The Perception of Formant Tuning in Soprano Voices

Rebecca R. Vos(1), Damian T. Murphy(1), David M. Howard(2), and Helena Daffern(1),
(1) *Department of Electronic Engineering, University of York, UK*
(2) *Department of Electronic Engineering, Royal Holloway University of London, UK*

ABSTRACT
Introduction: At the upper end of the soprano range, singers adjust their vocal tract to bring one or more of its resonances (Rn) toward a source harmonic, increasing the amplitude of the sound; this process is known as resonance tuning. This study investigated the perception of (R1) and (R2) tuning, key strategies observed in classically trained soprano voices, which were expected to be preferred by listeners. Furthermore, different vowels were com-pared, whereas previous investigations have usually focused on a single vowel.

Methods: Listeners compared three synthetic vowel sounds, at four fundamental frequencies (f0), to which four tuning strategies were applied: (A) no tuning, (B) R1 tuned to f0, (C) R2 tuned to 2f0, and (D) both R1 and R2 tuned. Participants compared preference and naturalness for these strategies and were asked to identify each vowel.

Results: The preference and naturalness results were similar for /ɑ/, with no clear pattern observed for vowel identification. The results for /u/ showed no clear difference for preference, and only slight separation for naturalness, with poor vowel identification. The results for /i/ were striking, with strategies including R2 tuning both preferred and considered more natural than those without. However, strategies without R2 tuning were correctly identified more often.

Conclusions: The results indicate that perception of different tuning strategies depends on the vowel and perceptual quality investigated, and the relationship between the formants and (f0). In some cases, formant tuning was beneficial at lower f0s than expected, based on previous resonance tuning studies.

KEYWORDS: soprano, resonance, tuning, perception, formant

Introduction

In female speech, the first and second formants typically lie between 310 and 860 Hz (D#4 - A5) and 920 and 2790 Hz (A5 – F7), [1]. The soprano range can extend above 1000 Hz, so there are frequencies at which the fundamental frequency (*f*0) may exceed the frequency range of one or both of the first two formants. Where this occurs, the absence of acoustic energy in the lower resonances' frequency ranges causes sound production to be less efficient, and because the first three to five formants are considered the most important for the perception of vowels this causes vowels to become harder to identify [2]. The wide spacing of harmonics at high *f*0 is also thought to contribute to the increasing inaccuracy of vowel perception with rising *f*0 [3].

Formant tuning

One strategy used by singers to increase the efficiency of the voice at high f0 values is known as formant tuning or resonance tuning [4], whereby the singer adjusts the shape of the vocal tract to change the frequencies of one or more of its first resonances. Altering the position of the first or second resonances (*R*1 and *R*2) increases the acoustic power transmitted by the voice, not only by ensuring that there is acoustic energy present in the frequency range of a vocal tract resonance, but also by matching the acoustic impedance of the source (glottis) and the filter (vocal tract) to produce a perceptually louder sound with less effort from the singer [5, 6].

It is well documented that classical male singers commonly converge formants 3, 4, and 5 [7] to create the singer's formant cluster, which increases the spectral energy in the region around 3 kHz [4] where the human ear is most sensitive [8]. Evidence of a true singer's formant cluster in sopranos, however, is extremely limited, and it would not necessarily provide the same acoustic benefits as for low voices. As sopranos sing at extremely high *f*0 values, there is already a considerable amount of spectral energy in this region due to the presence of high-amplitude early harmonics [9].

Sundberg [10] proposed that soprano singers were able to tune one or both of the first two vocal tract resonances near the harmonics of the voice source. This would enable the singer to make full use of the vocal tract resonances even at high fundamental frequencies, and increase the acoustic output power by increasing the vocal efficiency rather than requiring increased effort from the singer. Since then, studies on soprano singers have con-firmed evidence of resonance tuning, which is achieved by adjusting the shape of the vocal tract. An experiment by Garnier et al [11] investigated the resonance tuning strategies used by sopranos across their range. The study involved 12 sopranos (4 non-experts, 4 advanced, and 4 professionals) singing /ɑ/ vowels. They found that $R1$:$f0$ tuning was employed by all the professionals and advanced singers, and to a lesser extent by the non-expert singers. $R2$:2$f0$ tuning was seen in three professionals, two advanced, and two non-expert singers. Six of the singers used $R2$:$f0$ tuning at very high $f0$ values (above C6), and $R1$:2$f0$ tuning was only found in two of the singers (in the lower part of the range investigated).

It is now generally accepted that opening the jaw raises the first resonance [12], whereas the second resonance is controlled by changing the position of the tongue [13]. Shortening the vocal tract slightly by smiling raises all the resonance values [14].

Disadvantages of formant tuning

Although resonance tuning is an accepted phenomenon in soprano singing [10, 11, 15], and acoustic theory suggests vowel recognition would greatly diminish at high fundamental frequencies [3], in practice there is still some debate as to whether singers should “neutralize” vowels at high fundamental frequencies, choosing to focus on the sound quality produced, rather than the perceptual distinction between vowels, or make a special effort to keep them distinct, but potentially sacrifice some acoustic efficiency and ease of production [16].

The perception of resonance tuning

Although there is now clear evidence of the *practice* of resonance tuning [5,11,15], there is a lack of research into its *perception*. There have been a small number of studies on the perception of vowels at high frequencies [3, 17] that show that the likelihood of a sung vowel being misunderstood increases as a function of $f0$.

In 1991, Carlsson, Berndtsson and Sundberg published a perceptual study [18] in which synthesized singing tones were generated to represent a male voice, at fundamental frequencies ranging over a descending octave-wide chromatic scale from C4 (261 Hz) to C3 (131 Hz), representing the vowel /ɑ/. These tones were then treated in one of four ways. In “strategy A,” the first formant was tuned to the harmonic closest to 550 Hz. In “strategy B,” the second formant was tuned to the harmonic lying closest to 1000 Hz. In “strategy C,” either the first or second formant was tuned to the harmonic closest to 550 or 1000 Hz, depending on which option gave the smallest formant frequency deviation from these values. Finally, in “strategy D,” the formants remained at 550 and 1 kHz in all tones.

Sounds with tuned formants (using strategies A, B, or C) were presented together with the non-tuned tones (strategy D) in pairs, and 19 listeners were asked, “Which voice production do you find most correct?”

The tones with unchanged formant frequencies were preferred by all but one subject. The mere-exposure effect [19] (the psychological phenomenon whereby people prefer stimuli that they are more familiar with) could contribute to these findings, as due to the pairing methods used, subjects heard the sounds with unchanged tuning three times more often than the other tuning strategies. The protocol used in this study alters that used by Carlsson, Berndtsson and Sundberg, [18] to be suitable for the soprano voice, and removes the possibly confounding influences of the mere-exposure effect.

Based on the evidence of $R1$:$f0$ and $R2$:2$f0$ tuning by sopranos [11], the perception of these tuning conditions is investigated in this paper. The properties investigated include which tuning strate-gies are *preferred*, their *naturalness*, and which produce the mostly clearly *identifiable* vowel sounds. The hypothesis is that the strategies used most frequently by sopranos in practice will be preferred by subjects, perceived to be most natural, and correctly identified most often.

METHODS

Similar to the procedure used by Carlsson, Berndtsson and Sundberg [18], synthesized tones were created to replicate voiced sounds, for which the resonance frequencies could be con-trolled to represent different resonance tuning strategies. Tones with $f0$ typical for a soprano range were synthesized, and as resonance values have been shown to remain constant in singing up to the frequency where $f0$ = $F1$ [18] the average formant values in speech for women’s voices were used for the baseline resonance values (as defined by Peterson and Barney [1]). These are shown for the three vowels investigated in Table 1. As in [18], four resonance tuning strategies were tested:

- In “strategy A,” no resonance tuning is used, so the vowel resonances remain constant at the average values for the vowel.
- In “strategy B,” the first resonance is tuned to the fundamental, whereas the second and third resonances are kept constant at the average values for the vowel.
- In “strategy C,” the second resonance is tuned to the second harmonic, whereas the first and third resonances are kept constant at the average values for the vowel.
- In “strategy D,” the first resonance is tuned to the fundamental, and the second resonance is tuned to the second harmonic, whereas the third resonance is kept constant at the average value for the vowel.

TABLE 1
The First Three Formant Values for Three vowels, When Spoken by Female Voices

Vowel	F1	F2	F3
/a:/	850 Hz (G#5)	1220 Hz (D6)	2810 Hz (F7)
/u:/	370 Hz (F#4)	950 Hz (A#5)	2670 Hz (E7)
/i:/	310 Hz (D#4)	2790 Hz (F7)	3310 Hz (G#7)

Synthesized signal Glottal signal
The synthesized vowel sounds are produced using a Liljencrants-Fant glottal flow model to create a glottal signal. Typical parameter values for a female were used, from Reference 20, the details of which are given in the Appendix. Vibrato is also added to the voice source in order to make it sound more naturally sung than spoken. This consists of a 6 Hz sinusoidal modulation of the fundamental frequency, with an extent of 60 cents [21].

Vocal tract effects
The resonances of the vocal tract were treated as a series of connected single peak infinite impulse response (IIR) filters, using the "IIR peak" function in *MATLAB* (version R2016a, Natick, Massachusetts, The MathWorks Inc., (2016), and the glottal signal was passed through each filter in turn. The values used for the resonances are the formant values shown in Table 1 [1] with the bandwidths fixed at 50 Hz, noting that a study investigating formant bandwidth [72] which used averaged data from Fujimura and Lindqvist [23] and Fant [24] found that the bandwidth remains approximately constant at around 50 Hz for formant frequencies between 300 and 2000 Hz.

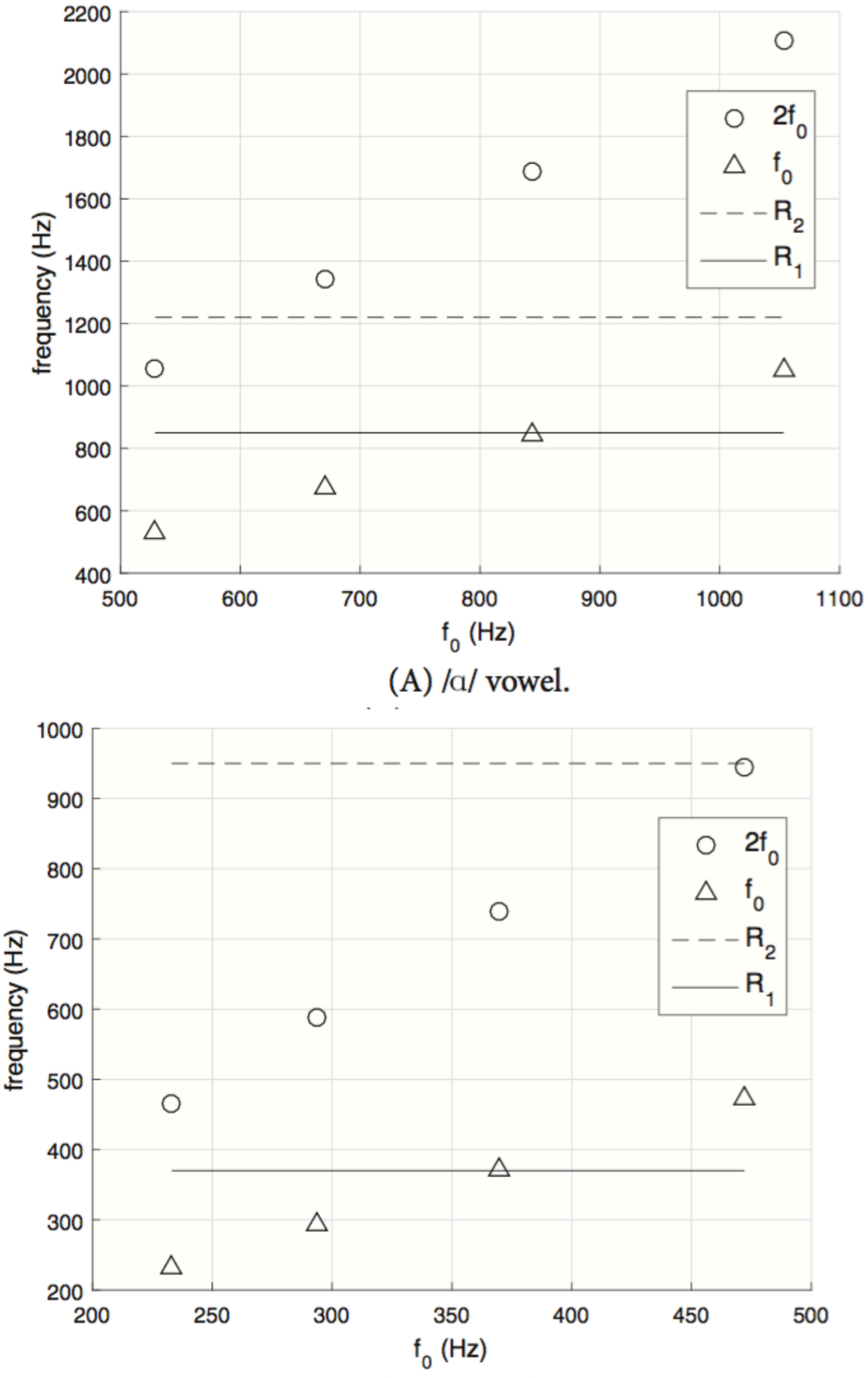
TABLE 2
The fundamental frequencies of the synthesised tones for each vowel sound

Pitch number vowel	1	2	3	4
/a:/	C5 529 Hz	E5 671 Hz	G#5 843 Hz	C6 1052 Hz
/u:/	A#3 233 Hz	D4 294 Hz	F#4 370 Hz	A#4 472 Hz
/i:/	A3 220 Hz	C#5 277 Hz	F4 349 Hz	A4 440 Hz

The resulting synthesized signal was then de-emphasized (attenuating the higher frequencies) so that the relative resonance amplitudes more closely resemble the human voice. The fundamental frequencies are chosen to be either side of the first resonance, as shown in Table 2.

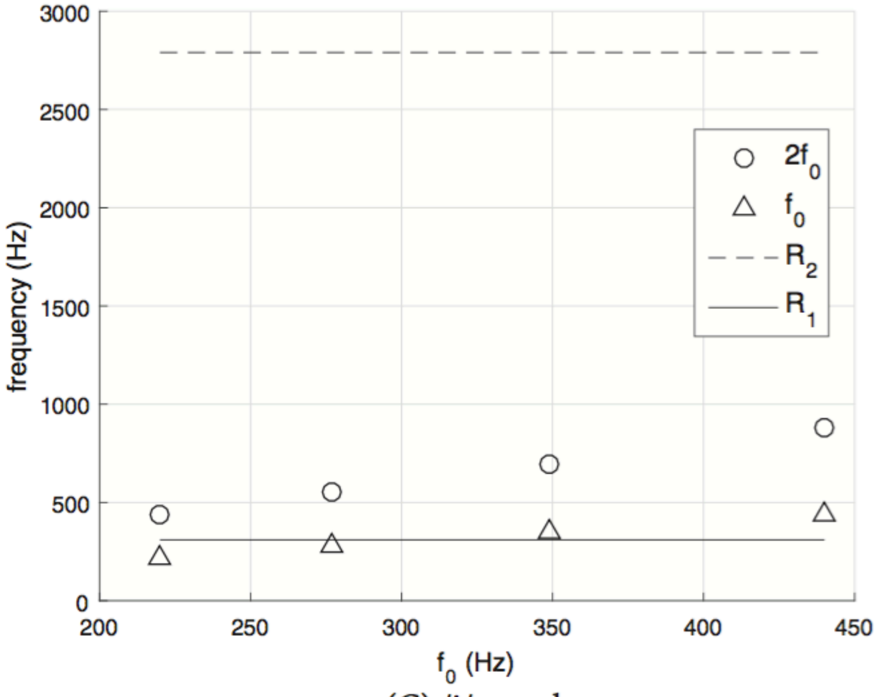
In order to make the synthesized voice sound more natural, and to prevent transient effects due to the sudden onset and offset of the sound, an amplitude window is applied, consisting of the relevant halves of a Hanning window in the first and last quarter of each tone.

In practice, a vocal tract resonance at a frequency just above a harmonic produces an inertive reactance, causing the vocal tract to assist the vibration of the vocal folds, which results in an increased acoustic power output. Conversely, when a vocal tract resonance is slightly below a harmonic, there is a compliant reactance, and the vocal tract no longer assists the vibration of the vocal folds, resulting in a reduced acoustic power output [25]. Therefore, to maximize the impact of resonance tuning, vocal tract resonances are tuned to just above the relevant harmonic frequencies. The relationship between the resonances and harmonics can be seen in Figure 1, where the harmonics are plotted against fundamental frequency, and the formant values in speech (the un-tuned values for R_1 and R_2) are represented by horizontal lines.



(A) /ɑ/ vowel.

(B) /u/ vowel.



(C) /i/ vowel.

Figure 1. Shows the values of the first and second formants in speech (*solid and dashed lines*, respectively) and the values of f_0 and $2f_0$ (first and second harmonics) for each vowel (triangle and circle respectively).

Subjects and distribution

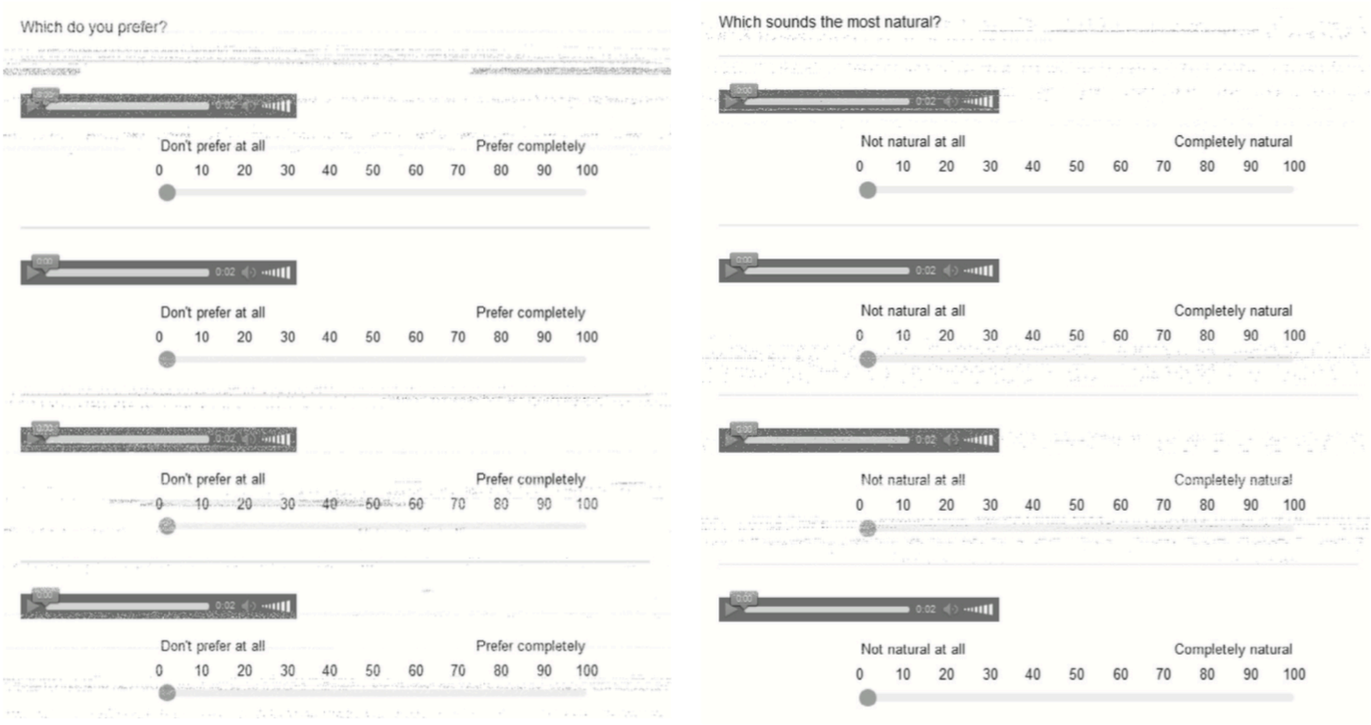
The listening test was distributed via e-mail and social media, and used the online survey software *Qualtrics* [26]. Forty five subjects took part; however, the results from 15 of these were discarded, either because they did not complete the entire test or because they reported serious hearing problems. Of the remaining 30 participants, 20 identified as male and 8 as female. They were aged 20– 75, with an average age of 33.7 years. The time taken (including breaks) varied from 13 minutes to 73 minutes (discounting two outliers), with an average time of 32 minutes.

Subjects were able to take the listening test on their own devices (excluding mobile devices). Fifteen subjects used closed-back headphones, seven used open-backed headphones and seven used earbuds. Subjects were instructed to take the test in a quiet environment with no distractions, and not to adjust the volume on their computer after starting the test. There may have been slight differences in audio quality between subjects; however, internet distribution allowed a greater number and variety of subjects to participate in the test, so was considered worthwhile. Schoeffler et al compared laboratory and web-based results of an auditory experiment and found no significant differences [27], demonstrating that this is an acceptable distribution method.

Procedure

Subjects first answered a questionnaire to ascertain demographic information, their level of vocal ability, singing training, and their music listening habits. This captured the subjects' own singing ability, as well as their experience of listening to professional singing. Nine subjects had some singing training (four of whom had professional training).

The listening test consisted of comparisons between sets of four tones using sliders. Each set contained tones with the same $f0$ and vowel, but treated with the four different tuning strategies: A, B, C, and D. The subjects could press the buttons to play the tones as many times as they wished. Each set of four tones was presented in a random order, and the order of tones presented in each question was also randomized to minimize the effects of program dependence. The three sets of questions considered the following perceptual aspects, *preference*, *naturalness*, and *vowel identification*.



(A) An example from the set of questions on *preference*.

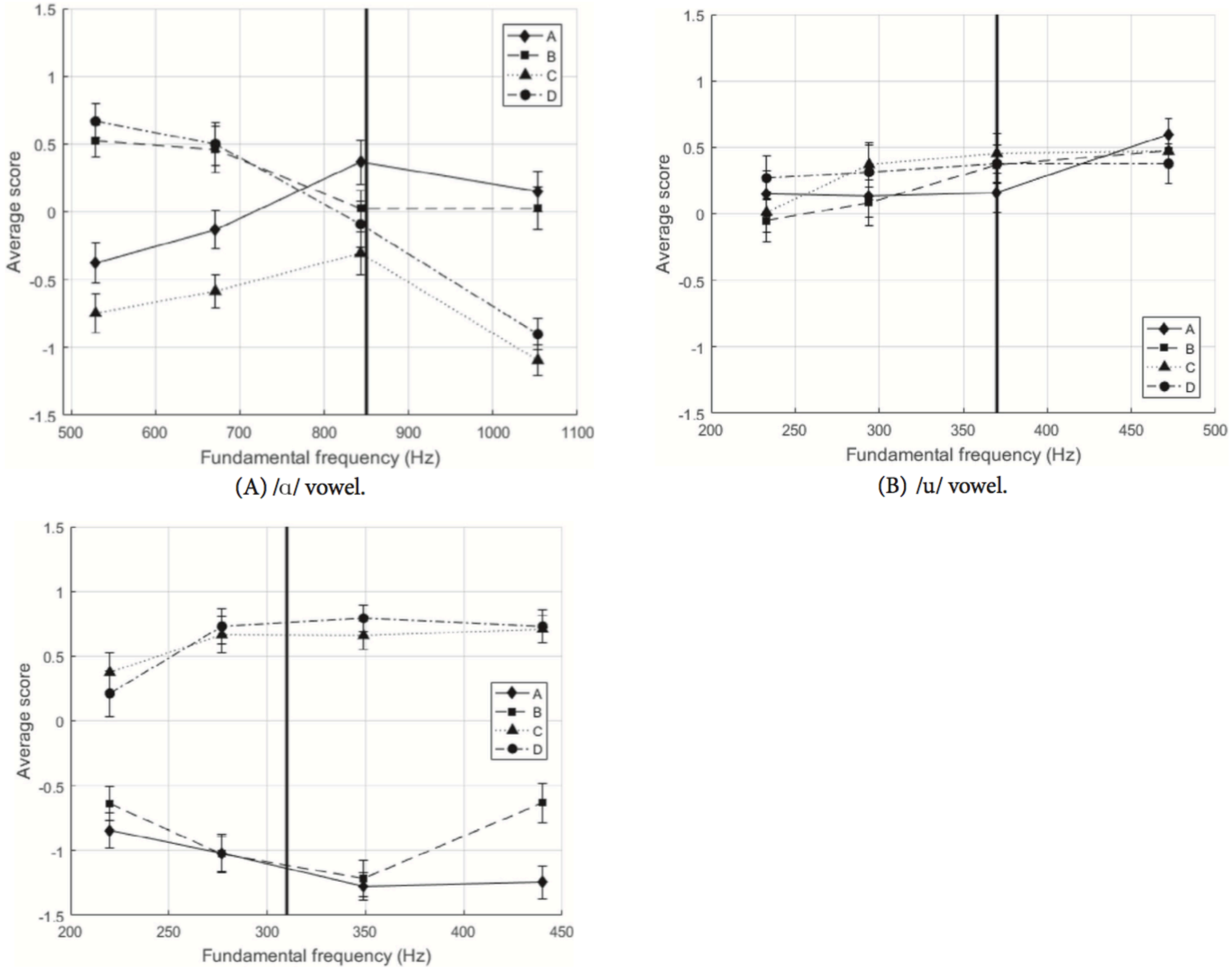
(B) An example from the set of questions on *naturalness*.

Figure 2. Shows the layout of the questions presented to participant on *preference*, *naturalness*, and *vowel identification*.

Examples of the three sets of questions are shown in Figure 2. Prior ethical approval was gained from the Physical Sciences Ethics Committee at the University of York.

RESULTS

Data collected from the questionnaire, together with the listening test answers, were collected in *Excel*, and then imported into *MATLAB* for analysis. Participants were asked to rate preference and naturalness on continuous sliding scales from 0 to 100, with 100 indicating the highest preference or naturalness. The resulting scores were first normalized to have a mean of 0 and a standard deviation of 1 across each participant, to reduce inter-subject variability. The mean score and the standard error of the mean across all participants were then calculated for each vowel, $f0$ and tuning strategy, so that the average normalized score could be plotted against $f0$ for each vowel.

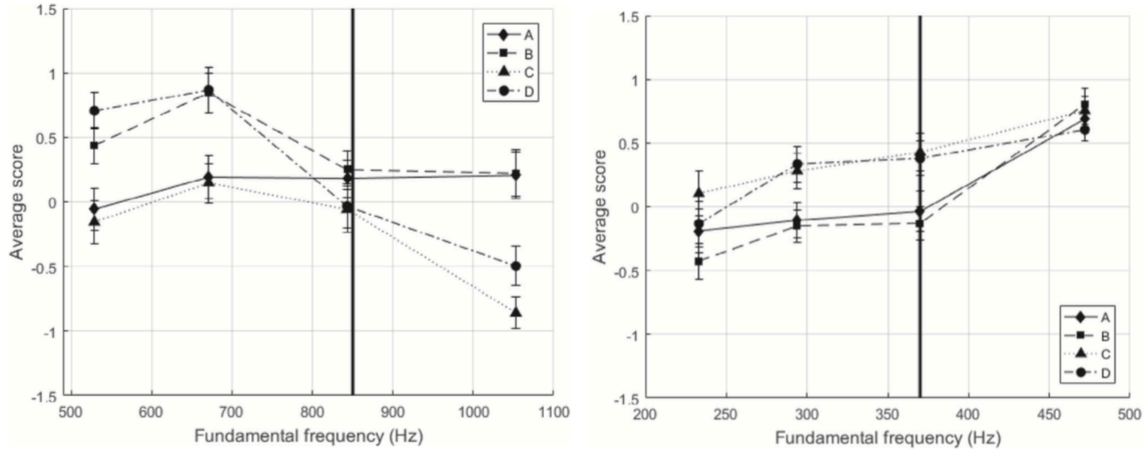


(A) /ɑ/ vowel.

(B) /u/ vowel.

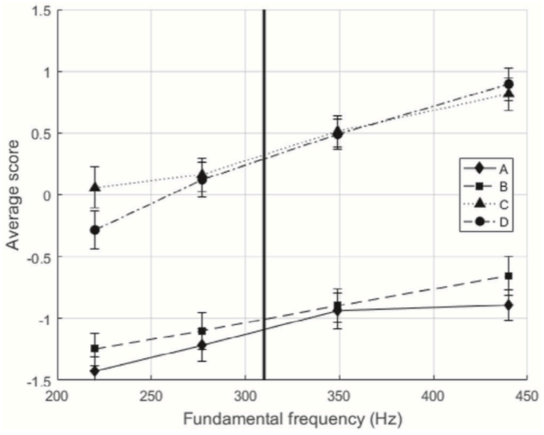
(C) /i/ vowel.

Figure 3. Shows the average scores for the different tuning strategies investigated for *preference*, with the standard error of the mean shown by error bars. The *thick vertical line* shows the frequency of the first formant in speech.



(A) /ɑ/ vowel.

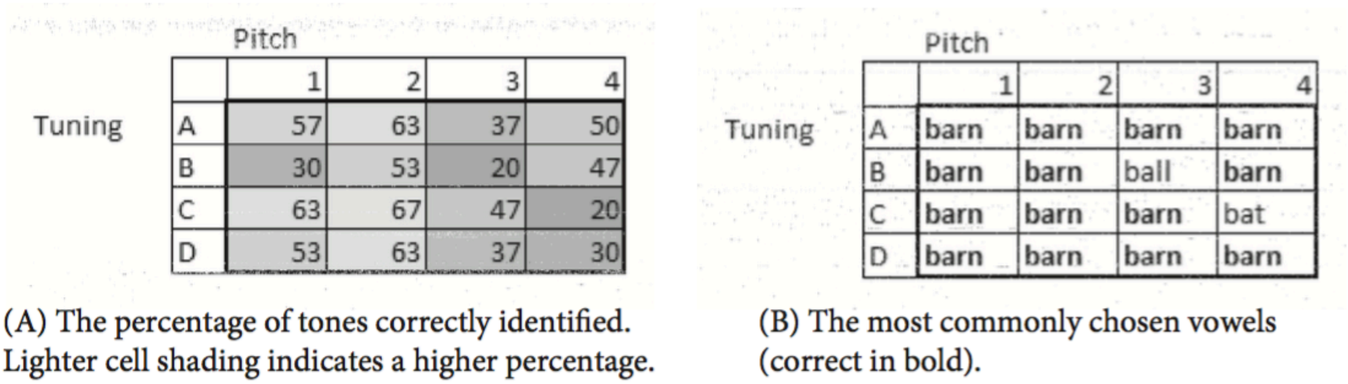
(B) /u/ vowel.



(C) /i/ vowel.

Figure 4. Shows the average scores for the different tuning strategies investigated for *naturalness*, with the standard error of the mean shown by error bars. The *thick vertical line* shows the frequency of the first formant in speech.

The results for preference and naturalness are shown in Figures 3 and 4, respectively. The question on vowel identification was analyzed by calculating the percentage of subjects who chose the correct vowel sound for each sound.



Pitch

Tuning	1	2	3	4
A	57	63	37	50
B	30	53	20	47
C	63	67	47	20
D	53	63	37	30

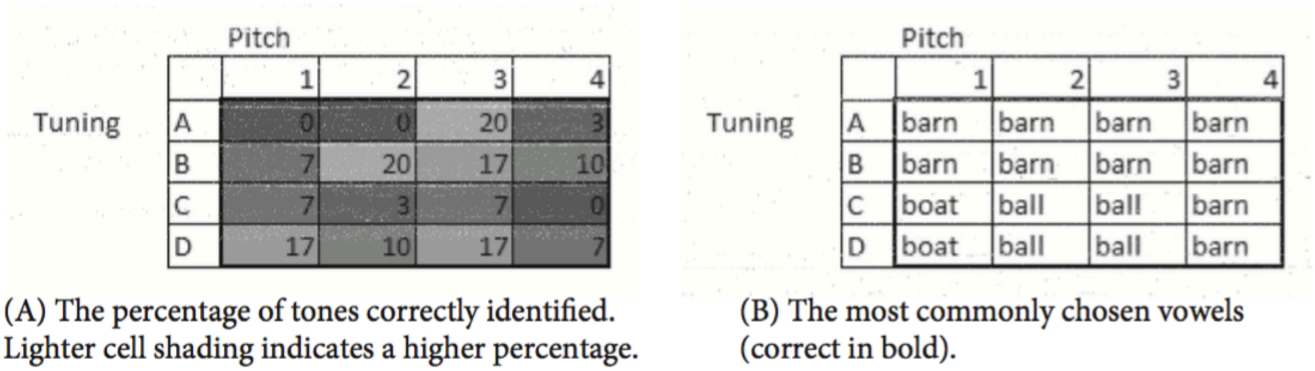
(A) The percentage of tones correctly identified. Lighter cell shading indicates a higher percentage.

Pitch

Tuning	1	2	3	4
A	**barn**	**barn**	**barn**	**barn**
B	**barn**	**barn**	ball	**barn**
C	**barn**	**barn**	**barn**	bat
D	**barn**	**barn**	**barn**	**barn**

(B) The most commonly chosen vowels (correct in bold).

Figure 5: Vowel identification results for the /a:/ vowel.



Pitch

Tuning	1	2	3	4
A	0	0	20	3
B	7	20	17	10
C	7	3	7	0
D	17	10	17	7

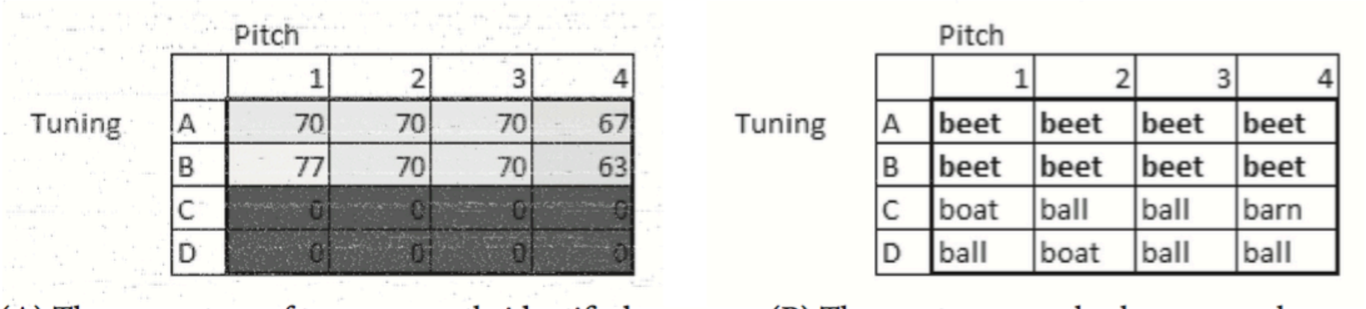
(A) The percentage of tones correctly identified. Lighter cell shading indicates a higher percentage.

Pitch

Tuning	1	2	3	4
A	barn	barn	barn	barn
B	barn	barn	barn	barn
C	boat	ball	ball	barn
D	boat	ball	ball	barn

(B) The most commonly chosen vowels (correct in bold).

Figure 6: Vowel identification results for the /u:/ vowel.



Tuning \ Pitch	1	2	3	4
A	70	70	70	67
B	77	70	70	63
C	0	0	0	0
D	0	0	0	0

(A) The percentage of tones correctly identified. Lighter cell shading indicates a higher percentage.

Tuning \ Pitch	1	2	3	4
A	**beet**	**beet**	**beet**	**beet**
B	**beet**	**beet**	**beet**	**beet**
C	boat	ball	ball	barn
D	ball	boat	ball	ball

(B) The most commonly chosen vowels (correct in bold).

Figure 7: Vowel identification results for the /i:/ vowel.

These values are shown in Figures 5A, 6A, 7A for each vowel and the most commonly chosen vowel sound is shown in Figures 5B, 6B, 7B.

/a:/ vowel

The results for the /a:/ vowel are similar for preference and naturalness, with strategies with $R1$ tuning (B and D) scoring highest at $f0$ values below $R1$, but strategies without $R2$ tuning (A and B) scoring highest at higher fundamental frequencies, and no clear relationship between tuning strategy and vowel identification.

The results for the vowel identification for the /a:/ vowel show that at $f0$ below $R1$ strategy, C ($R2$ tuning only) scored the highest, with strategies A and D (no tuning and both resonances tuned) just below. Strategy B ($R1$ tuning) was the most commonly misidentified. At $f0$ values above $R1$, no tuning (A) was the most correctly identified, and $R2$ tuning (C) the least.

/u:/ vowel

The results for the /u:/ vowel do not appear to show a clear difference between the different tuning strategies for preference; however, there is some separation for naturalness with strategies, with $R2$ tuning (C and D) scoring highest in the middle of the $f0$ range investigated. The vowel identification was generally very poor for this vowel (only 9% correct on average). There did not appear to be a clear pattern in these results, although tuning strategies involving $R2$ tuning (C and D) scored a little lower than those without (A and B) at most $f0$ values. Even the un-tuned tones were mostly incorrectly identified for the /u/ vowel. However, subjects were allowed to choose from 12 different vowel sounds, and the most often chosen vowel sounds were similar to the intended vowel (adjacent on the international phonetic alphabet (IPA) diagram—Figure 9). Where sounds were not identified as the intended vowel, the results for preference and naturalness are still valuable, as the subject was not told the intended vowel, and was simply asked to choose which sound he/she preferred or found the most natural. Considering these results compared with the other vowels seems to suggest that the /u/ vowel (the most closed and back vowel) is unusual, and perhaps fundamentally more difficult to identify or synthesize.

/i:/ vowel

The results for the /i:/ vowel are more revealing than the other vowels, with strategies with $R2$ tuning (C and D) scoring much higher than strategies without $R2$ tuning (A and B) for both preference and naturalness. However, this effect is reversed for the vowel identification, with approximately 70% of the tones without $R2$ tuning correctly identified, but none of the tones with $R2$ tuning.

Analysis of variance

The results for the questions on preference and naturalness are split by vowel, and analysis of variance (ANOVA) is carried out in *MATLAB*. The variables considered are tuning strategy (A, B, C, or D) and fundamental frequency. An interaction model is used to determine whether the variables interact significantly. Figure 8 shows the p values for each vowel, for both preference and naturalness questions. The chosen significance level was 5% ($p = 0.05$), and significant results are highlighted in gray. The ANOVA results for the questions on *preference* show that there was a significant difference between the results for different tuning strategies as well as different $f0$ values for the /ɑ/ vowel. There was also a significant interaction between these two variables, significant results were seen, which supports what is seen in Figure 3B, that is no clear pattern in the results. For the /i:/ vowel, there was a significant difference between tuning strategies, but not $f0$ values (and no interaction). Again, this supports what is seen in Figure 3C, a clear difference between the different tuning strategies, but no great variation in the results across fundamental frequencies.

For the naturalness results, no interaction between the variables was seen for any vowel, so the effects of tuning strategy and $f0$ can be considered separately. The results for all three vowels were the same: all three showed a significant difference in naturalness both between tuning strategies and fundamental frequencies.

These results imply that both the tuning strategy and $f0$ have a significant effect on the perception of synthesized singing sounds for *preference* and *naturalness*, although the exact relationship varies between vowels.

Discussion

In this section, the results for each vowel will be discussed, first in respect to the preference questions, then naturalness, and finally for vowel identification.

Preference

From Figure 3A, it can be seen that for the /a:/ vowel, at the lower two $f0$ values, strategies with $R1$ tuning (B and D) were preferred above strategies without $R1$ tuning (A and C). The four tuning strategies all scored similarly when $f0$ was equal to $R1$; however, when $R1$ was above $f0$ the results differ, with strategies without $R2$ tuning (A and B) preferred over those with $R2$ tuned (D and C). $R1$ tuning only (B) scored highly across the whole range of $f0$ values, which is indeed the method used most often by sopranos in this range.[11] $R2$ tuning only (C) scored the lowest across the whole range of $f0$ values, indicating that it was the least preferred tuning strategy. This is not surprising at lower fundamental frequencies, because $R2$ tuning is rarely observed in that region; however, above the normal range of $R1$ tuning, $R2$ tuning has been observed, although more commonly in conjunction with $R1$ tuning [11].

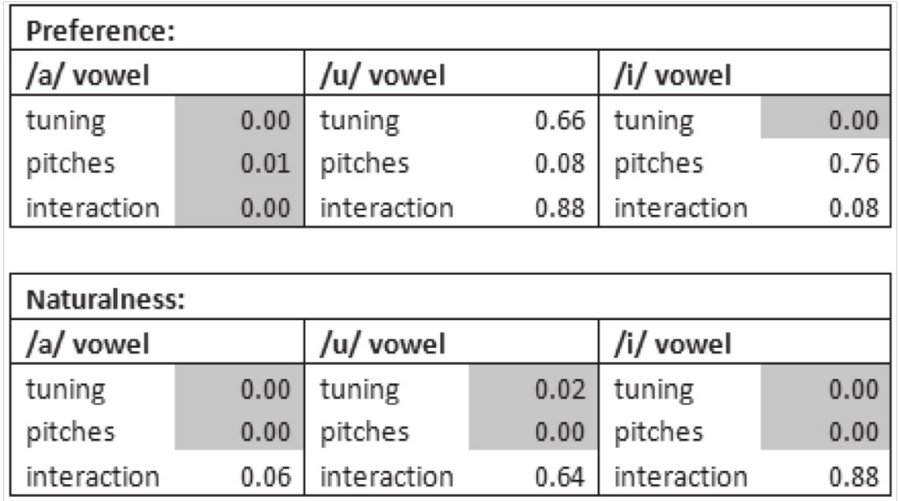
Interestingly, the results for the /u/ vowel (Figure 3B) show no significant difference in preference scores between the four tuning strategies used. There is a slight increase in score with $f0$ for all tuning strategies, which could simply indicate that the subjects preferred the higher pitched sounds, or that difficulty identifying vowel sounds might play a part. The ANOVA results (see figure 8) support this, indicating that for preference, neither tuning nor fundamental frequency had a significant effect.

For the /i/ vowel (Figure 3C), strategies with $R2$ tuning (C and D) were preferred over those without it (A and B) across all $f0$ values. The second formant for this vowel is very high (2790 Hz) compared with that of the other two vowels investigated (1120 Hz and 950 Hz for /ɑ/ and /u/, respectively). Therefore, when $R2$ is tuned to either the first or second harmonic, this represents a considerable increase in the amount of energy in the lower part of the spectrum, compared with an un-tuned $R2$. The very high scores in preference for tuning strategies with $R2$ tuning (C and D) indicate that this increase in low-frequency energy was preferred by listeners, which suggests that, in practice, listeners would prefer singers to lower the second resonance to similar frequencies as the other vowels. This lack of preference for un-tuned second resonances supports the evidence that at very high fundamental frequencies, professional singers often employ this technique [11], and that “sympathetically” written music may well take this into account, using vowels with lower formant values at high fundamental frequencies such as an /ɑ/ vowel [28].

Naturalness

From Figure 4A, as for preference, it can be seen that for the /ɑ/ vowel, strategies involving $R1$ tuning (B and D) were considered the most natural at $f0$ values below $R1$. However, as $f0$ rose above $R1$, the perceived naturalness of strategy D ($R1$ and $R2$ tuning) decreased, whereas strategy A(no tuning) remained roughly constant, so that at higher $f0$s strategies without $R2$ tuning (A and B) were perceived as more natural than those with $R2$ tuning (C and D). These results are surprising as they do not reflect the resonance tuning methods known to be used by singers for this vowel [11].

Although the current study only used synthesized samples, it is possible that as most of the subjects were not highly trained singers or listeners, they were not used to the timbre of opera, and therefore found the usual resonance tuning techniques used in opera (e.g., $R1$:$f0$) unnatural in general. Indeed, Smith and Wolfe [28] suggest that subjects who often listen to a certain type of vocal production, for example classical singing, may learn to use a different “formant map” for sopranos, giving them their own categorization of the vowel plane. In addition to this, “naturalness” is of course a subjective term, and in this experiment the subjects were left to decide for themselves what it meant, so there may have been some variation in this between subjects.



Preference:					
/a/ vowel		/u/ vowel		/i/ vowel	
tuning	0.00	tuning	0.66	tuning	0.00
pitches	0.01	pitches	0.08	pitches	0.76
interaction	0.00	interaction	0.88	interaction	0.08

Naturalness:					
/a/ vowel		/u/ vowel		/i/ vowel	
tuning	0.00	tuning	0.02	tuning	0.00
pitches	0.00	pitches	0.00	pitches	0.00
interaction	0.06	interaction	0.64	interaction	0.88

Figure 8: The *p* values from the analysis of variance results for preference and naturalness questions. Significant results are highlighted in *gray*.

For naturalness, as for preference, all four tuning strategies scored similarly for the /u/ vowel (Figure 4B). There was, however, some separation for the middle two *f*0 values, with strategies involving *R*2 tuning (C and D) scoring a little higher than those without (A and B). This is supported by the ANOVA results (Figure 8), which show that for naturalness, both tuning and fundamental frequency had a significant effect.

The results for both the preference and naturalness questions for the /i/ vowel are somewhat unexpected, considering that *R*2 tuning in isolation at these fundamental frequencies has not often been observed [11, 29]. However, these results must be considered in conjunction with the vowel identification results, in that the subjects were simply asked how natural the sounds were, but not told which vowel sounds they represented. It seems that the subjects found the sounds with *R*2 tuning more preferable and natural than those without, but not very well identified as an /i:/ vowel.

For the /i/ vowel (Figure 4C), tuning methods involving *R*2 tuning (C and D) consistently scored the highest, followed by those without (A and B). The average scores for naturalness remained fairly stable at all *f*0 values, and again a general increase in naturalness with *f*0 was seen. As for preference, these results suggest that lowering the high second formant has the greatest effect on naturalness, irrespective of whether *R*1 is tuned.

Vowel identification

The results for the /a:/ vowel (Figure 5) show that at *f*0 values below *R*1, strategy C (*R*2 tuning) scored the highest, with A and D (no tuning and both resonances tuned) just below. Strategy B (*R*1 tuning) was the most commonly misidentified. At *f*0 values above *R*1, this pattern changed to a completely different order (similar to preference and naturalness), with A the most correctly identified, and C the least. The average percentage of sounds correctly identified across all *f*0 values and tuning strategies was 46% (with a standard deviation of 16%).

The results for the /u:/ vowel (Figure 6) show that this vowel was correctly identified much less frequently than the /ɑ/ vowel (only 9% correct on average, with a standard deviation of 7%). There did not appear to be a clear pattern in these results, although tuning strategies involving *R*2 tuning (C and D) scored a little lower than those without *R*2 tuning (A and B) at most *f*0 values. This could be due to the importance of the position of the second formant in distinguishing this vowel, meaning that at all *f*0 values, tuning of *R*2 distorted the vowel sound. Tuning strategies A and B were most commonly identified as an /ɑ/ vowel across all *f*0 values; however, strategies with *R*2 tuning (C and D) were most commonly identified as /o/ (as in "boat") at the lowest *f*0, /ɔ/ (as in "ball") at the middle two *f*0 values, and /ɑ/ at the highest *f*0. This suggests that tuning *R*2 causes the vowel to sound more open (Figure 9); however, the poor identification of even the un-tuned sample suggests that there may have been issues with the synthesis of this vowel sound.

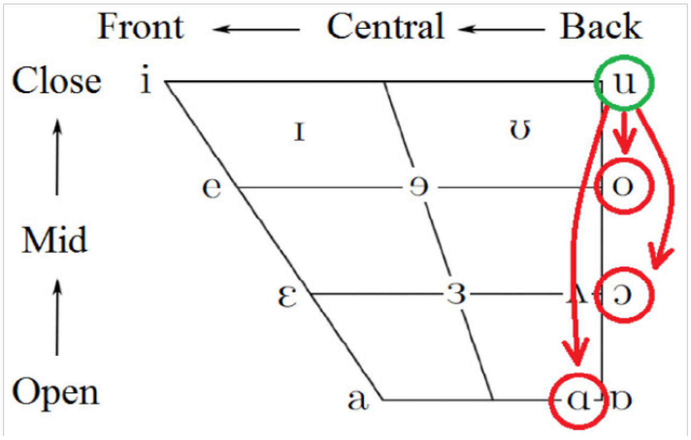


Figure 9: A simplified map of the IPA monophthong vowels and the ways in which the /u:/ vowel (top right) was most commonly misidentified.

The results for the /i:/ vowel (Figure 7) show a very clear pattern, where strategies without $R2$ tuning (A and B) were correctly identified in around 70% of tones (with a standard deviation of 4%); however, strategies with $R2$ tuning (C and D) were never correctly identified. One explanation of this might be provided by Benolken and Swanson [17], who suggest that some vowels that have similar first formant values, like the /i:/ and /u:/ vowels (only 60 Hz apart), are differentiated by their second formants, so altering the second formant results in a dramatic loss in identifiability. The sounds with $R2$ tuning (C and D) were most commonly identified as /ɔ/ (as in "ball"), /o/ (as in "boat"), or /ɑ:/ (as in "barn"), showing that the perceived vowel sound changed from front to back (Figure 9).

Overall impressions

There were marked and unexpected differences between the results for the three vowels for the three perceptual attributes investigated. The /i:/ vowel produced the most notable differences across tuning strategies for all three perceptual attribute, with strategies involving $R2$ tuning scoring the highest for both preference and naturalness but the lowest for vowel identification. Based on the findings of Henrich et al, [30], Carlsson, Berndtsson and Sundberg [18], and Sundberg [4]. it was predicted that the strategy with no resonance tuning (A) would score the highest for all three of the perceptual attributes investigated at fundamental frequencies below the first resonance, as there is little evidence of singers using resonance tuning within this frequency range. However, the opposite of this was found: at $f0$ values below $R1$, strategy A was generally one of the lowest scoring, whereas stra-egy D (both resonances tuned) scored highly for both preference and naturalness. The results, therefore, suggest that for certain vowel sounds, if physically possible, it might be beneficial to employ resonance tuning over a wider range of fundamental frequencies than had previously been thought. At fundamental frequencies below the first resonance, *lowering* $R1$ slightly to coincide with the fundamental would increase the acoustic power transmitted, therefore reducing the effort required by a singer to communicate effectively to an audience.

At fundamental frequencies above $R1$, it was expected that $R1$:$f0$ tuning (strategy B) would score highly for all three perceptual attributes, as this is the most commonly observed in practice, and $R2$:2$f0$ tuning (strategy C) would score the lowest, as it is rarely observed in isolation [30]. Indeed, Wolfe et al. [6] suggest that $R2$ tuning might be unintentional, based on the theory that as the fundamental frequency rises, $R1$ is tuned to the fundamental by increasing the opening of the mouth, and as both $R1$ and $R2$ rise with increased mouth opening, $R2$ is raised as a side effect of raising $R1$. This would suggest that $R2$ tuning in isolation (C) should score quite low for both preference and naturalness; however, for some vowels and $f0$ values, this was not the case. For example, for preference, $R2$ tuning (C) scored highly for the /i:/ vowel. However, the second resonance is known to be very sensitive to changes in the shape of the tongue [31], so it is possible that listeners perceived the differences in the sounds as due to different tongue shapes.

strategies without $R2$ tuning (A and B) behaving similarly, as well as strategies with $R2$ tuning (C and D). This seems to suggest that the presence or absence of $R2$ tuning had the greatest influence on the listeners' perception of the sounds, and further investigation is required to fully understand this result.

Although most previous studies have focused on single vowels (most commonly /ɑ/), this study found that the rankings of different tuning strategies are highly dependent on the vowel, as extremely different patterns are observed across the three vowels investigated, /ɑ:/, /i:/, and /u:/. In addition to this, resonance tuning (by any of the three strategies investigated here) does not necessarily improve the *preference*, *naturalness*, or *vowel identification*, as in some cases strategy A (no tuning) scored the highest, even at fundamental frequencies above $R1$. For example, for the /i:/ vowel, no tuning (A) scored lower than the other tuning strategies for naturalness and preference, but improved the *vowel identification*. In addition to this, some tuning strategies might improve one perceptual quality, while having little effect on or detracting from another quality. For example, $R1$ tuning alone
(B) scored poorly for both preference and naturalness for the /i/ vowel, but resulted in good vowel identification.

This suggests that choosing the most appropriate resonance tuning techniques is, therefore, a balancing act for singers, as they must tailor the resonances of their vocal tract according to their performance aims, and decide whether to prioritize a pleas-ing voice quality over the clarity of the text in a particular situation, or perhaps sacrifice a little naturalness to achieve a higher volume in another. Deciding when and how to use resonance tuning is, therefore, an exercise in compromise in terms of performance for the ease of the singer and perception of the listener. The practical implications of the findings of this study, however, hinge on the assumption that singers are capable of controlling their vocal tract resonances with great precision, an interesting question for further research.

Conclusions

This study investigated the impact of specific resonance tuning techniques on perception through a listening test that compared synthetic vowel sounds. This allowed the resonance tuning of the sound samples to be directly manipulated and controlled. The results showed no general patterns for the perception of the different tuning strategies investigated, and in fact this appears to be highly dependent on the vowel synthesized. This suggests that, in practice, resonance tuning is likely an exercise in compromise for a singer, as employing a certain resonance tuning strategy might improve one perceptual attribute while worsening another.

These findings bring to light some of the complex relation-ships between the production and perception of vowel sounds, and the different requirements of different vowels. The next steps will consider the complex relationships between different perceptual attributes of resonance tuning using recorded voices as well as synthetic sounds. Future developments of this work also need to consider the importance of context on perception, for instance within a word or musical phrase.

Acknowledgments

Some initial results from this research were discussed during a visit by the first author to GIPSA-Lab (Grenoble, France) in 2015. The authors would also like to thank Chris Mellor (University of York, UK) for his advice on the statistical analysis per-formed in this study, and the reviewers for their advice in refining the manuscript.

Appendix:
Liljencrants-Fant (LF) model details

The LF parameters used (setting R_d = 1) were:

$$F_a = 400 \text{ Hz};\ R_k = 0.30;\ R_g = 1 \qquad (A.1)$$

where F_a in the cut-off frequency (accounting for the degree of spectral tilt), R_k specifies the relative duration of the falling branch from the peak at time T_p to the discontinuity point T_e, and R_g is a parameter that increases with a shortening of the rise time T_p (Figure A1).

$$R_a = t_a / t_0 \qquad (A.2)$$

$$R_g = t_0 / 2t_p \qquad (A.3)$$

$$R_k = (t_e – t_p) / t_p \qquad (A.4)$$

$R_d = (t_d / t_0)(1 / 110)$

$= (U_0 / E_0)(F_0 / 110)$ (A.6)

$\sim(0.5 + 1.2\ R_k)((R_k / 4\ R_g) + R_a)/0.11$

The parameters of the LF glottal model are calculated from the equations:

$tc = 1 / f0$ (A.7)

$tp = t0 / 2\ Rg$ (A.8)

$ta = 1 / 2\ fa$ (A.9)

$OQ = (1 + R_k) / 2R_g$ (A.10)

$t_e = t_0\ (1 + R_k) / 2R_g$ (A.11)

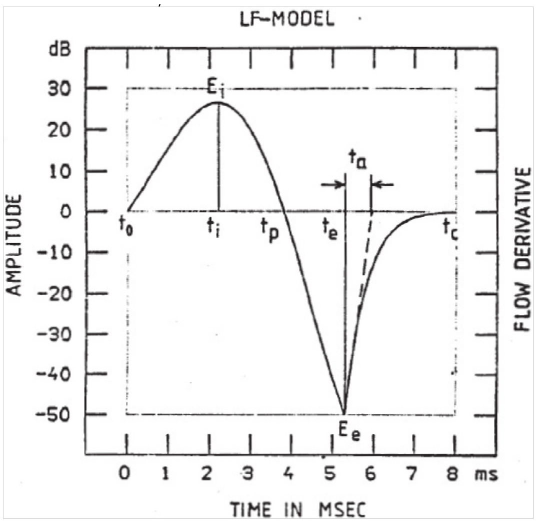


FIGURE A1. Shows the parameters of the Liljencrants-Fant (LF) model.

References

1. Peterson GE, Barney HL. Control methods used in a study of the vowels. *J Acoust Soc Am*. 1952;24:175.
2. Sawusch JR. Effects of duration and formant movement on vowel perception. In: *Proceedings ICSLP 96. Fourth International Conference on Spoken Language 1996*, Vol. 4. Philadelphia, PA: IEEE; 1996:2482–2485.
3. Scotto di Carlo N, Germain A. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica*. 1985;42:188–197.
4. Sundberg J. Vocal tract resonance in singing. *J Sing*. 1988;44:11–31.
5. Garnier M, Henrich N, Smith J, et al. Vocal tract adjustments in the high soprano range. *J Acoust Soc Am*. 2010;127:3771–3780.
6. Wolfe J, Garnier M, Smith J. Vocal tract resonances in speech, singing, and playing musical instruments. *HFSP J*. 2009;3:6–23.
7. Sundberg J. Articulatory interpretation of the singing formant. *J Acoust Soc Am*. 1974;55:838–844.
8. Hunter EJ, Titze IR. Overlap of hearing and voicing ranges in singing. *J Sing*. 2005;61:387–392.
9. WeissR,BrownWJr,MorisJ.Singer'sformantinsopranos:factorfiction? *J Voice*. 2001;15:457–468.
10. Sundberg J. Formant technique in a professional female singer. *Acta Acust United Acust*. 1975;32:89–96.
11. Garnier M, Henrich N, Smith J, et al. The tuning of vocal resonances and the upper limit to the high soprano range. In: *Proceedings of the International Symposium on Music Acoustics ISMA 2010*. Katoomba, New South Wales, Australia: International Symposium on Music Acoustics Location; 2010:11– 16.
12. Sundberg J, Skoog J. Dependence of jaw opening on pitch and vowel in singers. *J Voice*. 1997;11:301–306.

13. Sundberg J, Rossing TD. The science of the singing voice. *J Acoust Soc Am*. 1990;87:462–463.

14. TartterVC.Happytalk:perceptualandacousticeffectsofsmilingonspeech. *Percept Psychophys*. 1980;27:24–27.

15. Joliveau E, Smith J, Wolfe J. Acoustics: tuning of vocal tract resonance by sopranos. *Nature*. 2004;427:116.

16. MillerR.*OntheArtofSinging*.Oxford,UK:OxfordUniversityPress;1996.

17. Benolken MS, Swanson CE. The effect of pitch-related changes on the perception of sung vowels. *J Acoust Soc Am*. 1990;87:1781.

18. Carlsson-Berndtsson G, Sundberg J. Formant frequency tuning in singing. *J Voice*. 1992;6:256–260.

19. Zajonc RB. Mere exposure: a gateway to the subliminal. *Curr Dir Psychol Sci*. 2001;10:224–228.

20. Fant G. The LF-model revisited. Transformations and frequency domain analysis. *Speech Trans Lab Q Rep, Royal Inst Tech Stockholm*. 1995;2:40.

21. Sundberg J. Acoustic and psychoacoustic aspects of vocal vibrato. *STL- QPSR*. 1994;35:45–68.

22. Hawks JW, Miller JD. A formant bandwidth estimation procedure for vowel synthesis [43.72. ja]. *J Acoust Soc Am*. 1995;97:1343–1344.

23. Fujimura O, Lindqvist J. Sweep-tone measurements of vocal-tract characteristics. *J Acoust Soc Am*. 1971;49(2B):541–558.

24. Fant G. The acoustics of speech. In: *Proceedings of the 3rd International Congress on Acoustics Stuttgart*, Vol. 1. New York, NY: Elsevier; 1961:188–201.

25. Titze IR. A theoretical study of f_0-f_1 interaction with application to resonant speaking and singing voice. *J Voice*. 2004;18:292–298.

26. Qualtrics[computerprogram].Provo,Utah,USA,Copyright2015.Available at: http://www.qualtrics.com.

27. SchoefflerM,StöterF-R,BayerleinH,etal.Anexperimentaboutestimating the number of instruments in polyphonic music: a comparison between internet and laboratory results. In: *ISMIR*. Canada: The International Society of Music Information Retrieval; 2013:389–394.

28. Smith J, Wolfe J. Vowel-pitch matching in Wagner's operas: implications for intelligibility and ease of singing. *J Acoust Soc Am*. 2009;125:196–201.

29. Vos RR, Daffern H, Howard DM. Resonance tuning in three girl choristers. *J Voice*. 2017;31:122.e1–122.e7. 30. Henrich N, Smith J, Wolfe J. Vocal tract resonances in singing: strategies used by sopranos, altos, tenors, and baritones. *J Acoust Soc Am*. 2011;129:1024.

31. Lindblom BE, Sundberg JE. Acoustical consequences of lip, tongue, jaw, and larynx movement. *J Acoust Soc Am*. 1971;50(4B):1166–1179.

32. Fant G, Liljencrants J, Lin Q. A four-parameter model of glottal flow. *STL-QPSR*. 1980;4:1–13.