Objective Measures of Listening Effort: Effects of Background Noise and Noise Reduction

Anastasios Sarampalis
University of California at Berkeley

Sridhar Kalluri
Brent Edwards
Starkey Hearing Research Center,
Berkeley, CA

Ervin Hafter
University of California at Berkeley

Purpose: This work is aimed at addressing a seeming contradiction related to the use of noise-reduction (NR) algorithms in hearing aids. The problem is that although some listeners claim a subjective improvement from NR, it has not been shown to improve speech intelligibility, often even making it worse.
Method: To address this, the hypothesis tested here is that the positive effects of NR might be to reduce cognitive effort directed toward speech reception, making it available for other tasks. Normal-hearing individuals participated in 2 dual-task experiments, in which 1 task was to report sentences or words in noise set to various signal-to-noise ratios. Secondary tasks involved either holding words in short-term memory or responding in a complex visual reaction-time task.
Results: At low values of signal-to-noise ratio, although NR had no positive effect on speech reception thresholds, it led to better performance on the word-memory task and quicker responses in visual reaction times.
Conclusions: Results from both dual tasks support the hypothesis that NR reduces listening effort and frees up cognitive resources for other tasks. Future hearing aid research should incorporate objective measurements of cognitive benefits.

Hearing-impaired (HI) listeners, despite understanding speech in quiet almost as well as normal-hearing (NH) listeners, have great difficulties when speech is presented in background noise (e.g., Plomp, 1994). This is true even when amplification is provided by means of a hearing aid such that the speech is within the range of audibility; this problem is a widely reported reason for hearing aid owners to stop using their devices (Kochkin, 2000). Furthermore, this difficulty becomes more pronounced as the degree of hearing loss increases (Killion, 1997).

Advances in digital hearing aid technology have allowed the widespread use of signal processing algorithms such as spectral feature enhancement, multiband compression, directional microphones, and noise reduction (NR), mainly with the aim of improving speech intelligibility, particularly in adverse listening conditions. The benefits, or lack thereof, of these algorithms on speech intelligibility are, understandably, well documented (e.g., Dillon & Lovegrove, 1993; Hickson, 1994; Levitt, Neuman, Mills, & Schwander, 1986; Ricketts, Lindley, & Henry, 2001). Objective measurements of benefits beyond those seen with speech tests, however, are not so prevalent. In particular, NR algorithms, which will be the main focus here, aim to counteract the effects of noise on speech perception and sound quality by improving the signal-to-noise ratio (SNR). These algorithms exist in many forms, but in general, they all work by adjusting

1092-4388/09/5205-1230

the gain in different frequency regions according to some measure of the level of noise in each of these regions (Dillon & Lovegrove, 1993; Levitt, 2001). They generate and update continuously an estimate of the noise and use it to determine the SNR in each frequency band. Regions with a low SNR are dominated by noise and contribute little to speech intelligibility, and so their gain can be reduced with little cost. As such, NR algorithms perform best when the noise and signal occupy different spectral regions.

Despite the proliferation of NR algorithms in hearing aids, objective demonstrations of improvements in speech intelligibility can be described at best as elusive (Dillon & Lovegrove, 1993; Edwards, 2004). In fact, NR can even be detrimental to performance if processing artifacts are introduced. Nevertheless, hearing aid users often express a preference for these algorithms, reporting improved sound quality, better ease of listening, and sometimes even a perceived improvement in speech understanding. Keidser (1996), for example, demonstrated that HI listeners consistently express preference for the types of amplification that offer reduced gain at frequencies where the SNR is poorest, as the typical NR algorithm would do. To explain this dichotomy, Hafter and Schlauch (1992) proposed that NR algorithms did not improve speech reception thresholds (SRTs) in the laboratory because they are essentially redundant with internal processes in the human participant, thus providing no additional information about the signal. In other words, digital NR algorithms perform a function similar to that of the listeners' auditory and cognitive systems and, as such, do not improve speech understanding. Their proposal was that participants might still like the NR because, in doing for them what they could do for themselves, it lightens their cognitive load. From this perspective, NR might not affect the SRT but may release attentional resources to be used for other, simultaneous tasks. Although this reduction in cognitive load might not affect performance in traditional speech tests conducted in the laboratory or clinic, it could be important in more natural settings, where multitasking is the norm and cognitive demands are greater. As such, Hafter and Schlauch (1992) warned that abandonment of NR on the basis of seemingly negative results in situations where full attention can be paid to audition could preclude its importance, say, in a classroom, where attention is split between audition and other tasks, and the reduction of attentional interference by NR might have positive implications on learning. What is more, improving ease of listening may lead to reduction of fatigue, which, in turn, may lead to improvements in intelligibility, if prolonged listening without NR tires listeners more quickly (Lim & Oppenheim, 1979). The purpose of the present study was to follow this line of reasoning by testing the role of NR in more complex cognitive environments in which the listener is faced with multiple demands for attention. Communication is a process that involves more than the auditory functions of the periphery, such as selectively attending to sound sources, storing information in memory, using context information to improve understanding, resolving ambiguities, and generating appropriate responses quickly. This complexity means that future hearing aid research will need to measure outcomes beyond those pertaining to audibility issues and more along the dimensions of cognitive benefits (Edwards, 2007). Such measures may demonstrate cognitive benefits even in situations where signal processing is not traditionally considered useful, such as directionality at high SNRs.

One of the most widely used methods in the cognitive literature for assessing the cognitive demands of a task is the *dual task*. It is based on theories of a single limited cognitive resource (e.g., Baddeley, 1998; Kahneman, 1973) that is shared among simultaneous processes. When a participant performs two simultaneous tasks, there is competition for resources. As the cognitive demand for one task increases, so does its share of allocated resources, resulting in reduced resources available for the competing task. As a consequence, the demands of one task can be inferred from changes in the performance in the competing task. The goal of the research presented here was to develop a dual-task paradigm to measure the cognitive demands of understanding speech in the presence of background noise, and, further, to evaluate whether use of NR reduces the effort expended on the task. Two such experiments with NH listeners are presented here in order to investigate these factors, without the confounding factors of cognitive and auditive issues of typical HI listeners. The first dual-task experiment was designed to investigate how noise and NR affect the ability to recall spoken words while the amount of contextual information was manipulated. We hypothesized that because the use of context is a top-down, effortful process (Pichora-Fuller, Schneider, & Daneman, 1995), changes in memory performance would be more likely to occur when contextual information can be used to disambiguate the signal. The second dual task was designed to assess the effects of NR on speed of processing while listening to speech in noise. If NR helps to reduce effort in noisy listening conditions, then we would expect to see improved performance in these secondary cognitive tasks, as compared with conditions without NR. If, on the other hand, NR does not increase the ease with which speech is understood in noise, performance on the cognitive tasks will be unaffected.

Experiment 1

In Experiment 1, the ability to remember words spoken in quiet or noise was tested using a dual-task paradigm. Listeners were asked to repeat the last word of sentences (primary task), with instructions to remember them for later recall (secondary task). The ability to understand the speech as well as to retain information was measured at different SNRs and with or without NR. Following the paradigm of Pichora-Fuller and colleagues (1995), we used the sentences from the revised Speech Perception in Noise (SPIN-R) test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984) to investigate the contribution of context and top-down processing in speech understanding.

Method

Participants. Twenty-five native speakers of American English took part in the first experiment. They had pure-tone thresholds lower than 15 dB HL at all audiometric frequencies and were paid for their participation. Their average age was 20 years (range: 18–26 years).

Stimuli. The SPIN-R sentences (Bilger et al., 1984) were used in the experiment. These sentences consist of eight lists of 50 sentences; half of the sentences contain context information that makes the last word predictable (e.g., "A chimpanzee is an ape."), whereas the other half do not (e.g., "She might have discussed the ape."). The sentences were presented either in quiet (at a level of 65 dB SPL) or in the presence of 4-speaker babble. When the babble was present, its level was set to 65 dB SPL, and the level of the sentences was adjusted to either −2 or 2 dB SNR. The 4-dB change in SNR was chosen to correspond to the approximate benefit observed with directional microphones (e.g., Dillon & Lovegrove, 1993). In babble, the materials were either left unprocessed or were processed using the Ephraim-Malah NR algorithm (Ephraim & Malah, 1984, 1985), which is a good example of current NR processing. All stimuli were preprocessed and recorded at a sampling rate of 22050 Hz. The two SNRs, with and without NR processing applied, and presentation in quiet amounted to five different conditions that were tested in the experiment.

The Ephraim-Mallah NR (Ephraim & Malah, 1984, 1985) algorithm assumes that the summed noise and speech samples making up the input waveform are independent Gaussian samples and then derives the minimum mean-square error estimator of the speech spectral amplitude given the (presumed constant) noise amplitude and the current noisy input amplitude. This estimator requires a variable related to the instantaneous SNR, which was determined based on a weighted average of information from the current and previous frames of analysis. One of the algorithm's basic properties is to attenuate the gain more at frequencies where the SNR is poorest. The maximum attenuation in the experiments presented here was 20 dB. Such an algorithm is not yet implemented in hearing aids but is available in other technologies, such as in the telecommunications industry, and may, in the future, be available in hearing aids (Edwards, 2007).

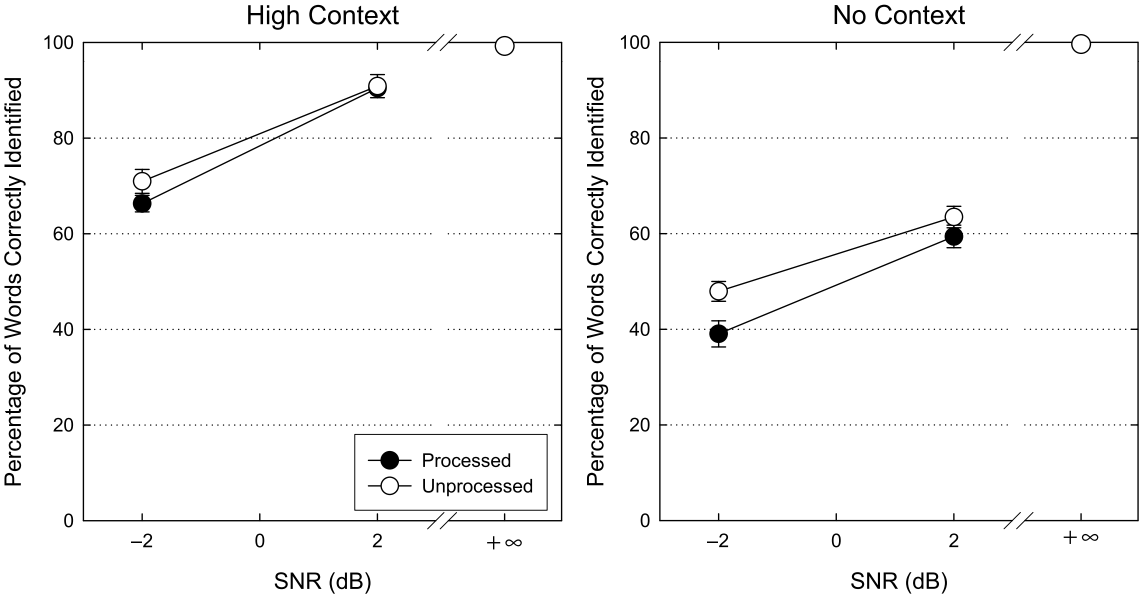
Procedure

Listeners were seated in a double-walled, sound-attenuating booth, and the sentences were presented to them diotically over headphones (Sennheiser HD580; Sennheiser Electronic Corporation, Old Lyme, CT). The experimenter was seated outside the booth and could hear the listener's responses through a pair of headphones. The procedure of the first experiment followed closely the paradigm used by Pichora-Fuller et al. (1995). After each sentence was presented, the listeners repeated the last word as they believed they heard it. They were encouraged to guess if they were uncertain, and, indeed, there were no cases in which no response was given. Responses were counted as correct when they were identical to the presented word. Participants were also asked to remember their responses, as they would be asked to recall them later. After every 8 sentences, a visual cue prompted the listeners to recall as many of the previously reported words as they could, verbally and in any order they preferred. For recall, responses were deemed correct when they were identical to the words reported previously. Five blocks, each with an experimental condition fixed within, were presented in random order, and each consisted of the first 48 sentences of one of the SPIN-R lists. A sixth list of 48 sentences was used prior to data collection as practice (without background noise).

Results

Figure 1 shows mean intelligibility as a function of SNR, for high- and no-context words. The parameter is NR processing. Performance was perfect with both types of words in quiet. In the presence of noise, identification performance with high-context words was 30 percentage points higher than with no-context words. Because the keywords in the two lists are the same, the benefit of context information demonstrates the contribution of top-down processing on understanding speech. With both lists of words, performance went down approximately 20 percentage points, with a 4-dB lowering of SNR. With no-context words, performance was consistently better without the NR algorithm. With high-context words, performance was similar with or without processing, although performance was slightly better without the NR algorithm at the low SNR. A three-way repeated-measures analysis of variance (ANOVA) was performed on the results in conditions containing noise, with SNR,

Figure 1. Speech intelligibility as a function of signal-to-noise ratio (SNR), averaged across 25 listeners in Experiment 1. The left and right panels show performance for material having contextual information and for material lacking contextual information, respectively. Data with noise reduction (NR) processing are plotted with filled symbols, and those without NR processing are plotted with open symbols. The error bars denote 1 standard error of the mean (*SEM*).

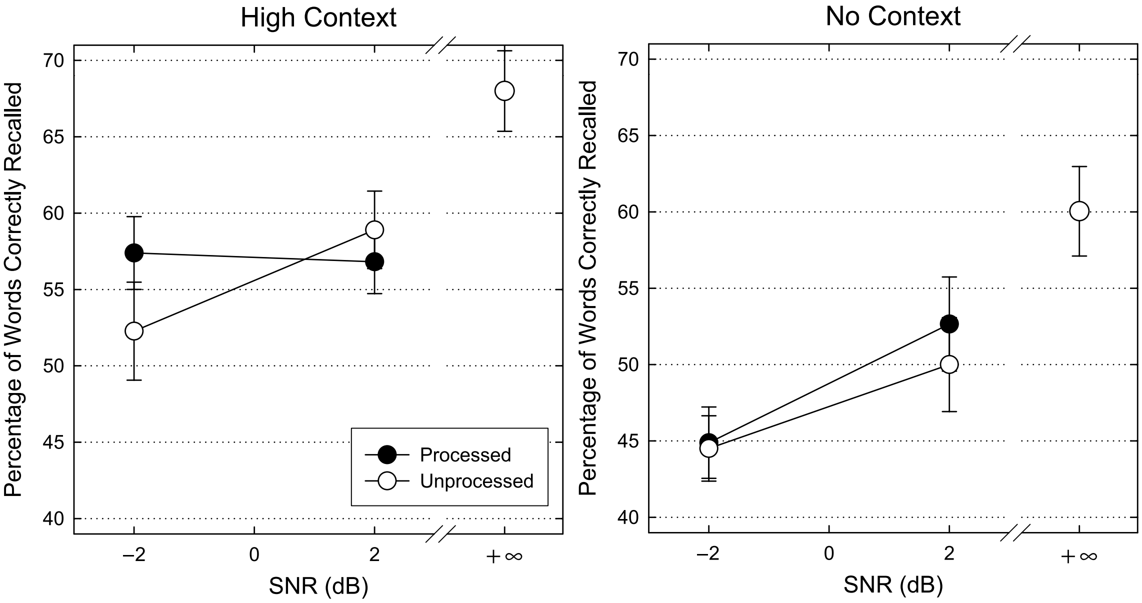


processing, and context as factors. The three-way interaction was not significant. From the three two-way interactions, only the SNR × Processing interaction was significant, $F(1, 24) = 20.48$, $\mathrm{MS_e} = 109.52$, $p < .05$. Four planned comparisons were performed between conditions with and without NR, one for each SNR and word type. No significant differences were found between performance with the NR algorithm for the 2-dB SNR for either type of word. On the other hand, with an SNR of –2 dB, performance was significantly better without the NR with both high- and no-context words, $t(24) = 2.147$, $p < .05$ and $t(24) = 3.058$, $p < .005$, respectively. These results were not unexpected. As previously stated, NR algorithms do not improve speech understanding, and care must be taken in their design to ensure that they improve sound quality while not decreasing speech intelligibility.

Figure 2 shows mean recall performance as a function of SNR. The plots are organized as in Figure 1. Overall, recall was better for high-context than for no-context words. This was true for sentences presented in quiet and in noise. With no-context words, recall performance in quiet was approximately 60% correct and fell in the presence of noise. What is more, a drop in SNR of 4 dB resulted in a further drop in recall performance of approximately 5–8 percentage points. There was little effect of NR on recall performance, with a small benefit of processing at the 2-dB SNR. With high-context words, the results were qualitatively different. In the absence of NR processing, performance fell with the introduction of noise and with a decrease in SNR, by an amount that was similar to that seen with unprocessed no-context sentences. On the other hand, with NR processing, performance did not vary with SNR and, at the lower SNR, was 5 percentage points higher than in the unprocessed condition. A three-way repeated-measures ANOVA was performed on the recall data, with SNR, processing, and type of word as factors. In this case, the three-way interaction was significant, $F(1, 24) = 15.125$, $\mathrm{MS_e} = 78$, $p < .05$, suggesting that the effects of processing were different for the two types of words and two SNRs. Four planned comparisons between scores with and without NR were performed, one for each SNR and word type. No significant change in performance was seen with NR for the two SNRs with no-context words. Similarly, no significant change was seen with the high-context words at the 2-dB SNR. However, recall performance was significantly better with NR than without NR for a –2-dB SNR with the high-context words, $t(24) = 2.362$, $p < .05$.

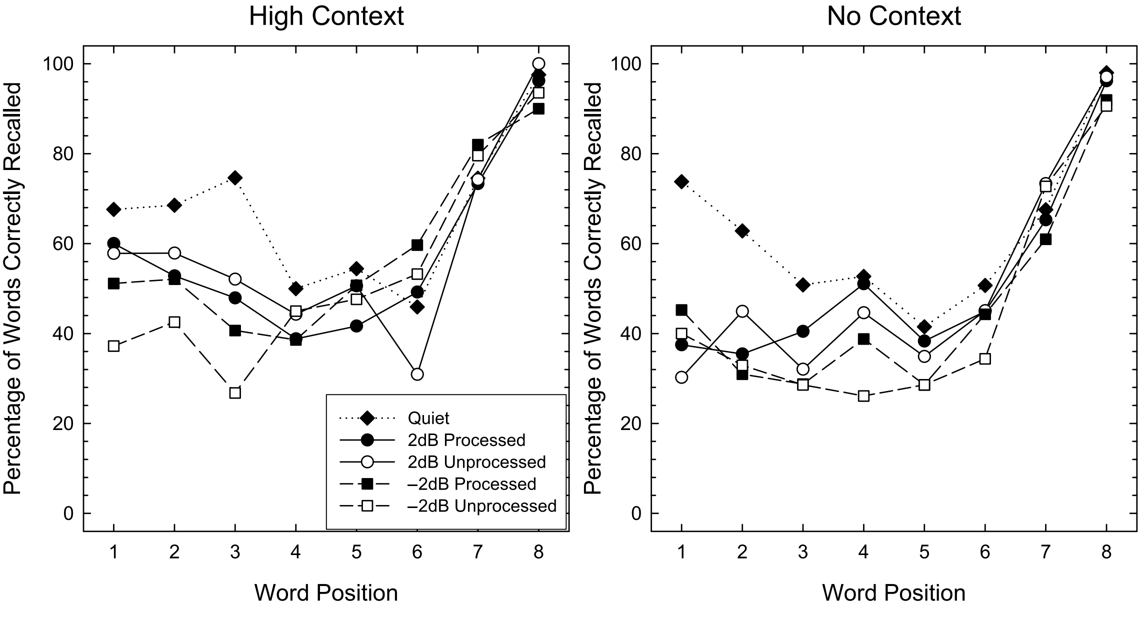
In order to assess how rehearsal was affected by the presence of noise and NR, the mean recall scores were replotted as a function of word position (the position in the eight-item list of to-be-remembered words). When free recall data are plotted in this way, two effects usually appear. The first is better performance for the items at the end of the block (recency effect). The second effect is increased recall for the first few items in the list (primacy effect). According to Atkinson and Shiffrin (1968), the primacy effect can be explained by assuming that the

Figure 2. See caption in Figure 1, but this figure illustrates free recall performance.



first few items in the list receive more rehearsal than later ones and are encoded in long-term memory. Figure 3 shows mean recall performance as a function of word position. Each plot shows data for one type of word, whereas the parameter is presence of noise and NR. All of the curves in the two plots show a very clear recency effect, with performance close to 100% for the last item and little difference between conditions. This was expected, as the recency effect is generally thought to reflect short-term storage capacity (Glanzer & Cunitz, 1966) and is largely unaffected by changes in task demands. On the other hand, the size of the primacy effect was highly variable across conditions. With no-context words, the effect was absent when the sentences were

Figure 3. Free recall performance in Experiment 1, as a function of word position, averaged across 25 listeners. The left and right panels show performance for sentences with and without context, respectively. The parameter is presence of noise and NR processing.



presented in noise, regardless of the presence of NR. The early parts of the curves were essentially flat, and performance was no better for early items than it was for middle ones. Conversely, primacy with high-context words showed a dependency both on SNR and NR processing. More specifically, the ability to recall early words declined as the SNR decreased. The effect of NR on the primacy effect was dependent on the SNR. Although there was no change due to NR when the SNR was 2 dB, there was an approximately 15-percentage-point improvement in recall of the first three items due to the NR when the SNR was –2 dB.

These observations were confirmed by two two-way ANOVAs, one for each word type, with word position (8 levels) and noise condition (5 levels) as factors. The two-way interaction with no-context words was nonsignificant, whereas the two-way interaction with high-context words was significant, $F(28, 672) = 1.977$, $\mathrm{MS_e} = 52.562$, $p < .005$. To ensure that the significant interaction could not be accounted for by the quiet scores being different from all the conditions with noise, we also ran two-way ANOVAs excluding the results in quiet. These confirmed the significant two-way interaction with high-context words, $F(21, 420) = 1.69$, $\mathrm{MS_e} = 95.22$, $p < .05$, and the nonsignificant interaction with no-context words. Thus, significant changes in mean recall performance seen in Figure 2 were all reflecting changes in the ability to rehearse the to-be-remembered items. These findings indicate that in the absence of NR, the ability to rehearse the content-rich spoken words was reduced as the SNR was decreased. When context information was available, rehearsal was facilitated by providing NR processing, at least at the lowest SNR tested here.

Experiment 2

Experiment 2 investigated the effects of noise level and NR on speed of processing. As in the previous experiment, listeners were asked to repeat speech presented in quiet or in noise while performing a simultaneous cognitive task. In this experiment, however, the secondary, cognitive task required participants to respond quickly to complex visual stimuli.

Method

Participants. Twenty-five native speakers of American English took part in the experiment, some of whom had participated in Experiment 1. They had pure-tone thresholds lower than 15 dB HL at all audiometric frequencies and were paid for their participation. Their average age was approximately 21 years (range: 19–27 years).

Stimuli. The Institute of Electrical and Electronics Engineers (IEEE) sentences (IEEE, 1969; recorded by Galvin and Fu at the House Ear Institute) were used to measure speech intelligibility performance. The corpus consists of 720 sentences of comparable length and difficulty, all recorded in the same male voice (e.g., “The fruit peel was cut in thick slices.”). The sentences were presented either in quiet (at a level of 65 dB SPL) or in the presence of 4-speaker babble. When the babble was present, its level was set to 65 dB SPL, and the level of the sentences was adjusted to –6, –2, or 2 dB SNR. When presented in babble, the materials were either left unprocessed or were processed using the Ephraim-Malah NR algorithm (Ephraim & Malah, 1984, 1985). All stimuli were preprocessed and recorded at a sampling rate of 22050 Hz.

A visual reaction-time task was given concurrently to the auditory task in order to measure speed of processing. During this task, a computer monitor located approximately 50 cm from the participant presented a display that consisted of two boxes. The boxes measured 8 cm and 6 cm in height and width, respectively, and were separated by 1.5 cm. At quasi-random intervals, a digit between 1 and 8 appeared on either one of these two boxes. The participants used a keyboard provided to press the arrow button that pointed toward the digit, if the digit was even, or away from the digit, if the digit was odd. They were instructed to perform this task as fast as they could while maintaining a high level of accuracy. Each digit remained on the screen until the participant pressed one of the two arrow keys, or for a maximum of 2.5 s. The next digit appeared after a randomly chosen interval of time, uniformly distributed between 0.5 and 2.0 s after the previous digit had disappeared. Accuracy scores and reaction times were recorded for each trial.

Procedure

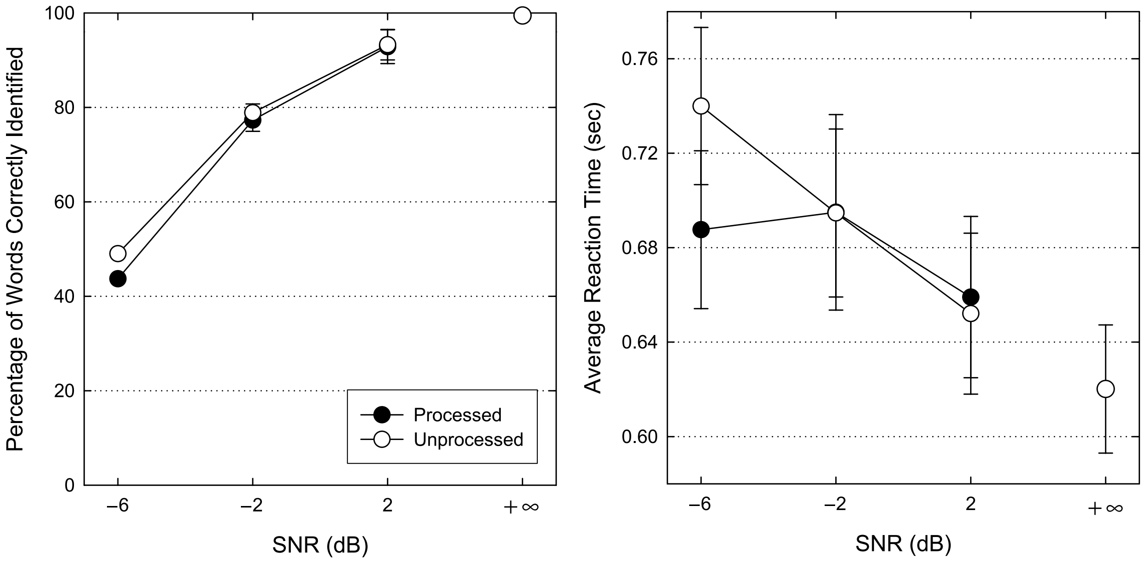
Listeners were seated in a double-walled, sound-attenuating booth throughout the experiment. A block of presentations started when the listener indicated that he or she was ready to begin. At that point, the visual display appeared, and the audio signal began playing diotically over a pair of headphones (Sennheiser HD580). The audio signal consisted of 25 sentences, presented at a rate of 1 sentence every 8 s (sentences were approximately 3 s in duration). The first sentence started 9 s after the beginning of the block. The timing of the visual stimuli was uncorrelated to that of the auditory stimuli. In conditions with noise, the noise started at the beginning of the block and was present throughout. The listeners were asked to repeat the sentences or the parts that they could understand after each sentence was presented. All words except for the indefinite and definite articles were considered keywords. The participants were instructed to do both the auditory and the visual task simultaneously, paying equal amounts of attention to the two. Two blocks of 25 sentences were presented for each of the seven experimental conditions. The experimental conditions were randomized in order. Prior to data collection, listeners practiced the two tasks, both individually and simultaneously until they reported being comfortable with them. An experimenter, seated outside the soundproof chamber, scored the participants' responses to the sentences.

Results

Figure 4 shows mean speech intelligibility and reaction time performance (left and right panel, respectively) as a function of SNR. Consider first the speech intelligibility data. Performance in quiet was perfect for all 25 listeners. The introduction of noise at an SNR of 2 dB brought performance down by approximately 7 percentage points. Thereafter, each 4-dB drop in SNR worsened performance. At –6 dB SNR, performance was, on average, around 50% correct. There was little noticeable change in performance with the NR algorithm at the two higher SNRs, but it was worse by approximately 5 percentage points at the lowest SNR. A two-way ANOVA, with SNR and processing as factors, was calculated on the results. The two-way interaction was found to be nonsignificant, as was the main effect of processing. The effect of SNR was significant, however, $F(2, 48) = 207.556$, $MS_e = 87.76$, $p < .001$.

On the visual task, listeners were able to perform relatively fast (620 ms, on average) when the sentences were presented in quiet. When the sentences were presented without NR, the introduction of noise resulted in slower reaction times. What is more, with every 4-dB drop in SNR, reaction times worsened by approximately 44 ms, a small yet reliable change in performance. When the sentences were presented with NR, a different pattern of reaction times was seen. At the two highest SNRs, performance was virtually identical to performance without processing. At the lowest SNR (–6 dB), however, performance was better with NR processing than without it. A two-way ANOVA with SNR and processing as factors was calculated, as were planned comparisons between conditions with and without NR for each SNR. The two-way interaction and the effect of processing were not found to be significant. There was, however, a significant main effect of SNR, $F(2, 48) = 10.716$, $MS_e = 0.004$, $p < .001$. In the three protected planned comparisons, no significant difference in reaction times was found between the conditions with and without NR at the two higher SNRs (2 and –2 dB). However, at the lowest SNR (–6 dB), reaction time performance was significantly better with the NR than without it, $t(24) = 4.64$, $p < .001$. This finding suggests that at this low SNR, use of an NR algorithm may free up cognitive resources that would otherwise be involved in extracting speech from noise, allowing them to be allocated to other simultaneous processing tasks. Although, at first, it may be hard to reconcile the lack of a two-way interaction with the significant difference at –6 dB SNR and lack of significant difference at the other SNRs, it is not impossible,

Figure 4. Mean speech intelligibility performance (left panel) and mean reaction times (right panel) as a function of SNR, averaged across 25 listeners in Experiment 2. The parameter is presence of NR processing. Error bars denote 1 *SEM*.



statistically, nor was it due to an error in calculation. Planned comparisons tend to be more powerful tests than ANOVA tests for interactions, and, as such, these seemingly incongruent patterns are sometimes observed.

Discussion

The results of the two experiments indicate that the presence of background noise during a listening task can have negative effects not only on the listening task but also on the listener's ability to perform simultaneous, cognitive activities, such as rehearsing to-be-remembered words or responding speedily in a complex visual task. Furthermore, these results suggest that the benefit of using a digital NR algorithm is not in making speech more intelligible but, rather, in reducing the cognitive effort involved in the task. This can be seen as an improvement in performance in a simultaneous task.

Shared Attention While Extracting Speech From Noise

Evidence that noise in an auditory signal affects performance in nonauditory tasks is not new. Broadbent (1958), for example, degraded speech by introducing distortions of different types and degrees and showed that although intelligibility performance remained unchanged, the distortions resulted in changes in performance in a concurrent measure of effort: a high-speed, visual tracking task. Similarly, Rabbitt (1966, 1968) conducted a series of studies that investigated the effects of noise during an auditory task on listening effort. He showed that the ability to remember correctly-identified words was impaired when pulse-modulated white noise was added, even when the noise was not sufficient to reduce recognition performance (Rabbitt, 1966). What is more, he found that memory for items heard early in a list, regardless of whether they were presented in quiet or noise, was impaired when later items in the same list were presented in noise (Rabbitt, 1968). Rabbitt (1968, 1991) suggested that the process of recognizing the items in noise drew on cognitive resources necessary for the rehearsal and encoding of the earlier items. The increased effort involved with recognizing the later items interfered with the ability to retain the earlier ones (see also Heinrich, Schneider, & Craik, 2008). The results from Experiment 1 presented here are in line with such an assertion: Items presented in noise were less likely to be remembered successfully.

Rabbitt (1991) also suggested that HI listeners have to rely more heavily on cognitive resources to recognize speech than do NH listeners because of the inherently degraded acoustic input. In line with this hypothesis, lists of words that were recognized equally well by the two groups of listeners were better recalled by NH listeners than by those with mild hearing loss. Rakerd, Seitz, and Whearty (1996) also demonstrated this by asking groups of NH and HI listeners to retain lists of visually presented digits for a period of 60 s. During this retention period, individuals were either asked to passively listen to speech-shaped noise or to listen to a passage of speech presented in quiet. They were informed that questions on the content of the passage would be asked later. All groups of listeners forgot more digits in the speech condition than in the noise condition. However, listeners with congenital or early-onset hearing loss demonstrated a significantly greater cost of speech processing; they forgot more digits in the speech condition than did the NH controls. Rakerd et al. (1996) concluded that peripheral deficits associated with hearing loss limit access to acoustic information even when amplification is provided, increasing the demand for cognitive resources and reducing the ease of listening. Similar conclusions were reached by Feuerstein (1992), who measured the effects of simulated unilateral hearing loss on speech intelligibility and on performance in a simultaneous visual reaction-time task. Feuerstein found that listening effort was significantly increased when hearing loss was simulated, as indicated by an increase in reaction times in the secondary task, and concluded that in certain listening situations, unilateral hearing loss can reduce ease of listening. Other work using similar dual tasks supports a hypothesis of increased effortfulness of listening due to hearing impairment (e.g., Hicks & Tharpe, 2002; McCoy et al., 2005; Pichora-Fuller et al., 1995), audio distortion with NH listeners (e.g., Mackersie, Boothroyd, & Prida, 2000), and aging (e.g., Murphy, Craik, Li, & Schneider, 2000).

Cognitive Effort and Processing in Hearing Aids

Typically, research into the benefits of hearing aid signal processing has focused on its effects on traditional speech reception measures that measure changes in intelligibility but often fall short of providing a complete picture of the processes involved during communication. Such measures can be insensitive to changes in the role of top-down processes that make the perception and comprehension of degraded signals (due to either external or internal noise) possible. Several studies, however, have demonstrated and called for the use of measures that are sensitive to changes in these processes in different listening situations. Gatehouse and Gordon (1990), for example, used response times to speech stimuli as an additional measure of benefit from amplification for listeners with mild to moderate hearing loss. The authors demonstrated that the response time measures exhibited the benefits of amplification in situations where the traditional speech reception measures failed to do so because of ceiling effects or large variability. Similarly, Baer, Moore, and Gatehouse (1993) investigated the effects of a digital algorithm designed to enhance spectral contrast of speech in noise, using listeners with sensorineural hearing loss. They measured both intelligibility and response times and found that although both demonstrated benefits from using the algorithm, the benefits were about twice as large for the response time measures than for the intelligibility scores. Baer et al. (1993) suggested that the greatest benefits of the processing may be in reducing listening effort rather than in improving speech intelligibility.

Results from the two experiments presented here also underline the importance of using cognitive measures in hearing aid research. Traditional speech reception measures showed no sensitivity to the benefits of a digital NR algorithm. On the other hand, performance in both cognitive tasks showed dependence on processing, with improvements seen at the lowest SNR when an NR algorithm was used to counteract the effects of noise. As in the experiment by Baer et al. (1993), the benefits of signal processing were not in improving speech recognition scores but, rather, in reducing listening effort. Although these results were obtained with NH listeners, they have implications for the use of signal processing in hearing aids. Alain, McDonald, Ostroff, and Schneider (2004) and Pichora-Fuller et al. (1995) argue that one of the results of aging is a switch from automatic processing of sounds to more top-down, controlled processing. It is reasonable, then, to assume that the use of simultaneous measures of cognitive processing is even more pertinent when working with elderly or HI individuals, where changes in the peripheral auditory system result in inherently degraded signals. In the present article, we demonstrate two ways that cognitive measures may be used to increase sensitivity to the benefits of signal processing without a significant cost in the duration of testing.

Clearly, extracting speech from noise—whether it be an added acoustic noise or an inferred internal noise associated with hearing impairment—can reduce performance in such simultaneous cognitive tasks as word recall or complex visual reaction time. The primary conjecture of the present experiment was that this effect would be reduced in the presence of a noise-reduction algorithm, a prediction that proved true at the lowest SNR used. It is generally accepted that there are no positive effects of NR on intelligibility, and such algorithms have even been shown to reduce intelligibility at low SNRs. Given this, the positive results found here for NR at low SNRs for the secondary cognitive task give credence to the proposal in Hafter and Schlauch (1992) that all new solutions for hearing aids should be tested in settings that emulate the high demands of more natural situations. Typical laboratory settings may fail to capture the complexity of comprehending speech in everyday life and so provide an incomplete test of hearing aid processing. Consider, for example, the task of an HI student in a noisy classroom, where extracting the teacher's words from noise is only the first stage of encoding, comprehending, and eventually learning the content of a lesson. If NR processing makes listening less effortful, therefore facilitating comprehension, then it is important not to dismiss the value of such algorithms in the absence of SRT effects.

In the interest of making the task of extracting speech from background noise as demanding as possible, the noise that was used in the experiments presented here (4-speaker babble) had a long-term spectrum that was very similar to that of the target speech. This arguably provides a very stringent test for NR algorithms, as they operate best when the target and noise occupy different spectral regions. It is very likely, then, that the benefits of NR on cognitive load may be greater in situations where the noise has a spectral shape that is more dissimilar to speech, as in the case of industrial or environmental noise.

A possible concern with an interpretation based on reduced listening effort is that the cognitive benefits of the NR algorithm are seen only at SNRs where listening performance is also somewhat reduced as a result of processing. Could it be that our participants were only getting better at the cognitive task because they were trading off their performance with the listening task? There are several reasons why we believe this not to be the case. First, in Experiment 1, if this alternative explanation were true, one would expect to see an improvement in recall scores with no-context sentences more so than with high-context sentences, as the drop in speech intelligibility with the NR algorithm is greater with no-context sentences. Moreover, it would be reasonable to expect participants to be more prone to giving up on the listening task and favoring the secondary task, when the primary task is most difficult, as it was with no-context sentences. This effect is clearly not observed with our data. A second reason to reject a trade-off explanation is that where a significant benefit in RT scores is seen with the NR algorithm in Experiment 2, there is only a small and nonsignificant change in speech intelligibility. What is more, the extent to which RT scores improved with the NR algorithm at –6 dB SNR is certainly not commensurate to the change in speech intelligibility, when the rest of the data are taken into account.

Implications for Directional Microphones

It is intriguing to consider these effects in the context of other hearing aid processing strategies designed to improve the SNR, particularly the use of directional microphones. These have greater sensitivity to sounds originating from an area in front of the user and have been demonstrated to increase the SNR by about 4 dB (Amlani, 2001; Dillon & Lovegrove, 1993), as compared with omnidirectional microphones. Results from both experiments in the present article demonstrate that increasing the SNR not only improves the ability to understand speech but, more importantly, reduces listening effort. Indeed, a 4-dB increase in SNR, the equivalent of the benefit offered by directional microphone technology, resulted in a significant increase in the ability to retain words spoken in noise and to respond quickly to simultaneous complex visual stimuli. Consequently, it is reasonable to suggest that use of directional microphone technology in behind-the-ear hearing aids can reduce effort in noisy listening environments. Such technology is generally thought to be beneficial for hearing aid users only in improving speech intelligibility in noise and so, when the SNR is high enough, hearing aids with directional microphones switch to the omnidirectional mode on the assumption that directionality offers no benefit. However, as Pichora-Fuller et al. (1995) have demonstrated, HI listeners rely heavily on top-down processing to understand speech, even at such high SNRs. Thus, the present results suggest that the use of directional microphones can be of merit even when speech intelligibility is at ceiling because they help reduce listening effort.

Summary and Conclusions

It is widely accepted that the noise-reduction algorithms present in many hearing aids are ineffective in improving the intelligibility of speech at low SNRs, and yet they are often preferred by listeners. We examined this conflict in terms of an attentional effort hypothesis that says that NR, by doing some of the processing normally done by a listener, may free resources for other, simultaneous tasks.

Results from two dual-task experiments suggest that extracting speech at low SNRs reduces the listeners' abilities to rehearse heard material and to respond quickly to complex visual stimuli. At SNRs where speech reception thresholds are shown to be unimproved by NR, presence of the algorithm can allow the listener to remember more of the words in a presentation and to be quicker in responding to a complex visual task.

Acknowledgments

This research was funded by a University of California Discovery grant. The authors would like to thank Bill Woods for help in the implementation of the signal processing and discussions about the article.

References

Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2004). Aging: A switch from automatic to controlled processing of sounds? *Psychology and Aging, 19,* 125–133.

Amlani, A. M. (2001). Efficacy of directional microphone hearing aids: A meta-analytic perspective. *Journal of the American Academy of Audiology, 12,* 202–214.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 742–775). New York: Academic Press.

Baddeley, A. D. (1998). The central executive: A concept and some misconceptions. *Journal of the Neuropsychological Society, 4,* 523–526.

Baer, T., Moore, B. C. J., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *Journal of Rehabilitation Research and Development, 30,* 49–72.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32–48.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon.

Dillon, H., & Lovegrove, R. (1993). Single microphone noise reduction systems for hearing aids: A review and an evaluation. In G. A. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance* (pp. 353–372). Boston: Allyn and Bacon.

Edwards, B. (2004). Hearing aids and hearing impairment. In S. Greenberg, A. N. Popper, W. A. Ainsworth, & R. R. Fay (Eds.), *Speech processing in the auditory system* (pp. 339–421). New York: Springer.

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification, 11,* 1–15.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics Speech and Signal Processing, 32,* 1109–1121.

Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics Speech and Signal Processing, 33,* 443–445.

Feuerstein, J. F. (1992). Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort. *Ear and Hearing, 13,* 80–86.

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology, 24,* 63–68.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

Hafter, E. R., & Schlauch, R. S. (1992). Cognitive factors and selection of auditory listening bands. In A. Dancer, D. Henderson, R. J. Salvi, & R. P. Hammernik (Eds.), *Noise-induced hearing loss* (pp. 303–310). Philadelphia: B.C. Decker.

Heinrich, A., Schneider, B. A., & Craik, F. I. M. (2008). Investigating the effects of continuous babble on auditory short-term memory performance. *Quarterly Journal of Experimental Psychology, 61,* 735–751.

Hicks, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *Journal of Speech, Language, and Hearing Research, 45,* 573–584.

Hickson, L. M. H. (1994). Compression amplification in hearing aids. *American Journal of Audiology, 11,* 51–65.

Institute of Electrical and Electronics Engineers (IEEE). (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics AU-17,* 225–246.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Keidser, G. (1996). Selecting different amplification for different listening conditions. *Journal of the American Academy of Audiology, 7,* 92–104.

Killion, M. C. (1997). Hearing aids: Past, present, future: Moving toward normal conversations in noise. *British Journal of Audiology, 31,* 141–148.

Kochkin, S. (2000). MarkeTrak V: Why my hearing aids are in the drawer: The consumer's perspective. *Hearing Review, 53,* 34–41.

Levitt, H. (2001). Noise reduction in hearing aids: A review. *Journal of Rehabilitation Research and Development, 38,* 111–121.

Levitt, H., Neuman, A., Mills, R., & Schwander, T. (1986). A digital master hearing aid. *Journal of Rehabilitation Research and Development, 23,* 79–87.

Lim, J. S., & Oppenheim, A. V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE, 67,* 1586–1604.

Mackersie, C. L., Boothroyd, A., & Prida, T. (2000). Use of a simultaneous sentence perception test to enhance sensitivity to ease of listening. *Journal of Speech, Language, and Hearing Research, 43,* 675–682.

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology, 58A,* 22–33.

Murphy, D. R., Craik, F. I. M., Li, K. Z. H., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging, 15,* 323–334.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97,* 593–608.

Plomp, R. (1994). Noise, amplification, and compression: Considerations of 3 main issues in hearing-aid design. *Ear and Hearing, 15,* 2–12.

Rabbitt, P. M. A. (1966). Recognition: Memory for words correctly heard in noise. *Psychonomic Science, 6,* 383–384.

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology, 20,* 241–248.

Rabbitt, P. M. A. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngologica Supplementum, 476,* 167–176.

Rakerd, B., Seitz, P. F., & Whearty, M. (1996). Assessing the cognitive demands of speech listening for people with hearing losses. *Ear and Hearing, 17,* 97–106.

Ricketts, T., Lindley, G., & Henry, P. (2001). Impact of compression and hearing aid style on directional hearing aid benefit and performance. *Ear and Hearing, 22,* 348–361.

Received May 30, 2008

Accepted March 20, 2009

DOI: 10.1044/1092-4388(2009/08-0111)

Contact author: Anastasios Sarampalis, who is now with the Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, the Netherlands. E-mail: a.sarampalis@rug.nl.