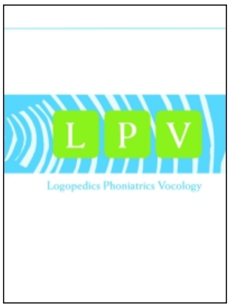

Logopedics Phoniatrics Vocology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ilog20

Using visual feedback to tune the second vocal tract resonance for singing in the high soprano range

Marie Jeanneteau , Noel Hanna , André Almeida , John Smith & Joe Wolfe

To cite this article: Marie Jeanneteau , Noel Hanna , André Almeida , John Smith & Joe Wolfe (2020): Using visual feedback to tune the second vocal tract resonance for singing in the high soprano range, Logopedics Phoniatrics Vocology, DOI: 10.1080/14015439.2020.1834612

To link to this article: https://doi.org/10.1080/14015439.2020.1834612

LOGOPEDICS PHONIATRICS VOCOLOGY
https://doi.org/10.1080/14015439.2020.1834612

RESEARCH ARTICLE

Using visual feedback to tune the second vocal tract resonance for singing in the high soprano range

Marie Jeanneteau, Noel Hanna, André Almeida, John Smith and Joe Wolfe

School of Physics, University of New South Wales, Sydney, Australia

ABSTRACT

Purpose: Over a range roughly C5–C6, sopranos usually tune their first vocal tract resonance (*R*1) to the fundamental frequency (f_o) of the note sung: *R*1:f_o tuning. Those who sing well above C6 usually adjust their second vocal tract resonance (*R*2) and use *R*2:f_o tuning. This study investigated these questions: Can singers quickly learn *R*2:f_o tuning when given suitable feedback? Can they subsequently use this tuning without feedback? And finally, if so, does this assist their singing in the high range?
Methods: New computer software for the technique of resonance estimation by broadband excitation at the lips was used to provide real-time visual feedback on f_o and vocal tract resonances. Eight sopranos participated. In a one-hour session, they practised adjusting *R*2 whilst miming (i.e. without phonating), and then during singing.
Results: Six sopranos learned to tune *R*2 over a range of several semi-tones, when feedback was present. This achievement did not immediately extend their singing range. When the feedback was removed, two sopranos spontaneously used *R*2:f_o tuning at the top of their range above C6.
Conclusions: With only one hour of training, singers can learn to adjust their vocal tract shape for *R*2:f_o tuning when provided with visual feedback. One additional participant who spent considerable time with the software, acquired greater skill at *R*2:f_o tuning and was able to extend her singing range. A simple version of the hardware used can be assembled using basic equipment and the software is available online.

ARTICLE HISTORY
Received 6 October 2019
Revised 17 August 2020
Accepted 5 October 2020

KEYWORDS
Resonance; vocal tract; formant; acoustic impedance; whistle voice; soprano; high pitch; vocal exercise; resonance tuning

Introduction

The trained soprano voice may have a large range of pitch and an impressively powerful output. One technique involved in both is the use of resonance tuning. The vocal tract exhibits several acoustic resonances *R*1, *R*2, *R*3, etc. In speech, the lowest frequency resonance, *R*1, usually occurs between about 300 and 800 Hz (about D4 to G5) and is mainly varied by changing the jaw height and mouth opening. The next resonance, *R*2, can be varied between about 800 and 2200 Hz (roughly G5 to C7), mainly by moving the tongue constriction backwards or forwards [1].

Observations of sopranos singing [2,3] and direct measurements of resonances [4] showed that sopranos modify their articulation as they ascend over a range about C5 to C6 so as to adjust *R*1 to match the sung pitch f_o; a technique called *R*1:f_o resonance tuning. Whether trained or not, virtually all sopranos seem to use *R*1:f_o tuning in the context of an ascending scale [5]. However, depending upon the target vowel, most singers find it difficult to tune *R*1 much above 1000 Hz (about C6) [5,6]. Relatively few sopranos sing much above C6; is this simply because the vocal folds are often unable to produce a sufficiently high f_o? Or is it a consequence of a suitable resonance not being available or some combination of these? (Figure 1).

The few who do sing above C6 often use *R*2 rather than *R*1; i.e. they switch to *R*2:f_o tuning [8,9]. This allows them to sing rather higher, sometimes up to or beyond C7: one octave above "high C". The sopranos who sing in this range can also use the M3 or "whistle" laryngeal mechanism of phonation [7,10]; although, the transition need not coincide with a change in resonance tuning [9]. Sopranos presumably have learnt *R*1:f_o tuning using the sound of their own voice as feedback and noticing that this tuning produces a louder sound for less effort, and perhaps a more stable sound. To increase lip opening and decrease jaw height as the pitch rises would thus become a habitual gesture [11] (some sopranos simultaneously tune *R*2 using *R*2:2f_o tuning when using *R*1:f_o tuning [12], but this also involves opening the lip aperture as the pitch rises).

To transfer from *R*1:f_o to *R*2:f_o tuning and to proceed above their *R*1 limit involves a counter-intuitive step: singers need to close their lips somewhat at the transition and to begin using the tongue constriction for tuning control. Perhaps as a consequence, few use *R*2:f_o tuning. These observations suggested the questions addressed by the present study: if given suitable visual feedback, can singers quickly learn *R*2:f_o tuning? Can they then use this tuning without the feedback? And finally, if so, does this assist their singing in the high range?

CONTACT John Smith john.smith@unsw.edu.au School of Physics, University of New South Wales, Sydney 2052, Australia

Supplemental data for this article can be accessed here.

In order to use visual feedback to teach resonance tuning, it is desirable to determine the resonance frequencies precisely and to show them in real time (meaning with no or little noticeable delay) in a readily comprehensible visual display. Most studies on speech or singing at low frequencies use the spoken or sung sound to estimate the resonance. For a sustained sung note at frequency f_o, the output sound from the larynx will also contain the harmonics $2f_o$, $3f_o$, $4f_o$, etc. These are filtered by the resonances of the tract to produce broad peaks in the spectral envelope known as formants F1, F2, F3, etc. [13,14]. In the low frequency range, the strongest harmonic in F1 gives an approximate estimate of $R1$, and similarly for Fi and Ri. Analysis of the sound using techniques including linear predictive coding can improve these estimates. However, these methods become increasingly inaccurate as f_o exceeds 350 Hz and virtually impossible above 500 Hz [15]; for a sustained signal at f_o the resolution cannot be much better than $f_o/2$. Similarly, inverse filtering also becomes increasingly unreliable as f_o increases [16].

The resolution can be significantly improved if a broadband source is used to excite the vocal tract (in such a source, the frequency components are much more closely spaced than f_o). One approach uses a non-harmonic voice source, in this case vocal fry or ingressive speech [17]. Another possibility is to use the broadband noise produced during whispering [18]. However, these methods cannot be used in real time whilst actually singing. Another approach involves vibrating the neck externally near the glottis using a broadband mechanical source [2,19]. However, the mechanical shaking of the larynx can perturb the participants.

A different approach is to excite the vocal tract at a single target frequency with a small loudspeaker placed just outside the open lips [20,21]. The singer then closes the glottis and changes their vocal tract configuration until the loudest resonance is heard. They then sing whilst trying to maintain the same tract configuration. This has the advantage of not significantly perturbing the singer. However, if the resonance changes when the glottis opens, this method does not provide any biofeedback for correction; the singer does not know if their resonance is above or below the target frequency and by how much.

One method of precisely determining the tract resonances during singing is to inject a broadband acoustic current at the open lips and to measure the resultant response [22,23] (the broadband sound has no pitch and has been likened to that of the wind). This does not unduly disturb the singer and can provide real time information on the resonance frequencies.

Normally, a singer has the sound of the voice as feedback, but this provides only indirect feedback on the resonance $R2$. One aim of this study was to use the technique of broadband excitation at the lips to provide visual feedback by simultaneously displaying the frequencies of the tract resonances and the voice harmonics in real time, and to use this feedback to teach $R2{:}f_o$ tuning to singers. Another aim was to observe whether singers thus trained would utilise this tuning when the visual feedback was absent.

Materials and methods

Overview

For most of this study, participants stood or sat with a small measurement device placed at their lower lip. They looked at a computer screen that displayed, in real time, the frequencies of the resonances of the vocal tract and of the harmonics of the voice. They were instructed how to use this visual feedback to modify their articulation.

Participants

Sopranos from an established amateur choir and a major national choir, having a self-assessed working range with an upper limit between about A5 and D6, were invited to participate in the study, whose design was approved by the institution's Human Ethics Committee. The invitation briefly described the procedure and explained that participation was voluntary and that data would be anonymised. They were not paid for their participation. Eight sopranos, from 27 to 64 years old, accepted the invitation. Seven were female and one a male (like the women, the male participant sang in mechanism M2, which he called falsetto, rather than head voice). Four were experienced amateurs with an intermediate level of singing (participants I1 to I4), three were advanced (A1 to A3) and one (P1) was professional (see Table 1).

Estimating tract resonances

The resonances of the vocal tract can be estimated under "ecological" phonation conditions, i.e. with the participant phonating in a normal fashion, by exciting the vocal tract with a broadband source at the lips [22]. This gesture is not perturbing: indeed, many singers will indeed be familiar with holding a microphone close to their lips. The technique involves placing a small, flexible source of broadband acoustic current with an attached small microphone at the lips of the participant, as shown in Figure 2. The flexible tube connected to the loudspeaker allows the current source and microphone to remain in the same position relative to the participant's lip. This allows participants to produce reasonably natural vocal gestures with little disturbance. The current source and microphone comprise a simple impedance head – a device used to measure the ratio of sound pressure to acoustic flow and thus to characterise resonances. This technique has the advantage that the tract resonances can be

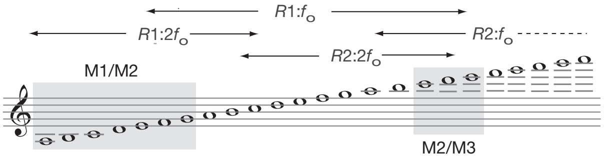

Figure 1. Possible approximate regions of overlap between laryngeal mechanisms and the four different resonance tunings used by women singers [5]. The M1/M2 transition corresponds to the *primo passaggio* and the M2/M3 transition to *secondo passaggio* [7].

measured in most situations with greater precision than the formants they usually produce, particularly at high pitch.

For these experiments, it is useful to make a calibration measurement for each participant with the lips closed rather than an absolute calibration. The software then displays $\gamma(f)$, essentially the ratio of the sound spectrum measured with vocal tract and voice present to that without. For this calibration, the source and microphone are first placed at the participant's lower lip with the lips closed; the "impedance head" is thus loaded with its radiation impedance Z_{rad} at that position, baffled by the face, i.e. $Z_{closed} = Z_{rad}$. The spectral envelope of the source signal from the computer is then adjusted, in an iterative process, so that the acoustic pressure measured, p_{closed}, is approximately flat (i.e. independent of frequency) with this particular load [25].

Next, using the same source signal from the computer, the acoustic pressure p_{open} is measured with the lips open. In this state, the acoustic load on the "impedance head" is the parallel combination of the vocal tract impedance at the lips, Z_{in}, and the radiation impedance, again baffled by the face, i.e. $Z_{open} = Z_{in}||Z_{rad}$.

To determine resonant frequencies, the magnitude of the ratio $\gamma(f) = p_{open}/p_{closed}$ is displayed. The source has a high output impedance compared with the load (i.e. it approximates an ideal current source), so the output current U is approximately the same with the lips open and closed ($U_{open} \cong U_{closed}$). Making this approximation, it follows that

$$\gamma \equiv p_{open}/p_{closed} \equiv U_{open}Z_{open}/U_{closed}Z_{closed} \cong (Z_{in}||Z_{\mathrm{rad}})/Z_{rad}$$

$$\gamma \cong Z_{in}/(Z_{rad} + Z_{in}) \tag{1}$$

At resonance, $Z_{in} \cong -Z_{rad}$, which corresponds approximately to maxima in $\gamma(f)$. In this fashion, it is possible to estimate the resonance frequencies of the tract as measured at the lips, and these are expected to be similar to the resonances if measured near the vocal folds. A real time display of the magnitude and phase of $\gamma(f)$ is provided as part of the visual feedback in this study. In earlier studies, $R1$ and $R2$ were estimated and plotted on the vowel plane in "real time" [22] and used for language training [23]. For some other studies not involving training, a delayed display of $\gamma(f)$ was adequate.

Hardware

The "impedance head" used for this experiment was built with a 1.0 m length flexible tube of 5 mm internal diameter placed next to a 1/4 inch pressure microphone (Brüel and Kjaer 4944-A). We used a conditioning amplifier (Brüel and Kjaer Nexus 2690) whose signal was digitised at 44.1 kHz and 16-bit resolution by an audio interface (MOTU 828) connected to a computer (iMac 2.7 GHz). The computer output used the same interface and the analogue signal was amplified by a 20 W audio amplifier (built in-house) which drove a 1 inch compression driver (BM-D450 MKII, P.Audio, Nakhon, Thailand) connected to the tube via an aluminium horn of length 40 mm. Details of simpler and cheaper version of this apparatus that should be adequate for resonance training are available in [26].

Table 1. Details of the experimental participants and the limits of their singing.

	Participant								
	I1	I2	I3	I4	A1	A2	A3	P1	MJ[a]
Gender	F	F	F	M	F	F	F	F	F
Experience (years)	12	4	2	1	33	8	10	14	8
Experience [24]	8	8	7	8	5	5	7	4	8
Before training									
Highest comfortable note	A5	G5	B5	F5	C6	C6	A5	D6	A5
Upper singing limit	C6	C6	E6	C6	E6	Eb6	C6	E6	C6
During training with feedback									
Glottal and velar control	Y	Y	Y	N	N	Y	Y	Y	Y
Lowest $R2$ (miming)	A5	D6	–	G5	F#5	D5	D#5	E5	D#5
Lowest $R2$ (singing)	A5	D#5	D#5	G5	–	F5	E5	G5	E5
Highest $R2$ (singing)	C#6	C6	–	G5	–	D6	D#6	–	F#6
After training									
$R2$:f_o tuning without feedback	N	N	N	N	N	Y	Y	N	Y

Experience is specified using the taxonomy of Bunch and Chapman [24]: 4 = regional/touring, 5 = local community (often semi-professional), 7 = full-time voice student, and 8 = amateur. Y and N signify "yes" and "no", respectively. "–" indicates that the subject could not perform the exercise.

[a]Author MJ had a separate, long-term use of the software.

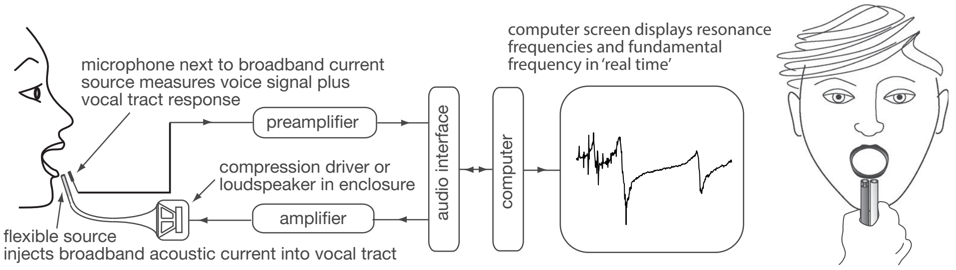

Figure 2. Schematic diagram of the apparatus (not to scale). The acoustic current source and the microphone are actually attached side by side, but are here shown displaced for clarity.

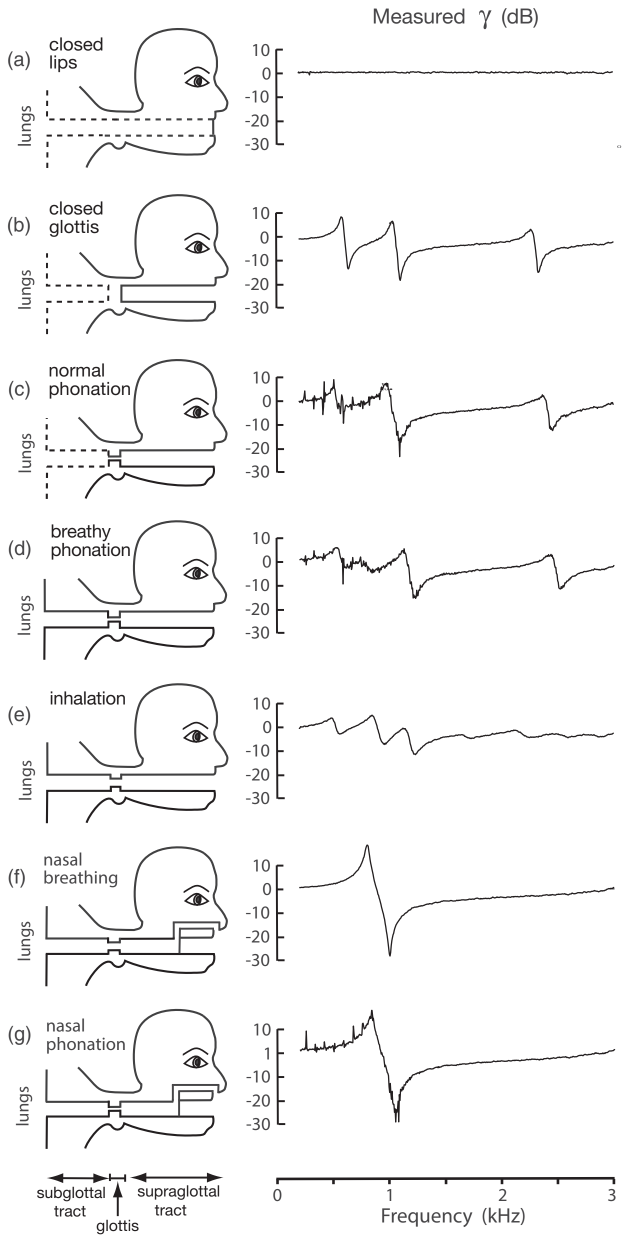

Figure 3. Examples of how different vocal tract configurations can be identified from the measurements at the lips. At right are the measured ratios $\gamma(f)=p_{open}/p_{closed}$ as one of the authors produces seven different gestures. The schematics on the left suggest the vocal configuration in terms of a 1D model of the tract; they are highly simplified and not to scale.

Software

Panels like that shown in Figure 4 provided real-time visual feedback. The open-source software used to display these was written using MATLAB and is available for download [27], with more details given in [28,29]. Several panels allow the user to generate the broadband signal, to calibrate with the lips closed, to record the audio response and to display the impedance ratio $\gamma(f)$ and f_o in real-time. Both the magnitude and phase of $\gamma(f)$ are displayed; the phase can be useful in detecting resonances in the presence of noise, but was not necessary in these training exercises.

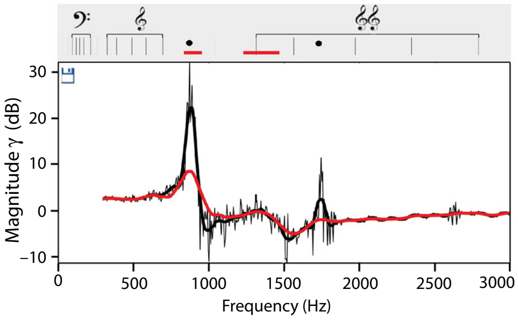

Figure 4. This screen grab shows the magnitude of the impedance ratio $\gamma(f)$ when a soprano sings at A5. Superposed on the raw $\gamma(f)$ signal (thin, noisy black line) is a smoothed version (thick black line) and the smoothed signal with the voice harmonics removed (thick grey line). Above the graph, black dots show the detected harmonics and the horizontal grey bars indicate the range of the maxima in $\gamma(f)$ associated with the resonances (their sensitivity can be adjusted by the user). The harmonics and resonances are shown as pitches on the three musical staves (tilted sideways) and clefs at the top of the display (the super-treble clef (𝄞) is two octaves above the treble and its staff extends from 1320 to 2800 Hz). In this example, the singer is tuning the first resonance $R1$ to the fundamental frequency f_o ($R1$:f_o tuning).

Estimating vocal tract configuration

This technique not only allows estimates to be made of the resonance frequencies of the supraglottal tract, it can also reveal other useful information on the vocal tract configuration (degree of glottal opening, raised/lowered velum, etc.). Such information is useful as participants will sometimes inadvertently produce undesirable configurations as they learn to tune $R2$; it is important to recognise and to understand these configurations. Figure 3 shows some examples of how different vocal tract configurations will affect $\gamma(f)$; they have been measured on one of the authors after calibrating p_{closed}.

Figure 3(a) shows the expected result for closed lips; any slight deviations from a ratio of one being due to subsequent small changes in relative position between lips and impedance head, or digitisation error. Large deviations indicate that a new calibration would be advantageous.

Figure 3(b) shows the result for a closed glottis and raised velum; three supraglottal resonances ($R1$, $R2$, and $R3$) are clearly visible. This configuration is desired for the miming exercise described below.

Figure 3(c) shows what happens for normal phonation with a raised velum; the glottis is still sufficiently small to isolate the subglottal tract and harmonics of f_o are now visible. This is the desired configuration for singing with feedback.

Figure 3(d) shows the result for a breathy phonation; the glottis is now sufficiently open for an additional resonance to appear around 750 Hz as a result of interaction between supra- and subglottal tracts [30]. A display like this can be used as feedback to discourage the breathy voice.

Figure 3(e) shows the result when inhaling. The glottis is now fully open and a new set of resonances appears: these are associated with the combined length of the subglottal and supraglottal tracts [30].

Figure 3(f,g) shows the consequence of a lowered velum; in each case, a single strong resonance around 900 Hz

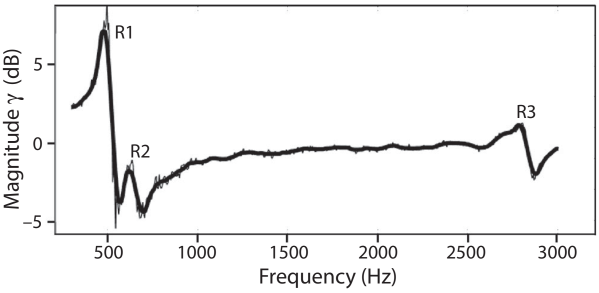

Figure 5. Participant A2 could produce the lowest *R*2 during miming. In this example, she has reached D#5 (~620 Hz).

appears as a consequence of the velum tending to seal the mouth cavity from the other tracts. A lowered velum prohibits $R2 : f_o$ tuning.

For the resonance tuning training described in this paper, $\gamma(f)$ should look qualitatively like Figure 3(c).

Before attending the experimental tuning session, participants were asked to practise two exercises. The first involved learning how to open and close the glottis. The second involved using nasalisation to learn how to raise and to lower the velum. Details are given in the supplementary file.

Initial control experiment

Participants stood or sat, as they felt more comfortable, and were asked to hold in their hand the small flexible tube with a microphone attached. They were asked to adjust the position of the microphone and the flexible tube so that it and the microphone rested gently on their lower lip throughout the experiment. One approach was to hold both tube and microphone in a fist and touch the thumb to the chin as a point of reference – see Figure 2. The tube was connected to a loudspeaker that excited the vocal tract with a synthesised broadband signal consisting of a sum of sine waves from 300 to 3000 Hz at 11 Hz spacing.

An initial "calibration" procedure was then made with the lips closed. A subsequent measurement taken with the lips closed should then look like Figure 3(a).

In a first step, singers were asked to perform one or more chromatic scales on the vowel in "hard" [ɑ], each scale ascending from a starting note of their choice, with the appropriate pitch cued on a glockenspiel. This determined the highest pitch that the singer could produce under these normal conditions. Singers were then asked to sustain each note of the same ascending scale for 3 s, still on an [ɑ] vowel, with no change in pitch or loudness and with minimum vibrato. The broadband current source was active during these measurements, so that the singer's habitual resonance tuning strategy could be quantified. The scale could be produced in one or more breaths as necessary until the singer reached the upper limit of their range.

Training session

For this, participants looked at a computer screen that displayed the visual feedback described below and were coached to produce the necessary control. The training session lasted one hour and comprised the following five exercises.

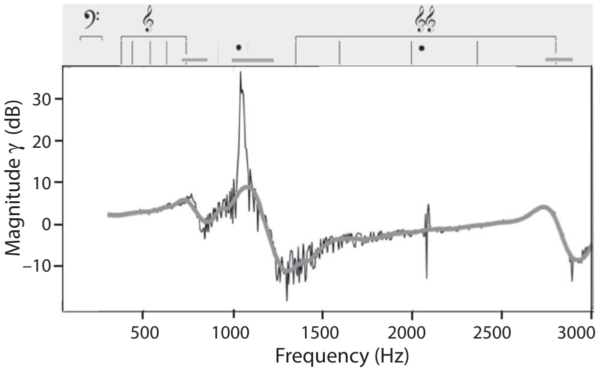

Figure 6. A screen grab of participant A2 using *R*2:f_o tuning singing at C6 with feedback. The plot shows the magnitude of the impedance ratio γ(*f*) as described in Figure 4.

Open/closed glottis

In order to obtain strong vocal tract resonances in the supraglottal tract, a closed or nearly closed glottis is required. To verify whether participants could switch easily between closed and opened glottis, they were asked to repeat their practice exercise whilst miming (i.e. without phonating) with the apparatus at their lips, whilst observing $\gamma(f)$ and comparing it with those in Figure 3. If participants were able to fully close their glottis, only two or three strong, broad peaks (corresponding to *R*1, *R*2, and/or *R*3) would appear on $\gamma(f)$ over the frequency range displayed (cf. Figure 3(b)). An open glottis can be identified by the presence of an additional resonance as the subglottal tract becomes involved – this can be seen in Figure 3(e) during inhalation or Figure 3(d) for a wide glottis leading to a breathy voice [30].

Once their glottis was closed, the experimenter checked the ability of participants to control the position of the peaks in $\gamma(f)$ by changing their mouth aperture, rounding their lips or changing their tongue position whilst miming.

The introduction of the harmonic voice signal produced by phonation can now be demonstrated for an open and a closed glottis – Figure 3(d,c), respectively.

Raised/lowered velum position

Nasalisation was not studied in this project and participants needed to avoid it in order to produce *R*2. In practice, this meant avoiding a $\gamma(f)$ spectrum with a single peak around 900 Hz (cf. Figure 3(f,g)), indicating nasalisation and a lowered velum.

Lowering the second resonance R2 whilst miming

Participants were shown that the peak positions of *R*1 and *R*2 change when singing a long steady note softly at a comfortable pitch using the vowels of the five words "who'd", "heard", "hard", "hoard", and "heed". They were then asked to try and keep *R*2 as low as possible whilst miming with a closed glottis: this requires rounding the lips and keeping the tongue down and towards the back of the mouth, whilst ensuring that the tongue does not contact the velum. The aim was to reduce *R*2 sufficiently so that it fell within their singing range. Figure 3(b) shows an example of the desired $\gamma(f)$ spectrum.

Learning R2:f_o tuning whilst singing

An example of the display used as feedback in the tuning training session is shown in Figure 4. Here, the display shows the measured $\gamma(f)$ and a smoothed version with the voice harmonics removed. The fundamental frequency of the voice f_o and its harmonics are shown by dots on musical staves, and the approximate bandwidths of *R*1 and *R*2 are indicated by horizontal bars. Participants are thus provided with direct visual feedback to try and match *R*2 to f_o (they also, of course, had the sound of their own voices as feedback).

Once participants had shown that they could control the position of *R*2 whilst miming, the next step was to try tuning *R*2 approximately to f_o. Thus, participants would select a pitch for f_o (often F5 or G5 was a starting point) and, whilst miming with a closed glottis, then adjust *R*2 so that it matched the intended f_o. Participants would then try and hold the same mouth, tongue, and lip positions, think about the pitch they were going to sing, and open their glottis. If *R*2 remained in approximately the desired position the participant could then sing and see if they were matching *R*2. Participants then repeated this procedure and tried singing at different values of f_o covering the range G5 to D#6.

For these exercises, tuning means the systematic changing of *R*2 to follow approximately the singing frequency f_o; participants were not immediately expected to match *R*2 exactly with f_o. Tuning f_o within the bandwidth of *R*2, to within about 100 Hz [31,32], should offer benefits such as increased sound output, and perhaps stability. Exact tuning presumably takes extensive practice for a singer when only the produced sound is available to provide feedback. Furthermore, estimation of *R*2 from maxima in $\gamma(f)$ also introduces some uncertainty, particularly with the simplified approach used here.

Learning R2:f_o tuning at very high pitch

Preliminary experiments on author MJ showed that at very high pitch (starting around B5) sometimes *R*2 would increase on opening the glottis. For this reason, participants were asked to tune *R*2 about a semitone below the target f_o whilst miming with closed glottis before trying to sing.

Results and discussion

Initial measurements before training

All of the eight participants used *R*1:f_o tuning for much of their range. Although some simultaneously used *R*2:2*fo* tuning over part of their range, no participant used *R*2:f_o tuning at any pitch.

Control of glottis and velum position

Six of the eight singers were able to control both glottis aperture and velum position efficiently after two or three attempts. The two other singers, I4 and A1, had difficulty in closing the glottis. Once the glottis was closed, they had difficulty keeping the velum raised and as a result nasalised most of the time, especially below G5. These two participants were sometimes asked to revise their velum control by pinching their nose at times during the *R*2:f_o training session.

Lowering R2 whilst miming

Seven of the eight singers were able to lower their second resonance *R*2 whilst miming. Only I3 could not, as this participant found it difficult to close the glottis without lowering the velum. The lowest *R*2 reached during miming was D5; Figure 5 shows an example with *R*2 lowered to D#5 (~620 Hz).

R2:f_o tuning with feedback

Once participants could lower *R*2 sufficiently whilst miming, they could try tuning *R*2 approximately to f_o whilst singing. Six of the eight participants were able to lower *R*2 below A5 (~880 Hz). For the five participants who could control glottis opening easily, *R*2 when miming was equal or below that when singing, except for I2 who could not lower it below D6 when miming. These results are summarised in Table 1.

Figure 6 shows an example of successful *R*2:f_o tuning whilst singing at C6.

Six of the eight participants were able to position *R*2 at the intended f_o before phonation. Their comments on the procedure were:

- I1 found "good to see the *R*2 position on the screen" and added that "seeing is clearer than feeling";
- I2 felt that "the experiments need[ed] more practising" as "the hardest thing for [her] is to keep/control the position of tongue and mouth";
- A2 who did the exercise particularly easily (see Figure 6) commented that it was "not painful" (this comment surprised us, but it appears this participant had feared that using a different tuning might be uncomfortable or worse).

Some participants found it difficult to follow the procedure. Their comments are summarised as follows:

- A1 could not do the exercise at all because she found it difficult to lower *R*2 without nasalising and also difficult to move the velum up and down to isolate the nasal passage;

Table 2. The average resonance frequencies before and after the transition period from miming (f_{R1m} and f_{R2m}) to singing (f_{R1s} and f_{R2s}) when feedback was present.

	Miming (before)		Singing (after)			
Singer	f_{R1m}/f_o	f_{R2m}/f_o	f_{R1s}/f_o	f_{R2s}/f_o	f_{R1s}/f_{R1m}	f_{R2s}/f_{R2m}
I1	0.70 ± 0.04	**1.12**±0.03	0.70 ± 0.10	1.27 ± 0.13	0.99 ± 0.12	1.13 ± 0.09
I2	0.59 ± 0.11	**0.96**±0.19	0.60 ± 0.05	1.20 ± 0.09	1.03 ± 0.12	1.28 ± 0.28
I4	0.55	**0.99**	0.68	**1.03**	1.24	1.04
A2	0.82 ± 0.07	**1.02**±0.10	0.75 ± 0.05	**1.11**±0.02	0.91 ± 0.04	1.10 ± 0.11
A3	0.55 ± 0.05	**0.95**±0.06	0.72 ± 0.06	**1.12**±0.03	1.30±–0.13	1.19 ± 0.07
P1	0.69 ± 0.03	**1.06**±0.07	**1.11**±0.02	1.55 ± 0.09	1.63 ± 0.10	1.46 ± 0.10

The standard deviations for each singer were calculated from their measurements at different pitches f_o, where f_o is the measured frequency during singing. Values in bold indicate when $R1$ or $R2$ were within 12% of f_o. Singer I4 could only sing through the transition at one pitch.

- I3 could barely do the exercises because she found it difficult to close the glottis without simultaneously lowering the velum. She had to do the nose-pinching exercise often during training (however, like all sopranos studied in this lab, she always tuned $R1{:}f_o$);
- I4 often lowered the velum as he switched from glottis closed to open, thus nasalising somewhat when starting to sing;
- P1 could position $R2$ at the desired frequency before phonation, but always shifted to $R1{:}f_o$ tuning when opening the glottis to sing.

The transition from miming to singing

Table 2 summarises the results during the transition from miming with a closed glottis (subscript m) to singing (subscript s) for the six participants who were able to control both glottis and velum. The ratio f_{R2m}/f_o during miming in Table 2 was generally close to one, indicating that all six participants were then tuning $R2$ close to f_o, where f_o was the actual value measured during singing. The ratio f_{R1m}/f_o was in the range 0.55–0.82 indicating the absence of $R1{:}f_o$ tuning.

Table 2 shows different behaviours during the transition from miming to singing at time = 0; defined as the instant when voice sound from the participant was first detected in the audio response. Some participants (I1, I4, A2) seemed to maintain their phonatory and articulatory shape during the transition, as indicated by only a small increase in the ratio f_{R2s}/f_{R2m}, e.g. see Figure 7(c). Other participants (I2, A3) increased $R2$ around the moment of transition, e.g. see Figure 7(b). Changes in $R1$ and $R2$ are reflected in the ratios f_{R1s}/f_{R1m} and f_{R2s}/f_{R2m} in Table 2.

One participant (P1) anticipated the transition and increased $R1$ and $R2$ by an average of 63% and 46%, respectively. She thus effectively switched from $R2{:}f_o$ to $R1{:}f_o$ tuning at the moment singing commenced (see Figure 7(a)).

Interestingly, the ratio f_{R2s}/f_{R2m} was greater than one for all six participants, i.e. f_{R2} increased during the transition, perhaps as a consequence of the glottal opening and changes in its geometry. There is also the possibility that the value of $R2$ experienced by the vocal folds might be higher than that displayed on the screen, and participants might then use acoustic feedback. However, changes in $R1$ were not as consistent: the ratio f_{R1s}/f_{R1m} increased for three participants and decreased for the other three.

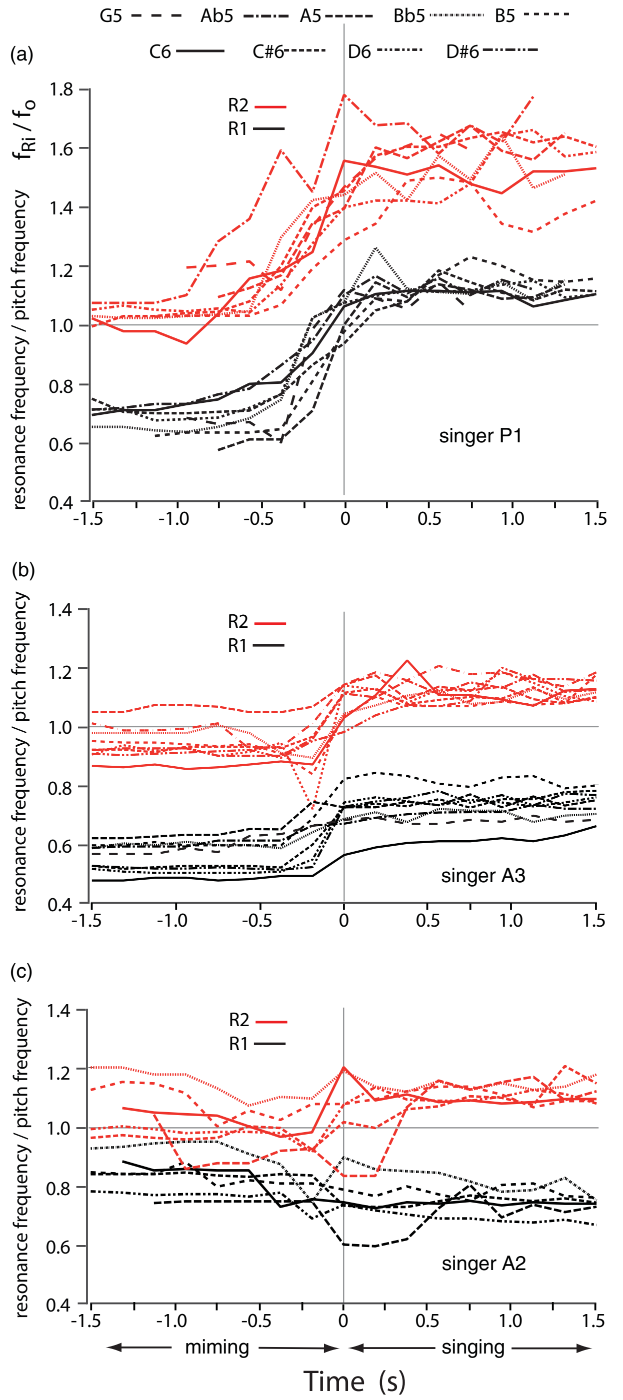

Figure 7. The ratios $R1/f_o$ (black lines) and $R2/f_o$ (grey lines) during the transition from miming with a closed glottis (negative times) to singing (positive times). Time = 0 indicates when each singing measurement started, i.e. at the beginning of the first cycle of broadband signal following the instant when the microphone detected the start of the output voice sound. A new cycle of broadband signal started every 186 ms. (a) Participant P1 successfully tunes $R2{:}f_o$ whilst miming, but switches to $R1{:}f_o$ tuning when preparing to sing. (b) A3 successfully tunes $R2{:}f_o$ whilst miming, but raises $R2$ when singing to about 12% above f_o. (c) A2 similarly increases $R2$ by 10% during the transition from miming to singing. See Table 2 for statistics.

Nevertheless, inexact tuning is probably useful, provided that f_o lies within a bandwidth of Ri. Participants can then use acoustic feedback to refine their tuning.

Participants A1 and P1, who had the highest comfortable range before training, were unable to perform $R2$:f_o tuning during singing when feedback was present. Their comments below suggest that it was difficult to override their already successful singing strategies in such a short training session.

- A1 felt "that too many years singing, training to open the mouth the higher you sing, made it difficult to do the opposite" and "voice box seemed to grip as [she] tried to imagine [herself] singing higher";
- P1 found it "very difficult to 'train' [her] mouth (when opening the glottis) not to change shape into a more comfortable singing position" and "felt constricted the higher the sound got". Again, when trying to open the glottis and produce sound, she felt "constricted around the throat."

The effects of this short training session using $R2$:f_o tuning with feedback were varied. Two participants (I1 and A3) increased their range above C6; however, the range of one participant (I2) remained the same, and the range of two participants (I4 and A2) actually decreased.

Subsequent tuning in the absence of feedback

After the training session, participants were asked to sing an ascending chromatic scale whilst $\gamma(f)$ was measured as described previously for the initial control experiment. For this exercise, visual feedback of $\gamma(f)$ was not provided to the participants and they were not specifically asked to use $R2$:f_o tuning. Four of the six participants who demonstrated $R2$:f_o tuning when feedback was present during singing did not use it when feedback was absent. However, two participants (A2, A3) did spontaneously switch from $R1$:f_o to $R2$:f_o tuning at the very top of their range; at Eb6 and D6, respectively – see Figure 8. The pitch at which this transition occurred was similar to that observed in earlier studies [8,9].

Although a single one-hour training session was sufficient for six participants to learn $R2$:f_o tuning when visual feedback was present, only two participants subsequently used $R2$:f_o tuning when feedback was absent, and then only at the very top of their range. It thus seems possible that longer periods of training could help sopranos employ $R2$:f_o tuning to extend their range, providing that they can also learn an appropriate laryngeal mechanism to increase f_o.

Results of long-term use of visual feedback

The eight participants in the main part of this study only had a single, one-hour, training session with the software. However, the first author of this paper (MJ in Table 1) used the software almost daily for an estimated 30 min (non-consecutive) over four months. She reported that the main challenge lies with proprioceptive sensitivity and ability. The visual feedback allows detection of the phonatory and articulatory shape that may be optimal for a specific pitch, but kinaesthetic learning is necessary to analyse, keep and later reproduce this optimal position, and to avoid physiological or habitual reflexes. This step will presumably constitute the longest part of learning to use $R2$:f_o tuning.

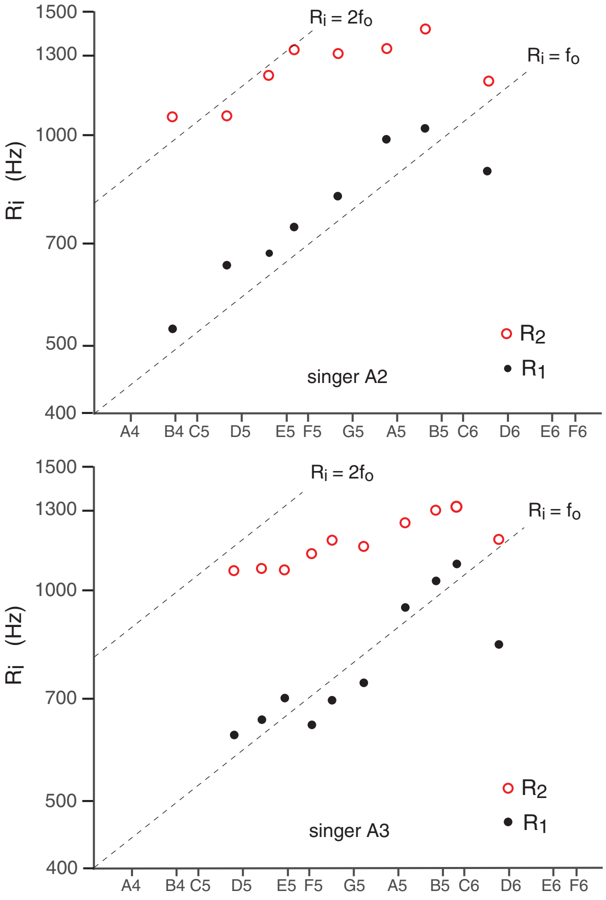

Figure 8. Examples where two participants spontaneously switched from $R1$:f_o to $R2$:f_o tuning at the very top of their range above C6 after training when feedback was absent. In each case $R2$ (open grey circle) has dropped in frequency to replace $R1$ (full circle) close to the $R1$=fo tuning line (dashed diagonal line).

As an amateur (see Table 1), she had no previous specialist knowledge of the vocal system. Although she had sung for eight years, she had had no formal singing training. However, this perhaps also made her flexible since she was not influenced by any specific vocal technique or school. Prior to the study, her comfortable and limiting upper pitches were A5 and C6, respectively. It took her about 4 weeks of practice to have competent and independent management of glottis opening, velum position, and vocal tract gesture.

Using the visual feedback as described above over four months, she was able to use either $R1$:f_o or $R2$:f_o tuning over the one octave range from E5 to E6. Without feedback she extended her comfortable range by a fifth (A5 to E6), and her new upper limit is F#6. Her upper limit using $R1$:f_o

is C#6, so part of this new range may have been enabled by learning $R2{:}f_o$ tuning. She noticed a different sound quality, with timbre changes and increased SPL when using $R2{:}f_o$. However, an electroglottograph (model EG-2, Glottal Enterprises, Syracuse, NY) showed none of the large changes in EGG signal that have previously been observed for transitions to the M3 laryngeal register [9].

Possible extensions of visual feedback

An earlier study showed that the provision of visual feedback of *R*1 and *R*2 could help teach vowel pronunciation in a foreign language [23]. The approach described herein – using visual feedback of *R*1 and/or *R*2 – could be used to teach unfamiliar tunings to singers. It could include the *R*1:2*f*o tuning used by female singers in musical theatre [33] and Bulgarian singing [34]. It might also be used to teach the *R*1:2*f*o and *R*1:3*f*o tuning used by some male voices [5,35].

Conclusions

In a one-hour session, six of eight participants learned how to open and close the velum and glottis independently whilst provided with visual feedback. Having done so, these six also learned how to shape the vocal tract for $R2{:}f_o$ tuning, whilst provided with visual feedback. Without visual feedback, none of the six used $R2{:}f_o$ tuning systematically at high pitch: one hour of training was insufficient. However, two of the participants (A2 and A3) spontaneously switched from $R1{:}f_o$ to $R2{:}f_o$ tuning at the very top of their range; at Eb6 and D6, respectively – see Figure 8. After four months using the feedback for about half an hour per day, the first author learned to perform $R2{:}f_o$ tuning without feedback and also extended her comfortable and possible singing range by a fifth or more.

Acknowledgements

The authors thank our volunteer participants and the Australian Research Council for its support.

Disclosure statement

The authors report no conflict of interest.

Notes on contributors

Marie Jeanneteau was an intern in the lab. She is currently completing her Master's Degree in Fluid Mechanics in the French School of Engineering at INP-ENSEEIHT.

Noel Hanna is a leading education professional at UNSW Global. His research has focussed on the fundamental physics of voice production and respiration.

André Almeida is currently a full-time researcher in the School of Physics at UNSW Sydney. He researches on the physics of sound production in reed and lip instruments and the voice.

John Smith is an honorary associate professor in the School of Physics at UNSW Sydney. He currently researches the acoustics of the vocal tract, musical instruments, and their interactions.

Joe Wolfe is a professor of physics at UNSW Sydney. He leads a lab researching music acoustics and voice acoustics. He has also created extensive on-line resources on these topics.

ORCID

Marie Jeanneteau http://orcid.org/0000-0001-5465-2387
Noel Hanna http://orcid.org/0000-0002-3284-3219
André Almeida http://orcid.org/0000-0001-6075-7281
John Smith http://orcid.org/0000-0003-3605-2753
Joe Wolfe http://orcid.org/0000-0001-8030-0561

References

[1] Clark J, Yallop C, Fletcher J. An introduction to phonetics and phonology. Malden (MA): Blackwell; 2007. p. 265.
[2] Sundberg J. Formant technique in a professional female singer. Acustica. 1975;32:89–96.
[3] Sundberg J. The science of the singing voice. DeKalb (IL): Northern Illinois Univ. Press; 1987.
[4] Joliveau E, Smith J, Wolfe J. Acoustics: tuning of vocal tract resonance by sopranos. Nature. 2004;427(6970):116.
[5] Henrich N, Smith J, Wolfe J. Vocal tract resonances in singing: strategies used by sopranos, altos, tenors, and baritones. J Acoust Soc Am. 2011;129(2):1024–1035.
[6] Joliveau E, Smith J, Wolfe J. Vocal tract resonances in singing: the soprano voice. J Acoust Soc Am. 2004;116(4 Pt 1):2434–2439.
[7] Wolfe J, Garnier M, Henrich Bernadoni N, et al. Chapter 5: the mechanics and acoustics of the singing voice: registers, resonances and the source-filter interaction. In: Russo FA, Ilari B, Cohen CJ, editors. The Routledge companion to interdisciplinary studies in singing. Vol. 1. New York: Routledge; 2020.
[8] Garnier M, Henrich N, Smith J, et al. Vocal tract adjustments in the high soprano range. J Acoust Soc Am. 2010;127(6):3771–3780.
[9] Garnier M, Henrich N, Crevier-Buchman L, et al. Glottal behavior in the high soprano range and the transition to the whistle register. J Acoust Soc Am. 2012;131(1):951–962.
[10] Miller DG, Schutte HK. Physical definition of the "flageolet register". J Voice. 1993;7(3):206–212.
[11] Sundberg J, Skoog J. Dependence of jaw opening on pitch and vowel in singers. J Voice. 1997;11(3):301–306.
[12] Smith J, Wolfe J, Henrich N, et al. Diverse resonance tuning strategies for women singers. In: Bresin R, Askenfelt A, editors. Proceedings of the Stockholm Music Acoustics Conference 2013. Berlin, Stockholm: Logos Verlag; 2013. p. 306–310.
[13] ANSI. ANSI S1.1-1994. American National Standards on Acoustics. Melville (NY): Acoustical Society of America; 2004.
[14] Titze IR, Baken RJ, Bozeman KW, et al. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. J Acoust Soc Am. 2015;137(5):3005–3007.
[15] Monsen RB, Engebretson AM. The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear prediction. J Speech Lang Hear Res. 1983;26(1):89–97.
[16] Arroabarren I, Carlosena A. Inverse filtering in singing voice: a critical analysis. IEEE Trans Audio Speech Lang Process. 2006;14(4):1422–1431.
[17] Miller DG, Sulter AM, Schutte HK, et al. Comparison of vocal tract formants in singing and non-periodic phonation. J Voice. 1997;11(1):1–11.

[18] Kallail KJ, Emanuel FW. An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. J Phonetics. 1984;12(2):175–186.

[19] Pham Thi Ngoc Y, Badin P. Vocal tract acoustic transfer function measurements: further developments and applications. J Physique IV. 1994;C5:549–552.

[20] Coffin B. On hearing, feeling and using the instrumental resonance of the singing voice. NATS Bull. 1974;31:26–30. Reprinted in: Coffin B. Historical vocal pedagogy classics. Maryland: Scarecrow Press; 1989. p. 189–200.

[21] Coffin B. Coffin's sounds of singing: principles and applications of vocal techniques with chromatic vowel chart. 2nd ed. Metuchen (NJ): The Scarecrow Press; 1987.

[22] Epps J, Smith J, Wolfe J. A novel instrument to measure acoustic resonances of the vocal tract during speech. Meas Sci Technol. 1997;8(10):1112–1121.

[23] Dowd A, Smith JR, Wolfe J. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. Lang Speech. 1998; 41(1):1–20.

[24] Bunch M, Chapman J. Taxonomy of singers used as subjects in research. J Voice. 2000;14(3):363–369.

[25] Smith JR, Henrich N, Wolfe J. The acoustic impedance of the Boehm flute: standard and some non-standard fingerings. Proc Inst Acoust. 1997;19:315–330.

[26] Music Acoustics. Vocal tract resonances and broadband excitation; 2020. Available from: http://newt.phys.unsw.edu.au/jw/broadband.html

[27] Jeanneteau M, Hanna N, Almeida A, et al. Real time resonance estimation open source software; 2019. https://doi.org/10.5281/zenodo.2616871

[28] Jeanneteau M, Hanna N, Almeida A, et al. Open-source software for estimating vocal tract resonances using broadband excitation at the lips. Proceedings of the 19th International Congress of Phonetic Sciences; Melbourne, Australia; 2019. p. 2971–2975.

[29] Jeanneteau M, Hanna N, Smith J, et al. A real-time feedback system for vocal tract tuning. Proceedings of the International Symposium on Music Acoustics (ISMA 2019); 2019 Sep 13–17; Detmold, Germany; 2019. p. 26–31.

[30] Hanna N, Smith J, Wolfe J. How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips. J Acoust Soc Am. 2018;143(5):2639–2650.

[31] Hawks JW, Miller JD. A formant bandwidth estimation procedure for vowel synthesis [43.72.Ja]. J Acoust Soc Am. 1995;97(2): 1343–1444.

[32] Hanna N, Smith J, Wolfe J. Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation. J Acoust Soc Am. 2016;139(5):2924–2936.

[33] Bourne T, Garnier M. Physiological and acoustic characteristics of the female music theater voice. J Acoust Soc Am. 2012; 131(2):1586–1594.

[34] Henrich N, Kiek M, Smith J, et al. Resonance strategies in Bulgarian women's singing. Logoped Phoniatr Vocol. 2007; 32(4):171–177.

[35] Henrich Bernardoni N, Smith J, Wolfe J. Vocal tract resonances in singing: variation with laryngeal mechanism for male operatic singers in chest and falsetto registers. J Acoust Soc Am. 2014;135(1):491–501.