TOWARDS AI-ASSISTED MUSIC GENERATION FOR MOOD ENHANCEMENT IN INDIVIDUALS WITH AUTISM SPECTRUM DISORDER (ASD)

DESN3002 – Final Report

AU, Catherine Rose
3035920170

Table of Contents:

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.

Severity level	Social communication	Restricted and repetitive behaviors
Level 1 - Demanding support	In the absence of support, deficits in social communication cause remarkable harm. Difficulty initiating social interactions and clear examples of atypical or unsuccessful responses to social overtures from others. They may appear to have reduced interest in social interactions.	Inflexible behavior causes significant interference with functioning in one or more contexts. Difficulty switching activities. Problems with organization and planning are obstacles to independence.
Level 2 - Requiring substantial support	Severe deficits in verbal and nonverbal social communication skills; apparent social impairments even in the presence of support; limitation in initiating social interactions and reduced or abnormal response to social overtures from others.	Inflexibility of behavior, difficulty coping with change, or other restricted/repetitive behaviors appear frequently and interfere with functioning in various contexts. Suffering and/or difficulty changing focus or actions.
Level 3 - Requiring very substantial support.	Severe deficits in verbal and nonverbal social communication skills cause severe impairment in functioning, severe impairment in initiating social interactions, and minimal response to social overtures from others.	Inflexibility of behavior, extreme difficulty coping with change, or other restricted/repetitive behaviors. Great distress/difficulty changing focus or actions.

Source: Adapted from American Psychiatric Association - APA (2013).

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).

Author Year	Study design	Population	Diagnostic	Interventions	Assessment	Main findings	Limitations/Comments
NON-CONTROLLED STUDIES							
Lim and Draper (22)	Pre/post evaluation	$N = 22$ Age = 3–5 y.o. Gender: 17 boys Recruitment: na	ASD Dg tool: na ID: na Severity: children were verbal or preverbal with immediate echolalia	**Three different conditions:** Music training: "music incorporated ABA Verbal Behavior"; sung instructions, songs with target words shown on pictures and target phrases with echoic training (A) Speech training: ABA method with the same instructions without singing (B) No specific intervention (C) The order of the three conditions was randomly assigned Settings: individual sessions three times a week for 2 weeks. Duration of each session: na	**Clinical symptom:** no **Interaction:** Observations of behavior in videotaped posttest sessions: Verbal production evaluation scale including semantics, phonology, prosody, and pragmatics **Other assessment:** no	Positive effect on speech production: Verbal production SMD = 9.227 (p = 0.000) for MT vs. no training	However, no significant difference in verbal production was observed between music training and speech training The participants scored much higher on echolalia production than in response to questions
Kalas (23)	Pre/post evaluation	$N = 30$ Age = 4–7 y.o. Gender: 28 boys Recruitment: community outpatient activity	ASD Dg tool: na ID: na Severity: CARS: 15 with severe ASD, 15 with mild/moderate ASD	**Two experimental groups:** With simple music condition: simple melody and accompaniment on keyboard With complex music condition: complex melody and complex accompaniment on keyboard (with syncopation or dotted rhythms) Settings: individual 10 min sessions twice a week for 3 weeks	**Clinical symptoms:** no **Interaction:** Behaviors observed in videotaped sessions: ESCS: Joint attention measured by the number of times the child responded to a bid for joint attention by following a pointing gesture toward the object or instrument of interest. RJA score ICC = 0.86 **Other assessment:** no	Statistically significant interaction between the music modality and level of functioning. $F_{(1)} = 20.089$, $p = 0.001$	The effect of simple vs. complex music depended on the level of functioning. Specifically, the simple music condition was more effective at eliciting RJA in children diagnosed with severe ASD, whereas the complex music condition was more effective at eliciting RJA in children diagnosed with mild/moderate ASD
Pasiali et al. (24)	Pre/post evaluation	$N = 9$ Age = 13–20 y.o. Gender: 4 boys Recruitment: special education facilities	ASD Dg tool: na ID: yes Severity: CARS: 3 with severe ASD, 4 with mild/moderate ASD	**Intervention:** Interactive music therapy: "musical attention control training" Settings: 45 min sessions once to twice a week for 6 weeks	**Clinical symptoms:** no **Interaction:** no **Other assessment:** TEA-Ch	Positive changes in scores on tests related to selective attention and attentional control/switching, but no difference in sustained attention	
Paul et al. (26)	Pre/post evaluation	$N = 3$ Age = 3–4 y.o. Gender: 3 boys Recruitment: na	ASD Dg tool: na ID: na Severity: CARS: 1 child with mild to moderate autism, 2 with severe autism	**Interventions** During activities (block matching, picture matching, clay play) (A) Spoken oral instructions (B) Sung oral instructions Settings: individual 15 min sessions once or twice a week, during 3 months. 9 sessions with each condition	**Clinical symptom:** no **Interaction:** Behavior observation of videotaped sessions: performance, frequency of social gesture and eye contact (ICC value correct-to-excellent) **Other assessment:** no	Visual graphic analyze: all participants scored higher in the sung condition compared to spoken condition for all measures	Small sample size No control group
Davis (25)	Pre/post evaluation	$N = 4$ Age = 6–7 y.o. Gender: 4 boys Recruitment: special education facilities	ASD Dg tool: na ID: na Severity: 2 patients (50%) are verbal	**Experimental group:** Interactive music therapy Three different conditions: cooperative music therapy (A), cooperative play (B) and independent play (C) Settings: individual 20 min sessions once or twice a week for 5 weeks	**Clinical symptoms:** no **Interaction:** Behaviors observed in videotaped sessions: joint attention behaviors: interaction (responding to joint attention) and requesting (initiation) behaviors **Other assessment:** no	Interaction: mean difference in scores for MT/independent play = 71 Requesting: mean difference in scores MT/independent play: negative value = −16.875	Increase in interaction behaviors for all subjects during cooperative play and MT compared with independent play, but discordant results for requesting behavior
CONTROLLED STUDIES							
Buday (27)	Crossover	$N = 10$ Age = 4–9 y.o. Gender: 8 boys Recruitment: public school	ASD Dg tool: na ID: 70% Severity: CARS: 5 with severe ASD, 5 with mild/moderate ASD	**Two conditions:** (A) Interactive music therapy: songs used to teach signs and speech (B) Rhythmic speech used to teach signs and speech Settings: individual 3 min sessions four times a week for 2 weeks (4 sessions in the first week with one condition and 4 sessions in the second week with the other condition)	**Clinical symptoms:** no **Interaction:** Behaviors observed in videotaped sessions: sign and speech imitating behaviors. ICC 98% **Other assessment:** no	For sign imitation: MT group: M = 5.1 Control group: M = 4.0 $\Omega^2 = 0.35$/estimated $d = 0.39$ $p < 0.05$ For speech imitation: MT group: M = 4.2 Control group: M = 3.2 $\Omega^2 = 0.42$/estimated $d = 0.30$ $p < 0.02$	Large effect of music vs. rhythm form on both sign and word learning Limitations: small sample size and only children who have shown an interest in music (obvious attention or enjoyment) were included. Crossover
Farmer (28)	RCT	$N = 10$ Age = 2–5 y.o. Gender: 9 boys Recruitment: na	ASD Dg tool: na ID: na Severity: na	**Experimental group:** Interactive music therapy sessions: guitar playing, songs, games, name and point at body parts, imitation of animals **Control group:** Sessions without music Settings: individual daily 20 min sessions for 5 days	**Clinical symptoms:** no **Interaction:** Behaviors observed in videotaped sessions: Frequency of appropriate verbal and gestural responses **Other assessment:** no	Graphical analysis: Positive effect on verbal responses and gestural responses	Sessions in different conditions (home, school, etc.)
Katagiri (29)	Controlled study	$N = 12$ Age = 9–15 y.o. Gender: na Recruitment: community outpatient activity	ASD Dg tool: na ID: na Severity: na	**Two experimental groups** with background music: prerecorded piano improvisations structured to reference four basic emotions with singing songs: music with lyrics **Active control group:** Verbal instructions Settings: individual 30 min sessions twice a week for 4 weeks **No-intervention control group:** no specific intervention	**Clinical symptoms:** no **Interaction:** no **Other assessment:** Behavioral task: explicit emotional labeling based on facial expression (photographs or schematic drawing) after a learning period	$F = 2.09$, df = 3, $p = 0.13$ Secondary exploratory analysis: Analysis of covariance using pretest scores as a covariate reveals that the background music condition may be the most effective $F = 8.28$ df = 3, $p = 0.01$	None of the four conditions was significantly more effective than the others in improving participants' understanding of the four emotions

Lim (30)	RCT	$N = 50$ Age = 3–5 y.o. Gender: na Recruitment: community outpatient activity	ASD Dg tool: na ID: na Severity: CARS or ADI-R: 25 with moderate/severe ASD, 25 with mild ASD	**Experimental group:** Music training: "Developmental Speech and Language Training through Music"; videotaped songs with target words shown on pictures (PECS) **Active control group**: Speech training: videotaped spoken stories with target words Settings: individual 10 min sessions twice a day for 3 days **No-intervention control group:** no specific intervention	**Clinical symptoms:** no **Interaction:** Behaviors observed in videotaped posttest sessions: Frequency of appropriate verbal responses **Other assessment:** no	Positive effect on speech production: Verbal production ($p < 0.001$) $d = 1.275$ for the MT group vs. no training group	However, no significant difference was observed between the MT group and the speech training group
Sandiford et al. (31)	RCT	$N = 12$ Age = 5–7 y.o. Gender: 11 boys Recruitment: local advertisement	ASD Dg tool: ADOS ID: na Severity: Only children with limited or no functional verbal communication were included	**Experimental group:** Melodic Based Communication Therapy (MBCT): in addition to the traditional word elicitation approach, a combination of listening, hand clapping, singing of the word by the therapist and the child Settings: individual 45 min weekly sessions for 5 weeks **Control group:** Traditional language therapy	**Clinical symptoms:** no **Interaction:** Behaviors observed in videotaped sessions (first and last sessions): number of imitative attempts ICC 96% **Other assessments:** Vocabulary testing based on the International Phonetic Alphabet (assessor is blinded to the intervention): number of verbal attempts and number of correct words Parent survey: number of words reported by the parent	Increase in the number of verbal attempts from weeks 1 through 4 and number of correct words after weeks 1 and 3 in experimental group, while the control group progressed significantly after weeks 4 and 5 No significant differences in the number of verbal attempts ($z = -1.4$, $p = 0.08$) or number of correct words ($z = -0.2$; $p = 0.05$) were observed between the experimental and control groups A significant number of new words were heard in the home environment for the experimental group ($z = -2.0$, $p = 0.04$), but no significant difference was observed between the two groups ($z = -0.75$; $p = 0.45$) Participants in the experimental group had more imitative attempts ($z = -2.2$; $p = 0.03$)	Small sample size Lack of follow-up: missing data for parent survey
LaGasse (32)	RCT	$N = 17$ Age = 6–9 y.o. Gender: 13 boys Recruitment: local advertisement	ASD Dg tool: na ID: na Severity: CARS2 (values na)	**Experimental group:** Music therapy group: Transformational Design Model, music experiences to facilitate the participants' desired social skills, music-making Settings: 50 min group (3–4 children) sessions, twice a week for 5 weeks **Control group:** Social skills group: cooperative play, including board games and word games. Group interaction included a ball game	**Clinical symptoms:** Parent-rated assessments: SRS (primary outcomes pretest: first group session, posttest: within 4 days of the last group session), ATEC **Interaction: Behaviors observed in videotaped posttest sessions:** eye gaze (ICC 0.934), joint attention (ICC 0.841), initiation of communication (ICC 0.935), response to communication (ICC 0.858), withdrawal behaviors (ICC 0.941) communication (initiation, response and withdrawal) in the first and last sessions **Other assessment:** no	SRS: Significantly greater improvement in the experimental group compared to control group ($p = 0.032$, partial $\eta^2 = 0.287$) ATEC: Nonsignificant difference between the two groups ($p = 0.0549$) Significant between-group differences in eye gaze toward persons ($p = 0.022$, partial $\eta 2 = 0.323$) and joint attention with peers ($p = 0.031$ partial $\eta 2 = 0.291$). No significant between-group differences in joint attention with adults, initiation of communication with another child, initiation of communication with an adult, response to communication, or social withdrawal behaviors	Joint attention increased with peers and decreased with adults. Promotion of peer-to-peer interaction Limitations: ATEC may be an inappropriate tool for measuring changes in social skills. Bias due to parental rating Small sample size Higher attrition in the control group. Missing data due to a lack of follow-up
Cibrian et al. (18)	RCT	$N = 22$ Age = 4–8 y.o. Gender: na Recruitment: special education facilities	ASD Dg tool: na ID: developmental age Mean = 5.72, SD = 1.2 Severity: na	**Two experimental groups:** Sound production based on child motor activities (touching, tapping or pinching a spandex fabric with animated background) and different motor exercises using tambourines Settings: individual sessions once a week during 8 weeks, duration of each session na	**Clinical symptom:** Developmental coordination disorder questionnaire **Interaction:** no **Other assessment:** Engagement in music survey: playing in touch questionnaire (clinician-rated) Timing synchronization	Improvements in motor coordination for 27% of participants, and in timing synchronization for all participants	DCD questionnaire may be inappropriate

y.o., years old; Dg, diagnostic; na, not available; ID, intellectual disability; ICC, Internal Consistency Coefficient; ABA, Applied Behavior Analysis; MT, music therapy; CARS, Children Autism Rating Scale; ESCS, Early Social Communication Scales; RJA, responses to joint attention; TEA-Ch, Test of Everyday Attention for Children; ADI-R, Autism Diagnostic Interview- Revised; PECS, Picture Exchange Communication System; ADOS, Autism Diagnostic Observation Schedule; SRS, Social Responsiveness Scale, completed by guardians; ATEC, Autism Treatment Evaluation Checklist, which was designed to evaluate new treatments through questions on speech and language skills, social skills, physical well-being, and sensory/cognition. The checklist has four areas: speech and communication (14 items), sociability (20 items), sensory/cognitive awareness (18 items), and health/physical behavior (25 items). A lower score indicates higher functioning. The scale was completed by guardians.

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.

```
from transformers import AutoProcessor, MusicgenForConditionalGeneration


processor = AutoProcessor.from_pretrained("facebook/musicgen-small")
model = MusicgenForConditionalGeneration.from_pretrained("facebook/musicgen-small")

inputs = processor(
    text=["80s pop track with bassy drums and synth", "90s rock song with loud guitars and heavy drums"],
    padding=True,
    return_tensors="pt",
)

audio_values = model.generate(**inputs, max_new_tokens=256)
```

```
sampling_rate = model.config.audio_encoder.sampling_rate
Audio(audio_values[0].numpy(), rate=sampling_rate)
```

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.

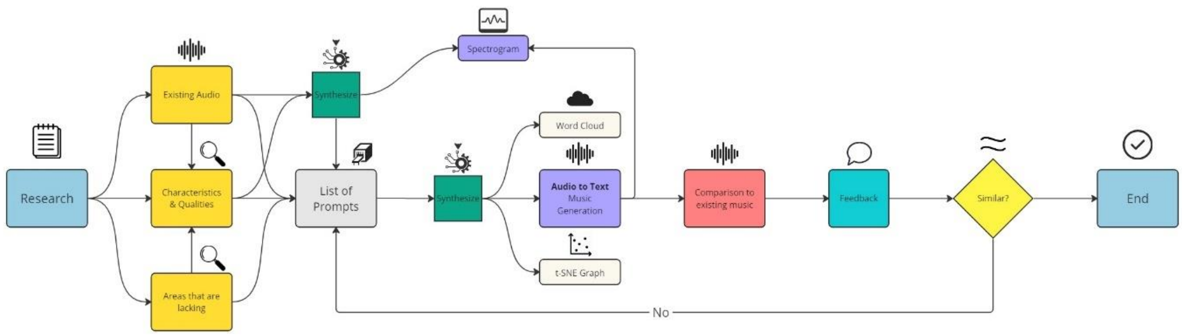

Figure 4

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.

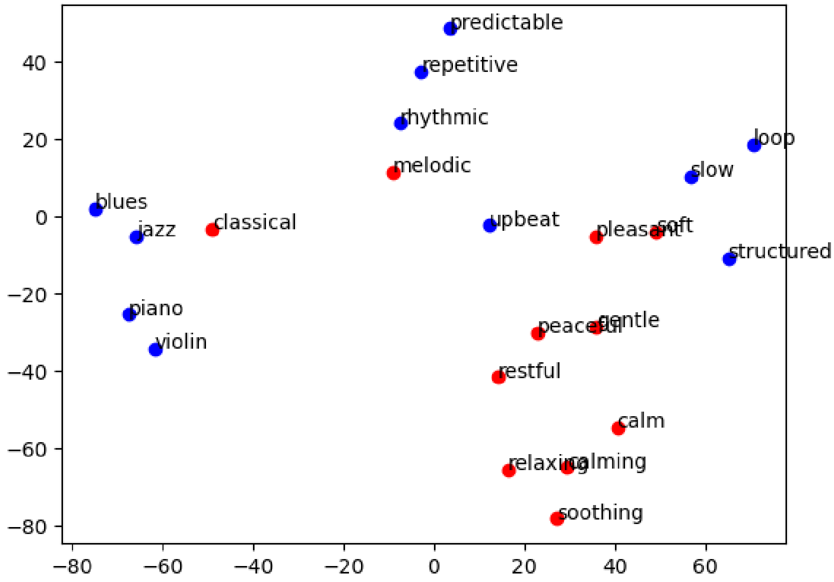

Figure 6

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.

```
audio_file = r'C:\Users\cathe\OneDrive\Desktop\DESN3002 Final Project\SHORTENED_Fingernails On A Chalkboard - Sound Effects.wav'
with wave.open(audio_file, 'rb') as wav_file:
    # Get audio file properties
    num_channels = wav_file.getnchannels()
    sample_width = wav_file.getsampwidth()
    num_frames = wav_file.getnframes()
    sample_rate = wav_file.getframerate()

    # Read audio data
    audio_data = wav_file.readframes(num_frames)
```

```
audio_np = np.frombuffer(audio_data, dtype=np.int16)
time = np.linspace(0, num_frames / sample_rate, num=len(audio_np))
```

```
# Plot raw wave signal
plt.figure(figsize=(10, 4))
plt.plot(time, audio_np, color= 'black')
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')
plt.title('Nails on Chalkboard')
plt.grid(True)
plt.show()
```

Figure 7

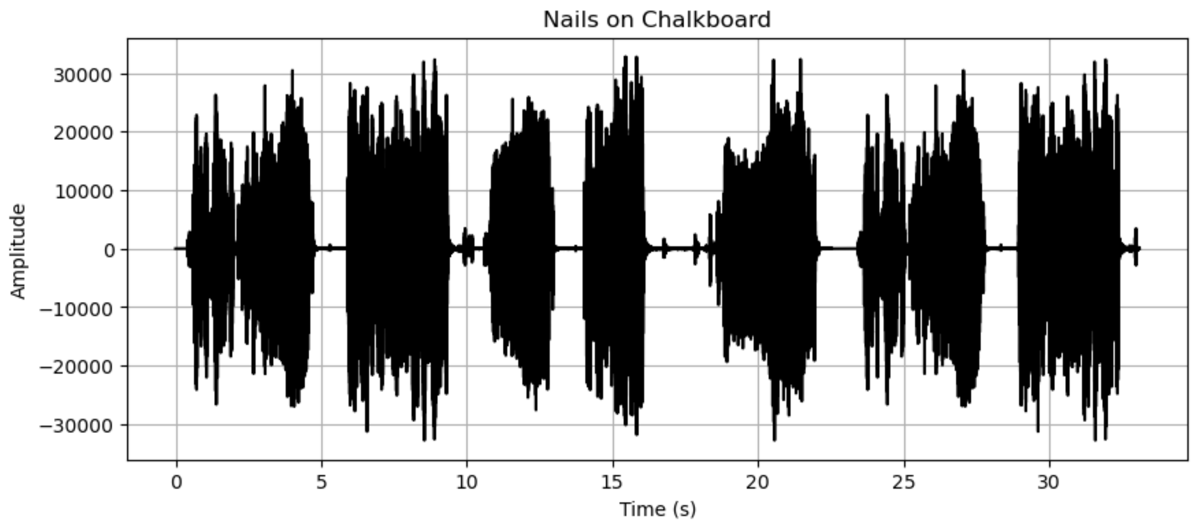

Figure 8

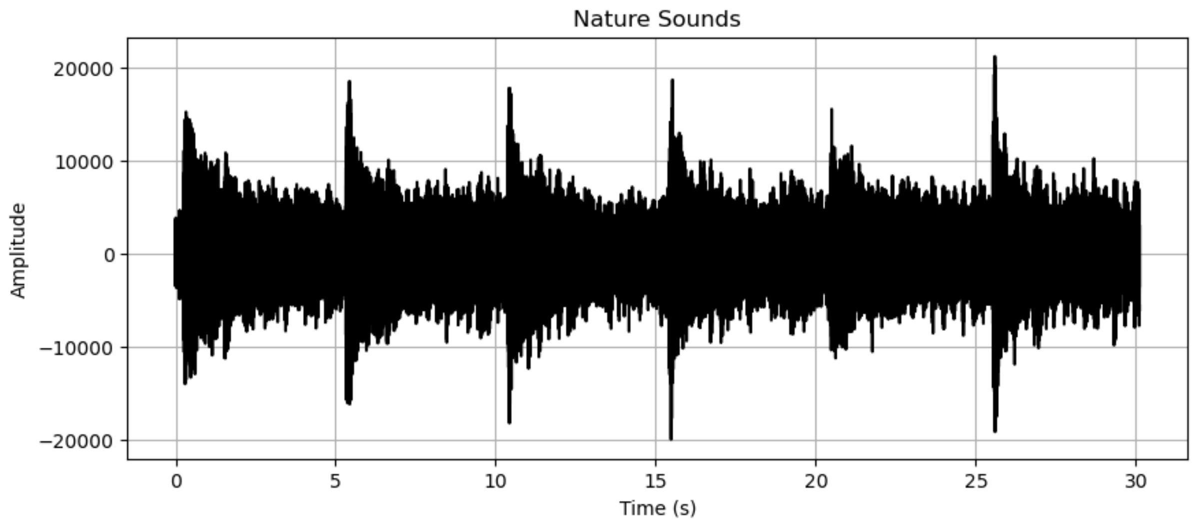

Figure 9

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.

```
audio_file ='path.wav'
y, sr = librosa.load(audio_file)
# Create spectrogram
spectrogram = librosa.feature.melspectrogram(y=y, sr=sr)

# Convert to decibels
spectrogram_db = librosa.power_to_db(spectrogram, ref=np.max)

# Plot spectrogram
plt.figure(figsize=(10, 4))
librosa.display.specshow(spectrogram_db, y_axis='mel', x_axis='time')
plt.colorbar(format='%+2.0f dB')
plt.title('Spectrogram')
plt.tight_layout()
plt.show()
```

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.

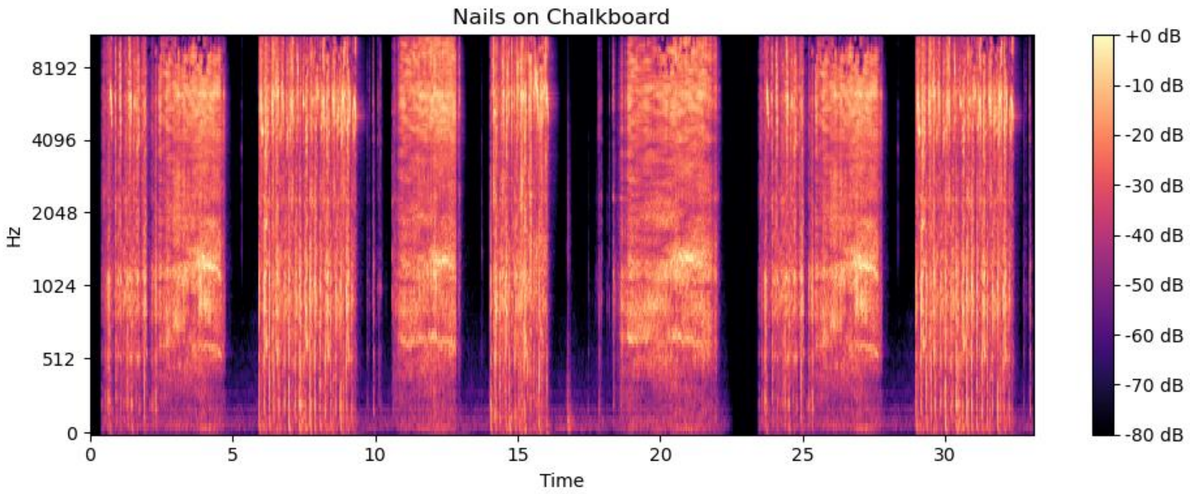

Figure 11

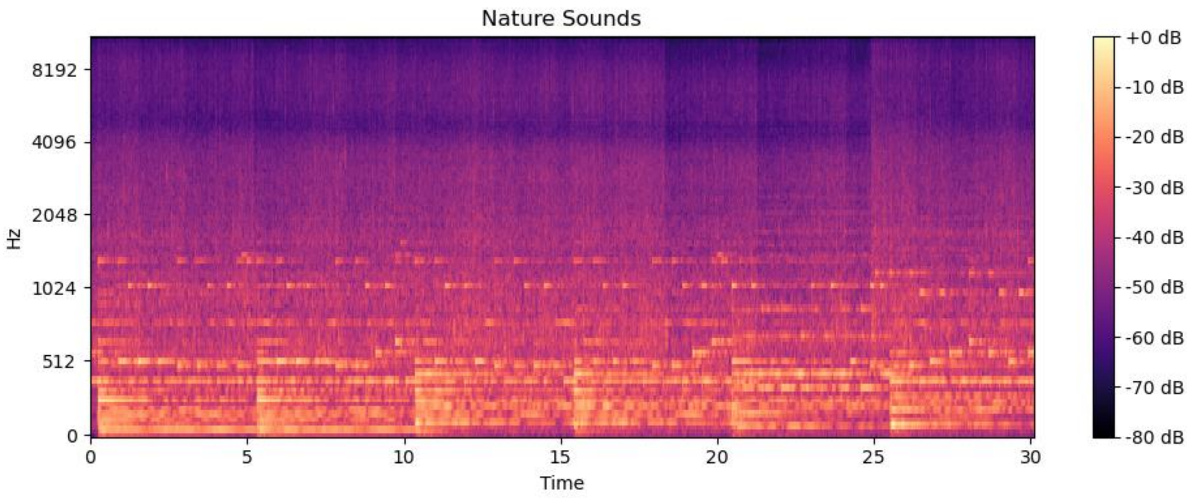

Figure 12

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.

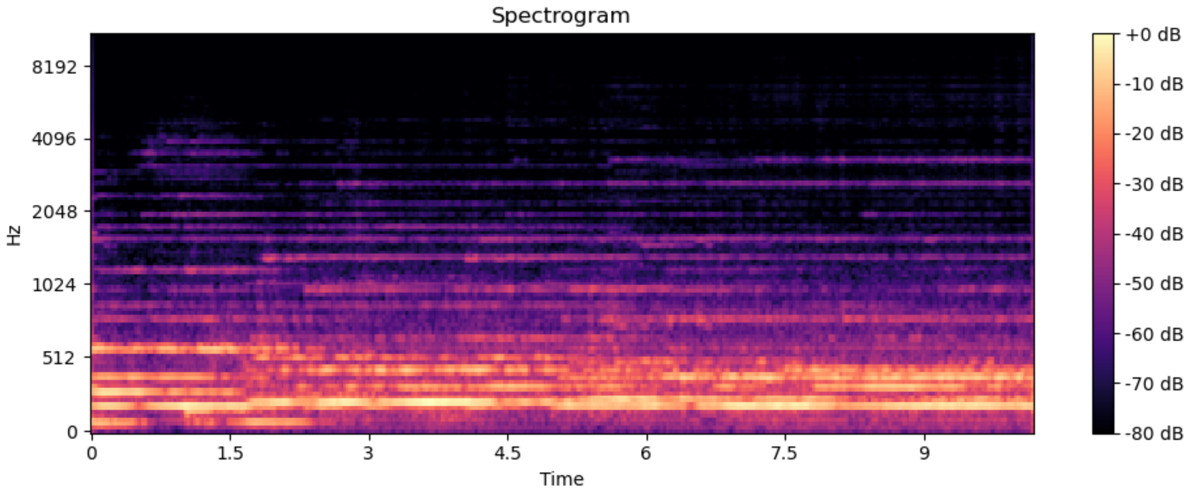

Figure 13

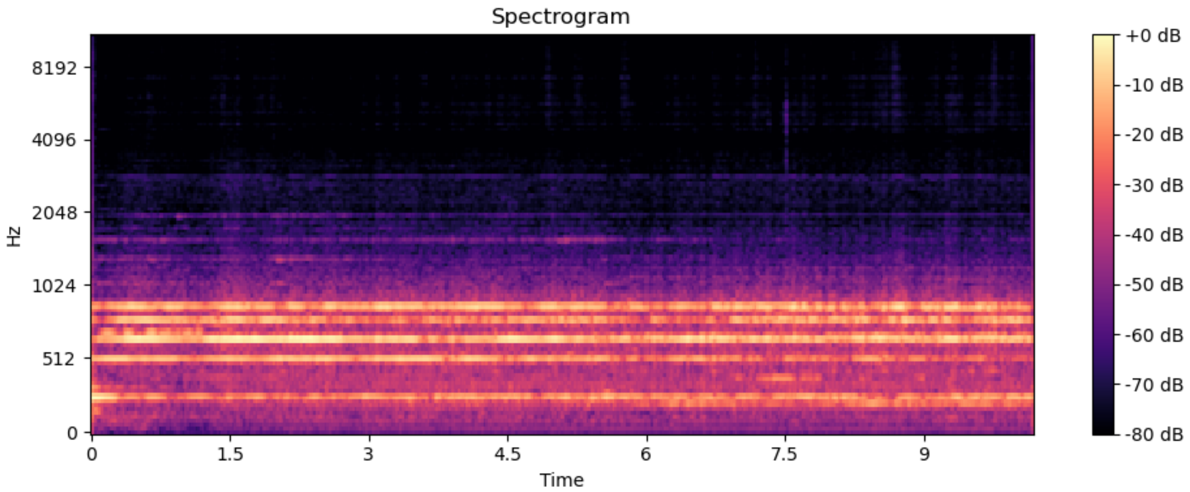

Figure 14

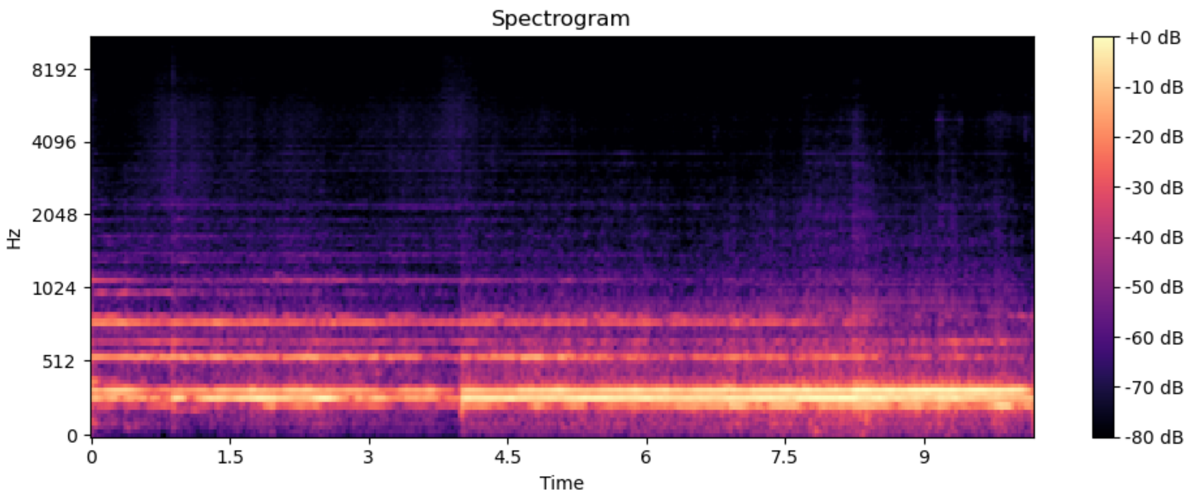

Figure 15

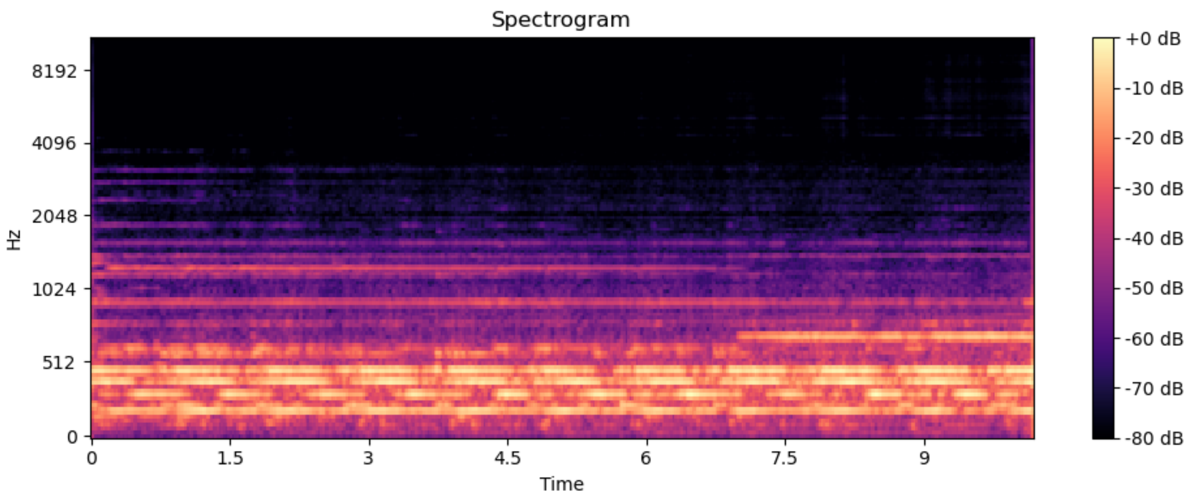

Figure 16

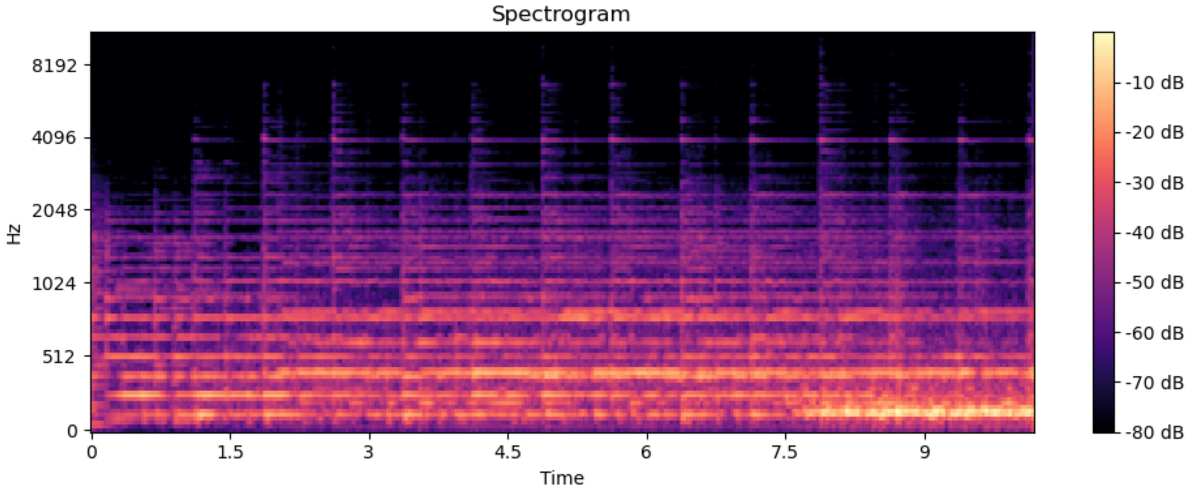

Figure 17

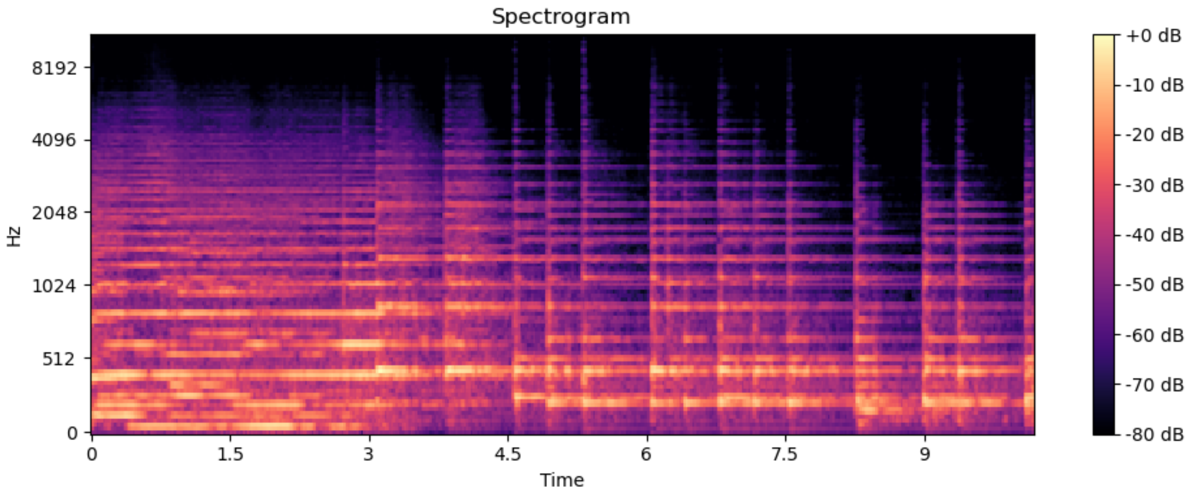

Figure 18

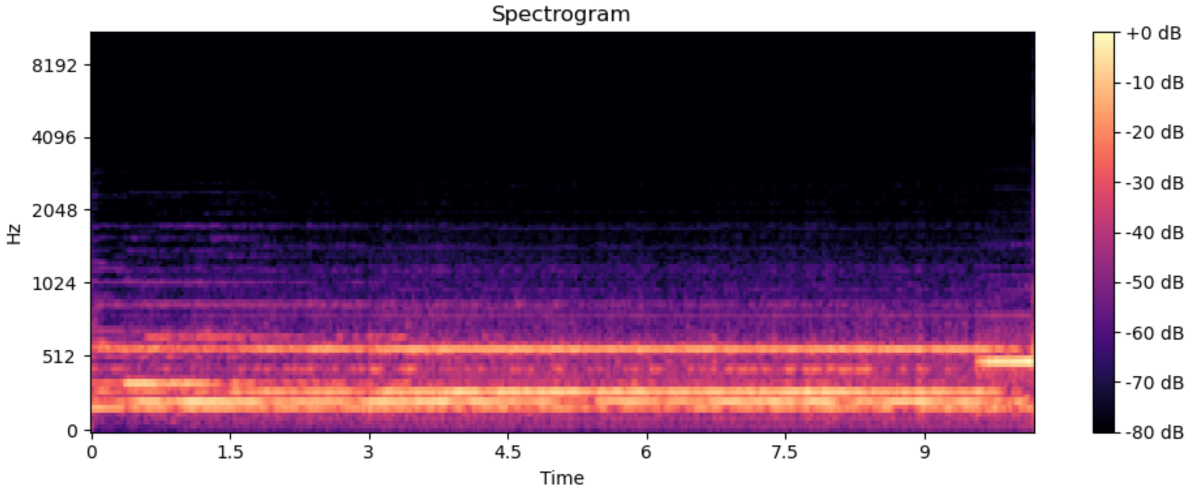

Figure 19

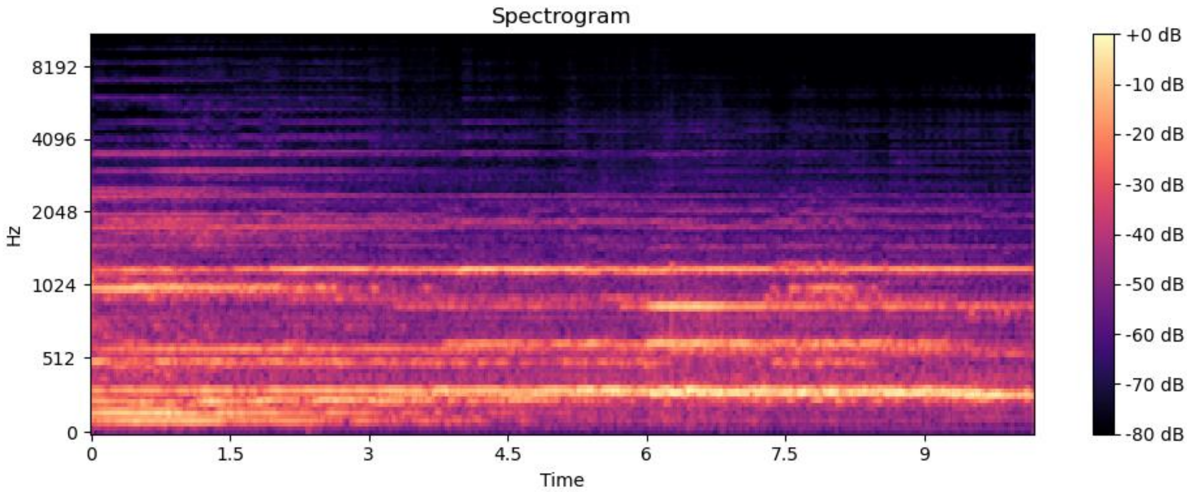

Figure 19

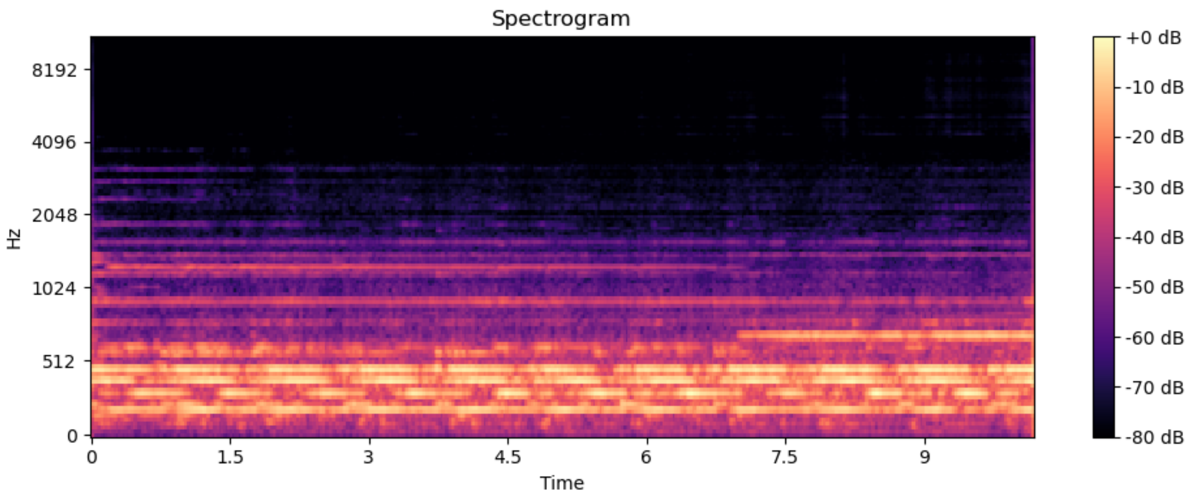

Figure 20

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.

```
audio_file3 =r'C:\Users\cathe\Downloads\LLM music\calm_peaceful_music_1.wav'
clip_to_cont, sr = librosa.load(audio_file3)
```

```
n_repeats = 2 # (each repeat adds 10 seconds)

overlap_dim = 325760
audio_all = _clip_0 = clip_to_cont
for i in range(n_repeats):
    _dim = _clip_0.shape[0]
    _dim_s = max(0, _dim - overlap_dim)
    _clip_0 = _clip_0[_dim_s:]
    inputs = processor(
        audio=_clip_0,
        sampling_rate=32000,
        text=["classical, calming, slow paced, gentle, peaceful, serene, tranquil, mellow, restful, pleasant, soothing, pleasurable, beautiful music"],
        padding=True,
        return_tensors="pt",
    )
    _clip_1 = model.generate(**inputs, do_sample=True, guidance_scale=3, max_new_tokens=256*2)[0,0].numpy()

    _dim_out = _clip_1.shape[0]
    audio_all = np.hstack((audio_all, _clip_1[_dim_out-overlap_dim:]))
    _clip_0 = _clip_1
```

```
audio_values_cont = model.generate(**inputs, do_sample=True, guidance_scale=3, max_new_tokens=256*2)
sample_cont = audio_values_cont[0, 0].numpy()
sampling_rate_cont = model.config.audio_encoder.sampling_rate

Audio(audio_all, rate=sampling_rate_cont)
```

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.

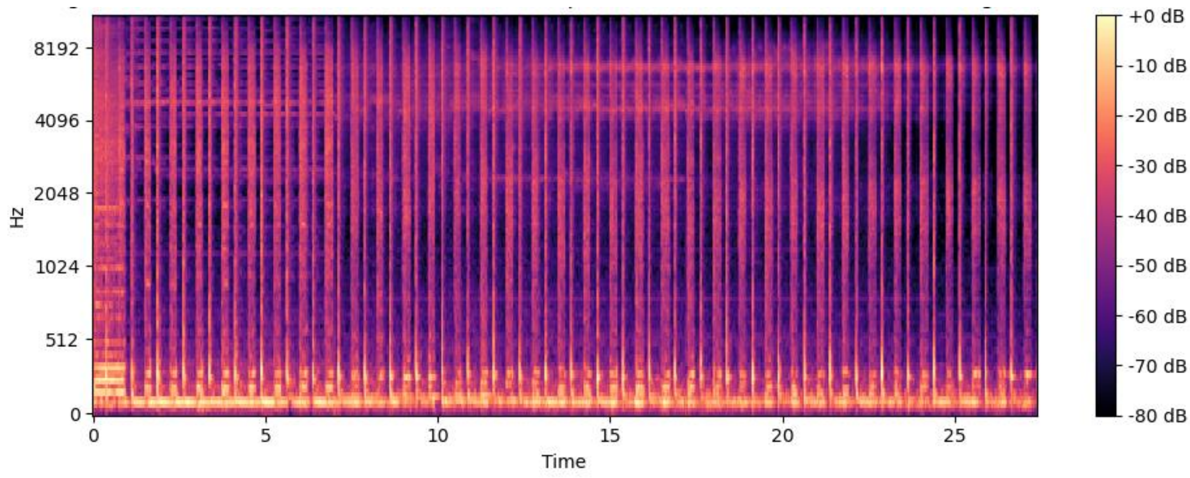

Figure 22

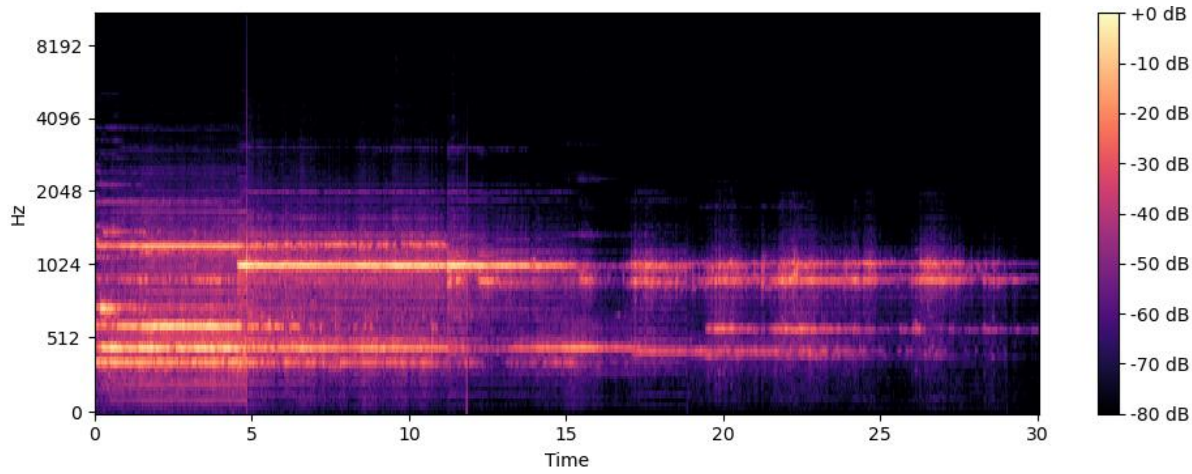

Figure 23

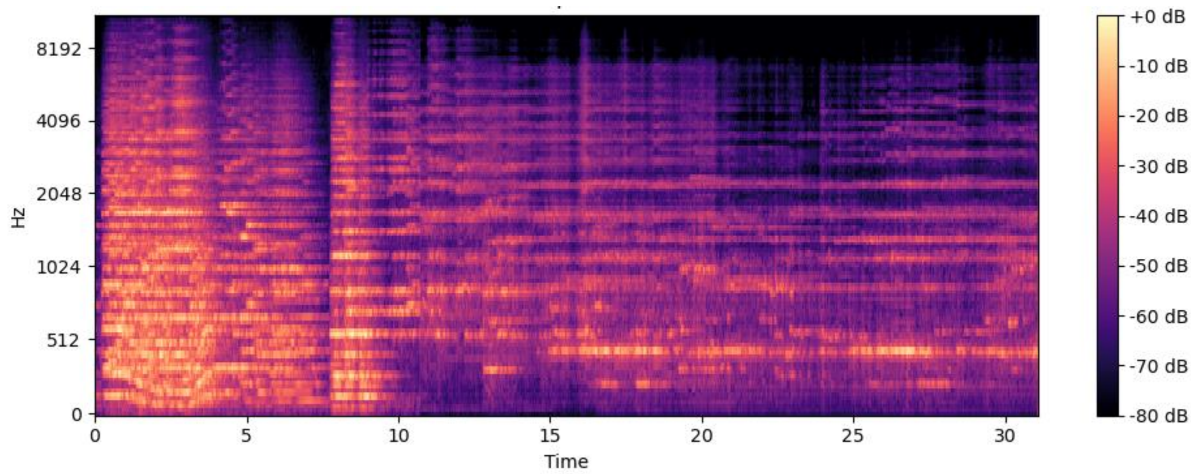

Figure 24

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.

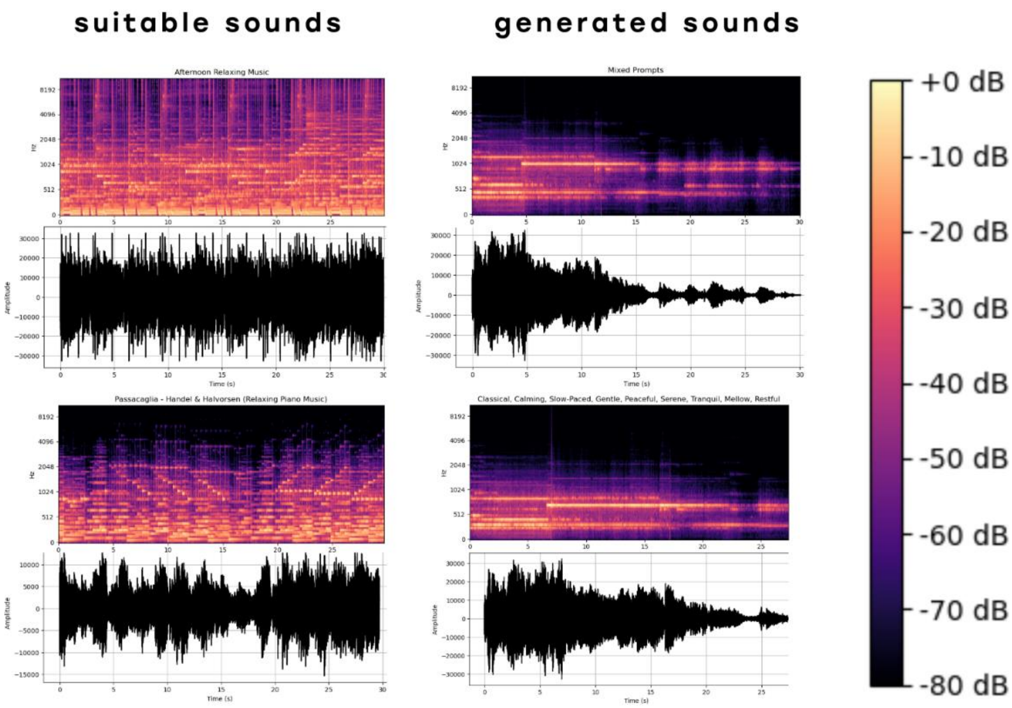

Figure 25

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.

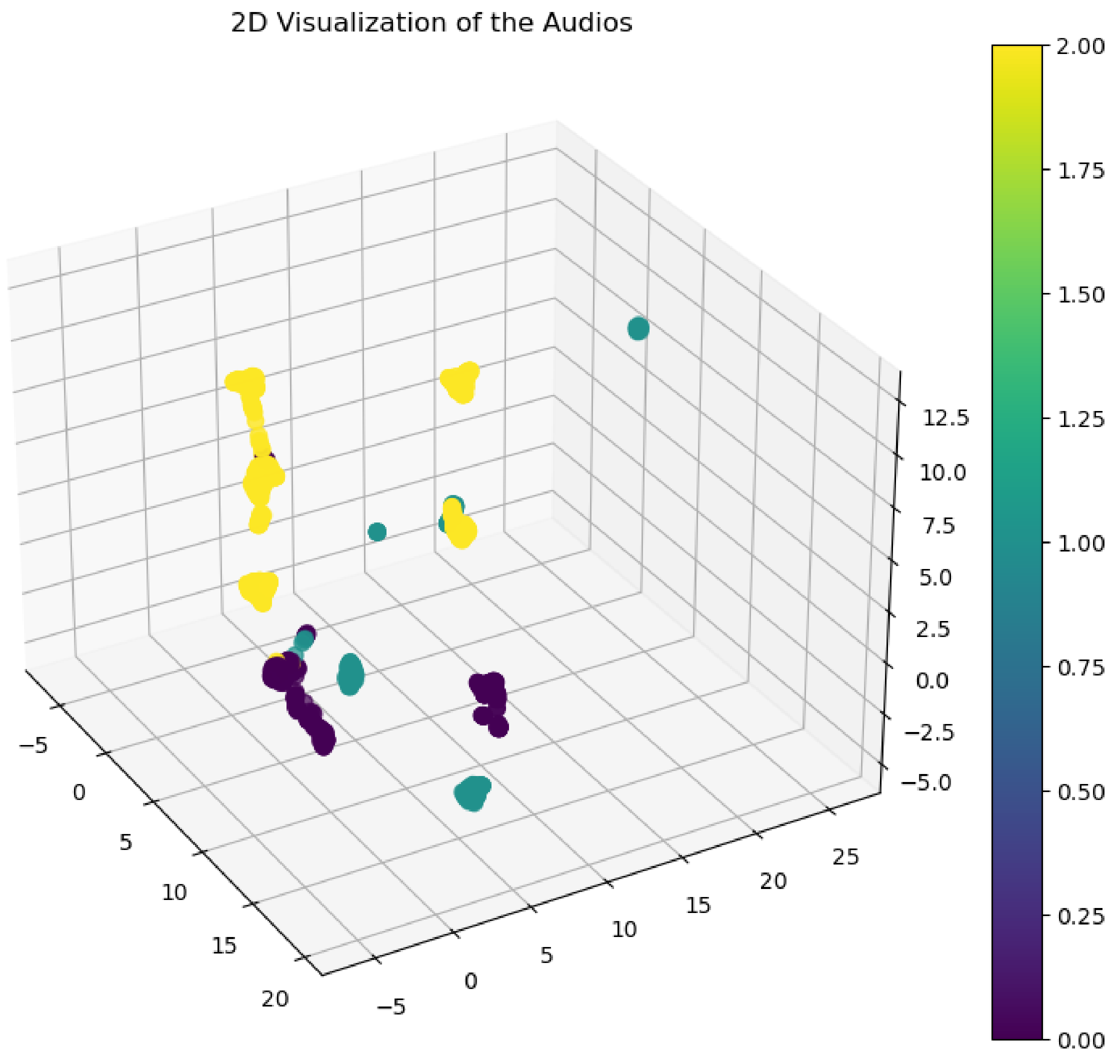

Figure 26

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.

Audio Title/Prompt	Before Audio	During Audio	After Audio	Conclusion
30Sec_EffectiveExample	6	8	8	Some mood improvement
30Sec_Less Suitable Example	5	3	4	Upset him during, slightly happier once over
10sec_GeneratedFromPrompt	6	6	6	No change
30sec_ContinuationOfGenerated	4	5	5	Less anxious, more focused
30Sec_ContinuationOfContinuated	5	5	7	Breathing slowed, less tense
30Sec_ContiuationOfEffective	6	8	9	Mood gradually improves, smiled

1 = extremely anxious/stressed , 5 = neutral , 10 = extremely calm/content

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