

Music and Inner Imagery

Vanilla Riis Mortensen

Keywords: music, visual mental imagery, emotions, VVIQ

1 Background

Visual imagery has been largely neglected within music psychology research despite importance for understanding music-evoked emotions[1]. In relation to music, visual mental imagery refers to the mechanism through which music evokes internal images in the listener.[2]

Large areas still remain almost unexplored, one of them being the correlation of emotion and visual imagery. Research suggests that mental imagery and emotional processes during music listening are closely intertwined[1], causing us to ask the following questions:

- Does the type of emotion matter?
- Does emotion come before imagery or do they occur at the same time?
- Are imagery and emotion always connected in the same way?

1.1 Effects of Musical Structure and Emotional Tone on Visual Imagery and Mind-Wandering

A study by Dahl et al. from 2022[3] looks at what kind of visual mental imagery people experience when listening to music, and how musical structure (tempo, timbre, harmony, instrumentation) influences what people visualize.

135 participants were randomly assigned to listen to one of four 2-minute songs with different characteristics. Participants were then asked if they experienced visual imagery. If yes, they were asked to describe it freely, then complete a card-sorting task choosing descriptors for the scene (e.g., “in the forest,” “tribal,” “sunny,” etc.) and rate emotions (valence, tension, nostalgia, etc.).

The songs *Deep Blue Day* and the Test Song, which had similar characteristics (calm, major key, slow tempo), created peaceful landscapes and positive imagery, and were rated as high valence and low tension. The song *Tanca*, which had characteristics like minor mode, rhythmic patterns, and darker timbres, created imagery of humans, rituals, battles, fire and conflict and was rated as low valence and high tension. The Modified Song with a mixture of Test Song + *Tanca* characteristics created blended imagery that transitioned from peaceful to ominous and was rated as intermediate.

The authors concluded that musical and acoustical characteristics strongly influence the imagery people experience, and people listening to the same music tend to visualize similar scenes.

Taruffi et al. explore the question of whether sad music and happy music make our minds wander differently in their study from 2017[2]. They did three experiments.

The first experiment was an online study with 216 participants. Participants listened to sad or happy instrumental music with their eyes closed. After each piece, they answered questions about mind-wandering strength (“Was your attention on the music or somewhere else?”), meta-awareness (“How aware were you of your own thoughts?”), content of thoughts (e.g. about the past/future, self-referential, movement) and form (words vs. imagery). Because sad music is usually slower and happy music faster, they repeated the experiment on 140 participants where they tested sad slow vs happy slow and sad fast vs happy fast. This let them test whether the emotion or the tempo drives mind-wandering. In the third experiment, 24 participants were fMRI-scanned while listening to 4 minutes of sad music and 4 minutes of happy music. The music was matched for tempo and loudness to ensure that the results were not affected by acoustic features but came from the emotional quality of the music (sad vs. happy).

The study revealed several findings. Sad music leads to inward, personal, reflective thinking, while happy music keeps people externally oriented and focused. The form of thoughts was dominantly imagery for both sad and happy music. Even when the tempo was the same, sadness still produced more mind wandering. Sad music increased activity in the core nodes of the Default Mode Network which is the network underlying daydreaming, self-reflection, autobiographical memory and internally generated thought.

Together, these studies demonstrate that musical features and emotional tone not only shape what we see in our mind but also how our mind works while we are listening.

1.2 Does Imagery Cause Emotion, or Vice Versa? Evidence From Recent Studies

In 2008, Juslin and Västfjäll proposed one of the most influential theories about how music triggers emotions.[1] They explained that music can induce emotion through multiple psychological mechanisms where visual imagery is one of the ways music makes people feel emotions. This means that the music is not directly creating the emotional feeling; instead, the image does. According to them, the music stimulates mental images such as landscapes, people, memories, colors, and narratives. These images themselves carry emotional content (positive, nostalgic, sad, uplifting) which spills over into your own emotional state.[4]

Day and Thompson challenged Juslin and Västfjäll’s theory that imagery causes emotion with their study in 2019[5]. In this study, 49 participants listened to 30 different short pieces of music, each 20 seconds long, taken from classical or pop genres, and were told to press a key as soon as they recognized an emotion in the music, felt an emotional reaction, and experienced a visual mental image. They found that reaction times were fastest for recognizing an emotion, second fastest for feeling the emotion, and slowest for experiencing a visual image. This suggests that people typically feel emotions before images appear. However, it is important to note that they measured simple, “low-level” emotions like pleasantness, energy, and tension. More complex emotions like nostalgia, which mixes happiness, sadness, longing, etc., might need more time to develop and could interact differently with imagery.[1]

In a study by Vroegh from 2018 (presented as a poster)[6], they studied the direction of the relationship between imagery and emotion. They did an online survey with 602 participants who listened to music and reported their emotional responses, how vivid/frequent their imagery was, and their attentional focus. Then Vroegh used structural equation modeling to test the causal relationships. They found that there were two different patterns depending on the emotion type.

When emotions were clearly positive, they would lead to imagery. However, when emotions were “mixed” (bittersweet, nostalgic, ambivalent), the imagery actually influenced the emotional response. This suggests that the relationship is not one-directional. It changes depending on whether emotions are simple and clear (e.g., happy) or complex (e.g., nostalgia, bittersweetness).

1.3 Hypotheses

The question I want to further explore is whether visual mental imagery comes before emotion, whether emotion comes before imagery, or whether the two processes co-occur during music listening. The findings across studies suggest that the answer may depend on the type of emotional response the music evokes.

Day and Thompson (2019) found that people generally feel an emotion before forming a visual image, but their study focused on simple emotions. In contrast, Vroegh (2018) showed that the direction of influence can reverse for more complex or mixed emotions: when emotions are clearly positive, they tend to lead to imagery, but when emotions are mixed or bittersweet, imagery may instead shape the emotional response. Taruffi et al. (2017) further demonstrated that sad music promotes inward, self-referential mind-wandering, while happy music keeps listeners outwardly focused, even when tempo is controlled.

Taken together, these studies suggest that happy music with simple, positive emotions may produce emotion first, while sad music with more complex emotional states could allow imagery to arise first.

Based on this reasoning, the study proposes the following hypotheses:

H1: During sad music, visual mental imagery will occur before emotional experience.

H2: During happy music, emotional experience will occur before visual mental imagery.

2 Methods

2.1 Conduction of the Online Test

Participants completed an online test developed in Unity. Before deploying a pilot test was conducted, on one person, to catch the most obvious mistakes or confusion points. Recruitment of participants was conducted through personal messages and a Facebook post containing a link to the test.

At the beginning of the test, participants were informed that they would not receive any personal reward for their participation, but that their responses would be greatly appreciated. They were also informed that the test would take approximately 10–15 minutes to complete. Participants were then given the option to continue or withdraw from the study.

Participants were first asked to report their gender and age. They then completed the Vividness of Visual Imagery Questionnaire (VVIQ)[7]. It was explicitly stated that having visual imagery, or having more vivid visual imagery, is not inherently more desirable than not experiencing imagery.

Next, participants were asked about their listening setup (headphones, speakers, or earbuds) and the environment in which they were listening (e.g., at home alone or with others, in a quiet or busy public space). They were then instructed to press the E key whenever they experienced an emotion and the I key whenever they experienced inner imagery. Participants were encouraged to close their eyes to enhance the listening experience, and after pressing “start,” they were given a five-second delay before the music began.

Participants then listened to one two-minute musical excerpts, either one sad or one happy, originally used by Taruffi et al. (2017), presented in randomized order. Afterwards, for each emotion they reported, participants were asked to select the type of emotion experienced, as well as its intensity and valence (on a five-point Likert scale), and were asked to shortly describe any visual imagery they experienced during listening.

Finally, participants were thanked for their participation and given the opportunity to leave an open-ended comment.

The test can be accessed here: <https://vanillariis.itch.io/sound-web-test>

Note: The test was designed so that if participants felt none of the available emotion categories (longing, sadness, nostalgia, happiness, anger, or neutral) accurately described their experience, they could select “Other.” This option opened a text field where participants were asked to describe the emotion in their own words. Unfortunately, due to a coding error, these written responses were not saved.

2.2 Musical Excerpts

The original intention was to use the exact musical excerpts from the source study, as these excerpts were matched for tempo and loudness, ensuring that any differences between conditions would reflect emotional tone rather than acoustic features. However, the expert ratings used in the original study could not be obtained, and the author did not respond to email inquiries requesting access to this information.

As an alternative, two musical pieces corresponding to the excerpts with the longest playback durations were selected. The pieces used were Hamlet by Ennio Morricone[8] (sad) and À La Folie by Michael Nyman [9] (happy).

Both pieces have been previously validated as effective in eliciting mental imagery and their respective emotional states, while minimizing potential bias arising from participants’ prior familiarity with the music.

3 Results

The data was analyzed using R studio[10].

3.1 Participants:

There was a total of 32 participants. 21 men, 10 females and 1 non-binary. In the age 18-64, with most participants being in the age 25-34. An overview of the age and gender of the participants can be seen in Figure 1.

17 of the participants got the Happy music and 15 of the participants got the Sad music.

Participant characteristics		
	N	%
GENDER		
Man	21	65.6
Non-binary	1	3.1
Woman	10	31.2
AGE GROUP		
18 - 24	9	28.1
25 - 34	15	46.9
35 - 44	1	3.1
45 - 54	4	12.5
55 - 64	3	9.4
Percentages are based on the total sample.		

Figure 1: *Overview of the age and gender of the participants.*

3.2 Visual Mental Imagery during music listening

28 of the 32 participants reported experiencing visual imagery. Of the four participants who did not, two had very low VVIQ scores, while the remaining two had average scores indicative of typical visual vividness. However, these two participants did not provide comments in the add-on section that might explain the absence of imagery despite their average VVIQ scores.

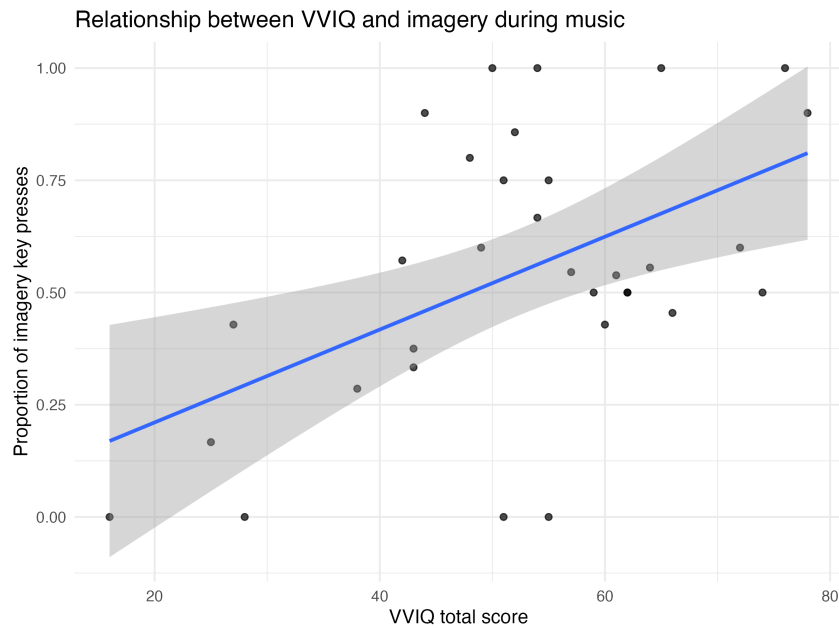


Figure 2: *The relationship between visual imagery ability (VVIQ total score) and the proportion of imagery key presses during music listening. Each point represents one participant, plotted by their VVIQ score on the x-axis and the proportion of imagery responses on the y-axis. The fitted regression line indicates a positive association between VVIQ and imagery responses: participants with higher VVIQ scores tended to report imagery more frequently during the music task.*

A Spearman rank-order correlation was conducted to examine the relationship between

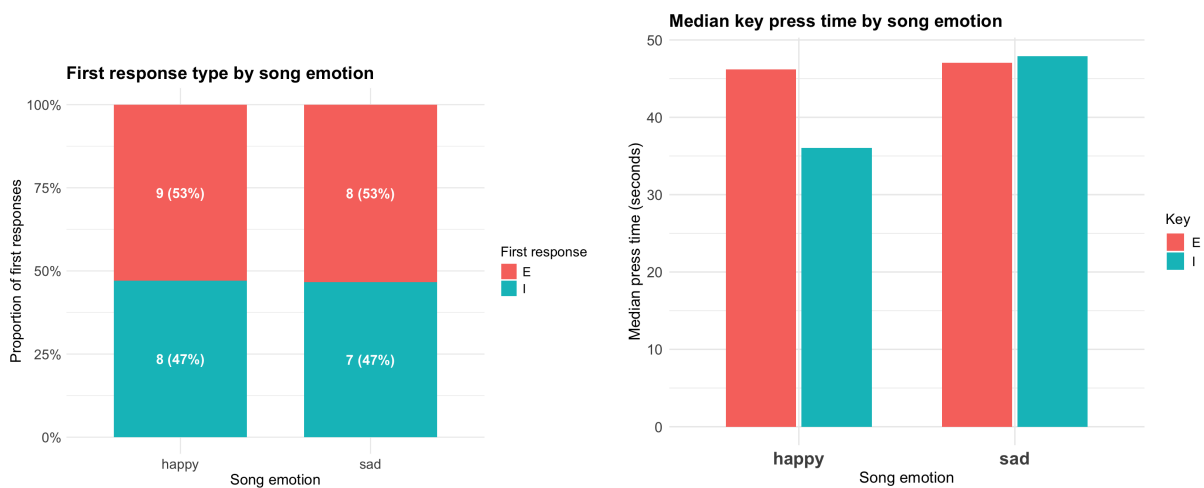
imagery ability (VVIQ total score) and imagery-related key presses during music listening. A visualization of the correlation can be seen in Figure 2.

To account for individual differences in overall response frequency, imagery responses were normalized by computing the proportion of imagery-related key presses relative to the total number of key presses for each participant.

A significant positive association was observed between trait imagery vividness (VVIQ) and the proportion of imagery-related responses during music listening (Spearman's $\rho = .39$, $p = .027$). Consistent with this finding, higher VVIQ scores were also associated with a lower emotion–imagery balance score, reflecting a relative preference for imagery over emotion responses.

This suggests that individuals with more vivid trait imagery tend to rely more strongly on imagery when experiencing music, rather than simply responding more frequently overall.

3.3 Imagery-first Versus Emotion-first



(a) The proportion of imagery-first versus emotion-first responses, for the sad and happy music. (b) Response timing for emotion and imagery for the sad and happy music.

Figure 3: The first figure shows if the participants first felt an emotion or first saw imagery during the music listening for the sad and happy condition respectively. The second figure shows the median key press time for all the responses during the happy and the sad condition.

The proportion of imagery-first versus emotion-first responses did not differ at all between happy and sad music as seen in Figure 3. However, analyses of response timing showed that imagery responses occurred slightly earlier overall during sad music, whereas emotional responses occurred much earlier during happy music.

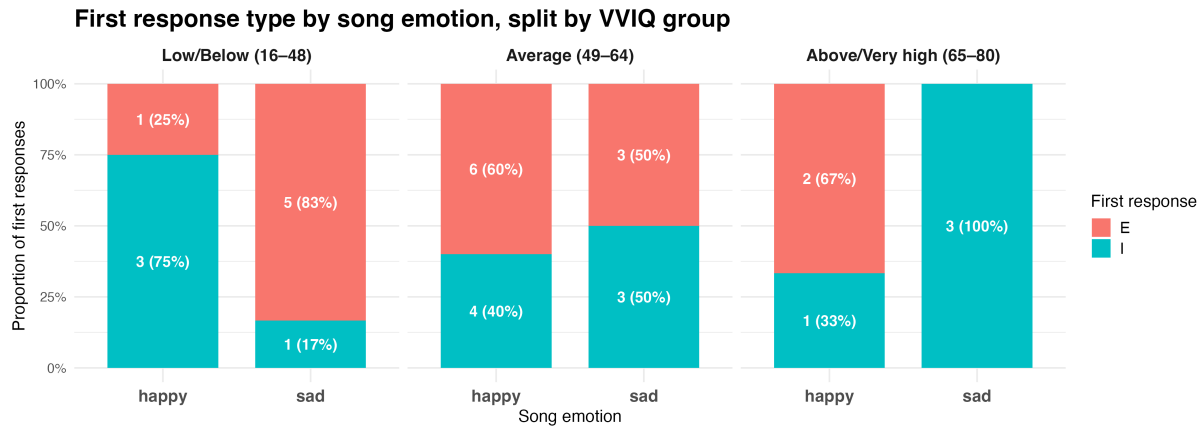


Figure 4: *The proportion of imagery-first versus emotion-first responses, for participants with low, average and high VVIQ scores, in the respective sad and happy condition.*

However, a more detailed analysis revealed a clear pattern when participants were divided into low/below average, average, and high/above average VVIQ groups. Participants with low VVIQ scores tended to report imagery before emotion during happy music, but emotion before imagery during sad music. Participants with average VVIQ scores showed an almost equal distribution between imagery-first and emotion-first responses in both conditions. In contrast, participants with high VVIQ scores tended to report emotion before imagery during happy music and imagery before emotion during sad music. A bar-chart showing the distributions for each category can be seen in Figure 4

3.4 Emotions in The Happy Versus Sad Music Conditions

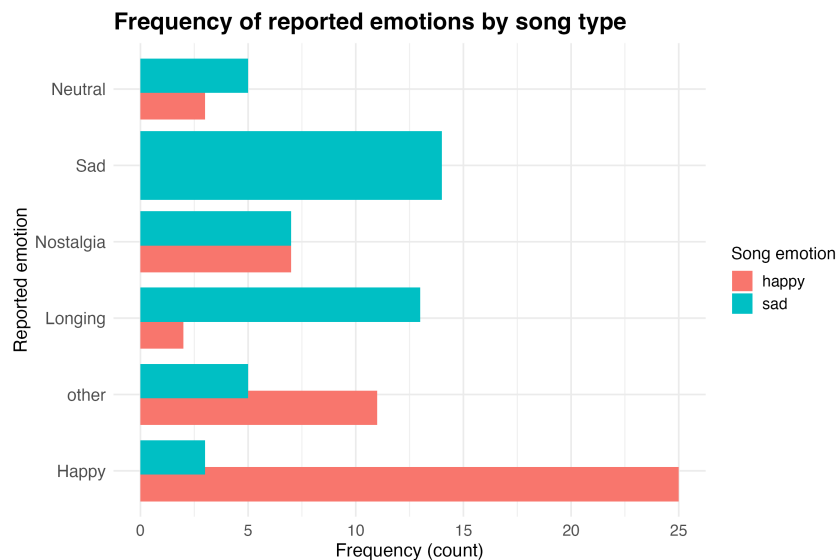


Figure 5: *This bar chart displays how often each emotion category was reported during the happy and sad music conditions. The y-axis lists the reported emotion categories. The x-axis shows the frequency. Bars are color-coded by song emotion (happy vs. sad).*

Figure 5 provides an overview of the emotions reported by participants under each musical condition. The distribution of emotional responses differs clearly between happy and sad music.

During happy music, participants most frequently reported happy emotions, with substantially fewer reports of and no reports of sad emotions. In contrast, during sad music, reports shift toward sad and longing emotions, both of which occur much more frequently than in the happy condition. Some also reported happy emotions during the sad music. Neutral responses were relatively infrequent in both conditions but appeared more often during sad music.

The Other category occurred in both conditions but was more frequent during happy music. Unfortunately, the specific descriptions of emotions classified as Other were not accessible. It is also noteworthy that one participant reported only Other emotions during listening; this participant alone accounted for five of the Other responses in the happy condition. If this participant were excluded, the number of Other responses would be nearly equal across the happy and sad conditions. Nostalgia was reported at similar rates in both conditions.

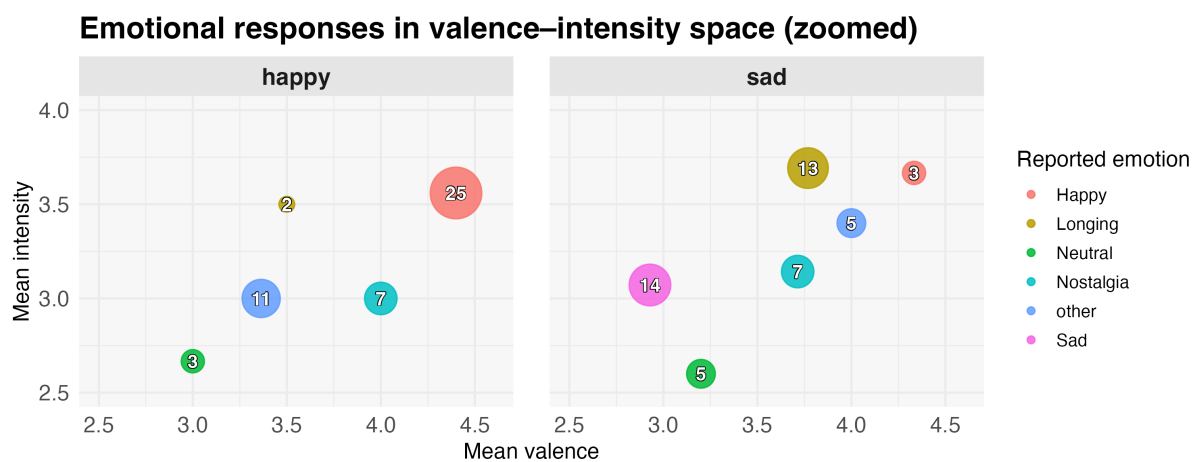


Figure 6: *This figure maps reported emotions into a two-dimensional affective space, separately for happy and sad music. The x-axis represents mean valence (how positive or negative the emotion was rated). The y-axis represents mean intensity (how strongly the emotion was felt). Each bubble represents one reported emotion category. Bubble size reflects the frequency with which that emotion was reported (numbers inside the bubbles show the exact counts).*

In Figure 6, all the emotions are mapped into a 2-dimensional space according to their mean intensity and mean valence.

In the happy music condition, Happy is located at high valence and relatively high intensity and is the most frequently reported emotion. Longing is also associated with relatively high intensity, although it is reported much less frequently. The remaining emotion categories (Nostalgia, Other, and Neutral) cluster at more moderate intensity levels, with Neutral being the least intense. Their valence ratings are generally positive, except for Neutral, although lower than those observed for Happy.

In the sad music condition, Longing is positioned at relatively high intensity and moderately high valence. Sad is characterised by low valence and moderate intensity. Reports of Happy emotion during sad music are less frequent but show relatively high intensity, exceeding that observed in the happy music condition. Nostalgia occupies an intermediate position, with positive valence in both conditions, though its valence is slightly lower and intensity slightly higher during sad music compared to happy music.

Across both conditions, Neutral emotions are positioned similarly, with near-neutral valence and low intensity, indicating consistency in how neutral experiences were rated regardless of musical context.

3.5 Visual Mental Imagery Themes

I reviewed all participant descriptions of visual imagery and extracted keywords for each. The resulting keywords for the happy and sad conditions are presented below.

Across both musical conditions, participants commonly reported imagery related to being transported back in time, entering a fantasy-like world, and experiencing natural landscapes and musical performance. However, the qualitative characteristics of the imagery differed markedly between conditions.

In the happy music condition, imagery was characterized by a high degree of movement and action. Scenes often involved battles that consistently resulted in victory, as well as flying, sunrises, dancing, and orchestral performances.

In contrast, imagery in the sad music condition was notably more static and somber. Participants frequently described cloudy skies, darker visual tones, themes of death, grief, and loss, and solitary violin performance.

Happy Condition

- Battles
- Victory
- Horses riding
- Sunrise
- Dancing (Ballroom)
- Orchestra playing
- Dragons
- Flying
- Running water (rivers)
- Nature (Forest, countryside)
- Bridgerton

Sad Condition

- Cloudy sky
- Dark colors
- Medieval towns and castles
- Funeral
- Death
- Grief
- Loss
- Someone playing the violin
- Nature (landscapes)

3.6 Comments from the participants

At the end of the test, participants were given the opportunity to leave an open-ended comment. A small number of participants chose to do so. Several comments indicated that participants felt confused by the lack of clear feedback when pressing the buttons, which made it difficult to know whether their input had been registered correctly. One participant noted that some of their emotion or imagery key presses did not seem to be recorded as intended.

Additionally, one participant stated that the musical excerpts lacked sufficient nuance for them to experience a strong emotional response. Two participants reported that aspects of the test

were sometimes unclear, and one participant mentioned that the task was cognitively demanding and difficult to manage from a memory perspective.

4 Discussion

This study examined whether visual mental imagery or emotion comes first during music listening, and whether this order differs between happy and sad music. Based on earlier research, it was hypothesized that imagery would occur before emotion during sad music, while emotion would occur before imagery during happy music. The results only partly supported these expectations.

4.1 Individual Differences in Imagery Ability May Influence The Temporal Relationship between Emotion and Imagery

The proportion of imagery-first versus emotion-first responses did not differ between happy and sad music when the participants were in one big group. However, response timing showed that imagery responses tended to occur slightly earlier during sad music, while emotional responses occurred much earlier during happy music. This indicates that, even if the overall order does not change, the timing of imagery and emotion does vary depending on the emotional tone of the music.

When participants were divided based on their VVIQ scores, a different pattern emerged. The data suggest that participants with low VVIQ scores—that is, those who experience little or no visual mental imagery—tended to experience emotion before imagery during sad music, but imagery before emotion during happy music. For participants with high VVIQ scores, who report vivid visual mental imagery, the pattern was reversed: imagery tended to occur before emotion during sad music, and emotion before imagery during happy music. Participants with average VVIQ scores showed a more balanced pattern, falling between these two groups.

Previous research suggests that simple emotions are often felt quickly, while more complex emotions may develop more gradually[5]. However, these findings suggest that individual differences in imagery ability also influence the temporal relationship between emotion and imagery during music listening, and that this relationship may differ depending on both emotional tone and listener characteristics. However, it can be hard to tell if the data would still tell this story if there were more participants, and a final conclusion cannot be made from this small dataset.

4.2 Sad and happy music can engage listeners internally, but in different ways.

In the present study, sad music primarily evoked emotions such as sadness, longing and nostalgia, which are more complex emotional states and may allow visual imagery to emerge earlier or play a stronger role in shaping the emotional experience. However, happy music also elicited feelings of nostalgia in some participants, possibly because the music was associated with old times. In addition, some participants reported feeling happy during the sad music condition. Despite these overlaps, the overall pattern of reported emotions corresponded to the intended emotional character of the musical excerpts. That said, it is noteworthy that a substantial portion of the reported emotions in the happy condition could not be fully interpreted because the qualitative data for the “Other” category was lost. Based on the available ratings, these responses generally

showed neutral intensity and slightly positive valence, suggesting that they are unlikely to have strongly biased the overall results.

The qualitative imagery descriptions partly support the findings of Taruffi et al. (2017), who showed that sad music promotes inward, self-referential thinking, while happy music tends to keep listeners more externally oriented. In the present study, the imagery across the two conditions was somewhat similar, describing being in the old times, fantasy elements and natural landscapes. This may be related to the use of an acoustic, orchestral sound, which is often culturally associated with historical settings and fantasy narratives. However, imagery reported during sad music was generally more static and somber, often involving themes of darkness, loss, and solitude. In contrast, imagery during happy music was more action-oriented, featuring movement, victory, and social or fantastical scenes. However, these vivid and imaginative scenes suggest that happy music did not necessarily lead to an outward attentional focus. Instead, participants appeared to become absorbed in internally generated fantasy worlds. This challenges a simple inward–outward distinction and suggests that both sad and happy music can engage listeners internally, but in qualitatively different ways. This extends Taruffi et al.’s interpretation by suggesting that happy music does not necessarily reduce internal engagement, but may instead shift the quality of internal experience from reflective to imaginative and narrative-driven.

4.3 Limitations of The Test

It became clear that the button presses would have benefited from some form of feedback to confirm that responses were registered correctly. One participant also described the test as cognitively demanding, suggesting that it might have been easier to complete if only one task was required at a time, such as reporting either emotion or imagery rather than both simultaneously. However, as this concern was raised by only a single participant, it does not appear to represent a major issue.

It should also be noted that the test allowed participants to proceed without providing responses, as no mandatory response checks were implemented. This could present a potential problem if participants accidentally skipped items or forgot to register a response. Such checks should therefore be included in future iterations of the test. Fortunately, no participants in the present study appeared to have missed any required responses.

5 Conclusion

In relation to the proposed hypotheses, the results provide only partial support. At the group level, neither hypothesis was clearly confirmed, as no consistent imagery-first pattern emerged for sad music (H1), nor a consistent emotion-first pattern for happy music (H2). However, response timing and analyses based on imagery ability suggest that these hypothesized patterns may depend on individual differences in imagery ability rather than applying uniformly across listeners.

6 Future Work

Future studies should include a larger and more diverse sample to better assess whether the observed patterns related to imagery ability are robust and generalizable. Improvements to the experimental design, such as clearer response feedback and mandatory response checks, would help reduce uncertainty in participant input. Additionally, future work could further explore how

the musical features of these pieces might influence imagery content and emotional experience during music listening.

7 Acknowledgments

The Author of this Mini-project acknowledges the use of ChatGPT[11] for grammar, sentence structure and coding.

References

- [1] Liila Taruffi and Mats Küssner. “A Review of Music-Evoked Visual Mental Imagery: Conceptual Issues, Relation to Emotion, and Functional Outcome”. In: (Oct. 2018).
- [2] Liila Taruffi et al. “Effects of sad and happy music on mind-wandering and the default mode network”. In: *Scientific Reports* 7.14396 (2017). DOI: 10.1038/s41598-017-14849-0.
- [3] Sofia Dahl, Antonio Stella, and Thomas Bjørner. “Tell me what you see: An exploratory investigation of visual mental imagery evoked by music”. In: *Musicae Scientiae* 27 (Nov. 2022), p. 102986492211248. DOI: 10.1177/10298649221124862.
- [4] Patrik N. Juslin and Daniel Västfjäll. “Emotional responses to music: The need to consider underlying mechanisms”. In: *Behavioral and Brain Sciences* 31.5 (2008), pp. 559–575. DOI: 10.1017/S0140525X08005293.
- [5] R. A. Day and W. F. Thompson. “Measuring the onset of experiences of emotion and imagery in response to music”. In: *Psychomusicology: Music, Mind, and Brain* 29 (2019), pp. 75–89.
- [6] Thijs Vroegh. *Investigating the directional link between music-induced visual imagery and two different types of emotional responses*. June 2018. DOI: 10.13140/RG.2.2.31760.30727.
- [7] Meadows Documentation. *Vividness of Visual Imagery Questionnaire*. URL: <https://docs.meadows-research.com/presets/vviq/> (visited on 01/08/2026).
- [8] Ennio Morricone. *Hamlet*. URL: <https://www.youtube.com/watch?v=vzcsTlwWKCQ> (visited on 01/08/2026).
- [9] Michael Nyman. *À La Folie*. URL: https://www.youtube.com/watch?v=_bSVY-L2LXw (visited on 01/08/2026).
- [10] Posit. *RStudio Desktop*. URL: <https://posit.co/download/rstudio-desktop/> (visited on 01/08/2026).
- [11] OpenAI. *ChatGPT*. 2026. URL: <https://chat.openai.com/>.