Article:

Vos, Rebecca Rose, Daffern, Helena orcid.org/0000-0001-5838-0120, Howard, David Martin orcid.org/0000-0001-9516-9551 et al. (1 more author) (2017) The Perception of Formant Tuning in Soprano Voices. Journal of Voice. ISSN 0892-1997

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Journal of Voice 00 (2017) 1–16

Journal of Voice

The Perception of Formant Tuning in Soprano Voices

Rebecca R. Vos[a], Damian T. Murphy[a], David M. Howard[b], Helena Daffern[a]

[a] The Department of Electronics and York Centre for Singing Science, University of York, Heslington, York, YO10 5DD
[b] Royal Holloway, University of London, Egham, Surrey, TW20 0EX

Abstract

0.1. Introduction

At the upper end of the soprano singing range, singers alter the shape of their vocal tract to bring one or more of the vocal tract resonances nearer to a harmonic of the voice source. This is a process known as *resonance tuning*, which increases the amplitude of the sound produced with little effort from the singer. This study investigates the perception of first and second resonance tuning, key strategies observed in classically trained soprano voices. It was expected that the most commonly-used strategies observed in singers would be preferred by listeners as part of a subjective test. This test also allows for comparison of different tuning strategies between vowels, whereas previous investigations have usually focussed only on a single vowel sound (usually /ɑ/).

0.2. Method

Synthetic vowel sounds are generated using the Liljencrants-Fant glottal flow model, passed through a series of filters to represent the vocal tract resonances. Listeners then compared the sounds, which included 3 vowels, at 4 fundamental frequencies (f_0), to which 4 different tuning strategies are applied: (A) the expected formant values in speech, (B) the first formant tuned to the fundamental, (C) the second formant tuned to the second harmonic, and (D) both first and second formants tuned to the first and second harmonics respectively. Participants were asked three sets of questions: comparing how much they preferred different tuning strategies, how natural they found different tuning strategies, and identifying the vowel for each sound.

0.3. Results

The results obtained varied greatly between vowels. The results for the /ɑ/ vowel were similar for preference and naturalness, but no clear pattern was seen for vowel identification. The results for the /u/ vowel did not appear to show a clear difference between the different tuning strategies for preference, and only a little separation for naturalness. The vowel identification was generally very poor for this vowel. The results for the /i/ vowel were striking, with strategies including R_2 tuning both preferred and perceived as being more natural than those without such tuning for both preference and naturalness. However, for vowel identification, strategies without R_2 tuning were most often correctly identified.

0.4. Conclusion

The results indicate that the perception of different tuning strategies alters depending on the vowel and the perceptual quality investigated (*preference*, *naturalness*, or *vowel identification*), and whether the first and second harmonic fall above or below the first or second formants. For some vowels and perceptual qualities, formant tuning was found to be beneficial at lower f_0 values than expected, based on current expectations of formant tuning in practice.

Keywords:

Soprano, Resonance, Formant

1. INTRODUCTION

In female speech, the first and second formants typically lie between 310 and 860 Hz and 920 and 2790 Hz respectively [1], (D#4 and A5, and A#5 and F7). The soprano range can extend to above 1000Hz, so there are frequencies at which the fundamental frequency (f_0) may exceed the frequency 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 since the first 3-5 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].

1.1. Formant tuning

A strategy used by singers to increase the efficiency of the voice at high f_0 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], creating the Singer's Formant Cluster (SFC), which increases the spectral energy in the region around 3kHz [4] where the human ear is most sensitive [8]. Evidence of a true SFC 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 could "tune" one or both of the first two vocal tract resonances to near the harmonics of the voice source. This would allow 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 confirmed 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 twelve sopranos (4 non-experts, 4 advanced, 4 professionals) singing /ɑ/ vowels. They found that R_1:f_0 tuning was employed by all the professionals and advanced singers, and to a lesser extent by the non-expert singers. R_2:2f_0 tuning was seen in 3 professionals, 2 advanced, and 2 non-expert singers. Six of the singers used R_2:f_0 tuning at very high f_0 values (above C6), and R_1:2f_0 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], while 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].

1.1.1. Disadvantages of Formant tuning

Whilst 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 "neutralise" 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].

Email address: rebecca.vos@york.ac.uk (Rebecca R. Vos)

1.1.2. The Perception of Resonance tuning

Although there is now clear evidence of the *practice* of resonance tuning (e.g. [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] which show that the likelihood of a sung vowel being misunderstood increases with f_0.

In 1991, Carlsson-Berndtsson and Sundberg published a perceptual study [18], in which synthesised 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 550Hz. In "strategy B", the second formant was tuned to the harmonic lying closest to 1000Hz. 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 1000 Hz 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 et al, [18] to be suitable for the soprano voice, and removes the possibly confounding influences of the mere-exposure effect.

Based on the evidence of R_1:f_0 and R_2:$2f_0$ tuning by sopranos [11], the perception of these tuning conditions is investigated in this paper. The properties investigated include which tuning strategies are *preferred*, their *naturalness*, and which produce the mostly clearly *identifiable* vowel sounds. The hypothesis being that the strategies used most frequently by sopranos in practise will be: preferred by subjects, perceived to be most natural, and correctly identified most often.

2. METHOD

Similar to the procedure used by G.Carlsson-Berndtsson et al. [18], synthesised tones were created to replicate voiced sounds, for which the resonance frequencies could be controlled to represent different resonance tuning strategies. Tones with f_0 typical for a soprano range were synthesised, and as resonance values have been shown to remain constant in singing up to the frequency where $f_0 = F_1$ [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], 4 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, while 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, while 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, while the third resonance is kept constant at the average value for the vowel.

2.1. Synthesised Signal

2.1.1. Glottal Signal

The synthesised vowel sounds are produced using a Liljencrants-Fant (LF) glottal flow model to create a glottal signal. Typical parameter values for a female were used, from [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 [21] sinusoidal modulation of the fundamental frequency, with an extent of 60 cents [21].

Vowel	F_1	F_2	F_3
/ɑ/	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)

Table 1: Shows the first three formant values for three vowels, when spoken by female voices [1].

Pitch number Vowel	1	2	3	4
/ɑ/	C5 529 Hz	E5 671 Hz	G#5 843 Hz	C6 1053 Hz
/u/	A#3 233 Hz	D4 294 Hz	F#4 370 Hz	A#4 472 Hz
/i/	A3 220 Hz	C#4 277 Hz	F4 349 Hz	A4 440 Hz

Table 2: Shows the fundamental frequencies of the synthesised tones for each vowel sound.

2.1.2. Vocal tract effects

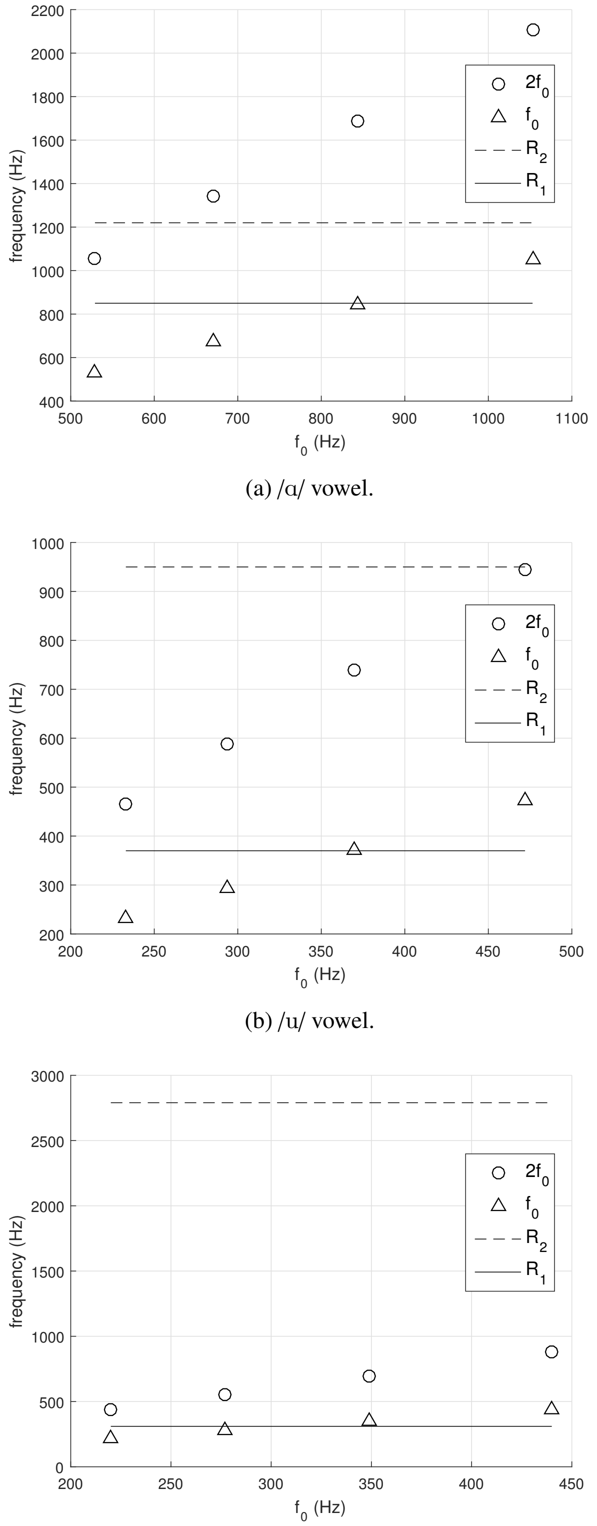
The resonances of the vocal tract were treated as a series of connected single peak IIR filters, using the `iirpeak` function in MATLAB, and the glottal signal was passed through each filter in turn. The values used for the resonances are the formant values shown in Table 1 [1], with the bandwidths fixed at 50Hz, noting that a study investigating formant bandwidth [22] which used averaged data from Fujimura & Lindqvist [23] and Fant [24], found that the bandwidth remains approximately constant at around 50 Hz for formant frequencies between 300 and 2000 Hz.

The resulting synthesised signal was then de-emphasised (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 synthesised 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 maximise 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 untuned values for R_1 & 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 $2f_0$ (1st and 2nd harmonics) for each vowel (triangle and circle respectively).

2.2. Subjects and Distribution

The listening test was distributed via email and social media, and used the online survey software Qualtrics [26]. 45 subjects took part, however 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 2 outliers), with an average time of 32 minutes.

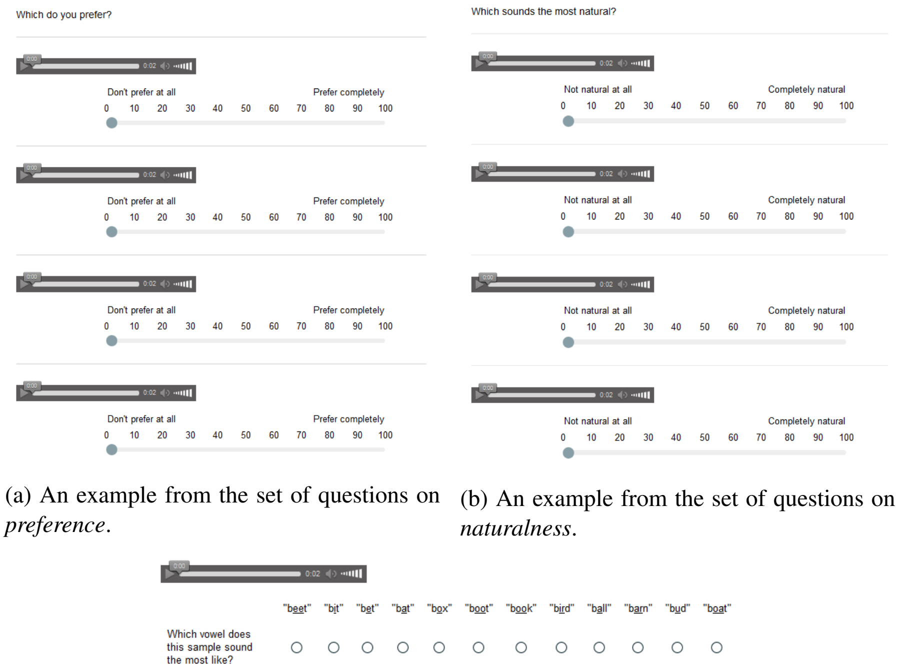
Subjects were able to take the listening test on their own devices (excluding mobile devices). 15 subjects used closed-back headphones, 7 used open-backed headphones, and 7 used earbuds. Subjects were instructed to take the test in a quiet environment with no distractions, and not to adjust the volume 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.

2.3. 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 subject's own singing ability, as well as their experience of listening to professional singing. Nine subjects had some singing training (four of which had professional training).

The listening test consisted of comparisons between sets of four tones using sliders. Each set contained tones with the same f_0 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 randomised, to minimise the effects of program-dependence. The three sets of questions considered the following perceptual aspects, *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.



Which do you prefer?

Which sounds the most natural?

Which vowel does this sample sound the most like?

0:00
0:02
"beet" "bit" "bet" "bat" "box" "boot" "book" "bird" "ball" "barn" "bud" "boat"
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

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

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

(c) An example from the set of questions on *vowel identification*.

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

3. 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 normalised to have a mean of zero 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, f_0 and tuning strategy, so that the average normalised score could be plotted against f_0 for each vowel. The results for preference and naturalness are shown in Figures 3 and 4 respectively.

The question on vowel identification was analysed by calculating the percentage of subjects that chose the correct vowel sound for each sound. These values are shown in Figures 5-7(a) for each vowel, and the most commonly chosen vowel sound is shown in Figures 5-7(b).

3.1. /ɑ/ vowel

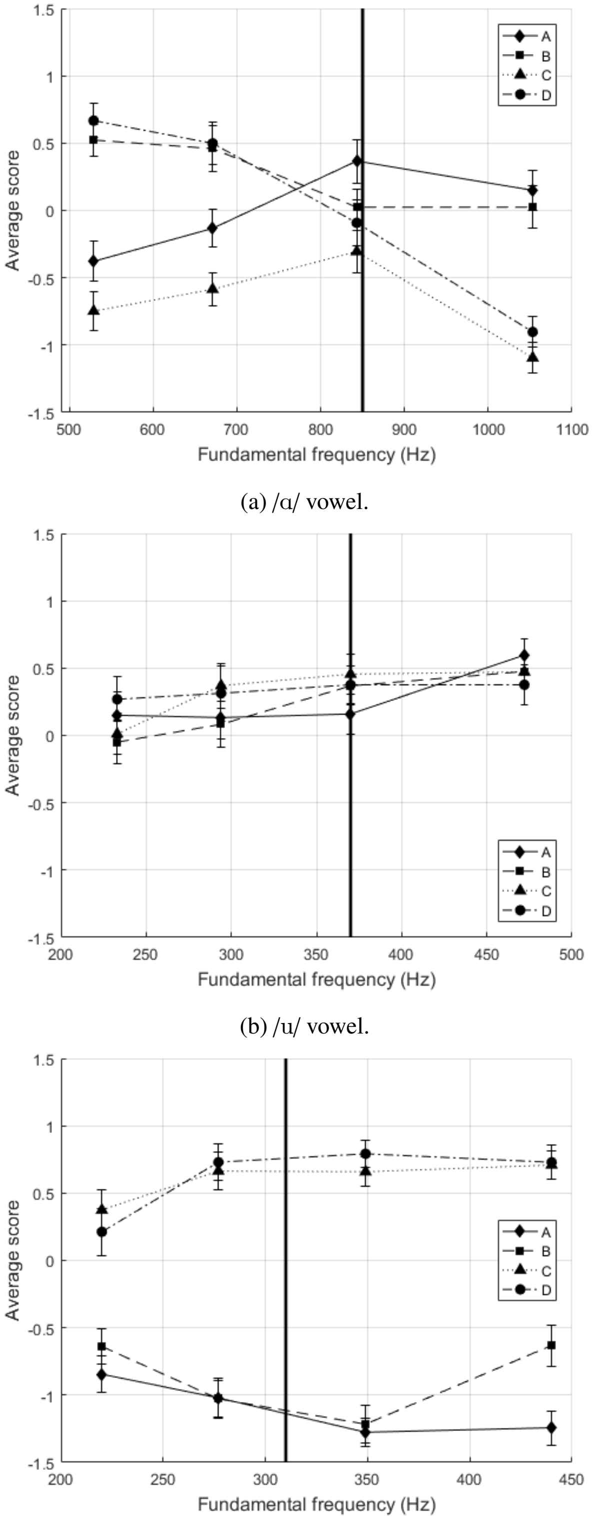
The results for the /ɑ/ vowel are similar for preference and naturalness, with strategies with R_1 tuning (B & D) scoring highest at f_0 values below R_1, but strategies without R_2 tuning (A & B) scoring highest at higher fundamental frequencies, and no clear relationship between tuning strategy and vowel identification. The results for the vowel identification for the /ɑ/ vowel show that at f_0 below R_1 strategy C (R_2 tuning only) scored the highest, with strategies A & 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 no tuning (A) was the most correctly identified, and R_2 tuning (C) the least.

3.2. /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 R_2 tuning (C & D) scoring highest in the middle of the f_0 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 R_2 tuning (C & D) scored a little lower than those without (A & B) at most f_0 values. Even the untuned 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 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, simply asked to choose which sound they preferred/found the most natural. Considering these results compared to 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 synthesise.

3.3. /i/ vowel

The results for the /i/ vowel are more revealing than the other vowels, with strategies with R_2 tuning (C & D) scoring much higher than strategies without R_2 tuning (A & B) for both preference and naturalness. However, this effect is reversed for the vowel identification, with approximately 70% of the tones without R_2 tuning correctly identified, but none of the tones with R_2 tuning.

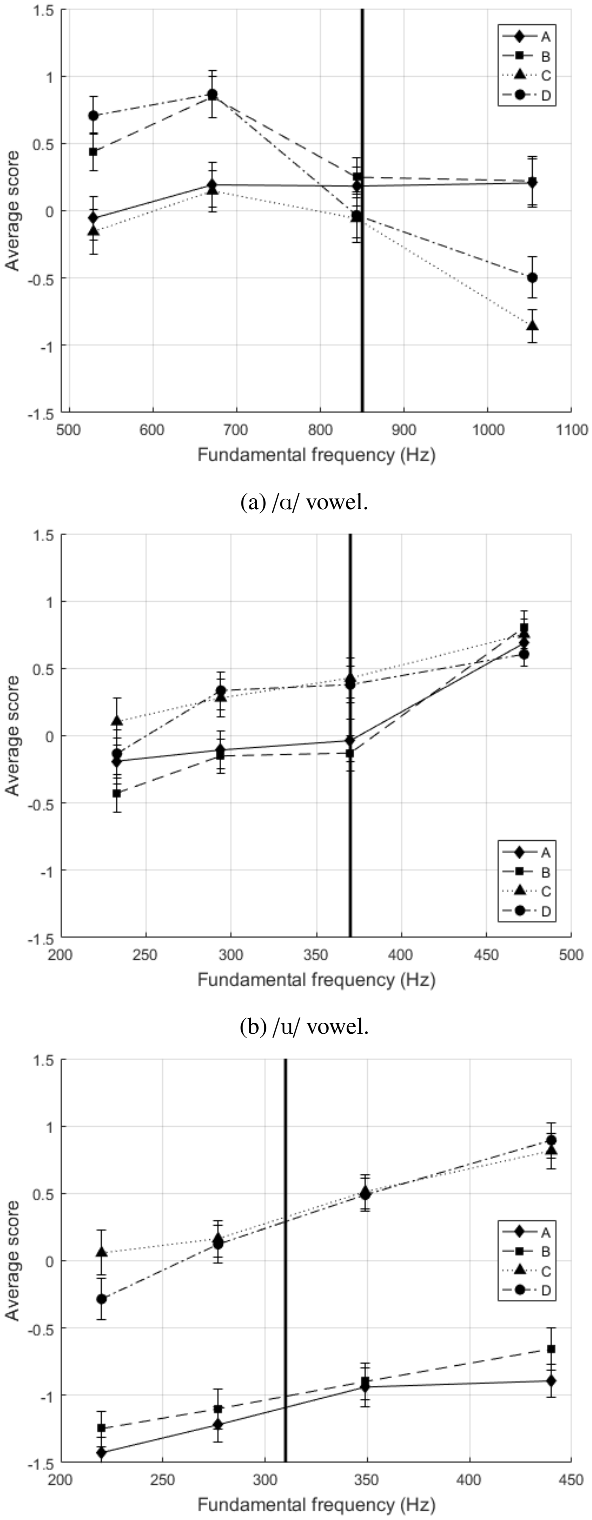


(a) /ɑ/ vowel.

(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.

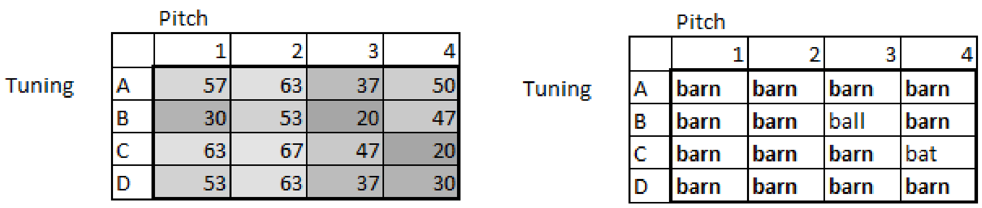


(a) /ɑ/ vowel.

(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.



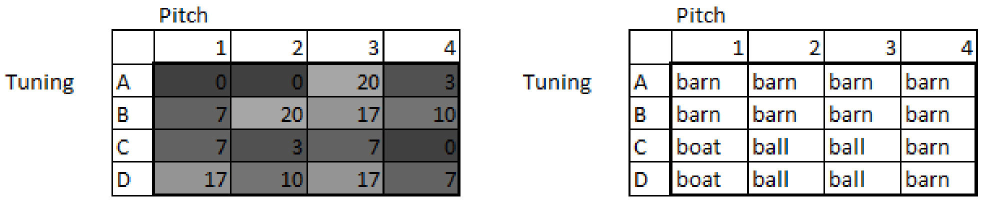
		Pitch			
		1	2	3	4
Tuning	A	57	63	37	50
	B	30	53	20	47
	C	63	67	47	20
	D	53	63	37	30

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

		Pitch			
		1	2	3	4
Tuning	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 /ɑ/ vowel.



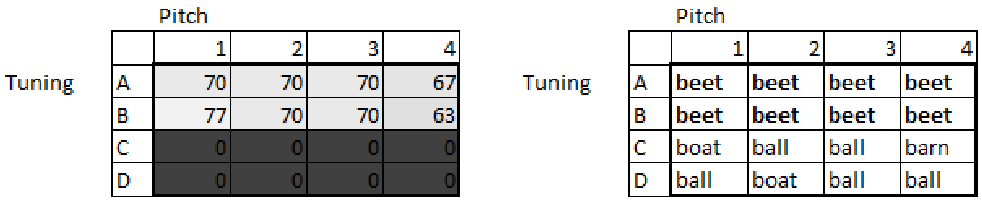
		Pitch			
		1	2	3	4
Tuning	A	0	0	20	3
	B	7	20	17	10
	C	7	3	7	0
	D	17	10	17	7

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

		Pitch			
		1	2	3	4
Tuning	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.



		Pitch			
		1	2	3	4
Tuning	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.

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

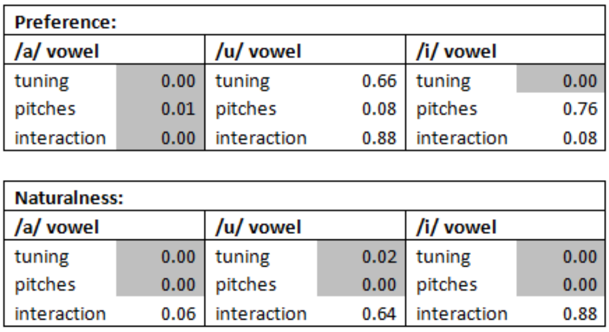
(b) The most commonly chosen vowels (correct in bold).

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

3.4. 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 grey.



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: Shows the *p*-values from the analysis of variance (ANOVA) results for preference and naturalness questions. Significant results are highlighted in grey.

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 f_0 values for the /ɑ/ vowel. There was also a significant interaction between these two variables, meaning that the subjects' preference for the sounds depended on a combination of both of these attributes. For the /u/ vowel no 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 f_0 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 f_0 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 f_0 and have a significant effect on the perception of synthesised singing sounds for *preference* and *naturalness*, although the exact relationship varies between vowels.

4. 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.

4.1. Preference

From Figure 3a, it can be seen that for the /ɑ/ vowel, at the lower two f_0 values, strategies with R_1 tuning (B & D) were preferred above strategies without R_1 tuning (A & C). The 4 tuning strategies all scored similarly when f_0 was equal to R_1, however when R_1 was above f_0 the results differ, with strategies without R_2 tuning (A & B) preferred over those with R_2 tuned (D & C). R_1 tuning only (B) scored highly across the whole range of f_0 values, which is indeed the method used most often by sopranos in this range [11]. R_2 tuning only (C) scored the lowest across the whole range of f_0 values, indicating that it was the least preferred tuning strategy. This is not surprising at lower fundamental frequencies, because R_2 tuning is rarely observed in that region, however above the normal range of R_1 tuning R_2 tuning has been observed, although more commonly in conjunction with R_1 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 f_0 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 for this (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 R_2 tuning (C & D) were preferred over those without it (A & B) across all f_0 values. The second formant for this vowel is very high (2790 Hz) compared to that of the other two vowels investigated (1120 Hz and 950 Hz for /ɑ/ and /u/ respectively). Therefore when R_2 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 untuned R_2. The very high scores in preference for tuning strategies with R_2 tuning (C & 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 untuned 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 [29].

4.2. Naturalness

From Figure 4a, as for preference, it can be seen that for the /ɑ/ vowel, strategies involving R_1 tuning (B & D) were considered the most natural at f_0 values below R_1. However as f_0 rose above R_1 the perceived naturalness of strategy D (R_1 & R_2 tuning) decreased, while strategy A (no tuning) remained roughly constant, so that at higher f_0s strategies without R_2 tuning (A & B) were perceived as more natural than those with R_2 tuning (C & 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 synthesised samples it is possible that since 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. $R_1 : f_0$) unnatural in general. Indeed Smith [29] suggests 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 categorisation 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..

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 & D) scoring a little higher than those without (A & 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, 28]. 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 & D) consistently scored the highest, followed by those without (A & 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.

4.3. Vowel Identification

The results for the /ɑ/ vowel (Figure 5) show that at f_0 values below R_1, strategy C (R_2 tuning) scored the highest, with A & D (no tuning and both resonances tuned) just below. Strategy B (R_1 tuning)was the most commonly mis-identified. 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 & D) scored a little lower than those without R_2 tuning (A & 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 & B were most commonly identified as an /ɑ/ vowel across all f_0 values, however, strategies with R_2 tuning (C & 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 (see Figure 9), however, the poor identification of even the untuned sample suggests that there may have been issues with the synthesis of this vowel sound.

The results for the /i/ vowel (Figure 7) show a very clear pattern, where strategies without R_2 tuning (A & B) were correctly identified in around 70 % of tones (with a standard deviation of 4 %), however, strategies with R_2

tuning (C & D) were never correctly identified. One explanation of this might be provided by Benolken [17], who suggests that some vowels which have similar first formant values, like the /i/ and /u/ vowels (only 60Hz apart), are differentiated by their second formants, so altering the second formant results in a dramatic loss in identifiability. The sounds with R_2 tuning (C & 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 (see Figure 9).

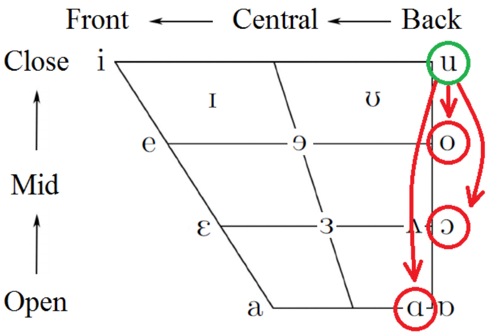


Figure 9: Shows a simplified map of the IPA monophthong vowels, and the ways in which the /u/ vowel (top right) was most commonly misidentified.

4.4. 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 attributes, with strategies involving R_2 tuning scoring the highest for both preference and naturalness but the lowest for vowel identification.

Based on the findings of Henrich [30], Carlsson [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 f_0 values below R_1, strategy A was generally one of the lowest scoring, whereas strategy 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* R_1 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 R_1, it was expected that R_1:f_0 tuning (strategy B) would score highly for all three perceptual attributes, as this is the most commonly observed in practice, and R_2:$2f_0$ tuning (strategy C) would score the lowest, as it is rarely observed in isolation [30]. Indeed, Wolfe [6] suggests that R_2 tuning might be unintentional, based on the theory that as the fundamental frequency rises, R_1 is tuned to the fundamental by increasing the opening of the mouth, and as both R_1 and R_2 rise with increased mouth opening, R_2 is raised as a side effect of raising R_1. This would suggest that R_2 tuning in isolation (C) should score quite low for both preference and naturalness, however, for some vowels and f_0 values this was not the case. For example, for preference R_2 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.

An interesting pattern seen in the results is that the strategies seemed to "pair up" for most of the perceptual attributes, with strategies without R_2 tuning (A and B) behaving similarly, as well as strategies with R_2 tuning (C and D). This seems to suggest that the presence or absence of R_2 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 focussed on single vowels (most commonly /ɑ/), this study found that the rankings of different tuning strategies is 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 R_1. 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, whilst having little effect on or detracting from another quality. For example, R_1 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 the singer, as they must tailor the resonances of their vocal tract according to their performance aims, and decide whether to prioritise a pleasing 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.

5. CONCLUSION

This study investigated the impact of specific resonance tuning techniques on perception through a listening test which 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 synthesised. 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 whilst worsening another.

These findings bring to light some of the complex relationships between the production and perception of vowel sounds, and the different requirements of different vowels. Next steps will consider the complex relationships between different perceptual attributes of resonance tuning utilising 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.

Appendix A. Appendix: LF model details

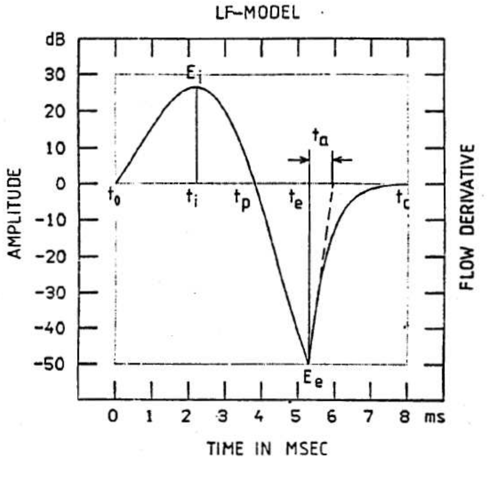


Figure A.10: Shows the parameters of the LF model.

The Liljencrants-Fant Model [32] parameters used (setting $R_d = 1$) were:

$$F_a = 400Hz, \;\; R_k = 0.30, \;\; R_g = 1 \tag{A.1}$$

Where F_a is 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 which increases with a shortening f the rise time T_p.

$$R_a = t_a/t_0 \tag{A.2}$$

$$R_g = t_0/2t_p \tag{A.3}$$

$$R_k = (t_e - t_p)/t_p \tag{A.4}$$

$$OQ = t_e/t_0 \tag{A.5}$$

$$\begin{aligned} R_d &= (t_d/t_0)(1/110) \\ &= (U_0/E_0)(f_0/110) \\ &\approx (0.5 + 1.2R_k)((R_k/4R_g) + R_a)/0.11 \end{aligned} \tag{A.6}$$

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

$$t_c = 1/f_0 \tag{A.7}$$

$$t_p = t_0/2R_g \tag{A.8}$$

$$t_a = 1/2\pi f_a \tag{A.9}$$

$$OQ = (1 + R_k)/2R_g \tag{A.10}$$

$$t_e = t_0(1 + R_k)/2Rg \tag{A.11}$$

References

[1] G. E. Peterson, H. L. Barney, Control methods used in a study of the vowels, The Journal of the Acoustical Society of America 24 (1952) 175.

[2] J. R. Sawusch, Effects of duration and formant movement on vowel perception, in: Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, Vol. 4, IEEE, 1996, pp. 2482–2485.

[3] N. Scotto di Carlo, A. Germain, A perceptual study of the influence of pitch on the intelligibility of sung vowels, Phonetica 42 (4) (1985) 188–197.

[4] J. Sundberg, Vocal tract resonance in singing, National Association of Teachers of Singing Journal 44 (4) (1988) 11–31.

[5] M. Garnier, N. Henrich, J. Smith, J. Wolfe, Vocal tract adjustments in the high soprano range, Journal of the Acoustical Society of America 127 (6) (2010) 3771–3780.

[6] J. Wolfe, M. Garnier, J. Smith, Vocal tract resonances in speech, singing, and playing musical instruments, HFSP journal 3 (1) (2009) 6–23.

[7] J. Sundberg, Articulatory interpretation of the singing formant?, The Journal of the Acoustical Society of America 55 (4) (1974) 838–844.

[8] E. J. Hunter, I. R. Titze, Overlap of hearing and voicing ranges in singing, Journal of singing: the official journal of the National Association of Teachers of Singing 61 (4) (2005) 387–392.

[9] R. Weiss, W. Brown Jr, J. Moris, Singer's formant in sopranos: fact or fiction?, Journal of Voice 15 (4) (2001) 457–468.

[10] J. Sundberg, Formant technique in a professional female singer, Acta Acustica united with Acustica 32 (2) (1975) 89–96.

[11] M. Garnier, N. Henrich, J. Smith, J. Wolfe, 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, 2010, pp. 11–16.

[12] J. Sundberg, J. Skoog, Dependence of jaw opening on pitch and vowel in singers, Journal of Voice 11 (3) (1997) 301–306.

[13] J. Sundberg, The Science of the Singing Voice, Northern Illinois University Press, 1987.
URL `http://books.google.co.uk/books?id=iYGNQgAACAAJ`

[14] V. C. Tartter, Happy talk: Perceptual and acoustic effects of smiling on speech, Perception & psychophysics 27 (1) (1980) 24–27.

[15] E. Joliveau, J. Smith, J. Wolfe, Acoustics: tuning of vocal tract resonance by sopranos, Nature 427 (6970) (2004) 116–116.

[16] R. Miller, On the art of singing, Oxford University Press, 1996.

[17] M. S. Benolken, C. E. Swanson, The effect of pitch-related changes on the perception of sung vowels, The Journal of the Acoustical Society of America 87 (1990) 1781.

[18] G. Carlsson, J. Sundberg, Formant frequency tuning in singing, Journal of Voice 6 (3) (1992) 256–260.

[19] R. B. Zajonc, Mere exposure: A gateway to the subliminal, Current directions in psychological science 10 (6) (2001) 224–228.

[20] G. Fant, The lf-model revisited. transformations and frequency domain analysis, Speech Trans. Lab. Q. Rep., Royal Inst. of Tech. Stockholm 2 (3) (1995) 40.

[21] J. Sundberg, Acoustic and psychoacoustic aspects of vocal vibrato, STL-QPSR 35 (2–3) (1994) 45–68.

[22] J. W. Hawks, J. D. Miller, A formant bandwidth estimation procedure for vowel synthesis [43.72. ja]., The Journal of the Acoustical Society of America 97 (2) (1995) 1343–1344.

[23] O. Fujimura, J. Lindqvist, Sweep-tone measurements of vocal-tract characteristics, The Journal of the Acoustical Society of America 49 (2B) (1971) 541–558.

[24] G. Fant, The acoustics of speech, in: In proceedings of the 3rd International Congress on Acoustics Stuttgart, Elsevier, New York, NY, volume 1, pages 188-201., 1961.

[25] I. R. Titze, A theoretical study of f¡ sub¿ 0¡/sub¿-f¡ sub¿ 1¡/sub¿ interaction with application to resonant speaking and singing voice, Journal of Voice 18 (3) (2004) 292–298.

[26] Qualtrics, [computer program] provo, utah, usa, copyright 2015.
URL http://www.qualtrics.com

[27] M. Schoeffler, F.-R. Stöter, H. Bayerlein, B. Edler, J. Herre, An experiment about estimating the number of instruments in polyphonic music: A comparison between internet and laboratory results., in: ISMIR, 2013, pp. 389–394.

[28] R. R. Vos, H. Daffern, D. M. Howard, Resonance tuning in three girl choristers, Journal of Voice.

[29] J. Smith, J. Wolfe, Vowel-pitch matching in wagners operas: Implications for intelligibility and ease of singing, J. Acoust. Soc. Am 125 (2009) 196–201.

[30] N. Henrich, J. Smith, J. Wolfe, Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones, The Journal of the Acoustical Society of America 129 (2011) 1024.

[31] B. E. Lindblom, J. E. Sundberg, Acoustical consequences of lip, tongue, jaw, and larynx movement, The Journal of the Acoustical Society of America 50 (4B) (1971) 1166–1179.

[32] G. Fant, J. Liljencrants, Q.-g. Lin, A four-parameter model of glottal flow, STL-QPSR 4 (1985) (1985) 1–13.