**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**, Juan David Leongómez**3**, Patrick E. Savage\***1,4**

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

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

**3**Universidad El Bosque, Bogotá, Colombia

**4**School of Psychology, University of Auckland, New Zealand

\*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 more widespread perceptual capabilities. 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 72 participants from Japan, Iran, and Canada (24 each). We hypothesize that there are cross-culturally consistent relationships between tempo changes and 1) visual density associations, and 2) arousal ratings. The aim of this study is to understand whether relationships between music, emotion, and visual imagery are cross-culturally consistent.

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 associations have been extensively studied, revealing evidence for interactions between musical elements and visual features (Athanasopoulos and Moran, 2013; Chiba et al., 2023; Dolschied et al., 2022; Giannos et al., 2021; Kussner 2014; Kussner and Leech-Wilkinson, 2014; Palmer et al., 2013; Pitteri et al., 2017; Rusconi et al., 2006; Vines et al., 2006; Walker, 1987; Whiteford et al., 2018). Emotions have also been documented to interact with musical elements such as pitch and tempo (Dalla Bella et al., 2001; Hevner, 1937; Ilie and Thompson, 2011; Jaquet et al., 2014; Motte-Haber, 1968; Ramos et al. 2011; 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 (Balkwill and Thompson, 1999; Balkwill et al., 2004; Cowen et al., 2020; Fritz et al., 2009; Thompson, 2010). 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, even though visual imagery is thought to be one of the underlying mechanisms for how musical emotion is induced (Juslin and Västfjäll, 2008).

We use metaphors from other domains to understand music as an abstract experience. Many music-related metaphors are related to other sensory experiences such as visuals. One of the ways of studying music-related visual imagery is through music to visual associations (Athanasopoulos, 2022) where visual metaphors such as height, size, roughness, horizontal placement, and color are presented to match with sonic stimuli. These cross-modal associations are shown to be mediated by factors such as musical training (Kussner and Leech-Wilkinson, 2014) and language (Dolschied et al., 2022). Additionally, research about music-based visual imagery has examined how musical features such as pitch, volume, and the representation of time in music are associated with visual imagery, and revealed correlations between musical and visual features such as pitch and spatial height (e.g. Athanasopoulos and Moran, 2013; Dolschied et al., 2022; Eitan and Timmers, 2010; Kussner 2014; Rusconi et al., 2006; Tan and Kelly, 2004), and horizontal time representation (Athanasopoulos and Moran, 2013; Kussner, 2014; Walker, 1987).

On the other hand, there are other visual features whose association to music is understudied. For example, there have only been a few studies that examined the cross-modal relationship between musical and visual textures such as the study about the association between musical dissonance and visual roughness conducted by Gianno et al. (2021). Musical texture can refer to melodic, rhythmic, and harmonic layering of musical material (Sarrazin, 2016) and is often conceptualized with notions that describe visual texture such as dense/sparse, thick/thin, rough/smooth, etc. Additionally, visual textures are either plane or volumetric and a plane simple visual texture is characterized by three dimensions of directionality, size, and density (Caivano, 1990). Density, as Caivano (1990) states, “depends on the relation between the texturing elements and the background”. By examining the association between rhythmic density (i.e., tempo) and visual density, we aim to investigate this understudied aspect of cross-modal mappings that could add to our understanding of music-induced visual imagery and music cognition.

Furthermore, musical features such as tempo and pitch have been found to impact the emotional qualities of musical excerpts (Gabrielsson and Juslin, 1996; Hevner, 1937; Ilie and Thompson, 2011; Jaquet et al., 2014; Juslin and Laukka, 2003; Ramos et al., 2011; 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 (Lindborg and Friberg, 2015; Palmer et al., 2013, 2016; Schloss et al., 2008; Tsang and Schloss, 2010; Whiteford et al., 2018; ). 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 and Thompson, 1999; Balkwill et al., 2004) 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. Bradley and Lang, 1994; Faith and Thayer, 2001; Gabrielsson & Lindström, 2010; Nielzén and Cesarec, 1981; Wedin, 1972) such as Russell’s arousal and valence dimensions (Russell, 1980). Arousal, in particular, is suggested to have cross-culturally general relationships with musical features such as tempo (see review in Singh & Mehr, 2023). Categorical emotions, on the other hand, are based on the basic emotion model (Darwin, 1872; Ekman, 1992; Izard, 2007) which consists of six or higher categories of emotion such as sadness, anger, happiness, fear, and disgust. Prior research has shown several categories of emotions which could mostly be detected by the listeners (Juslin and Laukka, 2004; Sloboda, 1992; Wells and Hakanen, 1991; Zentner et al., 2008). 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.

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 (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 (cf. Ozaki et al., In press). We delve deeper into the relationship between musical tempo, visual density, and emotional arousal 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**

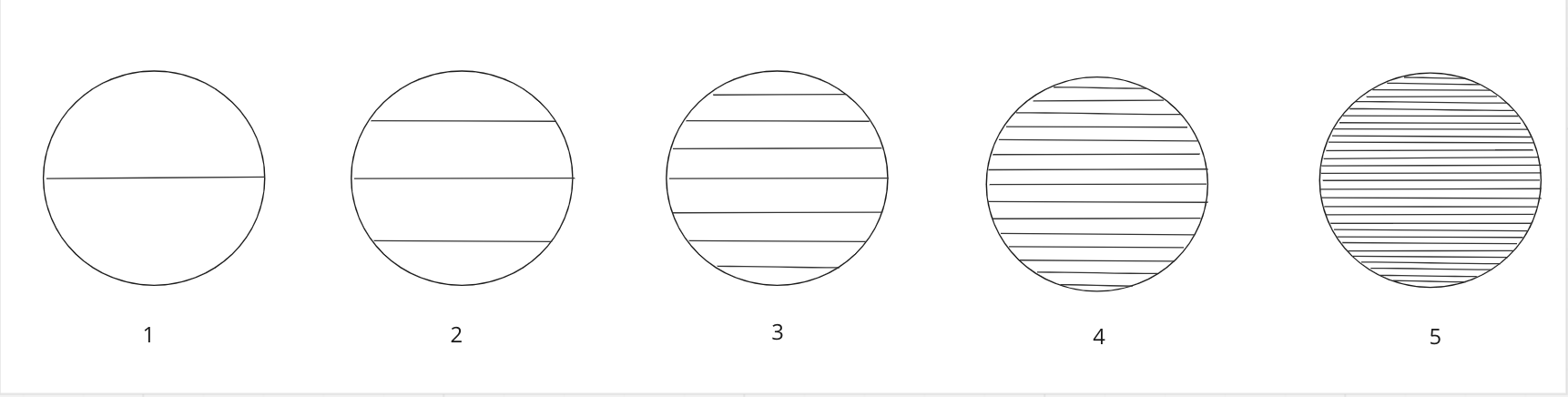
Our study aims to understand cross-cultural relationships between musical features, visual imagery, and emotions in Japan, Iran, and Canada. Based on our pilot data, we plan to compare these three populations through collection and analysis of 1) music to visual density associations, and 2) emotion appraisal of excerpts manipulated in pitch and tempo. It is worth noting that implicit associations are very useful in cross-cultural studies as translations and naturally occurring connotations around language can be misleading and are preferably avoided (Athanasopoulos, 2022). 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 metaphorically associated with any of our musical variables, which are tempo and rhythm complexity.

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 connections between the variables 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.

In our study, variation in rhythmic density (number of note onsets per second, or “tempo”[[1]](#footnote-1)) is captured through a selection of solo and group excerpts, which are also directly manipulated by speeding up/slowing down the recording tempo. Meanwhile, visual density is presented through horizontal straight lines against a white blank background and ranges from low to high on a scale from 1-5 (Fig. 1). For example, a higher number of lines within the same background area is equal to a higher density and vice versa. Visual density has yet to be empirically examined in cross-modal studies, however, it can provide some insight into music cognition across individuals and cultures since we use metaphors from visual textures to conceptualize musical complexity in terms of musical texture.

**Figure 1**. **Visual Density Stimuli.** Stimuli are adapted from Langlois et al. (2014) and arranged from very sparse (1) to very dense (5).

Empirical research from Antovic (2009a) and Eitan and Timmers (2010) has found diverse cross-modal mappings concerning musical pitch. To some extent, their results supported conceptual metaphor theory (CMT: understanding of a concept from a specific domain through the usage of metaphors in another domain) brought forward initially by Johnson and Lackoff (1980), but also suggested other potential mechanisms for these cross-modal mappings (Antovic, 2011). In this instance, do we make sense of rhythmic density in music through a visual density metaphor? If we associate excerpts with higher rhythmic density with images with higher visual density, we will have some evidence that suggests CMT might be the underlying mechanism. Musical features which are abstract in nature are conceptualized through more familiar notions and understood metaphorically. On the other hand, embodied cognition of time (here highlighted as the passage of time in music) is another potential metaphorical mechanism through which we conceptualize tempo in music. It is not fully possible to use linguistic reasoning to understand how musical features such as tempo interact with visual density e.g. whether and how tempo changes alter these music to visual density associations; however, they can be attributed and supported by notions of Time-Metaphors and Space-Time Metonymies by Lackoff and Johnson.

Lackoff and Johnson (1999) indicate that some metaphors for time are consistently found in different languages which explains how we relate motion to time, i.e., we move in relation to other objects and people while time is passing. Through this idea of time-motion metaphor, our study might suggest that rhythmic density to visual density associations refer to the passage of time and the density of musical events represent the amount of motion through time (higher density of musical events attributed to more motion, and subsequently associated with more lines and in this case, higher visual density).

**Table 1. *Peer Community In Registered Reports* study design template**

| **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=24 per cultural group [72 total]) will be done through snowball sampling (see appendix 1) | We will test our hypotheses by performing a cumulative link mixed model (CLMM) on our dependent variable (here, arousal) for high vs. low tempo versions of the 6 excerpts. | Sample size decided for each group by power analysis assuming an effect size of β = 1, power of 95%, and study-wide alpha = .05 controlling for our two hypotheses (i.e., p<.025). (cf. Appendix 4 for details) | If our predicted effect is significant in a contrast within each of the three cultural groups, we will conclude that tempo-arousal relationships are cross-culturally general. Otherwise, our results will be inconclusive (as equivalence testing is not feasible for our team using our CLMM analysis framework). | If we find a consistent tempo-arousal relationship, it would contradict cultural relativist theories. |
| 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 hypotheses**

We will use Cumulative Link Mixed Models (CLMM; cf. Appendix 4) to test the two hypotheses listed in table 1:

1. Increasing tempo consistently increases emotional arousal across cultures.

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

**2. Method**

We plan to recruit participants from Japan, Iran, and Canada (n=24 per group; see Power Analysis below), who are raised in these countries and whose first language is Japanese, Farsi, or English, respectively (e.g., excluding Canadians who speak French as a first language), 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. Solo pieces involve a solo instrument (Japan: *Shakuhachi*, Iran: *Ney*, and Canada: Fiddle) and group pieces involve multiple instruments playing simultaneously (Japan: *matsuribayashi* (festival ensemble), Iran: kamanche and dohol (we will replace the pilot stimuli for Iranian group music with another for better audio quality. The new excerpt consists of Persian kamanche and tombak from Lorestan province, and claps), Canada: fiddle, banjo, mandolin, bass, and guitar). 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. By comparing two excerpts manipulated equal amounts from the original rather than comparing the original with a modified version, both versions are equally different from the original and thus any differences due to the piece being heard faster/slower than intended should cancel each other out (if we compared the original with a modified version, the variable tempo would become confounded with the presence/absence of tempo manipulation). We selected these specific tempi 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. The musical stimuli were manipulated in Audacity and various tempo changes (10-50%) were examined in order to find the optimal change that was noticeable enough and natural-sounding. There was a consensus among the co-authors that tempi manipulations greater than 20% had excessive sound artifacts while manipulations less than 20% were not different enough from the original.

Since the solo and group excerpts differ not only in terms of solo/group but also in terms of tempo and other variables, it is not possible to compare solo vs. group recordings to draw strong conclusions about the effect of tempo or other variables. Instead, we will make such comparisons in a purely exploratory manner, and restrict our confirmatory hypothesis testing only to our controlled experiments manipulating tempo (i.e., the paired responses represent direct comparison of faster and slower versions of the same musical excerpt by the same participant).

**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 in their respective native language. Since the surveys and scales we are using are not always translated to foreign languages, the co-authors will translate and help in data collection by reviewing the translations and making sure the study is clear and feasible in their native language.Participants will be randomly presented with 24 excerpts (6 recordings each manipulated 4 ways to be higher/lower in tempo/pitch), and asked to rate their arousal from a scale of 1-5. They also select an image of 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. 1, 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. 1). 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. Each excerpt can be played on 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. The experiment takes about 30-40 minutes.

**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 (and replaced with new participants until we meet our target number of participants). 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 (Appendix 4) estimated that we would require a sample of at least 24 participants from each of the three cultural groups (i.e., *n* = 72 participants total) for our study to have 95% power to test both directional (one-tailed) hypotheses assuming an effect size of *β* = 1, as defined for a Cumulative Link Mixed Models (CLMM) (Borders et al., 2022; Taylor et al., 2022; for details see Appendix 4) 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; Albers & Lakens, 2018), we note that all of our pilot data groups demonstrated effect sizes greater than 1 (minimum: ~1.2 [effect of tempo on visual density for solo music]; maximum: ~3.2 [effect of tempo on visual density for group music]).

Our power analysis is based on a CLMM, to account for the fact that our dependent variables are from 5-point Likert scales, rather than normally distributed continuous data, and that the 6 paired responses from each participant are not independent of one another. Unlike metric models, CLMM have been shown to provide consistent results and are recommended for models of ordinal scales (Peng et al., 2023). Considering this, the effect size of *β* = 1 corresponds to the coefficient (thus, the predicted change between tempo conditions). However, CLMMs assume that the *real* dependent variable is a latent, unobserved distribution, from which we can only observe its categorisation (the 5 points in the Likerts scale). In this case, this latent distribution is modeled as a logit probability of observing a given score, so all model predictions and coefficients are in the units of that estimated latent, continuous variable (log odds). While power analysis can be performed analytically for certain statistical tests such as t-tests and one-way ANOVA, for more complex models and factorial designs, a Monte Carlo simulation-based approach is required (Lakens & Caldwell, 2021). In this approach, a statistical model is defined with a specific effect size of interest. Multiple synthetic datasets are then generated based on this model, and the proportion of datasets that result in significant results (i.e. rejecting the null hypothesis) is computed. This proportion represents the Monte Carlo estimate of the test's power to detect the desired effect size.

To do this, we simulated a population of 10,000 participants (6 paired responses per participant, for a total of 120,000 ratings of both visual density and arousal), and randomly selected 1,000 samples from this simulated population. The size of each sample was adjusted until we reached the desired statistical power. For each sample, models to test both arousal and visual density ratings were fitted with fixed effects for Tempo (Low, High), Instrumentation (Group, Solo), Participant country (Iran, Canada, Japan), and music country (Iran, Canada, Japan) and their interactions, as well as random intercepts and slopes between Tempo conditions for each participant, ensuring a fair representation of the experimental design (Barr et al., 2013; DeBruine & Barr, 2021). However, since our hypotheses are focused solely on the effects of manipulating tempo, the power analysis primarily examines the contrast between Tempo conditions, independent of Instrumentation, participant country, and music country levels.

3. Exploratory variables

In addition to our primary variables described above that will be collected for confirmatory hypothesis testing, we plan to also collect additional variables for post-hoc exploratory analyses and future hypothesis generation (see Appendix 2 for full details of all variables). In addition to standard participant demographic variables (e.g., age, gender, occupation), we plan to collect variables related to musical experience including the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014) and related to visual arts and emotions including synaesthesia, aesthetic preferences, and music-colour associations (following Palmer et al., 2013). These exploratory variables will not be used to draw any strong conclusions in our Stage 2 Registered Report.

**Data/code availability**

Our data and code are available at

<https://github.com/comp-music-lab/VisualEars>

Our musical stimuli are available at <https://osf.io/pkvw2/?view_only=0894653041ba4375afdcaa3a1989fe71>

**Acknowledgments**

We would like to thank Hideki Sakanashi, Mai Fujiwara, and Yasuo Shiozawa for the discussion; Yuto Ozaki for help making figures; and all participants in the pilot experiment and all Keio University CompMusic Lab members for their feedback and help. We thank Sam Schwarzkopf, Nadine Dijkstra, and Elena Karakashevska for detailed feedback on previous versions of the manuscript.

**Author contributions**

Conceptualization: Hadavi

Methodology: Hadavi, Savage, Kuroda, Shimozono

Investigation: Hadavi

Writing: Original draft: Hadavi

Formal analysis: Leongómez, Savage, Shimozono

Writing: Review and editing: Hadavi, Savage, Kuroda, Shimozono, Leongómez

Supervision: Savage

Project management: Savage, Kuroda

Funding acquisition: Savage, Kuroda

**Inclusion criteria**

We commit to the best practices in cross-cultural studies (Tan & Ostashewski, 2022; Savage, Jacoby, Margulis, et al., 2023), such as involving coauthors from diverse backgrounds from the initial planning phases of a study.

**Funding/Conflicts of interest:**

This project is supported by funding from the Yamaha Corporation, including several authors who are Yamaha employees. The authors affiliated with Yamaha will only have access to the same anonymized data that will also be available to the public. 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. Savage and Leongómez are Recommenders at *Peer Community In Registered Reports*.

**Ethical approval**

We have received approval (#546) from the Keio University Shonan Fujisawa Campus Research Ethics Committee.

**References:**

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, 74, 187–195. <https://doi.org/10.1016/j.jesp.2017.09.004>

Antovic, M. (2009). Musical metaphors in Serbian and Romani children: An empirical study. *Metaphor and Symbol*, 24(3), 184–202. <https://doi.org/10.1080/10926480903028136>

Antovic, M. (2011). Musical metaphor revisited: Primitives, universals and conceptual blending. Universals and Conceptual Blending, SSRN Electronic Journal. [http://doi.org/10. 2139/ssrn.1763503](http://doi.org/10)

Athanasopoulos, G. (2022). Musical Imagery From a Cross-Cultural Methodological Perspective. In *Music and Mental Imagery* (pp. 123-133). Routledge.

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>

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>

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>

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>

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>

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>

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>

Borders, J. C., Grande, A. A., & Troche, M. S. (2022). Power Analysis with Ordinal Outcomes: A Tutorial.

<https://osf.io/8sc5e>

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>

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>

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>

Chiba, G., Ozaki, Y., Fujii, S., & Savage, P. E. (2023). Sight vs. Sound judgments of music performance depend on relative performer quality: Cross-cultural evidence from classical piano and Tsugaru shamisen competitions. *Collabra: Psychology*, 9(1). <https://doi.org/10.1525/collabra.73641>

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>

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>

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920965119. <https://doi.org/10.1177/2515245920965119>

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>

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>

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*, *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>

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>

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>

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>

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>

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>

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>

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>

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

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>

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>

​​Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *The Behavioral and brain sciences*, *31*(5), 559–621. <https://doi.org/10.1017/S0140525X08005293>

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>

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. [https://doi.org10.3389/fpsyg.2014.00789](about:blank)

Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920951503. <https://doi.org/10.1177/2515245920951503>

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>

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>

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>

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)

Ozaki, Y., et al. (In press). Globally, songs and instrumental melodies are slower, higher, and use more stable pitches than speech: A Registered Report. *Science Advances*. Preprint: <https://doi.org/10.31234/osf.io/jr9x7>

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>

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>

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>

Peng, D.-C., Moreau, D., Cowie, S., & Addis, D. R. (2023). Cross-national replication of prosocial simulation effect using cumulative link mixed modelling. *Journal of Applied Research in Memory and Cognition*.<https://doi.org/10.1037/mac0000117>

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>

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)

Sarrazin N. R. & Open SUNY Textbooks. (2016). *Music and the child*. Published by Open SUNY Textbooks Milne Library (IITG PI) State University of New York at Geneseo. Retrieved May 25 2023 from <https://library.biblioboard.com/content/25c56965-e083-4fad-a071-c29a47e02c36>.

Savage, P. E., & Fujii, S. (2022). Toward a cross-cultural framework for predictive coding of music. *Nature Reviews Neuroscience*, 23, 641. <https://doi.org/10.1038/s41583-022-00622-4>

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. (2023). 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. (347-365, MIT Press). <https://doi.org/10.7551/mitpress/14186.003.0032>

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>

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. (2023). Universality, domain-specificity, and development of psychological responses to music. *Nature Reviews Psychology*, *2*, 333–346.<https://doi.org/10.1038/s44159-023-00182-z>

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>

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>

Taylor, J. E., Rousselet, G. A., Scheepers, C., & Sereno, S. C. (2023). Rating norms should be calculated from cumulative link mixed effects models. *Behavior Research Methods, 55*, 2175–2196. https://doi.org/10.3758/s13428-022-01814-7

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>

Thompson, W. (2010). Cross-cultural similarities and differences (Music and Emotion) (pp. 755–788).<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>

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. Cognition, 101(1), 80–113. <https://doi.org/10.1016/j.cognition.2005.09.0>03

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>

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>

*Appendix 1*

**[Accepted versions of Appendices 1-3 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 30-40 minutes and you will be compensated with $10.

Thank you very much.

*Appendix 2*

**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.)**

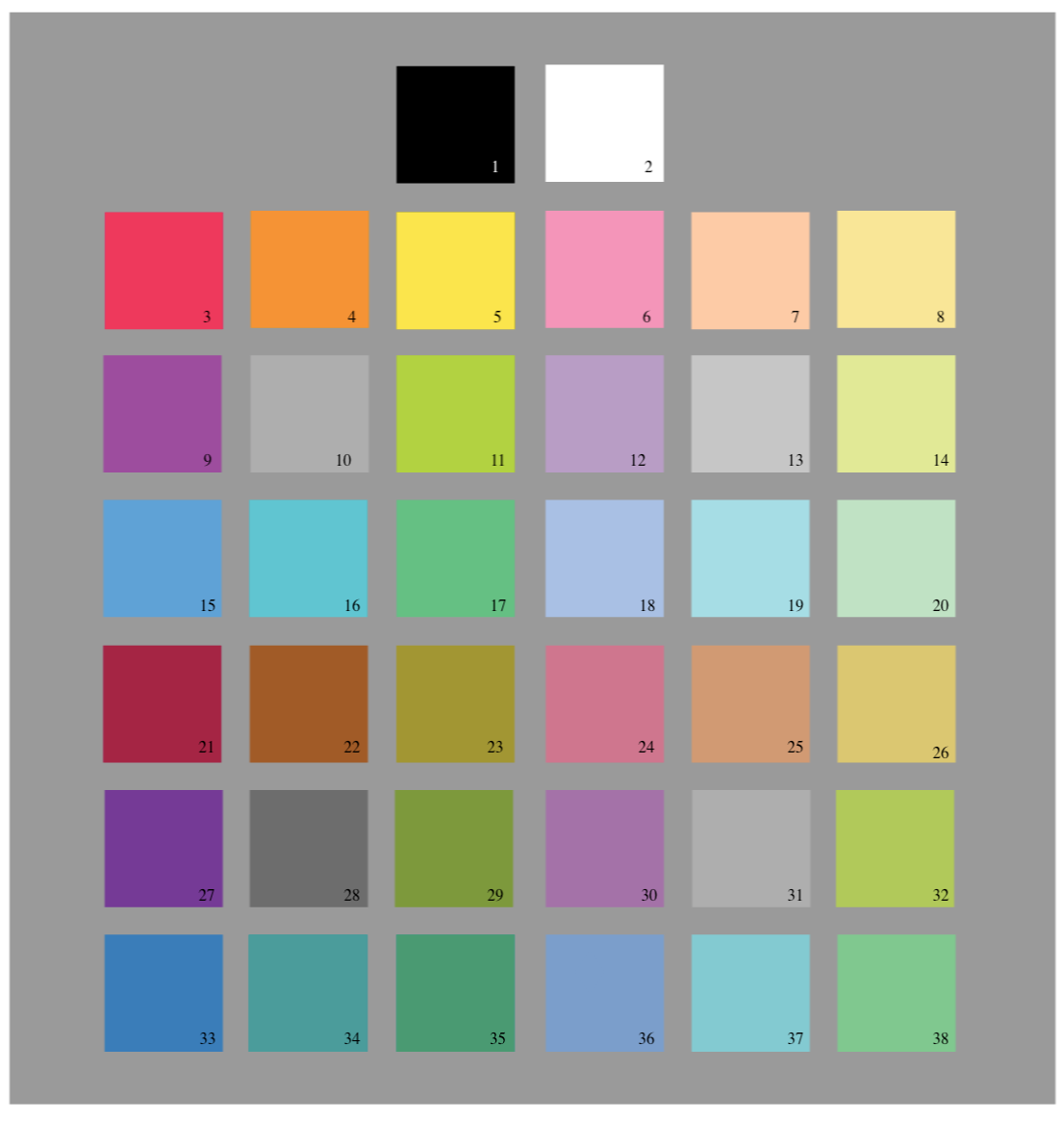
1. Please listen to 24 short musical excerpts linked below and enter your response to each excerpt by selecting the appropriate number.

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

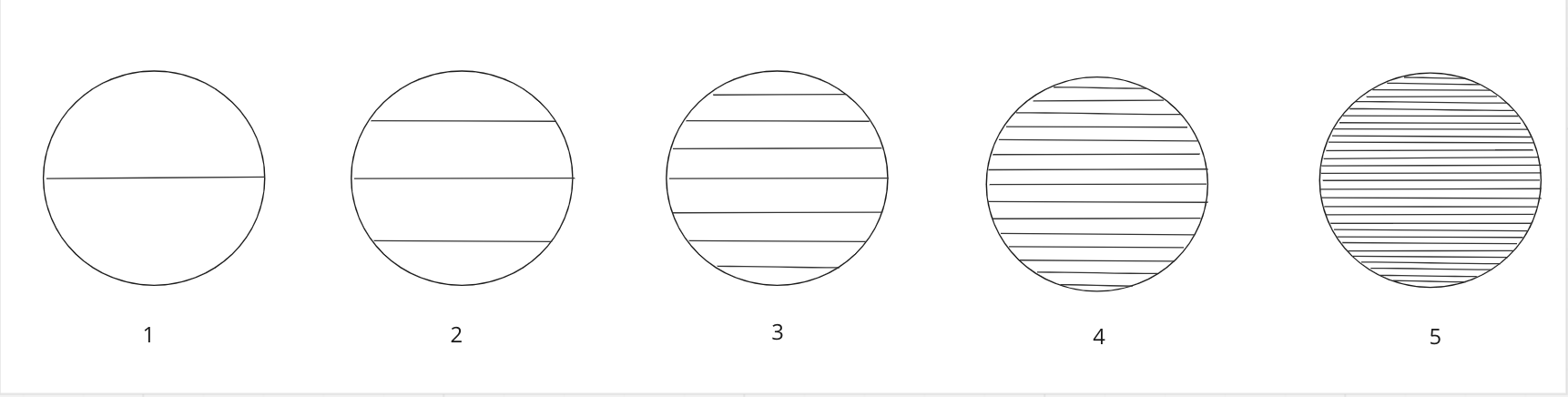
1. Using the colors in the image below, indicate one color that most represents each musical excerpt and come to your mind while listening to these excerpts.
2. Using the textures below,indicate one texture that most represents each excerpt.
3. Rate the arousal using scales 1-5: 1 being very subdued 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 specified** excerpts, enter Y if you have heard the piece before today, and N if you have not heard it before today.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Arousal  scales 1-5: 1 being very subdued and calm, and 5 being very excited and aroused | Most consistent texture | Most consistent color | 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 |
| Excerpt 1 |  |  |  |  |  |  |  |
| Excerpt 2 |  |  |  |  |  |  |  |
| Excerpt 3 |  |  |  |  |  |  |  |
| Excerpt 4 |  |  |  |  |  |  |  |
| Excerpt 5 |  |  |  |  |  |  |  |
| Excerpt 6 |  |  |  |  |  |  |  |
| Excerpt 7 |  |  |  |  |  |  |  |
| Excerpt 8 |  |  |  |  |  |  |  |
| Excerpt 9 |  |  |  |  |  |  |  |
| Excerpt 10 |  |  |  |  |  |  |  |
| Excerpt 11 |  |  |  |  |  |  |  |
| Excerpt 12 |  |  |  |  |  |  |  |
| Excerpt 13 |  |  |  |  |  |  |  |
| Excerpt 14 |  |  |  |  |  |  |  |
| Excerpt 15 |  |  |  |  |  |  |  |
| Excerpt 16 |  |  |  |  |  |  |  |
| Excerpt 17 |  |  |  |  |  |  |  |
| Excerpt 18 |  |  |  |  |  |  |  |
| Excerpt 19 |  |  |  |  |  |  |  |
| Excerpt 20 |  |  |  |  |  |  |  |
| Excerpt 21 |  |  |  |  |  |  |  |
| Excerpt 22 |  |  |  |  |  |  |  |
| Excerpt 23 |  |  |  |  |  |  |  |
| Excerpt 24 |  |  |  |  |  |  |  |

**Color**

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

**Texture**



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

2. Do you have synesthesia? (Synesthesia is when one sensory input is experienced simultaneously in other senses, for example, sounds are heard and experienced as colors ) Yes/No

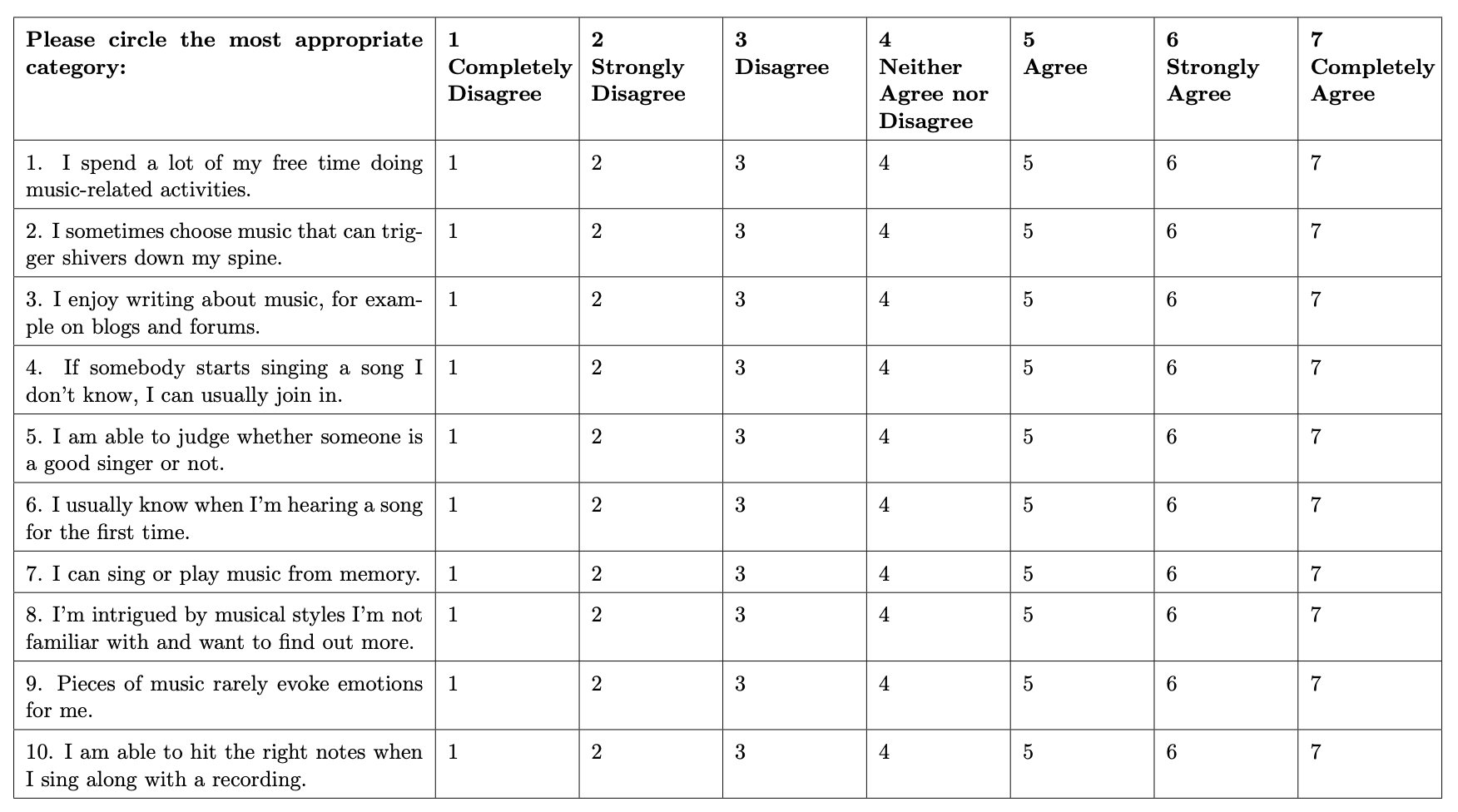
3. How many years of formal training in visual arts have you had?

4. Age: ….years/Prefer not to answer Gender: Male/Female/Other/Prefer not to answer

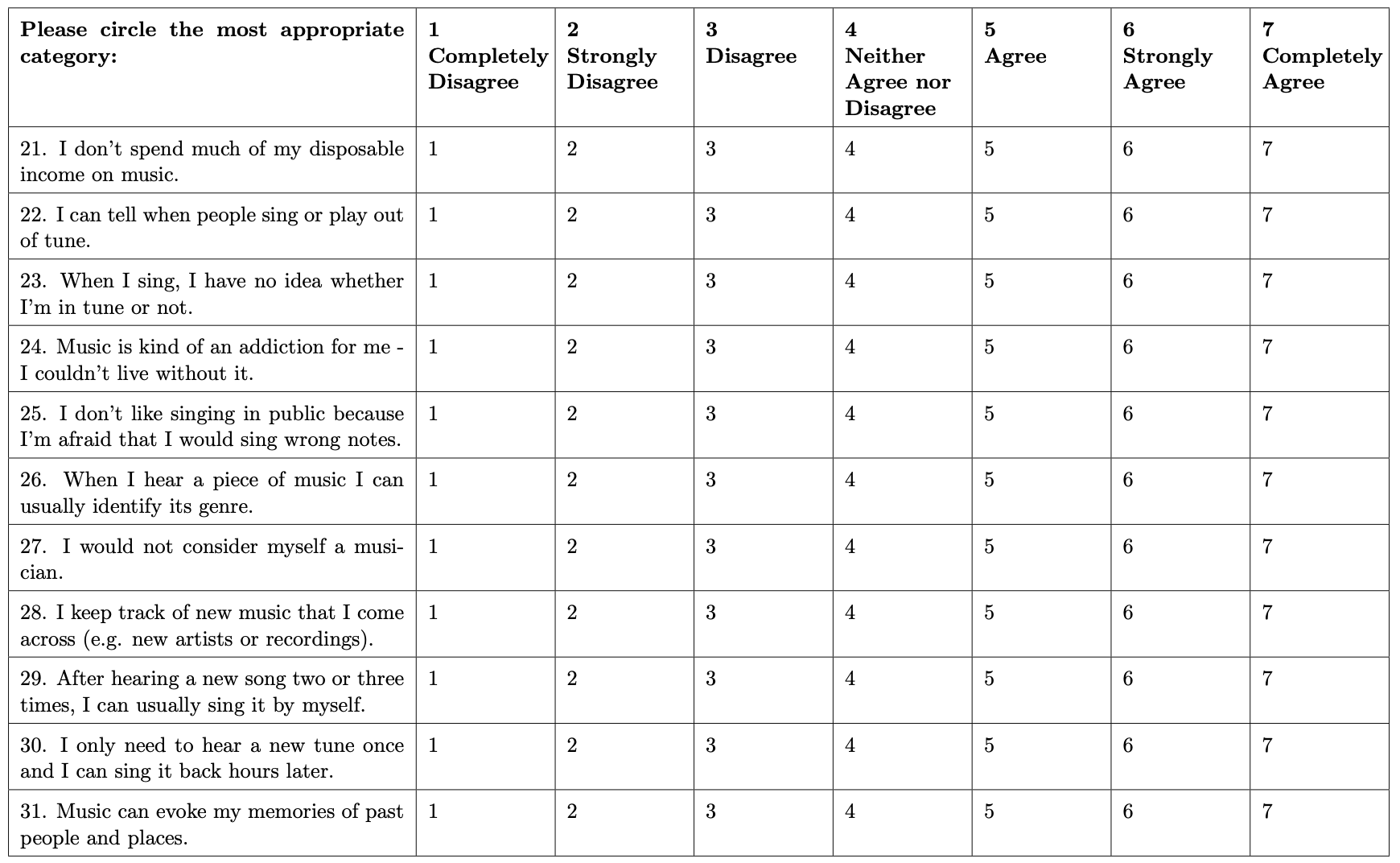
5. Please fill out the following questionnaire by selecting the most appropriate item.

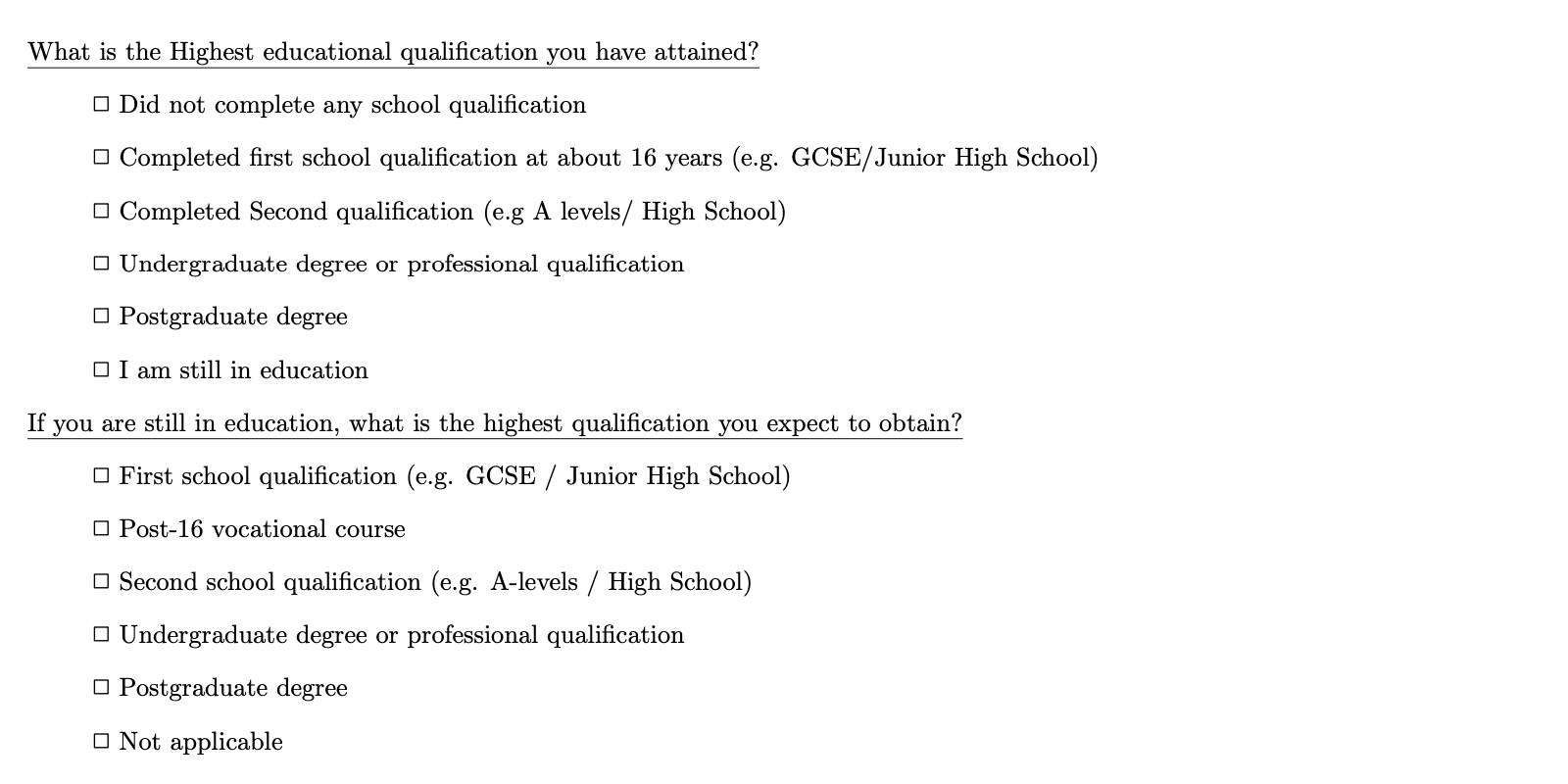
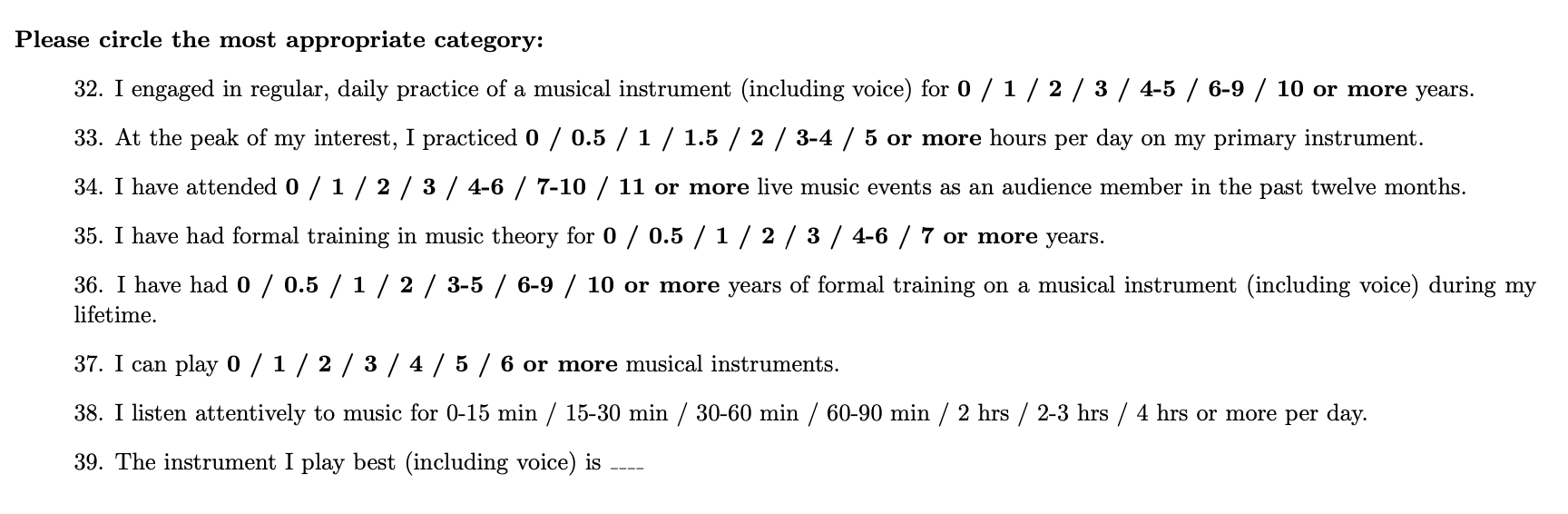
The Goldsmiths Musical Sophistication Index, v1.0

October 11, 2012









*Appendix 3: Informed consent form*

**Study Name**: Visualears Project

**Researchers**:

Shafagh Hadavi (PhD student, Keio University; [visualears.project@gmail.com](mailto:visualears.project@gmail.com); lead researcher)

Dr. Patrick Savage (Associate Professor, Keio University, [psavage@sfc.keio.ac.jp](mailto:psavage@sfc.keio.ac.jp); Principal Investigator)

**Purpose of the Research: To explore the relationship between music, visuals, and emotion.**

**What You Will Be Asked to Do in the Research**: You will be asked to use headphones, and listen to several excerpts of music and rate the emotional content (arousal/valence) on a scale from 1-5, and select from eight emotion categories. You will also be asked to pair each excerpt to various visual stimuli (visual textures, and a colour palette) and fill out a questionnaire about your musical and visual training.The experiment will take about 30-40 minutes.

**Risks and Discomforts**: We do not foresee any risks or discomfort from your participation in the research. You have the right to not answer any questions.

**Benefits of the Research and Benefits to You**: You will be compensated for an hour of your time based on the standard rates of Keio University for research participation. Cross-cultural studies involving cross-modal research questions has the potential to reveal insights into human perception and cognition and may as a result, help in improving our understanding and developing practical applications to enhance the quality of human experience in various fields such as music and art therapy.

**Voluntary Participation**: Your participation in the study is completely voluntary and you may choose to stop participating at any time without any penalty.

**Withdrawal from the Study**: You can stop participating in the study at any time, for any reason, if you so decide. Your decision to stop participating, or to refuse to answer particular questions, will not affect your relationship with the researchers, or any other group associated with this project. Note that once you have completed and submitted the study, it will no longer be possible to withdraw your data because responses will be anonymized.

**Confidentiality**: By agreeing to participate in this research, you give consent to the use of your responses to be evaluated and included in an OSF repository online as well as a journal publication/academic presentation/conference. The data is anonymous and will remain anonymous.

**Questions About the Research?** If you have questions about the research in general or about your role in the study, please feel free to contact through visualear.project@gmail.com. This research has been reviewed and approved by the Keio University SFC Research Ethics Committee. If you have any questions about this process, or about your rights as a participant in the study, please contact [rinri@sfc.keio.ac.jp](mailto:rinri@sfc.keio.ac.jp)

**Legal Rights and Signatures**:

I consent to participate in *Visualears Project*. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by agreeing to participate in this experiment.

**I agree I do not agree**

***Appendix 4: Detailed power analysis***

1. For comparability with past research we use the term “tempo” to describe rhythmic density, but note that tempo is often defined in terms of beats per minute assuming a fixed pulse, which does not apply to some of the musical stimuli in our study, such as the ametric Japanese shakuhachi music (cf. Savage & Fujii, 2022). Our “rhythmic density” / “tempo” variable is also analogous to the “temporal rate” variable used by Ozaki et al. (In press) to compare rhythmic density in music and speech. [↑](#footnote-ref-1)