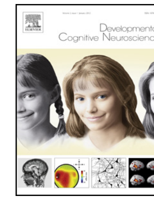
Developmental Cognitive Neuroscience 19 (2016) 223–232

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: http://www.elsevier.com/locate/dcn



Auditory evoked potentials to speech and nonspeech stimuli are associated with verbal skills in preschoolers

Soila Kuuluvainen[a,*], Alina Leminen[a,b], Teija Kujala[a]

[a] Cognitive Brain Research Unit, Institute of Behavioral Sciences, University of Helsinki, Finland
[b] Center of Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, Denmark

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form
23 December 2015
Accepted 1 April 2016
Available online 14 April 2016

Keywords:
Auditory
Event-related potential
Speech
Nonspeech
Verbal skills
Children

ABSTRACT

Children's obligatory auditory event-related potentials (ERPs) to speech and nonspeech sounds have been shown to associate with reading performance in children at risk or with dyslexia and their controls. However, very little is known of the cognitive processes these responses reflect. To investigate this question, we recorded ERPs to semisynthetic syllables and their acoustically matched nonspeech counterparts in 63 typically developed preschoolers, and assessed their verbal skills with an extensive set of neurocognitive tests. P1 and N2 amplitudes were larger for nonspeech than speech stimuli, whereas the opposite was true for N4. Furthermore, left-lateralized P1s were associated with better phonological and prereading skills, and larger P1s to nonspeech than speech stimuli with poorer verbal reasoning performance. Moreover, left-lateralized N2s, and equal-sized N4s to both speech and nonspeech stimuli were associated with slower naming. In contrast, children with equal-sized N2 amplitudes at left and right scalp locations, and larger N4s for speech than nonspeech stimuli, performed fastest. We discuss the possibility that children’s ERPs reflect not only neural encoding of sounds, but also sound quality processing, memory-trace construction, and lexical access. The results also corroborate previous findings that speech and nonspeech sounds are processed by at least partially distinct neural substrates.

1. Introduction

Even before the development of vocabulary, newborns and infants are biased towards listening to speech over equally complex nonspeech sounds (Vouloumanos and Werker, 2004, 2007). This bias lays the foundation to the development of later language skills via the increasing specialization of the cortex in processing speech (for reviews, see Kuhl, 2004; Kuhl et al., 2008). By preschool age, typically developing children distinguish speech and nonspeech sounds effortlessly, and master basic language skills necessary for learning in a formal school setting. However, little is known of the underlying processes of speech versus nonspeech sound encoding in preschoolers, as no comprehensive studies have been conducted in this age group. This study aims to investigate speech and nonspeech sound processing using cortical auditory event-related potentials (ERPs) and their association with neurocognitive task performance.

In children under 11 years of age, sounds typically elicit a pattern of so-called 'obligatory' ERPs labeled P1-N2-N4 or P100-N250-N450 according to their polarity (positive or negative) and latency (100, 250 or 450 ms; e.g. Pihko et al., 2005; Ponton et al., 2000; Shafer et al., 2015). They are identifiable already in neonates to harmonic tones presented at a slow rate, with P1 increasing in amplitude during the first three months, and N2 becoming increasingly robust between six and nine months of age (Kushnerenko et al., 2002). For syllables, P1 is identifiable already at the youngest age group of three-month-olds, whereas N2 emerges at around six months of age, both stabilizing in amplitude and latency by the age of two years (Shafer et al., 2015). P1 amplitude to syllables increases again at the age of five, remaining stable after that until the age of eight years, whereas N2 amplitude to syllables shows no clear developmental tendencies between ages two and eight years (Shafer et al., 2015).

In contrast to syllables, P1 amplitude to harmonic tones is of similar magnitude at ages four and nine years, decreasing by adulthood (Čeponienė et al., 2002) and it decreases steadily for pure tones from age seven to adulthood (Bishop et al., 2011; Sussman et al., 2008) N2 amplitude to harmonic tones decreases between ages four and nine (Čeponienė et al., 2002) and is stable for pure

* Corresponding author.
E-mail address: soila.kuuluvainen@helsinki.fi (S. Kuuluvainen).

http://dx.doi.org/10.1016/j.dcn.2016.04.001

tones between ages eight and eleven years (Sussman et al., 2008). No results were reported for N4 in these studies. Taken together, the results suggest that speech and nonspeech sound processing have different developmental trajectories, with turning points at around ages two and five years. However, for a more complete picture, the processing of speech and nonspeech sounds would have to be compared within the same participants. This has been done previously with school-aged children, but not with preschoolers.

Previous studies of speech and nonspeech processing in 8–10-year-old children have given variable results depending on stimulus properties. P1 amplitude was found to be larger for vowels than complex or simple tones (Bruder et al., 2011; Čeponienė et al., 2001) but smaller for syllables than nonspeech analogues (Čeponienė et al., 2005, 2008). The studies of Čeponienė et al. (2005, 2008) suggest that the child P1 is, in the absence of N1, fused together with P2, which in adults is enhanced to discrimination training (Tremblay et al., 2001) suggesting it reflects neural tuning to newly learned sound contrasts. Furthermore, the amplitude of P1 to prototypical vowels was found to correlate inversely with behavioral same-assessment of vowels and with reading speed, so that children with smaller P1s were more accurate in assessing two identical vowels "the same" and could also read more words per minute (Bruder et al., 2011). Therefore, the child P1 was suggested to reflect both sound detection and speech-nonspeech as well as the wideness of neural tuning curves to vowel prototypes (Bruder et al., 2011).

In the same studies, results for N2 amplitude were similarly variable. N2 was smaller (Čeponienė et al., 2001) or equal in size (Bruder et al., 2011) for vowels and simple tones when compared to complex tones, but larger for syllables than nonspeech analogues (Čeponienė et al., 2005, 2008). Since the amplitude of N2 elicited by tone pips was found to increase with repetition in nine-year-olds (Karhu et al., 1997), larger N2s to complex sounds than vowels were interpreted as memory-trace build-up for the unfamiliar stimuli (Čeponienė et al., 2001). In the studies using syllables, N2 and N4 behaved similarly, and were suggested to reflect higher-order sound analysis, such as the content recognition of syllables, scanning for access to semantic representations, or short-term memory retrieval (Čeponienė et al., 2001, 2005, 2008). As N4 was also larger for vowels than simple or complex tones, it is the only component, which has consistently had larger amplitude for speech than nonspeech sounds, and was thus interpreted as an index of sound "speechness" (Čeponienė et al., 2001, 2005, 2008).

A few studies of preschool children with clinical groups also stress the usefulness of ERPs as indexes of language development. For example, Lovio et al. (2010) reported diminished P1 peaks to syllables in 6-year-old children at risk for dyslexia, whereas Hämäläinen et al. (2013) reported abnormally large N2s to a short pseudo-word and its nonspeech counterpart in 6-year-old children who three years later had reading problems. Furthermore, in a longitudinal study, Espy et al. (2004) presented syllables and sinusoidal tones with long, 2.5-4.0 s inter-stimulus intervals (ISI), which produces the child N1 in addition to the P1-N2-N4 complex. Increased N1 amplitudes to both speech and nonspeech stimuli between ages 1 and 4 years were related to poorer pseudo-word reading at school, whereas decreased N2 amplitudes to nonspeech stimuli between ages 4 and 8 years predicted poorer word reading at school.

Here, our goal was to fill a gap in research by contrasting speech and nonspeech sound processing in preschoolers, using syllables and nonspeech stimuli that were carefully matched for acoustic properties with the speech stimuli. As, to our knowledge, there are no such previous studies in six-year-olds, our hypotheses are only tentative. If sound detection quality processing in preschoolers is akin to school-aged children, we will observe smaller P1 but larger N2 and N4 responses to syllables than nonspeech sounds (Bruder et al., 2011; Čeponienė et al., 2001, 2005, 2008). We will also analyze the relationship between cortical responses and neurocognitive task performance, expecting P1 amplitude to be associated with better phonological skills (Bruder et al., 2011), and larger speech than nonspeech N2/N4s to be associated with better cognitive functioning.

2. Methods

2.1. Participants

Originally, 94 typically developed monolingual Finnish-speaking children participated in a longitudinal study of preschool language abilities and later reading performance. The current study consists of the preschool data of 63 children (33 boys; 3 left-handed, 1 ambidextrous) that remained after the exclusion of the data of 31 children due to cancellation of participation (N = 12), a PIQ lower than the set limit of 85 (N = 1), later discovery of unclear family history of neurological problems (N = 1), excessive alpha band activity (N = 11) or motor artifacts (N = 8). The mean age of the children was 6 years 6 months (range 6 years 0 months–7 years 0 months), and they had an average of 80 (range 8–156) days of preschool teaching prior to the EEG experiment. All children were born full-term and had reportedly normal hearing. Most parents of the children had completed high school (fathers 73%, mothers 86%), and had college or university education (fathers 59%, mothers 71%), and were employed (fathers 90%, mothers 76%). The family background of these children is typical to the Finnish metropolitan capital area (Official Statistics of Finland (OSF), 2013).

The study was approved by the Ethical Board of Helsinki and Uusimaa Hospital District. Written consent was obtained from parents and oral consent from the child.

2.2. Stimuli and paradigm

The semi-synthetic CV syllables and their acoustically matched nonspeech counterparts used as stimuli were created using the Semisynthetic Speech Generation (Alku et al., 1999) method. Vowels/i/and/e/were compiled by extracting a glottal excitation from a natural speech signal, obtaining the desired formant structure with a digital all-pole filter, and adding the filtering effect of the vocal tract to the model. The F0 was 101 Hz for both vowels. For/i/, the lowest four formant frequencies were 410, 2045, 2260, and 3320 Hz and for/e/, 320, 2240, 2690, and 3275 Hz. The unvoiced plosives/k/and/p/were extracted from syllable/ke:/and the short word/pito/, and inserted to the beginning of the semi-synthetic vowels to create standard stimulus syllables/pi/and/ke/. The total duration of the standard stimulus was 170 ms (12 ms consonant and 158 ms vowel sections, including 5 ms rise and fall times), and its intensity set to approximately 55 dB SPL.

The nonspeech sounds were created by mimicking the glottal flow of the semi-synthetic syllables with a carefully controlled impulse train, so that the F0 was equal with the speech stimuli. Linear predictive coding (LPC; Rabiner and Schafer, 1978) of a prediction order of 10 was used to match the spectral envelope to that of the speech sound. Then the impulse train was used as an excitation to an all-pole filter which modeled only the second formant, i.e., the all-pole vocal tract consisted of a single resonance at 2240 Hz and 2045 Hz for the nonspeech counterpart of the/i/and/e/, respectively. LPC coding of a prediction order of 50 was used to model/p/and/k/, exciting it with random noise. The nonspeech syllables were formed by combining the corresponding nonspeech consonant and vowel counterparts (see Fig. 1). The speech and nonspeech stimuli were thus matched in terms of duration, F0, intensity