Behavioral/Cognitive

Microsaccadic Responses Indicate Fast Categorization of Sounds: A Novel Approach to Study Auditory Cognition

Andreas Widmann,[1] **Ralf Engbert,**[2] **and Erich Schröger**[1]

[1]Cognitive and Biological Psychology, University of Leipzig, D-04109 Leipzig, Germany and [2]Experimental and Biological Psychology, University of Potsdam, D-14476 Potsdam, Germany

The mental chronometry of the human brain's processing of sounds to be categorized as targets has intensively been studied in cognitive neuroscience. According to current theories, a series of successive stages consisting of the registration, identification, and categorization of the sound has to be completed before participants are able to report the sound as a target by button press after ∼300–500 ms. Here we use miniature eye movements as a tool to study the categorization of a sound as a target or nontarget, indicating that an initial categorization is present already after 80–100 ms. During visual fixation, the rate of microsaccades, the fastest components of miniature eye movements, is transiently modulated after auditory stimulation. In two experiments, we measured microsaccade rates in human participants in an auditory three-tone oddball paradigm (including rare nontarget sounds) and observed a difference in the microsaccade rates between targets and nontargets as early as 142 ms after sound onset. This finding was replicated in a third experiment with directed saccades measured in a paradigm in which tones had to be matched to score-like visual symbols. Considering the delays introduced by (motor) signal transmission and data analysis constraints, the brain must have differentiated target from nontarget sounds as fast as 80–100 ms after sound onset in both paradigms. We suggest that predictive information processing for expected input makes higher cognitive attributes, such as a sound's identity and category, available already during early sensory processing. The measurement of eye movements is thus a promising approach to investigate hearing.

Key words: audition; categorization; mental chronometry; microsaccade

Introduction

The mental chronometry of sound processing has been extensively studied by means of event-related brain potentials (ERPs), fluctuations in the EEG time locked to the occurrence of stimuli (Kutas et al., 1977; Ritter et al., 1979; Näätänen, 1992). One important finding is that sounds that have to be categorized as a target elicit a characteristic sequence of ERP components that are supposed to reflect the registration, selection, identification, and categorization of the sound (Ritter et al., 1979; Dien et al., 2004; Key et al., 2005; Joos et al., 2014), the latter reflected in the N2 component of the ERP occurring beyond 200 ms after stimulus onset (Patel and Azzam, 2005; Folstein and Van Petten, 2008). Here, we present three eye-movement studies yielding modulations of microsaccadic behavior that challenge the prevailing serial-stage model of sound categorization and offer a new approach to studying the brain's processing of sounds.

Microsaccades are rapid, small-amplitude movements of the eyes during fixation occurring with a frequency of 1–2 per second that are modulated by top-down cognitive processes (Engbert, 2006). After the presentation of visual and auditory stimuli microsaccades are inhibited for a short period of time and followed by a typical rebound with enhanced microsaccade rates (Rolfs et al., 2008). When rare, target visual stimuli have to be detected in a series of frequent standard stimuli in a so-called oddball paradigm, the microsaccadic inhibition is enhanced and the rebound delayed for targets (Valsecchi et al., 2007, 2009). Importantly, this pattern of results has even been reported for auditory oddball paradigms (Valsecchi and Turatto, 2009; Yuval-Greenberg and Deouell, 2011). However, the time course of the effects has not yet been examined in detail. Valsecchi and Turatto (2009) integrated within consecutive 100 ms time bins and observed a significant difference in the 200–300 ms bin, and Yuval-Greenberg and Deouell (2011) did not analyze the time course. Yet, visual inspection of the data obtained in these auditory oddball studies suggests that microsaccade behavior differed between target and standard sounds as early as 150 ms.

To examine this issue we replicated the experiment by Yuval-Greenberg and Deouell (2011) with minor modifications and a most conservative data analysis procedure (Experiment 1A). To exclude possible confounds of spatial attention including the problem of lower spatial eccentricity of targets compared with nontargets in this paradigm, we replicated the study using non-spatial pitch distractor and pitch target sounds (Experiment 1B).

We were interested to extend the current findings to a more natural saccade task, a simple form of score reading where tones and visual symbols had to be matched (Experiment 2). This experiment avoided the potential impact of unequal stimulus probabilities of the oddball paradigm on the results. We expected to find effects of sound category on microsaccadic behavior at ~150 ms. Considering the neural transmission delays, motor delay, and delay introduced by data analysis, this would imply that the brain's initial categorization of a sound as a target or a nontarget is present and affects behavior before 100 ms after sound onset. This would challenge the widely accepted serial-stage model of target detection in the auditory oddball paradigm and call for a different explanation of target detection.

Materials and Methods

Participants. Twelve participants took part in Experiment 1A. They were 20–34 years old (mean 24.4 years). Eight of the participants were female, four male. Eleven of the participants were right-handed, one left-handed. Fifteen participants took part in Experiment 1B. One participant had to be excluded from analysis as more than half of the trials had to be rejected due to blinks. The remaining subjects were 19–42 years old (mean 29.1 years). Seven of the participants were female, seven male. Thirteen of the participants were right-handed, one left-handed. Fourteen participants took part in Experiment 2. They were 20–41 years old (mean 25.9 years). Twelve of the participants were female, two male. Thirteen of the participants were right-handed, one left-handed. Participants were paid or received course credit for their participation. All participants reported normal hearing and normal or corrected-to-normal vision, no history of neurological diseases, and no CNS-active medications. All experiments were performed in accordance to the ethical guidelines of the German Psychological Society and the Declaration of Helsinki. All subjects gave their informed consent before their inclusion in the studies.

Apparatus, stimuli, and procedure in Experiments 1A and 1B. Participants were seated in a dimly lit sound booth (402-A; Industrial Acoustics Company). Participants' heads were stabilized with a chin and forehead rest. In Experiment 1A sine wave sounds with a frequency of 1 kHz and a duration of 50 ms (5 ms rise and fall times; 44.1 kHz sampling frequency) were presented with headphones (HD 25–1 II; Sennheiser; Sound Blaster X-Fi Titanium HD; Creative) with an interstimulus interval of 1.3 s. A 1 kHz pure tone was convolved with the head-related transfer functions of the left and right ear of a KEMAR dummy head (30 and 60° eccentricity, 0° elevation; Gardner and Martin, 1995) to generate the impression of spatial sounds. Frequent standard stimuli (80.4%) were presented at an eccentricity of 60° to the left (77.5 dB SPL at left ear, 72 dB SPL at right ear). Rare nontarget distractor sounds (9.8%) were also presented at an eccentricity of 60° to the left at a higher intensity (83.5 dB SPL at the left ear, 77.9 dB SPL at the right ear). Rare target sounds (9.8%) were presented at an eccentricity of 30° to the left at the same intensity as the standards (76.1 dB SPL at the left ear, 70.4 dB SPL at the right ear). In Experiment 1B frequent standard stimuli (80.4%) were sine wave sounds with a frequency of 500 Hz. Rare nontarget distractor sounds (9.8%) were sine wave sounds with a frequency of 454 Hz. Rare target sounds (9.8%) were sine wave sounds with a frequency of 550 Hz. Sounds were delivered binaurally (74.6 dB SPL at both ears). A 3 × 3 pixel black fixation dot was presented on a 19" CRT monitor (resolution 1024 × 768, 100 Hz, distance 53 cm; G90fB; ViewSonic) on gray background with a luminance of 61 cd/m². The experimental protocol was implemented using the Psychophysics toolbox (Kleiner et al., 2007) for MATLAB (The MathWorks) and Ubuntu Linux 12.04 (64 bit). Participants were asked to fixate the fixation dot and press a button of a gamepad (Microsoft Sidewinder) with their right thumb in response to target sounds. Participants were asked to respond as fast and accurately as possible.

Both experiments started with a training block of 32 trials including four distractor and four target sounds. The training block was repeated until three of four targets were correctly detected and not more than one false alarm was observed in response to standard and distractor sounds, respectively. Six experimental blocks of 184 trials including 18 distractors
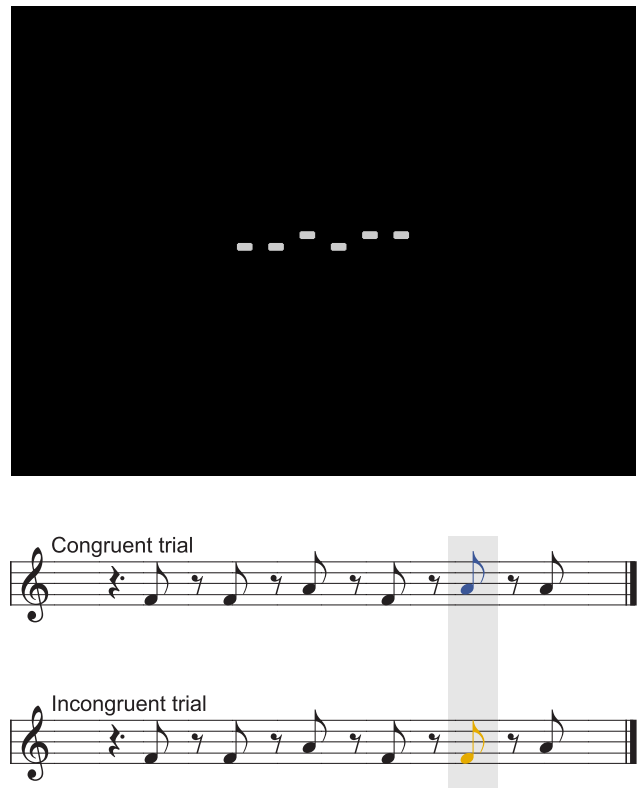

Figure 1. Prototypical visual display and corresponding auditory stimulation in Experiment 2. Each trial started with the display of a visual pattern. After 1 s a corresponding sound pattern was presented either congruent to the visual pattern in all elements (congruent trial) or deviating in a single element (incongruent trial). Sound duration and ISI were 300 ms. At the end of each trial, participants were asked to discriminate congruent from incongruent trials by button press.

and 18 target sounds in pseudorandomized order followed the training block. At least two standard sounds were presented between any two nonstandard sounds. Each block started with a nine-point calibration and validation procedure of the eye tracker. Reaction times, hit, and false alarm rate were reported to the participant after each block.

Stimuli and procedure in Experiment 2. The stimulation and procedure was a precise replication of the symbol-to-sound matching paradigm by Widmann et al. (2004). Participants were presented with visual patterns of four to six light gray rectangles (0.32° × 0.16° visual angle; 0.32° horizontal distance) on black background (Fig. 1) on the center of a 19" CRT screen. The upper corners of the rectangles were placed either above (0.36°) or below (0.04°) the horizontal meridian. The visual display remained on screen until the end of the trial. One second after the onset of the visual pattern, a sound pattern of either high (422 Hz) or low (352 Hz) sounds of 300 ms duration (5 ms rise and 5 ms fall times with raised cosine window; 65 dB SPL) and 300 ms interstimulus interval (ISI) started. The visual patterns predicted the corresponding sound patterns: for rectangles above the horizontal meridian high sounds were presented, while for rectangles below the horizontal meridian low sounds were presented. In half of the trials a single sound in the pattern was incongruent to the corresponding visual symbol. The task of the participants was to "read" the visual pattern while listening to the sounds and to detect the incongruent sounds. They were asked to press one button after the onset of the last sound and subsequently press one of two other buttons indicating whether there was an incongruent sound in the pattern.

In total 636 pseudorandomized patterns (84 four-element, 180 five-element, and 372 six-element patterns) were presented in 12 blocks of 53 trials. In 318 patterns one sound was incongruent to the corresponding visual symbol. The occurrence of an incongruent sound was pseudorandomized across trials and balanced for pattern length, relative position in the pattern, and pitch. The probability of an occurrence of an incongruent sound at the first or last position of a pattern was half the probability

of an occurrence at an inner position (as the first and last sound had to be discarded from analysis). Identical visual or auditory patterns never followed each other. One element (visual and auditory) had to be different from the other elements within the patterns (i.e., no patterns consisting of all equal visual or auditory elements were presented). The number of high and low pitch sounds was balanced across trials. The only differences in procedure to the original study were that the training blocks were not performed on a separate day but directly before the experiment and that each block started with a nine-point calibration and validation procedure of the eye tracker. For a more detailed description of stimulation and procedure, please, refer to the original study (Widmann et al., 2004).

Eye-movement recording. Binocular eye movements were recorded with an infrared EyeLink 1000 (SR Research) remote eye tracker with a sampling rate of 500 Hz. Stimulation and eye-tracker PCs were directly connected with a dedicated crossover Ethernet link. Sound onsets were marked by event messages sent over this connection. Proper sound onset timing was validated feeding the sound output and TTL signals (generated after sending the event message) into an oscilloscope. For TTL signals generated by the stimulation PC the TTL-to-sound delay was 0.6 ms with a maximum jitter of ±0.1 ms. For TTL signals generated by the digital output port of the eye tracker (initiated by commands sent over the Ethernet connection) the TTL-to-sound delay was 0.3 ms with a maximum jitter of ±0.2 ms. Proper time stamping of the event messages was confirmed by synchronizing to the eye-tracker PC clock. The mean delay between predicted and registered event message time was 0.1 ms (with SD = 0.3 ms due to rounding of the registered time to full milliseconds).

Eye-movement analysis. Using velocity thresholds (Engbert and Kliegl, 2003; Engbert and Mergenthaler, 2006), microsaccades can be detected in miniature eye movements and distinguished from the two other components denoted as drift, a slow movement component, and tremor, a high-frequency oscillatory component producing very small amplitudes (Ciuffreda and Tannen, 1995). Saccades and microsaccades are generated by the same neurophysiological systems, including the reticular formation and the superior colliculus (SC; Hafed et al., 2009; Hafed, 2011).

In Experiments 1A and 1B the eye-movement recordings were segmented into epochs from −300 to 900 ms relative to sound onset. The first two trials per block and the first trial after each nonstandard sound were excluded from analysis. Trials with misses and false alarms were not excluded from analysis (Vanrullen, 2011). Eighteen standard sounds were randomly selected from each block to have an equal SNR for all sound types. Epochs including saccades larger than 100 arc min were excluded from further analyses. In Experiment 2 epoch length was −300 to 700 ms relative to sound onset. The first and last sounds of each pattern were excluded from analysis (Widmann et al., 2004). Pairs of congruent and incongruent sounds matched for pitch and serial position in the pattern were randomly selected to achieve equal SNRs. As the task did not explicitly ask for fixation, in Experiment 2 no epochs were rejected due to saccades with amplitudes larger than 100 arc min. Consequently, as the measured rates include microsaccades and small saccades reflecting the overt orientation of visual attention, we use the term saccade rate in the context of Experiment 2.

Epochs with blinks were excluded from further analysis. Microsaccades were detected using the algorithm proposed previously (Engbert and Kliegl, 2003; Engbert and Mergenthaler, 2006). Microsaccades were detected in 2D velocity space using a threshold of $\lambda = 5$ times the median-based SD of the velocity time series per trial, separately for horizontal and vertical components. The resulting high-velocity eye-movement segments were considered as microsaccades only if (1) velocity was above threshold for a duration >6 ms, (2) there was binocular overlap between the two eyes of at least one sample, and (3) there was a minimum intersaccadic interval of 20 ms to the previous microsaccade. To compute microsaccade rates, the sum of microsaccades per participant, condition, and time point was calculated and normalized by the number of trials and the sampling rate. A causal smoothing kernel $\omega(\tau) = \alpha^2 \tau \exp(-\alpha \tau)$ for moving average calculations with a decay parameter of $\alpha = 1/20$ ms was applied (introducing a 20 ms delay of the maximum impact of a saccade on the saccadic rate; Dayan and Abbott,
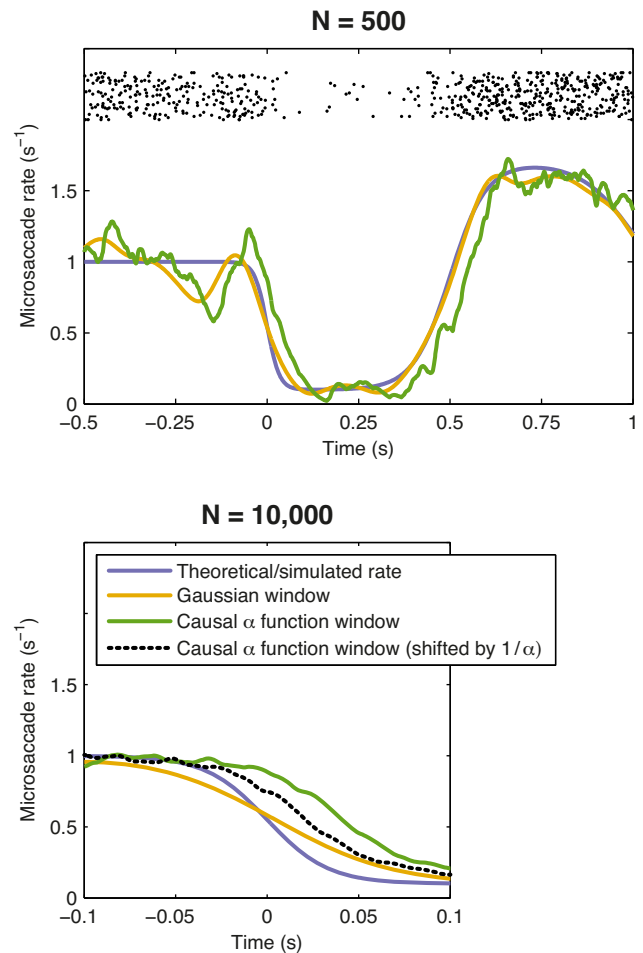

Figure 2. Top, Realizations of a time-dependent Poisson process (scatter plot) for a known time-varying rate (blue line). Rate estimation was performed by a Gaussian (yellow) and causal smoothing kernel (green). The plot is based on N = 500 simulated trials. Bottom, A plot of N = 10,000 simulated trials indicates that the Gaussian estimate (yellow line) introduces a bias toward an earlier onset of microsaccadic inhibition at t < 0, while the causal estimate (green line) represents a more conservative estimate of the onset of reduced rate. The delay introduced by the causal estimate can be accounted for by shifting the estimate by 1/α in time (black dashed line).

2001; Rolfs et al., 2008; Vanrullen, 2011). The microsaccade rate was averaged across participants separately for each sound type.

To investigate the numerical properties of our rate-estimation procedure based on the causal smoothing kernel, we simulated a time-dependent Poisson process as a model for microsaccadic onset times with time-dependent rate $r(t)$ and applied our algorithm to the simulated data (Fig. 2). Microsaccade onset times were generated, first, by simulation of a homogeneous (i.e., time-independent) Poisson process with constant rate $r_{max} > r(t)$ and, second, by subsequent thinning of sequences of events via rejection sampling: event i at time t_i is rejected with probability $p(t_i) = 1 - r(t_i)/r_{max}$. Resulting simulated events ($N = 500$ runs) are shown as the scatter plot in Figure 2, top. Next we applied the same rate-estimation procedure to the simulated Poisson process as used in the analyses of experimental data. While both Gaussian and causal smoothing kernels can be used to reconstruct the theoretical Poisson rate $r(t)$, the causal estimate introduces a delay of size $1/\alpha$ (Fig. 2, top). Based on $N = 10,000$ simulated trials we plotted the rate estimates around the time of the start of the reduced rate at $t = 0$ (as a model for microsaccadic inhibition). The curves (Fig. 2, bottom) indicate that the Gaussian kernel underestimates the onset of microsaccadic inhibition, while the causal kernel leads to a more conservative estimate. Shifting the causal kernel by $1/\alpha$ in time gives the most reliable estimate in our simulations (Fig. 2, black dashed curve).
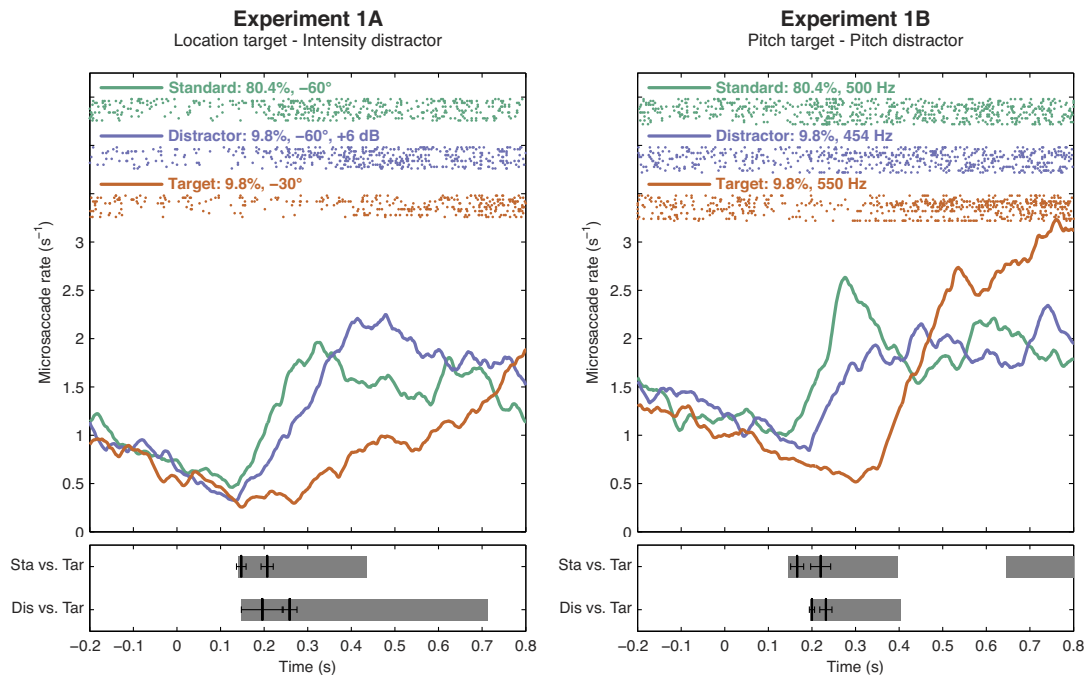
Figure 3. Microsaccade rates in response to standard (green), distractor (blue), and target sounds (red; upper row), as well as significant clusters (gray bars; lower row) in the standard versus target and distractor versus target comparisons. Black lines indicate the jackknife latency and SE estimate of a relative 20 and 50% peak amplitude criterion of the standard minus target and distractor minus target differences. The raster plots illustrate accumulated microsaccade data from 25 randomly chosen trials per condition and participant. Each line represents one participant; each dot corresponds to a microsaccade observed at the corresponding time point.

Statistical analysis. Nonparametric cluster-based randomization tests (Maris and Oostenveld, 2007) were computed for the difference in the microsaccade rate between standard and target sounds and distractor and target sounds (Experiments 1A and 1B) and congruent and incongruent sounds (Experiment 2), respectively. Cluster-based randomization tests were performed using the Fieldtrip MATLAB toolbox (Oostenveld et al., 2011). The onset latency of each significant cluster was determined. The latency and SEs of relative 20 and 50% peak amplitude criteria were estimated from the difference in the microsaccade rate of standard minus target sounds and distractor minus target sounds using the jackknifing technique (Miller et al., 1998; Kiesel et al., 2008). An α-level of 0.05 was applied.

Results
Experiment 1A: oddball paradigm with location target and intensity distractor
Behavioral performance
On average, 93.8% of target sounds were correctly detected (min = 77.7%; max = 100%; SD = 7.1%). The mean reaction time for correctly detected target sounds was 508 ms (min = 412 ms; max = 621 ms; SD = 75 ms). The average false alarm rate was 0.2% for standard (min = 0%; max = 0.7%; SD = 0.3%) and also 0.2% for distractor sounds (min = 0%; max = 0.9%; SD = 0.3%).

Microsaccade rates
Microsaccade rates are shown in Figure 3. The time course of microsaccade rate shows a clear microsaccadic inhibition effect across all experimental conditions. The microsaccadic activity was minimal at 128 ms for standard sounds, 136 ms for distractor sounds, and 150 ms for target sounds relative to sound onset. Following the microsaccadic inhibition, a rebound was observed for standard and distractor sounds peaking at 322 ms for standard sounds and 478 ms for distractor sounds. For target sounds no distinct rebound but rather a sustained inhibition of microsaccades was observed. The microsaccade rates in response to

standard and target sounds were significantly different in a cluster between 142 and 436 ms ($p < 0.001$). The jackknife estimates of the latency of relative 20 and 50% peak amplitude criteria of the standard minus target difference were 148 (SE = 11 ms) and 207 ms (SE = 14 ms). The microsaccade rates in response to distractor and target sounds were significantly different in a cluster between 148 and 712 ms ($p < 0.001$). The jackknife estimates of the latency of relative 20 and 50% peak amplitude criteria of the distractor minus target difference were 196 (SE = 48 ms) and 259 ms (SE = 17 ms), respectively.

Experiment 1B: oddball paradigm with pitch targets and pitch distractors
Behavioral performance
On average, 99.2% of target sounds were correctly detected (min = 95.4%; max = 100%; SD = 1.3%). The mean reaction time for correctly detected target sounds was 459 ms (min = 340 ms; max = 568 ms; SD = 64 ms). The average false alarm rate was 0.03% for standard (min = 0%; max = 0.1%; SD = 0.06%) and 0.8% for distractor sounds (min = 0%; max = 1.8%; SD = 0.8%).

Microsaccade rates
Microsaccade rates are shown in Figure 3. Microsaccade rates showed a microsaccadic inhibition effect in all conditions. The microsaccadic activity was minimal at 140 ms for standard sounds, 194 ms for distractor sounds, and 300 ms for target sounds relative to sound onset. Following the microsaccadic inhibition, a rebound was observed in all conditions peaking at 276 ms for standard sounds, 450 ms for distractor sounds, and 536 ms for target sounds. As in Experiment 1A for target sounds a sustained inhibition of microsaccades was observed. The microsaccade rates in response to standard and target sounds were significantly different in two clusters between 146 and 398 ms
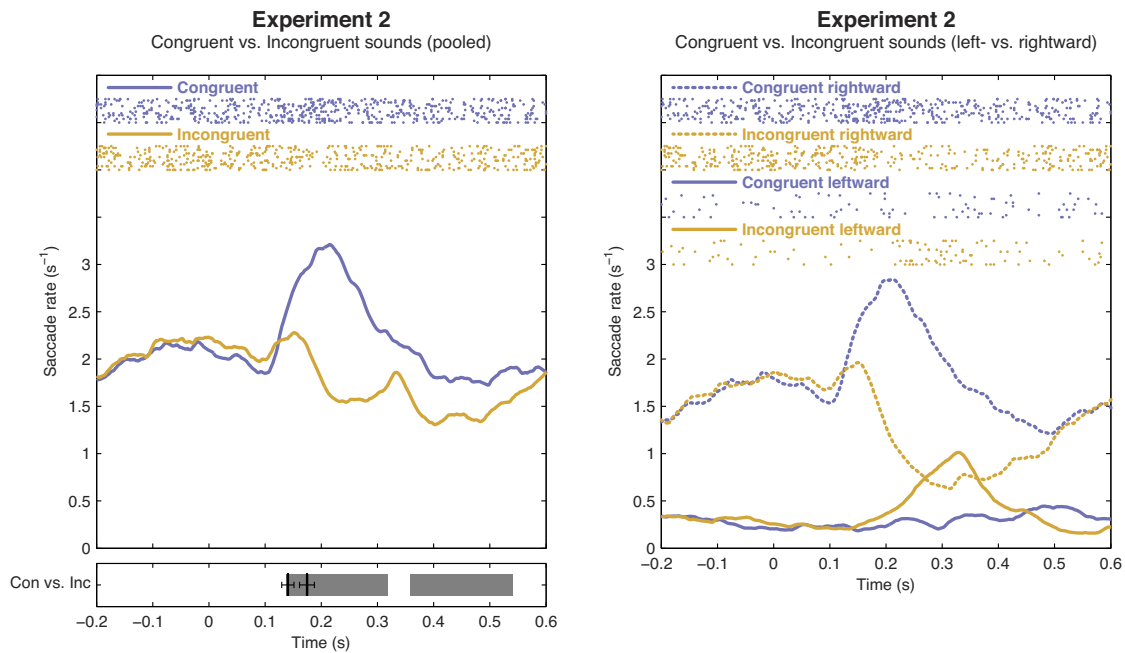
Experiment 2
Congruent vs. Incongruent sounds (pooled)

Experiment 2
Congruent vs. Incongruent sounds (left- vs. rightward)

Figure 4. Saccade rates in response to congruent (blue) and incongruent sounds (orange; upper row) to the corresponding visual trial, as well as significant clusters (gray bars; lower row) in the congruent versus incongruent comparison. Black lines indicate the jackknife latency and SE estimate of a relative 20 and 50% peak amplitude criterion of the congruent minus incongruent differences. The raster plots illustrate accumulated microsaccade data from 25 randomly chosen trials per condition and participant. Each line represents one participant; each dot corresponds to a (micro-)saccade observed at the corresponding time point.

($p < 0.001$) and 646–900 ms ($p = 0.012$). The jackknife estimates of the latency of relative 20 and 50% peak amplitude criteria of the standard minus target difference were 166 (SE = 15 ms) and 220 (SE = 14 ms). The microsaccade rates in response to distractor and target sounds were significantly different in a cluster between 196 and 404 ms ($p = 0.002$). The jackknife estimates of the latency of relative 20 and 50% peak amplitude criteria of the distractor minus target difference were 200 (SE = 6 ms) and 232 ms (SE = 14 ms).

Experiment 2: symbol-to-sound matching paradigm
Behavioral performance
Of the "last tone" responses 6.8% were given too early (before onset of the last sound) or later than 600 ms after the onset of the last sound. More last tone onset-detection responses were given within the response time window in congruent than incongruent trials (96.2 vs 90.2%; $t_{(13)} = 4.10$, $p = 0.001$). The mean reaction time of the last tone responses given within the reaction time window was 322.6 ms. The reaction time was lower in congruent trials compared with incongruent trials (310 vs 335 ms; $t_{(13)} = 4.58$, $p < 0.001$). Of trials with a last tone response given within the reaction time window 96.3% were categorized correctly with respect to congruency. The accuracy of trial congruency categorization was higher in congruent than incongruent trials (98.3 vs 94.4%; $t_{(13)} = 3.62$, $p = 0.003$).

Saccade rates
Saccade rates are displayed in Figure 4. Saccade rates did not show a pronounced microsaccadic inhibition after sound onset. There was a pronounced increase in the saccade rate in response to congruent sounds starting 102 ms and peaking 216 ms after sound onset. The saccade rates in response to congruent and incongruent sounds were significantly different in two clusters between 138 and 318 ms ($p < 0.002$) and 358–542 ms ($p = 0.022$). The jackknife estimates of the latency of relative 20 and

50% peak amplitude criteria of the congruent minus incongruent difference were 140 (SE = 11 ms) and 174 ms (SE = 13 ms).

In an additional descriptive analysis saccade rates of leftward- and rightward-directed saccades were computed separately. Leftward and rightward directions were defined as symmetric 135° sectors in the left and right hemifield, respectively. Therefore, saccades in a 45° sector to the top and bottom were excluded from analysis. The baseline rate was lower for leftward saccades than rightward saccades. The early difference between congruent and incongruent sounds in saccade rate was observed for rightward saccades only. A later increase in saccade rate ~300 ms post stimulus onset for incongruent but not congruent sounds was observed for leftward saccades only.

Discussion
Experiments 1A and 1B
Results show stereotypical microsaccadic inhibition in both experiments. The microsaccadic rebound was slightly delayed for distractor sounds compared with standard sounds. Target sounds showed sustained inhibition and a strongly delayed rebound. A significant difference in the microsaccade rate in response to standard and target sounds was found as early as 142–148 ms after sound onset with converging evidence from different statistical approaches. A significant difference in the microsaccade rate in response to distractor and target sounds was found as early as 148 ms after sound onset if distractor and target were defined by intensity and location and 196 ms if they were both defined by pitch.

For the interpretation of the observed latencies, we applied the most conservative estimates of the underlying neural delays based on observed overt eye-movement behavior. First, the motor delay between stimulation of the SC to the evoked saccade is assumed to be at least 20 ms (Robinson, 1972; Sparks, 1986). Second, we used causal filtering to compute the most conservative estimate of onset latency as a benchmark. The use of noncausal filters is

discouraged for the measurement of onset latencies, because it can result in systematic underestimation of onset latencies due to smoothing (Vanrullen, 2011; Rousselet, 2012; Widmann and Schröger, 2012). Causal filtering will, in contrast, result in systematic overestimation of onset latencies due to the filter delay (here ~20 ms) introduced in the signal. In an earlier study microsaccades rate latencies were corrected by the causal filter delay (Rolfs et al., 2008). Additionally, the simulation of the impact of causal filtering (Fig. 2) confirmed that the onset latency would not be systematically underestimated when corrected by the filter delay. That is, based on these results, an adjustment of the observed latencies by 20 ms appears justified. Third, the signal has to be transmitted from the auditory system to the SC of the saccadic system. The transmission delay and the pathway are yet unknown. A potential pathway candidate includes the caudate nucleus and the pars reticulata of the substantia nigra (SNPR; Hikosaka and Wurtz, 1983). Both have been shown to include neurons sensitive to auditory input. The caudate nucleus receives afferent projections from the cortex and the thalamus and projects to the SNPR. The SNPR has strong inhibitory projections to the SC saccade system (Hikosaka and Wurtz, 1983; Martinez-Conde et al., 2013). The transmission delay from auditory cortex to the SC necessarily must be larger than zero. We assume that a delay in the range of 5 to maximally 20 ms is a realistic estimate.

As a consequence, the observation of a significant difference in observed microsaccade rates as early as 142 ms after sound onset indicates that, in the oddball paradigm, information about sound category (target vs nontarget) is available already at ~80–100 ms after sound onset. This is a remarkable finding in contrast to earlier results, since it was previously assumed that processing of stimulus identity and category is reflected by the N2 and N2b ERP components occurring ~200–220 ms after sound onset (Patel and Azzam, 2005; Folstein and Van Petten, 2008). It was suggested that N2 reflects processes of attentive target discrimination (Ritter et al., 1979; Näätänen, 1992) and/or stimulus matching to a template (Gehring et al., 1992). Considering that the behavioral responses were as late as ~500 ms suggests that additional evaluation and selection stages have to be completed after the categorization of the sound as a target and before the execution of the button press response.

Recent research revealed that pre-attentive deviance detection in the auditory modality is a fast process including strong contributions from early subcortical structures (Alho et al., 2012; Grimm et al., 2012; Slabu et al., 2012). However, fast deviance processing is not sufficient to explain our observed pattern of results as also the modulation of microsaccade rate between rare "deviant" stimuli, the nontarget distractor, and the target showed an early significant difference with an onset latency of 148 ms (including causal filter delay, transmission delay, and motor delay). In fact, by using a three-tone (with rare target and nontargets) instead of a two-tone oddball paradigm (with only rare targets), one can exclude that the classification of the sounds as a target is solely based on its relative rareness, which is known to be encoded by the brain as early as 30 ms after sound onset in some cases (for review, see Grimm and Escera, 2012 for review).

The functional significance of the microsaccadic inhibition effect (for review, see Rolfs et al., 2008) has not yet received much experimental attention. Most likely the effect is due to an unspecific inhibition of the saccadic system by selective attention in response to task-relevant events. In a recently proposed computational model of fixational eye movements (Engbert, 2012), both slow fixational eye movements and microsaccades are produced by fluctuations of a self-generated activation map (Engbert et al.,

2011). In the model, spatial-selective attention modulates microsaccade probabilities via transient changes of a movement potential in the SC. Here we show that the model should even be extended to include modality-unspecific and nonspatial attention. Next, we tried to extend our findings to a paradigm on fast sound identification and categorization with directed saccadic behavior.

Experiment 2

Saccade rates in response to congruent sounds showed a strong increase of (rightward) saccades 100–200 ms after sound onset but incongruent sounds did not. The saccade rate between congruent and incongruent sounds showed a statistically significant difference as early as 138 ms after sound onset with converging evidence from different statistical approaches. Please note that this onset latency measure is the most conservative estimate and includes a transmission delay, motor delay (>20 ms), and filter delay (~20 ms) of 40 and up to 60 ms as in Experiment 1.

The most straightforward interpretation of the observed pattern of results is that covert and overt visual spatial attention is directed to the next visual symbol to prepare for the upcoming sound in case of congruent sounds. The predominance of rightward saccades reflects this overt orienting of attention. In case of incongruent sounds (comparable to targets in the oddball paradigm, as they define the response required at the end of the trial) visual spatial attention is not directed to the next stimulus, but rather target-related processes are initiated.

This interpretation is in line with evoked gamma-band response (eGBR) effects observed in the symbol-to-sound matching paradigm, where congruent (but not incongruent) sounds elicit an eGBR peaking ~40–50 ms after sound onset (Widmann et al., 2007). It seems likely that a template of the expected sound is pre-activated and can thus be compared with the actual sensory input at early levels of subcortical and cortical sensory processing. Match of expectation and experience result in an augmented eGBR (Herrmann et al., 2004). This is compatible with the present finding that sound identification including categorization into congruent versus incongruent occurs 80–100 ms after sound onset as reflected by the enhanced saccade rate, most likely due to allocation of overt attention to the next visual stimulus to prepare for the next expected sound as fast as possible. However, reinterpretation is required for the IR and N2 components observed in the symbol-to-sound matching paradigm (Widmann et al., 2004), as they seem to be elicited after the categorization of a sound.

We would like to note here that the observed effects of stimulus category on the saccadic rate in the symbol-to-sound matching paradigm most likely imply that the induced gamma-band effects reported by Widmann et al. (2007, 2012) are spurious and reflect contamination by saccadic potentials elicited by the contraction of extraocular muscles at the onset of saccades (Yuval-Greenberg et al., 2008; Yuval-Greenberg and Deouell, 2011). This, importantly, does not affect the validity of the ERP and eGBRs reported by Widmann et al. (2007, 2012).

Conclusion

Eye movements were measured in two different experimental paradigms comprising different dependent variables and different statistical approaches. Considering the neural transmission delays, motor delay, and delay introduced by data analysis, the initial categorization of a sound as a target or nontarget is present (and affects behavior) as early as ~80–100 ms after sound onset. This is before the N1 component of the auditory evoked potential, which is associated with transient detection (Näätänen, 1992) or pre-representational activation of afferent neurons an-

alyzing physical features of sensory input (Näätänen and Winkler, 1999; Joos et al., 2014).

This apparent contradiction—sound identity and category is processed before a sound representation is established— can be resolved when considering that early sensory processing is not only bottom-up stimulus driven but also guided in a top-down manner by prior information as suggested by the predictive coding theory (Friston, 2005, 2010). Predicted stimuli can be represented at lower levels of sensory and cognitive processing and can be compared with the actual sensory input at that level (cf. Schröger et al., 2014).

The measurement of (micro-) saccade rates appears to be a very useful tool for future investigation of the mental chronometry of auditory cognition.

References

Alho K, Grimm S, Mateo-León S, Costa-Faidella J, Escera C (2012) Early processing of pitch in the human auditory system. Eur J Neurosci 36: 2972–2978. CrossRef Medline

Ciuffreda KJ, Tannen B (1995) Eye movement basics for the clinician. St. Louis: Mosby.

Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT.

Dien J, Spencer KM, Donchin E (2004) Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophysiology 41:665– 678. CrossRef Medline

Engbert R (2006) Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. Prog Brain Res 154:177–192. CrossRef Medline

Engbert R (2012) Computational modeling of collicular integration of perceptual responses and attention in microsaccades. J Neurosci 32:8035– 8039. CrossRef Medline

Engbert R, Kliegl R (2003) Microsaccades uncover the orientation of covert attention. Vision Res 43:1035–1045. CrossRef Medline

Engbert R, Mergenthaler K (2006) Microsaccades are triggered by low retinal image slip. Proc Natl Acad Sci U S A 103:7192–7197. CrossRef Medline

Engbert R, Mergenthaler K, Sinn P, Pikovsky A (2011) An integrated model of fixational eye movements and microsaccades. Proc Natl Acad Sci U S A 108:E765–E770. CrossRef Medline

Folstein JR, Van Petten C (2008) Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 45:152–170. Medline

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815– 836. CrossRef Medline

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138. CrossRef Medline

Gardner WG, Martin KD (1995) HRTF measurements of a KEMAR. J Acoust Soc Am 97:3907–3908. CrossRef

Gehring WJ, Gratton G, Coles MG, Donchin E (1992) Probability effects on stimulus evaluation and response processes. J Exp Psychol Hum Percept Perform 18:198–216. CrossRef Medline

Grimm S, Escera C (2012) Auditory deviance detection revisited: evidence for a hierarchical novelty system. Int J Psychophysiol 85:88–92. CrossRef Medline

Grimm S, Recasens M, Althen H, Escera C (2012) Ultrafast tracking of sound location changes as revealed by human auditory evoked potentials. Biol Psychol 89:232–239. CrossRef Medline

Hafed ZM (2011) Mechanisms for generating and compensating for the smallest possible saccades. Eur J Neurosci 33:2101–2113. CrossRef Medline

Hafed ZM, Goffart L, Krauzlis RJ (2009) A neural mechanism for microsaccade generation in the primate superior colliculus. Science 323:940–943. CrossRef Medline

Herrmann CS, Munk MH, Engel AK (2004) Cognitive functions of gamma-band activity: memory match and utilization. Trends Cogn Sci 8:347–355. CrossRef Medline

Hikosaka O, Wurtz RH (1983) Visual and oculomotor functions of monkey substantia nigra pars reticulata. I. Relation of visual and auditory responses to saccades. J Neurophysiol 49:1230–1253. Medline

Joos K, Gilles A, Van de Heyning P, De Ridder D, Vanneste S (2014) From sensation to percept: the neural signature of auditory event-related potentials. Neurosci Biobehav Rev 42:148–156. CrossRef Medline

Key AP, Dove GO, Maguire MJ (2005) Linking brainwaves to the brain: an ERP primer. Dev Neuropsychol 27:183–215. CrossRef Medline

Kiesel A, Miller J, Jolicoeur P, Brisson B (2008) Measurement of ERP latency differences: a comparison of single-participant and jackknife-based scoring methods. Psychophysiology 45:250–274. CrossRef Medline

Kleiner M, Brainard D, Pelli D (2007) What's new in psychtoolbox-3? Perception 36, ECVP Abstract Supplement:14.

Kutas M, McCarthy G, Donchin E (1977) Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197:792– 795. CrossRef Medline

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190. CrossRef Medline

Martinez-Conde S, Otero-Millan J, Macknik SL (2013) The impact of microsaccades on vision: towards a unified theory of saccadic function. Nat Rev Neurosci 14:83–96. CrossRef Medline

Miller J, Patterson T, Ulrich R (1998) Jackknife-based method for measuring LRP onset latency differences. Psychophysiology 35:99–115. CrossRef Medline

Näätänen R (1992) Attention and brain function. Hillsdale, NJ: Lawrence Erlbaum Associates.

Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. Psychol Bull 125:826–859. CrossRef Medline

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869. CrossRef Medline

Patel SH, Azzam PN (2005) Characterization of N200 and P300: selected studies of the event-related potential. Int J Med Sci 2:147–154. Medline

Ritter W, Simson R, Vaughan HG Jr, Friedman D (1979) A brain event related to the making of a sensory discrimination. Science 203:1358– 1361. CrossRef Medline

Robinson DA (1972) Eye movements evoked by collicular stimulation in the alert monkey. Vision Res 12:1795–1808. CrossRef Medline

Rolfs M, Kliegl R, Engbert R (2008) Toward a model of microsaccade generation: the case of microsaccadic inhibition. J Vis 8(11):5.1–5.23. CrossRef Medline

Rousselet GA (2012) Does filtering preclude us from studying ERP time-courses? Front Psychol 3:131. CrossRef Medline

Schröger E, Bendixen A, Denham SL, Mill RW, Böhm TM, Winkler I (2014) Predictive regularity representations in violation detection and auditory stream segregation: from conceptual to computational models. Brain Topogr 27:565–577. CrossRef Medline

Slabu L, Grimm S, Escera C (2012) Novelty detection in the human auditory brainstem. J Neurosci 32:1447–1452. CrossRef Medline

Sparks DL (1986) Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus. Physiol Rev 66:118–171. Medline

Valsecchi M, Turatto M (2009) Microsaccadic responses in a bimodal oddball task. Psychol Res 73:23–33. CrossRef Medline

Valsecchi M, Betta E, Turatto M (2007) Visual oddballs induce prolonged microsaccadic inhibition. Exp Brain Res 177:196–208. CrossRef Medline

Valsecchi M, Dimigen O, Kliegl R, Sommer W, Turatto M (2009) Microsaccadic inhibition and P300 enhancement in a visual oddball task. Psychophysiology 46:635–644. CrossRef Medline

Vanrullen R (2011) Four common conceptual fallacies in mapping the time course of recognition. Front Psychol 2:365. CrossRef Medline

Widmann A, Schröger E (2012) Filter effects and filter artifacts in the analysis of electrophysiological data. Front Psychol 3:233. CrossRef Medline

Widmann A, Kujala T, Tervaniemi M, Kujala A, Schröger E (2004) From symbols to sounds: visual symbolic information activates sound representations. Psychophysiology 41:709–715. CrossRef Medline

Widmann A, Gruber T, Kujala T, Tervaniemi M, Schröger E (2007) Binding symbols and sounds: evidence from event-related oscillatory gamma-band activity. Cereb Cortex 17:2696–2702. CrossRef Medline

Widmann A, Schröger E, Tervaniemi M, Pakarinen S, Kujala T (2012) Mapping symbols to sounds: electrophysiological correlates of the impaired reading process in dyslexia. Front Psychol 3:60. CrossRef Medline

Yuval-Greenberg S, Deouell LY (2011) Scalp-recorded induced gamma-band responses to auditory stimulation and its correlations with saccadic muscle-activity. Brain Topogr 24:30–39. CrossRef Medline

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY (2008) Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58:429–441. CrossRef Medline