**Dear Dr. Schwarzkopf,**

**Thank you very much for your prompt screening and providing your valuable feedback. We appreciate your detailed review and advice and have made revisions accordingly. Kindly find our responses for each comment below.**

1. Hypotheses: Please ensure that all hypotheses listed in the text match those in your Design Table. Also consider if the statistics used to test the hypotheses are appropriate. Your text describes the hypotheses as testing for correlations, but your statistical approach does not explicitly test for correlations. I appreciate that you could informally phrase the research questions this way but the hypotheses should match and explicitly describe the analysis used. In the same vein, please also spell out your inference plan in the main text to match what is explained in the Design Table.

**We revised the hypothesis and inference plan in the text to match the table.**

2. Statistical approach: You propose a series of three paired t-tests for each cultural group (plus possible equivalent tests). Consider using an omnibus test to test the significance of the main effect of tempo (irrespective of group) and any interaction (in case the effect differs by culture). I believe this would be more sensitive. Alternatively, please provide a clear justification for the approach adopted here. You also write that the use of Likert scales mean that the parametric statistical tests are inadequate. You may consider using non-parametric alternatives like a Kruskal-Wallis and Wilcoxon signed-rank tests. However, I also appreciate that parametric statistics are commonly used for such ordinal data and this use may not necessarily pose a major problem. Nevertheless, if possible I would seek out the advice of a statistical expert.

**Thanks for this excellent advice. However, a quick search of options did not allow us to decide on the most appropriate way of doing this that would also allow us to perform the appropriate matching power analysis, and Japan is beginning a week-long “Golden Week” holiday that means we will be unable to recruit and consult an appropriate expert within a timeline that would be compatible with the scheduled review process. Therefore we have added the following footnote indicating this to the manuscript and hope that reviewers might be able to help us weigh in on this decision:**

***NB: We are also open to revising the planned analysis/power analysis to include potential alternative statistical approaches should they be considered more appropriate. Possible alternatives identified include:***

***-an omnibus test to test the significance of the main effect of tempo (irrespective of group) and any interaction (in case the effect differs by culture);***

***-non-parametric tests like a Kruskal-Wallis and Wilcoxon signed-rank tests;***

***-ordinal regression; and/or***

***-two-way within-sample ANOVA***

3. Exploratory analyses: In a RR Stage 1 submission, all planned tests should be formulated as preregistered hypotheses with a detailed statistical analysis plan, supported by power analysis. The RR format does not preclude carrying out exploratory analyses at Stage 2, provided they will be labeled explicitly. However, this does not belong in the Stage 1 manuscript. Please remove the *Exploratory Analyses* section or add these tests as fully preregistered hypotheses to the manuscript and Design Table, with appropriate power analysis. This same point applies to any comparison between solo and group excerpts. It is fine to describe these differences in your pilot data, but if you plan to conduct this comparison you will need to add specific hypotheses for this supported by power analysis, or only add these as exploratory analyses in Stage 2 but not mention them in Stage 1.

**We have excluded this section from our stage 1 submission and will add it to the stage 2 accordingly.**

4. Power analysis: On this note, you state in the *Power Analysis* section on page 6 that the minimally interesting effect size is d=0.04 but later on and in your Design Table you power the experiment for d=0.4. I assume this first instance is a typo? The power analysis is based on a conventional alpha=0.05 (Bonferroni corrected for number of hypotheses) but 95% power. This is fine and the decision is yours, but please note that many journals accepting RRs for publication have specific power requirements, most commonly 90% power with alpha=0.02.

**Thanks for catching the typo - both should be d=0.4. Thank you for noting the journal requirements - we are happy with our initial plan for now.**

5. Missing methods details:   
  
a) Please add details about the visual stimuli used in the experiment, such as the dimensions of the patterns, how they are generated, etc. The methods should make independent replication possible. Since your data collection is online, I appreciate that some parameters (like visual angle or luminance) are outside your control - but at least relatively parameters can be reported in detail. My suggestion is also to add a figure showing the texture patterns.I may be missing something but I also do not follow how X2+1 results in 31 lines (assuming X is the density scale variable X=5 in this case?)

**We have included more information about the visual textures including moving the Figure showing these from the Appendix to the main text, and adding the formula for density increase to facilitate replication.**

b) Under *Exclusion Criteria* please explain what you mean by participants who "most associate with the culture of the country" and how this is determined.

**Thanks for pointing out this ambiguous phrasing: we have omitted this phrase in the revised version.**

c) Under *Materials* it is unclear to me why there are 24 excerpts. With 6 excerpts (one solo one group, from each of the 3 countries) and two tempos, aren't there only 12 excerpts?

**The additional 12 excerpts were for the exploratory analysis of pitch, but since these have now been removed from the manuscript we have also removed them in the revised version.**

d) Also under *Materials*, please rephrase the first sentence to clarify that this is effectively an online survey participants "will fill out".

**Thank you for pointing this out. We revised the opening sentence to read “...will be conducted as an online survey and participants…”**

6. Error in Analysis Plan 1.2: Please remove the "emotion categories" from the first sentence in this section as this refers to potential exploratory analyses but is not preregistered. Also, under *1.2.2 Temporal-density correlation* you state that arousal ratings are being collected instead of visual density.

**Thank you for catching this typo. We have corrected the phrase to read “emotional arousal” rather than “emotion categories”.**

7. Phrasing of hypotheses: In the text you refer to the hypotheses as testing the effect of tempo "changes". Technically the experiment compares stimuli with different tempos not specifically the change in tempo. This is a minor point and you may want to keep it as is - just food for thought.

**We have revised the wording to reflect your advice. It now reads “Increasing tempo consistently increases emotional arousal across cultures…**” **Thank you for your suggestion.**

**We look forward to hearing from you and hope the explanations above clarified the manuscript. As we are going to be on leave till the next week due to the Golden week holiday in Japan, we gathered it would be more efficient to make as many changes as possible to get the draft ready and send on time for our scheduled review process instead of waiting a week before making further changes. We hope that this revised version would meet the screening requirements to proceed to the next step.**

**Thank you very much for your time and consideration.**

**Sincerely,**

**Shafagh Hadavi and Dr. Patrick Savage (On behalf of all authors)**

**Cross-cultural relationships between music, emotion, and visual imagery: A comparative study of Iran, Canada, and Japan [Stage 1 Registered Report]**

**Author names and affiliations**: Shafagh Hadavi\***1**, Junji Kuroda**2**, Taiki Shimozono**2**, Patrick E. Savage\***1**

**1**Graduate School of Media and Policy / Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan

**2***YAMAHA Corporation, Hamamatsu, Japan*

\*Correspondence to: [shafagh@keio.jp](mailto:shafagh@keio.jp); [psavage@sfc.keio.ac.jp](mailto:psavage@sfc.keio.ac.jp)

**Abstract**

Many people experience emotions and visual imagery while listening to music. Previous research has identified cross-modal associations between musical and visual features as well as cross-cultural links between music and emotion and between music and visual imagery. However, few studies have simultaneously investigated cross-cultural links between music, visual imagery, and emotion in order to distinguish the role of cultural experiences in contrast to innate perceptual capabilities. We hypothesize that there are cross-culturally consistent correlations between tempo changes and 1) visual density associations, and 2) arousal ratings. In this study, we investigate the relationship between emotional arousal and visual density induced by 6 musical excerpts differing in tempo and texture (solo vs group) in participants in Japan, Iran, and Canada. By distinguishing between universality and culture-specificity of these associations, we aim to understand whether relationships between music, emotion, and visual imagery are cross-culturally universal or culturally specific.

1. **Introduction**

Emotions and visual mental imagery are widespread responses to music in most cultures, yet the cross-cultural similarities and differences between musical features, emotions, and visual imagery remain poorly understood. Cross-modal audio-visual associations have been extensively studied and revealed evidence for consistency between musical elements and visual features (Dolschied et al., 2022; Kussner 2014; Rusconi et al., 2006; Walker, 1987; Giannos et al., 2021; Pitteri et al., 2017; Athanasopoulos and Moran, 2013; Kussner and Leech-Wilkinson, 2014; Palmer et al., 2013; Whiteford et al., 2018). Emotions have also been documented to interact with musical elements such as pitch and tempo (Hevner, 1937; Motte-Haber, 1968; Ilie and Thompson, 2011; Jaquet et al., 2014; Ramos et al. 2011; Dalla Bella et al., 2001; Webster and Weir, 2005). On the other hand, cross-cultural research in musical emotions has discovered both consistency and diversity in emotion appraisal in music (Cowen et al., 2020; Balkwill, and Thompson, 1999; Thompson, 2010; Fritz et al., 2009; Balkwill et al., 2004). However, few studies (e.g., Palmer, 2013) have investigated the relationship between musical features, visual imagery, and emotions to understand universality and diversity in cross-modal associations.

Audiovisual associations are shown to be mediated by “psychological and socio-cultural” elements (Taruffi and Kussner, 2022), musical training (Kussner and Leech-Wilkinson, 2014), language (Dolschied et al., 2022). Additionally, research about music-based visual imagery has examined how musical features such as pitch, volume, and time are associated with visual imagery, and revealed correlations between musical and visual features such as pitch and spatial height (e.g. Dolschied et al., 2022; Athanasopoulos and Moran, 2013; Kussner 2014; Tan and Kelly, 2004; Rusconi et al., 2006; Eitan and Timmers, 2010), and horizontal time representation (Athanasopoulos and Moran, 2013; Kussner, 2014; Walker, 1987). There are other visual features such as visual density that are yet to be examined that will add to our understanding of audiovisual associations across individuals and cultures, as well as some of the underlying mechanisms behind these associations.

Furthermore, musical features such as tempo and pitch have been found to impact the emotional qualities of musical excerpts (Hevner, 1937; Ilie and Thompson, 2011; Jaquet et al., 2014; Ramos et al., 2011; Gabrielsson and Juslin, 1996; Juslin and Laukka, 2003; see also Scherer and Oshinsky, 1977). For example, these studies found evidence for correlations between faster tempo and happy feelings (e.g. Hevner, 1937, Hunter et al., 2010), and higher arousal rates for faster tempo (Ramos et al., 2011), as well as higher pitch correlations with higher mean of valence ratings (Jaquet et al., 2014). However, research concerning the relationship between musical features, visual features, and emotions is still quite limited as most studies tackle only two of these domains. There are several studies that have discussed how music to visual associations are mediated by emotions (Schloss et al., 2008; Tsang and Schloss, 2010; Palmer et al., 2013, 2016; Whiteford et al., 2018; Lindborg and Friberg, 2015). One of the few cross-cultural studies involving music, visual features (color in this case), and emotions was done by Palmer et al. (2013) on US and Mexican participants. This study revealed that music to color associations are mediated by emotions across participants and cultures by asking their participants to pair classical music excerpts and colors from the Berkeley color project (Palmer and Schloss, 2010) as well as rating their emotional responses to each musical excerpt and color choices. However, studying cross modality in cultures outside of North America using music other than Western classical music would be useful in testing the generality of music-emotion-visual imagery interactions.

It is widely agreed upon that music can induce strong emotions in the listeners. Moreover, prior research indicates that participants are sensitive to emotional cues from another musical culture (Balkwill et al., 2004; Balkwill and Thompson, 1999) and can decode them without being familiar with the culture (Fritz et al., 2009). However, more cross-cultural studies on musical emotions are required to illuminate the universality of emotions (Juslin et al., 2016) as well as the interaction of musical features with emotions. A large review of cross-cultural responses to music highlights both universality and culture-specificity of musical emotions (Singh and Mehr, 2023). The majority of studies utilize either dimensional or categorical models of emotions. Dimensional emotions often involve two or three dimensions of affective states (e.g. Faith & Thayer, 2001; Gabrielsson & Lindström, 2010; Bradley and Lang, 1994; Wedin, 1972; Nielzén and Cesarec, 1981) such as Russel’s arousal and valence dimensions (Russel, 1980), with arousal in particular suggested to have cross-culturally general relationships with musical features such as tempo (see review in Singh & Mehr, 2023). One way to empirically tackle these ideas is to conduct cross-cultural studies and measure both emotional models in order to gain insight into the mechanisms of emotion detection across cultures whether the innate auditory and physiological qualities or the role of culture-specific elements are involved.

Importantly, most studies in the music cognition and perception literature incorporate mostly Western music and Western participants, and there is a need to test these findings in other cultures as well (Jacoby et al., 2020). Our study would contribute to the previous findings on the relation between cross-modal associations and emotion through the usage of traditional music from Iran and Japan in addition to Canadian folk music. Our choice of countries is based on our access to the local communities and having native speakers as coauthors who can facilitate the process of data collection. We delve deeper into this relationship through a comparative experiment in Japan, Iran, and Canada to discover the similarities and differences across these three seemingly different populations; countries with distinct language, cultural practices, scripts, and musical cultures. Our study will also make contributions to systematic empirical research on the significance of musical features in cross-modal associations (Eitan, 2017) by investigating relationships between musical tempo, visual density, and emotional arousal.

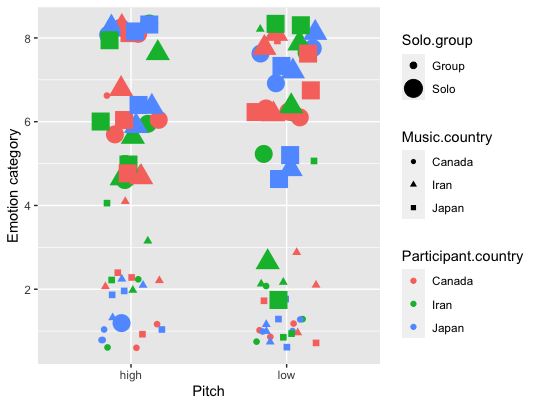
**1.1. Study Aims and Hypotheses**

Our study aims to understand cross-cultural relationships between musical features, visual imagery, and emotions in Japan, Iran, and Canada. Based on our pilot data, we plan to compare these three populations through collection and analysis of 1) music to visual density associations, and 2) emotion appraisal of excerpts manipulated in pitch and tempo. Previous studies are also mostly conducted on Western participants and in most instances Western music. Our aim is to study some of the underrepresented regions and musical cultures in order to make a meaningful contribution to cross-cultural studies in music cognition and perception. By comparing solo and group excerpts, we plan to detect whether visual density associations are meaningfully and metaphorically associated with any of our musical variables which are tempo, pitch, and musical texture(solo vs group). Many cross-modal studies have studied the relation between music and visual stimuli yet visual density has not been studied to our knowledge.

By making a distinction between cross-cultural consistency and diversity in these correspondences and emotion appraisals, we aim to understand whether 1) we can find any innate/physiologically derived connection that could explain our cross-modal associations, and 2) tempo mediates emotion appraisals and visual imagery.

Our hypotheses (Table 1) are listed as:

1. Increasing tempo consistently increases emotional arousal across cultures.
2. Increasing tempo consistently increases density of visual imagery across cultures.



**Table 1.**

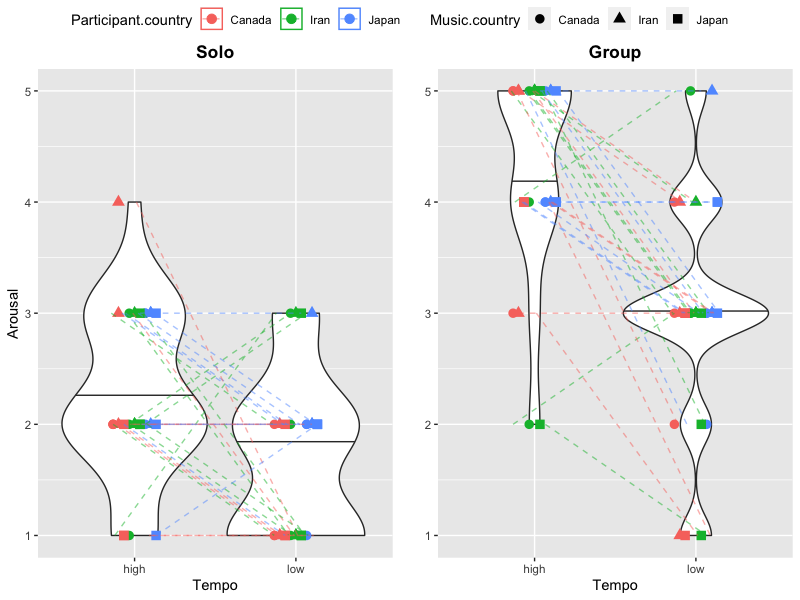
| **Question** | **Hypothesis** | **Sampling Plan** | **Analysis Plan** | **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis** | **Conclusion-Interpretation given**  **different outcomes** | **Theory that could be shown**  **wrong by the outcomes** |
| --- | --- | --- | --- | --- | --- | --- |
| **How are music, visual imagery, and emotion related across cultures?** | 1) Increasing tempo consistently increases emotional arousal across cultures. | Participant recruitment from Iran, Japan, and Canada (N=14 per cultural group [42 total]) will be done through snowball sampling. ( See appendix 1) | We will test our hypotheses by performing paired *t* tests[[1]](#footnote-0) on our dependent variables (here, arousal) for high vs. low tempo versions of the 6 excerpts. | We will test this hypothesis three times for each of our two hypotheses: once for each cultural group, with sample size decided for each group by power analysis assuming SESOI of d=.4, power of 95%, and study-wide alpha=.05 controlling for our two hypotheses (i.e., p<.025). | If our predicted effect is significant in each of the three cultural groups, we will conclude that tempo-arousal relationships are cross-culturally general. If any effect is not significant, we will use equivalence testing to test whether the effect is statistically equivalent to 0 (|d| < .4). If we find a significant effect in one or more cultures but a statistically equivalent result in one or more cultures, we will conclude the relationship is cross-culturally variable. If all three cultures give statistically equivalent results, we will conclude that tempo and arousal are consistently not related across cultures. If any tests are neither statistically significant nor statistically equivalent, we will interpret the results to be inconclusive. | If we find consistent tempo-arousal correlations, it would contradict cultural relativist theories. Conversely, if tempo and arousal are not consistently correlated, it would contradict universalist theories about musical emotion-arousal correlations.. |
| 2) Increasing tempo consistently increases density of visual imagery across cultures. | Same as H1, but for visual density instead of emotional arousal | | | | |
|

**1.2. Analysis Plan and pilot data results**

We will perform paired *t* tests to test our hypothesis on our dependent variables: arousal, density, emotion arousalacross tempo changes in all cultures. Our pilot data(N=9) supports our hypotheses; however, we would need to measure a larger sample for sufficient power. Our hypotheses include:

1.2.1 Effects of tempo on arousal. We will test this by comparing the overall means of the arousal ratings for all excerpts in two high and low tempo conditions across all participant cultures. (Fig. 1A)

1.2.2 Effects of tempo on density. We will test this by comparing the overall means of the arousal ratings for all excerpts in two high and low tempo conditions across all musical cultures and participant cultures. (Fig.

1B) 

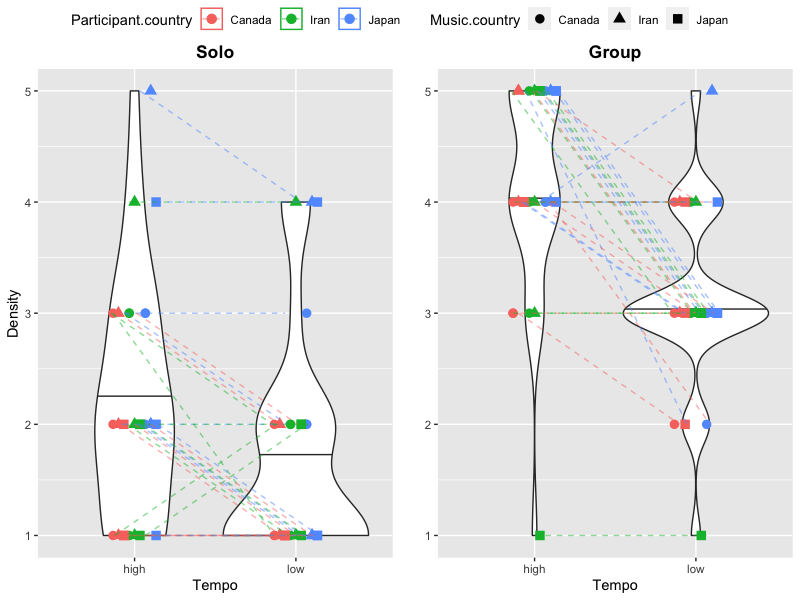


Fig 1A Fig 1B

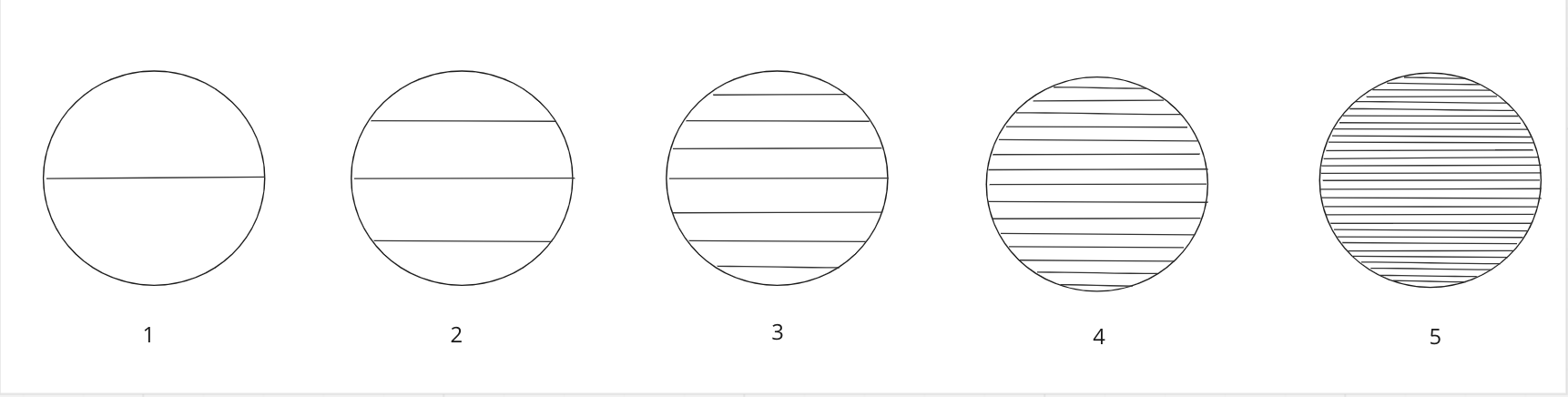
**Figure 1 (A and B): Pilot data for arousal and visual density ratings for 6 traditional instrumental excerpts(solo vs group) by nine participants.** Nine participants from Iran, Canada, and Japan (three from each culture) listened to 6 excerpts, one solo and one group excerpt from Iran, Canada, and Japan. The excerpts were all manipulated in tempo to reflect a 20% higher and lower tempo than the original. **Figure 1A:** Participants rated the arousal level for each solo and group excerpt on a scale from 1-5. Participants across all cultures associated excerpts manipulated into higher tempo with higher arousal levels and excerpts manipulated into lower tempo with lower arousal levels. The solo pieces which were slower in nature were also mostly associated with lower arousal and group excerpts were mostly rated higher for arousal since their tempo was higher. **Figure 1B:** Participants matched each excerpt with a visual image on a scale from 1-5 that they found most consistent with. The visual images consisted of visual textures differing in visual density represented by one horizontal line (scale 1) to 31 horizontal lines (scale 5). Participants across all cultures associated excerpts manipulated into higher tempo with higher density levels and excerpts manipulated into lower tempo with lower density levels. The solo pieces which were slower in nature were also mostly associated with lower visual density and group excerpts were associated with higher density levels.

**2. Method**

We plan to recruit participants from Japan, Iran, and Canada (n=14 per group; see Power Analysis below), who are raised in these countries and are immersed in the culture with any level of musical training. We selected 6 traditional instrumental excerpts between 10-20s, two from each country. We divided them into two categories of solo and group based on the instrumentation. The instruments used in the excerpts include different strings and woodwinds. The solo pieces have slower tempos (meter-less Japanese and Iranian solo, and vague meter for Canadian solo) and the group excerpts are noticeably faster (Japanese group at 91, Iranian group at 132, and Canadian group at 98 BPM). The tempo of all 6 excerpts is separately manipulated in a way that it is lowered and raised by 20% for each excerpt. We selected these specific tempo after finding in our preliminary pilot analyses that these provided the optimal balance between maximizing acoustic differences while minimizing audible recording artifacts created by the manipulation process.

**2.1.****Materials**

The experiment (Appendix 2) will be conducted as an online survey and participants will fill it out on their own device and space. Participants will choose their desired language, and start with the questions. They will be randomly presented by the 12 excerpts, and asked to rate their arousal from a scale of 1-5. They also select a visual texture (adapted from Langlois et al., 2014) which they think most represents the excerpt on a scale of 1-5. Participants will be presented with Fig 2 which is a series of 5 visual textures ascending in visual density represented by five circles with increasing numbers of parallel horizontal lines. Each circle has a diameter of 2cm. Texture one consists of one horizontal line in a circle. (Fig 2) The subsequent textures incorporate a steady increase in the number of horizontal lines leading to a steady increase in visual density levels. Y=2X+1 represents how the density increases, where Y is the number of lines in the next texture and X is the number of lines in the previous texture. These visual textures are generated through digital drawing. The excerpt will play on a loop while the participants are filling out the questions, and once ready to move to the next excerpt, a click on the next button can take the participants to the next excerpt.

**Fig 2.** Adapted from the stimuli by Langlois et al. (2014).

**2.1.1. Exclusion criteria for musical sample and participant recruitment**

We plan to recruit participants who are 18 years old and above, are brought up in Iran, Canada, and Japan respectively and their native language is Farsi, English, and Japanese.. Participants who do not pass the color and pattern recognition for reasons such as color blindness (cf. Appendix 2) before starting the experiment or those who do not complete each section will be excluded. There will be two attention check questions throughout to ensure accuracy of the responses.

**2.1.2. Independent variable**

Our independent variable is tempo, which is manipulated by raising the tempo to be either 20% higher or 20% lower than the original tempo.

**2.1.3. Dependent variables**

Our visual stimuli consists of a texture that varies in density on a scale of 1 to 5, 1 being the least dense with one line and 5 being the most dense. The number of lines increases density=X2+1 . Scale 1 has one line and scale 5 has 31 lines. Arousal ratings will be done using scales 1-5, 1 being very subdued and calm, and 5 being very excited or aroused.

**2.2. Power analysis**

Our power analysis estimated that we would require a sample of at least 84 paired responses for our study to have 95% power to test our two directional (one-tailed) hypotheses assuming a smallest effect size of interest (SESOI) of Cohen’s d=.04 (Brysbaert, 2019) while maintaining an overall false-positive rate of 5% (p < .025 after applying Bonferroni correction; see Data/Code availability statement for link to exact code). While effect sizes estimated from pilot data are by definition unreliable (Brysbaert, 2019), we note that all of our pilot data groups demonstrated effect sizes greater than d=0.4 (minimum: d=0.58 [effect of tempo on arousal for Iranian participants]; maximum: d=1.08 [effect of tempo on visual density for Canadian participants]). Note that for our specific paradigm, each participant produces 6 paired responses (one pair for each of the 6 musical stimuli), and each cultural group will be tested separately, so we will need at least 84 / 6 = 14 participants x 3 = 42 participants total across all three cultural groups.

Technically, this paradigm violates some of the assumptions of the statistical analysis. In particular: 1) our dependent variables are from 5-point Likert scales, rather than normally distributed continuous data; 2) the 6 paired responses from each participant are not independent of one another; and 3) by testing each hypothesis three times (one for each cultural group) and only confirming our predictions if all three tests are statistically significant, this has the effect of decreasing our true false positive rate while increasing our true false negative rate. We are not aware of a more appropriate analysis method that could incorporate these violations, but note that in practice points # 2 and #3 should partially balance one another out to avoid dramatically affecting our overall power analysis estimates. (See footnote 1 above for potential alternative analysis options.)

**Data/code availability**

Our data and code are available at <https://github.com/comp-music-lab/VisualEars>

**Acknowledgments**

We would like to thank Hideki Sakanashi, Mai Fujiwara, and Yasuo Shiozawa for the discussion; Yuto Ozaki for help making Figure 1; and all participants in the pilot experiment and all Keio University CompMusic Lab members for their feedback and help.

**Author contributions**

Project conception: Hadavi

Project design: Hadavi, Savage, Kuroda, Shimozono

Original draft: Hadavi

Statistical analysis: Savage, Hadavi

Editing and revisions: Hadavi, Savage, Kuroda, Shimozono

**Inclusion criteria**

We commit to the best practices in cross-cultural studies (Tan & Ostashewski, 2022; Savage, Jacoby, Margulis, et al., 2021), such as involving collaborators from diverse backgrounds from the initial planning phases of a study and implementing coauthorship mechanisms.

**Funding/Conflicts of interest:**

This project is supported by funding from the Yamaha Corporation, including several authors who are Yamaha employees. Our Registered Report approach means that we will commit to experimental design and analysis methods ahead of time and commit to publishing our results regardless of whether or not they support our hypotheses.

**Ethical approval**

We are planning to apply to the Keio University Shonan Fujisawa Campus Research Ethics Committee for approval once it has been otherwise accepted in principle (to maximize efficiency by avoiding having to submit revised protocols for approval).

**References:**

**Athanasopoulos, G., & Moran, N. (2013). Cross-Cultural Representations of Musical Shape. *Empirical Musicology Review*, 185–199.** [**https://doi.org/10.18061/emr.v8i3-4.3940**](https://doi.org/10.18061/emr.v8i3-4.3940)

**Athanasopoulos, G., Tan, S.-L., & Moran, N. (2016). Influence of literacy on representation of time in musical stimuli: An exploratory cross-cultural study in the UK, Japan, and Papua New Guinea. *Psychology of Music*, *44*(5), 1126–1144.** [**https://doi.org/10.1177/0305735615613427**](https://doi.org/10.1177/0305735615613427)

**Balkwill, L.-L., & Thompson, W. F. (1999). A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception*, *17*(1), 43–64.** [**https://doi.org/10.2307/40285811**](https://doi.org/10.2307/40285811)

**Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners 1: Recognition of emotion in music. *Japanese Psychological Research*, *46*(4), 337–349.** [**https://doi.org/10.1111/j.1468-5584.2004.00265.x**](https://doi.org/10.1111/j.1468-5584.2004.00265.x)

**Barrett, L. F. (2006). Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review*, *10*(1), 20–46.** [**https://doi.org/10.1207/s15327957pspr1001\_2**](https://doi.org/10.1207/s15327957pspr1001_2)

**Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *0*(0).** [**https://doi.org/10.1016/j.tics.2022.09.015**](https://doi.org/10.1016/j.tics.2022.09.015)

**Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.** [**https://doi.org/10.1016/0005-7916(94)90063-9**](https://psycnet.apa.org/doi/10.1016/0005-7916(94)90063-9)

**Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. Journal of Cognition, 2(1), 1–38.** [**https://doi.org/10.5334/joc.72**](https://doi.org/10.5334/joc.72)

**Cespedes-Guevara, J., & Dibben, N. (2022). The Role of Embodied Simulation and Visual Imagery in Emotional Contagion with Music. Music & Science, 5.** [**https://doi.org/10.1177/20592043221093836**](https://doi.org/10.1177/20592043221093836)

**Cespedes-Guevara, J., & Eerola, T. (2018). Music Communicates Affects, Not Basic Emotions – A Constructionist Account of Attribution of Emotional Meanings to Music. *Frontiers in Psychology*, *9*.** [**https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00215**](https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00215)

**Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. *Proceedings of the National Academy of Sciences*, *117*(4), 1924–1934.** [**https://doi.org/10.1073/pnas.1910704117**](https://doi.org/10.1073/pnas.1910704117)

**Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.** [**https://doi.org/10.1016/S0010-0277(00)00136-0**](https://doi.org/10.1016/S0010-0277(00)00136-0)

**Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., & Majid, A. (2022). Children’s associations between space and pitch are differentially shaped by language. *Developmental Science*, *n/a*(n/a), e13341.** [**https://doi.org/10.1111/desc.13341**](https://doi.org/10.1111/desc.13341)

**Eitan, Z. (2017). Musical Connections, Crossmodal correspondences. In Richard Ashley and Renee Timmers (eds), *The Routledge companion to music cognition* (pp.213-224). Taylor & Francis Group.ProQuest Ebook Central. https://ebookcentral.proquest.com/lib/keio/detail.action?docID=4890817.**

**Eitan, Z., & Timmers, R. (2010). Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*(3), 405–422.** [**https://doi.org/10.1016/j.cognition.2009.10.013**](https://doi.org/10.1016/j.cognition.2009.10.013)

**Faith, M., & Thayer, J. F. (2001). A dynamical systems interpretation of a dimensional model of emotion. Scandinavian Journal of Psychology, 42(2), 121–133.** [**https://doi.org/10.1111/1467-9450.00221**](https://psycnet.apa.org/doi/10.1111/1467-9450.00221)

**Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology : CB*, *19*(7), 573–576. https://doi.org/10.1016/j.cub.2009.02.058**

**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**](https://doi.org/10.1177/0305735696241007)

**Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford University Press.**

**Giannos, K., Athanasopoulos, G., & Cambouropoulos, E. (2021). Cross-Modal Associations Between Harmonic Dissonance and Visual Roughness. *Music & Science*, *4*, 20592043211055484.** [**https://doi.org/10.1177/20592043211055484**](https://doi.org/10.1177/20592043211055484)

**Griffiths, D., Cunningham, S., Weinel, J., & Picking, R. (2021). A multi-genre model for music emotion recognition using linear regressors. *Journal of New Music Research*, *50*(4), 355–372.** [**https://doi.org/10.1080/09298215.2021.1977336**](https://doi.org/10.1080/09298215.2021.1977336)

**Hevner, K. (1937). The Affective Value of Pitch and Tempo in Music. *The American Journal of Psychology*, *49*(4), 621–630.** [**https://doi.org/10.2307/1416385**](https://doi.org/10.2307/1416385)

**Hunter, P., Schellenberg, E., & Schimmack, U. (2010). Feelings and Perceptions of Happiness and Sadness Induced by Music: Similarities, Differences, and Mixed Emotions. *Psychology of Aesthetics Creativity and the Arts*, *4*, 47–56.** [**https://doi.org/10.1037/a0016873**](https://doi.org/10.1037/a0016873)

**Ilie, G., & Thompson, W. (2011). Experiential and Cognitive Changes Following Seven Minutes Exposure to Music and Speech. *Music Perception*, *28*, 247–264.** [**https://doi.org/10.1525/mp.2011.28.3.247**](https://doi.org/10.1525/mp.2011.28.3.247)

**Jacoby, N., Margulis, E. H., Clayton, M., Hannon, E., Honing, H., Iversen, J., Klein, T. R., Mehr, S. A., Pearson, L., Peretz, I., Perlman, M., Polak, R., Ravignani, A., Savage, P. E., Steingo, G., Stevens, C. J., Trainor, L., Trehub, S., Veal, M., & Wald-Fuhrmann, M. (2020). Cross-Cultural Work in Music Cognition. *Music Perception*, *37*(3), 185–195.** [**https://doi.org/10.1525/mp.2020.37.3.185**](https://doi.org/10.1525/mp.2020.37.3.185)

**Jaquet, L., Danuser, B., & Gomez, P. (2014). Music and felt emotions: How systematic pitch level variations affect the experience of pleasantness and arousal. *Psychology of Music*, *42*, 51–70.** [**https://doi.org/10.1177/0305735612456583**](https://doi.org/10.1177/0305735612456583)

**Juslin, P. N. (2003). Five facets of musical expression: a Psychologist’s perspective on music performance. *Psychol. Music* 31, 273–302. doi: 10.1177/ 03057356030313003**

**Juslin, P. N., Barradas, G. T., Ovsiannikow, M., Limmo, J., & Thompson, W. F. (2016). Prevalence of emotions, mechanisms, and motives in music listening: A comparison of individualist and collectivist cultures. *Psychomusicology: Music, Mind, and Brain*, *26*(4), 293–326.** [**https://doi.org/10.1037/pmu0000161**](https://doi.org/10.1037/pmu0000161)

**Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814.** [**https://doi.org/10.1037/0033-2909.129.5.770**](https://doi.org/10.1037/0033-2909.129.5.770)

**Küssner, M. (2014.). *Shape, Drawing and Gesture: Crossmodal Mappings of Sound and Music*. [Doctoral Dissertation, King’s College London]. King’s Research Portal. https://kclpure.kcl.ac.uk/portal/files/30755235/2014\_Kussner\_MatsB\_1051386\_ethesis.pdf**

**Küssner, M. B., & Leech-Wilkinson, D. (2014). Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing paradigm. *Psychology of Music*, *42*(3), 448–469.** [**https://doi.org/10.1177/0305735613482022**](https://doi.org/10.1177/0305735613482022)

**Küssner MB, Tidhar D, Prior HM, Leech-Wilkinson D. Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. Front Psychol. 2014 Jul 28;5:789. doi: 10.3389/fpsyg.2014.00789. PMID: 25120506; PMCID: PMC4112934.**

**Lakoff, G., & Johnson, M. (1980). *Metaphors We Live by*. Chicago: University of Chicago Press.**

**Langlois, T., Peterson, J., & Palmer, S. (2014). Visual Texture, Music, and Emotion. *Journal of Vision*, *14*(10), 437–437.** [**https://doi.org/10.1167/14.10.437**](https://doi.org/10.1167/14.10.437)

**Lindborg, P., & Friberg, A. K. (2015). Colour Association with Music Is Mediated by Emotion: Evidence from an Experiment Using a CIE Lab Interface and Interviews. *PLoS ONE*, *10*(12), e0144013.** [**https://doi.org/10.1371/journal.pone.0144013**](https://doi.org/10.1371/journal.pone.0144013)

**Motte-Haber H., de la. (1968). *Ein Beitrag zur Klassifikation musikalischer Rhythmen: Experimental-psychologische Untersuchungen*. Cologne, Germany: Arno Volk.**

**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**](https://doi.org/10.1371/journal.pone.0089642)

**Nielzén, S., & Cesarec, Z. (1981). On the perception of emotional meaning in music. *Psychology of Music, 9*(2), 17–31.** [**https://doi.org/10.1177/030573568192002**](https://psycnet.apa.org/doi/10.1177/030573568192002)

**Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877–8882.** [**https://doi.org/10.1073/pnas.0906172107**](https://doi.org/10.1073/pnas.0906172107)

**Palmer, S. E., Langlois, T. A., & Schloss, K. B. (2016). Music-to-Color Associations of Single-Line Piano Melodies in Non-synesthetes. *Multisensory Research*, *29*(1–3), 157–193.** [**https://doi.org/10.1163/22134808-00002486**](https://doi.org/10.1163/22134808-00002486)

**Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, *110*(22), 8836–8841.** [**https://doi.org/10.1073/pnas.1212562110**](https://doi.org/10.1073/pnas.1212562110)

**Pitteri, M., Marchetti, M., Priftis, K., & Grassi, M. (2017). Naturally together: Pitch-height and brightness as coupled factors for eliciting the SMARC effect in non-musicians. *Psychological Research, 81*(1), 243–254.** [**https://doi.org/10.1007/s00426-015-0713-6**](https://psycnet.apa.org/doi/10.1007/s00426-015-0713-6)

**Ramos, D., Bueno, J. L. O., & Bigand, E. (2011). Manipulating Greek musical modes and tempo affects perceived musical emotion in musicians and nonmusicians. *Brazilian Journal of Medical and Biological Research*, *44*, 165–172.** [**https://doi.org/10.1590/S0100-879X2010007500148**](https://doi.org/10.1590/S0100-879X2010007500148)

**Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, 99(2), 113–129.** [**https://doi.org/10.1016/j.cognition.2005.01.004**](https://psycnet.apa.org/doi/10.1016/j.cognition.2005.01.004)

**Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.** [**https://doi.org/10.1037/h0077714**](https://psycnet.apa.org/doi/10.1037/h0077714)

**Savage, P.E., Jacoby, N., Margulis, E.H., Daikoku, H., Anglada-Tort, M., Castelo-Branco, S.E.-S., Nweke, F.E., Fujii, S., Hegde, S., Chuan-Peng, H., Opondo, P., et al. (2021). Building sustainable global collaborative networks: Recommendations from music studies and the social sciences. In *The science-music borderlands: Reckoning with the past, imagining the future*, E. H. Margulis, D. Loughridge, and P. Loui, eds. (MIT Press). Preprint:** [**http://doi.org/10.31234/osf.io/cb4ys**](http://doi.org/10.31234/osf.io/cb4ys)

**Scherer, K. R., Oshinsky, J.S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motiv Emot* 1, 331–346.** [**https://doi.org/10.1007/BF00992539**](https://doi.org/10.1007/BF00992539)

**Schloss, K. B., Lawler, P., & Palmer, S. E. (2008). The color of music. Presented at the 8th Annual Meeting of the Vision Science Society, Naples, FL, May 2008.**

**Singh, M., & Mehr, S. (In press). Universality, domain-specificity, and development of psychological responses to music. *Nature Reviews Psychology*.**

**Tan, S.-L., & Kelly, M. E. (2004). Graphic Representations of Short Musical Compositions. *Psychology of Music*, *32*(2), 191–212.** [**https://doi.org/10.1177/0305735604041494**](https://doi.org/10.1177/0305735604041494)

**Tan, S. B., & Ostashewski, M. (Eds.). (2022). *DIALOGUES: Towards decolonizing music and dance studies*. International Council for Traditional Music. https://ictmdialogues.org/**

**Taruffi, L., & Küssner, M. B. (2019). A review of music-evoked visual mental imagery: Conceptual issues, relation to emotion, and functional outcome. *Psychomusicology: Music, Mind, and Brain*, *29*(2–3), 62–74.** [**https://doi.org/10.1037/pmu0000226**](https://doi.org/10.1037/pmu0000226)

**Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition & Emotion*, *22*(8), 1457–1470.** [**https://doi.org/10.1080/02699930701813974**](https://doi.org/10.1080/02699930701813974)

**Thompson, W. (2010). Cross-cultural similarities and differences (Music and Emotion) (pp. 755–788).** [**https://doi.org/10.1093/acprof:oso/9780199230143.001.0001**](https://doi.org/10.1093/acprof:oso/9780199230143.001.0001)

**Tsang, T., & Schloss, K. B. (2010). *Associations between Color and Music are Mediated by Emotion and Influenced by Tempo: (525772013-006)* [Data set]. American Psychological Association.** [**https://doi.org/10.1037/e525772013-006**](https://doi.org/10.1037/e525772013-006)

**Walker, R. (1987). Some differences between pitch perception and basic auditory discrimination in children of different cultural and musical backgrounds. *Bulletin of the Council for Research in Music Education, 91,* 166–170.**

**Webster, G. D., & Weir, C. G. (2005). Emotional Responses to Music: Interactive Effects of Mode, Texture, and Tempo. *Motivation and Emotion*, *29*(1), 19–39.** [**https://doi.org/10.1007/s11031-005-4414-0**](https://doi.org/10.1007/s11031-005-4414-0)

**Wedin, L. (1972). A multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology, 13*(4), 241–257.** [**https://doi.org/10.1111/j.1467-9450.1972.tb00072.x**](https://psycnet.apa.org/doi/10.1111/j.1467-9450.1972.tb00072.x)

**Whiteford, K. L., Schloss, K. B., Helwig, N. E., & Palmer, S. E. (2018). Color, Music, and Emotion: Bach to the Blues. *I-Perception*, *9*(6), 2041669518808535.** [**https://doi.org/10.1177/2041669518808535**](https://doi.org/10.1177/2041669518808535)

*Appendix 1*

**[Accepted version of the protocol will be translated into Farsi and Japanese.] This recruitment call will be distributed on Twitter and email.**

**Call for participants**

We are recruiting participants for a research study about music, visuals, and emotions. If you fit the criteria below, we would like to invite you to participate in our experiment. Please reach out at [visualears.project@gmail.com](mailto:visualears.project@gmail.com) if you are interested.

1. Your native language is English, you grew up in Canada, and you most associate with the Canadian culture.
2. You are 18 and above.
3. You have access to a computer/laptop/tablet/iPad and headphones.

The experiment will take about an hour and you will be compensated with $10.

Thank you very much.

*Appendix 2*

**[Accepted version of the protocol will be translated into Farsi and Japanese.]**

**Please wear headphones to do this experiment. It would be best to choose a place where you can complete the experiment without being distracted. Thank you for your participation!**

Please select the matching pattern and color. **( In this section, we present two items from our color and visual stimuli and ask the participants to match it with appearing items on their screen to ensure inclusion of the participant.)**

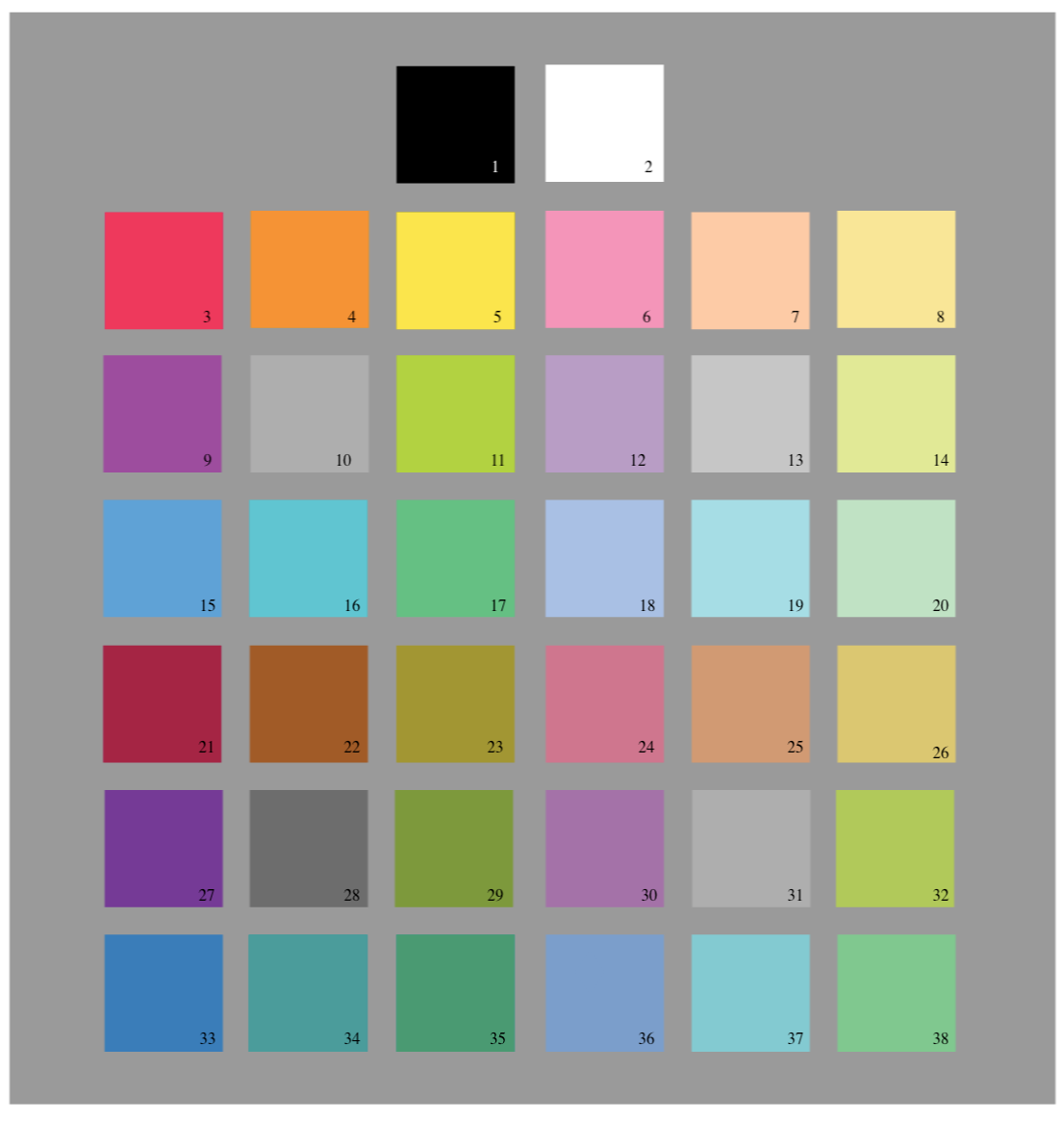
1. Please listen to 24 short musical excerpts linked below and enter your response to each excerpt by selecting the appropriate number.

<https://www.dropbox.com/scl/fo/sr4j2ijkilwod9k4grejh/h?dl=0&rlkey=y4mdwxhnmteyl8ijg279rctlg>

1. Using the colors in the image below, indicate one color that most represents each musical excerpt and come to your mind while listening to these excerpts.
2. Using the textures below,indicate one texture that most represents each excerpt.
3. Rate the arousal using scales 1-5: 1 being very subdued and calm, and 5 being very excited.
4. Rate the valence using scales 1-5, 1 being very negative and 5 being very positive.
5. Which of the following emotions best describes how the music makes you feel emotionally? 1-Happy 2-Excited 3-Angry 4-Afraid 5-Miserable 6-Sad 7-Tired 8-Relaxed
6. Rate your preference using scales 1-5, 1 for strongly dislike, 2, somehow dislike, 3, neither like nor dislike, 4, somehow like, and 5 strongly like.
7. For the **first 6** excerpts, enter Y if you have heard the piece before, and N if you have not heard it before.

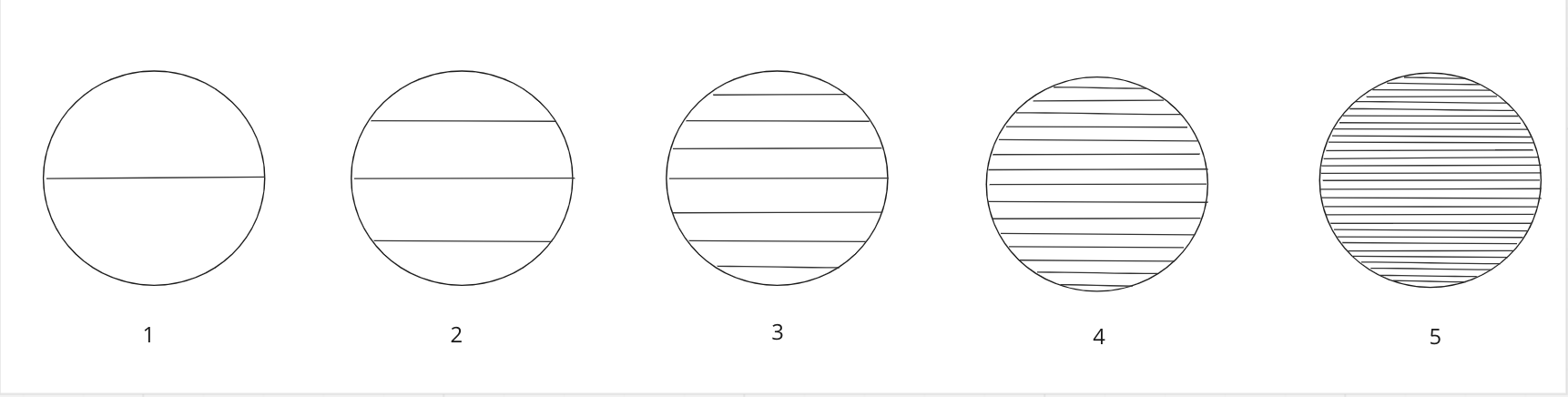
|  | Most consistent color | Most consistent texture | Arousal  scales 1-5: 1 being very subdued and calm, and 5 being very excited and aroused | Valence  scales 1-5: 1 being very negative and 5 being very positive | Emotion categories  1-Happy 2-Excited 3-Angry 4-Afraid 5-Miserable 6-Sad 7-Tired 8-Relaxed | Preference  scales 1-5 : 1 for strongly dislike, 2, somehow dislike, 3, neither like nor dislike, 4, somehow like, and 5 strongly like | Familiarity  Yes or No |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CA |  |  |  |  |  |  |  |
| PT |  |  |  |  |  |  |  |
| AT |  |  |  |  |  |  |  |
| XE |  |  |  |  |  |  |  |
| DG |  |  |  |  |  |  |  |
| EE |  |  |  |  |  |  |  |
| XP |  |  |  |  |  |  |  |
| DN |  |  |  |  |  |  |  |
| PL |  |  |  |  |  |  |  |
| EB |  |  |  |  |  |  |  |
| CP |  |  |  |  |  |  |  |
| DB |  |  |  |  |  |  |  |
| AZ |  |  |  |  |  |  |  |
| CL |  |  |  |  |  |  |  |
| ET |  |  |  |  |  |  |  |
| AC |  |  |  |  |  |  |  |
| EG |  |  |  |  |  |  |  |
| XA |  |  |  |  |  |  |  |
| AP |  |  |  |  |  |  |  |
| PE |  |  |  |  |  |  |  |
| 4XZ |  |  |  |  |  |  |  |
| DE |  |  |  |  |  |  |  |
| CN |  |  |  |  |  |  |  |
| PG |  |  |  |  |  |  |  |

**Colors**



Palmer et al. 2013, *PNAS*.(Berkeley Color Project)

**Textures**



Adapted from the stimuli by Langlois et al. (2014).

**GOLD MSI:** Please go to the appropriate version of Gold MSI below, and record your responses by selecting the most appropriate item. (Gold MSI will be translated into Farsi as well after In Principle Acceptance)

[GOLD MSI English version](https://media.gold-msi.org/test_materials/GMS/docs/gms_en.pdf)

[GOLD MSI Japanese version](https://media.gold-msi.org/test_materials/GMS/docs/gms_jp.pdf)

1. NB: We are also open to revising the planned analysis/power analysis to include potential alternative statistical approaches should they be considered more appropriate. Possible alternatives identified include:

   -an omnibus test to test the significance of the main effect of tempo (irrespective of group) and any interaction (in case the effect differs by culture);

   -non-parametric tests like a Kruskal-Wallis and Wilcoxon signed-rank tests;

   -ordinal regression; and/or

   -two-way within-sample ANOVA [↑](#footnote-ref-0)