

# Rapid Context-based Identification of Target Sounds in an Auditory Scene

Marissa L. Gamble and Marty G. Woldorff

## Abstract

■ To make sense of our dynamic and complex auditory environment, we must be able to parse the sensory input into usable parts and pick out relevant sounds from all the potentially distracting auditory information. Although it is unclear exactly how we accomplish this difficult task, Gamble and Woldorff [Gamble, M. L., & Woldorff, M. G. The temporal cascade of neural processes underlying target detection and attentional processing during auditory search. *Cerebral Cortex* (New York, N.Y.: 1991), 2014] recently reported an ERP study of an auditory target-search task in a temporally and spatially distributed, rapidly presented, auditory scene. They reported an early, differential, bilateral activation (beginning ~60 msec) between feature-deviating target stimuli and physically equivalent feature-deviating nontargets, reflecting a rapid target detection process. This was followed shortly later (~130 msec) by the lateralized N2ac ERP activation, reflecting the focusing of auditory spatial attention

toward the target sound and paralleling attentional-shifting processes widely studied in vision. Here we directly examined the early, bilateral, target-selective effect to better understand its nature and functional role. Participants listened to midline-presented sounds that included target and nontarget stimuli that were randomly either embedded in a brief rapid stream or presented alone. The results indicate that this early bilateral effect results from a template for the target that utilizes its feature deviancy within a stream to enable rapid identification. Moreover, individual differences analysis showed that the size of this effect was larger for participants with faster RTs. The findings support the hypothesis that our auditory attentional systems can implement and utilize a context-based relational template for a target sound, making use of additional auditory information in the environment when needing to rapidly detect a relevant sound. ■

## INTRODUCTION

The world in which we live is dynamic and complex. To make sense of our environment we must select specific and relevant information from among the background noise and less relevant information. This selection process is true for both the visual and auditory facets of our environment. Although the search for a particular piece of information in the environment has an extensive literature in visual search, research on the analogous process of auditory search is rather sparse. There is extensive research investigating the maintenance of selective auditory attention on a particular stream (Cherry, 1953), well known as the cocktail party problem. This selective attention to a particular auditory stream, and even the ability to segregate the stream into its component parts (Bregman, 1990), is a separable problem from the mechanisms by which a particular target can be picked out from among multiple auditory stimulus inputs. Without prior knowledge of the location or the timing of a particular auditory target, participants must search the environment for the relevant piece of information to guide their behavior. Here, we use the term auditory search to describe the behavior where an individual is sifting through all of the

current auditory inputs from the environment to find a particular relevant auditory event or object.

Behavioral studies have established that our ability to find a relevant piece of information in an acoustically complex environment depends on a number of factors. Cusack and Carlyon (2003), in a series of experiments using an auditory paradigm with temporally distributed stimuli, found that our ability to perceive and isolate an auditory target in a complex auditory field depends on the features of those particular stimuli in an asymmetric way. In particular, these researchers found that searching for the presence of a feature (e.g., a frequency-modulated tone among pure tones or a longer-duration sound among shorter-duration ones) is easier than searching for an absence of a feature (a pure tone among frequency-modulated tones or a shorter-duration sound among longer sounds). It has also been shown that personally salient, but task-irrelevant, auditory information can sometimes capture attention (Wood & Cowan, 1995; Moray, 1959). In addition, it has been reported that when a nontarget auditory stimulus was a feature singleton it interfered with the detection of relevant targets and slowed RTs, whereas when a target auditory stimulus was a feature singleton it facilitated target detection (Dalton & Lavie, 2004). Creating circumstances where the target was no longer a singleton, however, reduced these interference effects (Dalton &

Lavie, 2007). Finally, spatial distribution of the sound sources can influence the detection (Eramudugolla, McAnally, Martin, Irvine, & Mattingley, 2008) and perception (Gregg & Samuel, 2012; Shinn-Cunningham, Lee, & Oxenham, 2007) of auditory objects.

In addition to helping to understand the particular bottom-up physical factors that can influence target detection, high temporal resolution brain recording techniques, such as ERPs, can help delineate the cascade of top-down cognitive events that must occur for target identification and processing to occur. In a recent ERP study by Gamble and Luck (2011), participants performed an auditory search of two simultaneously presented auditory stimuli presented to the left and right ears and identified whether a predefined target was present or absent. To do this, participants had to find the designated target and orient their auditory attention toward it. An electrophysiological correlate of this orienting of attention was found in the form of a contralateral ERP negativity to the target over anterior electrode sites, approximately 200–300 msec after the presentation of the stimulus pair. This lateralized neural activity, termed the N2ac, is analogous to the visual ERP component, the N2pc (Luck & Hillyard, 1994), which has been widely used to study the orienting of visual spatial attention to a target during visual search. The N2ac was interpreted by Gamble and Luck (2011) as reflecting the focusing of spatial attention toward the detected auditory target and was thus proposed to provide a cognitive process marker that could be used to further study neural mechanisms involved in auditory search.

We recently leveraged the high temporal resolution of ERPs to expand upon the Gamble and Luck (2011) study to look more deeply at the processes underlying auditory search in a more complex auditory environment (Gamble & Woldorff, 2014). More specifically, we increased the number of spatially distributed stimuli from 2 to 10 and temporally distributed them over the course of 500 msec. This approach not only better simulated the varying spatial and temporal characteristics of sounds that tend to occur in auditory environments but the temporal distribution of the sounds also enabled us to selectively extract the neural responses to the relevant target sounds and to the physically equivalent, task-irrelevant, nontarget sounds and to compare them. This, in turn, enabled us to delineate key parts of the temporal cascade of neural events underlying auditory search, including those underlying both the target detection and spatial-attention orienting processes.

More specifically, in this more recent Gamble and Woldorff (2014) study, we again found a contralateral negativity to the relevant target over anterior electrode sites (i.e., the N2ac), beginning at ~150 msec poststimulus presentation, reflecting the orienting and focusing of auditory spatial attention to the target stimulus but now in a more expanded and ecologically valid paradigm. In addition, however, preceding the contralateral attentional-

orienting N2ac effect, we found a very early differential bilateral activity (starting at ~60 msec) between the responses to the target and to the physically identical nontarget stimuli, which we termed the early bilateral negativity (EBN) effect. The very early latency of this differentiation suggested the existence of a neural mechanism by which a template or representation of the auditory target stimulus is set up in the brain prior to stimulus occurrence. When an auditory target stimulus occurs, it can then be rapidly matched to the template, thereby enabling rapid identification.

In the current experiment, we focus specifically on examining the nature of this target-related template, as reflected by its ERP neural marker, the EBN effect. We had two main questions we wanted to address about the nature of our hypothesized target template and its ERP marker. First, in the experiment in Gamble and Woldorff (2014) where the EBN effect was first observed, a paradigm was used in which the auditory target was embedded in an array of sounds that were spread across time and space. Thus, one important first question was: Do the auditory stimuli need to be spatially distinct to show this target-nontarget early differential activation? The fact that this differential activation occurred in Gamble and Woldorff (2014) when the participant had no prior knowledge of the location of the auditory target or nontarget would suggest that it should not matter whether the auditory stimuli were arising from spatially distinct locations or were all arising from the same spot. To test this hypothesis, the current experiment simplified the paradigm by eliminating the spatial separation of the stimuli, instead using stimuli presented only in a single, central stream. Although Eramudugolla et al. (2008) have shown that spatial separation of auditory stimuli can make target identification easier, based on our reasoning above we still expected to see a similar early target-nontarget differentiation here that was similar to that observed in Gamble and Woldorff (2014).

A second critical characteristic we aimed to elucidate here concerned the nature of the template itself. In Gamble and Woldorff (2014), the auditory target stimuli were embedded in a rapidly presented stream of 10 sounds, eight of which were standards of the same pitch, with one feature-deviating target sound and one sensorially equivalent nontarget sound. Thus, the target and nontarget both occurred within a backdrop of a series of rapidly repeated standards. Thus, it may not be simply the “target-ness” of the designated auditory target sound that enables its rapid detection and very early neural differentiation, but possibly the deviancy aspect of the target is also important to this process. We thus wanted to also test this important question about the nature of the detection-related template.

To address these key questions concerning the process of rapid auditory target detection, in this study the stimulus paradigm included two different trial types presented at the midline, presented in random order: in-stream

trials, where 10 tones composed of eight standards, one target, and one nontarget were rapidly presented (as in Gamble & Luck, 2014), and alone trials, where the same target and nontarget tones were presented in isolation. The inclusion of these two trial types allowed us to test these key questions described above.

The predictions were as follows. First, if the feature deviancy of the auditory target stimulus is critical for its very rapid identification and selective processing, we should see this differential target and nontarget activation for the in-stream trials but not for the alone trials. In the in-stream trials, the target and nontargets were presented embedded in a stream of the repeated standards and would therefore be perceived as feature deviants. This inherent deviancy of the tone may thus provide a stepping stone from which early target identification may occur more easily. In contrast, in the alone trials, the auditory stimuli were presented in isolation with no context-creating stimuli from which to deviate. Thus, if the deviancy nature of the target sound is critical for its rapid identification, we should not see any target-specific enhancement of the EBN in the alone trials. Alternatively, if it is just “target-ness” alone that is necessary for early target identification, then we should see evidence of early target identification for both the in-stream and alone trials. More specifically, the targets in the in-stream trials and the targets in the alone trials were physically identical and would thus both, in their nature as targets, meet the criteria for eliciting early target identification if “target-ness” were sufficient to do so.

Considering that the early target-versus-nontarget effect on the EBN appears to be invoked to facilitate the rapid identification of the designated target in a rapidly streaming auditory scene, we hypothesized that individual differences in the size of this effect across participants might be related to the target detection task performance. In particular, we hypothesized that those with larger target-specific EBN effects for the in-stream trials, presumably reflecting better instantiation of the target template mechanism, would be faster and/or better at discriminating the targets.

Aside from the EBN, reflecting early target detection, we also expected to see the commonly recorded long-latency P300 ERP component. The P300 (or P3b), a large, typically centroparietal positive wave typically peaking at around 400–500 msec, is commonly found for detected targets in oddball paradigms and is generally thought to reflect late target-related processing (see Polich, 2007, for a review). In three stimulus “odd-ball” paradigms, a relevant target typically produces a parietally distributed P3b, whereas nontarget deviants can yield a more anteriorly distributed P3a. This P3a is larger for novel stimuli (Grillon, Courchesne, Ameli, Elmasian, & Braff, 1990), is sensitive to probability (Katayama & Polich, 1996), and is dependent on target stimulus ease of discrimination as well as perceptual load (Sawaki & Katayama, 2006, 2008). Given that, in our paradigm, the targets were inherently deviants, this

deviation/oddball should produce a robust P300 for the targets. Additionally, the large pitch difference between the target and standard, and the high dissimilarity of the target and nontarget, the rapid presentation rate, which may increase perceptual load, and the nontarget is a known and consistent distractor (i.e., not novel), and we therefore expected that there would be a significantly reduced P300 for the nontargets.

## METHODS

### Participants

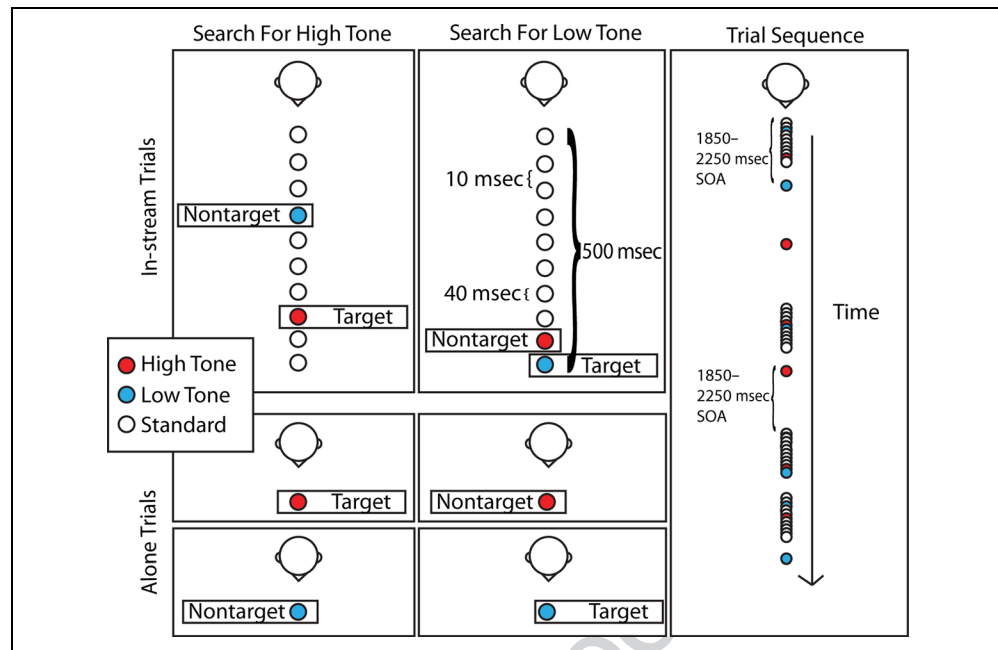
Twenty-three individuals (nine women) between 18 and 34 years old (mean = 21 years) participated for either course credit or payment compensation. All individuals gave informed consent through protocols approved by the Duke Institutional Review Board, and all reported normal hearing, normal or corrected-to-normal vision, and no history of neurological disorders. Of the initial participants, three individuals were excluded for either poor behavioral task performance or excessive artifacts in the EEG (e.g., muscle movements, blinks). Thus, the functional analyses were performed on the remaining 20 participants.

### Stimuli and Procedure

Three different tone stimuli, each 40 msec in duration with 3 msec rise and fall times, were used in the paradigm: a low tone of 500 Hz, a middle tone of 1396 Hz, and a high tone of 3000 Hz. These three tones were selected to be highly discriminable and distinct from each other. The middle tone was always presented as a pure sine wave, whereas the high and low tones could be either a pure sine wave or amplitude modulated. The amplitude modulation was accomplished by multiplying the original waveform by a 37.5-Hz envelope waveform. Each sound was presented bilaterally, via headphones, to the two ears.

To test the hypothesis that the target also needed to be a deviant to elicit early target-selective activity, two different trial types were included: in-stream trials and alone trials (see Figure 1). To create the in-stream trials, as in the previous study (Gamble & Woldorff, 2014), 10 tones were rapidly presented, separated by ISIs of 10 msec. Eight of these 10 sounds were middle tones (i.e., “standards”), one was a high tone, and one was a low tone. The first two tones in the trial were always standards, with the remaining tones (i.e., eight standards, the high tone, and the low tone) presented in a random order. In contrast, the new alone trials consisted of just a single low or high tone, presented by itself. Each experimental block was composed of 240 trials, half alone and half in-stream trials that were randomly intermixed. The intertrial stimulus onset asynchronies were jittered randomly between 1850 and 2250 msec.

**Figure 1.** Experimental paradigm. Each subject participated in two different attentional search conditions: search for high tone and search for low tone. In each condition, two trial types were presented in randomized order: in-stream trials and alone trials. For each trial type, participants were told to search for the designated target (either the high or low tone) and ignore the nontarget (either the high or low tone).



There were two main attentional task conditions—namely search for high tone and search for low tone, along with and one passive-listening condition, for a total of three task conditions, which were presented over the course of six different experimental blocks, counter-balanced within and across participants. Each condition manipulated how the participant was interacting with the same stimuli. For both of the attend conditions, the participant actively listened to the auditory stimuli, with the task to search for the designated target tone among the presented stimuli, discriminate its tonal quality, and then respond by making a button press to indicate whether the target was either pure or amplitude modulated. For the attend-high condition, the target was the high tone and the nontarget the low tone. For the attend-low condition, the target was the low tone and the nontarget the high tone. For the passive-listening condition, all the same auditory stimuli were presented, but participants were instructed to ignore them and to instead read a book of their own choosing. We will not be discussing the results of the passive-listening condition here.

### Recording and Analysis

The EEG was recorded using a customized, extended-coverage 64-electrode elastic cap (Duke64 layout, Electrocap International, Eaton, OH) with a Synamps Neuroscan system (Charlotte, NC). EEG was sampled at 500 Hz, with an online bandpass filter of 0.01–100 Hz and a gain of 1000. EEG was recorded online referenced to the right mastoid and was later re-referenced offline to the algebraic average of the left and right mastoids. Vertical eye movements (VEOG) were monitored by placing electrodes beneath the left and right eyes, referenced to electrodes

placed above the left and right eyes, respectively. Horizontal eye movements (HEOG) were monitored by placing electrodes on the outer left and right canthi, referenced to each other. Independent component analysis was used to identify and remove blink-generated activity from the data (Jung et al., 2000). Trials with excessive muscle activity were also rejected from inclusion in the averages. Finally, the selectively averaged ERP waveforms (see below) were filtered offline with a 9-point running average filter to attenuate any 60-Hz line noise contamination.

The isolation of key electrophysiological markers reflecting target identification and discrimination required several steps of analysis of the EEG data. The attend-high-target and attend-low-target conditions were collapsed into an attend condition, thus creating target and nontarget ERP waveforms from the same physical stimuli. Collapsing the data this way yielded five different condition/trial types: target\_alone, nontarget\_alone, target\_in\_stream, nontarget\_in\_stream, standards\_in\_stream.

Although the alone trial types needed no further manipulation, the in-stream trials did. In particular, because of the rapidity with which the stimuli were presented in these streams, the resultant time-locked averaged ERPs to the stimuli within these streams overlapped substantially, thereby introducing substantial contamination from the neural responses onto the neighboring auditory stimuli's neural responses. To address this contamination of the overlapping stimulus responses, we took a two-step approach that not only removed overlap but also isolated the deviance-related activity of the target and nontarget stimulus responses. First, we employed a modified version of the ADJAR filter technique (Woldorff, 1993) to estimate the contribution of the overlap and to then remove it from the ERP waveforms. More specifically, the time-locked



averages to the targets contained overlap contribution from the standards and the nontargets, whereas the time-locked averages to the standards contained overlap from both the targets and the nontargets. By convolving an initial best estimate of the target activity (i.e., the large target-minus-standard difference wave) with the adjacent-event distribution of the target relative to the standards, we were able to create an estimate of the target activity overlap on the standard tone ERP average. We then corrected the standard tone ERPs by subtracting the target overlap estimate from them, leaving “target-corrected standards” that then only contained overlap from the nontarget responses. Because the original target ERPs also would have had nontarget overlap, we then subtracted target-corrected-standards waveforms from the original target waveforms, thereby both subtracting out this nontarget overlap and selectively extracting the deviance-related activity for the targets. This overlap-correcting sequence was also performed to obtain the overlap-correct deviance-related nontarget activity (by performing the nontarget minus nontarget-corrected-standards subtraction). These corrected difference waves were then used for analysis for the in-stream trials below.

For our main analyses of the EBN effect, the mean amplitude between 60 and 120 msec for electrodes in a frontocentral ROI for the responses to the targets and nontargets was subjected to paired  $t$  tests, one for the in-stream trials and one for the alone trials. This specific latency range and the electrode ROI were selected based on the time range from the Gamble and Woldorff (2014) study when the EBN was significantly different for the targets and nontargets and the electrode locations for which these effects were largest. We also applied the analyses to a 40-msec window (80–120 msec) around the peak of the EBN (100 msec), determined from the waveform derived from collapsing across the targets and nontargets across trial types. It should be noted, however, that separate analyses were used for the in-stream and alone trial types because of the substantial differences in the overall activity levels and extraction process for the two types. Below we have laid out the results from each analysis to address our specific hypothesis for the EBN. An additional set of statistical analyses were run on the mean amplitude between 300 and 600 msec over a parietal-electrode ROI, separately for the in-stream and alone trials, to measure effects on the P300.

## RESULTS

### Behavioral

Despite being a challenging task, participants were able to accurately discriminate the target stimulus as being pure versus AM on 89.8% of the trials, with an average RT of 595 msec for correct trials. A  $2 \times 2$  ANOVA between the target pitch (high vs. low tones) and trial type

(in-stream vs. alone trials) on response accuracy yielded a significant main effect for trial type,  $F(1, 19) = 5.75, p = .03$ , where participants were somewhat more accurate in the alone trials (mean = 94.3%) than on the in-stream trials (mean = 89.7%). There was no significant main effect for target pitch, nor any significant interactions with the pitch factor.

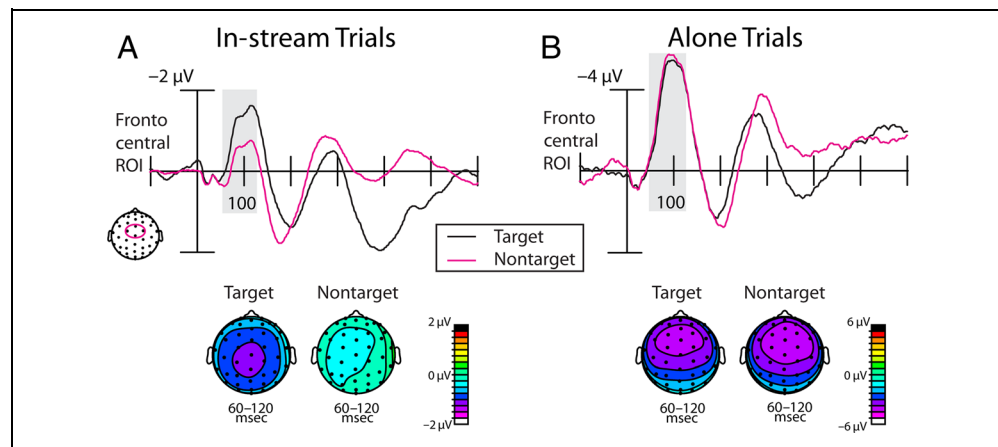
A  $2 \times 2$  ANOVA with the factors target pitch (high and low tones) and trial type (in-stream vs. alone) was also performed on the RTs. RTs were calculated relative to the onset of the target stimulus for both in-stream and alone trials. There was a marginally significant main effect of trial type,  $F(1, 19) = 4.32, p = .051$ , where participants were faster for the in-stream trials (mean = 598 msec) than the alone trials (mean = 622 msec). All other main effects and interactions on the RTs were not significant. Because there were no significant differences between the designated targets when they were high versus low tones, all the following analyses were collapsed across this factor.

### *The EBN Effect: The Role of Target Deviancy*

To address the question of whether the early target–nontarget differential activation is a result of just the “target-ness” of the target tone or depends on it also being a deviant stimulus within a stream, we examined whether the in-stream and alone trials both yielded this early target–nontarget ERP difference. As shown in Figure 2, the target and nontarget responses within the in-stream trials rapidly differentiated, with the target minus standard waveform becoming more negative starting around 60 msec until approximately 120 msec and the nontarget minus standard waveform being substantially less negative in the same time range, replicating the EBN effect observed in Gamble and Woldorff (2014). The differences in this early negativity were analyzed statistically by performing a  $t$  test on the mean amplitude between 60 and 120 msec on the frontocentral ROI consisting of four electrodes near Cz, for the target minus standard and nontarget minus standard difference waves. This analysis, which addresses the target deviancy hypothesis, showed that the target minus standard difference wave was significantly larger than the nontarget minus standard difference wave,  $t(19) = 3.43, p = .003$  (see Figure 2). An additional analysis at the peak of the EBN, from 80 to 120 msec, was also highly significant,  $t(19) = -3.257, p = .004$ .

In contrast, in the alone trials,<sup>1</sup> both the target and nontarget waveforms showed similarly large early negativities peaking at around 100 msec. A  $t$  test on the mean amplitudes between 60 and 120 msec at a frontocentral ROI for the targets and nontargets in the alone trials yielded no significant difference,  $t(19) = 1.28, p = ns$ . The analysis at the peak of the EBN (80–120 msec) was also not significant,  $t(19) = 0.849, p = ns$ .

**Figure 2.** In-stream and alone trials attend conditions: frontocentral ROI. (A) In-stream trials showed a larger and more negative EBN waveform for target stimuli compared to nontarget stimuli. The voltage maps show a clear central distribution for the target compared to the nontarget in the latency represented by the vertical gray bar. (B) Alone trials showed no discernable difference in the ERP waveforms or the voltage maps for the target or the nontarget stimuli.



### Individual Differences in the EBN Effect

Additional analyses examined the role of individual differences in the EBN effect on task performance. In particular, we divided into separate bins those participants who were the fastest versus the slowest (median split) at performing the target identification on the in-stream trials (see Figure 3B), splitting the participant pool into two groups of 10. The fast responders had an average RT of 524 msec for the in-stream trials, whereas the slow responders had an average RT of 665 msec. Although average RTs differed between the two groups by 141 msec, the accuracy did not statistically differ (fast responders, mean = 91.9%; slow responders, mean = 85.8%),  $t(18) = 1.78, p = ns$ .

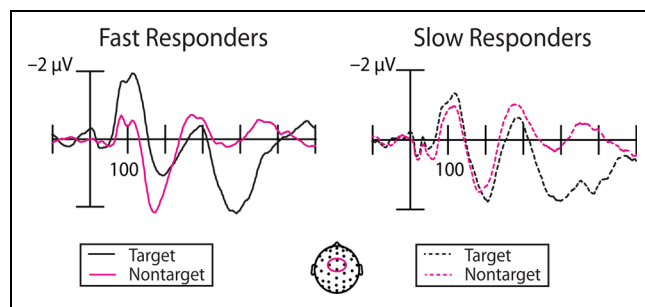
As Figure 3B shows, however, the fast responders showed a large difference between the responses for the target and the nontarget EBNs, whereas the slow responders showed a much smaller difference between the target and nontarget waveforms, consistent with our hypothesis concerning this relationship. To ascertain whether these two groups differed statistically, we ran a one-tailed, independent-sample  $t$  test on the target minus nontarget difference wave between 60 and 120 msec,  $t(18) = -1.548, p = .069$ . Although this time range was marginally significant, the same analysis at the peak of the EBN (80–120 msec) was significant, showing that the EBN effect for the fast responders (mean =  $-1.24$ ) was larger than that of the slow responders (mean =  $-0.35$ ;  $t(18) = -1.97, p = .032$ ), indicating that participants with greater differences in this early neural responses to the designated target and nontarget stimuli performed better on the task.

### Longer Latency Effects: P300 Responses

**P300s on in-stream trials.** Following the EBN in the in-stream trials, there was a rather large positive peak in the P2 range for both the target minus standard and nontarget minus standard difference waveforms, which was then followed by a large positive P300 response for the

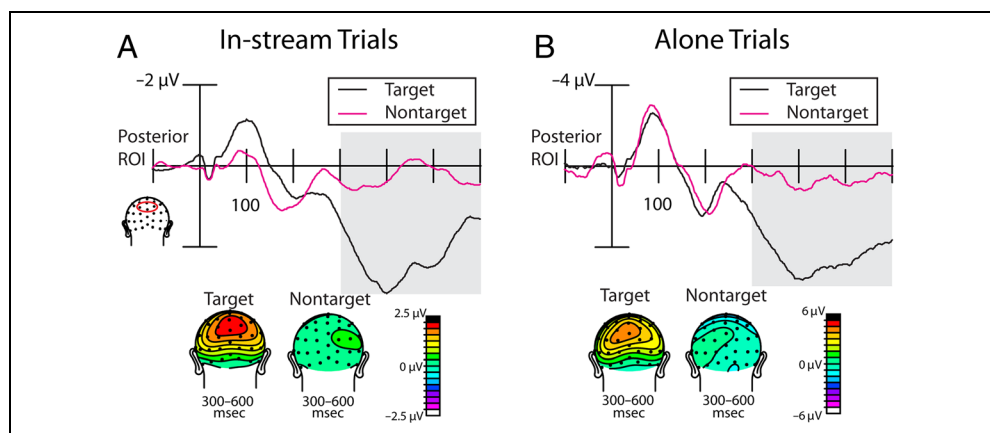
target minus standard waveform but essentially no such response for the nontarget minus standard waveform. To analyze the effects in the long-latency P300 range, we ran a  $t$  test on the mean amplitudes between 300 and 600 on a posterior-electrode ROI for the in-stream target minus standard and nontarget minus standard difference waves (Figure 4A). These analyses revealed a highly significant effect,  $t(19) = 8.16, p < .001$ , showing that the target P300 wave was substantially larger (more positive) than that of the nontarget. Additional specific comparisons indicated that a P300 was present for both the targets,  $t(19) = 10.40, p < .001$ , and the nontargets,  $t(19) = 2.32, p < .05$ , but was just substantially smaller for the nontargets.

**P300 on alone trials.** Figure 4B shows the P300 effect for target and nontarget stimuli on the alone trials, which shows a large positive waveform starting around 300 msec for the target stimuli, whereas the nontarget ERP again remained close to baseline. A  $t$  test on the mean amplitudes of the target and nontarget responses in the alone trials, between 300 and 600 msec at a posterior electrode



**Figure 3.** Fast and slow responders for in-stream trials. Separation of participants into fast responders and slow responders based on their RTs on the in-stream trials (median split) showed a different ERP response profile: fast responders showed a much larger difference in neural response to the targets relative to the nontargets than the slow responders.

**Figure 4.** In-stream and alone trials: attend conditions, posterior ROI. (A) In-stream trials showed a larger positive polarity P300 wave for the target stimuli only. Voltage maps for the target, nontarget, and target-minus-nontarget difference waves show a large positivity centered around parietal electrodes in the latency represented by the vertical gray bar. (B) Waveforms for the alone trials showed a similar pattern, with a large positive peak for the target but not the nontarget. Voltage maps showed a clear distribution of the positive polarity waveforms over parietal scalp characteristic of P300 responses.



ROI, showed a significant difference between the target waveforms and the nontarget waveforms,  $t(19) = 6.806$ ,  $p < .001$ . A one-sample  $t$  test for the target,  $t(19) = 6.10$ ,  $p < .001$ , and the nontarget,  $t(19) = 1.47$ ,  $p = ns$ , indicated that there was a P300 present for only the task-relevant targets for the alone trials.

## DISCUSSION

In this study, we aimed to investigate more deeply the mechanisms associated with auditory target identification and discrimination in a temporally dynamic auditory environment. In our previous study (Gamble & Woldorff, 2014), the search for a particular target in a temporally and spatially complex auditory environment yielded the contralateral N2ac effect (Gamble & Luck, 2011), reflecting the focusing of spatial auditory attention toward a detected target sound. In that study, prior to the N2ac, which onset at 130 msec, we had also observed an EBN, onsetting at ~60 msec and focused over central electrodes, that was substantially larger for the targets compared to the nontargets. This very early differential activity, as marked by the EBN effect, occurred so rapidly after the stimulus onset that we surmised the presence of a target template in the brain to which each incoming auditory stimulus was compared, thereby enabling a very rapid target detection mechanism. This template can apparently be invoked in a complex auditory environment, allowing for this rapid target detection, which is then followed by a rapid focusing of auditory spatial attention toward that target.

Although our previous study enabled the delineation of the cascade of events involved in auditory target processing during auditory search, it was unclear from this first experiment what the nature of the target template was that enabled such rapid target identification and differentiation from the nontargets. To further investigate

these questions, in the current experiment participants listened to randomized sequences in a midline-presented series of auditory stimulus trials that included 10 stimulus in-stream trials, with one target, one nontarget, and eight standard stimuli, and single-stimulus alone trials consisting of just one target or nontarget in isolation. Participants were required to search for, identify, and discriminate the target tonal quality. These particular paradigm parameters were used to address two questions: (1) Would the EBN effect be present when there was only one spatial input channel? (2) Is “target-ness” sufficient to elicit this target/nontarget EBN effect or is it necessary that the target also be a feature deviant. In turn, more overarchingly, what do the answers to these questions indicate about the nature of the target-related template that this early effect implicates as being employed during auditory search in complex auditory environments?

### The EBN Effect When There Is Only One Spatial Stimulus Stream

The Gamble and Woldorff (2014) experiments showed the target/nontarget EBN effect when there was a spatial separation of the auditory stimuli. This means not only were the target and nontarget stimuli always in different auditory spatial channels but also the participant had no prior knowledge of where these stimuli were going to be presented. In the current experiment, with all stimuli presented in the midline, we eliminated the uncertainty of location, but we still kept in place the variability and uncertainty of the timing of the target stimulus. Given that rapid target identification would also be valuable in this paradigm as well, we expected we would still see the target/nontarget differential EBN effect, which is exactly what we observed. The uncertainty and variability in location of the stimuli do not appear to be necessary for the occurrence of this early effect; rather, the target/

nontarget differential EBN effect, reflecting the early target-specific identification, was still present and robust when location, but not timing, of the target stimulus was known.

### **The Target Template: The Role of Target Feature Deviancy**

One of the major goals of this study was to determine whether this EBN effect was a result of a simple template that was being set for the target that facilitated its rapid detection or whether it required the combination of the relevant auditory stimulus being both a designated target and a deviant in relation to an auditory context background, in particular by occurring within a series of repeating standard sounds. To test this, we examined whether the early differential EBN effect would be observed for targets occurring both as deviants within a rapid stream (in-stream trials) and when occurring by themselves (as alone trials), with these trial types presented in random order. The EBN effect also occurring in the alone trials would suggest that the only requirement to see this effect is for the auditory stimulus to be a designated target and would therefore support a simple target-specific template mechanism.

The results, however, indicated that there was no any discernable EBN difference between the targets and the nontargets in the alone trials, but there was again a robust difference for the in-stream trials. The lack of an effect in the alone trials, but a clear target–nontarget differential activation in the in-stream trials, indicates that a necessary requirement for this early differential EBN activation is that the relevant auditory stimulus has to be not only a target but also a feature deviant within an ongoing auditory context. Thus, the underlying mechanism that results in this very early latency processing differentiation (starting at 60 msec poststimulus) must be more complex than a simple template for the designated target tone. This early target identification system apparently also relies on the deviant nature of the relevant stimulus within an auditory context, perhaps as a springboard for its identification as a target. In other words, the deviant nature of the auditory target stimulus may be making the sound more salient, thereby increasing participants' ability to detect it (Kayser, Petkov, Lippert, & Logothetis, 2005).

### **The Relational Template**

It makes sense that, to achieve the goal of detecting and processing a specific piece of information in a complex and dynamic auditory environment, it would be beneficial to use as much of the information available as possible to isolate the relevant information. The fact that we see this rapid detection and early processing differentiation in the in-stream trials and not the alone trials suggests that the process may be a more general "relational template" that relies on the relationship of the

target and its features relative to an ongoing auditory context, in particular the ongoing standards, but presumably also including the irrelevant nontarget stimuli. The target-defined feature in the current experiment occurs as a deviation within an ongoing auditory context, thereby helping to distinguish the target from its surroundings and to enable its identification earlier and more easily.

It could be argued that the early effect we are seeing is the well-known Nd or processing negativity. These effects, however, are generally found in very different paradigms than that which was employed here. In particular, Nd paradigms typically consist of two or more streams of frequently repeated tones, and the Nd ERP effect is the result of a comparison between the stimulus when it is part of an attended versus an unattended stimulus stream. In contrast, in the present experiment, the target occurred as a single occurrence, with no repetition, within a brief stream of other stimuli. Additionally, pitch-based versions of these Nd attention effects tend to start later in time and last substantially longer (e.g., Degerman, Rinne, Särkkä, Salmi, & Alho, 2008; Hansen & Hillyard, 1980, 1983), whereas our EBN effects appear as a very early and temporally focal ERP effect beginning at 60 msec poststimulus. Moreover, and perhaps most importantly, the fact that we see this early differentiation only in the in-stream trials, and not in the alone trials intermixed into the same run, strongly argues against the EBN effect being a form of Nd.

In addition, the fact that we observed an association between RTs and the size of the EBN effect supports the idea that this effect reflects the detection and identification of the relevant target as facilitated by the template mechanism. If the relational template is engaged to facilitate rapid target detection and discrimination, then successful engagement should yield faster RTs and clearer and larger target versus nontarget differentiation. This is exactly what we see with the individuals who responded the fastest and who showed a larger difference between the target and nontarget EBN responses. Less effective use of the relational template, reflected by a weaker and less clear distinction between the target and the nontarget, resulted in slower/poorer performance on the task. This variation between individuals thus provides further evidence that effective implementation of the relational template mechanism results in facilitation of the detection and discrimination of designated target stimuli.

What is still unclear from the current experiments, however, is whether this early differential response, and therefore the resulting neural response of the template mechanisms, is due to target enhancement, nontarget suppression, or possibly a combination of both. It would seem that the most efficient strategy for early target detection would be to enhance the processing of the target. However, the only way for this to occur is to have a stable and clear representation of the target for establishing



a template. If there is a stable and robust representation of the target, especially in conjunction with its deviancy from an ongoing auditory context, participants are primed for the particular stimulus parameters and can therefore more efficiently process the target stimulus. An alternative (or additional) strategy, however, could be to also suppress the processing of stimuli that one knows is not the target (i.e., particularly the nontarget deviant). By suppressing the information that is clearly not the target, one could presumably more easily detect the target when it occurs. Either strategy, or a combination thereof, could ultimately facilitate the ability to detect and discriminate the target. Regardless, the present results showing clear early differentiation between a target and nontarget stimulus, but only within an ongoing auditory context and not when occurring alone, implicate the ability of the brain to maintain and make use of a relational template to facilitate the detection of relevant target sounds.

### Long-latency Processing: P300 Responses

Despite there being an early EBN effect for the in-stream trials but not for the alone trials, the late-level processing reflected by the P300 was present for both trial types, but mainly just to the targets; the P300 was absent or very small for the nontargets. The fact that there was no clear P3a for the nontargets indicates a couple of things. First, the paradigm successfully allowed participants to focus their attention to the relevant stimuli without the irrelevant deviants capturing much attention, which would have been reflected in a P3a to the nontargets (Sawaki & Katayama, 2006). The faster rates of presentation in the current paradigm, compared to the one stimulus per 1–2 sec for standard odd-ball paradigms, likely served to increase perceptual load and allows for the more consistent suppression of irrelevant information (Woldorff, Hackley, & Hillyard, 1991). Additionally, the fact that we see a clear and robust P300 for targets regardless of whether they were presented in the in-stream trials or the alone trials is consistent with the idea that P300 reflects higher-level operations involved in target processing (see Polich, 2007, for a review) and not necessarily dependent on previous processing steps. The fact that we see an EBN effect for the in-stream trials and not the alone trials but see P300s for both are indicative that the processing reflected by the P300 is not contingent on the relational template processing mechanism.

### Conclusions

The auditory environment can be a rich landscape, full of both informative and distracting sounds. The present results indicate that the process by which we successfully select relevant information and simultaneously reject irrelevant information can occur particularly rapidly by the instantiation in the brain of a context-based relational template. This template results in an early bilateral differ-

entiation between a relevant target and an irrelevant nontarget that utilizes the deviancy aspect of the target stimulus within a local auditory context to aid in rapid target sound detection and identification.

Reprint requests should be sent to Marissa L. Gamble, Center for Cognitive Neuroscience, P.O. Box 90999, Durham, NC 27708, or via e-mail: marissa.gamble@duke.edu.

### Note

1. These trials were presented in isolation, with no surrounding auditory stimuli, such as the standards. Thus, the statistical analyses and figures could be done on the raw waveforms, with no need of any ADJAR correction or subtractions.

### REFERENCES

- Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*, 975–979.
- Cusack, R., & Carlyon, R. (2003). Perceptual asymmetries in audition. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 713–725.
- Dalton, P., & Lavie, N. (2004). Auditory attentional capture: Effects of singleton distractor sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 180–193.
- Dalton, P., & Lavie, N. (2007). Overriding auditory attentional capture. *Perception & Psychophysics*, *69*, 162–171.
- Degerman, A., Rinne, T., Särkkä, A.-K., Salmi, J., & Alho, K. (2008). Selective attention to sound location or pitch studied with event-related brain potentials and magnetic fields. *The European Journal of Neuroscience*, *27*, 3329–3341.
- Eramudugolla, R., McAnally, K., Martin, R., Irvine, D., & Mattingley, J. (2008). The role of spatial location in auditory search. *Hearing Research*, *238*.
- Gamble, M. L., & Luck, S. J. (2011). N2ac: An ERP component associated with the focusing of attention within an auditory scene. *Psychophysiology*, *48*, 1057–1068.
- Gamble, M. L., & Woldorff, M. G. (2014). The temporal cascade of neural processes underlying target detection and attentional processing during auditory search. *Cerebral Cortex (New York, N.Y.: 1991)*.
- Gregg, M. K., & Samuel, A. G. (2012). Feature assignment in perception of auditory figure. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 998–1013.
- Grillon, C., Courchesne, E., Ameli, R., Elmasian, R., & Braff, D. (1990). Effects of rare non-target stimuli on brain electrophysiological activity and performance. *International Journal of Psychophysiology*, *9*, 257–267.
- Hansen, J., & Hillyard, S. (1983). Selective attention to multidimensional auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 1–19.
- Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, *49*, 277–290.
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T., Mckeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*, 163–178.

- Katayama, J., & Polich, J. (1996). P300, probability, and the three-tone paradigm. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *100*, 555–562.
- Kayser, C., Petkov, C., Lippert, M., & Logothetis, N. (2005). Mechanisms for allocating auditory attention: An auditory salience map. *Current Biology*, *15*, 1943–1947.
- Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1000–1014.
- Moray, N. (1959). Attention in dichotic-listening—Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56–60.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148.
- Sawaki, R., & Katayama, J. (2006). Stimulus context determines whether non-target stimuli are processed as task-relevant or distractor information. *Clinical Neurophysiology*, *117*, 2532–2539.
- Sawaki, R., & Katayama, J. (2008). Difficulty of discrimination modulates attentional capture by regulating attentional focus. *Journal of Cognitive Neuroscience*, *21*, 359–371.
- Shinn-Cunningham, B. G., Lee, A. K. C., & Oxenham, A. (2007). A sound element gets lost in perceptual competition. *Proceedings of the National Academy of Sciences*, *104*, 12223–12227.
- Woldorff, M. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology*, *30*, 98–119.
- Woldorff, M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, *28*, 30–42.
- Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 255–260.

Uncorrected Proof