forum acusticum 2026
8-12 September · Graz, Austria

SPATIAL RELEASE FROM MASKING IN TURKISH SPEECH INTELLIGIBILITY

Kivanc Kitapci[1], Öğütnaz Çoban[2], Melike Kayhan[1], Asya Larisa Coutinho Kitapci[3]

[1]*Department of Interior Architecture and Environmental Design, Faculty of Architecture and Design, TOBB Economics and Technology University, Ankara, Türkiye*
[2] *Department of Audiology, Institute of Health Sciences, Ankara Medipol University, Ankara, Türkiye*
[3] *Building Sciences, Graduate School of Natural and Applied Sciences, Middle East Technical University, Türkiye*

Abstract

This study investigated the effects of Spatial Release from Masking (SRM) on speech intelligibility in normal-hearing native Turkish speakers using the Turkish Modified Rhyme Test (MRT). Thirty participants (15 females, 15 males; mean age: 23.0 years, *SD* = 2.3) completed the MRT under nine spatial conditions defined by the combination of target word azimuth angle (0°, -50°, +50°) and speech-shaped noise (SSN) azimuth angle (0°, -50°, +50°) at a fixed 0 dB signal-to-noise ratio (SNR). Accuracy and reaction time of the participants were also recorded. A two-way repeated-measures ANOVA revealed a significant main effect of word angle (p = .016) but not SSN angle (p = .206). No significant effect of spatial separation between target and masker was observed. Among the five distinctive features, graveness yielded the lowest accuracy and sibilation the highest (p = .022), with no significant interaction between distinctive features and SSN angle. The data establish a normative baseline for Turkish MRT and draw attention to spatial configuration and distinctive features as variables that deserve consideration in multilingual speech intelligibility research.

Keywords: *modified rhyme test, speech intelligibility, spatial release from masking, speech-shaped noise, Turkish*

1. Introduction

Speech intelligibility in challenging acoustic conditions cause significant perceptual load on the auditory system. Even listeners with normal hearing frequently struggle to follow a conversation when background noise is present, a challenge that becomes especially important in everyday settings such as classrooms, crowded social environments, and clinical spaces. Various factors influence a listener's ability to follow a target speaker while a competing background noise is present. Examples of the influential factors are the spectral and temporal properties of the masker, the signal-to-noise ratio (*SNR*), and the relative positions of the sound sources in space [1].

One of the well-established factors claimed to improve speech intelligibility in noise is spatial release from masking (SRM): when the target speech and the masking signal originate from different directions, speech intelligibility tends to improve compared to a co-located target-masker pair. Angular separation between target and masker has been linked to better intelligibility across various listening contexts [2,3]. The auditory system differentiates the signals received by both ears by utilizing the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) mechanisms to improve the separation of the target from the masker [4].

This study is part of a broader research project funded by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) that examines how energetic masking, information masking, and binaural unmasking shape speech intelligibility across languages and communication channels. As a pilot study of this broader research project, the current study focuses on normal-hearing native Turkish speakers to establish a theoretical foundation on how SRM affects speech intelligibility in Turkish. Speech-shaped noise (SSN) was selected as the masker as it produces energetic masking spectrally matched to the target speech, allowing the spatial effects of the masker to be examined in isolation from more complex information masking conditions that will be addressed in subsequent stages of the project.

Originally introduced by House et al. [5], the Modified Rhyme Test (MRT) is a closed-set word recognition task in which listeners choose among six monosyllabic words that differ by a single phoneme. The structured format of the MRT is especially suitable for detecting differences in various acoustic conditions, and it has been utilized in both audiological assessment and noise research. In the current study, a Turkish version of the MRT has been used for phonemic investigation of masking effects in Turkish [6].

Corresponding author: *kkitapci@etu.edu.tr*

In the literature, the number of SRM studies conducted with localized materials, in our case, Turkish, is limited compared to English. Turkish draws on a phonological system unlike English, with vowel harmony governing vowel sequences, a heavily agglutinative morphology, and a consonant inventory that diverges in several respects from Germanic languages [7]. These structural differences raise the possibility that Turkish listeners may respond to SRM in ways that do not replicate findings from English oriented studies. Understanding such effects in Turkish-speaking listeners is important for both basic auditory science and audiological practice in Turkey. The present study therefore analyzed the effects of target word direction and SSN direction on word recognition accuracy and reaction time in normal-hearing native Turkish speakers using the Turkish MRT. We hypothesized that spatial separation between target and masker would improve speech intelligibility consistent with SRM, and that distinctive features would influence lower intelligibility.

2. Methodology

2.1. Participants

Thirty native Turkish-speaking adults (15 females, 15 males; mean age: 23.0 years, *SD* = 2.3) participated to the study. All participants reported normal hearing with no history of ear disease, noise exposure, or audiological complaints. Hearing status was confirmed via self-report. Participants provided written informed consent, and the study was conducted in accordance with the Declaration of Helsinki.

2.2. Materials

The Turkish MRT was used as the speech intelligibility measure [6]. The test consists of 48 monosyllabic Turkish word groups, each containing six rhyming items differing by a single consonant in initial or final position. Following the phonetic classification scheme of Arıöz and Günel [6], the 48-word groups were assigned to five categories based on their Turkish distinctive features: graveness, nasality, sustention, compactness, and sibilation. Each target word was embedded in the carrier phrase "Şimdi duyacağınız kelime..." (The word you will hear is...) prior to presentation. To control for level differences across items, all stimuli were RMS-normalized in MATLAB before the experiment.

2.3. Masker

The masker was a speech-shaped noise (SSN) computed in MATLAB from the long-term average spectrum of the target speech corpus. This approach ties the masker's spectral profile directly to the stimuli, so that observed performance differences across spatial conditions are not confounded by unequal spectral overlap between target and masker. The SSN was spatialized to the same three angles as the target speech using the same binaural recording setup. Target and masker were mixed at 0 dB SNR in all conditions.

2.4. Experimental Setup

Binaural recordings of the target speech stimuli were made using a Neumann KU-100 dummy head microphone from three fixed angles: 0° (frontal), -50° (50L), and +50° + (50R), yielding nine spatial combinations in total. Testing was conducted in a sound-treated audiometric booth within the acoustic laboratory. The experiment was delivered via jsPsych on a calibrated computer system.

2.5. Procedure

Prior to the main experiment, participants completed a practice block of six trials to familiarize themselves with the task interface and response procedure. On each trial, a target word was presented aurally together with the SSN masker. Simultaneously, the target word and five rhyme-group alternatives were displayed on screen, and participants selected the word they heard. Both target word identity and SSN angle were randomized across trials. A 5-second countdown timer was displayed; non-responses within this window were recorded as missing. Accuracy (correct/incorrect) and reaction time (RT, ms, from stimulus onset to response) were logged automatically. Each participant completed all 48 trials.

2.6. Statistical Analysis

Data were analyzed using IBM SPSS Statistics (version 31.0). Two repeated-measures ANOVAs were conducted: (1) a two-way ANOVA with word angle and SSN angle as within-subject factors, and (2) a one-way ANOVA with distinctive features as the within-subject factor. Mauchly's test of sphericity was applied for all repeated-measures analyses; no violations were detected, and Sphericity Assumed values are therefore reported throughout. Post-hoc pairwise comparisons were conducted with Bonferroni correction. The significance threshold was set at $\alpha = .05$.

Additionally, a 5 × 3 repeated-measures ANOVA with distinctive features and SSN angle as within-subject factors was conducted to examine their interaction. Since each participant completed 48 trials drawn from a randomized pool of 432 possible target–masker combinations, not all distinctive feature × SSN angle cells were represented for every participant. Six participants had at least one empty cell and were therefore excluded from this analysis, leaving N = 24.

3. Results

3.1. Overall Performance

Mean word recognition accuracy across all conditions was 81.53% (SD = 7.31%) and mean reaction time was 1941.4 ms (SD = 258.2 ms). Correct responses were associated with shorter reaction times (M = 1851.5 ms, SD = 244.8 ms) compared to incorrect responses (M = 2365.0 ms, SD = 402.1 ms).

3.2. Effect of Word Angle and SSN Angle

Word angle emerged as the only statistically significant factor in the two-way repeated-measures ANOVA, $F(2, 58) = 4.47$, $p = .016$, partial $\eta^2 = .134$, while neither the main effect of SSN angle, $F(2, 58) = 1.63$, $p = .206$, nor the Word Angle × SSN Angle interaction, $F(4, 116) = 0.40$, $p = .807$, reached significance. Mauchly's test confirmed that sphericity was not violated for any effect (all $p > .05$), and Sphericity Assumed values are therefore reported. Mean accuracy for all nine Word Angle × SSN Angle combinations is presented in Table 1, and ANOVA results are summarized in Table 2.

Bonferroni-corrected post-hoc comparisons revealed that word recognition accuracy was significantly lower when the target word was presented from 50° left (W50L: M = 77.0%, SE = 1.96) compared to 0° (W0: M = 84.1%, SE = 1.99; p = .028). Performance in the W50R condition (M = 84.3%, SE = 2.29) did not differ significantly from W0 (p = 1.000) or W50L (p = .055).

Table 1. Mean accuracy % (SD) for each Word Angle × SSN Angle combination (N = 30)

Word Angle	S0° M (SD)	S50L° M (SD)	S50R° M (SD)
W0°	82.1 (22.6)	89.4 (14.7)	80.8 (17.7)
W50L°	76.2 (23.3)	78.9 (20.3)	76.1 (18.3)
W50R°	85.3 (18.8)	84.9 (13.7)	82.8 (18.0)

Table 2. Two-way repeated-measures ANOVA: Word Angle × SSN Angle

Effect	df	Mean Square	F	p	partial η^2
Word Angle	2.58	1549.52	4.47	**.016**	.134
SSN Angle	2.58	478.30	1.63	.206	.053
Word × SSN	4.116	143.95	0.40	.807	.014

Note. df = degrees of freedom. Bold p value indicates significance at α = .05.

3.3. Effect of Spatial Separation

To understand whether increasing spatial separation between target and masker translated into better intelligibility, scores were grouped by the angular distance between the two sources: co-located (0°), partially separated (50°), and maximally separated (100°). A one-way repeated-measures ANOVA revealed no significant effect of separation group. Descriptive statistics and test results are presented in Table 3.

Table 3. Mean accuracy % (SD) by spatial separation group (N = 30)

Separation Group	M %	SD
Co-located	81.26	11.56
Partial (50°)	82.11	8.87
Maximum (100°)	80.48	9.61

Note. One-way repeated-measures ANOVA: F (2, 58) = 0.25, p = .779, partial η^2 =.009. Mauchly's test: p =.054. Sphericity Assumed values reported.

3.4. Effect of Distinctive Features

A one-way repeated-measures ANOVA on accuracy across five distinctive features revealed a significant main effect, F (4, 116) = 2.97, p = .022, partial η^2 = .093. Mauchly's test confirmed that sphericity was not violated (p = .212), and Sphericity Assumed values are therefore reported. Graveness words yielded the lowest accuracy (M = 76.3%, SD = 14.7) and sibilation words the highest (M = 85.6%, SD = 13.4). Bonferroni-corrected post-hoc comparisons did not yield any statistically significant pairwise differences, with the closest contrast being graveness versus sibilation (p = .081). Descriptive statistics per distinctive feature and ANOVA results are presented in Table 4 and Table 5, respectively.

Table 4. Mean accuracy % (SD) per distinctive feature (N = 30)

Distinctive Feature	M (%)	SD
Graveness	76.30	14.70
Nasality	78.30	17.20
Sustention	83.00	11.80
Compactness	85.20	11.40
Sibilation	85.60	13.40

Table 5. One-way repeated-measures ANOVA: Distinctive Feature

Effect	df	Mean Square	F	p	partial η^2
Distinctive Feature	4.116	516.27	2.972	**.022**	.093

Note. df = degrees of freedom. Bold p value indicates significance at α = .05.

3.5. Distinctive Feature × SSN Angle Interaction

A 5 × 3 repeated-measures ANOVA with distinctive feature and SSN angle as within-subject factors was conducted on a subsample of 24 participants with complete data across all cells. Mauchly's test confirmed sphericity was not violated for any effect (all p > .05), and Sphericity Assumed values are reported. Neither the main effect of distinctive feature nor SSN angle nor their interaction reached significance. Estimated marginal means are presented in Table 6.

Table 6. Mean accuracy % (SE) per distinctive feature × SSN angle (N = 24)

Distinctive Feature	S0° M(SE)	S50L° M(SE)	S50R° M(SE)
Graveness	71.3 (5.8)	74.6 (4.3)	77.4 (4.8)
Nasality	82.4 (4.3)	84.4 (4.5)	76.3 (5.7)
Sustention	84.1 (5.0)	88.4 (4.1)	79.1 (4.6)
Compactness	87.2 (3.4)	81.8 (5.5)	84.2 (5.6)

Sibilation	89.2 (4.3)	80.4 (6.1)	78.7 (5.8)

Note. distinctive_feature: F (4, 92) = 2.02, p = .098; SSN_angle: F (2, 46) = 0.97, p = .389; interaction: F (8, 184) = 0.73, p =.666.

4. Discussion

4.1. Effect of Word Angle

Among all the variables tested, word angle is the only statistically significant factor. Listeners were notably less accurate when the target word received from -50° compared to 0°; however, the accuracy were similar when the target word received from +50°. The data indicates a directional asymmetry rather than a general lateralization effect, possible caused by the right-ear advantage [8, 9]. It was claimed that the left hemisphere tends to be dominant in speech processing; therefore, signals arriving at the right ear follow a more direct processing pathway. A word arriving from the left, by contrast, is routed initially to the right hemisphere, and must then be transferred across hemispheres. This sequence may introduce enough additional processing load to decrease intelligibility.

4.2. Effect of SSN Angle

Contrary to our hypothesis, SSN angle did not show a statistically significant main effect on word recognition accuracy (p = .206). Descriptively, however, a trend consistent with SRM was observed: accuracy was highest when the SSN was presented from 50° left (M = 84.4%) and lowest from 50° right (M = 79.9%), representing a 4.5 percentage point difference. The absence of statistical significance may be partly attributable to the fixed 0 dB SNR used in the present study. At this SNR, the energetic masking load is substantial, and individual variability in binaural processing may have obscured the spatial benefit. Research consistently shows that SRM tends to be largest when listeners are working near their intelligibility threshold [3], and a 0 dB SNR may simply have been too demanding for spatial cues to exert a measurable influence on performance. It is also noteworthy that while SSN angle alone did not reach significance, the significant effect of word angle demonstrates that spatial orientation did influence performance in the present paradigm. Under the current conditions, the direction of the target signal rather than the masker was the more critical spatial determinant of recognition accuracy. This finding is consistent with studies emphasizing the role of target lateralization in speech intelligibility [1] and highlights that spatial release from masking in the present context may operate primarily through target-side processing advantages rather than masker-side suppression. Future studies employing adaptive SNR procedures or multiple SNR levels would allow a more sensitive examination of SSN direction effects in Turkish-speaking listeners.

The absence of a separation-group effect further reinforces this interpretation. Even when scores were collapsed into co-located, partial, and maximum separation conditions, no performance advantage emerged for greater spatial separation, F (2, 58) = 0.25, p =.779. Therefore, under the current 0 dB SNR conditions, spatial release from masking did not scale with the degree of target-masker separation.

4.3. Effect of Distinctive Features

The analysis of the data on the distinctive features showed that graveness was the hardest distinctive feature to identify, while sibilation was the easiest. The present data, collected from normal-hearing participants, seem to be supporting the results of the study of Arıöz and Günel [6], which underlines that sustention and sibilation tended to show the most variable scores among hearing-impaired listeners. Although no pairwise comparison passed Bonferroni correction, the approximately 10% gap between the highest and lowest intelligible distinctive featurees has practical implications for clinical MRT interpretation.

Although the distinctive feature × SSN angle interaction is not statistically significant, (F[8, 184] = 0.73, p = .666), the descriptive data for sustention is important. Intelligibility was approximately 9% higher when SSN arrived from +50° (M = 88.4%) compared to -50° (M = 79.1%). This may indicate a directional sensitivity specific to sustained consonants, though the present subsample (N = 24) limits firm conclusions. Future work with larger samples could examine whether this result is of significance.

4.4. Limitations and Future Directions

The present study employed a fixed 0 dB SNR, which represents a relatively demanding listening condition. This choice was deliberate, as it allowed spatial effects to be examined under a consistent and challenging masking load. However, it is possible that a range of SNR levels would reveal a clearer SRM benefit, since spatial advantages tend to be more significant when listeners are closer to their intelligibility threshold. Future work incorporating adaptive SNR procedures or multiple fixed levels would allow a more complete picture of how noise direction affects Turkish MRT performance. A higher number of participants and cross-linguistic comparisons would strengthen the generalizability of these findings, as speech intelligibility has been shown to vary across languages under different acoustic conditions [10], and remain priorities for the next phase of the work.

5. Conclusions

This study examined the effects of SRM and distinctive features on Turkish speech intelligibility in normal-hearing listeners. The following conclusions are drawn: Word angle significantly affected recognition accuracy (p = .016), with performance being lower when the target was presented from -50°, consistent with a right-ear advantage for speech perception. SSN angle did not produce a significant main effect (p = .206), and no Word Angle × SSN Angle interaction was observed (p = .807). When scores were further collapsed by degree of spatial separation between target and masker, no performance advantage emerged for greater separation, F (2, 58) = 0.25, p = .779, suggesting that spatial release from masking did not scale with target-masker distance under the current 0 dB SNR conditions.

Among the five distinctive features, graveness consistently showed the highest intelligibility, while sibilation was the least intelligible ($p = .022$). The distinctive feature × SSN angle interaction was similarly non-significant ($F [8, 184] = 0.73$, $p = .666$), though sustention showed a descriptive pattern with better performance when SSN arrived from -50° than from +50°, that merits follow-up in larger samples. In conclusion, the results suggest that both spatial configuration of the target and masker signals, and distinctive features of languages appear to influence speech intelligibility on some degree. However, studies utilizing a larger variety of SNR conditions and conducting listening tests on more languages are required to understand the effects of maskers on multilingual speech intelligibility.

6.Acknowledgments

This work is supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) under project number 125M560.

References

[1] Bronkhorst, A. W., The cocktail-party problem revisited: Early processing and selection of multi-talker speech, Attention, Perception, & Psychophysics, 77(5), 1465–1487, (2015).

[2] Bronkhorst, A. W. and Plomp, R., The effect of head-induced interaural time and level differences on speech intelligibility in noise, Journal of the Acoustical Society of America, 83(4), 1508–1516, (1988).

[3] Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K., The role of perceived spatial separation in the unmasking of speech, Journal of the Acoustical Society of America, 106(6), 3578–3588, (1999).

[4] Biberger, T. and Ewert, S. D., The effect of room acoustical parameters on speech reception thresholds and spatial release from masking, Journal of the Acoustical Society of America, 146(4), 2188–2200, (2019).

[5] House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D., Articulation-testing methods: Consonantal differentiation with a closed response set, Journal of the Acoustical Society of America, 37(1), 158–166, (1965).

[6] Arıöz, U. and Günel, B., Modified rhyme test for evaluating Turkish speech intelligibility, Gazi Medical Journal, 26(2), 75–81, (2015).

[7] Özsoy, A. S., Türkçenin Yapısı [The Structure of Turkish], Boğaziçi Üniversitesi Yayınevi, Istanbul, (2004).

[8] Hugdahl, K. and Westerhausen, R., Speech processing asymmetry revealed by dichotic listening and functional brain imaging, Neuropsychologia, 93, 466–481, (2016).

[9] Wächtler, M., Sandmann, P., and Meister, H., The right-ear advantage in static and dynamic cocktail-party situations, Trends in Hearing, 28, (2024).

[10] Galbrun, L. and Kitapci, K., Speech intelligibility of English, Polish, Arabic and Mandarin under different room acoustic conditions, Applied Acoustics, 114, 79–91, (2016).