large scale data acquisition of simultaneous MRI and speech," Appl. Acoust. **83**, 64–75.

Arnela, M., and Guasch, O. (**2013**). "Finite element computation of elliptical vocal tract impedances using the two-microphone transfer function method," J. Acoust. Soc. Am. **133**, 4197–4209.

Arnela, M., Guasch, O., and Alías, F. (**2013**). "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations," J. Acoust. Soc. Am. **134**, 2946–2954.

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

Dabbaghchian, S., Arnela, M., and Engwall, O. (**2015**). "Simplification of vocal tract shapes with different levels of detail," in *Proc. of 18th International Congress of Phonetic Sciences (ICPhS)* (Glasgow, Scotland, UK).

Espinoza, H., Codina, R., and Badia, S. (**2014**). "A Sommerfeld non-reflecting boundary condition for the wave equation in mixed form," Comput. Methods Appl. Mech. Eng. **276**, 122–148.

Fant, G. (**1970**). *Acoustic Theory of Speech Production*, 2nd ed. (Mouton, Paris), Chap. 1.

Flanagan, J. L. (**2008**). *Speech Analysis, Synthesis and Perception*, 3th ed. (Springer-Verlag, New York), Chap. 3 and 7.

Guasch, O., and Magrans, F. X. (**2004**). "The global transfer direct transfer method applied to a finite simply supported elastic beam," J. Sound Vib. **276**, 335–359.

Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (**2007**). "Vowel formants from the wave equation," J. Acoust. Soc. Am. **122**, EL1–EL7.

Kako, T., and Touda, K. (**2006**). "Numerical method for voice generation problem based on finite element method," J. Comput. Acoust. **14**, 45–56.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York), Chap. 7.

Monson, B. B., Lotto, A. J., and Ternström, S. (**2011**). "Detection of high-frequency energy changes in sustained vowels produced by singers," J. Acoust. Soc. Am. **129**, 2263–2268.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York), Chap. 7.

Motoki, K., Miki, N., and Nagai, N. (**1992**). "Measurement of sound-pressure distribution in replicas of the oral cavity," J. Acoust. Soc. Am. **92**, 2577–2585.

Story, B. H. (**2008**). "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," J. Acoust. Soc. Am. **123**, 327–335.

Švancara, P., and Horáček, J. (**2006**). "Numerical modelling of effect of tonsillectomy on production of Czech vowels," Acta Acust. **92**, 681–688.

Takemoto, H., Adachi, S., Mokhtari, P., and Kitamura, T. (**2013**). "Acoustic interaction between the right and left piriform fossae in generating spectral dips," J. Acoust. Soc. Am. **134**, 2955–2964.

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

Vampola, T., Horáček, J., Laukkanen, A. M., and Švec, J. G. (**2013**). "Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on ct measurement," Logopedics Phoniatrics Vocology 1–10.

Vampola, T., Horáček, J., and Švec, J. G. (**2008**). "FE modeling of human vocal tract acoustics. Part I: Production of Czech vowels," Acta Acust. **94**, 433–447.

Vampola, T., Horáček, J., and Švec, J. G. (**2015**). "Modeling the influence of piriform sinuses and valleculae on the vocal tract resonances and anti-resonances," Acta Acust. **101**, 594–602.

Van Hirtum, A., and Fujiso, Y. (**2012**). "Insulation room for aero-acoustic experiments at moderate Reynolds and low Mach numbers," Appl. Acoust. **73**, 72–77.