Speech-on-Speech Masking: Effect of Maskers with Different Degrees of Linguistic Information

Le masquage par la parole : l'effet de bruits masquants contenant différents degrés d'information linguistique

Anoop Basavanahalli Jagadeesh
Ajith Kumar Uppunda

KEYWORDS

LINGUISTIC MASKING

INFORMATIONAL MASKING

SPEECH-ON-SPEECH

LISTENING EFFORT

SUBJECTIVE RATING

Anoop Basavanahalli Jagadeesh and Ajith Kumar Uppunda

All India Institute of Speech and Hearing, Mysuru, India

Editor:
Josée Lagacé

Editor-in-Chief:
David H. McFarland

Abstract

The current study measured speech recognition and subjective listening effort scores while systematically varying the amount of linguistic information in maskers. Linguistic information in the maskers was varied by (a) increasing the number of speakers in the speech babble maskers and (b) time-reversing them. In Experiment 1, we measured speech recognition performance (signal-to-noise ratios required for 50% accuracy of sentences) for 16 participants. The speech (sentences) recognition scores were obtained in 15 background conditions: speech babble maskers with 2 to 8 speakers (7 conditions), time-reversed babble maskers (7 conditions), and a speech-spectrum noise. For Experiment 2, another 15 participants rated the effort (7-point rating scale) required to understand sentences in the same maskers as Experiment 1. This was done at a signal-to-noise ratio of 0 dB. Results showed that fewer speakers in the babble maskers (a) caused the greatest masking effects and (b) required the greatest listening effort ratings. Speech babble maskers resulted in significantly higher masking effects than reverse babble maskers only for the 2- and 3-speaker babble conditions. However, the listening effort scores were substantially higher for the speech babble maskers than reverse babble maskers in most of the conditions. Results suggest that both magnitudes of masking and the listening effort scores are related to the linguistic information in the masker.

Abrégé

Dans la présente étude, les scores de reconnaissance de stimuli verbaux et d'effort à l'écoute ont été mesurés en variant systématiquement la quantité d'information linguistique contenue dans des bruits masquants. L'information contenue dans les bruits masquants a été modifiée en (a) augmentant le nombre d'interlocuteurs et (b) en inversant ces bruits. Dans l'expérience 1, les scores de reconnaissance de la parole (rapport signal-sur-bruit permettant de comprendre 50% des phrases) de 16 participants ont été mesurés. Ceux-ci ont été obtenus dans 15 environnements bruyants : 7 bruits de verbiage qui incluaient de 2 à 8 interlocuteurs, ces 7 mêmes bruits de verbiage inversés, de même qu'un bruit à spectre vocal. Dans l'expérience 2, 15 autres participants ont noté (sur une échelle de 7 points) l'effort qu'ils ont eu à déployer pour comprendre des phrases dans les mêmes 15 environnements bruyants. Pour cette deuxième expérience, le rapport signal-sur-bruit était de 0 dB. Les résultats ont montré qu'un bruit de verbiage qui inclut moins d'interlocuteurs (a) cause un effet masquant plus important et (b) génère des scores d'effort à l'écoute plus élevés. Les bruits de verbiage ont conduit à un effet masquant plus important que les bruits de verbiage inversés, et ce, seulement lorsqu'ils contenaient de 2 à 3 interlocuteurs. Toutefois, les scores d'effort à l'écoute étaient considérablement plus élevés pour les bruits de verbiage que pour les bruits de verbiage inversés, et ce, dans la majorité des environnements bruyants. Les résultats suggèrent que les scores de reconnaissance de stimuli verbaux et d'effort à l'écoute sont associés à la quantité d'information linguistique contenue dans un bruit masquant.

Speech perception is affected in different ways by different types of maskers depending on their spectral, temporal, linguistic, and/or energetic characteristics. Speech recognition is reportedly better in temporally modulated noise compared to steady-state noise due to "dip-listening" or "glimpsing" (Festen & Plomp, 1990; Jin & Nelson, 2006; Summers & Molis, 2004). Speech maskers are dynamic signals and provide multiple opportunities for listeners to glimpse (i.e., gaps between words and sentences or the presence of weak speech segments such as /f/ and /θ/, for example) the target speech signal. Yet, speech recognition is more challenging in the presence of speech backgrounds compared to non-speech backgrounds (Bronkhorst, 2000; Carhart et al., 1975; Hoen et al., 2007; Lu et al., 2016) because speech maskers cause perceptual confusions with the target speech due to their linguistic similarities. This excessive masking of speech by competing speech(es) is termed *informational masking* (Brungart, 2001; Brungart et al., 2001; Kidd et al., 2008). As speech maskers are highly variable in their linguistic content, the type and amount of linguistic confusions that these maskers create are also highly variable and random.

In a masking scenario, total masking is a sum of at least two major types of masking (Kidd & Colburn, 2017): energetic masking and informational masking. Energetic masking is associated with the physical attributes of the target and maskers. Informational masking, on the other hand, is caused by the uncertainty or confusability between the target and the masker (Hafter & Schlauch, 1989; Leek et al., 1991; Neff & Green, 1987). In speech-on-speech masking, informational masking is calculated as the difference in the magnitude of masking under a speech masker—often speech babbles with varying numbers of speakers—and a non-speech masker with identical spectral content (i.e., speech-spectrum noise [SSN] or modulated noise; Balakrishnan & Freyman, 2008; Bronkhorst, 2000; Brungart et al., 2013; Freyman et al., 2004). The reduction in speech recognition under speech maskers, compared to non-speech maskers, despite similar spectral characteristics of maskers, is often attributed to confusion with the linguistic information present in the speech maskers (Mattys et al., 2009; Rhebergen et al., 2005; Simpson & Cooke, 2005). These linguistic confusions are due to the acoustic phonetic information (Hoen et al., 2007) and/or the lexical semantic information (Brungart & Simpson, 2004) present in the target speech and babble masker.

The amount of informational masking in speech-on-speech masking depends on the amount of intelligible linguistic information in the masker (Simpson & Cooke, 2005). The intelligibility and linguistic information of the speech masker (speech babble) are inversely related to the number of simultaneous speakers in the babble (Rosen et al., 2013; Simpson & Cooke, 2005). Thus, the linguistic information in a babble masker is greatest when there are fewer simultaneous (usually less than four) speakers. Previous studies have confirmed this inverse relationship between the number of talkers in the babble and the magnitude of informational masking (i.e., Lu & Cooke, 2008; Simpson & Cooke, 2005).

Maskers which approximate the features of a babble can also cause greater masking effects than noise maskers. For example, a time-reversed babble masker lacks any lexical semantic information but still causes greater masking effects than an SSN. This is because the time-reversed babble has spectral and temporal features similar to that of the speech babble (Arai, 2010; Rhebergen et al., 2005). The excessive masking observed for the reverse babble maskers is thought to be because of (a) greater linguistic confusion and uncertainty due to the presence of acoustic phonetic information and (b) increased forward masking effects due to their unusual temporal envelope (Rhebergen et al., 2005). A time-reversed babble masker also provides excellent control in order to study the effects of intelligibility on informational masking.

Most of the work on informational masking has involved speech recognition tasks. Listening effort, a metric describing the difficulty and effort involved in comprehending speech, can also be used to measure informational masking (Rennies et al., 2019). Typically, a listener expends little energy to understand speech in a quiet environment. However, the addition of a competing signal places extra demands on the cognitive resources of a listener. Such a cognitive effort expended by a listener when parsing a target from a competing message is referred to as *listening effort* (Howard et al., 2010; Peelle, 2017; Pichora-Fuller et al., 2016). The Ease of Language Understanding model (Rönnberg et al., 2013) and the Framework for Understanding Effortful Listening model (Pichora-Fuller et al., 2016) describe, in detail, the interactions among speech comprehension, cognitive resources, and background noise. According to these models, a competing signal introduces a mismatch between the incoming signal—perhaps, due to distortion—and the long-term phonological/lexical representations at the level of the phonological loop, part of the working memory construct (see Baddeley, 2003, for a detailed review on the different components of working memory). Such distortions necessitate the allocation of additional cognitive resources for parsing the target speech from the background speech. Speech babble maskers with fewer talkers, and more linguistic information, result in a higher cognitive processing load compared to that of non-speech maskers (Koelewijn et al., 2012) and hence more listening effort is required for speech recognition.

Considerable differences in listening effort scores are reported in spite of similar recognition scores (Brungart et al., 2013). Increased listening effort can have negative consequences on sustained speech comprehension, perhaps due to listener fatigue (Gosselin & Gagné, 2010; Peelle, 2017). Therefore, listening effort has equal importance to speech recognition performance as a metric. Previous studies have mostly used speech recognition scores alone to estimate the effects of informational masking.

In our study, we intend to supplement the outcomes of speech recognition tasks with listening effort scores. We aim to systematically vary the amount of information in the masker and observe the effects of informational masking on two different, but related, metrics of speech perception: recognition and subjective listening effort rating (Rudner et al., 2012). The information in the maskers was altered by varying the number of speakers in the babbles and by time-reversing them. Comparing the performance among the babble maskers—with a varying number of talkers—helps in quantifying the change in lexical-semantic information of the masker. The time-reversed versions of these babble maskers help quantify the change in acoustic-phonetic information. These maskers can also help estimate the real-world difficulties faced by listeners. While the time-reversed babble maskers can simulate a non-native babble (sounds "speech-like" but does not have semantic information), the regular speech babble masker simulates the typical cocktail party scenario. We hypothesize that both increasing numbers of speakers in the babble maskers and their time-reversal will reduce the overall linguistic information contained in the masker. This results in lesser amounts of informational masking and, therefore, causes the tasks (speech recognition as well as listening effort rating) to become progressively easier.

Experiment 1: Speech Recognition Performance

Method

Participants

Sixteen participants (6 women, 10 men) aged 18 to 27 years (M = 24.1 years) volunteered for Experiment 1. All participants were native speakers of the Kannada language and had at least 12 years of formal education. Each participant had pure-tone air conduction hearing thresholds of 15 dB HL or better at octave frequencies between 250 and 8000 Hz. None of the participants had any history or complaints of otological or neurological problems. All participants signed informed consents before starting the testing. The Ethics Committee of the All India Institute of Speech and Hearing, Mysuru, reviewed and approved the research according to their bio-behavioural guidelines (Ref No: Ph.D/AUD-2/2016-17).

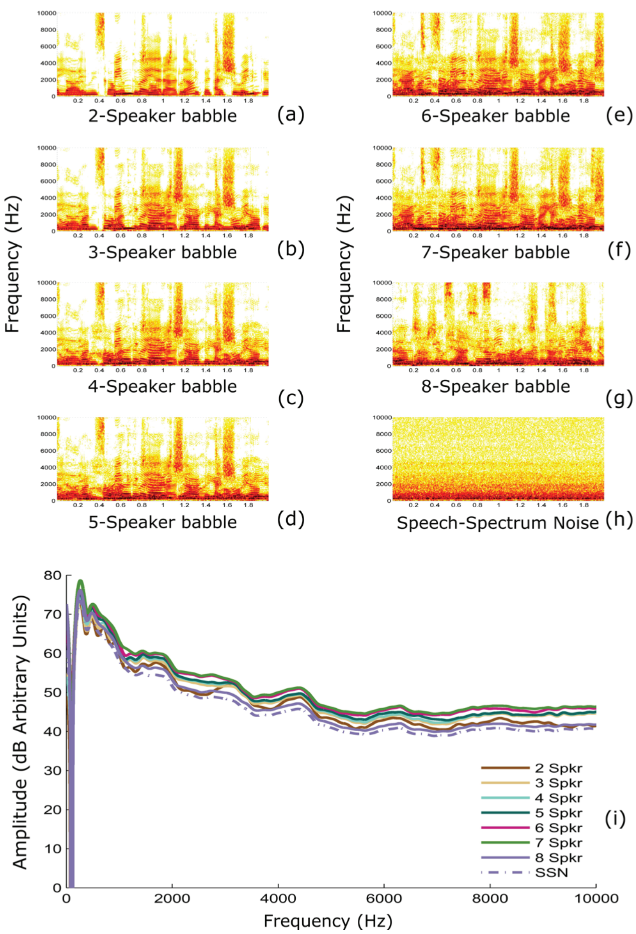
Stimuli

Target stimuli consisted of 15 phonemically balanced lists from the Kannada sentence identification test (Geetha et al., 2014). Each list included 10 low-predictability sentences spoken by a native female speaker of the Kannada language. Each sentence contained four keywords. All lists were matched for difficulty level. This meant that the signal-to-noise ratio (SNR) required for 50% accuracy (in the presence of SSN) was observed to be comparable (-5 dB) for all the lists (see Geetha et al., 2014, for further details regarding the generation and validity of the sentence lists). All lists were digitally stored in a computer with a 16-bit resolution and a sampling rate of 44100 Hz.

Three types of maskers were created for the study: speech babble maskers (SB), reversed babble maskers (RB), and a steady-state SSN. Eight female native speakers of Kannada read random sections in Kannada newspapers for 3.5 minutes. Speakers were instructed to read the passages in their regular speech rates, stress, and intonation patterns. The spoken samples were recorded using a Behringer B-2 Pro dual-diaphragm condenser microphone (Behringer, Germany) kept 5 cm from the speakers' mouth. The recordings were done using the Adobe Audition 3.0 software installed in a Lenovo-Z50 personal computer and connected to a Motu Microbook II external sound card interface. Spoken samples were recorded at a sampling frequency of 44100 Hz. Each individual recording was pruned for silent gaps of greater than 100ms. The pruned recordings were then amplitude (Root Mean Square) normalized. Two randomly chosen tracks were first mixed to obtain the 2-speaker babble. The 3-, 4-, 5-, 6-, 7-, and 8-speaker babble maskers were then created by successively adding randomly selected individual tracks to the previously mixed signal.

The RB maskers and the SSN were created from the previously generated SB maskers. Each of the seven SB maskers was temporally reversed to obtain the seven corresponding RB maskers. Finally, the SSN was created from the 8-speaker babble using a custom Matlab script (Gnanateja, 2016). The SSN had the same long-term average spectrum as the 8-speaker babble masker. Thus, there were 15 maskers for the experiments—seven speech babble maskers, seven reversed babble maskers, and the SSN masker. **Figure 1** depicts the spectra and the spectrograms of the different maskers used in the current study. **Figure 1** shows similar spectral compositions for all maskers.

Figure 1



Panels (a) through (g) represent the spectrograms of the 2-speaker to 8-speaker speech babble maskers for a 2-second section. The reverse babble maskers are not represented separately as they have identical spectrograms as the speech babble maskers, albeit reversed in time. With an increase in the number of speakers from panel (a) to (g), the salience of the spectral and formant information progressively decreases. Panel (h) represents the spectrogram of the speech-spectrum noise masker. Panel (i) shows the average long-term speech spectra of all the masker conditions considered for the study. Spkr = Speaker, SSN = Speech-Spectrum Noise.

Procedure

The experiments were carried out in a sound-treated room with ambient noise levels acceptable according to standards (American National Standard Institute, 2003). All target stimuli (sentences) were presented binaurally at 70 dB SPL. Stimuli were presented using a Lenovo-Z50 personal computer connected to Sennheiser HD 380 pro (Wedemark, Germany) headphones. The SNR required for achieving 50% correct identification (referred to as SNR-50 henceforth) of the speech stimuli was obtained for each participant, across all the 15 maskers. A separate sentence list was used to calculate SNR-50 for each of the 15 noise conditions. Each sentence had four keywords, resulting in 40 keywords per list. The SNR in each list was reduced from +10 to -8 dB across the 10 sentences in 2 dB steps. The SNR was manipulated by increasing the masker levels in 2 dB steps from 60 dB SPL to 78 dB SPL while keeping the target (sentence) level constant at 70 dB SPL. The masker began 0.5 s before the onset of each sentence and remained 0.5 s after the offset of the sentence. The mixing of the maskers with the sentences was done using custom Matlab scripts (Gnanateja, 2012). The mixing was done such that each target sentence was mixed with a random section within the masker. Also, a particular list was pre-selected to be mixed with a particular masker, thus yielding 15 lists, each mixed with a different masker. Further, the selection order of these 15 lists, as well as the order of presentation of the 10 sentences within each of these lists, was pseudo-randomized to minimize order effects.

We instructed participants to repeat the whole target sentence verbatim and to guess the possible words when the SNR of the presented speech was difficult. The total number of correctly identified keywords was noted for each list. The SNR-50 was calculated for each list using the Spearman-Karber equation (Finney, 1952; Tillman & Olsen, 1973),

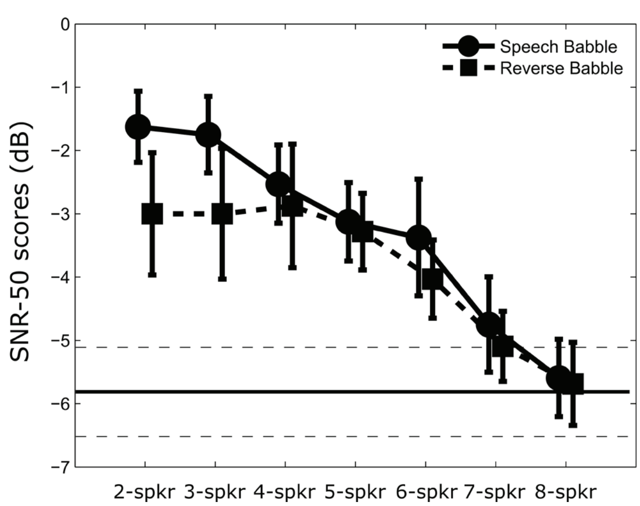
$$\text{SNR-50} = i + 1/2(d) - [(d)(\#correct) / (W)]$$

where *i* is the initial presentation level (+10 dB), *d* is the decrement step size (2 dB), *W* is keywords per decrement (4 in this case), and *#correct* is the total number of correct keywords repeated by the participants. This formula is designed to obtain the statistical 50% point in various biological and medical experiments and was hence suggested as a method to measure spondee thresholds (Tillman & Olsen, 1973). Because there were 15 masked conditions, the Spearman-Karber equation helped in calculating the speech recognition thresholds (i.e., the SNR-50 scores) quickly.

Results

We used JASP (Version 0.7.5.6; JASP Team, 2016) to carry out all statistical analyses. **Figure 2** shows the means and standard deviations of the SNR-50 scores for the different masker conditions. The figure shows a general tendency for the SNR-50 scores to improve when the number of speakers in the masker was increased for both the SB and RB masker conditions. A two-way 2 (Masker Type: SB & RB) X 7 (Number of Speakers in the Babble: 2 to 8) repeated-measures Analysis of Variance (RM-ANOVA) was performed on the data. The RM-ANOVA (corrected for violations of sphericity assumptions using Greenhouse-Geisser correction wherever necessary) revealed significant main effects of the masker types, $F(1, 15) = 35.60$, $p < .001$, $\eta_p^2 = .70$, and number of speakers, $F(4.16, 62.40) = 77.12$, $p < .001$, $\eta_p^2 = .88$, and a significant interaction between the masker types and the number of speakers, $F(3.77, 56.60) = 3.41$, $p < .001$, $\eta_p^2 = .27$. Further, to test for effects of the number of speakers on SNR-50 within the two maskers (SBs and RBs), a one-way RM-ANOVA was performed among the 2- to 8-speaker babble conditions. Additionally, the SSN condition was included in the ANOVA models. Therefore, each one-way RM-ANOVA compared scores across eight conditions (seven babble conditions plus the SSN condition).

Figure 2



Means and standard deviations of the SNR-50 (signal-to-noise) scores across the different masker conditions (number of speakers in the babble). The circles represent the SNR-50 scores for the speech babble maskers, the squares for the reverse babble maskers. The solid line at the bottom indicates the mean SNR-50 scores for the speech-spectrum noise masker, and the dashed lines show the standard deviations for the same. Spkr = Speaker.

The one-way RM-ANOVA revealed a significant main effect of the number of speakers for both the SB, $F(3.98, 59.83) = 122.0$, $p < .001$, $\eta_p^2 = .89$, and RB masker conditions, $F(4.01, 60.07) = 50.67$, $p < .001$, $\eta_p^2 = .77$. **Table 1** shows the follow-up posthoc pairwise comparisons (adjusted for multiple comparisons using Bonferroni's correction). SNR-50 scores in the presence of 2- and 3-speaker SB maskers were significantly poorer compared to all other conditions. There were no significant differences in the SNR-50 scores among the 4- through 6-speaker SB masker conditions. The 7- and 8-speaker SB maskers resulted in significantly better SNR-50 scores than the other SB maskers. However, there were no statistically significant differences in SNR-50 scores between the 7- and 8-speaker SB masker conditions and SSN masker condition. For the RB maskers, the 2- through 6-speaker conditions resulted in similar masking effects. SNR-50 scores in the 7- and 8-speaker masker conditions were significantly better than in the 2- to 6-speaker RB conditions. Additionally, there were no statistically significant differences in the SNR-50 scores between the 7- and 8-speaker babble conditions and the SSN.

We also compared the corresponding individual speaker condition pairs across the SB and RB maskers. Paired samples *t* tests revealed SNR-50 scores for 2-speaker, $t(14) = 5.21$, $p < .001$, and 3-speaker, $t(14) = 5.09$, $p < .001$, RB masker conditions to be significantly better than the corresponding SB masker conditions. Apart from these, all other paired comparisons revealed no significant differences.

Overall, Experiment 1 showed that the masking effect was highest when the number of speakers in the babble masker was less than four. The SB maskers caused significantly greater masking effects than the RB maskers for the 2- and 3-speaker conditions. The SNR-50 scores were comparable for both babble masker types when the numbers of speakers in the babble were between four and eight. The SSN caused significantly lesser masking effects than the babble maskers, except the 7- and 8-speaker conditions.

Experiment 2: Listening Effort Rating

Results of Experiment 1 showed that speech recognition performance was modulated by the type of masker and the number of speakers in the masker. SNR-50 scores differed significantly between the low (2- or 3-speaker) and high (7- or 8-speaker) number of speaker conditions. Also, performances under the RB maskers did not show significant differences until the 6-speaker condition. There were also no significant differences between the SB and RB masker conditions from the 4-speaker condition and above. However, despite comparable performances on the speech recognition task, it is possible that the effort expended to achieve similar performances could be different (Brungart et al., 2013). Listening effort can be a particularly useful metric, especially when recognition performances reach saturation levels (Gagné et al., 2017; Rennies et al., 2014). Furthermore, listening effort rating scores are indicated to be more influenced by working memory compared to speech recognition performance (Rudner et al., 2012). This points towards subjective ratings and speech recognition having different but complementary psychophysiological mechanisms.

Table 1

Test Statistic Value, Statistical Significance, and Effect Size of the Posthoc Pair-Wise Comparisons for the SNR-50 Scores Across the Masker Conditions

Comparisons		Speech babble			Reverse babble		
		t	*p*	Cohen's *d*	*t*	*p*	Cohen's *d*
2 speaker	3 speaker	0.75	1.000	0.19	0.00	1.000	0.00
	4 speaker	4.42	.014	1.10	-0.43	1.000	-0.11
	5 speaker	8.78	< .001	2.20	1.15	1.000	0.29
	6 speaker	7.83	< .001	1.96	5.57	.002	1.39
	7 speaker	15.92	< .001	3.98	8.21	< .001	2.05
	8 speaker	19.71	< .001	4.93	8.79	< .001	2.20
	SSN	18.04	< .001	4.51	9.76	< .001	2.44
3 speaker	4 speaker	6.48	< .001	1.62	-0.36	1.000	-0.09
	5 speaker	8.20	< .001	2.05	1.09	1.000	0.27
	6 speaker	6.24	< .001	1.56	3.64	.087	0.91
	7 speaker	14.01	< .001	3.50	8.08	< .001	2.02
	8 speaker	20.01	< .001	5.00	9.21	< .001	2.30
	SSN	17.50	< .001	4.37	8.91	< .001	2.23
4 speaker	5 speaker	3.72	.057	0.93	1.40	1.000	0.35
	6 speaker	3.45	.100	0.86	4.65	.011	1.16
	7 speaker	9.93	< .001	2.48	9.54	< .001	2.39
	8 speaker	15.44	< .001	3.86	9.30	< .001	2.33
	SSN	15.00	< .001	3.75	10.75	< .001	2.69
5 speaker	6 speaker	1.46	1.000	0.37	3.77	.067	0.94
	7 speaker	7.68	< .001	1.92	11.07	< .001	2.77
	8 speaker	15.32	< .001	3.82	12.35	< .001	3.09
	SSN	13.55	< .001	3.39	12.26	< .001	3.06
6 speaker	7 speaker	4.99	.005	1.25	5.51	.002	1.38
	8 speaker	9.36	< .001	2.34	6.66	< .001	1.66
	SSN	10.50	< .001	2.63	6.90	< .001	1.73
7 speaker	8 speaker	3.18	.173	0.80	4.84	.008	1.21
	SSN	3.60	.074	0.90	5.26	.003	1.32
8 speaker	SSN	2.78	.391	0.70	1.17	1.000	0.29

Note. Significance value adjusted for multiple comparisons using Bonferroni's correction. The comparisons are made across the maskers with different number of speakers. SNR = signal-to-noise; SSN = speech-spectrum noise.

Therefore, we hypothesized that during the sentence recognition task of Experiment 1, different maskers could require different listening effort scores even when the SNR-50 scores are comparable. To test this hypothesis, and to supplement the results of Experiment 1, we measured the subjective efforts involved in listening to sentences under each of the masker conditions as used in Experiment 1. This was done by measuring the subjectively rated listening effort using the same sentences and maskers used in Experiment 1. Thus, the results can be easily interpreted in line with these results of Experiment 1. Furthermore, we also ensured that there was complete intelligibility of all the sentences used for the listening effort task. Subjective ratings are a valid and sensitive approach to observe the difficulties and effort involved in a listening task (Johnson et al., 2015; Letowski & Scharine, 2017). This is also ecologically valid, as listening difficulty contributes to long-term comprehension and operator fatigue faced under different masker conditions.

Method

Participants

Fifteen participants aged 20 to 29 years (*M* = 24 years), all of whom were not a part of Experiment 1, volunteered for Experiment 2. A different set of participants was chosen to avoid habituation or familiarity effects. Familiarity effects were expected in the second experiment because we used the same target sentences as were used in Experiment 1. Inclusion criteria were identical to those used in Experiment 1. All participants signed informed consent forms according to the bio-behavioural guidelines of the All India Institute of Speech and Hearing, Mysore.

Stimuli

The same 15 lists used in Experiment 1 were chosen for this experiment as well. Three target sentences, randomly chosen from each list, were selected for each of the masker conditions to ensure that the participants had enough samples to judge the listening effort. Participants listened to the sentences mixed in SB maskers (2- to 8-speakers), RB maskers (2- to 8-speakers), and SSN at 0 dB SNRs. The choice of the 0 dB SNR, instead of SNR-50, was based on the results from the first experiment and a pilot study where we observed that at 0 dB SNR, all the participants had 100% correct identification of sentences across all the masker conditions. The SNR corresponding to SNR-50 was not chosen because the audibility and intelligibility of the target sentences are compromised (only 50%). The SNR of 0 dB ensured complete audibility of the target sentences across all masker conditions, and the only variable that was varied was the information in the maskers. The sentences were presented in a pseudo-random order at the same intensity (70 dB SPL) as in the previous experiment.

Procedure

In a sound-proofed room, all participants were asked to listen to the sentences in the presence of the different maskers and rate the effort involved in understanding the target sentences. Experiment 2 was done on all participants on a different day. All participants were first given three to five practice sentences to gain familiarity with the task and rating scales. The sentences used for the practice trials were different from the ones used in the actual experiment. The rating was similar to the 7-point scale used by Krueger et al. (2017). Each of the seven categories was also assigned a number as in the Krueger et al. study. A *no effort* rating was given 1, *very little effort* was given 3, *little effort* was given 5, *moderate effort* was given 7, *considerable effort* was given 9, *significant effort* was given 11, and *extreme effort* was given 13. These numbers were not visible to the listeners and were used only for analysis.

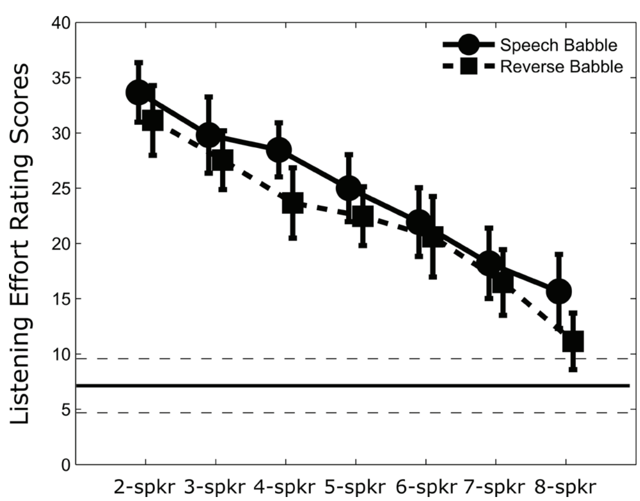
Once familiar with the procedure, each of the 45 sentence tokens was presented in a pseudo-random order, and the participants were asked to rate the effort required to perceive the target sentences. However, they were not required to repeat the sentences as it was already ensured that all stimuli had 100% identification at the SNR chosen. Once all stimuli were rated, the scores of the three sentences of each masker condition were added up to get a single *effort score* for each of the masker conditions.

Results

The statistical analyses were similar to those reported in Experiment 1. **Figure 3** shows the means and standard deviations of the listening effort rating scores across the different masker conditions. The rating scores, similar to the speech recognition scores of Experiment 1, showed a clear trend of decreasing listening effort with increasing number of speakers for both SB and RB conditions. The SSN required less effort than the two types of babble maskers. A two-way 2 (Masker Type: SB and RB) X 7 (Number of Speakers) RM-ANOVA was performed. The RM-ANOVA showed a significant main effect of both masker type, $F(1, 14) = 43.70$, $p < .001$, $\eta_p^2 = .76$, and number of speakers, $F(2.03, 28.51) = 168.26$, $p < .001$, $\eta_p^2 = .92$, on listening effort as well as a significant interaction between the two, $F(6, 84) = 3.59$, $p = .003$, $\eta_p^2 = .20$.

Further, separate one-way RM-ANOVAs were carried out on the seven speaker conditions within both SB and RB masker types. The SSN condition was also included in the ANOVA models. One-way RM-ANOVAs (corrected for violations of sphericity assumptions, wherever necessary) revealed significant main effects of the number of speakers for both the SB maskers, $F(2.64, 37.05) = 183.30$, $p < .001$, $\eta_p^2 = .93$, and RB maskers, $F(3.26, 45.64) = 153.20$, $p < .001$, $\eta_p^2 = .92$. Additional pairwise comparisons (adjusted for multiple comparisons using Bonferroni's correction) for each masker condition (SB or RB), showed that the babble maskers with fewer speakers required significantly greater effort than babble maskers with more speakers. All comparisons were significant ($p < .05$) except for those between the 3- and 4-speaker SB masker conditions. Similarly, in the RB masker conditions, significantly greater efforts were required for maskers with fewer speakers than for those with more speakers. This was true for most comparisons except for those between 4- vs. 5-speaker ($p = 1$), 5- vs. 6-speaker ($p = .216$), and 4- vs. 6-speaker ($p = .311$) conditions. **Table 2** summarizes the pairwise comparisons for both SB and RB maskers, including the SSN.

Figure 3



Means and standard deviations of the listening effort rating scores across the different masker conditions (number of speakers in the babble). The circles represent the effort scores for the speech babble maskers, the squares for the reverse babble maskers. The solid line at the bottom indicates the mean listening effort rating scores for the speech-spectrum noise masker, and the dashed lines show the standard deviations for the same. Spkr = Speaker.

We also compared the listening effort ratings between the corresponding number-of-speaker conditions between the SB and RB maskers. Paired samples *t* tests for the corresponding speaker conditions revealed significant differences (SB maskers required greater effort than RB maskers) between the listening effort when listening to the SB and RB maskers for all conditions ($p < .05$) except the comparisons for the 6-speaker ($p = .136$) and 7-speaker ($p = .072$) conditions.

Overall, the results from Experiment 2 suggest that the listening effort rating was influenced by the amount of linguistic information prevalent in the masker. Greater efforts were needed when there were fewer speakers in the masker. This was true for both SB and RB maskers. Most often, SB maskers required greater listening effort than RB maskers. The SSN masker was rated as requiring the least effort.

Discussion

The present study reports two experiments that measure the effect of linguistic content in a masker on the (a) sentence recognition accuracy and (b) listening effort involved in the perception of those sentences. The amount of linguistic information in a masker was varied by increasing the number of speakers in the masker (from two to eight) as well as by time reversal. A steady-state SSN masker, with nearly identical spectral information as that of the 8-speaker SB masker, was also considered as the control (energetic) masker condition.

Speech Recognition Performance Under Different Maskers

In Experiment 1, SNR-50 scores improved as a function of the number of speakers in the masker for both SB and RB maskers. SNR-50 scores became similar to that of the SSN when the number of speakers in the babble maskers reached eight. Other investigators have also reported improvements in speech recognition scores when the number of speakers in the babble increase (Boulenger et al., 2010; Hoen et al., 2007; Van Engen & Bradlow, 2007). When there are fewer speakers in the masker, the masker contains more recognizable linguistic information (phoneme identity and lexical items). This results in greater competition for attention between the target speech and babble masker. Because of the higher competition for the "limited" amount of attention resources, the SNR-50 scores are likely to be affected. Increasing the number of speakers results in an acoustically dense background, thus reducing the access to the linguistic information from the background (particularly the lexical-semantic information).

Results also showed that SB maskers with two and three speakers caused significantly higher masking (approximately 1.5 dB of SNR loss) compared to the corresponding RB maskers. The poorer performance in the SB masker condition is likely due to the additional presence of lexical-semantic information, despite similar spectral information (see **Figure 1**). Freyman et al. (2001) also reported significantly lesser masking effects by the time-reversed babble. In a prose recall task, Bell et al. (2008) further suggested that the amount of disruption depended on the semantic properties contained in the irrelevant

Table 2

Test Statistic Value, Statistical Significance, and Effect Size of the Posthoc Pair-Wise Comparisons for the Listening Effort Rating Scores Across the Different Masker Conditions

Comparisons		Speech babble			Reverse babble		
		t	*p*	Cohen's *d*	*t*	*p*	Cohen's *d*
2 speaker	3 speaker	6.12	< .001	1.58	5.08	.005	1.31
	4 speaker	7.45	< .001	1.92	10.43	< .001	2.69
	5 speaker	9.13	< .001	2.36	8.77	< .001	2.26
	6 speaker	11.82	< .001	3.05	9.19	< .001	2.37
	7 speaker	13.70	< .001	3.54	12.08	< .001	3.12
	8 speaker	14.77	< .001	3.81	19.72	< .001	5.09
	SSN	22.51	< .001	5.81	20.22	< .001	5.22
3 speaker	4 speaker	1.54	1.000	0.40	8.47	< .001	2.19
	5 speaker	4.58	.015	1.18	6.14	< .001	1.59
	6 speaker	6.87	< .001	1.77	5.93	.001	1.53
	7 speaker	10.16	< .001	2.62	9.74	< .001	2.52
	8 speaker	10.09	< .001	2.61	17.08	< .001	4.41
	SSN	17.00	< .001	4.39	20.82	< .001	5.38
4 speaker	5 speaker	7.60	< .001	1.96	1.66	1.000	0.43
	6 speaker	13.16	< .001	3.34	2.92	.311	0.76
	7 speaker	15.27	< .001	3.94	6.76	< .001	1.75
	8 speaker	17.05	< .001	4.40	12.98	< .001	3.35
	SSN	22.01	< .001	5.68	15.88	< .001	4.10
5 speaker	6 speaker	9.28	< .001	2.40	3.11	.216	0.80
	7 speaker	11.74	< .001	3.03	9.72	< .001	2.51
	8 speaker	16.24	< .001	4.19	15.71	< .001	4.06
	SSN	18.14	< .001	4.68	15.51	< .001	4.00
6 speaker	7 speaker	7.90	< .001	2.04	6.00	.001	1.55
	8 speaker	14.55	< .001	3.76	10.99	< .001	2.84
	SSN	15.54	< .001	4.01	11.59	< .001	2.99
7 speaker	8 speaker	4.01	.046	1.04	8.79	< .001	2.27
	SSN	12.13	< .001	3.13	10.52	< .001	2.72
8 speaker	SSN	8.84	< .001	2.282	5.68	.002	1.47

Note. Significance value adjusted for multiple comparisons using Bonferroni's correction. The comparisons are made across the maskers with different number of speakers. SSN = speech-spectrum noise.

speech. The SB maskers cause informational masking at both phonetic as well as lexical levels, whereas the RB maskers only cause informational masking via their phonetic information (Rhebergen et al., 2005). Hence, the lexical-semantic information of the SB maskers (2- and 3-speaker conditions) causes the additional 1.5 dB of masking.

Interestingly, in our study, speech recognition scores were quite similar across the two babble maskers when the number of speakers in the babble was four and above. This effect appears to be similar for both SB and RB maskers. Hence both the SB and RB maskers result in similar masking effects above the 4-speaker condition. Further, any residual effects of the lexical-semantic information available in the SB maskers seem to be offset by the unusual nature of the RB masker along with their excessive forward masking effects. However, it should be noted that the effort involved in performing the tasks under the 4- to 8-speaker conditions are not necessarily similar.

Another interesting finding from the study is that the speech recognition performances with 7- and 8-speaker conditions, for both RB and SB maskers, were like that of the SSN condition (see Figure 2). This indicates that as the number of speakers in the masker approaches eight, the salience of both semantic and phonetic information in the acoustically dense babbles reduce. Hence, there is minimal additional masking present, and the nature of the masking effect observed is predominantly energetic. However, other studies do show that with eight speakers or even slightly higher numbers of speakers in the babble, the performance does not reach the level of noise masker. Simpson and Cooke (2005) showed that consonant identification does not reach the levels of noise (only energetic masking) until the babble had at least 16 speakers. Although we did observe that the listening effort was significantly lower for the SSN maskers compared to the 8-speaker conditions of both SB and RB maskers, this finding needs to be studied in further detail.

Listening Effort Rating Under Different Maskers

Experiment 2 examined the effort required to listen to sentences presented in the 15 masker backgrounds. Similar to the results of Experiment 1, the maskers with fewer speakers required greater listening effort. Also, listening in the presence of SB maskers was significantly more effortful than the RB masker. The results of Experiment 2, therefore, supplement the results of Experiment 1.

A competing/irrelevant signal causes listeners to employ extra effort in perceiving the target signals (Baddeley, 2000; Ellermeier et al., 2015; Ellermeier & Zimmer, 2014; Li et al., 2004; Neath, 2000; Schneider et al., 2007). Our results showed that when the audibility and intelligibility of the target sentences were accounted for, the listening effort was significantly greater for the maskers with fewer speakers than in speaker conditions with a higher number of speakers. This indicates that the listening effort is positively related to the overall linguistic information present in the maskers.

Similar to the results of Experiment 1, Experiment 2 showed that the 2- and 3-speaker SB conditions were the most difficult with the 2-speaker SB being rated as the most effortful condition to perceive. Also, SB maskers required significantly greater listening effort than RB maskers. As explained earlier, the 2- and 3-speaker SB maskers contain robust lexical-semantic information. These meaningful, yet irrelevant, sentences often distort the target at the level of the phonological store of the working memory system. Our previous study (see Basavanahalli Jagadeesh & Kumar, 2019) further reinforces these assumptions. Therefore, significantly greater effort needs to be expended in parsing the target sentence from the background. Additionally, studies have shown that an irrelevant and dynamic background, like speech, with clear audibility and meaning, causes a switch in attention away from the target sound (Neath, 2000). Greater effort, therefore, will also be required to bring the focus back to the target signal.

In Experiment 2, there were significant differences in the listening effort rating scores between the 7- and 8-speaker conditions as well as the SSN. This indicates that even though the SNR-50 scores were similar across the SSNs and the 7- and 8-speaker conditions, the effort required in successfully performing the tasks varied significantly. This points towards the complementary nature of the two metrics. It also appears that the SNR-50 reveals the gross differences between the different masker conditions, while the listening effort reveals the subtler differences between the maskers. Variations in listening effort despite comparable performances in the speech perception tasks have already been demonstrated (Sarampalis et al., 2009). Two possible assumptions could explain this result. It is likely that the 7- and 8-speaker conditions still contain small but noticeable amounts of linguistic information in them. While this may not be enough to cause reductions in speech perception performance, it is still more effortful (or possibly annoying) to parse the targets from the background.

Further, the presence of slightly more evident modulations in the babble maskers (compared to noise) could cause greater modulation masking than the noise maskers. *Modulation masking* refers to the masking that occurs due to the inherent modulations present in the maskers (Stone & Canavan, 2016; Stone et al., 2012; Stone & Moore, 2002). In contrast to the concept of *dip-listening* where the presence of modulations in the masker leads to better speech perception, modulation masking results in a reduction in speech recognition performance because of the modulations present in the noises. The presence of spurious modulations in the noise can cause greater masking effects than a steady noise. We hypothesize that since these spurious modulations are more robust in the 8-speaker maskers than the SSN, it is likely to cause greater masking effects.

Other Factors That Contribute to Informational Masking

Multiple factors, other than the linguistics of the target and masker, also contribute to speech-on-speech masking. These include onset-related cues, spatial separation, the perceptual similarity between the target and maskers (gender and/or fundamental frequency), intensity of the two competing speech signals, and attention (Bregman, 2009). When the target and masker have significant overlap with respect to these factors, the masking effect will be more substantial.

However, we believe that we have accounted for most of these factors in our choice of stimuli and methodology. We accounted for any effects of spatial cues by presenting both the target and masker diotically. Gender and fundamental frequency related cues were accounted for, to a certain degree, as both the stimuli and the maskers were spoken by young adult women. Dialectal or accent-related variations were also largely controlled since such differences are shown to aid in more effective parsing of the target stimulus (Freyman et al., 2001). Stimulus (target) onset cues were also largely controlled by ensuring that the onset of the target was 0.5 s after the onset of the masker. The maskers themselves were selected from different random sections of the different babble maskers for each sentence, further reducing any possible onset related cues. We have also discussed the potential role of attention while explaining the confusions expected while parsing target and stimulus in a low-speaker babble condition.

A complex element associated with speech-on-speech masking, however, is the possibility of dip listening. Dip listening refers to the glimpsing of target signals during the momentary drops in the maskers' levels (Miller & Licklider, 1950). When the masker contains only two speakers (2SB or 2RB), the occurrences of dips in the maskers are more. This should, theoretically, lead to greater chances for the target stimuli to be heard and processed. Yet, the 2SB masker condition suffered the greatest deteriorations, both with respect to speech perception and listening effort. Rennies et al. (2019) further supposed that, despite the extra information available during these glimpses, it takes significantly greater effort to reassemble the target from the glimpses into a stream of coherent speech.

Another perspective of the same glimpsing phenomenon, however, could be related to the presence of glimpses in the target sentences themselves. Dips in the target sentences can lead to a clearer perception of the babble maskers. The perception of the babble, in turn, leads to greater masking effects. Additionally, as soon as the babble masker's speech is detected, the central executive of the working memory system is likely to immediately shift attention to the babble in an attempt to make sense of the information in the babble. Again, if the masker has significant lexical-semantic information, the attention is likely to be sustained longer on the babble maskers than when the babble maskers have no semantic information (Vachon et al., 2017). This counteracting effect of dips in the babble maskers and the target sentences do not appear to cancel each other. Evidently, the possible benefits of dip-listening are more than offset by the significant informational masking that is created in such a linguistically confusing listening scenario.

The glimpsing phenomenon is also pertinent in the conditions where the number of speakers in the maskers is high. The higher number of speakers in the babble maskers tend to fill in the dips, causing greater energetic masking (Freyman et al., 2004). However, it appears that increased energetic masking does not compensate for the loss of linguistic information in the masker. Brungart (2001) also argued that, overall, the informational masking tends to dominate the energetic masking in the overall masking effect. While we acknowledge the contribution of the other factors to the speech-on-speech masking scenario, it seems that the contribution of the informational component to the overall masking is significantly greater than the energetic component.

Conclusion

The results of our study show that the amount of informational masking was related to the robustness of the linguistic information present in the maskers. Varying the linguistic information in the maskers manifested as an increase in the listening effort as well as a reduction in performance in the sentence recognition task. Greatest masking effects and listening efforts were observed for babble maskers with fewer speakers while the SSN elicited the least masking effects and listening effort scores. We also recognize that factors other than just the linguistic information contribute to the overall masking. Furthermore, the use of ecologically valid maskers such as the speech babble and the time-reversed babble could give a realistic idea of the problems faced by listeners with compromised auditory functions (e.g., ageing and/or hearing loss) in identifying speech in the background of speech. However, the interpretation of any speech perception task could depend on the type of target stimuli used. Hence, future research directed towards understanding informational masking effects for different types of stimuli is essential and necessary. Furthermore, we recommend the use of the babble maskers (both forward and reversed) in clinical settings along with the noise maskers to better simulate the problems faced by listeners, particularly the elderly and hearing impaired, in their daily lives.

References

American National Standard Institute. (2003). *Maximum permissible ambient noise levels for audiometric test rooms* (Rev. ed). (ANSI S3.1-1999).

Arai, T. (2010). Masking speech with its time-reversed signal. *Acoustical Science and Technology, 31*(2), 188–190. https://doi.org/10.1250/ast.31.188

Baddeley, A. D. (2000). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin and Review, 7*(3), 544–549. https://doi.org/10.3758/BF03214369

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Balakrishnan, U., & Freyman, R. L. (2008). Speech detection in spatial and nonspatial speech maskers. *The Journal of the Acoustical Society of America, 123*(5), 2680–2691. https://doi.org/10.1121/1.2902176

Basavanahalli Jagadeesh, A., & Kumar, A. U. (2019). Effect of informational masking on auditory working memory: Role of linguistic information in the maskers. *Hearing, Balance and Communication, 17*(4), 270–279. https://doi.org/10.1080/21695717.2019.1630980

Bell, R., Buchner, A., & Mund, I. (2008). Age-related differences in irrelevant-speech effects. *Psychology and Aging, 23*(2), 377–391. https://doi.org/10.1037/0882-7974.23.2.377

Boulenger, V., Hoen, M., Ferragne, E., Pellegrino, F., & Meunier, F. (2010). Real-time lexical competitions during speech-in-speech comprehension. *Speech Communication, 52*(3), 246–253. https://doi.org/10.1016/j.specom.2009.11.002

Bregman, A. S. (2009). Auditory scene analysis. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 729–736). Elsevier. https://doi.org/10.1016/B978-008045046-9.00246-1

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica, 86*(1), 117–128.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., Iyer, N., Thompson, E. R., Simpson, B. D., Gordon-Salant, S., Schurman, J., Vogel, C., & Grant, K. (2013). Interactions between listening effort and masker type on the energetic and informational masking of speech stimuli. *Proceedings of Meetings on Acoustics, 19*(1), 1–9. https://doi.org/10.1121/1.4800033

Brungart, D. S., & Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty. *The Journal of the Acoustical Society of America, 115*(1), 301–310. https://doi.org/10.1121/1.1628683

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America, 110*(5), 2527–2538. https://doi.org/10.1121/1.1408946

Carhart, R., Johnson, C., & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. *The Journal of the Acoustical Society of America, 58*(S1), S35. https://doi.org/10.1121/1.2002082

Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K., & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America, 138*(3), 1561–1569. https://doi.org/10.1121/1.4928954

Ellermeier, W., & Zimmer, K. (2014). The psychoacoustics of the irrelevant sound effect. *Acoustical Science and Technology, 35*(1), 10–16. https://doi.org/10.1250/ast.35.10

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America, 88*(4), 1725–1736. https://doi.org/10.1121/1.400247

Finney, D. J. (1952). *Probit analysis: A statistical treatment of the Sigmoid response curve* (2nd ed.). Cambridge University Press.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America, 109*(5 Pt. 1), 2112–2122. https://doi.org/10.1121/1.1354984

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America, 115*(5 Pt.1), 2246–2256. https://doi.org/10.1121/1.1689343

Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing, 21*, 1–25. https://doi.org/10.1177/2331216516687287

Geetha, C., Kumar, K. S. S., Manjula, P., & Pavan, M. (2014). Development and standardisation of the sentence identification test in the Kannada language. *Journal of Hearing Science, 4*(1), 18–26. https://doi.org/10.17430/890267

Gnanateja, G. N. (2012). Speech in noise mixing, signal to noise ratio [Computer software]. Retrieved from http://www.mathworks.com/matlabcentral/fileexchange/37842-speech-in-noise-mixing-signal-to-noise-ratio

Gnanateja, G. N. (2016). Speech spectrum shaped noise [Computer software]. Retrieved from https://www.mathworks.com/matlabcentral/fileexchange/55701-speech-spectrum-shaped-noise

Gosselin, P. A., & Gagné, J.-P. (2010). Use of a dual-task paradigm to measure listening effort. *Canadian Journal of Speech-Language Pathology and Audiology, 34*(1), 43–51.

Hafter, E. R., & Schlauch, R. S. (1989). Factors in detection under uncertainty. *Journal of the Acoustical Society of America, 86*(S1), 12. https://doi.org/10.1121/1.2027300

Hoen, M., Meunier, F., Grataloup, C.-L., Pellegrino, F., Grimault, N., Perrin, F., Perrot, X., & Collet, L. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Communication, 49*(12), 905–916. https://doi.org/10.1016/j.specom.2007.05.008

Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *International Journal of Audiology, 49*(12), 928–932. https://doi.org/10.3109/14992027.2010.520036

JASP Team. (2016). JASP (Version 0.7.5.6) [Computer software]. https://jasp-stats.org/

Jin, S., & Nelson, P. B. (2006). Speech perception in gated noise: The effects of temporal resolution. *The Journal of the Acoustical Society of America, 119*(5 Pt 1), 3097–3108. https://doi.org/10.1121/1.2188688

Johnson, J., Xu, J., Cox, R., & Pendergraft, P. (2015). A comparison of two methods for measuring listening effort as part of an audiologic test battery. *American Journal of Audiology, 24*(3), 419–431. https://doi.org/10.1044/2015_AJA-14-0058

Kidd, G., Jr., & Colburn, H. S. (2017). Informational masking in speech recognition. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The auditory system at the cocktail party, Springer handbook of auditory research* (pp. 75–109). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources, Springer handbook of auditory research* (pp. 143–189). Springer International Publishing. https://doi.org/10.1007/978-0-387-71305-2

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing, 33*(2), 291–300. https://doi.org/10.1097/AUD.0b013e3182310019

Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., & Holube, I. (2017). Relation between listening effort and speech intelligibility in noise. *American Journal of Audiology, 26*(3S), 378–392. https://doi.org/10.1044/2017_AJA-16-0136

Leek, M. R., Brown, M. E., & Dorman, M. F. (1991). Informational masking and auditory attention. *Perception & Psychophysics, 50*(3), 205–214. https://doi.org/10.3758/BF03206743

Letowski, T. R., & Scharine, A. A. (2017). *Correlational analysis of speech intelligibility tests and metrics for speech transmission.* US Army Research Laboratory (No. ARL-TR-8227). https://doi.org/10.13140/RG.2.2.26581.93921

Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance, 30*(6), 1077–1091. https://doi.org/10.1037/0096-1523.30.6.1077

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America, 124*(5), 3261–3275. https://doi.org/10.1121/1.2990705

Lu, Z., Daneman, M., & Schneider, B. A. (2016). Does increasing the intelligibility of a competing sound source interfere more with speech comprehension in older adults than it does in younger adults? *Attention, Perception, & Psychophysics, 78*(8), 2655–2677. https://doi.org/10.3758/s13414-016-1193-5

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243. https://doi.org/10.1016/j.cogpsych.2009.04.001

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America, 22*(2), 167–173. https://doi.org/10.1121/1.1906584

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7*(3), 403–423. https://doi.org/10.3758/BF03214356

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics, 41*(5), 409–415. https://doi.org/10.3758/bf03203033

Peelle, J. E. (2017). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing, 39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L, & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing, 37*(S1), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Rennies, J., Best, V., Roverud, E., & Kidd, G., Jr. (2019). Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and perceived listening effort. *Trends in Hearing, 23*, 1–21. https://doi.org/10.1177/2331216519854597

Rennies, J., Schepker, H., Holube, I., & Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America, 136*(5), 2642–2653. https://doi.org/10.1121/1.4897398

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, 118(3 Pt 1), 1274–1277. https://doi.org/10.1121/1.2000751

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7*(31), 1–17. https://doi.org/10.3389/fnsys.2013.00031

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America, 133*(4), 2431–2443. https://doi.org/10.1121/1.4794379

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology, 23*(8), 577–589. https://doi.org/10.3766/jaaa.23.7.7

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research, 52*(5), 1230–1240. https://doi.org/10.1044/1092-4388(2009/08-0111)

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18*(7), 559–572. https://doi.org/10.3766/jaaa.18.7.4

Simpson, S. A., & Cooke, M. (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of *N*. *The Journal of the Acoustical Society of America, 118*(5), 2775–2778. https://doi.org/10.1121/1.2062650

Stone, M. A., & Canavan, S. (2016). The near non-existence of "pure" energetic masking release for speech: Extension to spectro-temporal modulation and glimpsing. *The Journal of the Acoustical Society of America, 140*(2), 832–842. https://doi.org/10.1121/1.4960483

Stone, M. A., Füllgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America, 132*(1), 317–326. https://doi.org/10.1121/1.4725766

Stone, M. A., & Moore, B. C. J. (2002). Tolerable hearing aid delays. II. Estimation of limits imposed during speech production. *Ear and Hearing, 23*(4), 325–338. https://doi.org/10.1097/00003446-200208000-00008

Summers, V., & Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level. *Journal of Speech, Language, and Hearing Research, 47*(2), 245–256. https://doi.org/10.1044/1092-4388(2004/020)

Tillman, T. W., & Olsen, W. O. (1973). Speech audiometry. In J. Jerger (Ed.), *Modern developments in audiology* (2nd ed., pp. 37–74). Academic Press.

Vachon, F., Labonté, K., & Marsh, J. E. (2017). Attentional capture by deviant sounds: A noncontingent form of auditory distraction? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(4), 622–634. https://doi.org/10.1037/xlm0000330

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America, 121*(1), 519–526. https://doi.org/10.1121/1.2400666

Authors' Note

Correspondence concerning this article should be addressed to Anoop Basavanahalli Jagadeesh, Department of Audiology, All India Institute of Speech and Hearing, Mysuru - 570006, India. Email: **anoop2187@gmail.com.**

Acknowledgments

This research was funded by All India Institute of Speech and Hearing Research Fellowship (No. SH/ACA/Ph.D./Admissions/2016-17). Portions of this work were presented at the World Congress of Audiology, Cape Town, South Africa, October 2018. The authors acknowledge the Director, All India Institute of Speech and Hearing, Mysore, for providing the research infrastructure.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.