Neuropsychologia 57 (2014) 50–58

Contents lists available at ScienceDirect

Neuropsychologia

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

Early development of polyphonic sound encoding and the high voice superiority effect

Céline Marie [a,b], Laurel J. Trainor [a,b,c,*]

[a] Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1
[b] McMaster Institute for Music and the Mind, Hamilton, Ontario, Canada
[c] Rotman Research Institute, Baycrest Centre, Toronto, Ontario, Canada

ARTICLE INFO

Article history:
Received 11 September 2013
Received in revised form
25 February 2014
Accepted 26 February 2014
Available online 12 March 2014

Keywords:
Auditory development
Auditory scene analysis (ASA)
Infant
Mismatch negativity (MMN)
Polyphonic sound
Music

ABSTRACT

Previous research suggests that when two streams of pitched tones are presented simultaneously, adults process each stream in a separate memory trace, as reflected by mismatch negativity (MMN), a component of the event-related potential (ERP). Furthermore, a superior encoding of the higher tone or voice in polyphonic sounds has been found for 7-month-old infants and both musician and non-musician adults in terms of a larger amplitude MMN in response to pitch deviant stimuli in the higher than the lower voice. These results, in conjunction with modeling work, suggest that the high voice superiority effect might originate in characteristics of the peripheral auditory system. If this is the case, the high voice superiority effect should be present in infants younger than 7 months. In the present study we tested 3-month-old infants as there is no evidence at this age of perceptual narrowing or specialization of musical processing according to the pitch or rhythmic structure of music experienced in the infant's environment. We presented two simultaneous streams of tones (high and low) with 50% of trials modified by 1 semitone (up or down), either on the higher or the lower tone, leaving 50% standard trials. Results indicate that like the 7-month-olds, 3-month-old infants process each tone in a separate memory trace and show greater saliency for the higher tone. Although MMN was smaller and later in both voices for the group of sixteen 3-month-olds compared to the group of sixteen 7-month-olds, the size of the difference in MMN for the high compared to low voice was similar across ages. These results support the hypothesis of an innate peripheral origin of the high voice superiority effect.

1. Introduction

From birth, infants are immersed in a noisy world containing simultaneous and overlapping sounds (e.g., multiple talkers, environmental sounds, music). The capacity to process the complex sound wave reaching the ear and determine what auditory objects (sources emitting streams of sounds) are present, and where they are located in the environment, is crucial for following voices and learning linguistic and musical structure. Although several studies have demonstrated infants' ability to perceptually organize sequential, non-overlapping sounds (e.g., Demany, 1982; McAdams & Bertoncini, 1997; Smith & Trainor, 2011; Trainor & Adams, 2000; Winkler, Háden, Ladinig, Sziller, & Honing, 2009), only a few experiments have examined the perception of simultaneous sounds in infancy (Folland, Butler, Smith, & Trainor, 2012; Marie & Trainor, 2013). In the present study we examine whether 3-month-old infants can encode two streams of tones with simultaneous note onsets in auditory cortex and, if so, whether they are like older infants and adults in showing a more robust encoding of the stream with higher pitch.

Auditory scene analysis (Bregman, 1990) requires spectrotemporal analysis of the incoming sound wave in order to integrate components that belong to single objects and separate components that belong to different objects. These processes of integration and separation apply to both simultaneous and sequential aspects of the sound input. Bregman (1990) hypothesized that although learning can affect some aspects of auditory scene analysis, these processes largely occur automatically and without conscious awareness. Given the importance of relatively low level, bottom up processes in auditory scene analysis, it is plausible to predict that the ability to segregate simultaneous sounds should be present early in development.

We investigated whether two simultaneous tones elicit two memory traces in auditory cortex in adults using the mismatch negativity (MMN) component of event-related potential (ERP) electroencephalographic (EEG) and magnetoencephalographic (MEG) recordings (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005;

* Corresponding author at: Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1. Tel.: +1 905 525 9140x23007; fax: +1 905 529 6225.
E-mail address: ljt@mcmaster.ca (L.J. Trainor).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.02.023
0028-3932

Fujioka, Trainor, & Ross, 2008). In general, MMN is elicited when there is an occasional change or deviant stimulus (deviants can reflect a change in pitch, intensity, timbre, or pattern, etc.) in an ongoing sequence of standard stimuli (see Näätänen, Paavilainen, Rinne, & Alho, 2007; Picton et al., 2000 for reviews). MMN is generated in secondary auditory cortex. It appears at the scalp as a frontal negativity, reverses in polarity at posterior sites when using an average reference, and peaks between 120 ms and 250 ms after stimulus onset, depending on stimulus complexity. MMN is thought to reflect the formation of memory traces in auditory cortex. Importantly, it occurs in response to unpredicted sounds, is only present when deviant sounds in an ongoing stream are relatively rare, and increases in amplitude with decreasing probability of a deviant sound. Fujioka et al. (2005) presented binaurally two simultaneous melodies (called voices or streams) on each trial. The two melodies fit together harmonically, and were each composed from the first 5 notes of the Western major scale. Which melody was in the high voice and which was in the low voice varied across conditions of the experiment. On 25% of trials, one note (deviant stimulus) was changed in the higher-pitched melody and on another 25% of trials, one note (deviant stimulus) was changed in the lower-pitched melody, leaving 50% standard trials. Significant MMN was found for both changes, suggesting that separate memory traces are formed for each melody in auditory cortex. This result was replicated with simpler stimuli in which a single high tone (466.2 Hz, B-flat$_4$, international standard notation) repeated in one voice and a single low tone (196.0 Hz, G3) in other, with 25% of repetitions containing a pitch change (two semitones) in the higher tone and 25% in the lower tone (Fujioka et al., 2008). The stimuli were presented binaurally and, again, MMN was found for both deviant stimuli. Furthermore, we tested 7-month-old infants on the same simultaneous tone stimuli (Marie & Trainor, 2013) and found that they also showed MMN responses to pitch changes in both the high and low tones, suggesting that they also form two memory traces for simultaneous tones in auditory cortex.

Interestingly, the MMN amplitude in both Fujioka et al. (2005) and Fujioka et al. (2008) was consistently larger for deviant stimuli in the higher-pitched compared to lower-pitched voice. This *high voice superiority effect* is consistent with the musical practice of most often placing the most important melody line in the highest-pitched voice. It is also consistent with behavioral studies indicating that listeners find it easiest to detect pitch changes in the highest-pitched of several simultaneous streams (Crawley, Acher-Mills, Pastore, & Weil, 2002; Palmer & Holleran, 1994; Zenatti, 1969). In order to test whether the high voice superiority effect results from experience with Western music, we tested adult musicians who played either a soprano range or a bass range instrument (Marie, Fujioka, Herrington, & Trainor, 2012). The logic was that if this effect reflects long-term exposure to Western musical compositional practice, musicians who have played an instrument in the bass register for many years should show a low voice superiority effect. We found that musicians who played a soprano range instrument showed a high voice superiority effect as expected; however, musicians who played a bass range instrument did not show a low voice superiority effect, but rather a non-significant trend for a high voice superiority effect. This suggests that although the high voice superiority effect might be somewhat modifiable by many years of experience, it is difficult, if not impossible, to reverse the high voice superiority effect. Furthermore, the 7-month-old infants in Marie and Trainor (2013) also showed a high voice superiority effect.

We investigated a possible peripheral innate origin to the high voice superiority effect using a computational model of the ear (Ibrahim & Bruce, 2010; Zilany, Bruce, Nelson, & Carney, 2009) whose input is a sound wave and whose output is a representation of neural firing patterns in the auditory nerve (Trainor, Marie, Bruce, & Bidelman, 2014). Inputting the stimuli of Fujioka et al. (2008) and Marie and Trainor (2013), we found that the pitch salience (Bidelman & Heinz, 2011) of the higher tone was greater than that of the lower tone, particularly in mid and lower pitch ranges. These results suggest that the high voice superiority effect might have an innate origin in physiological properties of the cochlea.

Although the data from 7-month-old infants is consistent with an innate origin for the high voice superiority effect, there is evidence of enculturation or *perceptual narrowing* by this age, so it is possible that by 7 months of age, infants have already learned the Western cultural norm of placing the most important melody in the highest voice. Perceptual narrowing refers to the phenomenon whereby very young infants process perceptual features that are relevant as well as those that are not relevant in their environment, whereas older infants become specialized for features that are relevant in their environment (e.g., Lewkowicz & Ghazanfar, 2009; Scott, Pascalis & Nelson, 2007). Specifically, older infants, like adults in their culture, become worse at processing perceptual features that only matter in foreign contexts. For example, at 6 months of age, infants are equally good at processing linguistic consonant contrasts from their native and foreign languages, but by 10 months of age they, like adults, become worse at perceiving consonant contrasts that are not used in their native language (e.g., Japanese adults and older infants have difficulty distinguishing /l/ and /r/ whereas younger Japanese infants do not; Werker, 1989). Similarly, 4-month-olds are able to discriminate monkey faces as well as human faces, and foreign-race faces as well as own-race faces, but between 6 and 8 months infants are becoming specialized for discriminating faces from their own species and own race (e.g., Slater et al., 2010). As well, the ability to discriminate voices from a foreign species (rhesus monkey) decreases between 6 and 12 months of age (Friendly, Rendall, & Trainor, 2013). In the musical domain, more studies are needed, but the evidence suggests that infants become specialized for processing the tonal and rhythmic structures in the music of their culture after 6 months of age (Gerry et al., 2012; Hannon & Trehub, 2005; Hannon & Trainor, 2007; Lynch, Eilers, Oller, & Urbano, 1990; Lynch & Eilers, 1992; Trainor & Trehub, 1992; Trainor, 2005; Trainor & Corrigall, 2010; Trainor & He, 2013; Trainor, Marie, Gerry, Whiskin, and Unrau, 2012; Trainor & Unrau, 2012; Trainor & Hannon, 2013). Therefore, in the present study, we chose to test whether infants younger than 6 months show a high voice superiority effect, as there is little evidence of perceptual narrowing in any domain before this age.

By 3 months of age, adult-like MMN responses to pitch changes can be measured. The ERP responses at 2 months of age to occasional changes (deviant stimuli) in pitch in a repeating sequence of sounds are dominated by an increase in the amplitude of a frontally positive slow wave (e.g., Friederici, Friedrich, & Weber 2002; He, Hotson, & Trainor, 2007, 2009a,b; Leppänen, Guttorm, Pihko, Takkinen, & Lyytinen, 2004; Morr, Shafer, Kreuzer, & Kurtzberg, 2002), a response that decreases with increasing age. However, by 3 months of age an MMN with adult-like morphology is clearly evident, which decreases in latency and increases in amplitude with increasing age (e.g., Choudhury & Benasich, 2011; He et al., 2007; Kushnerenko et al., 2002; Trainor, 2012).

Related to the development of auditory scene analysis for simultaneous tones, the integration of harmonics into a single pitch percept appears not to have a cortical representation until after 3 months of age as evidenced by ERP studies on the pitch of the missing fundamental (He & Trainor, 2009), and the ability to use harmonicity as a cue for auditory object separation also appears to emerge around this age as evidenced by ERP studies of object-related ERP responses to mistuned harmonics