VOCAL TRACT RESONANCES: A PRELIMINARY STUDY OF SEX DIFFERENCES FOR YOUNG AUSTRALIANS

Tina Donaldson, Diana Wang, John Smith and Joe Wolfe
School of Physics, University of New South Wales, Sydney 2052

Abstract: We report direct measurements of the first two resonance frequencies of the vocal tracts of young women university students producing the vowels of Australian English. The resonances are determined from the response of the tract to a broad band, external, acoustic source. From these data we construct a vowel resonance map for these Australian women and compare it with the corresponding data for a sample of young Australian men, also university students.

1. INTRODUCTION

Each vowel sound in a language or dialect is characterised by a set of formants, which are broad maxima of acoustic power in the speech spectrum [1,2]. These formants are produced by resonances of the vocal tract, which in turn depend on its geometry, including the height of the jaw and the position and shapes of the tongue and lips. A plot that locates each vowel by the frequencies of the two formants or resonances with the lowest frequencies is called a vocal plane or vowel map. The formant frequencies of Australian English have been measured for male speakers [3,4]) and female speakers [5-7]. For reasons that we explain below, formants are more difficult to measure objectively in women than in men. Furthermore the precision of measurements can be improved considerably if the resonances of the tract rather than the formants of speech are measured. Recently the vocal tract resonances have been measured directly for a sample of young Australian men, who were students at the University of New South Wales in Sydney [8]. Here we measure directly, for the first time, the vocal tract resonances for vowels in Australian English as spoken by young Australian women. The sample was taken from students at the same university.

We begin with a brief overview of the source-filter model of voiced speech. In this model [1] (see Figure 1), the vibration of the vocal folds produces a periodic, harmonic-rich signal at the fundamental frequency f_o. This signal is transmitted to the radiation field outside the mouth by the vocal tract, which has a frequency dependent gain. Resonances of the vocal tract produce peaks in the gain spectrum that in turn give rise to maxima in the envelope of the speech spectrum. The broad peaks in the output sound spectrum are called formants. For non-nasalised speech, the human vocal tract may be approximated as a tube that is nearly closed at the glottis or vocal folds and open at the mouth. The radiated power of speech is increased (all else equal) when the tract acts as an impedance matcher from the low acoustic impedance of the radiation field at the mouth to the higher impedance at the glottis: in other words, for resonances with a pressure anti-node near the glottis and a node near the mouth. If the vocal tract were a tube of length L with uniform cross section, these resonances would occur at wavelengths $\lambda \cong 4L$, $4L/3$, $4L/5$ etc. Taking $L \sim 170$ mm, the resonance frequencies would be approximately 500, 1,500, 2,500 Hz, etc. (In fact a tract pronouncing the vowel [ɜ] as in "heard" has resonances at approximately these frequencies.) However, changing the shape of the tract varies considerably the frequencies of the resonances.

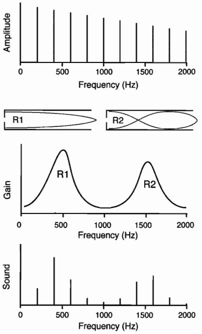

Figure 1. The source-filter model for voiced speech. The harmonic-rich signal from the vocal folds (top) is transmitted to the radiation field (bottom) via the tract. The tract most effectively matches the impedance at its resonances. The middle sketch represents the tract as a uniform cylinder and shows the pressure amplitudes. In practice, the resonance frequencies are modified by moving the jaw, tongue and lips.

The frequency *R1* of the first resonance is determined largely by the height of the jaw and thus the mouth opening. (As a tube is increasingly flared at the open end, the lowest resonance frequency rises, and the spacing of resonances is decreased.) The frequency *R2* of the second resonance is more strongly determined by the position at which the tongue constricts the mouth (high for tongue constriction forward and conversely). In languages (such as English and other European languages) that do not use lexical tone, *R1* and *R2* largely determine the vowel sound, while *R3* and *R4* mainly carry information characteristic of the speaker.

The vocal folds vibrate at a fundamental frequency f_o, which is typically in the range 80 to 200 Hz for men and about 150 to 300 Hz for women. f_o is also the spacing between harmonics in the speech sound. It is this spacing that limits the resolution in determination of the formants — the peaks in the spectral envelope — and which consequently makes determination of formants for women in general less precise than for men. Signal processing algorithms for determining the formants require parameters input by the experimenter, and when precisions substantially smaller than f_o are sought, the values of these parameters affect the values of formants measured. Fig 1 illustrates the difficulty of obtaining precise values of the formant frequencies when the harmonics are spaced by an f_o of 200 Hz, typical of women's speech. In this study, we overcome this problem by employing an external source of acoustic current at the mouth to excite the vocal tract while the subjects phonate. This allows determination of the resonances of the tract with a typical resolution of ± 10-20 Hz [8].

2. MATERIALS AND METHODS

The method is an adaptation of one described previously [8-10]. Briefly, a computer (Macintosh IIci) uses a analogue/digital card (National instruments NB-A2100) to synthesise a waveform as the sum of sine waves with frequencies from 200 to 4,500 Hz, with a spacing of 5.4 Hz. This waveform is amplified and passed to a loudspeaker that is matched via an exponential horn to a pipe of inner diameter 6 mm. The end of this pipe, filled with acoustic absorbing material, is an acoustic current source, whose characteristic output impedance is about 16 GPa s m^{-3} or 16 GΩ. This source is placed vertically so that the end of the pipe just touches the subject's lower lip. A microphone (8 mm diameter), whose signal is recorded by the same A/D card and computer, is attached to the end of the pipe.

For each subject, a calibration procedure is conducted, during which the amplitudes of the individual sine waves are adjusted so that the microphone signal measured with the subject's mouth closed is independent of frequency. During this calibration, the acoustic pressure p_{cal} at the microphone is $u_{cal}Z_{rad}$, where u_{cal} is the acoustic current and Z_{rad} the impedance of the radiation field at the (closed) mouth, as baffled by the subject's face. (The disturbance of the radiation field by the presence of the source and microphone shifts the measured resonance frequency by 11 Hz or less, which does not exceed the precision of the measurements.) Because of the high output impedance of the source, the current produced during a measurement is almost identical to that produced during calibration. During phonation, the microphone signal is the sum of that due to the subject's voice (which consists of harmonics of f_o) and that produced by the interaction of the injected acoustic current with the subject's vocal tract. The acoustic impedance of the subject's tract Z_{tract} is in parallel with Z_{rad}, so the broad band component of the acoustic pressure is thus $p_{meas} = u_{cal}Z_{rad}Z_{tract}/(Z_{rad} + Z_{tract})$. We plot the ratio γ of the microphone signals for measurement and calibration. For the broad band component of the signal, this yields

$$\gamma = \frac{p_{meas}}{p_{cal}} = \frac{Z_{tract}}{Z_{rad} + Z_{tract}} = \frac{1}{1 + Z_{rad}/Z_{tract}}$$

Making the assumption that the frequency variation of Z_{rad} is much less than that of Z_{tract}, γ has maxima when Z_{tract} has maxima.

The subjects were nine Australian women, aged from 18 to 20, who were first year physics students at the University of New South Wales in Sydney. Their data are thus suitable for comparison with those for males [8]. All had been born in Australia or had lived in Australia for longer than seven years and were recognised by the investigators as having unremarkable Australian accents. The vowels were presented in a /h_d/ or /h_t/ context. The words used (with phonetic vowel symbols in brackets) were "heed" [i], "hid" [ɪ], "head" [ɛ], "had" [æ], "hard" [ɑ], "hot" [ɒ], "hoard" [ɔ], "hood" [ʊ], "who'd" [u], "hut" [ʌ] and "heard" [ɜ]. They were asked to pronounce and to sustain the each of the words for four seconds, whilst each measurement was made. The series was then repeated.

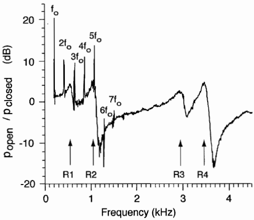

Figure 2. The ratio of the spectra measured with the mouth open to that with the mouth closed (p_{open}/p_{closed}) for the vowel /ɒ/ (in "hot"). Several harmonics of the voice signal with fundamental frequency f_o = 215 Hz can be seen. The maxima in the broad band signal corresponding to the resonances R1, R2, R3 and R4 are indicated by arrows.

3. RESULTS AND DISCUSSION

Figure 2 shows the magnitude of the measured ratio $\gamma = p_{meas}/p_{cal}$ for one of the subjects pronouncing the vowel [ɒ] in "hot". The narrow peaks are the harmonics of the fundamental $f_o = 215$ Hz. The broader peaks in the broad band signal at about 550, 1050, 2950 and 3450 Hz are due to the resonances of the tract. (In this example, note how the first resonance is more easily identified than the first formant.)

Figure 3 shows *R1* plotted against *R2*. (Plots of *F1 vs F2*, with axes inverted, are traditional in acoustic phonetics because phoneticians have traditionally plotted jaw height vs position of the tongue constriction.) The relative positions of the vowels are similar to those in the comparable resonance plot for young Australian men [8]. The relative positions are also similar to those reported for the formants of Australian English [5]. Apart from the intrinsic differences between resonances and formants, we measured sustained vowels in this study, whereas Cox [5] measured them in normal speech. The substantial overlap between "hard" [ɑ] and "hut" [ʌ], and between "heed" [i] and "hid" [ɪ] may seem surprising until one realises that these pairs are usually distinguished in normal speech by duration. The data are also included in the Table.

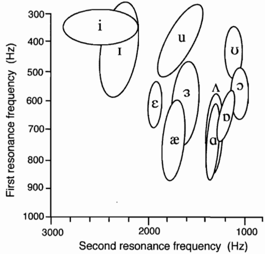

Figure 3. The distribution of (R2,R1) for the vowels of English as spoken by this sample of young Australian women. The centre of each ellipse is the mean of (R2,R1). The slope of the major axis indicates the regression of R2 on R1, and the semiaxes are the standard deviations in those directions.

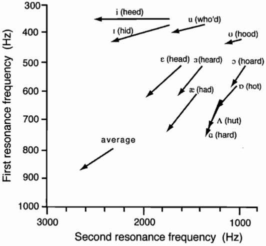

Figure 4. The displacement of the average resonance data for women reported herein from those reported for Australian men [8]. The displacement averaged over all vowels is also indicated.

In all cases, the frequencies *Ri* for the female subjects were higher than those for males, except for *R1* in "heed" [i]. Figure 4 shows that the average displacement of women's from men's data is thus approximately away from the origin of the vowel plane. The average value of the increase in *R* for the women's data was 12% (65 Hz) for *R1* and 20% (290 Hz) for *R2*. The displacements are comparable with the average increases in reported formant frequencies for Australian English (20% in *F1* and 15% in *F2* [5]) and Greater American English (16% in *F1* and 25% *F2* [11]). One possible explanation for the difference is a difference in the average lengths of male and female vocal tracts. However, social effects may be important, too, and people may learn to produce the resonances appropriate for their sex and f_o: a person with a relatively long vocal tract could readily raise *R1* and *R2* for each vowel simply by opening the mouth and advancing the tongue by a small amount.

Table. The mean and standard deviation for the resonant frequencies of the vowels spoken by Australian university students, male and female.

vowel	/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ɒ/	/ɔ/	/ʊ/	/u/	/ʌ/	/ɜ/
word	heed	hid	head	had	hard	hot	hoard	hood	who'd	hut	heard
R1 female	350±60	420±80	600±60	730±110	740±130	650±70	570±70	430±70	390±80	710±100	600±40
R1 male	350±40	370±50	510±50	610±60	630±60	590±60	510±50	420±40	370±50	630±60	510±40
R2 female	2490±390	2300±250	1930±90	1740±150	1330±70	1200±100	1060±110	1110±120	1670±250	1300±130	1620±160
R2 male	1730±200	1720±170	1610±120	1440±120	1200±110	1030±80	940±130	980±210	1350±230	1180±140	1380±90

ACKNOWLEDGMENTS

We thank our volunteer subjects. This paper is based on an undergraduate research project by TD and DW.

REFERENCES

[1] Fant, G., *Speech Sounds and Features.* MIT, Cambridge, Mass (1973).

[2] Clark, J. and Yallop, C., *An Introduction to Phonetics and Phonology*, Basil Blackwell Ltd, Oxford (1990).

[3] Bernard, J., 'Towards the acoustic specification of Australian English' *Zeitschrift fur Phonetic, Sprachwissenschaft und Kommunikationsforschung, 23*, 113-128 (1970).

[4] Bernard, J.R. and Mannell, R.H., "A study of /h_d/ words in Australian English", Working Papers of the Speech, Hearing and Language Research Centre, Macquarie University (1986).

[5] Cox, F., *An Acoustic Study of Vowel Variation in Australian English*, Doctoral dissertation, Macquarie University, Sydney (1996).

[6] Cox, F., 'Vowel change in Australian English', *Phonetica, 56*, 1-27 (1999).

[7] Cox, F., and Palethorpe, S., 'The changing face of Australian English vowels', In English in Australia. eds D.Blair and P.Collins. John Benjamins Amsterdam/Philadelphia pp 17-44, (2001).

[8] Epps, J., Smith, J.R., and Wolfe,J., 'A novel instrument to measure acoustic resonances of the vocal tract during speech', *Measurement Science and Technology, 8*, 1112-1121 (1997)

[9] Dowd, A., Smith, J.R., and Wolfe, J., 'Real time, non-invasive measurements of vocal tract resonances: application to speech training', *Acoustics Australia, 24*, 53-60 (1996).

[10] Dowd, A., Smith, J.R., and Wolfe, J., 'Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time', *Language and Speech, 41*, 1-20 (1997).

[11] Petersen, G.E., and Barney, H.L., 'Control methods used in a study of vowels', *J. Acoustic Soc. Am. 24*, 175-184 (1952).