Proceedings of Meetings on Acoustics

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 3aSCb: Components of Informational Masking

3aSCb4.   Interactions between listening effort and masker type on the energetic and informational masking of speech stimuli

Douglas Brungart*, Nandini Iyer, Eric R. Thompson, Brian D. Simpson, Sandra Gordon-Salant, Jaclyn Schurman, Chelsea Vogel and Kenneth Grant

 *Corresponding author's address: Audiology and Speech Center, Walter Reed National Military Medical Center, Bethesda, MD 20889, douglas.brungart@med.navy.mil

  In most cases, normal-hearing listeners perform better when a target speech signal is masked by a single irrelevant speech masker than they do with a noise masker at an equivalent signal-to-noise ratio (SNR). However, this relative advantage for segregating target speech from a speech masker versus a noise masker may not come without a cost: segregating speech from speech may require the allocation of additional cognitive resources that are not required to segregate speech from noise. The cognitive resources required to extract a target speech signal from different backgrounds can be assessed by varying the complexity of the listening task. Examples include: 1) contrasting the difference between the detection of a speech signal and the correct identification of its contents; 2) contrasting the difference between single-task diotic and dual-task dichotic listening tasks; and 3) contrasting the difference between standard listening tasks and one-back tasks where listeners must keep one response in memory during each stimulus presentation . By examining performance with different kinds of maskers in tasks with different levels of complexity, we can start to determine the impact that the informational and energetic components of masking have on the listening effort required to understand speech in complex environments.

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

In principle, the task of segregating a target speech signal from an interfering speech masker should be more difficult than the task of segregating a target speech signal from a noise masker. This increased difficulty stems from the fact that the interfering speech masker requires the listener not only to detect the acoustic and phonetic element of the target speech in the presence of an interfering sound, but also to segregate those detected elements from the acoustic and phonetic elements of a masking signal that also might plausibly have come from the target talker. We often refer to the increased difficulty in speech perception that occurs because the listener may have difficulty segregating the acoustic or phonetic elements of a target speech signal from those of a similar-sounding speech masker as "informational masking." [1]. This type of interference is distinguished from "energetic masking," which occurs when an acoustic signal is rendered undetectable by the presence of an interfering sound that occurs in the same time period and in the same frequency channel as the target speech.

Because speech maskers are much more likely to produce informational masking effects than noise maskers, one might expect listeners in general to perform worse in speech perception tasks that involve speech maskers than in those that involve modulated noise maskers that produce a comparable amount of energetic masking. And there are definitely examples of cases where the target and masking speech signals are contextually similar where listeners perform much better with modulated noise maskers than with interfering speech. One example of this is the Coordinate Response Measure (CRM), which requires listeners to extract a target speech signal from a mixture of speech signals that only differ from the target in terms of the inclusion of a different call sign [1]. However, in most cases where the target speech signal is masked by an irrelevant speech masker, listeners actually perform slightly better with a speech masker than they do with a modulated noise masker that produces an equivalent amount of energetic masking [2, 3, 4]. This raises the question of how listeners are able to overcome the additional challenge of segregating speech from an irrelevant speech masker without suffering any apparent decrement in their overall speech perception abilities.

The fact that there is no apparent decrement in performance in listening tasks involving irrelevant speech maskers rather than modulated noise maskers does not necessarily imply that there are no additional costs related to the informational masking effects that may occur in these tasks. It may be the case that the surprisingly high levels of performance that listeners are able to achieve with irrelevant speech maskers require the allocation of additional cognitive resources (attention, working memory, etc.) that may not be required for the theoretically simpler task of extracting speech from a modulated noise masker. In other words, it is possible that listeners are able to achieve high levels of performance in listening tasks that involve irrelevant speech maskers, but that these performance levels are only achievable when the listeners are able to devote a great deal of listening effort to the task [5]. If this is true, then one might expect the advantages seen for irrelevant speech maskers in simple speech perception tasks (i.e. those that simply require listeners to attend to a stimulus and repeat back the contents of the speech message) to be reduced or eliminated in more complicated tasks that require listeners to monitor multiple speech signals or to store the contents of the speech messages in working memory.

In this brief paper, we describe the results of three experiments that show that different types of informational and energetic maskers can produce very different levels of relative performance when the complexity of the listening task increases from the very simple (i.e. identify the presence or absence of the target talker) to the very complex (i.e. listen and respond to two simultaneous speech signals or hold the contents of each message in memory and respond to the sentence that was heard prior to the most recent stimulus presentation). The results are discussed in terms of their potential applicability to the design of improved hearing aid algorithms for hearing impaired listeners.

EXPERIMENT 1: DETECTION, GENDER ID, AND COMPREHENSION IN A DICHOTIC SPEECH TASK

The purpose of Experiment 1 was to evaluate how systematic increases in the complexity of a listening task influenced the relative levels of performance listeners were able to achieve with speech and noise maskers. The listening tasks were all based the Coordinate Response Measure (CRM) corpus, which consists of phrases containing one of eight call signs, one of eight numbers, and one of four colors, combined into a carrier phrase of the form "Ready <call sign> go to <color> <number> now." [1]. The

experiment used a 2-AFC paradigm that required listeners to either a) listen to two sequential stimuli, one containing a target phrase from the CRM corpus and one containing only a masker, and identify which one contained the target phrase (detection task); b) listen to two sequential stimuli, one containing a normal CRM phrase one containing a time-reversed CRM phrase, and determine which one contained the non-reversed CRM phrase (discrimination task); or c) listen to two sequential stimuli containing phrases from the CRM corpus and determine which one contained the color-number combination displayed on the computer screen prior to the start of the trial (identification task).

Each of these tasks was tested with two types of maskers. The speech masker consisted of a randomly-selected CRM phrase that was spoken by a talker of the same sex as the target talker but was selected to have a different call sign, color and number than the target phrase. The noise masker was generated by randomly selecting a speech masker and then spectrally shaping a Gaussian noise to match the magnitude spectrum that speech masker. In all cases, different maskers were randomly selected for use in each of the two intervals of the 2-AFC tasks.

The signal-to-noise ratios (SNRs) used in the experiment ranged from -56 dB to 8 dB in 4 dB steps. Only the SNR was varied within a block of trials: the type of task (Detection, Discrimination or Identification) and noise type remained fixed within a 30-trial block. A total of 12 paid listeners (ranging in age from 19-25), all with normal audiometric thresholds, participated in the experiment. On average, the listeners participated in 20 trials for each combination of SNR, masker type, and listening condition.
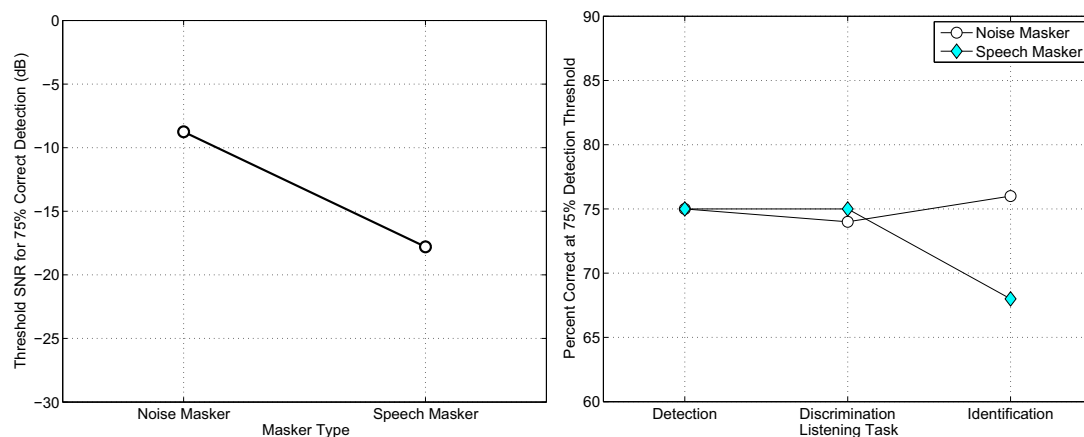

FIGURE 1: Results from Experiment 1. The left panel shows the Threshold SNR required for 75% correct performance in the detection task with each type of masker. The right panel shows the percent correct responses in each listening task when the SNR was set to the 75% correct detection threshold with each masker type. These values are interpolated from the logistic fits of the psychometric curves for each task.

Results

Figure1 shows the results of Experiment 1. The left panel of the figure shows the threshold SNR values for 75% correct responses in the identification task with each type of masker. These data indicate that listeners were able to detect the presence of a target phrase at a 10-dB lower SNRs with the speech masker than with the noise masker (-18 dB versus -8 dB).

The data in the right panel of the figure show how performance varied as a function of task complexity when the SNR was fixed at the threshold level required to obtain 75% correct responses in the detection task. For both the speech and noise maskers, there was no degradation in performance with the task changed from a task requiring listeners to detect the presence of a low-level speech signal in the stimulus to one that required them to discriminate between to intervals and identify the one that did not contain the time-reversed target phrase. However, when the task was changed to the highest-complexity "identification" task, which required listeners to identify the interval that contained a particular color-number combination, there was a substantial degradation in performance with the "speech" masker, but not with the "noise" masker. This result suggests that the lower detection thresholds that listeners are able to achieve with speech maskers may require listeners to deploy additional cognitive resources that are not required for the detection of a speech signal in noise. When the task complexity increases,

these cognitive resources may need to be redeployed, and this may lead to a relative decrease in the level of performance listeners are able to achieve with a speech masker. In other words, listeners are able to perform better overall in tasks with speech maskers than they can in tasks with continous noise maskers, but they are working harder to achieve this level of performance.

EXPERIMENT 2: EFFECT OF MASKER TYPE AND TASK COMPLEXITY ON SPEECH PERCEPTION IN MONAURAL AND DICHOTIC LISTENING TASKS

Experiment 2 was designed to examine the impact that different types of informational and energetic maskers would have on tasks that required listeners to monitor more than one channel of simultaneous speech. As in Experiment 1, the different tasks were based on the Coordinate Response Measure (CRM) task, which requires listeners to attend to a multi-talker speech signal and respond with the color and number contained in a target phrase that is addressed to a pre-determined call sign. In this particular experiment, the level of complexity in the CRM task was systematically manipulated by having listeners perform one of four different closely related tasks:

- **Monaural:** A stimulus containing a CRM phase with the call sign "Baron" was mixed with one of five different types of maskers (described below) and presented monaurally to one of the listener's ears. The presentation ear (left or right) was chosen randomly prior to each block of trials, and it was identified to the listener in written instructions provided prior to the start of the block. The listener's task was to use a mouse to select the color and number contained in the CRM phrase from a colored grid displayed on the computer screen after each trial.

- **Target in Known Ear:** A stimulus containing a CRM phase with the call sign "Baron" was presented to one ear, and a second stimulus containing a CRM phrase with the call sign "Charlie" was presented in the second ear. These two phrases were always spoken by two different talkers of the same sex, and the sex of these two talkers was randomly selected from trial to trial. Each of the two CRM phrases was mixed with a different sample of the same type of masker, and the levels of the maskers were adjusted so the CRM phrases were presented at the same SNRs in both ears. The locations of the "Baron" and "Charlie" call signs (left or right) were fixed in each block of trials, and the location of the "Baron" call sign was identified to the listener in the instructions provided prior to the start of the block. The listener's task was to ignore the ear with the call sign "Charlie" and use a mouse to select the color-number combination contained in the CRM phase addressed to the target call sign "Baron".

- **Target in Unknown Ear:** The stimuli in this condition were identical to those in the "Known Ear" condition, except that the locations of the "Baron" and "Charlie" call signs (left or right) were randomly selected in each trial. The listener's task was to listen to both CRM phrases, ignore the phrase addressed to the call sign "Charlie," and use a mouse to select the color-number contained in the CRM phase addressed to the target call sign "Baron".

- **Respond to Both Ears:** The stimuli in this condition were identical to those in the "Known Ear" condition. However, in this condition, the listeners were instructed to attend to both CRM phrases and to use a mouse to identify the color-number combination spoken by the "Baron" CRM phrase in one grid of colored numbers (titled "Baron") and to identify the color-number combination spoken by the "Charlie" CRM phrase in a second grid of colored numbers. These two responses could occur in any order.

Each of these four task conditions was tested with five different types of maskers:

- **Continuous Speech-Shaped Noise:** A speech-shaped noise with the same spectrum as the corresponding CRM phrase was generated by taking the FFT of the CRM phrase, setting its phase spectrum to match that of a random Gaussian noise of the same length, and taking the real part of the inverse FFT of the random-phase signal.

- **Irrelevant Speech:** An irrelevant speech signal was generated by selecting a random passage of speech read from the book "Wealth of Nations" by Adam Smith. This speech signal was always read by a different talker who was the same sex as the talkers used to generate the CRM phrases.

- **Time-Reversed Speech:** Identical to the irrelevant-speech condition, except that the speech masker was time-reversed prior to presentation to the listener.

- **Modulated Speech-Shaped Noise:** A modulated speech-shaped noise was generated by selecting a random passage of irrelevant speech (as described above), generating a speech-shaped noise with the same spectrum as this speech sample using the same inverse-FFT method used for the speech-shaped noise masker , and then modulating this speech-shaped noise with the envelope of the irrelevant speech passage. This envelope was calculated by convolving the absolute value of the irrelevant speech waveform with a 6.4 ms rectangular window.

- **Babble:** An irrelevant speech babble was generated by combining four different randomly-selected passages of speech read from the "Wealth of Nations", with two read by male talkers and two read by female talkers

A total of 12 paid volunteers participated in the experiment, including 5 males and 7 females. All had normal audiometric thresholds, and their ages ranged from 22 to 57 (Mean age 25 years). The trials were divided into blocks of 44 trials, which were divided across 19 SNRs ranging from -27 dB to +21 dB, plus a no-masker condition. There was some variation in how many blocks each subject completed in each condition, but each subject completed at least 269 blocks of trials in the experiment, with a minimum of 9 blocks in any one condition. Across all the listeners, a total of 182600 trials were collected in the experiment.
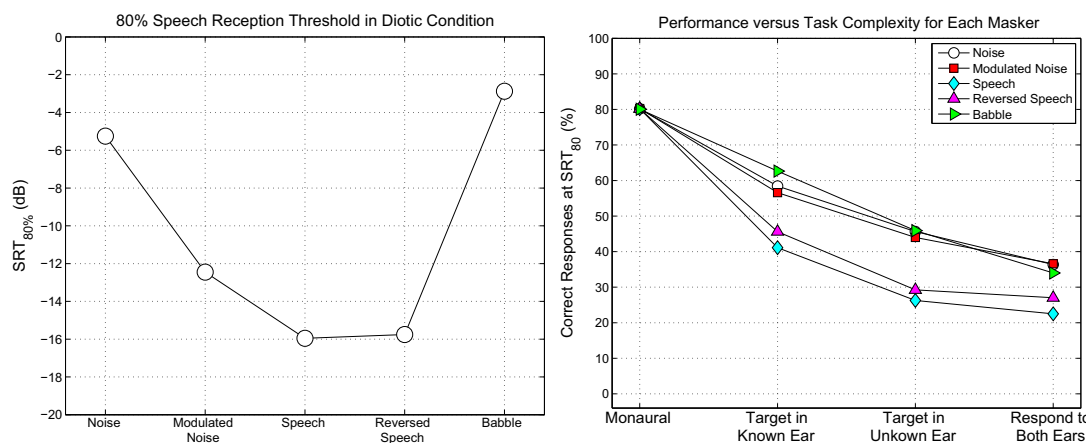
Results

FIGURE 2: Results from Experiment 2. The left panel shows the Speech Reception Threshold (SRT) required for 80% correct performance in the monaural task with each type of masker. The right panel shows the percent correct responses in each listening task when the SNR was set to the $SRT_{80\%}$ value in the monaural condition with each masker. These values are interpolated from the logistic fits of the psychometric curves for each task.

The left panel of Figure 2 shows the minimum SNR value required to obtain 80% correct responses in the Monaural condition with each type of masker. These values were obtained by generating a logistic fit to the overall psychometric functions obtained for masker type. These results show that overall performance was best with the Speech and Reversed speech masker (which had the lowest SRT_{80} values of roughly -16 dB), slightly worse with the Modulated Noise masker, and much worse with the Continuous Noise and Babble Maskers. These results are not surprising, as they are consistent with other studies that have examined speech perception performance with different types of maskers. The Speech, Reversed Speech and Modulated Noise maskers tend to have dips in the interfering waveform that allow listeners to extract relevant speech information even in cases where the SNR is very low, which leads to very low SRT_{80} values relative to the Continuous Noise and Babble maskers.

For the purposes of this paper, the more interesting results are shown in the right panel of Figure 2, which shows performance in the more complex tasks of the experiment when the SNR value with each masker type was fixed at the SNR value that produced 80% correct responses in the baseline Monaural condition. These results show that performance decreased for all the masker types as task complexity increased, but that this decrease was much greater for the Speech and Reversed Speech maskers than it was for the other three types of maskers.

These results provide important insights into the role that limited cognitive resources may play in extraction of target speech from informational and energetic maskers. The Noise and Babble maskers produce a relatively large amount of masking in the simple monaural task, as indicated by their relatively high SRT_{80} values, but they appeared to require relatively few cognitive resources to achieve optimal performance, as indicted by the relatively small decreases in performance that occurred when task complexity increased in those conditions. In contrast, the Speech and Reversed Speech maskers produced relatively small amounts of masking in the monaural condition, as indicated by their relatively low SRT_{80} values, but they appeared to require substantially more cognitive resources to achieve optimal performance than the noise or babble conditions, as indicted by the relatively large degradations in performance when the complexity of the listening task increased. Based on these four conditions, it might appear that there is a general trend for masking waveforms that produce relatively little masking in a simple monaural listening task (and thus allow listeners to achieve relatively high performance levels at very low SNRs) to require relatively more cognitive resources to achieve optimal performance (and thus produce relatively large decreases in performance with increasing task complexity). However, this general trend is clearly contradicted by the Modulated Noise masking condition, which produced relatively low amounts of masking in the monaural condition (with an SRT_{80} value similar to that of the Speech and Reverse Speech maskers), but was just as robust to increases in task complexity as the Noise and Babble Maskers. This implies that the relatively high levels of performance that listeners were able to achieve at low SNR values in the monaural condition with the Speech and Reversed Speech maskers were not simply the result of dip listening due to amplitude fluctuations in the envelopes of these signals, which also occurred in the Modulated Noise masking condition. Rather, it seems that the extraction of speech from a competing speech or reversed speech masker at a very low SNR value is only achievable if the listener is able to deploy additional cognitive resources to the listening task. When the task complexity is increased, these resources have to be redeployed to accommodate this increased task complexity, and performance with the speech or reversed speech maskers goes down. In other words, listeners are able to obtain a higher level of performance at lower SNRs with speech or reversed speech than they can with equivalent noise maskers, but they must expend additional listening effort in order to achieve this results. Presumably, these additional cognitive resources are required to cope with informational masking effects related to the extraction of the phonetic and acoustic elements of the target speech signal from those of a confusable speech-like masker[1].

A last intriguing question raised by these results is why the additional cognitive resources required to extract target speech from a single speech or reversed speech masker do not seem to be required for the extraction of speech from a babble masker. The reasons for this apparent discrepancy are not entirely clear, but they may be related to the fact that the babble masker required a much higher SNR than the other speech maskers to achieve 80% correct responses. At this SNR (-3 dB), the target talker was more intense than any of the other maskers in the stimulus. This may have allowed the listeners to adopt a relatively simple listening strategy (i.e., listen to the loudest talker) that did not require as many cognitive resources as attending to the quietest talker in a two-talker stimulus.

EXPERIMENT 3: EFFECT OF MASKER TYPE AND CONTEXTUAL CUES ON SPEECH PERCEPTION IN ONE-BACK LISTENING TASKS

Experiment 3 was designed to compare the relative impact that speech and noise maskers have on speech recognition performance under conditions where task complexity was increased by requiring the listeners to perform a secondary memory task (a one-back task) when making their verbal responses. In addition, sentences of varying contextual cues where used to determine if the ability to use contextual cues in the sentences has a differential effect on speech recognition with added memory demands and if this ability interacts with the type of masker. In order to evaluate the effects of contextual cues, a set of 200 R-SPIN sentences with highly probable contextual cues (HP sentences) [7] were recorded by one female speaker who is a native speaker of English. Within each of these R-SPIN sentences, three to seven words were selected to serve as keywords for scoring purposes. A second set of sentences with highly improbable contextual cues were created from the HP sentences by taking the same keywords and randomly ordering them into a new set of sentences with highly improbable contextual cues. These

---

[1]It is worth noting that other experiments with non-speech target signals have also shown that informational masking effects can be amplified in situations where listeners are required to divide their attention across multiple frequency regions in a monaural stimulus [6]

anomalous-probability (AP) sentences were grammatically and syntactically correct, but completely meaningless. For example, a typical AP sentence would be "His doctor drank a lost risk", whereas a typical HP sentence would be "His plans meant taking a big risk."

The maskers used in the experiment were generated by having the same female speaker who recorded the R-SPIN sentences record a series of selected passages from Grimm's "Fairy Tale Classics". Two masking conditions were tested. In the "speech" masking conditions, two 45s long passages of these fairy tale recordings were randomly selected and looped to play continuously through the course of each 20-trial block. In the "noise" masking conditions, each of these speech maskers was used to generate a speech-shaped noise with the same average long-term spectrum, and these two speech-shaped noise maskers were mixed together and looped to play continuously through the block of trials.

A total of seven paid volunteer listeners participated in the experiment, including five females and two males ranging from ages 22 to 26 years. All listeners had hearing sensitivity within normal limits. A measure of working memory, the Listening SPAN (LSPAN), was administered prior to the experimental testing to measure each subject's memory storage and processing abilities. Subjects were asked to listen to a group of sentences and determine if each sentence was true or false. The subject was also directed to remember the last word in each sentence to be repeated at the end of the group of presented sentences. As the LSPAN progressed, the number of sentences in each group increased, requiring the subject to remember an increasing number of final words. If a subject could not successfully complete the 2-sentence trial of the LSPAN, they would not be eligible to participate in the R-SPIN portion of the experiment. The R-SPIN HP and AP sentences were divided into 20 blocks of trials with 20 sentences in each block. Each subject participated in all 20 blocks. A total of four different conditions were tested: 1) HP sentences with a noise masker; 2) HP sentences with a babble masker; 3) AP sentences with a noise masker; 4) and AP sentences with a babble masker. These blocks of trials were presented in random order to each subject. For each condition, the first block of trials for each subject was a training condition, where the first sentence was presented at an SNR of 4 dB, and the listeners responded by immediately repeating back the sentence as soon as they heard it (a 0-back) response. After each trial, the SNR was adjusted by increasing the SNR 2 dB for each incorrect keyword and decreasing the SNR by 0.5 dB for each correct keyword. This tracking procedure was designed to select the SNR in order to ensure that the listeners were performing at an 80% correct performance level in the 0-back task for each of the listening conditions.

After this first training block, the other four blocks of trials in each condition were testing blocks in which the first ten sentences were 0-back trials where the listeners responded immediately to each stimulus presentation, and the last ten sentences were 1-back trials that required the subject to listen to two sequential speech signals, hold the contents of each message in memory, and respond to the sentence that was heard prior to the most recent stimulus presentation. In each of these testing blocks, the initial SNR was set to equal the final SNR of the previous block in the same listening condition, and the SNR on the first ten 0-back trials was adaptively tracked using the same procedure described above. Then the SNRs on the last ten 1-back trials of the testing block were selected to exactly match the SNRs used on the previous ten 0-back trials, without adaptive tracking (i.e. trials 11-20 used the same SNRs as trials 1-10). This means that 0-back and 1-back conditions were always tested at exactly the same set of SNR values.

Results

Figure 3 shows the results from Experiment 3. The left panel shows the SNR required to achieve 80% correct responses in the baseline 0-back trials of each condition. These SRT_{80} values were derived from the average SNR values that occurred across all the trials tested in the adaptive track after the first 20 training trials in each condition. In this experiment, it appears that the SRT_{80} of the 2-talker speech masker was roughly 2 dB higher than the SRT_{80} for the noise masker, which is not surprising because it contained two independent speech signals that were produced by the same talker as the target speech (thus providing a high probability of informational masking effects) and had independent amplitude fluctuations that likely filled in many of the gaps that would typically provide opportunities for dip listening in a single-talker masking waveform. It also appears that the low-context (anomalous) sentences required roughly a 4 dB higher SRT than the high-context sentences, regardless of masker type.

The right panel of Figure 3 shows the percent correct responses in the 0-back and 1-back trials in the test trials in each condition of the experiment. As expected, the 0-back conditions all produced roughly
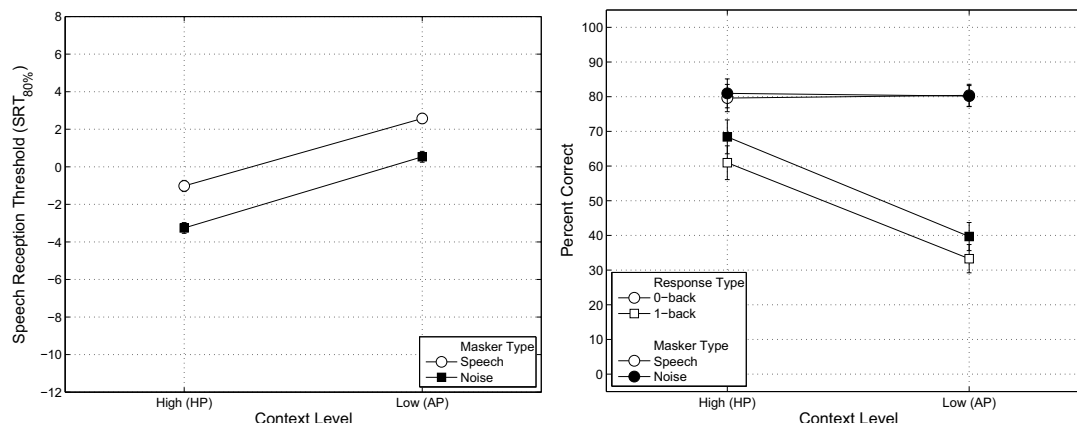
FIGURE 3: Results from Experiment 3. The left panel shows the Speech Reception Threshold (SRT) required for 80% correct performance in the 0-back with each type of masker. The right panel shows the percent correct responses in each listening task when the SNR was set to the $SRT_{80\%}$ value in the 0-back and 1-back conditions with each type of masker.

80% correct responses. The key question in this experiment, however, was how performance differed in the 1-back conditions where the memory component increased the amount of listening effort required for the task. From these results, it is clear that the 1-back task was much easier with the high-context sentences than it was for the low-context sentences. This is an interesting result, because it implies that the advantages of contextual cues are substantially underestimated even by the relatively large (4 dB) difference in SRT found in the 0-back listening tasks with the low- and high-context maskers (left panel of Figure 3). High-context speech material is both easier to understand and easier to remember over time, which means that listeners who are able to take advantage of these cues have a tremendous advantage over listeners who cannot.

The second result clearly shown in the right panel of Figure 3 is that performance in the 1-back task with a speech masker was significantly worse than performance with a noise masker, even when the SNR was adjusted to generate equivalent performance in the baseline 0-back task. This again clearly demonstrates that speech segregation tasks tend to require more cognitive resources than tasks that require the extraction of speech from noise. In other words, listeners are able to obtain equivalent or better performance in speech perception tasks with speech maskers than they can in speech perception tasks with noise maskers, but that this performance comes at the cost of additional listening effort.

CONCLUSIONS

In this paper, we have examined three experiments that compared performance with speech and noise maskers in a variety of listening tasks with differing levels of complexity. In general, the results of these experiments are consistent with previous experiments that have shown that listeners tend to perform better in simple listening tasks with a single irrelevant speech masker than with a continous noise masker that produces a comparable amount of energetic masking [2, 3, 4]. However, in all cases, the results have also shown that, when the SNR is of the stimulus is adjusted to produce equivalent performance in a relatively simple listening task, performance tends to decrease more rapidly with increasing task complexity with speech maskers than it does with noise maskers. We propose that this more rapid decrease in performance with additional task complexity indicates that the relatively high level of performance that listeners are generally able to obtain in listening tasks with irrelevant speech maskers are only achievable if the listeners are able to focus their cognitive resources on those listening tasks without being distracted by other competing task requirements. In other words, they are able to obtain better than expected performance with speech maskers, but they are only able to do so by devoting additional effort to the listening task. This additional effort is likely related to the "informational masking" effects that listeners must overcome when they attempt to extract the acoustic and phonetic elements of a speech signal from those of a potentially confusable speech masker.

If it is indeed the case that there are significant variations in the amount of effort listeners require to

extract information from different types of maskers, then it is clear that auditory researchers need to be a little more careful when they design psychoacoustic experiments to evaluate the impact of hearing impairment or the effectiveness of signal processing algorithms designed to improve the performance of hearing impaired listeners. If the assessment of auditory performance is limited to conventional 0-back listening tasks, it is quite possible that researchers who find equivalent performance in two listening conditions may be entirely missing the fact that listeners may be working harder to achieve that level of performance in one of those conditions. This could easily explain why listeners sometimes report much greater preference for a hearing aid algorithm (for example, one incorporating noise reduction) even when laboratory tests fail to show a significant improvement in intelligibility in that condition [5]. We believe that the one-back task described in Experiment 3 might provide a particularly elegant way of evaluating the impact of listening effort on different listening conditions, because it uses a secondary task that is directly integrated into the listener's primary listening task and does not rely on a secondary non-auditory task to distract the listener away from the auditory stimulus. Future research is needed to determine how performance in the 1-back task varies in other types of complex listening tasks involving different types of maskers and more complex listening environments with spatially separated competing speech signals.

ACKNOWLEDGMENTS

REFERENCES

[1] D. Brungart, "Informational and energetic masking effects in the perception of two simultaneous talkers", Journal of the Acoustical Society of America **109**, 1101–1109 (2001).

[2] J. Festen and R. Plomp, "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**, 1725–1736 (1990).

[3] M. K. Qin and A. J. Oxenham, "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers", Journal of the Acoustical Society of America **114**, 446–454 (2003).

[4] J. Bernstein and K. Grant, "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners", The Journal of the Acoustical Society of America **125**, 3358 (2009).

[5] A. Sarampalis, S. Kalluri, B. Edwards, and E. Hafter, "Objective measures of listening effort: Effects of background noise and noise reduction", Journal of Speech, Language and Hearing Research **52**, 1230 (2009).

[6] G. Kidd Jr, V. Richards, C. Mason, F. Gallun, and R. Huang, "Informational masking increases the costs of monitoring multiple channels", The Journal of the Acoustical Society of America **124**, 223 (2008).

[7] R. Bilger, J. Nuetzel, W. Rabinowitz, and C. Rzeczkowski, "Standardization of a test of speech perception in noise", Journal of Speech, Language and Hearing Research **27**, 32 (1984).