Preschoolers' real-time coordination of vocal and facial emotional information

Jared M.J. Berman [a], Craig G. Chambers [b], Susan A. Graham [a],*

[a] Department of Psychology, University of Calgary, Calgary, Alberta T2N 1N4, Canada
[b] Department of Psychology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

ARTICLE INFO

ABSTRACT

An eye-tracking methodology was used to examine the time course of 3- and 5-year-olds' ability to link speech bearing different acoustic cues to emotion (i.e., happy-sounding, neutral, and sad-sounding intonation) to photographs of faces reflecting different emotional expressions. Analyses of saccadic eye movement patterns indicated that, for both 3- and 5-year-olds, sad-sounding speech triggered gaze shifts to a matching (sad-looking) face from the earliest moments of speech processing. However, it was not until approximately 800 ms into a happy-sounding utterance that preschoolers began to use the emotional cues from speech to identify a matching (happy-looking) face. Complementary analyses based on conscious/controlled behaviors (children's explicit points toward the faces) indicated that 5-year-olds, but not 3-year-olds, could successfully match happy-sounding and sad-sounding vocal affect to a corresponding emotional face. Together, the findings clarify developmental patterns in preschoolers' implicit versus explicit ability to coordinate emotional cues across modalities and highlight preschoolers' greater sensitivity to sad-sounding speech as the auditory signal unfolds in time.

Introduction

Detecting and recognizing another's emotion is critical to communicative success. Consider, for example, the utterance "It's snowing!" The meaning of this statement varies considerably depending

* Corresponding author.
   E-mail address: susan.graham@ucalgary.ca (S.A. Graham).

upon whether the speaker is excited (perhaps a skier) or upset (perhaps a stranded commuter). In many cases, the emotional disposition of the speaker is communicated alongside the linguistic information in the speech stream via nonlinguistic modulations of the voice (e.g., variations in pitch level, pitch contours, and speech rate; see Banse & Scherer, 1996; Frick, 1985), which together are often referred to as "emotional prosody" or "vocal affect." Although there is a growing literature examining children's attention to vocal affect (e.g., Berman, Chambers, & Graham, 2010; Quam & Swingley, 2012; Sauter, Panattoni, & Happé, 2013), the question of how children at different ages dynamically coordinate vocal affect with the other perceptual information source that commonly signals an individual's emotional disposition, namely facial expression, is not well understood. Here, we examined 3- and 5-year-olds' ability to link vocal affect in unfolding speech with a corresponding emotional face, focusing on differences related to emotional valence and the time course of information processing. Our assessment of this ability is based on implicit/unconscious physiological measures (namely saccadic eye movements to displayed faces) as well as overt/explicit measures (children's points toward what they consciously judge to be the "matching" face for an unfolding utterance).

The building blocks for children's effectiveness and flexibility in coordinating the cues to emotion that are provided by different sensory–perceptual modalities are evident early in development (see Walker-Andrews, 1997, for a review). For example, 3-month-olds will match vocal affect with one of two dynamic emotional faces, but only when the individual depicted is their mother (Kahana-Kalman & Walker-Andrews, 2001). By 7 months of age, however, infants can use unfamiliar faces and voices to discriminate across emotional categories (Flom & Bahrick, 2007; Walker-Andrews, 1986). Furthermore, event-related potential (ERP) studies have found patterns supporting the view that there is more consistent discrimination of and recognition for multimodal versus unimodal cues to emotional state prior to 7 months of age (Grossmann, Striano, & Friederici, 2006, 2007).

When cues to emotion are conveyed via a single modality (face or voice), we see a gradual development of emotion recognition abilities over the childhood years. For example, although neonates can discriminate between different facial expressions (Field, Woodson, Greenberg, & Cohen, 1982), children's ability to associate the label *happy* with happy facial expressions and the label *sad* with sad facial expressions emerges just before 4 years of age. The ability to identify other facial emotions, such as fear and surprise, develops more slowly during childhood (Widen, 2013). In the auditory domain, children can accurately judge a speaker's emotional state based on vocal affect by age 4 if the emotional content of the speech itself is either neutral or filtered out (e.g., Morton & Trehub, 2001; Nelson & Russell, 2011). However, when there are competing cues regarding the relevant emotion conveyed via either the meaning of a sentence (e.g., Morton & Trehub, 2001) or the situational context (Gil, Aguert, Le Bigot, Lacroix, & Laval, 2014), preschoolers' ability to recognize vocal affect is compromised. Research has also shown that preschoolers' success at identifying vocal affect varies across emotional categories, with greater success in recognizing sad vocal affect compared with happiness or fear (Nelson & Russell, 2011), and that the ability to recognize emotion categories from auditory information continues to develop well beyond the age of 5 years (Sauter et al., 2013).

Our goal in the current study was to examine aspects of the multimodal coordination of auditory and visual cues to emotion during the preschool years. However, rather than focusing on the added benefit of multimodal input or the integrative processing of simultaneously perceived facial and speech-based cues, we instead considered how cues provided by one modality (speech information) allow a child to identify a relevant percept in another modality (an emotional face). This "orienting" ability is a core component of emotion recognition in many day-to-day scenarios involving a context with multiple individuals and where the perceiver is not already focused on an individual producing emotional cues via facial expression or vocal affect. This arises in part because auditory information can be easily perceived without directed attention due to the way in which sound travels. As a result, vocal affect is often detected and interpreted by a perceiver first, leading the perceiver to redirect visual attention to establish the identity of the "source" individual.

An important aspect of our study is the inclusion of both *implicit* and *explicit* measures. Much of the research on vocal emotion has focused primarily on young children's detection and recognition of emotional cues using explicit measures that are less able to capture the dynamic nature of the auditory signals that convey emotion, namely the fact that spoken utterances unfold in time. This issue is essential to consider in view of evidence from studies of adults demonstrating that the time required