The Perception of Formant Tuning in Soprano Voices

*Rebecca R. Vos, *Damian T. Murphy, †David M. Howard, and *Helena Daffern, *York and †Surrey, UK

Summary: Introduction. At the upper end of the soprano range, singers adjust their vocal tract to bring one or more of its resonances (R_n) toward a source harmonic, increasing the amplitude of the sound; this process is known as *resonance tuning*. This study investigated the perception of (R_1) and (R_2) tuning, key strategies observed in classically trained soprano voices, which were expected to be preferred by listeners. Furthermore, different vowels were compared, whereas previous investigations have usually focused on a single vowel.

Methods. Listeners compared three synthetic vowel sounds, at four fundamental frequencies (f_0), to which four tuning strategies were applied: (A) no tuning, (B) R_1 tuned to f_0, (C) R_2 tuned to $2f_0$, and (D) both R_1 and R_2 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 R_2 tuning both preferred and considered more natural than those without. However, strategies without R_2 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 (f_0). In some cases, formant tuning was beneficial at lower f_0s than expected, based on previous resonance tuning studies.

Key Words: Soprano–Resonance–Formant–Tuning–Perception.

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 1000 Hz, 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 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

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, 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 could "tune" one or both of the first two vocal tract resonances 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 12 sopranos (4 nonexperts, 4 advanced, and 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 nonexpert singers. R_2:$2f_0$ tuning was seen in three professionals, two advanced, and two nonexpert 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] 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]

Accepted for publication March 23, 2017.
From the *The Department of Electronics and York Centre for Singing Science, University of York, York, UK; and the †Royal Holloway, University of London, Surrey, UK.
Address correspondence and reprint requests to Rebecca R. Vos, The Department of Electronics and York Centre for Singing Science, University of York, Heslington, York YO10 5DD, UK. E-mail: rebecca.vos@york.ac.uk
Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2017.03.017

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 (eg, 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 with f_0.

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 1000 Hz in all tones.

Sounds with tuned formants (using strategies A, B, or C) were presented together with the nontuned 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 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 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 controlled to represent different resonance tuning strategies. Tones with f_0 typical for a soprano range were synthesized, 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 Reference 18, four resonance tuning strategies were tested:

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

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)

- 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.

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[21] 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 iirpeak 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[22] 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.