Analysis of engagement behavior in children during dyadic interactions using prosodic cues☆

Rahul Gupta *, Daniel Bone, Sungbok Lee, Shrikanth Narayanan

Signal Analysis and Interpretation Laboratory (SAIL), University of Southern California, 3710 McClintock Avenue, Los Angeles, CA 90089, USA

Abstract

Child engagement is defined as the interaction of a child with his/her environment in a contextually appropriate manner. Engagement behavior in children is linked to socio-emotional and cognitive state assessment with enhanced engagement identified with improved skills. A vast majority of studies however rely solely, and often implicitly, on subjective perceptual measures of engagement. Access to automatic quantification could assist researchers/clinicians to objectively interpret engagement with respect to a target behavior or condition, and furthermore inform mechanisms for improving engagement in various settings. In this paper, we present an engagement prediction system based exclusively on vocal cues observed during structured interaction between a child and a psychologist involving several tasks. Specifically, we derive prosodic cues that capture engagement levels across the various tasks. Our experiments suggest that a child's engagement is reflected not only in the vocalizations, but also in the speech of the interacting psychologist. Moreover, we show that prosodic cues are informative of the engagement phenomena not only as characterized over the entire task (i.e., global cues), but also in short term patterns (i.e., local cues). We perform a classification experiment assigning the engagement of a child into three discrete levels achieving an unweighted average recall of 55.8% (chance is 33.3%). While the systems using global cues and local level cues are each statistically significant in predicting engagement, we obtain the best results after fusing these two components. We perform further analysis of the cues at local and global levels to achieve insights linking specific prosodic patterns to the engagement phenomenon. We observe that while the performance of our model varies with task setting and interacting psychologist, there exist universal prosodic patterns reflective of engagement.
© 2015 Elsevier Ltd. All rights reserved.

Keywords: Engagement; Prosody; Global level cues; Local level cues; Classifier decision fusion

1. Introduction

During childhood an individual develops critical social, physical, psychological and cognitive skills and abilities. This development is affected by several factors including society (Walker et al., 2007; Davie et al., 1972), family (Biller, 1993; Egeland and Farber, 1984) and peers (Dodge et al., 2003). Furthermore, developmental changes are reflected in behavioral aspects such as joint attention (Akhtar et al., 1991; Tomasello and Farrar, 1986), and ability to

---

engage, among others (Cloward et al., 1960; Göncü, 1999; Morrissey-Kane and Prinz, 1999). Quantitative assessment of these behavioral aspects, while very challenging, can provide tools for understanding important aspects of child development, both typical and atypical. In turn, this can further inform intervention methods targeted toward assisting healthy child development. Several approaches for quantifying a child's behavioral aspects such as social and language development (Volkmar et al., 1993; Coplan and Gleason, 1990) exist and we note that their specific type and nature is very dependent on, and tailored to the corresponding behavior of interest. Given the vast heterogeneity and variability in developmental trajectories, especially in the presence of neuro-cognitive and behavioral disorders, there is an imminent need for quantitative methods and analysis tools that can help shed further light into developmental behavioral processes and mechanisms.

Understanding and quantitatively characterizing the engagement patterns of a child, a core behavioral construct, can be useful for both diagnostics and intervention design. Child engagement is defined as the child being involved with his/her environment in a contextually appropriate manner (McWilliam and Casey, 2008). Engagement is a complex internal state externalized and reflected in several modalities including face and body language (Xu et al., 2010; Sanghvi et al., 2011), speech (Yu et al., 2004; Manning et al., 1994) and physiology (Nes et al., 2005). Several studies suggest that a greater engagement has a constructive impact on a child's development (McWilliam et al., 2003; Taylor et al., 2003). For instance, investigations by de Kruif and McWilliam (1999) suggest positive multivariate relationships between developmental age and observed child engagement, where the developmental age is determined over personal–social, adaptive, communication, motor and cognitive domains (Newborg et al., 1984). Göncü (1999) has reported on the impact of a child's engagement during social activities to his/her development and underscores the importance of an interdisciplinary approach to such an endeavor. The importance of engagement is also emphasized in the study of children with developmental disorders like autism (Delano and Snell, 2006; Poulsen and Ziviani, 2004) aiming to enhance behavioral intervention methods (Kasari et al., 2010; Rogers, 2000). Furthermore, improved engagement is associated with success of organizations like child care centers (Maher Ridley et al., 2000) and schools (Skinner and Belmont, 1993). Methods of intervention exist to improve child engagement in different settings such as school (Skinner et al., 1990), parent–child interactions (Casey and McWilliam, 2005) and play with peers (Cielinski et al., 1995).

Given that child development and engagement are strongly coupled, several schemes have been proposed to measure engagement. Yatchmenoff (2005) proposes to quantify engagement in child protective services by categorizing it into five dimensions of receptivity, expectancy, investment, mistrust and working relationship. Kishida and Kemp (2006) have proposed measures of engagement specific to practitioners as opposed to researchers motivated by their relation to practicality, sensitivity to the participants and ability to measure across the span of activity types. Libbey (2004) measured engagement of children in schools by defining a few school connectedness measures based on conceptual interrelatedness. Other studies such as in Read et al. (2002) view engagement as a dimension of a higher order construct and propose measures of engagement as a subcomponent to analyze the construct. However, these studies do not extend to the cases involving natural interaction with children and are often limited to artificial settings. Moreover, given that engagement is a latent internal state inferable only using observed cues, a majority of studies rely on a subjective measurement of engagement. Such measures are susceptible to several uncertainties introduced by variability in interaction settings, interpersonal differences, inconsistencies across subjective judgments and even the operational definition of engagement.

We aim to address the need for an objective engagement behavior quantification method that is robust to the variations in environmental parameters. We perform a study in which children interact with a psychologist while performing different socio-cognitive tasks. The child–psychologist dyadic interaction provides an opportunity to investigate interaction engagement under various settings introduced by differences in task conditions. In the study, we develop a computational system based solely on the observed vocal cues, specifically vocal prosody during the dyadic interaction. Furthermore, it is hypothesized that the engagement of the child can be predicted from the acoustic prosodic cues of both the child and the psychologist and data-driven methods can be designed to capture this relationship.

Our approach is inspired by several previous studies that link speech prosody to human behavior based constructs such as emotion (Austermann et al., 2005; Lee et al., 2011), approach-avoidance (Rozgic et al., 2011; Xiao et al., 2012), entrainment (Lee et al., 2014), blame/acceptance (Black et al., 2013), and empathy (Kempe, 2009; Aziz-Zadeh et al., 2010). These studies present techniques that computationally model prosodic patterns which are otherwise difficult to quantify perceptually. We build upon our previous work in Gupta et al. (2013, 2012) and present a data-driven approach to identify global prosodic patterns (task level statistics) as well as those that last over much shorter time spans (local