Research Article

Detecting the temporal structure of sound sequences in newborn infants ☆,☆☆

Gábor P. Háden [a,*], Henkjan Honing [b,c], Miklós Török [d], István Winkler [a,e]

[a] Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar Tudósok körútja 2, H-1117 Budapest, Hungary
[b] Institute for Logic, Language and Computation, University of Amsterdam, Science Park 107, 1098 XH Amsterdam, The Netherlands
[c] Amsterdam Brain and Cognition, Nieuwe Achtergracht 129, 1018 WS Amsterdam, The Netherlands
[d] Department of Obstetrics–Gynaecology and Perinatal Intensive Care Unit, Military Hospital, Podmaniczky u. 111, H-1062 Budapest, Hungary
[e] Institute of Psychology, University of Szeged, Egyetem u. 2, H-6722 Szeged, Hungary

ARTICLE INFO

ABSTRACT

Most high-level auditory functions require one to detect the onset and offset of sound sequences as well as registering the rate at which sounds are presented within the sound trains. By recording event-related brain potentials to onsets and offsets of tone trains as well as to changes in the presentation rate, we tested whether these fundamental auditory capabilities are functional at birth. Each of these events elicited significant event-related potential components in sleeping healthy neonates. The data thus demonstrate that the newborn brain is sensitive to these acoustic features suggesting that infants are geared towards the temporal aspects of segregating sound sources, speech and music perception already at birth.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Extracting temporal regularities from sound sequences and detecting their violations are fundamental capabilities for correctly perceiving objects in the acoustic environment (Griffiths and Warren, 2004; Winkler et al., 2009a), including interpreting speech and music (Honing, 2013; Patel, 2008) and also form the basis of synchronized communication with others (Jaffe et al., 2001; Jungers et al., 2002). Speech dynamics provide information about the emotional state and intents of the speaker, and structure information within and between sentences and allow marking agreement and turn taking (O'Connell and Kowal, 2008). In music, temporal information defines beat, metrical structure, and tempo, allows the extraction of expressive timing in a performance, helps coordination between players, and conveys

emotions (Honing, 2013). Because of its fundamental nature, one may assume that this capability appears early during infancy. The current study tested in newborn infants the detection of the three most basic temporal features of sound sequences: onset, presentation rate change, and offset.

Behavioral studies testing the processing of temporal features of sound sequences found that 2-month old infants detect 15% tempo accelerations in isochronous sequences at the base rate of 600 ms inter-onset interval (IOI), but not at faster or slower IOIs (Baruch and Drake, 1997). They also coordinate movements with the tempo of external sounds (Bobin-Bègue et al., 2006). By 6 months, infants form long-term memories of tempo (Trainor et al., 2004) and by 9 months they can distinguish between happy and sad music (Flom et al., 2008). The sensitivity to temporal features of the stimulation was also proposed to be an important predictor of later performance in tests of verbal development in young infants (Benasich and Tallal, 2002; Chonchaiya et al., 2013) though the mechanisms underlying these effects are unclear (Protopapas, 2014). Much less is known about auditory temporal processing in newborns. Previous studies showed that neonates segregate interleaved tonal sequences by pitch (Winkler et al., 2003), prefer infant-directed to adult-directed speech and singing (Cooper and Aslin, 1990; Masataka, 1999), and discriminate languages based on rhythmic class (Nazzi et al., 1998; Ramus et al., 2000; Nazzi and Ramus, 2003). These capabilities probably involve detecting auditory temporal cues.

In neonates, sound processing can be tested by measuring event-related potentials (ERP) elicited by acoustic events. Two parallel event detector systems have been described in adults (Näätänen, 1990; Näätänen et al., 2011): one sensitive to sudden changes in sound energy and another triggered by violations of some regular feature of a sound sequence. The former is based on adaptation/refractoriness of afferent neurons (in adults, the N1 ERP response; Näätänen and Picton, 1987; May and Tiitinen, 2010), the latter probably on prediction errors in the brain (the MMN component; Näätänen and Picton, 1987; Garrido et al., 2009; Winkler, 2007). Although no true equivalent of either of these adult ERP responses has been obtained in newborn infants, neonatal ERP responses to large energy changes and violations of simple acoustic regularities have been described (Alho et al., 1990; for a review, see Kushnerenko et al., 2013). Previous ERP studies testing temporal features of sound sequences in young infants showed that occasional shortenings of a regular 300 ms long pre-stimulus interval to 100 ms are detected at 2 months of age (Otte et al., 2013; for similar results in 10-month olds, see Brannon et al., 2004, 2008) and that newborns can distinguish between the downbeat and other positions within a rhythmic sequence (Winkler et al., 2009a,b).

Thus, whereas we know that young infants use temporal cues while making complex linguistic and musical discriminations, the developmental origins of the underlying processing capabilities have not yet been established. Here we assess the developmental origins of processing the temporal structure of sound sequences by testing whether the neonate brain is sensitive to the onset and offset of sound trains that roughly estimate the structure of sentences or short musical phrases as well as to presentation rate change within them. To this end we compare responses elicited by significant events (onset and rate change) in our sound sequences to events that are physically the same but appear in a different context. We assume that the onset of the train elicits a response; however if the rate change elicits a response it signals that the change was indeed detected. Finally if we see a response at the offset of the train in a position where the continuation of the train could be expected we can assume that the offset itself was detected.

2. Methods

ERPs were recorded from 30 (16 male) healthy, full-term newborn infants during day 1–3 postpartum. The mean gestational age was 39.7 weeks (SD = 1.00), birth weight was 3450 g (SD = 372.46), and the average Apgar score was 9/9.8 (SD = 0.52/0.48). An additional 7 (2 male) participants were recorded, but discarded due to excessive electrical artifacts (<100 artefact free epochs per condition). Informed consent was obtained from one or both parents. The mother of the infant could opt to be present during the recording. The study was conducted in full accordance with the World Medical Association Declaration of Helsinki and all applicable national laws and it was approved by the Medical Research Council—Committee of Scientific and Research Ethics (ETT-TUKEB), Hungary. The experiment was carried out in a dedicated experimental room at the Department of Obstetrics–Gynaecology and Perinatal Intensive Care Unit, Military Hospital, Budapest, Hungary.

Trains of complex tones (Fig. 1) uniform within but varying in pitch (F0) across trains (8 different pitches taken from the C major scale: C3, D3, E3, F3, G3, A3, B3, and C4, viz. 130.80, 147.15, 163.50, 173.96, 196.20, 217.13, 245.25, and 261.60 Hz, respectively) were delivered to the infants at ~65dB$_{SPL}$. Each tone consisted of the F0 and its first five harmonics, the spectral power of the higher harmonics being set at 1:2, 1:3, 1:4, 1:5, and 1:6, with respect to that of the F0 component. Tone duration was 50 ms, including 5 ms rise and 5 ms fall times (raised cosine ramps). 170 trains were presented in two stimulus blocks. For each train, a pitch was selected randomly (with equal probability; no pitch repetition allowed). Trains consisted of 8–24 (randomly selected, equal probability) tone repetitions and a silent interval. The N element trains were split into two parts: the first part consisting of N/2 (rounded

down) + 0/1 (random) elements, the second part consisting of N-n$_{(first\ part)}$ elements. Tones in the first part of the train were presented at the "slow" rate (average IOI = 200 ms, 150 ms offset to onset), and in the second part at the "fast" rate (average IOI = 100 ms; 50 ms offset to onset), followed by a silent interval (average IOI 1050 ms, 1000 ms offset to onset). All time intervals were taken from normal distributions with a standard deviation of 5%. The amount of jitter is below the adult JND for tempo discrimination (Quené, 2007; Grondin et al., 2011) and was chosen to dampen steady state responses arising from the fast stimulation. The length of individual trains was about 1.2–3.6 s. Sounds were presented binaurally using the E-Prime stimulus presentation software (Psychology Software Tools, Inc., Pittsburgh, PA, USA) via ER-1 headphones (Etymotic Research Inc., Elk Grove Village, IL, USA) connected via sound tubes to self-adhesive ear-couplers (Natus Medical Inc., San Carlos, CA, USA) placed over the infants' ears.

EEG was recorded with Ag/AgCl electrodes attached to the F3, Fz, F4, C3, Cz, and C4 locations (international 10–20 system) using a direct-coupled amplifier (V-Amp, Brain Products GmbH., Munich, Germany) at 24-bit resolution and a sampling rate of 250 Hz or 1000 Hz (1000 Hz sampling rate was used for 16 participants; due to experimenter error, these recordings were off-line down-sampled to 250 Hz). The reference and ground electrodes were attached to the nose and the forehead respectively. Signals were off-line filtered between 1 and 30 Hz and epochs from −100 to 500 ms with respect to the event onset (tone or expected tone, see next paragraph) were extracted for each sound. The 100 ms pre-stimulus interval served as the baseline for amplitude measurements and illustrations. Epochs with an absolute voltage change outside the 0.1–100 μV range throughout the epoch were rejected from the analyses as artefacts. Data from infants with less than 100 artefact-free epochs per condition were dropped from the analyses. The mean number of epochs and per condition is given in Table 1.

Responses were measured at the "train onset" (the first element of the train), at "presentation rate change" (the first element after a short, 100 ms IOI interval in the train), and at "expected tone" (100 ms after train offset; see Fig. 1 b.). Responses were also measured for "slow control" tones, tones between train onset and the rate change but separated from both by at least 2 elements; and "fast control" tones, tones between rate change and train offset but separated from both by at least 2 elements. (e.g. in a 20 element train where the rate change occurs at the 11th element, elements 4–8 are slow controls whereas elements 14–17 are fast controls). Responses to train onset were compared to slow control responses whereas rate change and train offset responses were compared to fast control responses. Average response amplitudes were measured from separate time windows for each type of the three events. The latencies of the two highest-amplitude (early and late) difference peaks were determined from the mean group response averaged over all six electrode locations. The corresponding window was defined as the continuous segment of data points on both sides of the respective peak within which the difference amplitude exceeded 30% of the corresponding peak amplitude (see Table 1. and Fig. 2. for the latency ranges). This method allows comparison between responses to event types and their respective controls when the latencies of peaks vary or no clear peaks are visible.

Effects were tested with separate dependent-measures analyses of variance (ANOVA) of the structure Stimulus type [Event vs. Control] × Frontality [F vs. C electrode line] × Laterality [left vs. midline vs. right] for the three timing events (Train Onset, Presentation Rate Change, Expected tone) and the two (early and late) measurement windows. Greenhouse–Geisser correction ε factors (where appropriate) and the partial η^2 effect sizes are given in Table 1.

3. Results

ERP responses for train onsets, presentation rate changes, and expected tone (train offsets) are shown together with the corresponding