The effect of whisper and creak vocal mechanisms on vocal tract resonances

Yoni Swerdlin, John Smith,[a] and Joe Wolfe
School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia

(Received 12 May 2009; revised 20 January 2010; accepted 21 January 2010)

The frequencies of vocal tract resonances estimated using whisper and creak phonations are compared with those in normal phonation for subjects who produced pairs of these phonations in the same vocal gesture. Peaks in the spectral envelope were used to measure the frequencies of the first four resonances ($R1-R4$) for the non-periodic phonations, and broadband excitation at the mouth was used to measure them with similar precision in normal phonation. For resonances $R1-R4$, whispering raises the average resonant frequencies by 255 Hz with standard deviation 90 Hz, 115 ± 105, 125 ± 125, and 75 ± 120 Hz, respectively. A simple one dimensional model of the vocal tract is presented and used to show how an effective glottal area can be estimated from shifts in resonance frequency measured during the same vocal gesture. Calculations suggest that the effective glottal area thus defined increases to 40 ± 30 mm^2 during whispering. Creak phonation raised significantly only the first and third resonant frequencies, by 45 ± 50 and 65 ± 120 Hz respectively. It thus appears possible to use creak phonation to determine resonances with more precision than is available from the spectral envelope of voiced speech, and this supports its use in teaching resonance tuning to singers. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3316288]

PACS number(s): 43.70.Gr, 43.70.Aj, 43.70.Bk [CHS] Pages: 2590–2598

I. INTRODUCTION

The acoustic resonances (Ri) of the human vocal tract are of interest for several reasons. When excited by various mechanisms, these resonances give rise to peaks in the spectral envelope of the output sound (e.g., Fant, 1970). In speech, the peaks in the spectral envelope with the two lowest frequencies usually identify the vowels in Western languages and contribute to regional accents. Further, their variation in time is important to the identification of many consonants. The peaks in the spectral envelope that occur at higher frequencies are important in determining the timbre and identity of the voice (e.g., Fant, 1973; Clark *et al.*, 2007).

The resonances are also important in music, for reasons not directly related to phonetics (Wolfe *et al.*, 2009). Following suggestions by Sundberg (Lindblom and Sundberg, 1971; Sundberg, 1977), it has been demonstrated that some singers "tune" the lowest resonance to a frequency near the fundamental (f_0) of the note sung (Joliveau *et al.*, 2004a, 2004b), thereby obtaining extra output power for a given vocal effort. Other singers have been shown to tune the first resonance to the second harmonic (Henrich *et al.*, 2007). Furthermore, it is proposed that these resonances can also influence the vibratory behavior of the vocal folds (Titze 1988, 2004, 2008). Indeed it is possible that composers have aided the acoustics of the soprano voice at high pitch when setting text to music by appropriately matching sung pitch to resonance frequency (Smith and Wolfe, 2009). Vocal tract resonances also play an important role in determining the timbre or pitch of wind instruments, e.g., the didjeridu (Tarnopolsky *et al.*, 2005, 2006), the saxophone (Chen *et al.*, 2008), and the clarinet (Chen *et al.*, 2009). Indeed, experienced musicians have been observed to play with the relatively small glottis (Mukai, 1992) that would enhance vocal tract resonances.

The frequencies of the resonances may be estimated in a number of ways. The spectral maxima in the output sound will occur at frequencies close to those of the tract resonances that produce them, so one method involves estimating the resonances from the sound spectrum of speech or song. In normal phonation, however, the tract is predominantly excited by periodic vibration of the vocal folds. Consequently, the frequency domain is sampled at multiples of the fundamental frequency f_0, so it is difficult to determine unambiguously the frequencies of the resonances with a resolution much finer than $f_0/2$. f_0 is typically 100–300 Hz in conversational speech, but may be considerably higher in singing where the resolution is correspondingly much worse (Monsen and Engebretson, 1983). The estimation of resonance frequencies from spectral peaks in normal phonation is further complicated by the frequency dependence of the source function at the glottis, which is in general unknown.

One possible method of improving the frequency resolution involves vibrating the neck near the glottis using a broadband mechanical source (Coffin, 1974; Sundberg, 1977; Pham Thi Ngoc and Badin, 1994). This has the advantage that it stimulates the tract from an area near the glottis, but its disadvantages are that the transfer functions between the mechanical signal and the acoustical signal at the glottis are unknown, and that it involves perturbing the subjects.

A potentially more precise method of estimating the frequencies of resonances of the tract during normal phonation involves exciting it with a known, external, acoustic flow at the mouth (Epps *et al.*, 1997; Dowd *et al.*, 1998). A broadband source of acoustic flow and a microphone are posi-

---

[a] Author to whom correspondence should be addressed. Electronic mail: john.smith@unsw.edu.au

tioned at the subject's lower lip. During normal phonation, the microphone pressure signal is the sum of the widely spaced harmonics of the periodic voice signal and the pressure produced by the injected acoustic flow interacting with the vocal tract and the radiation field. However, it has the disadvantages that the tract is measured from the mouth rather than the glottis, and it is measured in parallel with the external radiation field.

Another method of estimating resonances involves exciting the tract by a non-periodic vocal mechanism, thereby producing a spectrum whose peaks may be determined with greater precision than is possible for normal speech. Whispered speech is produced by turbulent flow through a relatively small, nearly constant aperture formed between the vocal folds. In creak phonation, also called the creak voice, vocal fry or mechanism 0, the vocal folds open in an aperiodic way (Hollien and Michel, 1968; Gobl, 1989). Researchers in acoustic phonetics have used whisper or creak phonation to obtain information about the resonances, with potential relevance to normal speech. Another practical use of creak phonation concerns the use of resonance tuning in singing: Singers may use spectral analysis of their creak phonation to learn to tune a resonance of the vocal tract (Miller *et al.*, 1997).

The whisper and creak methods have a possible limitation in that the different phonation types involve changes in the geometry of the tract around the glottis. Further, even if the geometry of the entire tract (glottis excepted) were fixed, the frequencies of the resonances should vary due to different average areas of the glottis. It is thus possible that the resonance frequencies are different for the different modes of phonation. Consequently, measurements of Ri made during whisper or creak phonation might not be exactly comparable with normal phonation.

Indeed, researchers have found that, on average, the resonances of whispered speech usually occur at significantly higher frequencies than those of normally phonated speech (Kallail and Emanuel, 1984a, 1984b; Jovocic, 1998; Matsuda and Kasuya, 1999). In contrast, the resonances produced by creak and normal phonations have been found to be similar (e.g., Miller *et al.*, 1997), although Ladefoged *et al.* (1988) and Ananthapadmanabha (1984) reported slight increases in $R1$ during creak phonation and Moosmüller (2001) found that, for women, $R2$ is slightly lowered in the creaky voice.

The above measurements are subject to the limitation that, while the resonances associated with whispered speech and creak phonations can be determined precisely from the spectral peaks in the sound, this is not usually possible for normal phonation, as explained above. Further, the studies cited above all compare averages of the resonance frequencies measured in separate vocalization gestures.

In the present study, using ten young Australian women as subjects, the resolution of such studies is increased by introducing two experimental improvements. The first is to use acoustical excitation at the mouth to estimate the acoustical resonances of the vocal tract more precisely during normal phonation. The second is to compare them with estimates of the resonances using whisper or creak phonation in the same vocal gesture. Finally a simple mathematical model is developed to estimate the increase in effective glottal area during these phonation modes. Some of the experimental results given have been briefly presented earlier (Swerdlin *et al.*, 2008).

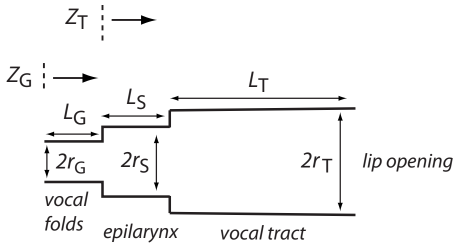

FIG. 1. A schematic (not to scale) indicating the simple 1D cylindrical model of the vocal tract. Arrows indicate the planes corresponding to the impedances Z_G and Z_T.

II. THEORY

A. Simple one dimensional model

In normal speech, the vocal tract is open at the lips and alternately closed and slightly open at the glottis as the vocal folds vibrate. In whispering, the glottis is permanently partly open. Barney *et al.* (2007) used a mechanical model in addition to an equivalent circuit model and showed that increased glottal opening raised the frequency of $R1$. One might explain this qualitatively as follows: A closed glottis produces a node in the acoustic flow. Provided that the subglottal tract has no strong resonances at the frequencies of interest, a slightly open glottis behaves approximately as an inertance in the frequency range of interest and so reflects a wave with phase changes in pressure and flow that are, respectively, slightly greater than 0 and slightly less than 180°. This displaces the node of acoustic flow toward the mouth, raising the resonant frequency. Because of the inertia of the air in the glottis, the effect decreases with increasing frequency: At sufficiently high frequency, the air in the glottis acts to "seal" the glottis and thus turns the slight opening into a termination that is effectively closed. Hence one expects that the increase in frequency will be greatest for the lowest resonance.

A very simple one dimensional (1D) model is shown in Fig. 1. To simplify the mathematical treatment, the vocal tract is modeled as a simple cylinder of effective length L_T and radius r_T. The radiation impedance at the lip opening is incorporated by including an end correction in L_T. The impedance Z_T looking from the junction of tract and glottal region is given by

$$Z_T = jZ_{T0}\tan(kL_T), \tag{1}$$

where $k=2\pi f/c$, f denotes the frequency, and c denotes the speed of sound. Wall losses will be neglected. The cross-sectional area S_T of the tract is given by $S_T=\pi r_T^2$. The characteristic impedance Z_{T0} of the tract is given by $Z_{T0}=\rho c/S_T$, where ρ is the density of air. The constricted region between the vocal folds is also modeled as a simple cylinder of effective length L_G and effective radius r_G. Again, end effects are incorporated in the effective length. The effective radius includes the open quotient and the influence of the subglottal region via the glottis. Initially the epilaryngeal region is ne-

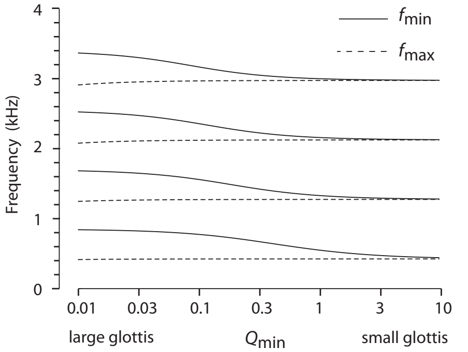

FIG. 2. A semi-logarithmic plot of the dependence of the frequencies $f_{\min}$ and $f_{\max}$ on the parameter $Q_{\min}$ $(=(r_T^2/r_G^2)(L_G/L_T))$. The frequencies $f_{\min}$ and $f_{\max}$ are those where minima and maxima, respectively, will occur in Z_G, the impedance looking out from the vocal folds toward the lip opening. Curves were calculated by solving Eq. (5) or Eq. (7), and assuming that $L_T = 200$ mm.

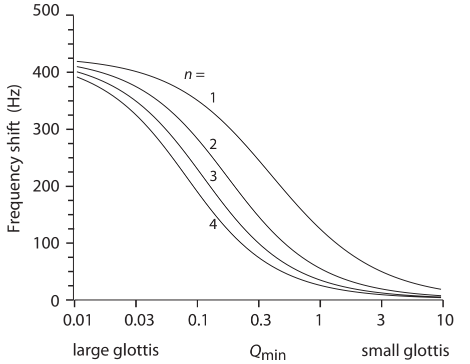

FIG. 3. A semi-logarithmic plot of the shift in resonance frequency from its value with a closed glottis as a function of the parameter $Q_{\min}$ $(=(r_T^2/r_G^2)\times(L_G/L_T))$. n indicates the order of the resonance Rn. Curves were calculated by solving Eq. (5) and assuming that $L_T = 200$ mm.

glected. The impedance Z_G seen from the glottis through the constricted vocal folds would then be given by

$$Z_G = Z_{G0}\frac{Z_T\cos(kL_G) + jZ_{G0}\sin(kL_G)}{Z_{G0}\cos(kL_G) + jZ_T\sin(kL_G)}, \tag{2}$$

where the characteristic impedance of the vocal fold constriction is given by $Z_{G0} = \rho c/S_G$ and the glottal cross-sectional area is given by $S_G = \pi r_G^2$. The frequency of the nth minimum is determined primarily by the Z_T terms and will occur when $kL_T \approx n\pi$. For the situation considered here $L_G \ll L_T$. Then $kL_G \ll 1$ for small n and consequently $\sin(kL_G) \approx kL_G$ and $\cos(kL_G) \approx 1$. Equation (2) then simplifies to

$$Z_G \approx jZ_{G0}\frac{Z_{T0}\tan(kL_T) + Z_{G0}kL_G}{Z_{G0} - Z_{T0}\tan(kL_T)kL_G}. \tag{3}$$

Z_G will exhibit minima when

$$Z_{T0}\tan(kL_T) = -Z_{G0}kL_G. \tag{4}$$

After the substitution $x = kL_T$, Eq. (4) can be written in the form

$$\tan x = -Q_{\min}x, \tag{5}$$

where

$$Q_{\min} = (r_T^2/r_G^2)(L_G/L_T). \tag{6}$$

Similarly Z_G will exhibit maxima when

$$\tan x = Q_{\max}/x, \tag{7}$$

where

$$Q_{\max} = (r_T^2/r_G^2)(L_T/L_G) = Q_{\min}(L_T^2/L_G^2). \tag{8}$$

These transcendental equations determine x, and thus the frequencies $f_{\min}$ and $f_{\max}$ at which the extrema occur in Z_G. The tan function is periodic, and Eqs. (5) and (7) will thus exhibit multiple solutions that correspond to the various resonances of the system—see Fig. 2. The minima in Z_G will correspond to maxima in the transfer function between the glottis and the mouth, and will consequently be associated with peaks in the spectral envelope of the output sound.

A new value of $Q_{\min}$ can thus be calculated from a change in the resonance frequency. $Q_{\min}$ depends on the relative areas and lengths of the glottis and vocal tract [see Eq. (6)]. Small values of $Q_{\min}$ correspond to a glottis whose area is a larger fraction of the vocal tract area and/or whose effective length is a smaller fraction of the vocal tract length. For very small values of $Q_{\min}$, corresponding to a cylinder that was ideally open at the glottis, the maxima and minima would be evenly distributed with frequency at the expected harmonic frequencies. If $Q_{\min}$ increases due to a decrease in glottal effective area, the frequency $f_{\min}$ decreases and eventually becomes very similar to $f_{\max}$ for large values of $Q_{\min}$ (Fig. 2).

An increase in the effective glottal area from its low value in normal speech will thus cause an increase in the resonance frequencies. The value of $Q_{\min}$ at which a given shift in $f_{\min}$ occurs moves to lower values of $Q_{\min}$ as the order of the resonance increases—see Fig. 3. This figure also shows that the frequency shift due to a decrease in $Q_{\min}$ associated with a small glottis is predicted to become smaller as the order of the resonance increases. Similarly a decrease in the effective glottal length will cause an increase in the resonance frequencies.

The effect of changes in the geometry of the epilaryngeal region can now be included in the mathematical treatment using the same approach and approximations as that used above. $Q_{\min}$ and $Q_{\max}$ can then be replaced with $Q'_{\min}$ and $Q'_{\max}$ and are given by

$$Q'_{\min}\frac{r_T^2}{L_T}\left(\frac{L_G}{r_G^2} + \frac{L_S}{r_S^2}\right), \tag{9}$$

$$Q'_{\max} = r_T^2\left(\frac{L_S}{L_T}\frac{1}{r_S^2} + \frac{L_T}{L_G}\frac{1}{r_G^2}\right), \tag{10}$$

where the epilaryngeal region has an effective length L_S and radius r_S. In general the second term in Eq. (9) will be less important. However, Eq. (9) does predict that a decrease in r_S (while other parameters remain constant) will increase $Q'_{\min}$ and thus decrease the resonance frequencies.

III. MATERIALS AND METHODS

A. The subjects

Ten Australian women, aged between 20 and 30 years, volunteered to participate. All were native speakers of Australian English, were judged to have similar Australian accents, and none reported or showed evidence of speech problems or abnormalities. Nine had lived in Australia for all their lives, and the other for half her life. Each subject was given a brief explanation of the University's ethics policy, signed a consent form, and was then given a lesson (typically 3 min) on how to produce creak phonation. The instruction "Hum your lowest note and then go lower" began the instruction, and the experimenter gave demonstrations and feedback. One subject was not able to produce the creak voice reliably and consequently only her results for the whisper were recorded.

Women were chosen as subjects because their higher fundamental frequency generally improves the precision of resonance estimates using external broadband excitation. This is because it is then easier to separate the speech signal from the response to the broadband signal. This is the opposite result to methods that use the speech signal alone, where the precision decreases with increasing fundamental frequency.

B. Resonance frequencies in different phonation modes

The technique reported by Epps *et al.* (1997) and Dowd *et al.* (1998) was used to estimate the vocal tract resonances using broadband external excitation. The excitation signal was synthesized from harmonics of a signal with a frequency of 5.383 Hz (i.e., 44 100 Hz/2^{13}). The harmonics that fell between 200 Hz and 4.5 kHz were summed, with relative phases chosen to improve the signal to noise ratio (Smith, 1995). This signal was amplified and delivered to an enclosed loudspeaker (150 mm diameter), which was attached to an exponential horn of 600 mm length and coupled to a flexible tube (300 mm length with inner radius 6 mm), and which contained acoustic fiber to reduce resonances—see Fig. 4. This source of acoustic flow was placed at the subject's lower lip. Next to the source, a small electret microphone (Optimus 33-3013) recorded both the sound of the voice, and the sound of the acoustic source interacting with the subject's vocal tract and the radiation field.

In an initial calibration stage, a measurement is made with the subject's mouth closed, i.e., when the measurement device is effectively loaded only by the impedance of the radiation field Z_{rad} at the lips and baffled by the subject's face. The relative amplitude of harmonics in the synthesized signal is then adjusted so that the measured pressure signal at the lips is independent of frequency. Measurements are then made during vocalization with the mouth open. The impedance measured is then $Z_{\parallel}$, the impedance of the vocal tract Z_{tract}, in parallel with Z_{rad}. The variable γ, the ratio of the pressure measured during vocalization to that measured with the mouth closed, in response to the same acoustic flow, is then calculated. Because the output impedance of the acoustic flow source is large, this ratio equals the ratio of the impedances in the two cases; i.e.,

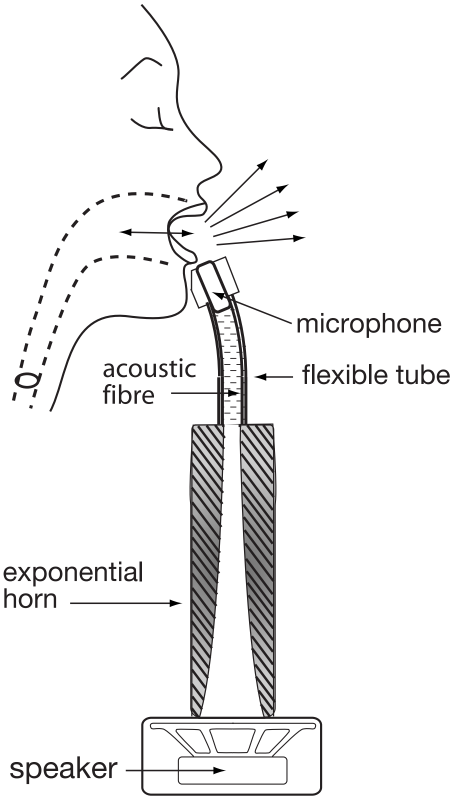

FIG. 4. Schematic (not to scale) showing how an external broadband signal is used to estimate the resonance frequencies of the vocal tract. The vocal tract is measured in parallel with the external radiation field. The microphone (8 mm diameter) is located immediately adjacent to the source of acoustic flow. It thus measures not only the sound produced by the subject, but also the response to the broadband signal interacting with the vocal tract. Initial calibration measurements are made with the subject's mouth closed.

$$\gamma = Z_{\parallel}/Z_{\text{rad}} = Z_{\text{tract}}/(Z_{\text{tract}} + Z_{\text{rad}}). \tag{11}$$

At resonance, the imaginary components of Z_{tract} and Z_{rad} are equal and opposite, so the denominator is very small and maxima in γ identify resonances. The relationship between maxima in the transfer functions from glottis to external radiation field and maxima in the impedance measured just outside the lips is complicated, but they generally agree for our experimental conditions (Smith *et al.*, 2007). Experiments with simple physical models of the vocal tract suggest that a resolution around ±20 Hz is possible.

For whisper and creak phonations, the power spectra were calculated using a window of 8192 points and a sampling rate of 44.1 kHz, and edited and displayed using the program AUDACITY (http://audacity.sourceforge.net). Resonance frequencies were estimated visually from the maxima in the spectral envelope. Examples are shown in Fig. 5.

C. The experimental sessions

The sessions were conducted in a "quiet room" inside the acoustics department. It was designed specifically for acoustic experiments. The walls and ceiling are treated to

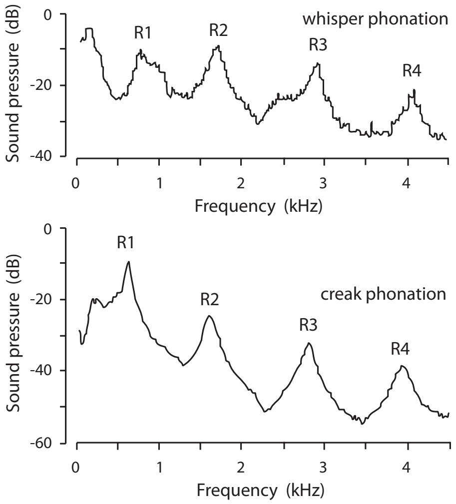

FIG. 5. Examples of measurements showing the spectra measured for the vowel in heard using whisper phonation (top) and creak phonation (bottom).

reduce external sounds by around 30 dB and surfaces treated to reduce reverberation. Background noise was always below 35 dBA.

Subjects were asked to produce one of five vowels, being those in the English words "head" (ɛ), "hard" (ɑ), "who'd" (u), "hoard" (ɔ), and "heard" (ɜ). The desired vowel was indicated to the subject by showing one of these words on a card.

The estimated values of resonance frequencies for a particular vowel are not important to the primary aim, which is to determine, for a given vowel gesture, the differences among the frequencies of the resonances during normal phonation, creak phonation, and whisper phonation. The context of the vowel was completely artificial: Subjects produced a particular vocal tract articulation and held it constant for several seconds. This would be a limitation in a study of accent, but here it is not a disadvantage. Rather, it allows the subject to concentrate on using the same articulation for each mechanism.

Each example of each vowel was produced in the order normal-whisper-normal-whisper or normal-creak-normal-creak. Subjects were asked to take a deep breath and, in a single gesture, to produce about 2.5 s of each of the four phonations without changing the position of tongue and mouth—see Fig. 6. During the second normal phonation, the vocal tract resonances were measured by broadband excitation. The whole gesture was digitally recorded and a 2 s sample of each of the whisper or creak segments was subsequently analyzed.

All subjects were able to perform this procedure comfortably. None reported being perturbed by the broadband signal, which had a sound level of about 70 dBA at the subject's ears, or by having the flexible tube touch their lower lip.

Once the resonances of each of the five vowels had been measured using both whisper and creak phonations, the sequence of measurements was repeated twice, giving a total of 30 vocal gestures for each subject. Our method involving external broadband excitation means that the impedance of the tract is measured in parallel with the external radiation field and consequently a weak tract resonance may not always be capable of resolution at low frequencies. It was also occasionally difficult to identify a particular resonance from the recorded sound in whisper and creak phonations. Both problems reduced the number of samples available for analysis.

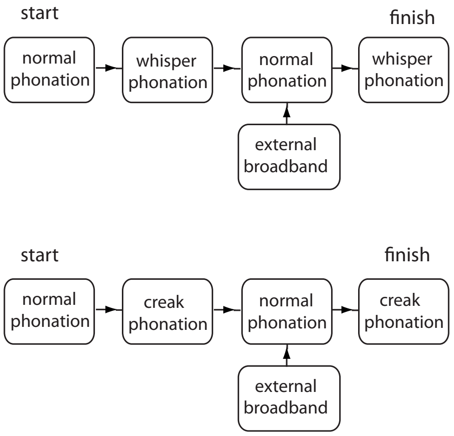

FIG. 6. Schematic showing the sequences used to compare whisper or creak phonation with normal phonation in a single vocal gesture.

IV. RESULTS AND DISCUSSION

A. Phonetic values

The average values of the first four resonances $R1-R4$ for the normal voice are given in Table I. They agree with those given by Donaldson *et al.* (2003) for young Australian women. Because the vowels studied were sustained, they are not necessarily the same as ordinary spoken vowels. How-

TABLE I. The measured resonant behavior of the tract for the ten subjects during normal, whisper, and creak phonations. Data in this and subsequent tables are presented as mean ± standard deviation (number of samples).

Vowel	$R1$ (Hz)	$R2$ (Hz)	$R3$ (Hz)	$R4$ (Hz)
		Normal phonation		
Head	605 ± 55 (53)	1860 ± 133 (52)	2800 ± 220 (55)	3960 ± 345 (52)
Hard	780 ± 95 (41)	1370 ± 45 (53)	2895 ± 130 (48)	3950 ± 150 (53)
Who'd	435 ± 70 (50)	1480 ± 355 (53)	2695 ± 95 (35)	3755 ± 200 (37)
Hoard	590 ± 60 (55)	1135 ± 80 (55)	2910 ± 100 (45)	3865 ± 215 (55)
Heard	625 ± 60 (52)	1555 ± 100 (54)	2810 ± 65 (48)	3910 ± 330 (53)
		Whisper phonation		
Head	875 ± 55 (50)	1960 ± 235 (50)	2915 ± 260 (52)	4015 ± 365 (51)
Hard	1010 ± 60 (55)	1520 ± 120 (48)	2990 ± 120 (54)	4000 ± 130 (50)
Who'd	755 ± 230 (19)	1655 ± 395 (49)	2840 ± 120 (45)	3905 ± 150 (42)
Hoard	885 ± 80 (41)	1200 ± 65 (51)	3035 ± 190 (54)	3930 ± 205 (56)
Heard	870 ± 60 (52)	1700 ± 105 (58)	2905 ± 180 (54)	3945 ± 355 (56)
		Creak phonation		
Head	665 ± 65 (52)	1890 ± 140 (51)	2870 ± 105 (51)	4040 ± 190 (43)
Hard	800 ± 60 (50)	1365 ± 70 (50)	2975 ± 113 (50)	3965 ± 185 (48)
Who'd	480 ± 55 (50)	1440 ± 325 (49)	2750 ± 145 (50)	3800 ± 250 (48)
Hoard	640 ± 70 (52)	1126 ± 80 (52)	3000 ± 145 (52)	3845 ± 220 (50)
Heard	665 ± 75 (51)	1550 ± 120 (51)	2850 ± 100 (52)	3925 ± 240 (51)

ever, this study is concerned with how the values of Ri depend on the phonation mechanism, rather than their absolute values.

The average values of the resonance frequencies $R1-R4$ for creak phonation given in Table I are similar to the average values for the normal voice. However, the average values of $R1-R4$ for whisper phonation were always higher than the average $R1-R4$ for the normal voice. The difference for the first resonance between whisper and normal phonations was large; when averaged across all vowels, the difference was 270 Hz and the frequency ratio of whisper to normal was 1.45. The effect was reduced for the second resonance; the difference being 125 Hz and the ratio 1.09. These values are similar to those found by Jovocic (1998) for Serbian vowels, with the exception of /u/ where Jovocic (1998) found that the resonance frequencies decreased significantly during whispering. Although the average values of $R3$ and $R4$ for whispering were always slightly higher than those for the normal voice, the differences are not often substantially larger than the experimental uncertainties.

These differences between the average values of the resonance frequency for whispered and normal phonations are either similar (Jovocic, 1998) or somewhat larger than some reported previously (Kallail and Emanuel 1984a, 1984b; Matsuda and Kasuya, 1999), and also show a similar decrease for higher resonances. Where comparison is possible, the absolute values for the increase in $R1-R3$ with whispering are consistent with an earlier study on female subjects (Kallail and Emanuel, 1984a), except that a considerably higher value for the shift in $R1$ in who'd using whisper phonation was found. However, the difference might be partly because Kallail and Emanuel rejected over 30% of their samples because of incorrect identification by a listening panel, whereas this project is primarily concerned with acoustical rather than perceptual aspects.

B. Stability of vocal tract configuration

Sensitive comparisons between the tract resonances in the different phonation modes can be made using data measured during the same vocal gesture. Consequently, it is important to confirm first that the tract remained effectively in the same configuration during each sequence. Table II shows the average differences between pairs of resonance frequencies estimated "before" and "after" the period of whispering in the same vocal gesture. Some differences were negative and some positive. A paired t-test was applied to these pairs of data to determine whether there was a statistically significant difference between the before and after measurements. Of the values in the table, two values ($R2$ for head and who'd) are significantly different from zero at the 5% level, which is a little more than one would expect in 20 tests. ($R2$ for these two vowels was also significantly different for creak phonation.) It is therefore possible that there is a slight non-random variation in the value of $R2$ (by tens of hertz or a few percent) between the initial and final whispers in each sequence. The very good reproducibility for whispering is perhaps because subjects would be experienced in occasionally making transitions between whispered and normal speech in various conversations.

TABLE II. The stability of vocal gestures. The table presents the average difference between the pairs of resonance frequencies (for either whisper or creak phonation) that were measured immediately before and after each normal phonation within each sustained vocal gesture. The symbol * indicates that the difference was significant at the 5% level or lower as indicated by a paired t-test.

Vowel	$\Delta R1$ (Hz)	$\Delta R2$ (Hz)	$\Delta R3$ (Hz)	$\Delta R4$ (Hz)
		Whisper phonation		
Head	−5 ± 40 (23)	60 ± 85 (23)*	20 ± 75 (24)	−20 ± 75 (25)
Hard	0 ± 25 (27)	−10 ± 80 (23)	−10 ± 80 (25)	−5 ± 75 (22)
Who'd	−5 ± 15 (8)	85 ± 235 (24)*	−15 ± 95 (22)	10 ± 125 (20)
Hoard	−5 ± 35 (19)	−15 ± 40 (25)	5 ± 100 (25)	−25 ± 85 (27)
Heard	−5 ± 45 (25)	20 ± 75 (29)	20 ± 70 (27)	5 ± 75 (27)
All vowels	0 ± 35 (102)	30 ± 125 (124)*	5 ± 85 (123)	−10 ± 85 (121)
		Creak phonation		
Head	−10 ± 20 (26)*	60 ± 55 (25)*	25 ± 60 (25)*	20 ± 70 (19)
Hard	0 ± 15 (25)	10 ± 30 (25)*	30 ± 70 (25)*	−10 ± 70 (23)
Who'd	−10 ± 35 (25)	55 ± 55 (24)*	−30 ± 70 (25)*	−20 ± 80 (24)
Hoard	−10 ± 20 (26)*	−10 ± 40 (26)	15 ± 75 (26)	0 ± 75 (25)
Heard	−5 ± 25 (25)	0 ± 60 (25)	−5 ± 70 (26)	−15 ± 60 (25)
All vowels	−5 ± 25 (127)*	25 ± 55 (125)*	5 ± 70 (127)	−5 ± 70 (116)

Table II also shows the changes between pairs of resonance frequencies for creak phonation made during the same vocal gesture. Here there are larger differences, again with both positive and negative signs, and an increased number are significantly different at the 5% level. This is perhaps a consequence of the subjects being less familiar with creak phonation than whisper phonation. The differences are still relatively small, of the order of 10 Hz for $R1$, which is around the limit of resolution of the resonance estimates.

Are the resonances for normal speech different in our sequences when immediately preceded by whisper or creak phonation in the same vocal gesture? This was tested for each subject and vowel by comparing the average values of Ri measured for the normal speech in each sequence involving whispering with those involving creak phonation—see Table III. The differences in Ri associated with an intervening segment of whispering vs creak phonation are not significant. When averaged over all resonances, the difference was only 0.3 ± 7.3% (168): The effect of context was small.

TABLE III. The influence of the immediately preceding phonation mode on normal phonation. The table presents the fractional difference in average resonance frequency measured during normal phonations immediately before and after a whisper vs a creak phonation for a particular subject/vowel combination. Data were normalized by dividing by the average resonance frequency for that subject/vowel combination.

$\Delta R1/R1$	$\Delta R2/R2$	$\Delta R3/R3$	$\Delta R4/R4$
0.011 ± 0.094 (44)	0.012 ± 0.050 (45)	−0.002 ± 0.06 (39)	0.010 ± 0.072 (40)

TABLE IV. The estimated difference $\Delta R1$ in the first resonance frequency of the ten different subjects when changing from normal to whispered phonation, or from normal to creak phonation, measured in the same vocal gesture. Results from the five vowels studied have been combined. Subject No. 10 was not able to produce a satisfactory creak phonation.

	$\Delta R1$ (Hz)	
Subject	Normal to whisper	Normal to creak
1	255 ± 75 (12)	75 ± 50 (14)
2	185 ± 45 (9)	10 ± 20 (15)
3	235 ± 130 (5)	35 ± 45 (13)
4	260 ± 65 (9)	85 ± 30 (14)
5	235 ± 85 (7)	25 ± 30 (15)
6	215 ± 60 (13)	55 ± 35 (13)
7	290 ± 70 (12)	50 ± 45 (15)
8	235 ± 90 (4)	25 ± 70 (10)
9	360 ± 135 (12)	50 ± 50 (12)
10	245 ± 25 (9)	⋯

C. Resonance shifts due to whispering

Comparison between the average values for resonance frequencies measured during whispering and the average values measured during normal phonation (Table I) shows that $R1$ for whispering is distinctly higher than $R1$ for normal speech for all vowels. However, it is not immediately apparent that the other Ri are significantly higher during whispering. It is now possible to make use of the facts that pairs of estimates of the resonance frequencies for both normal and whispered phonations were made during the same vocal gesture, and that Table II indicates that the only properties of the tract that changed significantly over time during a vocal gesture were those associated with the change in phonation. The resonance frequencies for whispering in each individual gesture are taken to be the average of the values measured immediately before and immediately after each normal phonation in that gesture. The value of $R1$ measured during whispering was always found to be higher than the value of $R1$ measured during normal phonation in each individual vocal gesture; this was true for all subjects and vowels studied—see Table IV. The situation was similar for $R2$ with the value for whisper being higher than that for normal phonation for 115 of the 119 vocal gestures studied. $R3$ was higher for whisper than normal phonation in 94 of the 109 gestures, and $R4$ was higher for whisper in 88 of 118 gestures.

Table V shows the average values of the difference between pairs of values of the resonance frequency measured during whispering and during normal phonation, when measured during the same vocal gesture. It can be seen that all the resonance frequencies of the tract are significantly higher during whispering, and that the difference usually decreased for the higher resonances.

When averaged across subjects, the differences are always positive and always statistically significant at the 5% level for all vowels and all resonances, according to paired t-tests. Further, the magnitude of the difference decreases with the order of the resonance as predicted by Eq. (5)—see Fig. 3. (Shifts due to creak phonation are discussed later.)

D. Effective glottal dimensions during whispering

The estimated resonance frequencies for normal and whispered speech measured during the same vocal gesture can be used to estimate changes in glottal dimensions. For example, r_{GW}, the effective glottal radius during whispering, can be estimated. In the absence of appropriate information, the calculation presented here first assumes that the effective glottal length remains unaltered. Equations (5) and (6) can be rearranged to allow calculation of the effective length of the

TABLE V. The average differences ΔR in resonance frequency between whisper and normal phonations, or creak and normal phonations, measured in the same vocal gesture. The symbol * indicates that the difference was significant at the 5% level as indicated by a paired t-test.

Vowel	$\Delta R1$ (Hz)	$\Delta R2$ (Hz)	$\Delta R3$ (Hz)	$\Delta R4$ (Hz)
		Whisper phonation		
Head	255 ± 65 (21)*	140 ± 95 (20)*	145 ± 130 (24)*	105 ± 135 (25)*
Hard	220 ± 75 (20)*	155 ± 100 (23)*	120 ± 165 (22)*	45 ± 90 (22)*
Who'd	330 ± 190 (8)*	55 ± 140 (23)*	120 ± 105 (18)*	125 ± 150 (17)*
Hoard	280 ± 80 (19)*	70 ± 55 (25)*	115 ± 115 (20)*	45 ± 115 (27)*
Heard	250 ± 70 (24)*	150 ± 65 (28)*	110 ± 95 (25)*	75 ± 95 (27)*
All vowels	255 ± 90 (92)*	115 ± 105 (119)*	125 ± 125 (109)*	75 ± 120 (118)*
		Creak phonation		
Head	60 ± 45 (26)*	35 ± 80 (25)*	40 ± 95 (25)*	20 ± 85 (19)
Hard	40 ± 55 (21)*	−5 ± 50 (24)	60 ± 155 (23)*	25 ± 105 (22)
Who'd	45 ± 40 (24)*	20 ± 40 (24)*	−60 ± 145 (16)*	25 ± 170 (18)
Hoard	40 ± 45 (26)*	−15 ± 50 (26)	90 ± 105 (22)*	10 ± 110 (25)
Heard	35 ± 50 (24)*	10 ± 70 (25)	70 ± 90 (23)*	−10 ± 85 (24)
All vowels	45 ± 50 (121)*	10 ± 60 (124)	65 ± 120 (109)*	15 ± 110 (108)

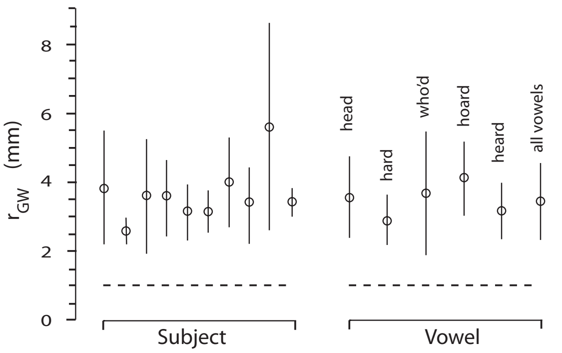

FIG. 7. Values of r_{GW}, the "effective" glottal radius during whispering for different subjects and vowels. They were calculated using the resonance frequencies for normal speech and whispering measured during individual vocal gestures. Values were calculated using Eqs. (10) and (11), assuming that $r_T=15$ mm and $L_G=3$ mm. The effective glottal radius in normal speech, r_{GN}, was assumed =1 mm as (indicated by the dashed lines on the figure). Error bars indicate standard deviations.

vocal tract (not including the glottis) in normal speech from the measured resonance frequency. Thus for the nth resonance of the tract

$$L_T = \tan^{-1}(-k_N r_T^2 L_G / r_{GN}^2)/k_N, \qquad (12)$$

where $k_N = 2\pi f_N / c$ and f_N is the nth resonance frequency measured in normal speech. This requires that values have to be assumed for r_T, L_G, and also r_{GN}, the effective glottal radius in normal speech. Providing that L_T does not change during the transition to and from whispering, then

$$r_{GW} = [-k_W r_T^2 L_G / \tan(k_W L_T)]^{1/2}, \qquad (13)$$

where $k_W = 2\pi f_W / c$ and f_W is the nth resonance frequency measured during whispering.

Thus, if estimates are available for the lowest resonance frequency in two different phonation modes, assumption of the geometry in one mode allows an estimate of the effective glottal area in the other mode. This estimation, based on the simple cylindrical model described above, also assumes that the rest of the tract geometry remains unchanged. Certainly, different vowels will produce different values of r_T for the upper vocal tract; however, the simple model is primarily concerned with the transition from glottis to the lower vocal tract, where r_T does not vary substantially from vowel to vowel.

Figure 7 presents the values of r_{GW} calculated from the measured values of $R1$ during a single vocal gesture using Eqs. (12) and (13). With one exception, the values were consistent across all ten subjects. The data for this one subject (subject 9—second from the right) were atypical (see Table IV) and were not used in further calculations of glottal radius. Figure 7 also indicates that a similar range of values of r_{GW} was associated with each of the five vowels studied; the average value being $3.4 \pm 1.1(79)$ mm. The increased glottal opening is consistent with observations made via laryngeal endoscopy (Matsuda and Kasuya, 1999). The glottal area during whispering was thus found to be $40 \pm 30(79)$ mm^2, a range that is consistent with directly measured glottal areas (Sundberg *et al.*, 2009).

The calculated values of r_{GW} will of course depend on the values assumed for r_T, L_G, and r_{GN}. However, the values used for the calculations shown in Fig. 7 produce a value of $Q_{\min} \approx 4$, where the dependence of frequency shift on $Q_{\min}$ (and thus on the initial assumptions of r_T, L_G, and r_{GN}) is relatively small—see Fig. 3.

There is also evidence that the supra-glottal region is constricted during whispering (Tsunoda *et al.*, 1997; Matsuda and Kasuya, 1999). The inclusion of such supra-glottal narrowing would lead to a smaller estimated value for r_{GW}—see Eq. (9).

The values shown in Fig. 7 assumed that the effective length of the glottis remained unchanged during the transition to whispering. In practice, for most tract geometries, an increase in r_{GW} is likely also to increase the effective L_G because of an increase in the end effect associated with a larger aperture. Thus a given change in frequency will be associated with a greater change in r_{GW}. To model this effect properly would require a more detailed model of the glottal geometry. However, an estimate may be obtained by continuing the simple cylindrical model and incorporating an end effect at the glottis/tract boundary by replacing L_G with $L_G + 0.85 r_{GN}$ in Eq. (12) and $L_G + 0.85 r_{GW}$ in Eq. (13). This produces a quadratic equation in r_{GW}.

$$L_T = \tan^{-1}[-k_N r_T^2 (L_G + 0.85 r_{GN}) / r_{GN}^2]/k_N, \qquad (14)$$

$$\tan(k_W L_T) r_{GW}^2 + 0.85 k_W r_T^2 r_{GW} + k_W r_T^2 L_G = 0. \qquad (15)$$

As expected, this approach yields an appreciable larger value; $r_{GW} = 6.3 \pm 3.5(79)$ mm. For geometries lacking circular symmetry, the influence of end effects is likely to be smaller.

In terms of this very simplified model the measured increase in estimated resonance frequency from normal phonation to whispering is consistent with a plausible increase in glottal aperture. The real anatomy is obviously much more complicated, but changes of similar order would be expected.

E. Resonance shifts due to creak phonation

The average values for resonance frequencies measured during creak phonation are slightly higher than the average values measured during normal phonation for all vowels (Table I). However, the standard deviations in $R1$ are large. The differences in the averages are smaller for the higher resonances, while the standard deviations remain large. However, it is again possible to examine pairs of resonances for different phonation modes measured during the same vocal gesture. (The resonance frequencies for creak phonation in each individual gesture are taken to be the average of the values measured immediately before and immediately after each normal phonation in that vocal gesture.)

Table V shows that the average frequency shift from normal phonation to creak is positive, small, and significant for the first and third resonances, the exception being the third resonance of who'd. The differences are usually not significant for the second and fourth resonances. The values of $R1$, $R2$, $R3$, and $R4$, measured during a single vocal gesture, were found to increase from normal to creak phonation in 84%, 59%, 81%, and 64%, respectively, of the gestures measured. The differences are generally positive, and this is consistent with the results of Ladefoged *et al.* (1988) and Ananthapadmanabha (1984), and inconsistent with the (small) decreases reported by Moosmüller (2001). However, these researchers used the peaks in the spectral envelope of normal phonation to estimate the resonances, which implies additional imprecision in the estimate of the resonance.

The observed average small increase in resonance frequency can be associated with a decrease in $Q_{\min}$ during glottal phonation, and this is consistent with a decrease in the ratio of glottal length to glottal area.

V. CONCLUSIONS

The resonance frequencies for whispered phonation for all subjects and vowels were found to be substantially higher than for the normal voice measured during the same vocal gesture, although the difference was greater than that found by other investigators. The increases are largest for $R1$ and decrease with increasing frequency. Calculations using a simple cylindrical model of the vocal tract, and assuming that the effective radius of the glottis is 1.0 mm for normal speech, yield a reasonable value of $40\pm 30\ \text{mm}^2$ for the effective glottal area during whispering. The lowest resonance frequencies of creak phonation were found to differ from those of normal speech by an average of 45 ± 50 Hz. This difference will usually be smaller than half of the fundamental frequency f_0, and then creak phonation might determine resonances with more precision than is available from the peaks in the spectral envelope of voiced speech, and be useful in teaching resonance tuning to singers (Miller *et al.*, 1997).

ACKNOWLEDGMENTS

We thank our volunteer subjects and the Australian Research Council for support. We would also like to thank Maëva Garnier and the reviewers for their comments on the manuscript.

Ananthapadmanabha, T. V. (**1984**). "Acoustic analysis of voice source dynamics," Speech Transm. Lab. Q. Prog. Status Rep. **25**, 1–24.

Barney, A., De Stefano, A., and Henrich, N. (**2007**). "The effect of glottal opening on the acoustic response of the vocal tract," Acta Acust. Acust. **93**, 1046–1056.

Chen, J. M., Smith, J., and Wolfe, J. (**2008**). "Experienced saxophonists learn to tune their vocal tracts," Science **319**, 776.

Chen, J. M., Smith, J., and Wolfe, J. (**2009**). "Pitch bending and *glissandi* on the clarinet: Roles of the vocal tract and partial tone hole closure," J. Acoust. Soc. Am. **126**, 1511–1520.

Clark, J., Yallop, C., and Fletcher, J. (**2007**). *An Introduction to Phonetics and Phonology* (Basil Blackwell, Oxford).

Coffin, B. (**1974**). "On hearing, feeling and using the instrumental resonance of the singing voice," NATS Bulletin **31**, 26–30.

Donaldson, T., Wang, D., Smith, J., and Wolfe, J. (**2003**). "Vocal tract resonances: A preliminary study of sex differences for young Australians," Acoust. Aust. **31**, 95–98.

Dowd, A., Smith, J. R., and Wolfe, J. (**1998**). "Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time," Lang Speech **41**, 1–20.

Epps, J., Smith, J. R., and Wolfe, J. (**1997**). "A novel instrument to measure acoustic resonances of the vocal tract during speech," Meas. Sci. Technol. **8**, 1112–1121.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

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

Gobl, C. (**1989**). "A preliminary study of acoustic voice quality correlates," STL-QPSR **30**, 9–22.

Henrich, N., Kiek, M., Smith, J., and Wolfe, J. (**2007**). "Resonance strategies used in Bulgarian women's singing style: A pilot study," Logoped. Phoniatr. Vocol. **32**, 171–177.

Hollien, H., and Michel, J. F. (**1968**). "Vocal fry as a phonational register," J. Speech Hear. Res. **11**, 600–604.

Joliveau, E., Smith, J., and Wolfe, J. (**2004a**). "Tuning of vocal tract resonance by sopranos," Nature (London) **427**, 116.

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

Jovocic, S. T. (**1998**). "Formant feature differences between whispered and voiced sustained vowels," Acustica **84**, 739–743.

Kallail, K. J., and Emanuel, F. W. (**1984a**). "Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects," J. Speech Hear. Res. **27**, 245–251.

Kallail, K. J., and Emanuel, F. W. (**1984b**). "An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects," J. Phonetics **12**, 175–186.

Ladefoged, P., Maddieson, I., and Jackson, M. (**1988**). *Investigating Phonation Types in Different Languages* (Raven, New York).

Lindblom, B. E. F., and Sundberg, J. E. F. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179.

Matsuda, M., and Kasuya, K. (**1999**). "Acoustic nature of the whisper," in Proceedings of the Eurospeech '99, pp. 133–136.

Miller, D. G., Sulter, A. M., Schutte, H. K., and Wolf, R. F. (**1997**). "Comparison of vocal tract formants in singing and non-periodic phonation," J. Voice **11**, 1–11.

Monsen, R. B., and Engebretson, A. M. (**1983**). "The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction," J. Speech Hear. Res. **26**, 89–97.

Moosmüller, S. (**2001**). "The influence of creaky voice on formant frequency changes," Forensic Linguistics **8**, 100–112.

Mukai, M. S. (**1992**). "Laryngeal movement while playing wind instruments," in Proceedings of the International Symposium on Musical Acoustics, Tokyo, Japan, pp. 239–242.

Pham Thi Ngoc, Y., and Badin, P. (**1994**). "Vocal tract acoustic transfer function measurements: Further developments and applications," J. Phys. IV **C5**, 549–552.

Smith, J., Henrich, N., and Wolfe, J. (**2007**). "Resonance tuning in singing," 19th International Conference on Acoustics, Madrid, Spain, Paper No. MUS-06-003-IP.

Smith, J., and Wolfe, J. (**2009**). "Vowel-pitch matching in Wagner's operas: Implications for intelligibility and ease of singing," J. Acoust. Soc. Am. **125**, EL196–EL201.

Smith, J. R. (**1995**). "Phasing of harmonic components to optimize measure signal-to-noise ratios of transfer functions," Meas. Sci. Technol. **6**, 1343–1348.

Sundberg, J. (**1977**). "The acoustics of the singing voice," Sci. Am. **236**, 82–91.

Sundberg, J., Scherer, R., Hess, M., and Müller, F. (**2009**). "Whispering—A single subject study of glottal configuration and aerodynamics," J. Voice

Swerdlin, Y., Smith, J., and Wolfe, J. (**2008**). "How whisper and creak phonation affect vocal tract resonances," J. Acoust. Soc. Am. **123**, 3242.

Tarnopolsky, A. Z., Fletcher, N. H., Hollenberg, L. C. L., Lange, B. D., Smith, J., and Wolfe, J. (**2005**). "The vocal tract and the sound of the didgeridoo," Nature (London) **436**, 39.

Tarnopolsky, A. Z., Fletcher, N. H., Hollenberg, L. C. L., Lange, B. D., Smith, J., and Wolfe, J. (**2006**). "Vocal tract resonances and the sound of the Australian didjeridu (yidaki): I. Experiment," J. Acoust. Soc. Am. **119**, 1194–1204.

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

Titze, I. R. (**2004**). "Theory of glottal airflow and source-filter interaction in speaking and singing," Acta Acust. Acust. **90**, 641–648.

Titze, I. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.

Tsunoda, K., Ohta, Y., Soda, Y., Niimi, S., and Hirose, H. (**1997**). "Laryngeal adjustment in whispering: Magnetic resonance imaging study," Ann. Otol. Rhinol. Laryngol. **106**, 41–43.

Wolfe, J., Garnier, M., and Smith, J. (**2009**). "Vocal tract resonances in speech, singing and playing musical instruments," Human Frontier Science Program Journal **3**, 6–23.