

Eötvös Loránd University
Faculty of Humanities

THESES OF THE PHD DISSERTATION

Phonetic Analysis of Sung Vowels

by

Andrea Deme

Doctoral School of Linguistics

Head of doctoral school: Prof. Dr. Vilmos Bárdosi CSc

Applied Linguistics Doctoral Program

Head of doctoral program: Prof. Dr. Mária Gósy DSc

Committee:

Chair: Prof. Dr. Mária Gósy DSc

Opponents: Prof. Dr. Klára Vicsi DSc

Dr. habil. Judit Bóna PhD

Secretary: Dr. Dorottya Gyarmathy PhD

Members: Dr. György Szaszák PhD;

Prof. Dr. Gábor Olaszky DSc, Prof. Dr. Anna Adamikné Jászó DSc

Supervisor: Dr. habil. Alexandra Markó PhD

Co-supervisor: Prof. emer. Dr. Johan Sundberg PhD

Budapest, 2015

Introduction

The research presented in the dissertation is based on the apparent contradiction between the requirements of vowel articulation and the requirements of the production of high-pitched singing voice. In accordance with this contradiction, in the following experiments both the production and perception of sung vowels is examined on Hungarian vowels.

Vowel quality, i.e. the basis of vowel differentiation is based upon the resonance frequencies, or in other words, formants of the vocal tract during articulation (see Fant 1960). Among these resonances the first and second formants (F1 and F2) have a distinguished role, since these resonances strongly correlate with certain articulatory maneuvers, and thus they are also regarded to be definitive of vowel quality. F1 correlates inversely to (tongue) height or directly to openness (i.e. jaw opening). (For example the close /i:/ has a lower F1 than the open /a:/.) F2 corresponds to backness: F2 of back vowels is lower than that of front vowels. (For example, the back /u:/ has a lower F2 than the front /i:/, see e.g. Gósy 2004). As a result, the F1 of close vowels may be relatively low in frequency, or, by all means, lower than that of the fundamental frequency (f_0) or pitch that is often used by soprano singers.

Therefore the question arises: what happens if sopranos sing at an f_0 that is higher than the F1 of the vowel in speech? Does F1 disappear from the spectrum of the vowel? This scenario is very unlikely, since it would result in a great loss of acoustic energy (i.e. loudness), in a drastic change in timbre, and it would also mean that certain vowel qualities also disappear at high f_0 s. In the opera performances, however, we do not encounter most of these phenomena: the singing voice of a soprano is loud (without any further amplification), homogenous in timbre, and the sung text is also seemingly comprehensible. Hence, in high pitched singing the modification of vowel articulation is presumed.

The prevalent suggestion in the literature regarding sung vowel production is that in those cases, when the f_0 would exceed the frequency of the F1 of the vowel in speech, singers tend to tune the F1 to (or slightly above) the raised high f_0 (notation: $F1 : f_0$). Consequently, sung vowels are produced more and more open if the pitch is raised (Sundberg 1975, Johansson et al. 1982, Sundberg & Skoog 1995, Bresch & Narayanan 2010, Hertegård & Gauffin 1993, Joliveau et al. 2004, Garnier et al. 2010). As a result of the $F1 : f_0$ tuning, singing voice may gain even 20-30 dB in sound pressure level, i.e. loudness (Sundberg 1979, 1987), which seems to be a strong argument in favor of the widespread suggestion on formant tuning. Empirical evidence, however, is very difficult to gather to support it due to the spectral undersampling of the vocal tract transfer function at high f_0 (see e.g. de Cheveigné & Kawahara 1999). The spectral undersampling effect makes the traditional ways of formant frequency estimation (i.e. the Fourier analysis of the radiated sound) considerably unreliable. Therefore, determination of the formants in high-pitched sung vowels requires alternative solutions (e.g. articulatory modeling, the usage of external excitation during the recordings etc.). Yet, these methods may also exhibit

more problems or even errors in the analysis (see e.g. Bresch & Narayanan 2010, Hertegård & Gauffin 1993).

Since the $F1 : f0$ strategy changes the articulatory configuration of certain vowels at a high $f0$ (and the acoustic manifestation of vowels also changes as a result of the spectral undersampling effect), the question arises, how high-pitched sung vowels are perceived. The question is, to some extent, neglected in the literature, and thus we find only a limited amount of results on particular vowel qualities in particular languages. According to the apparently most accepted assumptions, i) the identification of sung vowels (i.e. the identification of the intended vowel quality) decreases with pitch-raising and ii) the vowels produced with a higher $F1$ (with a greater degree of openness) are also perceived as more open vowels (Scotto di Carlo & Germain 1985, Benolken & Swanson 1990, Hollien et al. 2000). On the basis of speech perception (but without firm evidence), it is also often suggested that sung vowels are perceived at a higher rate in accordance with their intended (but not actually realized) quality if uttered in CVC context. The sole proof for this suggestion is a study by Smith and Scott (1980). However, as most of the above studies are not well-controlled regarding some important factors, the conclusions of the authors on the perception of sung vowels are to be handled with precautions and the issues of the identification of high-pitched sung vowels are still an open question.

Aims, questions, hypotheses

The aim of the studies presented here is to describe the acoustic and perceptual properties of high-pitched sung vowels in Hungarian. On the basis of previous literature and the associated problems, the following questions were formulated (and organized in four groups). Questions regarding the acoustics of sung vowels:

1. What are the acoustic consequences of high-pitched singing in Hungarian vowels, particularly in those vowels in which $F1$ would be exceeded by the high $f0$? What are the properties of the radiated sound (i.e. the frequency of the spectral maxima) and the $F1$ and $F2$ resonances of the vocal tract during sung vowel articulation?

Since the investigation of these questions requires an alternative measurement method to by-pass the issue of spectral undersampling, another problem to resolve in the dissertation is that some previously introduced measurement methods for formant data extraction must be adapted and combined in high-pitched singing.

Questions regarding the perception of sung vowels:

2. How are high-pitched sung vowels perceived? How does vowel identification change according to pitch-raising? Are open vowels perceived at higher rate in accordance with their intended quality at high $f0$ than more close ones? What are the characteristics of the “errors” (i.e. mismatches between the perceived and intended vowel qualities) accompanying pitch-raising? Is it true that production and perception exhibit a direct and linear relationship, thus vowels produced with a greater degree of openness (because of the $F1 : f0$ tuning at higher $f0$ s) are also perceived as more open?

3. How does consonantal context affect the identification of high-pitched sung vowels? Do the acoustic cues originated in coarticulatory formant transitions affect vowel identification? How does consonantal context affect vowel identification, if the carrier sequence is a sense word, hence higher level (top-down) perceptual processes are also activated during vowel identification and support lower level (bottom-up) perceptual processes?

4. How does the voicing onset affect the identification of sung vowels? Psychoacoustic experiments have provided evidence that musical instruments are better identified if the sound contains the onset. Is this argument also applicable in the case of sung vowels? That is, does the presence of the voicing onset lead to the result that vowels are identified at a higher rate in accordance with their intended quality?

5. Do the perceptual tendencies of naive listeners differ from those of singers? Do singers (due to their experience in the production of singing voice) identify sung vowels at a higher rate in accordance with their intended (but not necessarily actually realized) quality? These questions (along with those above regarding consonantal context) address the issue whether any bottom-up or top-down perceptual compensation is possible in the identification of formant-shifted high-pitched sung vowels.

In connection with our questions, the following three hypotheses were formulated:

H1: Sopranos raise the F1 to (or slightly above) the raised high f_0 , if otherwise f_0 exceeded the F1 of the vowels. Accordingly, the lower limit of F1 : f_0 tuning is dependent on the degree of openness or, in other words, on the F1 of the vowels in speech.

H2: Due to the spectral undersampling of high-pitched sonorant sounds and the articulatory modifications of vowels accompanying pitch-raising, the identification of sung vowels decreases with increasing f_0 . As a result of the decreasing tonotopic difference between f_0 and F1 that also accompanies pitch-raising, sung vowels are at a higher rate identified as more close vowels if the f_0 is raised.

H3: Due to the spectral undersampling of high-pitched sonorant sounds (and coarticulatory transitions) and the articulatory modifications of vowels accompanying pitch-raising, consonantal context does not have a positive effect on vowel identification through transitional (i.e. acoustic) cues. The same applies to the onset of voicing. Consonantal context may only have any positive effect on sung vowel identification, if the carrier sequences in which the vowels are embedded are sense words and they allow much less possible choices than the number of the Hungarian vowels¹.

¹ Please note that in the present study “identification of sung vowels” and “correct identification of sung vowels” always refer to “identification according to the intended vowel quality”. It is a central claim of the dissertation that the acoustic properties, that is, the quality of vowels do change with the increasing f_0 , thus the produced quality of vowels becomes ambiguous. Consequently, the identification of these vowels cannot be considered neither “correct” nor “incorrect” as done by previous literature, but it is a question of analysis (along with the nature of the corresponding acoustic changes).

Experiments

Production and perception of Hungarian vowels in singing

The first aim of the first study was to analyze the acoustic parameters that are extractable from high-pitched vowels employing traditional (Fourier-transformation-based) acoustic analysis, that is, to detect the spectral maxima of the radiated sound, and to observe their changes that accompany pitch-raising. The second aim was to investigate the perception of sung vowels and its changes with increasing f_0 .

Subjects, material, methods

In the study all the 9 Hungarian vowel qualities were analyzed / b a : e e : i o : $\text{\text{ø}}$ u : y :² in mVn consonantal frames in one soprano's production (age: 50 years). The soprano produced the mVn sequences in speech (~200 Hz) and in singing (f_0 = 500 Hz, 550 Hz, 650 Hz). In the perception test 11 non-trained adults participated (4 men, 6 women, 21 to 25 years of age, M = 22 years). The task of the participants was to identify the vowels they heard and to designate it on an answer sheet. Stimuli were presented in a randomized order. Statistical analysis was performed on the perceptual data with the SPSS 13.0 software³. Acoustic analysis of the vowels spoken and sung by the soprano was carried out by the Praat software (Boersma & Weenink 2009): we estimated the frequency values of the spectral maxima of the radiated vowel sounds⁴.

Results

The results on the perception test are shown in Figure 1. According to the data, the identification of sung vowels (and the identification of distinctive vowel features) decreased with the f_0 -increase, but the decrease was not gradual. These results mean that vowels were identified in an increasingly smaller rate (according to the intended vowel quality)⁵ as the f_0 increased. (One data point on the figure represents all the responses of all the subjects at a given f_0 , that is, 9 vowels \times 4 fundamental frequencies \times 11 subjects = 396 responses correspond to 100%).⁶ Comparison of the identification rates of labiality and the degree of openness also revealed that labiality was significantly more resistant to f_0 -increase than openness (ANOVA, $F(2) = 8.34$; $p = 0.02$). It was also observed that responses became more and more scattered as the f_0 increased.

² As singers tend to have problems with singing short counterparts of long vowels, these were not included in our studies. Yet we can claim that by including only the long counterparts we studied all the 9 different qualities of the Hungarian vowel system (see e.g. Gósy 2004).

³ In the case of all the presented studies, the following statistical analysis was employed. Normal distribution of the data was tested with the Shapiro–Wilk-test. In the case of normal distribution, parametric tests were further employed: ANOVA, repeated measures ANOVA, t -test. In the case of non-normal distribution, non-parametric tests were employed: χ^2 -test. Correlations were tested with Pearson's test of correlation.

⁴ Linear Hz scales were transformed into nonlinear Bark scale after extraction.

⁵ „Intended vowel quality” always refers to the list of vowels the singer was asked to produce during the recordings.

⁶ The results on spoken stimuli are presented separately (here and elsewhere) as these stimuli are not simply lower in f_0 but also different in their mode of production.

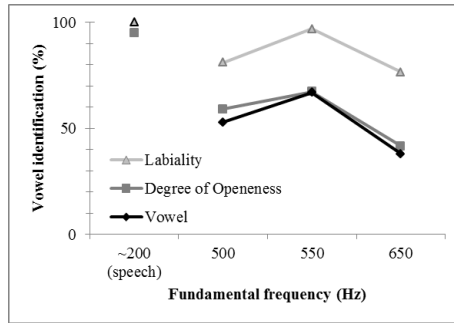


Figure 1: The identification of vowels and vowel features (degree of openness and labiality) as a function of f_0

Figure 2 shows the responses to the vowels which are produced with the second and third degrees of openness / ϵ e: \varnothing : o:/ . These vowels have both more open and more close neighbors in the Hungarian vowel system.

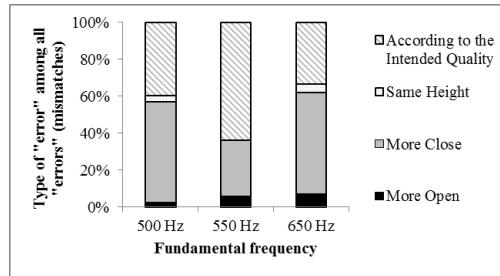


Figure 2: Responses to the vowels produced with the second and third degrees of openness / ϵ e: \varnothing : o:/ as a function of pitch

According to the prior literature on the perception of sung vowels, production and perception exhibit a direct and linear relationship: as the degree of openness increases with f_0 -increase in production (as a result of $F_1 : f_0$ tuning), vowels are also identified at a higher rate as more open ones. To address this issue a comparison presented in Figure 2 is the most adequate solution (but it is important to note that the above suggestion is not based on a similar analysis as the comparison is unprecedented in prior literature). The comparison of the responses in Figure 2 revealed that if vowels were not identified according to the intended quality, they were mostly identified as more close vowels.

Figure 3 shows the first three spectral maxima (that are, in sense (i), the first three formants F_1 , F_2 , F_3 . see Fant 1960) obtained by the frequency analysis of the sound pressure function. Although at lower f_0 s the frequency value of these prominences may also correspond to the resonances of the vocal tract (that are the formants in sense (ii), see also Fant 1960), at higher f_0 (as a result of the spectral undersampling effect) they only reflect the voice source harmonic/partial that falls within the bandwidth of a resonance, and thus becomes enhanced.

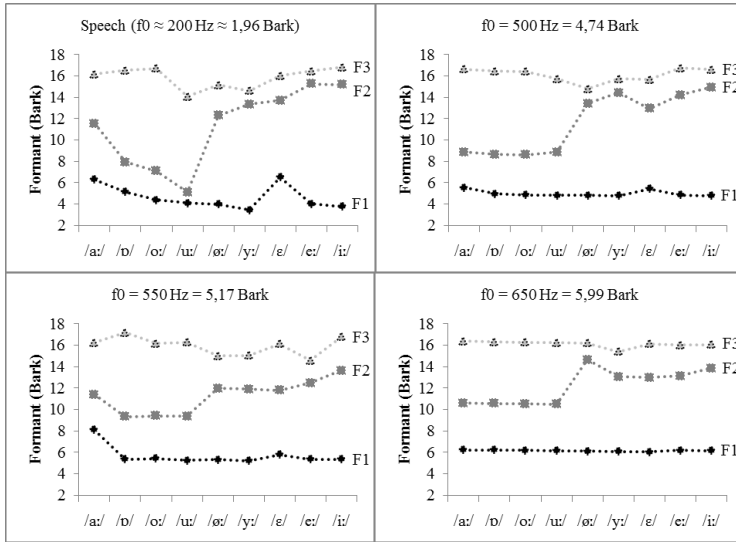


Figure 3: Formant values as spectral maxima (F1, F2, F3) of the spoken and sung vowels

The acoustic data suggest that in those cases, when the f_0 would have exceeded the F1 of the vowels (i.e. at 500 Hz in the present experiment), the singer tuned the F1 to the raised high f_0 . The tuning is suggested on the basis of indirect evidence: from 500 Hz on during the f_0 -increase, f_0 and F1 became indistinguishable, while f_0 became the most prominent partial in the spectrum which means that f_0 was enhanced by a vocal tract resonance (F1) on the f_0 s at hand.

Discussion

The perceptual results (obtained in a well-controlled comparison) falsified the claims of previous literature (suggested mainly on the basis of not well-controlled comparisons and speculation): although vowels produced with an increased degree of openness were identified to a smaller extent according to the intended quality, they were not identified as more open ones but as more close ones, as the f_0 increased. The most generalizable tendency found was the increased degree of scatter in the responses reflecting that the produced quality of the sung vowels changed gradually (not categorically). On the basis of the above observations we concluded that the direct and linear relationship suggested between the production and perception of sung vowels by previous literature is not correct, and these interdependencies are more complex than suggested. Acoustic data obtained through the analysis of the radiated vowel sounds provided indirect evidence of the F1 : f_0 tuning.

The effect of consonantal context on the identification of sung vowels: the effect of the manner of articulation

In the second study we addressed the question whether the perceptual tendencies found in the previous study may be accounted for by the nasal context. Although the

literature on the perception of nasalized vowels in speech suggests that nasalization may result in the perception of more close vowels, Krakow et al. (1987) also showed that nasalization in a predictable context may also be compensated for in the perceptual system. To test the research question, vowel perception in nasal context was compared to voiced and unvoiced obstruent contexts.

Subjects, material, methods

For the perception test the soprano participant of the first study produced the 9 Hungarian vowels in *mVn* /m-n/ nasal context, and in *zsVzs* /ʒ-z/ and *sVs* /ʃ-ʃ/ fricative contexts in speech and in singing at the f0 of 500, 550, 600, 650 Hz. Again, a perception test was carried out on the recorded material (similar to the previous one) involving 15 participants (10 women, 5 men, 21-29 years of age, M = 23 years).

Results

Figure 4 shows the most important results: the responses to the vowels produced with the second and third degrees of openness. According to the data when /ɛ e: ø: o:/ were not identified according to the intended quality, they were identified as more close vowels irrespective of the context. Moreover, nasal context showed a slightly (but not significantly) greater ratio of “correct” identification of vowels.

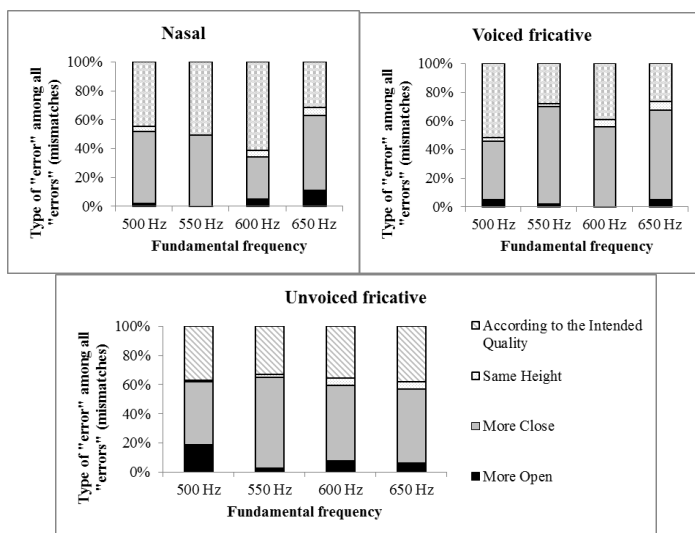


Figure 4: Responses to the vowels produced with the second and third degrees of openness /ɛ e: ø: o:/ as a function of pitch

Discussion

Based on the results we concluded that perceptual tendencies of sung vowels observed in nasal context did not differ remarkably from those of other obstruent contexts. The results strengthen our claim that the suggestion of previous literature on direct and linear relationships between the production and perception of sung vowels

is not correct, and vowels produced with a greater degree of openness at higher f_0 s are not necessarily perceived as more open vowels.

The effect of consonantal context on the identification of sung vowels: vowel identification in consonantal context, in isolation, and in the absence of the voicing onset

In our third study we re-examined the claims of Smith and Scott (1980) regarding the positive effect of consonantal context on the identification of high-pitched sung vowels. In their study Smith and Scott (1980) argued that consonantal context supports the identification of the inbetween vowel due to dynamic cues (i.e. formant transitions). In their experiment, however, they tested conditions that were not balanced in number and thus their argument is not grounded satisfactorily. The positive role of the voicing onset was hypothesized on the basis of prior psychoacoustic studies that showed the positive role of the onset of complex sounds in recognition of musical instruments, i.e. timbre recognition (see e.g. Berger 1964).

Subjects, material, methods

The material consisted of the three most spaced vowels of the Hungarian vowel system /a: i: u:/ in a soprano's production (age: 27 years) in speech and in singing in three octaves covering the f_0 -range of the soprano (from 175 to 988 Hz, on the musical notes *h*, *f*, *h'*, *f'*, *h''* and *f''*). The singer produced the vowels in isolation (*V* condition) and in *bVb* /b-b/ context (*CVC* condition). The stimuli for testing the absence of the onset (*Cut* condition) were prepared by a sound editing program (the onset was eliminated by logarithmic fading-in effect). The stimuli were tested in a perception test involving 22 participants (21-35 years of age, $M = 29$ years).

Results

The main results of the study are summarized in Figure 5. The data showed that the identification of vowels was only affected by the f_0 (ANOVA: $F(1) = 33.33$, $p < 0.001$), but not by the *condition* or the *vowel quality* factors. The relationship between f_0 and the ratio of identification was characterized by a medium strength inverse correlation (Pearson's $r = -0.631$, $p < 0.001$).

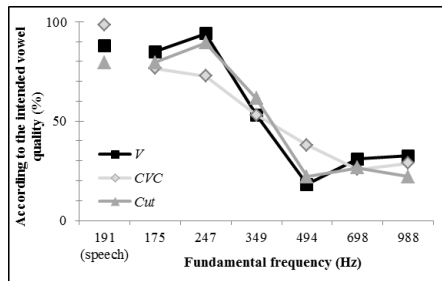


Figure 5: Identification of the three vowels /a: i: u:/ according to the intended vowel quality in the three conditions as a function of pitch

Discussion

The results did not verify the claims that consonantal context, i.e. formant transitions and the onset of voicing play a positive role in the identification of the intended vowel quality in singing. These findings are consistent with our suggestion that acoustic vowel qualities gradually change with pitch-raising, thus the intended vowel qualities cannot be cued merely by acoustics and the changes cannot be compensated for at the level bottom-up perceptual processes.

The effect of consonantal context on the identification of sung vowels: vowel identification in sense and nonsense words

In the fourth study we further investigated the role of consonantal context in sung vowel identification. As in their study Smith and Scott (1980) compared two conditions that were not balanced in number, on the basis of their results we suggested that the intended vowel quality may be perceived at a higher rate if vowels are uttered in sense words. In the fourth study we thus investigated whether the intended vowel qualities are better identified in sense words than in nonsense words, even if the number of possible answers is more balanced, i.e. similarly great in number. We suggested that under these conditions the advantage observed in sense words disappears (due to the acoustic changes of vowels at high f_0). The second question of the study addressed further possible factors that may positively affect the identification of the intended vowel quality. We compared the identification of sung vowels in non-singers and singers to establish whether the practice in production that is attributed to singers may also be incorporated in the perceptual processes and turn into a perceptual advantage.

Subjects, material, methods

In the study a two-phase listening test was conducted involving 20 *singers* (26-45 years of age, $M = 31$ years) and 20 *non-singers* (26-52 years of age, $M = 29$ years). The material consisted of the vowels / v a : e i : o : y :/ in *sense* words (szVr / s-r/ , e.g. *szár* ‘stalk’) and in *nonsense* words (dVr / d-r/ , e.g. *dár*) sung by the soprano of the preceding study on the f_0 s also indicated in the previous section. In the first phase each participant listened to the stimuli of the *sense* condition, and then in the second phase they listened to the *nonsense* stimuli. In both phases the subjects were presented to all the 9 Hungarian vowels as possible candidates (so that they could also provide illegal responses), but they were also provided the list of legal responses in the case of the *sense* condition and the C_C frame in both phases. Thus, in the case of the *sense* condition listeners may have had a slightly higher chance to guess the vowel quality “right” than in the case of nonsense condition (13% and 11% chance, respectively).

Results

The main findings are summarized in Figure 6. Statistical analysis supported the apparent tendency that neither *condition* (*sense* vs. *nonsense*) nor the *type of the listeners* (*non-singers* vs. *singers*) had a significant effect on the identification of sung vowels. The main tendency in both conditions and groups resemblances the

tendencies observed previously, i.e. that the identification of the intended quality decreased gradually (as the scatter of the responses increased) as a function of pitch.

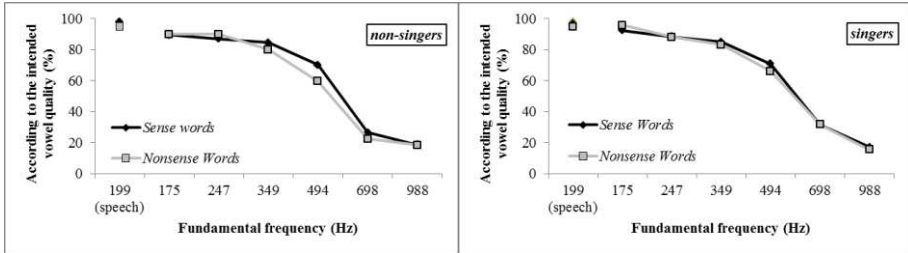


Figure 6: Vowel identification according to the intended vowel quality in the two conditions and in the two groups as a function of pitch

Discussion

The experiment provided evidence that despite the fact that sense words may facilitate a higher rate of identification of the intended vowel qualities (see Smith & Scott 1980), this positive effect disappears if the number of possible choices is more balanced in the sense and nonsense carrier sequences or in the task where vowels are to be identified in isolation (which resembles vowel identification in nonsense words). The second conclusion is that despite their experience in production, singers do not enjoy any perceptual advantage in the identification of the intended vowel quality in singing. This finding also supports the claim that the acoustic modifications of vowels accompanying pitch-raising cannot be compensated for at the level of bottom-up perceptual processes.

Acoustic properties of the Hungarian vowels in singing in the light of a new methodological solution, and the interdependencies of acoustics and perception in Hungarian vowels in singing

The fifth study focused again on both the production and perception of the Hungarian vowels in singing. First, we combined some previously introduced recording protocols and measurement methods to by-pass the spectral undersampling problem of high pitch and to measure vocal tract resonances (i.e. formants) independently of the voice source. Second, the recorded material was tested in a perception experiment and some interrelations between production and perception were concluded. The study was uniquely employing both the experimental results and perception theory of Traunmüller (1981), and acoustic data to interpret perceptual results. Traunmüller (1981) suggested that perceived openness depends on the tonotopic (i.e. tonality) distance between f_0 and F_1 , and that above a critical f_0 (370 Hz), for psychoacoustic reasons, principally just vowels produced with the first or the fourth and fifth degrees of openness may appear in perception. However, the consequences of the changing relations between the prominent features of the vowels (i.e. f_0 and F_1) resulting from $F_1 : f_0$ tuning are not predictable from Traunmüller's theory.

Subjects, material, methods

The test material comprised of the 9 Hungarian vowel qualities again. The vowels were sung on the 6 musical notes indicated in the previous sections and uttered also in speech by 3 professional soprano singers (28, 28 and 45 years of age). The acoustic analysis consisted of two partly independent approaches. First, we applied manual inverse filtering (DeCap, Svante Granqvist) on the radiated vowel sounds. During inverse filtering we made use of a parallelly recorded Electroglottograph (EGG) signal that captured the vocal fold movement during phonation (see also Hertegård & Gauffin 1993). Second, to validate the data obtained through inverse filtering, we applied an Electrolarynx at the singers' neck (above the larynx) which resampled the vocal tract transfer function and we also detected the prominences of this signal after it radiated through the singers' mouth (see also Sundberg 1975). Then a listening experiment was conducted on the stimuli produced with natural phonation with 21 adult listeners (26-45 years of age, $M = 29$ years).

Results

According to the acoustic data the lower limit of $F1 : f_0$ tuning was dependent on openness or the $F1$ of the vowels in speech (first degree of openness /i: y: u:/: note $f' = 349$ Hz; second and third degrees of openness /e: ø: o/: note $f'' = 698$ Hz; fourth degree of openness /ε ɒ a/: note $h'' = 988$ Hz, see Figure 7).

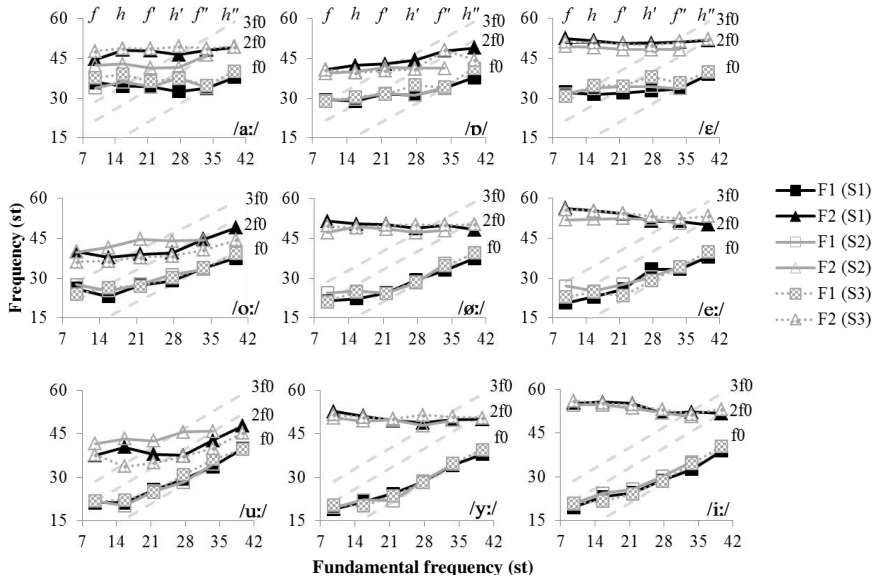


Figure 7: Formant values as vocal tract resonances (F1, F2) of the three sopranos as a function of the f_0 (the dashed gray lines represent the first three harmonics of the voice source: f_0 , $2f_0$, $3f_0$)

Moreover, the data also showed that it is not just the F1 but also the F2 that was affected by pitch-raising. The slightly decreasing F2 in front vowels and the

drastically increasing F2 in back vowels reflect the centralization of vowels in the front–back dimension accompanying f_0 -increase. This centralization, however, is also the result of F1 : f_0 tuning. The upward tuning of F1 requires an increase in the degree of jaw opening, but the increase also impacts the horizontal position of the tongue, since the velar constriction becomes more and more difficult to form as the jaw opening increases. As the singer reaches the physiological limits of her compensatory efforts to form the constriction, the place of articulation necessarily neutralizes. The above acoustic changes lead to the reduction of the (psycho)acoustic vowel space with the f_0 -increase (see Figure 8).

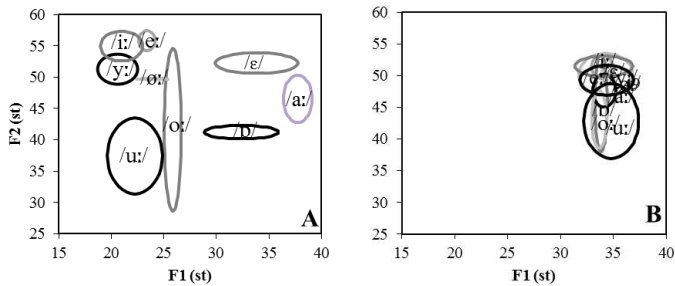


Figure 8: Vowel realizations in the (psycho)acoustic vowel space in speech (A) and in singing at $f'' = 698$ Hz (B). The ellipses are defined by the mean of the extracted formant frequencies (center), and the standard deviation (± 2 SD in both axes)

According to the perceptual data the identification of the intended qualities decreased as the f_0 increased⁷. When the vowels were “misidentified” below $f_0 = 370$ Hz (or even at 494 Hz, or h') in the most cases vowels were identified as more close, in line with our previous results. At the note f'' (988 Hz) vowels were identified as / ɔ a / in nearly 100% of all cases. The inbetween f_0 s showed an increased number of vowels produced with the second and third degrees of openness (see Figure 9). Consequently, the results in singing contradicted the previous findings for speech, as in singing above the critical 370 Hz not only vowels produced with the first and fourth degrees of openness appeared in perception. In this sense, singing seems to have a greater degree of diversity in terms of vowel qualities which is (most probably) also the result of F1 : f_0 tuning: F1 : f_0 tuning gives rise to an enhanced f_0 component and thus creates a prominence relation between f_0 and F1 that are unusual in speech. The increased number of / ɔ a / responses at f'' may also be attributed to the same cause: at f'' both f_0 and $2f_0$ are enhanced by the first two vocal tract resonances (F1 and F2) and this prominence relation resemblances that of / ɔ / and / a / in speech.

Discussion

With a new combination of some previously introduced methods we successfully demonstrated (in the entire vowel set of a particular language and in three singers’

⁷ The identification tendencies of the vowels produced by the three singers did not differ remarkably.

production) that the lower limit of $F1 : f_0$ tuning is dependent on the vowels' degree of openness or their $F1$ in speech. Moreover, the data also demonstrated the reduction of the front-back dimension in the (psycho)acoustic vowel space with pitch-raising. Perceptual data interpreted in light of Traunmüller (1981)'s theory appeared to be explicable, including the fact that vowels below 600 Hz are very likely to be perceived as more close vowels (as showed in our previous studies), but vowels sung at the f_0 of 988 Hz (f'') are mostly perceived as /b a:/. Moreover, on the basis of Traunmüller (1981) we also claim that the increased number of /b a:/ responses at high f_0 are not the outcome of a perceptual opening tendency accompanying the opening tendency in production. Our results thus do not support the suggestion of the previous literature on direct and linear interrelations between production and perception.

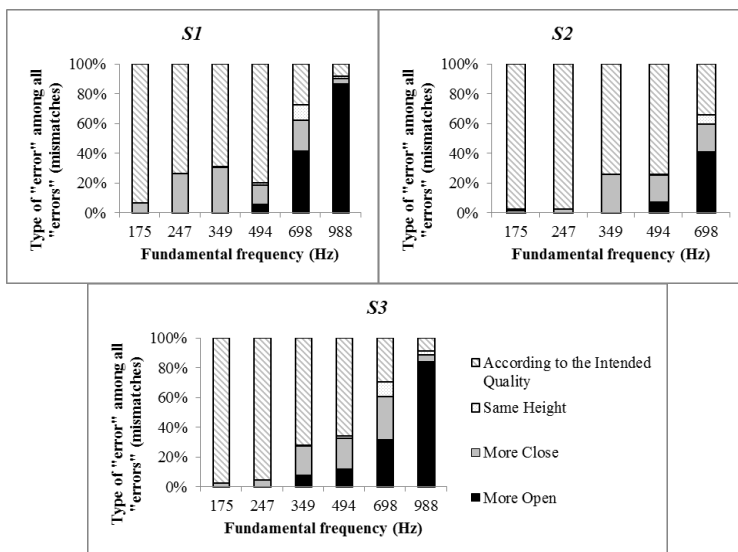


Figure 9: Responses to the vowels produced with the second and third degrees of openness /ε e: ø: o:/ as a function of pitch

Discussion and conclusions

H1: The results verified our first hypothesis: we found that sopranos tended to tune $F1$ to the raised high f_0 , if the f_0 had exceeded the $F1$ of the vowels. Accordingly, the lower limit of $F1 : f_0$ tuning was dependent on the degree of openness or the $F1$ of the vowels in speech. To perform resonance frequency detection we also successfully combined and tested some alternative recording protocols and measurement techniques.

H2: The second hypothesis was partly verified. On the one hand, we found evidence for the decrease of the identification of the intended vowel qualities with f_0 -increase. The results showed that (as a result of the acoustic changes) the perceptual

quality of the vowels gradually became more ambiguous with the f_0 -increase. On the other hand, we found that the decreasing tonotopic difference between f_0 and F_1 (accompanying the f_0 -increase) resulted in the percept of more close vowels only below h' (494 Hz). Above 370 Hz (approx. $f_{\#}'$) the scatter of responses increased, while above 700 Hz (f'') / p a/ dominated the array of responses. These tendencies are most probably not the result of a perceptual opening tendency accompanying the opening tendency in production (as suggested by previous literature), but the result of the increased tonotopic distance between F_1 and f_0 , and the $F_1 : f_0$ tuning.

H3: Our results on perception contradicted the claims of Smith and Scott (1980): we found evidence that the consonantal context does not exert a positive effect on the identification of the intended vowel in singing through transitional acoustic cues, and the same applied to the voicing onset of isolated vowels. Based on the results we concluded that the identification of the intended vowel quality may only be supported by probability and the activation of top-down processes: the identification may only be supported by context (word and sentence embedding) and phonology (phonotactics and phonological neighborhood density). This claim is also supported by the finding that despite their experience in production, singers do not exhibit any perceptual advantage in the identification of the intended vowels.

Theses

1. The $F_1 : f_0$ tuning suggested by previous literature was demonstrated in the entire set of the Hungarian vowels. It was also demonstrated that vowels become more and more open in production, and the vowel space reduces and shifts towards the position of / a/ as the f_0 increases.

2. We successfully measured resonance frequencies in high-pitched sung vowels independently of the voice source by the use of inverse filtering with EGG and by means of external excitation (i.e. the analysis of Electrolarynx recordings).

3. We falsified the previously suggested claim that the opening tendency in the production of sung vowels (accompanying f_0 -increase) is also accompanied by an opening tendency in perception. The only generalizable tendency that is characteristic of the perception of vowels with f_0 -increase is a gradually increasing ambiguity of vowel qualities.

- 3.a We only found the notable increase of open vowels / p a/ as responses when the relative distance between f_0 and $2f_0$ components and their enhancement were similar to that of the frequency and energy of the first to prominent components of / a/ in speech. This requires both the frequency of the f_0 to be around 900-1000 Hz and the first two partials (f_0 , $2f_0$) to coincide with the first two resonances (F_1 , F_2) which means that the partials are both enhanced by formants. (In our studies we demonstrated these correspondences at the musical note $f'' = 988$ Hz.)

- 3.b. Below 1 kHz the interdependencies of production and perception are not characterized by direct and linear relations. The most dominant tendency found in perception below 5-600 Hz is the identification of the intended vowels as more close; the ratio of more open responses only starts to increase above this f_0 . Based on our

results we suggested that the percept of more open vowels is not the direct consequence of the opening tendency in production, and it requires both the $F1 : f_0$ tuning and the spectral undersampling of high-pitched sonorant sounds (resulting in the correspondence of $F2$ and $2f_0$).

4. We also suggested that it is also the result of $F1 : f_0$ tuning that vowels produced with the second and third degree of openness may appear in perception. Without the $F1 : f_0$ tuning it is only the appearance of vowels with the first and fourth degree of openness that is predicted by the literature of speech perception.

5. We concluded that the consonantal context does not play a positive role in the identification of the intended vowels in singing through acoustic cues. However, another aspect of context may have a positive effect, that is, vowel embedding in real words and sense texts, and high probability of the carrier words (resulting from a smaller number of phonological neighbors).

6. We concluded that singers and non-singers identify sung vowels according to very similar perceptual tendencies, and in spite of their advantage in production, singers do not exhibit any advantage in the identification of the intended vowels.

7. We concluded that the modification of the acoustic structure of sung vowels increases with the increasing f_0 , and this cannot be compensated for only by bottom-up perceptual processes; the compensation also requires the activation of top-down perceptual processes.

References

- Benolken, Martha S. – Swanson, Charles E. 1990. The effect of pitch-related changes on the perception of sung vowels. *Journal of the Acoustical Society of America* 87(4). 1781–1785.
- Berger, Kenneth W. 1964. Some factors in the recognition of timbre. *Journal of the Acoustical Society of America* 26(10). 1888–1891.
- Boersma, Paul – Weenink, David 2009. *Praat: Doing phonetics by computer (Verison 5.3)*. <http://www.praat.org>. (Downloaded on 12. January 2014.)
- Bresch, Erik –Narayanan, Shrikanth 2010. Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. *Journal of the Acoustical Society of America* 128(5). EL335–EL341.
- de Cheveigné, Alain – Kawahara, Hideki 1999. Missing-data model of vowel identification. *Journal of the Acoustical Society of America* 105(6). 3497–3508.
- Fant, Gunnar 1960. *Acoustic Theory of Speech Production*. Mouton & Co., The Hague, Netherlands.
- Garnier, Maëva – Henrich, Natalie – Smith, John – Wolfe, Joe 2010. Vocal tract adjustments in the high soprano range. *Journal of the Acoustical Society of America* 127(6). 3771–3780.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Hertegård, S. and J. Gauffin 1993. Voice source – vocal tract interaction during high-pitched female singing. In: *Proceedings of the Stockholm Music Acoustics Conference 1993 (SMAC 1993)*. Stockholm, Royal Swedish Academy of Music. 177–182.
- Hollien, Harry – Mendes-Schwartz, Ana P. – Nielsen, Kenneth 2000. Perceptual confusions of high-pitched sung vowels. *Journal of Voice* 14(2). 287–298.
- Johansson, Christine – Sundberg, Johan – Wilbrand, Hermann 1982. X-ray study of articulation and formant frequencies in two female singers. In: Askenfell, Anders – Felicetti, S. –

- Jansson, Erik – Sundberg, Johan (eds.) *Proceedings of the Stockholm Music Acoustics Conference 1983 (SMAC 1983)*, Stockholm, Royal Swedish Academy of Music. 203-218.
- Joliveau, Elodie – Smith, John – Wolfe, Joe 2004. Vocal tract resonances in singing: the soprano voice. *Journal of the Acoustical Society of America* 116(4). 2434–2439.
- Krakow, Rena A. – Beddor, Patrice S. – Goldstein, Louis M. – Fowler, Carol A. 1987. Coarticulatory influences on the perceived height of nasal vowels. *Status Report on Speech Research*, Haskins Laboratories. 31–53.
- Scotto di Carlo, Nicole – Germain, Aline 1985. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica* 42(2).188–97.
- Smith, Lloyd – Scott, Brian L 1980. Increasing the intelligibility of sung vowels. *Journal of the Acoustical Society of America* 67(5). 1795–1797.
- Sundberg, Johan 1975. Formant technique in a professional female singer. *Acta Acustica united with Acustica* 32(2). 89–96.
- Sundberg, Johan 1979. Perception of singing. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 20(1). 1–48.
- Sundberg, Johan 1987. *The science of the singing voice*. Northern Illinois University Press, DeKalb Illinois.
- Sundberg, Johan –Skoog, Jörgen 1995. Jaw opening, vowel and pitch. *Speech Transmission Laboratory Quarterly Progress and Status Report (STL-QPSR)* 36(2–3). 43–50.
- Svante Granqvist. é. n. *DeCap*. tolván.com. (Downloaded on 26. April 2013.)
- Trautmüller, Hartmut 1981. Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* 69(5). 1465–1475.

The author's publications related to the topic

- Deme Andrea 2011. Az énekelt magánhangzók fonetikai elemzése. In: Parapatics Andrea (eds.) *Félúton 6: A hatodik Félúton konferencia (2010) kiadványa*. 25-42.
- Deme Andrea 2011. Az énekelt magánhangzók észlelése réshangkörnyezetben. In: Váradi T (eds.) *V. Alkalmazott Nyelvészeti Doktorandusz Konferencia*. MTA Nyelvtudományi Intézet, Budapest. 16-28.
- Deme Andrea 2012. Az énekelt magánhangzók fonetikai elemzése. In: Parapatics Andrea (eds.) *Doktoranduszok a nyelvtudomány útjain: 6. Félúton konferencia*. ELTE – Eötvös Kiadó (Talentum), Budapest. 33-46.
- Deme, Andrea 2013. On the Hungarian sung vowels. *The Phonetician* 105-106(1–2). 73-87.
- Deme Andrea 2014a. A mássalhangzó-környezet és a zöngéindulás hatása az énekelt magánhangzók észlelésére. *Beszédkutatás* 22. MTA Nyelvtudományi Intézet, Budapest. 155-171.
- Deme, Andrea 2014b. Intelligibility of sung vowels: the effect of consonantal context and the onset of voicing. *Journal of Voice* 28(4). 523.e19-523.e25.
- Deme, Andrea 2014c. Formant strategies of professional female singers at high fundamental frequencies. In: Fuchs, Susanne –Grice, Martine –Hermes, Anne –Lancia, Leonardo – Mücke, Doris (eds.) *Proceedings of the 10th International Seminar on Speech Production Cologne (ISSP)*, Köln. 90–93.
- Deme Andrea (in press). Speech perception at its best: Extracting linguistic information from phonetically underspecified input. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS) 2015*.