DURMS
Durham Undergraduate
Research in Music & Science

# Emotion Detection in Very Brief Excerpts of Unfamiliar Film Music

Anthony Sutch
*Durham University*

## ABSTRACT

This study aimed to discover how rapidly humans could detect specific emotions in excerpts of unfamiliar film music. It replicated elements of Filipic et al.'s (2010) study, specifically in relation to using a gating paradigm of excerpt lengths, testing human emotion detection rate at 1000ms, 250ms and the novelly short 150ms. 16 participants listened to these excerpts and selected which emotions were conveyed from the four emotion categories: happy, sad, tender, and tense–or choose the 'inconclusive' option. The ratio of 'correct' versus 'incorrect' or 'inconclusive' selections then informed the percentage detection rate for the emotion in each excerpt which, if above the median, meant sufficient detection occurred, or if below the median, meant that the emotion was insufficiently detected. This study found that its participants can detect emotion sufficiently at 1000ms and 250ms, but not sufficiently at 150ms. Moreover, it found that tense, to a considerable extent, and happy, to some extent, were sufficiently detected. Tender and sad, however, were not sufficiently detected, due partly to their interchangeability.

## 1. INTRODUCTION

Humans can detect emotions in sound extremely quickly as evidenced by multiple previous studies. Rapoport (2002) discovered that after 160ms of vocal sound, humans could differentiate between excitement and calmness, and similarly Wambacq et al. (2004) concluded that it took 200ms for their participants to detect vocal emotion. Humans can also rapidly detect emotions in instrumental music, as seen in Peretz et al.'s (1998) study that deduced that the difference between happy and sad piano music could be identified after hearing excerpts only 250ms long. Overall, an emotional response to music can be triggered by very short stimuli. The emotions perceived by the participants of these studies (and this experiment) were outwardly expressed by the excerpted music, as opposed to subjective internally felt emotions (Gabrielsson, 2001).

This study replicates elements of Filipic et al.'s (2010) study, specifically the second experiment conducted relating to judging emotion in brief musical excerpts. They used pieces of very moving and emotionally neutral instrumental music, ranging from solo piano music to orchestral music, in a gating paradigm ranging from 250ms to 5000ms. The dynamic level was equal for every excerpt, and they deduced that basic acoustic features like loudness did not affect the extent of emotion identification; rather they found that emotion perception in music was based on a combination of features including timbre and performance cues. From their results, it was concluded that participants could determine the emotional qualities of music for clips as fast as 250ms, which "extends the observation of fast-acting cognitive and emotional processes from face and voice perception to music perception" (Filipic et al., 2010).

Because the replicated study found that emotional judgements could be made at the shortest interval they used, 250ms, this experiment utilised even shorter excerpts of 150ms. Additionally, this experiment focuses on specific emotions, employing a discrete model, unlike the replicated project which used a dimensional model focusing on valence and arousal (Eerola & Vuoskoski, 2011). This will allow for more specific findings relating to the difference between certain emotions and their relative detection rates. This experiment also made an extension to the replicated project by using unfamiliar film music instead of instrumental music. This is because it is a relatively neutral genre in terms of opinion and therefore people are less likely to have preconceived notions towards it which would affect their emotional judgement. Moreover, it is argued that film music is inherently emotive and therefore there is definite emotional intent and limited amounts of emotional ambiguity for the excerpts of the study (Noad, 2008). Recognising that familiarity with music is a factor in participants' emotion recognition, this experiment–unlike the replicated project and previous music and emotion research which too often utilises well-known music–used soundtracks that were relatively unfamiliar and only recognisable by a film aficionado (Eerola, 2019).

The research question for this study asks how accurately can human participants detect emotions in short excerpts of music? More specifically, this investigation enquires out of a few distinct emotions which are the

most and least easily recognised? And which emotions are the most interchangeable or mixed up by the participants?

## 2. METHOD

The study consisted of a single online questionnaire created using Qualtrics which was filled out by 16 participants who took an average of 13.6 minutes to complete. They were recruited through volunteer sampling, after responding to either an email or social media post describing the study and asking for participants. The sixteen participants had an average age of 21.7 (SD = 7.9) and 63% were Female, 31% Male and 6% Non-Binary. They predominantly described themselves as music-loving nonmusicians at 31%, with both amateur and serious amateur musicians encompassing 25% of the participants respectively, and finally two described themselves as nonmusicians and one as a professional musician.

After filling out these questions, they listened to 48 musical excerpts taken from 16 different soundtracks. A practice excerpt came before the real study so they could be prepared for the fast speed of the excerpts and know to pay attention as they could only listen to the excerpt once – as was prewarned to them. The soundtracks of unfamiliar film music were taken from a dataset published in Eerola and Vuoskoski's study (2011). Each track was assigned an emotion and participants of said study had to describe the emotion they heard which contributed to an overall rating for each track. The tracks selected for this experiment were the highest rated tracks for each emotion, all of which had a mean rating of 7 or higher out of 9. The short excerpts were then selected randomly from within the soundtracks and edited to the correct length on Audacity. There were 3 excerpts per track, one was 1000ms long, another 250ms, and a final excerpt 150ms, which formed a gating paradigm. Like the replicated project, this study ensured that excerpts from the same tracks at different durations did not occur consecutively – to "avoid successive stimulations of the same memory entry" (Filipic et al., 2010, p. 335). Unlike said project however, this experiment did not play the excerpts in blocks relating to their speed going for shortest to longest, instead the order in which the participants heard the extracts was randomised using a random number generator. (Table 1).

**Table 1**

*A table showing the mean rating and emotions of 16 soundtracks turned into very short excerpts and the randomised order said excerpts were played to the study's participants*

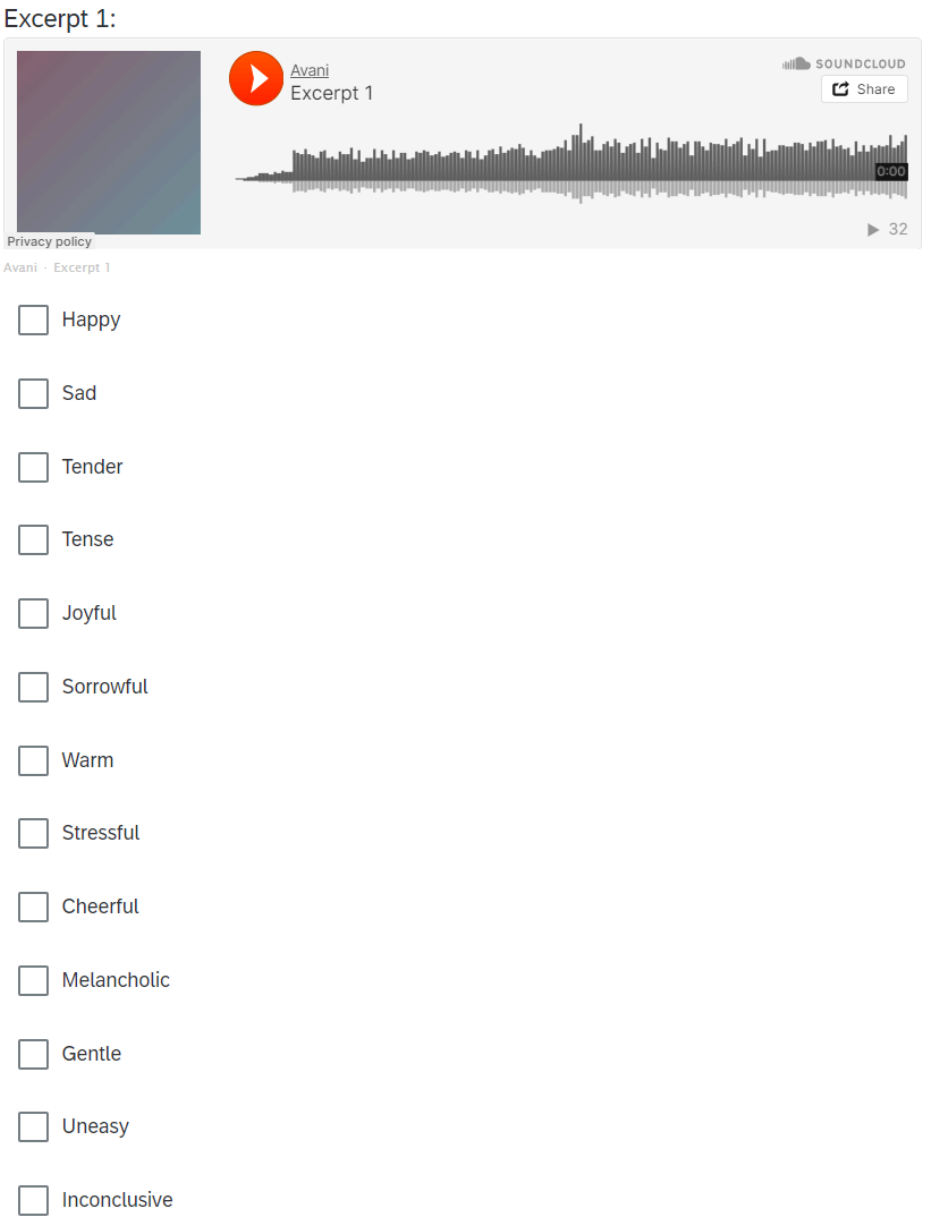
| Track Number | Mean Rating /9 | Emotion (Ranking) | Excerpt Number (1000ms) | Excerpt Number (250ms) | Excerpt Number (150ms) |
|---|---|---|---|---|---|
| 001 | 7.33 | Happy (1) | 16 | 36 | 48 |
| 005 | 7.17 | Happy (2) | 22 | 38 | 23 |
| 004 | 7.17 | Happy (3) | 28 | 13 | 15 |
| 003 | 7.17 | Happy (4) | 14 | 2 | 26 |
| 031 | 7.67 | Sad (1) | 24 | 46 | 5 |
| 032 | 7.50 | Sad (2) | 33 | 32 | 25 |
| 033 | 7.50 | Sad (3) | 11 | 19 | 45 |
| 034 | 7.33 | Sad (4) | 35 | 4 | 47 |
| 061 | 7.83 | Tender (1) | 42 | 17 | 39 |
| 062 | 7.60 | Tender (2) | 3 | 37 | 6 |
| 063 | 7.40 | Tender (3) | 27 | 1 | 29 |
| 065 | 7.17 | Tender (4) | 10 | 7 | 12 |
| 303 | 7.00 | Tense (1) | 21 | 34 | 41 |
| 302 | 7.00 | Tense (2) | 9 | 44 | 30 |
| 301 | 7.00 | Tense (3) | 31 | 40 | 43 |
| 304 | 7.00 | Tense (4) | 20 | 8 | 18 |

The targeted emotions for this experiment were: happy, sad, tender, and tense, with 4 soundtracks pertaining to each. Said emotions were chosen due to their distinctness from each other, so that if they were interchangeable in the experiment it would reveal something significant about human emotion detection. After hearing an

excerpt once, the participants had to select which emotion they thought was conveyed from a list including the above four emotions and two synonyms for each emotion (Figure 1). The synonyms were included, as multiple options reduced the chance of participants selecting the 'correct' answer by guessing and prevented the participants noticing the clear distinct four emotions and thereby going with one for the sake of it. Furthermore, an 'inconclusive' option was included, along with encouragement to use it in the pre-test instructions, to discourage participant guessing. They had to select at least one option and were advised to select up to three options. This meant that the more correct options they selected, the more indicative of that particular emotion the excerpt was, and oppositely the more incorrect or 'inconclusives' they selected, the less strongly that emotion was detected overall.

**Figure 1**

*An image displaying the emotions and synonyms used as options for the participant's selection in this experiment*



## 3. RESULTS

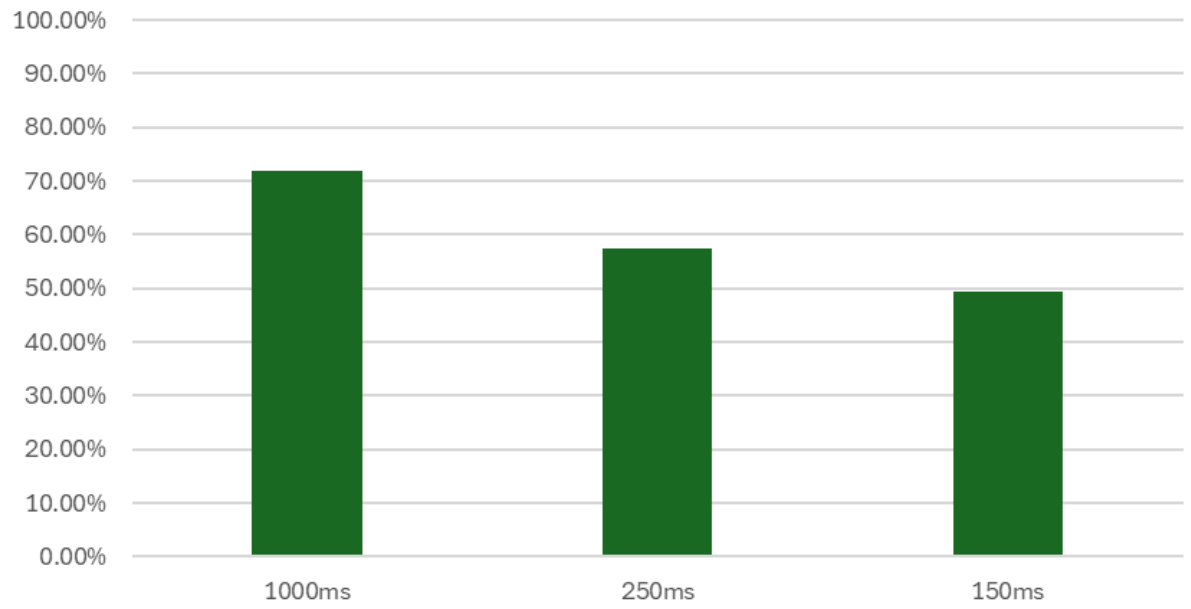
The hypotheses for this study were that the 150ms duration excerpts would have the largest amount of insufficiently detected emotions and that the 1000ms excerpts would have the least incorrect or 'inconclusive' answers. Additionally, happy and sad would be the most recognisable emotions because they are the simplest emotions, and therefore potentially the easiest to sonically represent. Similarly, tender and tense were considered the least obviously recognised emotions as they are more complex and thus more difficult to represent.

The proportion of correct answers selected versus the proportion of incorrect and 'inconclusives' was calculated to give a percentage accuracy of correct emotion detection for each excerpt. Filipic et al. (2010) used the median to categorise items rated 'neutral' if below the median and 'moving' if above. Therefore, for this experiment, if the percentage accuracy was above 50% for the emotion targeted in a particular excerpt, it was considered 'sufficiently detected', whilst below 50% was considered 'insufficiently detected'.

Overall, in terms of how accurately the participants could detect emotions in short amounts of music, the excerpts of 1000ms had an average detection accuracy of 71.8%, 250ms excerpts had a detection rate of 57.3%, whilst the 150ms excerpts had a correct detection accuracy of 49.4% (Figure 2).

**Figure 2**

*A graph showing the overall accuracy of emotion detection (%) for each duration*




In terms of emotion detection, tense had a detection rate of 92.7%, happy of 60.3%, tender of 47.0% and sad of 37.8% (Figure 3). For the tense excerpts, the 150ms clips had practically the same detection rate as the 250ms ones. For every other emotion, however, there was a correlation between detection percentage and excerpt length–with happy going from 73.8% for 1000ms to 58.1% for 250ms to 48.9% for 150ms, sad from 56.4% to 34.5% to 22.5%, and tender from 61.1% to 45.2% to 34.8% for the same speeds (Figure 4).

**Figure 3**

*A graph showing the overall accuracy of emotion detection (%) for each emotion category*

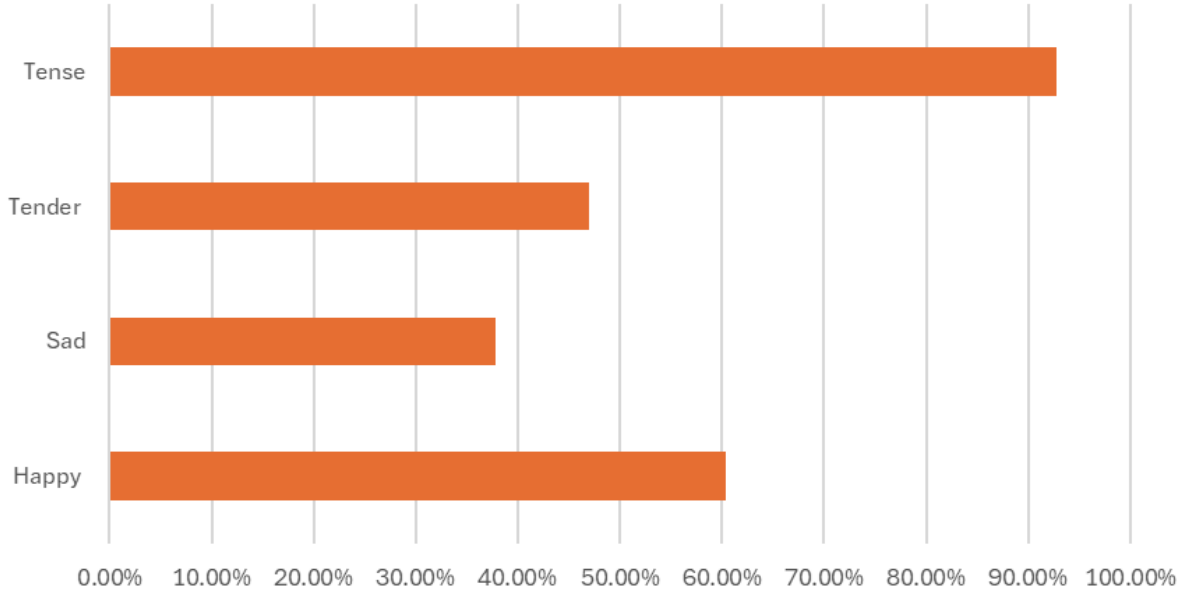
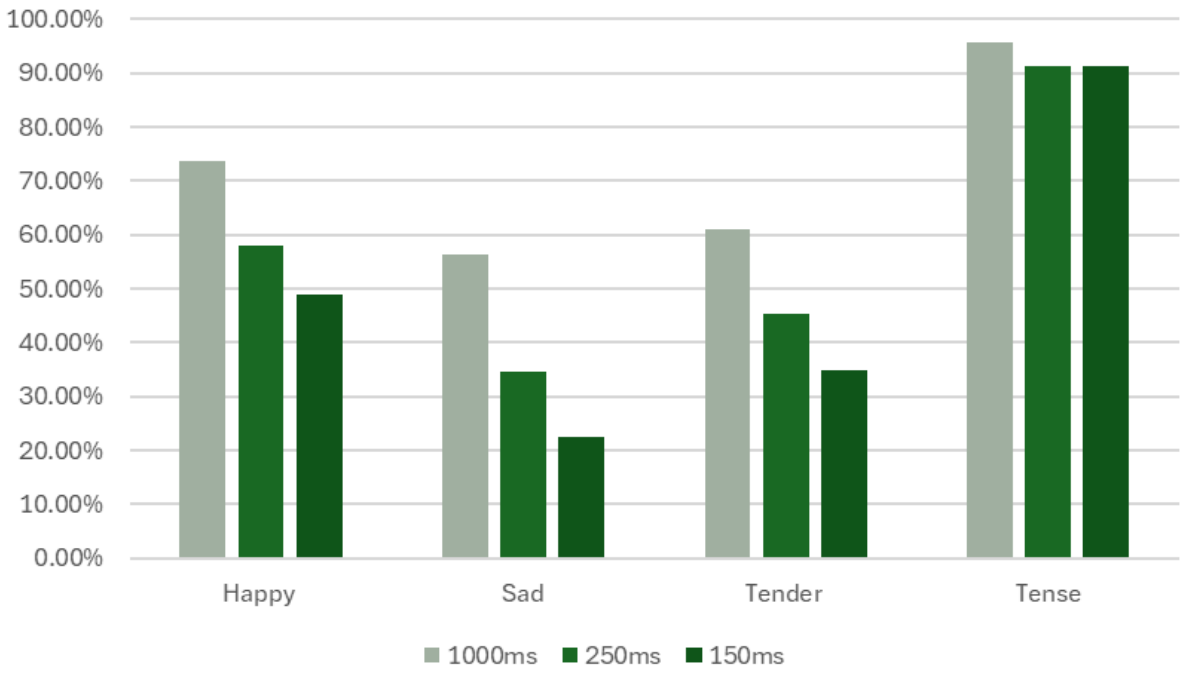



**Figure 4**

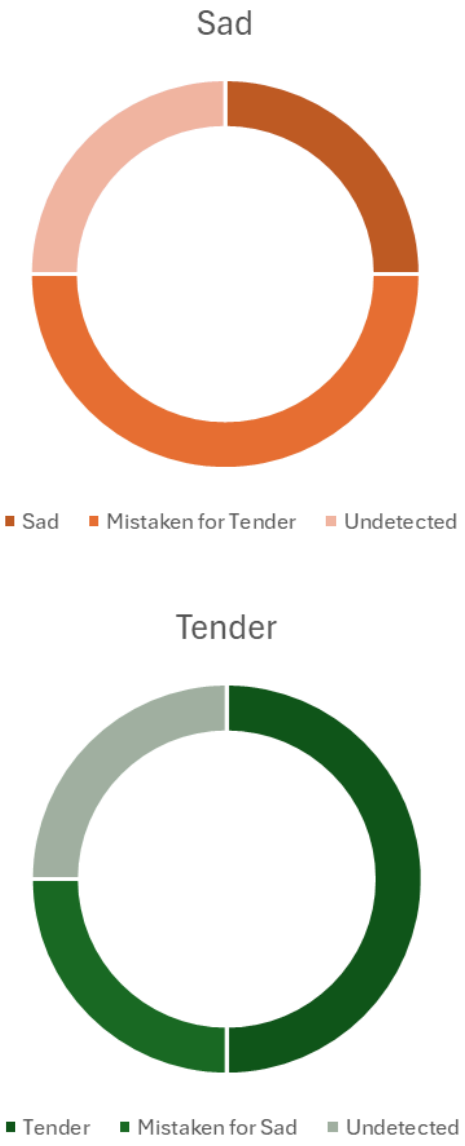
*A graph showing the average accuracy (%) of emotional detection for each duration (ms) within each emotion category*



In terms of emotion interchangeability, sad and tender were mixed up the most by the participants. Interchangeability was measured by tallying the instances that one of the words belonging to another emotion category had the same or higher number of selections as the most chosen word from the 'correct' emotional category. Sad was mistaken for tender 50% of the time, in 6 of the 12 sad excerpts, and tender was mistaken for sad in 25% of the tender clips (Figure 5). This mostly occurred in responses to the 250ms and 150ms clips, rather than the 1000ms excerpts, which fits with the trend of decreasing detection rate as the excerpt length shortens.

**Figure 5**

*Charts displaying the overall designations of the 12 sad and 12 tender excerpts*



# 4. DISCUSSION AND CONCLUSION

In summary, the participants could detect expressed emotions easily in 1000ms of music and to some extent in 250ms, which was hypothesised and similarly found in the replicated study. As the result for 150ms was below the median however, the participants could not conclusively detect emotion in 150ms of music, due to more inaccurate emotions and 'inconclusives' selected than correct identifications. The longest duration excerpts having the least insufficiently detected results and the shortest having the most creates a trend that sufficient emotion detection decreases as the length of music decreases. This corroborates with Filipic et al.'s (2010) findings which displayed the same trend spanning from 5000ms to 250ms, and Nordström and Laukka's (2019) study which also found this tendency to be true.

Similarly to Filipic et al. (2010), this experiment found that when an emotion was identified at a short duration it was very likely identified at a longer duration too. In terms of specific emotions, the tense emotion in music was detected by the participants to an extremely large extent at all three excerpt speeds and was the most recognised emotion by far in this experiment, opposing the initial hypothesis. Five audio clips had a 97% detection rate for tense, with two of these being 150ms long which contributes to 150ms and 250ms having similar detection rates in tense excerpts (91.3% and 91.2%). This indicates that the participants could probably hear 'tense' in even shorter excerpts, indicating that only the smallest amount of auditory information is needed for its identification. This is likely due to the high survival value of humans being able to perceive auditory cues of a threatening environment that are emulated in tense music–for example sudden dynamic changes and dissonance (Loui, 2022).

The participants could detect happy to some extent overall, with the detection rate for 1000ms and 250ms over the median, and 150ms just under the median. Therefore, happy can be conclusively detected in clips of music 250ms or faster, but not conclusively in audio shorter than that.

Tender was not sufficiently detected in brief excerpts overall, as the detection percentage was below the median for both 250ms (45.2%) and 150ms (34.8%). Tender, however, can be identified in 1000ms clips with a 61.1% detection rate which is above the median.

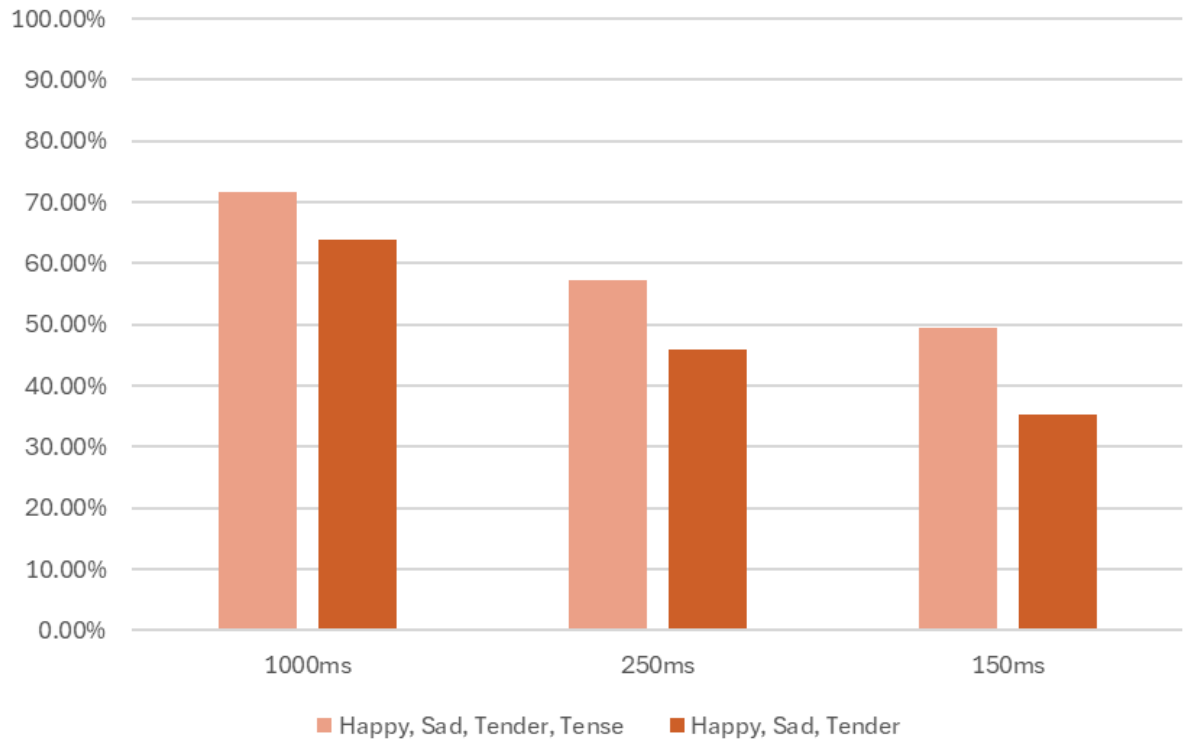
Contrary to the hypothesis, sad was the least easily recognised emotion and cannot be sufficiently detected to a large extent in short excerpts. The clips for 250ms and 150ms were comfortably below the median, the latter of which had one result with a 10% detection rate. The 1000ms excerpts targeting sad, however, were detected above the median.

The interchangeability of sad and tender reveals that in the very brief excerpts of 250ms and 150ms there is not enough auditory information for the participants to process the difference between those emotions. Therefore, the participants could not sufficiently detect each emotion, and they were the least easily recognised emotions in the study.

The discrepancy between the average detection rate for the tense excerpts compared to the other emotions is due to the latter being basic emotions, whilst tense is the extreme on a bipolar scale usually associated with dimensional models of music emotion study (Eerola & Vuoskoski, 2011). Removing tense from the study has large ramifications on the average emotion detection percentages for the durations. Compared to when tense is included in the averages, detection accuracy falls from 71.8% to 63.8% for 1000ms, from 57.3% to 45.9% for 250ms, and from 49.4% to 35.4% for 150ms (Figure 6). This is significant for 250ms as it takes it below the median, implying that basic emotions are insufficiently detected at 250ms by the participants in this study. Moreover, the detection rate for 150ms goes from close to the boundary and needing further study, to comfortably in the range of insufficient emotion detection. Without replacing tense with a similar basic emotion, like fear or anger, however, these conclusions cannot be conclusively drawn as they do not represent the 'detectability' of a wide range of basic emotions.

**Figure 6**

*A graph showing the overall accuracy of emotion detection (%) for each duration (ms) both including and excluding tense excerpts*




Nordström and Laukka's (2019) study found that anger, happiness, and sadness could be observed after ≤100ms of stimuli and other emotions including fear and tenderness needed longer to be detected but were still recognised relatively fast (after ≤250ms). This opposes this experiment's findings as neither happy or sad were conclusively detected by the participants at 150ms and tender did not have a lower detection rate than sad. Tense is somewhat applicable to anger, possibly agreeing with Nordström and Laukka's results; however, it is similarly applicable to fear which contradicts this study's results massively implying tense, or fear, is less easily detected than happy and sad.

Due to the limited pool of previous studies using audio excerpts as brief as 150ms, more data in this area is needed to conclusively say that humans cannot sufficiently detect emotion in short durations of music. This speaks to the wider point that for the general validity this study needs to be repeated, to increase the participant pool. Moreover, an extension to this study could be to look at other emotions; however, these emotions would need to be distinctive enough so that if interchangeability occurred it would have significant drawable conclusions. Additionally, the impact of gender or culture on the detection of specific emotions in short excerpts could be considered. For the latter, the findings of Balkwill & Thompson (1999) and Laukka et al. (2013) could be tested as part of an experiment in this field.

## REFERENCES

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception: An Interdisciplinary Journal*, *17*(1), 43–64. https://doi.org/10.2307/40285811

Eerola, T. (2019, April 16). *Music and emotion stimulus sets consisting of film soundtracks*. [Data Set]. Open Science Framework. https://osf.io/p6vkg/wiki/home/

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review*, *17*(3), 335–341. https://doi.org/10.3758/pbr.17.3.335

Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, *5*(1_suppl), 123–147. https://doi.org/10.1177/10298649020050s105

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449. https://doi.org/10.1037/a0031388

Loui, P. (2022, February 10). *Consonance and dissonance*. Music Technology Online Repository. https://mutor-2.github.io/ScienceOfMusic/units/04/

Noad, B. J. (2008). What are the theoretical relationships between music, film music, and emotions, and what are the implications for a semiotic analysis of affect in narrative film? In *Bridging the gap between ideas and doing research: Proceedings of the 2nd annual postgraduate research conference*, 63–71. https://hdl.handle.net/1959.11/4043

Nordström, H., & Laukka, P. (2019). The time course of emotion recognition in speech and music. *The Journal of the Acoustical Society of America*, *145*(5), 3058–3074. https://doi.org/10.1121/1.5108601

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. https://doi.org/10.1016/s0010-0277(98)00043-2

Rapoport, E. (2002). Singing, mind and brain - unit pulse, rhythm, emotion and expression. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology* (pp. 451–468). Springer.

Wambacq, I. J. A., Shea-Miller, K. J., & Abubakr, A. (2004). Non-voluntary and voluntary processing of emotional prosody: An event-related potentials study. *NeuroReport*, *15*(3), 555–559. https://doi.org/10.1097/00001756-200403010-00034