Event-related potentials for interaural time differences and spectral cues

Michael J. Hautus[a], Blake W. Johnson[b] and Lincoln J. Colling[b]

Dichotic pitches and mistuned harmonics can each lead to the perception of one or two auditory objects. Comparison of event-related potentials for the perception of one versus two objects reveals an early negative and a late positive component. The relationship of these components with auditory segregation was further investigated using stimuli containing monaural spectral cues to pitch, binaural timing cues to pitch, or a combination of both, interleaved with control stimuli (no pitch). Stimuli containing timing cues or a combination of timing and spectral cues reliably elicited both components, which were of larger amplitude when both cues were present. For stimuli containing only spectral cues, the early component was attenuated in amplitude and no measurable late component was detected. *NeuroReport* 20:951–956 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins.

NeuroReport 2009, 20:951–956

Keywords: auditory-evoked potential, auditory segregation, dichotic pitch, interaural time difference

[a]Research Centre for Cognitive Neuroscience, Department of Psychology, University of Auckland, Auckland, New Zealand and [b]Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia

Correspondence to Michael J. Hautus, Research Centre for Cognitive Neuroscience, Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand
Tel: +649 373 7599 x85924; fax: +649 373 7450;
e-mail: m.hautus@auckland.ac.nz

Received 16 March 2009 **accepted** 3 April 2009

Introduction

Little is currently known about the neural structures that implement the grouping rules of auditory scene analysis. However, a promising neurophysiological clue comes from recent studies, which have measured event-related potential (ERP) responses from human listeners presented with sound mixtures that engage the segregational processes of auditory scene analysis. Alain and colleagues [1,2] reported that complex harmonic sounds containing a mistuned harmonic elicited a characteristic ERP component at a latency of 150–250 ms after the onset of stimulus. They termed this component the 'object-related negativity' (ORN) because it was associated with the perception of two concurrent sound objects. The ORN seems to reflect relatively automatic perceptual processes because it can be elicited independently of whether participants are attending to or ignoring the sounds. However, when listeners are required to actively respond to the mistuned sounds, the ORN is followed by a P300-like response (termed a 'P400') at a latency of about 400 ms, suggested to index neural mechanisms associated with higher level, controlled processes involved in making perceptual decisions about perceptual objects [2,3].

Johnson, Hautus, and colleagues [4–9] measured ERPs to a type of dichotic pitch created by presenting listeners with two broadband noises that are identical except for an interaural time difference for a narrow band of frequencies. The auditory system interprets the binaural lag as a perceptual grouping cue and splits the sound into two concurrent but spatially separated perceptions: a ball of noise and a faint pitch. These authors reported that, in comparison with otherwise identical control stimuli containing no interaural lag, dichotic pitch stimuli elicited an enhanced negativity at a similar latency to the mistuning-related ORN. The dichotic pitch ORN was also followed by a P400 when listeners were required to actively discriminate the sounds [5,6].

Taken together, the studies described above support the interpretation that the ORN indexes the operation of a fairly automatic, preattentive set of auditory mechanisms that can use at least two physically disparate grouping cues (i.e. inharmonicity and interaural time difference) to parse sounds into two or more concurrent auditory sound sources. The goal of this study is to establish whether a broader generalization can be made about the ORN and P400 and their relationship with auditory segregation by investigating a third class of cues, monaural spectral cues to pitch. These cues can be integrated into the framework of dichotic pitch independent of the timing cues inherent in these stimuli, making it possible to also consider the interaction of these monaural pitch cues with the binaural interaural time difference cues.

Methods

Participants

Thirteen volunteers (eight males; mean age = 24.3 years) with no reported history of auditory or neurological illness

0959-4965 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0b013e32832c92bf

took part in the study. All reported being right handed. The project was approved by the University of Auckland Human Participants Ethics Committee.

Auditory stimuli

Three stimulus types that contained either interaural time difference or timing cues [dichotic pitch (DP)], monaural spectral cues to pitch [monaural pitch (MP)], or a combination of both cues [combination pitch (CP)] were created using a complementary filtering method similar to the study by Dougherty *et al.* [10]. A fourth stimulus – atonal noise – was used which did not contain timing or spectral cues.

Two identical 500-ms broadband Gaussian noises (sampling rate, 44.1 kHz) were created. One noise was bandpass filtered (eighth order Butterworth), with a center frequency of 500 Hz (MP), 600 Hz (DP), or 700 Hz (CP) and a 3-dB bandwidth of 50 Hz. The other noise was notch filtered, using the same corner frequencies as the bandpass filter. The filters were designed in one of two ways depending on whether or not the resultant stimulus contained a spectral cue. For stimuli with no spectral cues (DP and noise), the filter functions of the notch and bandpass filters summed to one for all frequencies; this resulted in a flat spectrum after recombination of the outputs of the two filters. For stimuli with spectral cues (MP and CP), the filters were still complementary, but the height of the bandpass filter was three times that of the notch filter; this resulted in stimuli with a spectral peak within the passband of the bandpass filter.

To produce the timing cues, the bandpass-filtered noise was reproduced and one copy was delayed by 0.5 ms. Two spectrally identical noises were obtained by adding the notch-filtered noise to each copy of the bandpass-filtered noise. For the CP stimuli, the bandpass-filtered noise was both time-delayed to one ear and had a higher noise power density than the remaining frequencies. The CP stimuli demonstrate the combined effect of two cue types for a particular (single) pitch percept.

Before presentation, the stimuli were bandpass filtered (fourth order Butterworth) with corner frequencies of 400 and 800 Hz. Rise and fall times were 4 ms ($\cos^2$) and stimuli were generated on two channels of a 16-bit converter (National Instruments, DAQPad 6052E, Austin, Texas, USA). Stimulus levels were 70 dB SPL at the eardrum and delivered diotically through insert earphones (Etymotic Research Inc., Model ER2, Elk Grove, Illinois, USA).

Screening

All participants underwent a screening procedure to ensure that they could accurately perceive dichotic pitch. The procedure was a single-interval task in which participants indicated whether the stimulus presented was dichotic pitch or atonal noise. Participants were required to score above 80% correct on 60 consecutive trials.

Task and procedures

The stimuli were presented in four blocks of 192 trials with one stimulus presented per trial. Of the 192 trials, 96 contained one of the three pitch stimuli (32 trials for each of the DP, CP, and MP stimuli), whereas the other 96 contained the atonal noise stimulus. The order of the stimuli was randomized within each block.

Participants were required to indicate on each trial, by means of a button box placed under their right hand, whether a pitch or a noise stimulus had been presented. Pitch and noise stimuli were equiprobable within a block. The inclusion of a task was necessary to ensure that participants attended to the stimuli, strengthening the likelihood of detecting a P400 component.

Participants were advised to respond as soon as possible and no response could be made after 4000 ms. The intertrial interval was selected from a rectangular distribution between 1500 and 3500 ms. To minimize eye movements, ocular artifacts, and alpha activity, participants were instructed to fixate on a cross that was displayed on a video monitor placed 45 cm in front of them.

Electroencephalogram data acquisition and preprocessing

Electroencephalogram (EEG) recordings were acquired, with a common vertex (Cz) reference, using 128-channel Ag/AgCl electrode nets (Electrical Geodesics Inc., Eugene, Oregon, USA). Continuous EEG recordings (200 MΩ input impedance; 250 Hz sampling rate; 0.1–100 Hz bandpass filter) were conducted inside an electrically shielded room. Electrode impedance was kept below 45 kΩ [11]. EEG files were segmented into 600 ms epochs that included a 100 ms prestimulus baseline. Ocular artifacts were corrected [12]. Four ERPs corresponding to the four stimulus types – MP, CP, DP, and noise – were calculated by averaging together individual epochs for each participant. ERPs were digitally filtered using a bidirectional three-pole Butterworth filter [13] with corner frequencies of 0.1 and 30 Hz, and then re-referenced to the average reference.

Data analysis

Behavioral data

To determine whether a salient pitch was evoked by each of the pitch stimuli, group d' estimates were calculated by averaging d' values attained for each participant [14].

Event-related potential data

Grand averages were calculated by averaging individual participant's ERPs for each stimulus type. Statistical analyses for the ORN were conducted on mean voltage over a 100–210 ms poststimulus time window, collapsed over a subset of eight frontocentral electrodes (corresponding approximately to FCz, FC1, FC3, C4, FC4, FC2, Cz, C3 in the extended 10–20 system [15]). Statistical analyses for the P400 were conducted on the mean voltage over a 340–440 ms poststimulus time window, collapsed over eight right-frontal electrodes (corresponding approximately to FCz, Cz, CPz, FC2, FC4, C2, C4, CP2). These electrode regions of interest were based on electrodes showing maximal ORNs and P400s. ORNs for DP, CP, and MP, and P400s for DP and CP, showed near-identical topographies and were consistent with the topographic distributions of the ORN and P400 reported earlier [4–7]. Time windows were chosen, by visual inspection, to encompass the component of interest without overlapping onto any obviously different, nearby components.

The effects of each of the two cues (timing and spectral) were calculated by comparing average voltages by means of a spectral cue (present or absent) by timing cue (present or absent) repeated-measures analysis of variance (ANOVA), using the Greenhouse–Geisser correction to control for inhomogeneities of variances. Post-hoc comparisons used Bonferroni–Holm corrected paired samples *t*-tests.

Results

Behavioral data

Group d′ values and 95% confidence intervals for the MP, CP, and DP stimuli were 2.38 ± 0.61, 3.28 ± 0.69, and 2.75 ± 0.54, respectively. These values indicate that all pitch stimuli were detected at well above chance. The hit rate (correctly reporting the pitch present) was almost always above 90% for all stimulus types across participants.

Event-related potential data

Stimulus-type effects

ERPs, calculated for correct responses only, exhibit a typical P1–N1–P2–N2 complex with mean latencies of about 64, 112, 200, and 296 ms after the onset of stimulus (Fig. 1). A late positive peak was also observed 364 ms after stimulus onset. The difference waveforms reveal an ORN for each of the three segregated stimulus types: DP [$t(12) = -3.333$, $P = 0.006$], MP [$t(12) = -3.284$, $P = 0.007$], and CP [$t(12) = -9.201$, $P < 0.001$]. A P400 was observed for the DP [$t(12) = 2.635$, $P = 0.022$] and CP [$t(12) = 5.139$, $P < 0.001$] difference waveforms, but not for the MP difference waveform [$t(12) = -0.723$, $P = 0.484$].

Cue-type effects: the object-related negativity component

ANOVA revealed main effects for timing cues [$F(1,12) = 21.954$, $P = 0.001$] and spectral cues [$F(1,12) = 19.174$, $P = 0.001$], but no interaction [$F(1,12) = 1.026$, $P = 0.331$].

In the presence of spectral cues (CP and MP), the addition of timing cues (CP; $M = -0.460$, $SD = 0.654$) resulted in ERPs that were more negative than ERPs recorded without timing cues [MP; $M = 0.130$, $SD = 0.701$; $t(12) = -4.656$, $P = 0.002$]. In the absence of spectral cues (DP and noise), the addition of timing cues (DP; $M = -0.055$, $SD = 0.932$) again resulted in ERPs that were more negative than ERPs recorded without timing cues [noise; $M = 0.394$, $SD = 0.662$; $t(12) = -3.333$, $P = 0.018$].

In the presence of timing cues (CP and DP), the addition of spectral cues (CP) produced ERPs that were more negative than ERPs recorded without spectral cues [DP; $t(12) = -3.323$, $P = 0.012$]. In the absence of timing cues (MP and noise), we again found that the addition of spectral cues (MP) produced ERPs that were more negative than those recorded without spectral cues [noise; $t(12) = -3.284$, $P = 0.007$].

Cue-type effects: the P400 component

ANOVA revealed main effects for timing cues [$F(1,12) = 27.459$, $P < 0.001$] and spectral cues [$F(1,12) = 7.789$, $P = 0.016$], and an interaction [$F(1,12) = 12.015$, $P = 0.005$]. This latter finding indicates that the effect of spectral cues was different under the two conditions of timing cues.

In the presence of spectral cues (CP and MP), the addition of timing cues (CP; $M = 1.088$, $SD = 0.858$) elicited ERPs that were more positive than ERPs recorded without timing cues [MP; $M = 0.034$, $SD = 0.706$; $t(12) = 5.154$, $P = 0.001$]. In the absence of spectral cues (DP and noise), the addition of timing cues (DP; $M = 0.445$, $SD = 0.815$) also elicited ERPs that were more positive than those recorded without timing cues [noise; $M = 0.123$, $SD = 0.640$; $t(12) = 2.636$, $P = 0.044$].

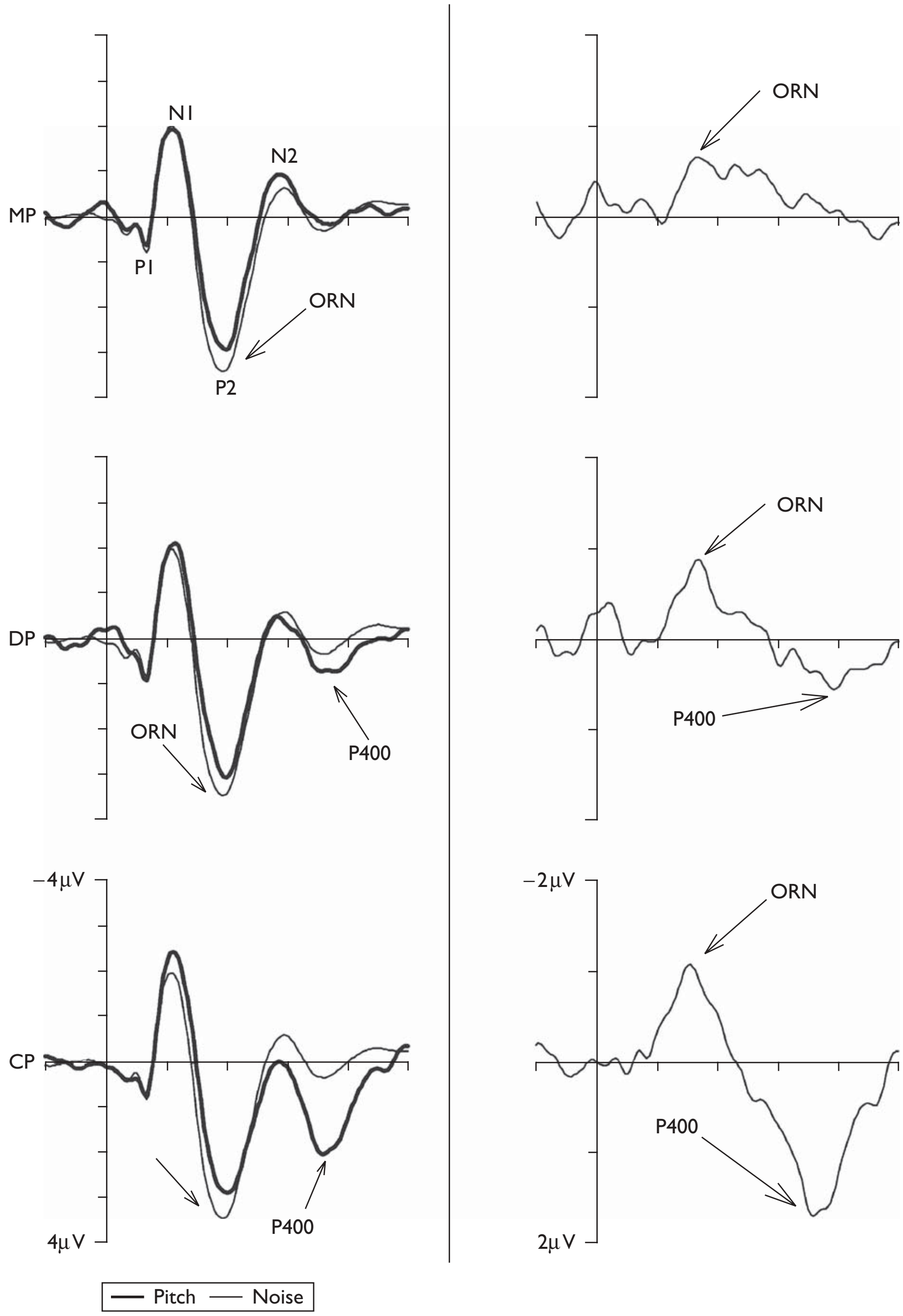
In the presence of timing cues (CP and DP), the addition of spectral cues (CP) resulted in ERPs that were more positive than those recorded without spectral cues (DP). In the absence of timing cues (MP and noise), however, the addition of spectral cues (MP) did not produce ERPs that were different from those recorded without spectral cues (noise).

Discussion

Key findings

There were four main findings. First, all three stimulus types (MP, DP, and CP) evoked an ORN when compared with noise alone. This suggests that the ORN is sensitive

Fig. 1



Left panel: grand-averaged event-related potential (ERP) waveforms for each of the three pitch versus noise comparisons: monaural pitch (MP) versus noise, dichotic pitch (DP) versus noise, and combination pitch (CP) versus noise. Right panel: difference waveforms computed by subtracting the relevant pitch ERP from the noise ERP. The vertical scale is 1 μV/division (negative up) and the horizontal scale is 100 ms/division. The axes intersect at the time of stimulus onset. ORN, object-related negativity.

to the processes of segregation initiated by the presence of both timing and spectral cues, and additionally, monaural and binaural cues to pitch. Second, an ORN was evoked by the addition of timing cues regardless of whether the timing cues resulted in the creation of a pitch, that is, there was a difference between CP ERPs and MP ERPs even though CP and MP stimuli do not differ with respect to their spectral shape; the main difference is whether timing cues were present (CP) or absent (MP). This is consistent with the interpretation that the ORN is of larger magnitude when more cue types (current study), or stronger cues [4], are present.

Third, spectral cues were also capable of eliciting an ORN regardless of whether or not they were paired with timing cues (the MP versus noise and CP versus DP comparisons). Finally, a P400 was only evoked when timing cues were present in the stimulus, and, although the P400 is known to vary systematically with pitch salience [1], it appeared not to be evoked by spectral cues alone (MP). It should be noted that although MP was the least salient of the three pitch stimuli it was still detected at well above chance and, therefore, would be expected to elicit a P400, if a P400 can be evoked by spectral cues alone.

Effect of cue type on the object-related negativity component

It is clear that timing cues are able to produce an ORN even when both stimuli produce a pitch percept. An ORN, between 100 and 210 ms after the onset of stimulus, was found when comparing DP ERPs to noise ERPs, and also when comparing CP ERPs to MP ERPs; these two ORNs were not significantly different in magnitude. This suggests that the observed ORN reflects the processing of the timing cues because neither of these comparisons contained a change in spectral cues.

We also aimed to determine whether or not the auditory system is sensitive to pitch cues in the absence of timing cues. An ORN was observed when comparing MP ERPs to noise ERPs; these stimuli are spectrally dissimilar and differ, that is, MP produces a pitch percept whereas noise does not. This clearly indicates that the ORN is associated with a third cue for object segregation, spectral shape.

Effect of cue type on the P400 component

The P400 is dependent on attention, occurring only when participants actively attend to a stimulus by responding during perceptual segregation tasks [1]. Furthermore, the P400 is context dependent, and occurs only when presenting stimuli requiring perceptual segregation interleaved with control stimuli [6]. These authors have suggested that the P400 may be sensitive to changes in a number of perceptual objects present in the auditory scene.

The current results confirm the findings of Alain *et al.* [1] in suggesting that the P400 varies systematically with stimulus salience. Although timing cues alone were able to generate a P400, pairing timing cues with spectral cues, thus producing a more salient percept, led to a larger P400. Interestingly, for the least salient of the three pitch stimuli, MP, no P400 was observed. These findings provide further evidence that the ORN indexes the processing of physical cues, and the P400 reflects top-down perceptual processes [1,9].

The case against specificity

On the basis of their work with mistuned harmonic stimuli, Alain *et al.* [2] suggested that the ORN reflects the operation of fairly specialized auditory mechanisms that register a mismatch between a harmonic template extracted from a complex harmonic stimulus and the harmonic frequency expected based on the fundamental of the complex sound. However, the fact that the ORN can also be elicited by dichotic pitch argues against such specificity: As the noises that evoke dichotic pitch have no harmonic structure to match to, the dichotic pitch ORN must be based purely on binaural processing of an interaural timing difference. The present results also show that the ORN can be elicited by monaural cues derived from spectral shape. Again, there is no harmonic template to match to; these monaural pitches were also narrow-band noises.

Conclusion

The ORN is sensitive to a fairly wide class of cues to object segregation and therefore reflects the activity of rather general mechanisms of auditory scene analysis. The ORN seems to reflect the operation of mechanisms involved in solving a perceptual problem – that of parsing the information in an acoustic wave that originates from separate sources in the world – rather than the operation of mechanisms devoted to the processing of any particular cue to these sources.

Acknowledgement

This work was supported by a grant from the University of Auckland Research Committee.

References

1 Alain C, Arnott SR, Picton TW. Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J Exp Psychol Hum Percept Perform* 2001; **27**:1072–1089.
2 Alain C, Schuler BM, McDonald KL. Neural activity associated with distinguishing concurrent auditory objects. *J Acoust Soc Am* 2002; **111**:990–995.
3 Alain C. Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear Res* 2007; **229**:225–236.
4 Clapp WC, Johnson BW, Hautus MJ. Graded cue information in dichotic pitch: effects on event-related potentials. *Neuroreport* 2007; **18**:365–368.
5 Hautus MJ, Johnson BW. Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *J Acoust Soc Am* 2005; **117**:275–280.
6 Johnson BW, Hautus M, Clapp WC. Neural activity associated with binaural processes for the perceptual segregation of pitch. *Clin Neurophysiol* 2003; **114**:2245–2250.
7 Johnson BW, Hautus MJ, Duff DJ, Clapp WC. Sequential processing of interaural timing differences for sound source segregation and spatial localization: evidence from event-related cortical potentials. *Psychophysiology* 2007; **44**:541–551.
8 Johnson BW, Hautus MJ, Hayns AL, Fitzgibbon BM. Differential cortical processing of location and pitch changes in dichotic pitch. *Neuroreport* 2006; **17**:389–393.
9 Johnson BW, Muthukumaraswamy SD, Hautus MJ, Gaetz WC, Cheyne DO. Neuromagnetic responses associated with perceptual segregation of pitch. *Neurol Clin Neurophysiol [electronic resource]: NCN* 2004; **2004**:33.

10 Dougherty RF, Cynader MS, Bjornson BH, Edgell D, Giaschi DE. Dichotic pitch: a new stimulus distinguishes normal and dyslexic auditory function. *Neuroreport* 1998; **9**:3001–3005.
11 Ferree TC, Luu P, Russell GS, Tucker DM. Scalp electrode impedance, infection risk, and EEG data quality. *Clin Neurophysiol* 2001; **112**:536–544.
12 Jervis BW, Nichols MJ, Allen EM. The assessment of two methods for removing eye movement artefact from the EEG. *Electroencephalogr Clin Neurophysiol* 1985; **61**:444–452.
13 Alarcon G, Guy CN, Binnie CD. A simple algorithm for a digital three-pole Butterworth filter of arbitrary cut-off frequency: application to digital electroencephalography. *J Neurosci Methods* 2000; **104**:35–44.
14 Macmillan NA, Kaplan HL. Detection theory analysis of group data. Estimating sensitivity from average hit and false-alarm rates. *Psychol Bull* 1985; **98**:185–199.
15 Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol* 2001; **112**:713–719.