



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Schickhofer, Lukas; Malinen, Jarmo; Mihaescu, Mihai

Compressible flow simulations of voiced speech using rigid vocal tract geometries acquired by MRI

Published in: Journal of the Acoustical Society of America

DOI: 10.1121/1.5095250

Published: 01/04/2019

Document Version Publisher's PDF, also known as Version of record

Published under the following license: CC BY

Please cite the original version:

Schickhofer, L., Malinen, J., & Mihaescu, M. (2019). Compressible flow simulations of voiced speech using rigid vocal tract geometries acquired by MRI. *Journal of the Acoustical Society of America*, *145*(4), 2049-2061. https://doi.org/10.1121/1.5095250

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Compressible flow simulations of voiced speech using rigid vocal tract geometries acquired by MRI

Lukas Schickhofer, Jarmo Malinen, and Mihai Mihaescu

Citation: The Journal of the Acoustical Society of America **145**, 2049 (2019); doi: 10.1121/1.5095250 View online: https://doi.org/10.1121/1.5095250 View Table of Contents: https://asa.scitation.org/toc/jas/145/4 Published by the Acoustical Society of America

ARTICLES YOU MAY BE INTERESTED IN

Inverse estimation of the vocal tract shape based on a vocal tract mapping interface The Journal of the Acoustical Society of America **145**, 1961 (2019); https://doi.org/10.1121/1.5095409

Aerodynamic impact of the ventricular folds in computational larynx models The Journal of the Acoustical Society of America **145**, 2376 (2019); https://doi.org/10.1121/1.5098775

Characterization of inter-speaker articulatory variability: A two-level multi-speaker modelling approach based on MRI data

The Journal of the Acoustical Society of America 145, 2149 (2019); https://doi.org/10.1121/1.5096631

Relation of second formant trajectories to tongue kinematics The Journal of the Acoustical Society of America **145**, EL323 (2019); https://doi.org/10.1121/1.5099163

Pitch-shift responses as an online monitoring mechanism during level tone production The Journal of the Acoustical Society of America **145**, 2192 (2019); https://doi.org/10.1121/1.5096977

Cross-modal association between vowels and colours: A cross-linguistic perspective The Journal of the Acoustical Society of America **145**, 2265 (2019); https://doi.org/10.1121/1.5096632



Compressible flow simulations of voiced speech using rigid vocal tract geometries acquired by MRI

Lukas Schickhofer,^{1,a)} Jarmo Malinen,² and Mihai Mihaescu¹

¹Department of Mechanics, Linné FLOW Centre, KTH Royal Institute of Technology, Stockholm, SE-10044, Sweden

²Department of Mathematics and Systems Analysis, Aalto University, Aalto, FI-00076, Finland

(Received 4 July 2018; revised 14 January 2019; accepted 7 March 2019; published online 16 April 2019)

Voiced speech consists mainly of the source signal that is frequency weighted by the acoustic filtering of the upper airways and vortex-induced sound through perturbation in the flow field. This study investigates the flow instabilities leading to vortex shedding and the importance of coherent structures in the supraglottal region downstream of the vocal folds for the far-field sound signal. Large eddy simulations of the compressible airflow through the glottal constriction are performed in realistic geometries obtained from three-dimensional magnetic resonance imaging data. Intermittent flow separation through the glottis is shown to introduce unsteady surface pressure through impingement of vortices. Additionally, dominant flow instabilities develop in the shear layer associated with the glottal jet. The aerodynamic perturbations in the near field and the acoustic signal in the far field are examined by means of spatial and temporal Fourier analysis. Furthermore, the acoustic sources due to the unsteady supraglottal flow are identified with the aid of surface spectra, and critical regions of amplification of the dominant frequencies of the investigated vowel geometries are identified. © 2019. Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/ 4.0/). https://doi.org/10.1121/1.5095250

[JFL]

Pages: 2049-2061

CrossMark

I. INTRODUCTION

Voice production and the expression of sounds via speech are crucial human traits and essential for our communication. The primary source of this process is the oscillation of the vocal folds, leading to a pulsation of the sub- and supraglottal pressure. This causes periodic pressure fluctuations in the vocal tract, which are propagated into the far field through the mouth opening.

The human voice production is therefore often successfully modeled as a source-filter arrangement with the glottis acting as an acoustic source and the vocal tract as a filter (Fant, 1971; Stevens, 2000; Titze and Alipour, 2006). As a result of reflections, constructive and destructive interference of acoustic waves, the signal exiting the upper airways is modulated in such a way that certain distinct tonalities appear in the spectral envelope of acoustic pressure fluctuations. Various parts of the upper airways produce resonant patterns at different frequencies, leading to pronounced peaks in the far-field acoustic spectrum. These peaks are referred to as formants, and their position in the Fourier spectrum characterizes the produced sound. In the case of vowels, particularly the first two formants, F_1 and F_2 , identify the spoken tone. During speech, the air volume of the vocal tract cavities is continuously manipulated through muscular flexion and extension, allowing the vocalization of various sounds through articulation. Through phonation, the harmonic component of the spoken sound, as a result of the periodic vocal fold modulation of the glottal airflow, is the dominant soundgenerating mechanism at lower frequencies (Klatt and Klatt, 1990). In this range it overshadows any inharmonic noise component that occurs due to flow turbulence. However, application of the quasisteady approximation for aerodynamic sound shows that the acoustic contribution of flow structures and turbulence becomes important at frequencies higher than 2 kHz (Howe and McGowan, 2007; Zhang et al., 2002). The acoustic contribution in the high-frequency range of several kilohertz is due to the fact that the vortical structures present in the supraglottal region are small and of quadrupole source type that becomes negligible compared to the dominant sources of monopole- and dipole-types at lower frequencies (Zhao et al., 2002). Zhang and Neubauer (2010) therefore hypothesized that variations in the supraglottal flow field barely impact low-frequency acoustics. Although the unique tonal signature of a spoken sound is primarily determined by the first two formants, the higher harmonics and frequencies are responsible for the personal quality of the voice (Sundberg, 2005). For a complete understanding of voice production covering the full range of audible frequencies and dominant modes between 20 Hz and 20 kHz, both harmonic and inharmonic factors of broadband type must therefore be taken into account. Disturbances introduced by the flow instabilities need to be accurately computed if the high-frequency part of the voice is to be considered. Moreover, it is unclear to what extent both intra- and supraglottal flow structures contribute to the overall acoustic signal and, thereby, to phonation in general (Zhang, 2016).

Computational fluid dynamics (CFD) of phonation has emerged as a valuable tool to treat such questions. Indeed,

0

^{a)}Electronic mail: schic@mech.kth.se

there exists no method to spatially and temporally resolve the flow and pressure fields in the vocal tract during human phonation *in vivo*. Experiments both on elastic, selfoscillating glottal physical models (Murray *et al.*, 2014; Neubauer *et al.*, 2007; Syndergaard *et al.*, 2017; Triep and Brücker, 2010), as well as on *ex vivo* canine larynx models (Khosla *et al.*, 2007; Oren *et al.*, 2014) give clear indications on the complex, unsteady, three-dimensional (3D) flow field arising in the supraglottal region. These studies show effects such as secondary flows, recirculation regions, asymmetric attachment of the glottal jet to one of the airway walls, flow instabilities, and transition to turbulence. The shedding of vortices and downstream convection of instabilities was observed to increase during the closing phase of the glottal cycle or while the vocal folds take on a divergent shape.

Haigermoser (2009) and Koschatzky *et al.* (2011) used near-field flow data from particle image velocimetry (PIV) to estimate the contribution of the acoustic source terms for rectangular cavity flows. Lodermeyer *et al.* (2018) consequently applied a similar approach on PIV data of the downstream region of a self-oscillating, physical vocal fold model and calculated the acoustic source terms of Lighthill's equation. All these observations support the conclusion that any accurate description of flow and acoustics in the vocal tract needs to be sufficiently well-resolved in time and space to capture these phenomena that are excluded from traditional phonation models based on the Bernoulli equation.

CFD has been applied within the biomechanics of phonation to fill that gap by computing the unsteady flow field and correctly predicting the flow characteristics (de Luzan *et al.*, 2015; Mihaescu *et al.*, 2010). Simplified geometries of the glottal region have been used in these studies to investigate the frequency content of vorticity generated downstream of the glottis. Šidlof *et al.* (2015) and Zörner *et al.* (2016) used CFD simulations for the calculation of nearfield aerodynamic fluctuations, which were then imposed on the acoustic far-field through Lighthill's analogy and acoustic perturbation equations. Thus, they were able to link the aerodynamic sources in the glottal area to the acoustic region.

Recently, magnetic resonance imaging (MRI) measurements can be used for the analysis and setup of anatomically realistic computational geometries for numerical simulations of the flow inside the human upper airways. Story et al. (1996) used MRI data for the extraction of cross-sectional area functions of the vocal tract during phonation of steady vowel sounds and thereby measured the internal resonant air volume. Blandin et al. (2015) and Arnela et al. (2016) could show that for frequencies below 4-5 kHz vocal tract simplifications show only weak influence on the acoustic response due to the predominant plane wave propagation in the lowfrequency range. However, for higher-order modes, they found a strong dependence on the geometrical representation of the vocal tract. It was found that only detailed, 3D geometries of the vocal tract are able to correctly emulate the high-frequency behavior. Realistic geometries can properly replicate the physical and physiological conditions inside a human vocal tract and mimic its behavior as an acoustic filter to give accurate low- and high-frequency tonalities in the far-field sound spectrum and their associated sound pressure levels (SPLs).

Furthermore, it has been shown that numerical simulations in the upper airways provide valuable insight into the physical consequences a surgical intervention might have on a patient (Mylavarapu et al., 2013). CFD has become a promising method to assess the effectiveness of invasive measures to alleviate medical conditions such as obstructive sleep apnea through computation of parameters such as the airway resistance and the pressure drop along the pharyngeal domain (Mihaescu et al., 2011). In speech production, numerical flow simulations based on MRI data have the potential to motivate clinical decision-making for the therapy of vocal disorders. An accepted theory with voiceless sounds, whispered vowels, and fricatives is that the noise generated through the turbulent glottal jet and its interaction with the surrounding airway constitutes the source for sound (Stevens, 2000). For patients whose sustained vocal fold oscillation is inhibited due to paralysis, paresis, or cancer, one can use this information to enable whisper-like sounds via treatments. A major motivation for this work and research on the biomechanical and acoustical aspects of the human upper airways is therefore not only the advancement of knowledge about flow-induced voice generation, but also the large prevalence of vocal disorders. Voice disorders account for up to 25% of all occupational diseases (Sliwinska-Kowalska et al., 2006). About 30% of the common population are estimated to be affected by a voice production disorder at some point in their life with an increased incidence of more than 60% among professions with extensive voice use, such as teachers and professors (Mittal et al., 2013).

The goal of this study is an accurate computation of the whole audible frequency range of voiced speech under articulation of natural vowels. This requires resolution of both the harmonic component due to the regular pressure fluctuations by glottal modulation, as well as the inharmonic broadband component due to flow instabilities and turbulence, which is typically neglected by traditional models of voice acoustics based on wave equations. In order to capture possible high-frequency cross-modes, 3D MRI geometries are applied.

II. METHODS

A. Geometry and mesh

The computational geometries for the Finnish vowels [a], [e], [i], [o], and [u] were acquired by MRI from a 30-yrold, healthy male subject following the experimental arrangements described by Aalto *et al.* (2011). The imaging was carried out using Siemens Magnetom Avanto 1.5T scanner (Siemens Medical Solutions, Erlangen, Germany) running 3D VIBE (volumetric interpolated breath-hold examination) MRI sequence (Rofsky *et al.*, 1999).

The subject was firmly aligned by soft material inside the head and neck MRI coil in order to allow only movement of the muscles involved in the voice generation process (cf. Fig. 1). Using the scanner settings detailed in Aalto *et al.* (2014, Sec. 3.2), the imaging of a single vowel geometry



FIG. 1. (Color online) Vocal tract geometry of the dataset for the pronunciation of vowel $[\alpha]$ with the anatomical directions and planes referred to in this study. The approximate positions of larynx, pharynx, and the far-field monitoring point *P* are indicated.

took ca. 8 s using 1.8 mm isotropic voxels. Due to restrictions of the MRI technology, the computational geometries do not include teeth. The subject produced prolonged vowels at the fundamental frequency $f_o = 104$ Hz for ca. 12 s in supine position. In addition to the reference pitch signal, the subject could also hear his own, analogically denoised utterance with a delay of approximately 90 ms through the standard Siemens MRI headphones (Siemens Medical Solutions, Erlangen, Germany). During MRI, the acoustic noise level within the scanner is ca. 90 dB (SPL), and the headphones provide 26 dB nominal damping of this external noise. The vowel production time exceeds the time required for MRI in order to guarantee the stability of phonation during imaging and obtain a speech sample right before and after the imaging that is free of acoustic scanner noise.

The surface model of the air-tissue interface was extracted from 3D MRI voxel data by an earlier version (Aalto et al., 2013) of the automatic vocal tract segmentation software published by Ojalammi and Malinen (2017). The 3D numerical mesh of the fluid domain is unstructured and consists of polyhedral cells ranging from the airway channels to the far-field domain. The geometry in the intraglottal region is the result of averaging through the MRI recording, as the vocal fold motion is too fast to be resolved in time. Nevertheless, the whole domain is anatomically accurate and no geometrical modifications have been applied. Additional local mesh refinement has been applied in the intra- and supraglottal region, where velocity and pressure gradients in the flow are expected to be largest close to the shear layers of the glottal jet (cf. Fig. 2). Three consecutive refinement stages are present in the supraglottal region with a decrease of the base cell size by 20% at each stage. The grids for the different vowel articulations used in this study have a resolution with an approximate cell count of 7.5×10^6 based on the grid convergence study detailed in Sec. IIF. The chosen mesh sizes are significantly above the minimum requirement of the solver of 40 cells per wavelength for the highest relevant frequencies of phonation, and are therefore suitable for extracting the important acoustic information. The airflow is considered as compressible and modeled in the simulations by the ideal gas



FIG. 2. (Color online) Sections of the mesh in the supraglottal region along the sagittal and transversal plane with indication of the glottal constriction.

law $p = \rho RT$, relating gas pressure p and density ρ via the specific gas constant of air R = 287.06 J/(kgK) at room temperature $T \approx 300 \text{ K}$.

B. Vowel formant and Helmholtz resonance data

Two kinds of comparison data were used to validate the results of the CFD computations: formant frequency data extracted from the speech sample recorded during the MRI and data obtained through solution of the Helmholtz eigenvalue problem based on the same computational geometry (cf. Sec. IIF). Following the notation introduced by Titze et al. (2015), the formants are generally denoted by F_i , with index i = 1, ..., n. Unless explicitly stated otherwise, the formant data F_i in this study has been extracted through the unsteady flow simulations. As explained by Aalto et al. (2011), noise free samples were recorded inside the scanner right before and after the imaging using a two-channel acoustic sound collector and waveguides. The frequency response of the instrumentation was numerically compensated from the speech samples, and the formant analysis was carried out by linear predictive coding (LPC) in MATLAB (MathWorks, Natick, MA). The peaks of the resulting spectral envelope were inspected manually, and values for F_1 and F_2 were determined by comparisons, when necessary, with corresponding speech data that were recorded from the same subject in an anechoic chamber (Aalto et al., 2012, Table II). The lowest eigenvalues from the Helmholtz problem were computed by the finite element method (FEM) with piecewise linear shape functions and tetrahedral meshes. A Dirichlet boundary condition is used at the mouth, leading to discrepancy with the values of F_1 and F_2 measured from speech as discussed in Aalto et al. (2014, Sec. 5.2). The resonance frequencies corresponding to F_1 and F_2 can be found in Aalto et al. (2012, Table I).

C. Glottal flow and boundary conditions

Various waveforms have been suggested as a realistic model of the glottal source (Fant *et al.*, 1985; Fujisaki and Ljungqvist, 1986; Rosenberg, 1971). For a comparative

analysis of several different waveform models the reader is referred to the work by Fujisaki and Ljungqvist (1986). In the current study, the Rosenberg waveform $f_R(t) = f_p(t)$ + $f_n(t)$ of the glottal pulse is applied to account for the dynamic behavior of the vocal folds (Rosenberg, 1971). It models the volume velocity through the glottal constriction and consists of the function

$$f_p(t) = \frac{A_0}{2} \left[1 - \cos\left(\frac{t}{t_p}\pi\right) \right], \quad 0 \le t \le t_p, \tag{1}$$

for its positive slope and the function

$$f_n(t) = A_0 \cos\left(\frac{t - t_p}{t_n} \frac{\pi}{2}\right), \quad t_p \le t \le t_p + t_n, \tag{2}$$

for its negative slope with the pulse amplitude A_0 . The relative duration of the glottal opening t_p with respect to the glottal period t_0 is referred to in the following as glottal opening quotient. The glottal pulse with the parameter values used for this study is shown in Fig. 3. These are the fundamental frequency of $f_o = 120 \,\text{Hz}$, characteristic for an average male voice, and the glottal opening quotient of t_p/t_0 = 0.5525. The ratio of the time in which the glottis is opening to the total pulse length is $t_p/(t_p + t_n) = 0.85$. The resulting open quotient $(t_p + t_n)/t_0 = 0.65$ of this study lies in the range of typical values for voiced speech (Mittal et al., 2013). The pressure amplitude at the glottal entrance is chosen to be the mean subglottal pressure $P_0 = P_s \approx 300 \,\text{Pa}$. This value is at the lower end of typical physiological subglottal pressures of voiced speech and coincides with a relatively low glottal frequency of the waveform and a volume velocity amplitude of $A_0 = 2.2 \times 10^{-4} \text{ m}^3/\text{s} (0.22 \text{ l/s})$ in Eqs. (1) and (2) (Titze and Alipour, 2006).

The vocal tract is assumed static and rigid. The interface between airway walls and airflow is modeled as solid walls with a Dirichlet boundary condition, $u_i = 0$, referred to as *no-slip boundary condition*, for the velocity components of the fluid. Free stream and acoustically non-reflecting boundary conditions are imposed at the inlet of the tubular vocal tract (i.e., at the subglottal end) and at the outlet surfaces of the computational domain located about 15 cm away from the mouth in the acoustic region, where the fluid characteristics are specified by the Mach number, atmospheric pressure, and room temperature.

D. Numerical method

For a direct computation of both the velocity field in the vocal tract and pressure fluctuations in the near and far field, compressible flow simulations using the finite volume method (FVM) are performed in this work. Thus, the behavior of the fluid is described by the Navier-Stokes equations in compressible form. Body forces like gravity are not considered, as their effect on airflow in the upper airways can be neglected. With the current study, handling turbulence by means of the large eddy simulation (LES) approach, the large, energy-containing scales are resolved down to the Taylor microscale in the inertial subrange. The resolution



FIG. 3. (Color online) Rosenberg glottal waveform for a typical source signal of a male voice. The waveform of the unperturbed signal is given in (a) with points of interest of the cycle marked for the opening phase (I), peak glottal flow (II), closing phase (III), and full closure (IV). The frequency content is shown in (b).

down to scales in the inertial subrange is an important feature, in particular when acoustic pressure fluctuations are of interest. LES applies a filtering operation to the flow variable $\tilde{\psi}(\vec{r},t) \equiv \int_V \psi(\vec{r}',t) G(\vec{r},\vec{r}') d\vec{r}'$ by use of the filter function $G(\vec{r},\vec{r}') = G(\vec{r} - \vec{r}')$ that acts as a convolution kernel. Thus, the quantity can be decomposed into a filtered part and a subfiltered one. The cutoff value of resolved scales is dependent on the numerical cell size with the subgrid scales of the flow being low-pass filtered by the computational grid. Therefore, the filter width is directly related to the spatial resolution of the mesh. The unresolved scales of the flow below that value are computed by the wall-adapting local eddy-viscosity (WALE) subgrid scale model after Nicoud and Ducros (1999), which relates the subgrid scale viscosity to the grid filter width, density, and resolved velocity field. For the time-dependent computations of this study, the spatially filtered Navier-Stokes equations are calculated for a numerical mesh size small enough to accurately handle the complex geometry of the upper airways and resolve the most important energy-containing flow structures.

Following the approximation by Pope (2001), the Taylor microscale is estimated as

$$\lambda_T \approx [15\nu \times 0.1d/(0.1w_{\rm max})]^{1/2},$$
(3)

with a kinematic viscosity of air of $\nu \approx 1.79 \times 10^{-5}$ m²/s, a glottal width of $d \approx 2.5 \times 10^{-3}$ m, and a maximum axial velocity of $w_{\rm max} \approx 21.98 \,{\rm m/s}$. The integral length scale l=0.1d is assumed to be one order of magnitude smaller than the characteristic length scale of the geometry in the glottal region, which is the glottal width. Furthermore, the turbulence intensity is estimated at 10%. With the established value of the Taylor microscale in Table I, the spatial dimensions of the grid are chosen to meet its order and are consequently refined toward the supraglottal region in steps of 60%, 40%, and 20% of the maximum volume size of 9.62×10^{-4} m in the far field. This way we ensure that the criterion of the Taylor microscale is met by the grid resolution in the whole domain. Additional characteristic parameters of the flow are given in Table I. Their values lie in the range of a typical male voice (Stevens, 2000).

Grid spacing Δx and time step Δt are chosen to satisfy the Courant-Friedrichs-Levy (CFL) criterion. The convective Courant number $CFL_c = u_i \Delta t / \Delta x$ and the acoustic one $CFL_a = c_0 \Delta t / \Delta x$ are related via the Mach number $CFL_c / \Delta x$ $CFL_a = u_i/c_0 = M$. Since the Mach number of the considered flow regime is small ($M \ll 1$), the convective Courant number becomes the critical parameter for the choice of grid and time step sizes to capture all the aerodynamic fluctuations in the source region ($CFL_c \ll CFL_a$). The maximum time step of the computations is chosen as $\Delta t_{\text{max}} = 20 \,\mu\text{s}$, sufficiently small to resolve the full Fourier spectrum of frequencies relevant for phonation. The near-wall regions are resolved accordingly such that a non-dimensional, perpendicular wall distance of $y^+ = (yu^*)/\nu = (y\sqrt{\tau_w}/\rho)/\nu \le 1$, depending on the wall shear stress τ_w and the kinematic viscosity ν , is obtained. The fluctuations in velocity and pressure are computed in a coupled manner. The convective flux is computed by spatial discretization via a second-order bounded centraldifference scheme. Integration in time is performed via an implicit scheme and with a temporal discretization of second order consisting of an average of ten inner iterations to

TABLE I. Values of flow parameters for the investigated cases of voiced speech.

Definition	Parameter	Value
Reynolds number	Re	100-3000
Integral length scale	l	$\sim 2.5 \times 10^{-4} \text{ m}$
Taylor microscale	λ_T	$\sim 1.8 \times 10^{-4} \ \text{m}$
Fundamental frequency	fo	120 Hz
Peak glottal jet velocity	u _{max}	$\sim 22 \text{ m/s}$
Peak volumetric flow	A_0	$\sim 2.2 \times 10^{-4} \text{ m}^3/\text{s}$

ensure convergence within each time step. The mean flow fields are extracted with a statistically significant sample size of more than 50 full glottal cycles and after any initial spurious reflections in the numerical domain have vanished. The high number of cycles is chosen to resolve also the low-frequency component of the acoustic spectrum below the source frequency. The fluctuations in the flow variables are then computed by subtracting the mean values from the total instantaneous values, $\psi' = \psi - \bar{\psi}$.

E. Frequency-wavenumber spectrum

For an analysis of disturbances and their propagation velocities, two-dimensional Fourier transformation of the fluctuating pressure is performed along the centerline in the near- and far-field domains [see, e.g., Nussbaumer (2012)]. The time-space Fourier transform of data is computed as

$$\hat{f}(k,\omega) = \int_{\mathcal{S}} \left(\int_0^\infty f(x,t) e^{-i(\omega t + kx)} dt \right) |dx|,\tag{4}$$

with $k \in K$ and $\omega \in [0,\infty]$. Furthermore, the phase velocity of disturbances is defined as $c_{\varphi} = f\lambda = f2\pi/k$. Since the wavenumber k is subsequently plotted in the range $[-\pi,\pi]$, we use $c_{\varphi} = f\pi/k$. Thus, the group velocity of disturbances can be computed via $c_g = \pi \partial f/\partial k$, with the slope of a frequencywavenumber diagram $\partial f/\partial k$.

F. Validation and verification

The LES approach has been initially validated for a simplified geometry of the intraglottal domain based on the M5 model introduced by Scherer *et al.* (2001), which is well defined through surface design equations. The results for the intraglottal pressure computed numerically are shown in Fig. 4 for both the flow and non-flow wall depending on the asymmetric deflection of the glottal jet toward one side. There is an overall very good agreement between the predicted and experimental pressures for all three transglottal



FIG. 4. (Color online) Comparison of numerically determined intraglottal pressure profiles with the experimental data from Scherer *et al.* (2001) for subglottal pressures of 3 cmH₂O, 5 cmH₂O, and 10 cmH₂O (1 cmH₂O = 98.0665 Pa).

pressures considered. The definition of flow and non-flow wall, respectively, has been the same as in the original study by Scherer *et al.* (2001). The largest deviations of the numerical results from the experimental reference data are observed for the flow wall at a transglottal pressure of 10 cmH₂O with 0.009%, 1.1%, and 7.8% for the relative minimum, mean, and maximum values, respectively. The smallest deviations occur for the non-flow wall at 5 cmH₂O with 0.001%, 0.8%, and 2.4% for the relative minimum, mean, and maximum values, respectively.

Acoustic validation of the applied method with respect to the resolved formants is performed as well. For this purpose, using the numerical approach and the boundary conditions stated in Secs. IIC and IID, flow and acoustics data corresponding to the vocal tract geometries relevant for the three vowels $[\alpha]$, [i], and [u] are calculated. Fouriertransformed signals of predicted far-field pressure fluctuations in a monitoring point *P* outside of the vocal tract (as depicted in Fig. 1) are compared with

- (i) resonance frequencies computed from the Helmholtz eigenvalue problem, based on the 3D wave equation solved in the same vowel geometries,
- (ii) experimental formant data extracted from speech recordings carried out simultaneously with the MRI recordings.

The result can be seen for the three vowels [a], [i], and [u] in Fig. 5. The dominant spectral envelope peaks from the simulations of the unsteady flow match with the frequency values obtained from the Helmholtz equation and speech recordings. Particularly the first formant F_1 is captured well for all vowel geometries. The second formant F_2 shows good agreement with the experimentally obtained formants, except for the vowel [i] in Fig. 5(b), where the formant does not appear as a pronounced peak in the farfield spectrum. There is a narrow constriction caused by the tongue position in the high front vowel [i], and it is likely to produce turbulence and large fluctuations in the flow, leading to high broadband noise that overshadows the peak at F_2 . However, a peak of the curve of local maxima is found at a frequency of ~1900 Hz, which most likely corresponds to the formant F_2 and is plotted together with the formant data of all other vowels in Fig. 11. That the Helmholtz resonance data and the experimental formant corresponding to F_2 of [a] do not match is due to the overly simple boundary condition in the Helmholtz equation as discussed in Aalto et al. (2014, Sec. 5.2). Overall, the acoustic information is well captured through the direct compressible flow simulations and agrees with the experimental reference data. The solver used for this study has also been applied to study aeroacoustics related to fluidstructure interaction at low Reynolds numbers (Schickhofer et al., 2016).

Furthermore, the numerical solutions to the field variables are verified by a grid sensitivity study based on five sizes varying from a coarse mesh of 10^5 cells to a fine one of approximately 7.5×10^6 cells for the geometry of vowel [a]. The grid convergence index (GCI) and discretization errors



FIG. 5. (Color online) Far-field spectra of the vowels [a] in (a), [i] in (b), and [u] in (c). The Fourier spectrum of the acoustic pressure fluctuations obtained from the LES simulations is shown with the peak envelopes, the discrete values of the formant frequencies F_1 and F_2 obtained through experiments, and the Helmholtz resonance frequencies. Uncertainty of the experimental data is indicated by the grey regions.

are calculated for the chosen grid sizes according to the procedure described by Celik *et al.* (2008), which has been shown to be equally valid for non-structured grids as the ones used for this study. The averaged relative GCI along a centerline through the glottis and supraglottal regions is



FIG. 6. (Color online) Instantaneous normalized velocity magnitude in the sagittal plane section with indications of the near-field probes P1-P4 and the far-field probe.

 $\overline{\text{GCI}}_p = 1.7\%$ for the mean pressure and $\overline{\text{GCI}}_u = 9.7\%$ for the mean velocity of the finest grid. For the current study, the highest grid resolution of the sensitivity study was chosen for all analyzed vocal tract configurations.

III. FLOW FIELD

A. Scales of flow fluctuations

The resolved frequency range for velocity and pressure fluctuations throughout the vocal tract is investigated by Fourier transformations of the time history data obtained at discrete points as indicated in Fig. 6. The pressure fluctuations, $p' = p - \bar{p}$, and the velocity fluctuations, $u'_i = u_i - \bar{u}_i$, are computed by subtracting the mean values from the total instantaneous values. The locations of the probe points are chosen such that P1 is positioned close to the glottal constriction, while P2 and P3 are located in the *laryngopharynx*, or supraglottal domain, respectively, where flow instabilities close to the glottal jet and transition to turbulence in its shear layers are likely to occur. The probe P4 is well inside the oral cavity, while the far-field probe is located at a distance of 8.5 cm from the mouth opening.

The scaling of velocity and pressure distributions depends on the physical nature of the flow and whether there

are free-stream conditions or boundary and shear layers present (George et al., 1984). Therefore, a change in the underlying scaling law is observed for spectra at locations away from the aerodynamic source region (cf. Fig. 7). Probe signals of the pressure fluctuations in the near field show a steeper slope of roughly $f^{-11/3}$, whereas spectra at probe positions in the supraglottal region and toward the acoustic far field follow $f^{-7/3}$. The $f^{-11/3}$ component arises from shear interaction in the flow, while the $f^{-7/3}$ component dominates where the approximation of free-stream conditions and turbulence-turbulence interaction holds. The velocity fluctuations in the near field follow roughly the scaling of $f^{-5/3}$, where the approximation of isotropic turbulence holds. Moving away from the source (i.e., glottis), one can observe deviations from this scaling law as can be noticed, e.g., for the velocity spectrum at location P3, located in the neighborhood of a larger rotating flow structure of low velocity.

Additionally, constrictions and expansions of the narrow airway, as well as the shear layers of the glottal jet are likely inducing anisotropy to the flow, which shows in deviations of the velocity spectra at locations P2 and P3 from the scaling of $f^{-5/3}$ for the case of isotropic turbulence in the inertial subrange. Furthermore, the spectra tend to exhibit clear peaks of the source frequency and its higher harmonics up to roughly 1000 Hz at all considered probe positions. This can be seen by comparing the investigated spectra at the various positions in Fig. 7. The amplitude of the glottal frequency is damped out throughout the upper airways as the distance from the glottis is increasing, and becomes less pronounced outside of the vocal tract in the acoustic domain (cf. Fig. 12).

B. Flow instabilities

The resulting flow through the glottis under realistic conditions of the surrounding airway shows secondary flow patterns and strong recirculation regions in the supraglottal domain. This behavior is visualized in Fig. 8 in terms of the axial velocity component at four different stages of the glottal cycle corresponding to the vocal tract configuration for vowel [a]. Flow of large negative axial velocity is moving along the anterior airway wall, opposing the flow direction of the glottal jet. Furthermore,



FIG. 7. (Color online) Power spectral density of the velocity fluctuations (a) and pressure fluctuations (b) in the near field at probe positions *P*1, *P*2, *P*3, and *P*4 (as indicated in Fig. 6). The data of the unsteady flow were extracted from the case of vowel [a].



FIG. 8. (Color online) Axial velocity in coronal and transversal planes of the supraglottal region at instances (a)–(d) of the points of reference (I)–(IV) of the glottal cycle depicted in Fig. 3. For animations of the depicted flow features, please refer to Mm 1 and Mm 2.

the flow field in the supraglottal region above the vocal folds shows a large production of vorticity due to the velocity gradient in the shear layers of the glottal jet. The movement of layers of air at different mean velocities causes the occurrence of Kelvin-Helmholtz instabilities at their interface, leading eventually to the roll up of vortices. These perturbations are formed close to the glottal exit, as depicted in Fig. 9, and convected downstream through the upper airways. A notable increase in number, range, and magnitude of scales and developed flow instabilities can be seen in Figs. 9(c) and 9(d), during the closing phase of the glottal waveform and eventual complete inhibition of flow during closing. For animations of the depicted flow features, please refer to Mm 1 and Mm 2 for axial velocity and to Mm 3 and Mm 4 for vorticity and isosurfaces.

- Mm. 1. Animation of the axial velocity component in the coronal plane section during a full glottal cycle. This is a file of type "gif" (13.1 MB).
- Mm. 2. Animation of the axial velocity component in the sagittal plane section during a full glottal cycle. This is a file of type "gif" (18.1 MB).
- Mm. 3. Animation of perpendicular vorticity component and isosurfaces in the coronal plane section. This is a file of type "gif" (18.6 MB).
- Mm. 4. Animation of perpendicular vorticity component and isosurfaces in the sagittal plane section. This is a file of type "gif" (23.8 MB).

In order to establish the energy content and spectral features of disturbances in the near field and outside of the vocal tract in the far field, frequency-wavenumber spectra are shown in Figs. 10(a) and 10(b). Dashed lines are indicating the acoustic cone, which consists of lines at a slope corresponding to the speed of sound. Thus, any pressure disturbances lying on this cone can be considered purely acoustic. While the highest energy contribution in the far field is contained inside the acoustic domain of the spectrum at low wavenumbers, the near field shows the presence of disturbances propagating at approximately uniform velocity downstream of the vocal tract. The group velocity of these disturbances in the source region, indicated by their relative slope in the frequency-wavenumber spectra, is computed from the simulation data as $c_g = 9.8$ m/s. This agrees well with the theoretical result from linear stability analysis of the flow subjected to perturbations growing in time [see, e.g., Panton (2006)]. For a first approximation, we consider nearly parallel inviscid mean flows in the x-direction with \vec{u} $=(\bar{u}+u',\bar{v}+v',w')$ and $p=\bar{p}+p'$ and subjected to arbitrary disturbances composed of normal modes,

$$u'_{i} = \hat{u}_{i}(\mathbf{y})e^{i(\alpha x + \beta z - \alpha ct)}$$
⁽⁵⁾

and

$$p' = \hat{p}(\mathbf{y})e^{i(\alpha x + \beta z - \alpha ct)},\tag{6}$$



with the wavenumbers α and β in x- and z-direction, respectively, and the complex wave speed $c = c_R + ic_I$. The

FIG. 9. (Color online) Isosurfaces of λ_2 -criterion ($\lambda_2 = -5 \times 10^6 \text{ s}^{-2}$) and vorticity component perpendicular to sagittal and transversal planes of the supraglottal region at instances (a)–(d) of the points of reference (I)–(IV) of the glottal cycle shown in Fig. 3.



FIG. 10. (Color online) Frequency-wavenumber spectra of the near-field perturbations (a) and far-field perturbations (b) for phonation of the vowel [a]. Cuts of the spectra are shown at the discrete frequencies f_0 , F_1 , and F_2 for the near field in (c) and the far field in (d). The acoustic cone is indicated as a dashed line, while the group velocity of Kelvin-Helmholtz instabilities is marked with the dashed-dotted lines in black (estimate from simulations) and red (linear stability analysis).

normal-mode disturbances are thus traveling waves of amplitudes that depend on the y-direction, perpendicular to the mean flow direction. Under these assumptions, the condition for stability is $c_I < 0$. Additionally, the dispersion relation of the system can be solved for the angular frequency ω and the exponent of the temporal growth rate $e^{\alpha c_I t}$ becomes

$$\alpha c_{I} = \left[\frac{\alpha^{2} \rho_{1} \rho_{2} \bar{u}^{2}}{\left(\rho_{1} + \rho_{2}\right)^{2}} - \frac{\alpha g(\rho_{1} - \rho_{2})}{\rho_{1} + \rho_{2}} - \frac{\gamma \alpha^{3}}{\rho_{1} + \rho_{2}} \right]^{1/2} \approx \frac{\alpha \bar{u}}{2}.$$
(7)

Furthermore, the phase velocity of the fastest moving disturbances caused by Kelvin-Helmholtz instabilities is predicted as half of the mean velocity of the jet,

$$c_{\varphi} = \frac{\omega}{k} = \frac{\alpha c_R}{k} = c_R = \frac{1}{2}\bar{u},$$
(8)

and thus ~ 11 m/s for the considered conditions of this study with the mean velocity \bar{u} of the glottal jet in streamwise direction. Following Squire (1933), any unstable 3D disturbance has a corresponding two-dimensional disturbance that is more unstable. Thus, the above approximation for propagating instabilities holds also for the 3D cases considered in this study.

Cuts through the frequency-wavenumber diagrams reveal the energy content at discrete frequencies [cf. Figs.

10(c) and 10(d)]. The broadband noise component is significantly larger in the near field than in the far field at all considered tonalities f_0 , F_1 , and F_2 . Moreover, the inharmonic fluctuations become more prominent toward higher frequencies of 1000 Hz and above.



FIG. 11. (Color online) Formant data for all investigated vowel geometries extracted through unsteady LES data and compared to the experimental values.

IV. ACOUSTICS

A. Formants

The considered vocal tract geometries for the production of the vowels [a], [e], [i], [o], and [u] are investigated in terms of their resonance frequencies (i.e., formants), by computing the spectra of pressure fluctuations in the acoustic domain. Figure 11 gives an overview of the results for F_1 and F_2 . The overall trend of formant frequencies from high to low vowels, as described by Stevens (2000), is well captured by the current direct approach and agrees with the experimental data. From the high vowel [i], over the intermediate vowel [e], to the low vowels [a], [o], [u], the formant F_2 shows a steady decrease. An exception is the frequency F_2 for vowel [e], which is slightly higher than for vowel [i] and possibly due to overestimation from the numerical spectrum. Formant F_1 on the other hand shows an increase toward the intermediate and then further decrease toward the low vowel range. Figure 12 displays the resulting power spectral density for the near- and far-field pressure fluctuations of the vowel [a] at various locations throughout the vocal tract. It can be recognized from Fig. 12 that the critical region of sound amplification in cases of the first formants F_1 and F_2 is the lower domain of the airways, and they are equally dominant in all spectra due to standing waves resonating in the vocal tract volume. The third formant F_3 , however, is less pronounced relative to F_1 and F_2 at locations P1 to P3. Toward P4 and outside the airway tract, a clearer and more significant mode develops close to the approximate location of F_3 . It is possible that the accumulation of flow perturbations over the length of the airways contributes to this effect. As such, unsteady flow might not only cause measurable increase in broadband noise, but also significantly add to the energy content of higher formant modes. The sudden increase of the relative energy content at frequencies in the range of formant F_3 in the far-field domain is probably due to sound radiation from the mouth. According to Stevens (2000), the mouth opening can be



FIG. 12. (Color online) Distribution of pressure fluctuations along the vocal tract during pronunciation of the vowel [α] at locations *P*1, *P*2, *P*3, and *P*4 (as indicated in Fig. 6). The dominant frequencies F_1 , F_2 , and F_3 are indicated.

considered as a simple source radiating uniformly in all directions as good approximation for frequencies up to 4 kHz. This leads to a radiation characteristic of magnitude $|R(f)| = f\rho/(2r)$ at a distance r. Thus the sound radiation rises with frequency and introduces a positive increasing spectral tilt of 6 dB per octave under ~4 kHz, which would explain the energy shift toward higher frequencies in the far field of Fig. 12. The frequencies from near-field turbulent flow can be distinguished in the broadband region. Although smaller in amplitude due to dissipation, the noise component of the spectrum, which is induced by the turbulence and its associated quadrupole sources, is detected outside the internal airways. As expected, the overall amplitude of power spectral density continuously decreases throughout the upper airway channels toward the mouth opening.

While the harmonic, low-frequency component of the spectrum up to about 1000 Hz shows a fairly small decrease from position P1 to P4 inside the upper airways, the magnitude of fluctuations in the inharmonic, high-frequency range above 1000 Hz is reduced significantly. Thus, the broadband sound component in Fig. 12 is decreased with increased distance from its source in the supraglottal region, while the formant modes can be detected as peaks of nearly constant amplitude throughout the vocal tract.

B. Fourier surface spectra

Fourier analysis of the surface pressure fluctuations at the vocal tract walls highlights the areas of amplification through resonance, as well as acoustic surface sources generated by the unsteady pressure loads on the airway walls caused by the intermittent flow interacting with the walls. The surface acoustic spectra shown in Fig. 13 reveal the SPLs at the discrete vocal tract modes for the vowels [i], [e], [a], [o], and [u] in order of tongue position from high to low vowels. The SPL at the surface is normalized by its maximum value for each of the vowel geometries to allow for better comparison. This is a way to visualize the impact of the various parts of the airway geometry on the production and filtering of the vowel formants. Because it is based on total fluctuating pressure at the walls, the surface Fourier transformation takes into account the acoustic fluctuations through resonance effects and standing wave patterns, as well as the fluctuations caused by near-field flow motion.

As a general trend it can be seen that lower formant frequencies are amplified predominantly in the *laryngopharynx*, *oropharynx*, or back cavities. The oral cavity seems to further contribute to the amplification of formant F_1 in the configurations of the lower vowels [0] and [u], where it takes up a large volume, while the lower part of the *pharynx* is the crucial region for F_1 with the high vowel [i]. The normalized SPLs of the first formant F_1 of the vowels [0], [0], and [u] are fairly evenly distributed with peak values in the lower pharynx for [0] and [0] that move toward the mouth opening for [u].

The mode at F_2 tends to be related to higher sound pressures in domains of smaller spatial dimensions such as the oral and nasal cavities (i.e., the front cavities) in the airway



FIG. 13. (Color online) Normalized surface SPLs of the near-field acoustic pressure for the frequencies F_1 and F_2 of the vowels [i], [e], [a], [o], and [u] in (a)–(e). The SPL at the surface of the airway walls has been normalized by its maximum for each of the vocal tract cases.

of vowel [a]. In the case of the vowel [u], the formant F_2 is related predominantly to the back cavity of the pharynx.

V. CONCLUSIONS

In this study, the mechanisms of sound pressure propagation and amplification through realistic upper airway geometries of several natural vowels were investigated through means of LES computations. By applying the Rosenberg glottal pulse model for the pulsating flow and pressure at the level of the vocal folds, the acoustic source signal was propagated throughout the upper airways and subsequently modified by the airway volume. Both the harmonic, low-frequency component, as well as the inharmonic, high-frequency component of pressure fluctuations arising during phonation are thus computed directly in the near and far fields of the computational domain.

The particular case of vowel [a] was considered with respect to the production and convection of Kelvin-Helmholtz instabilities in the supraglottal region, which shows a sharp increase during the glottal closing phase and propagates through the vocal tract at a maximum phase velocity equal to half the glottal jet velocity (cf. Sec. III B). The Reynolds number in the flow reaches a maximum at the peak of the waveform with fast acceleration of the flow through the narrow glottis short before the closing phase starts. A rise of vortical structures shed from the glottal constriction is observed during the closing phase. Flow structures develop particularly in the supraglottal jet shear layers due to flow instabilities caused by large velocity gradients. The pressure fluctuations outside the acoustic cone are larger in amplitude for higher frequencies and their slope in the frequency-wavenumber spectrum follows the estimated group velocity from linear stability analysis.

Flow separation and vortex shedding are known to take place in the intraglottal region, especially for a divergent glottal shape (Mihaescu *et al.*, 2010). Vortices and instabilities might therefore propagate from the intraglottal region into the supraglottal region and contribute to the near-field pressure fluctuations. This effect, which is due to the viscous airflow and particular divergent shape of the vocal folds during closing, is similarly reproduced in our time-averaged glottal geometry (cf. Figs. 6, 8, and 9).

The frequency contribution and energy content of the supraglottal flow structures at higher frequencies of 1000 Hz and above might bleed into higher formant tonalities of F_3 and above (cf. Sec. IV A) and thus impact not only the broadband noise component at very high frequency but also the position and intensity of higher formants relative to lower ones. In the current study, the acoustic contribution of broadband sound to the voice spectrum is resolved to the upper end of the audible range. The high-frequency component of the sound spectrum is important for voice quality and the personal characteristics of a voice, referred to as timbre (Sundberg, 1994). It is also of relevance for professional singers, whose voices commonly exhibit a frequency peak of high energy in the spectral envelope at about 3000 Hz, referred to as singer's formant (Sundberg, 2001). Furthermore, Kreiman and Gerratt (2012) pointed out the perceptual importance of the inharmonic broadband component to voice quality. The noise-toharmonic ratio (NHR) was found to be an important indicator of general intelligibility and *breathiness* of a voice (de Krom, 1995; Hillenbrand, 1988).

It could be further shown that both spectra of near- and far-field pressure perturbations have significant peaks at the dominant frequencies F_1 and F_2 (cf. Figs. 5 and 12). For the investigated vowels [i], [e], [a], [o], and [u] the formants are accurately predicted and show good agreement with the simultaneous sound recordings obtained during the MRI measurements (cf. Sec. IV A).

Localized resonances of individual formants have been studied using predominantly one-dimensional acoustic models. Fant (1971) was the first to find cavity affiliations of vowel formants by simplifying the vocal tract as a system of tubes or Helmholtz resonators. Thus, the formants could be interpreted as resonances of the various cavities of the vocal tract. Consequently, Fant and Pauli (1974) used the onedimensional Webster's horn equation to extract the spatial distribution of kinetic and potential energies at particular resonance modes inside the vocal tract configurations of various vowels. In the current study, the specific areas of amplification of the formants are identified by means of Fourier transformation of the near-field pressure fluctuations at the airway surfaces (cf. Sec. IV B). They give 3D information on the relevant areas of sound generation in the upper airways that cannot be obtained through one-dimensional models. Furthermore, they entail both sound due to resonance and surface pressure fluctuations caused by impacting flow, which is not captured by models based on wave equations.

Two factors are essential for the acoustic characteristics of a healthy voice: the voice source, which requires functioning vocal folds, on the one hand, and the vocal tract with parts of the nasal tract acting as resonators on the other hand (Sundberg, 2005). If parts of the vocal tract are affected by paralysis (e.g., of the tongue), tumors, or tissue inflammation, its ability to produce acoustic modes in the far-field spectrum might be compromised due to changes of the related resonance volume. As an example, inflammation might affect the pharynx through acute epiglottitis, in which case swelling of the airway can lead to stenosis. This potentially lethal condition inhibits regular vocal articulation as well as breathing (Omori, 2011). Another condition is incomplete velopharyngeal closure, which leads to increased nasal airflow and resonance causing nasality (or hyperrhinophony) of the speech sounds (Hirschberg et al., 1995). Due to dysfunction of the oral motor mechanism, as it may occur in paresis, the velopharyngeal port in the nasopharynx does not close completely, which disturbs the pharyngo-oral airflow necessary for the pronunciation of oral sounds. In all these cases, regular speech is inhibited due to deviations of the airflow and the associated resonance volume.

A complete, geometrically detailed mapping of vocal tract resonances at the formant frequencies as presented in Sec. IV B gives information on the 3D cavity-association of vowel sounds in a healthy state. Spatially and temporally resolved velocity and pressure fields show the impact of geometrical features as well as of the time-varying source signal. This information can be of use as reference data for clinical purposes and comparison with similar data from patients with impaired vocal function, making it more

important to resolve the complete frequency range relevant for human phonation through high-order LES computations.

ACKNOWLEDGMENTS

This work is supported by the Swedish Research Council (Vetenskapsrådet) through the Grant No. VR 621-2012-4256. Computational resources were provided by the Swedish National Infrastructure for Computing (SNIC). The members of the PUMA group lead by Risto-Pekka Happonen, Department of Oral and Maxillofacial Surgery at University of Turku, Riitta Parkkola, Jani Saunavaara, and Daniel Aalto are thanked for their cooperation in the MRI data acquisition. The MRI data collection procedure was accepted by the Ethics Committee of the Hospital District of Southwest Finland.

- Aalto, D., Aaltonen, O., Happonen, R.-P., Jääsaari, P., Kivelä, A., Kuortti, J., Luukinen, J.-M., Malinen, J., Murtola, T., Parkkola, R., Suanavaara, J., Soukka T., and Vainio M. (2014). "Large scale data acquisition of simultaneous MRI and speech," Appl. Acoust. 83, 64–75.
- Aalto, D., Helle, J., Huhtala, A., Kivelä, A., Malinen, J., Saunavaara, J., and Ronkka, T. (2013). "Algorithmic surface extraction from MRI datamodelling the human vocal tract," in 6th International Conference on Biomedical Electronics and Devices, Barcelona, Spain, pp. 257–260.
- Aalto, D., Huhtala, A., Kivelä, A., Malinen, J., Palo, P., Saunavaara, J., and Vainio, M. (2012). "How far are vowel formants from computed vocal tract resonances?," arXiv:1208.5963.
- Aalto, D., Malinen, J., Palo, P., Aaltonen, O., Vainio, M., Happonen, R.-P., Parkkola, R., and Saunavaara, J. (2011). "Recording speech sound and articulation in MRI," in *International Conference on Biomedical Electronics and Devices*, Rome, Italy, pp. 168–173.
- Arnela, M., Dabbaghchian, S., Blandin, R., Guasch, O., Engwall, O., Van Hirtum, A., and Pelorson, X. (2016). "Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds," J. Acoust. Soc. Am. 140(3), 1707–1718.
- Blandin, R., Arnela, M., Laboissière, R., Pelorson, X., Guasch, O., Hirtum, A. V., and Laval, X. (2015). "Effects of higher order propagation modes in vocal tract like geometries," J. Acoust. Soc. Am. 137(2), 832–843.
- Celik, I. B., Ghia, U., Roache, P. J., Freitas, C. J., Coleman, H., and Raad, P. E. (2008). "Procedure for estimation and reporting of uncertainty due to discretization in CFD applications," J. Fluids Eng. 130(7), 1–4.
- de Krom, G. (1995). "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments," J. Speech Lang, Hear. Res. 38(4), 794–811.
- de Luzan, C. F., Chen, J., Mihaescu, M., Khosla, S. M., and Gutmark, E. (2015). "Computational study of false vocal folds effects on unsteady airflows through static models of the human larynx," J. Biomech. 48(7), 1248–1257.
- Fant, G. (1971). Acoustic Theory of Speech Production (Walter de Gruyter, Berlin), Vol. 2.
- Fant, G., Liljencrants, J., and Lin, Q.-G. (1985). "A four-parameter model of glottal flow," STL-QPSR 24(4), 1–13.
- Fant, G., and Pauli, S. (1974). "Spatial characteristics of vocal tract resonance modes," in *Speech Comm. Sem.*, Aug., Stockholm, pp. 1–3.
- Fujisaki, H., and Ljungqvist, M. (1986). "Proposal and evaluation of models for the glottal source waveform," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'86*, Vol. 11, pp. 1605–1608.
- George, W. K., Beuther, P. D., and Arndt, R. E. (1984). "Pressure spectra in turbulent free shear flows," J. Fluid Mech. 148, 155–191.
- Haigermoser, C. (2009). "Application of an acoustic analogy to PIV data from rectangular cavity flows," Exp. Fluids 47(1), 145–157.
- Hillenbrand, J. (1988). "Perception of aperiodicities in synthetically generated voices," J. Acoust. Soc. Am. 83(6), 2361–2371.
- Hirschberg, J., Dejonckere, P., Hirano, M., Mori, K., Schultz-Coulon, H.-J., and Vrtička, K. (1995). "Voice disorders in children," Int. J. Pediatr. Otorhinolaryngol. 32, S109–S125.

- Howe, M., and McGowan, R. (2007). "Sound generated by aerodynamic sources near a deformable body, with application to voiced speech," J. Fluid Mech. 592, 367–392.
- Khosla, S., Muruguppan, S., Gutmark, E., and Scherer, R. (2007). "Vortical flow field during phonation in an excised canine larynx model," Ann. Otol. Rhinol. Laryngol. 116(3), 217–228.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. 87(2), 820–857.
- Koschatzky, V., Moore, P., Westerweel, J., Scarano, F., and Boersma, B. (2011). "High speed PIV applied to aerodynamic noise investigation," Exp. Fluids 50(4), 863–876.
- Kreiman, J., and Gerratt, B. R. (2012). "Perceptual interaction of the harmonic source and noise in voice," J. Acoust. Soc. Am. 131(1), 492–500.
- Lodermeyer, A., Tautz, M., Becker, S., Döllinger, M., Birk, V., and Kniesburges, S. (2018). "Aeroacoustic analysis of the human phonation process based on a hybrid acoustic PIV approach," Exp. Fluids 59(13), 1–15.
- Mihaescu, M., Khosla, S. M., Murugappan, S., and Gutmark, E. J. (2010). "Unsteady laryngeal airflow simulations of the intra-glottal vortical structures," J. Acoust. Soc. Am. 127(1), 435–444.
- Mihaescu, M., Mylavarapu, G., Gutmark, E. J., and Powell, N. B. (2011). "Large eddy simulation of the pharyngeal airflow associated with obstructive sleep apnea syndrome at pre and post-surgical treatment," J. Biomech. 44(12), 2221–2228.
- Mittal, R., Erath, B. D., and Plesniak, M. W. (2013). "Fluid dynamics of human phonation and speech," Annu. Rev. Fluid Mech. 45, 437–467.
- Murray, P. R., Thomson, S. L., and Smith, M. E. (2014). "A synthetic, selfoscillating vocal fold model platform for studying augmentation injection," J. Voice 28(2), 133–143.
- Mylavarapu, G., Mihaescu, M., Fuchs, L., Papatziamos, G., and Gutmark, E. (2013). "Planning human upper airway surgery using computational fluid dynamics," J. Biomech. 46(12), 1979–1986.
- Neubauer, J., Zhang, Z., Miraghaie, R., and Berry, D. A. (2007). "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. 121(2), 1102–1118.
- Nicoud, F., and Ducros, F. (1999). "Subgrid-scale stress modelling based on the square of the velocity gradient tensor," Flow Turbul. Combust. 62(3), 183–200.
- Nussbaumer, H. J. (2012). Fast Fourier Transform and Convolution Algorithms (Springer Science and Business Media, Heidelberg, Germany), Vol. 2.
- Ojalammi, A., and Malinen, J. (2017). "Automated segmentation of upper airways from MRI—Vocal tract geometry extraction," in *Proceedings of* the 10th International Joint Conference on Biomedical Engineering Systems and Technologies—Volume 2: BIOIMAGING (BIOSTEC 2017), pp. 77–84.
- Omori, K. (2011). "Diagnosis of voice disorders," Japan Med. Assoc. J. 54, 248–253.
- Oren, L., Khosla, S., and Gutmark, E. (2014). "Intraglottal pressure distribution computed from empirical velocity data in canine larynx," J. Biomech. 47(6), 1287–1293.
- Panton, R. L. (2006). Incompressible Flow (Wiley, Hoboken, NJ).
- Pope, S. B. (2001). *Turbulent Flows* (Cambridge University Press, Cambridge, UK).
- Rofsky, N. M., Lee, V. S., Laub, G., Pollack, M. A., Krinsky, G. A., Thomasson, D., Ambrosino, M. M., and Weinreb, J. C. (1999). "Abdominal MR imaging with a volumetric interpolated breath-hold examination," Radiology 212(3), 876–884.

- Rosenberg, A. E. (1971). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. 49(2B), 583–590.
- Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (2001). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. 109(4), 1616–1630.
- Schickhofer, L., Dahlkild, A., and Mihaescu, M. (2016). "Aeroacoustics of an elastic element in unsteady flow of low Reynolds numbers," AIAA Paper No. 2016–2700.
- Šidlof, P., Zörner, S., and Hüppe, A. (2015). "A hybrid approach to the computational aeroacoustics of human voice production," Biomech. Model. Mechanobiol. 14(3), 473–488.
- Sliwinska-Kowalska, M., Niebudek-Bogusz, E., Fiszer, M., Los-Spychalska, T., Kotylo, P., Sznurowska-Przygocka, B., and Modrzewska, M. (**2006**). "The prevalence and risk factors for occupational voice disorders in teachers," Folia Phoniatr. Logop. **58**(2), 85–101.
- Squire, H. B. (1933). "On the stability for three-dimensional disturbances of viscous fluid flow between parallel walls," Proc. R. Soc. London A 142(847), 621–628.
- Stevens, K. N. (2000). Acoustic Phonetics (MIT Press, Cambridge, MA), Vol. 30.
- Story, B. H., Titze, I. R., and Hoffman, E. A. (1996). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. 100(1), 537–554.
- Sundberg, J. (1994). "Perceptual aspects of singing," J. Voice 8(2), 106–122.
- Sundberg, J. (2001). "Level and center frequency of the singer's formant," J. Voice 15(2), 176–186.
- Sundberg, J. (2005). "Vocal tract resonance," in *Professional Voice The Science and Art of Clinical Care*, edited by R. T. Sataloff (Plural Publishing, San Diego, CA).
- Syndergaard, K. L., Dushku, S., and Thomson, S. L. (2017). "Electrically conductive synthetic vocal fold replicas for voice production research," J. Acoust. Soc. Am. 142(1), EL63–EL68.
- Titze, I. R., and Alipour, F. (2006). *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Salt Lake City, UT).
- Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., Howard, D. M., Hunter, E. J., Kaelin, D., Kent, R. D., Kreiman, J., Kob, M., Löfqvist, A., McCoy, S., Miller, D. G., Noé, H., Scherer, R. C., Smith, J. R., Story, B. H., Švec, J. G., Ternström, S., and Wolfe, J. (2015). "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization," J. Acoust. Soc. Am. 137(5), 3005–3007.
- Triep, M., and Brücker, C. (2010). "Three-dimensional nature of the glottal jet," J. Acoust. Soc. Am. 127(3), 1537–1547.
- Zhang, Z. (2016). "Mechanics of human voice production and control," J. Acoust. Soc. Am. 140(4), 2614–2635.
- Zhang, Z., Mongeau, L., and Frankel, S. H. (2002). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. 112(4), 1652–1663.
- Zhang, Z., and Neubauer, J. (2010). "On the acoustical relevance of supraglottal flow structures to low-frequency voice production," J. Acoust. Soc. Am. 128(6), EL378–EL383.
- Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (2002). "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. 112(5), 2134–2146.
- Zörner, S., Šidlof, P., Hüppe, A., and Kaltenbacher, M. (2016). "Flow and acoustic effects in the larynx for varying geometries," Acta Acust. Acust. 102(2), 257–267.