

Feeling the Music: Measuring Enhanced Emotion in Musicians via CNN-based Sentiment Analysis

Gavin Warnakulasooriya
gavinwarn99@gmail.com

ABSTRACT

Empathy has long been associated with musical expertise, but quantifying empathy remains a difficult task. This study investigates the correlation between musical training and emotional reactivity via webcam based facial expression recognition. By using a Convolutional Neural Network (CNN), facial expressions of musically trained and untrained participants were analyzed as they watched emotionally intense videos, with and without audio. Musically trained individuals, defined as those with over 400 hours of total musical training, had significantly higher emotional reactivity than those without training. The musically trained group also showed a greater difference in their emotional response to videos with audio compared to videos without audio, suggesting an increase in emotional sensitivity to audio. Statistical analyses, using Mann-Whitney U tests and Cohen's d effect size calculations confirmed that the differences were statistically significant, and had a large effect size, thereby reinforcing the accepted hypothesis that musical training enhances emotional processing. The findings aid research on the cognitive and emotional benefits of musical training, providing quantitative backing that musicians have higher emotional reactivity (empathy). Additionally, the study highlights potential use cases of AI in facial sentiment analysis as an objective measure for psychological research concerning emotions. Future research could further explore causal links behind musical training and empathy and implications for the technology in music education, media analytics, and therapeutic methods.

Introduction

Past empirical studies have supported the idea that musical training can enhance one's empathy and emotional reactivity. Engaging with music fundamentally uses a level of emotional communication, which may train individuals to perceive and share emotions more effectively, a sign of empathy. Studies have often showed the positive social and emotional effects of musical training, like an improved recognition of emotional inflections in speech, an enhanced ability to recognize distress in infants, and higher scores on behavior scales (Kirschner & Tomasello, 2010; Rabinowitch et al., 2012; Wang et al., 2009). The findings of these past studies suggest that interaction with music can foster emotional awareness and build an increased capacity for empathy.

Defining musical training can be ambiguous, so a practical threshold is 400 hours of cumulative practice, corresponding to an intermediate level of proficiency in music. Past study has shown that individuals with at least 400 hours of training in music demonstrate stronger auditory skills, regardless of if they identify as a musician (Wu & Lu, 2021). The threshold ensures that the musician group has meaningful engagement with music, regardless of their own perceptions, allowing for a comparison to non-musicians with less than 400 hours of experience.

Facial Expression Recognition (FER) can classify emotions based on facial features using Convolutional Neural Networks (CNNs) (Kopalidis et al., 2024). The networks often include multiple convolutional layers combined with pooling layers, allowing them to detect small patterns in the face, such as small muscle movements. Dropout layers in the network prevent overfitting while categorical cross-entropy loss functions help provide stronger emotion classification.

One example of FER is the VGG-based FER model, which adapts a deep CNN (VGG-19) for emotion recognition. This network takes a facial image and processes it through multiple layers, outputting a probability for each emotion in specified categories. The model was trained on large datasets like FER2013 and has optimized its weights to classify expressions in an accurate manner. Studies indicate the CNN-based FER systems like the VGG-19 based model can achieve classification accuracies exceeding 87% (Liu et al., 2021). Advancements in FER allow researchers to measure emotional reactivity in an objective manner with precision, making it a powerful tool for psychological research.

The primary hypothesis of the study is that the musically trained individuals will demonstrate a greater degree of emotional reactivity than their non-trained counterparts when exposed to video stimuli. As an exploratory

measure, the study also investigates whether the musically trained group will exhibit a higher difference in reactivity to the videos when they view them with and without audio. If confirmed, the findings suggest that musical experience enhances both empathy and emotional responses to auditory stimulus.

Methods

Participant Sampling

This study employed a snowball sampling method, leading to a final group of 46 participants, divided into 23 musicians and 23 non-musicians. These participants ranged from 14 to 60 years old, with an equal distribution across gender, across various musical backgrounds. To validate groupings, participants completed a survey assessing their musical experience prior to participation in the experiment. Those with over 400 hours of cumulative musical training were classified as musically trained (Wang et al., 2009).

Procedure

Before exposure to stimuli, participants viewed a neutral gray screen for 10 seconds to establish baseline emotional levels. They then watched three emotionally provocative videos selected to elicit distinct emotional responses: a video of a distressed dog designed to evoke sadness and anger, a video of happy babies intended to elicit joy, and a jump-scare horror video to provoke surprise and fear. These videos were first presented without audio and then shown again with synchronized audio to examine differences in emotional reactivity based on auditory enhancement. A Python GUI using Tkinter and OpenCV controlled video playback and recorded facial expressions. The second round of video with audio serves as a repetition for all subjects as they are exposed to the same video and serves as the main category for additional exploration as to if musically trained individuals will change in reaction more than untrained individuals when exposed to audio. This second round is biased as in the second round, since the participants had already seen the videos, they would be less reactive by default. However, this means that any positive change in emotional reactivity was even more significant, thereby indicating a clear difference. The participants' video data was then analyzed by the python package "FER".

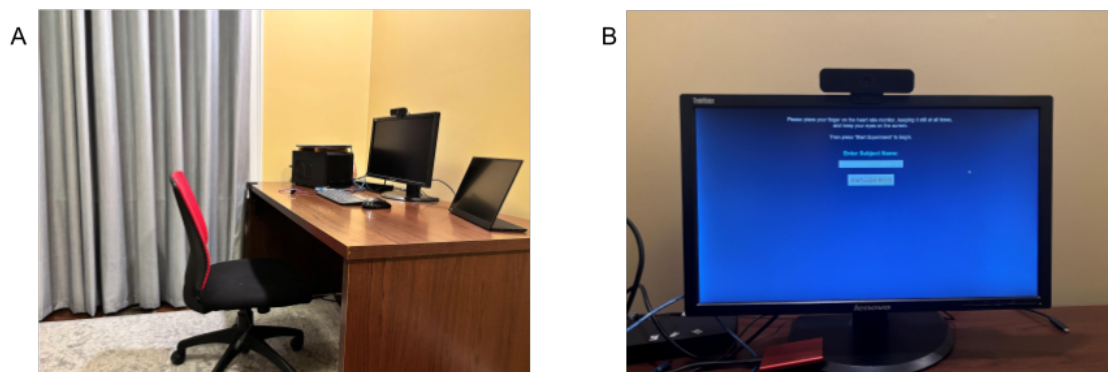


Figure 1. Setup. Desk, GUI. As seen in figure 1A, all participants were in the same room, with blackout curtains to minimize variability from external sunlight. As seen in figure 1B, participants then entered a custom Python GUI completely independent to minimize confounding factors via the presence of other people.

Table 1. Six video overviews

Video #	Description	Length	Target Emotion(s)
1	A sad video of a dog crying, no audio	~30 seconds	<i>sad, angry</i>
2	A happy video of babies laughing, no audio	~30 seconds	<i>happy</i>
3	A scary/ominous video of a chair with a figure jumping out, no audio	~20 seconds	<i>surprise, fear</i>
4	Same as video 1, with sad music.	~30 seconds	<i>sad, angry</i>
5	Same as video 2, with happy music and audible laughs	~30 seconds	<i>happy</i>
6	Same as video 3, with scary music and a scream	~20 seconds	<i>surprise, fear</i>

Variables

In the experiment, the independent variable was the participant's musical training status, categorizing individuals as either musically trained or untrained. The study included two dependent variables: the total empathy score, which measured the overall emotional reactivity of each participant, and the difference in empathy scores between the audio and non-audio conditions to assess changes in emotional responsiveness. The untrained group functioned as a negative control, allowing for the isolation of musical training as a factor influencing emotional reactivity. To maintain experimental rigor, several variables were held constant, including standardized procedures for all participants and a controlled environment in which all individuals were exposed to the same set of videos under identical conditions.

Controls

To establish controls for the FER model, both negative and positive controls were used. For negative control, several stock image of a person without facial features were run through the FER, returning 0 probability for all emotions, For a positive control, several images of happy and sad people were run through the model, which returned positive values for the 'happy' and 'sad' categories of the emotional probabilities. Thus, the performance of the FER model was verifiably accurate, and the experiment could continue.

Data Processing

The FER model extracted emotional data at 24 frames per second, classifying expressions into seven emotions: happiness, sadness, anger, surprise, fear, disgust, and neutrality (Liu et al., 2021). Baseline emotional scores (from the gray screen) were first averaged and then these average values were subtracted from video reaction scores to normalize individual differences. Then, a weighting algorithm was employed to ensure that the most significant emotional moments were given appropriate weighting within the dataset, meaning that participants' reactions during more significant moments during the videos were more heavily considered toward their total emotional score. This algorithm first segmented each participant's emotional response into discrete time intervals, collecting data at a rate of 24 frames per second. For each frame, the FER model classified emotions using a probability distribution across seven distinct emotional categories. From the FER's output data, the average baseline values were subtracted and then the result was multiplied by the weight (per emotional significance of moments). This process allowed for a precise, frame-by-frame quantification of emotional intensity, ensuring that the most emotionally salient moments contributed more significantly to the final reactivity score.

Following the quantification, only target emotions from Table 1 were isolated and then integrated to achieve an emotional score index for each video. To integrate the emotional responses over time, a Riemann sum integration method was applied, which allowed for the numerical approximation of the total emotional response by summing discrete emotional intensity values at finely spaced time intervals. Given the continuous nature of facial expressions, the Riemann sum approach was particularly effective in capturing fluctuations in emotional responses across time. The integration was performed using a right Riemann sum technique, where the function value at the right endpoint of each subinterval was used to approximate the area under the curve of emotional response over

time. This method provided a rigorous way to quantify total emotional reactivity, effectively transforming discrete emotional intensity measurements into a numerical reactivity score that could be compared between groups. Emotional index scores for each video were added to obtain a single total empathy score per each participant.

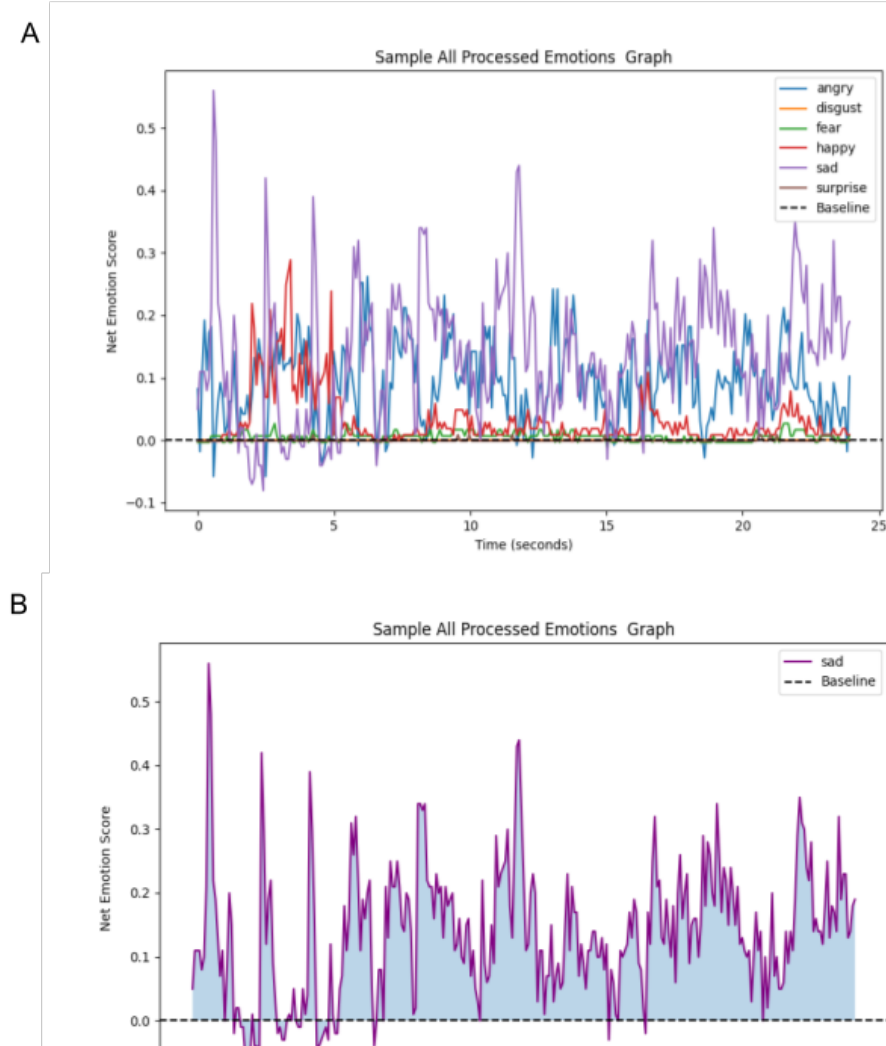


Figure 2. Processed emotional scores over time. A: All emotions, B: isolated emotion Reimann sum. As seen in figure 2A, emotions per the FER model can vary greatly over time, depending on the participants facial reaction to the stimuli. Per figure 2B, summing total values (in blue) can render a single score per each video.

Finally, for the exploratory section of the study, the net differences between audio and non-audio reactions were taken (e.g. emotional score index for video 4 - emotional score index from video 1) and summed to generate a single final difference score for the participant.

After this step, each subject would have both a total empathy score and a total difference in reactivity between audio and non-audio videos score. All analysis was conducted in a Python 3.9 environment with a Nvidia GTX 1650 and an Intel Core i7.

Results

Primary Outcome

The total empathy scores of all 46 participants were analyzed and categorized into two groups: musically trained and untrained individuals. To visually compare the differences in emotional reactivity between these groups, a bar graph and histogram were generated, illustrating the distribution of emotional responses across participants. These visual representations provide insight into the extent of variation in emotional reactivity and allow for a comparative analysis between musically trained and untrained individuals.

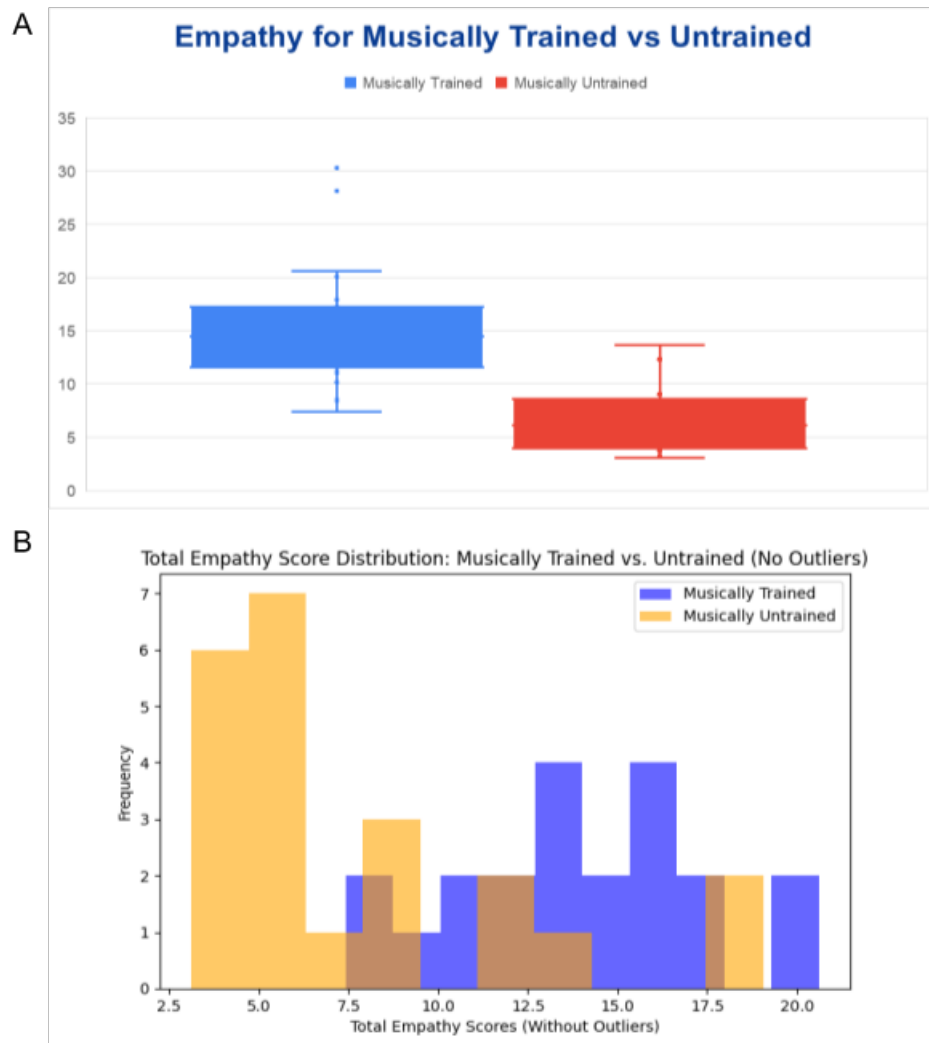


Figure 3. Comparison of total emotional scores between musically trained and untrained individuals

The box plot and histogram provide a clear depiction of the differences in emotional reactivity between musically trained and untrained individuals. The box plot illustrates that musically trained participants exhibited higher median empathy scores, with a greater range of values and a few extreme outliers, suggesting that their emotional reactivity varied more widely compared to non-musicians. In contrast, the untrained group demonstrated lower overall scores with less variability, indicating a more restricted range of emotional responses. The histogram further supports this trend, showing that the distribution of empathy scores in the musically trained group is more spread out, with a higher frequency of participants scoring at the upper end of the scale, while the untrained group clusters more heavily around lower scores. These results suggest that musical training is associated with heightened emotional sensitivity, reinforcing the hypothesis that musicians exhibit stronger emotional engagement with stimuli. The increased spread of scores within the musically trained group could also indicate that different levels or types of

musical training may further modulate emotional responsiveness, an avenue that warrants further exploration in future research.

Exploratory Outcome

From the difference scores evaluated, a histogram and line chart were produced to evaluate the differences between musically trained and untrained groups in terms of their difference in reaction to videos with audio and without audio.

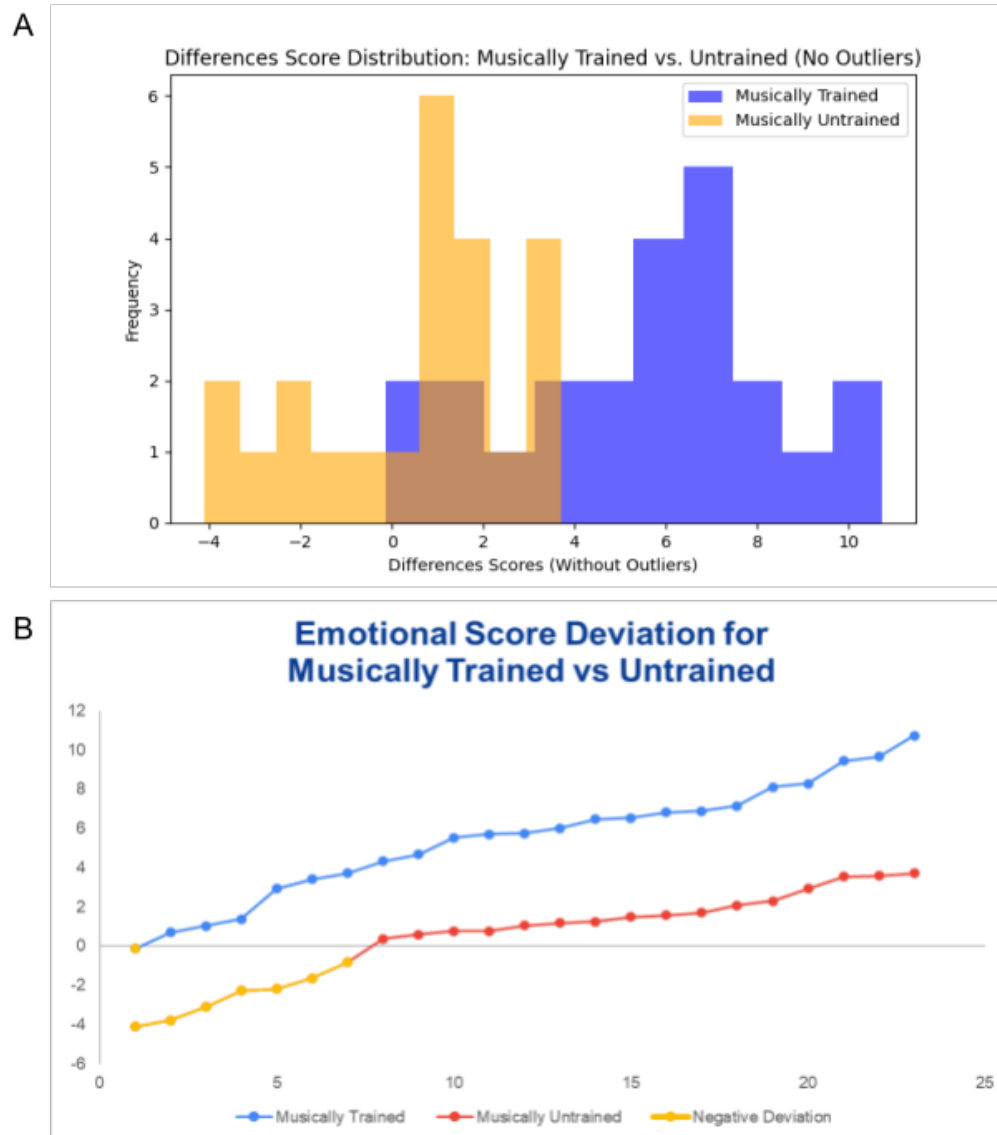


Figure 4. Comparison of Emotional difference scores for musically trained and untrained individuals.

The difference score distributions indicate that musically trained individuals show greater changes in emotional reactivity when exposed to media with added audio. This suggests a potential link between musical training and heightened sensitivity to emotional cues in auditory-enhanced stimuli.

Statistical Analysis

First, a Shapiro-Wilk test and Levene's test were applied to determine the type of inference test suitable for the study. Results of the tests can be seen below:

Table 2. Results of Shapiro-Wilk and Levene's tests.

Case	Test	Null Hypothesis	Result	Conclusion
Primary	Shapiro-Wilk	Data is Normal	Trained: W=0.9829 p=0.961 Untrained: W=0.8519 p=0.00367	Reject (Untrained is non-normal)
Exploratory	Shapiro-Wilk	Data is Normal	Trained: W=0.9724 p=0.745 Untrained: W=0.9283 p=0.101	Fail to Reject (Both are normal)
Primary	Levene's	Groups have equal variance	F=0.2511 p=0.619	Fail to Reject (Equal variance)
Exploratory	Levene's	Groups have equal variance	F=1.0920 p=0.301	Fail to Reject (Equal variance)

Because of the results of these tests, the Mann-Whitney U test was deemed most appropriate, due to its non-parametric quality. Finally, Cohen's d test measured the effect size of musical training on both emotional reactivity and differences with auditory stimuli (the primary and exploratory outcomes respectively).

Table 3. Results of Mann-Whitney U and Cohen's d tests

Case	Test	Null Hypothesis	Result	Conclusion
Primary	Mann-Whitney U	There is no significant difference between groups	U=395.00 p=7.11e-05	Reject null hypothesis (p<0.05)
Exploratory	Mann-Whitney U	There is no significant difference between groups	U=470.00 p=6.679e-06	Reject null hypothesis (p<0.05)
Primary	Cohen's d	N/A	d=1.4874	There is a large effect size of musical training on empathy
Exploratory	Cohen's d	N/A	d=1.8679	There is a large effect size of musical training on empathy score differences with an without audio

Due to p-values well below any reasonable threshold, we reject the null hypothesis for both the primary and exploratory cases. Thus, we confirm that musical training is associated with increased emotional reactivity and increased response to auditory stimuli. Finally, Cohen's d effect size calculation revealed a large effect size in both primary and exploratory analyses (d=1.4874, and d=1.8679, both greater than 0.8 threshold), reinforcing the large significance of musical training. Overall, these results confirm the original hypothesis that musicians exhibit more sensitivity to auditory and emotional cues, by a substantial degree.

Discussion

This study provides strong evidence that musical training enhances emotional reactivity, especially in response to emotional auditory stimuli. Musically trained individuals consistently showed significantly higher empathy scores than untrained participants, both in total emotional reactivity and in the degree of change when sound was introduced. These findings directly support the primary and exploratory hypotheses: first, that musically trained individuals would show greater emotional reactivity overall, and second, that they would exhibit a more pronounced difference between audio and non-audio conditions.

The ability of trained musicians to respond more deeply to emotional content, particularly when sound is present, may stem from enhanced auditory-emotional integration. Research by Coffey et al. (2017) demonstrates that musical training leads to increased connectivity between the auditory cortex and limbic structures like the amygdala. This neuroplasticity likely contributes to faster and more intense emotional processing of auditory cues. It also aligns with our observation that trained participants were not desensitized by video repetition but instead demonstrated greater reactivity when audio was added in the second round, an effect that would be unlikely unless they were particularly susceptible to sound-based emotional input.

These findings also expand the work of Rabinowitch et al. (2013) and Wu & Lu (2021), who found that musical engagement enhances empathy and prosocial behavior. By using an AI-driven, objective method (CNN-based FER), this study builds on their behavioral observations with real-time physiological evidence, strengthening the argument that musical training meaningfully alters emotional processing.

Beyond neural mechanisms, the implications for education and therapeutic practice are substantial. Structured musical instruction can serve as a medium for developing emotional intelligence and social-emotional learning, especially in early childhood or adolescent education (Schellenberg et al., 2015). The emotionally expressive and interpretive demands of music may help children learn to recognize and manage emotions - skills that are critical for lifelong mental well-being. Likewise, music therapy has shown efficacy in treating conditions like depression, anxiety, and autism spectrum disorder by offering patients a non-verbal channel to engage with their emotions (Aalbers et al., 2017).

From a technology standpoint, this study demonstrates how tools like FER can be used to quantify emotional sensitivity and measure emotional responses at scale. Prior studies on the topic were empirical, while this study created a more objective way to quantify emotional sensitivity, avoiding human biases. However, facial expression recognition remains imperfect: performance can be influenced by lighting conditions, camera angles, facial structures, and cultural display norms (Li & Deng, 2020; Liu et al., 2021). Future versions of this study could benefit from multimodal inputs, such as physiological signals from an EEG or heart rate sensors to improve reliability across populations.

Of course, this research has limitations. The small sample size ($n = 46$) and snowball sampling method restrict generalizability and may introduce selection bias. Emotional responses were also measured solely through visible facial changes, potentially missing individuals with intense but less overt expressions. In addition, we cannot yet determine the directionality of the observed effect: it remains possible that individuals with naturally high emotional sensitivity are more drawn to music, rather than music itself creating that sensitivity. Longitudinal or randomized intervention studies would be necessary to establish causality of this bidirectional effect.

Finally, this study opens the door to several future questions: Do different types of musical training (e.g., vocal vs. instrumental) or different genres (e.g., jazz vs. classical) lead to distinct emotional profiles? Does longer or earlier training yield stronger effects? Could combining facial analysis with fMRI or EEG provide clearer insight into the precise mechanisms by which music alters the brain's emotional processing pathways?

Conclusion

By employing CNN-based analysis, this research introduced a novel methodological approach by moving beyond observational techniques commonly used in psychological studies. Instead, it utilized objective, real-time facial expression data to quantify emotional reactivity with greater objectivity. This framework not only minimizes subjective bias but also offers a scalable and replicable method that can be applied in future psychological research, affective computing, and clinical assessments to evaluate emotional engagement in a more standardized manner.

In summary, this study confirms both aspects of our hypothesis: musically trained individuals exhibit greater emotional reactivity overall and demonstrate stronger sensitivity to auditory cues. These effects are

statistically robust and supported by both behavioral and neurological literature, suggesting that musical training enhances not only how we hear, but how deeply we feel.

References

- Coffey, E. B. J., Musacchia, G., & Zatorre, R. J. (2017). Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *Journal of Neuroscience*, 37(4), 830–838. DOI: 10.1523/JNEUROSCI.1265-16.2017.
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486–489. DOI: 10.5812/ijem.3505.
- Hoekstra, R., Kiers, H. A. L., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in Psychology*, 3, 137. DOI: 10.3389/fpsyg.2012.00137.
- Kirschner, S., & Tomasello, M. (2010). Joint music-making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31(5), 354–364. DOI: 10.1016/j.evolhumbehav.2010.04.004.
- Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in facial expression recognition: A survey of methods, benchmarks, models, and datasets. *Information*, 15(3), Article 135. DOI: 10.3390/info15030135.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. DOI: 10.3389/fpsyg.2013.00863.
- Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 13(3), 1195–1215. DOI: 10.1109/TAFFC.2020.2981446.
- Liu, Y., Wang, Z., & Yu, G. (2021). The effectiveness of facial expression recognition in detecting emotional responses to sound interventions in older adults with dementia. *Frontiers in Psychology*, 12, 642977. DOI: 10.3389/fpsyg.2021.642977.
- Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same population. *Tutorials in Quantitative Methods for Psychology*, 4(1), 13–20. DOI: 10.20982/tqmp.04.1.p013.
- Rabinowitch, T.-C., Cross, I., & Burnard, P. (2013). Long-term musical group interaction has a positive influence on empathy in children. *Psychology of Music*, 41(4), 484–498. DOI: 10.1177/0305735612440609.
- Schellenberg, E. G., Corrigan, K. A., Dys, S. P., & Malti, T. (2015). Group music training and children's prosocial skills. *PLoS ONE*, 10(10), e0141449. DOI: 10.1371/journal.pone.0141449.
- Wu, X., & Lu, X. (2021). Musical training in the development of empathy and prosocial behaviors. *Frontiers in Psychology*, 12, 661769. DOI: 10.3389/fpsyg.2021.661769.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Aalbers, S., Fusar-Poli, L., Freeman, R. E., Spreen, M., Ket, J. C. F., Vink, A. C., Maratos, A., Crawford, M., Chen, X. J., & Gold, C. (2017). Music therapy for depression. *Cochrane Database of Systematic Reviews*, 2017(11), CD004517. <https://doi.org/10.1002/14651858.CD004517.pub3>