Development and Evaluation of a Quantitative Analysis System for Robot-assisted Musical Therapy

Jia-You Lin, Yi-Hsiang Ma, Yeann Han, Sarah Cosentino, Atsuo Takanishi*

Autistic Spectrum Disorder (ASD) is a developmental disorder which impedes the normal development of social and behavioral skills and causes the affected subject to be anxious and unable to interact and communicate with others. Among the various forms of therapeutic approaches, music therapy is a commonly used method for improving their social skills. In our work, we developed a quantitative analysis system for accurate and timely feedback from the subjects, and aimed to create a high adaptability music therapy session with the use of the Waseda Anthropomorphic Saxophonist Robot No.5 (WAS-5). In the session, subjects are asked to perform a repetitive gesture to match the tempo of music played by WAS-5. Through an image processing algorithm for getting the human skeleton position, the tempo of the rhythmic movement can be calculated. WAS-5 will update the tempo accordingly to the feedback so that the subject can match their movement to the music tempo more easily. With the system, we expect to apply the robotic and motion capture technology in the field of psychology and child development which will allow for an adaptive and individually tailored therapy program with quantised evaluation standard.

I. Introduction

Autism spectrum disorder (ASD), as defined by the International Statistical Classification of Diseases and Related Health Problems 10th edition (ICD-10) published by WHO, is defined to be a neurodevelopmental disorder beginning in childhood and marked by the presence of remarkably abnormal social behavior such as grossly impaired communication and bizarre motor responses. About 1 in 59 children have been identified with ASD according to surveys from 11 Autism and Developmental Disabilities Monitoring Network sites. Although symptoms differ considerably from individual to individual, research by Lord and Jones indicates they can be categorised into two core dimensions: (1) social communication deficits on a community-level and (2) restricted/repetitive behaviors on an individual level. The restricted social repertoire and obsession behavior of children with ASD often leaves them frustrated in

DOI: 10.21552/delphi/2019/1/8

* Jia-You Lin, Yi-Hsiang Ma and Yeann Han are students at the Department of Integrative Bioscience and Biomedical Engineering, Waseda University, Japan. For correspondence: erin@huj.waseda.jp
Sarah Cosentino is an Associate Professor at the Department of Modern Mechanical Engineering, Waseda University, Japan. For correspondence: contact@takanishi.mech.waseda.ac.jp
Atsuo Takanishi is a Professor at the Department of Modern Mechanical Engineering and a core member of Human Robotics Institute (HRI), Waseda University, Japan. For correspondence: contact@takanishi.mech.waseda.ac.jp

The authors would like to express their gratitude to the Directorate General for Cultural Promotion and Cooperation at the Italian Ministry of Foreign Affairs for its support to RoboCasa. The authors would also like to express their gratitude to Tokyo Women’s Medical University, Waseda University Joint Institution for Advanced Biomedical Sciences (IWims), Waseda University Program for Leading Graduate Schools, STMicroelectronics, Life Performance Research, Humanoid Robotics Institute (HRI), and SolidWorks for their support to this work.

3 Catherine Lord and Rebecca M Jones, ‘Re-Thinking the Classification of Autism Spectrum Disorders’ (2012) 53 Journal of child psychology and psychiatry, and allied disciplines 490
social interactions which can result in disharmonies inside a social group. Additionally, they encounter difficulties understanding and responding to others’ emotions and behavior.

Currently there is no cure for the condition, however, some treatments have been proven successful in controlling symptoms. Music therapy, a commonly used treatment, is defined by Kenneth Bruscia as ‘a systematic process of intervention wherein the therapist helps the client to promote health, using music experiences and the relationships that develop through them as dynamic forces of change’. The music interaction involved in the therapy is a process using a non-verbal musical language to enhance the functional verbal production and the non-verbal social communication skills. In a music therapy practice and assessment session, clients, together with a therapist, will use musical exercises such as singing, playing instruments and moving their body to spontaneously improvise as a primary therapeutic experience. It helps the clients to link music stimulation with the advanced social interaction and even verbal instructions given by the therapist.

In the research of Edgerton the effectiveness of improvisational music therapy was verified by measuring the musical/nonmusical communicative performances of 11 autistic children aged between 6 and 9. Whipple has also established the significance of music therapy by mutually comparing 9 quantitative studies on treatment with or without music included. Among the studies, they selected different social

---

6 Bruscia (n 6)
aspects to verify the practicality of various types of therapy with different music selection. O’Loughlin\textsuperscript{9} investigated pre-linguistic communication behavior including intentions of eye contact or verbal action like singing and talking from the use of language-based songs. Litchman\textsuperscript{10} discovered the increase of vocabulary acquisition by the use of music as a background sound in learning environment.

Under the continual stimulation of music, a slight progress in learning patterns in musical frameworks will then spontaneously develop variability in dynamics, tempo, duration and accentuation.\textsuperscript{11} However, although many researchers focused on using quantitative analyses to verify the effect of different kinds of music therapy, the measurement techniques have not been combined to timely change the music during the practicum experiences. During the process of musical interaction, a therapist will constantly adapt the musical parameters, by continually observing the subjects’ motion and reaction to the music. With the use of a quantitative measurement system to replace the observation parts by the teacher in the interaction sequence, the reaction of children can be accurately and instantly monitored and used to change the musical parameters.

In our work, we developed a system that can be used to monitor repetitive body movements. We mounted the system in a humanoid saxophonist robot that is able to conduct a live performance according to the children’s reaction. We then carried out a preliminary experiment to verify the efficiency of the system. With a robot able to change its performance based on a patient’s reaction, there will be two expected benefits. First, with musical robots being used to reproduce the training sessions as human therapists, the insufficiency of trained therapists can be solved. Specifically, the symptoms of ASD vary a lot and require an individual treatment approach which causes problems in the training of qualified therapists. Musical robots with the ability to monitor a patient’s feedback, in the background, are therefore suited as human therapists. Secondly, research indicates that therapy conducted by robots is more effective as the ASD patients seemed to be less intimidated and anxious during their interaction with robots.\textsuperscript{12}

With the system, we expect to apply robotic and motion capture technology in the field of psychology and child development, and allow for an adaptive individually tailored therapy program with quantised evaluation standard.

\section*{II. Materials and Methods}

\subsection*{1. Hardware}

Robot assisted therapy (RAT) with autistic children has been applied in actual therapeutic sessions. Deficits in social interaction make it difficult for ASD individuals to fit in the community and reduce the chance of learning and communicating with others.\textsuperscript{13}

Therefore, recurrent interventions on therapeutic methods focus on the uses of life-like dolls or puppets to provide long-term and continual companionship. In addition, compared to interactions with human therapists, subjects are less intimidated and anxious during these interaction and can thus obtain a better learning experience.\textsuperscript{14}

The Waseda Anthropomorphic Saxophonist robot version 5 (WAS-5) has been in development since 2007. With 31 degrees of freedom (DoF) as shown in Figure 1, it can give live musical performances with a standard saxophone.

Mechanisms such as lungs and fingers allow it to produce air flow into the instrument and manipulate the length of the air column. Parts like lips, oral cavity, and tongue reproduce human’s complex muscles manipulation around a mouthpiece and reed with combination of hand and soft materials. In addition, with a waist, a gaze, eyebrows and eyelids the WAS robot is capable of mimicking a saxophonist’s body movements and facial expression. With this, we can
combine the benefits of robot and music therapy in the treatment of children with ASD, using our interactive humanoid saxophonist robot WAS-5.

2. Music Therapy Scenario

Tempo, as one of the musical elements, has been widely used as a controlled variable in music therapy since a rhythm-based section can be easily adapted for different songs and instruments. In addition, by reproducing rhythmic patterns, there are several expected effects on the social skills of the children, including on awareness, attention, empathic synchronicity and relationship building. The subject was asked to sit on a chair with a camera placed at a viewing distance of 50 [cm] and three markers were placed on the floor (Figure 2). The distance between each marker was set at the subject’s shoulder width which we measured in advance. The subjects were instructed to place both feet on the two markers on the left. And as the session began, the subject was asked to fix his left foot and move his right foot between the two markers on the right according to the tempo at which the WAS-5 robot performed.

The musical synchronisation process is concerned with the metrical level at which people tend to synchronise their own taps. For example, when listening to a musical tempo of 120 [bpm], one cannot only synchronise his movement with 120 [bpm], but also with half (60 [bpm]) or twice the tempo (240 [bpm]). In order to reduce the confusion of which tempo to follow, we focused only on the development and verification of the system, thus, the robot would only play two notes cyclically instead of an actual musical phrase with an explicit tempo. Specifically, the robot played the notes C4 and E4 in turn with different time intervals creating an obvious tempo variation. Along with the sound, the subject was required to step on the adjacent markers in line with the beat.

Researchers presented that humans prefer movement with tempo that is close to a neutral or spontaneous tempo, which is associated with natural body movement. There are many spontaneous rhythmic movements in daily life, for example, walking, applauding, heartbeat, and even tapping along to music. It is observed that a tempo close to the biological rhythms of the human body are thought to be com-

---


forting. In addition, Moelants concluded that there exists a correlation between the tempos of these kinds of spontaneous movements. Based on the above mentioned reasons, the tempo we used in the experiment was between 60[bpms] (tempo of heart beat) and 120[bpms] (tempo of walking).

3. Software

In Figure 3, the information and signal flow, hardware and software processing are presented. When listening to the music performed by the robot, the subject would move his right foot according to the tempo. At first, the tempo of music that the robot plays \( T_c \) was assigned as an initial value \( T_R \) between 60 to 120[bpms]. \( T_R \) here, represents the target tempo we expected the subject to catch up with. The images would then be captured by the camera at 60[fps] and at a resolution of 640 x 480 [pixels]. Each image was first analysed by OpenPose, a library authored by Gines Hidalgo for real-time human body key points detection, to calculate the distance between both feet. When the subject moves his foot in line with the tempo, it is represented in a ‘wave crest’ which mirrors the motion of stepping on the markers. Using the time difference between signal peaks of feet distance, the tempo of the actual movement \( T_M \) was measured. According to the subject’s motion tempo, the robot would perform the music differently. The tempo the robot plays at \( T_c \) depends on two parameters: \( T_R \) (the target tempo we expected subject to catch up with) and the control load \( u \) (a value to report how well the subject catches up with the target tempo). Two different control strategies were applied by giving different control load \( u \), and \( T_c \) of the next loop will be update as the summation of target tempo and previous control load.

\[
T_c = T_R + u
\]

(a) Constant strategy: the tempo the robot plays at will only be affected by the target tempo and updated according to the subject’s motion. That is, the robot will only play music without observing and getting feedback from subjects’ motion.

\[
u = 0
\]

(b) Leader strategy: the tempo the robot performs at will be a function of the subject’s motion tempo in order to lead the subject to the accurate tempo.

---

17 Theo Willemze, Prisma Praktisch Muiziebroek (Het Spectrum 1956)
19 Paul Fraisse, ‘Rhythm and Tempo’ (1982) 1 The Psychology of Music 149
It means that the robot will change its music according to the subject’s motion. For a more intuitive example, in order to accelerate a subject to move in a certain target tempo, the robot will constantly observe the difference between the target tempo and subject’s current movement tempo. If the difference is large meaning that the subject is not following, it will raise to a higher tempo, making the change more obvious to the subject. If the difference becomes smaller meaning that the subject is catching up, the music will slow down gradually to smoothen the process of following up the tempo. This kind of concept is commonly used in automatic process control applications in industry since it can regulate industrial process variables with acceptable performance in terms of disturbance rejection and avoid sudden termination. Here, we proposed a hypothesis assuming that the human can be seen as a plant (often used to refer to the physical properties of hardware in the system) in a close loop control system. The relation between input and output was given as a proportional–integral (PI) controller (a commonly used control strategy in industrial control systems). By a trial and error based of parameters adjustment, both $K_p$ and $K_i$ is assigned as 0.05. Note that parameter ‘$e$’ represents the difference between $T_R$ and $T_M$.

$$e = T_R - T_M$$  \hspace{1cm} (3)

The control load can be shown as

$$u = K_p e + K_i \int e dt$$  \hspace{1cm} (4)

Controlled load ‘$u$’ is sent to the robot with a TCP/IP protocol suite to update the time intervals of the next sound.

4. Experiment Protocol

We carried out an experiment to investigate the difference between the two control strategies. In the experiment, our main aim was to estimate the system properties before actually using it in a music therapy session. Therefore, we conducted it on one typical adult with a relatively long estimation time for each stage to exclude the effects of the differences between individuals. For each strategy, there were four stages with the initial value $T_R$ set as 60, 80, 100, 120 [bpm] in order. The music therapy scenario mentioned above was conducted with the subject catching up with the tempo while listening to the music played by the robot. Meanwhile, the tempo which the WAS-5 robot played remained the same for the constant strategy, and updated constantly according to the feedback for the leader strategy. Each stage lasted for 100 [s] before changing to the next $T_R$. Tempo of $T_R$, $T_M$, and $T_C$ were recorded at a sample rate of 60 [Hz].

III. Result

![Figure 4: Tempo as function of time for constant strategy](image)

*Source: Author’s elaboration*

![Figure 5: Tempo as function of time for leader strategy](image)

*Source: Author’s elaboration*

The result of the three time-varying tempos, $T_R$, $T_M$, and $T_C$ of both strategies are shown in Figure 4 and Figure 5. With solid lines connecting the data points,
the difference between the two methods can be observed. There existed a small delay of about 4.3 [s] in both of the strategies, which was caused by the delay when calculating the subject’s tempo. In the software part of the system (see gray box in Fig. 3), both the OpenPose and control strategy parts are given in real-time. On the other hand, for the part of tempo detection, there is a trade-off between reaction time and accuracy. To measure the actual tempo of a subject’s movement, excluding the effect of noise, we chose to sacrifice immediacy in favour of better accuracy.

In regard to the constant strategy, although the measured subject’s tempo $T_{M}$ is close to the given reference tempo $T_{R}$, we can still see small oscillation of signal representing the instability of the movement. The result of the leader strategy, by contrast, has a higher error margin at first but gradually stabilises. During the change of $T_{R}$, there is an overshoot (occurrence of a signal exceeding its target), which is also a common situation occurring in general signal processing and electronics field. This implies a possibility of considering a human as a plant with the ability of applying other control theories.

Furthermore, over the 100 [s] measurement time, the subject was observed to have different stability over his movement when the robot was playing at different tempos with different strategies. The error value between the target tempo and subject’s tempo showed how well he can follow the music, which is the accuracy of the movement. The STD in the other hand, presented the stability of one’s movement, which is the precision of the result. Table 1 and Table 2 show the average of $T_{M}$, error between the two tempos and STD of the measured value. It is found that with the use of leader strategy, both the error and STD appears to be smaller, meaning that the subject tends to have a better performance and higher stability following the music tempo when there is a quantitative close loop feedback to the robot during the musical interaction. With the system using robotics technology and a psychology-based music therapy scenario, a learning process with higher efficiency is expected to be done.

Discussion and Future Work

This work focuses on developing a quantitative system that is able to conduct music therapy using a humanoid saxophonist robot. The evaluation experiment was also conducted only to test the system and find out the efficiency of the control strategies we designed. However, there are many control parameters with the possibility to affect the music therapy which still remain to be investigated. Using this specific system, we can conduct experiments to find out the effectiveness of those parameters and optimise the

---

**Table 1: Error of the tempo detection system (constant strategy)**

<table>
<thead>
<tr>
<th>$T_{R}$</th>
<th>$T_{M}$</th>
<th>Error [%]</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>60.4</td>
<td>1.4</td>
<td>0.7</td>
</tr>
<tr>
<td>80</td>
<td>80.0</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>100</td>
<td>100.5</td>
<td>1.4</td>
<td>1.8</td>
</tr>
<tr>
<td>120</td>
<td>119.7</td>
<td>0.9</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Table 2: Error of the tempo detection system (leader strategy)**

<table>
<thead>
<tr>
<th>$T_{R}$</th>
<th>$T_{M}$</th>
<th>Error [%]</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>60.0</td>
<td>0.9</td>
<td>0.3</td>
</tr>
<tr>
<td>80</td>
<td>80.1</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>100</td>
<td>100.2</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>120</td>
<td>120.1</td>
<td>1.1</td>
<td>1.0</td>
</tr>
</tbody>
</table>


therapy session-based on the results. For example, as the robot only performed at different tempos (by playing C4 and E4 in turn), it would be interesting to investigate the difference when the robot plays different notes or even a short musical phase. Besides, the use of different musical parameters (pitch, volume, etc) can also be analysed. As the next step of our research, we would like to conduct actual music therapy sessions on ASD children. From the feedback of sensor’s data, we can monitor subjects’ reaction and study the difference in their movement from typical subjects. With the use of the system, the music therapy sessions become measurable, repeatable, and are expected to facilitate the researchers’ work. Furthermore, children with ASD are provided with a better treatment environment.

The currently used traditional musical therapy may cause several constraints on the lives of the patient and the family since it should be conducted regularly by a well-trained therapist who has a comprehensive understanding of the child’s current condition in order to plan an optimised session. On the other hand, the use of the WAS-5 robot, which can monitor and save physiological and interaction data for later assessment analysis, will allow therapy sessions to be more flexible, maximising the effectiveness and feasibility of continuous therapy. In the future, by applying the robot as a personalised device in the daily rehabilitation of ASD children, there is a possibility of creating new therapeutic methods with the advantage of preserving a long-term database, utilising high level adaptive functions and providing uninterrupted supervision. However, there are still limitations when employing a robot therapist. In spite of the scientific evidence showing there is no safety concern, perception and opinions of end-users are a very different matter. There are different factors for determining the psychological safety of using the robot\textsuperscript{22} that may affect patients’ willingness to do so. Most people still consider it ethically unacceptable to completely replace a human therapist by a robot, but do approve of using a robot to support the interaction between a human therapist and a child.\textsuperscript{23} This kind of application, which is called supervised autonomy is considered more likely to be accepted by the ASD children and their families. Its limitations should be addressed in further research and development.

In spite of the limitations, however, we hope the system we proposed in this paper has contributed to developing a more efficient and effective method for treating children with ASD symptoms. We also expect that combining robotics with music therapy will provide continuous and effective social training, helping children with ASD to conduct a constructive and independent life.

\textsuperscript{22} Hiroko Kamide and others, ‘New Measurement of Psychological Safety for Humanoid’, Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (ACM 2012)

\textsuperscript{23} Mark Coeckelbergh and others, ‘A Survey of Expectations about the Role of Robots in Robot-Assisted Therapy for Children with ASD: Ethical Acceptability, Trust, Sociability, Appearance, and Attachment’ (2016) 22 Science and Engineering Ethics 47