MANAGING REVERBERANT ACOUSTICS IN SINGING BY EXTENDING THE PLOSIVE CLOSURES IN VOWEL–PLOSIVE–VOWEL SEQUENCES

ALLAN VURMA
Estonian Academy of Music and Theatre, Tallinn, Estonia
EINAR MEISTER
Tallinn University of Technology, Tallinn, Estonia
LYA MEISTER
Tallinn University of Technology, Tallinn, Estonia
JAAN ROSS
Estonian Academy of Music and Theatre, Tallinn, Estonia
MARJU RAJU
Estonian Academy of Music and Theatre, Tallinn, Estonia
VEEDA KALA
Estonian Academy of Music and Theatre, Tallinn, Estonia
TUURI DEDE
Estonian Academy of Music and Theatre, Tallinn, Estonia

Abstract

Poor intelligibility of sung text often occurs in reverberant rooms due to masking by the reverberation tail of the singer's voice. This study investigates whether elongating the plosive closure phase can improve the recognition of voiceless plosives in vowel–plosive–vowel sequences sung in reverberant rooms. We hypothesize that a longer plosive closure allows the reverberation tail from the preceding vowel to decay before the plosive burst, thus reducing masking and enhancing plosive recognition. In Experiment I, 34 listeners heard stimuli (sung single-pitch vowel–plosive–vowel sequences) via headphones, with artificial reverberation and/or Brown Noise added to simulate different acoustics. Experiment II involved 33 listeners in a concert hall, where stimuli were played from a loudspeaker on the stage, and Brown Noise was played from a separate sound system. The plosive closure phase in the stimuli was edited using *PRAAT* software to durations 60 ms, 150 ms, or 260 ms. Recognition of plosives improved by up to 25 percentage points with longer closure phases, depending on the acoustic condition, burst intensity, and vowel pitch. Older listeners, and listeners seated in the back rows of the concert hall, showed poorer recognition. Extending the plosive closure phase generally did not improve plosive recognition in non-reverberant acoustics.

Keywords: sung text intelligibility, plosives, reverberation, room acoustics, masking

Managing reverberant acoustics in singing by extending the plosive closures in vowel–plosive–vowel sequences

In singing, particularly in classical style, in large halls or rooms with long reverberation, and also in the case of female voices and at high fundamental frequencies (f_o), poor intelligibility of sung text is often a problem (Gregg, 1991; Nelson & Tiffany, 1968; Phillips, 2002; Titze, 1982). However, the views of different singers and voice teachers concerning diction in singing and the vocal-technical means to improve it are controversial. For example, some vocal pedagogues have claimed that for good diction singers should pronounce consonants with greater intensity (Christy, 1967; Melton, 1953; Sharnova, 1947; Vennard, 1967; Ware, 1998), while others advise against this (Brown, 1946; Fuchs, 1964; Marshall, 1956). Sung text intelligibility may also depend on the presence and nature of other concomitant sounds in the room that can mask the singer's voice to a greater or lesser extent. These masking sounds can be autonomous, e.g., the orchestral accompaniment, or the room reflections of the singer's own voice—the reverberation caused by the room acoustics (Meyer, 2009). The grounds for the intelligibility of sung text are neurological processes that use direct bottom-up processing of acoustic information, alongside top-down processes in which a listener's language skills and the whole context around the text play a role in the predictions and expectations of what text will be sung next (Behrman, 2023). Thus, language proficiency or foreign accents may also affect intelligibility.

Although poor intelligibility can also occur in spoken text, this issue is often more pronounced in singing. For example, a study by Collister and Huron (2008) found that text intelligibility diminished by 76% in sung phrases compared to spoken phrases. Although the voice apparatus is the same in both cases, the two activities differ in many ways: in singing, as opposed to speaking, the pitch range used is typically wider, extends far higher, and is divided into discrete finite steps; in operatic style, particularly, singers typically produce higher voice intensity levels. Furthermore, sound durations need not precisely match the pattern of the spoken text. In singing,

pitches, note durations, and often also dynamics, are prescribed by the composer, though singers typically use some degree of freedom of interpretation, expressed by greater or lesser deviations from the written score (Friberg, 1991).

Poor sung text intelligibility may be related to both vowels and consonants (Appelman, 1986). In the present study, we focus specifically on voiceless plosives in vowel–plosive–vowel sequences sung in rooms with reverberant acoustics. The articulation of plosives starts with the closure phase (which is bilabial for /p/, alveolar for /t/, and velar for /k/) during which air pressure builds up behind the closure. This is followed by the burst phase, where the air pressure is released with an explosion-like noise. Additionally, in the presence of other adjacent speech sounds (vowels) the transition of vocal tract formants occurs (Behrman, 2023). The duration of a plosive closure phase can only be measured when another speech sound immediately precedes the plosive. In post-pausal positions the duration of the closure phase of a word-initial plosive cannot be measured because the acoustic information about when the closure phase starts and how long it lasts is lacking. The perceptual cues for identifying plosives include: (1) the frequency distribution of energy in the noise spectrum of the plosive burst, and (2) the trajectories of vocal tract formant transitions (Behrman, 2023). In addition, voice onset time (VOT)—the time gap between the beginning of the plosive burst and the onset of vocal fold vibration—can play a role in distinguishing between voiced and voiceless plosives (Behrman, 2023).

According to Vurma et al. (2023), the recognition of voiceless plosives in loud operatic singing as opposed to speaking can be hindered due to the stronger masking effect caused by the reverberation tail of vowels immediately preceding the plosive. As voice intensity increases, vowels tend to intensify proportionally more than plosives, especially when the f_o of the vowel is high. The experiments showed that the recognition of plosives improved when the plosive bursts were pronounced more strongly. However, this improvement occurred primarily in reverberant acoustics and/or with accompanying noise that simulates masking from orchestral accompaniment (Vurma et

al., 2023). To our knowledge, there are no published studies addressing whether and to what extent the duration of the plosive closure phase affects the recognition of plosives when singing in halls with reverberant acoustics, or concerning the standpoints of singers and voice teachers on this matter. Some published information indirectly related to this subject, however, does exist. For example, there are studies showing that the intelligibility of spoken text can be improved by inserting pauses at word boundaries (Petkov et al., 2016) or prosodic boundaries in synthetic speech (Best et al., 2015; Scharpff & van Heuven, 1988). Similarly, it has been reported that slowing down the speech enhances text intelligibility, especially in the case of older people (Gygi & Shafiro, 2014; Lessa & Costa, 2013; Schmitt & McCroskey, 1981). According to Koutsogiannaki (2016), the positive effect of longer pauses on text comprehension may arise from the listener's nervous system having more time for the top-down cognitive processes involved in text comprehension. However, conflicting findings suggest that pauses or altered reading speeds do not always improve text comprehension (Adams et al., 2012; Gordon-Salant & Fitzgibbons, 1997). As tempo in singing is mostly determined by the music, singers usually have little freedom to change it in order to improve sung text intelligibility.

We can also find some related statements of vocal pedagogues, although these tend to concern consonants in general. For example, according to Belisle (1967), “when plosives are given duration, they provide that element of separation between the vowel sounds which prevents those sounds from being merged.” Appleman (1986) argues that “the consonant must have enough duration to possess undeniable entity in the vocal line. It must complement the vowel, but although it is of shorter duration, it must never be of lesser importance than the sustained sound.” Nair (2016) claims that it is likely that “great singers produce consonants with a resonance that rivals the vowels, and lengthen the consonants to maintain diction parity with those vowels.” Eberhart's (1962) point of view, however, is the opposite, claiming that consonants should be pronounced clearly and sharply in singing but not made longer or more intense unless necessary for reasons of

interpretation or mood. Similarly, Di Carlo (2007) says that the duration of consonants in singing is actually slightly shorter than in speaking. However, neither Nair nor Di Carlo have presented any published empirical measurement data on this.

In the present study, we hypothesize that when singing vowel–voiceless plosive–vowel sequences in reverberant rooms, the recognition of plosives may improve if the singer slightly extends the closure phase of the plosive. We assume that this would allow the reverberation tail of the vowel immediately preceding the plosive more time to decay before the onset of the plosive burst, resulting in weaker masking of the burst by the vowel's reverberation tail. In this paper we present two perceptual experiments to test this hypothesis. Experiment I uses artificially simulated acoustics and presents the stimuli to listeners through headphones, ensuring identical auditory conditions for all participants. Experiment II involves listeners in a concert hall setting, with stimuli played from a loudspeaker on the stage. This enabled us to achieve a greater degree of ecological validity in the test design, but the acoustic input at the ears of the listeners here depended somewhat on the individual listener's location in the hall. Additionally, we explore how various factors—such as the pitch of vowels adjacent to the plosive, plosive type, burst intensity, listener gender and age, and the interaction of these with the plosive closure duration—affect plosive recognition. Investigating these interactions is important as it helps determine whether the presence and magnitude of expected improvements in plosive recognition due to extending the plosive closure duration depend on these factors. For these analyses, results from both experiments were combined into a single dataset.

The pitch of the vowel in the vowel–voiceless plosive–vowel junction may influence intelligibility, as it affects how acoustically informative vocal tract formant transitions can be as cues for plosive recognition. The higher the pitch, the greater the distance between the spectral partials of the vowel, thus rendering formant transition cues less informative. Additionally, formant tracking is commonly used by classically trained female singers (sopranos and mezzo-sopranos) at high

pitches, starting from the *secondo passaggio* region (usually at about F5), where the frequency of the first formant can be raised much higher than is typical of spoken vowels (Sundberg, 1987; Ware, 1998). With formant tracking, the singer tunes the first formant of the vocal tract to the rising first harmonic of the voice spectrum to increase the sound pressure level (SPL). This technique reduces vowel distinctiveness, making all vowels sound more like /a/ (since the first formant of this vowel is the highest of all five basic vowels; Sundberg, 1987). According to Di Carlo (2007), it is impossible to achieve good sung text intelligibility at fundamental frequencies above about 659 Hz (E5). The type of plosive could also be important, as smaller articulatory maneuvers are required to move between /a/ and /k/ than between /a/ and the other plosives. This may play a role at high pitches due to the formant tracking described above. Unlike /p/ and /t/, only /k/ can be articulated while simultaneously keeping the mouth wide open. Therefore, at high pitches, it is perhaps easier for female singers to produce /k/ in combination with adjacent vowels than to produce /p/ and /t/.

Additionally, listener age could influence text intelligibility due to presbycusis, an age-related decline in hearing sensitivity, especially at the higher frequencies crucial for perceiving plosive sounds. This decline tends to be more pronounced in men than in women. (Howard & Angus, 2017)

Experiment I (Stimuli with Artificially Added Acoustics)

Methods

To explore the impact of plosive closure phase duration on plosive recognition, we conducted perception tests where the participants were asked to identify a plosive within vowel–plosive–vowel sequences sung on a single pitch, with the duration of the plosive closure phase intentionally varied. In Experiment I, to simulate room acoustics, we presented stimuli with artificially added reverberation to participants via headphones. In both experiments, participant responses were analyzed using Generalized Linear Models (GLM) due to the binary nature of the responses (which excluded the use of ANOVA). GLM allows us to analyze relationships between a dependent variable and one or more independent variables by using a link function (binomial logistic

regression) to connect the dependent variable to the predictors, estimating both the significance and magnitude of each predictor's influence. Different predictors were used in separate GLMs depending on the focus of each statistical test.

Stimuli

To create the stimuli, we enlisted two classically trained professional singers—a mezzo-soprano and a tenor. The mezzo-soprano performs at international level in concerts, oratorios, and recitals (taxonomy category 2.4), while the tenor also performs internationally in principal operatic roles (categories 2.4 and 2.1; see Bunch & Chapman, 2000). They were asked to sing series of /aka/, /apa/, and /ata/ sequences at constant pitches within each series. The mezzo-soprano produced two series: on G4 (f_o = 392 Hz) and on F5 (f_o = 698.5 Hz), while the tenor sang one series on G3 (f_o = 196 Hz). We selected F5 for the mezzo-soprano as it lies in the *secondo passaggio* region of female singers (Ware, 1998), where achieving good text intelligibility is technically more challenging and strenuous than at lower pitches. At the same time, F5 is lower than the top high notes of sopranos, where the decline of text recognition is typically even greater (Sundberg, 1987). For comparison, the same mezzo-soprano also sang the stimuli on G4, which is in the comfortable middle register above speaking pitch (Ware, 1998). The tenor's G3 pitch, in a comfortable range for tenors and slightly above speaking pitch (Miller, 1986), allowed us to observe if plosive closure extension similarly affects plosive recognition at a significantly lower pitch and in a male singer. The voice category of the singers and chosen pitches were not critical for our study. However, we decided not to use any larger number of singers and stimuli, as this would have overloaded our listeners participating in the perception tests.

The G3 and G4 series were sung in two plosive burst intensity versions—a "weak burst", representing the singers' initial spontaneous production, and a "strong burst", with the singers slightly increasing the burst intensity (without a prescribed level of increase). The SPL difference between weak and strong burst intensities was 7 dB in the series on G3 (44 dB vs. 51 dB) and 2 dB in

the series on G4 (45 dB vs. 47 dB). The mezzo-soprano's F5 series included only one burst intensity (44 dB) due to the difficulty of significantly varying burst intensity at this high pitch. Besides the difference in burst intensity, burst durations were approximately 10 ms longer in the case of strong bursts (35 ms vs. 45 ms), adding subjective loudness to the burst (Howard & Angus, 2017). The SPL of adjacent vowels was about 67 dB in the G3 and G4 series and approximately 72 dB in the F5 series, consistent with the typical SPL increase of sung vowels at higher pitches (Sundberg, 1987).

The recordings were conducted in a recording studio with low reverberation (reverberation time T30 = 0.2 s) using the omnidirectional microphone DPA SC4061-FM (positioned at 3.5 cm from the corner of the singer's mouth), the audio interface Audient ID4, and a Dell Latitude 5400 laptop. T30 shows how long it takes for the sound to fade by 30 dB once it has ceased.

Choosing the Plosive Closure Phase Durations in the Stimuli. Next, we manipulated the plosive closure phase durations in the stimuli using *PRAAT* software (Boersma & Weenink, 2024), either extending closures by inserting slices of silence or shortening them by cutting portions of silence. Rather than relying on naturally sung variations by the singer, we used sound editing software to ensure precise control over the closure duration without altering other parameters that could affect recognition. We based the plosive closure duration values on statistical data from a database in a study by Vurma et al. (2023). In that study, 10 professional classically trained opera singers (two basses, two baritones, two mezzo-sopranos, two sopranos) sang Italian arias from their repertoire, and read the lyrics of these arias both in the manner of ordinary conversation and oratorically. The median value of the plosive closures pooled over all performances was 72 ms, quartiles 52 ms and 102 ms, and the longest value was 364 ms. For the stimuli in the present study, we decided to use the following plosive closure durations: 60 ms (slightly shorter than the median value in our database), 260 ms (corresponding to the outliers), and 150 ms (located in the midway between these two values). The burst durations and transitions to the following vowel were left intact for the sake of naturalness. Notably, the correlation between the closure and burst durations was very weak

(r = 0.13), which justifies focusing on manipulating closure duration alone. Vowel durations were fixed at 600 ms and 900 ms for Vowel 1 and Vowel 2 respectively.

Following this, we added artificial reverberation (simulating room acoustics) and/or Brown Noise (BN; simulating orchestral accompaniment) to the stimuli for Experiment I. In the case of Brown Noise, the decay of the long-term average spectrum (LTAS) with frequency (-6 dB/octave) is similar to that of a typical symphony orchestra (-9 dB/octave; Lindblom & Sundberg, 2007). Using the *PRAAT* Vocal Toolkit (Corretge, 2024) we applied presets Big Room 70% and Church 70%. These presets simulate environments where 70% of the overall sound energy comes from the reverberant field component. Based on our estimates, the "Big Room" preset corresponds to room acoustics with a T60 of around 1.3 seconds, while the "Church" preset corresponds to a T60 of about five seconds. (T60 shows how long it takes for the sound to fade by 60 dB.)

Estimating the Baseline of Plosive Recognition. To assess baseline plosives recognition and examine the potential influence of plosive closure duration without added acoustics, we conducted a pilot test with three participants. They were presented with "Clear" stimuli (without artificially added acoustics) through headphones at approximately 65 dBA (about 70 dBA for the F5 series). Participants recognized the stimuli in the G3 and G4 series with 100% accuracy. However, the recognition of the stimuli in the F5 series was about 90%. To avoid overloading the participants in the main tests, we included "Clear" stimuli only in the F5 series, assuming G3 and G4 series recognition would likely remain near 100%.

Summary of the Perception Test Design. To sum up, for Experiment I, we created three series of stimuli. The G3 and G4 series included: three plosives (p, t, k) × three plosive closure durations (60 ms, 150 ms, 260 ms) × two burst intensities (weak, strong) × five acoustic conditions (Big Room, Church, Brown Noise, Big Room+Brown Noise, Church+Brown Noise)—totaling 90 stimuli.

The F5 series included: three plosives (p, t, k) × three plosive closure durations (60 ms, 150 ms, 260 ms) × one plosive burst intensity × six acoustic conditions (Clear, Big Room, Church, Brown Noise, Big Room+Brown Noise, Church+Brown Noise)—totaling 54 stimuli.

Participants and Procedure

A total of 34 participants (11 male, 23 female, aged between 21 and 68 years, mean 44 years) took part in Experiment I. They were recruited through personal contacts of the research team from the Tallinn University of Technology and the Estonian Academy of Music and Theatre, using student and employee mailing lists, as well as posters and social media. Participants reported their age, gender, and native language; some had a musical background (e.g., EMTA students), though no further details were collected. Participation was voluntary, with no compensation offered.

Stimuli were presented using the *PRAAT* Listening Experiment platform on an HP EliteBook 2560p laptop equipped with an Ifi Zen Digital-to Analog Converter (DAC) and calibrated headphones HD560s. The SPL of the stimuli was approximately 65 dBA in the G3 and G4 series and around 70 dBA in the F5 series. Participants had to identify the plosive between the two vowels by clicking on the dedicated button on the computer screen (four forced choices: k, p, t, or "?" for uncertain). Stimuli were presented in a randomized order unique to each participant, with each stimulus occurring three times in random order during the test. Participants had unlimited time to make their decision, but stimuli could not be replayed. Tests were conducted in a quiet soundproof room, with each series (on G3, G4, and F5) taking about 15–20 minutes, and participants were able to take a few minutes rest between series.

The ethical risks of this study were low as the singers whose voice was used to create the stimuli, as well as the participants in the perception tests, remained anonymous. Moreover, the focus was not on evaluating individual performances but on the characteristics of abstract sounds.

Results: Experiment I

General Distribution of Responses

INSERT FIGURE 1 ABOUT HERE

Across all pooled responses, the proportion of "?" responses was only six percent, indicating that in most cases the participants were able to categorize the plosives. The most often mutually confused plosives were /k/ and /t/. In the G3 and G4 pitch series, the recognition rate was between 57.3% (for /t/ in the G4 series), and 80.8% (for /t/ in the G3 series). Among the three pitch series, the recognition was poorest in the F5 series (35.8% for /p/, 54.5% for /t/, and 73% for /k/; see Figure 1).

To assess agreement among the participants, we calculated inter-individual reliability using the intraclass correlation coefficient (*ICC*) [the model *ICC* (2, *k*), two-way random effects, the mean of *k* participants (*k* = 34), absolute agreement] as described by Shrout and Fleiss (1979). The level of agreement was excellent across all three plosive categories (*ICC* was between 0.97 and 0.98, $p < 0.001$). High *ICC* values indicate that the stimuli that were correctly recognized, and those that were not tended to be the same across the listeners.

Statistical Analysis: Experiment I

We used a Generalized Linear Model (GLM) to analyze how the probability of correct responses depended on the following independent factors: (1) plosive closure duration (with a base level of 60 ms, and levels of 150 ms and 260 ms); (2) acoustic condition (with a base level of Clear and additional conditions including Brown Noise (BN), Big Room (BR), Church (Ch), Brown Noise + Big Room (BN_BR), and Brown Noise + Church (BN_Ch); and (3) the interaction between closure duration levels and acoustic conditions. In the GLM, the base levels serve as reference points for comparison. The data comprised pooled responses from listeners across the three pitch series (G3, G4, and F5). Only stimuli with a weak burst were included, as the pitch series on F5 used only these stimuli.

The GLM confirmed that, compared to the baseline, all six acoustic conditions had a statistically significant negative effect on plosive recognition (Cond_BN (β = -1.37, p < .001), Cond_BR (β = -2.90, p < .001), Cond_Ch (β = -3.00, p < .001, Cond_BN_BR (β = -3.39, p < .001), and Cond BN_Ch (β = -3.52, p < .001)—see Table 1, and Figure 2). However, closure durations were only significant when interacting with the Big Room and Church acoustic conditions, at the 260 ms closure duration in both cases (β = 1.28, p < 0.001, and β = 0.56, p = 0.02, respectively). These interactions can be interpreted as follows: While in Clear acoustics (the baseline), the change in plosive' recognition was negligible when the closure duration increased from 60 ms to 260 ms, a statistically significant improvement in recognition was seen in Church acoustics (from 50% at 60 ms closure duration to 61% at 260 ms) and an even greater improvement was observed in Big Room acoustics (from 53% at 60 ms to 78% at 260 ms). However, the changes in BN, BN-BR, and BN-Ch acoustics were insignificant. The estimates (β) and z-values in Table 1 illustrate the magnitude of the effect of each corresponding factor; larger absolute values indicate greater effects. Positive estimate values suggest a positive relationship, while negative estimate values indicate a negative relationship. The predicted probability of plosive recognition was consistently high (over 80%) in acoustics without reverberation (Clear and BN conditions) but remained low (around 40%) for stimuli combined with added reverberation and Brown Noise (BN_BR and BN_Ch).

INSERT FIGURE 2 ABOUT HERE

Discussion

In Experiment I, we presented stimuli with artificially added reverberation through headphones, using the same technical equipment for all listeners. A GLM model was applied to generalize the outcome. Results indicated that in artificial reverberant acoustics, the recognition of voiceless plosives in sung vowel–plosive–vowel sequences improves as the plosives closure phase is extended. This improvement varies with the room's reverberation time, being less pronounced in artificial rooms with longer reverberation ("Church") compared to those with shorter reverberation

("Big Room"), and is absent in rooms with very short reverberation (the recording studio). These findings support our hypothesis: in environments with shorter reverberation, the vowel's reverberation tail decays more during the plosive closure, leading to weaker masking of the plosive.

However, the computational reverberation model may not adequately replicate sound behavior in real concert halls. To evaluate whether the results obtained in Experiment I extend to real acoustics, we conducted Experiment II in a concert hall, using the same stimuli but without artificially added reverberation.

Experiment II (The Stimuli in a Real Concert Hall)

Methods

Stimuli

In Experiment II, stimuli were played through a loudspeaker on the stage of a real concert hall with natural reverberation. The series on the G3 and G4 included: three plosives (p, t, k) × three plosive closure durations (60 ms, 150 ms, 260 ms) × two plosive burst intensities (weak, strong) × two acoustic conditions (Clear, Clear+Brown Noise)—a total of 36 stimuli. For the F5 series, the stimuli included: three plosives (p, t, k) × three plosive closure durations (60 ms, 150 ms, 260 ms) × one plosive burst intensity × two acoustic conditions (Clear, Brown Noise)—a total of 18 stimuli.

In Experiment II, it is important to note that all stimuli (both Clear and Brown Noise) reached listeners through the acoustics of the concert hall. Thus, the Clear stimuli in Experiment II were most comparable to the Big Room stimuli used in Experiment I (due to similar reverberation times) and, similarly, the Brown Noise stimuli in Experiment II corresponded most closely to the Brown Noise+Big Room stimuli in Experiment I.

Participants and Perception Test Procedure

Thirty-three participants (15 male, 17 female, one non-binary; ages 22–66, mean age 40 years) took part in the perception tests, 11 of whom had also participated in Experiment I. The

stimuli were played through a Genelec 8341a loudspeaker located in the center of the stage in the Great Hall of the Estonian Academy of Music and Theater. The loudspeaker was positioned 4 m from the front edge of the stage. The concert hall (width 15 m, length 30 m and height 13.8 m, reverberation time T60 = 1.8 s at 1000 Hz) has 470 seats—300 of which are located in the parterre and the remaining in two balconies (see Figure 3).

INSERT FIGURE 3 ABOUT HERE

Participants were divided into two groups, seated in rows 4 and 5 at the front and in rows 12 and 13 in the rear part of the hall. One seat was left empty between two adjacent participants. Each test series was conducted twice, with participants swapping seats between the front and back rows. The SPL of the stimuli (vowels) was about 90 dBA at 30 cm from the loudspeaker (which corresponds to approximately the *mf* loudness level of a singer's voice (Lamarche et al., 2010; Meyer, 2009), and about 65 dBA for participants in the front rows and 55 dBA for those at the back. When Brown Noise was added, it was played from loudspeakers above the stage. The SPL of the Brown Noise was close to that of the vowels, and it also depended on the seating position.

Participants recorded their responses (four choices—p, t, k, ?) in designated tables on their response sheets. Each stimulus was followed by a 2.5-second silence to allow time for writing. After every 18th stimulus, there was a longer break (about 10 seconds) for a short rest. The order of the stimuli was randomized but remained consistent across participants. As in Experiment I, each stimulus occurred three times randomly during the test. To avoid errors, a clearly visible sequence number of the currently played stimulus was projected onto the balcony rim. Completing all the test series in Experiment II took approximately 75 minutes.

Results: Experiment II

General Distribution of Responses

Recognition accuracy was highest for /t/ in the G3 series (81%) and lowest for /p/ in the F5 series (18%), where it was frequently misidentified as /t/ (57% of cases). The response "?" was selected in only 4% of cases (see Figure 4). In Experiment II, we again computed the *ICC* (k = 33). The level of agreement was excellent in all three sets (*ICC* was between 0.95 and 0.99, $p < 0.001$).

INSERT FIGURE 4 ABOUT HERE

Statistical Analysis: Experiment II

In Experiment II, we combined the results from the G3, G4, and F5 test series and applied a Generalized Linear Model (GLM) to analyze the effects of the following independent factors: (1) Duration, with a baseline level of 60 ms and additional levels of 150 ms and 260 ms; (2) Acoustic Condition, with a base level of Clear and an alternate level of Brown Noise (BR); and (3) the interaction between Duration and Acoustic Condition. The dependent variable was the probability of correct responses (Correct). Consistent with Experiment I, only stimuli with weak bursts were used since the singer was unable to produce strong versions of plosive bursts in the F5 series.

The GLM revealed statistically significant main effects as well as interactions between Duration and Condition (see Table 2). In the case of Clear stimuli increasing closure duration from 60 ms to 150 ms to 260 ms improved recognition correspondingly from 62% to 68% to 73% ($\beta = 0.25$, $p < .001$ for 150 ms, and $\beta = 0.5$, $p < .001$ for 260 ms). In contrast, the recognition of stimuli with added Brown Noise (BN) remained stable at approximately 44%, regardless of the plosive closure duration. This outcome is similar to the recognition rates observed for the related BN+BR stimuli (with artificially added Brown Noise and Big Room reverberation) in Experiment I (see Figure 2 for comparison).

INSERT FIGURE 5 ABOUT HERE

INSERT TABLE 2 ABOUT HERE

Seating in the Hall. In addition, we used data from Experiment II to investigate whether the listener's seating position in the concert hall affected the plosive recognition and if there was any interaction with the plosive closure duration. For this we employed another Generalized Linear Model where the dependent variable was the probability of correct responses; in this case the independent factors included (1) plosive closure Duration, with a base level of 60 ms, and additional levels of 150 ms and 260 ms; (2) listener location, with a base level of "at the front" (LocFront), and a level for "at the back" (LocBack); and (3) the interaction between duration and location. This model, too, was based on data pooled from the G3, G4 and F5 pitch series, including only the responses to stimuli with weak bursts and excluding those with accompanying Brown Noise, as no significant effect of closure duration was observed when noise was present.

Both plosive closure duration and listener location in the hall had statistically significant effects on recognition ($p < 0.05$ and better). In the back rows, recognition increased from 60% at a closure duration of 60 ms to 69% at 260 ms. In the front rows, recognition similarly increased from 65% at 60 ms to 77% at 260 ms (see Figure 6). However, this interaction did not reach statistical significance.

INSERT FIGURE 6 ABOUT HERE

Discussion

The results of perception tests and the probabilities of correct plosive recognition predicted by the Generalized Linear Model in Experiment II support the findings from Experiment I—elongating the plosive closure duration in vowel–consonant–vowel sequences improves the recognition of plosives also when sung in the acoustics of real concert halls. This effect was more pronounced in Concert Hall acoustics, which have moderate reverberation time, compared to Church acoustics with longer reverberation times. In addition, Experiment II enabled us to observe the influence of the balance between direct and reverberant sound on plosive recognition.

Participants were seated in two groups: those in the front rows, closer to the sound source, where the direct sound component was more prominent, and those in the back rows, where the balance shifted towards the reverberated sound component. (Howard & Angus, 2017; Meyer, 2009) For stimuli without Brown Noise, the benefit of extending the plosive closure duration was quite similar in the front and back rows, although the recognition of plosives in general was somewhat worse for listeners in the back rows, which was predictable— excessive reverberation is known to reduce intelligibility (Howard & Angus, 2017; Meyer, 2009). This effect was also observed when comparing plosive recognition in Church and Big Room acoustics in Experiment I.

Impact of Additional Factors on Plosive Recognition

Next, we explore whether some additional factors, such as vowel pitch, type of plosive, intensity of the plosive burst, and listener characteristics (age and gender), as well as their interactions with plosive closure duration, influence plosive recognition. The effect of our primary interest—the impact of plosive closure duration on plosive recognition—was consistent across both experiments and particularly significant in Big Room acoustics and in the Concert Hall, where the reverberation times were similar. Therefore, we combined the relevant datasets (Big Room results from Experiment I and Clear results from Experiment II) for further statistical analyses. As before, we focused on data with weak plosive burst intensity, except when specifically analyzing the impact of burst intensity. For these analyses, we also used the Generalized Linear Models for statistical evaluation. For all subsequent models, we define a base model with the proportion of correct recognitions as the dependent variable, with plosive closure duration (baseline of 60 ms, with additional levels of 150 ms and 260 ms) as the main predictor. Additional predictors were incorporated depending on the focus of each statistical test.

Pitch

In the Generalized Linear Model examining the influence of pitch, two additional predictors were added to the base model: (1) pitch, with a baseline level of G3 and levels G4 and F5, and (2) the interactions between duration and pitch. The model revealed that increasing closure duration improved recognition, with 150 ms (β = 0.76, p < .001) and 260 ms (β = 1.46, p < .001) showing higher accuracy compared to the baseline. In contrast, higher vowel pitch negatively affected recognition, with G4 (β = -0.20, p = .04) and F5 (β = -0.94, p < .001) both reducing accuracy compared to the baseline pitch level. All interactions were significant at p < 0.01, with most reaching p < 0.001, and were negative, indicating that the improvement in recognition accuracy with longer durations was smaller at higher pitch levels (G4, F5) (see Table 3 and Figure 7).

INSERT TABLE 3 ABOUT HERE

INSERT FIGURE 7 ABOUT HERE

Plosive Type

To examine the influence of plosive type, the predictor variable consonant (baseline level /p/ and additional levels /t/ and /k/), along with interactions between duration and consonant, were added to the base Generalized Linear Model. The results indicated that plosive closure duration had a statistically significant effect on plosive recognition, with p < 0.001 for all durations. However, only the recognition of /k/ was significantly different (improving by approximately 10 percentage points) from the baseline level /p/ (β = 0.30, p = 0.002; see Figure 8). Additionally, there was no significant interaction between duration and consonant, suggesting that extending the plosive closure duration led to similar improvements in recognition for all three plosives.

INSERT FIGURE 8 ABOUT HERE

Age and Gender of the Listener

The base Generalized Linear Model was expanded to include the predictor variables Age Group (baseline level "Old" and additional level "Young"), Gender (baseline level "Female" and

additional level "Male"), and various interactions (Duration × Age, Duration × Gender, Duration × Age × Gender). The analysis revealed that only the effects of plosive closure duration and age group were statistically significant, ($p < 0.01$), but not their interactions. Younger participants had approximately a ten-percentage-point higher probability of correctly recognized plosives compared to older participants, but the percentage of the improvement in plosive recognition from extending the closure duration was similar regardless of the individual participant's age.

Burst Intensity

To estimate the impact of burst intensity on plosive recognition, we integrated responses to the stimuli with both weak and strong bursts in pitch series G3 and G4, excluding data from pitch series F5, which used only weak burst intensities. In the corresponding Generalized Linear Model, the probability of correct responses was the dependent variable, while the independent factors were: (1) plosive closure duration (baseline of 60 ms and additional levels of 150 ms and 260 ms); (2) plosive burst intensity ("Weak" as the baseline, with "Strong" as an additional level); and (3) their interactions (Duration × Burst). According to the model results, both duration and burst intensity had statistically significant influences ($p < 0.001$). Figure 9 illustrates that longer plosive closure durations (260 ms versus 60 ms) improved plosive recognition by 17 percentage points for weak bursts, compared to 12 percentage points for strong bursts. This suggests that there was less room for improvement when the bursts were stronger. However, this interaction (with $p = 0.08$) did not reach statistical significance.

INSERT FIGURE 9 ABOUT HERE

General Discussion

In our study, both Experiments I and II found that in reverberant acoustics the recognition of voiceless plosives in sung vowel–plosive–vowel sequences improves when the closure phase of the plosive is slightly extended. This improvement can be attributed to a better sound level balance

between the plosive and the reverberation tail of the adjacent vowel, which can (partially or fully) mask the plosive. When the reverberation field has more time to decay before the plosive burst, the masking effect weakens, enhancing plosive recognition. Our outcome supports the position of those vocal pedagogues (e.g., Belisle, 1967; Appleman, 1986; Nair, 2016) who argue that for good text intelligibility in singing, consonants should be given sufficient duration. However, our study also shows that the benefit of elongating the plosive closure may disappear in rooms with a low reverberation, which aligns with the opposite standpoints by Eberhardt (1962) or Di Carlo (2007) that consonants should not made longer in singing. The rooms which typically have low reverberation and where singers might often practice are living rooms and small rehearsal rooms with soft furniture and curtains (Howard & Angus, 2017, Meyer, 2009). Therefore, a performance which sounds clear in a small rehearsal room is not necessarily intelligible in a bigger concert hall.

Other factors, such as the intensity and duration of pre-plosive vowels, might also influence plosive recognition and the benefit of elongated plosive closures. For instance, if the pre-plosive vowel has low intensity, its reverberation tail decays faster, falling beneath the audibility threshold sooner than in the case of a strong vowel. Similarly, because it takes time for the reverberation field to build up to a steady state, in the case of a short vowel the steady state level is not reached (Howard & Angus, 2017; Meyer 2009). Thus, the reverberation field of a short, weak pre-plosive vowel masks the subsequent plosive to a lesser extent. This could help explain why plosive recognition at F5 in our study was poorer compared to other pitches. At F5, the SPL of the vowel adjacent to the plosive was about 5 dB stronger than at the other pitches, while the intensity of the plosive bursts remained almost constant across the three pitches. The higher intensity of the vowels in the F5 series was used for better ecological validity, as an increase in intensity with pitch is typical to singing (Sundberg, 1987). In rooms with moderate reverberation, elongating the plosive closures can also improve plosive recognition in speech. However, since vowels in speech are typically shorter and less intense than in operatic singing, they produce a lower reverberation field, which reduces

plosive masking. Consequently, an optimal balance likely exists: plosive recognition and thus text intelligibility in reverberant rooms may improve when vowels are sung more quietly. However, singers still need to maintain the carrying power of their voice, related to the SPL of vowels.

The poorer recognition of plosives in the highest pitch series (F5) compared to the G4 and G3 series may also be influenced by the voice characteristics of the two different singers performing the stimuli. The G3 samples were sung by a tenor, while the G4 and F5 samples were performed by a mezzo-soprano. In the G4 and, especially, the F5 series, the wider spacing between the spectral partials in the vocal tract formant transitions between the plosive and adjacent vowels may have rendered the formant transition cues less informative. Additionally, articulation in high-pitch regions may be constrained by the need to shape the vocal tract for formant tracking and to keep the larynx in a low position to establish the singers' formant, which restricts the tongue's forward movement (Sundberg, 1987).

We were also interested in how the duration of plosive closure interacts with other factors that affect the recognition of plosives. If a particular factor significantly reduces or increases the recognition of plosives, the impact of closure duration on recognition may be less pronounced. For instance, if recognition is nearly at chance level when plosive closure durations are long, it will not decrease much further with shorter closures. Conversely, if recognition is close to 100% with short closures, extending the closure duration may not lead to much improvement. Our results reflect this interaction. At the high pitch of F5, recognition was relatively low and less influenced by closure duration (particularly between 60 ms and 150 ms) compared to the baseline pitch of G3. This indicates a negative interaction between Duration and Pitch (see Figure 7 and Table 3). In contrast, when examining stimuli with strong versus weak plosive bursts, we observed a positive interaction between Duration and Burst Intensity. Stimuli with strong bursts exhibited high recognition, approaching 100% at 260 ms, with less reliance on closure duration than the baseline "weak burst" (see Figure 9).

With all its various possibilities, the acoustic condition can be a stronger factor than the plosive closure duration in determining plosive recognition. For instance, elongating the plosive closure may not significantly improve recognition in a highly reverberant hall or with excessively loud accompaniment. Nevertheless, plosive closure duration remains an important factor from a performance perspective, as it is one of the few factors singers can control when articulating plosives, alongside plosive burst intensity and the exact place of articulation. In an informal experiment, the first author (a professional singer with over 30 years of experience) sang vowel–plosive–vowel sequences with intentionally controlled closure durations. The experiment demonstrated that plosive closure durations can be easily controlled and recalled without prior rehearsal. The standard deviation of closure duration ranged from about 7 ms for shorter closures to about 45 ms for longer ones, providing sufficient precision for optimizing plosive closure durations in singing practice.

Factors such as the acoustics of the performance venue, accompaniment, type of plosive, and pitch are either prescribed by the composer or beyond the singer's control. Likewise, listener-specific factors such as hearing impairments, seating position in the hall, or language skills are also outside the singer's influence. Our results confirmed that the age of the listener can indeed serve as a predictor of text intelligibility for a particular individual. The difference in plosive recognition between our younger and older participant groups with a borderline in the mid-forties was about ten percentage points. This difference would likely have been more pronounced if we had compared groups with a more distinct age gap, such as the participants in their twenties versus those in their sixties or seventies. Furthermore, the interaction between plosive closure duration and age could have been more apparent if listeners with severe hearing decline had been included in the perception tests.

We may also speculate as to whether extending the plosive closure phase could disrupt the legato—a commonly desired quality in classical singing aesthetics (Miller, 1996). However, it is

uncertain whether plosive closure duration primarily affects the perceived legato given that plosives inherently contain a closure phase (Behrman, 2023). Moreover, in reverberant acoustics, it is nearly impossible to perceive how long the plosive occlusion lasts in sung vowel–consonant–vowel sequences by listening alone, as the reverberation tail of the vowel preceding the consonant extends the perceived duration of that vowel. Several voice teachers (Appleman, 1986; Miller, 1996) attribute poor legato in singing to the detrimental habit of some singers to start and end all notes by swelling and diminishing them, causing unevenness and disruption of vocal lines. The concept of "pure vowels" (sung vowels must remain unaltered throughout their whole duration) often used by singers is linked to the same topic. The use of "pure vowels" is also typical of spoken Italian, but not e.g., of English (Miller, 1977). We may argue that the vowel–consonant–vowel sequences used in our perception tests are perceived as sung with a proper legato if the dynamics of the second vowel begin at the same dynamic level with which the first vowel ends, thereby maintaining the continuity of the musical phrase. Therefore, legato may be less related to the duration of the plosive closure phase between two vowels. However, further empirical study would be needed to test our assumptions.

The stimuli with Brown Noise offer additional insights. Plosive recognition seems to decline (though insignificantly) or remains low with longer plosive closures (e.g., see the lines labeled BN, BN_BR, BN_Ch in Figure 2, and BN in Figure 5). In the presence of Brown Noise, both the noise and the reverberation tail of the pre-plosive vowel masked the plosive. However, while the phonation of the vowel (i.e., the singer's voice) stopped when the plosive closure began, the sound source of the Brown Noise (the accompaniment) continued throughout the entire stimulus. Consequently, as the reverberation field of the vowel began to decay during the plosive closure, the level of the reverberation field caused by the Brown Noise continued to build up until it reached a stationary state level. Therefore, masking of the plosives by Brown Noise might increase at longer plosive closure durations, while masking by the reverberation tail of the pre-plosive vowel might decrease.

This would explain why, in some cases, plosive recognition improved with longer plosive closures, whereas when the same stimulus was presented with concurrent Brown Noise, recognition did not improve, or even declined.

To interpret our results for stimuli with added Brown Noise, which was intended to simulate masking of plosives by accompanying instruments or ensemble partners, we should also consider that the nature and spectral profile of real musical accompaniments differs substantially from Brown Noise in its spectrum and temporal structure. While a real accompaniment typically consists of musical tones with spectral partials only at single harmonic frequencies, Brown Noise has a continuous spectrum. Additionally, the temporal structure of real music varies widely. Moreover, the Long-Term Average Spectrum slope of Brown Noise is less steep compared to that of symphony orchestras (-6 dB/octave versus -9 dB/octave). All these factors may affect the masking of the singers' voice differently for each note and speech sound of the sung text. Since it was impossible to replicate all these variations in our perception tests, using Brown Noise still seemed the best solution to obtain general conclusions. Therefore, we may summarize that sounds from the accompaniment may reduce the influence of plosive closure duration on plosive recognition, but it is difficult to predict the magnitude of such influence in each individual case. Moreover, in some cases there might be no masking from the accompaniment at all, e.g., when a rest in the accompaniment aligns with the plosive.

We should also consider whether elongating the plosive closures could make the prosody of the sung text sound unnatural. Spoken text phonetics and prosody perceived at close distance in a small room with dry acoustics can differ substantially from those of the same text sung operatically, loudly and at high pitches in a large reverberant concert hall, heard at a long distance from the singer. In such cases, it is sometimes impossible to understand even the language the singer is using. In this context, the improvement in intelligibility can be justified even at the expense of a decline in naturalness. Additionally, pronunciation that seems optimal in a concert hall may sound exaggerated

in a small rehearsal room, and singers may benefit from feedback from a trusted and competent listener in the hall.

While our findings present generalized conclusions, some limitations remain, which would have been difficult to overcome without heavily overloading our participants. For example, the stimuli in our perception tests were based on the recordings of random samples sung by only two singers. Using other singers or sung samples would result in slightly different parameters of the stimuli (exact location of plosive articulation in the vocal tract, plosive burst intensities, the frequencies of vowel formants). Similarly, the chosen acoustic conditions represent only a subset of possibilities. This also applies to the real concert hall in Experiment II, where, in addition to the seating position in the front or back rows, each listener experiences slightly different acoustic conditions. However, while the specific numerical results reported are somewhat conditional and reflect a single trial, we are confident that our findings accurately characterize common trends in the situations examined in our study.

Conclusions

The results of this study demonstrated that, when singing vowel–plosive–vowel sequences in typical concert hall acoustics with reverberation, extending plosive closure phases can (similarly to increasing the intensity of plosive bursts) significantly enhance the recognition of plosives. This improvement occurs because the masking of the plosives by the reverberation tail of the pre-plosive vowel becomes weaker. However, the positive effect of elongating the plosive closures diminishes when the reverberation is too long to decay substantially during the plosive closure.

Conversely, if the reverberation is very short, the reverberation tail fades significantly even during a short plosive closure, making intelligibility unaffected by the occlusion duration; the plosives with both short and long closures are sufficiently clear in this case.

There are also several other findings in our work that align with acoustical theories:

- Recognition of plosives tends to decrease at higher pitches of sung vowel–consonant–vowel sequences.
- The presence of accompaniment increases the difficulty of recognizing plosives and, in certain cases, can negate the improvement gained by elongating the plosive closure durations.
- Stronger plosive bursts improve their recognition.
- The recognition of plosives deteriorates with age.
- The recognition of plosives deteriorates with increased distance from the singer in a reverberant hall.
- Interactions between the plosive closure duration and vowel pitch (and probably also between the plosive closure duration and the plosive burst intensity) are possible.

Further research is needed to explore whether singers actually use the lengthening of plosive closures in order to improve the intelligibility of sung text in reverberant acoustics, and whether the lengthening of the plosive closures worsens the legato and/or the natural prosody of the text in the context of a specific language.

References

ADAMS, E. M., GORDON-HICKEY, S., MORLAS, H., & MOORE, R. (2012). Effect of rate-alteration on speech perception in noise in older adults with normal hearing and hearing impairment. *American Journal of Audiology, 21*(1), 22–32. https://doi.org/10.1044/1059-0889(2011/10-0023

APPELMAN, R. (1986). *The science of vocal pedagogy.* Indiana University Press.

BEHRMAN, A. (2023). *Speech and voice science* (4th ed.). Plural Publishing.

BEST, V., MASON, C. R., SWAMINATHAN, J., ROVERUD, E., & KIDD, G., JR (2015, 18-22 May). Does providing more processing time improve speech intelligibility in hearing-impaired listeners? *169th Meeting of the Acoustical Society of America. Psychological and Physiological Acoustics:* Paper 1aPPb28

BELISLE, J. M. (1967). Some factors influencing diction in singing. *NATS Bulletin, 24*(2), 4–8.

BOERSMA, P. & WEENINK, D. (2024, July 12. University of Amsterdam.). *Praat: doing phonetics by computer.* http://www.praat.org/.

BROWN, R. M. (1946). *The singing voice.* The Macmillan Company.

BUNCH, M., & CHAPMAN, J. (2000). Taxonomy of singers used as subjects in scientific research. *Journal of Voice, 14*(3), 363–369. https://doi.org/10.1016/s0892-1997(00)80081-8

CHRISTY, V. A. (1967). *Expressive singing. Basic principles. A text for school or studio class or private teaching* (Vol.1). W. C. Brown Publishing Company.

COLLISTER, L. B., & HURON, D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review, 3*(3), 109–125. https://doi.org/10.18061/1811/34102

CORRETGE, R. (2024, July 12. University of Amsterdam.) *Praat vocal toolkit.* https://www.praatvocaltoolkit.com/index.html

DI CARLO, N. S. (2007). Language and diction: Effect of multifactorial constraints on intelligibility of opera singing (II). *Journal of Singing 63*(5), 559-567.

EBERHART, C. (1962). Diction. *NATS Bulletin, 18*(4), 8–9.

FRIBERG, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal, 15*(2), 56-71. https://doi.org/10.2307/3680917

FUCHS, V. (1964). *The art of singing and voice technique.* London House and Maxwell.

GORDON-SALANT, S., & FITZGIBBONS, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, *40*(2), 423–431. https://doi.org/10.1044/jslhr.4002.423

GREGG, J. W. (1991). On articulation – Part I. *Journal of Singing, 47*(5), 30-32.

GYGI, B., & SHAFIRO, V. (2014, April). Spatial and temporal modifications of multitalker speech can improve speech perception in older adults. *Hearing Research*, *310*, 76–86. https://doi.org/10.1016/j.heares.2014.01.009

HOWARD, D. M., & ANGUS, J. A. S. (2017). *Acoustics and psychoacoustics* (5th ed.). Routledge.

KOUTSOGIANNAKI, M. C. (2016). *Intelligibility enhancement of casual speech based on clear speech properties.* University of Crete. https://www.csd.uoc.gr/~mkoutsog/docs/PhD_Koutsogiannaki.pdf

LAMARCHE, A., TERNSTRÖM, S., & PABON, P. (2010). The singer's voice range profile: female professional opera soloists. *Journal of Voice*, *24*(4), 410–426. https://doi.org/10.1016/j.jvoice.2008.12.008

LESSA, A. H., & COSTA, M. J. (2013). The impact of speech rate on sentence recognition by elderly individuals. *Brazilian Journal of Otorhinolaryngology*, *79*(6), 745–752. https://doi.org/10.5935/1808-8694.20130136

LINDBLOM, B., & SUNDBERG, J. (2007). The human voice in speech and singing. In T. D. Rossing (Ed.), *Springer handbook of acoustics* (pp. 669–712). Springer. https://doi.org/10.1007/978-1-4939-0755-7

MARSHALL, M. (1956). Is exaggeration required for good English diction? *Diapason, 47*(3), 18.

MELTON, J. (1953, June 1). Do you put the words across? An interview by Annabel Comfort. *Etude*, June, 15. https://digitalcommons.gardner-webb.edu/etude/117/

MEYER, J. (2009). *Acoustics and the performance of music. Manual for acousticians, audio engineers, musicians, architects and musical instrument makers* (5th ed.). Springer.

MILLER, R. (1977). *English, French, German and Italian techniques of singing: A study in national tonal preferences and how they relate to functional efficiency.* The Scarecrow Press.

MILLER, R. (1986). *Structure of singing: System and art in vocal technique.* Shirmer.

MILLER, R. (1996). *On the art of singing*. Oxford University Press.

NAIR, A. (2016, June 1-5). *Why consonants matter – consonant resonance and its effect on vowels*. 45th Annual Symposium: Care of the Professional Voice.

NELSON, H. D., &TIFFANY, W. R. (1968). The intelligibility of song: Research results with a new intelligibility test. *NATS Bulletin, 25*(2), 22–33.

PETKOV, P. N., BRAUNSCHWEILER, N., & STYLIANOU, Y. (2016). Automated pause insertion for improved intelligibility under reverberation. *Proc. Interspeech 2016,* 145-149. https://doi.org/10.21437/Interspeech.2016-960

PHILLIPS, G. L. (2002). Diction: A rhapsody. *Journal of Singing, 58*(5), 405–409.

SCHARPFF, P., & VAN HEUVEN, V. (1988). Effects of pause insertion on the intelligibility of low quality speech. *Proceedings of* 7th *FASE Symposium*, 261–268. https://openaccess.leidenuniv.nl/bitstream/1887/2591/1/167_106.pdf

SCHMITT, J. F., & MCCROSKEY, R. L. (1981). Sentence comprehension in elderly listeners: The factor of rate. *Journal of Gerontology*, *36*(4), 441–445. https://doi.org/10.1093/geronj/36.4.441

SHARNOVA, S. (1947). Diction. *NATS Bulletin, 3*(6), 4.

SHROUT, P. E., & FLEISS, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

SUNDBERG, J. (1987). *The science of singing voice.* Northern Illinois University Press.

TITZE, I. (1982). Why is the verbal message less intelligible in singing than in speech. *NATS Bulletin, 38*(3), 37.

VENNARD, W. (1967). *Singing: The mechanism and the technic.* Carl Fisher.

VURMA, A., MEISTER, E., MEISTER, L., ROSS, J., RAJU, M., KALA, V., & DEDE, T. (2023). The intensities of vowels and plosive bursts and their impact on text intelligibility in singing. *The Journal of the Acoustical Society of America*, *154*(4), 2653–2664. https://doi.org/10.1121/10.0021968

WARE C. (1998). *Basics of vocal pedagogy: The foundations and process of singing*. McGraw-Hill.

Author Note

This study was supported by the Estonian Research Council grant PRG1552 and approved by the Research Ethics Committee of the University of Tartu. The authors have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Professor Allan Vurma, Estonian Academy of Music and Theater, Tatari 13, Tallinn 10116, Estonia. E-mail: allan.vurma@eamt.ee

TABLE 1. *Summary of the Generalized Linear Model coefficients in Experiment I—plosive closure duration, acoustic condition, and their interaction are used as independent variables.*

Predictor	Estimate (β)	Std. Error	*z*-value	*p*	
(Intercept)	3.01	0.16	19.29	< .001	***
Duration150	0.08	0.22	0.34	.74	
Duration260	-0.12	0.23	-0.54	.59	
Cond_BN	-1.37	0.18	-7.62	< .001	***
Cond_BR	-2.90	0.17	-17.12	< .001	***
Cond_Ch	-3.00	0.17	-17.69	< .001	***
Cond_BN_BR	-3.39	0.17	-19.93	< .001	***
Cond_BN_Ch	-3.52	0.17	-20.63	< .001	***
Duration260:Cond_BR	1.28	0.24	5.38	< .001	***
Duration260:Cond_Ch	0.56	0.24	2.40	.02	*

Note: *$p < .05$, **$p < .01$, ***$p < .001$

TABLE 2. *Summary of the Generalized Linear Model coefficients for Experiment II, with plosive closure duration, acoustic condition, and their interaction as independent variables.*

Predictor	Estimate (β)	Std. Error	*z*-value	*p*	
(Intercept)	0.49	0.05	10.00	< .001	***
Duration150	0.25	0.07	3.53	< .001	***
Duration260	0.50	0.07	6.82	< .001	***
Cond_BN	-0.76	0.07	-10.99	< .001	***
Duration150:BN	-0.24	0.10	-2.45	.01	*
Duration260:BN	-0.51	0.10	-5.07	< .001	***

Note: *$p < .05$, **$p < .01$, ***$p < .001$

TABLE 3. *Summary of the Generalized Linear Model coefficients for pooled Experiments I and II—plosive closure duration, pitch, and their interaction are used as independent variables.*

Predictor	Estimate (β)	Std. Error	*z*-value	*p*	
(Intercept)	0.75	0.07	10.44	< .001	***
Duration150	0.76	0.11	6.76	< .001	***
Duration260	1.46	0.13	10.97	< .001	***
PitchG4	-0.20	0.10	-2.04	.04	*
PitchF5	-0.94	0.10	-9.52	< .001	***
Duration150:Pitch_G4	-0.46	0.15	-3.03	.002	**
Duration260:Pitch_G4	-0.89	0.17	-5.27	< .001	***
Duration150:Pitch_F5	-0.63	0.15	-4.25	< .001	***
Duration260:Pitch_F5	-0.92	0.16	-5.63	< .001	***

Note: *$p < .05$, **$p < .01$, ***$p < .001$

Figure Captions

FIGURE 1. Mosaic plots showing the distribution of k, p, t, and "?" responses in Experiment I to the stimuli with k, p, and t (in columns).

FIGURE 2. The probability of correct responses predicted by the Generalized Linear Model as a function of the duration of the plosive closures by artificially added acoustic conditions (BN—Brown Noise, BR—Big Room, Ch—Church). Three pitch series are pooled. The line "Clear" is partly hypothetical; for the G3 and G4 series, it is based on the results of the pilot test with three participants.

FIGURE 3. The Great Hall of the Estonian Academy of Music and Theater with two groups of participants seated in rows 4 and 5, and 12 and 13.

FIGURE 4. Mosaic plots showing the distribution of k, p, t, and "?" responses (in Experiment II) to the stimuli with k, p, and t (in columns).

FIGURE 5. The probability of correct responses predicted by the Generalized Linear Model, as a function of plosive closure duration in the concert hall acoustics (Clear—stimuli where reverberation resulted from the hall's natural acoustics; BN—stimuli accompanied by Brown Noise played from the sound system above the stage). Results from all three pitch series are pooled.

FIGURE 6. The probability of correct responses predicted by the Generalized Linear Model as a function of plosive closure duration in the concert hall acoustics (Front—for listeners seated in the front rows; Back—for listeners seated in the back rows. Results from all three pitch series are pooled, and only stimuli without accompanying Brown Noise are included.

FIGURE 7. The probability of correct responses predicted by the Generalized Linear Model as a function of plosive closure duration, split by three pitch series. The data from Experiments I and II are pooled.

FIGURE 8. The probability of correct responses predicted by the Generalized Linear Model as a function of plosive closure duration split by three plosives. The data from Experiments I and II are pooled.

FIGURE 9. The probability of correct responses predicted by the Generalized Linear Model as a function of plosive closure duration split by burst intensities. The data in pitch series G3 and G4 from Experiments I and II are pooled.

Figure 1

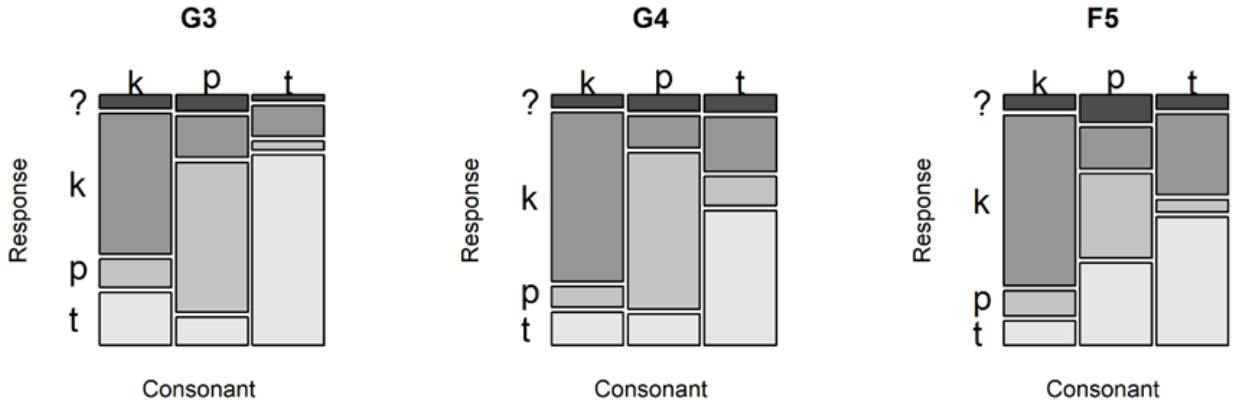


Figure 2

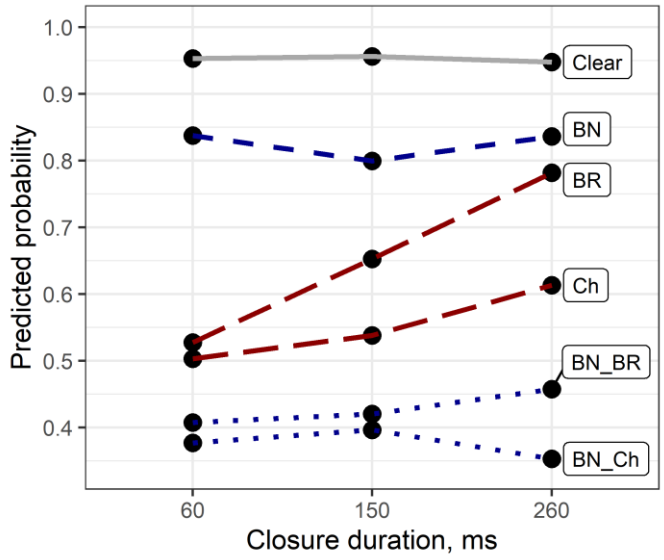

1.0
0.9
0.8
0.7
0.6
0.5
0.4
Predicted probability
60
150
260
Closure duration, ms
Clear
BN
BR
Ch
BN_BR
BN_Ch

Figure 3

Figure 4

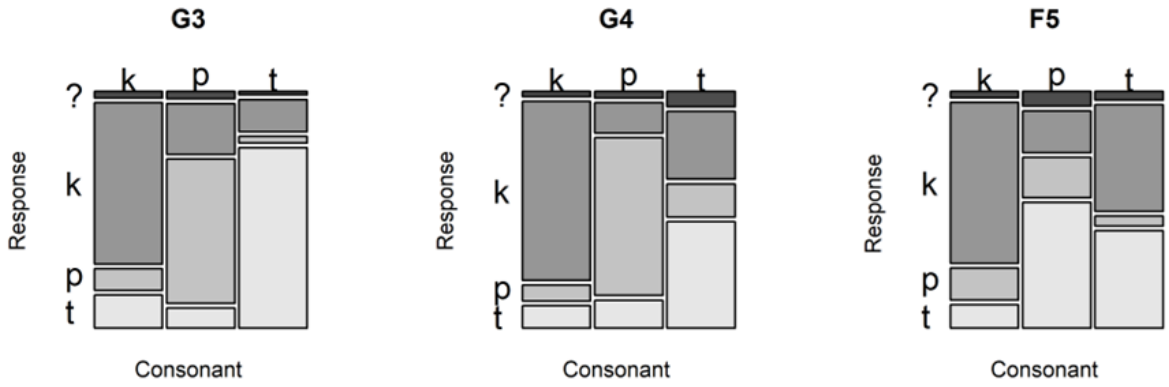


Figure 5

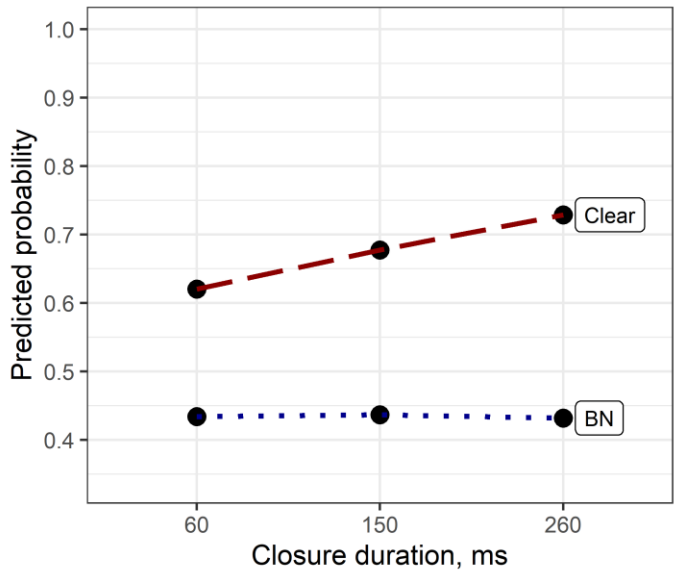

1.0
0.9
0.8
0.7
0.6
0.5
0.4
Predicted probability
Clear
BN
60
150
260
Closure duration, ms

Figure 6

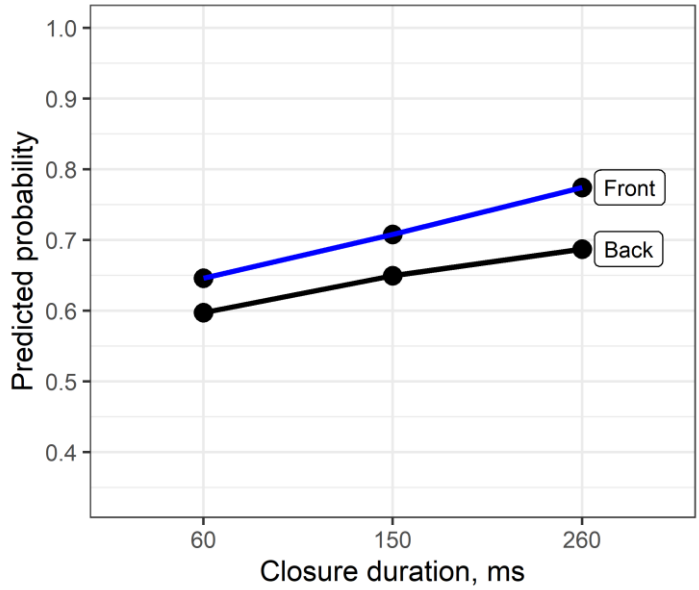

1.0
0.9
0.8
0.7
0.6
0.5
0.4
Predicted probability
Front
Back
60
150
260
Closure duration, ms

Figure 7

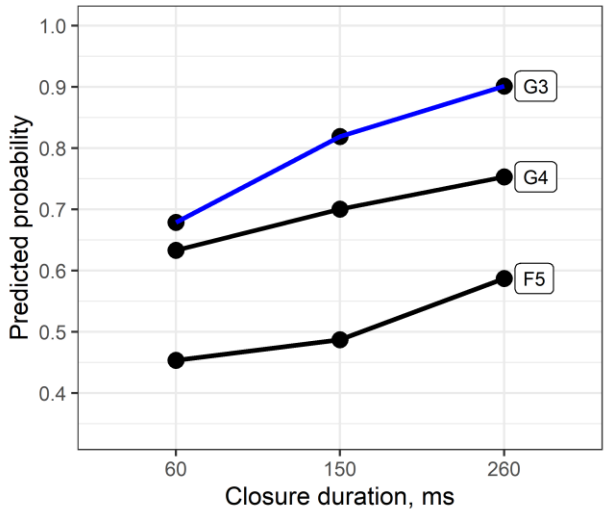

1.0
0.9
0.8
0.7
0.6
0.5
0.4
Predicted probability
G3
G4
F5
60
150
260
Closure duration, ms

Figure 8

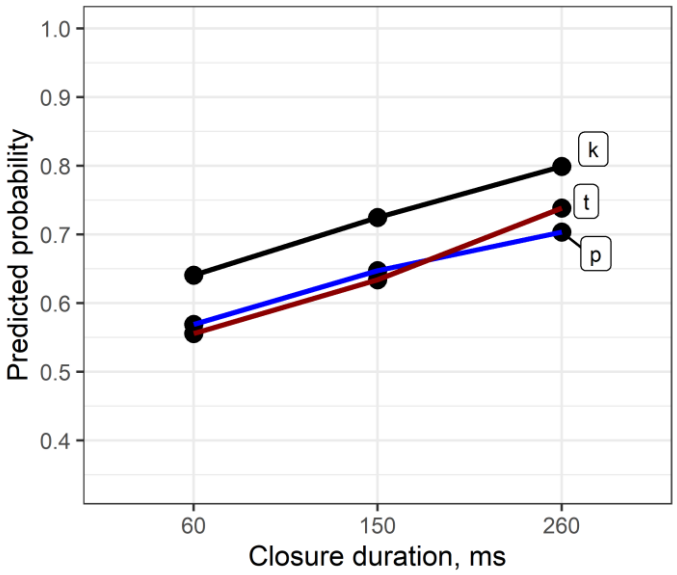
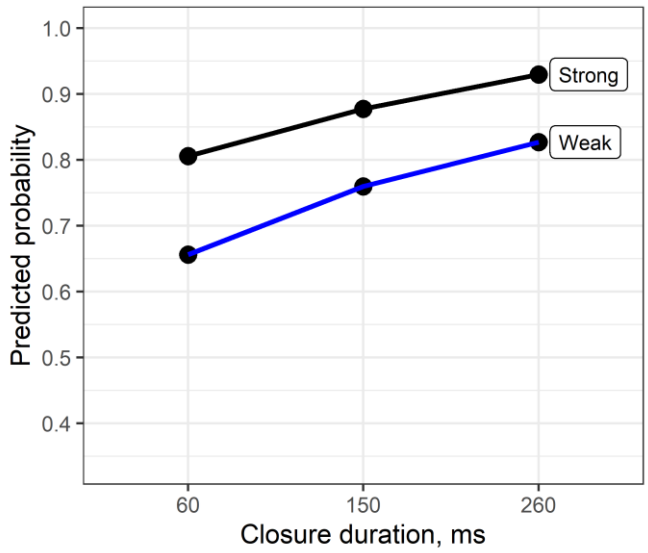

1.0
0.9
0.8
0.7
0.6
0.5
0.4
Predicted probability
k
t
p
60
150
260
Closure duration, ms

Figure 9


1.0
0.9
0.8
0.7
0.6
0.5
0.4
Predicted probability
Strong
Weak
60
150
260
Closure duration, ms