THE IMPACT OF THE INTENSITY RATIO BETWEEN VOWELS AND VOICELESS PLOSIVES ON THE INTELLIGIBILITY OF SUNG TEXT

Allan Vurma[a], Einar Meister[b], Lya Meister[b], Jaan Ross[a], Marju Raju[a], Veeda Kala[a], Tuuri Dede[a]

[a] Estonian Academy of Music and Theatre, Tallinn, Estonia
[b] Tallinn University of Technology, Estonia
allan.vurma@eamt.ee; einar.meister@taltech.ee; lya.meister@taltech.ee; jaan.ross@gmail.com; marju.raju@eamt.ee; veeda.kala@eamt.ee; tuurielo@gmail.com

ABSTRACT

In classical singing the intelligibility of the sung text is often less clear than in speaking. Our study aims to evaluate the hypothesis that in singing, compared to speaking, the intensity of voiceless plosives increases less than the intensity of vowels. The second goal is to determine whether pronouncing plosive bursts with greater intensity improves their intelligibility. Analysis of the performances of five Italian arias from the romantic period by five classically trained singers revealed that, in singing, the intensity of the vowels increased by 14.2 dB compared to the same sounds in speaking, whereas the intensity of the plosive bursts increased by only 7.1 dB. In the perception test with 60 participants, pronouncing plosive bursts more forcefully improved their recognition, but mainly only when the stimuli were presented with room reverberation and/or with pink noise imitating the masking sounds of the accompaniment.

Keywords: intelligibility, plosives, intensity, reverberation, masking

1. INTRODUCTION

In singing, it is the vowels that mainly determine the general sonority of the voice, while to secure text intelligibility a much greater functional load is on consonants [8, 17]. Research addressing consonants in singing is scarce. In the literature we can find some points of view expressed by voice teachers on how consonants should be pronounced to improve text intelligibility. These opinions, however, often contradict each other. For example, Ware [17], Christy [4], Vennard [16], Melton [11], and Sharonova [13] claim that consonants in singing should be pronounced with exaggerated power, while Fuchs [9], Marshall [10], and Brown [3] have stated that singers, on the contrary, should avoid this.

In the present study we focus on researching the intensity balance between vowels and voiceless plosives. While the intensity of vowels mainly depends on the level of subglottal air-pressure created by the lungs, the air-pressure to produce plosive bursts is created in the mouth cavity by rapidly decreasing its size, e.g., by quickly raising the larynx. The breathing system related to the lungs may not play a direct role here [1]. Therefore, the relative independence of the systems which determine the air pressure to produce vowels and the air pressure to produce /k/, /p/, and /t/ bursts may provide a theoretical basis to explain why, when changing the overall voice intensity, the intensity of vowels and voiceless plosives could change by a different amount.

In terms of text intelligibility, not only the plosive burst but also the glide of the formant frequencies from the *locus* of the plosive to the adjacent vowel may also play a role in providing a clue [6, 7]. Additional factors which could worsen the sung text intelligibility are the reverberation of the concert or opera hall and the sounds from the accompaniment and/or ensemble partner(s), which can mask the singer's voice [12].

This study consists of two stages. In the first stage, our aim is to find out whether or not in singing, compared to speaking, the intensity of voiceless plosives increases less than the intensity of vowels. The goal of the second stage is to investigate whether producing more intense plosive bursts improves their recognition.

2. STAGE I

2.1. Method I – Acoustic analysis and comparison of sung and read texts

Five singers (native Estonian speakers) were asked (1) to sing an Italian aria from a romantic period opera and (2) to read the text of the same aria. The recordings (with SPL calibration @30cm) were carried out in a low-reverberation recording studio using Soprano 1.0.20 software and a DPA SC4061-FM omnidirectional microphone (distance from the corner of the singer's mouth about 3 cm). Subsequently, the recordings were segmented into the level of single vowels and consonants (including plosive stops and bursts) using Praat [2]. The intensities of the vowels and plosive bursts were measured with a custom Praat script.

2.2. Results I

In Figure 1 we can see that all the speech sounds investigated were produced more intensely on average in singing than in reading. The average intensity of vowels was 94.5 dB (*SD* = 6.2 dB) and that of voiceless plosives 71.5 dB (*SD* = 6.3 dB) in singing, compared to 80.3 dB (*SD* = 5 dB) and 64.4 dB (*SD* = 5.3 dB) in speaking. The increase in intensity was 14.2 dB on average in the case of vowels but only 7.1 dB in the case of voiceless plosives. This difference is statistically significant ($t = 13.1$, $p < 0.001$).

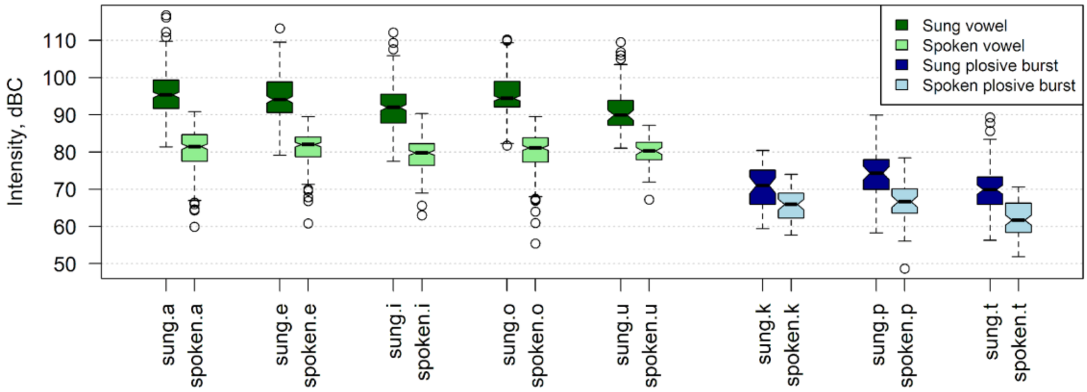

Figure 1. The boxplots of the intensities of sung and spoken vowels and voiceless plosives on an absolute scale (dBC weighting was used in the calibration).

2.3. Are the speech sounds intensity relationships similar in the case of Estonian and Italian singers?

The question may arise whether the aforementioned intensity relationships could depend on the mother tongue of the singer. To investigate that possibility, we also analysed three performances of Donizetti, Puccini and Verdi opera arias by three Italian singers recorded *a cappella* in a low-reverberation studio for another study in Bologna, the data of which has been made freely available in the internet repository (https://zenodo.org/record/3628247#.Y61fEHZBy5c accessed on 29.12.2022).

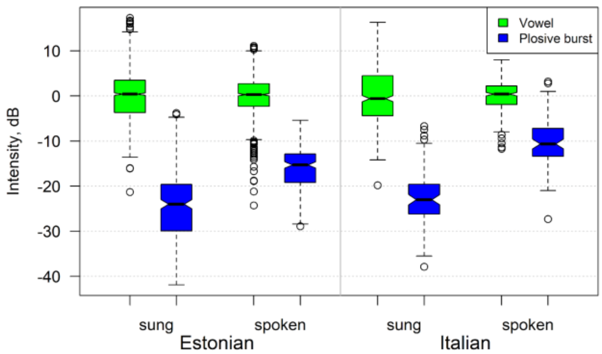

Figure 2. The intensity relationships between vowels and voiceless plosives in the case of Estonian and Italian vocalists.

To rate the intensity relationship between vowels and voiceless plosives in the case of reading, we investigated some random spoken excerpts in Italian (with great probability featuring native Italian speakers) from the internet (www.newsinslowitalian.com, accessed on 10.10.2022), the original purpose of which was to teach the Italian language.

In Figure 2, we can see that all the intensity relationships between vowels and voiceless plosives and between singing and reading are very similar in the case of Estonian and Italian singers and speakers. Therefore, we have good reason to believe that the issue addressed in our study is universal and does not depend to any great extent on the language background of the singer or speaker.

3. STAGE II

3.1. Method II – Perception tests

3.1.1. Stimuli

A classically trained mezzo-soprano sang the /a-k-a/, /a-p-a/ and /a-t-a/ sequences in a slow tempo on G4 ($f0 = 392$ Hz) several times, varying each time the intensity of the plosive burst. Versions with a weak, intermediate and strong burst were selected as the basis for the stimuli for the perception test. Additional versions were created in which the plosive bursts were removed and replaced with silence. On the basis of these 12 stimuli (3 plosives × 4 burst magnitudes), we produced five more series of stimuli with the addition of reverberation (to imitate the room acoustics) and/or pink noise (to imitate the masking sounds from accompanying instruments). For this purpose, we used the *Praat Vocal Toolkit* [5]. For the reverberation, we applied the script's pre-programmed choice, "church", with two different reverberation rates (25% and 50%). We may estimate that the reverberation time T60 of the acoustics modelled by the script was about five seconds. The complete paradigm of stimuli used for our perception test was 3 plosives × 4 burst intensities × 6 acoustic conditions = 72 stimuli.

3.1.2. Participants and the procedure

Altogether, 60 people (21 males, 38 females, one of another gender; aged between 11 and 74, average age 41) including the authors of this paper took the perception test. We tried to engage people over a wide age range, as could be the case in an audience at a random opera performance. The mother tongue of 55 of the participants was Estonian, but there were also participants whose mother tongues were Turkish, Chinese, Portugal, Russian and Spanish. We used the online platform *PsyToolkit* [14, 15] to administer the stimuli. The participants had to use their own laptop or mobile phone with headphones in quiet surroundings at a time convenient to them to run the test. For each participant, a unique order of stimuli was generated. Each stimulus was played only once, and the participants had to enter the consonant they heard between the vowels in the dedicated box on the screen. If recognition was not possible, the participant could enter a question mark in the box. It took between 7 and 24 minutes (14 minutes on average) to complete the test.

3.2. Results II

In Figure 3 we can see that if we look at the responses pooled over all six acoustic conditions the proportion of "correct" replies systematically improved with stronger plosive bursts only in the case of /k/, reaching almost 100% at the strongest burst (burst_3). According to the chi-square test, the dependence of the proportion between "correct" and "incorrect" answers from the intensity of the plosive burst was statistically significant ($\chi 2 = 404.6$, $p < 0.001$).

In the case of /p/, the proportion of "correct" identifications first increased with a more intense plosive burst but already reached its maximum level at medium burst intensity (burst_2) to become worse again at the strongest burst (burst_3). Also, in the case of /p/, the dependence of the proportion of "correct" and "incorrect" identifications on the burst intensity was statistically significant both over the extent of the whole curve ($\chi 2 = 135.9$, $p < 0.001$) and when the replies to only the two stimuli burst_3 and burst_2 ($\chi 2 = 8.1$, $p = 0.004$) were compared, indicating that the decay of the curve at burst_3 was not random.

In the case of /t/, the rate of "correct" recognitions (about 80%) did not depend to any great extent on the intensity of the burst and fell significantly to 65% only in the case of the stimuli where the burst was completely removed.

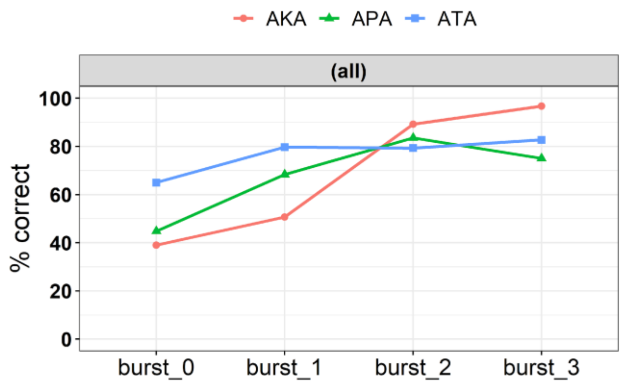

Figure 3. The dependence of the rate of "correct" identifications of plosives on the intensity of the plosive burst (burst_0 – burst removed, burst_1 – weak, burst_2 – medium, burst_3 – strong). The data is pooled over six acoustic conditions.

Despite the decline in the "correct" identifications also in the case of other stimuli with the plosive burst removed, the rate of "correct" identification never fell below the notional 33% borderline which distinguishes mere random guesses. Therefore, we may conclude that besides the plosive bursts, some other types of clues also play a role in identifying the voiceless plosives, and that the absence of the burst (or its complete masking by the accompaniment instruments) need not make the plosive impossible to identify (although the probability of "correct" identification diminishes).

Figure 4 describes the percentage of "correct" identifications split by six acoustic conditions. When the reverberation and/or pink noise were not added to the stimuli, their identification remained high (above 80%) regardless of the plosive burst intensity (the chi-square test was not able to show any statistically significant difference). The only exception is the stimuli with the removed burst (burst_0), in the case of which the level of "correct" identifications of /k/ fell to 22% (which is below the chance level of 33%). The decline in identification for /p/ and /t/ here was far smaller, but the difference in the proportion of "correct" and "incorrect" responses between the stimuli burst_0 and burst_1 was statistically significant in the case of all three plosives (respectively $\chi 2 = 63$, $p < 0.000$ in the case of /k/, $\chi 2 = 4.2$, $p = 0.04$ in the case of /p/ and $\chi 2 = 7$, $p = 0.008$ in the case of /t/).

In the case of the stimuli with added 50% reverberation and with pink noise the recognition generally improved with a stronger burst, although in the case of /p/ there was an optimum burst intensity for the recognition at burst_2.

In the case of /k/, if 50% reverberation, or 25% reverberation and pink noise were added to the stimulus, the recognition improved somewhat if the burst was removed (burst_0) compared with the stimuli with the weakest burst (burst_1). We may speculate that here two clues (the plosive burst and the formant transition from the *locus* of the plosive to the adjacent vowel) are competing. Masking from reverberation and the accompaniment can disturb the clues and create ambivalence. If the plosive burst is absent, the clue from the formant transition becomes dominant and the ambiguity decreases.

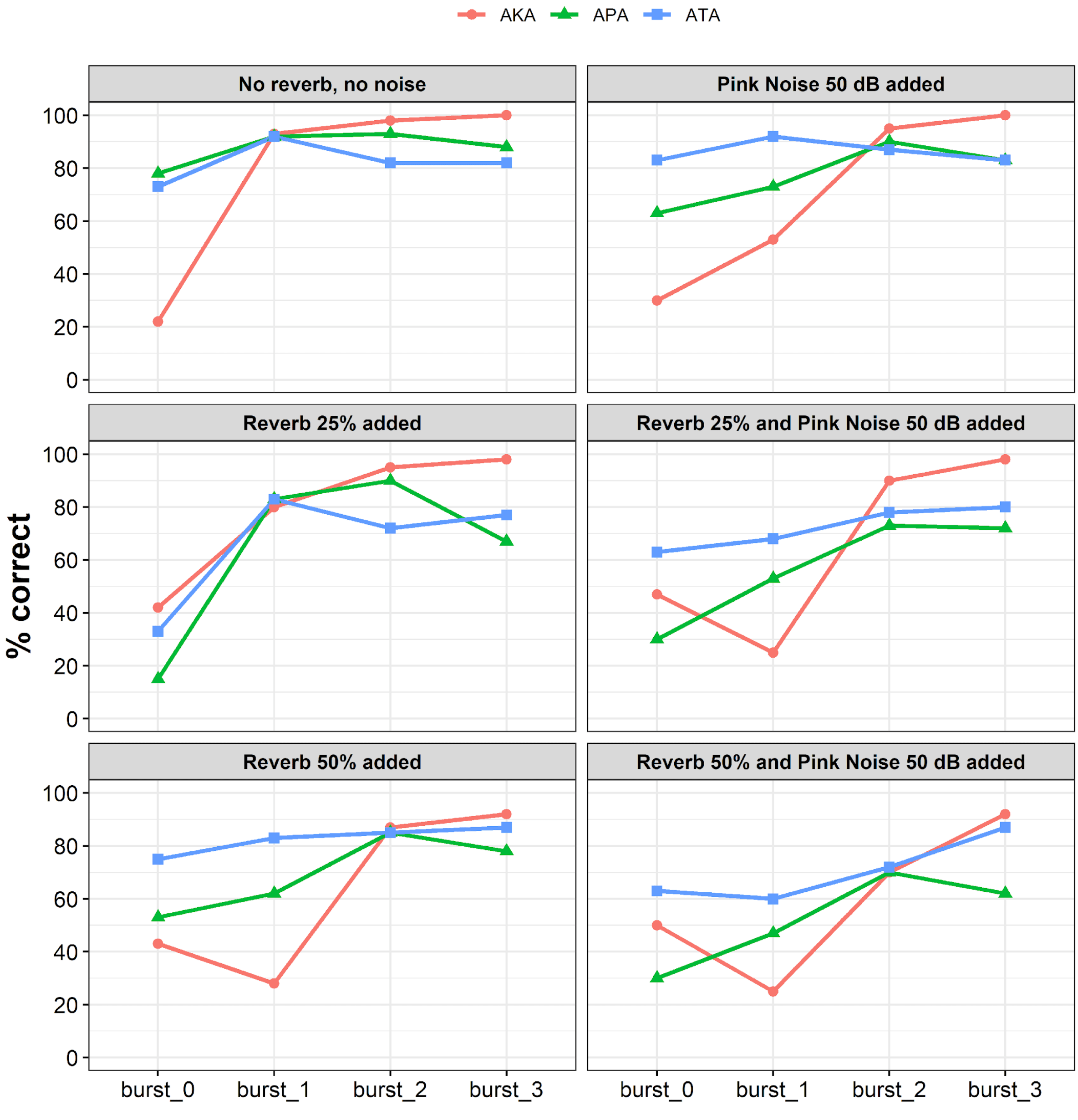

Figure 4. The dependence of the rate of "correct" identifications of plosives on the intensity of the plosive bursts at six acoustic conditions.

4. CONCLUSIONS

The results of the first stage of our study reveal that in operatic singing, when compared to speaking, the balance of intensity between vowels and voiceless plosive bursts changes in favour of the vowels, bringing a relative deterioration in the intensity of the plosives. The results of the second stage show that singing in acoustic conditions with low or absent reverberation and without masking from the accompaniment the intensity of plosive bursts is not critical to text intelligibility – a stronger burst may not improve the intelligibility. However, the total absence of the burst lowers the recognition, especially in the case of /k/. In the case of stronger reverberation and when singing with an accompaniment (or ensemble partners), more intense plosive bursts may improve recognition, although an optimum intensity level of the burst may exist (especially for the /p/) at which the intelligibility is the best.

5. ACKNOWLEDGEMENTS

This study was supported by the Estonian Research Council grant PRG1552.

6. REFERENCES

[1] Behrman, A. 2018. *Speech and voice science. Third edition.* San Diego, CA: Plural Publishing Inc.
[2] Boersma, P., Weenink, D. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.3., retrieved 15 November 2022 from http://www.praat.org//
[3] Brown, R. M. 1946. *The singing voice.* New York, NY: Macmillan Co.
[4] Christy, V. A. 1967. *Expressive singing, Volume 1, Basic principles, A Text for school or studio class or private teaching*. Dubuque, WC: C. Brown Publishing Co.
[5] Corretge, R. 2012-2023. Praat Vocal Toolkit. https://www.praatvocaltoolkit.com (Last viewed Dec 08, 2022).
[6] Deme, A. 2014. Intelligibility of sung vowels: the effect of consonantal context and the onset of voicing. *Journal of Voice,* 28, 4, 523.e19–523.e25
[7] Eek, A, Meister, E. 1996. Acoustics and perception of Estonian word-initial plosives [in Estonian]. *Keel ja* Kirjandus, 4, 241−253.
[8] Fletcher, H. 1929. *Speech and hearing*. New York, NY: D. Van Nostrand.
[9] Fuchs, V. 1964. *The art of singing and voice technique*. New York, NY: London House and Maxwell.
[10] Marshall, M. 1956. Is exaggeration required for good English diction? *Diapason*, 47, 18.
[11] Melton, J. 1953. Do you put the words across? An interview by Annabel Comfort. *Etude*, June, 15.
[12] Meyer, J. 2009. *Acoustics and the performance of music. Manual for acousticians, audio engineers, musicians, architects and musical instrument makers. 5th edition.* New York, NY: Springer.
[13] Sharonova, S. 1947. Diction. *The NATS bulletin*, 3, 4.
[14] Stoet, G. 2010. PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4)*, 1096–1104.
[15] Stoet, G. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44(1)*, 24–31.
[16] Vennard, W. 1967. *Singing: The mechanism and the technic*. New York, NY: Carl Fisher.
[17] Ware, C. 1998. *Basics of vocal pedagogy: the foundations and process of singing*. Boston, MA: McGraw-Hill.