The role of vocal tract resonances in singing and in playing wind instruments

John Smith* and Joe Wolfe

School of Physics, University of NSW, Sydney NSW 2052

ABSTRACT

The different vowel sounds in normal speech are produced by adjusting the position of tongue, lips and teeth so that the vocal tract resonates at certain specific frequencies. In voiced speech, these resonances interact with the harmonics of the lower frequency signal from the vibrating vocal folds to produce associated peaks, or formants, in the output spectrum. Singers sometimes use these resonances in musical rather than linguistic ways. For sopranos, the vibration frequency of their vocal folds may be much higher than the normal values for the lowest resonance, and consequently a reduced interaction would cause a loss of power. Direct measurements of the resonance frequencies of the vocal tract of classically-trained sopranos during singing show that they consistently increase them to match the frequency of their singing. This significantly increases the loudness and the uniformity of tone, at the expense of comprehensibility. The fundamental frequency of other singers is usually less than the value of the lowest resonance and so they would experience no advantage in tuning this resonance. However the power could be increased if the resonance frequency were tuned to a harmonic of the fundamental frequency. Our measurements indeed show that some altos, tenors and baritones use this strategy when appropriate. The role of the vocal tract resonances is quite different when playing a wind instrument. The sound is then generated by the vibrating lip or reed valve rather than by the vibrating vocal folds. The frequency of vibration is then primarily determined by one of the strong resonances of the wind instrument itself. Our measurements show that varying the resonances of the vocal tract can then still slightly alter the vibration frequency and change the harmonic structure or timbre of the produced sound. The described research has involved several members and associates of our Acoustics Laboratory.

INTRODUCTION

Although its original function simply involved the ingestion of food, the vocal tract in humans has evolved into a system that played a crucial role in the development of modern society. In this paper we discuss how techniques developed in our laboratory for measuring acoustic impedance rapidly and precisely, have enabled us to investigate the resonant behaviour of the vocal tract during speech, singing and whilst playing a wind instrument.

THE VOCAL TRACT

In normal speech and singing, the vocal folds are approximately a flow-driven relaxation oscillator located at the larynx. The oscillatory opening between them is known as the glottis. The vibrating vocal folds generate a harmonically rich signal with pitch frequency f_0, which is transmitted via the vocal tract into the surrounding air (Fant 1973) – see Figure 1. The fundamental frequency of vibration (which determines the pitch) is controlled by varying their tension and/or mass distribution. In normal speech f_0 is typically 110 Hz for men, 220 Hz for women and around 300 Hz for children. Its value is varied continuously during speech, and contributes to the prosody (e.g. a rise in pitch at the end of a sentence can indicate a question). Resonances within the vocal tract are controlled independently of f_0 by varying its shape using the position of the tongue, jaw and lips. These resonances serve to match the relatively high acoustic impedance of the lower tract near the vocal folds to the low impedance of the radiation field outside the mouth, and so produce broad peaks called formants in the spectral envelope of speech – see Figure 2

Vowels in Western European languages are generally identified by the frequencies of the first two formants (*F1*,*F2*) which are in turn determined by their associated tract resonances (*R1*,*R2*). In spoken Australian English, *R1* can range typically from 300 to 850 Hz, and *R2* can vary from 800 to 2900 Hz depending upon the vowel and gender (Epps, Smith and Wolfe 1997; Donaldson, Wang, Smith & Wolfe 2003).

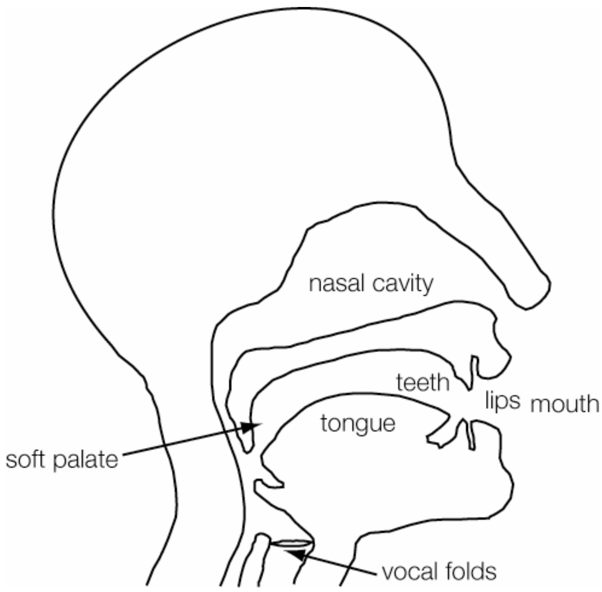

Figure 1. Simple schematic of the vocal tract.

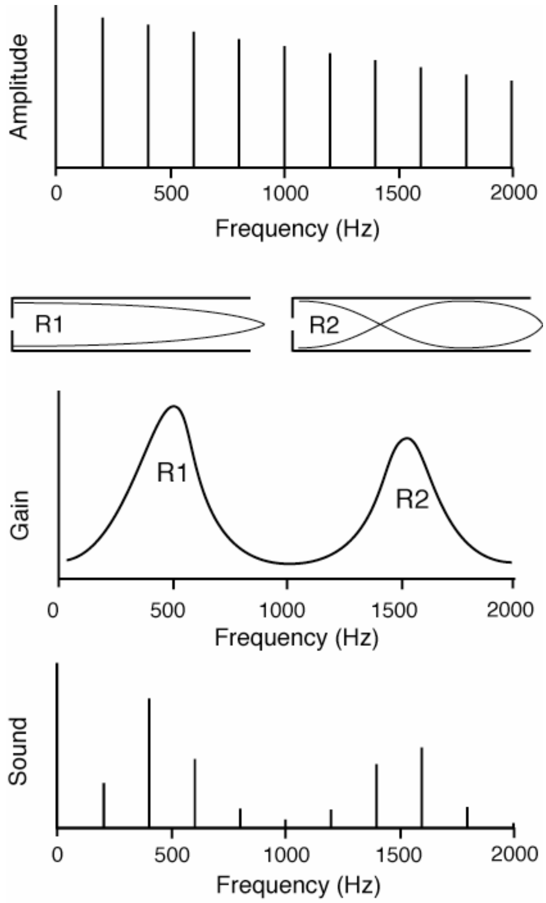

Figure 2. The source-filter model for voiced speech. The harmonic-rich signal from the vibrating vocal folds (top) is transmitted to the radiation field (bottom) via the tract. The tract most effectively matches the impedance near its resonances. The second sketch represents the tract as a uniform cylinder and shows the pressure amplitudes. In practice, the resonance frequencies are modified by moving the jaw, tongue and lips. (Donaldson *et al*. 2003).

MEASUREMENT OF VOCAL TRACT RESONANCES

The acoustic current source

Transfer functions are measured using a technique described in detail elsewhere (Dowd, Smith and Wolfe 1997; Epps *et al.*, 1997; Smith, Henrich and Wolfe*,* 1997; Wolfe *et al.*, 2001). A computer synthesises the broad band signal *s*(*f*) over the frequency range of interest from a set of harmonic components that have been chosen for the best compromise between frequency resolution and signal to noise ratio. The relative phases of these components are adjusted to improve the signal to noise ratio (Smith 1995). The electrical broad band signal produced via an analogue/digital interface (National Instruments NB-A2100) is amplified and used to drive an enclosed loudspeaker, which is matched by an exponential horn to an acoustic attenuator with a high value acoustic impedance – see Figure. 3. This serves as the output impedance Z_0 of the acoustic current source. The impedance of this attenuator is a compromise; if it is too large, the acoustic current *u*(*f*) will not produce an adequate pressure signal when applied to 'loads' with a small acoustic impedance, whereas if it is too small there is the possibility that the acoustic current might vary slightly when applied to different 'loads'. A small microphone (6.5 to 8 mm external diameter) is used to record the pressure *p*(*f*).

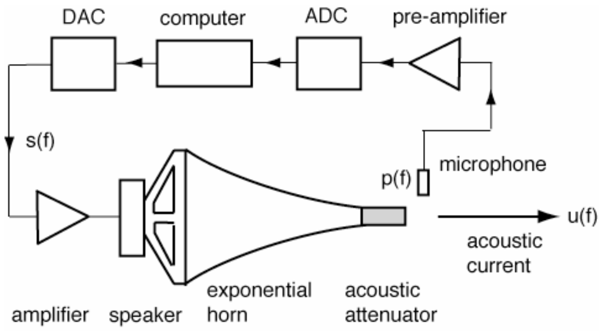

Figure 3. The acoustic current source.

Measurement of acoustic resonances

All the measurements presented herein involve the ratios of two pressure measurements in which the same acoustic current source is connected to two different acoustic 'loads'. Because the characteristics of the amplifier, loudspeaker, horn, and acoustic attenuator each depend upon frequency, the spectrum of the acoustic current *u*(*f*) can be quite different from the spectrum of the electrical signal *s(f)* produced by the computer. In the measurements reported here, the spectrum of the pressure $p_{\text{cal}}(f)$ as measured by the microphone in the initial 'calibration' procedure is adjusted to be independent of frequency – this helps equalise the signal to noise ratio across the measured spectrum. This is accomplished by adjusting the relative amplitude of the harmonic components in the synthesized electrical signal *s(f*). This procedure does not remove any frequency dependence in the response of the microphone itself, but any such dependence will cancel when the ratio $\gamma = p_{\text{meas}}/p_{\text{cal}}$ of the pressure measurements is considered.

The 'calibration' used for measurements of the vocal tract in speech and singing involves a measurement made at the same position, with the mouth closed – see Figure 4. This choice minimises the geometrical effects of the face: the ratio shows the effect of putting the open vocal tract in parallel with the radiation impedance, measured that the lips, with the face as a baffle.

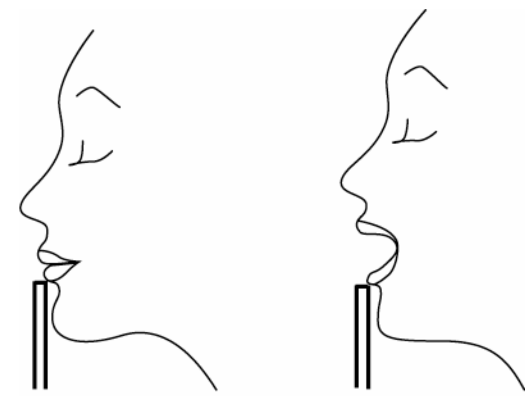

Figure 4. The configuration used to measure the response of the vocal tract showing how the current source and microphone, located on a flexible mount, are placed at the singer's lower lip. (Joliveau *et al*. 2004b)

Measurement of acoustic impedance

Measurement of acoustic impedance, rather than just the frequencies of resonance, requires calibration using a known reference impedance. The reference impedances used are long cylindrical pipes that are straight for at least the first 40 m. These pipes are effectively infinite for the frequencies in this study, and their impedances are purely resistive, and determined from the cross section.

VOCAL TRACT RESONANCES IN SPEECH

Figure 5 shows an example of the pressure measured just outside the mouth when the broad band acoustic current was injected just outside the mouth. The frequencies of the first two resonances largely determine the vowel and a 'map' of 11 English vowels, as spoken by young Australian women, is shown in Figure 6. Note how, in Figure 5, the fifth harmonic especially is enhanced compared to the steady decrease with frequency that we should expect from the source at the glottis. In this measurement, the higher harmonics are hidden by our measurement signal, but harmonics near the higher resonances are similarly boosted (data not shown). These higher formants are important in the quality of the voice, in some styles of singing, and in helping us recognise our friends' voices.

Our technique can measure the resonances with higher precision than is possible using the voice signal alone. This is because the injected acoustic current has a frequency resolution of 5 Hz, whereas the voice has harmonics spaced at f_0 (at least 80 Hz). Especially for high voices, the speech or singing signal can only give imprecise information about the frequencies of the formants or resonances. Measuring the resonance of the tract, rather than the formants that result from it, therefore has advantages that may be practical, as well as scientific.

In one study, we have used this technique to provide real time, visual feedback about the configuration of the vocal tract to subjects who were learning to pronounce foreign vowels.

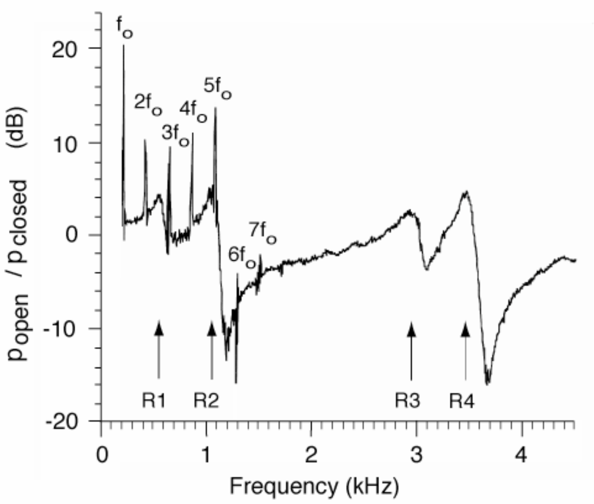

Figure 5. The ratio of the spectra measured with the mouth open to that with the mouth closed (p_{open}/ p_{closed}) for the vowel /ɒ/ (in "hot"). Several harmonics of the voice signal with fundamental frequency f_O = 215 Hz can be seen. The maxima in the broad band signal corresponding to the resonances *R1, R2, R3* and *R4* are indicated by arrows. (Donaldson *et al*. 2003).

This visual feedback has the advantage that is not subject to the perceptual 'categorisation' (the psychologists' term for what physicists would call quantisation) of the auditory signal (Dowd, Smith & Wolfe 1997). These subjects learned to imitate target vowels much more accurately than those who used auditory feedback alone. In the future, we hope to supply feedback to singers learning different musical styles.

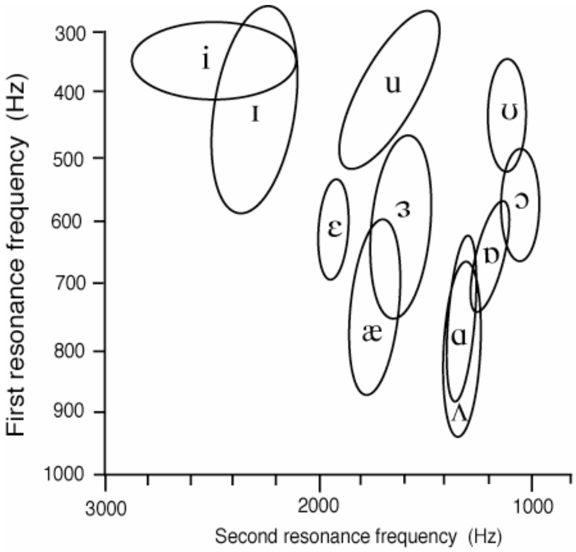

Figure 6. The distribution of (*R2,R1*) for the vowels of English as spoken by this sample of young Australian women. The centre of each ellipse is the mean of (*R2,R1*). The slope of the major axis indicates the regression of *R1* on *R2*, and the semiaxes are the standard deviations in those directions. The words used (with phonetic vowel symbols in brackets) were 'heed' [i], 'hid' [I], 'head' [ɛ], 'had' [æ], 'hard' [ɑ], 'hot' [ɒ], 'hoard' [ɔ], 'hood' [ʊ], 'who'd' [u], 'hut' [ʌ] and 'heard' [ɜ]. The substantial overlap between 'hard' [ɑ] and 'hut' [ʌ], and between 'heed' [i] and 'hid' [I] may seem surprising until one realises that these pairs are usually distinguished in normal speech by duration. (Donaldson *et al*. 2003).

VOCAL TRACT RESONANCES IN SINGING

In singing, the vibration frequency f_0 of the vocal folds is now determined by the desired pitch, so a new set of problems and possibilities emerge.

If the pitch frequency is not too different from that of normal speech, the singing voice will continue to gain support from the impedance transforming effect of the normal resonances *R1* and *R2*. However, once the pitch becomes sufficiently high, the resonances need not coincide with a harmonic of the pitch. Consequently, when the harmonics (nf_0, n an integer) are widely spaced, fixed resonance frequencies and varying f_0 lead to some pitch-vowel combinations being weaker than others, depending on whether the nf_0 fall near to or far from a resonance. If few of the nf_0 are aided by the impedance transforming effect of a strong resonance, the voice is weak and may have trouble competing with the accompaniment, especially in opera.

The obvious response would be for singers to shift their resonances from their usual values in speech and support the harmonics of f_0. However once the resonances in singing are shifted significantly, there will also be a consequential decrease in intelligibility. The exceptions will be for vocalise, where the voice is used without text, and the Australian National Anthem, where no-one seems to know the words anyway.

The musical range for sopranos ranges from C4 to C6 (approx. 250 to 1000 Hz), although some coloratura parts can go as high as G6 (~1500 Hz). The ear is more sensitive in the higher part of this frequency range so a soprano could significantly increase her sound output and timbral homogeneity by adjusting *R1* from its normal value to match f_0 (Sundberg (1977,1987).

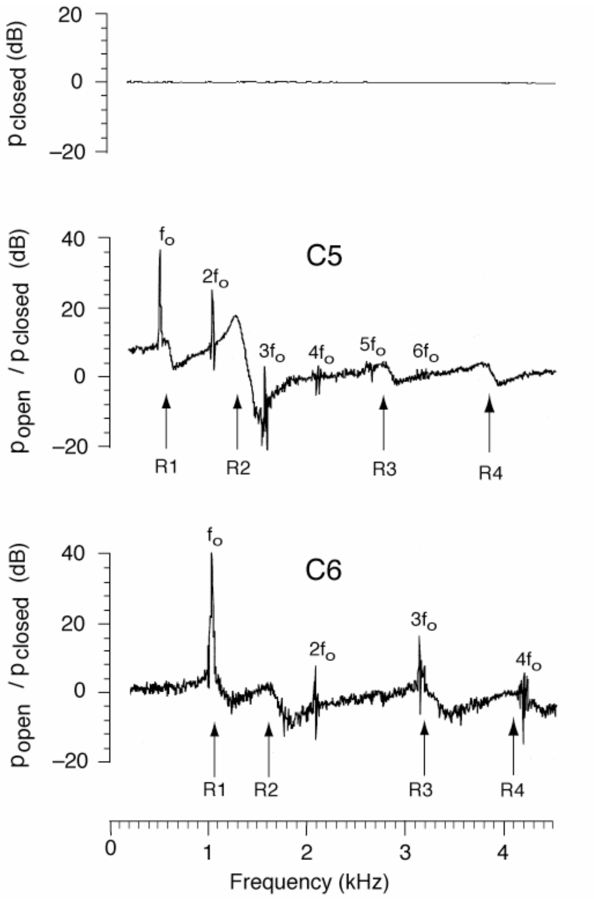

Figure 7. The upper figure presents the pressure spectrum measured for a soprano without singing and with the mouth closed (p_{closed}). The spectrometer had been previously calibrated in this configuration by adjusting the acoustic current so that the pressure spectrum measured with the mouth closed was independent of frequency with nominal value of 0 dB. The lower two figures present the ratio of the pressure spectrum measured with the mouth open to that measured with the mouth closed (p_{open}/ p_{closed}) when the subject sang the vowel /u/ (in *who'd*) on the notes C5 and C6. The harmonics of the (periodic) voice signal are indicated. The peaks in the broad band signal indicated by arrows correspond to the resonances *R1*, *R2*, *R3* and *R4*. Note that the spacing between harmonics makes it impossible to estimate the formant or resonance frequencies precisely from the voice signal alone. Note also that this singer has raised *R1* to 'tune' it to the note C6. Such results are summarised in Figure 8. (Joliveau *et al.* 2004b).

Our studies show that this is indeed the case (Joliveau, Smith and Wolfe 2004a,b) – see Figure 8. Furthermore the vowels are shown to converge so that their non-dimensional separation in (*R1*,R*2*) space and their overlap are comparable with the characteristic separation required to distinguish vowels (Dowd *et al.* 1997). The fundamental pitch of singers other than sopranos generally lies below the normal values of *R1*. However, tuning their resonances close to harmonics of the pitch could still be advantageous. Figure 9 illustrates how a baritone can tune *R2* to match $2f_0$.

At high sound levels, the vibratory behaviour of the vocal folds could also be influenced by the acoustic load presented by the vocal tract (e.g. Titze 1988). Singers thus might be able to avoid damage to their vocal folds by appropriate resonance tuning.

Some singers (usually tenors and altos) produce a singer's formant or vocal 'ring': a band of increased radiated power in the frequency range 3-4 kHz, a range where human ears are sensitive but where orchestral output power is reduced (Sundberg 1977, 2001). Figure 9 illustrates how the higher resonances (*R4*-R*6*) can be shifted to help increase output at high frequencies.

A spectacular instance of resonance modification occurs in harmonic singing. This involves making one resonance very strong and tuning it precisely to a high harmonic of f_0 – see Figure 10. All the other resonances are weakened. Thus one harmonic is made so much stronger than its neighbours that we can hear it as a separate note (Wolfe, Henrich and Smith 2004).

In singing, the control oscillator is the vocal folds and the strong resonances in the vocal tract downstream from there produce the strong difference in timbre that we recognise as different phonemes. We now proceed to a case in which similarly great changes in timbre are produced by analogous resonances in the tract, but in which the control oscillator is at the other end.

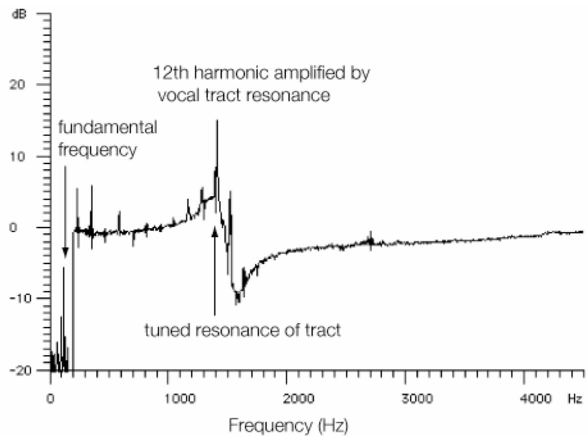

Figure 10. An example of the measured pressure spectrum during harmonic singing.

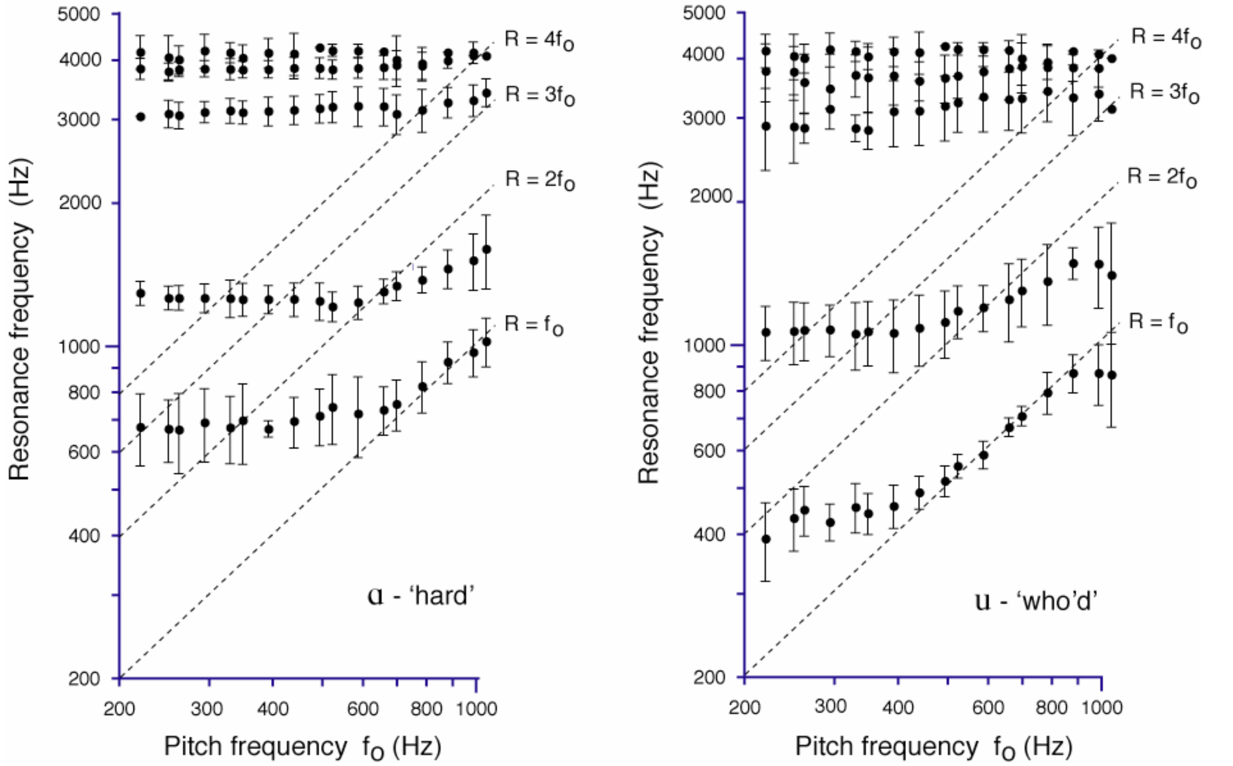

Figure 8. The first five average resonance frequencies of the vocal tract as a function of the fundamental frequency f_0 for the sustained vowels in two words /hard/ and /who'd/ sung softly by nine sopranos. The vertical bars indicate the standard errors. The dashed lines R = n f_0 indicate the exact tuning of a resonance to a harmonic of the pitch frequency f_0. The tuning of R1 to the fundamental can be clearly seen once f_0 exceeds the value of R1 at low pitch. The loss of resonance tuning for 'who'd' at the highest pitches is probably due to the physical impossibility of further tuning with the rounded lips used for these vowels.

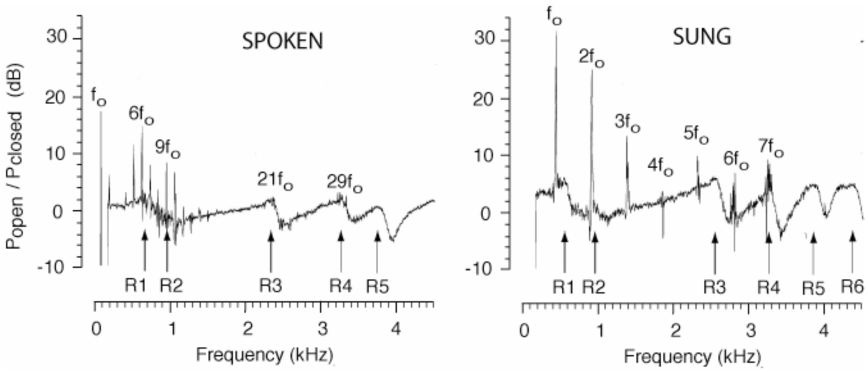

Figure 9. Measurements on the same subject, a tenor, speaking the vowel in the word 'hoard' (left), at approximately 100 Hz, and singing the same vowel in full voice (right), at G4 (approx. 400 Hz). R1 and R2 increased slightly when singing. However R3, R4 and R5 are shifted closer together and have become much stronger in the singing configuration. Consequently the 7th and 8th harmonics are considerably stronger than would be expected. The strong harmonics in this frequency range are consistent with reports of a singer's formant, and the indicated resonances seem to suggest a mechanism.

THE YIDAKI AND THE VOCAL TRACT

The didjeridu is an onomatopæic Western name for an instrument known as the yidaki in the Yolngu language of Northern Australia, whence it comes. It is traditionally made from the trunk of a small tree, hollowed out by termites. It is unusual in that it usually only plays one note, with occasional overtones in some styles. Its musical interest comes however from the spectacular changes in timbre, which far exceed those in orchestral wind instruments.

In a standard model (Backus, 1985) for lip valve instruments (trombone, tuba etc), the acoustic impedances of the bore and of the performer's vocal tract act in series on the lips and on the air flow through them. (The pressures on either side of the lips act in opposite directions, but the acoustic flow has opposite sign on each side, too, so the sum of the impedances appears in the equations.)

In orchestral lip valve instruments the narrow bore and a tight constriction at the mouthpiece together produce resonances whose acoustic impedance has values considerably exceeding those of the vocal tract, even at resonance. Consequently, the tract impedance has only a modest effect on the series combination. In the yidaki, by contrast, the effect is of primary importance.

For example, Figures 11 and 12 show the difference between sounds produced with the tongue close to the hard palate and with the tongue low in the mouth. On the same graphs are shown the acoustic impedance of the player's vocal tract, measured just inside the lips, during performance (Tarnopolsky *et al*., 2005).

The mechanism whereby this occurs is perhaps counter-intuitive. At resonance, the vocal tract presents a high impedance at the lips: in other words, a pressure antinode and flow node. If the impedance is sufficiently high (comparable with or greater than that of the instrument), then acoustic flow into the instrument is inhibited at that frequency. As a result, we see in Fig 11 the minima in the spectral envelope of the sound produced by resonances at 1.5 and 2.0 kHz, and the consequent formation of a clear formant between these two frequencies. The peak frequency depends strongly on the mouth geometry and can be varied by the player. In contrast, the low tongue configuration shown in Figure 12 has no comparably strong resonances in this range and so produces no comparable formant.

In orchestral lip valve instruments, the effect of the vocal tract geometry on register, pitch and timbre is not so spectacular, but it can still be musically important (Wolfe *et al.*, 2003). Current and future research will, we hope, further our understanding of its role in music perforamance and pedagogy.

CONCLUSIONS

In speech acoustics, the importance of the speech signal as both an object of study and as a non-perturbing probe of the vocal tract is obvious. Nevertheless, the capacity to measure the frequency response of the vocal tract, while in use, with a resolution of ~ 10 Hz, has improved our understanding of some techniques in singing and music performance. It has also produced techniques with important potential for practical application in speech training and therapy and may do likewise for singing and instrumental music in the future.

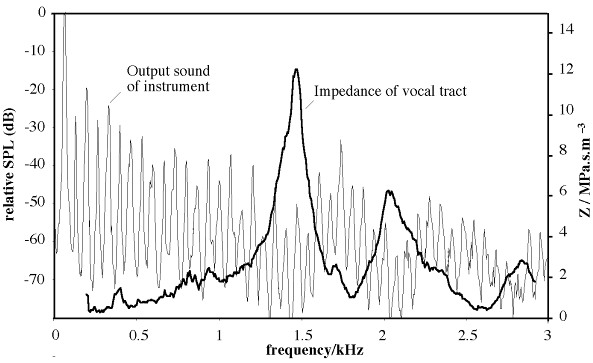

Figure 11. Spectrum of radiated sound and magnitude of vocal tract impedance measured just inside the player's lips during performance. The player performs the 'high drone', with the tongue close to the hard palate, that produces a characteristic, strong formant at 1.8 kHz. The strong peak in the impedance associated with a minimum in the spectrum of output sound is clearly evident. (Tarnopolsky *et al.* 2005).

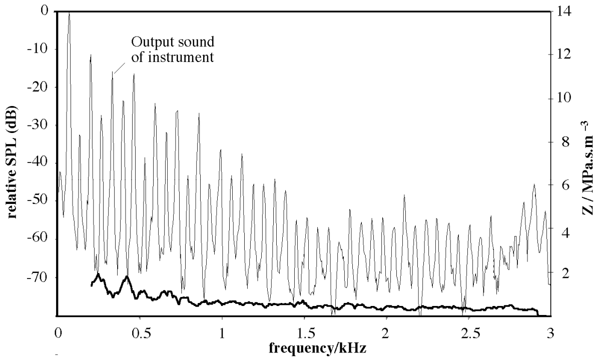

Figure 12. Spectrum of radiated sound and magnitude of vocal tract impedance measured just inside the player's lips during performance. The player performs with the tongue in the low position, a configuration that does not produce strong formants. (Tarnopolsky *et al.* 2005).

ACKNOWLEDGMENTS

The described research has involved several members and associates of our Acoustics Laboratory including Tina Donaldson, Annette Dowd, Julien Epps, Neville Fletcher, Nathalie Henrich, Lloyd Hollenberg, Elodie Joliveau, Benjamin Lange, Alex Tarnopolsky, and Diana Wang. We thank our volunteer subjects, UNSW and the Australian Research Council for support.

REFERENCES

Backus J. (1985), 'The effect of the player's vocal tract on woodwind instrument tone'. *J. Acoust. Soc. Am.* 78, 17-20.Berlioz, H (1882) *Traité de l'instrumentation et d'orchestration modernes*, 1844, Trans. M C Clarke. Novello, London.

Donaldson, T. Wang, D. Smith, J. & Wolfe, J. (2003) 'Vocal tract resonances: a preliminary study of sex differences for young Australians', *Acoustics Australia*, 31, 95-98.

Dowd, A, Smith, J R & Wolfe, J (1997) '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.

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

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

Joliveau, E, Smith, J & Wolfe, J (2004a), 'Tuning of vocal tract resonances by sopranos', *Nature* 427, 116.

Joliveau, E, Smith, J & Wolfe, J (2004b), 'Vocal tract resonances in singing: The soprano voice', *J.Acoust.Soc.Am.* 116, 2434–2439.

Smith, J. (1995), 'Phasing of harmonic components to optimize the measured signal-to-noise ratios of transfer functions' *Measurement Sci. and Tech*, 6, 1434-1348.

Smith, J R, Henrich, N, & Wolfe, J. (1977), 'The acoustic impedance of the Boehm flute: standard and some non-standard fingerings,' *Proc. Inst. Acoust.* 19, 315-320.

Sundberg, J. (1974), 'Articulatory interpretation of the "singing formant"', *J.Acoust.Soc.Am.*, 55, 838-844.

Sundberg, J (1977), 'The acoustics of the singing voice' *Scientific Am.*, March, 82-91.

Sundberg, J (1987), *The Science of the Singing Voice*, Northern Illinois Univ. Press, De Kalb, Ill.

Sundberg, J. (2001), 'Level and centre frequency of the singer's formant', *Journal of Voice* 15, 176-186.

Tarnopolsky, A, Fletcher, N, Hollenberg, L, Lange, B, Smith, J, & Wolfe, J. (2005), 'The vocal tract and the sound of a didgeridoo', *Nature*, 436, 39.

Titze, I. R. (1988), 'The physics of small-amplitude oscillations of the vocal folds', *J.Acoust.Soc.Am.* 83, 1536-1552.

Wolfe, J, Smith, J, Tann, J. & Fletcher, N. (2001). 'Acoustic impedance of classical and modern flutes' *J. Sound and Vibration*, 243, 127-144.

Wolfe, J, Smith, J, and Henrich, N. (2004), 'Harmonic singing *vs* normal singing' http://www.phys.unsw.edu.au/~jw/xoomi.html

Wolfe, J, Tarnopolsky, A Z, Fletcher, N H, Hollenberg, L C L & Smith, J. (2003), 'Some effects of the player's vocal tract and tongue on wind instrument sound'. *Proc. Stockholm Music Acoustics Conference (SMAC 03)*, (R. Bresin, ed) Stockholm, Sweden. 307-310.