

# Decoding of Emotion in Improvised Performance

Rowan Shields  
*Durham University*

## ABSTRACT

The following piece of research aims to identify how accurately listeners are able to decode the intended emotion of a performer in a short clip of improvised piano music. The paper aims to build on research by Gabrielsson and Juslin (1996) and Akkermans et al. (2018), using their study design, which focused on emotional decoding in pre-composed music, to explore the role of improvisation in emotional expression in music. Further information about emotional decoding is taken from studies such as Gilboa, Bodner, and Amir (2006) and Bodner (2014), which address how improvisation as a means of emotional expression (and the interpretation thereof) can be influenced by having a good understanding of emotion in music and cultural differences. The present study found that listeners often struggled to identify a number of emotions, in contrast to the similar studies which focused on pre-composed music. Additionally, weak positive correlations were found between high decoding accuracy and high scores on the Emotional and Musical Training Subscales of the Goldsmiths Musical Sophistication Index (Müllensiefen et al. 2014). Although significant correlations were not found here, as in prior studies (Akkermans et al. 2018), this positive correlation suggests that with further research this link may be demonstrated to a more meaningful degree.

## 1. INTRODUCTION

This research report gives details on a study in fulfilment of the Psychology of music module, which examines the ability of listeners to identify the intended emotions of a performer in an improvised piano excerpt. The study is based on two previous studies, one by Gabrielsson and Juslin (1996) and the other by Akkermans et al. (2018), which focused on the identification of emotions in pre-composed pieces. The primary aim of the present study is to determine how accurately listeners can identify the intended emotions of the performer, while the secondary aim is to compare the results with those of the previous studies and determine if improvised musical material makes decoding easier or more challenging.

The following study aims to explore whether listeners are able to identify the intended emotion of a performer in a 10-15 second clip of improvised piano. This study was based on a number of prior studies, in terms of design and subject area. Firstly, the structure of the study is closely linked to Akkermans et al. (2018), which is in turn a replication of Gabrielsson and Juslin (1996).

Gabrielsson and Juslin's (1996) study examined the ability of listeners to recognize the intended emotions of a performer in a musical excerpt. The study used excerpts of three pre-composed pieces, and the performers adapted their playing style to portray their intended emotion. They then had listeners rate the excerpts on scales of 0-10 for how well the clips represented 7 different emotions: Happiness, Sadness, Tenderness, Fearfulness, Solemnity, Anger, and Expressionless. They found that listeners were able to identify most emotions with high accuracy (83%), and that "basic emotions", such as happy, sad, and angry, were easier to identify than more complex ones, such as solemn or tender (Gabrielsson & Juslin, 1996). Additionally, they identified that often tender and sad emotions were confused. In their discussion, they also discussed the notion that certain instruments may be better or worse for portraying certain emotions, such as the flute struggling to portray anger, or the electric guitar struggling with solemnity (Gabrielsson & Juslin, 1996).

Akkermans et al. (2018) replicated Gabrielsson and Juslin's (1996) study using a much larger sample size ( $N=319$  participants, as opposed the 69 present in Gabrielsson and Juslin, 1996), as well as introducing several new factors, and found similar results. Some of the most relevant changes beyond sample size were the inclusion of piano on top of the instruments used in Gabrielsson and Juslin (1996), violin, flute, and voice, and their inclusion of several other factors to consider, most notably their use of the Goldsmiths Musical Sophistication Index (Müllensiefen et al. 2014), also known as the Gold-MSI, as a measure of musical experience, which they found to make a significant difference in terms of decoding.

Another study, which provides a basis for emotional decoding specifically of improvised music is Behrens and Green (1993). This study, like Gabrielsson and Juslin (1996) and Akkermans et al. (2018), had listeners rate

musical excerpts on a scale based on how well they represented various emotions, though this time a 4-point Likert scale ‘with 0 anchored by “improvisation does not express the feeling” and 4 anchored by “improvisation does express the feeling” (Behrens & Green, 1993). Additionally, listeners were only asked to rate the extracts as they related to 3 emotions: Sad, Angry and Scared, which they chose to do to ‘[limit the] amount of listening time’, and they specifically chose negative emotions ‘to reflect their emphasis on these emotions in our vocabulary’ (Behrens & Green, 1993; Lewis & Michalson, 1983). This study similarly found that listeners were generally able to identify the intended emotion of performers in short musical excerpts, and that the main factors affecting decoding accuracy were which instruments were used, and which emotion was trying to be portrayed (Behrens & Green, 1993).

One more study which addresses the improvised musical element is Gilboa, Bodner, and Amir (2006), entitled ‘Emotional Communicability in Improvised Music: The Case of Music Therapists’. This study was structured slightly differently, also having listeners listen to musical extracts, distributing 9 points between the six emotions considered: ‘anger, fear-tension, sadness, happiness-joy, calmness, and victory’ (Gilboa, Bodner & Amir, 2006). More points into an emotion indicated that the extract more accurately represented that emotion. Listeners were free to distribute these points however they saw fit, either putting all of them into one emotion, or spreading them out between several. This study was unique from the others as the improvisers were music therapists, which meant a good portion of their background and training was specifically related to emotion in music (Gilboa, Bodner & Amir, 2006). Additionally, the music therapists who did the recordings and other music therapists made up 21 of the 55 participants. This study’s finding aligned with prior research, finding that ‘emotions were successfully communicated by the [music therapists] and decoded by the listeners (Gilboa, Bodner & Amir, 2006). It is worth noting that this study found that the music therapists were better than the non-music therapists at decoding the intended emotions, but experience and technical ability of the music therapists who provided the recordings did not make a significant difference to how easily the listeners could decode their excerpts (Gilboa, Bodner & Amir, 2006). The study also found that the length of the excerpt did not affect whether it was able to be decoded by the listener, finding ‘that an improvisation can be as short as 16 seconds and still be quite clear, while a long improvisation does not necessarily guarantee that it will be correctly decoded (Gilboa, Bodner & Amir, 2006).

This study aims to address the improvisational element present in Behrens and Green (1993) and Gilboa, Bodner, and Amir (2006), while utilising the emotions and scales explored in Gabrielsson and Juslin (1996) and Akkermans et al. (2018). This is important as it will allow for direct comparison and exploration of whether improvised or pre-composed excerpts are easier to decode more directly than simply comparing the existing studies. With the above in mind, this study hypothesised that 1) Listeners would be able to successfully decode the performer’s intended emotion within the excerpt to a high degree of accuracy overall, with ‘basic emotions’ such as happy, sad, and angry being the easiest to identify, and 2) listeners will generally decode these improvised extracts more accurately than the pre-composed extracts on piano from Akkermans et al. (2018) as ‘In improvisation, the role of the composer and performer merge’ (Gabrielsson & Juslin, 1996) which means that elements of emotional encoding may be present not only in the playing technique, but also in aspects of compositional practice, such as implied key of the melody and the presence of certain intervals.. Additionally, as suggested by prior research (Akkermans et al. 2018; Gilboa, Bodner & Amir, 2006), there is a likelihood that higher scores on the Musical Training and Emotional subscales of the Goldsmiths Musical Sophistication index (implying a high level of musical training and emotional awareness of music) will have a significant positive impact on decoding accuracy.

## 2. METHODS

*Design.* The study took the form of a within participants design, in which the within-subjects independent variables were the 7 recordings in which the performer expressed each of the seven emotions (Solemn, Sad, Tender, Angry, Expressionless, Fearful and Happy), and the dependent variables were 0 – 10 agreement scales in which participants were asked to indicate how well each excerpt represented the above emotions. This design was intended to be as similar as possible to Akkermans et al. (2018), although it only used on instrument and one performer (and by extension, only one set of melodies, which were improvised instead of pre-composed). As such, the mixed measure elements of Akkermans et al. (2018) were not used for this study. Additionally, further demographic data was collected by having participants complete the questions for the Emotional and Musical Training subscales of the Gold-MSI.

*Participants.* Participants were recruited through convenience sampling methods, specifically word of mouth and social media. Although there were 31 respondents to the survey,  $N=22$  participants were considered for the study as the remaining 9 did not complete the survey entirely. Their level of musical training ranged from a score of 7 – 48 (Median = 36) on the Gold-MSI musical training subscale, and the range of scores on the emotional subscale was 27 – 42 (Median = 34.5). Further demographic information, such as gender or cultural background, was not collected for the purposes of this study, as prior research has not found that these factors have not had significant impacts on ability to perceive intended emotion in music (Bodner 2014; Akkermans et al. 2018).

*Materials.* The recordings of the excerpts were kindly recorded by a performer with an extensive background in both performance and composing, as well as experience improvising on piano and other keyboard instruments. The performer was presented with the seven emotions present in Akkermans et al. (2018), (Solemn, Sad, Tender, Angry, Expressionless, Fearful and Happy), and was asked to improvise for 10 – 15 seconds. These recordings were true improvisations, in the sense that they were created off the cuff, as opposed to being composed by the performer prior, thus differentiating them from pre-composed but unfamiliar stimuli. They were instructed to only play a melody line, as had been done in prior studies, but other than that were given free rein to realise their improvisation of the emotions however they saw fit. While allowing the performer to include a harmonic element may have provided another. Each of the clips was uploaded to and hosted on Soundcloud.com. The excerpts were not professionally recorded or edited due to practical constraints.

*Procedure.* The study took the form of an online survey, hosted on Qualtrics. First, respondents were informed of the purpose of the study, as well as how data would be handled, and confirmed they were 18 or older, to ensure that fully informed consent was obtained in line with ethical guidelines. I then asked the participants the questions from the Musical Training and Emotional Subscales of the Müllensiefen et al. (2014)'s 'Goldsmiths Musical Sophistication Index' (Gold-MSI), and the participants were subsequently assigned numerical scores based on their responses (Maximum score of 42 for Emotional Subscale and 49 for Musical Training Subscale). Participants were then asked to listen to the 7 different clips of improvised piano music, each in a randomised order to limit the impact of order effects. After listening to each clip, participants were asked to rate how well the clip represented the 7 different emotions on a scale of 0 – 10, with labels at 0 (not at all) 5 (Reasonably well) and 10 (completely).

### 3. RESULTS

The results were addressed in a number of different ways. The first was to identify in how many instances the intended emotion of a clip was identified correctly by participants – i.e., in how many instances the listeners rated the intended emotional expression the highest (as with Akkermans et al. (2018), two equally high emotions were considered incorrect). When this forced-choice metric was being considered, I found that on an individual level, on average participants were able to accurately identify the intended emotion 33% of the time, with the most accurately identified emotion being Happy at 68%, and the least being solemn, at 9%.

When the mean scores for each emotion were aggregated and then treated in the same way, I found that unlike in Akkermans et al. (2018), the aggregate scores were less successful at identifying the intended emotion, at only 29% accuracy. The two emotions that were accurately identified when aggregated were Fearful and Happy, though it is worth mentioning that Sad and Angry were both identified as close seconds, being confused with Solemn and Fearful respectively. These results are laid out in Table 1 and are reflected more simply in Figure 1 (see Table 1 and Figure 1).

Table 1. Mean Rating of Emotions by Intended Expression

	Solemn_Int	Sad_Int	Angry_Int	Tender_Int	Fearful_Int	Expressionless_Int	Happy_Int
Solemn_Rated	3.27	6.73	0.91	4.55	0.77	5.64	1.55
Sad_Rated	1.82	6.00	0.68	4.27	1.41	4.45	0.77
Angry_Rated	0.59	1.09	4.45	0.77	2.05	0.45	0.09
Tender_Rated	3.18	3.82	0.86	4.23	1.27	3.82	3.18
Fearful_Rated	0.77	3.14	4.50	6.05	6.05	2.64	0.68
Expressionless_Rated	4.14	2.64	1.68	3.05	1.73	3.50	2.41
Happy_Rated	3.91	0.91	2.05	2.09	2.55	1.23	6.45

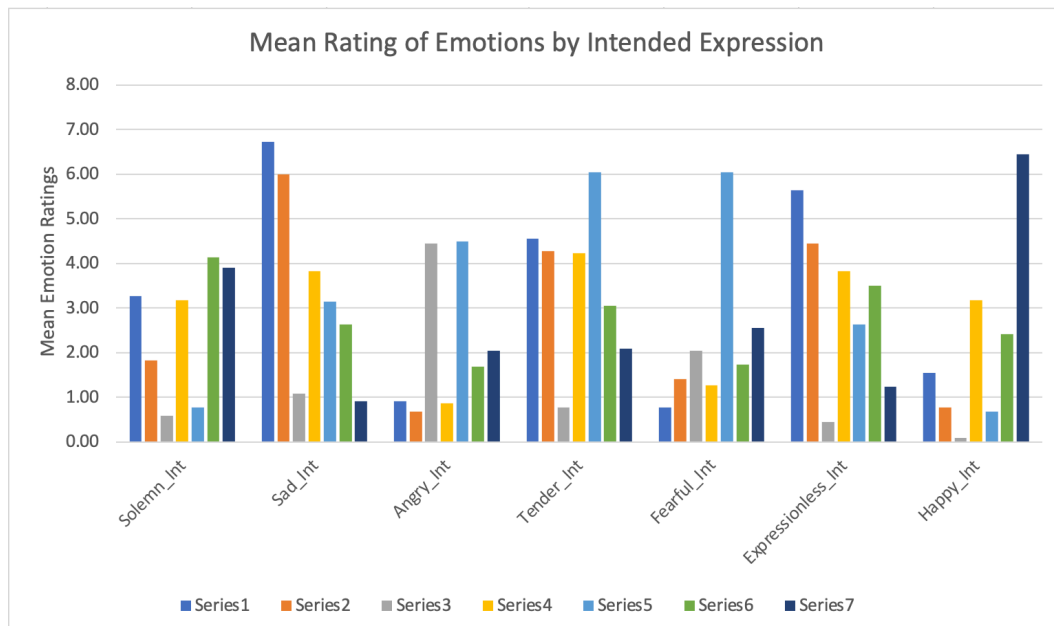


Figure 1. Mean Rating of Emotions by intended expression

Additionally, I carried out normal distribution hypothesis tests to identify whether participants were significantly likely to score a piece higher on a particular emotion if that emotion was intended by the performer. I did this in hopes of accounting for the possibility that participants may associate piano music with a particular emotion or emotion prior to testing. In these, I found that after aggregation only Anger reached the 0.05 significance level in this test, meaning participants aggregate mean for identifying the angry-intention clip was more than 2 standard deviations higher than their overall average anger scores. Anger being the highest in this regard is reflected in the individual significance tests, with the largest number of participants (32%) reaching this level of significance on an individual level. The results of these individual tests are laid out in Table 2, with green cells indicating a significant result, and red indicating a non – significant result. The bottom row shows the percentage of how of results which were significant by emotion.

Table 2. Individual tests for significance, comparing rating of emotions on track with corresponding emotion to average rating of that emotion on other tracks

	Solemn	Sad	Angry	Tender	Fearful	Expressionless	Happy
P1	X	✓	X	X	X	X	X
P2	X	X	✓	X	X	X	X
P3	X	X	X	X	X	X	X
P4	X	X	X	X	X	X	X
P5	X	X	✓	X	X	X	X
P6	X	X	X	X	X	X	X
P7	X	X	✓	X	X	X	X
P8	X	X	X	X	X	X	X
P9	X	X	X	X	X	X	X
P10	X	X	X	X	X	X	X
P11	X	X	X	X	✓	X	X
P12	X	X	✓	X	X	X	✓
P13	X	X	X	X	X	X	X
P14	X	X	X	X	✓	X	X
P15	X	X	✓	✓	X	X	X
P16	X	✓	✓	X	X	X	X
P17	X	X	X	X	X	X	X
P18	X	X	✓	X	X	X	X
P19	X	X	X	X	X	X	X
P20	X	X	X	X	X	X	X
P21	X	X	X	X	X	X	X
P22	X	X	X	X	X	X	X
	0%	9%	32%	5%	9%	0%	5%

When these results are compared to Akkermans et al. (2018), we can see that my aggregated scores are significantly lower, at 29%, compared to their 57%. However, my individual scores were slightly higher, at 33% as opposed to their 31%. This suggests that a larger sample size may distinctly change the outcomes of this study, as aggregation of a wider array of results may lead to more accurate aggregate decoding.

Finally, mean individual accuracy scores were compared to the scores achieved on the Gold-MSI Musical Training and Emotional subscales. There was no statistically significant correlation to be found between accuracy scores and either of their Gold-MSI scores, but weak positive correlations were found with both, with the correlation of Musical training and accuracy having an R value of 0.2, and Emotional subscale score and accuracy having an R value of 0.3. Figures 2 and 3 show these scores plotted against individual decoding accuracy, with a line of best fit (see Figures 2 & 3).

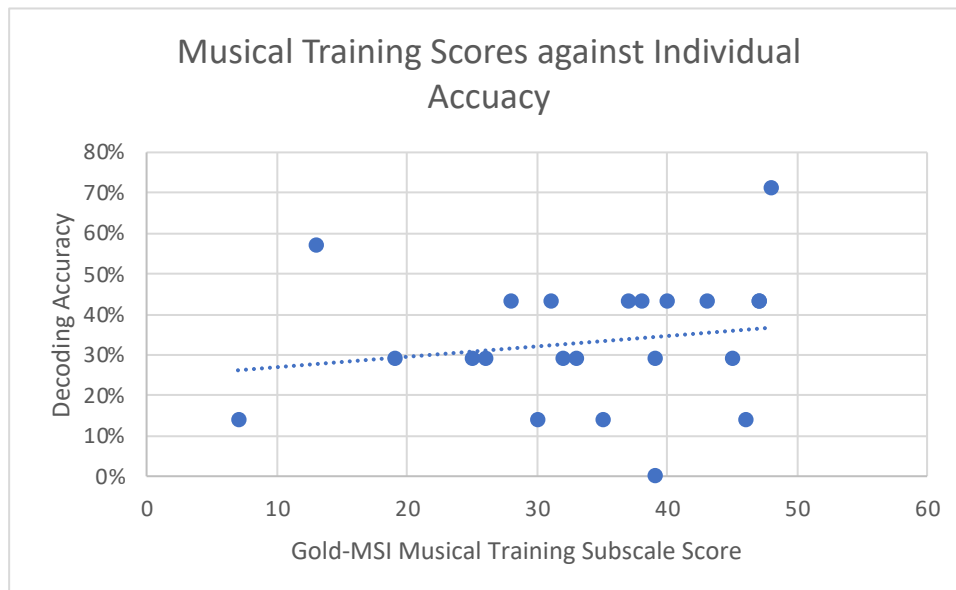


Figure 2. Musical Training Scores against Individual Accuracy with Line of Best Fit

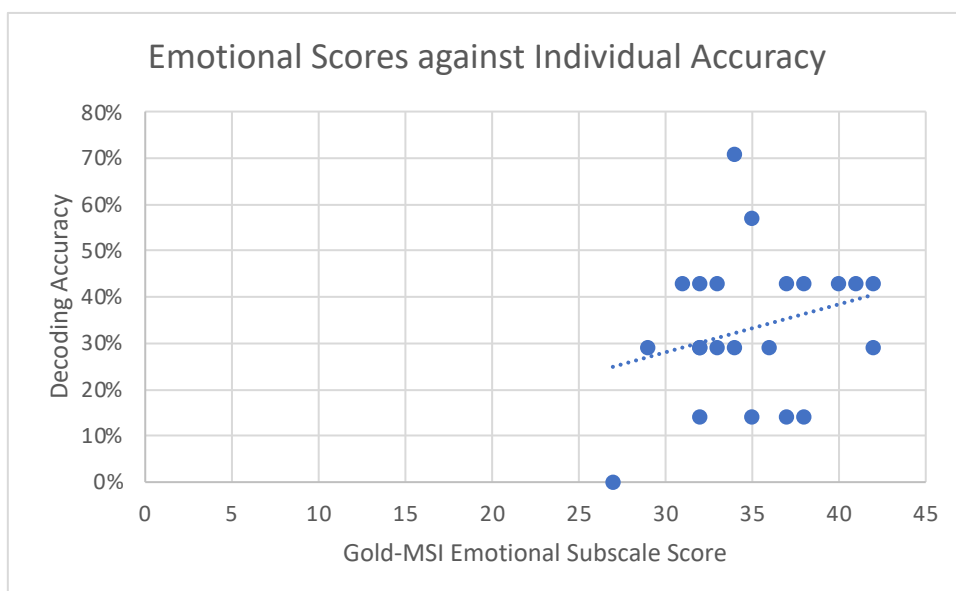


Figure 3. Emotional Scores against Individual Accuracy with Line of Best Fit

## 4. DISCUSSION

*Conclusions.* The results of the study suggests that listeners have a difficult time recognizing intended emotions in improvised piano music, with only 33% decoding accuracy overall. Also, of the ‘basic emotions’ studies such as Gabrielsson and Juslin (1996) identified as generally being easiest to identify, only Happy had more than 50% average decoding accuracy. As mentioned, the other ‘basic emotions’ were both consistently confused with other emotions. These findings contradict the prior research by Gabrielsson and Juslin (1996) and Akkermans et al. (2018)’s findings, which found much higher levels of decoding accuracy. These results suggest that the primary research hypothesis of this study was incorrect, and that listeners often struggle with identifying the intended emotion of a performer in an improvised piano extract.

Additionally, these findings suggest that adding an improvised element makes the task more challenging, which contradicts the secondary research hypothesis of this study. That being said, if scores are compared directly with Akkermans et al.(2018)’s findings on piano, as opposed to their aggregation of all studied instruments, this

study was only accurate on one fewer category, so when the aforementioned ‘close seconds’ Sad and Angry (two emotions that other studies have shown are usually easily identified) are considered, it is possible that further study with either a larger sample size or more variety of either performers or instruments may challenge this idea.

It is worth mentioning that the emotion that was accurately identified the second most, both on an individual and aggregated level, was fearfulness; one of the most often confused emotions in Akkermans et al. (2018). The comparatively relatively high individual decoding accuracy scores in my study suggest that when encoded properly, fearful is actually a fairly easy emotion to decode and can be done so with a high degree of consistency. Further research into what acoustical elements featured in the fearful extract in this study, and which of those may not have been present in Akkermans et al. (2018) or Gabrielsson and Juslin (1996), could provide a basis for understanding this discrepancy between the two studies.

Also worthy of note is the relationship- or lack thereof- between Musical Training and Emotional scores from the Gold-MSI. Prior research suggests that both musical training and awareness of emotion in music can have a significant positive impact on decoding accuracy (Akkermans et al. 2018; Gilboa, Bodner & Amir, 2006). This study was unable to find a significant relationship between the two, however, there were positive correlations, although weak. This, combined with the prior research suggests that a larger sample size may have been able to show a more distinct, significant impact by these two scores.

*Limitations.* One limitation of this study was the small sample size, which may have impacted the accuracy of the results. In future, I would hope to allow more time for a wider array of respondents to complete the study. This would be especially salient for comparisons to Akkermans et al. (2018) due to its large sample size.

Additionally, this study only examined the recognition of intended emotions in improvised piano music, which may limit the generalisability of the findings. To improve this, using clips from instruments such as voice and violin, which both were found to be more easily decoded in Akkermans et al. (2018), may have given more insight. Moreover, some emotions may be better portrayed by certain instruments, as suggested by Gabrielsson and Juslin (1996). Additionally, only one performer did recordings for the piano improvisations in this study. Having a wider array of performers who may interpret the prompts differently is also likely to have a tangible impact on the results. In future studies I would certainly aim to make these changes, but due to time and resource constraints, just using piano was the practical solution at this time.

Also, I may consider asking performers to provide slightly longer clips in any follow up research. On the one hand, I decided that short clips were appropriate, especially as Gilboa, Bodner, and Amir (2006) found that shorter clips were not significantly more challenging to decode than longer ones, the clips used in this study ranged in length from 10 seconds to 17 seconds, whereas their shortest excerpt was 16 seconds. In future, use of clips closer to the 20 second mark may be advisable, to ensure that all excerpts fit comfortably above the 16 second mark, which Gilboa, Bodner, and Amir (2006) found with confidence to be plenty of time to decode an emotion, assuming it was encoded effectively by the performer.

Acoustical modelling in the same way that prior studies had done was not completed at this time. I hope in further research to be able to study the musical elements that the performer, and in future any other performers, made use of to gain a deeper insight into what may help listeners to decode these intended emotions. Additionally, carrying out an analysis of the musical elements utilised by the performer may go some distance to explaining why a number of emotions were confused, particularly in instances where similar emotions were confused regularly, such as Sad being confused for Solemn or Angry being confused for Fearful.

Finally, I would include more demographic-related questions if I were to carry this study out again. I chose to omit these as prior studies have found little evidence to support that factors such as cultural background, age and gender have an impact on emotional decoding, and I felt that keeping the survey as short as possible would increase the likelihood that as many participants as possible would complete it. However, different trends may have been found in this study, so for future research on the topic it may be helpful to consider these factors in spite of the lack of prior evidence suggesting significant relationships.

In summation, this study found significantly lower decoding accuracy than expected when compared to its forerunners, suggesting that adding an improvised element to this task may have made decoding more challenging. However, the study provides a basis for further research into the topic, and with changes mentioned above, yet more interesting results could be found. With a wider participant pool, a greater variety of recordings,

and further study into what musical and acoustic features were present in the musical excerpts used in the study may give some important and interesting insights into the differences in both encoding and decoding between interpretation of pre-composed music and improvised music.

#### REFERENCES

- Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., Busch, V., Lothwesen, K., Elvers, P., Fischinger, T., Schlemmer, K., & Frieler, K. (2018). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, 33(6), 1099–1118. <https://doi.org/10.1080/02699931.2018.1541312>
- Behrens, G. A., & Green, S. B. (1993). The Ability to Identify Emotional Content of Solo Improvisations Performed Vocally and on Three Different Instruments. *Psychology of Music*, 21(1), 20–33. <https://doi.org/10.1177/030573569302100102>
- Bodner, E. (2014). Emotion Recognition in Improvised Music: the Case of the Multicultural Israeli Society. *Journal of Cross-Cultural Psychology*, 45(4), 618–627. <https://doi.org/10.1177/0022022113519854>
- Gabrielsson, A., & Juslin, P. N. (1996). Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience. *Psychology of Music*, 24(1), 68–91. <https://doi.org/10.1177/0305735696241007>
- Gilboa, A., Bodner, E., & Amir, D. (2006). Emotional Communicability in Improvised Music: The Case of Music Therapists. *Journal of Music Therapy*, 43(3), 198–225. <https://doi.org/10.1093/jmt/43.3.198>
- Lewis, M., & Michalson, L. (1983). *Children's Emotions and Moods*. Plenum Press.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>