Neural Encoding and Perception of Speech Signals in Informational Masking

Keri O'Connell Bennett,^{1,2} Curtis J. Billings,^{1,2} Michelle R. Molis,^{1,2} and Marjorie R. Leek^{1,2}

Objective: To investigate the contributions of energetic and informational masking to neural encoding and perception in noise, using oddball discrimination and sentence recognition tasks.

Design: P3 auditory evoked potential, behavioral discrimination, and sentence recognition data were recorded in response to speech and tonal signals presented to nine normal-hearing adults. Stimuli were presented at a signal to noise ratio of -3 dB in four background conditions: quiet, continuous noise, intermittent noise, and four-talker babble.

Results: Responses to tonal signals were not significantly different for the three maskers. However, responses to speech signals in the fourtalker babble resulted in longer P3 latencies, smaller P3 amplitudes, poorer discrimination accuracy, and longer reaction times than in any of the other conditions. Results also demonstrate significant correlations between physiological and behavioral data. As latency of the P3 increased, reaction times also increased and sentence recognition scores decreased.

Conclusion: The data confirm a differential effect of masker type on the P3 and behavioral responses and present evidence of interference by an informational masker to speech understanding at the level of the cortex. Results also validate the use of the P3 as a useful measure to demonstrate physiological correlates of informational masking.

(Ear & Hearing 2012;33;231-238)

INTRODUCTION

Background noise may reduce listeners' abilities to detect and recognize speech sounds. Decreased performance in a complex acoustic background results from contributions of both the auditory periphery and the information processing capabilities of the central auditory system. Although signalin-noise perception has been studied extensively using behavioral methodologies, the neural encoding of these signals in humans is not well understood. Recently, however, investigations of auditory evoked potentials (AEPs) have revealed masker-dependent deficits in cortical processing (Whiting et al. 1998; Martin et al. 1997, 1999; Martin & Stapells 2005; Androulidakis & Jones 2006; Kaplan-Neeman et al. 2006; Billings et al. 2009). Cortical AEP recorded in response to stimuli presented in various types of noise indicate that type of background noise as well as the signal to noise ratio (SNR) affect the responses.

Cortical AEPs recorded in broadband or filtered steadystate background noises or modulated noise indicate a differential effect of masker type on the responses. To evaluate neural encoding of signals in noise, the N1-P2 complex and the P3 peak have been evoked by both tonal and speech signals. For example, Martin et al. (1997) evoked cortical responses to speech signals masked by an unfiltered broad band noise as well as high-pass noise filtered at 4000, 2000, 1000, 500, and 250 Hz. Martin and Stapells (2005) analyzed the same waveforms evoked by the same speech signals and masker conditions as in the 1997 study but with low-pass filtered noise instead of high-pass noise. In both studies, the SNR was fixed but varied for each participant on the basis of the individual's behavioral masked thresholds for speech sounds. Results from both studies revealed significant masking above 1 kHz. When the filtered noise contained energy in the spectral region above 1 kHz, the consonant sound of the consonant-vowel syllable was significantly masked and the P3 response was eliminated. The disappearance of the waveform's peak was consistent with the reported decrease in behavioral discrimination for these syllables. In another study, Androulidakis and Jones (2006) evaluated the effects of modulated and unmodulated wide band noise on the N1 using a tonal stimulus. In the unmodulated condition, there were no measurable waveforms, likely because of the unfavorable SNR (-19 dB). However, in the modulated noise conditions, waveforms were present and responses had smaller amplitudes and longer latencies when compared with the quiet condition of tone only (no noise). The authors concluded that when a modulated noise is present, the N1-P2 is elicited when the tone is present and the noise has dropped to zero, suggesting a form of "dip-listening" or release from masking.

In addition to varied background noise type, AEP studies have also determined the effects of SNR between the masker and signal. In general, as the SNR decreases, or becomes unfavorable, the measured cortical waveforms become degraded. Whiting et al. (1998) used a steady-state broadband noise and a /ba/-/da/ contrast in an oddball paradigm to find that P3 peak latencies increased with increasing levels of noise. They also reported that the peak disappeared when the SNR was around 0 or -5 dB. Similarly, Kaplan-Neeman et al. (2006) also reported a differential effect of SNR, using a white noise masker, on the P3 peak. Specifically, the P3 latency elicited in response to speech syllables was significantly prolonged compared with the quiet condition for unfavorable SNRs (0, -3, and -6 dB). They suggested that prolonged latency recorded in the later stages of processing (P3) may be a reflection of difficulties in linguistic decision making as well as the reduced acoustic information resulting from unfavorable SNRs.

These previous studies suggest that masker properties affect the cortical AEPs in addition to behavioral discrimination tasks. They not only reveal the differential effect of SNR on the responses but also highlight the importance of both the spectral and temporal content of the maskers.

Masking can be categorized into two general types: energetic and informational. Energetic masking has been described as suppression of stimulus activity as a result of interference within the cochlea (Kidd et al. 2008). In contrast, informational

¹National Center for Rehabilitative Auditory Research, Portland Veterans Affairs Medical Center; and ²Department of Otolaryngology, Oregon Health & Science University, Portland, Oregon.

masking cannot be explained solely in terms of interactions in the auditory periphery. Durlach et al. (2003) described two factors involved in informational masking: uncertainty and similarity. Uncertainty is the difference between what the listener actually hears and what the listener expects to hear on a given trial. Similarity between a target and masker results in difficulties hearing them separately so as to discriminate the two sounds and correctly identify the target in the presence of the masker. Both uncertainty and similarity have been used to define informational masking and to develop theories and methods underlying its study.

Classic informational masking studies have routinely used tonal signals in the presence of a tonal masker. The rationale for the use of tones was described in a review article by Watson (2005), indicating that such signals linked informational masking studies to the large body of psychoacoustic research. Watson and Kelly (1981) used sequential tonal patterns to study informational masking. Neff and Green (1987) broadened that work to conditions of simultaneous targets and maskers, generating signals consisting of a tonal target in the presence of concurrent masker tones. The listeners' task was to detect a fixed tonal target in a background of complex multitone maskers that varied randomly on each presentation. This basic paradigm has been used in many subsequent studies, with modifications that have contributed further to understanding of informational masking, including issues involving spatial release from masking, training effects, stimulus or masker uncertainty, as well as contributions of frequency and intensity differences between targets and maskers (for review, see Kidd et al. 2008). Tonal signals presented in this informational masking paradigm have also been used to investigate the possible physiological correlates of informational masking (Gutschalk et al. 2008). In that 2008 study, responses to tonal signals in a background of masking tones were evaluated using magnetoencephalography. The results revealed that early responses (middle-latency steadystate responses) in an auditory detection task were present and robust whether participants perceptually detected the tone or not, whereas later responses (latencies of 50-250 msec) were correlated strongly with target detection. These results support the idea that cortical resources summoned in the later processing stages, within the auditory cortex, are involved in tasks in which a greater cognitive processing load is required, such as informational masking.

In the present study, the perception of tones and speech sounds in competing maskers was systematically investigated, under conditions thought to reflect either informational or energetic type maskers. Specifically, behavioral speech recognition and physiological responses (P3) were measured to signals presented in quiet, continuous speech-shaped noise, interrupted speech-shaped noise, and four-talker speech babble. To relate physiological correlates of masking and behavioral measures, AEPs, signal discrimination, and sentence recognition were measured. All maskers had nearly identical long-term spectra but varied either in their temporal characteristics or linguistic content. It was hypothesized that speech babble would provide more masking of a targeted speech signal than nonspeech maskers, suggesting an informational masking component beyond any energetic masking component. It was further hypothesized that P3 peak latency and amplitude, reflecting cognition and speech processing abilities (Polich et al. 1985; Picton 1992), would be affected differentially by speech babble in a manner analogous to the behavioral responses. The assessment of late potentials will provide a measure of neural encoding that may help to explain the variability in speech recognition across individuals listening in various kinds of noisy backgrounds.

MATERIALS AND METHODS

Subjects

Nine right-handed, normal-hearing listeners participated in this study (mean age = 25.7 yr, age range = 19–31 yr; four males and five females). Normal-hearing thresholds were defined as pure-tone thresholds of \leq 20 dB HL at octave frequencies from 250 to 8000 Hz with normal tympanometric measures (ASHA 1990). All participants were in good general health, had completed between 4 and 8 yr of higher education, and provided written informed consent. The experiment was completed with approval from the pertinent institutional review boards.

Signals and Maskers

Signals were presented monaurally to the right ear through Etymotic ER-2 insert earphones with a signal presentation level of 65 dB SPL and an overall root mean square (RMS) masker level of 68 dB SPL for a SNR of -3 dB. Participants were presented with eight different conditions consisting of two types of signals in four masker conditions. Masker conditions were (1) quiet with no competing masker, (2) a continuous noise masker, (3) an interrupted noise masker, and (4) a four-talker babble masker. The speech-shaped continuous masker (hereafter referred to as continuous) matched the long-term speech spectrum of the Institute of Electrical and Electronic Engineers (IEEE) sentence lists (IEEE 1969) which were used to obtain behavioral measures. The continuous noise was created by computing a Fast Fourier Transform based on a concatenated string of IEEE sentences, scrambling the values of the phase spectrum while maintaining the amplitude spectrum and performing a reverse Fast Fourier Transform. The speech-shaped interrupted noise (hereafter referred to as interrupted) had the same spectrum as the continuous noise, with random periods of noise and silence ranging in duration from 5 to 95 msec (Stuart & Philips 1996). These random periods of silence were chosen because they mimic changes in temporal patterns of speech. The fourtalker babble (hereafter referred to as babble) consisted of two female and two male talkers reading aloud for 10 min from printed prose (Lilly et al. 2011). All maskers except for the speech babble were low-pass filtered at 8 kHz. Inadvertently, the babble was not subjected to the low-pass filtering. Frequency analysis indicated that none of the target or masker stimuli had energy greater than 8.5 kHz. Overall RMS levels of the babble and the target signal with the broadest spectral content (sentences) showed a difference of 0.0025 dB in overall level. This difference is considered to be negligible and suggests that the minimal energy present above 8 kHz in the babble masker was unlikely to provide more energetic masking.

Two tone and two speech signals were generated to form the oddball contrasts for the AEP measures. The tone contrast consisted of 500 and 1000 Hz tones with rise/fall times of 9 msec; the speech contrast consisted of naturally produced female tokens /da/ and /ba/ from the University of



Fig. 1. Grand average waveforms for a subset of electrode sites demonstrate scalp topography differences between the deviant and standard waveforms obtained from the oddball paradigm. Waveforms are collapsed across signals (tone and speech) and masker type (continuous, intermittent, and babble). At the Pz electrode position, a robust P3 peak is present in response to the deviant signals (500 Hz and /da/), whereas the N1 peak, in response to the standard signals (1000 Hz and /ba/), is most robust at the Cz electrode. Electrode sites of TP9, TP10, FT10, and Iz depict inversion waveforms.

California- Los Angeles Nonsense Syllable test (Dubno & Schafer 1992). Tones and syllables were 150 msec in length. The natural speech syllables were truncated by removing all but the first 150 msec of the syllable including a 10 msec gradual offset which effectively reduced the vowel length and maintained the consonant in its entirety.

Auditory Evoked Potential Measurements

Cortical AEPs were obtained over two visits, each lasting about 2 h. P3 measurements were obtained using the tone (500-1000 Hz) and speech (/ba/-/da/) contrasts in separate oddball discrimination paradigms. The probability of presentation of the standards (/ba/ and 1000 Hz) was 0.8 and was 0.2 for the deviants (/da/ and 500 Hz). For each contrast, two blocks of 250 trials were presented for a total of 400 standard and 100 deviant presentations. Signal presentation was pseudorandomized so that a deviant was not the initial signal of a trial and no two deviants were presented consecutively. A stimulus onset asynchrony of 1100 msec (onset to onset) was used. Signal and masker conditions were randomized across subjects to reduce potential order effects. To prevent effects of frozen noise, three different variations of each background masker were produced and randomly mixed with the signals. Subjects were instructed to press a button in response to the deviant signal, placing emphasis on both speed and accuracy.

Simultaneous electrophysiology recordings to the deviant signal resulted in P3 AEPs.* Evoked potential activity was recorded using a 64 tin-electrode cap (Electro-Cap International, Inc.). P3 data were analyzed at the Pz electrode, where the P3 response was largest, using an average reference. Figure 1 depicts the scalp topography of the peak waveforms obtained from the standard and deviant signals of the oddball paradigm for the selected electrode sites. Cap position from nasion to inion was measured for each individual to ensure consistent cap placement between visits. The recording window consisted of 100 msec prestimulus and 700 msec poststimulus periods. Using the Neuroscan Scan 4.4 (Charlotte, NC) recording system, evoked responses were analog band-pass filtered on-line from DC to 100 Hz, amplified with a gain \times 500, and converted using an analog-to-digital sampling rate of 1000 Hz. Eye movement was monitored with electrodes located inferiorly and at the outer canthi of both eyes. Trials with eye-blink artifacts were corrected off-line, using Neuroscan software (Neuroscan, Inc 2007). After blink correction, trials containing artifacts exceeding \pm 70 μ V were rejected from averaging. The remaining sweeps were averaged and filtered off-line from DC to 30 Hz (filter slopes of 24 and 12 dB/octave, respectively). Averaging for the deviant signals were only completed for the sweeps that were considered hits (correct identification of the deviant signals). Peak amplitudes, relative to the baseline, and peak latencies, relative to the signal onset were determined by agreement of two judges. In the event that a disagreement between judges occurred, agreement was reached by evaluating the peak using temporal electrode inversion, global field power traces, and grand averages.

Behavioral Measurements

Participants completed a behavioral sentence-in-noise recognition task during a separate third session lasting 1 h or less. Signals were taken from the IEEE sentence lists (IEEE 1969) spoken by a female talker and presented in the four background noise conditions detailed earlier. The sentences-in-noise were presented monaurally to the right ear under ER-2 insert earphones at a level identical to that of the AEP measures; overall RMS sentence level was 65 dB SPL and masker level was 68 dB SPL (SNR = -3 dB). The IEEE sentence lists consist of low context sentences, each containing five keywords. Participants were instructed to repeat each sentence, and an investigator scored the percent correct key words identified over two 10-sentence lists for each masker condition (total per condition = 100 points).

^{*}The N1-P2 complex was measured from the standard signal of the oddball paradigm. These results are reported in Billings et al. (2011) and address release from masking as a function of different background noise types.



Fig. 2. Grand average P3 response waveforms in quiet and masker conditions (intermittent, continuous, and babble) to the deviant targets: 500 Hz tone (top) and the /da/ speech syllable (bottom). Responses are displayed for electrode Pz. Arrows mark approximate P3 peaks for the quiet and masker conditions. This graph shows robust responses in quiet and degraded responses (increased latency and decreased amplitude) under masking conditions. Speech-on-speech masking (/da/ presented in babble) results in the largest latency delays and amplitude reductions.

RESULTS

Auditory Evoked Potentials

The grand average waveforms for the P3 auditory evoked potential to the deviants of each contrast (500 Hz tone and /da/ speech syllable) as a function of masker condition are shown in Figure 2. This figure clearly depicts the robust waveform generated in response to the signals in quiet as well as the large difference in latency and amplitude resulting from all masking conditions for both signal types. The two panels also show the differential effects of masker types for the speech signal (lower panel) with little difference due to masker type for the tonal signal (upper panel). The mean amplitudes and latencies with SEs for the grand average waveforms are shown in Table 1.

Repeated measures analyses of variance (ANOVA), with $\alpha = 0.05$, determined the statistical significance of signal type (tone or speech) and masker type (continuous, intermittent, and babble) on latency, amplitude, and behavioral measures. The

measurements obtained in quiet were not used in the analysis. For P3 measures, Greenhouse-Geisser corrections (Greenhouse & Geisser 1959) were used where an assumption of sphericity was not appropriate. Four subsequent post hoc comparison statistics were made using *t*-tests between the three masker types and one between signal types, with Bonferroni corrections at $\alpha = 0.0125$, unless otherwise noted.

Latency Measures • Analysis of peak latencies indicated a significant effect of the signal ($F_{1.8} = 25.6, p = 0.001$) and masker type ($F_{1.2,9.5} = 38.1, p < 0.001$) and an interaction between signal and masker ($F_{12.16} = 16.4, p < 0.001$). Post hoc latency comparisons showed a differential effect of masker type for the deviant speech syllable /da/. When the target was speech (/da/) and the masker was speech babble, the latency was significantly prolonged when compared with the two speech-shaped noise maskers: continuous ($t_{[8]} = 7.2, p < 0.001$) and intermittent $(t_{ISI} = 7.5, p < 0.001)$. Latency measures between the continuous and intermittent noise for the speech signal was not statistically different ($t_{181} = 2.2, p = 0.061$). In addition, the latency measurements obtained from the speech signal /da/ in babble was significantly different than the 500 Hz tone signal in babble $(t_{181} = 6.9, p < 0.001)$. This effect is seen in the waveforms displayed in Figure 2, which depicts the longer latency for the informational masking (speech in babble) condition compared with all other waveforms.

Amplitude Measures • Analysis of peak amplitude of the P3 waveform indicated a significant effect of the signal $(F_{1181} =$ 7.7, p = 0.024) but failed to reach significance for masker type $(F_{12161} = 0.9, p = 0.414)$. There was a marginally significant interaction between signal and masker ($F_{[2,16]} = 3.6, p = 0.05$). Post hoc comparisons for the speech syllable/da/ indicated that the babble masker significantly reduced the amplitude of the waveform when compared with the continuous masker ($t_{181} =$ -3.9, p = 0.005) but not when compared with the intermittent type masker ($t_{181} = -1.6$, p = 0.142). Amplitude differences for the speech signal measured in the continuous and intermittent maskers failed to reach significance ($t_{181} = 1.1, p =$ 0.319). In addition, there was no significant difference between the amplitude of the speech and tone signals in the babble condition ($t_{181} = -2.5$, p = 0.036) at the Bonferroni-corrected α -level ($\alpha = 0.0125$).

Behavioral Measures

Table 1 shows the means and SEs for percent correct, reaction time, and *d*-prime (d') for each target-masker condition. Reaction time and d' were calculated from the button-press responses to the deviant signal in the oddball discrimination task. Reaction time was measured from signal onset and

TABLE 1. Auditory evoked potentials and behavioral measures

	Amplitude (μ V)		Latency (msec)		Percent Correct			Reaction Time (msec)		d'	
Masker	Tone	Speech	Tone	Speech	Tone*	Speech*	Sentence	Tone	Speech	Tone	Speech
Quiet	4.9 (0.7)	4.7 (0.6)	318.0 (10.0)	363.0 (7.7)	98.4 (0.5)	98.4 (0.3)	99.6 (0.2)	321.2 (14.8)	360.8 (18.5)	4.5 (0.2)	4.4 (0.2)
Intermittent	2.8 (0.4)	2.0 (0.5)	377.6 (10.1)	416.2 (7.4)	98.6 (0.3)	97.8 (0.4)	91.0 (1.6)	380.1 (13.7)	429.5 (11.7)	4.5 (0.2)	4.1 (0.2)
Continuous	2.3 (0.6)	2.3 (0.5)	398.6 (3.3)	431.6 (12.8)	98.1 (0.4)	97.3 (0.5)	77.3 (3.0)	428.4 (12.4)	430.9 (14.0)	4.3 (0.2)	4.0 (0.1)
Babble	2.6 (0.4)	1.4 (0.3)	394.3 (8.3)	504.2 (12.4)	97.7 (0.4)	89.6 (2.1)	36.8 (4.2)	420.9 (13.6)	513.7 (14.4)	4.1 (0.2)	2.6 (0.2)

Means and SEs for auditory evoked potential amplitude and latency values and behavioral tasks of percent correct, reaction time, and d'. *Percent correct is in PC_(max, yes/mo)



Fig. 3. Reaction time for the oddball discrimination task in quiet and masker conditions (intermittent, continuous, and babble) with SE of the mean. This bar graph displays prolonged reaction time for the speech /da/ condition when compared with the 500 Hz tone condition. In addition, the longest reaction time is apparent for the speech-on-speech condition (/da/ in babble).

the criterion free measure of d' was calculated from measurements of hit and false alarm rates. To compare the data obtained in the oddball discrimination task to the sentencein-noise task, measures of d' for tone and speech signals were converted to PC_[Max, yes/no] (Macmillan & Creelman 2005). **Reaction Time** • Figure 3 displays mean reaction times (in

msec) for the tone and speech signals for all masker conditions. Analysis of reaction times for the deviant 500 Hz tone and /da/ speech syllable indicated a significant main effect for signal type ($F_{[1,8]} = 15.7, p = 0.004$) and masker type ($F_{[2,16]}$ = 22.2, p < 0.001) as well as an interaction of signal and masker ($F_{12,161} = 20.5, p < 0.001$). When the speech target / da/ is presented in the babble masker, there is a statistically significant prolongation of reaction time when compared with the continuous ($t_{[8]} = -7.6$, p < 0.001) or intermittent ($t_{[8]} = -6.3$, p < 0.001) maskers. Reaction times for the speech signal in continuous and intermittent maskers were not statistically different ($t_{181} = 0.16$, p = 0.874). When reaction times of the babble condition are compared across signals (speech versus tone in babble), those times in response to the speech signal are significantly longer than those for the tone $(t_{181} = 5.5, p = 0.001)$. This suggests that the signal-by-masker interaction is reflected in the prolonged reaction time to the speech in babble condition.

Discrimination Task • Mean percent correct scores for syllable discrimination and sentence recognition tasks are presented in Figure 4. A repeated measures ANOVA was completed for the $PC_{[Max, yes/no]}$ discrimination scores and indicated a significant main effect of signal type ($F_{[1,8]} = 22.9, p = 0.001$) and masker type ($F_{[1,1,8,4]} = 13.2, p = 0.006$) as well as a significant signal by masker interaction ($F_{[1,8,3]} = 10.8, p = 0.010$). Specifically, post hoc comparisons indicated that discrimination abilities were significantly reduced when the target was the speech syllable /da/ in the babble when compared with the continuous ($t_{[8]} = 3.4, p = 0.009$) and intermittent maskers ($t_{[8]} = 3.7, p = 0.006$). Discrimination abilities for the speech signal in the continuous and intermittent maskers were not statistically different ($t_{[8]} = -1.2, p = 0.257$).

Recognition Task • For the sentence-in-noise recognition task, a repeated measures ANOVA was run in a 1 (signal type = sentences) \times 3 (masker type) design. Results revealed



Fig. 4. Percent correct scores for the oddball discrimination and sentence recognition tasks in quiet and masker conditions (intermittent, continuous, and babble) for the three target signals: 500 Hz tone, /da/ speech syllable, and sentences, with SE of the mean. This graph shows the poorest percent correct recognition for the speech-on-speech condition (/da/ in babble). †The discrimination results depicted are PC [Max, vertrol scores.

a significant main effect of masker type ($F_{[2,16]} = 108.3$, p < 0.001). Further post hoc comparisons, with Bonferroni corrections for three comparisons ($\alpha = 0.0167$), revealed each of the three maskers to be significantly different from the others. Specifically, the comparisons were the continuous masker to the intermittent ($t_{[8]} = -5.0$, p = 0.001); the babble to the intermittent ($t_{[8]} = 12.2$, p < 0.001); and the babble to the continuous ($t_{[8]} = 9.9$, p < 0.001). These comparisons suggest a differential effect of masker type on sentence recognition in noise, with the intermittent masker providing the least amount of masking, followed by the continuous masker, and the babble providing the most masking, resulting in the worst recognition scores.

Physiological and Behavioral Comparisons

Sentence Recognition • Pearson correlations were calculated between the percent correct scores on the sentence-in-noise task and P3 latency measures for the /da/ speech syllable and 500 Hz tone signal, combined for each masker condition. Correlation results are shown in Figure 5. P3 latency measures to the speech (/da/) signal had a strong, negative correlation with the sentence-in-noise task and this model explained 56% of the variance (r = -0.749, $r^2 = 0.561$, p < 0.001). There was a weak nonsignificant correlation between the 500 Hz tone P3 latency values and the sentence-in-noise scores (r = -0.202, r^2 = 0.041, p = 0.143).

Reaction Time • Pearson correlations were also determined between the behavioral measures of reaction time and physiological P3 latency measures for the speech and tone signals, again combined for each masker condition. Figure 6 displays these correlation results with linear lines of best fit. A strong, positive correlation was found between reaction time and the speech (/da/) evoked P3 latency measures and explained 62% of the variance (r = 0.788, $r^2 = 0.620$, p < 0.001). A weak, nonsignificant correlation was determined between reaction time and the tone (500 Hz) evoked P3 latency measures (r = 0.125, $r^2 = 0.016$, p = 0.368).



Fig. 5. Individual data points were coded by masker type (intermittent, continuous, and babble) and plotted to display the relationship between the behavioral sentence-in-noise scores (% correct) and P3 latency measures (msec) for the /da/ speech syllable and the 500 Hz tone signals. Linear lines of best fit reveal a strong negative correlation between the physiological and behavioral measures for speech. Higher percent correct scores are associated with shorter P3 latencies.

DISCUSSION

The purpose of this study was to investigate how different masker types (informational and energetic) affect physiological measures (P3), behavioral measures (reaction time and percent correct), and the relationship between the two.

Herein, we present novel data describing the P3 auditory evoked response in the presence of a four-talker babble, presumed to include informational masking properties in addition to the energetic masking produced by speech-shaped continuous or intermittent noise. Recall that all maskers had nearly equivalent long-term spectra, and the intermittent masker contained some silent intervals. The four-talker babble contained few silent intervals but was clearly composed of speech and therefore would likely have been perceived as at least qualitatively similar to the speech targets. This similarity supports the view that the four-talker babble produced both informational and energetic masking.

Our results reveal that when the signal is a speech syllable, masker type plays an important role in cortical processing. This was evidenced by degradation of the evoked potential waveform in response to the speech syllable /da/ in babble relative to the



Fig. 6. Individual data points were coded by masker type (intermittent, continuous, and babble) and plotted to display the relationship between reaction time (msec) and P3 latency (msec) for speech /da/ and 500 Hz tone signals. Linear lines of best fit reveal a significant positive correlation between the physiological and behavioral measures for speech. Longer reaction times are associated with longer P3 latencies.

waveform for speech in continuous noise. Moreover, this difference in waveform degradation between the two masker types was not seen when the signal was the 500 Hz tone. This is important given that the most common complaint of hearing-impaired individuals is difficulty understanding speech in the presence of competing talkers. Speech perception in noise is typically assessed with behavioral measures, but the unknown neural underpinnings of this complicated process make generalizing about individual performance difficult. The results of this experiment demonstrate that AEPs may provide additional insight into speech perception, revealing deficits not recognized with current clinical measures. For example, these results show trends that amplitudes are smaller when speech is masked by speech, implicating a reduction in neural synchrony under informational versus energetic masking. One might expect that as the task requires more complex processing abilities, such as the conditions of speech in babble, an increase in amplitude would suggest more attentional resources allocated to signal detection in background noise, but these data indicate the opposite. Also, longer latencies measured under conditions of masking might signify greater uncertainty in cortical processing. Used together, AEPs and behavioral measures can help to improve understanding of where breakdowns occur along the auditory pathway and this differentiation may guide clinicians in choosing amplification devices or rehabilitation programs that are based on individual needs. Electrophysiological measures in a multitalker babble may also be useful for disorders of central auditory processing, as this study indicates a more detrimental effect of informational masking (thought to be primarily of central origin) on the cortical response.

Billings et al. (2011) evaluated the effect of masking, on obligatory auditory evoked responses, using the same noise types presented to the same participants as those studied here. The authors found that the N1 peak measurements of latency and amplitude, obtained in response to the standard signals in the oddball paradigm, were not significantly different between the four-talker babble and speech-shaped continuous masker conditions. In contrast, the findings presented here reveal that informational masking effects are reflected at the level of the cognitive P3 evoked response. Figure 7 shows individual subject data points comparing the N1 latency values from the standard signals, reported by Billings et al., to the P3 deviant signal latencies reported here. It should be noted that the target reported for the N1 is the /ba/ standard syllable, whereas the P3 target is the /da/ deviant syllable. This comparison highlights the significant effect of informational masking (speech in babble) on the P3 peak latency compared with energetic masking (speech in continuous) ($t_{[8]} = 7.2, p < 0.001$) and that this difference, between masker types, is not present in the N1 peak latency measures ($t_{[8]} = 0.05$, p = 0.96). The difference in latency between the earlier obligatory N1 recordings and the later cognitive P3 recordings in the informational masking condition (speech in babble) suggests that the physiological influence of the informational masker is more robust beyond the thalamocortical connections and primary and secondary auditory cortices that generate the N1 (Naatanen & Picton 1987; Eggermont 2007). The addition of informational masking engages cognitive processing at the level of the P3, suggesting that more complex processing is needed to perform the task (Picton 1992). These results are consistent with the cognitive factors that are thought to contribute to behavioral informational masking effects such as uncertainty, attention,



Fig. 7. Individual N1 latency values (reported by Billings et al., 2011) for the standard speech signal (/ba/) and the individual P3 latency values (in the present study) for the deviant speech signal (/da/), in the continuous and babble masker types. This figure displays the effects of an informational masker (babble) when compared with an energetic masker (continuous) on latency measures of the N1 and P3 peaks. There is a significant difference in mean latency values between the continuous and babble masker types for the P3 peak. This difference between masker type is not present in the obligatory N1 peak latency measures.

and memory (Kidd et al. 2008). These results warrant further research to determine if training or attentional effects can be used to alter the amount of informational masking that occurs within the auditory pathway. Further investigation may also help to determine if similarity-based informational masking (target and masker are similar) is a consequence of neural competition at levels of processing above the auditory periphery (Watson 2005).

Another intriguing result of the present study was the absence of a release from masking for the P3 and associated discrimination data, whereas release from masking was present in the sentence recognition task (see Fig. 4). Temporal characteristics of the noise, such as amplitude modulation, have been known to improve speech recognition in the presence of a background masker (Festen & Plomp 1990). In addition to masker modulation, other behavioral research studies suggest that an interrupted masker with periods of random silence may provide the normal-hearing listener with a release from masking, improving speech-perception-in-noise abilities (Stuart & Philips 1996). Consistent with the literature, the intermittent masker used here had the smallest effect (higher percent correct) on sentence recognition, possibly because of this release from masking. However, the syllable and tone discrimination task results did not show a release from masking for the intermittent noise because of ceiling effects (i.e., the continuous noise did little to mask the syllables, and so no release from masking was possible). In addition, the P3 waveforms, under the continuous and intermittent maskers, are consistent with the high discrimination scores. If there were a release from masking, one would expect the responses recorded to the intermittent masking to be more robust than the waveforms obtained in the continuous masker, with greater differences in amplitude and latency. This difference between syllable-in-noise and sentence-in-noise tasks may have occurred because the sentence recognition was an open-set task whereas syllable discrimination was a closed-set task. Influences of speech-on-speech

masking introduce problems of not only allocating resources to one sound or another but also tracking one particular sound over time. The behavioral data obtained from sentence recognition in noise required tracking of a voice over time. In contrast, the discrimination tasks (syllables in noise) may not have required allocation of more resources to track the sound over time particularly when the masker was not speech-like, resulting in this ceiling effect in discrimination scores. It is interesting that there seems to be a differential use of processing, even when uncertainty (closed-set task) is reduced from the signal in a speech-on-speech situation. This is evidenced by the significant decrease in discrimination scores and increase in P3 latency for the speech-in-babble condition.

The behavioral measures of reaction time and percent correct support the electrophysiological results. Specifically, the results showed that informational masking in the form of speech-on-speech masking plays an important role in recognition. This is interesting because four-talker babble has been considered to be more similar to a speech-shaped continuous masker and has been shown to impair recognition similarly (Miller 1947). However, our results suggest four-talker babble is a more effective masker of speech syllables than a speechshaped continuous or intermittent noise. This is likely due to the informational content of the babble, and the qualitative similarity of the babble masker to the speech signal. It is also notable that reaction time was a more sensitive measure than percent correct across masker type for both the speech syllable and tone signal. This is reflected by increases in reaction time with the addition of the maskers and is consistent with previous literature (e.g., Martin et al. 1997; Kaplan-Neeman et al. 2006). In contrast, for measures of percent correct discrimination, scores for the 500 Hz tone signal in all masker conditions were at ceiling (high recognition scores), as were most of the masker conditions with the speech /da/ signal. The one exception is the speech /da/ in babble (speech-on-speech condition), which was the only condition with a significant decrease in percent correct discrimination (poor recognition scores). This suggests that participants took longer to make a distinction between the speech syllables and were less likely to make a correct distinction (hit).

Correlations found in this study relating increases in reaction time and latency are consistent with previous studies (Martin et al. 1997; Whiting et al. 1998; Martin & Stapells 2005). Reaction time for informational masking (speech in babble) is significantly correlated with longer latencies and suggests longer neural conduction time (Picton et al. 1985). It is interesting that we also found a significant negative correlation between the sentence-innoise scores and speech signal P3 latencies, which indicates that as the task became more difficult, percent correct recognition decreased and the latency increased. This correlation seems to be driven by the measures obtained in the babble masker as they contained the longest latencies and poorest recognition scores.

CONCLUSIONS

The results of this study revealed a detrimental effect of an informational masker on speech perception, as expected, in behavioral measures but also at the cortical level of neural encoding. Specifically, the study revealed the following:

- 1. Informational masking interfered with neural encoding as evidenced by prolonged P3 latencies for the speech-onspeech informational masking condition.
- 2. Informational masking interfered with speech discrimination as evidenced by reduced syllable discrimination in the four-talker babble masker when compared with other masker types. Furthermore, speech recognition was significantly degraded in the four-talker babble masker for the behavioral sentence-in-noise task.
- 3. There was no release from masking for the interrupted masker type in the physiological data, but it is interesting that this release was present in the behavioral sentence-in-noise task.
- 4. Reaction times for the speech syllable and sentence recognition scores were each significantly correlated with P3 latency measures, indicating an important association between perception and underlying physiology.

Overall, these data confirm a differential effect of competing masker types, especially multitalker babble, on AEPs as well as on behavioral responses and document a correspondence between cortical activity and perception.

ACKNOWLEDGMENTS

This research was supported by a grant R01 DC 00626 from the National Institute on Deafness and Other Communication Disorders. Support was also provided by the Department of Veterans Affairs, Veterans Health Administration, Rehabilitation Research and Development Service [Career Development grants C6116W (to M.R.M.) and C6971M (to C.J.B.), and Senior Research Career Scientist award (to M.R.L.)]. The work was supported with resources and the use of facilities at the Portland VA Medical Center.

The contents of this article do not represent the views of the Department of Veterans Affairs or of the U.S. Government.

Address for correspondence: Keri O'Connell Bennett, AuD, National Center for Rehabilitative Auditory Research, Portland Veterans Affairs Medical Center-NCRAR, 3710 SW US Veterans Hospital Road, Portland, OR 97239, USA. E-mail: keri.oconnell2@va.gov.

Received April 4, 2011; accepted August 3, 2011.

REFERENCES

- American Speech-Language Hearing Association (ASHA). (1990). Guidelines for screening for hearing impairments and middle ear disorders. ASHA, 32(Suppl 2), 17–24.
- Androulidakis, A. G., & Jones, S. J. (2006). Detection of signals in modulated and unmodulated noise observed using auditory evoked potentials. *Clin Neurophysiol*, 117, 1783–1793.
- Billings, C. J., Bennett, K. O., Molis, M. R., et al. (2011). Cortical encoding of signals in noise: Effects of stimulus type and recording paradigm. *Ear Hear*, 32, 53–60.
- Billings, C. J., Tremblay, K. L., Stecker, G. C., et al. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear Res*, 254, 15–24.
- Dubno, J. R., & Schaefer, A. B. (1992). Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normalhearing listeners. *J Acoust Soc Am*, 91, 2110–2121.

- Durlach, N. I., Mason, C. R., Kidd, G. Jr, et al. (2003). Note on informational masking (L). JAcoust Soc Am, 1113, 2984–2987.
- Eggermont, J. J. (2007). Electric and magnetic fields of synchronous neural activity. In: Burkard, R. F., Don, M., Eggermont, J. J. (Eds). *Auditory Evoked Potentials* (pp. 3–21). Philadelphia: Lippincott Williams & Wilkins.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech perception threshold for impaired and normal hearing. J Acoust Soc Am, 88, 1725–1736.
- Greenhouse, W. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.
- Gutschalk, A., Micheyl, C., Oxenham, A. J. (2008). Neural correlates of auditory perceptual awareness under informational masking. *Plos Biol*, 6, e138.
- Institute of Electrical and Electronic Engineers. (1969). *IEEE Recommended Practice for Speech Quality Measures*. New York: IEEE.
- Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., et al. (2006). Identification of syllables in noise: Electrophysiological and behavioral correlates. *J Acousc Soc Am*, 120, 926–933.
- Kidd, G., Mason, C. R., Richards, V. M., et al. (2008). Informational Masking. In: Yost WA, Popper AR, Fay RR (Eds). Auditory Perception of Sound Sources (pp, 143–189). New York: Springer.
- Lilly, D. J., Hutter, M. M., Lewis, M. S., et al. (2011). A "virtual cocktail party"—Development of a sound-field system and materials for the measurement of speech intelligibility in multi-talker babble. J Am Acad Audiol, 22, 294–305.
- Martin, B. A., Sigal, A., Kurtzberg, D., et al. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. J Acoust Soc Am, 101, 1585–1599.
- Martin, B. A., Kurtzberg, D., Stapells, D. R. (1999). The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. J Speech Lang Hear Res, 42, 271–286.
- Martin, B. A., & Stapells, D. R. (2005). Effects of low-pass noise masking on auditory event-related potentials to speech. *Ear Hear*, 26, 195–213.
- Macmillan, N. A., & Creelman, C. D. (2005). Detection Theory: A User's Guide (2nd ed., pp. 369). New Jersey: Lawrence Erlbaum Associates.
- Miller, G. A. (1947). The masking of speech. Phsycol Bull, 44, 105-129.
- Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24, 375–425.
- Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept Psychophys*, 41, 409–415.
- Neuroscan, Inc. (2007). SCAN 4.4—Vol II, Edit 4.4: Offline Analysis of Acquired Data (Document number 2203, Revision E, pp. 141–148). Charlotte, NC: Compumedics Neuroscan.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. J Clin Neurophysiol, 9, 456–479.
- Polich, J., Howard, L., Starr, A. (1985). Stimulus frequency and masking as determinants of P300 latency in event-related potentials from auditory stimuli. *Biol Psychol*, 21, 309–318.
- Stuart, A., & Philips, D. P. (1996). Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear Hear*, 17, 478–489.
- Watson, C. S. (2005). Some comments on informational masking. Acust Acta Acust, 97, 502–512.
- Watson, C. S., & Kelly, W. J. (1981). The role of stimulus uncertainty in the discrimination of auditory patterns. In D. J. Getty & J. H. Howard Jr. (Eds). *Auditory and Visual Pattern Recognition* (pp. 37–59). Hillsdale, NJ: Erlbaum.
- Whiting, K. A., Martin, B. A., Stapells, D. R. (1998). The effect of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Ear Hear*, 19, 218–231.