Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 65 (2014) 109–118

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Feasibility of augmenting text with visual prosodic cues to enhance oral reading

Rupal Patel [a,b,*], Heather Kember [a], Sara Natale [a]

[a] Northeastern University, Department of Speech Language Pathology and Audiology, Boston, MA, United States
[b] Northeastern University, College of Computer and Information Science, Boston, MA, United States

Received 13 January 2014; received in revised form 19 June 2014; accepted 18 July 2014
Available online 30 July 2014

Abstract

Reading fluency has traditionally focused on speed and accuracy yet recent reports suggest that expressive oral reading is an important component that has been largely overlooked. The current study assessed the impact of augmenting text with visual prosodic cues to improve expressive reading in beginning readers. Customized reading software was developed to present text augmented with prosodic cues to convey changes in pitch, duration and/or intensity. Prosodic modulation was derived from the recordings of a fluent adult model and rendered as a set of visual cues that could be presented in isolation or in combination. To establish baseline measures, eight children aged 7–8 first read a five-chapter story in standard text format. In the subsequent three sessions, participants were trained to use each augmented text cue with the guidance of an auditory model. They also had the opportunity to practice reading aloud in each cue condition. At the post-training session, participants re-recorded the baseline story with each chapter read in one of the different cue conditions (standard, pitch, duration, intensity and combination). Post-training and baseline recordings were acoustically analyzed to assess changes in reading expressivity. Despite large individual differences in how each participant implemented the prosodic cues, as a group, there were notable improvements in marking pitch accents and elongating word duration to convey linguistic contrasts. In fact, even after only three training sessions, participants appeared to have generalized implementation of pitch and word duration cues when reading standard text at post-training. In contrast, while participants manipulated pause duration when provided with explicit visual cues, they did not transfer these cues to standard text at post-training. These findings suggest that beginning readers could benefit from explicit visual prosodic cues and that even limited exposure may be sufficient to learn and generalize skills. Further discussion focuses on the implications of this work on struggling readers and second language learners.

Keywords: Oral reading; Prosody; Reading software; Fluency; Children

1. Introduction

Fluent oral reading is a hallmark of skilled reading and involves a number of complex skills including rapid or semi-automatic word decoding and the extraction of syntactic and semantic information to facilitate the translation of text into speech (Adams, 1990). Although much of the literature defines reading fluency in terms of rate and accuracy (Daane et al., 2005), evidence suggests expression and ease of reading are also critical to oral reading fluency (Allington, 1983; Dowhower, 1991; Eason et al., 2013; Jenkins et al., 2003).

Reading with expression requires modulation of prosody – the rhythm and melody of speech. Prosody is used to signal linguistic contrasts and express emotions

* Corresponding author. Addresses: Northeastern University, Department of Speech Language Pathology and Audiology, 360 Huntington Ave., Room 204 FR, Boston, MA 02115, United States and Northeastern University, College of Computer and Information Science, 360 Huntington Ave., Room 204 FR, Boston, MA 02115, United States. Tel.: +1 (617) 373 5842.
E-mail address: r.patel@neu.edu (R. Patel).

http://dx.doi.org/10.1016/j.specom.2014.07.002
0167-6393/

and attitudes (Lehiste, 1970; Shattuck-Hufnagel and Turk, 1996). Speakers manipulate the fundamental frequency (F0) of their voice (perceived as pitch), together with changes in duration, vocal intensity (perceived as loudness), and voice quality to convey linguistic and affective goals (Xu, 1999, 2011). Prosody is also important for reading comprehension. Listeners use prosodic cues to segment oral language into meaningful syntactic units, or phrases within an utterance (Cutler et al., 1997; Shattuck-Hufnagel and Turk, 1996), which supports working memory and comprehension. In fact, even newborns and infants utilize prosody to attune to the rhythmic regularities of their native language (Morgan and Demuth, 1995).

Given the importance of prosody in spoken fluency and comprehension, we hypothesized that providing explicit visual cues to the underlying prosody would improve reading fluency in beginning readers. Although beginning readers can modulate conversational prosody, many children struggle to apply this skill when reading aloud, resulting in expressionless and labored speech even when they are proficient decoders. The lack of sufficient cues in written text may contribute to this apparent dichotomy between the presence of prosodic modulation in conversation yet its absence during reading. Readers must draw inferences about appropriate prosody from context, punctuation and grammar (Carlson, 2009; Miller and Schwanenflugel, 2006; Schreiber, 1987).

Furthermore, reading with expression is made even more challenging by the fact that prosodic control is developing simultaneously with reading acquisition (Cruttenden, 1985; Crystal, 1978; Local, 1980; Snow, 1994, 1998). The developmental trajectory of prosodic control begins with the modulation of cries (Gilbert and Robb, 1996; Lind and Wermke, 2002; Protopapas and Eimas, 1997; Wermke et al., 2002) and continues throughout childhood and even into adolescence (Cruttenden, 1985; Crystal, 1986; Local, 1980; Snow, 1994, 1998; Tingley and Allen, 1975; Wells et al., 2004). Young children often use different acoustic cues, or combinations of cues than older children to signal prosodic contrasts. For example, Patel and Grigos (2006) showed that while seven and eleven year olds marked yes/no questions with increased phrase final F0, four year olds tended to rely on duration cues. Perhaps rising contours are more motorically demanding (Snow, 1998) thus young children manipulate duration. Additionally, Grigos and Patel (2007) noted despite being able to mark prosodic contrasts, 7 year olds exhibited greater kinematic and acoustic variability than 11 year olds suggesting continued motor development.

Previous attempts to address the need to supplement written text with prosodic information have been limited to manipulations of spacing, punctuation, font and case. Some researchers have recommended formatting text to display intra-sentence phrasal boundaries to facilitate chunking of text into meaningful units (Cromer, 1970; Levasseur et al., 2006; O'Shea and Sindelar, 1983). Others have suggested manipulating punctuation (e.g. My friend? My friend! My friend.) and font case (e.g. I like SOME of my relatives versus I like some of MY relatives) to practice modulating intonation (Blevins, 2001). These approaches apply a set of grammatical rules to convey prosodic variation. There is however, considerable prosodic variation in natural speech that cannot be captured by simple mappings.

Our approach aimed to augment written text with visual prosodic cues that are derived from fluent adult recordings. The goal is to provide beginning readers with the scaffolding to read aloud expressively as they continue to master control of prosody. Toward this end, we developed a software program called ReadN'Karaoke that provides multimodal (visual and auditory) cues to three different components of prosody: pitch, duration, and vocal loudness. The first version (ReadN'Karaoke 1.0) directly manipulated text based on a fluent adult reader's F0, duration, and intensity variation. A user study with typically developing children showed significant increases in F0 and duration modulation when reading with the cues (Patel and McNab, 2011). Although these results were promising, participants also reported that manipulated words were sometimes hard to read; specifically word boundaries were often difficult to distinguish on pitch manipulated tokens (Patel and McNab, 2011). To address these concerns, a new visualization scheme was designed. Rather than manipulating text, ReadN'Karaoke 2.0, augments written text with overlaid cues to pitch, duration and intensity variation (Patel and Furr, 2011). Fig. 1 displays both the manipulated text from ReadN'Karaoke 1.0 alongside the augmented text from ReadN'Karaoke

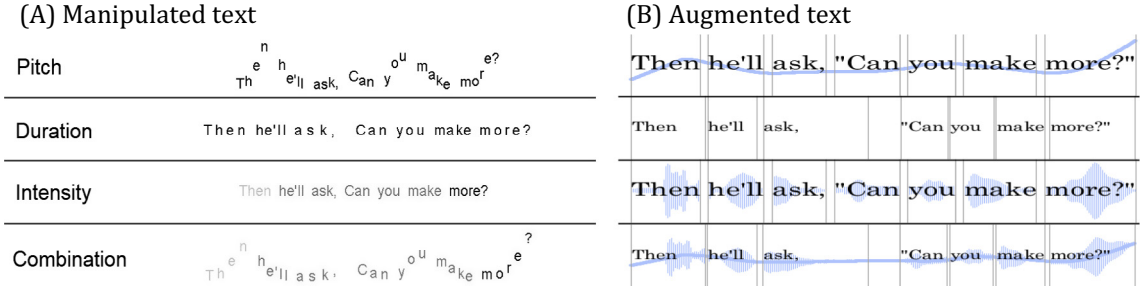

Fig. 1. The four different manipulated text formats rendered in ReadN'Karaoke 1.0 (left) and augmented text formats used in ReadN'Karaoke 2.0 (right).