

Major-minoriness in tonal music: Evaluation of relative mode estimation using expert ratings and audio-based key-finding principles

Psychology of Music

1–20

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03057356251326065

journals.sagepub.com/home/pomTuomas Eerola¹  and Michael Schutz²

Abstract

Mode is a foundational concept of Western music, serving as the basis for chords and harmonies, detecting and assessing cadences and form, and conveying musical emotion. Traditionally treated categorically, here we build upon recent work exploring this crucial musical construct on a continuum, an approach we refer to as 'relative mode'. Specifically, we formulate and evaluate a computational model calculating this property from either symbolic or audio representations of music by adapting common key-finding techniques traditionally used to identify mode categorically. Here, we use them to infer the relative mode based on differences between the potential strength of major and minor key candidates. The model evaluation is based on a corpus of excerpts from Preludes by Bach, Chopin, and Shostakovich previously assessed by expert music analysts. Our results suggest that the model (using only audio files) is able to predict relative mode to a degree closely aligning with experts (using both audio and notated scores). A pragmatic set of parameters for the model is identified, and the shortcomings and the applicability of the model to other eras and genres are discussed.

Keywords

mode, key, algorithm, tonal, Western music

Modality is a foundational construct in Western music theory, where it is core to tasks ranging from identifying chords to detecting cadences and assessing form. In addition, it plays a crucial role in music perception, particularly with respect to emotion. As mode (major/minor) forms a crucial aspect of key (C Major, c minor), scholarship on key-finding provides a useful point of departure. In that literature, automated approaches to key-finding (such as the Krumhansl–Schmuckler algorithm, Frankland & Cohen, 1996; Vos & Van Geenen, 1996) typically assess a

¹Durham University, Durham, UK²McMaster University, Hamilton, ON, Canada

Corresponding author:

Tuomas Eerola, Department of Music, Durham University, Palace Green, DH1 3RL, Durham, UK.

Email: tuomas.eerola@durham.ac.uk

passage's pitch-class distribution against key profiles derived from probe-tone experiments (Krumhansl, 1990; Krumhansl & Kessler, 1982).

Refinements include revising the profiles (Temperley, 1999), altering calculation principles (Shmulevich & Yli-Harja, 2000; Temperley, 2002), re-weighting elements (Schmuckler & Tomovski, 2005), accounting for nuances in minor keys (Vuvan et al., 2011), exploring time-varying approaches (Toivianen & Krumhansl, 2003) and optimising key-finding by combining different approaches (Albrecht & Shanahan, 2013). Moving beyond key-finding algorithms requiring symbolic representations, the next tranche of improvements utilise audio, typically by applying either templates akin to key profiles (Izmirli, 2005) or mathematical models to chromagrams – decompositions of energies across the spectrum into pitch-classes (Chuan & Chew, 2005, 2006, 2014). Others incorporate temporal contingencies between the detected local keys (Nápoles López et al., 2019).

Although key-finding itself is largely resolved (>90% correct in global key-finding results, Nápoles López et al., 2019), our intention here is not to *identify keys* but to estimate *mode on a continuum* from minor to major – a concept we call 'relative mode'. Treating mode as a continuum (rather than a discrete category) offers a useful degree of nuance in analysing musical structure. Among other benefits, we see value in enhancing our understanding of its effects on perception (such as predicting emotional responses). The ability to compute this from digital audio offers an invaluable tool for a wide range of computational, perceptual, and analytical tasks.

Modality: Discrete or continuous?

Our thinking regarding mode as a continuous versus categorical construct brings to mind research on categorisation, which distinguishes between 'natural' and 'artificial' categories (Prentice & Miller, 2007; Rosch et al., 1976). Applying that framework to musical analysis, Zbikowski (2002) uses the terms Type 1 (implicit/naturally arising) and Type 2 (explicit identification of features). That framework is helpful in thinking about modality – a term that can be either a Type 1 or Type 2 category depending upon context. For example, when applied to *chords and scales*, modality functions as a Type 2 category, with clear definitions regarding the distance between root and third of a chord (i.e., four semitones for major; three for minor). However, as a descriptor of *musical passages*, modality functions more like a Type 1 category with fuzzy boundaries (e.g., some minor key passages end on major chords; some major passages 'borrow' harmonies from parallel minor keys). Our position is that when discussing passages of music, mode functions more like the Type 1 category, with some passages exhibiting a larger collection of major (or minor) features than others.

Although to the best of our knowledge modality has not previously been discussed explicitly as a continuous construct, the renowned composer Arnold Schoenberg notes the Lydian and Mixolydian modes are 'major-like' (Schoenberg & Stein, 1969), as they are, in a sense, 'variations' on a major scale (with raised fourth and lowered seventh scale degree, respectively). Schoenberg similarly referred to the Dorian, and Phrygian modes as 'minor-like', as their structures are 'variations' on the natural minor scale (with a lowered second, and an ascending harmonic minor scale, respectively). Although these modes originate from periods of music history prior to modern conceptions of tonality, as 'our twenty-first century ears are accustomed to major and minor scales, we sometimes hear the modes as alterations of these more familiar scales' (Clendinning & Marvin, 2016; for historical functions of keys, see Long, 2020). Assessing this intuition experimentally, Temperley and Tan (2012) found ratings of happiness – which here could be taken as an index of valence – identify the Ionian (major) scale as the most happy and the Aeolian as one of the least happy (with Phrygian even 'sadder' than

Aeolian). Temperley (2007) also treats some aspects of tonality as a continuum – for example, ‘tonalness’ as the degree to which passages reflect the common-practice tonality; ‘tonal clarity’ as a ratio between the most probable and the second most probable key. Although these concepts are mode-agnostic, they treat key information as a continuous (vs. discrete) category. Similarly, White (2012) proposed a key-finding algorithm that expands into temporal analysis through interpolation of local profiles over time, which is also an interesting and relevant related process (see also Quinn, 2010 for a related localised key-finding algorithm). These insights are reminders that Type 2 categorisations of modality (i.e., clear-cut binary classification of major and minor) are just one way of considering the pitch structures – albeit one with implications for conceptual coherence for Western common period practice music.

Key-finding approaches

The traditional categorical treatment of modality guides contemporary approaches to key-finding, which essentially translate a pool of continuous measures to a singular identifier. Here, we adapt that approach to generating automated predictions of relative mode using the conceptual framework employed by *MIR toolbox* (Lartillot et al., 2008). *MIR toolbox* is a MATLAB toolkit for extracting musical features from audio, which has been used to build models of emotion and mood recognition (Lin et al., 2011), inform music recommendation systems such as Spotify (Vasu & Choudhary, 2022), and identify the relative ‘ground truth’ for musical features crucial for emotion expression (Beveridge & Knox, 2009). Although it is the most commonly used tool for placing stimuli along the major/minor spectrum, the quality of these evaluations has not been formally assessed and general explorations have documented issues with the *MIR toolbox*’s consistency and reliability (Kumar et al., 2015; Zhou et al., 2023). We do note, however, recent work from one of us exploring this issue suggests Librosa (McFee et al., 2015) to be a more suitable tool for chromagram-based tasks (Swierczek & Schutz, 2025).

For many applications (including the vast majority related to musical emotion), identification of the chroma component of the nominal key (i.e., the ‘D’ in ‘D minor’) is less important than the modality component (i.e., the ‘minor’ in ‘D minor’). In other words, knowing that a passage is ‘major’ is more important than whether it is in A Major versus D Major (the widespread adoption of equal temperament has essentially removed differences in interval sizes between major keys). For these reasons, tools reliably quantifying the relative modality of music from audio files would be valuable for both the research and applied communities.

Aims and approach

Our goal in this series of analyses is to explore the challenges, possibilities, and opportunities for using a continuous treatment of mode derived computationally from audio files. Given the complexity of mode in music created outside scientific experiments (Battcock & Schutz, 2019), the ability to automatically compute relative mode would aid efforts to better predict listener responses to audio files. It would also provide a perceptually informed set of tools for music theorists, music information retrieval specialists, and music cognition scholars to better understand and more accurately quantify a crucial aspect of musical structure.

Any exploration of this nature requires credible ‘ground truth’ clarifying the theoretically correct values. Our approach draws upon a data set consisting of relative mode ratings from a team of five music theorists, who each analysed eight measure excerpts from sets of preludes composed between 1610 and 1988 (Delle Grazie et al., 2025). That study offers expert ratings of all 72 pieces analysed in this manuscript: preludes from Bach’s *Well Tempered Clavier* (Book 1), Chopin’s *24 Preludes*, and Shostakovich’s *24 Preludes* (Op 34).

Here, we use those expert ratings to assess the quality of relative mode evaluations computed algorithmically. Building on a large corpus of materials used in recent studies of specialised 24 Prelude sets (Delle Grazie et al, 2025) as well as expansions of those materials specifically for this project, we conducted three experiments. The first computes relative mode from symbolic representation (MIDI) for all 72 preludes, using this as a theoretical upper limit for evaluating the effectiveness of various key-finding profiles and means of assessment. The second assesses parameters crucial to the chromagram interpretation (required when moving to digital audio files), using one example of each of the 72 pieces from our full corpus. The third evaluates the effectiveness of our approach using the full corpus containing 18 recordings of Bach's *Well Tempered Clavier* (Book 1), 14 of Chopin's *Preludes* (Op 28), and 10 of Shostakovich's *Preludes* (Op 34), computing relative mode from 42 recordings of these three sets of preludes (1,008 pieces in total; detailed in Appendix 4).

These analyses hold three benefits for the research community. First, they provide a proof-of-concept for the ability to quantify relative modality from audio files, which promises important applications for predicting emotional responses to music – particularly for passages that are not notated. Second, they offer an exploration of different approaches to automated mode identification by comparing competing models (stage one), assessing differences between symbolic representation and actual audio recordings (stage two), and evaluating the robustness of this approach with respect to performer interpretation and/or recording parameters (stage three). Third, they document a procedure for quantifying relative modality (i.e., treating mode as a continuous rather than discrete variable). Together, they lay the groundwork for future analysis of musical emotion that is not reliant on symbolic representation – which is important as many musical cultures do not use the score-based notation system common in the Western classical tradition.

Relative mode estimation model

Our proposal for the *relative mode estimation* (RME) model is based upon common key-finding algorithms, which generally have three steps: (1) transforming of spectrum (audio) or note events (MIDI) into pitch-class distributions (PCDs), (2) summing the energies across analysis windows for the pitch-class distributions, and (3) calculating similarity between empirical key profiles and PCDs. Here, we also propose a fourth step, (4) calculating RME by taking the difference between the most likely major and minor key profiles. The schematic of the full process is shown in Figure 1.

Technical details

The first three steps are well known in the literature. The first (pitch-class extraction) is done simply by tallying the notes from scores in the case of symbolic data (e.g., Krumhansl, 1990) or extracting the chroma (i.e., pitch-classes) in the case of digital audio (Gómez, 2006). There are several techniques for extracting this cue from audio, such as the *Constant Q Transform* or CQT (Brown, 1991) and enhancements to the CQT (Müller & Ewert, 2011), as well as constant Q harmonic coefficients (Rafii, 2022), discussed in detail later.

The second step compares PCDs from the first with reference key profiles derived from a series of empirical experiments (Krumhansl, 1990). Numerous studies have developed alternative key profiles, some arising from the analysis of music and diagnostic operations involving key finding (Aarden, 2003; Albrecht & Shanahan, 2013; Bellmann, 2005; Sapp, 2011; Temperley, 2007). We will evaluate these alternative profiles later to assess their utility in capturing relative mode.

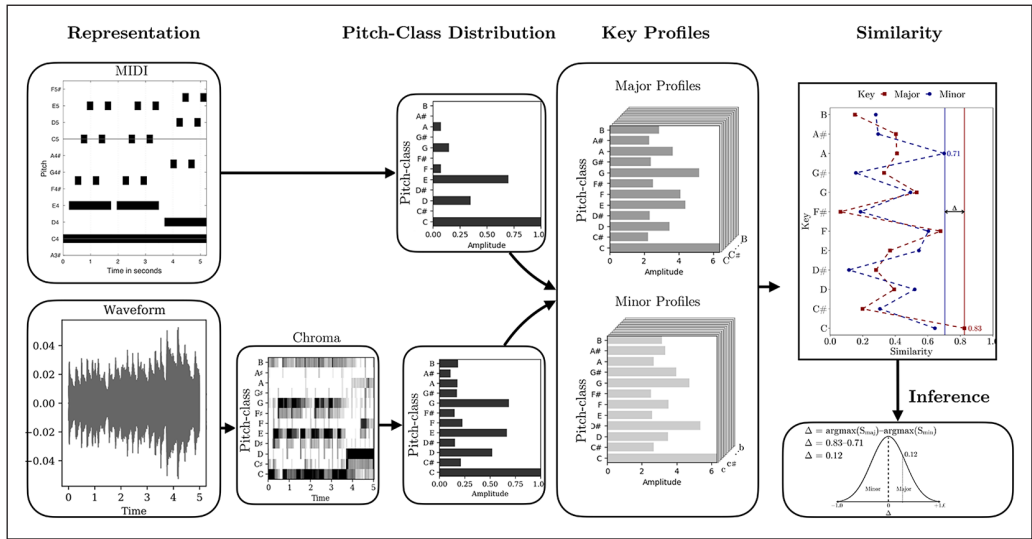


Figure 1. Overall Stages of Relative Mode estimation (RME) Involving (1) the Extraction of PCDs, (2) Comparison of PCDs to Key Profiles, (3) Calculation of Similarity Between PCDs and Key Profiles, and (4) Drawing of An Inference From the Difference Between Maximum Major and Minor Key Similarities. The Excerpt Represents the Opening Bar and Half From C Major Prelude From WTC I by J. S. Bach.

The third step compares PCDs and key profiles using a similarity measure. The Pearson correlation coefficient has been used most frequently and originally as this metric (Krumhansl, 1990), but other similarity metrics based on Cosine and Euclidean distances have also been proposed (Albrecht & Shanahan, 2013; Temperley, 2007). We will formally evaluate these in subsequent experiments.

The fourth and final step compares the strength of candidate keys in major and minor, using the difference (Δ) between the maximum major key strength (S_{maj}) and the maximum minor key strength (S_{min}). Here, we define key strength as the highest similarity of the major and minor 12 keys, calculating this according to:

$$\Delta = \max(S_{\text{maj}}) - \max(S_{\text{min}}).$$

This measure will be positive if the strongest major key has a higher maximum than the strongest minor key and vice versa. Although this approach to RME can in theory apply to passages of any length, we envisage it to be most useful when it is applied within a relatively short window (e.g., 3 s). This windowed analysis is sensitive to the fluctuating nature of relative mode as music unfolds in time. This prevents passages that modulate from clearly major to clearly minor from yielding RME values suggesting a lack of clear tonality.

Demonstrating our approach

To exemplify this concept, the same example excerpt shown in Figure 1 (Bach C Major Prelude from WTC I) is processed through two variants of the RME (explained in detail later) using a one-bar window of analysis at a time (Figure 2). The first (darker line) utilises so-called *Simple* key profiles (Sapp, 2011), and the second line represents key profiles established by Albrecht and Shanahan (2013). The figure illustrates the score (Panel A), the pitch-class

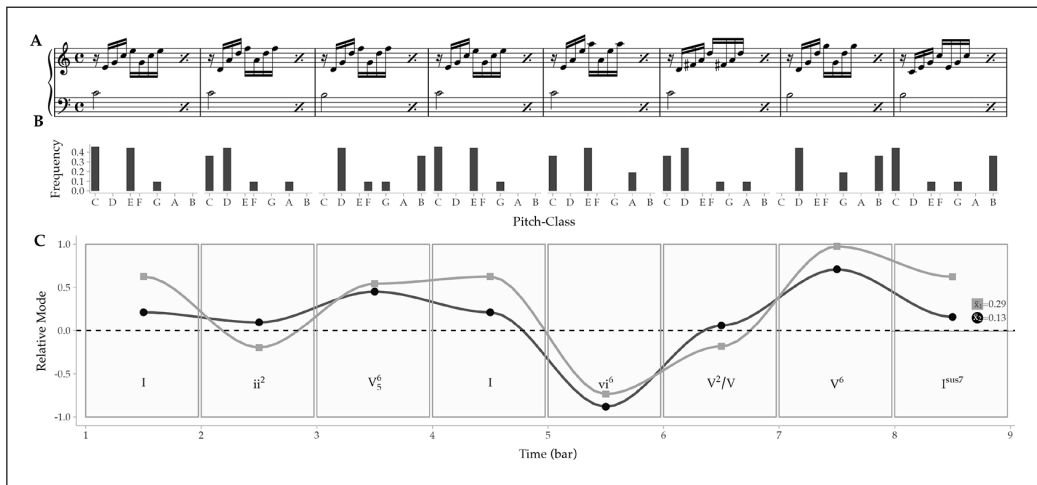


Figure 2. The First 8 Bars of C Major Prelude by Bach (Panel A) With Pitch-Class Extraction (Panel B) and Relative Mode Estimated Using Windowed Analysis (Panel C).

The two lines in panel C show the output of RME using two key profiles, *Simple* (grey line) and *Albrecht* (black line), both using the Cosine similarity metric. For clarity in this example, the analysis window size (illustrated with rectangles) is set to four beats. A conventional harmonic analysis has been added as an annotation, and the final bar contains the mean values from both relative mode estimations (0.13 and 0.29 for *Albrecht* and *Simple*, respectively).

distribution within the window (Panel B), and the result of the RME for each window (markers) with an integration of the relative mode over time (lines) (Panel C). The lowest panel also shows the harmonic analysis (roman numerals) and the means of the RMEs across the excerpt in the final bar. For the simple key profile, this mean is 0.13 and 0.29 for the process relying on Albrecht's variant key profiles.

Both versions of the RME clearly suggest that the excerpt is in major, but two bars (second, fifth) flirt with minor. The central difference between these two key profiles is that the Simple profile does not quite fit with the straightforward C major profile of Bars 1 and 4, as the profile expects dominant more than the third (we will explore the impact of different key profiles in detail later).

Reliable information on relative modality is central to assessing the accuracy of our approach. For this, we used a novel data set developed through a collaboration between MS and a team of music analysts. Their data consist of evaluations from five expert musicians in a three-phase procedure inspired by widely recognised protocols for minimising error in evaluation of complex constructs (Kahneman et al., 2021). As full discussion of that process and data set is forthcoming (Delle Grazie et al., 2025), here we focus on summarising only those aspects crucial to understanding the data we use here as ground truth for assessing the accuracy of computational estimates of mode.

To arrive at continuous estimates of modality, Delle Grazie et al. asked five raters to independently evaluate 'relative mode' for the first eight measures (corresponding to a normative musical phrase) of the same preludes by Bach, Chopin, and Shostakovich used in our study (as well as numerous other prelude sets). Each rater received packets containing the notation for each excerpt (with identifications of composer removed) consisting of excerpts presented in an order randomised uniquely for each rater. Raters also had access to performances of these excerpts taken from commercially available audio used in numerous perceptual experiments (Anderson & Schutz, 2022; Delle Grazie et al., in press; Kelly et al., 2021), performed by Pietro De Maria (2015) for Bach, Vladimir Ashkenazy (1993) for Chopin, and Konstantin Scherbakov (2003) for Shostakovich.

In Phase 1, each rater independently evaluated excerpts on a scale from 1 (*entirely minor*) to 7 (*entirely major*) while noting key structural features. In Phase 2, raters reviewed each other's notes and evaluations for a selection of 12–20 excerpts, before discussing as a group and being given the option of confidentially updating their ratings. In Phase 3, raters independently reviewed personalised packets grouped by rating (e.g., each received a packet of excerpts they had evaluated as '1', '2', and '3', etc.) with a final opportunity for adjustment. Although this method resulted in well-explained and meaningful differences of opinion for some pieces, it nonetheless led to strong agreement. Delle Grazie et al. found an average correlation of .90 for the 72 pieces used here in our study, with some differences between the three composers in question ($r_{Bach} = .89$, $r_{Chopin} = .95$, $r_{Shostakovich} = .84$). For reasons articulated at length in their manuscript, they believe these correlations likely represent the maximum level of agreement achievable by expert assessors.

Experiment I: Assessing model parameters using symbolic data

Our first experiment chose the best reference key profile and similarity measure. To do so, we built pitch-class distributions from symbolic notation as a reference point. Specifically, we used the first eight measures of the 72 preludes, extracting pitch distributions using the MIDI toolbox (Eerola & Toivainen, 2003), employing exponential weighting of the note durations proposed by Parncutt's (1994) durational accent model, adjusting durations to be perceptually plausible. We minimised the issues of note durations and modulations by using a 3 s, non-overlapped window (which we believe is useful in capturing nuance, although for simplicity in this proof-of-concept, here, we later averaged these running windows to yield one RME value per piece). We applied six variations of key profiles and three variations of similarity metrics – 18 approaches in total. This included six variations of calculating PCDs: the classic empirical key profiles from probe-tone experiments by Krumhansl and Kessler (1982), the empirical key profiles extracted from the Essen corpus by Aarden (2003), key profiles established by Bellmann (2005), the variant profiles proposed by Temperley (2007), the simple weightings offered by Sapp (2011), and altered key profiles proposed by Albrecht and Shanahan (2013). We compared the PCDs to each of these six types of profiles using three similarity metrics (Pearson, Cosine, and Euclidean). We used MATLAB for PCD extraction and R for all other steps.

Table 1 shows the correlations between the expert ratings and model outputs using different key profiles and similarity metrics. The key profiles and correlation metrics generally perform comparably, delivering correlations between .820 and .850 with only minor differences between them (although we note lower performance for Krumhansl and Aarden profiles, and slightly better performance using the Cosine metric for all of them). It is worth noting that the simple key profile performs at least at the same level as many of the complex and optimised key profiles. Neither window length nor method of collapsing a series of values (e.g., mean vs median) affected performance.

Figure 3 plots the output of our RME model against expert ratings (slightly jittered to avoid overlaps of excerpts with the same means). The plot suggests a linear relationship between the expert ratings and model predictions, with some differences across the composers: excerpts by Bach and Chopin tend to be further away from the centre barring some notable exceptions (e.g., Bach F-major Prelude). Excerpts by Shostakovich tend to be more evenly spaced across the rated dimension, which is also reflected in the model. The correlations for each composer range from .832 for Bach and .868 for Chopin to .791 for Shostakovich with the overall being $r = .835$.

These results suggest that the RME can predict the expert ratings reasonably well in this small sample ($N = 72$). To contextualise, the obtained correlation coefficient ($r = .84$) is only slightly lower than the correlation between raters, $r = .90$. In terms of variance predicted,

Table 1. Correlations Between Expert Ratings and Continuous Mode Estimation Model for 72 MIDI Encoded Preludes From Bach (Book I), Chopin (Book I), and Shostakovich (Book I).

Key profile	Pearson	Cosine	Euclidean
Krumhansl	.692	.709	.711
Aarden	.755	.754	.747
Temperley	.820	.827	.824
Bellmann	.844	.850	.846
Simple	.833	.840	.830
Albrecht	.840	.842	.835
Mean	.797	.804	.799

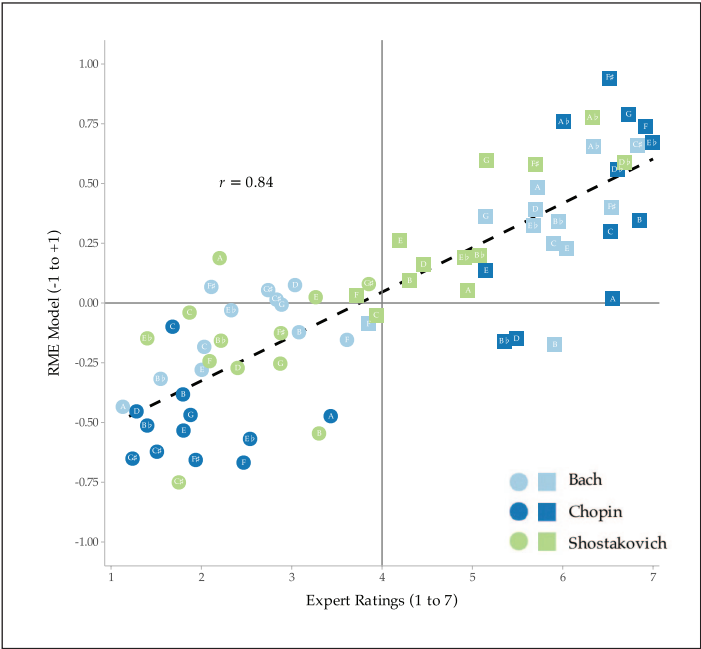


Figure 3. Expert Ratings Versus Model Predictions Across the 24 Preludes From Three Composers (N=72) Using ‘Simple’ Key Profile and Cosine Similarity Metric.

the model achieves approximately 70%, although this is only suggestive of the model prediction rate as this is a small sample and no formal cross-validation was built into the evaluation.

Experiment 2: Assessing model parameters using a small sample of audio data

The first experiment used pitch-class distributions derived from symbolic notation (MIDI representation). The second assessed the efficacy of that conceptual approach applied to audio. In addition to key profiles and similarity metrics, this experiment considers the impact of different chroma extraction techniques used in extracting the PCDs from audio.

Given the large model parameter space (the acoustic analysis options, the wrapping of the spectrum, the role of the key profiles, and similarity metrics), we first ran iterations of the analyses on a small subset of the materials. We chose one recording from each prelude and composer (72 total) in a pseudo-randomised fashion (specific choices appear in Appendix 4). This set included mostly piano renditions recorded between 1934 and 2020, as well as seven recordings of the Bach preludes on harpsichord (Galling, Hamilton, Kirkpatrick, Landowska, Leonhardt, and two by Newman). Together, these offer a range of variations regarding instruments, tuning, audio quality, and performance characteristics (tempo, dynamics, timing variation). Altogether we drew upon recordings from 42 different albums by 36 different artists (for a full list, see Appendices 1 to 3). We analysed 72 of these 1,008 recordings, assessing different spectral decomposition parameters, key profiles, and similarity metrics on an RME model implemented in Python using the *Librosa* library (McFee et al., 2015). We used set parameters (threshold: 0, octaves:7, bins/octave:36 bins, hop length:8,192, f0 minimum: 65.4Hz) for three variant extraction techniques. These included constant Q transform (CQT, see Schörkhuber & Klapuri, 2010), constant energy normalised variant (CENS, Müller & Ewert, 2011), and constant Q harmonic coefficients (CQHC, Rafii, 2022) (more details at https://github.com/tuomaseerola/relative_mode). Correlations between our resulting RME and expert ratings appear in Table 2.

Table 2 suggests RME calculations from audio are only slightly inferior to those from symbolic data (Experiment 1): .018 (Pearson), .013 (Cosine), and .018 (Euclidean) in the averaged correlations between the model and expert ratings when averaged over the six key profiles. Consistent with Experiment 1, we found improvements on the classic empirical profile by Krumhansl in all newer variants including the simple key profile.

In terms of the chroma extraction, CQT leads to the best results (max $r = .859$, mean $r = .786$ shown in Table 2), and the variant extraction techniques led to inferior overall predictions (CENS $r_{max} = .842$ and $r_{Mean} = .782$, and CQHC $r_{max} = .824$ and $r_{Mean} = .741$). As the difference between CQT and CENS is marginal (.786 vs .782), we speculate the newer extraction techniques may offer improvements only when the sonic materials (timbre) are more variable. There are minor variations in the model correlations across the subsets of the composers, with Bach generally close to the overall mean ($r = .811$), and Chopin receiving higher correlations ($r = .881$) and Shostakovich lower ($r = .775$).

Although numerous combinations of parameters are worthy of exploration, our work thus far succeeds in identifying the parameters consistently performing well in this small data set. Therefore, it appears RME can be reliably discerned from audio (at least in this repertoire).

Experiment 3: Model performance in a large sample of audio data

In our final experiment, we utilise all 1,008 examples (18×24 for Bach, 14×24 for Chopin, 10×24 for Shostakovich) recorded between 1934 and 2020. Several Bach recordings utilise harpsichord rather than piano (details appear in Appendices 1 to 3). This diverse range of recordings allows for the exploration of audio-based RME computation across different approaches to tempo, dynamics, instrument, expression, and recording quality. As in Experiment 2, we utilise the RME model that capitalises simple key profile, CQT chroma extraction (with tuning calibration and the Cosine similarity metric), applying the model to all recordings.

Figure 4 displays the mean correlations and the confidence intervals (derived from bootstrapping with 1,000 resamples), along with the practical ceiling based on agreement among expert raters. At first glance, it appears that older recordings may suffer from small artefacts (e.g., tape hiss, narrow bandwidth, unusual tuning reference) harming performance within Bach. However, upon closer investigation, we did not find any recording characteristics (tempo, timbre, dynamics, compression, instruments, microphone positions) consistently affecting performance.

Table 2. Correlations Between Expert Ratings and Continuous Model for 72 Recorded Performances of Preludes From Bach (Book I), Chopin (Book I), and Shostakovich (Book I).

Key profile	Pearson	Cosine	Euclidean
Krumhansl	.648	.636	.624
Aarden	.708	.716	.699
Temperley	.850	.854	.847
Bellmann	.833	.838	.832
Simple	.846	.859	.853
Albrecht	.828	.838	.832

To identify the consistent noise factors in the RME analysis, we extracted dynamics, several timbral descriptors (brightness, spectral centroid, spectral flux, RMS amplitude, roughness) and tempo descriptors for each excerpt and added these as additional predictors to the regression with RME model predicting the expert ratings. However, no single audio descriptor could contribute significantly (more than 1%–2% of variance accounted) to the model that already had a highly successful predictor (RME) within it. A more extensive analysis of the potential additional considerations would benefit from a larger set of materials and from systematic alterations of the most plausible variations of these factors. However, we consider that best left for future research.

Our model appears fairly robust across *general* variations in the acoustic and performance qualities of recordings; however, performance clearly varies by *specific* recording (Figure 4). For example, Shostakovich preludes recorded by Jascha Nemtsov (2009) have the lowest correlations with expert ratings and one of the largest confidence intervals. Further inspection suggests our model performs worse on short excerpts, which most often occur in Shostakovich (e.g., eight bars in fast tempo for E major prelude performed by Nemtsov lasts 6 s, F minor for 7 s). Altering the length of the window (from 1.5 to 3 s) and the overlap of the window (from 25% to 75%) mitigates the issue for the shortest examples but do not bring significant improvements to the data set as a whole. To keep the overall architecture of this model parsimonious, we do not propose changes to the model at this stage but signpost this as an area for future improvement.

Figure 5 summarises the success of the model across individual performances for each prelude. The vast majority (96% of 1,008 recordings) of the 72 preludes fall within the least square line of fit when we look at the distribution of RME values (the thick lines denoting QI range). For Bach’s preludes, the exception is B major which is rated as moderately major by the experts but moderately minor by the model. For Chopin, the linear fit captures the majority of the Preludes with notable exceptions for C and the E minor preludes (which experts consider minor but the model evaluates as borderline).

It is also worth highlighting that expert ratings of Bach and Chopin are clearly bifurcated but more evenly spread on a continuum for Shostakovich. The lack of excerpts falling into the middle of the continuum for Bach and Chopin may artificially increase the linear correlation. However, Shostakovich’s preludes covering the full distribution of RME ratings suggest this might not in fact be an issue in the other subsets (we note this as an aspect warranting future exploration).

Summary across the experiments

To summarise, we compiled correlation coefficients for all three analyses into one summary table (Table 3). All analyses use the same parameters (3-s non-overlapping windows, simple key profiles, and Cosine similarity metric).

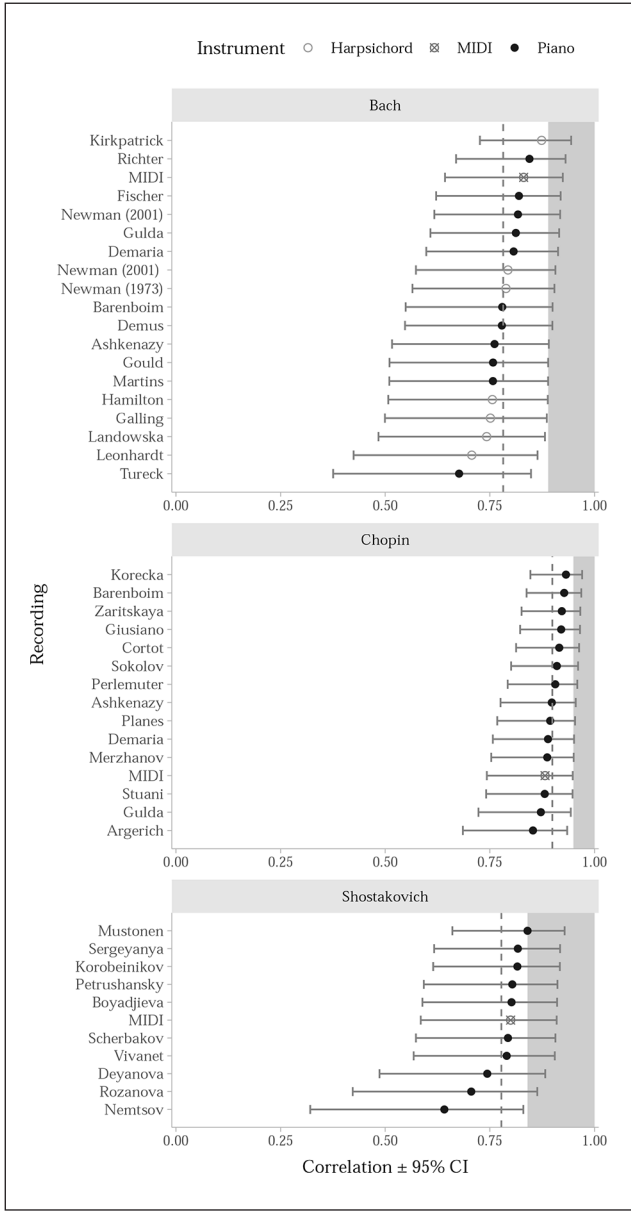


Figure 4. Correlation Between Model Predictions and Expert Ratings Across Recordings and Data Sets. Filled (piano) and empty (harpsichord) circles denote the instrument and squares symbolic (MIDI) denote the representations. The dashed line shows the model average, and the shaded area shows the feasible maximum (defined as the average correlation between the experts for each data set).

Data in Table 3 suggest the larger and more diverse sample of recordings in the present Experiment 3 can be predicted approximately at the same level ($r = .820$) as the other, simplified data sets (overall $r = .835$ for Experiment 1 and $r = .859$ for Experiment 2) when averaging. Mean correlations (and 95% CIs) suggest a fairly consistent pattern within composers, with the

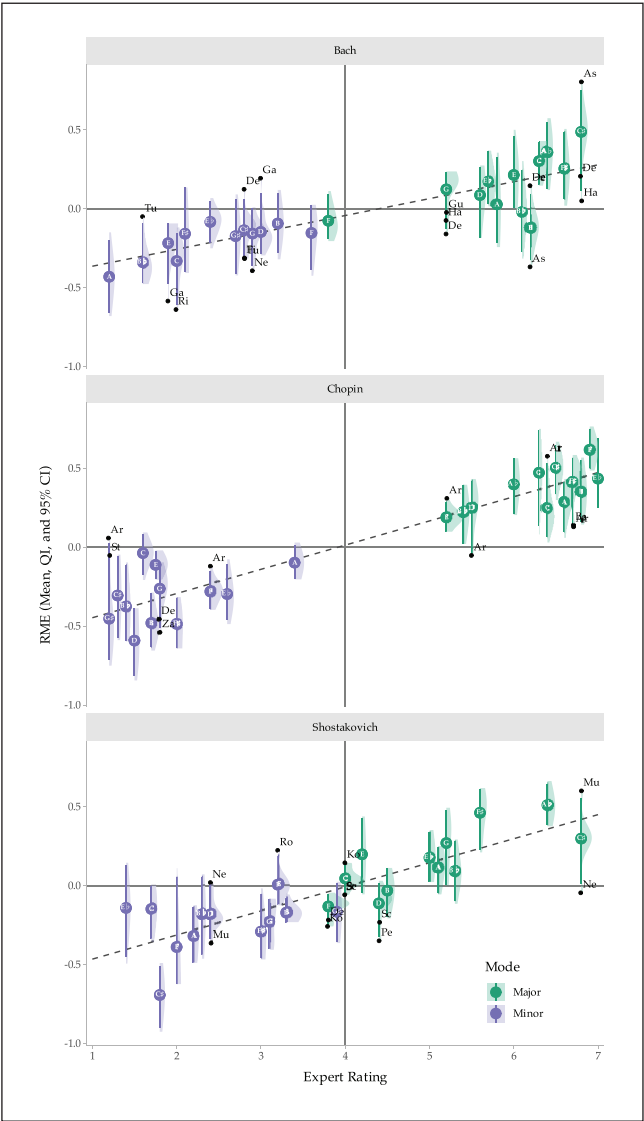


Figure 5. RME Model Predictions Summarised Across Preludes, Recordings and Datasets ($N = 1,008$). The distribution of the values in the Y-axis indicates the mean, quantile intervals (Q1) and 95% confidence intervals (CI) and outliers are shown by the first two initials of the performer last name (Mu = Mustonen, Ne = Newman, Sc = Scherbakov, Ro = Rozanova, Pe = Petrushansky, see Appendices 1 to 3). X-axis shows the mean expert ratings for the preludes. The least square fit has been added to demonstrate the linear relationship between the model predictions and the expert ratings. The outliers (1.5 Q1 range) have been added with the initial letters of the performer.

RME model more accurate in capturing expert ratings of Chopin versus Bach and Shostakovich Preludes (with marginal differences between Bach and Shostakovich). In the more detailed analysis reported in supporting materials (https://github.com/tuomaseerola/relative_mode), we summarised the main aspects that occasionally led to poor fit between the model prediction and

Table 3. Results Across Composers and Variants Using Correlations Between the Expert Ratings and the RME Model (Means and 95% CIs).

Composer	Exp. 1	Exp. 2	Exp. 3
	MIDI (N=72)	Audio (N=72)	Audio (N=1,008)
Bach	.832 [.638, .942]	.839 [.659, .928]	.782 [.555, .901]
Chopin	.868 [.729, .937]	.915 [.811, .963]	.900 [.779, .956]
Shostakovich	.791 [.631, .882]	.799 [.584, .909]	.777 [.548, .898]
Overall	.835 [.666, .920]	.859 [.784, .910]	.820 [.628, .919]

expert ratings (primarily excerpt length and window size/overlap). However, the model overall performs in a robust manner, explaining 70%–74% of the expert ratings.

Bearing in mind the mean correlation between the expert ratings of .890 (Bach), .940 (Chopin), and .840 (Shostakovich), the model accuracy could be said to be near the ceiling of algorithmic performance for a complex construct on which even trained musical experts will not always agree. When considered in that context, only a small amount of improvement (.05–.10) could be expected. However, the actual process of assessing the relative mode by the experts and the algorithm is probably different, as experts can use their extensive knowledge of harmonic patterns, cadences, and phrases to their advantage in this process, whereas the algorithm is merely counting the pitch-classes.

As mentioned earlier, the *mirmode* function in *MIR toolbox* (version 1.8.1, Lartillot et al., 2008) provides a similar estimation, but this is known to be unreliable (Kumar et al., 2015; Zhou et al., 2023). Nonetheless, as the *MIR toolbox* is widely used in the field of music cognition, we felt colleagues might benefit from knowing its performance in this type of task. Therefore, we applied this function with default parameters to the full data in Experiment 3. This achieves a correlation of .474 with all recordings ($r = .394$ with Bach recordings, $r = .630$ with Chopin recordings, and $r = .289$ with Shostakovich recordings). This suggests the effectiveness of our RME model is not independent of the manner in which it is implemented.

Discussion

Using a novel corpus of 1,008 audio recordings featuring different performances of 24-piece Prelude sets by Bach (18), Chopin (14), and Shostakovich (10), here we propose a process automating RME through an approach explaining a substantial amount of variance of expert ratings (>70%). To place this in context, other automated extractions of scalar ratings of music excerpts usually deliver lower prediction rates: other studies document audio-based cues predict approximately 35% of ‘wanting to move’ ratings (Witek et al., 2014) and 52% of tension ratings (Barchet et al., 2024), as well as 45% of valence and 61% of arousal ratings (Malheiro et al., 2016; Saari et al., 2016). Therefore, we find our model’s performance in predicting a complex and fairly novel construct such as relative mode promising.

In addition to absolute performance, we find the outcome of our model encouraging for three reasons. First, reasonable analysts will disagree about complex musical constructs, and here, we find a correlation between our model and expert ratings close to that achieved between the experts themselves. Second, we chose the most parsimonious parameters for the model that delivered consistent results, suggesting future refinements might build upon our ‘generic’

parameters. Third, our model appears robust across many common variations in recorded performances. For example, the expressive (timing, tempo, dynamics, tuning, balance between the two hands in piano, etc.) and technical (room size, reverberation, microphone placement, recording quality, compression, and instrument type) factors are known to affect the tonal contents of a signal (Müller & Ewert, 2011).

Although the concept of relative mode is not currently in widespread use, its foundation can be seen in the writings of notable theorists (Clendinning & Marvin, 2016; Temperley & Tan, 2012; Schoenberg & Stein, 1969). This is particularly true of research on the role of mode in different musical eras, as its nature and function have changed remarkably. Therefore, we believe RMEs hold significant potential benefits for enhancing our understanding of music. Moreover, as mode is widely recognised as the ‘super cue’ of musical emotion (Eerola et al., 2013), a more granular approach to classification beyond the traditional binary of major/minor could prove useful for a wide range of musical scholars.

We see particular value for a more nuanced treatment of modality within the field of music cognition, where mode is often treated as a binary category – particularly in experimental stimuli which often employ highly controlled and simplified examples (Baumgartner et al., 2006; Costa et al., 2004; Dalla Bella et al., 2001; Gagnon & Peretz, 2003; Gosselin et al., 2007; Gosselin et al., 2006). However, the process of obtaining RMEs from expert analysers is time-intensive and requires not only a careful assemblage of scores, recordings, and expert theorists but also extensive discussion of nuances and distinctions (Delle Grazie et al., 2025). Consequently, software tools for automating this process would be highly desirable for scholars in a variety of domains.

Limitations

Despite the encouraging results, we note several instances of mismatch between our RME and expert ratings. These often relate to the length of the excerpts, that is, when extremely short excerpts do not behave well with a fixed windowing. In some rare cases, the poor fit between the experts and the model may reflect complexities in the passages leading to disagreement amongst experts in ways the algorithm is not exploring (i.e., patterns of chords and cadences). However, we observed too few cases to construct alternative modelling strategies for these examples. Consequently, future research using a more diverse range of relative mode ratings with different types of excerpts and compositional strategies would be useful in expanding and improving this approach.

Another limitation relates to the musical scope of our materials: here, we used relatively homogeneous Western classical music materials by three composers from the late 17th to early 20th century. Consequently, more research is needed to explore a broader scope of material spanning different classical music time periods and genres. It would also be useful to explore the viability of the relative mode concept and the algorithm in materials outside Western classical music. Although the efforts to obtain expert data on a wider span of classical music are ongoing (Delle Grazie et al., 2025), it remains to be seen to what degree the model can predict ratings in such materials. The notion could also be explored with non-Western tonal systems which have well-codified associations with modes such as North Hindustani ragas (Mathur et al., 2015) or Korean court music (Nam, 1998).

Implications for future analyses

Major and minor form foundational concepts in Western music theory with a well-understood role in conveying musical emotion. Western music theory as well as music cognition generally

treat major and minor as a binary category with explicitly defined features (i.e., Type 2 categories). However, treating it instead as a Type 1 category (i.e., a collection of features with fuzzy boundaries) holds potential to better capture its actual use. Given the primacy of mode in emotional responses to music, a more granular understanding of its structure holds tremendous potential. This more nuanced perspective could have multiple uses from probabilistic modelling of musical structure to predicting emotions expressed by music.

In addition to implications for our understanding of mode as a construct, we believe this approach of exploring classic categorical music-theoretic concepts on a continuum could be invaluable for other topics in music theory and music cognition. For instance, much of music theory has been driven by notational constraints that articulate elements as categorical (key signatures, cadences, pitches, and phrases). Yet many of these constructs are operationalised in a continuous fashion – segment boundaries can be seen as probabilistic rather than binary (Abdallah et al., 2015; Krumhansl, 1996); and absolute pitch is more of a continuum of a pitch memory than a clearly categorical phenomenon (Schellenberg & Trehub, 2003). Reviewing some of the classic categorical concepts as continuums may expand the toolbox of music analysts and allow those who strive to capture perception of music to capitalise richer sets of data.

Acknowledgements

The authors would like to thank the Durham Institute of Advanced Study for support of this project. The authors would also like to thank Jamie Ling, Cameron Anderson, Massimo Delle Grazie, and Brian McNulty for their assistance in organising files and preparing materials. Support from the Western University Visiting Research Chair program (Fall of 2022) facilitating a collaboration with Jonathan De Souza led to the expert rating data instrumental to this project.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project began during a fellowship at the Durham Institute of Advanced Study awarded to M.S. in the winter of 2023. It was supported by the Social Sciences and Humanities Research Council of Canada, the Canadian Foundation for Innovation, and the McMaster Arts Research Board.

ORCID iD

Tuomas Eerola  <https://orcid.org/0000-0002-2896-929X>

References

- Aarden, B. (2003). *Dynamic melodic expectancy* (PhD thesis). The Ohio State University.
- Abdallah, S., Gold, N., & Marsden, A. (2015). Analysing symbolic music with probabilistic grammars. In D. Meredith (Ed.), *Computational music analysis* (pp. 157–189). Springer.
- Albrecht, J. D., & Shanahan, D. (2013). The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode. *Music Perception*, 31(1), 59–67. <https://doi.org/10.1525/MP.2013.31.1.59>
- Anderson, C. J., & Schutz, M. (2022). Exploring historic changes in musical communication: Deconstructing emotional cues in preludes by Bach and Chopin. *Psychology of Music*, 50(5), 1424–1442. <https://doi.org/https://doi.org/10.1177/03057356211046375>
- Barchet, A. V., Rimmele, J. M., & Pelofi, C. (2024). TenseMusic: An automatic prediction model for musical tension. *PLOS ONE*, 19, Article e0296385. <https://doi.org/10.1371/journal.pone.0296385>

- Battcock, A., & Schutz, M. (2019). Acoustically expressing affect. *Music Perception: An Interdisciplinary Journal*, 37(1), 66–91.
- Baumgartner, T., Esslen, M., & Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, 60(1), 34–43.
- Bellmann, H. G. (2005). About the determination of key of a musical excerpt. In R. Kronland-Martinet, T. Voinier & S. Ystad (Eds), *Proceedings of computer music modeling and retrieval* (pp. 76–91). Springer.
- Beveridge, S., & Knox, D. (2009). An exploration of the effect of structural and acoustical features on perceived musical emotion. In Proceedings of the 4th audio mostly conference, publisher association for computing machinery (ACM) (pp. 93–98). ACM.
- Brown, J. C. (1991). Calculation of a constant Q spectral transform. *Journal of the Acoustical Society of America*, 89(1), 425–434.
- Chuan, C.-H., & Chew, E. (2005). Polyphonic audio key finding using the spiral array CEG algorithm. In 2005 IEEE international conference on multimedia and expo (pp. 21–24). IEEE.
- Chuan, C.-H., & Chew, E. (2006). Audio key finding: Considerations in system design and case studies on Chopin's 24 Preludes. *EURASIP Journal on Advances in Signal Processing*, 2007, 1–15.
- Chuan, C.-H., & Chew, E. (2014). The KUSC classical music dataset for audio key finding. *The International Journal of Multimedia & Its Applications*, 6(4), 1–16.
- Clendinning, J. P., & Marvin, E. W. (2016). *The musician's guide to theory and analysis*. WW Norton & Company.
- Costa, M., Fine, P. A., & Ricci Bitti, P. (2004). Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception*, 22(1), 1–14.
- 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), B1–10.
- Delle Grazie, M., Anderson, C. J., De Souza, J., & Schutz, M. (2025). Analysis from multiple perspectives (AMP): Applying decision hygiene to analysis of musical structure. *Psychology of Music*. Manuscript submitted for publication.
- Delle Grazie, M., Anderson, C. J., & Schutz, M. (in press). Breaking with common practice: Exploring modernist musical emotion. *Psychology of Music*.
- Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, 4, Article 487. <https://doi.org/10.3389/fpsyg.2013.00487>
- Eerola, T., & Toiviainen, P. (2003). *MIDI toolbox: MATLAB tools for music research*. University of Jyväskylä.
- Frankland, B., & Cohen, A. J. (1996). Using the Krumhansl and Schmuckler key-finding algorithm to quantify the effects of tonality in the interpolated-tone pitch-comparison task. *Music Perception*, 14(1), 57–83.
- Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to 'happy-sad' judgements in equitone melodies. *Cognition and Emotion*, 17(1), 25–40.
- Gómez, E. (2006). Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing*, 18(3), 294–304. <https://doi.org/10.1287/ijoc.1040.0126>
- Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, 45, 236–244.
- Gosselin, N., Samson, S., Adolphs, R., Noulhiane, M., Roy, M., Hasboun, D., & Peretz, I. (2006). Emotional responses to unpleasant music correlates with damage to the parahippocampal cortex. *Brain*, 129(10), 2585–2592.
- Izmirli, Ö. (2005). Template based key finding from audio. *ICMC, 2005*, 211–214.
- Kahneman, D., Sibony, O., & Sunstein, C. (2021). *Noise: A flaw in human judgment*. William Collins.
- Kelly, B. O., Anderson, C. J., & Schutz, M. (2021). Exploring changes in the emotional classification of music between eras. *Auditory Perception & Cognition*, 4(1–2), 121–131. <https://doi.org/10.1080/25742442.2021.1988422>
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.
- Krumhansl, C. L. (1996). A perceptual analysis of Mozart's piano sonata k. 282: Segmentation, tension, and musical ideas. *Music Perception*, 13(3), 401–432.

- Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89(4), 334–368.
- Kumar, N., Kumar, R., & Bhattacharya, S. (2015). Testing reliability of Mirtoolbox. In *2015 2nd international conference on electronics and communication systems (ICECS)* (pp. 710–717). IEEE. <https://doi.org/10.1109/ECS.2015.7125004>
- Lartillot, O., Toivainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme & R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Springer.
- Lin, Y.-C., Yang, Y.-H., & Chen, H. H. (2011). Exploiting online music tags for music emotion classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 7(1), 1–16.
- Long, M. K. (2020). What do signatures signify? The curious case of seventeenth-century English key. *Journal of Music Theory*, 64(2), 147–201.
- Malheiro, R., Panda, R., Gomes, P., & Paiva, R. P. (2016). Emotionally-relevant features for classification and regression of music lyrics. *IEEE Transactions on Affective Computing*, 9(2), 240–254.
- Mathur, A., Vijayakumar, S. H., Chakrabarti, B., & Singh, N. C. (2015). Emotional responses to Hindustani raga music: The role of musical structure. *Frontiers in Psychology*, 6, Article 513.
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and music signal analysis in python. *Proceedings of the 14th Python in Science Conference*, 8, 18–25.
- Müller, M., & Ewert, S. (2011). Chroma toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In A. Klapuri & C. Leide (Eds.), *Proceedings of 12th international conference on music information retrieval* (pp. 215–220). ISMIR.
- Nam, U. (1998). Pitch distributions in Korean court music: Evidence consistent with tonal hierarchies. *Music Perception*, 16(2), 243–247.
- Nápoles López, N., Arthur, C., & Fujinaga, I. (2019). Key-finding based on a hidden Markov model and key profiles. In *6th international conference on digital libraries for musicology* (pp. 33–37). ACM.
- Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception: An Interdisciplinary Journal*, 11(4), 409–464.
- Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, 16(4), 202–206.
- Quinn, I. (2010). Are pitch-class profiles really ‘key for key’? *Zeitschrift der Gesellschaft für Musiktheorie* [Journal of the German-speaking Society of Music Theory], 7(2), 151–163.
- Rafii, Z. (2022). The constant-Q harmonic coefficients: A timbre feature designed for music signals [Lecture Notes]. *IEEE Signal Processing Magazine*, 39(3), 90–96. <https://doi.org/10.1109/MSP.2021.3138870>
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382–439.
- Saari, P., Eerola, T., Barthet, M., Fazekas, G., & Lartillot, O. (2016). Genre-adaptive semantic computing and audio-based modelling for music mood annotation. *IEEE Transactions on Affective Computing*, 7(2), 122–135. <https://doi.org/10.1109/TAFFC.2015.2462841>
- Sapp, C. S. (2011). *Computational methods for the analysis of musical structure* (PhD thesis). Stanford University.
- Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, 14(3), 262–266.
- Schmuckler, M. A., & Tomovski, R. (2005). Perceptual tests of an algorithm for musical key-finding. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 1124–1149.
- Schoenberg, A., & Stein, L. (1969). *Structural functions of harmony*. WW Norton & Company.
- Schörkhuber, C., & Klapuri, A. (2010). Constant-q transform toolbox for music processing. In *7th sound and music computing conference* (pp. 3–64). CORE.
- Shmulevich, I., & Yli-Harja, O. (2000). Localized key finding: Algorithms and applications. *Music Perception*, 17(4), 531–544.

Swierczek, K., & Schutz, M. (2025). Evaluating musical feature extraction with versions. *Psychology of Music*. Manuscript submitted for publication.

Temperley, D. (1999). What’s key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, 17(1), 65–100.

Temperley, D. (2002). A Bayesian approach to key-finding. In C. Anagnostopoulou, M. Ferrand & A. Smaill (Eds.), *Music and artificial intelligence: Second international conference ICMAI 2002* Edinburgh, Scotland, UK, September 12–14, 2002 Proceedings (pp. 195–206). Springer.

Temperley, D. (2007). *Music and probability*. MIT Press.

Temperley, D., & Tan, D. (2012). Emotional connotations of diatonic modes. *Music Perception: An Interdisciplinary Journal*, 30(3), 237–257.

Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, 32(6), 741–766.

Vasu, K., & Choudhary, S. (2022). Music information retrieval using similarity based relevance ranking techniques. *Scalable Computing: Practice and Experience*, 23(3), 103–114.

Vos, P. G., & Van Geenen, E. W. (1996). A parallel-processing key-finding model. *Music Perception*, 14(2), 185–223.

Vuvan, D. T., Prince, J. B., & Schmuckler, M. A. (2011). Probing the minor tonal hierarchy. *Music Perception: An Interdisciplinary Journal*, 28(5), 461–472.

White, C. W. (2012). Changing styles, changing corpora, changing tonal models. *Music Perception: An Interdisciplinary Journal*, 31(3), 244–253.

Witek, M. A., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLOS ONE*, 9(4), Article e94446.

Zbikowski, L. M. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford University Press.

Zhou, J., Anderson, C. J., & Schutz, M. (2023). Accuracies in algorithmic predictors of musical emotion. *Proceedings of the Acoustics Week in Canada*, 51(3), 78–79.

Appendix I

Recordings Used – Bach’s Well Tempered Clavier Book.

Performer	Year	Label	Instrument
Joerg Demus	1956/1992	MCA Records	Piano
Sviatoslav Richter	1970/1992	BMG Classics	Piano
Glenn Gould	1963/64/65/93	Sony Classical	Piano
João Carlos Martins	1964/1994	Labour Records	Piano
Friedrich Gulda	1972/1995	Decca	Piano
Rosalyn Tureck	1953/1999	Deutsche Grammophon	Piano
Anthony Newman	2001	KHAEON World Music	Piano
Vladimir Ashkenazy	2006	Decca	Piano
Daniel Barenboim	2006	Warner Classics	Piano
Edwin Fischer	1989/2007	EMI Records	Piano
Pietro De Maria	2015	Decca	Piano
Ralph Kirkpatrick	1963	Deutsche Grammophon	Harpsichord
Malcolm Hamilton	1964	Everest	Harpsichord
Anthony Newman	1973	Columbia Masterworks	Harpsichord
Wanda Landowska	1987	RCA Victor Red Seal	Harpsichord
Gustav Leonhardt	1973/1989	BMG Classics	Harpsichord
Anthony Newman	2001	Vox Cum Laude	Harpsichord
Martin Galling	2006	Vox Records	Harpsichord

Appendix 2

Recordings Used – Chopin’s 24 Preludes.

Performer	Year	Label	Instrument
Friedrich Gulda	1954	Decca	Piano
Irina Zaritskaya	1989	Naxos	Piano
Vladimir Ashkenazy	1993	Decca	Piano
Martha Argerich	2002	Deutsche Grammophon	Piano
Daniel Barenboim	1976/2003	EMI Classics	Piano
Vlado Perlemuter	2006	Nimbus Records	Piano
Philippe Giusiano	2006	Mirare	Piano
Pietro De Maria	2008	Decca	Piano
Alain Planès	2011	Harmonia Mundi	Piano
Grigory Sokolov	2013	Naïve	Piano
Giampaolo Stuanì	2014	OnClassical	Piano
Maria Korecka-Soszkowska	2019	DUX Recording Producers	Piano
Alfred Cortot	2020	Archipel Records	Piano
Victor Merzhanov	1975/2021	Vista Vera	Piano

Appendix 3

Recordings Used – Shostakovich’s 24 Preludes.

Performer	Year	Label	Instrument
Konstantin Scherbakov	2003	Naxos	Piano
Boris Petrushansky	2006	Stradivarius	Piano
Elena Rozanova	2008	Harmonia Mundi	Piano
Lilia Boyadjieva	2009	Artek	Piano
Jascha Nemtsov	2009	Profil Medien	Piano
Timur Sergeyanya	2011	Northern Flowers	Piano
Andrei Korobeinikov	2012	Mirare	Piano
Andrea Vivanet	2021	Gramola Records	Piano
Olli Mustonen	2015	Decca	Piano
Marta Deyanova	1985	Nimbus Records	Piano

Appendix 4

A Random Sample of Recordings Used in Experiment 2.

Bach's WTC I	Chopin's 24 preludes	Shostakovich's 24 preludes
Ashkenazy A Major	Argerich C Major	Boyadjieva C Minor
Ashkenazy A Minor	Argerich G Minor	Boyadjieva Bb Minor
Barenboim G# Minor	Ashkenazy C Minor	Boyadjieva F Minor
De Maria Ab Major	Ashkenazy Ab Major	Korobeinikov B Major
De Maria G Minor	Ashkenazy Db Major	Korobeinikov Db Major
Demus G Major	Barenboim C# Minor	Korobeinikov F# Major
Fischer F# Minor	Barenboim G# Minor	Nemtsov B Minor
Galling F# Major	Cortot D Major	Nemtsov C# Minor
Gould F Minor	Cortot A Major	Nemtsov F# Minor
Gulda F Major	De Maria D Minor	Petrushansky D Major
Hamilton E Minor	De Maria A Minor	Petrushansky G Major
Kirkpatrick E Major	Giusiano Bb Major	Rozanova D Minor
Landowska Eb Minor	Giusiano Eb Major	Rozanova G Minor
Leonhardt Eb Major	Gulda Bb Minor	Scherbakov Eb Major
Martins D Minor	Gulda Eb Minor	Scherbakov Eb Minor
Newman B Major	Korecka-Soszkowvska B Major	Scherbakov Ab Major
Newman B Minor	Korecka-Soszkowvska E Major	Scherbakov G# Minor
Newman C# Major	Merzhanov B Minor	Sergeyanya E Major
Newman C# Minor	Merzhanov E Minor	Sergeyanya A Major
Newman D Major	Perlemuter F Major	Vivanet E Minor
Richter C Minor	Planes F Minor	Vivanet A Minor
Richter Bb Minor	Sokolov F# Major	Ashkenazy C Major
Tureck C Major	Stuani F# Minor	Ashkenazy Bb Major
Tureck Bb Major	Zaritskaya G Major	Ashkenazy F Major

Table contents refer to the performer, chroma, and mode, sampled from the full set of recordings ($N=1008$, see Appendices 1 to 3).