# A Proof-of-Concept Demonstration Using Musical Stimuli for The Measurement of Psychological Individual Differences

**Adrianne John Galang, Monica Renee Policarpio, April Meryll Xu, Ma. Jenina Nalipay, & Rafael Cabredo**

**Abstract**

Given that there are well documented differences in terms of preferences for music, this study explored whether very basic melodic stimuli can be used as correlates of some psychological individual differences such as personality traits. We created a set of short musical stimuli which we manipulated in terms of basic melodic properties such as pitch range, key, and average pitch. In a university sample with a final count of 103 valid cases, we collected preference ratings, pupil dilation data, and questionnaire data on personality and demographic variables. We then used a machine learning technique to build decision trees that would provide insights on what patterns might be useful to pursue further in subsequent work. We found that there were suggestive patterns regarding how preferences for the set of musical stimuli related to self-reported personality traits. Despite very important limitations to the current results, we believe that there is potential for refining this technique to yield a potentially useful paradigm in the measurement of psychological individual differences.

*Keywords: decision trees, mixed effects, individual differences, musical stimuli, exploratory analysis*

## Introduction

The study of personality often involves the measurement of behaviors through self-reports (John, Naumann, and Soto 2010), through assessing patterns of reasoning (James, 1998), or by inferring apperception through projective techniques (Murray, 1938). One strategy that has not been fully investigated is the use of nonverbal stimuli to study individual differences in preference. This is made plausible by experiments that have shown that perceptual preferences are related to certain personality traits (Barron and Welsh, 1952; Chamorro-Premuzic, Fagan, & Furnham 2010; Geen, 1984). An attempt in measuring individual difference using visual stimuli preference is seen in the Barron-Welsh Art Scales (Barron and Welsh, 1952). The possible advantage this confers is its non-verbal means of administration which can cater to special populations and young children.

However, the properties of visual stimuli, especially when sufficiently complex, may be difficult to quantify and thus may not be easily subjected to certain computational methods. Musical stimuli, on the other hand, are readily formalized with a level of precision not practical for visual stimuli. This can allow a more data-driven investigation not available with previous methodologies. At the same time, the fact that we have a well-articulated framework for specifying musical stimuli holds out the potential that findings from such procedures will be more readily interpretable than that using other modalities as stimuli.

The objective of the current study is to explore the potential of basic musical stimuli and machine learning algorithms as tools for investigating the psychology of personality. We are not particularly interested in how people respond to actual music in their daily lives, and thus we are unconcerned about genre-based classifications or preferences relating to complete pieces of music. For the meantime, we are confining ourselves to people's responses to very basic and very short musical stimuli. In this preliminary study, we are also less interested in coming up with a good predictive model of preferences. Rather, we are using a machine learning technique as a heuristic for theory-generation. We prefer to be circumspect about the utility of the specific models generated in this study and leave their validation to future projects collecting fresh independent data from a new sample. Should this strategy prove fruitful, it could yield insights into a new paradigm of non-self-report methods of assessing interesting psychological variables.

There is existing work that has explored the use of machine learning in predicting personality from social network site use. For example, Ortigosa, Carro, and Quiroga (2014) hypothesized that users having similar personality will show similar behavioral patterns when interacting through virtual social networks, and these patterns can be mined to predict user personality. To test this hypothesis, they developed a Facebook application to collect information about 20,000 users regarding their personality traits and their interactions within Facebook. Automatic classifiers, which are essentially predictive models, were trained by using different machine-learning techniques to look for interaction patterns that provide information about users' personality traits based on collected data. These classifiers were found to be able to predict user personality using parameters related to user interaction, such as number of friends or wall posts. Results show that the classifiers have a high level of accuracy, making the proposed approach a reliable method for predicting user personality. There have been other research along similar lines that have used similar text-mining strategies to discover interesting relationships between social media language use and personality (Schwartz, et al., 2013).

Kosinski, Stillwell, and Graepel (2013) have already shown that individual characteristics like male sexual orientation, political ideology, and to some degree personality trait scores on 20 items derived from the International Personality Item Pool questionnaire (Goldberg et al., 2006), can be predicted using data on Facebook "likes" from a large volunteer sample (approximately 58,000). Although the predictive ability of "likes" data was not better than the questionnaire instrument itself, they go on to show in a later study (Youyou, Kosinski, and Stillwell, 2015) that Facebook "likes" data can perform better than close others (e.g., friends) at predicting self-rated personality traits in a sample of 86,220 participants. At the very least, this line of work is showing that preference data, in this case, Facebook likes, can on its own be an alternate way to index personality traits.

### *The relationship between musical preferences and personality*

Some important work in the study of musical preferences and personality was undertaken by Rentfrow and Gosling (2003). Because the concept of musical genre (e.g., "country", "classical", "heavy metal") has always been imprecise and problematic, Rentfrow and Gosling sought to reduce the complexity to coherent dimensions through exploratory factor analysis of ratings of 14 musical genres. Four dimensions emerged from their data, which they termed as Reflective and Complex, Intense and Rebellious, Upbeat and Conventional, and Energetic and Rhythmic. They go on to show that these four dimensions have significant correlations with personality traits and social attitudes. This study is notable because it represents one of the few systematic attempts to comprehensively describe musical preference. It also goes some way to addressing the conceptual problems of studying musical genres, although problems remain especially given that the starting items still assume certain cultural knowledge of the identified genres.

Cattell and Saunders (1954) represent a very early attempt to explore possible connections between the formal parameters of music and personality. Using 120 musical excerpts played on piano, they had 196 university level students and 188 patients from a psychiatric institution listen to and indicate their liking for or dislike of the musical excerpts. Cattell and Saunders' preliminary interpretations of the factor structure of the preference ratings were based on the 16 personality factors identified by Cattell (Cattell, 1956). Thus, for example, one factor loaded the preference for Debussy's "Girl with the Flaxen Hair" with Mendelsshon's "Violin Concerto, Mov. 2" and was interpreted as having "a quality of warmth and gentleness, such as might be expected to appeal to the cyclothyme Factor A or H" (p. 19).

Dobrota and Ercegovac (2014; and see also Ercegovac, Dobrota, and Kuščević, 2015) more recently extended the investigation described by Cattell and Saunders (1954) by testing for actual correlations between preference for musical features. A sample of 202 students completed a scale of optimism and pessimism, and items from the International Personality Item Pool (IPIP) for measuring personality, as well as rating fragments of actual music. They observed gender differences in that females tended to have a preference for music in the major key and fast tempo. Extraversion, Agreeableness, and Conscientiousness correlated with preferences for the combination of major key with fast tempo, while only Openness predicted preferences for music in the minor key and slow tempo. Only the result for Openness remained significant once all predictors were tested in a hierarchical regression. These results however have a clear limitation because the way they selected their stimuli leads to a confound between tempo and musical key: all fast tempo songs used were also exclusively in the major key, while all slow tempo songs were in the minor key.

Similar research exploring links between personality questionnaire scores and preference for certain musical aspects have been conducted, but they either used too few musical excerpts (48 in the case of Ladinig and Schellenberg, 2012), or the relationship between preferences and musical aspects were confounded with familiarity (Dunn, 2009).

## *Importance of isolating links between specific musical features and personality*

Several information processing accounts of personality assume that individual differences in behavior and emotionality are in part due to dispositional factors that affect how a person perceives the world, beyond the mere physical facts of stimuli (e.g., Higgins and Scholer, 2008; Mischel and Shoda, 1995). Situations are presumed to have elements that interact with personality to produce differential outcomes (Rauthmann, et al., 2014). Musical stimuli are presumably not exempt from this relation, given the role it plays in many traditional, religious, and secular societies. To be able to describe this relationship in a precise way would help advance the understanding of personality and social psychology.

Although there are some existing theories regarding aesthetic preference, such as Berlyne's arousal-based framework and newer cognitive appraisal approaches (Silvia, 2005), the current study will make few theoretical assumptions and will be more empirical and atheoretical in nature, following the pragmatic exploratory strategy employed in the creation of the Minnesota Multiphasic Personality Inventory (MMPI). The MMPI is considered by many to be the gold standard for empirically developed psychometric scales (Wiggins, 2003). By being guided almost exclusively by the inter-correlations between items and relevant target criteria, scales are generated based on data rather than theory.

For our own initial exploration, instead of directly predicting personality outcomes, we wanted to examine whether there were any links at all between features of melodies and personality traits and other individual difference variables, so we tried to predict people's responses to melodies using both the intrinsic features of the melodies themselves, as well as how these interacted with the individual characteristics of respondents.

We also want to evaluate the utility of non-self-report assessments of preference. There is some evidence that pupil dilation might be associated with forms of psychological response to music, such as arousal or liking (e.g. Gingras et al., 2015; Bianco et al., 2019; see also Laeng et al., 2016 for musical chills), so we decided to record pupillary responses to our melody stimuli to explore the possibility that they could provide an independent measure of our outcome that we can use to validate their self-rated liking for the melodic segments.

In the project described here, we used a decision tree analysis method that was capable of handling multilevel data (Fokkema, Edbrooke-Childs, and Wolpert, 2021) to predict how positive each respondent's rating was for each melody. Of particular interest would be patterns that indicate interactions between musical features and individual difference variables.

## Methodology

### *Sample*

From an initial university sample of 150, we excluded cases where no valid data from the melody preference procedure were recovered, and a further 5 cases based on their score on the amusia screening test. We also excluded one case of a 60 y.o. college instructor who indicated that they had 800 citations for their published work, since we thought that they were a multidimensional outlier on several counts in the context of a mostly young adult student sample. The final sample count was 103 participants (58 females) with an average age of 19.7 years (s.d.=1.6).

### *Materials and Instruments*

#### *Stimuli*

We used 16 music files that were generated from an earlier study using a digital music editor. All melodies were melodic segments consisting of 2 measures with 4 quarter notes for each measure (assuming a time signature 4/4).

The study focused on three musical features: pitch range (how far apart the highest pitch was from the lowest pitch), key or mode, and average pitch. There were 2 variants for each feature, which when combined yields 8 versions of each melody. Pitch range was manipulated by creating a version of the melody whose range was restricted to within 3 to 4 half-steps, or by creating one that was at least 5 whole steps in range, all the while trying to maintain as much as possible the melodic "shape". Average pitch was manipulated by making versions 5 half-steps apart in terms of pitch (see Figure 1). Sixteen melodies were created by the authors, which were then manipulated according to the 3 attributes to produce 128 melody tracks. The data collected from the earlier study provided a preliminary look at the predictive properties of the 16 music files with its counterpart versions. One music file (along with its counterparts) has subsequently been shown in previous data to be a weak predictor due to its favorable response from a majority of the participants and thus removed. For this study the remaining 120 melody tracks were used.

*Table 1. Variables manipulated in stimuli*

| ***Variable*** | ***Variant 1*** | ***Variant 2*** |
|---|---|---|
| *Pitch range* | *Large interval* | *Small interval* |
| *Mode* | *Major* | *Minor* |
| *Average pitch* | *Higher* | *Lower* |

A. Higher pitch, major key, large pitch interval

B. Lower pitch, minor key, short pitch interval

*Figure 1. Two versions of the same basic melody segment used as stimuli.*

The musical stimuli were presented, and responses were recorded using the Superlab stimulus presentation software (www.cedrus.com). All participants used noise-canceling closed-ear headphones during the procedure. For each musical stimulus the participant decided how much they "Liked" the melody: the participant could choose from a visual analog scale with a range of -3 to +3, and the zero point corresponded to "I don't know".

### *Pupillometry*

We used a 30 Hz Tobii eye-tracking bar to measure pupil dilation during the melody preference procedure. To ensure proper capture of pupil information, a fixation point was presented for each trial before the melody stimuli was played. Respondents were instructed to keep their eyes on the fixation point as they listened to each stimulus and to try to avoid blinking. As in Gingras et al. (2015) we measured the average pupil size 500 milliseconds before the onset of the melody to serve as a baseline, and we then computed the average pupil size during the melody itself. We then computed a pupil response measure based on the proportion (i.e., in percent) by which the baseline pupil measurement changed in response to the stimulus (see Figure 2).

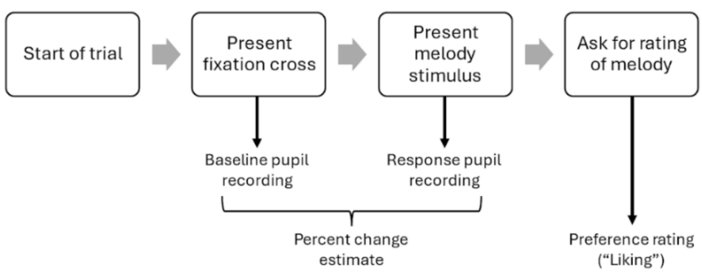


*Figure 2. Diagram of trial structure for melody preference task and pupil measurement.*

*Amusia screening*

There is the distinct possibility that some respondents might have very atypical responses to musical stimuli and thus would be in a class of their own in terms of their preferences. This might be true for conditions such as congenital amusia (Ayotte, Peretz, and Hyde, 2002). We therefore screened participants on the online version of the Montreal Battery of Evaluation of Amusia (MBEA; Pretez, Champod, and Hyde, 2003) which assigns respondents a score out of 100. There was no definite cut-off for this instrument although Peretz and colleagues observed that in their sample people who self-identified as amusic got scores below 78. We opted for a slightly less conservative threshold of 70.

*Self-report measures*

We administered a range of individual differences measures. The 60-item HEXACO Personality Inventory Revised (HEXACO-60; Lee & Ashton, 2004) is a set of self-reported trait questionnaires intended to have a 6-factor structure. For our study we focused on the six dimensions of Honesty-Humility, Emotional Stability (i.e., being less perturbed by negative affect), Extraversion (i.e., sociability and dominance), Agreeableness (i.e., yielding to others to preserve harmony), Conscientiousness (i.e., industriousness and orderliness), and Openness to Experience (i.e., rich intellectual and cultural interests) as predictors for our analyses. For our data, the MacDonald's omega estimates for each scale, respectively, are as follows: .82, .83, .80, .81, .80, and .79 which suggests acceptable reliability.

The Triarchic Psychopathy Measure (TPM; Blagov, Patrick, Oost, Goodman, and Pugh, 2016) was designed to measure subclinical psychopathy along three dimensions: being unrestrained by anxiety or fear and therefore prone to risk taking (Boldness); willingness to be antagonistic (Meanness); and being disorganized and having poor impulse control (Disinhibition). The internal consistency of the measures of these three dimensions was good with MacDonald's omega estimates as follows: .87, .88, and .84.

The Creativity Achievement Questionnaire (CAQ; Carson, Perterson, & Higgins, 2005) is a self-report checklist of creative engagements and achievements on 10 different domains. Answers on the CAQ are weighted depending on how notable the achievement was. We used the overall score for the whole inventory as the variable for our analyses. Because it was a checklist that characteristically produced skewed data and whose domain subsets were not expected to be correlated, we did not estimate its internal consistency.

The PANAS-X (Watson & Clark, 1994) is a self-report measure of affect. For the version we used, respondents were instructed to self-report about their current affective states. We intended the PANAS-X scores to provide contextual information about states that might influence the measurement of responses to the melody stimuli. To maximize the predictive capacity of the PANAS-X and account for as much of the variability in the dependent variables that was due to momentary affective states rather than stable personality factors, we subjected the PANAS-X data to an exploratory factor analysis to find out which groups of items seemed to capture substantial individual differences in affective states. Factor analysis is a technique employed to uncover an underlying set of domains or dimensions when given a much larger set of observed variables (Revelle, 2015). The specific factor analytic technique used was the minimal residual method implemented by the "psych" package (Revelle, 2020) for the statistical analysis software R (R Core Team, 2021). Results showed that four dimensions seemed plausible, but we only utilized the two most substantial components as control variables.

We also included items from the Goldsmiths Musical Sophistication Inventory (Gold-MSI; Müllensiefen et al., 2013) that measures the following musical experiences: years of daily practice, hours of practice during their point of highest engagement, years of training in music theory, number of musical instruments they play, and the amount of time they spend listening attentively to music. We did not include the item on the amount of training on musical instruments they've had since there was a large number of missing data points. Although we also asked about the musical instruments they played and what were their favorite songs we did not include these in our analyses.

For all internal consistency estimates we used the "psych" package (Revelle, 2020).


*Figure 3. Diagram of procedure steps for each session.*

## *Procedures*

Prospective respondents were informed that the study was about music and personality. The site for the data gathering procedures changed across the conduct of the study and we were not able to regulate the climate control conditions of the rooms (we discuss the probable impact of this in the Limitations section). Respondents also scheduled their participation at different times in the day, and we tried to account for this by also including time of day in the model. After an orientation where they provided their informed consent and received their compensation in the form of gift vouchers, the respondents first completed the MBEA, after which they completed the self-report section of the study. They were then presented with the melody preference procedure, and they were led through some practice trials before being allowed to proceed. Respondents took around 30 minutes to complete the procedure, and there was a 5-minute break at the midpoint of the set of trials (see Figure 3).

Because of problems with the eye tracking device, participants sampled later in the course of the project did not have any pupil measurements associated with their sessions. They were still included in the dataset if all their other data points were intact.

### *Analysis*

In addition to the manipulated features, we also calculated the sample (or approximate) entropy for each melody as a rough estimate of the melody's complexity (i.e., how unpredictably the pitch changes). For each melody we coded each note as its distance in half-steps above the lowest note in the melody, which would yield a series of 8 integers. We then used the "pracma" package in R to compute sample entropy for each series (Borchers, 2023). Entropy estimates ranged from -0.15 to 0.24.

We chose decision tree modeling as our main analytic approach mainly because it gave us the exploratory power of machine learning algorithms while yielding models that were more readily interpretable than other methods, and thus can directly aid in theorizing and hypothesis generation (Basak and Krishnapuram, 2005). Although decision trees techniques are well-known as algorithmic methods for data analysis, only lately have they been deployed with multilevel and longitudinal data in mind. Accounting for the non-independence of outcome measures, which is a feature of the specific decision tree package we used and is not present in other packages, minimizes both Type I and Type II errors, compared to other comparable decision tree techniques.

As with other decision trees, the method implemented in the GLMM trees package (Fokkema, Edbrooke-Childs, and Wolpert, 2021) partitions the data set into subgroups based on which predictor variable will maximize the difference in the outcome variable between two groups of cases. For each resulting subgroup, it will continue to split the cases into further groups using statistical tests until a predetermined alpha criterion is no longer met. What is distinctive about GLMM trees is that the statistical test it uses to partition the data is based on a generalized linear mixed model, which takes into account the non-independence of cases.

While useful at generating preliminary insights into the data, results of individual decision trees cannot be taken as strong evidence for a particular set of associations since it tends to capitalize on chance findings and amplify them, whether or not they turn out to be stable or meaningful. The unprocessed but anonymized data (unfiltered) as well as the analysis code can be found at https://osf.io/y5fzm.

## Results and Discussion

We had a set of loose criteria for determining whether a tree was worth interpreting from the point of view of the study objectives. Specifically, we wanted to avoid trees that had such a large number of nodes that generalizations become difficult to infer. We were also more interested in trees that featured interactions between individual difference variables and stimulus features since these might represent the kind of stable personal preferences we would be interested to investigate in further studies.

In order to produce interpretable trees, we varied two parameters. One was the alpha threshold used to make partitions. With the default being .05, setting a smaller (i.e., stricter) alpha of .001 had the effect of reducing the number of nodes, which is one of the advantages of GLMM trees using statistical tests to make partitions (also true for similar methods, like conditional inferences trees, such as in Venkatasubramaniam et al., 2017). A second parameter was the size of terminal nodes (i.e., the groups of outcomes that the tree identified as homogenous based on the predictors). This value can affect the resulting depth of the constructed tree, since it acts as a stopping rule that prevents further partitions if any of the resulting nodes become smaller than the specified threshold (Montesinos-López et al., 2022). Setting this to a minimum size of 100 observations (as opposed to having no minimum) had the effect of also reducing nodes and the overall depth of the tree, again aiding in its interpretability. We emphasize that these thresholds were chosen as heuristic values to maximize intelligibility of the results and thus were largely subjective choices made by the researchers.

We produced multilevel decision tree models of respondent ratings by entering the musical features (key, pitch range, average pitch, entropy) as trial level predictors, and entered the various individual difference measures (except for the PANAS-X), survey items about their past musical training and experience, the respondents' gender and age, as well as the time of day when they performed the study, as person level predictors. We also accounted for the fact that the melody stimuli are not completely independent from each other owing to the fact that they were produced from 15 initial melody segments by entering this as an additional grouping variable aside from individual persons. For the PANAS-X, we performed principal components analysis and extracted the first two factors as indices of positive and negative emotional states and used them as control variables, along with self-reports given during debriefing that indicated whether the respondent felt some discomfort, fatigue, or sleepiness during the session (we treated this as a dichotomous variable of "disutility").

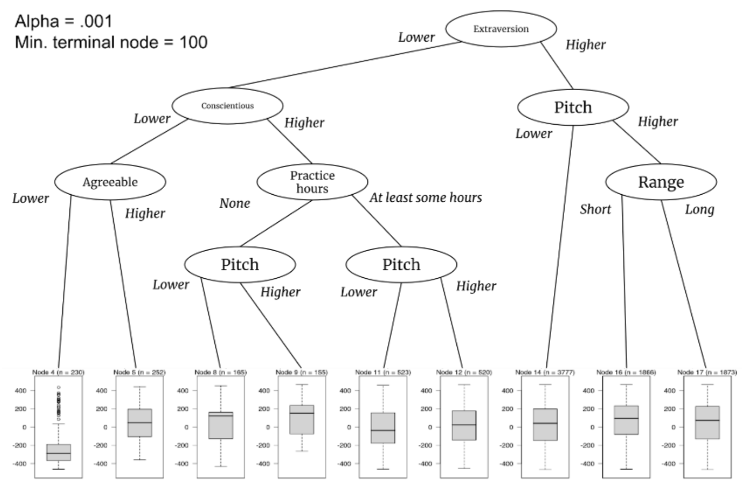


*Figure 4. Decision tree predicting preference ratings ("liking") using all available valid data.*

In interpreting the results, we focus on interactions that involve trial level stimuli features (i.e., mode, range, pitch, entropy) and individual difference variables, since other kinds of decision paths represent general tendencies for the average respondent and are therefore not helpful in exploring individual differences.

We fit models with preference ratings as the outcome and also other models with the pupil measurements as outcome, as a possible independent source of preference data. It became quickly apparent that models trying to predict pupil data resulted in extremely convoluted decision trees that defeated our abilities to explain them. Filtering the data to only cases with the most intact pupil measurement data (i.e., less missing data due to failures to detect the pupil) did not lead to better models since this also drastically reduced the sample size and thus increased the proportion of variance due to error. In other cases, none of the paths in the trees implied interactions between melody features and individual differences, and so were uninteresting as far as the objectives of this project are concerned. Therefore, all results we report here are predicting self-rated "liking" for melody stimuli.

The right-hand side of Figure 4 indicates a preference among persons with higher Extraversion scores for higher pitched versions of the stimuli (terminal node 14 vs. 16 and 17). This might be consistent with accounts of high-pitched music as more emotionally arousing taken together with arousal based understandings of Extraversion (e.g., Geen, 1984). In addition, there seems to be some discrimination between high pitched melodies with greater pitch range and those with less range, but the difference seems very subtle. For those with lower Extraversion scores (left hand of the tree), those who also had higher scores on Conscientiousness also displayed a tendency to discriminate between low- and high-pitched stimuli with a preference for the latter, although this seems further moderated by whether the person has ever engaged in intensive musical practice in the past. It is not clear however why that should be the case, and the size of the terminal nodes indicates that this is a set of trials associated with a small number of respondents.

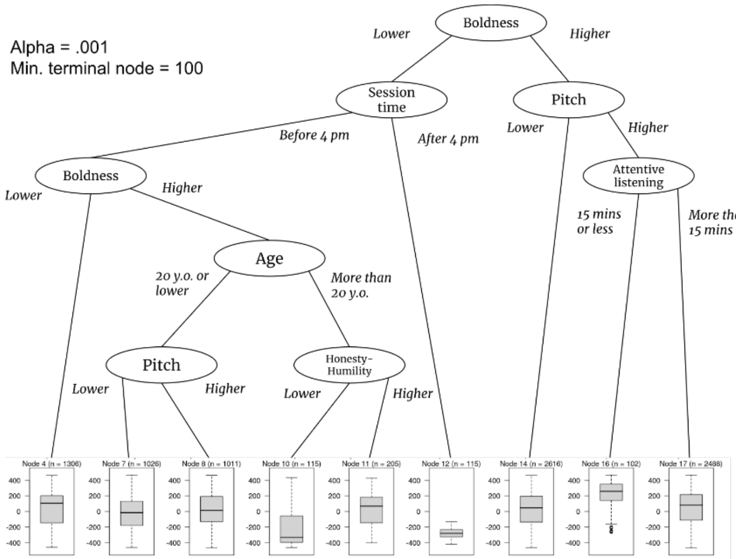


*Figure 5. Decision tree predicting preference ratings ("liking"), using data filtered on inter-item standard deviation.*

We also computed a simple index of protocol quality, inter-item standard deviation (ISD; Marjanovic et al. 2015), for the HEXACO and TPM scales to identify cases where there might be evidence of inattentive or random responding. Following Marjonovic and colleagues, in the absence of more empirically derived thresholds for our population, we used a conventional threshold of flagging the top 5% of the sample for each scale in terms of ISD. We then created a dataset that excluded any respondent where 3 or more of their personality trait scale responses exhibited ISDs in the top 5% of the sample, and this led to 4 participants being filtered out. Figure 5 shows the decision tree produced from this dataset. While it shares some elements with the tree in Figure 4 especially in terms of the role of average pitch, the traits involved are different, and it is slightly less straightforward to interpret. The left-hand side representing lower Boldness[1] scores seem to involve a non-linear interaction in that it implies that intermediate levels of Boldness (terminal nodes 7, 8, 10, and 11) lead to lower preference ratings compared to both lower or higher Boldness. Within this set of intermediate Boldness, the only interaction with a melody feature involved average pitch and age, and this seemed extremely subtle. Potentially more comprehensible is the right-hand side corresponding to higher Boldness scores where we see a preference for higher pitched stimuli, a trend even more accentuated in those who did not engage in much deliberate and attentive music listening as part of their daily life. This seems to parallel the results for Extraversion reported above, and given the conceptual overlaps between Boldness and Extraversion (i.e., in terms of being reward oriented and socially dominant) this might be an encouraging trend. However, we also see here how substantially the model changes even with just the exclusion of 4 respondents, so again we emphasize caution in drawing conclusions from these.

Beyond the two models shown above, other combinations of tree parameters and data filtering choices led to trees that we considered infeasible as far as interpretability was concerned, usually because they contained a confusing proliferation of nodes, or because there were no interactions between melody features and traits.

It is important to note that although we were not formally testing hypotheses, we did have some intuitions about what results might emerge, partly guided by the literature but also partly based on our own implicit theories of how musical experience works. We largely expected that the traits that would have the most relevance to preferences would be Neuroticism, Extraversion, and Openness, and that pitch range and key would interact with the first two traits, while complexity (i.e., entropy) would matter for the latter, and possibly also for the measure of creative achievement. At least for our data none of these expectations were borne out specifically, even though Extraversion and pitch range did interact to some degree, and instead average pitch seemed to be a melody feature with better discriminating power.

## Limitations and future directions

During the course of data gathering, the site for the measurement procedures had to be moved between at least two sites owing to scheduling difficulties. It was also the case that we had very little control over the degree of chilliness and comfort of the rooms since climate control was being handled centrally in the building. It was evident from debriefings that a common complaint among participants was the cold.

Possibly even more deleterious to participant well-being was the length of the procedures and the eye-strain that was experienced by some as a consequence of needing to keep their eyes unblinking for a few seconds at a time for each trial. In conjunction with the cold, several experienced dryness of the eyes and thus contributed to the discomfort of the situation. Aside from possibly influencing the data and the resulting models, this was clearly a failure on our part to anticipate the physical conditions of our measurement sessions. A shorter protocol, possibly achieved through a planned missing data design, would be advisable for future investigations.

In the middle of sampling, we experienced problems with the eye tracking device. More and more measurements resulted in failures to acquire pupil data until the device became unusable. This resulted in much of the later sessions not having actual pupil measurements. To keep our procedures consistent, we asked participants to still engage in looking at the fixation points as if pupil measurement was being done.

Despite trying to ensure that we had a much wider variety of stimuli than any comparable previous study that we know of, it is still a fact that for the current project we are basing inferences on variations on just 15 basic melodies, each of which was only a few seconds long. Any interpretation of our results has to be mindful that they apply very specifically to segments of music that share characteristics with our own stimuli. Generalizing to the wider universe of songs and compositions found in the actual world of real music is neither plausible nor is it the objective of this project.

It is worth reiterating that slight changes to which cases were filtered out of the sample had a tendency to greatly change the models produced. Such instability is to be expected since the sample size at the person-level was not large compared with the trial-level data, and so we caution against overinterpretation of these specific results. Instead, they should be taken as examples of what might be possible in terms of future methods and theories and these preliminary results may be used to guide the design of further exploratory work.

Going further, beyond the specific focus of this study on personality traits, other kinds of psychologically interesting individual differences could be explored as potential targets of exploration. It is quite likely that certain kinds of variation in perceptual experience might covary with melody preferences, in the same way that we would expect congenital amusia to significantly affect the experience of music.

## Conclusion

The results of our modeling attempts indicate that there are plausible links between psychological individual differences and preferences for specific melodic features within our sample of stimuli. It was perhaps slightly surprising to us that this was mainly driven by average pitch, but this seems to bear out debriefing reports from our pilot experiments where participants claimed to have a preference for the higher pitched melodies. It seems reasonable to suppose that differential preferences regarding simple melodic stimuli are driven partly by variation in previous experiences with musical works and also by more basic perceptual sensitivities for especially meaningful sound profiles. For example, it is possible that perceived emotion in music parallels the detection of emotion in human voices (Juslin and Laukka, 2003) such as the association of high pitch with the perception of positive affective content (Friedman et al., 2018). Since Extraverts are thought to be sensitive to potentially positive stimuli because they might signal reward (Corr, 2004), they might also have a stronger preference for higher pitched melodies (while low Extraversion persons would possibly be indifferent). Still, while the personality traits observed to be related to pitch preference seems to make some theoretical sense, it is best to refrain from overinterpreting this for now. Nevertheless, we believe that our demonstration should provide ample motivation for continuing this line of investigation, given the possible theoretical and practical outcomes that might be produced.

## Notes

[1]It is also worth pointing out terminal node 12. Although this represents a small minority of trials, it might warn researchers off from conducting similar procedures in the evening.

## References

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125 (Pt 2), 238–251.

Barron, F., & Welsh, G. (1952). Artistic perception as a possible factor in personality style: Its measurement by a figure-preference test. *Journal of Psychology, 33,* 199-203.

Basak, J., & Krishnapuram, R. (2005). Interpretable hierarchical clustering by constructing an unsupervised decision tree. *IEEE transactions on knowledge and data engineering*, *17*(1), 121-132.

Bianco, R., Gold, B. P., Johnson, A. P., & Penhune, V. B. (2019). Music predictability and liking enhance pupil dilation and promote motor learning in non-musicians. *Scientific reports, 9(1),* 17060.

Blagov, P. S., Patrick, C. J., Oost, K. M., Goodman, J. A., & Pugh, A. T. (2016). Triarchic psychopathy measure: Validity in relation to normal-range traits, personality pathology, and psychological adjustment. *Journal of Personality Disorders*, 1-11.

Borchers, H. W. (2023). Package 'pracma'. Retrieved from https://cran.r-project.org/web/packages/pracma/pracma.pdf..

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal, 17*(1), 37-50.

Cattell, R. B. (1956). Validation and intensification of the Sixteen Personality Factor Questionnaire. *Journal of Clinical Psychology, 12,* 205–214.

Cattell, R. B., & Saunders, D. R. (1954). Musical preferences and personality diagnosis: I. A factorization of one hundred and twenty themes. *The Journal of Social Psychology*, 39 (1), 3-24.

Chamorro-Premuzic, T., Fagan, P, & Furnham, A. (2010). Personality and uses of music as predictors of preferences for music consensually classified as happy, sad, complex, and social. *Psychology of Aesthetics, Creativity, and the Arts, 4,* 205-213.

Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience & Biobehavioral Reviews, 28(3),* 317-332.

Dobrota, S., & Reic Ercegovac, I. (2014). Student's musical preferences: The role of music education, characteristics of music and personality traits. *Croatian Journal of Education*, *16(2)*, 363-384.

Dunn, P. G., de Ruyter, B., & Bouwhuis, D. G. (2011). Toward a better understanding of the relation between music preference, listening behavior, and personality. *Psychology of Music*, *40(4)*, 411-428

Ercegovac, I. R., Dobrota, S., & Kuščević, D. (2015). Relationship between music and visual art preferences and some personality traits. *Empirical Studies of the Arts,* 33(2), 207-227.

Fokkema, M., Edbrooke-Childs, J., & Wolpert, M. (2021). Generalized linear mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. *Psychotherapy Research*, 31(3), 329-341.

Friedman, R. S., Neill, W. T., Seror III, G. A., & Kleinsmith, A. L. (2018). Average pitch height and perceived emotional expression within an unconventional tuning system. *Music Perception: An Interdisciplinary Journal, 35(4),* 518-523.

Geen, R. G. (1984). Preferred stimulation levels in introverts and extraverts: Effects on arousal and performance. *Journal of Personality and Social Psychology, 46,* 1303-1312.

Gingras, B., Marin, M. M., Puig-Waldmüller, E., & Fitch, W. T. (2015). The eye is listening: Music-induced arousal and individual differences predict pupillary responses. *Frontiers in Human Neuroscience*, *9*, 1-12.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84-96.

Higgins, E. T., Scholer, A. A. (2008). When is personality revealed?: A motivated cognition approach. John, O. P. (Ed); Robins, R. W. (Ed); Pervin, L. A. (Ed). *Handbook of personality: Theory and Research* (3rd ed.), (pp. 182-207). New York, NY, US: Guilford Press.

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods, 1,* 131-163.

John, O. P., Naumann, L. P., & Soto, C. J. (2010). Paradigm shift to the integrative Big Five trait taxonomy. In Oliver P. John, Richard W. Robinson, & Lawrence A. Pervin (Ed.), *Handbook of Personality: Theory and Research.* (114-158). New York, NY: Guilford Press.

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.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802 - 5805.

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts*, *6(*2), 146–154..

Laeng, B., Eidet, L. M., Sulutvedt, U., & Panksepp, J. (2016). Music chills: The eye pupil as a mirror to music's soul. *Consciousness and Cognition, 44,* 161-178.

Lee, K. & Ashton, M.C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, 39 (2), 329-358.

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences, 84*, 79-83.

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer Nature.

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. J. (2013). *Goldsmiths Musical Sophistication Index (Gold-MSI) V1. 0: Technical Report and Documentation Revision 0.3*. London: Goldsmiths, University of London.

Murray, H. A. (1938). *Explorations in Personality*. New York, NY: Oxford University Press.

Ortigosa, A., Carro, R. M., & Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. *Journal of Computer and System Sciences, 80(1)*, 57-71.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*(1), 58-75.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., . . . Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology, 107*(4), 677-718.

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: a five-factor model. *Journal of Personality and Social Psychology, 100(6)*, 1139–1157.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84(6),* 1236-1256.

Revelle, W. (2015). Constructs, components, and factor models. *An Introduction to Psychometric Theory with Applications in R. http://personality-project. org/r/book (accessed October 24, 2024).*

Revelle, W. (2020). Package 'psych. Retrieved from https://cran.r-project.org/web/packages/psych/psych.pdf.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS One, 8(9)*, 1–16.

Silvia, P. J. (2005). Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology, 9*(4), 342-357.

Venkatasubramaniam, A., Wolfson, J., Mitchell, N., Barnes, T., JaKa, M., & French, S. (2017). Decision trees in epidemiological research. *Emerging Themes in Epidemiology, 14*, 1-12.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer--based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 10361040.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form.* Ames: The University of Iowa.

Wiggins, J. S. (2003). *Paradigms of Personality Assessment*. NY: The Guilford Press.

**ADRIANNE JOHN R. GALANG** completed his Ph.D. training in personality and social psychology at the University of the Philippines Diliman. After having held teaching positions in both UP Diliman and De La Salle University, he pursued a series of postdoctoral fellowships, first at Nanyang Technological University Singapore, and then at Masaryk University in the Czech Republic, before returning to take up a faculty position at UP under its Balik Ph.D. program. He was the recipient of the 2016 Dr. Sybil Eysenck Award for Young Researchers from the International Society for the Study of Individual Differences (ISSID). AJ mainly works in quantitative psychology, with expertise in psychometrics and behavioral experiment design. His research and publications have ranged over topics such as creativity, risk taking, theoretical psychology, metascience, and personality measurement.

**MONICA RENEE G. POLICARPIO** is a psychological researcher and educator from De La Salle University. In completing her M.S. degree in the same institution, she explored intertemporal decision making and individual differences in anticipatory pleasure. Her primary interests include self-regulation, reasoning, judgment, and decision making. Her publications and projects are about self-control, skin color bias, cognitive biases, moral reasoning, and the non-religious. Monica is currently advancing her expertise through a Ph.D. program in Educational Psychology, with a focus on metacognitive accuracy and its implications for decision-making in learning.

**APRIL MERYLL XU** graduated with a bachelor degree in computer science specializing in software technology from De La Salle University.

**MA. JENINA NALIPAY is** currently an Assistant Professor at the Department of Counselling and Psychology, Hong Kong Shue Yan University. She has doctorate degrees in Clinical Psychology from the University of Santo Tomas and Education, with specialization in Curriculum and Instruction, from the Chinese University of Hong Kong. Her research interests include teacher and student mental health and motivation.

Dr. **RAFAEL CABREDO** earned his PhD in Information Science and Technology from the University of Osaka, where his research focused on music emotion analysis, physiological sensors, and machine learning techniques. He currently serves as a faculty member at De La Salle University, where he has held various administrative roles. Dr. Cabredo's research interests primarily fall within the areas of Intelligent Systems and Multimedia Information Retrieval. His work includes projects on process automation using artificial intelligence, music and audio information extraction and analysis, as well as game design and development.

**Acknowledgments:** *The project was initiated through an Interdisciplinary Research Grant from De La Salle University that was awarded to the first author. We would like to recognize the valuable assistance of Desilu Finix Adarme, Carlo Gabriel Flores, Alyssa Marie Dar Juan, and Temple Zane Barrios, who helped in the data gathering and data processing procedures even under very challenging conditions. We thank German De Ramos for indulging our untutored questions about music theory. Among the colleagues and scholars who supplied comments on our early work, we are especially grateful to Ma. Regina Justina Estuar and Gregorio Del Pilar, as well as the two anonymous reviewers who lent their time and expertise to improving our paper.*