Exploring Speech Therapy Games with Children on the Autism Spectrum

Mohammed E. Hoque[1], *Joseph K. Lane*[1], *Rana el Kaliouby*[1], *Matthew Goodwin* [1,2], *Rosalind W. Picard*[1]

[1] MIT Media Lab, 20 Ames St., Cambridge, MA 02139
[2] Groden Center, 86 Mount Hope Ave, Providence, RI, 02906
{mehoque,jklane,kaliouby,mgoodwin,picard}@media.mit.edu

Abstract

Individuals on the autism spectrum often have difficulties producing intelligible speech with either high or low speech rate, and atypical pitch and/or amplitude affect. In this study, we present a novel intervention towards customizing speech enabled games to help them produce intelligible speech. In this approach, we clinically and computationally identify the areas of speech production difficulties of our participants. We provide an interactive and customized interface for the participants to meaningfully manipulate the prosodic aspects of their speech. Over the course of 12 months, we have conducted several pilots to set up the experimental design, developed a suite of games and audio processing algorithms for prosodic analysis of speech. Preliminary results demonstrate our intervention being engaging and effective for our participants.

Index Terms: autism spectrum disorder, prosody, computerized speech intervention

1. Introduction

Autism spectrum disorders are a collection of neuro-developmental disorders characterized by qualitative impairments in social interaction and social relatedness, as well as difficulties in acquiring and using communication and language abilities, and a restricted range of interests and preference for consistency and predictability in daily routines [1]. Approximately one third to one half of individuals on the autism spectrum have significant difficulty using speech and language as an effective means of communication [2]. Some examples include reduced engagement in turn taking, irregular patterns of speech rate and inflection of voice, inappropriate pauses in reciprocal conversations, literal interpretation of figurative language, and difficulty understanding the social cues of the listener, i.e., monologuing [3]. These difficulties in speech production and processing can result in interpersonal interactions being overwhelming, confusing, stressful and are often misinterpreted as a general disinterest to engage in social interactions.

In this paper, we introduce an on-going exploratory speech intervention where we engage our participants in customized interactive games to help improve their speech intelligibility. For example, several of our participants speak so fast that they are hard to understand by their teachers and peers. To address this difficulty, we customized a turtle race game, where a participant controls one of the turtles by speaking at a slower speech rate. The objective of the game is to finish the race first by speaking at a slow speech rate. Our hypothesis, in this intervention, is that real-time visualizations of speech properties, which often act as social mirrors, can influence social communication. With this intervention, our objective is not to replace the traditional speech therapist. Instead, we propose our games as an easily customizable and freely available supplement to speech-language therapies to help individuals with speech difficulties.

1.1. Prosody and Autism Spectrum Disorder

The inflection of voice, patterns of pauses, relative duration of syllables, relative loudness, and rhythm are often termed as prosodic aspects of speech. These features are not entirely predictable at word or sentence level; instead, they are obtained by analyzing phoneme sequences [4]. In other words, prosodic features are independent of words and can not be deduced from lexical channels.

In reciprocal conversations, prosody defines communicative functions (syntactic, pragmatic, and affective) and enhances or changes the meaning of what is said [6] Prosody is embedded into conversational turns and often works as a "signaling device" to influence the next set of actions that people would normally take given a social context. However, individuals on the autism spectrum, despite having a good understanding of grammar and phonology, often exhibit extremely poor usage of social language [7]. Fine *et al.* [8] showed that individuals on the spectrum are unable to assess the social context of a situation, and as a result, they either do not use appropriate patterns of intonation or they systematically demonstrate the misuse of linguistic system.

Because of the implications of speech-language abilities on social communication, there is a longstanding interest in the assessment of speech production and processing abilities in autism. These interests, along with recent developments in speech technology, have made it possible to develop precise and objective measures of speech and language ability. For instance, Van Santen *et al.* [9] used computer-based speech technologies to quantify expressive prosody and generate acoustically controlled speech stimuli for measuring receptive prosody. Also, spectral speech tools such as Speech Visualization [10] allow real-time viewing and analysis of speech in controlled settings for loudness, pitch, intonation, timing, rate, and rhythm. However, none of these tools are easily accessible nor they provide interactive interfaces which can be rapidly changed based on participants preferences.

The remaining part of this paper is divided into three sections. In section 2, we present our experimental design including an explanation of our proposed computerized speech therapy, evaluation schema and participants' details. Section 3 provides preliminary results of our intervention; Section 4 concludes the paper and describes future work.

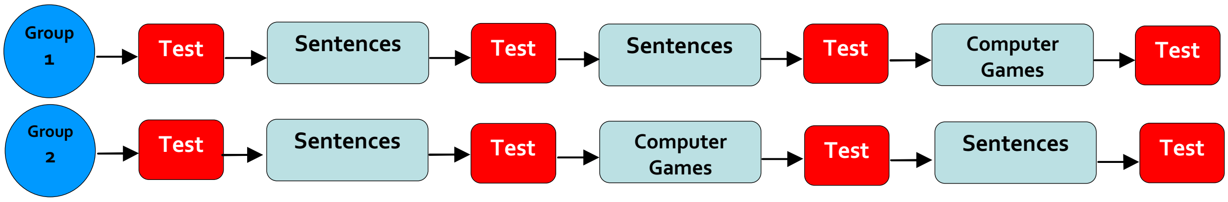

Figure 1: Experimental design.

2. Experimental Design

As shown in Figure 1, our study contains two groups running in parallel and undertaking both traditional and computerized speech intervention. An initial evaluation of the participant's performance was completed to quantify baseline speech production characteristics. Groups 1 and 2 go through traditional speech therapy (described in section 2.2), for two weeks followed by a test to evaluate any changes in performance incurred during this phase. Group 1 then undergoes an additional two weeks of traditional speech therapy while Group 2 experiences two weeks of computerized speech therapy. Another post-evaluation is conducted to record any improvement. Then, each group receives the other style of intervention for another two weeks. A final post-evaluation is performed to record any change in performance. Recording performance at each point of the intervention enables us to observe a cause-and-effect relationship between the traditional and our proposed speech therapy. It may also be useful in measuring specific impact of the speech interventions on the participants.

2.1. Participants

This study was conducted at the Groden Center, a non-profit school that provides evaluative, therapeutic, and educational programs for individuals diagnosed with autism and other developmental disorders. Eight participants with diagnosis of Autism Spectrum Disorder and one with a diagnosis of Down syndrome were recruited. Five of them were assigned to Group 1 and the remaining four participants were assigned to Group 2. Assignment of the participants into two groups was based on the similar speech difficulties. The participants' age, gender, diagnosis, and objectives for speech-language therapy as listed in their school profile are shown in Table 1.

2.2. Sentences (Traditional Speech therapy)

In traditional speech therapy, therapists usually employ flash cards with objects drawn on them and prompt the participants to explain the content of it. To allow for a fair comparison between the traditional speech therapy and the computerized speech therapy, instead of using flash cards, we have modeled around 170 everyday utterances per group containing 3-4 words per utterance. Examples of such utterances are, "*Boys are playing*", "*I want Pizza*", "*I am happy today*" etc. During the intervention, a therapist prompts the participant to repeat any of the 3/4 word sentences chosen from a list of 170 predetermined sentences. If the participant exhibits appropriate inflection of voice, volume, and engagement, the therapist moves on to another sentence. Otherwise, further instruction is given to prompt the participant to try again.

Table 1. Demographic information of the participants, as listed in their school, recruited in this project. PDD= Pervasive Development Disorder; ASD= Autism Spectrum Disorder; ADHD= Attention Deficit Hyperactivity Disorder; ODD = Oppositional Defiant Disorder; MR = Mental Retardation

#	Age	Sex	Diagnosis	Speech Goal
1	15	M	PDD, Global speech and language delay	speak faster and louder
2	14	M	ASD	speak louder
3	15	F	Bi Polar, Mild MR, ADHD, ODD	speak louder and slower
4	15	F	ASD	speak slower
5	16	M	Axis 1-Mood, Severe MR	take turns by using appropriate social language
6	19	M	ASD, anxiety, NOS	speak faster
7	17	M	Down Syndrome	speak clearer
8	8	F	ASD	speak louder
9	8	M	ASD	speak slower

2.3. Computerized Speech Therapy

In computerized speech therapy, just as in the traditional speech therapy, the therapist prompts the participant to repeat one of 3/4 word sentences from the same list of 170 predetermined sentences. The only difference is that instead of getting direct feedback from the therapist, participants get feedback from interactive games. The speech therapist per group remains the same throughout the intervention for both traditional and computerized therapy.

Kaypentax [11] has been specializing in building games for more than two decades to help people with their speech disorders. We decided to start with their off-the-shelf product to quickly validate the efficacy of our approach, instead of spending time upfront developing games. However, after a series of pilot studies and careful observation of those games, we were able to understand the underlying mechanism and recreate the same set of games using the freely available software called Scratch (scratch.mit.edu), developed at the MIT Media Lab. This was more desirable as Scratch is free, platform independent, and provides users with an easy-to-use interface to change the background, characters, and game parameters. It was also well received by our participants since in autism there tend to be a preference towards customization.

As shown in Table 1, since most of the participants have difficulties with amplitude modulation and speech rate, we chose games that incorporate amplitude and speech rate. In games that employ amplitude, participants are required to modulate the volume of their voice to control objects in the game. Similarly, in games that employ speech rate, participants must control the rate of their speech in order to perform well in the games.

2.4. Experimental Setup

Our experimental setup has evolved over time through a few months of pilot studies with teachers, staff, and occasionally with participants. In the beginning, we had participants use a table-top microphone attached to a laptop as an input device. However, we then realized that participants would lean forward to the microphone to compensate for their low volume. Therefore, we started using an Audio Technica AT892 MicroSet Wired Headset, which hangs non-intrusively from the ear, ensuring that most participants maintain a consistent distance between their mouth and the microphone. This microphone was interfaced with the laptop using a Tascam US-144 audio and midi interface.

We found that the participants were often distracted from the games by the keyboard in front of them. Additionally, we felt that interaction between the therapist and the participants would be more effective if they were engaged in face-to-face communication. To address these issues, we introduced an external monitor in our experimental setup to allow the teacher to control the laptop while maintaining face-to-face contact with the participant. This set up is shown in Figure 2.

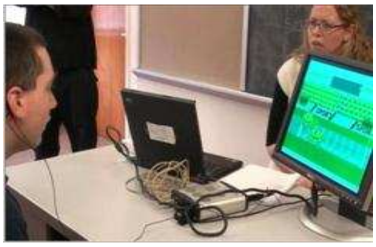

Figure 2: The experimental setup of teacher and the participant facing each other. The teacher controls the game by using the laptop and the content of the game is displayed on the external monitor.

2.5. Test/Evaluation

We are aware of the concern that good performance in speech games or traditional speech therapy may not translate well into natural social situations. To address this concern, at the beginning of each session, we engage the participants in natural conversation about topics that are of interest to them. After the session, the therapist engages the participant on the same topic, while the participant is attending to the game.

Each speech session is recorded using Audacity, an open source audio editor. Using semi-automated speaker segmentation methods, we extract the utterances for each participant. The extracted utterances are then processed by Sona-Speech's Multi-Dimensional Voice Program (MDVP) and *Praat* [12] speech processing software. In MDVP, the four measured parameters are the Relative Average Perturbation (RAP) to measure variability of the pitch period; Shim to measure the variability of peak-to-peak amplitude; Noise Harmonic Ratio (NHR) used to quantify the amount of noise present in the analyzed signal; and finally the Voice Turbulence Index (VTI), which measures the higher frequency inharmonic to harmonic ratio and provides a measure of noise commonly generated by incomplete or loose movement of the vocal chords.

Using *Praat*, prosodic features related to pitch (minimum, maximum, mean, rate of change etc.), intensity (minimum, maximum, mean, mode), speaking rate (syllables per second), and pauses (average number of pauses per utterance, average duration of pauses, maximum duration of pauses per utterance) are automatically extracted for utterances per participant. Additionally, the speech therapist keeps track of how many prompts were needed during each session. This is done using a simple hand held golf clicker making it easier for the teacher to attend to the participants. The therapist's subjective analysis of the participant's performance is also recorded. The analysis from the MDVP, the prosodic speech features, the number of prompts, and the therapist's reports are all used to assess the participant's performance across the speech study.

3. Preliminary Analysis and Results

Through a few pilot studies of our proposed experimental design, we have observed interesting results supporting our hypothesis.

The participants enjoy interacting with the games so much that they often continue to play the games even after the allocated time. This contrasts to their traditional speech therapy sessions, half of which were discontinued due to cognitive overload. All nine participants immediately understood the objective of the game, and after a few trials, interacted very well with the game. In traditional speech therapy, the participants were often distracted, bored, annoyed, and restless by being asked to repeat a set of sentences. The participants often had to be given some sort of reward (e.g., chocolates) to keep them on task. On the other hand, during the computerized intervention, participants were excited and engaged while trying out different set of games eliminating the needs for supplementary rewards throughout any of the sessions. Additionally, in the computerized intervention, the participants' interest seemed to increase proportionally as they gradually understood the objectives of the games. The gradual improvement of the participants' ability to meaningfully control their voice surpassed the expectations of some of the teachers who interact with the students on a daily basis. The teachers felt that the computerized sessions were not only more enjoyable for the students, but also less distressing for the teachers as they no longer had to repeatedly focus on "reinforcement" to keep the participants on task.

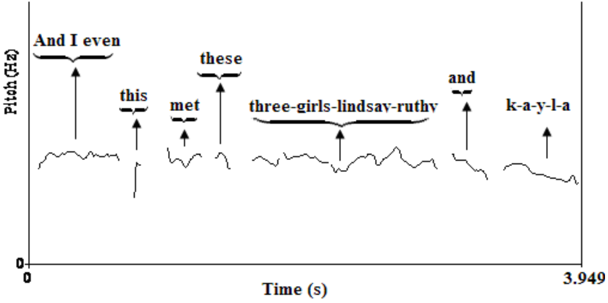

Figure 4(a): Fundamental frequency of a natural utterance by participant #6

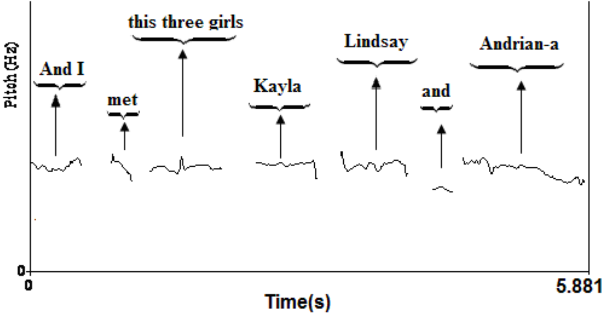

Figure 4(b): Fundamental frequency of the same (almost) utterance, while playing the turtle race game

We have noticed that participants are able to use appropriate speech rate, intonation, and pauses to explain the very same sentence in conversation while attending to the game, an achievement that the therapists did not think was possible. For example, one of the participant's - let's call him Tom, has a tendency of speaking quickly making it difficult to understand him. While engaging in a natural conversation with him, Tom said a sentence whose fundamental frequency is plotted along with its transcription in Figure 4(a). The results of Tom saying the same sentence while attending to the computer game are plotted in figure 4(b). One can see in 4(b) that the pitch elements, corresponding to different words, are disjoint. This phenomenon corresponds to the participant's ability to pause between each word, thereby drastically improving intelligibility. Figure 4(a, b) shows that participant took 3.9 seconds to complete the sentence compared to 5.9 seconds to express the same sentence while attending to the game. This is an indication of the participant's ability to pace his speech in context. We have also computationally measured the speech rate (syllables per second) of the two utterances to quantitatively validate our assertion. For the utterance displayed in Figure 4(a), there were 2.78 syllables per second, whereas for Figure 4(b), there were 2.04 syllables per second. This confirms that Tom did speak with a slower speech rate when playing the games.

In Figure 5 (a, b), a comparison of prosodic properties for two participants in natural conversation and while attending the computer games is plotted. There is a significant reduction (particularly for #3), as shown in Figure 5(a), in the number of pitch breaks during computer intervention. Having fewer pitch breaks in a conversation corresponds to having better control of one's pitch. Figure 5(b) demonstrates comparison of other pitch characteristics for two participants during their natural speech and speech while attending to the games.

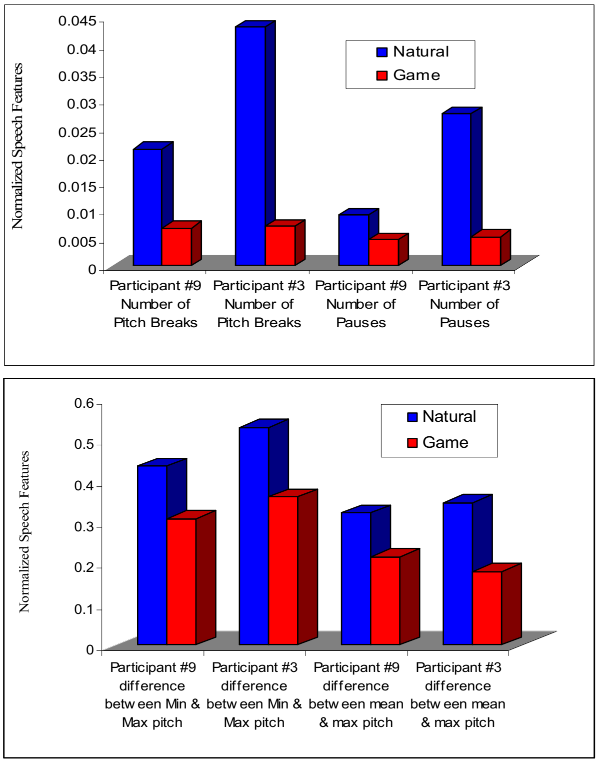

Figure 5(a, b): Comparison of prosodic properties of speech of two participants in natural conversation and the conversations during computer intervention (game)

4. Conclusion

In this paper, we have introduced an exploratory novel speech intervention to help individuals on the autism spectrum with unintelligible speech. We have pursued traditional speech therapy along with customizable interactive speech-enabled games, in parallel. We measure performance in every two weeks to observe any changes that our participants exhibit due to the speech intervention that the individual is subjected. In order to measure the performance, we have designed three different criteria. First, we analyze the recorded speech using Sona-Speech's Multi-Dimensional Voice Program (MDVP). Then, speech therapists provide information related to number of prompts that were required per session and their subjective analysis of how the participant performed. Finally, *Praat* speech processing software is used to record prosodic statistics of speech.

Our initial observations through a set of pilot studies, employing the proposed experimental setup, suggest that participants are more engaged, competitive, and cheerful during the computerized intervention compared to the traditional speech intervention. Our proposed approach aims to enable speech-language therapists, teachers, and parents to assess and teach verbal expressions in a new and enjoyable way that is individually-tailored for each person's interest, sensory, and perceptual capabilities. By helping individuals on the autism spectrum with their communication needs, we aim to increase their competency, confidence, and engagement in social interactions thereby improving their quality of life by enabling them to integrate more naturally into society.

5. Acknowledgements

This research was supported by Nancy Lurie Marks Family Foundation (NLMFF).

6. References

[1] American Psychiatric Assoc. *The Diagnostic and Statistical Manual of Mental Disorders. Vol. 4, 1994*

[2] S. E. Bryson. "Brief report: Epidemiology of autism" *Journal of Autism and Developmental Disorders*, vol. 26, pp. 165-167, April 1996.

[3] L. Capps, J. Kehres and M. Sigman. "Conversational Abilities among Children with Autism and Children with Developmental Delays". *Autism*, vol. 2, pp. 325–344, 1998.

[4] C. Shih and G. Kochanski, "Prosody and Prosodic Models," *7th International Conference on Spoken Language Processing*, Denver, Colorado, 2002.

[5] E. Couper-Kuhlen. *An Introduction to English Prosody*. London: Edward Arnold, 1986.

[6] A. Cruttenden. *Intonation*. Cambridge University Press, 1986.

[7] D. Bishop, J. Chan, C. Adams, J. Hartley, and F. Weir. "Conversational responsiveness in specific language impairment: evidence of disproportionate pragmatic difficulties in a subset of children". *Development and Psychopathology*, vol. 12, pp. 177-99, 2000.

[8] J. Fine, G. Bartolucci, G. Ginsberg, and P. Szatmari. "The Use of Intonation to Communicate in Pervasive Development Disorders". *Journal of Child Psychology and Psychiatry*, vol. 32, pp. 771-82, 1991.

[9] J. V. Santen, et al A.Kain. "Synthesis of Prosody Using Multi-Level Sequence Units". *Speech Communication*, vol. 46, pp. 365-375, 2005.

[10] Locutour. "Speech Visualization: An Easy to Use Spectral Speech Tool", Learning Fundamentals Inc.

[11] KayPentax. "Sona-Speech", Internet: www.kayelemetrics.com, [January 17, 2009].

[12] P. Boersma and D. Weenink. "Praat: Doing Phonetics by Computer." Internet: www.praat.org, [January 7, 2009].