Estimation of vocal tract and trachea area functions from impedance spectra measured through the lips

Anne Rodriguez, Noel Hanna, André Almeida, John Smith and Joe Wolfe

School of Physics, University of New South Wales
n.hanna@unswalumni.com

Abstract

Determining the area function $A(x)$ of the airway between the lips and vocal folds from external measurements is a classic inverse problem. $A(x)$ is estimated by fitting the acoustic impedance measured through the lips. Excellent fits are possible with about eight cylindrical segments representing the tract. In examples where $A(x)$ has only small slope, moderately good agreement is found on the scale of about a centimetre. Calculations of the impedance loading the glottis are affected by the epilaryngeal tube, which has less effect on the impedance through the lips. The frequencies of these extrema are better estimated than their magnitudes.

Index Terms: vocal tract, acoustic impedance, speech production, subglottal tract

1. Introduction

Measurements of the acoustic impedance spectrum of the vocal tract can now be made through the lips rapidly, precisely and over a large frequency range. Best results are obtained with the glottis closed, but good results are possible during phonation and breathing. With no other information, how much can they tell us about the shape of the tract, the trachea and the acoustic load they impose on the glottal source?

The vocal tract is often analysed as an acoustic duct with varying cross section $A(x)$ along its length x – essentially a one dimensional (1D) model. Discrete $A(x)$ models have a long history. Fant [1] used a simple, two-cylinder model, to demonstrate the acoustic phonetic model of vowel placement: the position of the discontinuity between the two segments (of lengths l_1 and l_2) determined the frontness or backness of the vowel, while the area ratio (A_1/A_2) largely determined the vowel height. Sets of several or more cylinders approximating $A(x)$ are now commonly used, and used hereafter in this paper.

Determining the vocal tract $A(x)$ from the acoustic signal has long been a goal of speech science (e.g. [2,3]) but the speech signal, sparsely sampled in the frequency domain, does not contain enough information to do this reliably. Pulse reflectometry measures the impulse response of the tract at the lips while sealed around a measurement device; from this $A(x)$ can be reconstructed with several assumptions [4], this approach is used clinically.

The acoustic impedance is the (complex) ratio of acoustic pressure (p) to flow (U): $Z(f) = p/U$. Here we examine whether recent advances in impedance spectrometry are sufficient to estimate $A(x)$ reliably. Although the difficulty of the inversion problem is similar to pulse reflectometry, the advantage of this approach is the improved signal:noise ratio, which allows more detailed acoustic information to be used as an input.

In the present paper, we investigate solutions to the inversion problem in terms of the number of parameters that can reasonably be extracted from good data and their goodness of fit. It uses a known target $A_t(x)$ and a one-dimensional duct model to generate a target acoustic impedance spectrum $Z_t(f)$, which is then fitted by calculating $Z(f)$ from estimated functions $A(x)$ whose parameters are varied to minimise the squares of differences between $Z(f)$ and $Z_t(f)$. We concentrate on the vocal tract and use area functions $A_t(x)$ for the American English vowels /i/, /o/, /ɔ/, /æ/ and /ɛ/ from [5]. An intended application of the $A(x)$ is the back propagation calculation of the acoustic signal: obtaining the flow and pressure at the vocal folds, as described in a companion paper in these proceedings [6]. For this reason, the vocal tract shapes are modified to provide a consistent lip aperture.

Recent measurements of $Z_t(f)$ during a neutral vowel gesture show the effects of yielding, lossy walls [7], and measurements during inhalation show the effects of the subglottal duct (trachea) [8].

2. Materials and Methods

2.1. Acoustic impedance measurements

Measurements of $Z_t(f)$ through the lips are made as described in [7]. Briefly, the subject seals the lips around the end of a cylindrical waveguide on which are mounted three microphones, which are input via conditioning amplifiers and an audio interface to a computer.

At the opposite end is a loudspeaker in parallel with a short, narrow pipe that allows exhaled air to exit. The computer generates a sum of sine waves in the range 200 to 4000 Hz, with magnitudes and phases adjusted to improve signal:noise ratio, via an audio interface and amplifier; this drives the speaker. With suitable calibration, impedance at the lips may be calculated at each frequency from the three microphone signals.

This method measures the input impedance $Z_t(f)$ in a fraction of a second but, on its own, no corresponding $A(x)$.

2.2. Simulated impedance through the lips and glottis

Whether for measurements at the lips or a calculated load at the glottal source, impedance spectra for the $A(x)$ are calculated using the transfer matrix method [*e.g.* 9]. Starting with a load impedance Z_L, the impedance at the opposite end of the adjacent cylindrical section is calculated as

$$Z = \frac{\rho c}{\pi r^2}\left(\frac{\pi r^2 \cdot Z_L \cos(\Gamma l) + 2j\rho c \cdot \sin(\Gamma l)}{j\pi r^2 \cdot Z_L \sin(\Gamma l) + 2\rho c \cdot \cos(\Gamma l)}\right) \quad (1)$$

where ρ is the density of air, c the speed of sound, r the radius, l the length and Γ is the complex wave number

$$\Gamma = \frac{\omega}{c} - j\alpha \approx \frac{\omega}{c} - j\frac{(1.2\times10^{-5}\ s^{1/2})\sqrt{\omega}}{r} \quad (2)$$

where ω is the angular frequency and α is an attenuation coefficient [9], which accounts for losses to the tube walls by viscous drag and thermal conduction. For a closed, immoveable termination (a reasonable approximation to a closed glottis), the load impedance is infinite.

Figure 1 shows the radius $r(x)$, for several vowels from the published $A(x)$ in [5], assuming cylindrical segment geometry and modified to provide a consistent lip aperture.

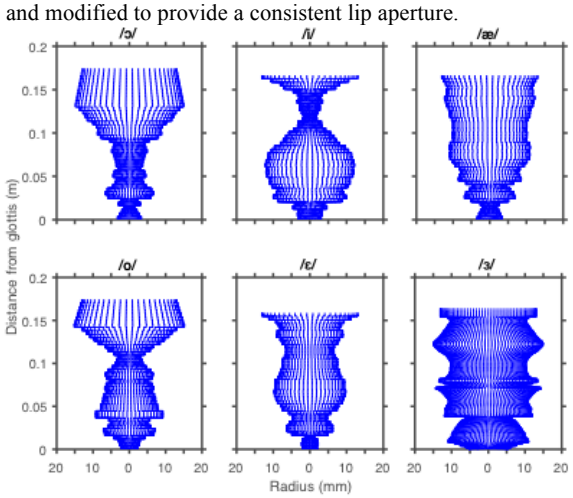

Figure 1. *Shows r(x) for the synthetic vowels /ɔ/, /i/, /æ/, /ɒ/, and /ɛ/ modified from [5], and the vowel /ɜ/ from [10].*

For the case of the open glottis during inhalation, we use a very simple model for the trachea following [11]. Because the trachea branches rapidly, a single open cylinder with a large flange is used.

The $Z(f)$ that loads the glottal source was similarly calculated by iteration of the transfer matrix method, this time starting with the radiation impedance at the open lips and working back to the glottis.

2.3. Inversion and the optimal number of elements

We begin with an $A_t(x)$ (subscript t for target) from Figure 1 and calculate $Z_t(f)$. We then consider a candidate $A(x)$ function having n cylindrical elements with lengths and radii l_i and r_i, all of which may be adjusted during the fitting. From these, we calculate $Z(f)$ through the lips as described above. Scaled sums of squares of errors for magnitude, phase and combination respectively are calculated:

$$S_m = \frac{1}{N}\sum\left(\frac{\log(|Z|)-\log(|Z_t|)}{\log(|Z_t|)}\right)^2 \quad (3)$$

$$S_\phi = \frac{1}{N}\sum\left(\frac{\phi-\phi_t}{\pi}\right)^2 \quad (4)$$

$$S_\Sigma = \frac{1}{N}\sum\left(\frac{\log(|Z|)-\log(|Z_t|)}{\log(|Z_t|)}\right)^2 + K\left(\frac{\phi-\phi_t}{\pi}\right)^2 \quad (5)$$

where there are N frequencies. These are minimised using the MATLAB non-linear least squares fitting function *fit*. The range goes from 1 to 4 kHz for theoretical fits and targets, but over the measured range of 200-4000 Hz when fitting real data. The purpose of the semi-log scale for magnitude is to give impedance maxima and minima similar weight in the sum.

Minimising S_m or S_ϕ gives similar solutions for $A(x)$. We also minimised S_Σ. (The denominators in S_m and S_ϕ give two sums of order unity. K is a weighting factor set to 1/20, chosen to approximate the ratio S_m/S_ϕ in Figure 2). This gives an improvement to the A(x) fits; see Figure 2.

Inside a loop for the minimisation of S, l_i and r_i are varied within the constraints 0.1 to 100 mm and 0.1 to 30 mm respectively and with initial values from 5 to 50 mm and 10 mm.

3. Results and Discussion

3.1. Synthetic vocal tract models

3.1.1. Ideal number of cylindrical segments

The number n of cylindrical segments was varied from 1 to 10, and a minimum found for each n, so that S may be plotted as a function of n, as in Figure 2. This figure uses data from the vowel /ɔ/. For these data, and for the others studied, each successive added element makes an improvement up to about n = 7 or 8, but there is no improvement thereafter for S_m or S_ϕ. For the results that follow, n = 8.

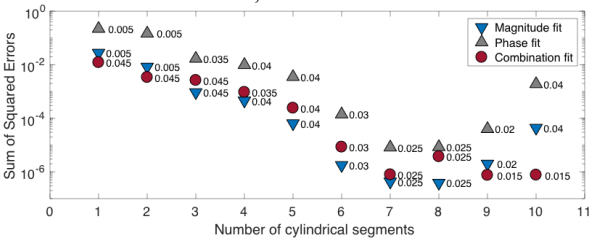

Figure 2. *Scaled sum of squares vs number of cylindrical segments n for fits for the vowel* /ɔ/, calculated at the lips. *Fits minimise differences in impedance magnitude, phase or combination of both. The value at each point shows the initial l_i that gave the lowest SSE.*

Figure 3a compares the radii $r_t(x)$ and the $r(x)$ of the n=8 model from Figure 2. On a longitudinal scale of about a centimetre, the geometrical features are roughly reproduced.

Fig 3b compares $Z_t(f)$ and $Z(f)$ through the lips. Here, the fit is encouragingly good – but this is expected for a fit whose difference has been minimised by adjusting 16 parameters.

The match between the load $Z_g(f)$ on the glottis calculated using $A(x)$ and $A_t(x)$ is not as good, of course: this quantity was not fitted. It is worth noting, however, that the frequencies of the peaks in Z_g, which are closely related to the formants outside the mouth, are fitted fairly well. One can understand some of the limitations in estimating the magnitude of impedance at the glottis Z_g: varying the cross section of a short constriction just above the glottis would be expected to vary a large inertance loading the glottis and in series with the tract: this would have a large effect on Z_g, but relatively little on the impedance at the lips. The fit is poorest at high frequency, in part because the 8-element fit has a lower spatial resolution than the 43-element target. It is also important to point out that, at frequencies above about 3 kHz, the approximation that wavelength >> radius, necessary for 1D plane wave propagation, is less reliable, so one might expect the 1D model to begin to fail.

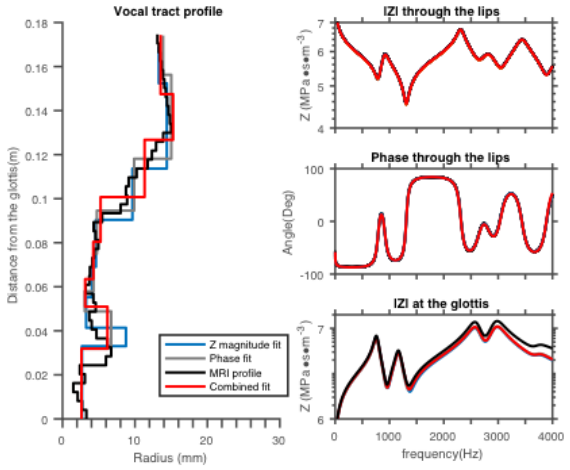

Figure 3. *(left) compares radii: target $r_t(x)$ (black) and fitted $r(x)$ for the vowel /ɔ/ having 43 and 8 elements respectively. (upper and middle right) compare the $Z(f)$ magnitude and phase through the lips calculated from these $r_t(x)$ and $r(x)$. (lower right) compares the calculated $Z_g(f)$ at the glottis.*

3.1.2. *Sensitivity to initial parameters*

Inversion problems in general have potentially many solutions. In this case, how much do they differ? To test this, we tried two methods. One was to vary the initial shape $A_0(x)$ (by changing l_i *or* r_i) and to compare the final $A(x)$. Increasing the initial l_i without constraining the sum of l_i improved the fit up to a certain point, then reduced the quality of fit. The results shown in Figure 2 show those whose initial lengths gave the best fit. Changing the initial r_i had little effect. Another was to compare the $A(x)$ determined by minimising S_m, S_ϕ and S_Σ. Although the sums of squares (quality of fit) varied, the $A(x)$ were qualitatively very similar in most cases (see Figures 4 and 5).

3.2. *In vivo* vocal tract measurements

The calculations above assumed rigid-walled ducts and losses appropriate to dry, hard walls. It is interesting to look at real measurements of $Z(f)$, which, at low frequencies, show features associated with walls of finite mass and rigidity. These require further parameters, for the specific mass, stiffness and losses in the wall. However, they also show new features: a new maximum and minimum in $Z_t(f)$, and so are a more demanding set to fit.

The data from one of the subjects in [7, 8] were used here. For this subject (S7), MRI measurements of the vocal tract and the upper section of the trachea were available for the sustained vowel /ɜː/ [10], the neutral vowel in the word 'heard' in Australian English.

Figure 4 shows a measurement of $Z(f)$ through the lips with closed glottis. The shape is in qualitative (and even semi-quantitative) agreement with the impedance of a uniform 17 cm cylinder above about 300 Hz. However, it fails at low frequency, where a new maximum and minimum appear. The maximum at about 200 Hz is attributed to the mass of the tissue surrounding the vocal tract vibrating on the compliance ('spring') of the enclosed air. (An impedance minimum at about 20 Hz due to that of the same mass vibrating on the compliance of its own tissue is not visible in the measurement at this range [7].) Also superposed is a fit using the technique described above, but with three new parameters: the compact mass (inertance) and specific loss associated with its motion, and an attenuation coefficient (the α in equation (2)) which is increased to model the increased losses of wet, irregular wall losses in living tissue (discussed in [7]). These elements are added as a compact resonant circuit in parallel with the impedance calculated through the lips.

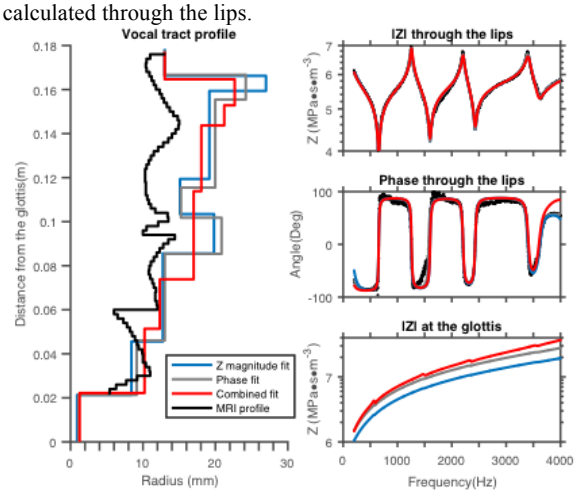

Figure 4. *Measured (black) and fitted impedance magnitude and phase spectra for a male subject with glottis closed. Fitted r(x) compared against r(x) from MRI derived geometry [10]. Otherwise, as Figure 3.*

Although this adds even more parameters, the effect of each of the parameters is readily apparent and almost independent.

- the mass adjusts the frequency of the first impedance peak
- the tissue loss adjusts the magnitude of the first impedance peak
- the wall loss adjusts the bandwidth of all extrema

The first element at the lip has a radius fixed at 13.1 mm (equal to that of the three-microphone impedance head used for measurements) and the upper bound for the length was decreased to 40 mm from 100 mm.

Note, if the very low frequency range below 50 Hz were to be considered, then an additional tissue compliance parameter would also be necessary for the fit.

Figure 4 shows the *r*(*x*) based on the fits compared with *r*(*x*) derived from MRI. Note that neither corresponds to the target $Z_t(f)$: the MRI is based on a 2D midsagittal image and so is also only an approximation of the true unknown target $A_t(x)$, since it makes the unrealistic assumption of 1D geometry. The major shape features are roughly reproduced, although the fitted radii are generally larger. The total length of the tract is different in the two cases, however, the 2 cm immediately above the glottis has little effect on the impedance through the lips. Its effect on the impedance seen by the glottis is discussed below.

Measurements made on a female subject performing a similar gesture are shown in Figure 5. Note the relatively high impedance of the first minimum compared to the second. This implies that the vocal tract is less well approximated by a uniform cylinder, and possibly that the resonance due to the total tract length terminated at the closed glottis is not as strong as that at the back of the mouth or pharynx, i.e. due to a constriction less sound power reaches and is reflected back from the glottis. Indeed, both fits to amplitude and phase give a large mouth cavity with a constricted pharynx.

Also note the effect of the epilaryngeal tube, the area just above the vocal folds that includes the aryepiglottic folds, on

the impedance at the glottis. The magnitude and phase fits both infer very narrow epilaryngeal tubes, which act as a large inertance on the glottal load (see also Figure 4). However, the combination fit produces a wider epilaryngeal tube, which has the effect of greatly increasing the relative prominence of the first three resonances. Note that here the true $A_t(x)$ is unknown so we cannot (yet) know which is better.

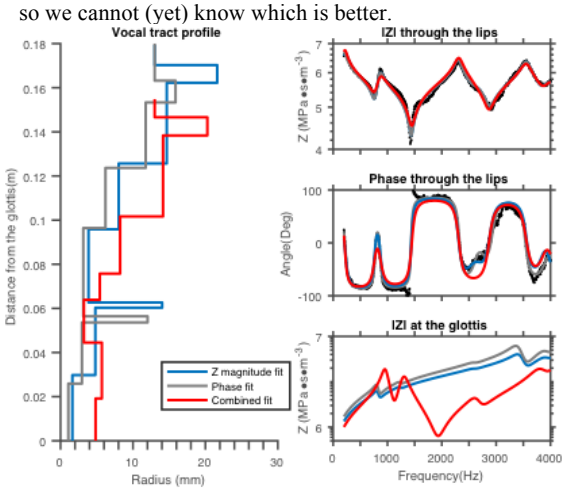

Figure 5. As Figure 4 *for a female subject with glottis closed. r(x) are estimated from fits, $r_t(x)$ is unknown.*

3.2.1. *Including the subglottal tract*

A preliminary trial of fitting to measurements with open glottis during inhalation, showed that at least three additional cylindrical elements are needed to represent the glottis and subglottal tract, which is terminated with an ideal flange, to represent the rapid branching as suggested by [11]. Without considering yielding walls, and using the same loss value as for the vocal tract, this adds six more parameters; three radii and three lengths. However, it gives several more features to fit: because it roughly doubles the length of the non-uniform vocal tract, the number of purely acoustic minima and maxima in $Z(f)$ is approximately doubled.

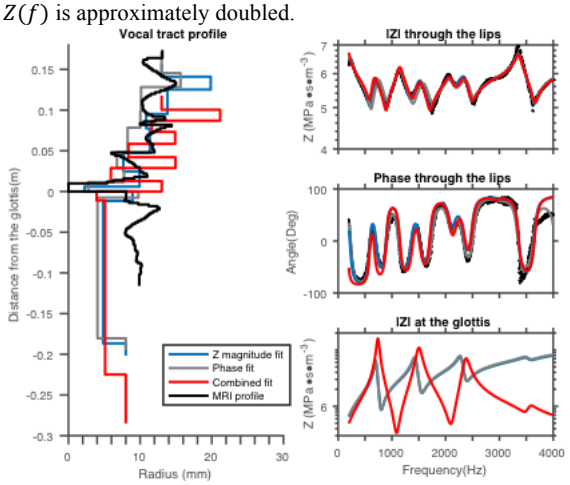

Figure 6. *As Figure 4 for a male subject during inhalation to allow estimation of the subglottal geometry. The end of the black MRI data is due to the MRI image ending about 12 cm below the larynx.*

An example is shown in Figure 6 using the MRI data from [10], which includes several cm of the trachea $r(x)$ below the glottis. Here, although both the magnitude and phase-based fits are reasonable in terms of length, since the acoustic length of that subject's subglottal tract has previously been reported as 195 or 216 mm using the frequencies of the measured impedance extrema and the calculations in [8] and [11]. However, the combined fit is not, showing that more work is needed.

4. Conclusions

For a rigid, dry vocal tract model closed at the glottis, seven or eight cylindrical elements seem to be the optimum model for $Z(f)$. Such a fit gives a rough representation of the shape on a resolution of about a centimetre, and the fitted $Z(f)$ through the lips matches very well. The fit for $Z(f)$ at the glottis is reasonable for the frequencies of its extrema, but not good for their magnitudes, due mainly to the strong influence of the epilaryngeal tube.

To fit *in vivo* measurements, it is necessary to increase the attenuation coefficient α to account for losses on wet, yielding tissues, as has been previously reported [7]. At low frequency, yielding walls introduce the tissue inertance and the tissue losses. These are included here as free parameters required to reproduce the extra low frequency minimum and maximum in $Z(f)$ measured through the lips.

In a preliminary trial, an open cylinder below the glottis was added to reproduce the extra extrema in Z observed when the glottis was open during inhalation, which allowed acoustic excitation of both vocal tract and trachea [8].

5. Acknowledgements

We thank the ARC for support and our volunteer subjects. This project was approved by the UNSW ethics committee.

6. References

[1] Fant, G., Acoustic Theory of Speech Production, Mouton, 1960.

[2] Mermelstein, P., "Determination of the Vocal-Tract Shape from Measured Formant Frequencies", J. Acoust. Soc. America 41, 1283 1967; doi: 10.1121/1.1910470Exam

[3] Ladefoged, P., Harshman, R., Goldstein, L., and Rice, L., J. "Generating vocal tract shapes from formant frequencies" Acoust. Soc. America 64, 1027, 1978, doi: 10.1121/1.382086

[4] Marshall, I. Rogers, M. and Drummond, G., "Acoustic reflectometry for airway measurement. Principles, limitations and previous work" Clin. Phys. Physiol. Meas. 12(2): 131-141, 1991.

[5] Story, B. H., Titze, I. R., and Hoffman, E. A., "Vocal tract area functions from magnetic resonance imaging" J. Acoust. Soc. America, 100(1),537-554, 1996.

[6] Almeida, A., Lehoux, H., Hanna, N., Smith, J. and Wolfe, J., "Estimating pressure and flow at remote locations in a vocal tract from microphone measurements elsewhere", Speech Science and Technology Conference, 2018

[7] Hanna, N., Smith J. and Wolfe, J., "Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation" J. Acoust. Soc. America, 139(5):2924–2936, 2016.

[8] Hanna, N., Smith J. and Wolfe, J., "Acoustic response of the subglottal tract measured by impedance spectrometry through the lips" J. Acoust. Soc. America, 143(5):2639-2650, 2018.

[9] Fletcher, N. H. and Rossing, T. D., The physics of musical instruments. Springer-Verlag, New York, 1998.

[10] Hanna, N., Amatoury, J., Smith, J., and Wolfe, J., "How long is a vocal tract? Comparison of acoustic impedance spectrometry with magnetic resonance imaging" Proc. Mtgs. Acoust. 28, 060001, 2016 doi: 10.1121/2.0000400

[11] Lulich, S.M., Alwan, A., Arsikere, H., Morton, J.R., Sommers, M.S., "Resonances and wave propagation velocity in the subglottal airways," J. Acoust. Soc. Am. 130(4):2108–2115, 2011.