An acoustic glottal source for vocal tract physical models

Antti Hannukainen¹, Juha Kuortti¹, Jarmo Malinen¹,² and Antti Ojalammi¹

¹ School of Science, Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland
² School of Electrical Engineering, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

E-mail: jarmo.malinen@aalto.fi

Received 24 May 2017, revised 24 July 2017
Accepted for publication 11 August 2017
Published 17 October 2017

Abstract

A sound source is proposed for the acoustic measurement of physical models of the human vocal tract. The physical models are produced by fast prototyping, based on magnetic resonance imaging during prolonged vowel production. The sound source, accompanied by custom signal processing algorithms, is used for two kinds of measurements from physical models of the vocal tract: (i) amplitude frequency response and resonant frequency measurements, and (ii) signal reconstructions at the source output according to a target pressure waveform at the mouth position. The proposed source and the software are validated by computational acoustics experiments and measurements on a physical model of the vocal tract corresponding to the vowels [a] of a male speaker.

Keywords: acoustic measurements, speech research, vowels, glottal source, physical models

(Some figures may appear in colour only in the online journal)

1. Introduction

Acquiring comprehensive data from human speech is a challenging task that is crucial for modelling speech production and developing speech signal processing algorithms. The possible approaches can be divided into direct and indirect methods. Direct methods concern measurements carried out on test subjects either by audio recordings, acquisition of pressure, flow velocity, or electrical signals (as takes place in electroglottography), or using different methods of medical imaging during speech. Indirect methods concern simulations using computational models (such as those described in [1] and the references therein) or measurements from the physical models of the vocal tract [1]. Typically, computational and physical models are created and evaluated based on data that has first been acquired by direct methods. The main advantage of indirect methods is the absence of the human component that leads to experimental restrictions and unwanted variation in the data quality.

The purpose of this article is to describe an experimental arrangement, its validation and some experiments on one type of physical model for vowel production, involving acoustic resonators corresponding to vocal tract (VT) configurations during prolonged vowel utterance. Resonant frequencies extracted from the measured frequency responses are used for the development and validation of acoustic and phonation models such as the one introduced in [1]. The synthetic vowel signals are produced by reconstructing typical glottal pulse waveforms at the vocal folds position of the model. These signals are intended to benchmark the glottal inverse filtering (GIF) algorithms, as was done in [9, 10]. Large amounts of measurement data are required for these applications, which also imposes practical requirements on the measurement arrangement.

The anatomic geometry for the VT physical models was acquired by magnetic resonance imaging (MRI) with simultaneous speech recordings [2, 3]. The MRI voxel data was processed into surface models, as explained in [4], and then printed in ABS plastic by rapid prototyping, as explained below in section 2.5. In itself, the idea of using 3D printed VT models in speech research is by no means new: see, e.g. [5–7]. Just creating physical models of the VT is not enough for model experiments. A suitable acoustic signal source is required with custom instrumentation and software. As these experiments involve a niche area in speech research, directly

1 Physical models are understood as artefacts or replicas of parts of the speech anatomy in the context of this article.
applicable commercial solutions do not exist and constructing a custom measurement suite looks to be an attractive option. We point out that high-quality acoustic measurements can be carried out on physical models of the VT without using a custom-built sound source, see [7, figure 3], where the sound pressure is fed into the model through the mouth opening, and the measurements are carried out using a microphone at the vocal fold position. However, excitation from the vocal folds position is desirable because the face and the exterior space acoustics are issues. Thus, we propose an acoustic glottal source design, shown in figure 1, which resembles the loudspeaker-horn constructions shown in [5, figure 1], [6, figure 3], [8, figure 2a]. All such loudspeaker-horn constructions can be regarded as variants of the compression drivers that are often used as high impedance sources for horn loudspeakers. Unfortunately, most commercially available compression drivers are designed for frequencies over 500 Hz, whereas a device based on a loudspeaker unit can be scaled up to lower frequencies, as required in speech research.

The principle of operation of the proposed acoustic glottal source is fairly simple. The source consists of a loudspeaker unit and an impedance matching horn, as shown disassembled in figure 1(b). The purpose of the horn is to concentrate the acoustic power from the low-impedance loudspeaker to an opening of diameter 6 mm, which is the high-impedance output of the source. There are, however, a number of conflicting design objectives that need to be taken into account in a satisfactory way. The device should not be impractically large at least, and it should also be usable for the measurements of physical models of human VTs in the frequency range of interest, specified as 80 . . . 7350 Hz in this article. To achieve these goals in a meaningful manner, we use a design methodology involving (i) heuristic reasoning based on mathematical acoustics, together with (ii) numerical acoustics modelling of the main components and their interactions. Numerical modelling of all the details is not necessary for a successful outcome, but it helps in understanding the problems requiring attention. Just optimising the source performance by trial and error would lead to overly time-consuming measurements and laboratory work.

The design and construction process was incremental, and it consisted of the following steps, which were repeated when necessary:

(i) The choice of the acoustic design and the main components, based on the general principles of acoustics, horn design and feasibility.
(ii) The finite element (FEM)-based modelling of the horn acoustics to check the overall validity of the approach, to detect and correct the expected problems in construction.
(iii) The construction of the horn and the loudspeaker assembly together with the required instrumentation.
(iv) A cycle of measurements and modifications, such as the placement of acoustically soft material and silicone sealings in various parts based on, e.g. the FEM modelling.
(v) The development of MATLAB software for producing properly weighted measurement signals for sweep experiments that compensate most of the remaining non-idealities.
(vi) The development of MATLAB software for reproducing the Liljencrants–Fant (LF) glottal waveform excitation at the glottal position of the physical models.

The proposed source is used to measure the frequency responses of the physical models of the VT during the utterance of Finnish vowels [a, i, u], obtained from a 26-year-old male (in fact, one of the authors of this article). The measured amplitude frequency responses are compared with the spectral envelope data from the vowel samples shown in figure 10, recorded in an anechoic chamber from the same test subject. In addition to these responses, the vowel signal is produced by acoustically exciting the physical models by a glottal pulse waveform of the LF type, numerically reconstructed at the output of the source. The produced signals for the vowels [a, i, u] each have good audible resolution, yet they also have the distinct ‘robotic’ sound quality that is typical of most synthetically produced speech.

So, as to the physical dimension of the measured signals, this article is restricted to sound pressure measurements using microphones. If acoustic impedances are to be measured instead, some form of acoustic (perturbation) velocity measurement needs to be carried out. The velocity measurement can be carried out, e.g. by hot wire anemometers [11], impedance heads consisting of several microphones [12], or even by a single microphone using a resistive calibration load coupled to a high impedance source [13]; see [14, table 1] for various approaches. In general, carrying out velocity measurements is
much more difficult and expensive than just measuring sound pressure. Determining the pressure-to-pressure responses of VT physical models is, however, sufficient for the purposes of this article, since (i) the resonant frequencies can be determined from pressures, and (ii) the GIF algorithm can be configured to run on sound pressure data.

The organisation of this article is as follows. The background in mathematical acoustics, acoustic transmission lines, horns, and deconvolution theory is reviewed in section 2. The reader may prefer to start reading from section 3 where the design objectives for the acoustic source, approaches including computational modal analysis, the construction of the instruments, and the general experimental arrangement are explained. In section 4, the acoustic robustness of the proposed design is studied in terms of numerical modal perturbation analysis, so as to quantify to what extent the source acoustics may perturb the resonant behaviour of the VT physical model being tested. The frequency response of the acoustic source is far from flat, and the source is not even quite linear. Thus, measured calibration weights and source compensation by deconvolution and impulse response estimation are required; see section 5. With the aid of these approaches, it is possible to accurately measure some of the speech-relevant acoustic characteristics of the physical models of the VT, both in the frequency and time domain. Such results are given in section 6.

2. Background

We review the relevant aspects from mathematical acoustics, horn design, signal processing, and MRI data acquisition.

2.1. Acoustic equations for horns

Acoustic horns are impedance matching devices that can be described as surfaces of the revolution in a three-dimensional space. Thus, they are defined by the strictly nonnegative continuous functions \( r = R(x), \) where \( x \in [0, \ell], \ell > 0 \) is the length of the horn and \( r \) denotes the radius of the horn at \( x. \) The end \( x = 0 \) \((x = \ell)\) is the input end (respectively, the output end) of the horn. It is typical, though not necessary, for the function \( R(\cdot) \) to either increase or decrease.

There exists a large amount of literature on horns for loudspeakers; see, e.g. [15–18]. As a general rule, the matching impedance at the end of a horn is inversely proportional to the opening area. For uniform diameter waveguides, the matching impedance coincides with the characteristic impedance given by \( Z_0 = \rho c/A_0, \) where \( A_0 \) is the intersectional area. The constant \( c \) denotes the speed of sound and \( \rho \) is the density of the medium. In this article, the values \( \rho = 1.3 \text{ kg m}^{-3}, c = 345 \text{ m s}^{-1} \) are used.

To describe the acoustics of an air column in a horn, we use two (partial) differential equations. The three-dimensional acoustics is described by the lossless Helmholtz equation in terms of the velocity potential

\[
\lambda^2 \phi_\lambda = c^2 \Delta \phi_\lambda \text{ on } \Omega \quad \text{and} \quad \frac{\partial \phi_\lambda}{\partial \nu}(\mathbf{r}) = 0 \text{ on } \partial \Omega \setminus \Gamma_0
\]  

(1)

where the acoustic domain is denoted by \( \Omega \subset \mathbb{R}^3 \) with the boundary \( \partial \Omega. \) A part of the boundary, denoted by \( \Gamma_0, \) is singled out as an interface to the exterior space. The boundary condition on \( \Gamma_0, \) modelling the interaction with the surrounding space, can be chosen in many ways, depending on the application. For the purposes of this section and section 3.1, we use the Dirichlet boundary condition

\[
\phi_\lambda(\mathbf{r}) = 0 \text{ on } \Gamma_0.
\]  

(2)

Equations (1) and (2) have a countably infinite number of solutions \((\lambda_j, \phi_j) = (\lambda_j, \phi_j) \in \mathbb{C} \times H^1(\Omega) \setminus \{0\}, j = 1, 2, \ldots, \) and each of the solutions is associated with a Helmholtz resonant frequency \( f_j(\Omega) \) by \( f_j = \text{Im}[\lambda_j/2\pi]. \)

In addition to the acoustic resonance, the acoustic transmission impedance of the source is important. Because it is more practical to deal with single-input single-output (SISO) impedances, we use the lossless Webster resonance model to define it. In terms of Webster’s velocity potential, the transmission impedance is given for any \( s \in \mathbb{C} \) by

\[
s^2 \psi_s = \frac{c^2}{A(x)} \frac{\partial}{\partial x} \left( A(x) \frac{\partial \psi_s}{\partial x} \right) \text{ on } [0, \ell],
\]

\[-A(0) \frac{\partial \psi_s}{\partial x}(0) = \hat{i}(s), \text{ and } R_L A(\ell) \frac{\partial \psi_s}{\partial x}(\ell) = \rho s \phi_s(\ell)
\]  

(3)

where \( A(x) = \pi R(x)^2 \) is the intersectional area of the horn, \( \rho \) is the density of air and \( R_L \geq 0 \) is the termination resistance at the output end \( x = \ell. \) The frequencies and Laplace transform domain \( s \) variables are related by \( f = \text{Im} s/2\pi. \) The function \( \hat{i}(s) \) is the Laplace transform of the (perturbation) volume velocity used to drive the horn, and the output is given as the Laplace transform of the sound pressure \( \hat{p}(s) = \rho s \phi_s(\ell). \) The transmission impedance of the horn, terminated to the resistance \( R_L > 0, \) is given by

\[
Z_{R_L}(s) = \hat{p}(s)/\hat{i}(s) \text{ for all } s \in \mathbb{C}.
\]  

(4)

Note that when solving equation (3) for a fixed \( s, \) we may, by linearity, choose \( \hat{i}(s) = 1 \) when plainly \( Z_{R_L}(s) = \rho s \phi_s(\ell). \) As the impedance of a passive system, the transmission impedance satisfies the positive real condition

\[\Re Z_{R_L}(s) \geq 0 \text{ for all } s \in \mathbb{C}^+ := \{ s \in \mathbb{C} : \Re s > 0 \}.\]

2.2. Suppression of transversal modes in horns

By transversal modes we refer to the resonant standing wave patterns in a horn, where significant pressure variation takes place perpendicular to the horn axis, as opposed to purely longitudinal modes. The purpose of this section is to argue why transversal modes are undesirable from the point of view of acoustic source design.

As a well-known special case, consider a waveguide of length \( \ell \) that has a constant diameter, i.e. \( A(x) = A_0. \) Then the transmission impedance given by equations (3) and (4) can be given the explicit formula

\[\text{Because the external termination resistance } R_L \text{ is the only loss term in equation (3), we call the model lossless.}\]
2.3. Minimisation of backward reflections

When a horn is excited from its input end, some of the excitation energy is reflected back to the source with delays. For horns of finite length $\ell$, there are two kinds of backward reflection. Firstly, the geometry of the horn may cause distributed backward reflections over the length of the horn. Secondly, there may be backward reflections at the output end of the horn, depending on the acoustic impedance seen by the horn at the termination point $x = \ell$ in equation (3). We next consider only the backward reflections of the first kind since only they can be affected by the horn design.

Because the acoustics of the horn described by equations (1)-(3) is internally lossless, minimising the transmission loss (TL) amounts to minimising the backward reflections that take place inside the horn. This is a classical shape optimisation problem in designing horns. Modern numerical approaches are based on topology optimisation techniques as presented in e.g. [18, 21], where other design objectives are also usually taken into account.

We take another approach and use analytic geometry and physical simplifications of the wave propagation to design the function $r = R(x)$ on $[0, \ell]$, following Paul Voigt, who proposed a family of tractrix horns in his patent ‘Improvements in horns for acoustic instruments’ in 1926, see [22]. His invention was to use the surface of revolution of the tractrix curve, given by

$$x = a \ln \frac{a + \sqrt{a^2 - r^2}}{r} - \sqrt{a^2 - r^2}, \quad r \in [0, a]$$

where $a > 0$ is the parameter specifying the radius of the wide end. Obviously, equation (6) defines a decreasing function $x \mapsto R(x) = r$ mapping $R : [0, \infty) \to (0, a]$ with $R(0) = a$ and $\lim_{x \to \infty} R(x) = 0$, which defines the tractrix
horn by its radius at $x$. The length $\ell > 0$ of the horn is solved from $R(\ell) = b$ where $0 < b < a$ is the required radius of the narrow end.

The tractrix horn is known as the pseudosphere of the constant negative Gaussian curvature in differential geometry. That it acts as a spherical wave horn is based on a geometric property of equation (6). More precisely, it can be seen from figure 2(a) that a spherical wavefront with a curvature radius of $a$, propagating along the centreline of the horn, meets the tractrix horn surfaces at right angles. Disregarding, for example, the viscosity effects in the boundary layer near the horn surface, the right angle property is expected to produce minimal backward reflections for spherical waves. This behaviour is similar to how a planar wavefront behaves in a constant diameter, uncurved waveguide.

Table 1. The number of tetrahedrons of the three FEM meshes used for the resonance computations in sections 3.1 and 4. The degrees of freedom indicate the size of the resulting system of linear equations.

| Mesh                        | Tetrahedrons | d.o.f. |
|-----------------------------|--------------|--------|
| Impedance matching cavity   | 71 525       | 15 246 |
| Cavity joined with VT       | 175 946      | 38 020 |
| VT                          | 97 847       | 21 745 |

2.4. Regularised deconvolution

A desired sound waveform target pattern will be reconstructed at the source output by compensating the source dynamics in section 5.2. Our approach is based on the constrained least squares filtering used in digital image processing [23, 24].
Suppose that a linear, time-invariant system has the real-valued impulse response \( h(t) = h_0(t) + h_e(t) \), which is expected to contain some measurement error \( h_e(t) \). When the input signal \( u = u(t) \) is fed to the system, the measured output is obtained from

\[
y(t) = (h_0 * u)(t) + v(t) \quad \text{with} \quad v = h_e * u + w \quad \text{for} \quad t \in [0, T].
\]

As usual, the convolution is defined by

\[
(h_0 * u)(t) = \int_{-\infty}^{\infty} h_0(t - \tau) u(\tau) \, d\tau,
\]

and our task is to estimate \( u \) from equation (7) given \( y \) and some incomplete information about the output noise \( v \). We assume \( u, v \in L^2(0, T) \) and that \( h_0 \) is a continuous function. We define the noise level parameter by \( \epsilon = \| v \|_{L^2(0,T)} / \| y \|_{L^2(0,T)} \) and require that \( 0 < \epsilon < 1 \) holds.

Unfortunately, equation (7) is typically unsolvable even for smooth \( y \), since the noise \( v \) may fail to be continuous, whereas the convolution operator \( h_0 * \) is smoothing. Instead of solving equation (7), we solve an estimate \( \hat{u} \) for \( u \) from the regularised version of equation (7), given for \( y \in L^2(0, T) \) by

\[
\text{Arg min}_{\hat{u}} \left( \kappa \| \hat{u} \|^2_{L^2(0,T)} + \| \hat{u}' \|^2_{L^2(0,T)} \right)
\]

with the constraint

\[
\| y - h_0 * \hat{u} \|_{L^2(0,T)} = \epsilon \| y \|_{L^2(0,T)}.
\]

Here \( T > 0 \) is the sample length, \( \kappa > 0 \) is a regularisation parameter and \( \epsilon \) is the noise level introduced above in view of \( v \) in equation (7). Obviously, it is not generally possible to choose \( \epsilon = 0 \) in equation (8) since \( y = h_0 * u \) could be unsolvable in \( L^2(0, T) \).

The Lagrangian function of equation (8) has the form

\[
L_e(\hat{u}, \mu) = \kappa \| \hat{u} \|^2_{L^2(0,T)} + \| \hat{u}' \|^2_{L^2(0,T)} - \mu \left( \| y - h_0 * \hat{u} \|^2_{L^2(0,T)} - \epsilon^2 \| y \|^2_{L^2(0,T)} \right).
\]

Using the variation \( \delta L_e \) we get

\[
\frac{d}{d\eta} L_e(\hat{u}_\eta, \mu) \bigg|_{\eta = 0} = 2 \text{Re} \left( \kappa \langle w, \hat{u} \rangle_{L^2(0,T)} + \langle w', \hat{u}' \rangle_{L^2(0,T)} - \mu \langle h_0 * w, y - h_0 * \hat{u} \rangle_{L^2(0,T)} \right) = 0
\]

for all test functions \( w \in \mathcal{D}([0, T]) \). Thus

\[
\kappa \langle w, \hat{u} \rangle + \langle w', \hat{u}' \rangle - \mu \langle h_0 * w, y - h_0 * \hat{u} \rangle = 0
\]

which, after partial integration and adjoining the convolution operator \( h_0 * \) gives

\[
\kappa \hat{u} + \hat{u}^{(4)} - \mu (h_0 * y) = 0
\]

leading to the normal equation

\[
\hat{u} = \left( \gamma + \left( \frac{d^4}{dt^4} \right) + (h_0 *)^* \right)^{-1} (h_0 *)^* y
\]

(9)

together with the constraint \( \| y - h_0 * \hat{u} \|_{L^2(0,T)} = \epsilon \| y \|_{L^2(0,T)} \). For each \( \kappa \), we have

\[
\| v_{\kappa,0} \|_{L^2(0,T)} = \| y \|_{L^2(0,T)} \quad \text{and} \quad \lim_{\kappa \to \infty} \| v_{\kappa,\mu} \|_{L^2(0,T)} = 0.
\]

By continuity and the inequality \( 0 < \epsilon < 1 \), there exists a \( \mu_0 = \mu_0(\epsilon, \kappa) \) such that \( \| v_{\kappa,\mu_0} \|_{L^2(0,T)} = \epsilon \| y \|_{L^2(0,T)} \) as required. We conclude that \( \hat{u} \) given by equation (9) with \( \gamma = 1/\mu_0 \) is a solution of the optimisation problem (8), and hence, the regularised solution of equation (7) depending on parameters \( \epsilon, \kappa > 0 \). In practice, the values of these regularising parameters must be chosen experimentally based on the problem data \( y \) and \( v \).

In the frequency plane, equations (9) and (10) take the form

\[
\hat{u}^\wedge(\xi) = \frac{H(\xi)^\wedge y(\xi)}{\gamma (\kappa + \xi^4) + |H(\xi)|^2}
\]

where \( H(s) = \int_0^\infty e^{-st} h_0(t) \, dt \) is the transfer function corresponding to \( h_0(t) \) and

\[
\hat{v}_{\kappa,\mu}(\xi) = G_{\kappa,\mu}(i\xi)^\wedge y(\xi)
\]

where

\[
G_{\kappa,\mu}(s) = \left( 1 + \frac{\| H(s) \|^2}{\kappa + s^4} \right)^{-1}.
\]

Note that \( |G_{\kappa,\mu}(i\xi)| < 1 \), and the last equation indicates that the high frequency components of \( y \) and \( v_{\kappa,\mu} \) are essentially identical. By Parseval’s identity, the value of \( \gamma = 1/\mu_0 \) can be solved from \( \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{v}_{\kappa,\mu}(\xi)|^2 \, d\xi = \epsilon^2 \| y \|^2_{L^2(0,T)} \).

2.5. VT anatomic data and physical models

The three-dimensional anatomic data of the VT is used for computational validation of the sound source and for carrying out actual measurements.

VT anatomic geometries were obtained from a (then) 26-year-old male (one of the authors of this article) using 3D MRI during prolonged productions of all Finnish vowels [2]. A speech sample was recorded during the MRI, and it was processed for formant analysis by the algorithm described in [3]. Three of the MR images corresponding to the Finnish quantal vowels [a, i, u] were processed into 3D surface models as explained in [4]. A spherical boundary condition interface was attached to the front of the mouth opening of [a] to produce the computational geometries shown in figure 3.

A Stratasys uPrint SE Plus 3D printer was used to produce the VT physical models in ABS plastic shown in figure 4. These models are in a natural scale with a wall thickness of 2 mm; they extend from the glottal position to the lips, and are equipped with an adapter (visible in figure 4) for coupling them to the sound source shown in figure 1(a).
3. Design and construction

Based on the considerations of section 2, the following design objectives were deemed desirable:

(i) The TL from the input to output should be as low as possible.

(ii) There should be no strong transversal resonant modes inside the impedance matching cavity of the device.

(iii) The frequency response $\omega \mapsto |Z_{RL}(i\omega)|$ of the transmission impedance should be as flat as possible for the relevant termination resistances.

It is difficult—if not impossible—to optimise all these characteristics in the same device. Fortunately, DSP techniques can be used to cancel out some of the remaining undesirable features, and instead of the requirement (iii), it is more practical to pursue a more modest objective:

(iii’) The frequency response $\omega \mapsto |Z_{RL}(i\omega)|$ should be such that its lack of flatness can be accurately precompensated by causal, rational filters.

We next discuss these objectives and their solutions in light of section 2.

The tractrix horn geometry was chosen so as to minimise the TL for the reasons explained in section 2.3. In this work, we use $a = 50.0 \text{ mm}$, $b = 2.2 \text{ mm}$ and $\ell = 153.0 \text{ mm}$ as the nominal values for equation (6). These physical dimensions were chosen based on the practicality and availability of suitable loudspeaker units.

In contrast to loudspeaker or gramophone horns, which essentially have point sources at their narrow input ends, the sound source is now located at the wide end. Hence, it would be desirable to generate the acoustic field with the spherical surface source, the curvature radius of $a$ and the centrepoint at the centre of the wide end. This requirement is impossible to satisfy using commonly available loudspeaker units, but a reasonable outcome can be obtained by placing the loudspeaker (with a conical diaphragm) at an optimal distance from the horn opening, as shown in figure 2(b). This results in a design where the impedance matching cavity of the source is an extended horn, consisting of the tractrix horn that has been extended by a cylinder of diameter $2a = 100.0 \text{ mm}$ and height $h = 20.0 \text{ mm}$, as shown in figure 2(b). Thus, the total length of the impedance matching cavity is $\ell_{tot} = \ell + h = 173.0 \text{ mm}$. This dimension corresponds to the quarter wavelength resonant frequency at $f_{low} = 1648 \text{ Hz}$.

Figure 4. Physical VT models of articulation geometries corresponding to [i, i, i] in the natural scale. Adapter sleeves have been glued to the glottis ends for coupling to the sound source. A 1 coin of diameter 23.3 mm is shown in the last picture.

Figure 5. The pressure distributions of some resonance modes of the impedance matching cavity above the loudspeaker unit. The lowest mode is at 1648 Hz, and it is purely longitudinal. The lowest transversal mode is at 1994 Hz, and is due to the cylindrical part joining the loudspeaker unit to the tractrix horn. At 4218 Hz, a transversal mode appears where strong excitation exists between the cylindrical part and the horn. The lowest transversal mode that is solely due to the tractrix horn geometry is found at 5229 Hz.
obtained by solving the eigenvalue problem equation (1) by a finite element method (FEM) with the outcome shown in figure 5. For frequencies much under \( f_{\text{low}} \), the impedance matching cavity need not be considered as a waveguide but just as an attenuated delay line.

Since the shape of the impedance matching cavity has already been specified, there is no geometric freedom left to improve anything. Thus, it is unavoidable to relax the design objective (iii) in favour of the weaker requirement (iii'). As discussed in section 2.2, requirement (iii') can be satisfactorily achieved if overly strong transversal modes of the impedance matching cavity can be avoided or suppressed by placing structures or acoustically soft material inside the cavity. Understanding the modal behaviour helps in the optimal placement of the attenuating material inside the cavity.

### 3.1. Modal analysis of the impedance matching cavity

The first step in treating transversal modes of the impedance matching cavity is to detect and classify them. For this purpose, the Helmholtz equation (1) was solved by FEM in the geometry of the cavity, producing resonances up to 8 kHz. Some of the modal pressure distributions are shown in figure 5. As explained in section 4 below, the acoustic resonances of the VT geometry shown in figure 2(c) were also computed in a similar manner, and their perturbations were evaluated when coupled to the cavity, as shown in figure 3.

The triangulated surface mesh of the impedance matching cavity was created by generating a profile curve of the tractrix horn in MATLAB. A surface of revolution was created from the profile in Comsol where the cylindrical part and the loudspeaker profile were included. Similarly, the surface mesh of the VT during the phonation of the Finnish vowel [i] was extracted from the MRI data [4]. This surface mesh was then attached to the spherical interface \( \Gamma_0 = \Gamma_e \) (see figure 2(b)) in order to produce the geometry shown in figure 2(c). In the same manner, the surface meshes of the impedance matching cavity and the VT were joined together to produce the geometry shown in figure 3. Finally, tetrahedral volume meshes for FEM computations were generated using GMSH [25] with the details given in table 1.

The Helmholtz equation (1) with the Dirichlet boundary condition (2) at the output interface \( \Gamma_0 = \Gamma_e \) was solved by FEM using piecewise linear elements. In this case, the problem reduces to a linear eigenvalue problem whose lowest eigenvalues give the resonant frequencies and modal pressure distributions of interest. Some of these are shown in figure 5.

The purely longitudinal acoustic modes were found at frequencies 1648 Hz, 2540 Hz, 3350 Hz, 3771 Hz, 4499 Hz, 5061 Hz, 5745 Hz, 6671 Hz, 7088 Hz, 7246 Hz and 7737 Hz. All of these longitudinal modes have a multiplicity of 1. Transversal modes divide into three classes: (i) those where excitation is mainly in the cylindrical part of the impedance matching cavity, (ii) those where the excitation is mainly in the wide end of the tractrix horn, and (iii) those where both parts of the impedance matching cavity are excited to an equal extent. The resonance due to the cylindrical part appears at frequencies 1994 Hz, 3094 Hz, 4150 Hz, 6063 Hz, 6262 Hz, 6872 Hz, 6942 Hz, 7334 Hz and 7865 Hz, and they all have a multiplicity of 2 except the resonance at 4150 Hz, which is simple. There is a longitudinal resonance at 4150 Hz as well. There are only four frequencies corresponding to the transversal modes (all with a multiplicity of 2) in the tractrix horn: namely, 5229 Hz, 5697 Hz, 6764 Hz and 6781 Hz. The mixed modes of the third kind were observed at 4218 Hz (2), and 5200 Hz (3) where the number in the parenthesis denotes the multiplicity.

Based on these observations, the acoustic design of the impedance matching cavity was deemed satisfactory as the transversal dynamics of the tractrix horn only shows up above 5.2 kHz.

### 3.2. Details of the construction

The horn geometry was produced using the parametric tractrix horn generator OpenSCAD script [26]. The horn was 3D printed by Ultimaker Original in PLA plastic with a wall thickness of 2 mm and a fill density of 100%. The inside surface of the print was coated by several layers of polyurethane lacquer after which it was polished. The horn was installed inside a cardboard tube, and the space between the horn and the tube was filled with ca. 1.2 kg of plaster of Paris in order to suppress the resonant behaviour of the horn shell itself and to attenuate the acoustic leakage through the horn walls.

The loudspeaker unit of the source is contained in a hardwood box with a wall thickness of 40 mm shown in figure 1. The exterior dimensions of the box are 215 mm \( \times \) 215 mm \( \times \) 145 mm. In order to reduce the acoustic leakage, the box was sealed air tight by applying silicone sealant to all joints from inside. The horn assembly and cylindrical space above the loudspeaker cone form the impedance matching cavity shown in figure 2(b) when the two components (shown separately in figure 1(b)) are joined together. There is another acoustic cavity under the loudspeaker unit whose dimensions are 135.0 mm \( \times \) 135.0 mm \( \times \) 70.0 mm. This cavity was tightly filled with acoustically soft material, i.e. polyester fibre, to reduce resonance.

The walls of the cylindrical part of the impedance matching cavity were covered by felt to control the standing waves there. Acoustically soft material was placed inside the impedance matching cavity shown in figure 2(b) by a trial and improvement method, based on repeated frequency response measurements and reasoning based on the resonant modes shown in figure 5. The main purpose of this work was to suppress the overly strong transversal modes in the cavity. As a secondary effect, the purely longitudinal modes were also suppressed to some extent. This resulted in a high but tolerable increase in the TL of the cavity.

### 3.3. Electronics and software for measurements

We used a 4” two-way loudspeaker unit (of a generic brand) whose diameter matches the opening of the tractrix horn. Its nominal maximum output power is 30 W (RMS) when coupled to a 4Ω source. The loudspeaker is driven by a power
amplifier based on TBA810S IC. There is a decouplable AC mA-meter in the loudspeaker circuit that is used to set the output level of the amplifier to a fixed reference value at 1 kHz before the measurements. The power amplifier is fed by one of the output channels of the sound interface ‘Babyface’ by RME, connected to a laptop computer via a USB interface.

The acoustic source contains an electret reference microphone (of a generic brand, ⊙9 mm, biased at 5 V) at the output end of the horn. The reference microphone is embedded in the waveguide wall, and there is an aperture of ⊙1 mm in the wall through which the microphone detects the sound pressure. A narrow aperture is required so as not to overdrive the microphone by the high sound intensity, and it is positioned ca. 13.5 mm below the vocal folds position of the 3D printed VT model (depending on the anatomy).

The measurements near the mouth position of the 3D-printed VTs are carried out by a signal microphone. As a signal microphone we use either a similar unit as the reference microphone or a Brüel & Kjær measurement microphone model 4191 with the capsule model 2669 (as shown in figure 1(a)) and a preamplifier Nexus 2691. The B&K unit has a noise floor over 20 dB lower than the electret units, although this is of no significance when measuring the resonant frequencies of the VT model in a noisy office environment. Measurements in the anechoic chamber yield much cleaner data when the B&K unit is used, and this is advisable when studying acoustic loads with a higher TL leading to lower signal levels. Then, extra attention must be paid to all other aspects of the experiments so as to achieve the full potential of the high-quality signal microphone.

The reference and the signal electret microphone units were picked from a set of 10 units to ensure that their frequency responses within 80 Hz . . . 8 kHz were practically identical. It was observed that the differences in the frequency and phase response of any two such microphone units were very small. Furthermore, these microphones are practically indistinguishable from the Panasonic WM-62 units (with nominal sensitivity −45 ± 4 dB re 1 V/Pa at 1 kHz) that were used in the instrumentation for MRI/speech data acquisition reported in [4].

The final results given in section 6 were measured using the Brüel & Kjær II model 4191 at the mouth position. The results shown in [3, figure 5] were measured using the electret unit. In this article, the electret microphone measurements at the mouth position were only used for comparison purposes.

Bias voltages for the electret microphones are produced by a custom preamplifier with two identical channels based on LM741 ICs. The amplifier has a nonadjustable 40 dB voltage gain in its passband that is restricted to 40 Hz . . . 12 kHz. Particular attention was paid to reducing the ripple in the microphone bias as well as the cross-talk between the channels. The input impedance 2.2 k Ω of the preamplifier is a typical value for electret microphones, and the output is matched to 300 Ω for the two input channels of the ‘Babyface’ unit.

Signal waveforms and sweeps were produced numerically as explained in section 5 for all experiments. All computations were carried out in MATLAB (R2016b) running on a Lenovo Thinkpad T440s, equipped with a 3.3 GHz Intel Core i7-4600U processor and a Linux operating system. The experiments are automatically run using MATLAB scripts, and MATLAB is interfaced to ‘Babyface’ through Playrec (a MATLAB utility, [27]).

3.4. Measurement arrangement

An outline of the arrangement for measuring a VT model is shown in figure 6. Both the amplifiers, the digital analogue converter (DAC) and the computer are located outside the anechoic chamber. The arrangement inside the anechoic chamber contains two microphones: the reference unit near the glottal position embedded inside the source, and the signal microphone in front of the mouth opening. The signal microphone must be kept in the same position in all measurements for reproducibility.

Because of the high TL of the VT model (in particular, in the VT configuration corresponding to [i]) and the relatively low sound pressure level produced by the source (compared to the extremely high sound pressure from human vocal folds), one may have to carry out signal microphone measurements at signal levels only ca. 20 . . . 30 dB above the hearing threshold. The laboratory facilities require coaxial microphone cables of length 10 m to be used in order to prevent hum. Another significant source of disturbance is the acoustic leakage from the source directly to the signal microphone. This leakage can be reduced by ca. 6 dB by enclosing the sound source in a box made of insulating material, and preventing sound conduction by placing the source on silicone cushions resting on a stone block (not shown in figures 1 and 6).

4. Computational validation using a VT load

When an acoustic load is coupled to a sound source containing an impedance matching cavity, the measurements unavoidably concern the joint acoustics of the source and the load. Hence,
precautions must be taken to ensure that the characteristics of the acoustic load truly are the main component in the measurement results. In the case of the proposed design, the small intersectional area of the source output leads to high acoustic output impedance which is consistent with an acoustic current source. Also, the narrow glottal position of the VT model helps to isolate the two acoustic spaces from each other.

We evaluate this isolation by computing the Helmholtz resonances of the joint system shown in figure 3 and compare the results with (i) the formant frequencies measured from the same test subject during the MR imaging, and (ii) the Helmholtz resonances of the VT geometry shown in figure 2(c). The VT part in both of the computational geometries is the same, and it corresponds to the vocal [q]. The vowel [q] was chosen out of [a, i, u] because its three lowest formants are most evenly distributed in the voice band.

In numerical computations, the domain \( \Omega \subset \mathbb{R}^3 \) for the Helmholtz equation (1) consists of the VT geometry of [q] either as such (leading to ‘VT resonances’ in table 2) or joined to the impedance matching cavity at the glottal position (leading to ‘VT + source resonances’ in table 2). The acoustic modes and resonant frequencies were computed from equation (1) by FEM, and some of the resulting resonant frequencies and modal pressure distributions are shown in figure 3. The FEM meshes have already been described in section 3.1.

When computing the resonant structure of the joint system shown in figure 3 we use \( \Gamma_0 = \Gamma_s \) and instead of equation (2) we use the Robin boundary condition

\[
\lambda \phi_\lambda (r) + c \frac{\partial \phi_\lambda}{\partial n} (r) = 0 \quad \text{on} \quad \Gamma_0
\]

making the interface absorbing. Using the more complicated equations (12) instead of (2) is preferable since the linear dimension of \( \Gamma_s \) is of the same order as the wavelengths of interest as opposed to the much smaller \( \Gamma_v \). When computing VT resonances without the impedance matching cavity (the top row in table 2), the interface \( \Gamma_v \) (see figure 2(c)) at the glottal opening must have a boundary condition as well, and equation (12) is used with \( \Gamma_0 = \Gamma_v \cup \Gamma_s \). In both cases, the resulting quadratic eigenvalue problems were solved by transforming them to larger, linear eigenvalue problems as explained in [28, section 3]. For a similar kind of numerical experiment involving VT geometries but without a source, see [29].

The formant values given in table 2 (bottom row) have been extracted by Praat [30] from post-processed speech recordings during the acquisition of the MRI geometry as explained in [4, 5]. The extraction was carried out at 3.5 s from the start of the phonation with a duration of 25 ms.

Given in semitones, the discrepancies between the first two rows in table 2 are −2.3, −0.1 and 0.05. Similarly, the discrepancies between the last two rows in table 2 are −2.4, 0.4 and −0.9. The largest discrepancy concerning the first formant \( F_1 \) is partly explained by the challenges in formant extraction from the speech sample pair of the MRI data used. In [2, table 2], the value for \( F_1 \) from the same test subject was found to be \( 580 \pm 23 \) Hz by averaging over ten speech samples during MRI and using a more careful treatment for computing the spectral envelope, based on the MATLAB function

\text{arburg}.

We conclude that for Helmholtz resonances corresponding to \( F_2 \) and \( F_3 \) of the VT model [q], the perturbation due to acoustic coupling with the impedance matching cavity are small fractions of the comparable natural variation in spoken vowels. So as for the lowest formant \( F_1 \), it seems that the impedance matching cavity actually represents a better approximation of the true subglottal acoustics contribution than the mere absorbing boundary condition imposed at the glottis position of a VT geometry. Furthermore, the three lowest resonant modes of the VT (corresponding to formants \( F_1, F_2, F_3 \)) appear where the impedance matching cavity remains in the ‘ground state’; see figure 3. This supports the claim that the narrowing of the acoustic space at the vocal folds position effectively keeps the impedance matching cavity of the source and the VT model only weakly coupled.

|                | \( F_1 \) | \( F_2 \) | \( F_3 \) |
|----------------|---------|---------|---------|
| VT resonances  | 519     | 1130    | 2297    |
| VT + source resonances | 594     | 1136    | 2290    |
| Formant frequencies | 683     | 1111    | 2417    |

5. Calibration measurements and source compensation

5.1. Measurement and compensation of the frequency response

We describe the production of an exponential frequency sweep\(^6\) with uniform sound pressure at the output of the source. These sweeps are used for the frequency response measurements of VT models. The defining property of exponential sweeps is that each increase in frequency by a semitone takes an equal amount of time. In this work, the frequency interval of such sweeps is \( 80 \ldots 7350 \) Hz with a duration of \( 10 \) s. All measurements leading to curves in figures 7 and 8 were carried out using the dummy load shown in figure 1(c) as the standardised reference.

If one plainly introduces a constant voltage amplitude sweep to the loudspeaker unit, the sound pressure at the source output (as seen by the reference microphone) will vary over \( 20 \) dB over the frequency range of the sweep, as shown in figure 8(a). The key advantage in producing sound pressure of a constant amplitude is that excessive external noise contamination can be avoided at frequencies in which the output sound pressure would be low. Standardising the sound pressure of the source also makes the source acoustics less visible in the measurements, reducing the perturbation effect at \( F_1 \) observed in section 4.

The essentially flat sound pressure output shown in figure 8(c) can be obtained from the source by applying

\(^6\) Also known as the logarithmic chirp.
the amplitude weight $w$ shown in figure 8(b) to the input voltage of the loudspeaker unit. As is to be expected, both the weighted and unweighted voltage sweeps have almost identical phase behaviour, as can be seen in figure 7(a). In contrast, the voltage sweep and the resulting sound pressure at the reference microphone are out of phase in a very complex frequency-dependent manner; see figure 7(b). Such phase behaviour cannot be explained by the relatively sparsely located resonances of the impedance matching cavity.

An iterative process requiring several sweep measurements was devised to obtain the weight $w$, and it is outlined below as algorithm 1. The parameters of the algorithm were tuned by trial and error so as to improve the convergence to a satisfactory compensation weight. During the iteration, different versions of the measured sweeps have to be temporally aligned with each other. The required synchronisation is carried out by detecting a $1$ kHz cue of length $1$ s, positioned before the beginning of each sweep. This is necessary because of the wildly variable latency times in the DAC/software combination.

Algorithm 1. Computation of the equalisation weight $w$.

1: procedure CALIBRATECOMPENSATION(n, t)
2: $w \leftarrow [1, 1, \ldots, 1]$
3: for $k \leftarrow 0 \ldots N$ do
4: $x \leftarrow w \cdot \text{ExponentialChirp}(t)$
5: $y \leftarrow \text{Play}(x)$
6: $H \leftarrow \text{ComputeEnvelope}(y)$
7: $d \leftarrow \text{Dynamics}(H)$
8: $r \leftarrow \text{Regularization}(d)$
9: $w \leftarrow |H| \cdot w / r$
10: return $w$

We consider the calibration successful if the measured dynamics at the final iteration stage is below $1$ dB.

The system comprising a power amplifier, loudspeaker and acoustic load is somewhat nonlinear, which becomes evident using wide frequency ranges and high amplitude variations. Even though the two curves in figures 8(a) and (b) are obviously related, they do not add up to a constant that would be independent of the frequency. Not even the dynamical ranges
of these curves coincide, as would happen in a linear and time-invariant device. In spite of the nonlinearity, it is possible to use a very slowly increasing sweep to produce an accurate voltage gain from the output of the DAC to the output of the reference microphone preamplifier over a very wide range of frequencies. One example of such a voltage gain function is shown in figure 8(a), but its inverse is not a good candidate for the compensation weight.

5.2. Reference waveform tracking

The second goal is to reconstruct a predefined waveform as the sound pressure output, as observed by the reference microphone. In the context of speech, a good target waveform is the Liljencrants–Fant (LF) waveform [31], describing the flow through vibrating vocal folds; see figure 9(a) (left panel).

Because there is an acoustic transmission delay of ca. 0.5 ms in the impedance matching cavity in addition to much larger latencies in the computer hardware and software, a simple feedback-based PID controller is not feasible for any trajectory tracking problem. Instead, a feedforward control solution is required, where the response of the instrumentation is cancelled out by regularised deconvolution, so as to obtain an input waveform that produces the desired output. For this, we use a version of constrained least squares filtering, whose mathematical description is given in section 2.4.

Regularised deconvolution requires the impulse response of the whole measurement system that corresponds to the convolution kernel $h_0$ in equation (7) to be estimated. The response is estimated using the sinusoidal sweep excitation described in [32], and the result of the measurement can be seen in figure 7(c). Because the deconvolution contains regularisation parameters $\gamma$ and $\kappa$, it tolerates some noise in the estimated impulse response.

We proceed to describe how the mathematical treatment in section 2.4 can be turned into a workable signal processing algorithm in discrete time. All signals (including the estimated convolution kernel $h_0$) are discretised at the sampling rate $44 100$ Hz used in all signal measurements. We denote the sample number of a discretised signal, say, $x[n]$ by $N = 44 100 \cdot T$, where $T$ is the temporal length of the original (continuous time) signal $x(t)$, $t \in [0, T]$ and sampling is carried out by setting, e.g.

$$x[n] = \frac{1}{T_s} \int_{(n-1)T_s}^{nT_s} x(t) \, dt$$

where $1 \leq n \leq N$ and $T_s = s/44 100$.

The measured (discrete) impulse response $h_0[n]$ is extended to match the signal length $N$ by padding it with zeroes, if necessary.

In discrete time, the regularised deconvolution in equations (9) and (10) takes the matrix/vector form

$$\hat{u} = (\gamma (\kappa I + R^T R + \gamma^{-1} H^T H)^{-1} H^T \, y$$

and

$$v_{\gamma, \kappa} = (\kappa I + R^T R + \gamma^{-1} H^T H)^{-1} (\kappa I + R^T R) \, y.$$  (13)
The components of the $N \times 1$ column vectors $\mathbf{u}, \mathbf{y}, \mathbf{v}, \mathbf{v}_{\gamma, \kappa}$ are plainly the discretised values $\hat{u}[n], \hat{y}[n], \hat{v}_{\gamma, \kappa}[n]$ for the $n = 1, \ldots, N$ of signals $u, y, v_{\gamma, \kappa}$, respectively, given in equations (9) and (10) where $\mu = \gamma^{-1}$. The second order difference $N \times N$ matrix $R$ is the symmetric matrix whose top row is $[2, -1, 0, \ldots, 0, -1]$, making it circulant. The non-symmetric $N \times N$ matrix $H = [h_{jk}]$ is constructed by setting $h_{jk} = h_{0j}((N + j - k) \mod N + 1)$ for $1 \leq j, k \leq N$. Because all of the matrices $R = R^T, H, H^T$ are now circulant, so is the symmetric matrix $\gamma (\kappa I + R^T R) + H^T H$ in equation (13).

Hence, the matrix/vector products in equation (13) can be understood as circular discrete convolutions that can be implemented in $N \log(N)$ time using the fast Fourier transform (FFT). This leads to a very efficient solution for $\mathbf{u}$ given $\mathbf{y}$, even for long signals.

Defining the transfer functions $\hat{R}(z), \hat{H}(z)$ and the transforms $\hat{y}(z), \hat{v}_{\gamma, \kappa}(z)$ for $z = e^{i\theta}$ as

$$\hat{R}(z) = -z^{-N} - z^{-1} + 2 - z^{-1} - z^{N},$$

$$\hat{H}(z) = \sum_{n=0}^{N} h_{0n}z^n + \sum_{n=-N}^{-1} h_{0n}z^n,$$

$$\hat{y}(z) = \sum_{n=1}^{N} \hat{y}[n]z^n,$$

we observe that the latter of equation (13) takes the form of the discrete Fourier transform (DFT)

$$\hat{v}_{\gamma, \kappa}(z_k) = \frac{\kappa + |\hat{R}(z_k)|^2}{\kappa + |\hat{R}(z_k)|^2 + \gamma^{-1}|\hat{H}(z_k)|^2},$$

realised in MATLAB code, where $z_k = e^{2\pi k/N}$ and $k = 1, \ldots, N$ enumerates the discrete frequencies. By Parseval’s identity, we interpret the residual equation (11) in discretised form as

$$\sum_{k=1}^{N} |\hat{v}_{\gamma, \kappa}(z_k)|^2 = \epsilon \sum_{k=1}^{N} |\hat{y}(z_k)|^2$$

which, together with equation (14), gives an equation from which $\gamma = \gamma(\epsilon, \kappa)$ can be solved for each $0 < \epsilon < 1$ and $\kappa > 0$. This is done using the MATLAB fminbnd function to ensure that $\gamma > 0$. The values for $\epsilon, \kappa$ are chosen based on the experiments.

6. Results

Two kinds of measurements were carried out on the 3D printed VT physical models. Firstly, the magnitude frequency response was measured to determine the spectral characteristics such as the lowest resonant frequencies. Secondly, a realistic glottal excitation signal was fed into the physical model of the VT to simulate the vowel acoustics in a spectrally correct manner.

6.1. Sweep measurements

The spectral density of the power is obtained from physical models of the VT by the sweep measurements. The sweep is constructed as described in section 5.1 to obtain a constant sound pressure at the reference microphone when the source is terminated by the dummy load. The signal from the measurement microphone at the mouth position of the physical model is then transformed into an amplitude envelope (a similar approach can be found in [33, figure 2]) by an envelope detector (i.e. computing a moving average of the nonnegative signal amplitude). Finally, this output envelope is divided by a similar envelope from the reference microphone at the source output. The resulting amplitude envelopes are shown in the top curves of figure 10, and the resonance data is given in table 3.
6.2. Glottal pulse reconstruction

For reproducing nonsinusoidal target signals at the source output, the general method described in sections 2.4 and 5.2 is used. In this work, we use the LF waveform shown in figure 9(a) (left panel) as the target signal, since it resembles the signal produced by the vocal folds during phonation. The regularised convolution is able to produce the desired target waveform, as can be seen in figure 9(b) (right panel). For the results shown in figure 9, the impulse response of the instrumentation and all signals were measured when the source was terminated by the vowel geometry [1].

7. Discussion

After many design cycles and improvements, the proposed acoustic glottal source appears to be well suited for its intended use. It remains to discuss the shortcomings and possible improvements of the design and algorithms.

The three most serious shortcomings in the final design are (i) the high TL in the impedance matching cavity due to the required attenuation by polyester fibre, (ii) acoustic leakage through the source chassis, and (iii) the usable low frequency limit at ca. 80 Hz. Since the proposed design is scalable, the latter two deficiencies are most easily treated by increasing the physical dimensions, chassis wall thickness, and, hence, the mass of the source. A 6” or even an 8” loudspeaker unit with lower bass resonant frequencies could be considered, equipped with separate concentric tweeters to produce the higher frequencies. Overly increasing the size of the source, however, makes it impractical for demonstration purposes.

The transversal resonance was checked by adding polyester fibre to the wide parts of the impedance matching cavity, which results in a marked increase in TL. Considering the amplitude response dynamics of ca. 35 dB of the source shown in figure 8, the output volume remains relatively low in uniform amplitude sweeps that are produced, as explained in section 5.1. Even though the VT physical models have an additional TL on the order of 20...40 dB depending on the vowel and test subject, it is possible to carry out the frequency response of the formant position measurements without an anechoic chamber or a high quality measurement microphone at the mouth position, and the results are quite satisfactory; see [3, figure 5]. To obtain high-quality frequency response data or carry out the waveform reconstructions presented in section 6, one has to do the utmost to reduce acoustic leakage, hum and noise level, including the use of the Brüel & Kjær II measurement microphone in the anechoic chamber. Then secondary error components emerge, as can be observed, e.g. as roughness between the formant peaks in figure 10. We point out that the quality of the microphone used at the mouth opening does not affect the measured frequencies of the formant peaks. However, the microphone position or the paraboloid concentrator shown in [3, figure 4] does have a small yet observable effect—particularly at the lowest resonance frequency of the physical model.

An attractive way of getting a louder sound source is to use Smith slits [34, 35] for checking the transversal resonances within the wide part of the impedance matching cavity. The required design work is best carried out using the computational design optimisation methods introduced in [18, 21].

This article does not concern impedance measurements, rather the response between two acoustic pressures. For impedance measurements, the perturbation velocity should be measured at the output of the source for which a number of approaches, based on microphones, have been proposed [8, 12, 14]. In the current design, hot wire anemometry at the reference microphone position would be most suitable; see [11, 36]. Even the smallest Microflown unit (see [37–39]) commercially available, placed in the middle of the source output channel with a diameter of 6 mm, would cause severe back reflections.

We have used two different response compensation techniques in section 5: amplitude weighting for sinusoidal sweeps, and regularised deconvolution for more complicated waveforms. Using deconvolution to produce sweeps tens of seconds long is not practical, since the dimension of equation (13) would be too high. As opposed to weighted sweeps, regularised deconvolution takes into account the phase response of the full measurement system. The deconvolution is a linear operation, whereas the measurement system shows signs of amplitude nonlinearity in figure 8. This is one of the reasons why tracking more challenging targets than the LF waveform (e.g. the ramp signal) will not give as good an outcome. The compensation weight reconstruction method in section 5.1 does not rely on linearity at all, and its performance can be improved by increasing the sweep length.

One of the challenging secondary objectives is to design dummy loads of a reasonable physical size for the source, which would present a constant resistive load over a wide range of frequencies. The dummy load shown in figure 1(c) consists of a tractrix horn tightly filled with polyester fibre, and it has the property of not being resonant to an observable degree. Two particularly inspiring examples on the construction of the resistive acoustic loads are given in [8] (42 m of insulated steel pipe with an inner diameter of 7.8 mm) and [14] (97 m of straight PVC pipe with an inner diameter of 15 mm). The practical challenges in such approaches are considerable.

We conclude by discussing the numerical efficiency of the discretised deconvolution proposed in section 5.2. In order to obtain an $N \log N$ algorithm, the $N \times N$ matrices $R$ and $H$ had to be circulant. Another way to proceed is allowing $R$ to be the usual tridiagonal, symmetric, second order difference matrix, and $H$ to be the upper triangular matrix obtained from the impulse response, which are both noncirculant Toeplitz matrices. Then, the symmetric matrix $\gamma (ed + R^* R) + H^* H$ in equation (13) is a slightly perturbed Toeplitz matrix, and the required (approximate) solution of the linear system can be carried out by Toeplitz-preconditioned conjugate gradients at a superlinear convergence speed; see, e.g. [40]. Again, an $N \log N$ algorithm is obtained if the matrix/vector products are implemented by FFT.
8. Conclusions

A sound source was proposed for the acoustic measurements of vocal tract physical models, produced by fast prototyping methods from magnetic resonance images. The source design requires only commonly available components and instruments, and it can be scaled to different frequency ranges. Heuristic and numerical methods were used to understand and to optimise the source design and performance. Two kinds of algorithms were proposed to compensate the source nonoptimality: (i) an iterative process for producing uniform amplitude sound pressure sweeps, and (ii) a method based on regularised deconvolution for replicating the target sound pressure waveforms at the source output. The sound source, together with the two compensation algorithms, written in MATLAB code, were deemed successful based on measurements of the vocal tract geometry corresponding to the vowels \([q]\) of a male speaker.

Numerical modal analysis of electromagnetic cavity resonators has many similarities with the analysis of the proposed acoustic system. Such similarities are further accentuated in designs for UHF and SHF radio frequency bands, where the typical wavelengths are of the same magnitude as the instrument used. In radio technology, high-quality measurement equipment is widely available, and standardised measurement procedures exist, leading to good replicability, comparability, and accuracy of results. In experimental studies of the acoustics of speech production, for example, the technology is less mature, and there are more low-level practical challenges. These include the isolation of electromagnetic fields, and the (resistive) termination of the transmission lines is almost trivial, whereas the analogous problems within acoustics are instructive, even for an engineer who has an extensive background in RF design.

Acknowledgments

The authors wish to thank the Department of Signal Processing and Acoustics, Aalto University (Prof P Alku, Lab. Eng. I Huhtakallio, M Sc M Airaksinen) and Digital Design Laboratory, Aalto University (M Arch A Mohite) for the consultations and facilities.

The authors have received financial support from the Instrumentarium Science Foundation, Magnus Ehrnrooth Foundation, Niilo Helanter Foundation, and the Vilho, Yrjö and Kalle Väisälä Foundation.

ORCID iDs

Juha Kuortti @ https://orcid.org/0000-0001-7264-5420
Jarmo Malinen @ https://orcid.org/0000-0001-9117-8450

References

[1] Aalto A, Murtonen T, Malinen J, Aalto D and Vainio M 2017 Modal locking between vocal fold oscillations and vocal tract resonance Interspeech 2017 (August 20–24) pp 3487–91
[2] Aalto D et al 2014 Large scale data acquisition of simultaneous MRI and speech Appl. Acoust. 83 64–75
[3] Kuortti J, Malinen J and Ojalami A 2018 Post-processing speech recordings during MRI Biomed. Signal Process. Control 39 11–22
[4] Ojalami A and Malinen J 2017 Automated segmentation of upper airways from MRI vocal tract geometry extraction Proc. Bioimaging 277–84
[5] Zhu D T W, Li K, Epps J, Smith J and Wolfe J 2013 Experimental evaluation of inverse filtering using physical systems with known glottal flow and tract characteristics J. Acoust. Soc. Am. 133 EL388–62
[6] Epps J, Smith J and Wolfe J 1997 A novel instrument to measure acoustic resonances of the vocal tract during phonation Meas. Sci. Technol. 8 1112–21
[7] Takekoto 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 3724–38
[8] Wolfe J, Smith J, Tann J and Fletcher N 2001 Acoustic impedance spectra of classical and modern flutes J. Sound Vib. 243 127–44
[9] Alku P, Story B H and Airas M 2006 Estimation of the voice source from speech pressure signals: evaluation of an inverse filtering technique using physical modelling of voice production Pol. Phoniatria Logopedica 58 102–13
[10] Alku P 2011 Glottal inverse filtering analysis of human voice production—a review of estimation and parameterization methods of the glottal excitation and their applications Sadhana 36 623–50
[11] Neuschaefer-Rube M K C 2002 A method for measurement of the vocal tract impedance at the mouth Med. Eng. Phys. 24 467–71
[12] Chu D, Li K, Epps J, Smith J and Wolfe J 2013 Experimental evaluation of inverse filtering using physical systems with known glottal flow and tract characteristics J. Acoust. Soc. Am. 135 EL358–62
[13] Singh R and Schary M 1978 Acoustic impedance measurement using siren sweep excitation and known volume velocity technique J. Acoust. Soc. Am. 64 995–1003
[14] Dickens P, Smith J and Wolfe J 2007 Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution J. Acoust. Soc. Am. 121 1471–81
[15] Dinsdale J 1974 Horn loudspeaker design Wiedl. World 4 19–24
[16] Edgar B C 1981 Tractrix horn contour Speaker Build. 2 9–15
[17] Delgado R and Klipsch P W 2000 A revised low-frequency horn—a variable mouth acoustic horn Int. J. Numer. Methods Eng. 59 1–6
[18] Udawalpolara R, Wadbro E and Berggren M 2011 Optimization of a variable mouth acoustic horn Int. J. Numer. Methods Eng. 85 591–606
[19] Lukkari T and Malinen J 2015 A posteriori error estimates for Webster’s equation in wave propagation J. Math. Anal. Appl. 427 941–61
[20] Lukkari T and Malinen J 2013 Webster’s equation with curvature and dissipation (arXiv:1204.4075)
[21] Yedeg E L, Wadbro E and Berggren M 2016 Layout optimization of thin sound-hard material to improve the far-field directivity properties of an acoustic horn Struct. Multidiscip. Optim. 55 795–808
[22] Voigt P 1927 Improvements in horns for acoustic instruments GB Patent Specification 16794/26
[23] Hunt B 1972 Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory IEEE Trans. Autom. Control 17 705–3
[24] Phillips D L 1962 A technique for the numerical solution of certain integral equations of the first kind J. ACM 9 84–97

Meas. Sci. Technol. 28 (2017) 115902
A Hannukainen et al
[25] Geuzaine C and Remacle J-F 2009 Gmsh: a 3-D finite element mesh generator with built-in pre- and post-processing facilities Int. J. Numer. Methods Eng. 79 1309–31
[26] Süß L 2014 Tractrix horn generator OpenSCAD script http://thingiverse.com/thing:218241
[27] Humphrey R 2008 Playrec http://playrec.co.uk/ (Accessed: 11 October 2015)
[28] Hannukainen A, Lukkari T, Malinen J and Palo P 2007 Vowel formants from the wave equation J. Acoust. Soc. Am. 122 EL1–7
[29] Arnela M, Guasch O and Ali 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
[30] Boersma P and Praat D W 2017 Praat: doing phonetics by computer http://praat.org version 6.0.21 (Accessed: 16 September 2015)
[31] Fant G, Liljencrants J and Lin Q-G 1985 A four-parameter model of glottal flow STL-QPSR 4 1–13
[32] Müller S and Massarani P 2001 Transfer-function measurement with sweeps J. Audio Eng. Soc. 49 443–71
[33] Wolfe J, Chu D, Chen J and Smith J 2016 An experimentally measured source-filter model: glottal flow, vocal tract gain and output sound from a physical model Acoust. Australia 44 187–91
[34] Smith B H 1953 An investigation of the air chamber of horn type loudspeakers J. Acoust. Soc. Am. 25 305–12
[35] Oclee-Brown M J 2009 New methodology for the acoustic design of compression driver phase plugs with concentric annular channels Audio Eng. Soc. 57 771–87
[36] Pratt R, Elliot S and Bowsher J 1977 The measurement of the acoustic impedance of brass instruments Acustica 38 236–46
[37] van der Eerden F, de Bree H-E and Tijdeman H 1998 Experiments with a new acoustic particle velocity sensor in an impedance tube Sens. Actuators 69 126–33
[38] de Bree H-E, Leussink P, Korthorst T, Jansen H, Lammerink T S J and Elwenspoek M 1996 The Microflown: a novel device for measuring acoustic flows Sens. Actuators A 54 552–7
[39] de Bree H-E 1997 The Microflown PhD Thesis University of Twente
[40] Malinen J 1998 Properties of iteration of Toeplitz operators with Toeplitz preconditioners BIT Numer. Math. 38 356–71