Designing Emotionally Navigable Harmonic Spaces for Interactive Music Generation

Ariana Pereira[1], Choenden Kyirong[1], Raquel Lucena[1] and Rafael Ramirez[1]

[1]Universitat Pompeu Fabra (UPF), Carrer de la Mercè, 12, Ciutat Vella, 08004, Barcelona

Abstract

MoodSwing is an interactive music-generation interface within the Musitopia platform in which users draw trajectories across a two-dimensional emotional canvas to generate evolving harmonic progressions. Designing such a system requires positioning chords within a valence–arousal space so that user navigation produces music that is both emotionally meaningful and musically coherent. This work investigates whether chord-to-emotion mappings can be computationally designed to balance these two objectives simultaneously. We present a design pipeline combining CLAP audio embeddings of synthesized chord audio, affective modeling using the `song_sent_scores` toolkit, rule-based harmonic evaluation, and simulated annealing optimization. Emotional chord organizations were derived from both isolated chord embeddings and contextual chord embeddings in order to evaluate how harmonic context influences affective positioning within latent emotional space. The resulting harmonic organizations were evaluated computationally using voice-leading and structural similarity metrics, and perceptually through a listener study comparing human judgments of emotional affect and harmonic smoothness against the embedding-derived arrangements. Results suggest that emotionally coherent harmonic organization and musically smooth harmonic organization represent partially competing objectives, though optimization techniques can identify diverse harmonic structures that balance both constraints simultaneously. Furthermore, comparisons between isolated and contextual chord embeddings reveal that many harmonic-emotional relationships remain stable across contexts, while certain harmonies exhibit substantial contextual sensitivity. These findings contribute toward the design of emotionally navigable harmonic interfaces for interactive music systems and highlight the importance of multi-objective approaches that integrate affective modeling, harmonic coherence, and perceptual evaluation.

Keywords

Generative models, Music Therapy, Computational Creativity, Music Information Retrieval

1. Introduction

Emotional expression is a fundamental aspect of musical experience [1] and plays a central role in the research and design goals of Musitopia [1] [2], a platform focused on interactive music-based therapies and emotionally informed creative tools. A primary objective of the platform is to make musical expression more accessible to users without formal music theory training by enabling interaction through emotional intent rather than symbolic musical knowledge. People use drawing for emotional expression by externalizing internal feelings, allowing them to process complex emotions without relying on words. Through intuitive mark-making, anyone can associate abstract moods with something tangible such as colors [3] aiding in self-reflection, catharsis, and emotional regulation.

One such system developed within the platform is *MoodSwing*, an interactive music-generation interface in which users draw trajectories across a two-dimensional emotional canvas to generate evolving harmonic progressions in real time. Within *MoodSwing*, users are not passive recipients of algorithmically generated music but active participants in a co-creative process. By manipulating the trajectory, color, and thickness of their drawings, users make expressive decisions that directly influence the harmonic, melodic, and temporal characteristics of the generated music. The visual interface encourages exploration and experimentation, allowing users to iteratively refine musical outcomes

ICCC'26: International Conference on Computational Creativity, June 29–July 03, 2026, Coimbra, Portugal

ariana.pereira01@estudiant.upf.edu (A. Pereira); ngawangchoeden.kyirong01@estudiant.upf.edu (C. Kyirong); raquel.lucena@upf.edu (R. Lucena); rafael.ramirez@upf.edu (R. Ramirez)

https://pereiraa10.github.io/ (A. Pereira); https://www.choendenkyirong.com/ (C. Kyirong)

0009-0005-1176-6081 (A. Pereira); 0009-0007-2490-4388 (R. Lucena); 0000-0002-9421-8566 (R. Ramirez)

[1]https://musitopia.appskynote.com/

through intuitive gestures rather than formal compositional techniques. In this way, MoodSwing functions as a creative partner that translates emotional intent into musical structure while preserving meaningful user control over the resulting composition. This interaction framework enables users to externalize emotions, explore alternative musical interpretations of the same emotional state, and engage in creative expression through a modality that is accessible regardless of prior musical training. The resulting musical outputs can be revisited, modified, and regenerated through additional drawing interactions, supporting an iterative creative workflow in which users explore multiple emotional and musical possibilities from the same initial idea. Additionally, the interface is organized according to the circumplex model of affect [4], representing emotional states along continuous valence and arousal dimensions. Within this space, chords are positioned according to their associated emotional characteristics, allowing users to navigate harmony spatially through emotional movement rather than traditional compositional rules.

Designing such a system introduces a central research challenge: how can chords be positioned within emotional space so that generated music remains both emotionally coherent and musically convincing? Harmonic relationships that appear emotionally meaningful do not necessarily produce smooth or perceptually satisfying musical transitions, while harmonically coherent progressions may fail to preserve emotionally interpretable structure.

An initial heuristic design for MoodSwing was informed by prior findings in music psychology associating major harmonies with positive affect and minor or dissonant harmonies with more negative or tense emotional responses [5, 6, 7]. However, such mappings are often culturally dependent, discretized, and difficult to translate into continuous interactive systems. Furthermore, these symbolic associations do not fully capture the contextual and perceptual complexity of harmonic emotion.

This work explores whether emotionally informed harmonic spaces can instead emerge from learned audio representations while simultaneously incorporating principles of musical coherence. By integrating CLAP audio embeddings [8], affective modeling, rule-based harmonic evaluation, and probabilistic optimization techniques, into a unified framework for creative interaction, we develop a multi-objective computational approach to explore whether harmonic structures can be organized according to emotional similarity while remaining musically navigable and coherent. By jointly optimizing emotional consistency and harmonic continuity, we construct and evaluate harmonic spaces that support emotion-driven musical exploration, providing both empirical insight into the relationship between affect and harmony and a practical foundation for emotionally guided generative music systems. Through this framework, we investigate the extent to which emotional organization can serve as a meaningful interface for creative musical interaction.

2. Related Work

The design of an emotionally navigable harmonic canvas draws on three intersecting traditions: dimensional models of affect, empirical research on harmony and emotion, and computational systems that condition music generation on affective targets. Russell's circumplex model of affect has become the dominant framework in music emotion research [9, 10] owing to its continuous geometry, which supports systems that treat emotion as a space to be traversed rather than a set of discrete categories. Gabrielsson and Lindström [11] provide the canonical synthesis of how structural musical features map into this affective space.

While much of this literature focuses on broad features such as tempo and mode, vertical harmony has more recently received dedicated empirical attention at the chord level. Lahdelma and Eerola [12] characterized the affective profiles of fourteen chord qualities, substantiating long-standing intuitions [13] that major chords read as positive, minor as melancholic, and diminished and augmented as ambiguous or anxious, while revealing more nuanced characteristics for seventh chords. Affective-priming studies [6, 7, 14] demonstrate that these chord–emotion associations operate automatically, and Zhang et al. [15] show that the perceived valence of a chord is modulated by its harmonic context. This finding directly motivates our comparison of isolated and contextual embedding conditions.

Affect-driven music generation systems have a long history of mapping affective targets onto musical parameters at the score level: Wallis et al. [16] controlled valence through chord-progression choice and arousal through accompaniment density; Ehrlich et al. [17] embedded this paradigm within a closed-loop brain–computer interface for emotion mediation; AffectMachine-Classical [18] drives composition through a probabilistic chord-transition matrix anchored to valence–arousal targets. Dash and Agres [19] survey this space, distinguishing simple rule sets (typically tempo to arousal, mode to valence) from complex sets that incorporate harmonic structure. The work presented here departs from this tradition in two ways. First, rather than hand-specifying the chord–valence relationship, we recover it empirically from CLAP [20, 21] audio embeddings using the `song_sent_scores` toolkit [8]. Second, the artifact being designed is not a generated piece of music but a spatial layout where affective coordinates of a fixed chord vocabulary allows users to traverse the harmonic space through gestural interaction.

3. Methodology

3.1. Harmonic Circumplex Design

The interactive harmonic navigation system implemented within Musitopia's MoodSwing can be seen in Figure 1. The circumplex serves as the core interaction model of the platform, allowing users to navigate harmonic material by drawing on the canvas to form emotional trajectories rather than through conventional symbolic music theory concepts. The initial spatial organization of the chords was heuristically designed by combining dimensional models of affect [4] with harmonic-emotion associations reported in prior music psychology literature [11, 12].

3.2. Chord–Emotion Embedding Derived through Audio Synthesis

To evaluate the optimal emotional positioning of the 21 chords we included in the initial heuristic mapping, all chords from the original heuristic mapping were rendered in MIDI using root positioning centered around C3 with an additional octave doubling of the root note in the bass to reinforce harmonic stability. The MIDI was synthesized and converted to .mp3 files using a soft piano-like instrument sound selected for its neutral and emotionally legible character, representative of the ambient music style central to the domain of wellness applications like Musitopia. Performance parameters such as velocity, duration, and timing had consistent values across all audio renderings to minimize expressive variability. Two forms of audio material were prepared for analysis:

1. **Single-chord audios:** 21 isolated chords synthesized for 4 seconds each
2. **Full-cycle chord audios**: continuous audios of all 21 chords being played for 4 seconds each in succession of each other. 21 different audio files were generated cycling through different chords as starting points.

All audio files were analyzed using the `song_sent_score` [8] toolkit, which estimates valence and arousal coordinates using CLAP (Contrastive Language–Audio Pretraining) [20] embeddings. For single-chords audio recordings, a single coordinate pair was extracted per file. For full-cycle audios, embeddings were computed using a 4-second chunking window, corresponding to the duration of each chord segment in the progression. Because the contextual sequences were rendered continuously with reverberation and natural decay, neighboring chords partially overlapped acoustically. This "bleeding" between segments was intentionally preserved to better approximate how harmonic transitions are perceived in realistic listening conditions rather than as perfectly isolated events. The coordinate pairs from the chords analyzed in harmonic context from the full cycle audio files were obtained by averaging coordinates across all chunks of audio which related to the same chord.

Figure 2 is an example of one of emotional trajectory outlined by the coordinates of one example full cycle audio file, derived from the CLAP embeddings. Within all of the audio files analyzed, most samples are clustered within the lower-right region of the emotional plane, roughly corresponding to calm and relaxed affect. This concentration was likely influenced by the consistent timbre of the

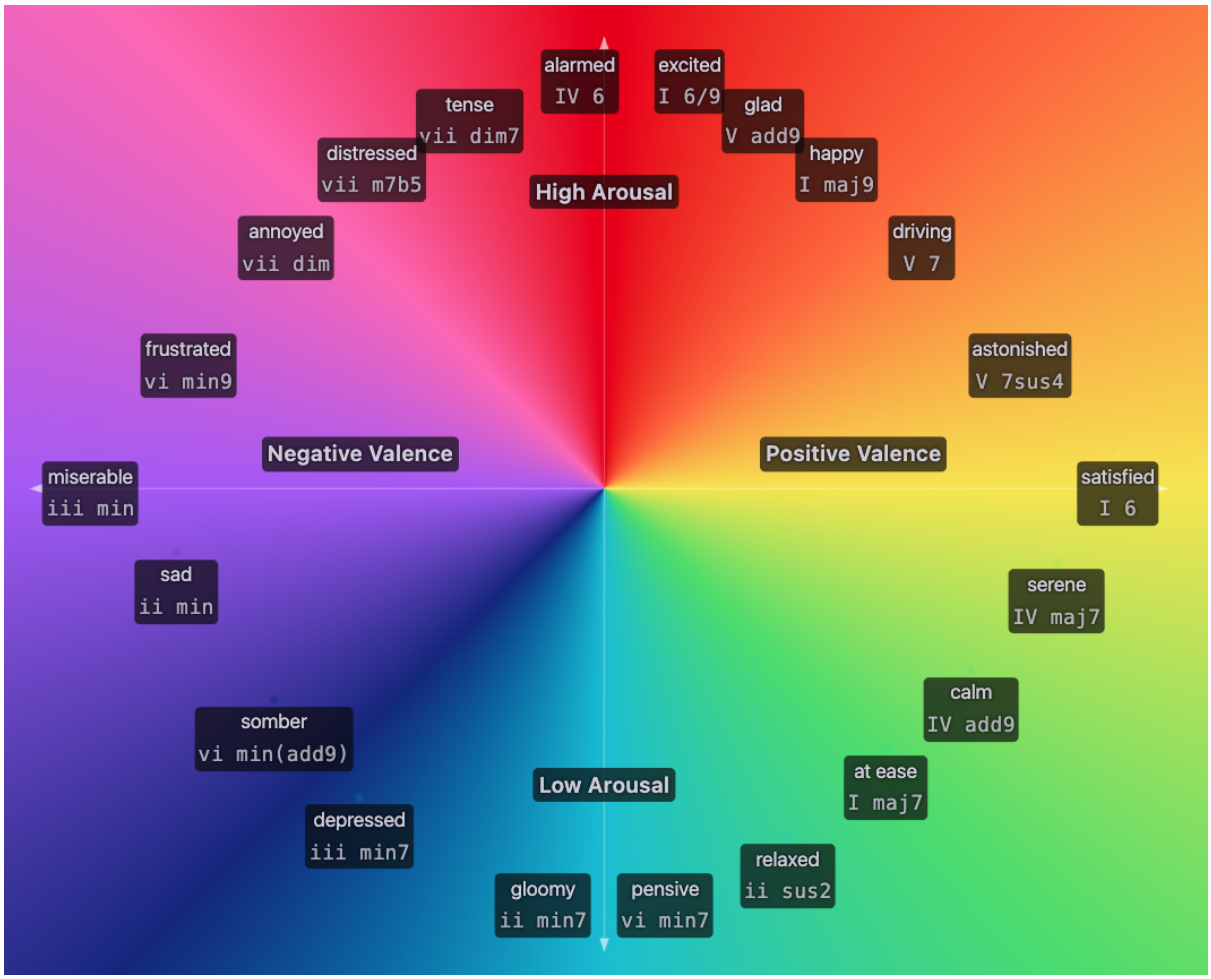


Figure 1: Heuristic harmonic circumplex used in MoodSwing for emotionally harmonic navigation. The horizontal axis represents valence and the vertical axis represents arousal, while mood labels indicate the affective associations of different chord regions.

piano synthesis. To facilitate a relative comparison between chords, all coordinates were subsequently normalized independently along both x- and y-axes into the range [0,1] relating to valence and arousal respectively.

3.3. Circular Ordering and Sequence Derivation

To derive an optimized emotional ordering of chords, centroid-based angular analysis was performed on the normalized valence–arousal coordinates so chords could be sorted radially around the emotional space into a sequence which could fit well into the existing circumplex canvas design of MoodSwing (as shown in Figure 1. This procedure generated two optimized cyclic chord progressions, one derived from the single-chord embeddings (“Single”), and a second ordering derived from the in-contextual embeddings (“In Context”).

To compare the similarity of the two circular chord progressions, a circular index distance metric was introduced:

$$D(A,B) = \min_{k\in\{0,\ldots,N-1\}} \sum_{c\in C} \left(\min\left(|p_A(c) - p_{B_k}(c)|,\ N - |p_A(c) - p_{B_k}(c)|\right)\right)^2$$

Where

$$p_A(c)$$

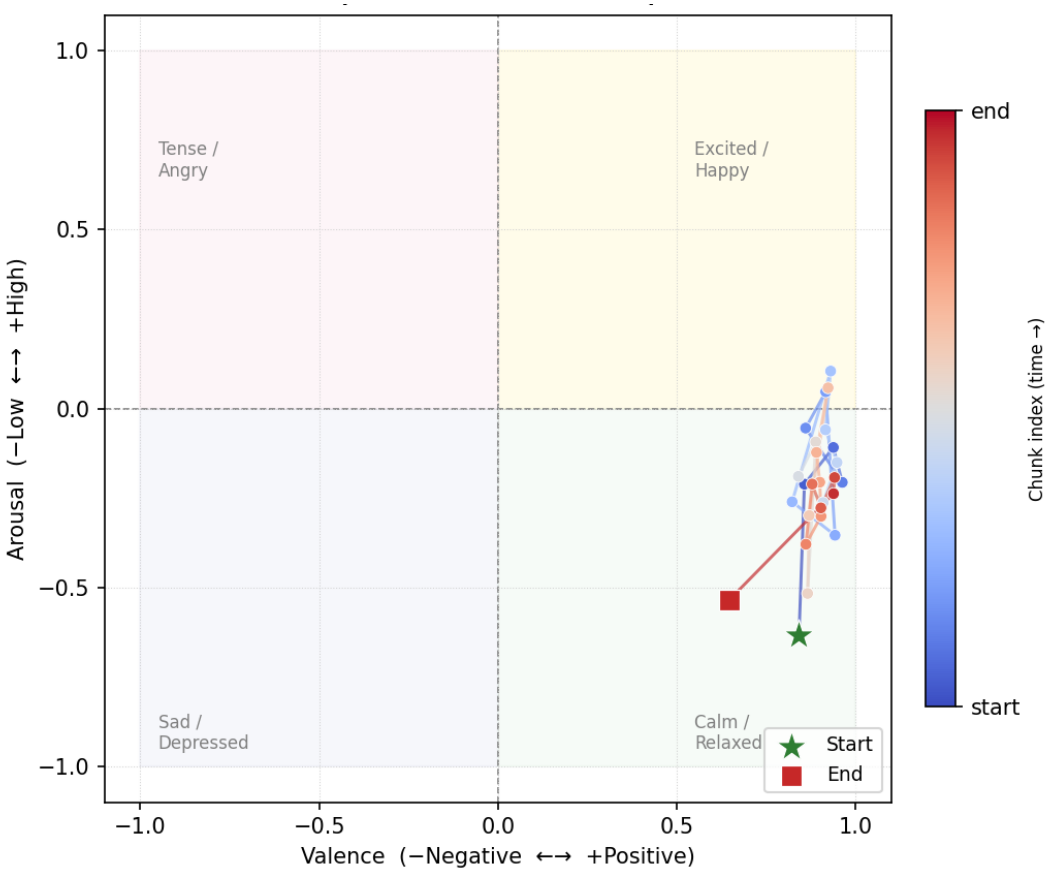


Figure 2: An example of the song sent score valence and arousal coordinates evaluated over every chord within one full-cycle chord audio file

is the position of the chord c in progression A, and

$$p_{B_k}(c)$$

is the position of the chord c in the k-th cyclic rotation of progression B.

This metric was chosen since squaring makes the metric sensitive to gross structural reorganization, not just fine-grained reordering and taking into account that cyclical chord progressions do not have a canonical starting point. As a baseline comparison, a randomly ordered chord progression was also included in the analysis.

3.4. Computational Musicality Assessment

Following the emotional optimization stage, the resulting chord sequences were evaluated for perceived musicality using a rule-based transition scoring framework. The goal was to investigate whether emotionally organized harmonic structures derived from the CLAP audio embeddings also exhibit properties commonly associated with smooth or coherent harmonic movement.

The scoring system evaluated each chord to chord transition using three complementary metrics, designed to capture different intuitions about what makes a harmonic change feel smooth or satisfying. Chords were evaluated in an octave- and voicing-agnostic manner.

1. **Pitch Motion Distance (M)** measures how much the individual notes of one chord need to move to reach the notes of the next. The intuition is rooted in the voice-leading principle of *parsimony*: transitions where notes that move by smaller intervals tend to sound smoother.
   A lower value of M indicates less overall movement — and therefore smoother voice leading. Duplicate mappings are permitted: two source notes may both map to the same target note if that minimizes motion.

2. **Common Tone Retention (*C*)** rewards transitions that preserve shared pitch classes between adjacent chords. When notes are held in common between two chords, the ear perceives a sense of continuity across the transition. The count of common tones retained is later normalized by the number of notes in the source chord, so that larger chords are not unfairly rewarded.
3. **Tendency Tone Resolution (*R*)** rewards harmonically conventional resolutions of *tendency tones* which are notes that carry an inherent sense of tension and are expected by the ear to move in a specific direction. Four interval types were identified as tendency tones based on common-practice tonal theory [22], each with its expected resolution direction: major sevenths (resolving upward by a semi-tone, or downward by a tone); dominant and minor sevenths (resolving downward by a semitone or tone); suspended fourths (resolving downward by a semitone or tone); and suspended seconds (resolving upward by a semitone or tone).

The three metrics were combined into a single musicality score for each transition between a S source chord and T target chord. The motion component was inverted and normalized by a constant $\lambda = 24$ (corresponding to the approximate worst-case total motion for a four-note chord), so that less motion contributes a higher score. Common tone retention was normalized by chord size. The weighted sum is:

$$\text{MusicalityScore}(S,T) = w_m\left(1 - \frac{M(S,T)}{\lambda}\right) + w_c\left(\frac{C(S,T)}{|S|}\right) + w_r\, R(S,T)$$

where $w_m = 1.0$, $w_c = 2.0$, and $w_r = 1.5$ are the weights assigned to pitch motion smoothness, common-tone retention, and tendency-tone resolution respectively. Common-tone retention is weighted most heavily, to mimic a stylistic approach that is commonplace in ambient relaxation music. $|S|$ is the number of pitch classes in the source chord.

Because the chord sequences in this study form closed loops (the final chord transitions back to the first), each progression was evaluated as a circular structure. The overall total musicality score for a progression is the average transition score across all N steps, including the wraparound. A higher total musicality score indicates that the progression, as a whole, exhibits smoother and more musically coherent harmonic motion. The system also recorded the individual per-component averages (motion, common-tone, and tendency) for each progression to allow fine-grained analysis of which factors drove the overall score.

3.5. Musicality and Emotion Mapping Optimization via Simulated Annealing

To further investigate an optimized harmonic organization within this musicality framework, the transition scoring system described in the previous section was reframed as a combinatorial optimization problem. Rather than evaluating chord-to-chord transitions independently, the goal was to determine whether the full collection of 21 chords could be arranged into a cyclic progression that simultaneously maximized smooth musical movement and preserved emotionally intuitive structure based on the *In Context* and *Single* baseline progressions.

Composite Scoring Function

Each candidate chord progression was assigned a total score composed of two parts: the summed musicality of each chord transition, and a structural similarity bonus measuring how closely the progression resembled the emotionally informed baseline progressions. The total optimization score for a progression P of length N is: progression P of length N is:

$$\text{TotalScore}(P) = \underbrace{\sum_{i=1}^{N} \text{MusicalityScore}\big(P_i,\ P_{(i \bmod N)+1}\big)}_{\text{sum of local musicality scores}} + w_s \cdot N \cdot \text{CIDSim}(P,\ \mathscr{R})$$

where $w_s = 1.0$ is the similarity weight, $\mathscr{R} = \{\text{In Context}, \text{Single}\}$ is the set of reference progressions, and CIDSim is the structural similarity score defined below. The factor of N scales the global similarity term so that it remains comparable in magnitude to the summed local musicality scores.

Structural Similarity Scoring via Circular Index Distance

To compare the structural similarity of two chord progressions, a metric called *Circular Index Distance* (CID) was used. Because a looping chord progression has no fixed starting point, the comparison process tests every rotational alignment before measuring similarity.Smaller positional changes indicate stronger structural similarity, while larger displacements indicate that the emotional organization of the progression has changed substantially. The resulting displacement value was then converted into a normalized similarity score between 0 and 1, where a value near 1 indicates that two progressions share very similar large-scale structure, while values closer to 0 indicate substantially different ordering patterns.

The optimizer compared each candidate progression against the emotional reference progressions and retained the higher similarity score. This allowed the system to freely discover solutions that resembled either emotional organization strategy rather than forcing convergence toward a single target.

Optimisation via Simulated Annealing

Finding the best ordering of 21 chords is computationally intensive because the large number of possible permutations. Because evaluating every possible sequence is infeasible, the study used *simulated annealing* (SA), a probabilistic optimization technique commonly used for large combinatorial search problems.

Simulated annealing can be understood as a guided trial-and-error process. The algorithm begins with a random chord ordering and repeatedly proposes small modifications to the sequence. After each modification, the system evaluates whether the new progression improves the overall score. If the new solution performs better, it is accepted immediately. However, the algorithm also occasionally accepts worse solutions early in the search process. This behavior is important because it helps the optimizer escape locally optimal solutions and continue exploring alternative harmonic structures. The probability of accepting worse solutions gradually decreases over time according to a "temperature" parameter.

Two mutation operators were applied as well to increase diversity in the searched pool of candidates while still preserving the requirement that every chord appears exactly once within the loop:

- **Swap operations:** two chords exchange positions, allowing small local refinements.
- **2-opt reversals:** an entire section of the progression is reversed, enabling larger structural rearrangements.

Because simulated annealing can converge to different solutions depending on its starting point, the optimizer was executed multiple times using different random initializations. Duplicate solutions representing simple rotational shifts of the same loop were removed, and the highest-scoring unique progressions were retained for analysis. The final optimized progressions were then compared against four baseline ordering strategies (*Random*, *In Context*, *Single*, and *Heuristic*) using both overall musicality scores and the structural similarity metric described above.

3.6. Human Perception Study

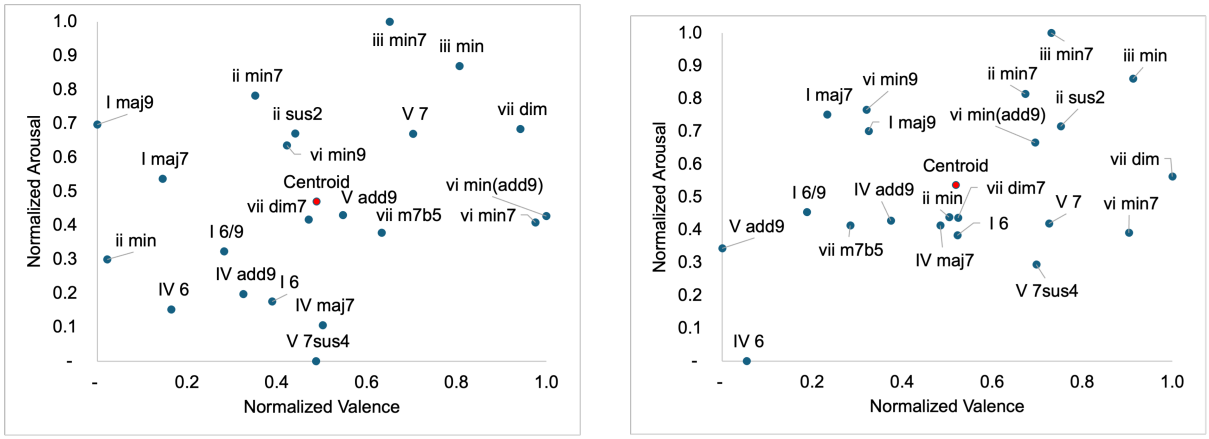
To validate the computational findings, a human evaluation study was conducted using an online survey and was separated into two stages. In the first stage, participants listened to 10 chord transitions evenly drawn from some of the best and worst scoring transitions in the baseline chord progressions, and rated each transition on a seven-point smoothness scale, where higher values indicated greater perceived musical coherence and continuity to see if human perception of musicality aligned with the computational framework.

In the second stage, participants evaluated 10 selected chords drawn from the 21-chord vocabulary. For each chord, participants rated perceived emotional qualities using two seven-point Likert scales

corresponding to energy (arousal) and pleasantness (valence). This assessment examined whether listeners perceived emotional relationships similar to those predicted by the embedding-based analysis. These ratings were used to compare human judgments of transition quality against the outputs of the computational musicality metrics and optimization procedures.

4. Results

4.1. Emotional Embedding Structure



(a) Full-cycle in context chord audio embeddings normalized in the valence and arousal planes.

(b) Single chord audio embeddings normalized in the valence and arousal planes.

Figure 3: Comparison of CLAP chord embedding structures.

The valence–arousal coordinates extracted from the synthesized chord audio revealed a coherent but imperfect emotional structure across the latent space. As shown in the emotional trajectory visualization (Figure 2), the chord embeddings exhibited a general cyclical progression through the affective plane, although the resulting geometry did not form a perfectly circular distribution. Instead, most coordinates were concentrated toward the lower-left region of the raw valence–arousal space in both the isolated single-chord condition and the in-context progression condition. Using the original CLAP output range of [-1,1], the embeddings demonstrated substantially greater variation in arousal than valence, demonstrating that the model was more sensitive to energetic qualities than positive or negative affective qualities within the synthesized audio.

In Figure 3, the normalized embeddings appear well dispersed across the latent space with no major outlier chords. The centroids of both embedding sets (in-context and single) were located near the center of the normalized plane. Visually, the in-context embeddings appeared slightly more evenly distributed than the isolated chord embeddings.

Distinct clustering behavior also emerged between chord categories. In contrast to previous work in affective associations with chord qualities, minor-based harmonies tended to cluster toward the upper-right region of the normalized plane, corresponding to relatively high arousal and high valence, while major harmonies clustered closer to the lower-center region with generally lower arousal values. These distributions differed substantially from the original heuristic mapping and established research which shows that typically minor chords are associated with low valence emotions.

4.2. Computational Musicality Evaluation and Optimization

The emotional chord orderings were next evaluated using the proposed rule-based musicality scoring system. Table 1 summarizes the resulting musicality scores for the chord progressions optimized for musicality and similarity, as well as the baseline chord progressions.

Table 1
Comparison of Baseline and Optimized Chord Progressions using Musicality and Similarity Metrics

Progression	Avg. Total Score	Avg. Motion	Avg. Common Tone	Avg. Tendency	Avg. Similarity
Optimized #1	4.35	0.93	1.45	1.29	0.69
Optimized #2	4.34	0.94	1.51	1.21	0.68
Optimized #3	4.33	0.93	1.42	1.29	0.69
Random	3.32	**0.89**	1.07	0.86	0.50
In Context	3.77	0.88	1.04	0.86	1.00
Single	**3.84**	0.88	0.96	**1.00**	1.00
Heuristic	3.36	**0.89**	**1.13**	0.71	0.62

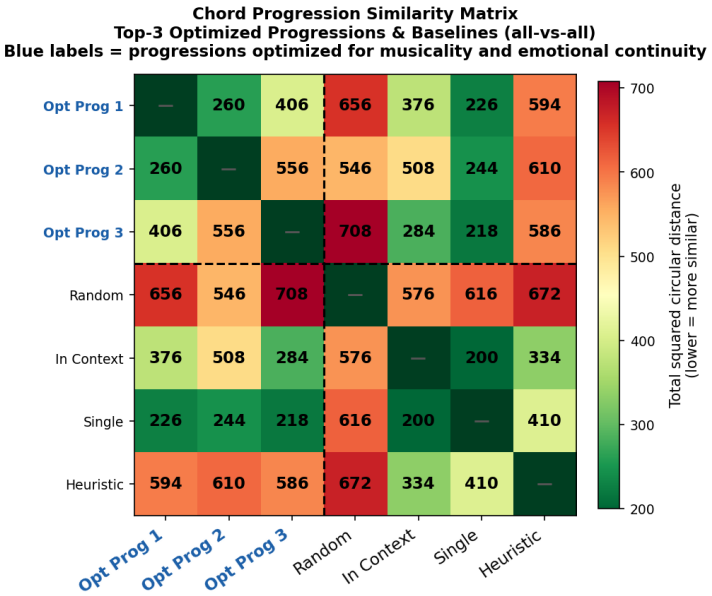


Figure 4: Comparison of the sequential order of chords across baseline and musicality-optimized chord progressions evaluated using the sum of squared circular index distance

Among the baseline approaches, the *Single* progression achieved the highest overall musicality score, followed closely by the *In Context* progression. However, this advantage was largely driven by the structural similarity component included in the total score. When considering only the voice-leading components (motion distance, common-tone retention, and tendency-tone resolution), the *Single* progression performed only marginally better than the *Random* baseline. The *Heuristic* progression produced mixed results. Although it achieved the strongest performance for both average motion distance and common-tone retention, its overall score was reduced by comparatively weak tendency-tone resolution. Conversely, the *Single* progression demonstrated the strongest tendency-tone behavior, suggesting that emotionally grouped chord orderings may naturally preserve some aspects of functional harmonic movement even when not explicitly optimized for voice-leading quality. These differing outcomes highlight that the musicality metrics capture multiple dimensions of harmonic organization that do not necessarily align. Some progression strategies favored smoother pitch movement, while others better preserved harmonic resolution tendencies or emotionally coherent ordering. This suggests that emotional coherence and harmonic coherence may represent related but distinct organizational principles within harmonic space.

The simulated annealing optimization process substantially improved overall musicality scores relative to all baseline progressions. Notably, the top three optimized solutions achieved nearly identical total scores despite having visibly different circular chord orderings, as illustrated in Figure 4. This indicates that multiple high-quality harmonic organizations exist within the constraints of the scoring

framework rather than a single globally optimal solution. Although the optimized progressions differed structurally from one another, they consistently maintained higher similarity scores than the *Random* and *Heuristic* baselines. This demonstrates that the optimization framework was able to discover progressions that balanced both musical smoothness and emotional structural coherence, producing chord organizations that were simultaneously more musically connected and more aligned with the emotionally informed reference structures.

Figure 4 further illustrates the structural relationships between the baseline and optimized chord progressions using circular index distance. The three optimized progressions were structurally similar to one another overall, yet each maintained distinct ordering patterns, demonstrating that the optimization framework admits multiple valid high-scoring solutions rather than converging toward a single harmonic arrangement. Interestingly, some optimized progressions were more structurally similar to the emotionally derived *In Context* and *Single* progressions than to other optimized solutions. For example, the distance between *Opt Prog 2* and *Opt Prog 3* was greater than the distance between either progression and the emotionally informed baseline orderings. This suggests that several structurally diverse chord organizations can still satisfy both the musicality and emotional coherence objectives of the optimization framework.

Among the baseline approaches, the *Single* and *In Context* progressions exhibited the strongest structural similarity to one another, indicating that both emotional embedding strategies produced relatively consistent large-scale harmonic organization. In contrast, the *Random* progression showed substantially weaker similarity to all other approaches, as expected from a non-structured ordering. The *Heuristic* progression occupied an intermediate position, displaying greater similarity to the emotionally derived progressions than to the random baseline. This suggests that the heuristic ordering strategy captures some degree of emotional or harmonic structure that is also reflected within the audio embedding space, despite not being explicitly optimized using the embedding representations.

Of the optimized solutions, *Optimized #3* demonstrated the strongest overall similarity to the emotionally derived baseline progressions. This indicates that the optimization process was capable of discovering chord organizations that not only improved computational musicality metrics, but also preserved structural characteristics present in the original emotion-based embedding arrangements.

4.3. Human Perceptual Evaluation

A total of 20 participants completed the perceptual evaluation study, of which 17 participants identified as musicians with a median of 10 years of musical training, while 3 participants reported no formal musical background. To assess the consistency of participant ratings, Intraclass Correlation Coefficients (ICC(2,1)) were computed for the three perceptual evaluation tasks: musicality ratings of chord transitions, arousal ratings of individual chords, and valence ratings of individual chords. The resulting agreement levels were generally low, indicating substantial variability in how participants perceived and interpreted the musical material. Such variability is expected in studies involving subjective judgments of emotional affect and musical preference, where personal listening history, training, and emotional associations can strongly influence responses.

The musicality transition ratings showed essentially no inter-rater agreement ($\text{ICC}(2,1) = 0.0026$, 95% CI $[-0.02, 0.09]$, $p = 0.3865$), suggesting that participants differed considerably in how they evaluated the smoothness or coherence of chord transitions. Emotional ratings demonstrated slightly stronger consistency, particularly for arousal perception. Arousal ratings produced the highest agreement among the three tasks ($\text{ICC}(2,1) = 0.2492$, 95% CI $[0.11, 0.55]$, $p < 0.001$), while valence ratings remained comparatively weak but statistically above chance ($\text{ICC}(2,1) = 0.0607$, 95% CI $[0.01, 0.22]$, $p = 0.0012$).

4.3.1. Musicality Ratings

As shown in Table 2, the average perceptual musicality rating across all chord transitions was 4.35 (out of a scale from 1 to 7), indicating that the selected transitions were generally perceived as more musical than not, but overall were not significantly correlated with the computational musicality score of each

of the selected transitions.

Table 2
Comparison of Computational Musicality Scores and Perceptual Ratings for Chord Transitions

Evaluated Chord Transition	Computational Musicality Score	Perceptual Score
I6 → viidim	0.75	4.60
iiimin7 → iimin	1.29	4.40
viidim → iiimin	1.54	4.45
I6/9 → iimin	1.59	4.90
iimin7 → Imaj9	1.88	4.55
V7 → iiimin	3.38	3.80
iisus2 → vimin9	3.92	4.55
Imaj9 → Imaj7	3.92	3.90
Imaj7 → I6/9	3.96	4.40
vimin9 → IV6	4.63	4.65

Several individual transitions demonstrated strong disagreement between the computational and perceptual evaluations. The transition Imaj9 ⟶ Imaj7 received one of the highest computational musicality scores due to its extensive common-tone overlap, however it received one of the lowest perceptual ratings with an average score of 3.9, below the overall mean. Conversely, the transition I6/9 ⟶ iimim received one of the lowest computational scores amongst the set of perceptually-evaluated chord transitions, but achieved the highest perceptual rating with an average score of 4.9.

4.3.2. Valence–Arousal Ratings

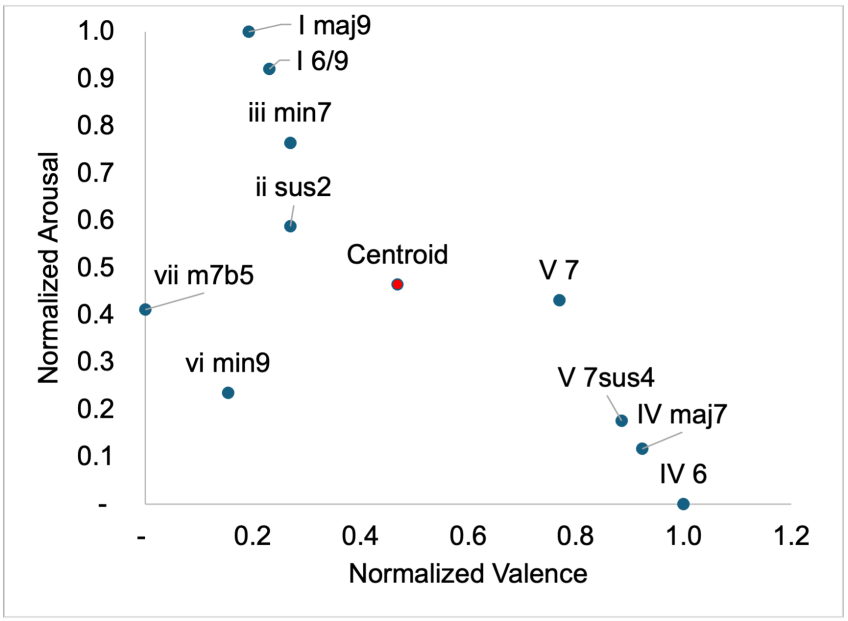


Figure 5: Perceptual ratings of 10 chords for pleasantness (valence) and energy (arousal), mapped into normalized valence and arousal planes.

Participants also evaluated 10 chords according to perceived pleasantness (valence) and energy (arousal). When mapped into a normalized valence–arousal plane as shown in Figure 5, clusters of minor chords formed in the lower valence space and major chord clusters in the higher valence regions, which aligns with previous music psychology research and grounds the validity of the perceptual evaluation.

Euclidean distance comparisons between the perceptual coordinates and the CLAP-derived coordinates in Table 3 further indicated that the perceptual mapping of chords largely differed from the

Table 3
Euclidean Distance Between Perceptual Ratings and Embedding-Based Emotional Coordinates

Chord	Perceptual vs In Context	Perceptual vs Single
V 7sus4	0.43	0.22
IV maj7	0.42	0.53
I maj9	0.36	0.33
ii sus2	0.19	0.50
iii min7	0.45	0.52
vi min9	0.48	0.56
vii m7b5	0.63	0.28
IV 6	0.85	0.95
I 6/9	0.60	0.47
V 7	0.25	0.04
Average Distance	0.47	0.44

embedding-based emotional coordinates, but the Single chord embeddings more closely aligned with the perceptual ratings than the In Context chord embeddings (with a marginal difference of 0.03).

5. Discussion

This study investigated whether it is possible to design a chord-to-emotion mapping that is both emotionally informed and musically grounded for use within *MoodSwing*, an interactive harmonic exploration feature within the Musitopia platform. To address this question, the work combined three complementary approaches: audio embedding-based emotional modeling, rule-based harmonic evaluation, and computational optimization. Together, these approaches explored how harmonic emotion representations can support interactive musical interfaces that allow users to creatively navigate emotional expression through music.

5.1. Emotional Organization in Audio Embedding Space

One of the central findings of this work is that emotionally meaningful harmonic structure can emerge from audio embedding models without explicitly encoding symbolic music theory relationships. Both the *Single* and *In Context* CLAP embedding approaches produced coherent emotional organizations of chords despite relying solely on learned audio representations.

Although the resulting latent spaces did not form perfectly circular emotional geometries, the embeddings nonetheless demonstrated stable large-scale structure across both isolated and contextual listening conditions. The relatively low circular index distance between the *Single* and *In Context* progressions suggests that many chords retained consistent emotional positioning even when embedded within larger harmonic sequences. This indicates that the embedding space captured persistent harmonic-affective relationships rather than responding purely to local contextual variation. At the same time, some harmonies exhibited substantially greater positional shifts between the two embedding strategies, suggesting that certain chord qualities derive a stronger portion of their emotional identity from surrounding harmonic context. This finding supports the broader idea that harmonic emotion is not solely an intrinsic property of isolated chords, but also emerges through sequential and contextual relationships.

Interestingly, chord qualities traditionally associated with negative affect, such as minor and diminished harmonies, did not consistently occupy low-valence regions within the embedding-derived emotional planes. In contrast, these relationships appeared more clearly within the perceptual evaluation results. Many extended minor harmonies instead clustered within relatively high-valence and high-arousal regions. This suggests that the audio embedding model may have been influenced more strongly by timbral qualities, chord density, voicing complexity, or spectral similarity than by culturally learned symbolic harmonic associations.

Because this work intentionally focused on a vocabulary of extended jazz harmonies, future studies may benefit from interval-based analyses or alternative harmonic representations to better isolate which musical features most strongly influence emotional positioning within embedding spaces.

5.2. Balancing Emotional Coherence and Musicality

A key contribution of this study is the finding that emotional coherence and harmonic coherence appear to represent partially competing organizational objectives within harmonic space. The emotionally derived chord orderings produced coherent affective structure, but they did not achieve the highest scores under the rule-based musicality framework. Conversely, the optimization pipeline was able to substantially improve computational musicality while still preserving structural similarity to the emotionally informed embedding arrangements.

Importantly, the optimization results did not converge toward a single ideal progression. Multiple structurally distinct chord orderings achieved nearly identical musicality scores while maintaining relatively strong similarity to the emotional embedding-derived progressions. This suggests that there may not be one most optimal emotionally coherent harmonic pathway, but rather a diverse family of musically plausible emotional trajectories.

This observation is particularly important for the design goals of *MoodSwing*. An emotionally navigable musical interface should not constrain users to a single optimized emotional pathway, but instead support exploration across many valid harmonic trajectories that balance emotional expression and musical smoothness differently. The variability observed in the perceptual evaluation further reinforces this interpretation. Listener responses demonstrated substantial disagreement regarding both emotional perception and harmonic smoothness, suggesting that musical affect and musicality are inherently multidimensional and listener-dependent phenomena. As a result, designing emotionally expressive harmonic systems becomes less a problem of finding a universally optimal solution and more a problem of balancing competing perceptual and musical constraints. In this sense, the design of emotionally informed harmonic interfaces can be understood as a multi-objective optimization problem in which affective organization, harmonic coherence, and perceptual preference coexist as complementary but sometimes conflicting design goals.

5.3. Human Perception and the Limitations of Rule-Based Musicality

The perceptual evaluation revealed substantial disagreement between the computational musicality metrics and listener judgments. The weak negative correlation between the calculated transition scores and the human ratings suggests that the rule-based framework did not fully capture how listeners perceive harmonic smoothness or musical coherence.

Several examples illustrated this mismatch. Transitions with extensive common-tone overlap frequently received high computational scores despite being perceived by listeners as static, repetitive, or musically uninteresting. Conversely, some transitions with relatively little pitch overlap received favorable perceptual ratings due to implied functional relationships, stylistic familiarity, or contextual harmonic interpretation not represented within the scoring model.

These findings highlight the limitations of reducing musicality to a small collection of symbolic heuristics. Human harmonic perception appears to depend not only on pitch movement and tendency-tone resolution, but also on factors such as voicing interpretation, timbral continuity, implied functional expectation, stylistic familiarity, and culturally learned listening patterns.

At the same time, the relatively strong performance of random chord orderings suggests that extended jazz harmonies form a densely interconnected harmonic space in which many transitions satisfy basic smoothness constraints. This may partially explain why some emotionally optimized and randomly generated progressions achieved comparable rule-based musicality scores despite producing very different perceptual experiences.

More fundamentally, the findings raise questions about the extent to which emotional affect can be meaningfully assigned to isolated harmonic objects. The present study evaluated emotional associations

using individual chords and short chord transitions presented outside of a broader musical context. While this simplification was necessary to construct and evaluate the harmonic space, emotional responses to music are rarely determined by harmony alone. The affective character of a chord emerges through its relationship to preceding and subsequent musical events, its function within a progression, performance characteristics such as timbre and dynamics, and the cultural and stylistic conventions through which listeners interpret musical meaning. As a result, emotional mappings should not be understood as fixed properties of harmonic structures but rather as probabilistic tendencies that may vary across listeners, musical traditions, and listening contexts.

The substantial variability observed in participant responses is therefore not simply a limitation of the experimental design but may reflect a fundamental characteristic of musical emotion itself. Listeners bring different musical backgrounds, cultural experiences, preferences, and learned associations that shape how they interpret harmonic material. Consequently, any emotional harmonic space represents a snapshot of collective tendencies within a particular population rather than a universal model of emotional meaning. The value of such mappings lies not in establishing definitive emotional labels for chords, but in providing a structured framework through which users can explore and navigate emotionally suggestive harmonic relationships.

These results suggest that emotional associations with harmony are inherently contextual and cannot be fully generalized from isolated chord presentations. Future versions of the optimization framework may therefore benefit from modeling emotion across longer musical timescales, incorporating phrase-level and progression-level context rather than individual harmonic events alone. Rather than relying exclusively on symbolic harmonic heuristics—which may introduce biases derived from specific Western theoretical traditions—future systems could integrate learned perceptual models, sequential contextual evaluation, listener-informed preference modeling, and culturally diverse training data. Such approaches may better capture the dynamic and situated nature of musical emotion while aligning computational optimization more closely with human musical experience.

5.4. Design Implications for Affective Musical Interfaces

The broader contribution of this work lies in its implications for the design of affective musical interfaces such as *MoodSwing* for Musitopia. By mirroring the continuous, non-discrete nature of emotional experiences such as the circumplex model, this study explores harmonic relationships as trajectories through an emotionally organized latent space that users can actively navigate.

This directly informs the development of affective musical interfaces like *MoodSwing* within the Musitopia platform, where chords are positioned spatially on an interactive canvas according to learned emotional relationships. Within such a system, it is crucial to adequately capture the affective nature of music through harmonic movement while simultaneously managing within musical coherence constraints shaped by optimization and perceptual design. The findings suggest that emotionally informed harmonic organization is computationally achievable, but that emotional organization alone is insufficient for producing musically convincing harmonic motion.

Likewise, optimizing exclusively for musical smoothness risks collapsing emotionally meaningful structure. Effective affective music interfaces therefore require hybrid approaches that integrate emotional embeddings, perceptual modeling, and harmonic optimization simultaneously.

More broadly, this work demonstrates how computational design approaches can support new forms of emotionally guided music interaction. Systems such as *MoodSwing* may ultimately enable users without formal music theory training to engage with harmony as an expressive emotional medium, opening new possibilities for music therapy, emotional self-expression, adaptive composition, and interactive creative exploration.

6. Conclusion

This study explored whether chord-to-emotion mappings can be designed in a way that is simultaneously emotionally informed and musically grounded for use within the *MoodSwing* feature of Musitopia. To in-

vestigate this question, the work combined audio embedding-based emotional modeling, computational optimization, rule-based harmonic evaluation, and perceptual listener analysis.

The results demonstrate that CLAP audio embeddings are capable of generating coherent emotional organizations of harmonic material without explicitly encoding symbolic music theory rules. Furthermore, the comparison between isolated chord embeddings and contextual chord embeddings revealed that many emotional harmonic relationships remain stable across listening conditions, while some harmonies exhibit strong contextual sensitivity. At the same time, the study found that emotional coherence and harmonic smoothness represent partially competing objectives. Progressions derived directly from emotional embedding spaces did not achieve the strongest musicality scores, while progressions optimized for musical smoothness risked diverging from emotionally coherent structures. However, the optimization design framework proposed here demonstrated that these objectives can be balanced simultaneously, producing chord progressions that improve harmonic coherence while preserving emotionally meaningful organization.

The perceptual evaluation further revealed that computational musicality metrics do not fully align with human judgments of harmonic smoothness or emotional interpretation. These findings highlight the limitations of purely rule-based musicality systems and suggest that future affective music interfaces may require more perceptually grounded evaluation strategies.

Future work will expand the perceptual evaluation using a larger and more musically diverse participant pool in order to obtain more stable estimates of emotional perception and musical preference. Additional work may also explore alternative audio embedding models [**?**], different instrumental timbres, interval-based harmonic representations, and broader temporal structures involving rhythm and melody.

Ultimately, this research contributes toward the development of interactive systems that allow users to navigate music through emotion rather than formal music theory alone. By combining affective embeddings, harmonic optimization, and perceptual design principles, future iterations of *MoodSwing* may support richer forms of emotionally guided musical creativity and exploration.

Acknowledgments

This paper is part of Maria de Maeztu Units of Excellence Programme (CEX2021-001195-M), IMPA project PID2023-152250OB-I00 funded by MCIU/AEI/10.13039/501100011033/FEDER, UE, Musitopia project, funded by the Fundació Barcelona Music Lab, and supported by Music Tech Europe Academy.

Declaration on Generative AI

During the preparation of this work, the author(s) used Claude and ChatGPT in order to perform grammar and spelling checks. Further, the author(s) used Claude for code to generate figures. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. Micallef Grimaud, T. Eerola, Emotional expression through musical cues: A comparison of production and perception approaches, PLOS ONE 17 (2022) 1–24. URL: https://doi.org/10.1371/journal.pone.0279605. doi:10.1371/journal.pone.0279605.

[2] M. Ortega, R. Lucena, R. Ramírez, Musitopia: Bridging human expertise and ai for digital music therapy, 2026, pp. 22077–22081. doi:10.1109/ICASSP55912.2026.11464753.

[3] H.-C. Weng, L.-Y. Huang, L. Imcha, et al., Drawing as a window to emotion with insights from tech-transformed participant images, Scientific Reports 14 (2024) 11571. URL: https://doi.org/10.1038/s41598-024-60532-6. doi:10.1038/s41598-024-60532-6.

[4] J. Posner, J. A. Russell, B. S. Peterson, The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology, Development and Psychopathology 17 (2005) 715–734. doi:10.1017/S0954579405050340.

[5] C. Labbé, W. Trost, D. Grandjean, Affective experiences to chords are modulated by mode, meter, tempo, and subjective entrainment, Psychology of Music 49 (2021) 915–930. URL: https://doi.org/10.1177/0305735620906887. doi:10.1177/0305735620906887. arXiv:https://doi.org/10.1177/0305735620906887.

[6] B. Sollberger, R. Reber, D. Eckstein, Musical chords as affective priming context in a word-evaluation task, Music Perception - MUSIC PERCEPT 20 (2003) 263–282. doi:10.1525/mp.2003.20.3.263.

[7] N. Steinbeis, S. Koelsch, Affective priming effects of musical sounds on the processing of word meaning, Journal of Cognitive Neuroscience 23 (2011) 604–621. doi:10.1162/jocn.2009.21383.

[8] F. G. Pedrosa, song_sent_scores: Computational design for charting dynamic emotion in songs with a multimodal circumplex framework (2025).

[9] P. N. Juslin, Handbook of Music and Emotion: Theory, Research, Applications, Oxford University Press, 2010.

[10] T. Eerola, J. K. Vuoskoski, A comparison of the discrete and dimensional models of emotion in music, Psychology of Music 39 (2011) 18–49. URL: https://doi.org/10.1177/0305735610362821. doi:10.1177/0305735610362821. arXiv:https://doi.org/10.1177/0305735610362821.

[11] A. Gabrielsson, E. Lindström, The role of structure in the musical expression of emotions, Handbook of music and emotion: Theory, research, applications (2010) 367–400. doi:10.1093/acprof:oso/9780199230143.003.0014.

[12] I. Lahdelma, T. Eerola, Single chords convey distinct emotional qualities to both naïve and expert listeners, Psychology of Music 44 (2016) 37–54.

[13] L. B. Meyer, Emotion and Meaning in Music, University of Chicago Press, Chicago, IL, 1956.

[14] I. Lahdelma, J. Armitage, T. Eerola, Affective priming with musical chords is influenced by pitch numerosity, Musicae Scientiae 26 (2020). doi:10.1177/1029864920911127.

[15] J. Zhang, L. Li, L. Wei, H. Wang, Moderating effects of chord progressions on the emotional experience of major and minor chords, Acta Psychologica 253 (2025) 104690. doi:10.1016/j.actpsy.2025.104690.

[16] I. Wallis, T. Ingalls, E. Campana, J. Goodman, A rule-based generative music system controlled by desired valence and arousal, in: Proceedings of 8th international sound and music computing conference (SMC), 2011, pp. 156–157.

[17] S. K. Ehrlich, K. R. Agres, C. Guan, G. Cheng, A closed-loop, music-based brain-computer interface for emotion mediation, PLOS ONE 14 (2019) 1–24. URL: https://doi.org/10.1371/journal.pone.0213516. doi:10.1371/journal.pone.0213516.

[18] K. Agres, A. Dash, P. Chua, Affectmachine-classical: A novel system for generating affective classical music, 2023. doi:10.48550/arXiv.2304.04915.

[19] A. Dash, K. Agres, Ai-based affective music generation systems: A review of methods and challenges, ACM Computing Surveys 56 (2024). doi:10.1145/3672554.

[20] B. Elizalde, S. Deshmukh, M. Al Ismail, H. Wang, Clap learning audio concepts from natural language supervision, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.

[21] H.-H. Wu, O. Nieto, J. P. Bello, J. Salamon, Audio-text models do not yet leverage natural language, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.

[22] R. Hutchinson, Music theory for the 21st-century classroom, 2021. URL: https://musictheory.pugetsound.edu/mt21c/.