Neural representation of a melodic motif: Effects of polyphonic contexts

Madeline Huberth [a,*], Takako Fujioka [a,b]

[a] Center for Computer Research in Music and Acoustics, Department of Music, Stanford University, United States
[b] Stanford Neurosciences Institute, Stanford University, United States

ARTICLE INFO

ABSTRACT

In music, a melodic motif is often played repeatedly in different pitch ranges and at different times. Event-related potential (ERP) studies have shown that the mismatch negativity (MMN) reflects memory trace processing that encodes two separate melodic lines ("voices") with different motifs. Here we investigated whether a single motif presented in two voices is encoded as a single entity or two separate entities, and whether motifs overlapping in time impede or enhance encoding strength. Electroencephalogram (EEG) from 11 musically-trained participants was recorded while they passively listened to sequences of 5-note motifs where the 5th note either descended (standard) or ascended (deviant) relative to the previous note (20% deviant rate). Motifs were presented either in one pitch range, or alternated between two pitch ranges, creating an "upper" and a "lower" voice. Further, motifs were either temporally isolated (silence in between), or temporally concurrent with two tones overlapping. When motifs were temporally isolated, MMN amplitude in the one-pitch-range condition was similar to that in the two-pitch-range upper voice. In contrast, no MMN, but P3a, was observed in the two-pitch-range lower voice. When motifs were temporally concurrent and presented in two pitch ranges, MMN exhibited a more posterior distribution in the upper voice, but again, was absent in the lower voice. These results suggest that motifs presented in two separate voices are not encoded entirely independently, but hierarchically, causing asymmetry between the upper and lower voice encoding even when no simultaneous pitches are presented.

1. Introduction

Listeners of music must parse multiple simultaneous sounds and connect them to form perceptual objects such as melody, harmony, and rhythm. In contrast to music that features a single dominant melody against a background texture, some music is structured with two or more lines of independent melody that occur concurrently. The resulting musical texture, called *polyphony*, requires the listener to simultaneously follow each melodic line, or *voice*, as it unfolds in time while integrating simultaneous notes as harmony. In some polyphonic musical forms, such as the *fugue*, a single melodic motif (i.e., a short succession of notes that has its own identity based mainly on melodic contour), is repeatedly played in the different voices in different pitch ranges (Walker, 2001). While recognizing each motif may seem an easy task, it is often the case that the next motif starts before the previous motif concludes. In this situation, listeners have to not only track the two motifs at the same time, but also each with a distinct time stamp in reference to the entry point. This type of music can be played even in a single instrument, as evidenced in the rich repertoire of canons composed for keyboard instruments, for which voices are differentiated by pitch, but not much by timbre. How our brain works for such a complex challenge entails important questions about our auditory perceptual and cognitive functions. The automatic process of separating auditory streams in the acoustic environment and integrating them into perceptual objects is referred to as auditory scene analysis (Bregman, 1990). Generally, sounds that have similar acoustic features are perceived as one object, coming from one source, whereas sounds with acoustically different features are perceived as segregated objects, coming from different sources. Polyphonic music gives us an interesting opportunity to study interactions between stream segregation and memory encoding where both functions are presumably active and possibly interacting with each other when processing multiple concurrent voices.

Neural correlates of auditory stream segregation have been examined through event-related potentials (ERP), and in particular, by using the mismatch negativity (MMN) response. The MMN is typically observed as a negative voltage deflection in the difference wave obtained by subtracting the ERPs to frequent "standard" stimuli from that of infrequent "deviant" stimuli (Näätänen,

* Corresponding author at: 660 Lomita Court, Stanford, CA 94305, United States.
E-mail address: mhuberth@ccrma.stanford.edu (M. Huberth).

1990, 1992; Näätänen, Gaillard, & Mäntysalo, 1978; see Näätänen, Paavilainen, Rinne, & Alho, 2007 for a review). Often peaking between 100 and 200 ms with its largest amplitude at the fronto-central electrodes, MMN has been elicited in response to infrequent changes in frequency, intensity, timbre, spatial location, and duration. Generally, the greater the perceptual difference between a deviant and standard sound, the larger and earlier the MMN (Näätänen et al., 2007). The process is considered to be largely automatic, as the MMN can be elicited in passive listening without requiring listeners' attention to sounds. One view of the MMN's generation is that it involves a memory-reliant comparison mechanism in which a sound violates a previously detected auditory regularity; MMN generation relies on an auditory memory trace that stores a template of standard sounds, against which an irregularity in incoming sounds is detected (Näätänen, Jacobsen, & Winkler, 2005). The MMN response has also been recently associated with a modification of the preattentive model of the acoustic environment (Winkler, Karmos, & Näätänen, 1996), where anticipatory predictions of future events are being made and updated constantly based on recent past events (Friston, 2005; Garrido, Kilner, Stephan, & Friston, 2009; Garrido et al., 2008; Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009; Winkler, Denham, & Nelken, 2009).

MMN studies have also shown that the auditory memory trace can encode more complex and abstract sound patterns in addition to basic acoustical parameters. For example, changes in musical interval direction between standard and deviant tone pairs irrespective of the interval size elicit an MMN (Paavilainen, Jaramillo, & Näätänen, 1998; Saarinen, Paavilainen, Schröger, Tervaniemi, & Näätänen, 1992). These studies, and others (Sussman, Gomes, Nousak, Ritter, & Vaughan, 1998; Sussman, Ritter, & Vaughan, 1998; Tervaniemi, Saarinen, Paavilainen, Danilova, & Näätänen, 1994), exemplify that sensory memory traces, as reflected by MMN, integrate information about closely temporally spaced stimuli into a unitary sensory event and reflect auditory grouping processes.

The MMN generation mechanism also appears to track different features of stimuli in parallel, separate memory traces. Näätänen, Pakarinen, Rinne, and Takegata (2004) tested five types of deviants that each occurred 10% of the time and varied in their stimulus feature, thus making the frequency of standard stimuli to be 50% but maintaining the frequency of non-deviant stimuli for each feature to be 90%. The results demonstrated an MMN response to each feature deviant, suggesting that acoustic features are processed in parallel (Huotilainen et al., 1993; Vuust et al., 2011). Interestingly, however, not all combinations of stimulus dimensions may be processed entirely independently. For example, the amplitude of MMN to a multiple-feature deviant sometimes approximates the summation of the responses to each single-feature deviation measured separately (Takegata, Paavilainen, Näätänen, & Winkler, 1999), but not always linearly, showing reduced encoding strength (Paavilainen, Valppu, & Näätänen, 2001; Wolff & Schröger, 2001).

In multiple potential auditory streams, each of which contain different standard and deviant stimuli, separate memory traces may form for each stream (Nager, Teder-Sälejärvi, Kunze, & Münte, 2003; Shinozaki et al., 2000). Similarly to multi-feature encoding, there is some evidence for the cost of encoding separate objects simultaneously. For example, in Shinozaki et al. (2000), the MMN peak to a deviant in one stream was delayed when the other stream was additionally present. In Nager et al. (2003), while MMN amplitude between the one- and two-stream conditions showed no significant difference, the addition of a third spatial stream attenuated MMN. Thus the MMN provides a useful tool for studying neural mechanisms underlying complex, layered sequential auditory perception, such as streaming and even music, since the

oddball sequence can be made naturally resembling dense and intricate musical textures.

In music, a melody can be considered a single auditory stream, requiring sequential integration of pitch and duration information. With regards to melody encoding, the MMN has been shown to reflect detection of contour (general direction of pitch change) and interval (precise pitch distance) changes in melodic patterns, as found in electroencephalography (EEG) and magnetoencephalography (MEG) studies (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Tervaniemi, Rytkönen, Schröger, Ilmoniemi, & Näätänen, 2001; Trainor, McDonald, & Alain, 2002). While both nonmusicians (Trainor et al., 2002) and musicians (Tervaniemi et al., 2001) detect contour changes as reflected in MMN, MMN is larger in musicians than nonmusicians in response to contour changes (Fujioka et al., 2004), suggesting that musical training enhances the ability to automatically register abstract changes in the relative pitch structure of melodies.

In addition to the MMN, the P3a response is often elicited in response to deviants in a train of frequent standards, even if the listeners' attention is not directed to the sound. The P3a response is thought to reflect an early attention process resulting from a representational change in working memory operating mainly in the frontal lobe (Polich, 2007), and has been found to exhibit multiple subcomponents. When found with a frontocentral distribution and a ~230 ms latency, it is thought to represent inadvertent capture of attention, while a later subcomponent, peaking at ~315 ms and right-frontally dominant, reflects the actual orienting of attention (Escera, Alho, Winkler, & Näätänen, 1998). Although P3a amplitude may not relate directly to MMN amplitude (Horváth, Winkler, & Bendixen, 2008), it is also hypothesized that the MMN not only reflects processing at a preattentive level, but that the process underlying MMN also triggers the switching of attention to potentially important events in the unattended auditory environment (Giard, Perrin, Pernier, & Bouchet, 1990; Näätänen et al., 1978) as if the process related to P3a can take the output of the MMN-related process as its input (Escera et al., 1998). Although task manipulations can affect the magnitude of the P3a (Comerchero & Polich, 1999; Katayama & Polich, 1999), a stimulus can elicit a P3a even if it is ignored (Schwent, Hillyard, & Galambos, 1976). Particularly in response to music, the P3a response appears to be quite robust, occurring in response to harmonic and melodic deviant stimuli in both active (Janata, 1995; Trainor, Desjardins, & Rockel, 1999) and passive paradigms (Seppänen, Pesonen, & Tervaniemi, 2012). While some studies report that its amplitude in response to musical stimuli grows larger with musical training (Putkinen, Tervaniemi, Saarikivi, Ojala, & Huotilainen, 2014), others find no such training effects (Trainor et al., 1999).

MMN has been used in our previous studies to examine how simultaneous melodies are encoded (Fujioka, Trainor, & Ross, 2008; Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005). These studies were the first to use a series of two tones that occurred together as a dyad but continued to form two different streams, unlike other preceding studies described above that used a single tone alternating in frequency or spatial location to imply the multiple streams (Nager et al., 2003; Shinozaki et al., 2000; Sussman, Ritter, & Vaughan, 1999). Using MEG, Fujioka et al. (2005) examined melodic encoding in both musicians and nonmusicians by presenting two simultaneous five-tone melodies (i.e., five two-note intervals), resulting in an 'upper' and a 'lower' voice, differentiated by pitch. The deviants were introduced on the fifth tone on 50% of trials such that 25% of deviants were presented in the upper voice and 25% were presented in the lower. An MMN was elicited by deviants in each voice, indicating that each of the two melodies was represented in its own memory trace in the auditory cortex. In addition to the larger MMN found in musicians, a larger and earlier MMN was found for the upper voice deviants compared to the lower