International Journal of Psychophysiology 107 (2016) 16–28

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Promoting the perception of two and three concurrent sound objects: An event-related potential study

Zsuzsanna Kocsis [a,b,*], István Winkler [a,c], Alexandra Bendixen [d], Claude Alain [e,f]

[a] Institute of Psychology and Cognitive Neuroscience, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2., Budapest, H-1117, Hungary
[b] Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Egry József u. 1., Budapest, H-1111, Hungary
[c] Institute of Psychology, University of Szeged, Egyetem u. 2., Szeged, H-6722, Hungary
[d] Cognitive Systems Lab, Institute of Physics, Technische Universität Chemnitz, Reichenhainer Str. 70, Chemnitz, D-09126, Germany
[e] Rotman Research Institute, Baycrest Centre, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada
[f] Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada

ARTICLE INFO

Article history:
Received 17 February 2016
Received in revised form 24 June 2016
Accepted 29 June 2016
Available online 1 July 2016

Keywords:
Object-related negativity (ORN)
Concurrent sound segregation
Three simultaneous sound objects
Cue combinations

ABSTRACT

The auditory environment typically comprises several simultaneously active sound sources. In contrast to the perceptual segregation of two concurrent sounds, the perception of three simultaneous sound objects has not yet been studied systematically. We conducted two experiments in which participants were presented with complex sounds containing sound segregation cues (mistuning, onset asynchrony, differences in frequency or amplitude modulation or in sound location), which were set up to promote the perceptual organization of the tonal elements into one, two, or three concurrent sounds. In Experiment 1, listeners indicated whether they heard one, two, or three concurrent sounds. In Experiment 2, participants watched a silent subtitled movie while EEG was recorded to extract the object-related negativity (ORN) component of the event-related potential. Listeners predominantly reported hearing two sounds when the segregation promoting manipulations were applied to the same tonal element. When two different tonal elements received manipulations promoting them to be heard as separate auditory objects, participants reported hearing two and three concurrent sounds objects with equal probability. The ORN was elicited in most conditions; sounds that included the amplitude- or the frequency-modulation cue generated the smallest ORN amplitudes. Manipulating two different tonal elements yielded numerically and often significantly smaller ORNs than the sum of the ORNs elicited when the same cues were applied on a single tonal element. These results suggest that ORN reflects the presence of multiple concurrent sounds, but not their number. The ORN results are compatible with the horse-race principle of combining different cues of concurrent sound segregation.

1. Introduction

In everyday situations, we are often surrounded by sounds emanating from multiple sources. Although the acoustic energy from these sources sums into a complex acoustic wave, our auditory system is proficient in parsing this mixture into separate sound sources (i.e., auditory streams [Bregman, 1990] or perceptual objects [e.g., Kubovy and Van Valkenburg, 2001]). Early behavioral studies have identified several cues that contribute to the separation of concurrent sound sources. These include differences in frequency periodicity (Hartmann, 1985; Hartmann et al., 1986; Moore et al., 1986) and in location (Bronkhorst and Plomp, 1988) as well as onset asynchrony (Bregman and Pinker, 1978; Rasch, 1978). Other cues, such as amplitude and frequency modulation, have not been investigated to the same extent, and the results are somewhat inconsistent. For instance, some studies show that slow amplitude or frequency modulations are effective segregators for tones (McAdams, 1984a, 1984b; Dolležal et al., 2012). Along the same lines, other studies showed a lower likelihood for segregation when the tonal elements of a complex sounds are modulated at the same rate than when the rate differs across partials (Bregman et al., 1985; Bregman et al., 1990). In contrast, another study reported no benefit from having different rates of modulation in parsing two different vowels except when the harmonics of one vowel were modulated while harmonics of the other vowel remained stationary (Summerfield et al., 1992). Carlyon (1991) also showed that listeners could not reliably discriminate between coherent and incoherent frequency modulation of complex tones. Thus, further research is needed

* Corresponding author at: Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2., Budapest, H-1117, Hungary.
E-mail addresses: kocsis.zsuzsanna@ttk.mta.hu (Z. Kocsis), winkler.istvan@ttk.mta.hu (I. Winkler), alexandra.bendixen@physik.tu-chemnitz.de (A. Bendixen), calain@research.baycrest.org (C. Alain).

http://dx.doi.org/10.1016/j.ijpsycho.2016.06.016
0167-8760/

to better understand the role of frequency and amplitude modulation in concurrent sound segregation.

Many studies have investigated the impact of individual cues on segregating concurrent sounds. In comparison, relatively few studies have considered how the information from different cues is integrated (McDonald and Alain, 2005; Kocsis et al., 2014; Weise et al., 2012). To date, the effects of multiple segregation cues on concurrent sound perception have been investigated by combining two or more *convergent* cues in a way to promote the perception of two concurrent sound objects. That is, either multiple cues were applied to the same tonal element (e.g., having the mistuned tonal element presented at a different location or with a temporal delay) or identically to two different tonal elements, thus promoting the two elements to be grouped together and, again promoting the perception of two concurrent sounds (the complex tone resulting from grouping the two manipulated tonal elements and the complex tone resulting from grouping the unmanipulated tonal elements). An unresolved issue is whether the presence of multiple *divergent* cues, that is, cues promoting different groupings of tonal elements (e.g., mistuning one partial while presenting another partial with temporal delay) could lead to the perception of three (or more) concurrent sound objects. In the present study, divergent cues were operationalized as manipulations of two different tonal elements each promoting the segregation of the target element both from the rest of the harmonic complex and from the other manipulated element (see Fig. 1). Cues acting on the same tonal element were always convergent, whereas cues acting on different tonal elements were always divergent with respect to each other.

The present study aims to investigate whether manipulations of two different tonal elements of a harmonic complex tone could promote the perception of three concurrent sound objects and elicit separate object-related negativity components (ORN; Alain et al., 2001) of the event-related brain potential (ERP). The ORN peaks between 150 and 180 ms from cue onset with maximal amplitude at frontal and frontocentral electrodes. With nose reference, it inverts polarity at the mastoids (Alain et al., 2002), consistent with generators located in the superior temporal gyrus near Heschl's gyrus (Alain et al., 2001; Arnott et al., 2011). ORN has been shown to be larger at the mastoid electrodes during active listening (when listeners were required to judge whether they heard one or two concurrent sounds) than during passive listening (listeners had no task related to the sounds), which indicates attentional modulation of the ORN amplitude (Alain et al., 2001). The ORN can be elicited by many different cues inducing concurrent sound perception including inharmonicity (Alain et al., 2001, 2002; Bendixen et al., 2010), onset asynchrony (Lipp et al., 2010; Weise et al., 2012), dichotic pitch (Johnson et al., 2003; Hautus et al., 2009), differences in the fundamental frequency (Δf_0) of speech sounds (Snyder and Alain, 2005; Alain et al., 2005), and simulated echo (Sanders et al., 2008a, 2008b). There are also reports of ORN being elicited by a combination of some of the above cues, such as inharmonicity and location difference (McDonald and Alain, 2005) or inharmonicity and onset asynchrony (Weise et al., 2012; Kocsis et al., 2014).

Du et al. (2011) showed that the combined effect of location and Δf_0 on the amplitude of the magnetic equivalent of the ORN response closely matched the sum of the ORN responses elicited by the single cues (i.e., location <u>or</u> Δf_0 alone). However, Kocsis et al. (2014) found sub-additive effects of combining inharmonicity, onset asynchrony, and source location difference. Kocsis et al. (2014) used either one of the three single-cue manipulations or combined two or all three cues for segregation. The manipulations affected either one or two tonal elements in a congruent manner (i.e., cues promoting the two tonal elements to be grouped into a single sound object by e.g., same percentage of mistuning or same temporal delay applied to two different tonal elements). In different blocks of trials, participants either watched a subtitled, muted movie (no response required), or were asked to focus on the stimuli and to press a button indicating whether they heard one or two concurrent sound objects. Participants performed generally well (above 87%) in identifying two objects in most conditions. The main finding was that cue combinations always elicited numerically smaller ORN amplitudes than the sum of the ORN amplitudes separately elicited by the comprising cues. That is, the ORN amplitude showed subadditivity to various combinations of different cues promoting the perception of two concurrent sound objects. This suggests that ORN reflects the overall read-out of the auditory system's assessment of the presence of two objects as opposed to indexing the processing of the different cues.

In the present study, we compared the effects of convergent and divergent cues on perceptual and neural (ORN) indicators of concurrent sound segregation. In Experiment 1, we investigated the synergic effect of various cues (i.e., harmonicity, temporal delay, AM, and FM) in conjunction with a location cue (applied to the same tonal element) on the perception of two concurrent sound objects. We also tested whether applying divergent segregation cues on two different tonal elements would promote the perception of three auditory objects. We used four different cues in conjunction with location difference to assess their potential strength and tested how perception of two vs. three sound objects occurs with different cue combinations. In Experiment 2, we recorded the electroencephalogram (EEG) in a passive listening condition to test whether the same stimuli elicit significant ORN responses and to examine whether the ORN amplitudes show additivity (or super/sub-additivity) for cue combinations. This allows us to distinguish two functional interpretations of the ORN component: If ORN elicited by the divergent manipulations (three-objects conditions) is as large as the summed amplitudes of the ORNs elicited separately by the

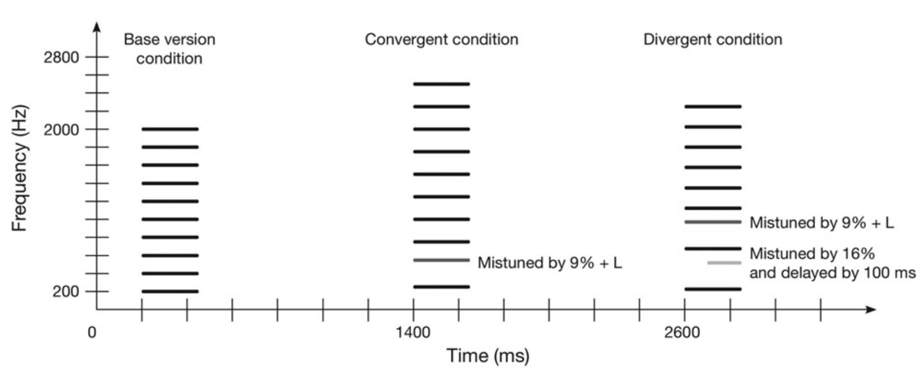

Fig. 1. Schematic depiction of three consecutive trials illustrating stimuli in the base version, convergent, and divergent conditions, which are set up to promote the perception of one, two, and three concurrent objects, respectively. The x axis depicts time, while the y axis depicts frequency. The black horizontal lines depict the frequency components of the harmonic complex. The lighter grey lines depict tonal elements that were manipulated (manipulation type marked on the figure) but presented at the same location as the remaining harmonics. The darker grey lines depict tonal elements that were manipulated and presented from a different location ("+ L" marked by the manipulation type). Note that the $f0$ varied from stimulus to stimulus (200–378 Hz). Therefore, the 10 pure tones of individual stimuli covered different frequency ranges. The frequency range was equalized across conditions. Stimulus timing is compatible with Experiment 2. In Experiment 1, the next stimulus was delivered 300 ms after the listener's response.