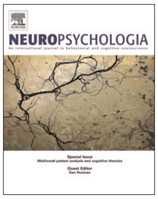
Neuropsychologia 68 (2015) 38–50

Contents lists available at ScienceDirect

Neuropsychologia

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



Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept

Gavin M. Bidelman [a,b,*], Claude Alain [c,d,e]

[a] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[b] School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Avenue, Memphis, TN 38105, USA
[c] Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, ON, Canada
[d] Department of Psychology, University of Toronto, Toronto, ON, Canada
[e] Institute of Medical Sciences, University of Toronto, ON, Canada

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form
18 December 2014
Accepted 22 December 2014
Available online 24 December 2014

Keywords:
Auditory scene analysis
Streaming
Sensory systems
Event-related potentials (ERPs)
Frequency-following response (FFR)
Object-related negativity (ORN)
Sound classification
Reaction time

ABSTRACT

Natural soundscapes often contain multiple sound sources at any given time. Numerous studies have reported that in human observers, the perception and identification of concurrent sounds is paralleled by specific changes in cortical event-related potentials (ERPs). Although these studies provide a window into the cerebral mechanisms governing sound segregation, little is known about the subcortical neural architecture and hierarchy of neurocomputations that lead to this robust perceptual process. Using computational modeling, scalp-recorded brainstem/cortical ERPs, and human psychophysics, we demonstrate that a primary cue for sound segregation, i.e., harmonicity, is encoded at the auditory nerve level within tens of milliseconds after the onset of sound and is maintained, largely untransformed, in phase-locked activity of the rostral brainstem. As then indexed by auditory cortical responses, (in)harmonicity is coded in the signature and magnitude of the cortical object-related negativity (ORN) response (150–200 ms). The salience of the resulting percept is then captured in a discrete, categorical-like coding scheme by a late negativity response (N5; ~500 ms latency), just prior to the elicitation of a behavioral judgment. Subcortical activity correlated with cortical evoked responses such that weaker phase-locked brainstem responses (lower neural harmonicity) generated larger ORN amplitude, reflecting the cortical registration of multiple sound objects. Studying multiple brain indices simultaneously helps illuminate the mechanisms and time-course of neural processing underlying concurrent sound segregation and may lead to further development and refinement of physiologically driven models of auditory scene analysis.

1. Introduction

In social gatherings, human listeners must perceptually integrate the sounds originating from one person's voice, e.g., fundamental frequency (f_0) and harmonics, and segregate those of other talkers. The segregation of concurrent speech sounds is promoted by differences in a number of acoustic cues including harmonicity, spatial location, and onset asynchrony (Alain, 2007; Carlyon, 2004; Ciocca, 2008). In the laboratory, the perception of concurrent sound objects can be parametrically induced by mistuning one spectral component (i.e., harmonic) from an otherwise periodic harmonic complex tone. Sounds that are harmonically related (i.e., integer multiple of a fundamental) evoke unitary pitch percepts and are thus heard as belonging to a common source, whereas inharmonicity tends to cue the presence of multiple auditory objects. Low order harmonics that are mistuned by about 4–6% of their original value are “heard out” from the harmonic complex so that listeners perceive two sounds: a complex tone and another sound with a pure-tone quality (e.g., Alain et al., 2001; Darwin et al., 1994; Moore et al., 1986).

Evidence from animal neurophysiology and human scalp-recordings of event-related brain potentials (ERPs) suggest that concurrent sound segregation involves low-level processes that take place along the ascending auditory pathway including the auditory nerve (Sinex et al., 2003), cochlear nucleus (Sinex, 2008), the inferior colliculus (Sinex et al., 2005, 2002), and the primary and associative auditory cortices (Dyson and Alain, 2004; Fishman et al., 2014; Hautus and Johnson, 2005). The detection of a mistuned harmonic in an otherwise periodic signal may involve neurons sensitive to equal spacing between tonal elements (Roberts, 1998; Roberts and Brunstrom, 1998) which, in turn, could be

---

[*] Correspondence to: School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Avenue, Memphis, TN 38105, USA.
Fax: +1 901 525 1282.
E-mail address: g.bidelman@memphis.edu (G.M. Bidelman).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.12.020
0028-3932/

used to build a harmonic sieve or template (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009; Goldstein, 1973; Hartmann et al., 1990; Lin and Hartmann, 1998). That is, neurons sensitive to frequency periodicity could act as a series of filters that allow harmonically related partials (i.e., frequencies sharing a common f_0) to group together with the f_0 and mistuned partials to form separate neural representations and thus, the perception of multiple auditory objects (Darwin et al., 1994, 1995).

Furthermore, evidence from ERP studies suggests that the perception of concurrent sound objects may occur independently of attention (Alain and Izenberg, 2003; Alain et al., 2002; Dyson et al., 2005) and likely involves primary and associative auditory cortices (Arnott et al., 2011; Dyson and Alain, 2004). While the amplitude of the cortical ERPs can differ between passive and active listening conditions (Alain et al., 2001), studies that have directly examined the effects of selective attention have not revealed an effect of attention load on the object-related negativity (ORN) amplitude or latency, a component of the ERPs reflecting the early cortical processing of harmonic mistuning (Alain and Izenberg, 2003). Thus, while arousal may change the ORN, the response largely reflects pre-attentive processing of concurrent sound objects in the early auditory cortices.

In addition to the ORN, a second modulation is often observed (referred to as P400) and emerges when a listener is asked to indicate behaviorally whether they hear one or two concurrent sounds. Accordingly, the later P400 is likely related to attention-dependent cognitive processing of a second auditory object, while the ORN would likely reflect automatic, attention-independent processing of the mistuned harmonic (Alain, 2007; Alain et al., 2001; Hautus and Johnson, 2005). Together, the ORN and P400 provide stimulus- and perception/response-related cortical indices of listeners' ability to segregate concurrent sounds based on periodicity cues, which among other auditory functions, is important for understanding speech in noisy situations (e.g., Binns and Culling, 2007; Swaminathan and Heinz, 2012).

While the relationship between neural responses from the human auditory cortex and concurrent sound perception has been examined, the possible contribution of subcortical brainstem structures has so far remained unexplored in humans. Furthermore, the relationship between brainstem, cortical evoked responses, and perception remains to be determined. Here, we measured brainstem and cortical ERPs within the same participants elicited by harmonic stimuli that do and do not promote concurrent sound segregation. This approach has been helpful to tease apart neural events across the auditory system that index the encoding of physical stimulus characteristics from those related to perception (Bidelman et al., 2013). This approach also allows us to evaluate both primitive (i.e., pre-attentive or automatic) and more complex (i.e., abstract, endogenous) neural codes underlying concurrent sound segregation. Participants were presented with harmonic complex tones where all tonal elements were either in tune (i.e., integer multiple of the fundamental) or included a mistuned component. Human ERPs were also complemented by computational modeling, which allowed us to investigate whether or not neural correlates of concurrent sound perception are present as early as the auditory nerve (AN). Comparisons between AN, brainstem, and cortical responses allowed us to trace the brain correlates and hierarchical signal transformations underlying auditory scene analysis from the periphery through perception and in turn, assess whether these various neural responses correlate with listeners' concurrent sound perception. We hypothesized that inharmonicity would be well-represented in AN and brainstem activity, and that cortical evoked responses would better correlate with listeners' perception. These findings would be consistent with our previous reports, demonstrating that cortical activity reflects correlates of (speech) perception whereas brainstem responses largely code stimulus-related information (e.g., Bidelman et al., 2013).

2. Methods and materials

2.1. Participants

Ten English-speaking young adults (6 females; age: 25.8 ± 3.0 years) participated in the experiment. All were right-handed (Oldfield, 1971) and reported a collegiate level of education (18.7 ± 2.7 years). Audiometric screening confirmed normal hearing sensitivity (i.e., ≤ 25 dB HL) at octave frequencies between 250 and 8000 Hz. Participants reported no previous history of neurological or psychiatric illnesses. None had more than three years of formal musical training [for effects of musical training on concurrent sound segregation, see Zendel and Alain (2009, 2013)]. All participants gave written, informed consent in compliance with an experimental protocol approved by the Baycrest Centre Research Ethics Committee and were paid for their time.

2.2. Stimuli

Harmonic complex stimuli were similar to those used in our previous reports (e.g., Alain et al., 2001, 2012; Zendel and Alain, 2009). They were generated by combining 12 pure-tones at integer multiples of a $f_0 = 200$ Hz fundamental frequency added in sine phase. Each tone was 100 ms in duration including 10 ms of rise-fall ramping to reduce acoustic spectral splatter and minimize onset responses. Mistuning was achieved by selectively shifting the second harmonic of each complex upward by 1%, 2%, 4%, 8%, or 16% of its original value of 400 Hz (Fig. 1). An additional perfectly harmonic complex (0% mistuning) was included as a control stimulus. Stimulus presentation was controlled by MATLAB® 2011 (The MathWorks) routed to a GSI audiometer via a TDT RP2 interface (Tucker-Davis Technologies) and delivered binaurally at an intensity of 83 dB SPL through insert earphones (ER-3 A, Etymotic Research). Stimulus intensity was calibrated using a Larson-Davis sound pressure level (SPL) meter (Model 824) measured with a 2 cc, artificial ear coupler (AEC100l). Left and right ear channels were calibrated separately.

2.3. Electrophysiological recording and data preprocessing

Data acquisition and response evaluation were similar to previous reports from our laboratory (e.g., Bidelman et al., 2013, 2014a, 2014b). During ERP recording, listeners heard 200 randomly ordered exemplars of each token and were required on each trial, to indicate whether they heard “1-” or “2-sounds” (binary response) as quickly as possible. Participants were told that the stimuli would either sound like a single pitch (“1-sound”) or a pitch with an additional faint sounding pure-tone (“2-sounds”) (Alain et al., 2001). The next stimulus was then initiated between 400 and 600 ms after the participant's behavioral response (jittered in 20-ms steps, rectangular distribution). Thus, the interstimulus interval (ISI) for cortical ERP presentations was, on average, between ~900 and 1100 ms. An additional 2000 trials (ISI = 150 ms) were then collected in order to detect sub-microvolt brainstem ERPs (Bidelman et al., 2013). Given that early brainstem ERPs are largely unaffected by attention (Hillyard and Picton, 1979; Picton and Hillyard, 1974; Picton et al., 1971; Woods and Hillyard, 1978), participants watched a self-selected movie with subtitles during blocks of brainstem recording to facilitate a calm yet wakeful state. Electrophysiological recording lasted ~2 hrs.

The continuous electroencephalogram (EEG) was recorded between electrodes placed on the high forehead at the hairline