Formant strategies of professional female singers at high fundamental frequencies

Andrea Deme

Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary
Eötvös Loránd University, Hungary
deme.andrea@nytud.mta.hu

Abstract

When the soprano raises the fundamental frequency above the first formant of a vowel, a remarkable loss of acoustic energy and linguistic information occurs along with an abrupt change in the voice timbre. To avoid these effects, sopranos are assumed to tune their first formant to the raised fundamental frequency. The support for this claim is mostly based on formant data provided by indirect measurement methods and articulatory data, since direct acoustic data becomes more difficult (or even impossible) to obtain as the fundamental frequency gets higher. In the present study a new combination of measurement methods is introduced. The aim was to extract formant data of three sopranos in the entire set of the Hungarian vowel inventory in a wide pitch range. The results provide evidence for the technique of tuning the first formant to the raised high fundamental frequency in a substantial amount of data.

Keywords: formant tuning in soprano, inverse filtering, EGG, external excitation

1. Introduction

In high-pitched singing the fundamental frequency (f0 or h1) often exceeds the typical frequency region of the first formant (F1) of vowels. In these cases unaltered articulation of vowels would result in a remarkable loss of acoustic energy (i.e. sound level) and the loss of an acoustic cue (i.e. F1) which is regarded to be important for defining and identifying vowel qualitiy. Singers of the Western operatic style who are required to sing loudly (without any amplification) supposedly tend to compensate for these losses by changing the articulatory configuration of vowels while producing them at higher pitches. This way they can enhance the efficiency of resonance utilization. Several investigations have been conducted to provide data on the techniques singers use at high pitch regions to avoid the negative acoustic consequences of raising f0 above the F1 of the vowels in speech ($F1_{speech}$).

With respect to articulatory maneuvers Sundberg (1975), and Sundberg and Skoog (1995) concluded that pitch raising is accompanied by gradually increasing jaw opening in sopranos when f0 approaches the region of F1. Based on the traditional interpretation of the interrelation between articulatory movements and acoustics, the authors suggest that the articulatory data they obtained reflect the acoustic event of F1 “tuning” in an indirect way: they propose that F1 is tuned to f0 when f0 reaches the region of $F1_{speech}$ during pitch raising. Sundberg and Skoog (1995), however, also noted that the modification of jaw opening starts only at a higher f0 in the case of the close-mid vowels /eː oː/, whereas (in some singers) it is not clearly present in the close vowels /iː uː/. The authors claim that in the case of close vowels singers tend to tune their F1 as well as in all other cases, but this tuning is probably achieved by decreasing the degree of the tongue constriction (and not by modifying the jaw opening) for reasons of articulatory convenience. As opposed to Sundberg (1975), and Sundberg and Skoog (1995), a study by Bresch and Narayanan (2010) found that the dependency of F1 on f0 is only validated in the case of the close vowels /i u/. They also conclude that this dependency is probably more singer-dependent than it was suggested before. The authors base their claims on articulatory and formant data obtained from MRI-based vocal tract area function calculations.

Direct resonance data, on the other hand, are much more difficult to obtain for high-pitched singing. Due to wide harmonic spacing, the raised f0 causes undersampling of the vocal tract (VT) transfer function in the output sound. Consequently, analyzing the spectrum of the output signal is an inefficient way to determine formant frequencies (see e.g. Deme 2012). Therefore, the investigation of the acoustic characteristics of high-pitched sung vowels requires novel methodology to extract formant data in order to by-pass the problem of low resolution of the VT transfer function. The studies in this field concentrate mostly on the open vowel /a/ at particular pitch ranges. Sundberg (1975) used an external vibrator applied at the larynx for measuring five vowels at four fundamental frequencies below 700 Hz. Hertegård and Gauffin (1993) investigated the production of /a/ at 250, 390 and 750 Hz by inverse filtering the flow signal recorded with a Rothenberg mask. Joliveau et al. (2004) applied external excitation at the mouth for measuring four vowels in a pitch range below 1100 Hz. Garnier et al. (2010) used the same technique for investigating the vowel /a/ above 440 Hz. With regard to the cases when f0 > $F1_{speech}$ the above mentioned studies agreed on the tendency of F1: f0 tuning. However, Joliveau et al. (2004) also noted the lack of this tendency in vowels that use lip rounding above approximately 900 Hz. Some of these studies have also reported that the methods used might have also biased the data to some extent or required some modification of the singing technique intended to be observed: Hertegård & Gauffin (1993) noted that their subject complained that the mask hampered her in opening the mouth, while Joliveau et al. (2004) and Garnier et al. (2010) instructed their singer to sing very softly and without vibrato even at high pitches which might have also affected the control of formant tuning strategies.

The aim of the present study was to extend on the previous research by investigating the entire vowel set of a particular language in a wide pitch range (covering most of a soprano’s range), thus to discover the formant value changes accompanying pitch raising. This aim was achieved by the use of a new combination of some of the previously introduced methods, in a way that the problem of measuring formant frequencies at high f0s is resolved while the influence of the measurement method on the data obtained is minimalized. It was hypothesized that F1: f0 tuning occurs in each vowel in the cases of f0 > $F1_{speech}$ regardless of the degree of closeness or lip rounding. However, it was also suggested that F1: f0 tuning appears in correspondence with the closeness of the vowel during pitch-raising since the starting point of the tuning depends on the value of $F1_{speech}$.

2. Subjects, material, method

Three professional Western operatic female soprano singers were recorded producing nine Hungarian vowels /ɒ aː ɛ eː iː oː øː uː yː/ at six fundamental frequencies in three octaves from 175 Hz to 988 Hz (musical notes: F_3, B_3, F_4, B_4, F_5, B_5) and in speech. Although short counterparts of long vowels (/i o ø u y/) were not involved, and length is phonologically distinctive in Hungarian, the vowel set investigated here can be regarded as a representation of the entire vowel inventory. Since the duration of vowels is determined by the musical notes, length and quality distinction between short and long vowels become hard to interpret in singing. Each combination of vowels and fundamental frequencies was recorded twice in each singer's production which resulted in (3×9×7×2 =) 378 stimuli. Before recording one set of vowels, reference pitch was provided to the singer through headphones.

Two signals were recorded simultaneously. On the first channel the audio signal was captured by an omnidirectional microphone at 30 cm distance. On the second channel the signal of an electroglottograph (EGG) attached to the singer's neck was recorded. To resolve the problem that high-pitched sung vowels become undersampled due to wide harmonic spacing, low frequency external excitation (electrolarynx) was used during the recordings: while singing the vowel, the singers were instructed to phonate for a few seconds than turn the electrolarynx on, freeze the articulation and hold their breath (with closed glottis). This way at the end of each vowel the buzzer signal substituted the voice source and resampled the VT configuration. The output of this filtered buzz sound was also recorded with the microphone. An example of this "double sampling" is shown in Figure 1. The speech and EGG signals were time-aligned to compensate for the transit time of sound (from the larynx to the mouth) which was determined to be approximately 0.8 ms.

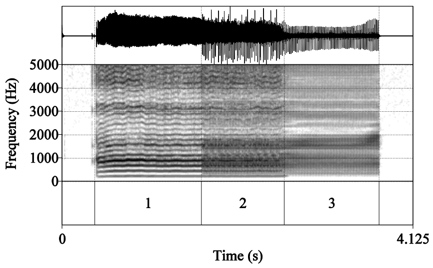

Figure 1: *Narrow-band spectrogram of the vowel /aː/ in speech sampled by the vibrating vocal folds (1) and the electrolarynx (3). In (2) the singer was phonating with the electrolarynx already switched on.*

The first (F1) and second (F2) formants of the vowels were determined by manual inverse filtering of the glottal flow signal (the pressure signal integrated) by means of the custom-made DeCap software (Svante Granqvist, KTH). The derivative of the EGG signal (dEGG) was also fed into DeCap and used to support the measurements (see Figure 2). The principle of the method is to compensate for the filter function of the VT, thus to restore the spectrum and waveform of the voice source by means of manually adjustable filters corresponding to the formants of the sound. In the case of correct adjustment the result is a smooth spectrum envelope and a smooth flow glottogram (with a ripple-free closed phase) that would be characteristic of the source signal (if it could have been captured at the glottis) (see Hertegård & Gauffin 1993, Sundberg et al. 2013). The accuracy of the analysis was increased by the use of dEGG which reflects the moment of vocal fold contact (as a peak in the signal), thus allowing to designate the maximum declination rate of the transglottal airflow and the starting point of the closed phase needed for adjusting the filters (see Henrich et al. 2004).

To verify formant value estimations the electrolarynx recordings were also filtered (in phase 3 displayed on Figure 1) and the results were compared with the formant and bandwidth frequencies obtained by the filtering of the flow. Therefore, the reliability of the measurements was enhanced. The two repetitions of each stimulus were measured separately and averaged in the analysis.

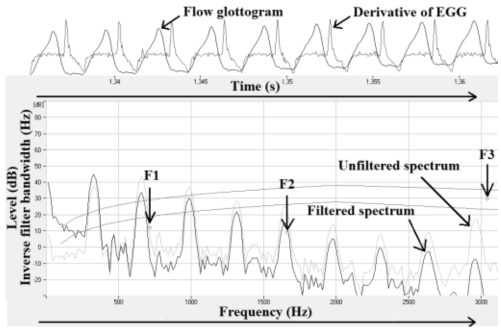

Figure 2: *Inverse filtering with DeCap (after Sundberg et al. 2013). The upper panel shows the inverse filtered signal and the dEGG (dark and light colors, respectively); the lower panel shows the audio spectrum, and the spectrum of the filtered flow (light and dark colors, respectively). The arrows with the labels F1, F2, F3 show the first three formants adjusted for inverse filtering.*

3. Results

One of the three sopranos did not manage to perform each of the stimuli at the highest fundamental frequency B_5 (988 Hz). Therefore, some data points are missing in Figure 3 and the displacement of vowels in the acoustic vowel space accompanying pitch raising was evaluated based on the comparison between speech and the second highest f0, i.e. F_5 (698 Hz) where no missing data were present.

3.1. Vowel formants in singing: F1

Figure 3 shows the first and second formants of the nine vowels as a function of the fundamental frequency for each singer separately and the harmonics of the f0 (dashed lines). On this figure, apart from moderate variability among the singers, clear-cut tendencies can be observed. When f0 approached (but not necessarily reached) the region of $F1_{speech}$, the singers tuned F1 to the raised f0 in each vowel. From that point on, the tuning of the F1 was parallel to the raising of the f0. In other words, no exception of F1 tuning was found at high f0s, and the lower limit of the tuning was dependent on closeness: it was the highest (698 Hz) for the open and open-mid vowels /aː ɒ ɛ/, lower (494 Hz) for the close-mid vowels /oː øː eː/, and the lowest (349 Hz) for the close vowels /uː yː iː/. Below these critical f0 values F1 was more independent of f0 in each vowel and it was realized in the vicinity of the value of $F1_{speech}$.

The results seem to be partly inconsistent with the findings of Joliveau et al. (2004), as they noted the failure of the F1: f0 tuning in the case of vowels produced with lip rounding. However, they observed this tendency around and above 1 kHz, whereas in the present study only 988 Hz was reached.

3.2. Vowel formants in singing: F2

The value of F2 was more independent of pitch, especially in the case of front vowels /aː ɛ/. In the front vowels /eː iː yː/ a slight decrease of F2 was observed starting at f0 where F1: f0 tuning also begins. In the back vowels /ɒ oː uː/, however, a stronger decrease of the F2 values was observed.

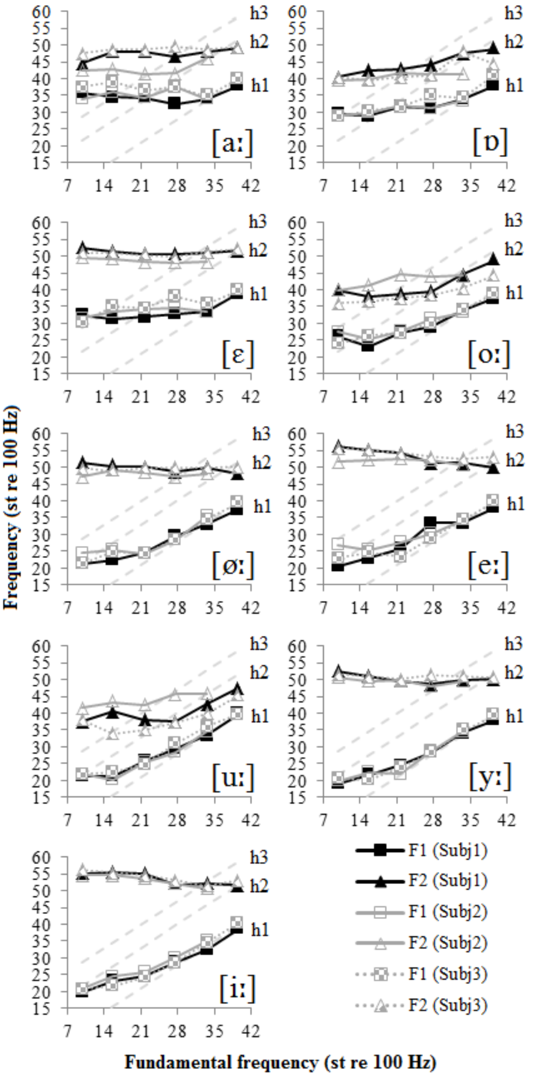

Figure 3: *Formant values (F1, F2) of the three sopranos as a function of the fundamental frequency. Each marker corresponds to a formant frequency of one singer at a particular fundamental frequency. The grey dashed lines represent the first three harmonics of the voice source (h1, h2, h3).*

These tendencies mean that front and back vowels start to converge at the higher f0s (namely from about F_4 or B_4, or approximately 350–500Hz), and that the degree of the change in F2 is dependent on the degree of backness: among the front vowels it is /eː iː/, while among the back vowels it is /oː uː/ that is affected the most.

3.3. The changes of the acoustic vowel space

In terms of the acoustic vowel space, the centralization tendency described in 3.1 and 3.2 results in the collapse of the front–back distinction and in a shift towards the position of /aː/ as seen in Figure 4. According to A of Figure 4, vowels in speech are well-separated along the two axes (corresponding to the front–back and close–open dimensions, respectively). In singing at F_5 (698 Hz), however, only the front–back distinction is preserved to some extent.

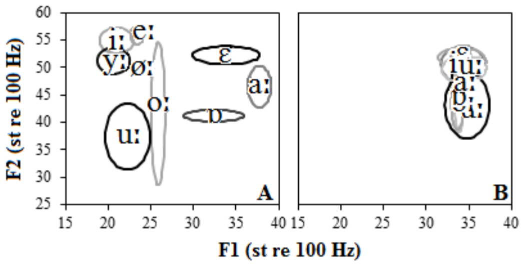

Figure 4: *The displacement of the vowels as a function of the fundamental frequency demonstrated by the acoustic vowel spaces of the three sopranos in speech (~200 Hz) (A) and in singing at F_5 = 698 Hz (B). The ellipse is defined by the mean of the extracted formant frequencies (center), and the standard deviation (± 2SD in both axes).*

4. Discussion and conclusions

The presented study focused on formant tuning strategies of sopranos trained according to the Western operatic singing technique. The singers of this technique are required to provide a homogenous timbre through their entire pitch range and a loud voice without amplification. When f0 is higher than F1 of a vowel in speech (as often happens in sopranos), the acoustic energy of the radiated sound decreases remarkably, thus even more vocal effort is being needed to achieve the desired sound level. Increasing the subglottal pressure alone to increase sound level would strain the vocal folds, hence it must be avoided. Effective utilization of resonances, however, can help the singer to supplement vocal effort. Consequently, resonance strategies are very important in sopranos' practice.

Based on the results, we can conclude that all professional sopranos in our study employed generalizable strategies for resonance tuning in a way that is consistent with the description in prior literature. It was demonstrated in a substantial amount of material (the entire set of the Hungarian vowels, in a wide pitch range, in 3 singers' production) that professional sopranos tend to tune their F1 to or slightly above the frequency of f0 if the raised f0 approaches the F1 of the vowels in speech. It should be emphasized, however, that the articulatory strategies singers might have applied to achieve this goal are not extractable from these acoustic data alone. In other words, the present data are only indicative of resonance strategies and not the articulatory strategies that the singers use in their practice. Accordingly, the debated issue whether singers lowered the tongue or increased the jaw opening to raise F1 in close vowels (see further in Sundberg and Skoog 1995) cannot be answered here.

The lower limit of F1: f0 tuning was dependent on the closeness of the vowel in accordance with the assumption that F1: f0 tuning starts when f0 approaches the region of F1 typical of the vowel in speech. Indeed, in close-mid and close vowels F1 tended to be increased steeply through the studied pitch range, parallel to pitch raising. In the present study one of the results of Joliveau et al. (2004), namely the lack of F1: f0 tuning in rounded vowels at high f0s was not replicated. This inconsistency might shed light on the connection between formant tuning and the pitch range of singers. In the study of Joliveau et al. the failure of the tuning was observed above 900 Hz, while most of the singers reached 1042 Hz during the experiment which means that the upper boundary of the singers (most probably) lied above the critical 900 Hz. By contrast, the singers in the present study reported that the upper boundary of their ranges was reached at the highest fundamental frequency studied (988 Hz). Garnier et al. (2010) noted that there is an upper limit of F1: f0 tuning close to 1100 Hz, above which the tuning of F2: f0 starts, if this region still belongs to the singer's pitch range. On this basis, two explanations can be suggested. One possibility is that the singers of the present study reached the top of their ranges at 988 Hz, thus they *did not need* any other strategy to reach higher pitches with the required voice quality and loudness; however, they managed to expand the limit of F1: f0 tuning in vowels with lip rounding to a small extent. The second explanation is that the singers reached the top of their ranges *because* they did not have any other strategies which would have been able to substitute F1:f0 tuning. The adequacy of these suggestions needs further investigation.

The changes of F2 differ systematically according to backness. Also, they were found to accompany F1: f0 tuning. While front vowels' F2 decreases slightly, F2 of back vowels increases to a greater extent. It can be suggested that the changes of F2 are the concomitant result of the increased jaw opening (as Joliveau et al. 2004 suggest). As the mandible is lowered, the constriction caused by the tongue is less and less narrow. To some extent this "loss" can be compensated by increasing the height of the dorsum. However, in their articulatory modeling study Lindblom and Sundberg (1971) concluded that at a sufficient jaw opening (at 23 mm in their data) this loss of constriction cannot be compensated for anymore; consequently, at this degree of opening no velar articulation is possible either. The authors claim that front vowels are also affected by jaw opening, but only to a lesser extent. The widening of the constriction results in higher F2 in back vowels and a slight lowering of F1 of the front vowels. Therefore, the decrease of the acoustic separation between front and back vowels is expected, as demonstrated in the present data. The changes of F1 and F2 resulted in reduced acoustic vowel space and a shift towards the position of /aː/ as the f0 increased.

What might be the consequences of these acoustic changes regarding perception? Although in the present data vowels converged and overlapped more and more with pitch raising, it was also observed that the front–back distinction was preserved at even moderately high f0s (i.e. 698 Hz). Accordingly, it can be expected that the distinction of the back and front vowels will uphold longer with pitch-raising also in perception (thus confusions in the identification task would include mainly back–back and front–front pairs), as demonstrated already in Deme (2012) and Deme (in press). Nevertheless, relating the exact formant data of the present study to perceptual tendencies in further detail may easily be misleading. We should bear in mind that the researcher faces a very difficult task in extracting formant data from the output sound in singing (particularly above about 400-500 Hz). The undersampling of the vocal tract transfer function is, however, not only influencing the efficiency of measurement techniques, but also speech perception. That is, at high pitch no exact formant frequencies are extractable for the auditory perception either: at a particular f0 only the same harmonics with different amplitude are available. For that reason, the author's ongoing work also includes vowel identification tests on the present material which correlates the perceptual tendencies to the raw acoustic output, as well as to the formant data obtained here. These two approaches are supposed to clarify jointly which of the acoustic parameters may account for the perceptual tendencies accompanying pitch raising. The preliminary results suggest that listeners might be able to extract formant values effectively (and rely on them in vowel identification) even at the fundamental frequency of F_5 (698 Hz). At B_5 (988 Hz), on the other hand, they simply seem to identify the first two harmonics as F1 and F2, respectively, in each of the vowel qualities produced by the singers.

Manual inverse filtering is a method that has already proved to be successful in formant value estimation in male singing voice on fundamental frequencies below 500 Hz (see Sundberg et al. 2013) and also in sopranos for the vowel /a/ (Hertegård & Gauffin 1993). In the present study inverse filtering was supported by the derivative of the electroglottograph signal (dEGG) according to the suggestion of Henrich et al. (2004) and Sundberg et al. (2013). The reliability of the measurements was enhanced by the use of external excitation (electrolarynx) which allows for measuring formant frequencies independently of the fundamental frequency of the voice. The combination of these methods appeared to be successful in the formant value estimation in singing, even at very high pitches. The present study reports the first attempt to describe the acoustic modifications caused by high pitched singing on the entire vowel set of a particular language.

5. Acknowledgements

The help and support of the following personnel and institutes is gratefully acknowledged. The recordings were made in the Kempelen Farkas Speech Research Lab of the Research Institute for Linguistics, Hungarian Academy of Sciences. The acoustic measurements were carried out in the Department of Speech, Music and Hearing of KTH Royal Institute of Technology, Sweden, with the help of prof. emer. Johan Sundberg, prof. Sten Ternström, and Svante Granqvist, PhD on a study visit funded by the Campus Hungary program of the Balassi Institute, Hungarian Scholarship Board Office.

6. References

Bresch. E. and Narayanan, Sh. (2010). "Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing". In: *Journal of the Acoustical Society of America* 128.5, pp. EL335–EL334.

Deme, A. (2012). "On the Hungarian sung vowels". In: *Phonetician* 1–2.105–106, pp. 73–87.

Deme, A. (in press) "Intelligibility of sung vowels: The effect of consonantal context and the onset of voicing". In: *Journal of Voice.*

Garnier, M., N. Henrich, J. Smith and J. Wolfe (2010). "Vocal tract adjustments in the high soprano range". In: *Journal of the Acoustical Society of America* 127.6, pp. 3771–3780.

Henrich, N.,Ch. d'Alessandro, B. Doval and M. Castellengo (2004). "On use of the derivative of electroglottographic signals for characterization of nonpathological phonation". In: *Journal of the Acoustical Society of America* 115.3, pp. 1321–1332.

Hertegård, S. and J. Gauffin (1993). "Voice source – vocal tract interaction during high-pitched female singing". In: *SMAC* 1993, 177–182.

Joliveau, E., J. Smith and J. Wolfe (2004). "Vocal tract resonances in singing: the soprano voice". In: *Journal of the Acoustical Society of America* 116.4, pp. 2434–2439.

Lindblom, B. E. and Sundberg, J. (1971). "Acoustical consequences of lip, tongue, jaw, and larynx movement". In: *Journal of the Acoustical Society of America* 50.4, pp. 1166–1179.

Sundberg, J. (1975). "Formant technique in a professional female singer". In: *Acta Acustica united with Acustica* 32.2, pp. 89–96.

Sundberg, J., F. Lã and B. P. Gill (2013). "Formant tuning strategies in professional male opera singers". In: *Journal of Voice* 27.3, pp. 278–288.

Sundberg, J. and J. Skoog (1995). "Jaw opening, vowel and pitch". *STL-QPSR* 36.2–3, pp. 43–50.