Vocal tract resonances in singing: The soprano voice[a)]

Elodie Joliveau, John Smith,[b)] and Joe Wolfe[c)]
School of Physics, University of New South Wales, Sydney NSW 2052, Australia

(Received 20 January 2004; revised 18 June 2004; accepted 15 July 2004)

The vocal tract resonances of trained soprano singers were measured while they sang a range of vowels softly at different pitches. The measurements were made by broad band acoustic excitation at the mouth, which allowed the resonances of the tract to be measured simultaneously with and independently from the harmonics of the voice. At low pitch, when the lowest resonance frequency $R1$ exceeded f_0, the values of the first two resonances $R1$ and $R2$ varied little with frequency and had values consistent with normal speech. At higher pitches, however, when f_0 exceeded the value of $R1$ observed at low pitch, $R1$ increased with f_0 so that $R1$ was approximately equal to f_0. $R2$ also increased over this high pitch range, probably as an incidental consequence of the tuning of $R1$. $R3$ increased slightly but systematically, across the whole pitch range measured. There was no evidence that any resonances are tuned close to harmonics of the pitch frequency except for $R1$ at high pitch. The variations in $R1$ and $R2$ at high pitch mean that vowels move, converge, and overlap their positions on the vocal plane ($R2$,$R1$) to an extent that implies loss of intelligibility. [DOI: 10.1121/1.1791717]

PACS numbers: 43.75.Rs [NHF] Pages: 2434–2439

I. INTRODUCTION

During normal voiced speech, 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). Resonances in the tract are controlled almost independently of f_0 by varying the position of the tongue, jaw, and lips. These resonances produce broad peaks in the spectral envelope of speech. Historically, the word ''formant'' has been used to describe both a resonance of the tract and a consequent peak in the spectrum of the output sound. However, these are physically quite distinct phenomena. To avoid confusion, the term ''formant'' will be used in this paper to describe a broad peak in the sound spectral envelope and Fi will be used to describe the frequencies at which these maxima occur. The term resonance will be used to denote an acoustic resonance of the vocal tract with resonance frequency Ri. Western vowels are generally identified by the frequencies of the first two formants ($F1$,$F2$) or those of their associated resonances ($R1$,$R2$).

Singers trained in the Western classical tradition often need to be heard in large auditoria, sometimes with loud orchestral accompaniment. Some, especially male singers, learn to produce a strong ''singers' formant'' and, thus, without extra effort, to produce greater power in the range around 3 kHz, a range where the competition from orchestras is reduced (Sundberg, 1974). This technique would be less effective for the high soprano range because the large spacing between the vocal harmonics means that few or no harmonics may coincide with such a resonance. In the high soprano range, however, the fundamental begins to enter the range at which human hearing sensitivity is greatest, so tuning $R1$ close to f_0 is a possibility that could produce a sound whose loudness varies less with pitch, and which is louder for constant effort (Sundberg, 1975, 1977, 1987). Indeed tuning $R1$ slightly above f_0 could maintain an inertive load on the vocal folds and consequently might enhance their vibration (Titze, 1988). As vowel identifiability is inevitably compromised once f_0 exceeds $R1$, this should not (further) reduce comprehensibility greatly. Sopranos are often taught to lower the jaw, to ''smile'' or to yawn as they ascend a scale; these actions increase mouth opening, which increases $R1$. Indeed, Sundberg and Skoog (1997) measured mouth openings increasing with f_0, consistent with the resonance tuning hypothesis.

It has proved difficult, however, to determine the degree to which this resonance tuning actually occurs during singing. The problem is that it is difficult to determine reliably the resonance frequencies of the tract from the sound alone, using either spectral analysis or linear prediction, once f_0 exceeds 350 Hz (Monson and Engebretson, 1983), and essentially impossible once f_0 exceeds 500 Hz.

Consequently, several indirect methods have been employed to see whether this tuning occurs. These have included use of an external vibrator held at the throat while the singer mimed singing (Sundberg, 1975), matching of sound spectra to various source-filter models (Sundberg, 1975), measurement of lip area and jaw opening followed by an articulatory model (Sundberg, 1975; Lindblom and Sundberg, 1971), and the use of various nonperiodic phonations such as vocal fry (Miller *et al.*, 1997). However, none of these techniques are capable of measuring the resonance frequencies precisely during natural singing (e.g., see Erickson and D'Alfonso, 2002).

Recently we have developed a new technique that uses an external broadband acoustic current source to excite the vocal tract resonances independently of the voice (Epps

[a)]A Brief Communication covering the tuning of only the average data for the lowest resonance has been published in Nature (London) **427**, 116 (2004).
[b)]Author to whom correspondence should be addressed. Electronic mail: john.smith@unsw.edu.au
[c)]Electronic mail: j.wolfe@unsw.edu.au

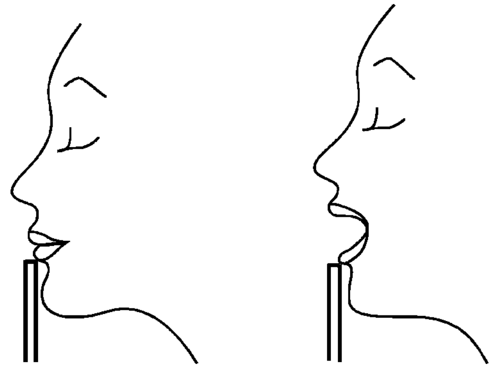

FIG. 1. The geometry (approximately to scale) in which the measurements were made showing how an acoustic current source and a microphone are placed so as to touch the singer's lower lip.

et al., 1997; Joliveau *et al.*, 2004). The present work aims to use this technique to determine the extent to which soprano singers tune their several vocal tract resonances to match harmonics of the sung pitch during normal singing.

II. MATERIALS AND METHODS

A. Measurements of vocal tract resonances

The resonances of the tract are measured directly, during sustained sung notes, using a technique described previously (Dowd *et al.*, 1997; Epps *et al.*, 1997). Briefly, a microphone and a small source of acoustic current, side by side on a flexible mounting, are placed just below the singer's mouth so that they gently touch the lower lip throughout the experiment—see Fig. 1. A computer (Mac IIci with analog interface card—National Instruments NB-A2100) synthesises the broad band signal from frequencies spaced at 5.38 Hz over the range of interest, here 0.2–4.5 kHz. The microphone is used to measure simultaneously the harmonics in the voice signal and the acoustic pressure produced when the broadband acoustic current interacts with the vocal tract. This current acts on the parallel combination $\mathbf{Z}_{\parallel}$ of the acoustical impedance of the vocal tract $\mathbf{Z}_t$ and that of the radiation field $\mathbf{Z}_r$ of the surrounding air. $\mathbf{Z}_r$ is inertive: it is a largely imaginary impedance whose value is low but which increases with frequency. It is measured for each singer by conducting a calibration experiment with her mouth closed, during which the acoustic current is adjusted so that the measured pressure spectrum p_{closed} is independent of frequency (see Fig. 2). Plots of the magnitude of the ratio $\gamma \equiv \mathbf{Z}_{\parallel}/\mathbf{Z}_r$ show peaks at the resonance frequencies of the tract (Dowd *et al.*, 1997). The acoustic current source has an output impedance much higher than the impedance of the load, so a plot of the magnitude of the ratio of the pressure measured during singing (p_{open}) to that measured during calibration (p_{closed}) shows peaks in the broad band signal corresponding to the vocal tract resonances, whose frequencies can be measured with a precision of order ± 11 Hz—see Fig. 2. It also shows the harmonics of the voice. Measurements with simple models of the vocal tract show that the resonance frequency measured in this fashion corresponds well with the resonances in the transpedance of the vocal tract (unpublished).

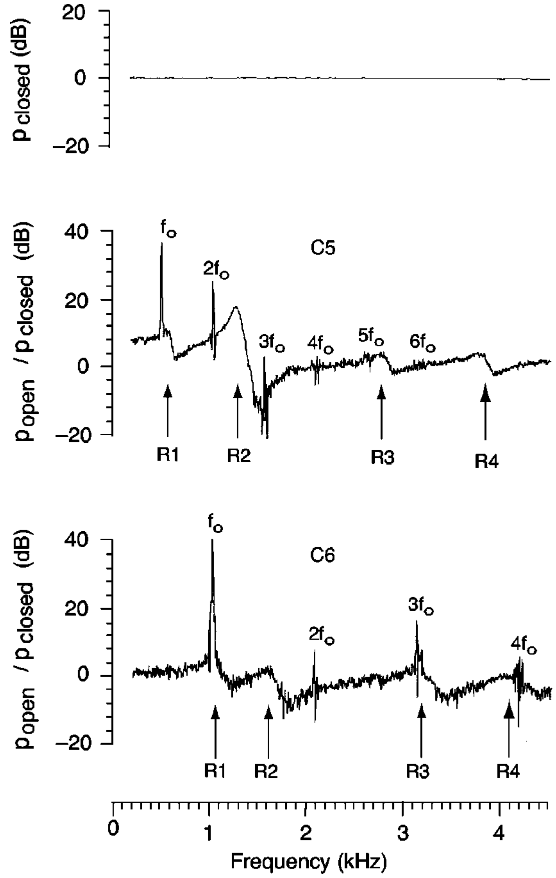

FIG. 2. The upper figure presents the pressure spectrum measured for a soprano without singing and with the mouth closed (p_{closed}). The spectrometer had been previously calibrated in this configuration by adjusting the acoustic current so that the pressure spectrum measured with the mouth closed was 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 ($p_{\text{open}}/p_{\text{closed}}$) 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 $R4$.

The presence of the current probe and microphone (total cross-sectional area approximately 120 mm^2) immediately below the lip slightly reduces the area for radiation and consequently the measured resonance frequency is expected to underestimate slightly the correct value. The magnitude of this difference was estimated by measuring the shift in resonance frequency that occurred when an additional current probe plus microphone was placed next to the first pair using a geometrically simplified, rigid model of a face and vocal tract. It was thus found that the error caused by the presence of the current probe lay within the resolution of these experiments (± 11 Hz).

B. The subjects

Nine sopranos participated (five were professionals with an average of 12 years classical training, four were students with an average of 7 years classical training). They all described their singing style as ''Western classical'' and sang predominantly in opera and choirs. All were born in Australia or had lived in Australia for at least 10 years and were judged by the investigators to speak educated, metropolitan, Australian English.

C. The experiments

Four vowels were chosen (ɑ,ɔ,u,ɜ) to ensure ease of singing and measurement, sampling of the phoneme space, and the effects of lip rounding. The word to be sung was presented in writing and had the form $h\langle\text{vowel}\rangle d$ (ɑ—*hard*; ɔ—*hoard*; u—*who'd*; ɜ—*heard*). Each subject then sang a note without vibrato that was sustained for 4 s. These notes comprised an ascending diatonic scale that covered their entire comfortable range for each vowel. The target pitch was provided by a glockenspiel before each note. They were asked to sing "softly, but in their trained singing style." They were asked to sing softly for the following reason. Singers can produce very high sound pressure levels immediately outside the mouth. The technique used requires a good signal-to-noise ratio and this requires that the sound pressure level produced by the injected broadband signal, at each injected frequency, should be comparable with that produced by the singing. As the broad band signal has many frequency components, the sound pressure level may become high enough to cause distortion in the microphones used. Although the distortion is small, the technique is rather sensitive to it. Subjects were asked to sing without vibrato because this causes a smearing of the harmonics of the voice signal that can obscure the tract resonances. No subjects complained or appeared to be worried by the presence of the broadband signal nor the request to sing *piano* and *senza vibrato*. Measurements were made in a quiet room with low reverberance. Reproducibility was measured by asking each singer to sing the vowel /ɑ/ in *hard* at pitch A4 before and after each scale. The standard deviation of the resonance frequencies of these reproducibility measurements, averaged across all singers, was thus found to be ±25 Hz ($R1$), ±60 Hz ($R2$), ±60 Hz ($R3$), and ±90 Hz ($R4$). The reproducibility for an individual singer was better than the variation among different singers.

III. RESULTS AND DISCUSSION

Figure 2 provides two examples of the measured ratio $p_{\text{open}}/p_{\text{closed}}$. One of the subjects was asked to sing the vowel /u/ (in *who'd*) at pitches C5 (523 Hz—near the middle of her range) and C6 (1046 Hz—near the top of her range). For C5, the first six harmonics of the voice are visible. It is apparent that, even at this modest pitch, it is difficult to estimate the tract resonances from the voice signal, simply because the spacing between harmonics (here 523 Hz) is too great. However, the resonances of the vocal tract are clearly seen in the broad band signal. $R1$ for this vowel and singer was about 420 Hz at low pitches around C4 (261 Hz). When this vowel was sung at C5, $R1$ has increased to a value slightly higher than the f_0 for C5. When the subject was asked to sing the same vowel at C6, her values of $R1$ to $R4$ have all increased. $R1$ and f_0 now coincide to within ~20 Hz, thus suggesting that it is possible for them to match quite closely. Again it is apparent that the resonances of the tract could not be determined from the voice signal at this pitch. From curves like those in Fig. 2, $R1$–$R5$ were determined for each singer for the four vowels over their comfortable pitch range. However, $R5$ was not always apparent in the frequency range studied, and $R1$ was not strong enough for reliable measurement on one singer.

Figure 3 shows the values for the vocal tract resonances for each vowel and pitch frequency f_0, averaged across all singers. The dashed lines represent $R = nf_0$, where $n = 1,2,...,6$. When a resonance coincides with the nth of these lines, the resonance is in tune with the nth harmonic of the voice. Consistent coincidence of the average resonance with one of the harmonics would suggest matching or tuning to that harmonic.

If there are several possible harmonics nearby, then it is insufficient to compare the resonance averaged over all singers with a single harmonic. This is because some singers might be tuning their resonance to the nth harmonic and other singers to the $(n+1)$th harmonic, and consequently the average would not coincide with a harmonic. This problem can be overcome by examining $|\Delta f|$, the absolute difference between a resonance and the nearest harmonic. Once $R1$ exceeds f_0, a resonance will always lie within $\pm f_0/2$ of a harmonic. If the resonance frequencies are distributed randomly with frequency rather than tuned, then the average $|\Delta f|$ would be expected to be close to $f_0/4$. On the other hand, where resonance tuning occurs, the average $|\Delta f|$ would be expected to be small with negligible dependence on f_0.

A. The tuning of *R*1

For each vowel and over the lower pitch range, f_0 is less than the value of $R1$ and the resonances $R1$ and $R2$ for each vowel are held approximately constant, independent of pitch (Fig. 3). This is the result that one would expect for speech, because $(R1,R2)$ characterizes vowels. However, once f_0 exceeds this value of $R1$, the value of $R1$ for the individual singers, and also for the average over all singers, increases with increasing f_0. This trend continues to 1 kHz for the vowels that do not use lip rounding (/ɑ/ in *hard* and /ɜ/ in *heard*), but for the vowels that use lip rounding (/ɔ/ in *hoard* and /u/ in *who'd*), the data fall below the tuning line near 1 kHz. This may be because, with the lips rounded, it is uncomfortable or anatomically impossible to raise $R1$ to 1 kHz.

The tuning line $R1 = f_0$ lay within the standard deviations of the data in the approximate range 600 to 1000 Hz. There was variation between the singers; some consistently tuned $R1$ to a frequency just above f_0 whereas others displayed no obvious pattern. The average value of $R1$ was, however, consistently slightly higher than f_0 and this supports the idea that vocal fold vibration is enhanced when the vocal tract presents an inertive load (Titze, 1988). There was no systematic difference between professionals and students—however, the lack of precise tuning in some subjects might be a partial consequence of singing softly.

It is possible that singers might tune $R1$ to match the second harmonic of f_0 at lower pitches. However, examination of $|\Delta f|$ over the range pitch A3 (220 Hz) to A4 (440 Hz)

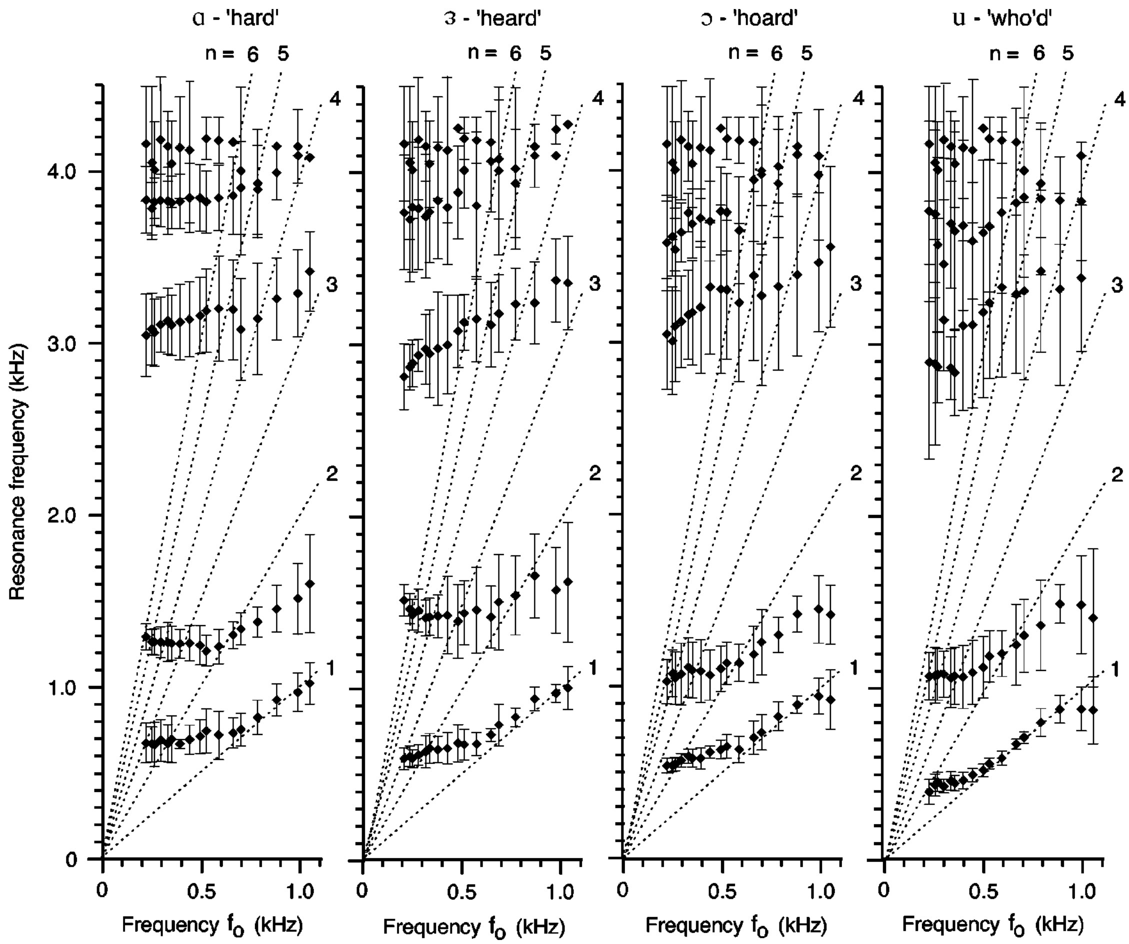

FIG. 3. The average vocal tract resonances for the four vowels studied, as a function of f_0. The vertical bars indicate the standard deviations. The dotted lines indicate the relationship $R=nf_0$, where n is the integer indicated.

found no evidence of systematic tuning for either the professionals or the students.

B. The tuning of R2

Once f_0 exceeds the normal value of $R1$, there is a systematic increase of $R2$ with f_0, but it is proportionally smaller than that measured for $R1$. This may be simply the result of the fact that, while $R2$ depends primarily on tongue configuration, it also depends somewhat on mouth opening (Fant, 1973): consequently the tuning of $R1$ can also vary $R2$. Figure 4 indicates that the overall trend is an increase in $|\Delta f|$ with f_0: as the separation of harmonics (f_0) increases the resonance tends to lie further from the nearest harmonic, as would be expected in the absence of tuning. The local

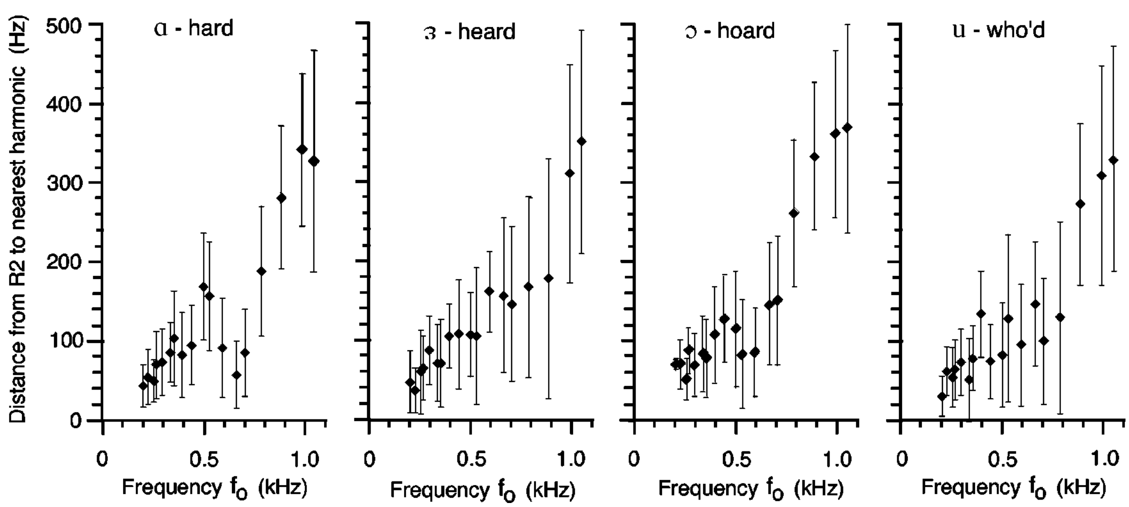

FIG. 4. The average value of $|\Delta f|$, the *absolute* value of the difference between $R2$ and the closest harmonic of f_0, plotted vs f_0 for each vowel. The vertical bars indicate the standard deviations.

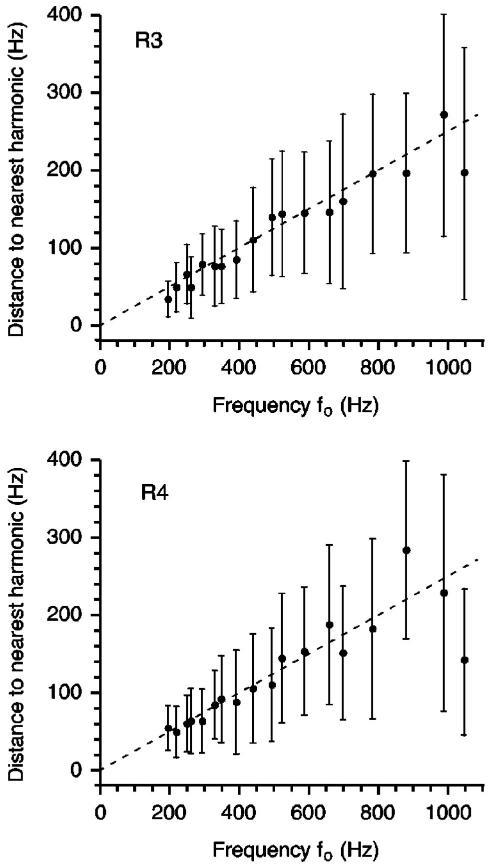

FIG. 5. The average value of $|\Delta f|$, the *absolute* value of the difference between $R3$ or $R4$ and the closest harmonic of f_0, plotted vs. f_0 and averaged over all vowels. The vertical bars indicate the standard deviations. The relationship $|\Delta f| = f_0/4$ that would be expected in the absence of tuning is indicated by a dashed line.

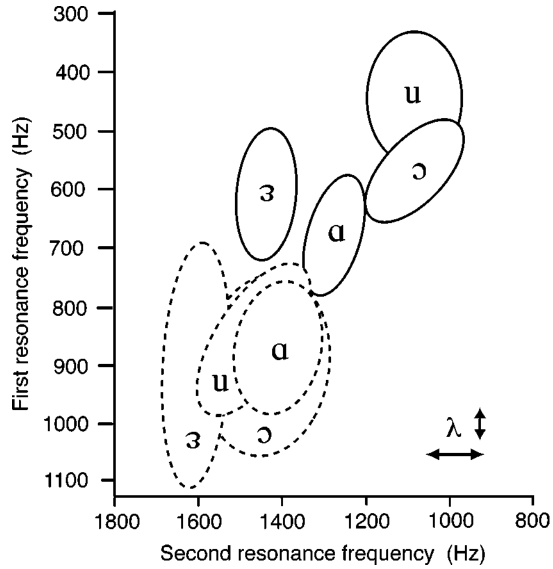

FIG. 6. Changes in position on the vowel plane as pitch is varied. The center of each ellipse gives the mean values of $R1$ and $R2$, and the slope of the major axis is the correlation between $R2$ and $R1$ for that vowel. The lengths of the semi-axes indicate the standard deviations calculated in those directions. The ellipses with solid borders are for the pitch range A3 to A4. At these low pitches, the vowels are well separated and their relative positions lie within the range typical for Australian English. The ellipses with dashed borders are for the pitch range G5 to B5. λ is a measure of the confusion length, the characteristic separation in the vowel plane at which vowels become confused.

minima around 700 Hz for ɑ and 600 Hz for ɔ correspond to the frequency at which the line $R = 2f_0$ crosses the curve $R2(f_0)$.

C. The tuning of *R*3–*R*5

The proportional variation in $R3$, $R4$, and $R5$ with frequency is not very strong (see Fig. 3), and showed no evidence of resonance tuning to match harmonics of f_0. Because the results were similar for all vowels studied, Fig. 5 shows the average $|\Delta f|$ as a function of f_0 for $R3$ and $R4$. In these data, the lines of best fit indicate that $|\Delta f| = 0.24f_0$, $0.23f_0$, and $0.20f_0$ for $R3$, $R4$, and $R5$, respectively. This suggests that the sopranos in this study did not significantly tune their higher resonances when singing softly.

Despite the absence of tuning in the higher resonances, there is a small but systematic increase in $R3$ and $R4$ for all vowels. $R3$ increased with increasing f_0 for all singers and vowels, with an average value of $R3$ on $f_0 = 0.48 \pm 0.39$. The variation of $R4$ with f_0 was more varied, and in some measurements $R4$ even decreased slightly with increasing f_0. However, the average regression of $R4$ on f_0 was similar and equal to 0.46 ± 0.38. These increases are observed across the whole range studied, and not merely across the range over which $R1$ is tuned. Perhaps these variations are related to changes in the vicinity of the larynx that are related to producing high f_0.

The singers' formant refers to a prominent peak in the sound spectrum around 3 kHz (Sundberg, 2001; Weiss *et al.*, 2001). This corresponds to the average measured value of $R3$. We have not performed a study of speech for sopranos and so cannot comment on whether $R3$, $R4$, and $R5$ are clustered differently during singing. Nor can we determine whether this peak is associated with a trained voice.

It must be remembered that the subjects in this study, although instructed to sing in their trained Western classical singing style, were singing softly without vibrato. The possibility thus exists that sopranos singing at full volume will indulge in additional and more precise control of resonances to increase the sound level produced. Such effects might be more important for singers at the very top of their profession.

D. Consequences for vowel intelligibility

Vowels are commonly represented on the $(F2,F1)$ or $(R2,R1)$ plane, where, in accordance with phonetic traditions, the axes are reversed. Figure 6 shows how tuning $R1$ close to f_0 moves the vowels studied in this plane. In the low range of f_0, the resonances of the vowels are well separated and their relative positions are consistent with those for normal speech (Donaldson *et al.*, 2003). In the high pitch range, all are shifted in the direction of increasing $R1$ and $R2$ and their separation is reduced, especially in the $R1$ direction. At high pitch, the vowels converge and overlap. Their separations become comparable with the confusion length λ, the characteristic separation in the vowel plane at which vowels become confused. [To date, this distance has only been measured for speech and for French vowels (Dowd *et al.*, 1997), whence the λ shown here. It is also likely that our perceptual categorization is modified when listening at high pitch.] This movement and convergence of vowels may contribute to the well-known difficulty in identifying vowels sung in the high range by sopranos (Berlioz, 1844; Scotto di Carlo and Germain, 1985; Benolken and Swanson, 1990) and might even be one of the reasons why opera houses often use surtitles even for operas in the native language of their audiences. Difficulty in comprehending the vowels of the high soprano voice cannot, however, be blamed on resonance tuning alone: once the spacing of harmonics approaches 1 kHz, it is virtually impossible for a human ear (or signal analysis) to determine features of the spectral envelope with the precision required to resolve vowels in the $(F1,F2)$ plane. This impossibility or near impossibility is appreciated by listening to recordings of "different" vowels at high pitch and, for this purpose, sound files have been placed on the web at http://www.phys.unsw.edu.au/~jw/soprane.html.

Because the price of (further) reduced intelligibility is not great, it is not surprising that sopranos elect to tune $R1$ close to f_0, thereby obtaining greater radiated power for a given effort, and also perhaps avoiding the effects on vocal timbre of having a fundamental whose amplitude varied strongly from note to note or vowel to vowel.

IV. CONCLUSIONS

A technique is demonstrated that can precisely measure the vocal tract resonances of sopranos during singing. The results presented are, however, subject to the constraint that the singers should sing softly without vibrato. In the lower part of their range, the vocal tract resonances of sopranos trained in the Western classical tradition have frequencies that vary little with pitch. However, once f_0 exceeds this normal value of $R1$, $R1$ is tuned so that $R1$ is close to f_0, except at very high pitches for vowels that involve lip rounding. Over the pitch range in which $R1$ is tuned, $R2$ increases a little, probably as an incidental consequence of the $R1$ tuning. $R3$ and $R4$ increase systematically, but not strongly over the whole soprano range. There is no evidence for the tuning of $R3$, $R4$, or $R5$ to harmonics of f_0.

ACKNOWLEDGMENTS

We thank our volunteer subjects and the Australian Research Council for support. We would also like to thank Professor Ron Scherer for his helpful comments.

Benolken, M., and Swanson, C. (**1990**). "The effect of pitch-related changes on the perception of sung vowels," J. Acoust. Soc. Am. **87**, 1781–1785.

Berlioz, H. (**1844**). *Grand traité d'instrumentation et d'orchestration modernes*, translated by M. C. Clarke (Novello, London, 1882).

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

Dowd, A., Smith, J. R., and Wolfe, J. (**1997**). "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.

Erickson, M. L., and D'Alfonso, A. E. (**2002**). "A comparison of two methods of formant frequency estimation for high-pitched voices," J. Voice **16**, 147–171.

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

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

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.

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.

Scotto di Carlo, N., and Germain, A. (**1985**). "A perceptual study of the influence of pitch on the intelligibility of sung vowels," Phonetica **42**, 188–197.

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

Sundberg, J. (**1975**). "Formant technique in a professional female singer," Acustica **32**, 89–96.

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

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois U.P., De Kalb, IL, 1987).

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

Sundberg, J., and Skoog, J. (**1997**). "Dependence of jaw opening on pitch and vowel in singers," J. Voice **11**, 301–306.

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

Weiss, R., Brown, W. S., and Morris, J. (**2001**). "Singer's formant in sopranos: fact or fiction?" J. Voice **15**, 457–468.