Investigating Therapist Vocal Nonverbal Behavior for Applications in Robot-Mediated Therapies for Autism

Wing-Yue Geoffrey Louie[1], Jessica Korneder[2], Ala'aldin Hijaz[1], Megan Sochanski[1]

[1] Intelligent Robotics Laboratory at Oakland University, Rochester, MI 48307 USA
[2] Applied Behavior Analysis Clinic at Oakland University, Rochester, MI 48307 USA

Abstract. Socially assistive robots (SARs) are being utilized for delivering a variety of healthcare services to patients. The design of these human-robot interactions (HRIs) for healthcare applications have primarily focused on the interaction flow and verbal behaviors of a SAR. To date, there has been minimal focus on investigating how SAR nonverbal behaviors should be designed according to the context of the SAR's communication goals during a HRI. In this paper, we present a methodology to investigate nonverbal behavior during specific human-human healthcare interactions so that they can be applied to a SAR. We apply this methodology to study the context-dependent vocal nonverbal behaviors of therapists during discrete trial training (DTT) therapies delivered to children with autism. We chose DTT because it is a therapy commonly being delivered by SARs and modeled after human-human interactions. Results from our study led to the following recommendations for the design of the vocal nonverbal behavior of SARs during a DTT therapy: 1) the consequential error correction should have a lower pitch and intensity than the discriminative stimulus but maintain a similar speaking rate; and 2) the consequential reinforcement should have a higher pitch and intensity than the discriminative stimulus but a slower speaking rate.

Keywords: Socially Assistive Robots, Nonverbal Behavior, Autism

1 Introduction

Socially assistive robots (SARs) have the potential to transform healthcare services provided to individuals and improve outcomes within a variety of healthcare settings. SARs have already been utilized in assisted living facilities [1], mental healthcare [2], and exercise programming [3]. In these settings, SARs successfully interacted with patients and improved their physical, mental, and emotional health. In general, current interventions utilizing SARs have focused on *what* content is delivered during an intervention but not *how* an intervention should be delivered to a patient to be effective.

During healthcare interactions both verbal and nonverbal behavior are necessary for effective as well as efficient communication and understanding to occur. For healthcare professional-patient interactions, verbal behaviors consist of the actual words spoken to patients and nonverbal behaviors include everything but the words that are spoken [4]. Examples of these nonverbal behaviors can include: gestures, facial expressions, body pose, interpersonal distance, appearance, vocal cues (i.e., prosody), and time-based cues (i.e., chronemics). Nonverbal communication strategies are important because they contribute positively to patient satisfaction, adherence, affect, and health outcomes during an intervention [5, 6].

The understanding of interactions with patients through nonverbal behavior is especially relevant in therapeutic fields such as Applied Behavior Analysis (ABA) [7]. This form of therapy is a widely used, evidence-based practice implemented for individuals with autism spectrum disorder (ASD). SARs are increasingly being combined with the principles of ABA to teach individuals with ASD social skills [8], imitation skills [9], and emotion recognition [10]. The verbal behaviors of these robot-mediated interventions are modeled after human therapist facilitated interventions where individuals with ASD have been successful in the acquisition of the skills targeted by the interventions. However, it remains unclear how robot nonverbal behaviors should be characterized throughout an intervention.

Our team's long-term research goal is to integrate SARs in healthcare settings to support the delivery of healthcare services to patients. Our current research efforts have focused on developing SARs to deliver ABA therapy to children with ASD to address the growing prevalence of ASD as well as labor challenges in delivering ABA services [11, 12]. Namely, we have developed a robot-mediated intervention that closely replicates existing ABA therapies and discrete trial training (DTT) teaching procedures implemented by human therapists at ABA clinics for teaching children with ASD to independently answer WH-questions [13]. The primary focus of our prior work was replicating the interaction flow and verbal behaviors of the human therapists during these interventions.

Our objective in this work is to study the context-dependent vocal nonverbal behaviors that ABA therapists exhibit when facilitating DTT therapy sessions with children with ASD. Specifically, we chose to investigate pitch, intensity and speaking rate of therapists because these are the vocal nonverbal behaviors that can be modeled with current state-of-the-art voice synthesizers [14]. We investigated the differences in therapist vocal nonverbal behaviors during different contexts within a DTT intervention session. Our primary hypotheses were that therapists: 1) display similar pitch, intensity, and speaking rate during the delivery of a consequential error correction and discriminative stimulus; and 2) display a higher pitch, higher intensity, and higher speaking rate during consequential reinforcement as compared to the delivery of a discriminative stimulus. This study will be important for informing the design of vocal nonverbal behaviors of SARs delivering ABA based interventions to individuals with ASD. Furthermore, this work serves as a model for studying context-dependent nonverbal behaviors during human-human interactions and utilizing these insights to design human-robot interactions (HRIs).

2 Related Work

Nonverbal behaviors have most commonly been studied within human-human interactions. The fields that have analyzed nonverbal behaviors among humans have included education [15], medicine [16], and therapeutic fields [17, 18]. The common trend amongst these studies has been the analysis of the overall nonverbal behavior of human participants during an entire human-human interaction. For example, in [16] the effect of a physician's overall eye contact on elderly patient understanding and adherence after a routine doctor visit was investigated. Overall, it was shown that when doctors use eye contact in conjunction with verbal communication patients had higher understanding and adherence to medical interventions.

In [17], the effect of music therapists' overall affect, interpersonal distance, and eye contact on older adult Alzheimer patients' affect and participation was investigated during group therapies. Namely, a study was conducted to measure Alzheimer patient affect and participation during group therapies under four therapist nonverbal behavior conditions: affect and interpersonal distance, affect alone, interpersonal distance alone, and no affect or interpersonal distance. Results indicated that the nonverbal behavior of therapists directly impacted the affect and participation of the older adult with Alzheimer's.

In [18], change in therapist nonverbal behavior was investigated during therapies with individuals with depression. Namely, therapist nonverbal behavior was coded at the beginning and end of therapy sessions with patients. After observing cognitive-therapy sessions, reliable changes in nonverbal behavior by the patient and therapist were observed from the start of the session to the end of the session. However, the study did not investigate whether these nonverbal behavior changes were associated with changes in the context of a therapist's communication goals during a therapy.

Additionally, nonverbal behavior literature in social robotics has also focused on studying human-human interactions and using these studies to serve as a model for HRIs. These studies have investigated and modeled human gaze [19], speech-based gestures [20], and dyadic interaction based facial expressions [21] during social interactions. Although recent research has been successful in modeling general nonverbal social behaviors during human-human interactions and applying these models to HRIs, there has been a lack of emphasis on investigating how the nonverbal behaviors of humans change as result of changes in their communication goals or contexts. Additionally, research regarding nonverbal behaviors in HRI have primarily focused on investigating motion-based cues.

Current research investigating nonverbal behaviors in human-human interactions for robotics or other fields have all focused on general nonverbal behaviors or general demeanor of the doctor/therapist during human-human interactions. There has been a lack of research towards investigating how human nonverbal behaviors change within an interaction due to changing communication goals. The communication goal context is important to the correct application of nonverbal behaviors. When nonverbal behaviors are used in inappropriate situations or contexts it can lead to negative attitudes towards message delivery, poor message comprehension, and lack of trust [22, 23]. Hence, nonverbal communication must be considered within the context in which it occurs because it guides a human's nonverbal behavior encoding and a listener's nonverbal behavior decoding [22]. A robot interacting with human users should be capable of applying the appropriate nonverbal behaviors in the correct communication goal contexts while considering the interaction partner's behaviors to improve HRIs.

In this work, we aim to close this gap by investigating context-dependent vocal nonverbal behaviors during human-human interactions. Specifically, we focus on how therapists adapt their vocal nonverbal behaviors during the delivery of ABA therapies to children with ASD. This study is a necessary step to inform the design of vocal nonverbal behaviors of SARs during robot-mediated ABA therapies for ASD.

3 ABA DTT Therapy

ABA therapy utilizes the principles of Behavior Analysis for individuals with ASD [24]. One such therapy is discrete trial training. Namely, discrete trial training structures a unit of instruction into three components: 1) discriminative stimulus, 2) behavior, and 3) consequence. A discriminative stimulus is a social and/or environmental cue which signals a behavior to occur. Behavior is the response to the discriminative stimulus. Consequences are the events that occur after a behavior. Whether naturally occurring interactions or therapeutic interactions, the consequences determine whether an individual will repeat the behavior or decrease the behavior from occurring again in the future. This analysis of behavior was transformed into a therapy which addresses behavior deficits and excess in children diagnosed with ASD [25]. Discrete trial therapy consists of therapists facilitating numerous discrete trials which address deficits such as imitation, visual performance, expressive and receptive language, and social interactions. For example, if a patient is unable to identify common objects in their environment, a DTT program would include the selection of the most relevant common objects for that patient and teach each object to mastery. The ABA therapist would ask the child a question (Discriminative Stimulus) such as “What is this?”, the patient’s behavior would follow, and depending on whether the response was correct or incorrect a reinforcer or error correction would end the trial. This process of teaching is an evidenced-based practice commonly used for individuals with ASD of all ages [26].

4 Model for Dyadic Social Interactions

In this work, we developed a general model to investigate dyadic social interactions between two agents (e.g., human-human or human-robot), Figure 1. We define a dyadic social interaction by the setting, roles of the participants, and overall goal of an interaction. Within a dyadic social interaction, a participant can then have multiple communication subgoals. We utilize Shannon & Weaver’s model of communication [27] to define each of the communication subgoals during an interaction. Namely, communication is defined as a process where a speaker delivers a message to a listener. A speaker encodes a message according to his/her distinct communicative subgoal; these messages can be encoded as verbal and/or nonverbal behaviors. This model can then be used to identify and classify basic units of nonverbal behaviors to be analyzed during real-world dyadic social interactions.

We applied this model to one-on-one DTT-based therapies delivered at an ABA autism clinic to children diagnosed with ASD. In this dyadic social interaction, the setting is a therapy room at a clinic, the roles of the participants are a therapist and child with

ASD receiving treatment, and the overall goal of an interaction is to teach a skill (e.g., greeting, WH-questions). The set of unique communicative subgoals of a therapist during ABA therapy are then the discriminative stimulus, consequential error correction, and consequential reinforcement. To achieve their communicative subgoals during an interaction, a therapist generates messages through verbal and nonverbal behaviors. Hence, the objective in this paper is to use this model to investigate how ABA therapist vocal nonverbal behaviors differ according to the different communicative subgoals (i.e., contexts) of a DTT therapy.

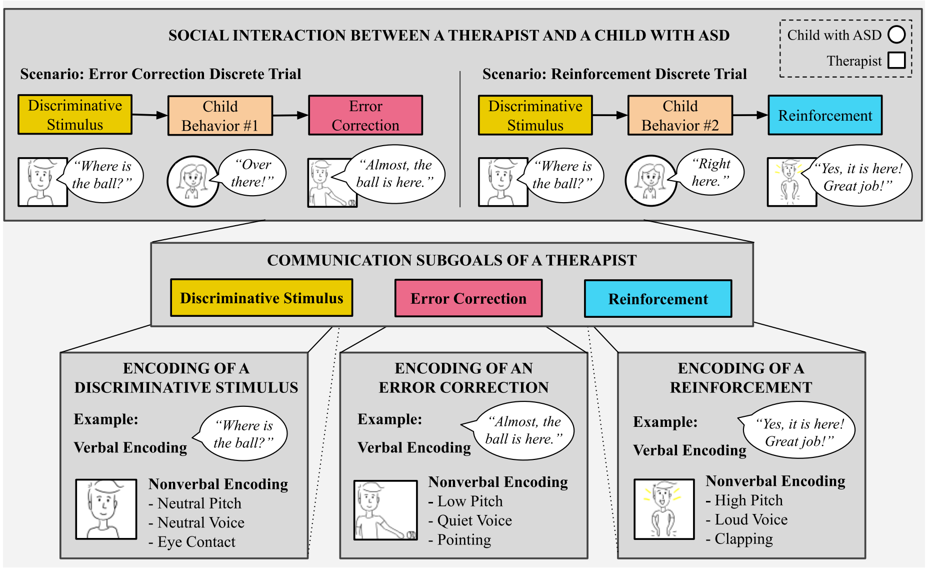

Fig. 1. Model for dyadic social interactions using discrete trial training therapy as an example.

5 Study Design

Our study focused on investigating the differences in vocal nonverbal behaviors of human ABA therapists during DTT therapies. Our primary hypotheses are that therapists:

1. *Pitch, intensity, and speaking rate during the delivery of a consequential error correction will not be significantly different from the delivery of a discriminative stimulus*
2. *Display a higher pitch, higher intensity, and higher speaking rate during consequential reinforcement compared to the delivery of a discriminative stimulus*

Our hypotheses were formulated according to expert therapists' expectations on their vocal nonverbal behavior during the therapies. Furthermore, we utilize the discriminative stimulus as the baseline for comparison because the discriminative stimulus subgoal is always used by a therapist to initiate a discrete trial with a patient. In order to evaluate these hypotheses, we conducted an analysis of vocal nonverbal behaviors utilized by therapists during one-on-one therapies with children with ASD.

5.1 Participants

A total of five ABA therapists from a university-based ABA clinic participated in this study delivering DTT therapies to children 2-9 years old with ASD. There was one female and four male children with ASD. The therapists ranged in age from 22-47 (μ=29.6) and had a range of 1-5 (μ=2.2) years of experience delivering ABA-based treatment. All therapists were female and have previously interacted with the children for four months to a year prior to the study.

5.2 Setting

The one-on-one therapy sessions were held at a university-based ABA clinic where the children were already receiving ABA services and the therapists implemented DTT programs already included in the children's on-going treatment program. The therapies targeted skills including: following one-step instructions, language acquisition, articulation, visual performance (i.e., matching), and gross motor imitation. The one-on-one sessions were each held in a private carpeted room 8ft x 10ft in size. The rooms each had three child sized chairs, a table, and storage containers with various items (e.g., toys, food, electronics, books). Each room also had pre-existing video recording equipment mounted 8ft high in the corner of the room. This video recording equipment was utilized to record the therapy sessions for our study.

5.3 Procedure

Informed consent from the therapists and children's parental guardians was obtained prior to the start of the study. Video recordings of one-on-one DTT sessions between a therapist and a child with ASD were then obtained. For each therapist, we collected video recordings until we obtained five trials of a discriminative stimulus-consequential error correction pair and five trials of a discriminative stimulus-consequential reinforcement pair. Since the video recordings were of real-world therapy sessions, the therapists' behaviors were dependent on the progress of the children receiving the therapies. Hence, more than five trials of discriminative stimulus-consequential error correction pairs were observed before obtaining five trials of discriminative stimulus-consequential reinforcement pairs. The converse also occurred. In either case only the first five trials of each pair were retained.

5.4 Data Collection

The video recorded sessions were segmented into the distinct communicative subgoals of the therapists during DTT sessions and prosodic data were collected for each of the distinct communicative subgoal segments. As discussed in Section 4, therapist distinct communicative subgoals during DTT sessions fall under three categories: discriminative stimulus, consequential error correction, and consequential reinforcement. These three categories were used to segment the video recordings of the sessions. Namely, the researchers reviewed the video recordings and categorized the therapist's speech into one of the three categories. Note that a communicative subgoal can be categorized based on a single word, a sentence, or multiple sentences spoken by a therapist. Data on the child's behavior was not collected. Figure 2 illustrates a therapy session segmented into distinct communicative subgoals.

Once the videos were segmented into distinct communicative subgoals we collected three prosodic parameters from each of the segments: mean pitch, mean intensity, and speaking rate. We utilized Praat [28], an application designed for phonetics research, to measure the therapists mean pitch (*Hz*) and mean intensity (*dB*) during each of the segmented distinct communication subgoals during the therapy. Speaking rate was defined as the number of syllables per a minute (*SPM*) spoken by the therapist. Namely, speaking rate was calculated by:

$$Speaking\ Rate = (S/T) \tag{1}$$

where S is the number of syllables in the therapist's speech during the distinct communication subgoal and T is the total amount of time the therapist took to communicate his/her speech.

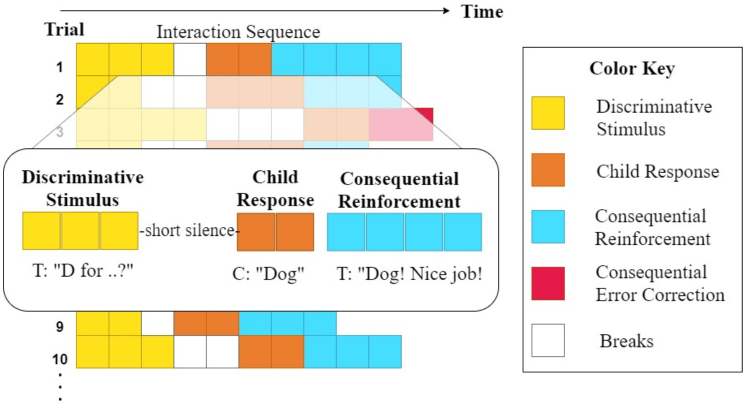

Fig. 2. An example of how a therapy session is segmented into distinct communicative subgoals.

5.5 Data Analysis

To test our hypotheses, we utilized multilevel models. Multilevel models are commonly utilized in speech research where repeated-measurement designs are used and multiple observations are nested within a participant [29]. Multilevel models account for the potential correlations between observations made within the same participant. In our study, a two-level multilevel model was used for each set of observations of the subgoal pairs (i.e., discriminative stimulus-consequential error correction and discriminative stimulus-consequential reinforcement) from the five participants. Namely, observations sampled from a participant are defined as the first level of the model and participants were defined as the second level. The treatment conditions were the two subgoals for each pair. The dependent variables were mean pitch, mean intensity, or speaking rate. We utilized IBM SPSS to run our statistical analyses.

6 Results

In total, we collected twenty-five discriminative stimulus-consequential error correction pairs and twenty-five discriminative stimulus-consequential reinforcement pairs from the therapy sessions. For each of these pairs we collected the mean pitch, mean intensity, and speaking rate for each of the subgoals in the pair. A multilevel model was constructed for each prosodic parameter (i.e., pitch, intensity, speaking rate) for each subgoal pair. A total of six multilevel models were constructed.

6.1 Discriminative Stimulus-Consequential Error Correction

Pitch - The relationship between the communication subgoals during an error correction trial and therapist pitch demonstrated significant variance in intercepts across participants, $x^2(1) = 12.72$, $p < 0.05$. The slopes did not significantly vary across participants, $x^2(1) = 1.56$, $p > 0.05$, and the slopes and intercepts did not significantly covary, $x^2(2) = 0.28$, $p > 0.05$. Only including the variance in intercepts across participants improved the fit of the model and, therefore, the final model used to interpret therapist pitch only included variance in intercepts.

From the final model, therapist pitch for the delivery of a consequential error correction was significantly different from the delivery of the discriminative stimulus, $F(1,45) = 4.40$, $p < 0.05$. This model suggests that therapist pitch is significantly lower during the delivery of a consequential error correction then discriminative stimulus, $b = -23.88$, $t(45) = -2.10$, $p = 0.04$.

Intensity - The relationship between the communication subgoals and therapist intensity demonstrated significant variance in both intercepts across participants, $x^2(1) = 33.833$, $p < 0.05$, and slopes across participants, $x^2(1) = 4.071$, $p < 0.05$. However, the slopes and intercepts across participants did not significantly covary, $x^2(2) = 4.2476$, $p > 0.05$. The final model only included the variance in intercepts and slopes across participants.

Therapist intensity for the delivery of a consequential error correction was not significantly different from the delivery of the discriminative stimulus, $F(1,4.97) = 5.35$, $p > 0.05$. However, this model suggests that therapist intensity was lower during the delivery of a consequential error correction and this relationship was not significant, $b = -4.70$, $t(4.97) = -2.31$, $p = 0.069$.

Speaking Rate - Similar to intensity, the relationship between the communication subgoal and therapist speaking rate demonstrated significant variance in intercepts across participants, $x^2(1) = 18.28$, $p < 0.05$, and the slopes across participants, $x^2(1) = 7.00$, $p < 0.05$. The slopes and intercepts across participants did not significantly covary, $x^2(2) = 0.27$, $p > 0.05$. The final model only included the variance in intercepts and slopes across participants.

Therapist speaking rate for the delivery of a consequential error correction was not significantly different from the delivery of the discriminative stimulus, $F(1,5.19) = 1.29$, $p > 0.05$. Although the model suggests that therapist speaking rate is higher during the delivery of a consequential error correction than discriminative stimulus, it was not significant, $b = 0.50$, $t(5.19) = 1.14$, $p = 0.305$.

6.2 Discriminative Stimulus-Consequential Reinforcement

Pitch - The relationship between the communication subgoal and therapist pitch during a reinforcement trial demonstrated significant variance in intercepts across participants, $x^2(1) = 16.18$, $p < 0.05$. However, the model including slopes across participants did not converge. The final model only included the variance in intercepts across participants. Such model simplification techniques are commonly used when multilevel models do not converge [30].

Therapist pitch for the delivery of a consequential reinforcement was significantly different from the delivery of the discriminative stimulus, $F(1,45) = 7.65$, $p < 0.05$. This model suggests that therapist pitch is significantly higher during the delivery of a consequential reinforcement than a discriminative stimulus, $b = 36.09$, $t(45) = 2.77$, $p = 0.008$.

Intensity - The relationship between the communication subgoal and therapist intensity demonstrated significant variance in intercepts across participants, $x^2(1) = 19.65$, $p < 0.05$. The slopes did not significantly vary across participants, $x^2(1) = 1.42$, $p > 0.05$, and the slopes and intercepts across participants also did not significantly covary, $x^2(2) = 0.65$, $p > 0.05$. The final model only included variance in intercepts across participants.

Therapist intensity for the delivery of a consequential reinforcement was significantly different from the delivery of the discriminative stimulus, $F(1,45) = 8.69$, $p < 0.05$. This model suggests that therapist intensity was significantly higher during the delivery of a consequential reinforcement, $b = 3.98$, $t(45) = 2.95$, $p = 0.005$.

Speaking Rate - Similar to pitch, the relationship between the communication subgoal and therapist speaking rate demonstrated significant variance in intercepts across participants, $x^2(1) = 10.03$, $p < 0.05$, but the model including slopes did not converge. The final model only contained variance in intercepts across participants.

The therapist speaking rate for the delivery of a consequential reinforcement was significantly different from the delivery of the discriminative stimulus, $F(1,45) = 24.65$, $p < 0.05$. This model suggests that therapist speaking rate is slower during the delivery of a consequential reinforcement than a discriminative stimulus, $b = -1.17$, $t(45) = -4.97$, $p < 0.001$.

7 Discussion

The results of our study only partially supported our hypotheses on the vocal nonverbal behaviors of therapists during the different communicative subgoals of a DTT-based therapy. As previously mentioned, our hypotheses were formulated according to the experience of expert practitioners and these discrepancies are likely because interpersonal communication is an automatic process that humans find difficult to describe explicitly [31]. This highlights the importance of studying context-dependent nonverbal behaviors during real-world human-human interactions so that they can be appropriately applied to the design of HRIs.

Our first hypothesis was only partially supported by the results. As expected, therapist speaking rate for the delivery of a consequential error correction was not significantly different from the delivery of the discriminative stimulus. In contrast to our expectations, therapist pitch and intensity was lower for consequential error correction than discriminative stimulus, but intensity was not statistically significant. Pitch and intensity were likely lower because therapists attempt to provide constructive feedback in a non-judgmental tone to the child. It is recommended that reprimands during DTT are provided in a quiet tone of voice [7]. Adhering to guidance on the use of reprimand helps to reduce escape behaviors, student alienations, and damaging the child-therapist relationship. Furthermore, a lower intensity reduces the emphasis on the failure to respond correctly. Focusing more noticeably on the positive behaviors and less on the incorrect behaviors helps to increase the positive behaviors and reduce the less desirable behaviors (i.e., differential reinforcement) [7].

Similarly, our second hypothesis was also only partially supported by the results of the study. As we expected, therapist pitch and intensity were higher for the consequential reinforcement than discriminative stimulus. The primary purpose for higher pitch and intensity was to display excitement, positivity, and to draw more attention to the positive reinforcement. Studies have shown that higher pitch and higher intensity are perceived as excitement and positivity by children [45]. This is important because the effectiveness of praise increases when presented in a manner acceptable to the individual (e.g., enthusiastically) [7]. Furthermore, when implemented correctly reinforcement can promote a positive relationship between the therapist and child [44]. However, the results demonstrated that therapist speaking rate was slower for the consequential reinforcement than the discriminative stimulus, which contrasted with our first hypothesis. Upon further analysis, it was observed that a slower speaking rate was often exhibited to emphasize and exaggerate the reinforcement. Studies have shown that slower speaking rates are used to emphasize communication points by a speaker [34]. Reinforcement is intended to draw attention to a correct behavior [7]. As such, when providing reinforcement therapists are likely to slow down their speaking rate to extend the verbal reinforcement to clearly show a positive reaction to the child.

8 Conclusions

The objective of this work was to study the vocal nonverbal behaviors of human therapists during the delivery of discrete trial therapies so that we can apply them to robot-mediated therapies. According to these findings, we make the following general recommendations for the design of the vocal nonverbal behaviors for a robot during robot-mediated discrete trial training therapies:

1. The discriminative stimulus should be utilized as the baseline for the vocal nonverbal behavior of the robot.
2. The consequential error correction should have a lower pitch and intensity than the discriminative stimulus but maintain a similar speaking rate.
3. The consequential reinforcement should have a higher pitch and intensity than the discriminative stimulus but a slower speaking rate.

As a next step, we plan to utilize these design recommendations to investigate whether robots modeling similar vocal nonverbal behaviors of human therapists will improve the efficiency and efficacy of discrete trial training therapies.

9 Acknowledgements

This work was supported by the National Science Foundation CRII Award (#1948224). We would like to thank all the participants from the Applied Behavior Analysis Clinic.

References

1. Papadopoulos, I., et al.: Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: A systematic review. BMJ Open 10(1), 1-13 (2020).
2. Rabbitt, S.M., Kazdin, A.E., Scassellati, B.: Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. Clinical Psychology Review 35, 35-46 (2015).
3. Martinez-Martin, E., Cazorla, M.: A socially assistive robot for elderly exercise promotion. IEEE Access 7, 75515-75529 (2019).
4. Blanch-Hartigan, D, et al.: Measuring nonverbal behavior in clinical interactions: A pragmatic guide. Patient Education and Counseling 101, 2209-2218 (2018).
5. Brown, A. B., Elder, J. H.: Communication in autism spectrum disorder : A guide for pediatric nurses. Pediatr Nurs. 40(5), 219–225 (2014).
6. Ambady, N., et al.: Physical therapists' nonverbal communication predicts geriatric patients' health outcomes. Psychology and Aging 17(3), 443–452 (2002).
7. Gable, R.A., et al.: Back to basics: Rules, praise, ignoring, and reprimands revisited. Intervention in School and Clinic 44(4), 195-205 (2009).
8. Begum, M., et al.: Measuring the efficacy of robots in autism therapy: How informative are standard HRI metrics. In: ACM/IEEE International Conference on Human-Robot Interaction, 335-342 (2015).
9. Feng, Y., et al.: A control architecture of robot-assisted intervention for children with autism spectrum disorders. Journal of Robotics, 1-12 (2018).
10. Salvador, M., et al.: Development of an ABA autism intervention delivered by a humanoid robot. In: International Conference on Social Robotics, 551–560 (2016).
11. Centers for Disease Control and Prevention (2020) Data & Statistics on Autism Spectrum Disorder. https://www.cdc.gov/. Accessed Jun 2020
12. Hurt, A. A., et al.: Personality traits associated with occupational "burnout" in ABA therapists. Journal of Applied Research in Intellectual Disabilities 26(4), 299–308 (2013).

13. Louie. W.-Y. G., Korneder, J.A., Abbas, I.: A pilot study for a robot-mediated listening comprehension intervention for children with ASD. In: IEEE International Symposium on Robot and Human Interactive Communication, 1-4 (2020).
14. Google Cloud (2020) Cloud Text-to-Speech - Speech Synthesis. https://cloud.google.com/. Accessed Jun 2020
15. Burroughs, N.F.: A reinvestigation of the relationship of teacher nonverbal immediacy and student compliance-resistance with learning. Communication Education 56(4), 453-475 (2007).
16. Gorawara-Bhat, R., Dethmers, D.L., Cook, M.A.: Physician eye contact and elder patient perceptions of understanding and adherence. Patient Education and Counseling 92, 375-380 (2013).
17. Cevasco, A.M.: Effects of the therapist's nonverbal behavior on participation and affect of individuals with Alzheimer's disease during group music therapy sessions. Journal of Music Therapy 47(3), 282-299 (2010).
18. Yarczower, M., Kilbride, J.E., Beck, A.T.: Changes in nonverbal behavior of therapists and depressed patients during cognitive therapy. Psychological Reports 69(3), 915-919 (1991).
19. Mutlu, B., et al.: Conversational gaze mechanisms for humanlike robots. ACM Transactions on Interactive Intelligent Systems 1(2), 1–33 (2012).
20. Yoon, Y., et al.: Robots learn social skills: End-to-end learning of co-speech gesture generation for humanoid robots. In: 19th IEEE International Conference in Robotics and Automation, 4303-4309 (2019).
21. Feng, W.., et al.: Learn2Smile: Learning non-verbal interaction through observation. In: IEEE International Conference on Intelligent Robots and Systems, 4131–4138 (2017).
22. Randall, A., et al.: Nonverbal behaviour as communication. In: H. Owen (eds.) The Handbook of Communication Skills, 73–119. Routledge, New York (2016).
23. Woodall, W.G., Burgoon, J.K.: The effects of nonverbal synchrony on message comprehension and persuasiveness. Journal of Nonverbal Behavior 5(4), 07–223 (1981).
24. Eikeseth, S., et al.: Intensive behavioral treatment at school for 4-to-7-year-old children with autism. Behavior Modification 26, 49–68 (2002).
25. Lovaas, O. I.: Behavioral treatment and normal educational and intellectual functioning in young autistic children. Journal of Consulting and Clinical Psychology 55, 3–9 (1987).
26. National Autism Center: National standards project: Findings and conclusions. NAC, Randolph (2009).
27. Claude, S. E.: A mathematical theory of communication. Bell System Technical Journal 27(3), 379-423 (1948).
28. Boersma, P., Weenink, D.: Praat: doing phonetics by computer. http://www.praat.org/, Amsterdam (2020).
29. Quené, H., Van Den Bergh, H.: On multi-level modeling of data from repeated measures designs: A tutorial. Speech Commun. 43(1–2), 103–121 (2004).
30. Barr, D. J., et al.: Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language 68(3), 255–278 (2013).
31. Vogeley, K., Bente, G.: Artificial humans': Psychology and neuroscience perspectives on embodiment and nonverbal communication. Neural Networks 23, 1077–1090 (2010).
32. Sigler, E. A., Aamidor, S.: From positive reinforcement to positive behaviors: An everyday guide for the practitioner. Early Childhood Education Journal 32(4), 249-253 (2005).
33. Quam, C., Swingley, D.: Development in children's interpretation of pitch cues to emotions. Child Development 83(1), 236–250 (2012).
34. Fosler-Lussier, E., Morgan, N.: Effects of speaking rate and word frequency on pronunciations in convertional speech. Speech Communication 29(2-4), 137-158 (1999).