**towards ai-assisted music generation for mood enhancement in individuals with autism spectrum disorder (ASD)**

DESN3002 – Final Report

AU, Catherine Rose

3035920170

A close-up of a logo

Description automatically generated

**Table of Contents:**

Table of Contents…………….…………….…………….…………….…………….………1

Background/Research…………….…………….…………….…………….………………2

Workflow/Pipeline…………….…………….…………….…………….…………….……..9

Prompts…………….…………….…………….…………….…………….………………...9

Method of Visualization…………….…………….………………………………………..11

Music Generation…………….…………….…………….…………….…………….…….14

Comparison + Feedback…………….…………….…………….…………….…………..22

Limitations…………….…………….…………….…………….…………….…………….25

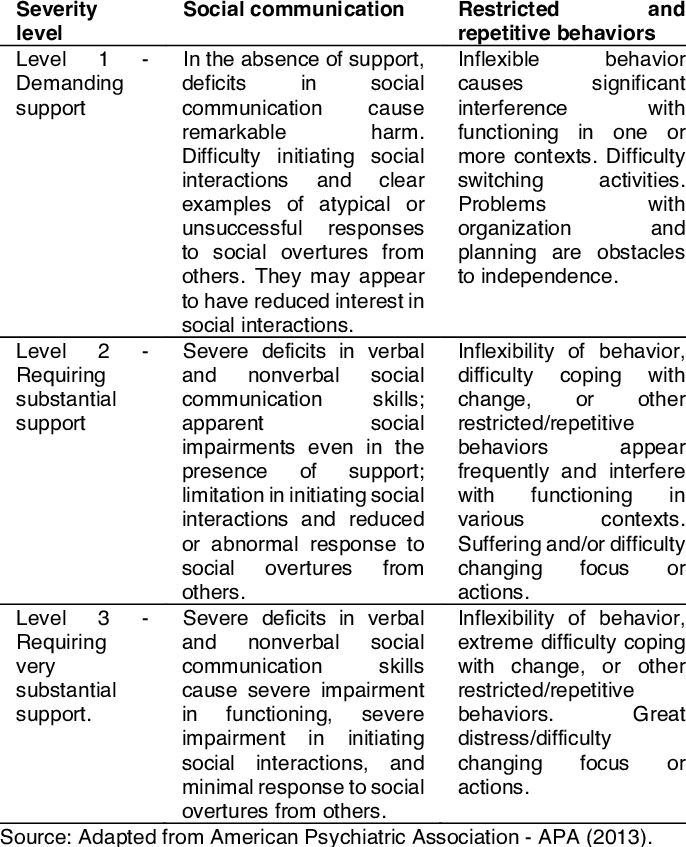
Closing…………….…………….…………….…………….…………….………………..25

References…………….…………….…………….…………….…………….…………...26

**Background/Research**

The prevalence of individuals with Autism Spectrum Disorder (ASD) has drastically increased since the 1990s. A study published in CDC’s Morbidity and Mortality Weekly Report (MMWR) found that one in 36 (2.8%) 8-year-old children are diagnosed with Autism (Centers for Disease Control and Prevention, 2023). In the years 2006 and 2021, the same study was carried out and was found that the ratio was one in 41 and one in 110 respectively (Johnson, 2023).

Autism Spectrum Disorder (ASD) is a developmental disability that can cause significant social, communication and behavioural challenges (Centers for Disease Control and Prevention, n.d.). This can affect them in lots of their daily activities, whether it’s at school, at home or at work. Additionally, there are different levels of severity an individual can have of ASD which affects how they function on a day-to-day basis. The severity is divided into three levels that address the individual’s sociability and repetitive patterns/tendencies.



*Figure 1 (Sabatini, Cobus, & Ito, 2023)*

One common experience a person with autism spectrum disorder is sensory sensitivities, it may be hypersensitivity or hyposensitivity. As the name suggests, this affects their sensory processing abilities, sight, sound, smell, taste and touch.

This study will look deeper into the auditory sensitivity, also known as, hyperacusis (Cross River Therapy, n.d.), that individuals with autism spectrum disorder experience. Up to 70% of ASD individuals’ daily lives will be affected by this (Cross River Therapy, n.d.). High-pitched ringing and sudden loud sounds are typical examples of what would stress and overwhelm an ASD individual. If the source of the problem fails to be controlled, it can cause their emotions to become stronger. The stress and overwhelming feelings can develop into anxiety, frustration, or even physical pain too.

When in a stressed state of mind, there have been some specialists that prescribe a duration of Music Therapy in hopes to calm the individual and relieve any of their anxiousness. Autistic individuals are able to develop a high level of concentration and calmness when listening to the appropriate music (Overcome with Us, n.d.). They may even become more attentive and focused on the current task they are carrying out and achieve more than they could without music.

Studies have been done to test the efficacy of Music Therapy on persons with autism spectrum disorder. They look closely at how it affects not only their mood and behaviour, but also their verbal and non-verbal skills, ability to concentrate on tasks and sleep quality. Music Therapy is a clinical and evidence-based use of music interventions (American Music Therapy Association, 2012) that is personalized for the listener to achieve their goals. The benefit of using music as a medium of a therapeutic activity for an individual with autism spectrum disorder, is that music is very malleable and easily manipulated to best suit the listener. Additionally, it is a non-invasive and non-threatening approach in trying to connect and form a relationship and level of trust with the ASD individual.

Music can be processed in both hemispheres of the brain (Wan et al., 2010) which allow for increased cognitive function. Language and verbal skills have also been found to improve (Koul & Shalev, 2022) once beginning music therapy treatment. For those individuals who are non-verbal, starting music therapy has prompted them to begin humming or singing to the music when played (Bhatia & Tandon, 2021).

A white paper with black text

Description automatically generated

A white paper with black text

Description automatically generated

*Figure 2 (Bhatia & Tandon, 2021)*

Figure 2 above demonstrates the relevance and currentness of music therapy being a therapeutic approach for individuals with autism spectrum disorder. It summarizes ten recent studies done that assess and observe the effects music therapy has on individuals of all different severities of autism.

Several of the studies had very encouraging outcomes suggesting that the participants’ mood and ability to better express their emotions increased (Shi, Lin, & Xie, 2016). However, many of the studies used excerpts of existing music that they believe would work well together to enhance the mood of the individual. This raises the question of whether or not having music that is composed solely for that one individual, could increase the ability of their mood enhancing even further. This would require for there to be prompts that describe what the music should sound like regarding elements such as tempo, genre, volume, mood and more.

Aside from the typical methods of composing music in a studio with instruments and vocals, there are more and more people experimenting with using Artificial Intelligence (AI) and Machine Learning to generate music and audio clips. Facebook for example, is experimenting with using AI to generate music using text prompts. MusicGen a pre-trained single Language Model (LM) that has been specifically tailored for conditional music generation. Given its well-established name in the industry, Facebook is a reliable source to begin experimenting with conditional music generation (AudioCraft, n.d.). Additionally, they have trained their model using a large dataset of music, therefore, allowing it to identify and apply numerous patterns to their generated music (AudioCraft, n.d.). One key limitation of this method of music generation is that it can only produce a maximum of 10 seconds per audio clip. This obstacle will be addressed later in the discussion.

The code shown Figure 3 below is the opening of the MusicGen model showing which libraries and tools are to be used to generate the audio and additionally, the code used to generate the playable sample of the generated music.

A computer screen shot of a computer code

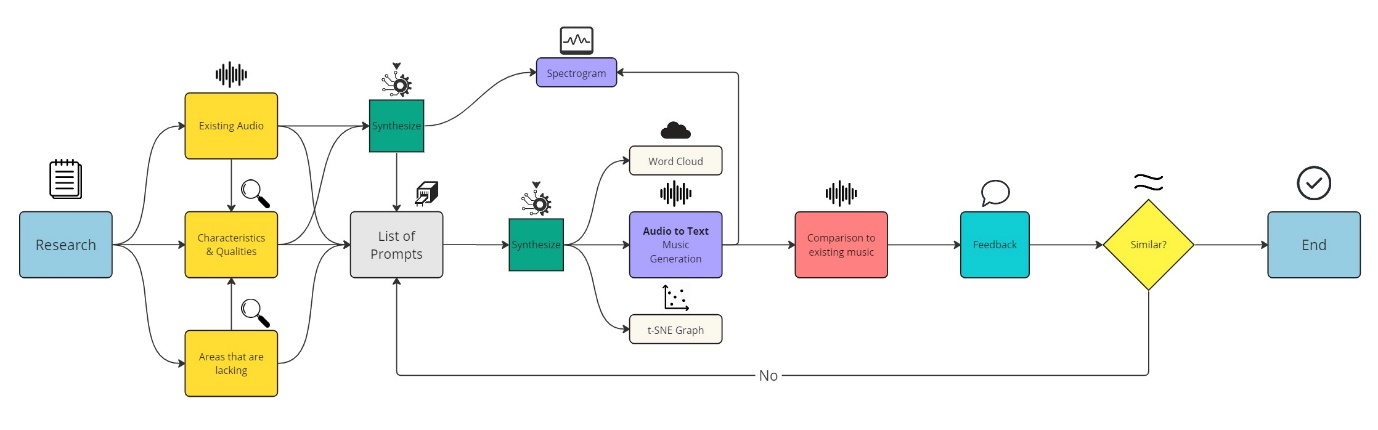
Description automatically generated



*Figure 3 (Facebook Research, n.d.)*

**Workflow/Pipeline**

Before generating music, a pipeline must first be established to ensure that there are no gaps in the process and that reproducibility can be possible. Figure X below shows the pipeline that will be carried out throughout the experimenting.



*Figure 4*

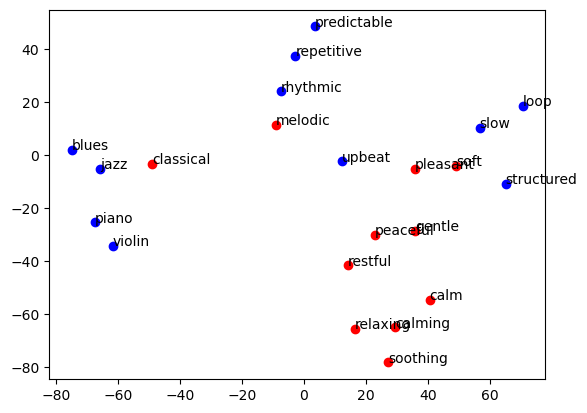
**Prompts**

With the research section completed, the next step is to compile a list of the adjectives that are used to describe the audio used in previous music therapy applications. This will provide a reliable foundation of the type of music that may be generated in the future as it has been proven to be effective for mood enhancement for individuals with autism spectrum disorder. The studies mentioned above are first manually analyzed to extract the key characteristics and qualities of the audio played for the individuals receiving music therapy. An additional layer of analysis was done by utilizing ChatGPT to further filter out the most commonly occurring qualities and characteristics of the music used in the studies. Both lists were then collated and processed into a Word Cloud that depicts the frequency of each characteristic in the studies, with the size of the word corresponding to how many times it occurred during the analysis process.



*Figure 5*

Besides generating a Word Cloud, the use of a T-distributed Stochastic Neighbour Embedding (t-SNE) graph was also adopted. This helps to dimensionally reduce the data and allow it to be plotted based on its similarities to one another.



*Figure 6*

Existing music and audio clips that are proven to help enhance the mood of individuals was also analyzed. This will aid in identifying whether or not the generated music is similar to the existing one, therefore, meaning it will be effective in enhancing the mood of the individual. Examples of less effective or desired songs were also analyzed to set a benchmark of what the generated music should not look like.

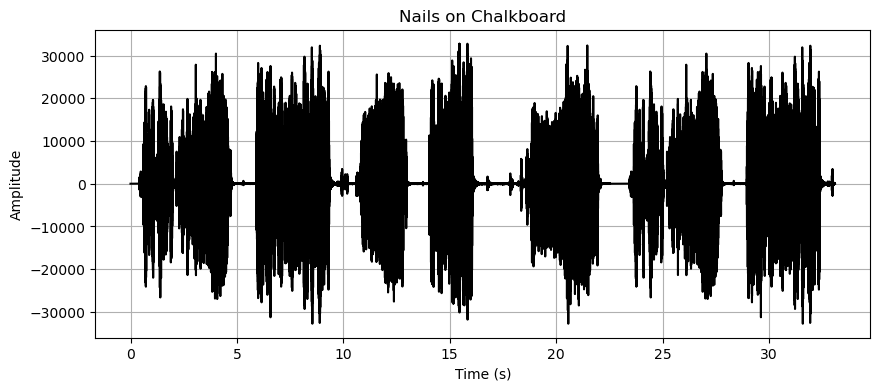
**Method of Visualization**

One method of visualizing audio is by creating an audio waveform. It plots the amplitude against time of the audio and creates a unique wave-like shape depicting the sound. The following code in Figure 7 shows the code used in Jupyter Notebook to generate the waveform. Figures 8 & 9 show this.

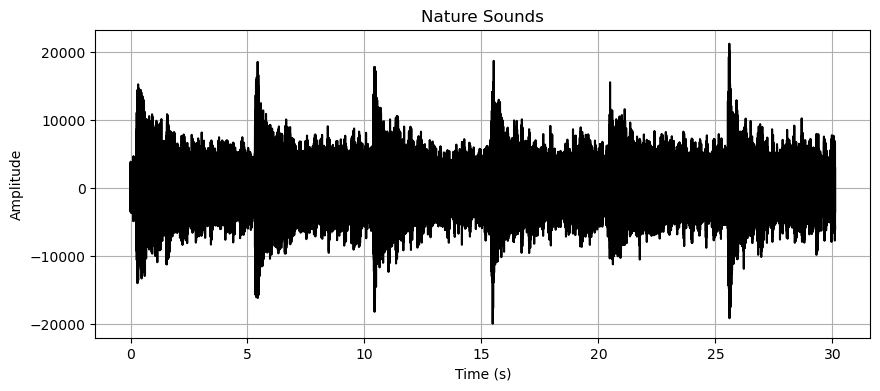
A screenshot of a computer

Description automatically generated

*Figure 7*



*Figure 8*



*Figure 9*

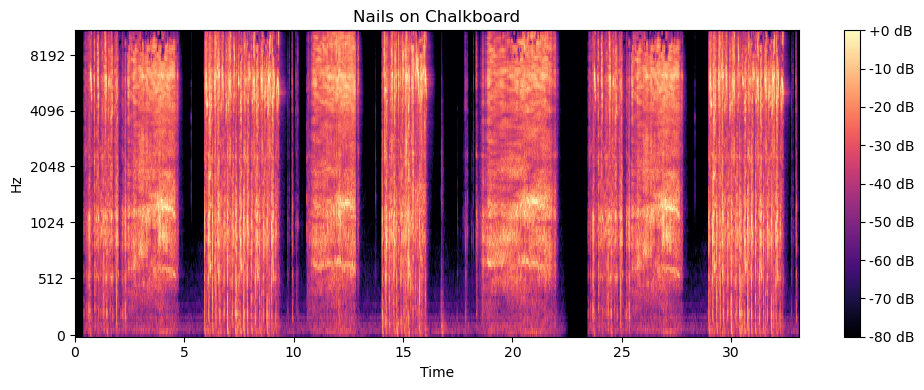
However, the audio wave forms are not able to highlight vast differences or patterns in the audio. A more effective method would be to utilize spectrograms. A spectrogram is able to visualize an audio clip and depict its time, frequency and amplitude all in one graph ("Spectrogram," n.d.). Spectrograms can be easily generated using Python or Jupyter Notebook. The way one would read a spectrogram is to look at the colours on the legend on the side and compare them to the ones generated in the graph. Typically, the darker the colour, the quieter the audio is. On the other hand, the brighter the colour is, i.e.: more yellow orange, the louder and sharper, the sound is. In order to load the audio into the notebook, the Python package, librosa, was used. It is used for music and audio analysis (McFee et al., 2020) and is able to isolate and extract key elements of an audio that could be needed for a script.

A close-up of a computer code

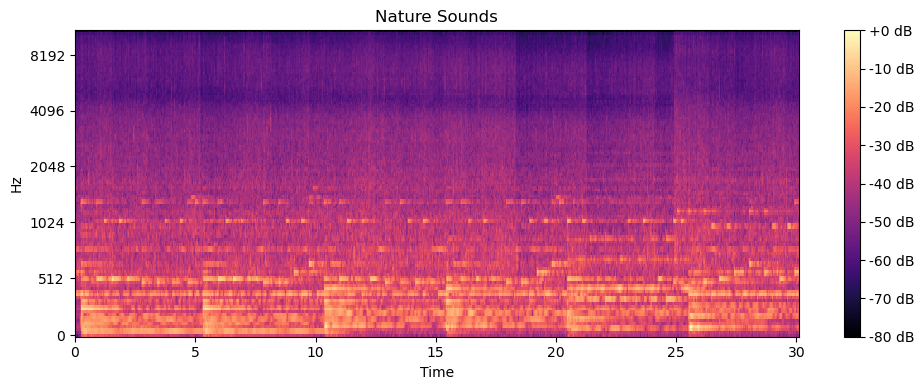
Description automatically generated

*Figure 10*

Once the audio is loaded into the file, the data is collected, and a spectrogram can be created. Spectrograms depicting the same audios used above for the audio waveform can be found below in Figure 11 & 12.



*Figure 11*



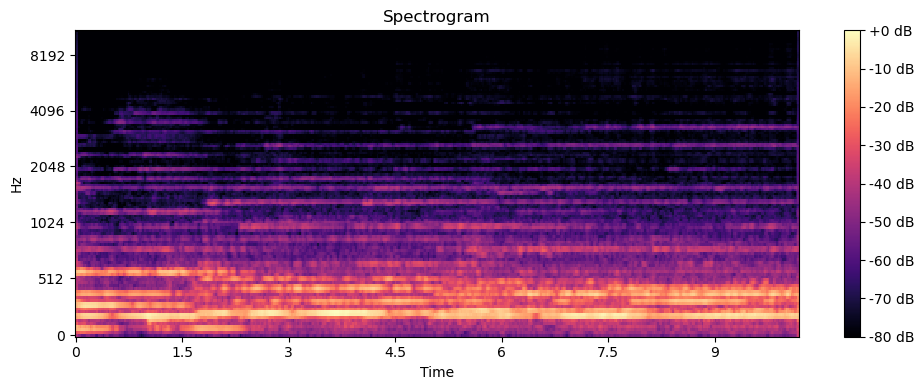
*Figure 12*

**Music Generation**

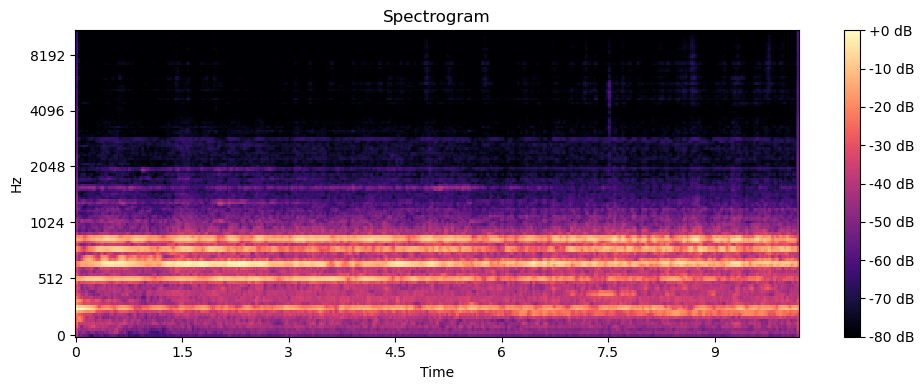
There are thousands of word combinations that can be used to generate music using the key words found in the studies. From now on the characteristics and qualities extracted from the studies will be referred to as prompts. To understand the accuracy and similarity of the model, the number of prompts fed into the model began at two prompts and then increased by one each time. Each combination of prompts was processed three time to ensure that the audios generated each time were similar sounding and that the sound wasn’t matched to the prompt purely by luck or chance. The combinations that produced the most similar sounding sounds to both the prompts given as well as the initial effective audio are:

* *“slow and gentle music”*
* *“classical, slow-paced and repetitive music”*
* *“peaceful, classical, slow and calm music”*

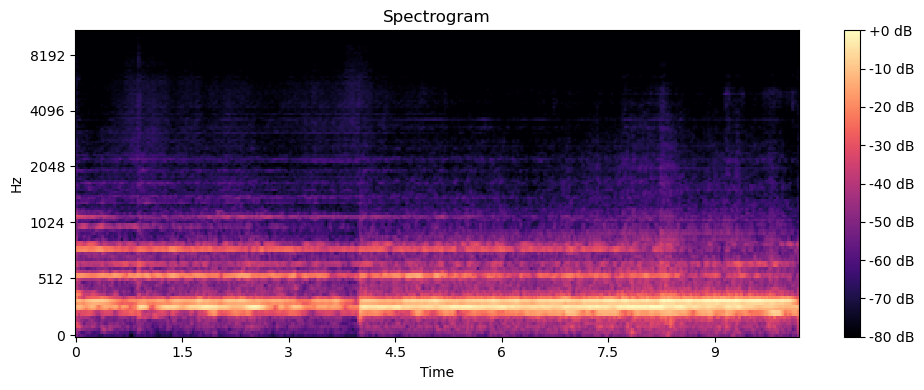
The spectrograms of the audios generated based on the prompts can be found below.



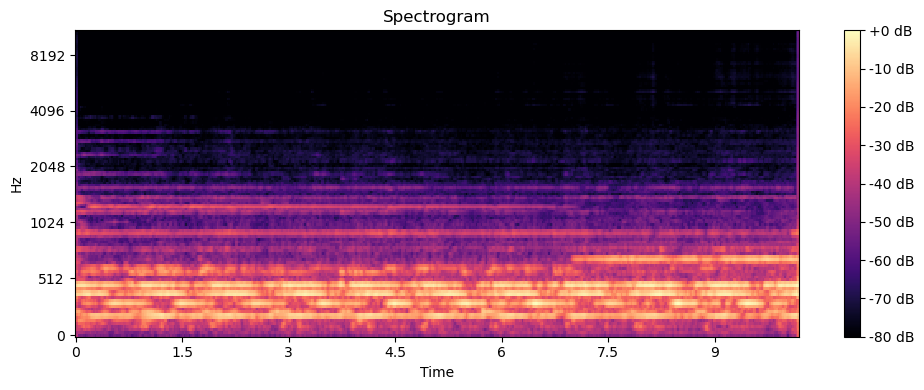
*Figure 13*



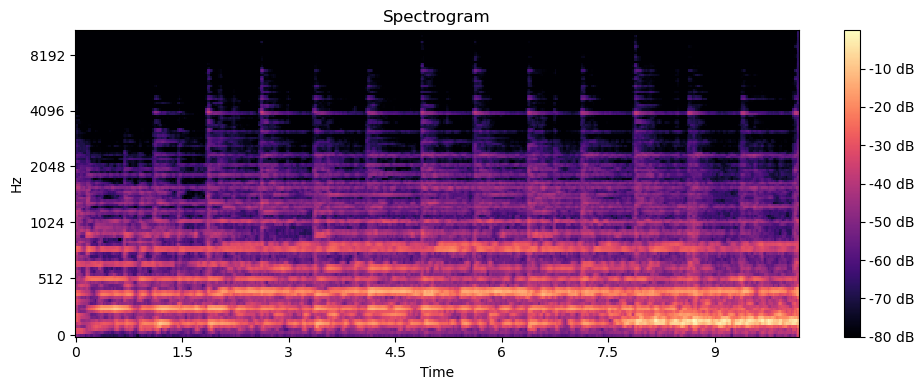
*Figure 14*



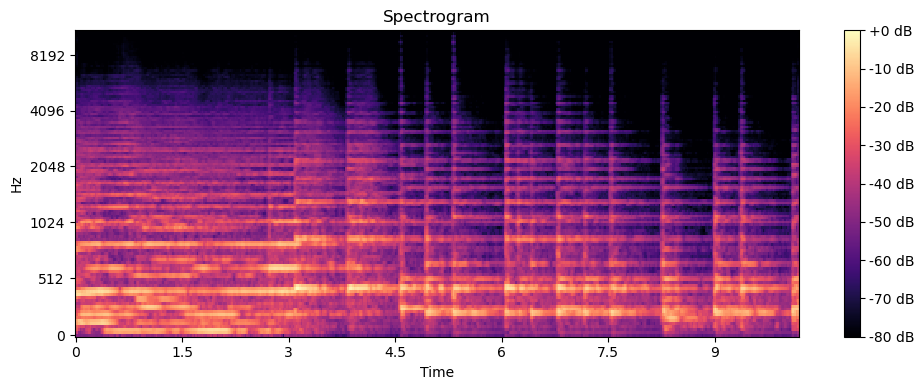
*Figure 15*



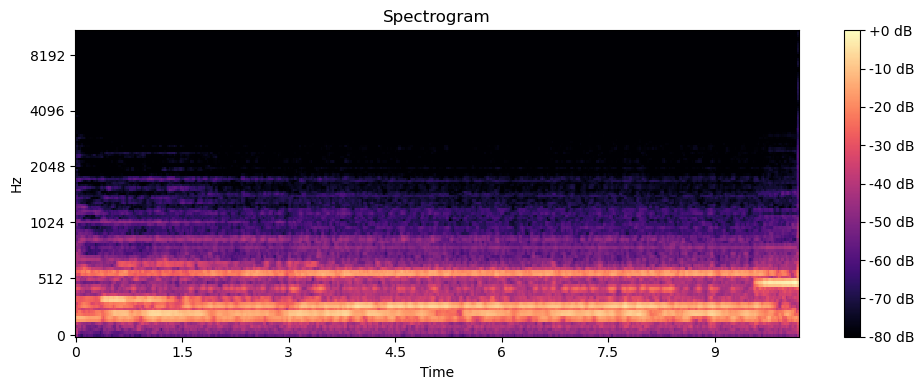
*Figure 16*



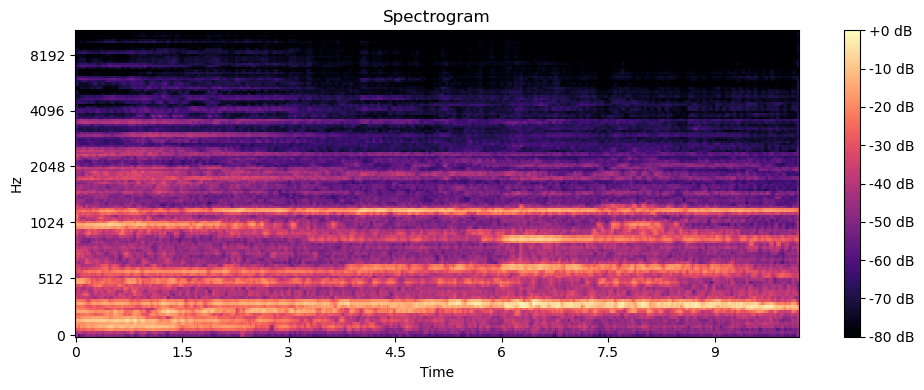
*Figure 17*



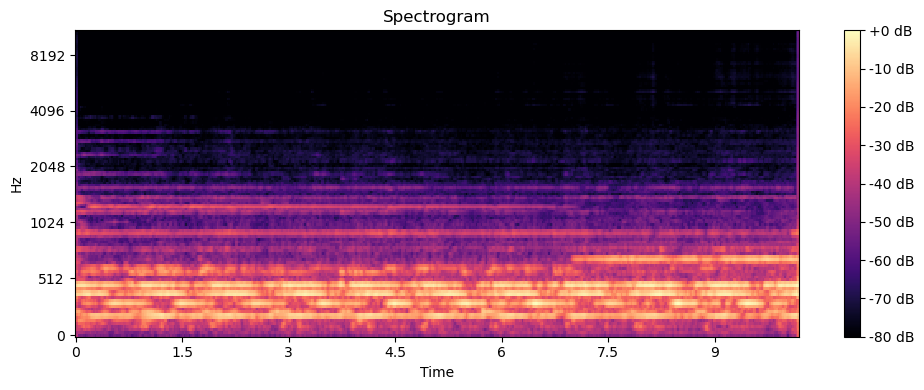
*Figure 18*



*Figure 19*



*Figure 19*



*Figure 20*

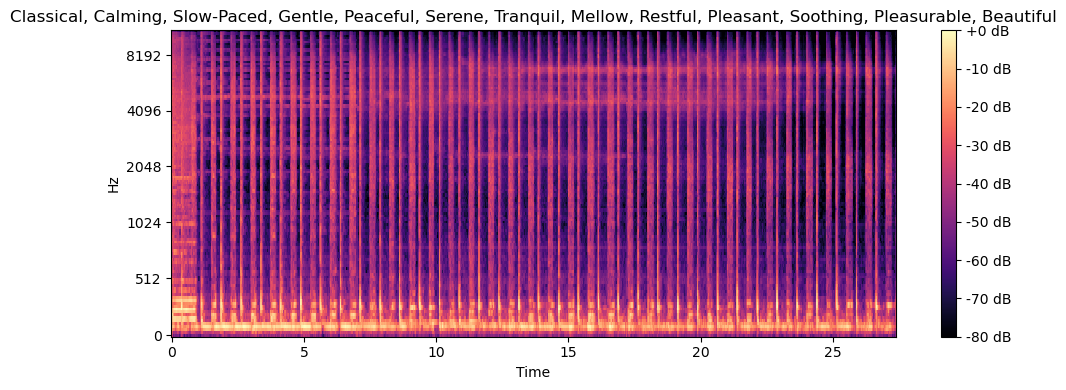
As previously mentioned, the maximum length of audio that could be generated using the MusicGen model is 10 seconds. To enhance the mood of an individual with autism spectrum disorder, an audio clip longer than 10 seconds would be needed. A method that was experimented with is the process of continuation. This allows for the 10 second audio clip to be loaded into the notebook and it will generate an additional 20 seconds based off of the loaded audio as well as another prompt. The code below in Figure 21 shows the Continuation method.

A screenshot of a computer

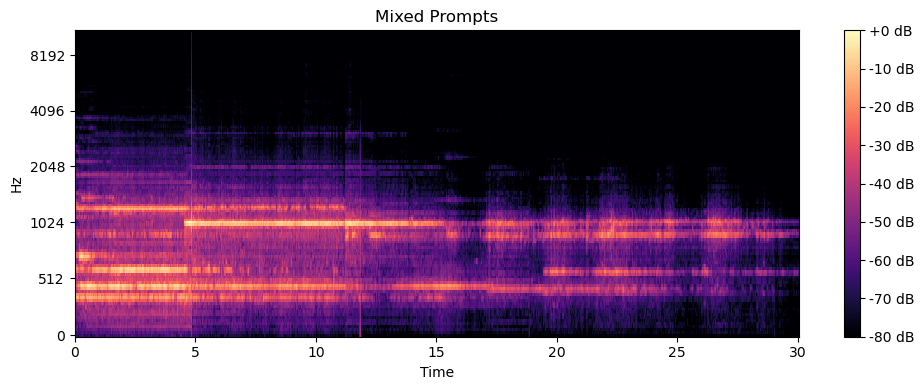
Description automatically generated

*Figure 21*

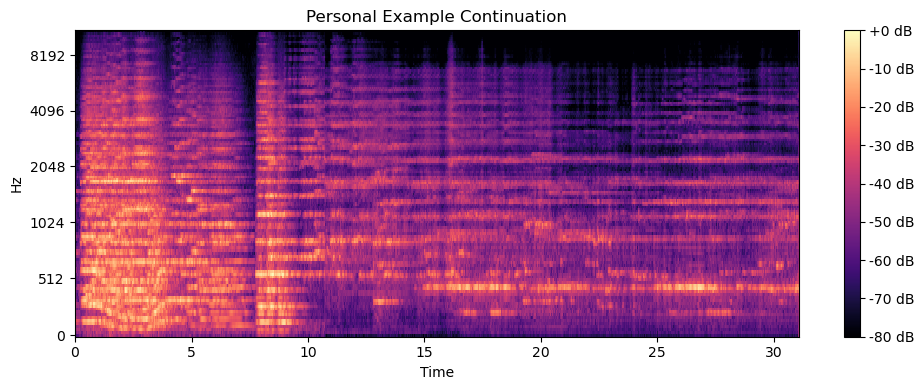
To maximize the capabilities of the continuation process, there were three different variations used. First, a 10 second generated audio clip using the prompts classical, slow paced and repetitive was loaded, and the notebook was given the same prompt as the initial audio. This created a similar sounding audio to the initial one. Second, the same audio file was loaded with librosa and fed prompts that were found in the word cloud but differing from the ones used to prompt the initial audio. Lastly, a 10 second clip from an existing audio from one of the studies was loaded and given the same prompts as the previous continuation. These three tests generated audios that were similar to their initially loaded ones and had hints of the prompts given. The three figures below depict the audios generated using continuation.



*Figure 22*



*Figure 23*

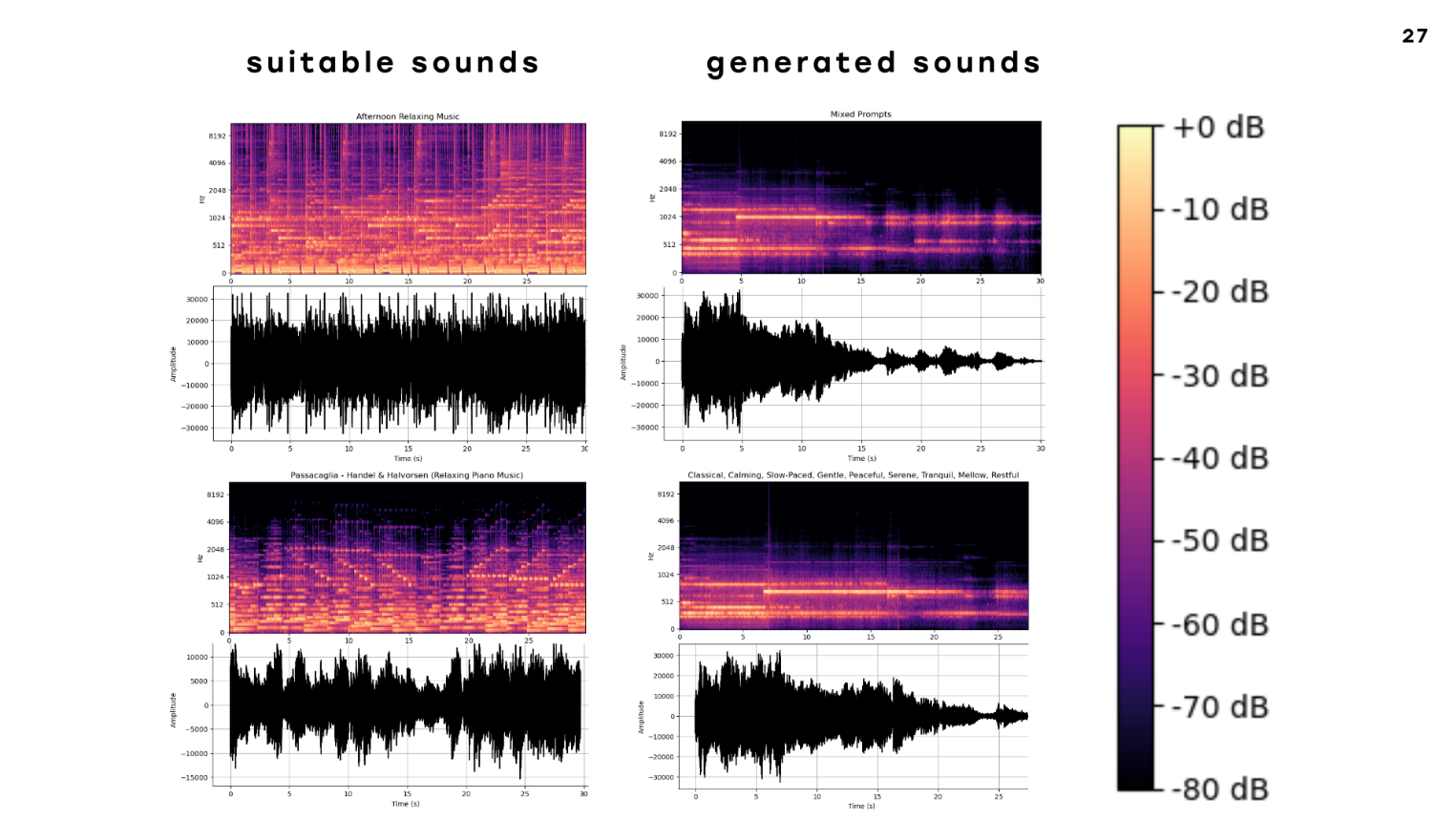


*Figure 24*

**Comparison + Feedback**

To determine whether or not the generated music is similar to the existing music used for music therapy, there will be three methods of analysis: comparisons of spectrograms, UMAPs and user feedback.

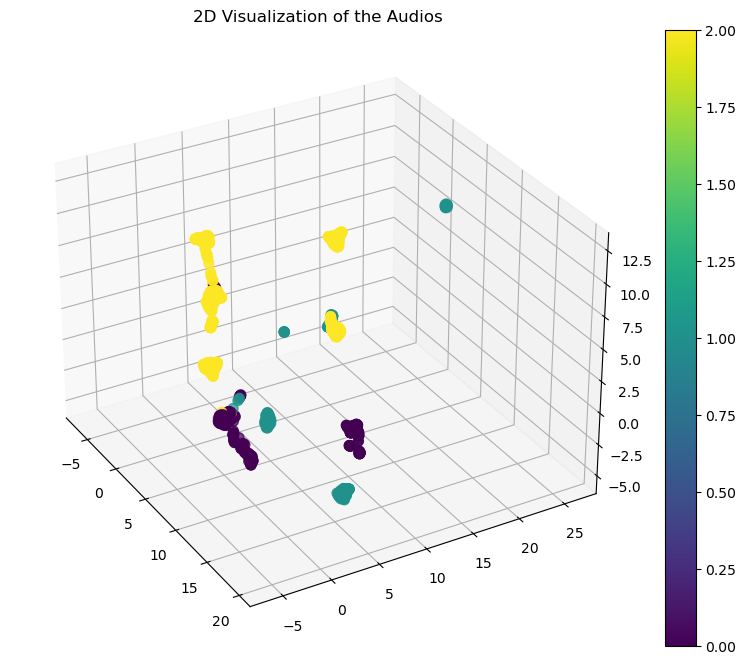
First, the spectrogram and audio waveforms of the existing audios and the generated ones will be compared side by side to identify similarities.



*Figure 25*

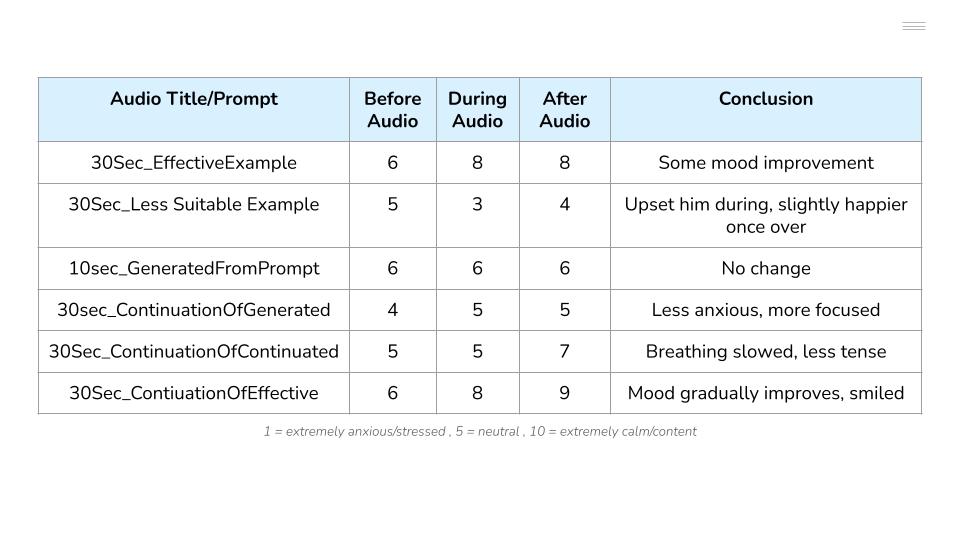
Figure 25 above shows that the spectrograms of the generated music have quite a significantly lower volume compared to the existing audio. Although they do not look that identical, the generated audio seems to in fact be more optimal for calming individuals with autism spectrum disorder. This is because a quieter, less spontaneous audio would be more preferable.

Secondly, a Uniform Manifold Approximation and Projection (UMAP) was plotted by imported the various audio clips. A UMAP is similar to a t-SNE graph, in the sense that they both take data and put it through the process of dimensionality reduction, allowing for the data to be visualized and for patterns and similarities to be highlighted. Below you can see the UMAP plot in Figure 26.



*Figure 26*

Lastly, an individual with autism spectrum disorder was played the various audio samples and their attitudes and actions were observed. This observation is merely to get an initial understanding of whether or not the generated music and its methodology has the potential to develop something much more in depth and complete. The participant was played six different audio clips ranging from 10 to 30 seconds. These included existing audio, generated audio and continued generated audio. Figure X below shows the results.



*Figure 27*

Based on the rating scale mentioned in Figure 27, it is clear that the “30Sec\_EffectiveExamle” was indeed successful at improving the participant’s mood and “30Sec\_LessSuitableExample” worsened the participant’s mood and they were only calmer after the audio had stopped. This allows for the assumption to be made that those two audios are good benchmarks for what an effective and less effective audio would be for enhancing the mood of an individual with autism spectrum disorder. Promisingly, the generated music was received well by the participant as their mood, anxiety and concentration were all positively affected.

**Limitations**

It is important to acknowledge that this exploration was very initial and has room for much greater depth and insight to be found. Some key limitations of the research was the lack of a controlled variable and environment for the user feedback section as well as having more participants and making observations over a longer period of time. Additionally, the maximum length of audio generated was 30 seconds, which is not ideal for the application of music therapy as it is likely a session of music therapy would last much longer, and a longer duration of music would be needed.

**Closing**

Overall, if the methodology can be reproduced multiple times to get concrete data, generating music through the use of Artificial Intelligence and Machine Learning for enhancing the mood of individuals with autism spectrum disorder shows a high level of promise. Aside from supporting individuals with ASD, this process could also be applied to generating music for music therapy for a wide range of applications. Alternatively, there are numerous other pre-trained models available that could generate music using other methods such as an audio to text model.

**Word Count (including titles, headings and in-text citations): 2448**

Link to presentation:

<https://new.express.adobe.com/published/urn:aaid:sc:AP:542e87c1-eff8-4f1f-a71f-5d3bde9871d3?promoid=Y69SGM5H&mv=other>

**References:**

1. Centers for Disease Control and Prevention. (2023, March 23). CDC Press Release: Autism Prevalence Increases to 1 in 34 Children. https://www.cdc.gov/media/releases/2023/p0323-autism.html

2. Johnson, C. Y. (2023, March 23). The 'autism epidemic': Why the new CDC numbers are not surprising. STAT. https://www.statnews.com/2023/03/23/autism-epidemic-cdc-numbers/

3. Centers for Disease Control and Prevention. (n.d.). Autism Spectrum Disorder (ASD). https://www.cdc.gov/autism/index.html

4. Sabatini, B., Cobus, D., & Ito, V. (2023). Sensory sensitivity and food selectivity in children with autism spectrum disorder (ASD): outlines for food therapy. Dataset Reports, 2. https://doi.org/10.58951/dataset.2023.64

5. Cross River Therapy. (n.d.). Sound Sensitivity and Autism. https://www.crossrivertherapy.com/autism/sound-sensitivity#:~:text=One%20of%20the%20common%20symptoms,anxiety%2C%20and%20even%20physical%20pain.

6. Cross River Therapy. (n.d.). Sound Sensitivity and Autism. https://www.crossrivertherapy.com/autism/sound-sensitivity#:~:text=One%20of%20the%20common%20symptoms,anxiety%2C%20and%20even%20physical%20pain.

7. Overcome with Us. (n.d.). How Music Can Enhance the Life of an Autistic Individual. https://overcomewithus.com/autism/how-music-can-enhance-the-life-of-an-autistic-individual#:~:text=Autistic%20people%20can%20develop%20a,lyrics%20when%20they%20are%20focused.

8. American Music Therapy Association. (2012). Music Therapy and Autism Spectrum Disorder. http://www.musictherapy.org/assets/1/7/MT\_Autism\_2012.pdf

9. Wan, C., Demaine, K., Zipsea, L., Norton, A., & Schlaug, G. (2010). From music making to speaking: Engaging the mirror neuron system in autism. Brain Research Bulletin, 82, 161–168. https://doi.org/10.1016/j.brainresbull.2010.04.010

10. Koul, R., & Shalev, A. (2022). Music-Based Interventions for Autism Spectrum Disorder: A Comprehensive Review. Frontiers in Psychiatry, 13, 905113. https://doi.org/10.3389/fpsyt.2022.905113

11. Bhatia, S., & Tandon, R. (2021). Music-Based Interventions for Autism Spectrum Disorder: A Systematic Review. Frontiers in Psychiatry, 12, 643234. https://doi.org/10.3389/fpsyt.2021.643234

12. Shi, Z.-M., Lin, G.-H., & Xie, Q. (2016). Effects of music therapy on mood, language, behavior, and social skills in children with autism: A meta-analysis. Chinese Nursing Research, 3(3), 137-141. https://doi.org/10.1016/j.cnre.2016.06.018

13. AudioCraft. (n.d.). MusicGen. https://audiocraft.metademolab.com/musicgen.html

14. Facebook Research. (n.d.). MusicGen. <https://github.com/facebook>

15. Spectrogram. (n.d.). In Wikipedia. Retrieved October 18, 2023, from https://en.wikipedia.org/wiki/Spectrogram

16. McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., ... & Moore, R. (2020). librosa: Audio and music signal analysis in Python. In Proceedings of the 14th Python in Science Conference (pp. 18-25). Zenodo. https://librosa.org/doc/latest/index.html