Selective listening of concurrent auditory stimuli: An event-related potential study

Aparna Rao\textsuperscript{a, *}, Yang Zhang\textsuperscript{a,b,**}, Sharon Miller\textsuperscript{a,1}

\textsuperscript{a} 115 Shevlin Hall, Department of Speech-Language-Hearing Sciences, 164 Pillsbury Drive SE, University of Minnesota, MN 55455, USA
\textsuperscript{b} Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN 55455, USA

** Corresponding author. Tel.: +1 612 624 7586; fax: +1 612 624 7586.
** Corresponding author. Tel.: +1 612 624 7818; fax: +1 612 624 7586.
E-mail addresses: raoxx098@umn.edu (A. Rao), zhang470@umn.edu (Y. Zhang), mill1979@umn.edu (S. Miller).

\begin{abstract}
This study employed behavioral and electrophysiological measures to examine selective listening of concurrent auditory stimuli. Stimuli consisted of four compound sounds, each created by mixing a pure tone with filtered noise bands at a signal-to-noise ratio of $-15$ dB. The pure tones and filtered noise bands each contained two levels of pitch. Two separate conditions were created: the background stimuli varied randomly or were held constant. In separate blocks, participants were asked to judge the pitch of tones or the pitch of filtered noise in the compound stimuli. Behavioral data consistently showed lower sensitivity and longer response times for classification of filtered noise when compared with classification of tones. However, differential effects were observed in the peak components of auditory event-related potentials (ERPs). Relative to tone classification, the P1 and N1 amplitudes were enhanced during the more difficult noise classification task in both test conditions, but the peak latencies were shorter for P1 and longer for N1 during noise classification. Moreover, a significant interaction between condition and task was seen for the P2. The results suggest that the essential ERP components for the same compound auditory stimuli are modulated by listeners' focus on specific aspects of information in the stimuli.
\end{abstract}

1. Introduction

Auditory selective attention refers to the ability to focus on pertinent sounds while ignoring others (Naätänen, 1988; Posner and Driver, 1992). It is an important skill needed for everyday living and communication. Conceptually, selective attention involves the dual cognitive processes of focus on targets or relevant aspects and suppression of irrelevant distractors. When selective attention is successful, focus on targets is enhanced and suppression of distractors is improved (Melara et al., 2002). The effects of attentional processes have been at the forefront of research in auditory neuroscience (e.g., Alho, 1992; Alain and Arnott, 2000; Hugdahl et al., 2003; Alho et al., 2006; Sabri et al., 2008). Methodologically, most of the studies have compared auditory responses in attentive conditions relative to inattentive conditions (e.g., watching a silent video or reading a book). It is not well understood how brain responses are affected by selectively attending or listening to different acoustic/perceptual sources in compound auditory stimuli.

The aim of the present study was to investigate brain mechanisms that support selective listening of concurrent auditory stimuli. The study was motivated by previous auditory neuroscience research, which demonstrated both separability and interference of different perceptual dimensions in speech stimuli (von Kriegstein et al., 2003, von Kriegstein and Giraud, 2004; Bonte et al., 2009; Zhang et al., 2009). For example, functional magnetic resonance imaging data showed that selectively listening to the speaker's voice information increased activation in the right superior temporal sulcus, whereas attention to the verbal content of the identical speech stimuli led to greater activation in the left superior temporal sulcus (von Kriegstein and Giraud, 2004). The time course of neural activation for selective listening has been studied with the ERP technique, which is known for its millisecond resolution (Luck, 2005). Kaganovich et al. (2006) reported differences in the amplitudes of the positive ERP peaks, P1, P2 and P3 when subjects attended to different informational dimensions of spoken words. The effects were attributed to a number of factors, including (a) the intrinsic significance of voice information to the auditory system, (b) the functional asymmetry of the human brain...
for extraction of linguistic and paralinguistic information, (c) talker interference for vowel categorization, and (d) attention-capturing qualities of human voice that affect late-occurring ERP components. However, it remains unknown whether the modulations on ERP components by selective listening are restricted to processing the special vocal and semantic dimensions in spoken language.

In the present ERP study, the brain mechanisms of selective listening were investigated by having participants attend to two distinct auditory sources concurrently available in compound nonspeech stimuli. Of particular interest was how subjects’ responses were affected by manipulations of cognitive effort and acoustic interference in two selective listening tasks and two test conditions. The key measures of our study were behavioral accuracy, reaction time, and the amplitudes and latencies of the auditory P1, N1, and P2 components. Our auditory stimuli consisted of pure tones (two levels of pitch) mixed with filtered noise bands (also two levels of pitch) with the intensity of tones set to +15 dB relative to the filtered noise. The stimulus design took into consideration of two facts: (a) The lack of equality in physical or perceptual properties among competing sounds is prevalent in the natural listening environment, and (b) the concurrent sounds do not necessarily integrate perceptually to form representational dimensions of one auditory object. Thus, our design allowed an examination of basic auditory processes of selective listening that are not restricted to linguistic stimuli. The use of nonspeech sounds also eliminated potential confounding factors such as semantic associations, word frequency, speaker familiarity, and speech quality.

Attention-induced changes in ERP components and their functional significances have been studied extensively. For example, attention could trigger changes as early as 50 ms after stimulus onset in the P1 peak of the ERP waveform (Woldorff and Hillyard, 1991). The positive (P1, P2 and P3) and negative peaks (N1 and N2) of the ERP waveforms are generally magnified in attended conditions when compared with unattended conditions (see reviews in Hillyard and Kutas, 1983; Giard et al., 2000; Näätänen, 1992). Unlike the previous studies, our primary interest in selective listening was to examine the specific role of cognitive effort in the presence of interference from a competing sound source. The study design adopted Garner’s interference paradigm (Garner, 1974) with two test conditions; the non-target sounds either varied in pitch (varying background condition) or were held constant (constant background condition). The constant background condition was introduced to control the interfering source to evaluate participants’ performance and the modulations on the ERP components. The intensity levels of the tones and the filtered noise bands in our stimuli were selected to avoid peripheral masking effects and to create a built-in perceptual bias for the tones. When participants classified the stimuli in separate blocks for the tone classification and noise classification tasks, the cognitive efforts needed to do the tasks were different while the compound auditory stimuli were physically identical. As the tones in our compound stimuli had a favorable signal-to-noise ratio of +15 dB, classification of tones was expected to be easier than classification of filtered noise. Similarly, classification of filtered noise would be more challenging as participants had to suppress the distracting tones which were higher in intensity (see Melara and Algol, 2003).

A domain-general hypothesis was tested. The interaction between cognitive effort and auditory interference during selective listening would hypothetically produce different modulation effects on the different ERP components for nonspeech stimuli as was shown for speech stimuli in previous studies. In particular, we predicted that there would not be a simple overall amplitude gain for the P1, N1, and P2 components as a function of task difficulty. Rather, the ERP components might reflect different operations at the cortical level in processing information about the auditory stimuli. If the auditory P1 response is affected by arousal and effort, an enhancement could be seen during classification of filtered noise for both test conditions. The N1 response is known to be affected by attentional effort (Hillyard et al., 1973; Woldorff and Hillyard, 1991; Näätänen, 1982) and stimulus salience (Melara et al., 2002; Escera et al., 1998). If stimulus bias takes precedence regardless of cognitive effort, a greater N1 may be expected during classification of tones. If cognitive effort is the overriding factor, an enhanced N1 would be observed during classification of filtered noise due to the difficulty of the task. The P2 component, which reflects suppression of irrelevant information and extraction of features for the response in the context of processing concurrent auditory sources (Garcia-Larrea et al., 1992; Tong and Melara, 2007), may be influenced by the inherent bias in the stimuli to favor classification of tones.

Our paradigm also allowed us to assess congruency effects, which provided information about how participants performed when the pitch of the target and the concurrent stimuli were similar (i.e., congruent) and when the concurrent stimulus conflicted with the pitch of the target stimulus (i.e., incongruent). Congruity effects were analyzed in both conditions for both tasks. We hypothesized better performance on congruent trials compared with incongruent trials for both classification of tones and classification of filtered noise (see MacLeod, 1991 for a review).

### 2. Materials and methods

#### 2.1. Participants

Ten individuals (4 males and 6 females) with normal hearing and no history of cognitive, speech, or language difficulties participated. All participants were right-handed with a handedness index score of +1 as determined using the Edinburgh handedness inventory (Oldfield, 1971). None of the participants had received formal musical training. The participants ranged in age from 18 years to 29 years, with the mean at 21 years. Participants were paid for their time. Informed consent was obtained from all participants in compliance with approval for the study by the Human Subjects’ Protection Program at the University of Minnesota.

#### 2.2. Stimuli and conditions

Four different complex sounds were created by mixing pure tones and filtered noise (see Table 1). The pure tones were 1000 Hz and 3000 Hz in frequency. The filtered noise bands were generated by low-pass filtering white noise at 2000 Hz (low-pitch filtered noise) and band-pass filtering between 2000 Hz and 4000 Hz (high-pitch filtered noise). The pure tones were set 15 dB above the level of the noise. The imbalance prevented simultaneous masking of pure tones by filtered noise. We did not aspire to achieve equal discriminability for the tone classification and noise classification.

<table>
<thead>
<tr>
<th>Noise band (low)</th>
<th>Pure tone (low) 1000 Hz</th>
<th>Pure tone (high) 3000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low tone, low noise</td>
<td>Low tone, low noise</td>
<td>High tone, low noise</td>
</tr>
<tr>
<td>High tone, low noise</td>
<td>Low tone, high noise</td>
<td>High tone, high noise</td>
</tr>
</tbody>
</table>

The handedness score was calculated using the formula (total score for right − total score for left)/(total score for right + total score for left).
tasks because this imbalance was necessary to answer our research questions. Listeners in everyday situations are confronted with stimuli that differ considerably in their acoustic and perceptual properties. In our study, the two tones (low- and high-pitch) and filtered noise bands (low- and high-pitch) were combined to generate four stimuli (i.e., 1000 Hz tone with low-pass filtered noise, 1000 Hz tone with band-pass filtered noise, 3000 Hz tone with low-pass filtered noise, and 3000 Hz tone with band-pass filtered noise). All stimuli were 100 ms in duration with 10 ms rise/fall times. Stimuli were presented through insert earphones (Etymotic Research ER-3A) binaurally at 60 dB above the participant’s threshold for 1000 Hz sine wave tone. The interstimulus interval was randomized between 1900 and 2100 ms.

The test conditions were depicted in Table 2. In the constant background condition, the targets (either tone or noise) varied in pitch level, and the non-targets in the compound stimuli held constant in pitch as background stimuli. For example, when participants were classifying the pitch level of the tones, the pitch of the filtered noise was held constant and vice versa. Separate blocks were created with high-pitched background stimuli and low-pitched background stimuli. In the varying background condition, all four compound stimuli were presented randomly in the same block. In addition to the target varying in pitch, the non-target background stimuli also varied in pitch.

### 2.3. Procedure

Participants were seated comfortably in an electrically and acoustically shielded booth and completed the experiment in two sessions. Each test session lasted 2 h with breaks as necessary. Participants were instructed to press a button on a keyboard which recorded responses with sub-millisecond accuracy (Direct IN-PCB keyboard from Empirisoft Corp.). The right button was pressed when the pitch was labeled as ‘high’ and the left button was pressed when the pitch was labeled as ‘low.’ The presentation order of the tone classification blocks and the noise classification blocks was counterbalanced with each test condition. ERP and behavioral data were recorded simultaneously for off-line analysis. The experiment started with short practice to ensure that the participants understood the listening tasks and test conditions. All participants were able to identify and discriminate the pitch levels of the tones and the pitch levels of the filtered noise bands, when the tones and noises were presented in isolation.

### 2.4. Behavioral data analysis

Behavioral accuracy and reaction times (RT) were measured as participants performed the pitch classification for the tones or noises of the compound stimuli. Sensitivity (d-prime) and response bias (c or criterion) were calculated using signal detection theory (Macmillan and Creelman, 1991). The signal detection theory accounts for both, the sensory process, which converts the sensory stimulation to an internal representation, and a decision process which relies on the output of the sensory process to make a decision. The sensory and decision processes are represented by sensitivity and criterion respectively. The two measures are independent of each other. Repeated-measures ANOVAs were performed separately on the dependent variables to compare performance across the two test conditions (i.e., constant background condition and varying background condition) and tasks (i.e., classification of tones and classification of filtered noise). To measure congruity effects, ANOVAs were performed separately on the dependent variables with condition, task and type of stimuli (congruent vs. incongruent) as factors.

### 2.5. ERP recording and analysis

The EEG data were collected using the Advanced Neuro Technology EEG system and a 64-channel Waveguard cap. The data were digitized using a sampling frequency of 512 Hz and filtered between 0.016 Hz and 200 Hz. Off-line analysis included artifact rejection of trials with potentials exceeding $+/- 50 \, \mu V$, low-pass filtering at 30 Hz and averaging trials using an epoch of 900 ms, which included a pre-stimulus baseline of 100 ms. The ground electrode was positioned at AFz and linked mastoids were used as the reference during off-line analysis. The average impedance of electrodes was below 5 Kohm. Active shielding technology on the Waveguard cap allows high-quality EEG recording in conditions where skin impedances are relatively higher than conventional standards (Vanhatalo et al., 2008).

The ERP waveforms were analyzed using peak measures of individual electrodes. Peak amplitudes and latencies for P1, N1 and P2 were extracted from the ERP waveforms for each subject. Latency ranges used for peak extraction were determined for each subject based on visual examination of the waveforms. On an average, the latency ranges used for the analysis were as follows: 35–80 ms for P1, 70–190 ms for N1 and 160–260 ms for P2.

---

**Table 2**
The conditions, blocks, stimuli and tasks performed by the participants are summarized below.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Block</th>
<th>Stimuli</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant background</td>
<td>1</td>
<td>1000 Hz tone, high-pitch filtered noise</td>
<td>Classification of pitch of tones: subject is asked to judge tones as “high” or “low”</td>
</tr>
<tr>
<td>Constant background</td>
<td>2</td>
<td>3000 Hz tone, high-pitch filtered noise</td>
<td>Classification of pitch of tones: subject is asked to judge tones as “high” or “low”</td>
</tr>
<tr>
<td>Constant background</td>
<td>3</td>
<td>1000 Hz tone, high-pitch filtered noise</td>
<td>Classification of filtered noise: subject is asked to judge filtered noise as “high” or “low”</td>
</tr>
<tr>
<td>Constant background</td>
<td>4</td>
<td>1000 Hz tone, low-pitch filtered noise</td>
<td>Classification of filtered noise: subject is asked to judge filtered noise as “high” or “low”</td>
</tr>
<tr>
<td>Varying background</td>
<td>5</td>
<td>1000 Hz tone, high-pitch filtered noise</td>
<td>Classification of pitch of tones: subject is asked to judge tones as “high” or “low”</td>
</tr>
<tr>
<td>Varying background</td>
<td>6</td>
<td>3000 Hz tone, low-pitch filtered noise</td>
<td>Classification of pitch of filtered noise: subject is asked to judge filtered noise as “high” or “low”</td>
</tr>
</tbody>
</table>
Electrodes were grouped to analyze differences in amplitude across hemispheres (right and left) and scalp regions (frontal, central, parietal, midline frontal, midline central and midline parietal). The electrodes were grouped as follows: frontal electrodes included F7, F5, F3, FT7, FC5, FC3 and the corresponding electrodes on the right hemisphere. Central electrodes included T7, TP7, C5, C3, CP5, CP3 and the corresponding electrodes on the right hemisphere. Parietal electrodes included P7, P5, P3, PO7, PO5, PO3 and the corresponding electrodes on the right hemisphere. Midline frontal electrodes included F1, Fz, F2, FC1, FC2, and FC2. Midline central electrodes included C1, Cz, C2, CP1, CP2 and CP2. Midline parietal electrodes included P1, Pz, P2 and POz. Repeated-measures ANOVA were conducted separately on peak amplitudes and latencies for the two conditions (i.e., constant background condition and varying background condition) and the tasks (i.e., classification of tones vs. classification of filtered noise). Additional factors included two levels for hemisphere (left and right) and three levels for electrode region (frontal, central and parietal) to measure differences across electrode sites (see above). Greenhouse-Geisser corrections were applied whenever necessary. Congruity effects were calculated by ANOVA conducted separately on ERP peak amplitudes and latencies with condition, task and type of stimuli (congruent vs. incongruent) as factors.

Global field power (GFP) was also calculated to cross validate the conventional ERP waveform analysis (Lehmann and Skrandies, 1984). Unlike the waveform peak analysis at selected electrodes, the GFP provides an objective assessment of spatial scalp distribution in terms of the standard deviation of potential values for all electrodes at any sampling point in the recording epoch. The post-stimulus differences in GFPs at each sampling point between the two tasks or between the two conditions were assessed relative to the potential distribution in the 100 ms window pre-stimulus baseline. Z-scores were obtained and Bonferroni-corrected. Sustained latency intervals of at least 10 ms or longer (Zhang et al., 2005, 2009) were highlighted where the GFP amplitudes differed significantly between the listening tasks or test conditions.

3. Results

3.1. Behavioral data

D-prime values revealed that participants performed better in the constant background condition when compared with the varying background condition ($F(1,9) = 8.33, p < 0.05$) and for classification of tones ($F(1,9) = 36.44, p < 0.05$) when compared with classification of filtered noise (Fig. 1). The interaction between condition and task was not significant ($F(1,9) = 2.08$, ns). As expected, d-prime values were higher for congruent stimuli than for incongruent stimuli ($F(1,9) = 20.36, p < 0.0001$). The difference in d-prime for congruent and incongruent stimuli was greater for the varying background condition when compared with the constant background condition ($F(1,9) = 51.55, p < 0.0001$), and for classification of filtered noise when compared with the classification of tones ($F(1,9) = 17.76, p < 0.0001$).

Criterion for signal classification did not differ significantly across conditions. While participants had comparable criterion for congruent and incongruent trials during classification of tones, they were more conservative in responding to incongruent trials than to congruent trials during classification of filtered noise ($F(1,9) = 15.34, p < 0.0005$).

Consistent with the d-prime data, reaction times were longer for noise classification than for tone classification ($F(1,9) = 7.15, p < 0.05$), confirming that noise classification was more difficult. Reaction times were also longer for the varying background condition relative to the constant background condition ($F(1,9) = 20.048, p < 0.01$), suggesting that varying background caused more interference in target categorization than the constant background (Fig. 2). The interaction between condition and task was significant ($F(1,9) = 5.63, p < 0.05$). As can be seen in Fig. 2, the difference in reaction times between conditions was smaller for classification of tones compared with classification of filtered noise. Overall, participants were slower in responding to incongruent trials than to congruent trials ($F(1,9) = 5.23, p < 0.05$). Reaction times were considerably longer for the congruent trials compared with the congruent trials during the varying background condition ($F(1,9) = 16.53, p < 0.001$). By contrast, reaction times were comparable for congruent and incongruent trials during the constant background condition.

3.2. ERP peak analysis

The grand-mean waveforms for the two conditions and two tasks are shown in Fig. 3. Peak amplitudes and latencies of the three major peaks, P1, N1 and P2, were extracted from individual data and subjected to statistical analysis. A summary of the statistical findings and mean amplitudes and latencies of P1, N1 and P2 across the scalp are shown in Table 3. Amplitudes of P1 showed a main effect of condition. Greater amplitudes were seen for the varying background condition relative to the constant background condition ($F(1, 356) = 31.43$, $F(1,9) = 20.048, p < 0.01$), suggesting that varying background caused more interference in target categorization than the constant background (Fig. 2). The interaction between condition and task was significant ($F(1,9) = 5.63, p < 0.05$). As can be seen in Fig. 2, the difference in reaction times between conditions was smaller for classification of tones compared with classification of filtered noise. Overall, participants were slower in responding to incongruent trials than to congruent trials ($F(1,9) = 5.23, p < 0.05$). Reaction times were considerably longer for the congruent trials compared with the congruent trials during the varying background condition ($F(1,9) = 16.53, p < 0.001$). By contrast, reaction times were comparable for congruent and incongruent trials during the constant background condition.

\[ p < 0.001 \). A main effect of task was also observed in P1. Amplitudes were greater when participants classified the pitch of filtered noise \((F(1,356) = 7.189, p < 0.01)\). Amplitudes were not significantly different for congruent and incongruent trials \((F(1,1279) = 0.08, \text{ns})\). In the constant background condition, amplitudes were greater for incongruent trials compared with congruent trials; in the varying background condition, amplitudes were greater for the congruent trials compared with the incongruent trials \((F(1, 1279) = 83.31, p < 0.0001)\). A summary of statistical findings involving congruity is shown in Table 4.

While P1 latencies did not differ for the two conditions, a main effect of listening task was seen. The P1 latencies were shorter when participants classified the pitch of filtered noise \((F(1, 356) = 8.23, p < 0.01)\). There were no significant interactions between condition and task in P1. The frontal and central electrode sites greater P1 than the parietal sites during classification of filtered noise, resulting in an interaction effect of task and electrode region \((F(2,356) = 9.566, p < 0.001)\). A main effect of congruity was not seen for latencies of P1 \((F(1,1279) = 2.5, \text{ns})\). Latencies were comparable for congruent and incongruent trials in the varying background condition.

Table 3

Mean amplitudes and latencies for P1, N1 and P2 are listed along with a summary of the main effects and interactions.

<table>
<thead>
<tr>
<th>ERP peaks</th>
<th>Means</th>
<th>Means</th>
<th>Means</th>
<th>Means</th>
<th>Summary of statistical findings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant background</td>
<td>Varying background</td>
<td>Classification of tones</td>
<td>Classification of filtered noise</td>
<td></td>
</tr>
<tr>
<td>P1 Amplitudes</td>
<td>0.95 µV</td>
<td>1.23 µV</td>
<td>1.056 µV</td>
<td>1.13 µV</td>
<td>Main effect: condition ((p &lt; 0.001))</td>
</tr>
<tr>
<td>Latencies</td>
<td>52.97 ms</td>
<td>53 ms</td>
<td>53.55 ms</td>
<td>52.41 ms</td>
<td>Main effect: task ((p &lt; 0.01))</td>
</tr>
<tr>
<td>N1 Amplitudes</td>
<td>–4.52 µV</td>
<td>–4.03 µV</td>
<td>–4.15 µV</td>
<td>–4.4 µV</td>
<td>Main effect: condition ((p &lt; 0.001))</td>
</tr>
<tr>
<td>Latencies</td>
<td>116.5 ms</td>
<td>116.22 ms</td>
<td>116 ms</td>
<td>116.71 ms</td>
<td>Main effect: task ((p &lt; 0.001))</td>
</tr>
<tr>
<td>P2 Amplitudes</td>
<td>2.69 µV</td>
<td>3.14 µV</td>
<td>3.00 µV</td>
<td>2.80 µV</td>
<td>Main effect: condition ((p &lt; 0.001))</td>
</tr>
<tr>
<td>Latencies</td>
<td>212.79 ms</td>
<td>225.23 ms</td>
<td>214.6 ms</td>
<td>223.43 ms</td>
<td>Main effect: task ((p &lt; 0.01))</td>
</tr>
</tbody>
</table>

Fig. 3. Grand-mean ERP waveforms are shown (A) for the constant background condition and (B) for the varying background condition. Waveforms were averaged over frontal, central and parietal sites. Specific electrode sites included in each region are explained in the text.
Table 4
Summary of statistical findings involving congruity for P1, N1 and P2.

<table>
<thead>
<tr>
<th>ERP peaks</th>
<th>Summary of statistical findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Amplitudes</td>
<td>Interaction: condition × congruity (p &lt; 0.0001)</td>
</tr>
<tr>
<td>Latencies</td>
<td>Interaction: condition × congruity (p &lt; 0.0001)</td>
</tr>
<tr>
<td>N1 Amplitudes</td>
<td>Main effect: congruity (p &lt; 0.0001)</td>
</tr>
<tr>
<td>Latencies</td>
<td>Interaction: condition × congruity (p &lt; 0.0001)</td>
</tr>
<tr>
<td>P2 Amplitudes</td>
<td>Main effect: congruity (p &lt; 0.0001)</td>
</tr>
<tr>
<td>Latencies</td>
<td>Interaction: task × congruity (p &lt; 0.001)</td>
</tr>
</tbody>
</table>

background condition; latencies were longer for incongruent trials than congruent trials in the constant background condition (F (1,1279) = 14.43, p < 0.0001). Once again, latencies were comparable for congruent and incongruent trials during classification of filtered noise. Consistent with behavioral results, the P1 latencies were longer for the incongruent trials than the congruent trials during classification of tones (F(1,1279) = 10.85, p < 0.001).

Main effects of condition and task were significant for the N1 amplitudes. Amplitudes were larger for the constant background condition (F(1, 356) = 95.53, p < 0.001) relative to the varying background condition and when participants focused on pitch of filtered noise relative to the tone classification task (F (1,356) = 53.839, p < 0.001). The interaction between condition and task was not significant (F(1,356) = 0.19, ns). N1 amplitudes were greater for the incongruent trials compared with the congruent trials (F(1,1279) = 152.22, p < 0.0001).

While N1 latencies did not differ for the two conditions, a main effect of task was found, showing longer latencies during classification of filtered noise relative to tone classification (F(1, 356) = 4.89, p < 0.05). Latencies were shorter for the incongruent trials compared with the congruent trials (F(1,1279) = 64.95, p < 0.0001). While latencies were comparable for the congruent and incongruent trials in the constant background condition, latencies were shorter for incongruent trials than congruent trials in the varying background condition (F(1,1279) = 11.21, p < 0.001).

Furthermore, a hemispheric effect for the electrode sites was seen for N1 peak amplitudes (F(1, 356) = 20.416, p < 0.001). For classification of tones, N1 amplitudes were greater over the right hemisphere when compared with the left hemisphere. However, N1 amplitudes over the two hemispheres were comparable for classification of filtered noise, which was more difficult than classification of tones. The significant interaction between task and hemisphere for N1 is shown in Fig. 4.

For P2 amplitudes, main effects of condition and task were seen. Amplitudes were larger for the varying background condition (F (1,354) = 53.797, p < 0.001) and for the classification of tones (F(1, 354) = 20.278, p < 0.001). A significant interaction was found between condition and task for P2 amplitudes (F(1,354) = 53.79, p < 0.001) (see Fig. 5A). While P2 amplitudes were comparable for the two conditions during classification of filtered noise, amplitudes were greater for the varying background condition relative to the constant background condition during classification of tones. P2 amplitudes were greater for incongruent trials compared with congruent trials (F(1,1277) = 109.38, p < 0.0001). In the constant background condition, amplitudes increased by a mean of 0.55 μV for incongruent trials compared with congruent trials; this increase was just 0.24 μV in the varying background condition (F (1,1277) = 92.93, p < 0.0001). An interaction between congruity and task revealed a similar pattern. For classification of tones, amplitudes increased by a mean of 0.67 μV for the incongruent trials compared with the congruent trials; this increase was just 0.12 μV for the classification of filtered noise (F(1,1277) = 95.38, p < 0.0001).

**Fig. 4.** Hemispheric difference in amplitude of N1 for classification of tones vs. classification of filtered noise.

**Fig. 5.** Interactions involving condition and task for (A) P2 amplitudes and (B) latencies.
P2 latencies revealed main effects of condition and task. Latencies were longer for the varying background condition ($F(1,354) = 174.33, p < 0.001$) and longer for classification of noise ($F(1,354) = 45.072, p < 0.001$). For P2 latencies, a significant interaction was found between condition and task ($F(1,354) = 174.33, p < 0.001$) (see Fig. 5B). The pattern was similar to the amplitude data with comparable latencies for the two conditions during classification of filtered noise and latencies that were longer for the varying background condition relative to the constant background condition during classification of tones. Latencies showed a main effect of congruity ($F(1,1277) = 10.01, p < 0.01$). Latencies were shorter for incongruent trials compared with congruent trials. An interaction between condition and congruity revealed comparable latencies for the incongruent and congruent stimuli for the constant background condition and shorter latencies for the incongruent trials compared with the congruent trials in the varying background condition ($F(1,1277) = 7.21, p < 0.01$). An interaction between task and congruity showed comparable latencies for congruent and incongruent trials during the classification of filtered noise and reduced latencies for incongruent trials relative to congruent trials during classification of tones ($F(1,1277) = 15.33, p < 0.0001$).

3.3. Global field power analysis

The global field power results of the grand-mean waveforms for classification of filtered noise and classification of tones are shown in Fig. 6. Because the GFP plots did not show polarity information, scalp potential topography was added to depict the ERP polarity for the GFP peaks. Latency intervals were highlighted where the difference of the two GFP waveforms were significantly different for consecutive sample points lasting at least 10 ms ($p < 0.05$). Consistent with the waveform analysis, there were significant GFP differences between conditions as well as between tasks for the P1, N1 and P2 components.

4. Discussion

4.1. Behavioral data

Listeners' sensitivity was poorer and reaction times were longer for the varying background condition when compared with the constant background condition. These results established that the varying background condition caused more interference and was more difficult for participants. Similarly, sensitivity was poorer and reaction times were longer when participants were required to classify the pitch of filtered noise when compared with classification of tones. This relative difficulty for the noise classification task was expected as there was perceptual built-in bias towards tone classification. Stimuli were designed with tones set 15 dB above the level of the filtered noise, favoring classification of tones. An interaction between condition and task was significant for reaction times, revealing long reaction times for the classification of filtered noise especially in the varying background condition. An interaction between condition and task was not significant for sensitivity, indicating that classification of filtered noise was more challenging of the two tasks in both conditions and participants could not overcome the in-built perceptual bias for tones. During classification of tones, participants were able to access target information and inhibit the effects of the filtered noise which was low in salience; however, this was not the case during classification of filtered noise, wherein subjects had to inhibit the effects of the tones which were highly salient. We were specifically interested in ERP findings in the presence of this asymmetry in task performance.

As expected, higher accuracy was achieved on classification of congruent trials than for classification of incongruent trials. When the pitch of the concurrent stimulus did not match the pitch of target, it caused a greater disruption in classification. This was true for both conditions. Participants experienced greater disruption on incongruent trials during classification of noise, which was the more difficult task. Congruity analysis on criterion once again indicated more conservative responses for incongruent trials during classification of filtered noise. Reaction times indicated slower responses to incongruent trials and the interaction between condition and congruity suggested slower responses to incongruent trials in the varying background condition compared with the constant background condition.

4.2. ERP data

ERP responses showed enlarged P1, reduced N1 and enlarged P2 for the varying background condition when compared with the constant background condition. ERP responses obtained during classification of filtered noise revealed greater P1 and N1 amplitudes. P2 amplitudes were larger during classification of tones for the varying background condition and were larger during classification of filtered for the constant background condition. These results were in accordance with our hypothesis, showing differential modulation effects on the ERP components depending on the source of information attended to, cognitive effort and interference from the non-target background sound. Further analysis of congruity revealed interesting findings on the kind of background sound that caused the greatest disruption to selective attention.

4.2.1. P1

Greater P1 amplitudes were noted in the varying background condition and for the classification of filtered noise. Studies have shown modulatory effects of attention during the early stages of processing in the visual (Mangun and Hillyard, 1990) and auditory domains (Woldorff and Hillyard, 1991). In our experiment, the varying background condition and the classification of filtered noise was more demanding, perhaps requiring greater attentional resources at an earlier stage of information processing. Classification along the salient informational source (i.e., tones) was less demanding, leading to a later and relatively reduced P1 for that condition. Kaganovich et al. (2006) found an enhanced P1 when
participants focused on the voice of the talker. In their study, stimuli were designed to be equally discriminable across the dimensions of vowel and talker, and their result was attributed to the salience of voice to the human auditory system. However, even if the task difficulty was controlled, there could be inherent inequality in accessibility of vocal and semantic information. It could be that classification of the talker required greater focus just like the greater cognitive effort needed for the classification of filtered noise in our study.

The origin of the auditory P1 includes thalamic pathways and cortical areas (Erwin and Buchwald, 1987; Liegeois-Chauvel et al., 1994), and the latency reflects delays in propagation through the peripheral and central auditory pathways (Eggermont et al., 1997). The earlier latencies seen in our study in the noise classification task suggests quicker propagation times that occurred with earlier selection when participants put forth greater cognitive effort.

An interaction between condition and congruity for P1 amplitudes suggests more effortful early processing for incongruent trials during the constant background condition. A greater P1 was seen for congruent trials compared with incongruent ones during the varying background condition. This may be an early indicator of attentional disruption caused by the non-target sound that was mismatched in pitch with the target sound. While participants were able to overcome this disruption in the easier constant background condition, it was not possible for the more challenging varying background condition. Behaviorally, the participants showed poorer sensitivity and slower reaction times for the incongruent stimuli in the varying background condition. For P1 latencies, the interactions between condition and congruity as well as between task and congruity suggest that the additional effort for the more challenging condition (varying background condition) and task (classification of filtered noise) facilitated quicker early processing of congruent stimuli than incongruent stimuli.

4.2.2. N1

The N1 is believed to reflect stimulus encoding and the formation of a trace of the eliciting stimulus in sensory memory (Naätänen and Picton, 1987). Reduced N1 amplitudes for the varying background condition when compared with the constant background condition may indicate the disruption of stimulus encoding and formation of the sensory memory trace for the more difficult condition. Additionally, N1 amplitudes were larger for the classification of noise. An enhanced N1 to attended stimuli has been the spotlight of several investigations (Hillyard et al., 1973; Woldorff and Hillyard, 1991; Naätänen, 1982) and two distinct accounts exist for the enhanced N1 seen to attended stimuli. In one view, an increased N1 to attended stimuli is attributed to the recruitment of additional neurons or to an improvement in the synchronicity of neuronal firing or both (Hillyard et al., 1973; Woldorff and Hillyard, 1991), thus pointing to an increase in the generator mechanism of the N1. A second hypothesis is that a sustained negativity (also known as the Processing Negativity) overlaps with the apparently exogenous N1 and extends beyond 400 ms based on interstimulus intervals (Naätänen, 1982). Further evidence comes from the scalp topography of processing negativity, which is believed to differ from the scalp topography of N1 (Naätänen et al., 1992). While we may not be able to disentangle the two sources in our study, both explanations seem plausible. An increased N1 triggered by greater attentional effort was seen when participants classified the pitch of filtered noise. Attention is known to increase neuronal responsiveness (Spitzer et al., 1988) and activity in brain regions responsible for processing those attributes (Corbetta et al., 1990). Attention to the noise bands may have led to stimulation of a greater region of the cortex and greater effort may have cause an increase in neural synchronicity as reflected by N1. On the other hand, the N1 peak enhancement may signal the overlap of endogenous components triggered by attention. In concurrence with the d-prime findings, the lack of significant interaction between condition and task for N1 amplitudes reveals that the classification of noise required greater attentional effort than classification of tones for both constant background and varying background conditions. The lack of interaction stemmed from the asymmetry in task performance in the constant background condition, and the asymmetry was maintained in the varying background condition (Melara and Mounts, 1993; Melara and Algom, 2003). Congruity analysis of N1 amplitudes further confirmed our explanation that greater attentional resources required for processing of incongruent stimuli led to higher peak amplitudes of N1. Longer N1 latencies for classification of noise suggests that a longer interval was required for processing filtered noise bands, although the propagation time coded by P1 was shorter for the same condition. The N1 is known to have its primary source generator in the auditory cortex, and the modulatory effects of attention have been shown on neurons in the primary cortex (Fritz et al., 2003, 2005). Congruity effects involving N1 latency revealed shorter latencies for incongruent trials although longer latencies were seen for the same with the P1 component. This was especially true for the varying background condition. Incongruent trials in our paradigm were associated with disruption of the P1 and increased effort reflected in the N1 latency range.

The interaction between condition and hemisphere revealed greater amplitudes over the right hemisphere to the N1 for tone classification relative to noise classification. During classification of filtered noise, the N1 amplitudes were comparable over electrodes on both hemispheres. In our study, we noticed the asymmetry with binaural stimulation. It is important to note here that asymmetric amplitudes on the scalp surface may not directly translate to asymmetric generators intracranially. Extensive research has confirmed that the N1 receives contributions from multiple neural generators in the frontal and temporal lobes (Naätänen and Picton, 1987; Woods, 1995) with the primary generators specifically localized in the superior temporal gyrus (Woldorff et al., 1993), Heschl’s gyrus, and the planum temporale (Godey et al., 2001). Hemispheric specialization for processing of auditory stimuli has been studied extensively with the classical view suggesting that the left hemisphere is specialized for processing speech. More recently, studies by Zatorre and Belin (2001) and Jamison et al. (2006) have provided evidence for hemispheric specialization for processing of nonlinguistic stimuli as well. Processing of temporal variations was lateralized to the left hemisphere and processing of spectral variation was lateralized to the right. In our study, hemispheric asymmetry was highlighted by the task of attending to one source of information versus the other; pitch processing of tonal stimuli, which had clear spectral characteristics, showed a preponderance over the right hemisphere (Tervaniemi and Hugdahl, 2003).

4.2.3. P2

A main effect of condition was observed in P2 amplitudes, which revealed greater values for the varying background condition relative to the constant background condition. In selective listening tasks, the P2 component has been linked to protecting against interference from irrelevant stimuli (Rif et al., 1991; García-Larrea et al., 1992; Melara et al., 2002), and to the process of comparing each stimulus trace to its perceptual representation in memory leading to stimulus categorization (Tong and Melara, 2007). Greater inhibition required for the varying background condition may have contributed to enlarged P2 amplitudes.
A main effect of task was also found. P2 amplitudes were greater for tone classification relative to noise classification. Even though it was also necessary to inhibit the interference of the filtered noise for tone classification, the tonal sounds in the compound stimuli were more salient to focus on. Thus we may speculate that the inhibitory process might play a minor role here because less cognitive effort was needed for tone classification. However, a number of factors may contribute to selective listening, including the intrinsic significance of tonal information to the auditory system, the degree of interference from non-target sounds, the functional asymmetry of the brain in extracting tonal information as against noise information, and the attention-capturing qualities of tonal stimuli. The significant interaction between condition and task supports the view that the P2 amplitudes are affected by both inhibition of interference and other factors that contribute to the extraction of tonal vs. noise information.

Post-hoc analysis further revealed that the magnitudes of P2 amplitudes were comparable across conditions for the classification of filtered noise. For the more difficult task (i.e., classification of filtered noise with the interfering non-target tones 15 dB above the noise), inhibitory mechanisms reflected by P2 were possibly functioning at a ceiling level for both conditions of varying background and constant background. By contrast, for the easier task (i.e., classification of tones), greater inhibitory processing was required for the varying background condition leading to higher amplitudes when compared to the constant background condition. Congruity effects are in agreement with the trends seen with condition and task effects of P2 amplitudes. Overall, P2 amplitudes were greater for incongruent trials compared with congruent trials. This can be explained by the greater inhibitory processing required when concurrent stimuli were conflicting. Additionally, greater gains were noted for incongruous stimuli than for congruous stimuli in the constant background condition compared with the varying background condition. The same was true for classification of tones compared with classification of noise. Therefore, it may be conjectured that in the more challenging situations reflected by P2 are constrained by ceiling effects.

Putative generators of P2 include the planum temporale and the association cortex (area 22) (Godfrey et al., 2001), with modulations by the reticular activating system (Rif et al., 1991). Overall, latencies were longer during classification of filtered noise compared with latencies during classification of tones. It is interesting to note that significant interaction between condition and task was found for P2 latencies as well. The results are consistent with the interpretation for a possible ceiling effect for inhibiting tonal interference during the classification of filtered noise. Although latencies were comparable for the two tasks for N1, latencies for P2 showed a significant difference, showing that processing indexed by the P2 mechanism took longer for the more difficult condition (i.e., classification of filtered noise). Our result is consistent with the findings that the pathways for the P2 mechanism are distinct from the N1 (see Crowley and Colrain, 2004). In particular, the maturation of course time of P2 is known to be shorter than the maturational time of N1 (Eggermont, 1988), providing evidence for two different pathways for the ERP responses (Ponton et al., 2000).

Congruity effects seen with P2 latencies revealed shorter latencies for incongruous trials compared with congruous trials, although we know that the incongruous trials were behaviorally more difficult to classify. Similarly the statistically significant interactions between condition and congruity and task and congruity seem to suggest that difficulty imposed by the factors of congruity, condition and task vary in their quality. As discussed earlier, the P2 under these circumstances may represent more than just inhibitory processing and access to representations.

4.3. Theoretical and clinical implications

Attention is known to modify processing at various levels in the auditory system (Hubel et al., 1959; Hillyard et al., 1973; Woldorff and Hillyard, 1991). At a cellular level, attention is known to enhance neuronal selectivity and responsiveness (Spitzer et al., 1988). The modulatory role of attention has been demonstrated in the primary, secondary and associated regions of the cortex (Polley et al., 2006; Fritz et al., 2003, 2007). Our selective listening results are consistent with the finding that neuronal responses changes are differentially modulated depending on the informational source attended to, cognitive effort needed to inhibit interference from non-target sounds, hemispheric asymmetry in extracting the attended information, and functional asymmetry of information coding for competing sources. In the auditory domain, this holds true not only for processing of informational dimensions in speech sounds but also for selective listening of concurrent sources of nonspeech information.

There has also been a long-standing interest in applying ERPs to diagnose auditory processing difficulties and to monitor progress with auditory training (see review in Martin et al., 2008). It remains an open question whether the pathological central auditory processing difficulties and training-induced improvements in clinical populations are associated with altered neural mechanisms for selective listening to different informational sources of the complex stimuli.

In summary, the results of our study demonstrate that focus on specific aspects of concurrent stimuli during selective listening differentially modulates ERP responses (P1, N1 and P2) to complex nonspeech stimuli. Taken together with previous research findings on speech stimuli (e.g., von Kriegstein et al., 2003, von Kriegstein and Giraud, 2004; Kaganovich et al., 2006), a proper interpretation of auditory brain responses in selective listening needs take into account the interaction of attention and factors such as perceptual salience, cognitive effort, acoustic/perceptual interference in the presence of other sources, and functional asymmetry of the brain for extracting the target information, regardless of whether the stimuli are speech or nonspeech sounds.

Acknowledgements

We would like to thank Liz Tusler, Tess Koerner, Asher Miller, Kayla Kelsey, Jason Sumontha and Dr. Edward Carney for their assistance in data collection. Thanks to Dr. Robert D. Melara and two anonymous reviewers for providing invaluable suggestions to improve this manuscript. Rao gratefully acknowledges financial support from the American Speech-Language-Hearing Foundation and the University of Minnesota. Zhang acknowledges lab startup funds from the University of Minnesota, Brain Imaging Research Awards (2008 and 2009) from the College of Liberal Arts, and support from CLA Associate Dean, Dr. Jo-Ilda Hansen.

References
