Reading KITTY: Pitch Range as an Indicator of Reading Skill

Alfredo Gomez
Harvey Mudd College
Claremont, CA
agomez@hmc.edu

Alessandra Otondo
Harvey Mudd College
Claremont, CA
aotondo@hmc.edu

Alicia Ngo
Harvey Mudd College
Claremont, CA
ango@hmc.edu

Julie Medero
Harvey Mudd College
Claremont, CA
jmedero@hmc.edu

Abstract

While affective outcomes are generally positive for the use of eBooks and computer-based reading tutors in teaching children to read, learning outcomes are often poorer (Korat and Shamir, 2004). We describe the first iteration of Reading Kitty, an iOS application that uses NLP and speech processing to focus children’s time on close reading and prosody in oral reading, while maintaining an emphasis on creativity and artifact creation. We also share preliminary results demonstrating that pitch range can be used to automatically predict readers’ skill level.

1 Introduction

Early elementary-aged children’s reading achievement is aided by reading practice in general, especially when that practice focuses on reading with appropriate prosody (Kim and Petscher, 2016). Because individualized instruction is time-intensive for teachers, many eBook readers and computerized reading programs are available for emerging readers but, while those tools are popular, they result in poorer reading interactions than paper books, which negatively impacts learning outcomes.

By focusing students’ time on practicing reading with appropriate prosody, with an end goal of an artifact creation, the Reading with Kinetic Typography (Reading Kitty) app aims to keep the positive affective elements of eBooks without sacrificing quality of interaction with texts. In the remaining sections, we outline first steps toward development of Reading Kitty. We also share preliminary analysis demonstrating that the prosodic analysis used in Reading Kitty can provide accurate and useful feedback.

2 The Reading Kitty app

Kinetic typography is a category of animations that uses moving text and images to emphasize the meaning and structure of text. Figure 1 shows a frame from a Kinetic Typography animation of a scene from the film Toy Story. In the animation, words are emphasized through text size, weight, and motion to reflect the way that they’re emphasized by the character that delivers the line.

By interacting with Reading Kitty, a child reader is able to create their own kinetic typography animations that reflect their interactions with a story. First, Reading Kitty guides a child reader through a close reading by presenting a series of questions customized to the story a child is reading. Those questions correspond to oral reading fluency benchmarks in the Common Core Standards Initiative (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). The child uses the answer to each close reading question as the basis for adjusting the appearance of words in the story.

Next, Reading Kitty has the child read aloud. The prosody of the child’s reading is used with the text characteristics selected by the child to generate a kinetic typography artifact, with better prosody resulting in more dynamic videos. Finally, feedback about the child’s oral reading is shared with the child’s teacher.

Figure 1: Kinetic typography frame (Reb, 2016).
3 Pitch Analysis

We want Reading Kitty to reward children who read with appropriate prosody with more interesting, dynamic animations of their readings. With this goal in mind, we first examine the feasibility of predicting pitch variation used by skilled readers. Classification models have been used previously to predict the locations of prosodic boundaries (Medero and Ostendorf, 2013). We adopt this idea by building a model to predict which words should be read with a high pitch. We train the model on the Boston Directions Corpus because this corpus is hand-labeled with ToBI pitch tones (Silverman et al., 1992).

We use features that have proven successful for predicting ToBI tones in Greek: part-of-speech (POS) tags, POS bigrams, syntactic chunk tags, the number of syllables in each word, and each word’s depth in an automatically-generated syntactic constituency tree (Zervas et al., 2004). For our data set, though, resulting error rate is too high to give reliable feedback to children. We look instead at pitch range across an utterance.

3.1 Pitch range

We analyze the CMU Kids Corpus (Consortium. and University., 1997), which contains 5,180 utterances (spoken sentences) from 76 children. The children range in age from 6 to 11 and are in first through third grade. These 76 children are split into 2 groups, skilled readers \((n = 44)\) and struggling readers \((n = 32)\). Each utterance in the corpus is accompanied by a transcript, along with word and phone-level forced alignments.

For each utterance, we extract the \(F_0\) values in Hz every .01s using PRAAT’s pitch tracker (Boersma, 2001). We then calculate the pitch range, in semitones, as described by Aoyama et al. (Aoyama et al., 2016) for each utterance. We average the ranges in semitones for each utterance across all readers in the skilled group and all readers in the struggling group.

Skilled readers show larger average pitch range than struggling readers across utterances (Fig. 2), which is promising for our goal of using pitch range to generate feedback for readers.

3.1.1 Paired Comparison

Finally, we perform a paired t-test. We pair a skilled reader with a struggling reader who read the same utterances. Since the participants in the two groups do not align exactly, we use the Hungarian method to find the pairings that maximize the number of shared utterances (Kuhn, 1955), resulting in 13 pairs with 41-91 shared utterances. We ignore utterances that are not shared by each pair. The skilled readers have an average pitch range of 17.03 semitones \((\sigma = 1.85\) semitones) and the struggling readers have an average pitch range of 14.83 semitones \((\sigma = 2.15\) semitones), which is statistically significant with \(p < .05\).

4 Conclusion

Reading Kitty represents a new way to bring evidence-based reading instruction in an electronic form. Our initial analysis provides promising evidence that pitch features can be used to give feedback on prosodic quality of oral readings. Further investigation into automating kinetic typography generation as well as field testing is required to incorporate these features to promote close-reading and creativity.

References


