5th International Symposium on Applied Phonetics (ISAPh 2024)
30 September - 2 October 2024, Tartu, Estonia

Coping with reverberant acoustics in singing by extending the plosive closures in vowel-plosive-vowel sequences

Allan Vurma[1], *Einar Meister*[2], *Lya Meister*[2], *Jaan Ross*[1], *Marju Raju*[1], *Veeda Kala*[1], *Tuuri Dede*[1]

[1]Estonian Academy of Music and Theatre, Estonia
[2]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, tuuri.dede@eamt.ee

Abstract

This study examines how the duration of voiceless plosive closures in vowel–plosive–vowel (VCV) sequences affects intelligibility of sung text, particularly in reverberant environments. We hypothesize that extended closure durations reduce the masking effect of reverberation, thus enhancing plosive recognition. Perceptual experiments were conducted with modified closure durations across various acoustic settings, using stimuli created from professional opera singers' performances. Results indicate that longer closure durations improve plosive recognition in typical concert hall acoustics with reverberation. The findings suggest that singers might improve text intelligibility by lengthening plosive closures.

Index Terms: singing, text intelligibility, plosive closures, room acoustics, masking, reverberation

1. Introduction

In classical singing, particularly in large halls with long reverberation times and with female voices or high fundamental frequencies (f_0), text intelligibility often declines [1][2][3][4]. This issue is the subject of ongoing debate among singers and teachers. Some advocate for stronger consonant articulation [5][6][7][8][9], while others oppose this approach [10][11][12]. Additionally, factors such as room acoustics and accompanying sounds can mask the singer's voice, further complicating intelligibility [13].

Sung text is typically less intelligible than spoken text. In a study [14] intelligibility in sung phrases dropped by 76% compared to spoken phrases. While singing and speaking use the same vocal apparatus, singing imposes constraints such as fixed pitches, durations, and dynamics, limiting flexibility needed for clear articulation. Poor intelligibility of sung text may be influenced by both vowels and consonants [15].

The intelligibility of sung text is influenced by neurological processes involving both bottom-up acoustic information and top-down cognitive processing. Top-down processes rely on factors such as language proficiency and contextual cues, which help listeners predict and interpret text. Limited language skills or a foreign accent can significantly hinder sung text intelligibility.

This study focuses on plosive consonants, examining the impact of plosive closure phase duration on their recognition in vowel–voiceless plosive–vowel (VCV) sequences sung in reverberant environments. The articulation of plosives begins with the closure phase (bilabial for /p/, alveolar for /t/, and velar for /k/) followed by the burst phase. Perceptual cues for identifying plosives include the frequency distribution of the burst's noise spectrum, voice onset time (VOT), and vowel formant transitions.

In loud operatic singing, reverberation from preceding vowel can mask voiceless plosives, reducing their recognition. Research suggests that stronger plosive bursts improve recognition, especially in reverberant or noisy conditions [16]. Lengthening the closure phase is another technique available to singers that could influence plosive recognition. However, empirical evidence on how extended plosive closure duration affects the intelligibility of sung text remains limited. Indirectly, studies on speech intelligibility provide some insights. For instance, inserting pauses between words or at prosodic boundaries in synthetic speech has been shown to improve intelligibility in reverberant environments [17][18]. Similarly, slowing down speech has been found to enhance clarity, particularly for older listeners [19][20].

In vocal pedagogy, opinions vary regarding consonant duration in singing. Some advocate for extending consonants to improve clarity, while others caution against unnecessary lengthening or intensification, except when required for artistic expression.

This study investigates how the duration of voiceless plosive closures in VCV sequences affects the intelligibility of sung text, particularly in reverberant environments. We hypothesize that extending the closure duration of voiceless plosives sung in reverberant acoustics, can improve plosive recognition. In these conditions, a longer plosive closure phase provides more time for the reverberation tail from the preceding vowel to decay before the onset of the plosive burst. Consequently, a longer closure phase results in weaker masking of the plosive burst and formant transitions by the reverberation tail. To test this hypothesis, perceptual experiments were conducted with variable plosive closure durations across different acoustic conditions.

10.21437/ISAPh.2024-22

2. Materials and method

2.1. Stimuli

This study comprises two perceptual experiments using VCV stimuli with modified closure durations in varied acoustic conditions. To determine appropriate closure durations for the stimuli, recordings of Italian-language Classical or Romantic opera arias were analyzed, performed by 11 professional classically trained singers. Additionally, the singers read the text of the arias aloud in two styles: (1) conversational speech and (2) an oratorical style. Recordings were conducted in a studio with minimal reverberation (T30 = 0.2 s) to minimize masking effects from room acoustics.

The recordings were segmented, and closure durations were measured using Praat [21]. The longest closure durations occurred during oratorical reading (mean = 93.8 ms, SD = 48.3 ms), while sung performances had the shortest durations (mean = 74.7 ms, SD = 48.2 ms), with conversational speech falling in between (mean = 84.7 ms, SD = 38.6 ms). In all styles, particularly singing, outliers extended up to 300 ms. Based on these findings, three closure durations were selected for the VCV stimuli:

- 60 ms, slightly below the median of sung performances,
- 150 ms, as an intermediate value,
- 260 ms, reflecting the outliers with extended closures.

Stimuli were derived from recordings of /a-k-a/, /a-p-a/, and /a-t-a/ sequences sung by two professional opera singers: a mezzo-soprano and a tenor. The mezzo-soprano produced two series: one at G4 (f_0 = 392 Hz) and another at F5 (f_0 = 698.5 Hz), while the tenor performed one series at G3 (f_0 = 196 Hz). The G3 and G4 series included two versions of plosive burst intensity: a “weak burst,” representing the singer's initial spontaneous production, and a “strong burst,” with the singers slightly increasing the burst intensity. The sound pressure level (SPL) difference between weak and strong bursts was 7 dB for the tenor's G3 series (44 dB vs. 51 dB) and 2 dB for the mezzo-soprano’s G4 series (45 dB vs. 47 dB). For the mezzo-soprano's F5 series, only one burst intensity (44 dB) was included due to the difficulty of varying burst intensity significantly at such a high pitch. The SPL of adjacent vowels was about 67 dB in G3 and G4 series and approximately 72 dB in the F5 series.

Closure durations were modified in Praat, while burst durations and vowel transitions remained unchanged to preserve naturalness. The correlation between closure and burst durations was very weak (r = 0.13), justifying the focus on manipulating closure duration alone. Vowel durations were standardized at 600 ms for V1 and 900 ms for V2.

Using the Praat Vocal Toolkit [22], stimuli sets were created with simulated reverberation conditions (Church (Ch) and Big Room (BR)) and with or without Brown Noise (BN) to simulate the masking effect of orchestral accompaniment. The final stimulus set consisted of three test series:

Series I: based on tenor recordings at pitch G3,

Series II: based on mezzo-soprano recordings at pitch G4,

Series III: based on mezzo-soprano recordings at pitch F5.

2.2. Experimental setup

2.2.1. Experiment I

The first perception tests were conducted in a soundproof booth, utilizing the Praat Listening Experiment platform. The experimental setup included a laptop, an external audio card, and Sennheiser HD 560s headphones. Test series I and II consisted of 90 stimuli each: 3 plosives (/k/, /p/, /t/) × 3 closure durations (60 ms, 150 ms, 260 ms) × 2 burst intensities (weak, strong) × 5 acoustic conditions (BR, Ch, BN, BN_BR, BN_Ch)). Series III excluded the burst intensity contrast but added a Clear (Cl) acoustic condition, resulting in 54 stimuli: 3 plosives (/k/, /p/, /t/) × 3 closure durations (60 ms, 150 ms, 260 ms) × 6 acoustic conditions (Cl, BR, Ch, BN, BN_BR, BN_Ch).

A total of 34 participants (11 males, 23 females, aged 21–68 years) completed the tests. Participants identified the plosive between the two vowels by clicking one of four response buttons (k, p, t, or ? for uncertainty) displayed on a computer screen. Stimuli were presented in a randomized order unique to each participant, with each stimulus repeated three times per test series. Each series lasted approximately 15–20 minutes.

2.2.2. Experiment II

The second experiment took place in a concert hall with natural reverberation (dimensions: width 15 m, length 30 m, height 13.8 m; reverberation time T60 = 1.8 seconds at 1000 Hz). G3 and G4 series included 36 stimuli each, varying: 3 plosives (/k/, /p/, /t/) × 3 closure durations (60 ms, 150 ms, 260 ms) × 2 burst intensities (weak, strong) × 2 acoustic conditions (Cl, BN). F5 series included 18 stimuli: 3 plosives (/k/, /p/, /t/) × 3 closure durations (60 ms, 150 ms, 260 ms) × 2 acoustic conditions (Cl, BN). The Clear stimuli were played through a Genelec 8341a loudspeaker placed at the center of the stage, while Brown Noise was emitted from a separate speaker above the stage.

A total of 33 participants (15 male, 17 female, one non-binary; aged 22–66 years) took part in the tests. Listeners 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. To account for seating location effects, each test series was conducted twice, with participants switching seats between the front and back rows. One seat was left empty between adjacent participants. Responses (k, p, t, ?) were recorded on response sheets.

The Clear stimuli in Experiment II were most comparable to the Big Room stimuli in Experiment I due to similar reverberation times. Similarly, the Brown Noise stimuli in Experiment II closely resembled the Brown Noise+Big Room stimuli from Experiment I.

2.2.3. Data analysis

The results from both experiments were analyzed using a Generalized Linear Mixed Effects Model (GLMM) for each pitch series. The dependent variable was the probability of correct responses (*Correct*). Fixed effects included *Duration*, *Acoustic condition*, *Consonant*, and *Burst*, while random effects included intercepts for *Age group* and *Gender*.

Analyses were performed using the *lme4* package [23] in R [24], with *post hoc* tests for fixed effects conducted using the *emmeans* package [25].

3. Results

3.1. Experiment I

Figure 1 illustrates the overall distribution of responses. Only 6% of all responses were categorized as "?", suggesting that participants were generally successful in identifying the plosives. The plosives /k/ and /t/ were the most frequently mutually confused. Recognition rates in the G3 and G4 series ranged from 57.3% to 80.8%. The F5 series had the lowest recognition rates, with /p/ identified at 35.8%, /t/ at 54.5%, and /k/ at 73%.

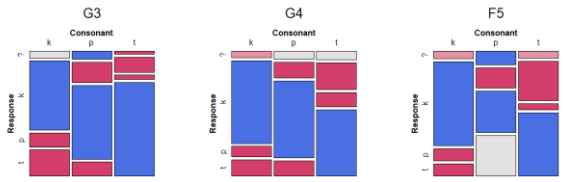

Figure 1. Mosaic plots illustrating the distribution of responses /k/, /p/, /t/, and "?"in Experiment I to the stimuli with /k/, /p/, and /t/.

Inter-subject reliability was assessed using the intraclass correlation coefficient (ICC), specifically the ICC (2, k) model, which employs a two-way random effects approach, the mean of k raters (k = 34), and absolute agreement, as outlined by Shrout and Fleiss. The level of agreement across all three sets was excellent, with an ICC of 0.97 (95% CI: 0.97–0.98; F[98, 1057] = 46.1; $p < 0.001$).

3.1.1. Statistical analysis

Separate Generalized Linear Mixed Models (GLMMs) were applied to analyze the proportion of correct responses across the G3, G4, and F5 test series. Fixed effects included closure duration (60, 150, 260 ms), acoustic condition (e.g., BN, BR, Ch, and combinations), consonant (/p/, /t/, /k/), and burst intensity (strong/weak, excluded in F5). Age group and gender were treated as random effects. See summaries of the GLMM models in the Appendix.

Significant main effects were observed for acoustic condition ($p < 0.001$), consonant ($p < 0.001$), and closure duration, which varied by acoustic condition. In the G3 and G4 series, burst intensity also had a significant impact ($p < 0.001$). Across all series, correct responses improved with longer closure duration in stimuli with BR or Ch reverberation, but recognition worsened as pitch increased ($\chi^2 = 559.6$, $p = 0.002$).

The largest improvement with longer closure duration occurred in BR acoustics. For BN conditions, recognition slightly worsened at 150 ms compared to 60 ms but improved at 260 ms, with significant changes observed in G3 and G4 ($p < 0.05$). Stimuli with BN_BR showed modest improvements (10 percentage points, $p < 0.001$) at 260 ms in G3, while BN_Ch conditions showed declines of about 10 percentage points at 150 and 260 ms in G3 and G4. No significant changes were observed in the F5 series for these conditions.).

Plosive type significantly affected recognition accuracy ($p < 0.001$), though the order of recognition varied across test series: /t/, /p/, /k/ in G3; /k/, /p/, /t/ in G4; and /k/, /t/, /p/ in F5. In BR acoustics, recognition improvements ranged from 9 percentage points (for /p/ with a strong burst in G4) to 43 points (for /k/ with a weak burst in G3). In Ch acoustics, improvements ranged from 6 points (/t/ with a strong burst in G4) to 16 points (/p/ with a weak burst in G4). These improvements, particularly between 60 ms and 260 ms, were statistically significant ($p < 0.001$; in F5 with Ch acoustics, $p < 0.01$). Differences were often significant between 60 ms and 150 ms and between 150 ms and 260 ms.

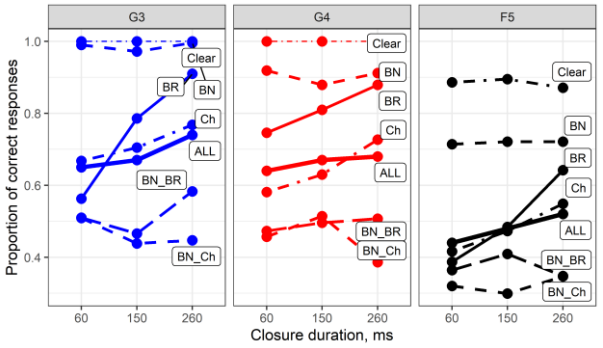

Figure 2. Predicted probability of correct responses from the GLMM as a function of plosive closure duration in the G3, G4, and F5 test series, grouped by artificially added acoustic conditions (BN: Brown Noise, BR: Big Room, Ch: Church).

Stronger bursts consistently improved recognition, increasing accuracy by up to 20 percentage points (e.g., BN_BR in G4) and averaging an 11-point improvement. This effect was highly significant in the G3 and G4 series ($\chi^2 = 253.1$, $p < 0.001$) but diminished with longer plosive closure durations.

Age group differences showed younger participants (<44 years, 17 individuals) outperforming older participants (>44 years, 17 individuals) by 7 percentage points for females and 10 points for males across all conditions. In BR acoustics, the gap widened to 14 points. This difference was statistically significant ($\chi^2 = 253.1$, $p < 0.001$), although both groups showed similar improvements with longer closures. No significant gender differences were found ($p > 0.05$).

3.2. Experiment II

The overall response distribution is shown in Figure 3. Recognition was highest for /t/ in the G3 series (81%) and lowest for /p/ in the F5 series (18%), where /p/ was most frequently confused with /t/ (57% of cases). The "?" response was used in only 4% of all cases.

Inter-subject reliability was excellent in all three sets ICC = 0.99 (95% CI = 0.9–0.98; F[36, 1184] = 93.7; $p < 0.001$).

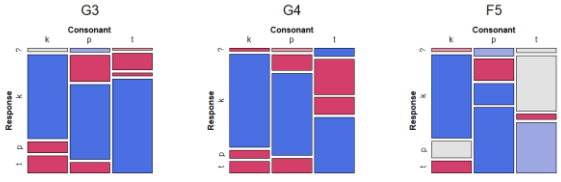

Figure 3. Mosaic plots illustrating the distribution of responses /k/, /p/, /t/, and "?"in Experiment II to the stimuli with /k/, /p/, and /t/.

3.2.1. *Statistical analysis*

In Experiment II, GLMMs for the G3 and G4 series used proportion of correct responses (*Correct*) as the dependent variable, with main effects for *Duration* (60, 150, 260 ms), *Acoustic condition* (Clear, BN), *Duration* × *Acoustic condition* interaction, *Location* (front, back), *Consonant* (/p/, /t/, /k/), and *Burst* (strong, weak). For the F5 series, *Burst* was omitted. *Age group* and *Gender* were included as random effects. See summaries of the GLMM models in the Appendix.

The recognition of plosives in a real concert hall acoustics resembled results from Experiment I, where longer plosive closure durations improved recognition, especially for stimuli without BN (Figure 4). However, the improvement in Experiment II was smaller, ranging from 3 to 16 percentage points, depending on consonant, burst strength, and seating location. All differences in response probabilities for Clear stimuli in the real concert hall were statistically significant ($p < 0.05$), except between 60 ms and 150 ms in the F5 series.

Adding BN worsened recognition by 10 to 45 percentage points, with no improvement at longer plosive closure durations, and recognition even decreased slightly, though not significantly. As in Experiment I, plosive type was a significant factor ($p < 0.001$), with recognition order: /t/, /p/, /k/ in G3, /k/, /p/, /t/ in G4, and /k/, /t/, /p/ in F5.

Stronger bursts consistently improved recognition in the G3 and G4 series. For Clear stimuli, the average recognition was 76% with weak bursts and 91% with strong bursts ($p < 0.001$). Burst intensity had a slightly greater effect in BN-acoustics, but recognition no longer depended on closure duration.

Age analysis showed that younger participants (20–39 years) recognized plosives 5 percentage points better than older participants (40–66 years), the difference was statistically significant ($\chi^2 = 29.9$, $p < 0.001$).

Recognition was influenced by plosive closure duration for both front and back row seating in the G3 and G4 series, with a greater benefit for back row listeners, particularly with weak bursts. For example, in the G3 series, recognition improved by 15 percentage points for back row listeners (11.3 percentage points for front row) with a 60 ms to 260 ms closure extension ($p = 0.001$). The overall recognition rate was 80% in the front rows and 74% in the back rows, with this difference being statistically significant ($\chi^2 = 31.7$, $p < 0.001$).

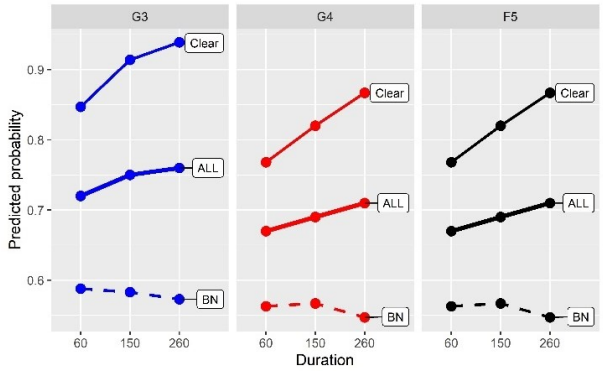

Figure 4. Predicted probability of correct responses from the GLMM as a function of plosive closure duration in the G3, G4, and F5 test series.

4. Discussion

Our study found that in reverberant acoustics, the recognition of voiceless plosives in sung vowel–plosive–vowel sequences improves when the plosive closure phase is extended. This enhancement occurs because a longer closure reduces the masking effect of the reverberation tail from the adjacent vowel. The benefit was more noticeable in environments with shorter reverberation times (Great Hall, BR) compared to longer reverberation (Church), and absent in rooms with very short reverberation (recording studio). These results align with our hypothesis: with longer reverberation, the vowel's reverberation tail decays less during the plosive closure resulting in a stronger masking of the plosive.

Plosive recognition also decreased with greater listener distance from the singer, with recognition being poorer in back rows than in the front. However, the benefit of extending the closure phase was slightly greater for those seated in the back rows, likely due to the dominance of reverberant sound in the back compared to direct sound in the front. In the F5 series, this front-back difference was negligible, possibly due to high-pitched singing affecting articulation.

Other factors, such as the intensity and duration of pre-plosive vowels, might also influence plosive recognition. If the pre-plosive vowel has low intensity, its reverberation tail decays faster below the audibility threshold compared to a strong vowel. Similarly, a shorter vowel does not allow the reverberation field to build up to a steady state, remaining weaker. Therefore, in both cases – when the pre-plosive vowel is weak or short – the ability of the vowel's reverberation tail to mask the plosive is reduced compared to a loud and long vowel. Consequently, an optimal balance likely exists: plosives 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.

Although plosive closure duration had less impact on recognition than some other factors (like acoustic condition), it remains important as it is one of the few factors singers can control. Other factors such as venue acoustics, plosive type, and pitch are determined by the composer or are beyond the singer's control.

To interpret our results for stimuli with added BN to imitate masking of plosives by accompanying instruments or ensemble partners, we should also consider that the nature and spectral profile of real music accompaniments differs substantially from BN. While a real accompaniment typically consists of musical tones with spectral partials only at their harmonic frequencies, the spectral distribution of BN is continuous and has a less steep Long-Term Average Spectrum slope compared to that of symphony orchestras (-6 dB/octave *vs.* -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. Therefore, we can summarize that sounds from the accompaniment may influence the impact of plosive closure duration on the recognition of plosives, but it is difficult to predict the magnitude of such influence in each individual case.

Lastly, we may consider whether elongating the plosive closures could make the prosody of sung text sound unnatural. The perception of spoken text phonetics and prosody from a short distance in a small room with dry acoustics can substantially differ from that of the same text sung operatically,

loudly and at high pitches in a big reverberant concert hall, heard from a long distance from the singer. In such cases, it is sometimes impossible to understand even the language the singer uses. Here, the improvement in intelligibility can be justified even at the expense of a decline in naturalness. Similarly, pronunciation that seems optimal in a concert hall may sound exaggerated in a small rehearsal room. Singers, in such situations, would benefit from receiving feedback from a trusted and competent listener in the hall.

Although our work has reported several generalized results, it still has some limitations, 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. If we had used different singers and chosen different samples from those they had sung as the basis of our stimuli in the perception tests, the parameters of the stimuli (exact location of plosive articulation in the vocal tract, plosive burst intensities, the frequencies of vowel formants) would have been somewhat different. Similarly, the selection of acoustic conditions applied was merely one of endless possibilities. This concerns also the real concert hall in Experiment II, where, besides the seating position in the front or back rows, the acoustic situation was somewhat different for each listener. However, while the specific numerical results reported are to a certain extent conditional on and reflect one particular trial, we believe they still accurately characterize common trends in the situations addressed in our study.

5. Conclusions

The results of this study showed that when singing vowel–plosive–vowel sequences in typical concert hall acoustics with reverberation, using longer plosive closure phases (similarly to increasing the intensity of plosive bursts) may significantly improve 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 influence of elongating the plosive closures is smaller when the reverberation is too long. 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; both short and long closures are sufficiently clear in this case. Additionally, recognition is better for younger listeners and those seated closer to the sound source.

The study highlights that plosive closure duration and burst intensity are aspects singers can adjust to improve intelligibility, but further research is needed to explore whether singers naturally use these adjustments and whether elongating plosive closures could negatively affect legato or the naturalness of the performance.

6. Acknowledgements

We are grateful for all participants in the two listening experiment.

The study has been supported by the Estonian Research Council (PRG 1552).

7. References

[1] J. W. Gregg, "On articulation – Part I". *Journal of Singing,* 47(5), 30-32, 1991.

[2] H. D. Nelsson, and W. R. Tiffany, "The intelligibility of song: Research results with a new intelligibility test". *NATS Bulletin*, 25(2), 22–33. 1968

[3] G. L. Phillips, "Diction: A rhapsody", *Journal of Singing*, 58(5), 405–409, 2002.

[4] I. Titze, "Why is the verbal message less intelligible in singing than in speech?", *NATS Bulletin*, 38(3), 37, 1982.

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

[6] J. Melton, "Do you put the words across? An interview by Annabel Comfort", *Etude*, June, 15, 1953.

[7] S. Sharnova, "Diction*", NATS Bulletin*, 3(6), 4, 1947.

[8] W. Vennard, *Singing: The mechanism and the technic*. Carl Fisher, 1967.

[9] C. Ware, *Basics of vocal pedagogy: The foundations and process of singing.* McGraw-Hill, 1998.

[10] R. M. Brown, *The singing voice*. The Macmillan Company, 1946.

[11] V. Fuchs, *The art of singing and voice technique*. London House and Maxwell, 1964.

[12] M. Marshall, "Is exaggeration required for good English diction?", *Diapason*, 47(3), 18, 1956.

[13] J. Meyer, *Acoustics and the Performance of Music. Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*. 5th edition, Springer, New York, 2009.

[14] L. B. Collister, and D. Huron, "Comparison of word intelligibility in spoken and sung phrases", *Empirical Musicology Review*, 3(3), 109–125, 2008. https://doi.org/10.18061/1811/34102

[15] R. Appelman, *The science of vocal pedagogy.* Indiana University Press, 1986.

[16] A. Vurma, E. Meister, L. Meister, J. Ross, M. Raju, V. Kala, and T. Dede, „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, 2023. https://doi.org/10.1121/10.0021968

[17] P. N. Petkov, N. Braunschweiler, and Y. Stylianou, „Automated pause insertion for improved intelligibility under reverberation". in *Proceedings INTERSPEECH 2016*, September 8–12, 2016, San Francisco, USA.

[18] V. Best, C. R. Mason, J. Swaminathan, E. Roverud, and G.Jr. Kidd, "Does providing more processing time improve speech intelligibility in hearing-impaired listeners?", *169th Meeting of the Acoustical Society of America*, Pittsburgh, Pennsylvania, 18–22 May 2015, Psychological and Physiological Acoustics: Paper 1aPPb28.

[19] B. Gygi, and V. Shafiro, "Spatial and temporal modifications of multitalker speech can improve speech perception in older adults", *Hearing Research*, 310, 76–86, 2014.

[20] A. H. Lessa, and M. J. Costa, "The impact of speech rate on sentence recognition by elderly individuals", *Brazilian Journal of Otorhinolaryngology*, 79(6), 745–752, 2013.

[21] P. Boersma, and D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.4.03, retrieved 4 January 2024, from http://www.praat.org/

[22] R. Correte, *Praat vocal toolkit*, [computer program], 2021–2022 https://www.praatvocaltoolkit.com/index.html (Last viewed Dec 08, 2022).

[23] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Usinglme4", *Journal of Statistical Software,* 67(1), 1–48, 2015.

[24] R Core Team, *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2021. https://www.R-project.org/.

[25] R. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means.* R package version 1.10.2, 2024. https://rvlenth.github.io/emmeans/

APPENDIX: Summaries of GLMMs

Experiment I: Test set G3

Formula: Correct ~ Duration + Burst + Duration * Cond + Consonant + (1 | Age_gr) + (1 | Gender)

Fixed effects:	β	SE	z	p	
(Intercept)	3.99	0.39	10.25	<0.001	***
Dur_150	-1.08	0.44	-2.46	-0.014	*
Dur_260	0.86	0.68	1.26	0.208	
Burst_s	0.26	0.05	5.08	<0.001	***
Cond_BR	-4.36	0.38	-11.34	<0.001	***
Cond_Ch	-3.91	0.39	-10.17	<0.001	***
Cond_BN_BR	-4.57	0.38	-11.90	<0.001	***
Cond_BN_Ch	-4.58	0.38	-11.92	<0.001	***
Consonant_p	0.22	0.06	3.63	<0.001	***
Consonant_t	1.27	0.07	19.44	<0.001	***
Dur_150:Cond_BR	2.12	0.46	4.66	<0.001	***
Dur_260:Cond_BR	1.20	0.70	1.72	0.085	.
Dur_150:Cond_Ch	1.25	0.46	2.75	0.006	**
Dur_260:Cond_Ch	-0.35	0.69	-0.51	0.608	
Dur_150:Cond_BN_BR	0.90	0.45	1.98	0.047	*
Dur_260:Cond_BN_BR	-0.56	0.69	-0.82	0.414	
Dur_150:Cond_BN_Ch	0.79	0.45	1.75	0.081	.
Dur_260:Cond_BN_Ch	-1.10	0.69	-1.60	0.109	

Reference levels: Duration = "60ms", Burst = "weak", Cond = "BN", Consonant = "k"

Experiment I: Test set G4

Formula: Correct ~ Duration + Burst + Duration * Cond + Consonant + (1 | Age_gr) + (1 | Gender)

Fixed effects:	β	SE	z	p	
(Intercept)	2.31	0.25	9.21	<0.001	***
Dur_150	-0.45	0.18	-2.45	0.014	*
Dur_260	-0.1	0.20	-0.49	0.625	
Burst_s	0.96	0.05	19.29	<0.001	***
Cond_BR	-1.36	0.17	-8.04	<0.001	***
Cond_Ch	-2.11	0.16	-12.83	<0.001	***
Cond_BN_BR	-2.54	0.16	-15.47	<0.001	***
Cond_BN_Ch	-2.61	0.16	-15.85	<0.001	***
Consonant_p	-0.29	0.06	-4.75	<0.001	***
Consonant_t	-0.78	0.06	-12.83	<0.001	***
Dur_150:Cond_BR	0.82	0.23	3.57	<0.001	***
Dur_260:Cond_BR	1.00	0.25	4.06	<0.001	***
Dur_150:Cond_Ch	0.66	0.22	2.98	0.003	**
Dur_260:Cond_Ch	0.75	0.23	3.23	0.001	**
Dur_150:Cond_BN_BR	0.54	0.22	2.48	0.013	*
Dur_260:Cond_BN_BR	0.23	0.23	1.01	0.311	
Dur_150:Cond_BN_Ch	0.68	0.22	3.1	0.002	**
Dur_260:Cond_BN_Ch	-0.2	0.23	-0.87	0.387	

Reference levels: Duration = "60ms", Burst = "weak", Cond = "BN", Consonant = "k"

Experiment I: Test set F5

Formula: Correct ~ Duration + Duration * Cond + Consonant + (1 | Age_gr) + (1 | Gender)

Fixed effects:	β	SE	z	p	
(Intercept)	2.99	0.26	11.32	<0.001	***
Dur_150	0.09	0.24	0.37	0.713	
Dur_260	-0.14	0.24	-0.60	0.555	
Cond_BN	-1.14	0.22	-5.28	<0.001	***
Cond_BR	-2.51	0.21	-11.82	<0.001	***
Cond_Ch	-2.39	0.21	-11.28	<0.001	***
Cond_BN_BR	-2.61	0.21	-12.23	<0.001	***
Cond_BN_Ch	-2.80	0.22	-13.03	<0.001	***
Consonant_p	-1.88	0.08	-23.29	<0.001	***
Consonant_t	-0.93	0.08	-12.22	<0.001	***
Dur_150:Cond_BN	-0.06	0.31	-0.18	0.858	
Dur_260:Cond_BN	0.17	0.30	0.58	0.563	
Dur_150:Cond_BR	0.31	0.30	1.03	0.303	
Dur_260:Cond_BR	1.19	0.30	4.01	<0.001	***
Dur_150:Cond_Ch	0.14	0.30	0.47	0.641	
Dur_260:Cond_Ch	0.67	0.29	2.29	0.022	*
Dur_150:Cond_BN_BR	0.10	0.30	0.33	0.743	
Dur_260:Cond_BN_BR	0.06	0.30	0.20	0.843	
Dur_150:Cond_BN_Ch	-0.19	0.31	-0.63	0.532	
Dur_260:Cond_BN_Ch	0.27	0.30	0.91	0.363	

Reference levels: Duration = "60ms", Cond = "Clear", Consonant = "k"

Experiment II: Test set G3

Formula: Correct ~ Duration * Cond + Consonant + Burst + Loc + (1 | Age_gr) + (1 | Gender)

Fixed effects:	β	SE	z	p	
(Intercept)	1.01	0.12	8.64	<0.001	***
Dur_150	0.65	0.13	5.12	<0.001	***
Dur_260	1.02	0.14	7.33	<0.001	***
Burst_s	0.83	0.06	13.64	<0.001	***
Cond_BN	-1.35	0.10	-13.35	<0.001	***
Consonant_p	-0.44	0.07	-6.31	<0.001	***
Consonant_t	0.57	0.08	7.49	<0.001	***
Loc_front	0.48	0.06	7.90	<0.001	***
Dur_150:Cond_BN	-0.67	0.15	-4.36	<0.001	***
Dur_260:Cond_BN	-1.08	0.16	-6.60	<0.001	***

Reference levels: Duration = "60ms", Burst = "weak", Cond = " Clear", Consonant = "k", Location = "back"

Test set G4

Formula: Correct ~ Duration * Cond + Consonant + Burst + Loc + (1 | Age_gr) + (1 | Gender)

Fixed effects:	β	SE	z	p	
(Intercept)	1.19	0.16	7.59	<0.001	***
Dur_150	0.329	0.10	3.09	0.002	**
Dur_260	0.68	0.11	6.19	<0.001	***
Burst_s	1.19	0.06	20.26	<0.001	***
Cond_BN	-0.94	0.10	-9.73	<0.001	***
Consonant_p	-0.60	0.07	-8.08	<0.001	***
Consonant_t	-1.76	0.07	-23.95	<0.001	***
Loc_front	0.40	0.06	6.88	<0.001	***
Duration_150:Cond_BN	-0.30	0.14	-2.19	0.029	*
Duration_260:Cond_BN	-0.74	0.14	-5.19	<0.001	***

Reference levels: Duration = "60ms", Burst = "weak", Cond = "Clear", Consonant = "k", Location = "back"

Test set F5

Formula: Correct ~ Duration * Cond + Consonant + Loc + (1 | Age_gr) + (1 | Gender)

Fixed effects:	β	SE	z	p	
(Intercept)	1.17	0.12	10.17	<0.001	***
Dur_150	0.04	0.13	0.26	0.792	
Dur_260	0.44	0.13	3.35	<0.001	***
Cond_BN	-0.49	0.13	-3.70	<0.001	***
Consonant_p	-2.55	0.10	-24.87	<0.001	***
Consonant_t	-1.25	0.09	-13.97	<0.001	***
Loc_front	0.01	0.08	0.10	0.924	
Dur_150:Cond_BN	-0.08	0.19	-0.42	0.671	
Dur_260:Cond_BN	-0.48	0.19	-2.55	0.011	*

Reference levels: Duration = "60ms", Cond = "Clear", Consonant = "k", Location = "back"

Signif. codes: '***' p<0.001, '**' p<0.01, '*' p<0.05, '.' p<0.1