International Journal of Psychophysiology 86 (2012) 229–237

Contents lists available at SciVerse ScienceDirect

International Journal of Psychophysiology

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

Processing of prosodic changes in natural speech stimuli in school-age children

R. Lindström [a,*], T. Lepistö [a,b], T. Makkonen [a,c], T. Kujala [a,d]

[a] Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland
[b] Department of Child Neurology, Helsinki University Central Hospital, Helsinki, Finland
[c] Finnish Center of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Finland
[d] Cicero Learning, University of Helsinki, Helsinki, Finland

ARTICLE INFO

Article history:
Received 4 July 2012
Received in revised form 2 September 2012
Accepted 26 September 2012
Available online 3 October 2012

Keywords:
Children
Emotional prosody
Speech
Event-related potentials
Mismatch negativity
Facial electromyography

ABSTRACT

Speech prosody conveys information about important aspects of communication: the meaning of the sentence and the emotional state or intention of the speaker. The present study addressed processing of emotional prosodic changes in natural speech stimuli in school-age children (mean age 10 years) by recording the electroencephalogram, facial electromyography, and behavioral responses. The stimulus was a semantically neutral Finnish word uttered with four different emotional connotations: neutral, commanding, sad, and scornful. In the behavioral sound-discrimination task the reaction times were fastest for the commanding stimulus and longest for the scornful stimulus, and faster for the neutral than for the sad stimulus. EEG and EMG responses were measured during non-attentive oddball paradigm. Prosodic changes elicited a negative-going, fronto-centrally distributed neural response peaking at about 500 ms from the onset of the stimulus, followed by a fronto-central positive deflection, peaking at about 740 ms. For the commanding stimulus also a rapid negative deflection peaking at about 290 ms from stimulus onset was elicited. No reliable stimulus type specific rapid facial reactions were found. The results show that prosodic changes in natural speech stimuli activate pre-attentive neural change-detection mechanisms in school-age children. However, the results do not support the suggestion of automaticity of emotion specific facial muscle responses to non-attended emotional speech stimuli in children.

1. Introduction

Speech prosody conveys information about the meaning of the sentence and the intention and the emotional state of the speaker, which all are important aspects of communication. Speech prosody is defined as simultaneous variation of three main auditory features of speech: parallel intensity, fundamental frequency (F_0), and duration changes (Botinis et al., 2001). Prosody perception is fundamental to speech comprehension: in spoken language the identification of the word boundaries is based on the prosodic cues and also the meaning of the sentence can change dramatically if the prosody changes (Kuhl, 2004). Even newborns are capable of extracting information from prosodic features and automatically respond to some prosodic cues suggesting that speech prosody is a key aspect in human communication (Sambeth et al., 2008). Even though learning to interpret the mood of the speaker is an important developmental task during childhood, very little is known about the psychophysiological basis of children's ability to process emotional prosodic changes in natural speech.

As speech prosody is based on rapid physical changes of speech stimuli, the auditory event related potentials (ERPs) are relevant tools for studying such basic processing. The auditory ERPs are also feasible tools for studying children as the recordings can be done even in absence of participants' attention and task performance (Taylor and Baldeweg, 2002). Mismatch negativity (MMN) is an ERP component that reflects automatic, pre-conscious detection of violations of auditory regularities and it is elicited by any discriminable change in auditory flow (for reviews, see Baldeweg, 2007; Näätänen et al., 2012, 2007; Sussman, 2007). It is classically recorded in an "oddball paradigm", in which frequent "standard" stimuli are occasionally replaced by "deviant" stimuli (Winkler, 2003). It is maximally negative over the fronto-central scalp areas. As the MMN is generated mainly in the auditory cortices it reverses polarity at the electrodes below the level of the Sylvian fissure (when nose is used as a reference; Alho, 1995; Näätänen et al., 2007).

In children the MMN is usually followed by a late discriminative negativity (LDN, or Late MMN) (Ceponiené et al., 1998; Korpilahti et al., 2001, 1995; for review see Cheour et al., 2001). The function of the LDN is not clearly known, but it is hypothesized to reflect more cognitive aspects of auditory change detection than the MMN (Ceponiené et al., 2004). If the stimulus deviation catches subject's attention the MMN is often followed by a positive deflection (P3a and/or P3b) reflecting further attentive discrimination of stimulus deviation (for a review see Escera and Corral, 2007; Polich, 2007).

* Corresponding author at: Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, P.O. Box 9, FIN-00014 University of Helsinki, Finland. Tel.: +358 4 526 48 980; fax: +358 9 191 29450.
E-mail address: riikka.h.lindstrom@helsinki.fi (R. Lindström).

0167-8760/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpsycho.2012.09.010

Prosodic cues such as stress patterns are automatically processed in adults (Honbolygó et al., 2004; Ylinen et al., 2009). Adult data also suggest that the prosodic features related to the emotional state of the speaker in natural speech are detected at the early, pre-attentive level of auditory processing as reflected by the MMN and its magnetic counterpart the MMNm (Kujala et al., 2005; Thönnessen et al., 2010). In Kujala et al. (2005) study MMN responses were elicited (at about 200–300 ms from the onset of the sound) by the stimuli with emotional connotations, which served as deviant stimuli. Thönnessen et al. (2010) also reported MMNm responses to happy and angry deviants even in a sequence of randomly acoustically and phonetically changing disyllabic pseudowords and showed that the MMNm was elicited by the violations of the abstract emotional category representations, not solely by the change in acoustical features of the stimuli. There is some evidence that also school-age children can detect emotional prosodic changes in natural speech stimuli at the pre-attentive level of auditory processing. Korpilahti et al. (2007) showed both early (at about 200 ms) and late (at about 650 ms) MMNs to one-word utterance pronounced with a tender or commanding voice. Grossman et al. (2005) reported ERPs to natural speech stimuli with different emotional prosody even in 7-month-old infants suggesting that human brain starts to detect emotionally loaded words very early during the development.

In addition to the ERP recordings, another way to study human psychophysiological correlates of processing emotional speech stimuli are facial electromyographic reactions (facial EMG). Rapid facial reactions, occurring within 1000 ms, are commonly observed in adults after seeing facial expressions (Dimberg et al., 2000; Dimberg, 1990a; McIntosh, 2006). A specific muscle area activation is associated with a specific emotion category: zygomaticus major (muscle responsible for raising the cheek in a smile) area with happy emotions and corrugator supercilii (muscle responsible for knitting the brows in a scowl) with negative emotions, for example (Dimberg and Thunberg, 1998; Dimberg, 1990a; McIntosh, 2006). In adult data, the rapid facial reactions to other stimulus types than facial expressions have also been reported. For example Hietanen et al. (1998) have shown rapid facial reactions to emotional speech stimuli and Magneé et al. (2007) to emotional face–voice pairs (for other EMG studies using sound stimuli see: Dimberg, 1990b, 1990c; Jäncke et al., 1996; Kjellberg et al., 1994). Rapid facial reactions are described to be as automatic and unconscious reactions because they occur rapidly (even within 300–400 ms after stimulus presentation; Dimberg and Thunberg, 1998; Dimberg, 1997), the subjects cannot avoid producing them (Dimberg et al., 2002), and the reactions can occur even if the stimuli are backward masked (Dimberg et al., 2000). There are two possible underlying mechanisms for these rapid facial reactions: 1) motor mimicry (an automatic tendency to mimic another person's emotional behavior probably mediated by the mirror neuron system) and 2) an affective, initial, reaction to the emotional stimulus (Beall et al., 2008; Magneé et al., 2007; Moody et al., 2007; Williams et al., 2001).

Although these rapid facial reactions are well documented in adults, only few studies have been conducted with children. Recently, Beall et al. (2008) showed rapid facial reactions to facial expressions in 7–12 year old children and demonstrated that the reactions were reflecting motor mimicry of the stimuli as well as the emotional reactivity towards the stimuli (for consistent results in adults see also Moody et al., 2007). However, to our knowledge, no previous studies have tested if children show rapid facial reactions to natural speech stimuli.

The aim of our study was to explore how school-age children process prosodic changes in naturally articulated words uttered with different emotional connotations. To this end we recorded the ERPs, rapid facial muscle reactions, and behavioral responses for different types of emotional prosodic changes of words in 7–12 years old children. These natural stimuli have a high ecological validity, enabling the inspection of responses reflecting very closely reactions in real-life situations. As the rapid facial reactions to facial expressions are reported being automatic and unconscious (Dimberg et al., 2000) we wanted to study if natural speech stimuli would also elicit such involuntary facial reactions. Therefore we recorded the facial EMG together with the auditory ERPs during non-attended experimental paradigm. By combing these approaches we aimed at obtaining comprehensive information about children's neural processing of prosodic changes in natural speech stimuli. In addition, our aim was also to determine how emotional prosodic changes are discriminated at the attentive level as reflected by behavioral measures.

On the basis of the studies by Grossman et al. (2005) and Korpilahti et al. (2007) we hypothesized that school-age children would detect emotional prosodic changes in natural speech at the pre-attentive level of auditory processing as reflected by the MMN. However, on the basis of Kujala et al. (2005) we expected that the more salient prosodic changes (the commanding deviant) are more easily pre-attentively detected as reflected by earlier and stronger MMN.

2. Materials and methods

2.1. Participants

18 healthy children participated in the experiment. Their mean age was 10 years (standard deviation, sd 1.4 years; range 7.5–12.3 years; 9 girls; 2 left handed). The children's cognitive abilities were assessed with the Wechsler Intelligence Scale for Children III (WISC-III, Wechsler, 1991), using three verbal and three performance subtests. The mean full IQ of the group was 111 (sd 11.1, range 93–134), the mean verbal IQ 115.3 (sd 16.4, range 83–144), and the mean performance IQ 108.4 (sd 10.2, range 88–136). All the children were monolingual Finnish speakers. According to parental reports they all had normal hearing, no present or past neurological disorders, language or learning difficulty or emotional problems, and no family history of developmental or psychiatric disorders. The children were recruited from three elementary schools in the Helsinki area. Parents received an introductory letter on the experiment through teachers and contacted the researcher if they were interested in the study. The parents signed an informed consent and the children gave a verbal approval prior to the experiment. The experiment was carried out according to the Declaration of Helsinki, and it was approved by the Ethical Committee of the Helsinki University Central Hospital.

2.2. Stimuli and procedure

The same stimuli and paradigm were used as in Kujala et al. (2005). The stimulus was a Finnish word, a female name “Saara” (none of the children participating in this study had a family member with this name) uttered by a female speaker with four different emotional connotations: neutral (stimulus length 577 ms), commanding (538 ms), sad (775 ms), and scornful (828 ms) (for acoustic waveforms, see Fig. 1). The stimuli were originally produced for a study investigating the identification of utterances with different emotional connotations (Leinonen et al., 1997). These stimuli were categorized by 73 male and female subjects in Leinonen et al. (1997) study and they all were more often identified correctly than as representing another optional emotion. The mean percentages of correctly identified stimuli in Kujala et al. (2005) study were: 75% (sd 38) for the neutral stimulus, 91% (sd 19) for the commanding stimulus, 66% (sd 48) for the scornful stimulus, and 63% (sd 42) for the sad stimulus. The peak loudness of the stimuli was adjusted to vary randomly within 5 dB (Leinonen et al., 1997) but no other basic acoustic parameters of the stimuli were modified to keep them as natural as possible. For each stimulus, the F_0 and intensity at different time points as well as the mean F_0 and intensity values were calculated from intensity and F_0 contours determined for the whole stimulus duration using Praat software (Paul Boersma &