Formant Tuning Strategies in Professional Male Opera Singers

Johan Sundberg

Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden

Filipa M. B. Lã

Department of Communication and Arts, INET-MD, University of Aveiro, Portugal

Brian P. Gill

Department of Music and Performing Arts Professions
Steinhardt School of Culture, Education and Human Development
New York University, USA

Running title: Male singers' formant strategies

Corresponding author:

Johan Sundberg, Professor, Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, SE-10044 Stockholm, Sweden

Phn:+468 790 7873

Email: pjohan@speech.kth.se

ABSTRACT

The term “formant tuning” is generally used for the case that one of the lowest formant frequencies coincides with the frequency of a source spectrum partial. Some authors claim that such coincidence is favorable and belongs to the goals of classical opera voice training, while other authors have found evidence for advising against it. This investigation analyzes the relationships between formant frequencies and partials in professional singers, who sang scales on the vowels /a/, /u/, /i/ and /ae/ in a pitch range including the *passaggio*, i.e., the fundamental frequency range of approximately 300 – 400 Hz, applying either of two singing strategies that are typically used (1) in classical and (2) in Non-classical singing, respectively. Formant frequencies of each note in the scales were measured by inverse-filtering the acoustic signal. In the classical style, the first formant tended to be lower than in the Non-classical style. Neither the first, nor the second formant tended to change systematically between scale tones, such that on some scale tones either or both formants was just below, just above or right on a spectrum partial. In many cases singers produced similar spectrum characteristics of the top tones of the scales with different first and second formant frequencies. Regardless of whether the first formant was slightly lower, slightly higher, or right on a partial, the properties of the voice source did not seem to be affected.

Keywords

Operatic singing, Non-classical Singing, Spectrum, Harmonics, Formant tuning

I. INTRODUCTION

Sopranos have been found to avoid the situation where F1 is lower than the fundamental frequency, F0; if F0 exceeds the normal value of F1, they increase F1 to a frequency very close to F0 (Sundberg, 1975) . The strategy of tuning F1 to a specific spectrum partial was strongly recommended by singing teacher Berton Coffin (1980), who developed a vowel chart presenting optimum vowels of a given pitch range (Coffin, 1980) . The terms “formant tuning” or “vocal tract tuning” have been introduced for this strategy, and it has been found to be applied not only by sopranos, but also by singers of other classifications. For male voices, vocal tract tuning is mostly observed in and above the *passaggio*, i.e., near the pitches of E4 to G4 - about 330 and 400 Hz (Doscher, 1994). Vocal tract tuning increases the sound level of the vocal output, and it is assumed to help the singer avoid vocal hyperfunction and register breaks (Coffin, 1980; Hertegård et al., 1990; Hirano et al., 1970; Miller and Schutte, 1990; Neumann et al., 2005). It is not specific to classically trained singers; it has been found also in some traditional singing (Henrich et al., 2011b).

Formant tuning has been receiving an ever increasing amount of attention over the last decade (Miller, 2008; Neumann et al., 2005; Schutte et al., 2005; Titze, 2004; Titze, 2008; Titze and Worley, 2009). The literature on formant tuning has recently been described in some detail elsewhere(Sundberg et al., 2011), and hence, only a brief overview will be given here (see Table 1).

Please insert Table 1 about here

As can be seen in Table 1, different methods have been used in this research. However, all these methods suffer from limitations: (1) it is difficult to interpret spectral data in terms of formant frequencies, since the spectrum is strongly dependent on the relationship between F0 and the frequencies of the formants, and when shifting from modal phonation to ingressive or vocal fry phonation, vocal tract shape, and hence the formant frequencies, may change; (2) results of listening tests with synthesized stimuli are very dependent on how natural the stimuli sound; (3) model work offers accurate predictions only if the model accurately reflects all relevant characteristics of the system; (4) broadband excitation at the lip opening during sustained vowel phonation implies that also the subglottal system is to some extent included in the resonator, since the glottis is open during part of the glottal vibratory cycle; and (5) in inverse filtering, the formant ringing during the closed phase is eliminated by tuning the frequencies of the inverse filters, so if some of this ringing was caused by factors other than formants, e.g., non-linear source–filter effects, these may be mistakenly eliminated.

In addition, the results of this research diverge, as can be seen in the same Table 1. There are at least three sharply conflicting ideas on vocal tact tuning in singing that have emerged in the world of vocal pedagogy and voice research - singers are recommended to or found to: (1) tune F1 and/or F2 to a harmonic partial; (2) keep F1 and F2 constant, independent of F0; and (3) tune F1 or F2 to a frequency just *above* its nearest partial, so that coincidence between partial and formant is avoided.

Against this background, it is clear that vocal tract tuning deserves more research. Thereby, a promising strategy seemed to be to expand our previous, preliminary study of the vowel /a/ (Sundberg et al., 2011), by including also vowels with lower F1 values.

II. MATERIALS AND METHODS

A. Participants and Protocol

A total of eight male classically trained singers, 23 to 42 years old, with varying levels of expertise, ranging from graduate student to international opera singer, voluntarily participated in this study (see Table 2). The singers performed ascending/descending nine note scales using vowels /ae/, /a/, /u/, and /i/. For each singer, three starting pitches were used, separated by semitone intervals and chosen so that the singer's *passaggio* range was reached (pitches E4 – G4, depending on voice type). When reaching their *passaggios*, they applied either of two different approaches: classical (sung twice) and Non-classical, as in e.g., musical theatre (sung once). The *Madde* vowel synthesizer software (by Svante Granqvist, KTH) was used to supply the reference tone. A total of 224[1] scales were recorded: as there were three starting pitches and four vowels, there were 192 classical and 32 Non-classical versions.

Please insert Table 2 about here

B. Equipment

All recordings were made in a sound treated studio in the Steinhardt School of Culture, Education and Human Development at New York University. A hybrid system consisting of a Laryngograph microprocessor and a Glottal Enterprise MS-110 computer interface was used to record audio and electrolaryngograph (EGG) signals simultaneously. Audio was picked up by a head-mounted electret microphone (Knowles EK3132). Sound level was calibrated by means of a 1 kHz sine wave, the sound pressure level of which was measured in dB(C) next to the recording microphone by means of a sound level meter, and the value observed was announced on the recording. Both signals were recorded using Speech Studio software (Laryngograph©) and stored as wav files.

C. Analysis

Listening test

---

[1] This number was mistakenly reported to be 288 in a previous publication Sundberg J., Lã F.M.B., Gill B.P. (2011) Professional male singers' formant strategies for the vowel a. Logopedics Phoniatrics Vocology. 36:156-167..

Of the recorded scales, 28 were randomly selected for a listening test, aimed at identifying typical examples of classical and Non-classical versions. The listening test was carried out with a panel of six experienced listeners (university singing teachers in USA with knowledge of vocal tract tuning). The subjects were given the following written instruction: "*Here you will listen to a set of ascending-descending scales sung by different male voices. We ask you to rate how successful the singer is with regard to vocal tract tuning (disregard other issues) in transitioning into the pitches in/above the passaggio in a Classical mode of singing*." The subjects gave their ratings along visual analogue scales (VAS). The test included 13 replicated stimuli and all 41 scales were presented in the same randomized order to all subjects, with a 5 second pause in between**,** lasting a total of approximately 11 minutes. The number of each stimulus was announced in the test file.

Formant frequency analysis

The formant frequencies in all recorded tones in the ascending part of the scale were measured by inverse filtering of the audio signal, using the custom made *Decap* software (by Svante Granqvist, KTH). This program displays waveform and spectrum in separate windows (see Figure 1). When analyzing an audio signal, it first converts the acoustical signal to a flow signal by integration and the frequencies and bandwidths of the inverse filters are set manually (see below). Then, it applies the classical equations for calculating the transfer function that corresponds to a given combination of formant frequencies and bandwidths (Fant, 1960). The input signal is then filtered with the inverse of this transfer function, thus eliminating the effects of the vocal tract transfer function on the input signal, and the resulting waveform and spectrum are displayed in quasi-real-time. Provided that the filters are correctly set, the output then displays the waveform and spectrum of the transglottal airflow, including non-linear source filter interaction, if any.

The program can also display a signal recorded on a second channel of the input file, with or without derivation and with an adjustable time delay.

In the upper window of Figure 1, both the transglottal airflow waveform, i.e., the flow glottogram, and the derivative of the EGG signal (dEGG) are displayed. The dEGG was delayed by a time interval of about 1 millisecond corresponding to the travel time of the sound from the glottis to the microphone. The lower window represents the input audio spectrum and the spectrum of the flow glottogram. The formant frequencies and bandwidths were adjusted according to three criteria: (1) ripple free closed phase; (2) voice source spectrum envelope as void of peaks and valleys as possible; and (3) synchrony between the negative dEGG peak and the maximum declination rate of transglottal flow during closure.

Please insert Figure 1 about here

After completing the tuning of the filters, their frequencies and bandwidths were saved in a formant data file. The top note of each of the scales starting on the highest pitch was then synthesized, using the frequencies and bandwidths saved in this formant file. This was done for each of the four vowels, using the *Madde* software. The bandwidths were about 50, 70, 100, 110, 120 Hz for the five lowest formants. The purpose of this synthesis experiment was to test the formant values used for the inverse filtering. The values were accepted only if the synthesized vowel and voice quality sounded similar to vowel sound analyzed, as determined by the authors.

The F0 signal was extracted from the EGG, using the FoX option included in the Soundswell signal workstation software (Hitech Development, Solna, Sweden). For each of the individual scale tones, F0 was averaged over a series of complete vibrato cycles by means of the histogram option of the same software. In this way, F0 and formant frequencies data were obtained, that were as reliable as possible.

III. RESULTS

Listening Test

A Pearson's Correlation was used to test the consistency of the participating experts' ratings. For the individual experts, the absolute difference between the first and second ratings of replicated stimuli varied between 4.4% and 11.9% and the correlation between 0.779 and 0.979 (see Table 3).

Please insert Table 3 about here

To assess whether there were significant differences between the ratings for the Classical and Non-classical examples included in the listening test, a non-parametric paired sample test -Wilcoxon- was performed. This test was used because the data showed a skewed distribution. Classical versions were scored significantly higher than Non-classical versions for all singers, except for the vowels /u/, as sung by singer 1, and the vowel /a/, as sung by singer 3 (see Table 4).

Please insert Table 4 about here

Formant Frequencies

Six pairs of scales of the same vowel sung by the same singer in both Classical and Non-classical versions were included in the listening test: three on /a/, one on /ae/, and two on /i/. No pairs for the vowels /u/ were chosen because none of the singers managed to sing a Non-classical version of the /u/ scale. To increase the analysis material, however, formant frequency measurements were made also on two more Classical versions of /i/ scales and four Classical versions of /u/ scales. Thus, a total of eighteen scales were selected for formant frequency analysis.

Figures 2 a, b and c show the formant frequencies obtained. Figure 2a pertains to the vowel /a/, described in our previous report (Sundberg et al., 2011), complemented by new data for vowel /ae/. As can be seen in the graph, the mean ratings of the Classical versions of these scales exceeded 75% of VAS length, while the mean ratings for the Non-classical versions varied between 10% and 32%. Especially in and above the *passaggio* range, F1 and F2 were lower in the Classical than in the Non-classical version of the /ae/, just as in the /a/. In the Classical version, all singers lowered F1 for the top scale tone or scale tones, while in the Non-classical version, F1 was just above, just below, or right on H2. F2 was just above, just below or right on H3 for the top pitches in the Classical versions. In these versions, all singers tuned F1 almost exactly to H2, at least in one of the scale tones. However, with few exceptions, neither F1 nor F2 changed markedly between scale tones, thus suggesting that the singers did not attempt to tune these formants to the frequencies of spectrum harmonics of the individual scale tones. In other words, the coincidence between formant and harmonic seemed to occur unintentionally. However, singer 2 tuned his F2 exactly to H3 in /a/, both in the Classical and Non-classical versions of his highest note.

Please insert Figure 2 about here

Figure 2b shows the F1 and F2 values observed for the vowel /u/, all in the Classical versions. Singers 1, 2, 3, and 4 all tuned F1 midway between H1 and H2 for the top pitches. Singers 1 and 3 kept F2 rather constant throughout the scale, with F2 coinciding with H3 on the penultimate note for singer 1 and the top note for singer 3. Singer 6 by contrast, changed F2 in the lower part of the scale such that it was close to H4 and H3. Singer 2 provided a clear example of vocal tract tuning; F1 tracked H2 for the 4 lowest scale tones, while for the higher tones he tuned F2 to H3.

The F1 and F2 values observed for /i/ are shown in Figure 2c. A common trait is that all singers reduced F2 more or less for the top pitches, both in the Classical and Non-classical versions. Singers 1 and 2 produced a similar difference between Classical and Non-classical for F1. However, the difference was quite small in the case of singer 1, and, interestingly, also the mean ratings of these examples were not very different, 48% and 78%. Singers 4 and 6 produced clear examples of vocal tract tuning, placing F1 just above H1 for the highest scale tones. Their F2, by contrast, remained basically independent of F0. Singers 1 and 2 increased F1 with increasing F0, singer 2 tuning it almost midway between H1 and H2 in the upper part of the scale.

Singers have been found to tune their formant frequencies with an accuracy of about 20 Hz (Henrich et al., 2011a). Therefore, if applying formant tuning they may not tune F1 exactly to a partial, but only to the vicinity of a partial. In the investigation just referred to, the criterion for formant tuning applied was that the distance between the formant and its closest partial should not be wider than 50 Hz. This corresponds to about 2 semitones in the range of the pitch of E4. If singers apply formant tuning with this accuracy, one would expect that the average of this distance would be close to zero semitones in our material. As formant tuning is assumed to be required for tones in the upper part of the *passaggio*, this should apply at least to the top note of the scales (Miller, 2008). Formant tuning may concern both F1 and F2. For these reasons we measured the frequency ratios, expressed in semitones, between these formants and their closest partials. The results are shown in the box plots in Figure 3. For the vowel /a/ in the classical version, the median distance was clearly negative for F1 and small and positive for F2. Thus, F1 was typically lower than its nearest partial. For /i/, F1 and F2 were about one semitone above and below the nearest partials, respectively. For the vowel /u/, the median was zero for both F1 and F2, but the scatter was considerable. For the Non-classical versions, the medians for /a/

and /i/ were close to zero. These results show that depending on the vowel, the formant may be below, above and right on a partial, although the proximity between F1 and F2 and their closest partials tended to be greater in the Non-classical versions.

Please insert Figure 3 about here

It also is relevant to compare the formant tuning data with the mean ratings of how successful the singers were with regard to vocal tract tuning, when they transitioned into the pitches in and above the *passaggio* in the Classical mode of singing. This rating can be assumed to apply to the top note of the scale. Moreover, formant tuning would imply that either F1 or F2, or both, are tuned to the close neighborhood of a partial. Hence, we calculated, for the top tone, the minimum of the distance between F1 and its nearest partial and the distance between F2 and its nearest partial, and then the minimum of these two values was selected. These minimum values can be compared with the mean ratings in Figure 4. No relationship can be seen with the mean ratings, but almost all values lie within a range of ±1.5 semitones from the closest partial. Thus, the ratings seemed independent of the frequency separation between F1 or F2 and their closest partial.

Please insert Figure 4 about here

Flow Glottogram Waveform

When formants coincide with the frequencies of harmonics, non-linear source filter interaction can be expected to occur (Titze, 2004). In such cases, the flow glottogram would show deviations from the typical shape, e.g., in terms of a ripple in the closed phase or a dent in the flow pulse. Such interaction would also imply that the source spectrum partial that coincides with a formant may become boosted or attenuated, so that such partials can be expected also to deviate from a smoothly falling source spectrum envelope.

Figure 5 shows a typical set of flow glottograms obtained from the top tones of the scales sung on the vowels /a/, /u/ and /i/. The samples exemplify the situation that F1 is just below, right on, or just above H1, H2, or H3. A small ripple in the close phase occurred for the /a/ when F1 was 58 Hz above H2 (1.5 semitone), for the /u/ when F2 was 60 Hz above H1 (1.1 semitone) and for the /i/ when F1 was 26 Hz above H1 (1.3 semitone). In the last mentioned case, the ripple did not seem related to F1 being just above a harmonic, since the ripple remained also when F1 was right on or just below a harmonic. It cannot be excluded that the ripple was caused by some external resonance; the ripple corresponded to a frequency of about 1200 Hz, and in the vowel /i/, this frequency is located in a spectrum region where the partials are weak, thus allowing an external resonance to have a noticeable effect on the flow glottogram.

Please insert Figure 5 about here

Spectra

In pedagogical practice, spectrum characteristics are sometimes used to provide a visual feedback to the student (Miller, 2008). Figure 6 illustrates how F1 and F2 were tuned for the top pitch of the vowel /a/ in five cases, all rated as good examples of vocal tract tuning. All these examples shared the characteristic of a rising audio spectrum envelope over partials 1, 2, and 3. However, this characteristic was achieved by means of different combinations of F1 and F2 and their distances to the closest partials. Singers 1, 2, and 4 placed F1 well above H1 and well below H2, while singer 3 and 6 tuned it just below and just above H2, respectively. Also with regard to F2, different strategies can be observed: singers 1, 3, and 6 placed F2 almost right on H3 while singers 2 and 4 placed it just below this same partial.

Please insert Figure 6 about here

One might imagine that singers possessing a strong singer's formant cluster would tend to tune F1 or F2 to a partial in order to promote a favorable spectral balance. Singers possessing a weak singer's formant cluster would promote spectrum balance by not tuning a formant to a partial. This would imply that the level difference between the strongest spectrum partial and the singer's formant cluster would show a relationship with the smallest distance between F1 or F2 and its closest partial. This relationship was examined: the coefficient of determination was 0.108. Thus, there was no correlation between formant tuning and the level of the singer's formant cluster.

Synthesis

The top tones were synthesized by means of the singing synthesizer *Madde*, as mentioned. This allowed for a more detailed examination of the contributions of non-linear source-filter interaction to the audio signal in these tones, since this synthesizer represents a completely linear source-filter model. It simply filters a source spectrum, which has an envelope falling with an adjustable number of dB/octave, and its transfer function is calculated from the given frequencies and bandwidths of the formants. Thus, the differences between the real and the synthesized versions of a vowel would show the contributions from non-linear factors. In these measurements, spectrum sections were taken at the upper turning point of the vibrato cycle, and F0 of the synthesis was adjusted to that F0 value.

Figure 7 compares the levels of the first ten audio spectrum partials produced by the singer and by the synthesizer, for four cases where either F1 or F2 coincided with a spectrum partial. In three of the four cases, the singer's fundamental was weaker than that of the synthesizer. It can be noted that in none of these cases, a particularly great discrepancy appeared for the partial that coincided with a formant (circled in the graph). Thus, the source spectrum seemed to be rather unaffected by the resonator in these cases.

Please insert Figure 7 about here

Figure 8 shows the audio spectra of these same tones. In the case of the /u/ sung by singer 3, the very strong third partial required that the bandwidth of the inverse filter corresponding to F2 was set to no more than 19 Hz, which is an unrealistically low value. A possible explanation is that this partial was boosted by non-linear source-filter interaction. For the other vowels, the bandwidths of the inverse filters were all within or very near the realistic limits, represented by the two parallel curves in the *Decap* graphs (see Figure 1).

It can also be noted that the spectrum peak constituting the singer's formant cluster was produced by clustering formants 3, 4, and 5 in the synthesis. The levels of these partials in the singers' spectra did not differ markedly from those of the synthesis. This implies that by and large, the singer's formant cluster could be completely explained by the classical source filter theory.

Please insert Figure 8 about here

IV. DISCUSSION

This investigation leans heavily on the assumption that inverse filtering is applicable also in the presence of a non-linear source-filter interaction. There seems to be no reason to doubt this assumption. It is a well established fact that the transfer function of the vocal tract can be accurately predicted, given the frequencies and the bandwidths of the formants (Fant, 1960). Inverse filtering merely computes this transfer function and filters the signal with the transfer function inverted. The transfer function will not be affected by a non-linear source filter interaction. Therefore, the source waveform and spectrum will faithfully reflect the effects of such interaction, provided that realistic formant frequencies and bandwidths were used when adjusting the inverse filter. For this reason, we were careful only to use formant frequencies and bandwidths typical of the vowels analyzed. It should also be noted that the inverse filter was adjusted such that the ripple during the closed phase was eliminated. During this phase, the folds prevents coupling of the vocal tract to the subglottal airways. In other words, under these conditions, the resonator system includes only the vocal tract. Moreover, the flow glottograms obtained showed a closed phase that started at the moment of vocal fold contact, as evidenced by the spike of the dEGG waveform. These facts support the assumption that our results represent reliable information and that the inverse filtering data accurately would reflect any significant effects of a non-linear source-filter interaction.

Nevertheless, is obviously important to compare our findings with those reported by others. Echternach (2010) converted MRI data for the vowel /a/ to area functions and calculated the formant frequencies of premiere tenors' *passaggio* (Echternach, 2010). He found that they progressively lowered F1 with rising F0. Henrich and associates observed a great inter-subject variation in their group of singers, which included both amateurs and professionals (Henrich et al., 2011a). In their male singers, they observed formant tuning in the top of the singers' ranges. We observed some examples of this.

The representativeness of the data is another important aspect of our findings. There are three reasons to assume that our subjects were good representatives of classically trained opera singers: (i) the examples selected for analysis were those receiving the highest and lowest mean ratings as Classical in the listening test, whereby those receiving the lowest mean ratings were regarded as typical of Non-classical; (ii) the subjects were all singing at professional/semi-professional level, some at the Metropolitan opera; and

(iii) all singers tried to follow the instructions given to provide the Non-classical and Classical versions.

Given these reasons for assuming that the data observed are representative, two main questions need to be considered: (i) is there a common tuning strategy that male singers employ to successfully navigate through their passaggio; and (ii) where are the formants in relation to the partials?

Regarding the first question, the answer is in the negative; although there were spectral similarities, there was no common formant tuning strategy that our male singers used to successfully manage their *passaggio*. Such variability was observed also by Henrich and associates (Henrich et al., 2011a). This suggests that perhaps, instead of a “formant tuning rule” that all male singers should apply to negotiate *passaggio* notes, they apply personal strategies, presumably tailored to their own anatomic-physiological characteristics.

On the other hand, three common denominators were found for vocal techniques perceived as Classical and Non-classical. First, in the top notes of the Classical examples, F1 and F2 were lower than in the Non-classical examples sung by the same singer, perhaps due to a lowering of the larynx, lip protrusion, or decreased jaw opening. Second, F1 in /a/ was lowered for the top notes, thus suggesting that a principle of formant detuning rather than formant tuning was applied. Third, all top notes on /a/ perceived as clearly representing the Classical singing technique shared the characteristic of a rising spectrum envelope over the three lowest spectrum partials of the audio signal, which possibly produces a desirable timbral effect in this style of singing.

It is noteworthy, however, that the sensitivity of the audio spectrum to the vibrato phase is quite large, particularly when a formant is tuned to the close proximity of a spectrum partial. Figure 9 presents an example of a spectrum taken at the peak and at the valley of the same vibrato cycle for an /a/ sung at the pitch of F#4. For example, assuming a bandwidth of 50 Hz for the vowel /a/ and an H2 located 25 Hz below F1, a vibrato peak-to-peak extent of 5% would lead to a sound level modulation of about 6dB for that partial. In other words, the spectrum envelope over the three lowest spectrum partials may vary greatly over a vibrato cycle. It is not clear how the criterion of a rising spectrum envelope over harmonics 1, 2, and 3 is applied under such conditions.

Please insert Figure 9 about here

With regard to the second question, where the formants are in relation to the partials, it is clear that F1 was never below H1. This finding is in agreement with earlier studies (Henrich et al., 2011a; Sundberg, 1975; Titze, 2008). On the other hand, in and above the *passaggio,* range we found only few examples of F1 and/or F2 being tuned to a partial. Mostly, the formant frequencies remained the same or similar between scale tones, so coincidence between F1 or F2 and a partial appeared rather to happen by chance. Thus, our results fail to support the claim that such tuning is an important principle in the Classical style of singing (see e.g., (Henrich et al., 2011a; Miller, 2008; Neumann et al., 2005). With regard to F1, it was tuned to a frequency well above, just above or right on H1 in /i/. In /a/ and /æ/ it was lowered to a frequency below H2 for the highest pitches, which is in accordance with earlier reports (Echternach, 2010; Hertegård et al., 1990; Neumann et al., 2005). In /u/, we found F1 to be tuned midway between H1 and H2, an observation markedly deviating from results reported by Henrich and associates(Henrich et al., 2011a). With respect to F2, it coincided with or was in the vicinity of H3 for /a/ and /æ/, whereas for /i/ and /u/ it seemed independent of F0 in all singers except one (singer 2). While Neumann and associates noted that F2 in /a/ was tuned to H4 in the high range of the ‘chest’ register (Neumann et al., 2005), we found that this situation happened at single pitches, when F0 was in the appropriate range for this to happen by coincidence, i.e., it occurred without any marked changes of F2 between scale tones.

As mentioned, F1 or F2 coincided with a harmonic in several scale tones. According to Titze (2008), this may cause instability because of non-linear source-filter interaction, at least when no vertical phase difference can be seen in the vocal fold vibration, such as in falsetto register (Titze, 2008). However, in none of these cases instability was noted, presumably because singers used only modal register. Apparently singers are able to avoid instabilities caused by such interaction, as suggested by Titze (2008). A relevant remaining question is how this can be done.

Non-linear source-filter interaction is likely to boost certain partials in the radiated spectrum. Inverse filtering merely implies that the effects of the vocal tract transfer function on the radiated sound are eliminated, as mentioned. A non-linear source-filter interaction would then manifest itself as irregularity of the source spectrum envelope obtained from the inverse filtering analysis. One example of this was illustrated in Figure 8, where F2 was almost identical with H3 in the /u/ sung by singer 3. In the inverse filtering, the very strong H3 was here compensated by an unrealistically narrow bandwidth of F2. No similar examples were observed in the Classical versions of the scale. On the other hand, irregularities were often noted in the source spectrum envelope of the Non-classical versions. This again raises the question what tricks singers can make to avoid non-linear source filter interaction.

It is noteworthy that the source spectrum envelope did not show any irregularities in the region of the singer’s formant cluster. The generation of this spectrum envelope peak was thus compatible with a normal voice source and the clustering of F3, F4, and F5 assumed in the inverse filtering analysis. Such clustering does not seem unrealistic; it was observed also in an acoustical model of the vocal tract, which contained a representation of the pyriform sinuses as well as of the larynx tube including a laryngeal ventricle (Sundberg, 1975). Thus, it seems that a singer's formant cluster can be produced without the help of non-linear source-filter interaction.

Formant tuning, meaning tuning of formants to partial, must be strongly influenced by F0 and the normal value of F1 and F2. The male *passaggio* is limited to the F0 range of 300 – 400 Hz, approximately, and F1 for the vowel /i/ is typically near 300 Hz. This automatically brings it to the vicinity of H1 in the *passaggio*. Likewise, F1 for the vowel /a/ is about 700 Hz which implies that distance between F1 and H2 will automatically be small in the *passaggio*. It is thought-provoking that in the Classical versions, the singers tended to decrease F1 in the *passaggio*, thus, contrary to formant tuning, expanding the separation between F1 and H2. In this case, the term formant detuning seems more appropriate than formant tuning.

V. CONCLUSIONS

The main results of the present investigation can be summarized as follows. (1) The Classical and Non-classical styles of singing differed with respect to formant frequencies in a consistent and clearly perceptible way, and for all vowels F1 and F2 tended to be lower in the Classical than in the Non-classical style. This difference was most pronounced at high F0. (2) In two cases out of a total of 18, examples of formant tuning were found, occurring at high pitches for the vowel /i/ and both at low and high F0 for the vowel /u/. (3) A rising spectrum envelope over the three lowest partials was a common denominator of the highest tones sung on /a/ and /ae/ in Classical style, even though it was produced with slightly differing combinations of formant frequencies and the spectrum varied greatly during the vibrato cycle. (4) F1 coincided with H2 at some scale tones in all singers' Classical as well as in their Non-classical versions of the scale. (5)

Almost without exception, inverse filtering analysis of the tones produced in the Classical style showed no clear signs of a non-linear source filter interaction, neither when F1 or F2 coincided with a spectrum partial, nor when F1 was slightly lower than a partial. Thus, in most cases, the major characteristics of the spectra produced by these singers with the Classical formant tuning strategy could be explained by the classical linear source-filter theory of voice production. On the other hand, in examples produced with a Non-classical formant tuning strategy, some source spectrum irregularities were found that may reflect such interaction.

ACKNOWLEDGEMENTS

The kind cooperation of the singers is gratefully acknowledged. A preliminary version of this investigation was presented at the *Physiology and Acoustics of Singing* conference at KTH, Stockholm, in August 2010, at the Seattle meeting of the Acoustical Society of America 2011, and at the Annual Symposium Care of the Professional Voice, June 2011. The authors are indebted to the Schering Health Care Ltd. & Bayer Portugal for providing the means to acquire the equipment, and to the Gomes Teixeira Foundation for support.

REFERENCES

Coffin B. (1980) Overtones of Bel Canto Scarecrow Press, New Brunswick.

Doscher B.M. (1994) The Functional Unity of the Singing Voice, The Scarecrow Press, Inc.,, London.

Echternach M. (2010) Untersuchungen zu Registerübergängen bei männlichen Stimmen (Investigations of register transitions in male voices) Projekt verlag Bochum, Freiburg.

Fant G. (1960) Acoustic Theory of Speech Production. 2nd edition ed. The Hague, Mouton.

Henrich N., Smith J., Wolfe J. (2011a) Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. Journal of Acoustic Society of America 129:1024–1035.

Henrich N., Kiek M., Smith J., Wolfe J. (2011b) Resonance strategies used in Bulgarian women's singing style: A pilot study. Logopedics Phoniatrics Vocology 32:171-177.

Hertegård S., Gauffin J., Sundberg J., " J. Voice 9, 149-155. (1990) A comparison of subglottal and intraoral pressure measurements during phonation. Journal of Voice 9:149-155.

Hirano M., Vennard W., Ohala J. (1970) Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. Folia Phoniatrica 22: 1- 20.

Miller D.G. (2008) Resonance in singing. Voice building through acoustic feedback. Inside View Press, Princeton, N.J.

Miller D.G., Schutte H.K. (1990) Formant Tuning in a Professional Baritone. Journal of Voice 4:231-237.

Neumann K., Schunda P., Hoth S., Euler H.A. (2005) The Interplay between Glottis and Vocal Tract during the Male Passaggio. Folia Phoniatrica et Logopaedica 57:308–327.

Schutte H.K., Miller D.G., M. D. (2005) Resonance strategies revealed in recorded tenor high notes. Folia Phoniatrica et Logopaedica 57.

Sundberg J. (1975) Formant technique in a professional female singer. Acustica 32:89-96.

Sundberg J., Lã F.M.B., Gill B.P. (2011) Professional male singers' formant strategies for the vowel a. Logopedics Phoniatrics Vocology. 36:156-167.

Titze I.R. (2004) A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice. Journal of Voice 18:292-298.

Titze I.R. (2008) Nonlinear source-filter coupling in phonation: Theory. Journal of Acoustical Society of America 123:2733–2749.

Titze I.R., Worley A.S. (2009) Modeling source-filter interaction in belting and high-pitched operatic male singing. Journal of the Acoustical Socitey of America 126:1530-1540.

FIGURE CAPTIONS

Figure 1. Example of Decap display. Upper panel shows the waveform of the filtered signal and the derivative of the EGG. Lower panel shows the input audio spectrum and the spectrum of the filtered flow. Formant bandwidths are given on an arbitrary scale along the ordinate. The arrows show the formant frequencies and bandwidths used for the inverse filtering. The two parallel curves represent realistic bandwidths according to Fant (1970).

Figure 2a. F1 and F2 observed for the vowel /a/ in the Classical and Non-classical versions of the scale (solid and dashed curves, respectively). Diagonal dashed lines refer to the frequencies of spectrum partials. The mean ratings are given in % of VAS length. The graphs for the /a/ vowel are taken from Sundberg & al, 2011.

Figure 2b. F1 and F2 observed for the vowel /a/ in the Classical and Non-classical versions of the scale (solid and dashed curves, respectively). Diagonal dashed lines refer to the frequencies of spectrum partials. The mean ratings are given in % of VAS length.

Figure 2c. F1 and F2 observed for the vowel /i/ in the Classical and Non-classical versions of the scale (solid and dotted curves, respectively). Diagonal dashed lines refer to the frequencies of spectrum partials calculated as n*MF0, where n is an integer and MF0 is F0 averaged over a set of adjacent complete vibrato cycles. The mean ratings are given in % of VAS length.

Figure 3. Distance between F1 and its closest partial and between F2 and its closest partial for the vowels /a, /i/, /u/ averaged across singers for the top pitch of the scale exercise. The frequency of the harmonic was calculated as n*MF0, where n is an integer and MF0 is F0 averaged over a set of adjacent complete vibrato cycles.

Figure 4. Mean ratings plotted as function of the minimum distance of F1 and F2 to their closest partial, for the top note of the scales. Filled and open symbols refer to tones intended as sung in Classical and Non-classical style.

Figure 5. Flow glottograms for the indicated relations between F0 and its closest harmonic observed for the indicated singers and vowels.

Figure 6. Audio and voice source spectra obtained from the Decap software for the vowel /a/ sung at the top tone of the scale by the indicated singers (see caption of Figure 1). The bold arrows show the frequencies and bandwidths of the inverse filters.

Figure 7. Levels of the first ten audio spectrum partials produced by the indicated singer and by the Madde synthesizer. The circles indicate the values observed for partials coinciding with either F1 or F2.

Figure 8. Audio and source spectra of the tones shown in Figure 5. The bold arrows represent the frequencies and bandwidths of the inverse filters (see caption of

Figure 1). In the case of the bottom left spectrum the bandwidth of the inverse filter corresponding to F2 was set to no more than 19 Hz, which is an unrealistic low value.

Figure 9. Audio spectra of the vowel /a/ as sung by singer 3 at the pitch of F#4 in Classical style. The left and right panels were taken at the peak and at the valley of the vibrato curve, respectively.