NOTICE: this is the author's version of a work that was accepted for publication in *Journal of Voice*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Intelligibility of sung vowels: the effect of consonantal context and the onset of voicing* [Journal of Voice, Vol. 28, Issue 4, pp. 523.e19-523.e25. (2014)] DOI: 10.1016/j.jvoice.2014.01.003

Intelligibility of sung vowels: the effect of consonantal context and the onset of voicing

Andrea Deme

Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary
Department of Phonetics, Eötvös Loránd University, Hungary

Abstract

Background: Studies addressing the identification of sung vowels concern mainly the effect of the fundamental frequency and conclude that correct vowel identification decreases with increasing pitch. In one experiment the impact of consonantal environment on the intelligibility of the vowels in high-pitched singing was also studied. The results of that experiment showed positive effect of the consonantal environment. This finding is in line with results that had been reported for speech in an earlier study. However, the data on singing are not as transparent as the authors suggest, and there are some conditions in the experiment that could also be controlled for more strictly. Therefore, the effect of the dynamic acoustic information encoded in the formant transitions at high fundamental frequencies is still an open question.

Objectives: The aim of the present study is to redesign and extend the above mentioned experiment to test whether the phonetic context and the onset of the vowel uttered in isolation (namely the onset of voicing) have a positive effect on vowel identification.

Methods: For this purpose, a vowel identification test was carried out. The stimuli included 3 Hungarian vowels /aː iː uː/ in 3 conditions (in /bVb/ context, in isolation and with eliminated onset) at 7 different fundamental frequencies from 175 Hz to 988 Hz (F3, B3, F4, B4, F5, B5, and speech). The stimuli were produced by one professional soprano singer.

Results: The results show that consonantal context does not specify vowel identity in singing as clearly as it has been demonstrated for spoken utterances. In addition, no effect of vowel onset (i.e. the onset of voicing) was found. Recognition percentages seemed only to be dependent on f_0 and vowel quality.

Conclusions: The unexpected results lend themselves to two possible explanations: the reduction of the consonants and the under-sampling of the formant transitions.

Key words: singing, vowel identification, consonantal context, onset of voicing, consonant.

1 Introduction

The issue of vowel perception in Western operatic singing has been addressed frequently in the literature. The question is particularly relevant, because in sopranos the fundamental frequency (f_0) often exceeds the frequency region of the first formant (F_1) of vowels that is typical in speech. For this reason, it is mostly the effect of f_0 on the intelligibility of sung vowels that is investigated. The data show that correct vowel identification decreases with increasing pitch [1] [2] [3] [4] [5]. The acoustic realization of high-pitched sung vowels and the observable acoustic differences between sung and spoken vowels originate from the articulatory differences of speech and singing and the acoustic properties of the high fundamental frequency. On the one hand, homogenous timbre and pitch-raising required in singing are provided by articulatory maneuvers that change the articulatory configuration

typical of the vowels in speech [6] [7]. On the other hand, the harmonic spacing gets wider as a result of high pitch, thus the harmonics convey the vocal tract transfer function to a lesser extent. Consequently, the vowel has a "limited resolution" in the output sound. As a result of the articulatory and acoustic changes, the decrease in the ratio of correct vowel identification with pitch increase is expected. Although the perceptual data show this tendency, it has also been found that the intelligibility of vowels can be preserved in certain conditions even at higher pitches [2] [3] [5]. From this apparent contradiction the following question arises: what are the cues that might support the human auditory system in these challenging discrimination tasks?

There is a generally accepted agreement that the phonetic context has an impact on the identification of the vowel in speech. As the well-known study of [8] demonstrated, the consonantal environment in CVC sequences specifies the vowel identity through the dynamic acoustic information provided by the formant transitions between the consonants and the vowel, thus the percentage of correct identification is higher for vowels uttered and perceived in CVC context than in isolation. Predominantly, this observation also tends to be accepted to sung vowels (see e.g. [3]); however, for high-pitched sung vowels it was tested only in one experiment. [9] investigated the effect of consonantal context on sung vowel identification in CVC sequences and in isolation. The authors claimed that they provided data on the positive effect of the neighboring consonant. Nevertheless, the results were obtained under not strictly controlled conditions. In [9] the vowels were recorded and compared in /bVd/ sequences and in isolation with low and high vertical larynx position (the position of the larynx was not monitored objectively, though). It should be noted that the place of articulation of the first and last consonant in the CVC sequence is not identical. Consequently, the formant transitions preceding and following the vowel might be remarkably different. This uncontrolled factor raises two questions. First, which consonant's impact was tested during the listening test in [9]? Second, do the preceding and following transitions impact the vowel perception to the same extent? Because of these unclarified questions, the perception data of [9] are difficult to interpret. Hence, it has to be concluded that the first approaches in testing the effect of consonantal context must include identical consonants in the carrier sequences. In addition, the differences presented in [9] are clear only above the fundamental frequency of F5 (698 Hz); below that the results are much less consistent.

[1] used similar sung material to [9] to test the effect of CVC context at lower pitches in countertenors. They compared the identification of the steady-state part of vowels with that of the vowels uttered in CVC context. Their findings are roughly in line with [9]. However, [9] and [1] have also provided evidence that vowels uttered in isolation might also preserve their distinctiveness to a certain extent. The data of [1] compared to the data of [9] reveal that identification percentages for vowels uttered in isolation are higher than those for the steady-state portion. Therefore, it can be concluded that vowels uttered in isolation can retain their intelligibility to a greater extent. Based on this comparison, it seems reasonable to suggest that not only the formant transitions but the onset of voicing or the vowel onset can also provide some extra information regarding vowel identity, thus supporting the human auditory system in vowel identification. As for the author's knowledge, there has been no study investigating this issue yet.

Other than the above experiments, there is only one paper that investigates the effect of certain consonant types on vowel identification in sung nonsense CVC sequences [5]. The study reports no clear differences between nasal vs. voiced and unvoiced fricative contexts, and suggests that according to the data no clear effect of the studied consonant types can be concluded.

The aim of the present study is to redesign and extend the experiment of [9] with particular modifications and restrictions concerning the control of the variables affecting vowel perception, and to investigate not only the effect of consonantal environment, but also the effect of vowel onset (the onset of voicing of vowels uttered in isolation). It is hypothesized that the identification of vowels is affected positively by the presence of the phonetic context provided by consonantal context or the

vowel onset, but this positive effect decreases with ascending pitch due to the under-sampling of the dynamic acoustic information of formant transitions and the onset of voicing.

2 Methods

The material of the study consists of one professional soprano singer's singing production. The singer was asked to produce 3 sustained vowels (the 3 most spaced vowels of the Hungarian vowel inventory) /aː iː uː/ in 2 conditions: in /bVb/ context (hereafter, "CVC") and in isolation (hereafter, "V"). She covered a pitch-range from 175 Hz to 988 Hz in singing (on the fundamental frequencies of F3 = 175 Hz, B3 = 247 Hz, F4 = 349 Hz, B4 = 494 Hz, F5 = 698 Hz, B5 = 988 Hz) and speech (average: 191 Hz). The consonants in the CVC sequence were chosen to be identical to control for the possible effect of the place of articulation. Therefore, to provide comparable results to [9] it had to be decided which of the consonants in the /bVd/ sequence has to be changed. Based on the author's previous observations, it was supposed that the first consonant is more pronounced in sung CVC sequences than the last one (thus possibly having stronger impact on perception as well). Therefore, the author decided to retain the first and change the last consonant. The exploration of the effect of the place of articulation of the consonant is a question of future studies. To test the effect of the onset of voicing or the natural vowel onset, a third condition was created by manipulation: the onsets of the vowels uttered in isolation were eliminated with exponential fading-in effect in Wavesurfer [10] (hereafter, "CUT"). Comparison of the V and CUT contexts enables the assessment of the effect of vowel onset. The recordings were made by an omnidirectional condenser microphone in a sound-treated room and digitized at 44.1 kHz.

In the perception test the 63 target stimuli (3 conditions × 3 vowels × 7 fundamental frequencies) and 15 distractor stimuli (containing other vowels in /bVb/ context and in isolation) were presented to each subject twice in a randomized order in Praat [11]. The loudness level of the stimuli was equated over the different samples. Before the test the subjects were informed that they would hear vowels produced at different pitches with or without consonantal context, and they were instructed to identify and to select the vowels they hear from the candidates displayed on the computer screen. The set of candidates the listener had to choose from consisted of 9 vowels from the Hungarian vowel inventory (which means the entire set of Hungarian vowels excluding only the phonologically short counterparts of long vowels): /ɒ aː ɛ eː iː oː øː uː yː/. The vowels were displayed alone and in orthographical form. To select the vowels the listeners had to click on the candidates by use of a mouse connected to the PC. The subjects listened to the stimuli binaurally through headphones. Correct responses and the errors of vowel identification were collected in confusion matrices for each fundamental frequency. (It should be noted here that the terms "correct identification" and "error" refer to the relationship between the response and the intention of the singer, which was guided by the wordlist presented to her. In this study there was no intention of assessing and defining vowel qualities the singer managed to produce.)

22 non-trained adult listeners participated in the perception test. The listeners were questioned about their health conditions: only listeners without any hearing disorders participated in the test. The names and personal data of the singer and all the listeners were discarded in the analysis of the data; signed consents of the participants were collected and submitted to the IRB.

The statistical analysis (χ^2-test, ANOVA, Pearson's correlation test, Tukey HSD) of the results was carried out in R [12].

3 Results

The consistency of the answers between the two repetitions of the same stimuli was assessed for each subject (Table 1). These reliability measures are the percentages of those cases in which listeners managed to identify the two stimuli identically (each subject provided 156 responses, thus the number

78 constituted 100%.). Therefore, these measures indicate whether the listeners identified the vowel presented in the stimulus or they were guessing. In the first case, the two replications should trigger consistent (though not necessarily "correct") responses, but in the second, the responses are more likely to differ, since when guessing, the subjects are less likely to be consistent.

In the reliability measures calculated for the subjects no outliers were found (χ^2 (21) = 13.5221; p = 0.89), each subject performed between 55% and 82%. Therefore, no subject's data had to be excluded from the analysis on suspicion of careless completion of the test. It has to be noted that the fairly low percentages (in Table 1) do not indicate that the participants were not reliable, as decreasing intelligibility of vowels accompanying pitch-raising is expected which might result in non-correct and less consistent responses at higher fundamental frequencies.

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22
Reliability measures (%)	82	76	66	74	59	63	71	63	72	71	69	71	73	62	72	71	55	56	67	76	71	68

Table 1: Reliability measures of the 22 listeners.

In Figure 1 the identification percentages of the vowels (y-axis) in the three conditions are presented as a function of fundamental frequency (x-axis); each data point represents the percentage of correct responses (of all listeners) related to the total number of responses in the particular condition (for all three vowels on average) per f_0.

Generally speaking, the rate of correct vowel-identification seems to decrease with increasing f_0 in each condition (Figure 1). A two-way ANOVA including the factors f_0 and *condition* showed a significant effect of f_0 on vowel intelligibility ($F(1)=33.33$, $p < 0.001$). Correlation tests revealed that the interrelation between f_0 and recognition percentages can be characterized by a strong negative correlation (Pearson's $r = -0.631$, $p < 0.001$; if the data is split into 3 groups according to the 3 different conditions CUT and CVC show a strong correlation: Pearson's $r = -0.703$, $p < 0.001$ and Pearson's $r = -0.609$, $p = 0.003$, respectively, while V shows a slightly weaker correlation: Pearson's $r = -0.590$, $p = 0.005$).

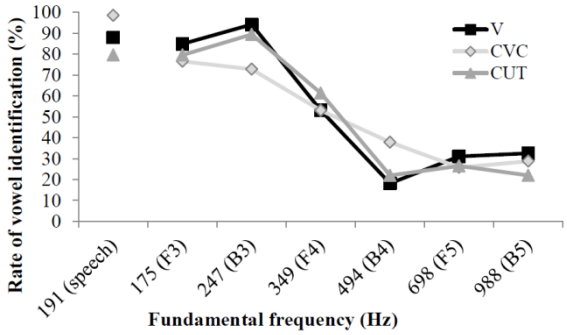

Figure 1: Vowel identification as a function of fundamental frequency.

However, the statistical analysis did not show evidence for the interrelation between the 3 conditions and vowel intelligibility: according to the two-way ANOVA neither condition by itself nor the combination of f_0 and condition had a significant effect on vowel identification.

The non-significant interrelation between the conditions and the identification percentages is originated in the inconsistent effect of the phonetic context, to which Figure 1 provides several examples. The function describing CVC is monotonically decreasing below F5, while the functions of V and CUT conditions are non-monotonic (see Figure 1), thus the relations concerning the intelligibility of the vowels in the 3 conditions might change with the fundamental frequency (e.g. the identification percentages are higher for V/CUT than for CVC at B3, but it is lower for V/CUT than for CVC at B4).

In general, the data showed greater differences between V vs. CVC and CUT vs. CVC conditions, than between V vs. CUT conditions. There are slight and inconsistent differences between the identification rates of the V and CUT groups.

Although the effect of phonetic context cannot be considered to be shown in singing, the results for speech are consistent with the findings reported in previous studies [8] and the hypothesis of the present paper: vowels uttered in consonantal context were characterized by the highest rate of correct vowel-identification, while vowels in isolation were less easy to identify, and vowels without onsets were the least perceivable (see Figure 1).

For a closer evaluation of the effect of formant transitions, the close front vowel /iː/ was selected at the fundamental frequency of F3 (175 Hz) where the formant transitions between /b/ and the vowel were the most obvious and best observable (see Figure 2). However, the identification percentages never showed the positive effect of the supposedly available dynamic acoustic information, i.e. the formant transitions (V: 86% > CUT: 80% > CVC: 50%, see Figure 4).

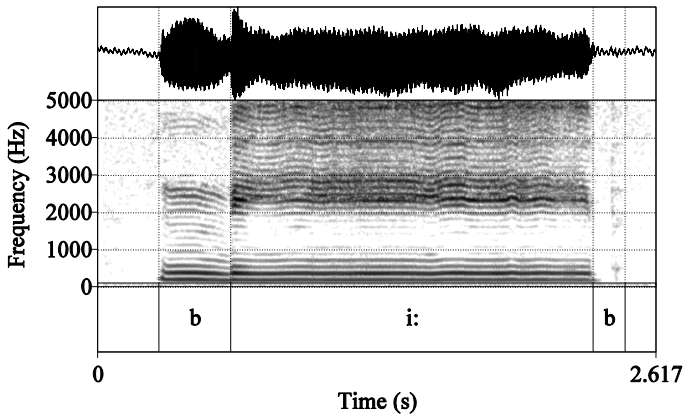

Figure 2: Narrow-band spectrogram of vowel /iː/ in consonantal context (/biːb/) at F3 (175 Hz).

Although the effect of consonantal context and vowel onset was not proven to have an impact on the intelligibility of vowels in singing, the identification rates turned out to be dependent on vowel quality. According to a ANOVA including 3 factors (f_0, *condition* and *vowel quality*), not only f_0, but vowel quality (a.k.a. vowel class or vowel type) ($F(2) = 22.673$, $p < 0.001$) and the combination of these two variables ($F(2) = 7.419$, $p = 0.002$) have a significant effect on vowel identification, as well. Based on this result, a closer analysis of the recognition percentages with respect to vowel qualities was also carried out.

Tukey HSD tests were conducted on all the possible pairwise contrasts among the three vowels which showed significant differences for /iː/ – /aː/ and /uː/ – /aː/ but not for /uː/ – /iː/ at the confidence level of 0.02. This means that there are significant differences only between vowels which differ in jaw opening. In accordance with this finding, it was also observed that the correlation between f_0 and vowel identification is also related to jaw opening, as this correlation persisted only for /iː/ and /uː/ (Pearson's $r = -0.781$ $p < 0.001$, Pearson's $r = -0.900$ $p < 0.001$, respectively), while in the case of /aː/ only a non-significant weak correlation appeared (Pearson's $r = -0.384$, $p = 0.086$) between f_0 and vowel identification rates. The correlation tests reveal that /iː/ and /uː/ got less and less identifiable with the increasing f_0, while the identification rates of /aː/ were less dependent on the value of f_0. This interrelation is also observable in Figures 3, 4 and 5.

The identification rate of /aː/ is relatively high below 494 Hz (B4) (Figure 3). There is a sudden decrease at 494 Hz (B4) where the lowest rates are observable in every condition. At this fundamental

frequency, /aː/ was identified as the more close vowel /ɒ/ in 46% of all responses. Above B4, there is an increase again and the rate of identification proceeds to increase slightly until the highest f_0. At F5 it was /ɒ/ (27% of all the responses), while at B5 it was /ɛ/ (8% of all responses) that occurred as the most frequent error. Both vowels have higher first formant and both of them are more close than /aː/ in Hungarian (see [13]).

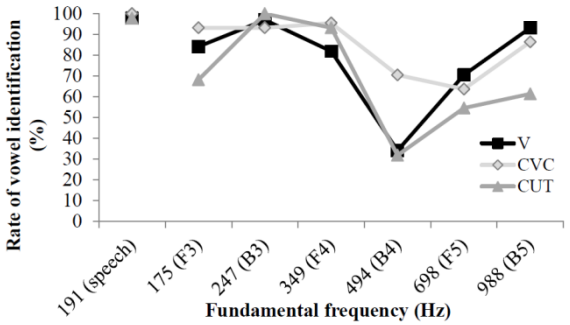

Figure 3: Rate of identification for /aː/.

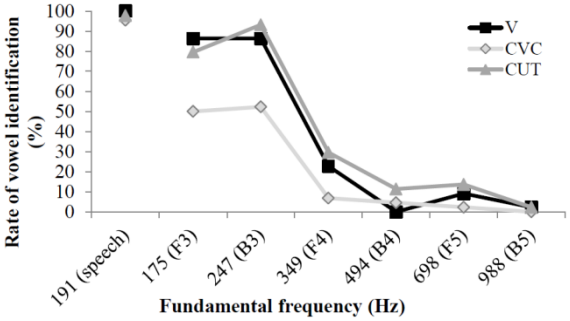

Figure 4: Rate of identification for /iː/.

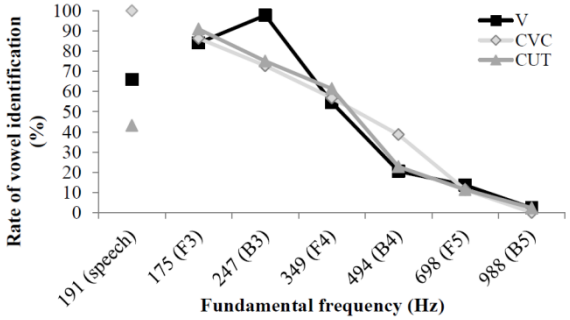

Figure 5: Rate of identification for /uː/.

The identification of /iː/ decreases suddenly at F4 (Figure 4). In most cases, the more open /eː/ occurred as a response above B3 (in 80% at F4, in 94% at B4 and in 67% at F5). At B5, on the other hand, /iː/ was identified as the more open /aː/ in 80%.

The identification percentages for /uː/ in speech meet the expectations about the differences of V and CVC conditions, and the presumptions presented above (about consonantal context and the onset), while /aː/ and /iː/ in speech did not show the same tendency (Figure 5). As for singing, no such tendencies were found. One-way ANOVAs conducted on the vowel groups separately (with the *condition* factor) confirmed that no significant effect of the condition in any of the groups is present, that is, no effect of the consonantal context or the vowel onset is shown for any of the vowels.

The tendencies for the perception of /uː/ in singing were very similar to those observed in the averaged data: the efficiency of the perception is basically characterized by smooth and gradually

decreasing functions with changing relations between V and CVC conditions. As for errors, /uː/ is identified mostly as the more open vowel /oː/ (in 13% at F3 and B3, in 42% at F4, in 62% and B4 and in 25% at F5), while at the highest f_0 it is the vowel /aː/ again that dominates the hierarchy of errors (with 86%).

For closer evaluation of the combined effect of f_0 and vowel quality on the recognition percentages of vowels (revealed by the ANOVA, see above), Tukey HSD tests were conducted on all possible pairwise contrasts among the three vowels within the groups of the three conditions. Within the group of CVC the vowels /iː/ and /aː/ were found to be significantly different at $p < 0.02$. However, none of the other combinations in any of the conditions differed significantly. These relationships can be observed in Figure 6, 7 and 8.

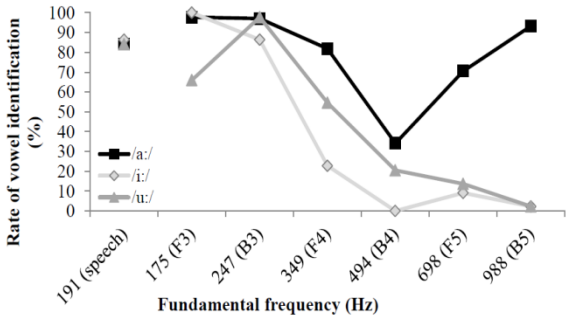

Figure 6: Rate of identification for vowels uttered and presented in isolation (condition "V").

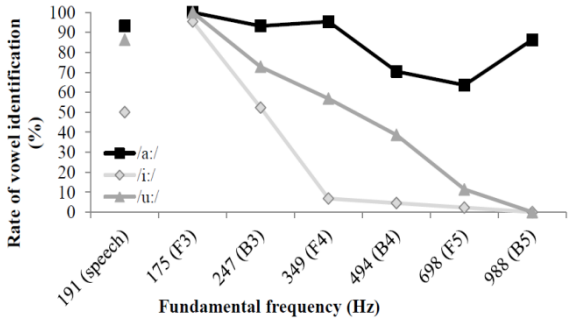

Figure 7: Rate of identification for vowels uttered and presented in consonantal context (condition "CVC").

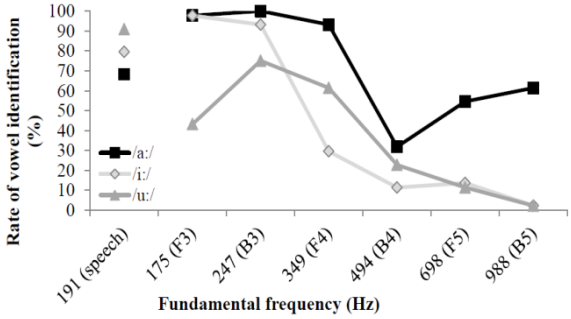

Figure 8: Rate of identification for vowels presented with deleted onset (condition "CUT").

The comparison of Figure 6, 7 and 8 show the only effect of the consonantal context that was observable in the data: the vowels in CVC condition showed the most clearly divergent tendencies according to vowel quality with /aː/ retaining its identity efficiently through the whole pitch range. In the case of vowels uttered and presented in isolation (Figure 6) recognition percentages showed a convergent (decreasing) tendency below 698 Hz (F5). From F5 upward, the tendencies for open (/aː/) and close vowels /iː/ and /uː/ begun to diverge. The recognition percentages in the case of vowels with consonantal context (Figure 7), on the other hand, started to diverge already at a much lower

fundamental frequency: 247 Hz (B3). In the case of vowels with eliminated onset (Figure 8) there were less clear-cut tendencies found for the close vowels /iː/ and /uː/ below 494 Hz (B4), whilst /aː/ tended to follow the tendency found in the V condition. Above B4, however, all results were in line with those found in the V condition, namely divergent tendencies for close and open vowels.

Tukey HSD tests were also conducted on all possible pairwise contrasts among the three conditions within the groups of each vowel. These comparisons were in line with previous findings (on vowels not grouped according to vowel quality), since they did not show the effect of condition on the identification percentages, i.e. the effect of consonantal context or vowel onset on the intelligibility of vowels.

4 Discussion

In the present study the effect of consonantal context and vowel onset on the identification of high-pitched sung vowels was investigated. Based on previous findings [9], it was hypothesized that the dynamic acoustic information encoded in the formant transitions between the consonant and the vowel support the human auditory system during vowel recognition. This reasoning was adapted to vowels uttered without context, since isolated vowels also provide some dynamic acoustic information (similar to coarticulatory transitions) in their onset of voicing. Therefore, it was proposed that the dynamic acoustic cue derivable from the onset can also help vowels to preserve their distinctiveness (to a certain extent) even at higher fundamental frequencies. The suggestion is also supported by the data of [9] and [1].

To test the assumptions 3 conditions were created in which vowels were i) uttered in CVC context, ii) uttered in isolation and iii) manipulated in a way that their onset was deleted. The stimuli were tested in a vowel identification task carried out with 22 listeners.

There are several conclusions that can be drawn from the results. Most importantly, it was not demonstrated that formant transitions and onset support vowel intelligibility in singing. Not only non-significant differences were found between the 3 conditions, but instances of negative effect of the neighboring consonants (at B3, generally) and the onset (in case of /aː/ at F4) were also seen. (Positive effect of consonantal context was seen at e.g. B4 in the averaged data, while the presence of the onset seemed to have some positive effect in the case of the close vowels /uː/ and /iː/ at B3). Additionally, the differences between V vs. CUT conditions are nearly always negligible (while, in at least some cases there are greater differences between V vs. CVC and CUT vs. CVC conditions); hence the effect of vowel onset is concluded to be even smaller than the impact the consonantal context can have. These findings contradict the hypotheses and lend themselves to two possible interpretations.

The first explanation is based on the harmonic structure of high-pitched singing voice. Wide harmonic spacing causing under-sampling of the vocal tract transfer function does not only influence the acoustics of the steady-state part of the vowel, but also the transitions. That is, at higher fundamental frequencies the acoustic information extractable from the transitions (as well as from the steady-state part) is reduced due to “low resolution”.

A second explanation is based on the instructions known as essential in the Western operatic (or bel canto) singing technique. When learning this technique singers are taught to maximize the duration of vowels and reduce the duration consonants as much as possible. However, there is no acoustic or articulatory data available describing the manner and nature of how singers manage to achieve this goal. According to a recent investigation, some obstruents in word-initial position tend to shorten with pitch-raising, but the voicing of the consonant is also affected remarkably [14]. The authors of [14] also mention that in many cases the reduction of the consonants involves less intensive turbulent noise or non-detectable bursts (in the case of stop consonants). However, the lack of turbulent noise or bursts does not only result in the decreased intelligibility of the consonant itself, but it also affects the coarticulatory formant transitions: acoustic transitions can only occur if the neighboring speech sounds

do have intensive acoustic components, just as coarticulation can only occur if at least two speech sounds are articulated subsequently. The results showing a seemingly incidental interrelation between the identification percentages and the conditions support both arguments.

The present findings contradict the results on the effect of neighboring consonants in sung vowel identification of [9] (and they also contradict the expectations that are based on the literature on spoken vowel identification): in our results no effect of /bVb/ consonantal context on vowel identification has been proved, whilst [9] claimed to have presented evidence for the positive effect of the /bVd/ context. How can these seemingly inconsistent results be accounted for? As it was already suggested in Section 1, the methods and the variables used in [9] were not described in detail and were not strictly controlled. It is not clear, for instance, how the height of the larynx was modified by the singer and monitored by the investigator throughout the recordings. In addition, the differences of the results obtained for the CVC and the isolated condition are more controversial than the authors interpret them. The number of subjects participating in the perception test can also account for the differences, as [9] involved only 10 listeners, while in the present study 22 participants were tested. The articulatory differences between Hungarian and (presumably) Australian English vowels can also be the cause of inconsistent results: the vowels represented by the same IPA symbol might somewhat differ with respect to the degree of centralization between the two languages. But most importantly, in [9] the effect of the consonants' place of articulation was handled differently than in the present investigation which makes the results of the two studies difficult to compare. The present study was well-controlled and carefully monitored through the recording and the testing phases as well, thus there are very good reasons to believe that our results can be considered as valid.

Though consonantal context and vowel onset were not shown to affect vowel recognition consistently, the analysis also showed that fundamental frequency and vowel quality have significant effect on the recognition percentages. Therefore, a detailed analysis of the effect of vowel qualities was also included. The main tendencies of vowel identification according to f_0 and vowel quality are roughly in line with those found in [2] [3] [5] and [15]:

1) In general, the identification task became more and more difficult (i.e. the percentages of correct identification decrease) with pitch-raising. With respect to vowel quality, there are two different tendencies (2 and 3 below).

2) The identification rates of the close vowels /iː/ and /uː/ decreased remarkably with pitch-raising. The vowels /iː/ and /uː/ have a low first formant, which is exceeded by f_0 early during pitch-raising. Therefore, /iː/ and /uː/ are affected by pitch-raising at even moderately high fundamental frequencies. Both vowels tend to be shifted towards vowels with higher first formants or greater jaw opening. In this respect, the results for /iː/ and /uː/ are also in line with theoretical speculations, which are based on the articulatory features of high-pitched singing: higher f_0 demands wider jaw opening irrespective of the articulatory demands of different vowel qualities.

3) The perceptual stability of the more open /aː/ persisted even at higher f_0 values. Considering that /aː/ is articulated by the widest jaw opening among all vowels and that pitch-raising is associated with the widening of the jaw, pitch-raising is supportive of the articulation of /aː/. Therefore, the persisting intelligibility of this vowel at higher fundamental frequencies is not surprising. This argument can be transposed into the acoustic domain as well. Since /aː/ has a relatively high first formant, increased pitch in singing exceeds this supposedly crucial acoustic cue only at higher f_0. Therefore, pitch-raising should not affect the perceptibility of /aː/ until higher fundamental frequencies. However, earlier empirical results reporting on the perceptual attributes of sung /aː/ at higher fundamental frequencies have also confirmed that this articulatory resistance does not necessarily result in perceptual stability as is often suggested on the basis of theoretical assumptions [5]. Indeed, the present study has also shown examples of lower recognition percentages for /aː/ at high fundamental frequencies, B4 and F5

(whilst for the even higher B5, for instance, the recognition percentages rise again). At both B4 and F5, /aː/ was perceived as the more close and labial /ɒ/. The occasional perceptual instability of /aː/ and the errors for more close vowels found in the present study are not in line with theoretical speculation (based on articulation) or some of the results reported in previous studies ([2][3][5][15]). However, they are perfectly in line with the results of [5] and they are also supported by the findings of [16]. While studying the effect of changing f_0 and first formant of a speech sound [16] concluded that the perception of openness resides in the tonotopic position of F_1 in relation to f_0, and that the decreasing distance between f_0 and the first formant (while other formants are kept constant) has the effect that vowels are perceived as more close. Hence, according to [16] it is reasonable to suggest that perceptual errors for more close vowels found in the present study are due to the changing distance (e.g. decreasing difference) between f_0 and first formant.

Comparison of the vowels' recognition percentages within the three conditions revealed the only effect of consonantal context. The differences in these recognition percentages tend to be dependent on the *condition* factor: vowels in consonantal context show the expected tendencies, namely the most divergent result for close and open vowels can be seen most clearly in CVC condition, and this context supported /aː/ to retain its identity the most efficiently through the whole pitch range.

The mode of production also seems to be important, as spoken stimuli were the easiest to identify (i.e. characterized by highest identification rates) in all cases, irrespective of the average fundamental frequency of speech.

To provide a more detailed picture of the relation between formant transitions and vowel identification, we plan to extend the investigation by including stops with different places of articulation in further studies. This extension will probably shed more light on the above mentioned inconsistencies.

5 Conclusions

The positive effect of consonants and dynamic acoustic information derivable from coarticulatory formant transitions and vowel onset could not be proven in the present paper. For the failure of validating the hypotheses two mutually non-exclusive explanations were presented. The results shed light on the necessity of the assessment of the articulatory and acoustic features of consonants in singing. Without consonants, there are no formant transitions either. Therefore, there is no reason to assess the effect of a phenomenon (i.e. the effect of formant transitions), as long as we do not know whether the phenomenon itself is present.

6 Acknowledgements

While conducting the research the author was funded by the Campus Hungary Program and hosted by the Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH Royal Institute of Technology. The work was supervised by Professor Johan Sundberg (KTH Royal Institute of Technology, Stockholm, Sweden), Professor Sten Ternström (KTH Royal Institute of Technology, Stockholm, Sweden), and Alexandra Markó, PhD (Eötvös Loránd University, Budapest, Hungary). The author is also grateful to dr. Svante Granqvist (KTH) for the profitable discussions.

References

[1] Gottfried TL, Chew SL. Intelligibility of vowels sung by a countertenor. *J Acoust Soc Am* 1986;79(1):124–30.

[2] Scotto di Carlo N, Germain A. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica* 1985;42(2):188–97.

[3] Hollien H, Mendes-Scwartz AP, Nielsen, K. Perceptual confusions of high-pitched sung vowels. *J Voice* 2000;14(2):287–298.
[4] Evgrafova K, Evdokimova V. Percetion of russian vowels in singing. *Baltic HLT Frontiers in Artificial Intelligence and Applications*. IOS Press, 2012;247:42–9.
[5] Deme A. On the Hungarian sung vowels. *Phonetician* 2012;1–2(105–106):73–87.
[6] Sundberg J. *The Science of the Singing Voice*. Illinois: Northern Illinois University Press; 1987.
[7] Garnier M, Henrich N, Smith J, Wolfe J. Vocal tract adjustments in the high soprano range. *J Acoust Soc Am* 2010;127(6):3771–80.
[8] Strange W, Verbrugge RR. Consonant environment specifies vowel identity. *J Acoust Soc Am* 1976;60(1):213–224.
[9] Smith L, Scott B. Increasing the intelligibility of sung vowels. *J Acoust Soc Am* 1980;67(5):1795–1797.
[10] Sjölander K, Beskov J. *Wavesurfer*. Computer software. http//: www.speech.kth.se/wavesurfer, 2009.
[11] Boersma P, Weenink D. *Praat: Doing phonetics by computer*. Computer program. http://www.praat.org, 2009.
[12] R Core Team. *R: A Language and Environment for Statistical Computing*. Computer program. http://www.R-project.org, Vienna, Austria, 2013.
[13] Hardcastle WJ, Laver J, Gibbon FE (eds.). *The Handbook of Phonetic Sciences.* 2nd Edition. Wiley-Blackwell, 2010.
[14] Deme A, Gráczi TE, Jankovics J. Obstruent voicing in singing. Talk presented at: 15th Summer School of Psycholinguistics; 2013 May 26–30; Balatonalmádi, Hungary.
[15] Sundberg J. Perceptual aspects of singing. *J Voice* 1994;8(2): 106–122.
[16] Traunmüller H. Perceptual dimension of openness in vowels. *J Acoust Soc Am* 1981;69(5): 1465–1475.