Intelligibility of Sung Vowels: The Effect of Consonantal Context and the Onset of Voicing

Andrea Deme, *Budapest, Hungary*

Summary: Background. Studies addressing the identification of sung vowels concern mainly the effect of the fundamental frequency (f_0) 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 was 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 three Hungarian vowels /aː iː uː/ in three conditions (in /bVb/ context, in isolation and with eliminated onset) at seven different fundamental frequencies from 175 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 (ie, 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 undersampling of the formant transitions.

Key Words: Singing–Vowel identification–Consonantal context–Onset of voicing–Consonant.

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.[1–5] 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 f_0. 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 Strange and Verbrugge[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[3]; however, for high-pitched sung vowels, it was tested only in one experiment. Smith and Scott[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 the study by Smith and Scott,[9] the vowels were recorded and compared in /bVd/ sequences and in isolation with low and high vertical larynx position (although the position of the larynx was not monitored objectively). 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

Accepted for publication January 8, 2014.

From the Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary.

Address correspondence and reprint requests to Andrea Deme, Research Institute for Linguistics, Hungarian Academy of Sciences, Benczúr u. 33, 1068 Budapest, Hungary. E-mail: deme.andrea@nytud.mta.hu

Journal of Voice, Vol. 28, No. 4, pp. 523.e19-523.e25
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.01.003

in the experiment[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 Smith and Scott[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 the study by Smith and Scott[9] are clear only above the f_0 of F5 (698 Hz); below that the results are much less consistent.

Gottfried and Chew[1] used similar material to Smith and Scott[9] to test the effect of CVC context at lower pitches in one countertenor's singing. 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 the study by Smith and Scott.[9] However, Smith and Scott[9] and Gottfried and Chew[1] have also provided evidence that vowels uttered in isolation might also preserve their distinctiveness to a certain extent. The data of Gottfried and Chew[1] compared with the data of Smith and Scott[9] reveal that the 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 article 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 versus voiced and unvoiced fricative contexts and suggests that according to the data, no clear difference in the effect of the studied consonant types can be concluded.

The aim of the present study was to redesign and extend the experiment of Smith and Scott[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 undersampling of the dynamic acoustic information of formant transitions and the onset of voicing.

METHODS

The material of the study consists of one professional soprano singer's singing production. The singer was asked to produce three sustained vowels (the three most spaced vowels of the Hungarian vowel inventory) /aː iː uː/ in two conditions: in /bVb/ context (hereafter, "CVC") and in isolation (hereafter, "V"). She covered a pitch range from 175 to 988 Hz in singing (on the f_0s 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 Smith and Scott,[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 conditions 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 (three conditions × three vowels × seven f_0) 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* by Boersma and Weenink.[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 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 nine vowels from the Hungarian vowel inventory (which means the entire set of the Hungarian vowels excluding only the phonologically short counterparts of long vowels /ɒː aː ɛː 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 f_0. (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.)

Twenty-two nontrained 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 Institutional Review Board.

The statistical analysis (chi square test, analysis of variance [ANOVA], Pearson correlation test, Tukey post hoc test) of the results was carried out in R.[12]

RESULTS

The consistency of the answers between the two repetitions of the same stimuli was assessed for each subject (Table 1). These