Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Speech Communication: Paper ICA2016-420

How long is a vocal tract? Comparison of acoustic impedance spectrometry with magnetic resonance imaging

Noel Hanna[(a)], Jason Amatoury[(b)], John Smith[(c)], and Joe Wolfe[(d)]

[(a)] School of Physics, University of New South Wales, Australia, n.hanna@unswalumni.com
[(b)] Neuroscience Research Australia (NeuRA) and School of Medical Sciences, University of New South Wales, Australia, j.amatoury@neura.edu.au
[(c)] School of Physics, University of New South Wales, Australia, john.smith@unsw.edu.au
[(d)] School of Physics, University of New South Wales, Australia, j.wolfe@unsw.edu.au

Abstract

Acoustic impedance spectrometry using the three-microphone, three-calibration technique has recently been applied to the vocal tract during phonation (Hanna *et al*., 2016. JASA, 139, 2924–2936). The qualitative and quantitative similarity of the impedance spectrum of the vocal tract with a simple cylindrical duct prompts the question: How well do geometric parameters derived from the measured impedance correspond to the vocal tract morphology? The main aim of this study is to compare, in one male subject (age 34, height 184 cm), the effective acoustic length of the vocal tract derived from impedance spectrometry with the anatomical length measured from a separate magnetic resonance imaging (MRI) scan. Three conditions were studied: 1) acoustic impedance measurements while the subject performed a neutral /ɜ/ vowel gesture with the lips sealed around the impedance measurement head, 2) MRI scan acquired while the subject performed the same gesture with a section of pipe between his lips of the same dimensions as the impedance head, and 3) MRI scan during closed- mouth nasal breathing. Even for the neutral vowel, the effective acoustic length is a (weak) function of frequency. Consequently, each of the acoustic tract resonances gives a slightly different effective length, with a range from 155 to 195 mm with glottis closed. Compared with the 1:3:5:7 ratios expected for cylindrical geometry, the higher resonances have slightly lower frequencies. This is perhaps because the cross-section in the region of the tract closer to the lips is on average greater than that of the region from the palate to glottis. However there is agreement between the length derived from the first acoustic resonance and the smoothed airway centroid length in the MRI of the mid-sagittal plane.

Keywords: vocal tract, resonance, impedance, magnetic resonance imaging

How long is a vocal tract? Comparison of acoustic impedance spectrometry with magnetic resonance imaging

1 Introduction

Recent measurements of the acoustic impedance spectrum of the vocal tract show qualitative and quantitative similarities between the vocal tract in a neutral configuration and a uniform cylindrical duct [1, 2]. This simple model, which can also accommodate non-rigid vocal tract walls, yields an effective vocal tract length by making the assumptions of 1D plane wave propagation and no large discontinuities or variations in the area function of the vocal tract. The average effective tract length (from a plane inside the lips to closed glottis) derived from impedance spectrometry was 173 mm [1], which is comparable with that determined from a similar location using acoustic pharyngometry (171 mm for male subjects of similar height [3]) and the expected geometrical length. This rough correspondence is helpful in teaching acoustic phonetics: indeed Fant's introduction to speech production [4] gives the example of a uniform cylindrical tract of length 17 cm, which gives resonances near 0.5, 1.5, 2.5 and 3.5 kHz that lie in the range of measured formants F1-F4, and close to the values for the neutral vowel /ɜ/.

To what extent does this simple model correspond to the vocal tract morphology when examined in detail? Although comprehensive studies of vocal tract geometry from magnetic resonance imaging (MRI) have been made (e.g. [5]) these are usually restricted to one configuration. However, the tract length and cross-sectional area varies depending on the vocal gesture, *i.e*. protrusion of the lips, position of the tongue constriction, mouth aperture *etc*., as shown with different vowel sounds [6]. Thus, in order to compare impedance spectrometry to the vocal tract morphology from MRI, measurements should be made on the same subject performing the same vocal gesture. In this study, we will use both methods to compare different estimates of the length, radius and cross-sectional area of the vocal tract.

Measurements during glottal closure should provide a complete acoustic reflection at a position that is readily determined from MRI. However, measurements during phonation, when the glottis is open, are expected to be more problematic, since the acoustic reflection would depend on the glottal aperture. Furthermore, because the tract is curved, the length derived from MRI depends on the method of measurement (e.g. defined anatomical path, airway centroid), which will give a range of lengths for a single MR image.

2 Material and methods

2.1 Experimental setup

The study was performed on one male subject (age 34, height 184 cm). Ethics approval was obtained from the Human Research Ethics Committees of the University of New South Wales. In both MRI and acoustic impedance spectrometry measurements, the subject was in a supine position, with the head aligned approximately with the Frankfort plane (the plane between the orbit of the eye and tragus of the ear) at 90° to the horizontal.

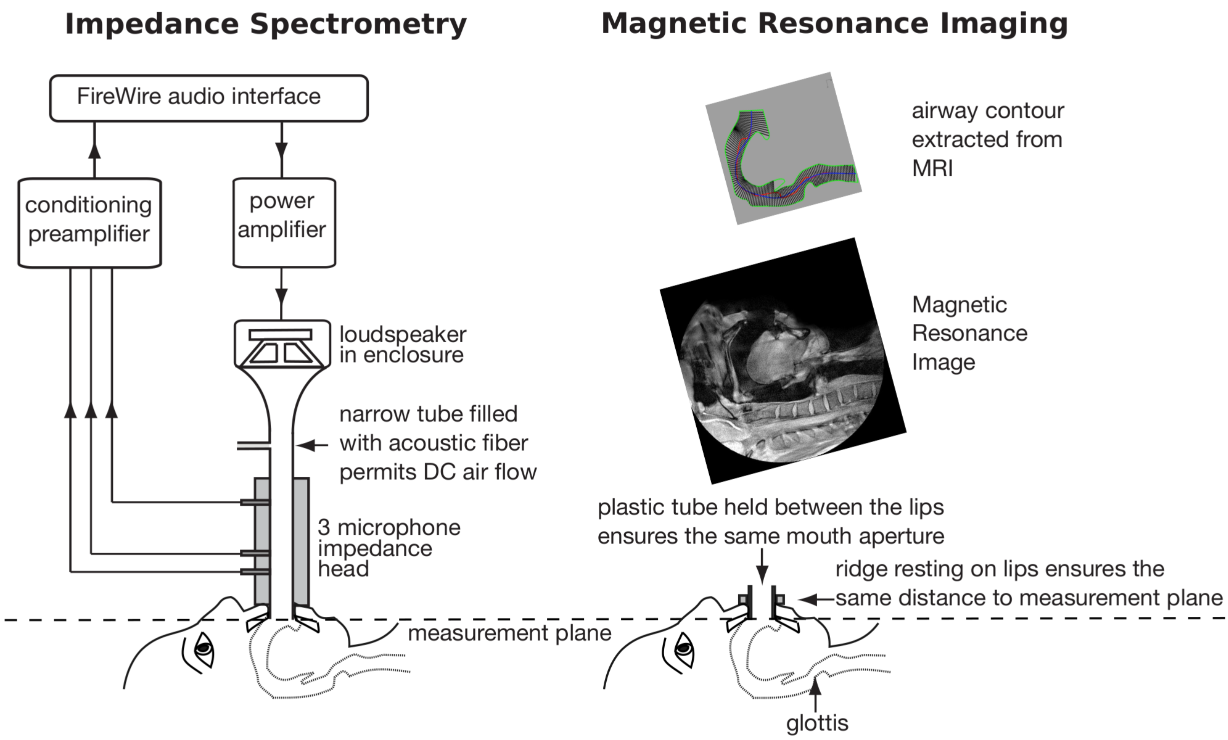

Figure 1: Schematics illustrating the experimental configurations

2.2 Magnetic Resonance Imaging (MRI)

MRI was performed using a 3T MRI scanner (Philips, Achieva 3T, Best, The Netherlands) and head-neck coil (Philips). A mid-sagittal image of the vocal tract was acquired using a fast field echo (FFE) 2D radial sequence with the following parameters: flip angle = 10°, repetition time/echo time = 4.3/1.83 ms, density of angles = 200%, 512×512 matrix, field of view = 230×230mm, and slice thickness = 8 mm.

MRI data were acquired while the subject 1) produced a sustained neutral /ɜ/ vowel in his modal (chest) speaking voice, 2) closed his glottis while miming the same vowel, and 3) closed his mouth and breathed through his nose. For measurements 1) and 2), a section of plastic pipe was held between the lips. Since the acoustic impedance measurement head constrains the mouth geometry, a section of plastic tube of the same diameter as the impedance measurement head was held between the subject's lips during the MRI to provide the same constraint as the acoustic measurements, as shown in Figure 1. The tube had a ridge 10 mm from its end, which rested on the outer edge of the lips. This ridge produced a visible indentation in the top and bottom lip, which was used to define the bilabial lip plane, and ensured that the tube protruded 10 mm inside the mouth from that plane, in the same way as the impedance head.

2.3 Impedance spectrometry

Impedance spectra were obtained in a separate session using a three-microphone impedance head as described in detail elsewhere [1, 7, 8]. Briefly, a calibrated broadband signal synthesized from sine waves between 200 Hz and 4 kHz with a spacing of 2.69 Hz (44.1 kHz/2^{14}) and phases chosen to improve the signal-to-noise ratio [9] was sent through the lips into the mouth of the subject through a cylindrical impedance head. The signals from the three microphones in this cylindrical impedance head are then used to determine the acoustic impedance (complex ratio of acoustic pressure to flow) at the measurement plane at the end of the impedance head, which was located inside the mouth and 10 mm from the outside edge of the lips. A schematic is given in Figure 1.

Four sets of measurements were made for the closed glottis and phonation conditions. Each measurement set consisted of 8 contiguous cycles, each lasting 370 ms. The frequencies of the impedance minima f_{Ri} and maxima $f_{\overline{\mathrm{R}}\mathrm{i}}$ were then identified as extrema in the impedance spectra [2].

2.4 Calculation of vocal tract length

The vocal tract length was calculated using the following methods:

2.4.1 MRI: distance between anatomical landmarks

Following Fitch and Giedd [5], the straight-line distances between the following sequential pairs of anatomical landmarks were made (Figure 2, black): bilabial lip plane, tongue tip, alveolar ridge, hard palate/velar junction, uvula, and superior surface of the glottis (taken at the base of the epiglottis; Figure 2).

2.4.2 MRI: length of airway centroid curve

The length of the airway centroid cureve, formed by the mean distance between the vocal tract walls, was measured from the bilabial lip plane to the base of the epiglottis. The smoothed centroid curve (using a 20 mm moving average filter) length was also measured. Figures 2 and 5 show the centroid in blue and the smoothed centroid in red.

2.4.3 Spectrometry: cylinder lengths derived from resonances and anti-resonances

Acoustic lengths were calculated from resonance and anti-resonance frequencies. For a uniform cylinder of length *l* that is closed at the far end, the acoustic resonances Ri, i = 1,2,3… occur at frequency f_{Ri}, such that

$$l = \frac{nc}{4f_{\mathrm{Ri}}}, \qquad n = 2i - 1 \tag{1}$$

where c = 354 m·s^{-1} is the speed of sound in saturated air at 35°C. Anti-resonances $\overline{\mathrm{R}}\mathrm{i}$, i = 1,2,3… occur at frequency $f_{\overline{\mathrm{R}}\mathrm{i}}$, such that

$$l = \frac{nc}{2f_{\bar{\mathrm{R}}\mathrm{i}}}, \qquad n = i \tag{2}$$

Hence, each resonance or anti-resonance provides an estimate for the length *l* from the measurement plane to the glottis. Previously, Hanna *et al*., calculated the mean of the lengths from f_{Ri} and $f_{\bar{\mathrm{R}}\mathrm{i}}$ for i=1,2,3 [1]. In the current study, each resonance is examined separately.

2.5 Calculation of vocal tract radius and cross-sectional area

2.5.1 MRI

The diameters (*D*) of the entire tract were measured in the mid-sagittal plane at 2 mm intervals along its length from the lips to the level of the T1 vertebra. D was defined as the airway transect in the mid-sagittal plane between the airway walls perpendicular to the smoothed airway centroid curve (shown by the black lines in Figure 5). The effective radius r = D/2, since the centroid curve is equidistant from the two walls.

Since no axial images were included in the MRI scan, it was not possible to reconstruct the vocal tract area function directly. Instead, for illustrative purposes we use the following relationship between the transect *D* and the area *A*

$$A \ = \ 0.44 \cdot D^2 + 61 \ \mathrm{mm}^2 \tag{3}$$

This empirical relationship between based on MRI data on a single subject [10] via Perrier *et al*. [11] was used to estimate the area function. Echternach *et al.* [12] report that this approach is in reasonable agreement with 3D MRI data for the purposes of generating formants.

2.5.2 Spectrometry

The geometric mean of the absolute impedance magnitude spectrum (horizontal line in Figure 4; an arithmetic mean on a log scale) was used as equivalent characteristic impedance. This yields an effective area *A* by its definition.

$$|\mathrm{Z}|_{\mathrm{mean}} \sim |\mathrm{Z}|_{\mathrm{characteristic}} = \frac{\rho c}{A} \tag{4}$$

Where ρ is the density of air and c the speed of sound. For air at 35°C, ρ = 1.146 kg·m^{-3}. This provides a single representative cross-sectional area of the tract. Then, assuming that the tract is cylindrical $r = \sqrt{(A/\pi)}$.

3 Results

3.1 Length of the vocal tract

3.1.1 MRI length

For the closed glottis measurements, the landmark path (black, Figure 2) gives a length of 198 mm, which is judged unrealistic because the tongue tip to alveolar ridge segment is at a large angle to the airway. While more representative than the landmark path, a line through the airway centroid (midpoint between airway walls) (blue, Figures 2 and 5) still includes some unnecessary deviations (in particular due to the epiglottis), giving a length of 168 mm. The smoothed centroid curve using a 20 mm moving average (red, Figure 2), which is more representative of the airway's overall curvature, gives a shorter length of 158 mm. The lengths calculated from the phonation measurements for the landmark path, centroid and smoothed centroid, were similar to the three above: 197, 163, and 156 mm, respectively.

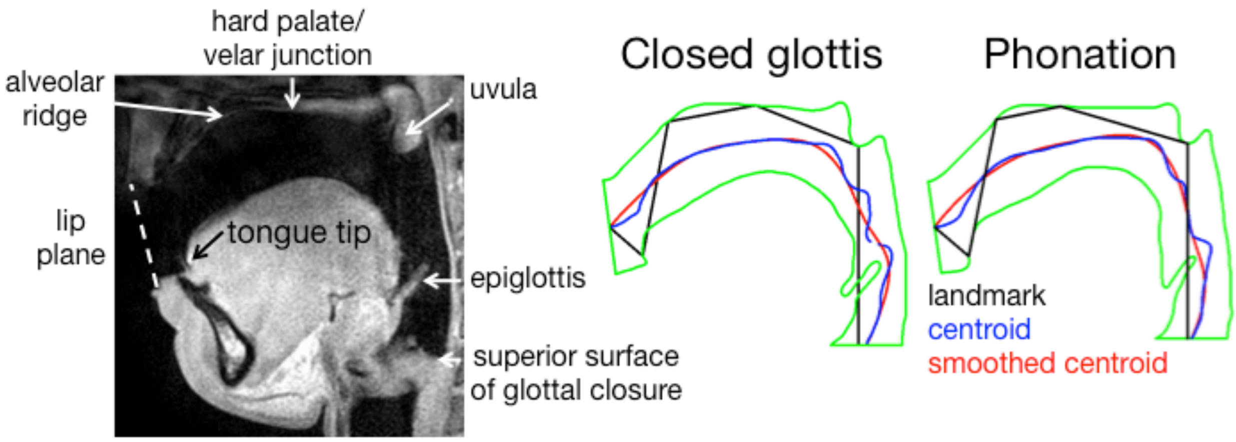

Figure 2: (Left) MRI during the closed glottis condition, identifying vocal tract landmarks. Mid-sagittal airway contours with line segments and airway centroid curves for closed glottis (middle) and phonation (right). The black line shows the landmark path: lip plane, tongue tip, alveolar ridge, hard palate/velar junction, uvula, and superior surface of the glottis, the blue shows the centroid curve and the red the smoothed centroid curve.

The length of the landmark path for nasal breathing was 151 mm, which is shorter than the adult male average landmark length of 161 mm found by Fitch and Giedd [5]. For nasal breathing, there is no single centroid path from lips to glottis since the velum provides complete (or near complete) closure off the bucal cavity. Thus, a centroid curve length measurement was not undertaken. Fitch and Giedd [5] drew curved lines by hand and obtained an average value for adult males of 155 mm, similar to the smoothed centroid length in the current study for both glottis closed and phonation.

3.1.2 Acoustic length

Figure 3 shows examples of measured impedance magnitude spectra with glottis closed (pale) and during phonation (dark).

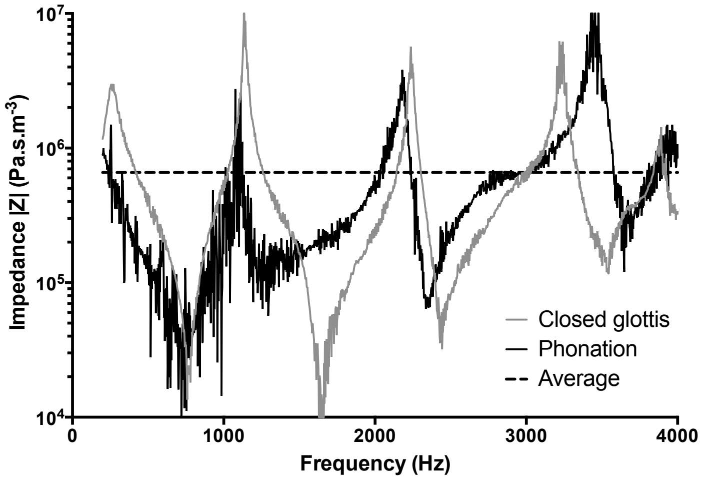

Figure 3: Example impedance magnitude spectra, each measured during one cycle lasting 370 ms, are shown for the closed glottis condition (pale) and phonation (dark). The horizontal dashed line is the geometric mean of the data for the closed glottis condition.

The vocal tract lengths calculated from the frequencies of each acoustic resonance f_{Ri} and anti-resonance $f_{\overline{\mathrm{R}}\mathrm{i}}$ (disregarding the acousto-mechanical resonances below 400 Hz, see [1]) are shown in Figure 4.

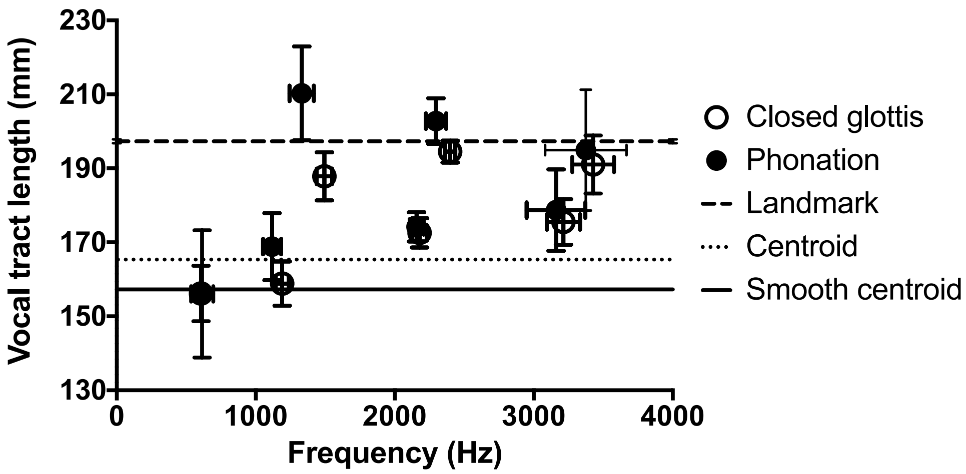

Figure 4: Calculated vocal tract lengths from each resonance (the 10 mm between measurement plane and lips has been added), represented as mean ± standard deviation. Horizontal lines show the MRI lengths in the closed glottis condition from lips to the superior surface of the glottis.

The bilabial lip plane used for the MRI measurements was taken where there was visible lip indentation due to the ridge of the plastic tube. This was 10 mm from the impedance head measurement plane, so, in Figure 4 and elsewhere, 10 mm is added to all the acoustic lengths determined from this measurement plane. (Note: if an acoustical length was determined from a formant, an acoustical end effect would need to be subtracted for comparison with the MRI lengths. The measurements here have no end effect.)

The average of the effective lengths (for all resonances) from glottis to lips is 177 mm with a standard deviation of 15 mm with the glottis closed, and 185±20 mm during phonation. In both cases the length derived from f_{Ri} (156±8 mm and 156±17 mm) are in agreement with the smoothed centroid length of 158 mm with glottis closed and 156 mm during phonation.

3.2 Radius and area of the vocal tract

3.2.1 MRI radius and cross-sectional area

Calculations of the airway radiii are plotted in Figure 5. Images showed about 100 mm of subglottal tract below the glottis and, for completeness, these data are included in Figure 5 but not used in calculations. The average cross-sectional area calculated from Equation (3) for the supraglottal vocal tract with glottis closed and phonation was 220 mm^2 and 250 mm^2, respectively.

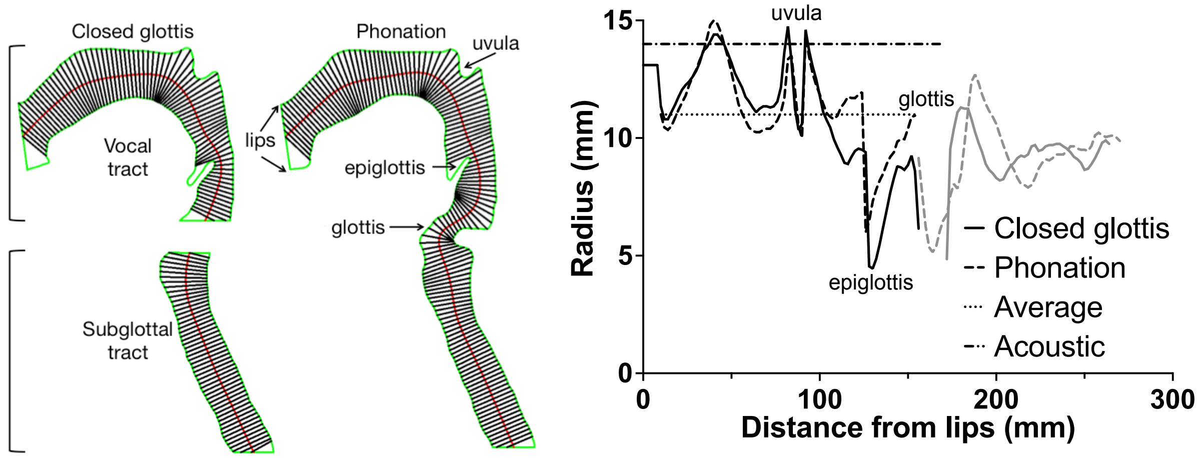

Figure 5: (Left) Airway contour segmentations (green) for closed glottis and phonation, including subglottal tract, with smoothed centroid (red) and perpendicular airway transect (black). (Right) Variation of vocal tract radius (supraglottal – black; subglottal – pale) with distance from the lips.

3.2.2 Acoustic radius and cross-sectional area

A single effective value of the area was calculated based on the geometric mean of impedance magnitude, |Z| = 658 kPa·s·m^{-3} (horizontal line in Figure 3) (699 kPa·s·m^{-3} during phonation). So A = 615 mm^2 from Equation (3) and r = 14 mm (shown in Figure 5) with the assumption that the tract is cylindrical.

Figure 5 suggests that the oral cavity of the vocal tract (between lips and velum) has a larger cross section than the pharyngeal cavity (from velum to glottis). One effect of this on the expected acoustic impedance spectrum is to raise the frequencies of the lowest resonances, which may explain that trend in Figure 4. It also tends to lower the characteristic impedance, helping to explain the larger acoustic radius in Figure 5.

4 Discussion

The idea of an effective vocal tract length obtained from an acoustic measurement is inherently simplified: any human tract is of course capable of producing a wide range of resonance frequencies, depending on the vocal gesture. Nevertheless, because of the approximately regular spacing of the resonances and anti-resonances for the neutral vowel, it is interesting to suggest an effective length for that vowel.

When considering the MRI images, a key decision in determining an effective length is deciding on the location of the glottis. During phonation, the glottis is defined as the aperture between the vibrating vocal folds. However, during a glottal stop or a mimed "closed glottis" gesture the position of the termination of the acoustic duct is ~10 mm above the vocal folds. For this study, in both cases, the glottis was defined in MRI images, as the base of the epiglottis.

Which geometric length is most useful to acousticians? Different measurement algorithms give different vocal tract lengths. However, one of the aims of determining a geometric vocal tract length is to provide input to acoustic models of the vocal tract. So, for the neutral vowel at least, it is interesting to know which algorithm best matches the acoustic length. The data presented here suggest that there is correspondence between the length derived from f_{R1} and the smoothed centroid length from MRI.

The attraction of the 1D uniform cylinder is its simplicity. The present study suggests that it is a fair approximation for the length when considering the neutral vowel. The effective radius obtained from the cylindrical duct model is a poorer approximation, as one might expect when fitting to a duct of irregular shape whose cross section is sometimes far from circular. Similarly for mid-sagittal MRI, a single radius measurement is assumed to approximate the cross section of the tract at that position.

A useful application of this work is in teaching: a discussion of acoustic phonetics can use an approximation of vocal tract length that is in fair agreement with the geometry. A more demanding application would be as a foundation for the method of Bunton *et al*., [13], which aims to estimate the area function of the vocal tract from lip opening when provided with an effective tract length. This study suggests that an approximation of the effective length can be obtained from acoustic measurements on the neutral vowel.

5 Conclusions

A range of effective or average values for vocal tract lengths, radii and area can be derived from both impedance and mid-sagittal MRI. For the neutral vowel, the reasonable agreement is helpful for simple calculations and modelling. The range of lengths derived from the first three acoustic resonances f_{Ri} and anti-resonances $f_{\overline{\mathrm{R}}\mathrm{i}}$ was 156 to 195 mm (156 to 210 mm during phonation), which largely overlaps those obtained from different geometric algorithms, and include the value of the length from the MRI smoothed centroid data of 158 mm (156 mm during phonation). The strongest agreement is obtained from the length derived from f_{R1} (156 mm) and the smoothed centroid lines for both closed glottis and phonation conditions.

Acknowledgments

We thank the Australian Research Council and NeuroSleep National Health and Medical Research Council of Australia (NHMRC) Centre of Research Excellence (#1060992). NH was awarded an ICA-ASA Young Scientist Conference Attendance Grant to present this paper.

References

[1] Hanna, N.N., Smith, J., Wolfe, J. Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation, *The Journal of the Acoustical Society of America*, 139, 2016, pp 2924–2936.

[2] Hanna, N.N., Smith, J., Wolfe, J. Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses, *Proceedings of the Australian Acoustical Society Conference*, Fremantle, Australia, 2012 pp 317–323.

[3] Xue, S.A., Hao, J.G. Normative standards for vocal tract dimensions by race as measured by acoustic pharyngometry, *Journal of Voice*, 20, 2006, pp 391–400.

[4] Fant, G. *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*, Mouton De Gruyter, The Hague, 1970, pp 6-12

[5] Fitch, W.T., Giedd, J. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America* 106, 1999, pp 1511–1522.

[6] Story, B.H., Titze, I.R., Hoffman, E.A. Vocal tract area functions from magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 100, 1996, pp 537–554.

[7] Dickens, P., Smith, J., Wolfe, J. Improved precision in measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution. *The Journal of the Acoustical Society of America*, 121, 2007, pp 1471–1481.

[8] Hanna, N.N. *Investigations of the acoustics of the vocal tract and vocal folds in vivo, ex vivo and in vitro* (Phd Thesis), UNSW, Australia and Université de Grenoble, France (https://tel.archives-ouvertes.fr/tel-01174056/), 2014, pp. 31-66

[9] Smith, J.R. Phasing of Harmonic Components to Optimize Measured Signal-to-Noise Ratios of Transfer-Functions, *Measurement Science and Technology*, 6, 1995, pp 1343–1348.

[10] Baer, T., Gore, J.C., Gracco, L.C., Nye, P.W. Vocal tract dimensions obtained from magnetic resonance images. *The Journal of the Acoustical Society of America*, 84, 1988, pp S125.

[11] Perrier, P., Boë, L.-J., Sock, R. Vocal tract area function estimation from midsagittal dimensions with ct scans and a vocal tract cast modeling the transition with two sets of coefficients. *Journal of Speech, Language, and Hearing Research,* 35, 1992, pp 53–67.

[12] Echternach, M., Birkholz, P., Sundberg, J., Traser, L., Korvink, J.G., Richter, B. Resonatory properties in professional tenors singing above the passaggio. *Acta Acustica united with Acustica*, 102, 2016, pp 298–306.

[13] Bunton, K., Story, B.H., Titze, I. Estimation of vocal tract area functions in children based on measurement of lip termination area and inverse acoustic mapping, *Proceedings of Meetings on Acoustics*, 19, 2013 pp 060054.