Exploring the Impact of Dynamic Tempo and Loudness Changes on Listeners' Perceived Emotion in Music

Siheng Fang,[1] Ève Poudrier[2]

[1] University of British Columbia, Faculty of Arts, Cognitive Systems Program
[2] University of British Columbia, School of Music
simonf07@student.ubc.ca

Abstract. This study examined how specific phrasing patterns (defined as dynamic changes in tempo and loudness) affect perceived emotional responses. Participants (N = 27) listened to controlled musical excerpts featuring eight different combinations of tempo and loudness and rated each excerpt's perceived expressiveness, energy (arousal), mood (valence), and preference relative to a baseline excerpt with no tempo or loudness changes. Results showed that forward phrasing patterns (increasing tempo and loudness towards climax, decreasing tempo and loudness after climax) led to higher perceived energy and more positive perceived mood, while reverse phrasing (decreasing tempo and loudness towards climax, increasing tempo and loudness after climax) led to lower perceived energy and more negative perceived mood. While matching phrasing patterns (tempo and loudness changes in the same direction) amplified emotional responses, contrasting patterns (e.g., increasing tempo with decreasing loudness towards climax, decreasing tempo with increasing loudness after climax) also yielded higher expressiveness ratings, suggesting that emotional perception may involve more complex processes than a simple addition of tempo and loudness effects. Moreover, tempo exerted a stronger influence than loudness on arousal and valence, contradicting earlier claims that loudness more strongly influences emotional perception.

Keywords: music cognition, perceived emotion, music phrasing patterns, tempo and loudness interaction, dynamic music features

1 Introduction

Music has the ability to evoke and influence human emotions. Even without formal training, most listeners can hear a piece of music and intuitively perceive feelings such as happiness, sadness, and excitement (Baird et al., 2018; Bresin & Friberg, 2011; Hunter & Schellenberg, 2010; Jäncke, 2008; Kamenetsky et al., 1997). Researchers in the field of music cognition seek to understand which features of music shape these emotional responses. Features such as tempo (how fast the music unfolds), loudness (how loud the music is performed), mode, and rhythm, have each been explored as potential factors influencing listeners' perceived emotions (Bresin & Friberg, 2011; Hunter & Schellenberg, 2010; Panda et al., 2020; Poudrier et al., 2023). However, there is a very limited number of studies that investigate the interplay between musical features, that is, its dynamic unfolding.

1.1 Musical Expressiveness

The expressiveness of music lies in how different auditory features are presented and how they vary over time. Fernández-Sotos et al. (2016) demonstrated that changes in tempo and rhythmic units modulate emotional arousal and valence. Higher tempos and complex rhythms (e.g., smaller beat subdivisions, such as sixteenth notes in a quarter-note based meter) increase arousal and evoke emotions such as happiness and surprise, while slower tempos and simpler rhythms produce lower arousal states, including calmness (positively valenced) or sadness (negatively valenced).

Kamenetsky et al. (1997) examined the effects of tempo and loudness on emotional expressiveness in music. Their study revealed that emotional expressiveness ratings were highest when both tempo and loudness varied, indicating the critical role of a dynamic interplay between these features in shaping perceived emotion. Interestingly, ratings were higher for variations in loudness alone compared to tempo alone, suggesting that listeners may be more sensitive to changes in volume than speed in conveying emotional nuance.

1.2 Phrasing Patterns and Perceived Intensity

The term "phrasing patterns," as defined in Bresin and Friberg's (2011) study, refers to the dynamic variations in tempo and loudness that are often employed by musicians during performance. Bresin and Friberg propose that forward phrasing – characterized by accelerando (increasing tempo) and crescendo (increasing loudness) followed by rallentando (decreasing tempo) and decrescendo (decreasing loudness) – is associated with low arousal emotions such as sadness or tenderness. Conversely, they argue that reverse phrasing, which begins with rallentando and decrescendo and then shifts to accelerando and crescendo, enhances high arousal states (e.g., aggression).

However, this claim appears inconsistent with research on musical climax and intensity perception (Patty, 2009; Kamenetsky et al., 1997; Fernández-Sotos et al., 2016). Studies indicate that increasing tempo and loudness typically build tension and raise arousal, while decreasing tempo and loudness reduce intensity and lead to lower arousal states. Musical climaxes, as described by Patty (2009), involve a buildup of energy followed by resolution, suggesting that forward phrasing should actually lead to higher arousal, contrary to Bresin and Friberg's claim. If forward phrasing naturally aligns with increasing arousal due to its accelerative and intensifying characteristics, then reverse phrasing should create lower arousal due to its decelerative and de-intensifying nature.

This study will test this hypothesis, aiming to challenge Bresin and Friberg's interpretation. The pattern of accelerando followed by rallentando is referred to as "forward tempo," while the pattern of rallentando followed by accelerando is referred to as “reverse tempo.” Similarly, the pattern of crescendo followed by decrescendo is identified as “forward loudness” and the pattern of decrescendo followed by crescendo as “reverse loudness.” To disentangle the influence of tempo and loudness, this study uses matched and mismatched phrasing patterns. Based on the naming convention, matched patterns will involve a combination of forward tempo and forward loudness or a combination of reverse tempo and reverse loudness, while mismatched patterns involve opposite directions in tempo and loudness (forward tempo and reverse loudness).

1.3 Research Question and Hypotheses

The research question of this project is *how specific combinations of dynamic tempo and loudness patterns influence listeners' perceived emotion*. We hypothesize that: (1) forward phrasing patterns will lead to higher arousal and reverse phrasing patterns will lead to lower arousal; (2) reverse phrasing will be perceived as more negative, while forward phrasing be perceived as more positive; (3) the effects of tempo and loudness will exhibit an additive relationship; and (4) loudness changes will more strongly influences the effect on emotional perception over tempo changes.

The first hypothesis is based on the findings of research on musical climax and intensity perception (e.g., Fernández-Sotos et al., 2016), and thus challenges the findings of Bresin & Friberg (2011). The second hypothesis is based on Patty (2009)'s study which suggests decreasing intensity towards climax will increase unpredictability, which then leads to more negative valence. Reverse phrasing aligns with this situation as its loudness and tempo both decrease towards climax.

The “additive relationship” in the third hypothesis means when tempo and loudness changes are matched (forward or reverse phrasing in both tempo and loudness), we expect to see an amplification of the emotional effect as compared to changes in tempo or loudness only (e.g., forward tempo and neutral loudness). If the relationship is linear, it can be speculated that mismatched patterns (e.g., forward tempo and reverse loudness) may cancel each other out and result in a neutral or ambiguous perception. Finally, the fourth hypothesis is based on the study findings of Kamenetsky et al. (1997) on the effects of tempo and loudness on emotional expressiveness.

2 Methods

2.1 Participants

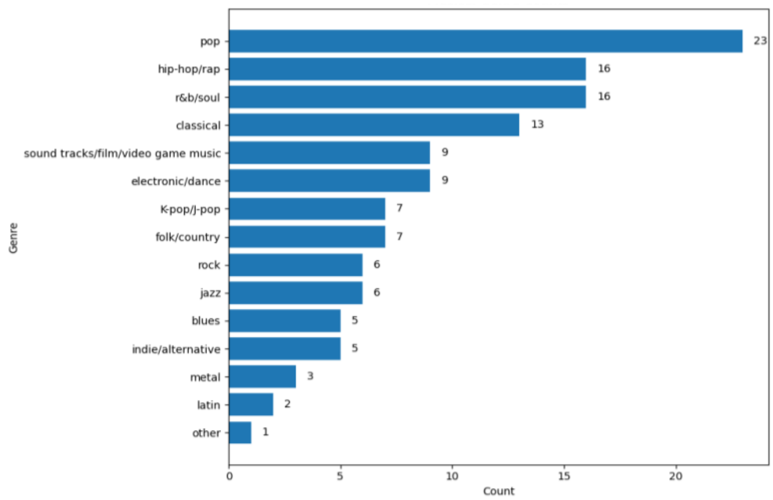
We used a survey to collect self-reported data from participants. The survey was approved by the Human Ethics Research Board of the University of British Columbia (UBC). Both passive and active recruitment techniques were used. Recruitment methods consisted of flyers and announcements in classes, as well as posters placed in various locations around UBC campus, including libraries, student centers, and lecture halls. A total of 42 participants' responses were collected. Participants were excluded if they did not complete at least 98% (the 2% incomplete is due to an optional raffle question at the end, not affecting the main survey report) of the experiment or reported low English fluency (i.e., reported "I have difficulty understanding English"). The final sample consisted of 27 participants' responses (male = 12, female = 14). One participant selected "prefer not to say" for gender; there were no participants that self-identified as non-binary.

Participants' age ranged from 17 to 35 years, with a mean age of 23.67 years (*SD* = 4.8; *Mdn* = 23). Participants also reported their years of formal musical training, ranging from 0 to 23, with a mean number of years of 4.64 (*SD* = 6.36; *Mdn* = 2). A total of 21 participants reported high English fluency, while six indicated moderate fluency. Participants also reported their familiarity with a wide range of musical genres. As shown in Figure 1, the most frequently selected genre was Pop (85%), followed by Hip-hop/Rap and R&B/Soul (59% each), and Classical (48%). Other popular choices included Soundtracks/Film/Game music, Electronic/Dance, K-pop/J-pop, and Folk/Country. These results suggest that participants brought diverse musical backgrounds to the study, which may have influenced how they perceived and responded to the musical stimuli.

Finally, listening devices used varied across participants. A greater number of participants used external speakers (*N* = 10), while six participants used noise-cancelling headphones or earbuds and another six participants used built-in computer or laptop speakers; four participants reported using standard headphones or earbuds, and one participant selected "other" (iPhone speaker).

Figure 1

Distribution of participants' musical genre preferences.



2.2 Materials

Only one source music excerpt was used across all trials (shown as Figure 2). Eight varied versions of this music piece were created using Logic Pro X in which all other features except tempo and note velocity (loudness) were controlled. The melodic climax shown on the downbeat of the second measure was used to control where tempo and loudness changes occur (from increasing to decreasing, or from decreasing to increasing). In Table 1, a three number sequence represents the values of each pattern's tempo (measured in bpm) and loudness (measured by the note velocity) at the start of the excerpt, climax, and the end of the excerpt. Tempo and loudness each gradually proceed from their value at the start to their value at climax, then they each gradually proceed from their value at climax to their value at the end. The baseline has a consistent tempo of 110 bpm and a consistent note velocity of 80 (-4.02 dB). A 30% change (increase or decrease relative to the baseline's tempo and velocity) at the climax is applied to each variation on its changing musical features.

Figure 2
Musical piece with red line indicating the musical climax (highest melody pitch).

Table 1
Phrasing pattern assignment: the three number sequence represents the values at start-climax-end.

Audio	Condition	Tempo (bpm)	Loudness (velocity)
Baseline	Neutral Tempo + Neutral Loudness	110-110-110	80-80-80
Variation 1	Forward Tempo + Neutral Loudness	110-143-110	80-80-80
Variation 2	Reverse Tempo + Neutral Loudness	110-77-110	80-80-80
Variation 3	Neutral Tempo + Forward Loudness	110-110-110	80-104-80
Variation 4	Neutral Tempo + Reverse Loudness	110-110-110	80-56-80
Variation 5	Forward Tempo + Forward Loudness	110-143-110	80-104-80
Variation 6	Reverse Tempo + Reverse Loudness	110-77-110	80-56-80
Variation 7	Forward Tempo + Reverse Loudness	110-143-110	80-56-80
Variation 8	Reverse Tempo + Forward Loudness	110-77-110	80-104-80

2.3 Procedure

The experimental survey was conducted online using Qualtrics. The order of experimental trials was randomized to control for order effects. Participants were not informed of the specific dynamic phrasing patterns in each trial to minimize bias.

Table 2
Description of measure.

Dependent Variables	Survey Question	Ratings (1~7)
Emotional Expressiveness	Compared to the baseline, how would you rate the emotional expressiveness of the variation?	1: much less expressive 2: moderately less expressive 3: slightly less expressive 4: about the same 5: slightly more expressive 6: moderately more expressive 7: much more expressive
Energy	Compared to the baseline, how would you rate the energy of the variation?	1: much lower 2: moderately lower 3: slightly lower 4: about the same 5: slightly higher 6: moderately higher 7: much higher
Mood	Compared to the baseline, how would you rate the mood of the variation?	1: much more negative 2: moderately more negative 3: slightly more negative 4: about the same 5: slightly more positive 6: moderately more positive 7: much more positive
Preference	Compared to the baseline, how would you rate your preference of the variation?	1: prefer the baseline much more 2: prefer the baseline moderately more 3: prefer the baseline slightly more 4: about the same 5: prefer the variation slightly more 6: prefer the variation moderately more 7: prefer the variation much more

First, participants provided informed consent and read the study instructions. They then completed a pre-experiment questionnaire, where they reported demographic information including age, gender, years of formal musical training, familiarity with different music genres, and English-language fluency. In a post-experiment survey, participants were also asked to report what type of listening device they used to complete the tasks. Participants also had the opportunity to provide their email to participate in a raffle.

Before beginning the experimental trials, participants completed a volume adjustment task where they were instructed to set their playback volume to a comfortable level using a short audio clip. Participants were instructed to keep the volume at the same level throughout the study.

The experiment consisted of eight trials, where participants listened to pairs of short musical excerpts. Each trial contained a baseline excerpt, followed by a varied excerpt that featured a specific dynamic phrasing pattern (see Table 1). Participants could listen to each excerpt as many times as needed to provide their ratings. For each trial, participants rated the perceived emotional expressiveness, energy, mood, and preference of the varied excerpt as compared to the baseline excerpt using four seven-point Likert scales. Table 2 shows a detailed description of the four scales with their associated questions and rating descriptors.

2.4 Measures

The independent variables for this study were patterns of tempo and loudness changes. The dependent variables are perceived emotional expressiveness, arousal (energy), valence (mood), and preference. Participants were asked to rate on seven-point Likert scales on the four dependent variables across the eight trials (n = 32). A rating of "4" means the varied and baseline excerpts are perceived to be about the same on that dependent variable. A higher rating score means the varied excerpt is perceived to be more expressive, more energetic, more positive in mood, or more preferred by the participants as compared to the baseline excerpt. A lower rating score means the varied excerpt is perceived to be less expressive, less energetic, more negative in mood, or less preferred by the participants as compared to the baseline excerpt.

2.5 Analysis

All data were analyzed using Python, with support from pandas, NumPy, statsmodels, SciPy, and Matplotlib libraries.

2.5.1 Condition Coding

After data cleaning (removing identifying information, responses below 98% completion, and response from participants who reported difficulty in reading English), each trial was coded based on the phrasing pattern applied to tempo and loudness (Table 3).

Table 3
Coding scheme of independent variables.

Independent variable	Pattern	Code	Description
Tempo	Neutral Tempo	T0	Stable tempo
	Forward Tempo	T1	accelerando to climax, rallentando afterwards
	Reverse Tempo	T2	rallentando to climax, accelerando afterwards
Loudness	Neutral Loudness	L0	Stable loudness
	Forward Loudness	L1	crescendo to climax, decrescendo afterwards
	Reverse Loudness	L2	decrescendo to climax, crescendo afterwards

The eight comparison trials were then each mapped to a code (Table 4).

Table 4
Code for each trial. Participants were asked to compare the variation with the baseline before rating.

Audio No.	Condition	Code
Variation 1	Forward Tempo + Neutral Loudness	T1L0
Variation 2	Reverse Tempo + Neutral Loudness	T2L0
Variation 3	Neutral Tempo + Forward Loudness	T0L1
Variation 4	Neutral Tempo + Reverse Loudness	T0L2
Variation 5	Forward Tempo + Forward Loudness	T1L1
Variation 6	Reverse Tempo + Reverse Loudness	T2L2
Variation 7	Forward Tempo + Reverse Loudness	T1L2
Variation 8	Reverse Tempo + Forward Loudness	T2L1

2.5.2 Statistical Analysis

A two-way repeated-measures ANOVA was conducted for each dependent variable (emotional expressiveness, energy, mood, and preference), with tempo (T0, T1, T2) and loudness (L0, L1, L2) treated as categorical within-subject factors. Interaction plots were also made to showcase the general trends.

Paired *t*-tests were also conducted, to identify directional relationships between tempo and loudness phrasing patterns. Given the number of comparisons conducted across multiple conditions and dimensions, the analysis involved a substantial number of simultaneous hypothesis tests. To mitigate the increased risk of Type I errors (false positives) associated with multiple comparisons, the Holm-Bonferroni correction was applied to the *p*-values obtained from all paired *t*-tests. Holm-Bonferroni's method provides a stepwise adjustment that controls the familywise error rate (FWER) – the probability of making at least one Type I error across all tests (Chen et al., 2017). Only results that remained significant after Holm-Bonferroni adjustment were interpreted as statistically reliable.

3 Results

Participants rated eight varied excerpts on four seven-point Likert scales. Participants' average rating for emotional expressiveness was the highest ($M = 4.78$; $SD = 1.65$), followed by mood ($M = 4.69$; $SD = 1.59$) and preference ($M = 4.59$; $SD = 1.73$). Participants' average rating of energy was the lowest, with the highest variance across all dependent variables ($M = 4.5$; $SD = 1.81$).

3.1 Influence of Tempo and Loudness

Table 5 presents the results of the two-way ANOVA examining the effects of tempo, loudness, and their interaction on emotional expressiveness, energy, mood, and preference. Across all four dependent variables, both tempo and loudness had a statistically significant main effect on participants' ratings of the four dependent variables. The interaction of tempo and loudness was statistically significant only for emotional expressiveness and mood.

Results are presented in Table 5. Note that the sum of squares values in the table reflect how much variance in each dependent variable is explained by tempo, loudness, and their interaction. Larger values indicate that a factor accounts for more variability in the outcome. The residual sum of squares represents the portion of variance that remains unexplained by the model – essentially the random or individual differences not accounted for by the experimental conditions. Overall, the effect of loudness was stronger on participants' ratings of emotional expressiveness and preference. On the other hand, tempo had a stronger effect than loudness on participants' ratings on energy and mood. Across dependent variables, tempo had the strongest effect on energy and the weakest effect on preference, while loudness had the strongest effect on preference and the weakest effect on emotional expressiveness.

Table 5.
Two-way ANOVA result. Significance level are as follows: '***' p<0.001; '**' p<0.01; '*' p<0.05.

Dependent Variables	**Parameter**	**Sum of Squares**	***df***	**Mean Square**	***F***	***p*-value**
Emotional Expressiveness	Tempo	71.86	2	35.93	13.06	<.001***
	Loudness	93.73	2	46.86	17.03	<.001***
	Tempo x Loudness	30.47	4	7.62	2.77	.028*
	Residual	572.30	208	2.75	–	–
Energy	Tempo	139.51	2	69.75	24.45	<.001***
	Loudness	103.50	2	51.75	18.06	<.001***
	Tempo x Loudness	26.15	4	6.54	2.28	.062
	Residual	595.93	208	2.87	–	–
Mood	Tempo	119.90	2	59.95	26.25	<.001***
	Loudness	95.20	2	47.60	20.85	<.001***
	Tempo x Loudness	29.22	4	7.30	3.20	.014*
	Residual	474.96	208	2.28	–	–
Preference	Tempo	38.20	2	19.10	6.56	.002**
	Loudness	128.88	2	64.44	22.11	<.001***
	Tempo x Loudness	27.18	4	6.80	2.33	.057
	Residual	606.07	208	2.91	–	–

3.2 Influence of Phrasing Patterns

To further examine the combined effects of tempo and loudness on participants' ratings, interaction plots were generated for each dependent variable (emotional expressiveness, energy, mood, and preference). These plots display the average participants' ratings for each tempo and loudness combination.

Standard error bars were included to reflect the variability in responses across participants. Overall, standard error values were moderate, ranging from 0.22 to 0.41. Within the dependent variables, T1L0 (forward tempo + neutral loudness) showed the highest agreement in participants' ratings of energy (*SD* = 1.13), while T1L2 (forward tempo + reverse loudness) showed the highest disagreement in ratings for preference (*SD* = 2.15).

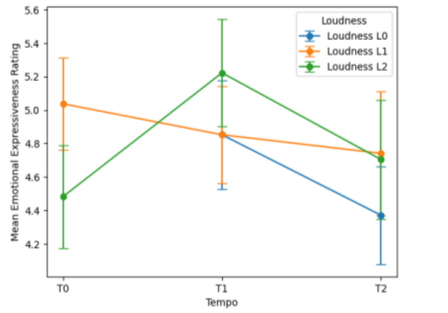
3.2.1 Emotional Expressiveness

Table 6 and Figure 3 presents the results for emotional expressiveness. The average of participants' rating for T1L2 (forward tempo + reverse loudness) on emotional expressiveness was the highest, followed by T0L1 (neutral tempo + forward loudness). Average rating for T2L0 (reverse tempo + neutral loudness) was the lowest.
When the tempo was neutral, forward loudness was rated higher than reverse loudness (T0L1 > T0L2). However, the relationship was reversed when the forward tempo pattern was combined with forward and reverse loudness (T1L1 < T1L2). When the tempo featured a reverse pattern, forward and reverse loudness gave rise to similar average ratings (T2L1 ≈ T2L2).

Table 6
Descriptive statistics for interaction plot of emotional expressiveness.

Code	Mean (*M*)	Standard Deviation (*SD*)	Standard Error (*SE*)
T0L1	5.04	1.43	0.28
T0L2	4.48	1.60	0.31
T1L0	4.85	1.68	0.32
T1L1	4.85	1.51	0.29
T1L2	5.22	1.67	0.32
T2L0	4.37	1.52	0.29
T2L1	4.74	1.93	0.37
T2L2	4.70	1.86	0.36

Figure 3
Interaction plot for emotional expressiveness.



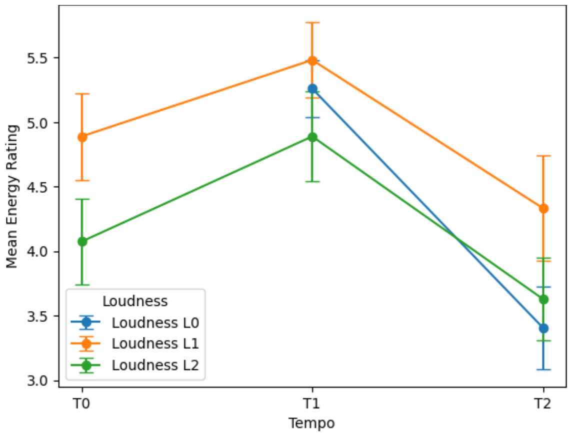
3.2.2 Energy

The results are shown in Table 7 and Figure 4. Participants' ratings were generally higher for excerpts with forward as compared with reverse tempo patterns, and their ratings were generally higher for excerpts with forward as compared with reverse loudness patterns. While the matched forward phrasing pattern (T1L1) showed the highest average rating, and the reverse tempo with neutral loudness (T2L0) showed the lowest average rating. In all tempo conditions, the highest average rating was with forward loudness. Reverse loudness (L2) and neutral loudness (L0) combined with tempo gave rise to an interaction, with T1L2 resulting in a lower average rating than T1L0, but T2L2 resulting in a higher average rating than T2L0, although this difference did not reach statistical significance.

Table 7
Descriptive statistics for interaction plot of energy.

Code	Mean (*M*)	Standard Deviation (*SD*)	Standard Error (*SE*)
T0L1	4.89	1.74	0.34
T0L2	4.07	1.73	0.33
T1L0	5.26	1.13	0.22
T1L1	5.48	1.53	0.29
T1L2	4.89	1.81	0.35
T2L0	3.41	1.67	0.32
T2L1	4.33	2.11	0.41
T2L2	3.63	1.67	0.32

Figure 4
Interaction plot for energy.



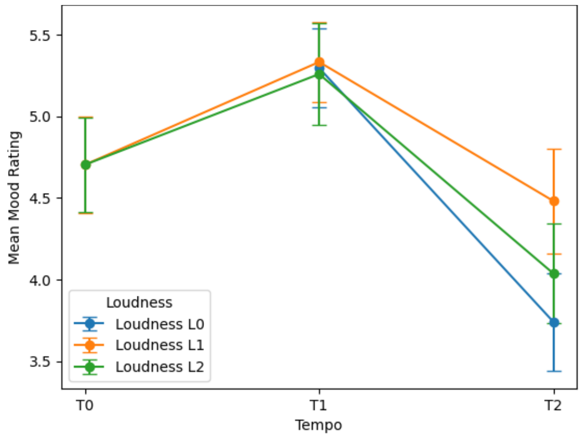
3.2.3 Mood

Table 8 and Figure 5 show that music excerpts with neutral tempo (T0) and forward tempo (T1), loudness changes do not affect participants' ratings of mood. For music excerpts with reverse tempo (T2), the average rating scores vary across the three loudness patterns, with forward loudness (T2L1) resulting in the highest average rating and the neutral loudness (T2L0) resulting in the lowest average rating.

Table 8
Descriptive statistics for interaction plot of mood.

Code	Mean (*M*)	Standard Deviation (*SD*)	Standard Error (*SE*)
T0L1	4.70	1.54	0.30
T0L2	4.70	1.51	0.29
T1L0	5.30	1.27	0.24
T1L1	5.33	1.27	0.24
T1L2	5.26	1.63	0.31
T2L0	3.74	1.56	0.30
T2L1	4.48	1.67	0.32
T2L2	4.04	1.58	0.30

Figure 5
Interaction plot for mood.



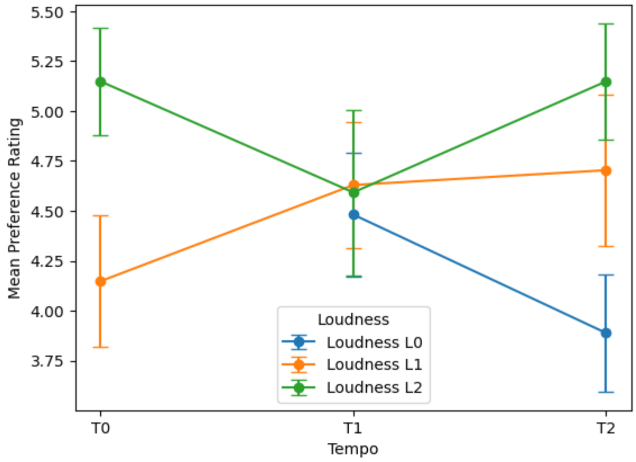
3.2.4 Preference

Table 9 and Figure 6 suggested that when the tempo of the music excerpts was in the forward pattern (T1), loudness did not significantly affect participants' preference ratings. However, when the music excerpts featured neutral tempo (T0) or reverse tempo (T2) were combined with reverse loudness (L2), participants' average ratings were significantly higher.

Table 9
Descriptive statistics for interaction plot of preference.

Code	Mean (M)	Standard Deviation (SD)	Standard Error (SE)
T0L1	4.15	1.70	0.33
T0L2	5.15	1.41	0.27
T1L0	4.48	1.60	0.31
T1L1	4.63	1.64	0.32
T1L2	4.59	2.15	0.41
T2L0	3.89	1.53	0.29
T2L1	4.70	1.98	0.38
T2L2	5.15	1.51	0.29

Figure 6
Interaction plot for preference.



3.3 Paired *t*-tests Result

Table 9 presents only the paired *t*-test comparisons that remained statistically significant after Holm adjustment. These comparisons highlight key differences between specific phrasing patterns across energy, mood, and preference measures.

Table 9
Significant paired *t*-test results after Holm adjustment. Note that a positive *t*-score indicates the left side of the comparison is higher; a negative *t*-score indicates that the right side of the comparison is higher. Significance level are as follows: '***' $p<0.001$; '**' $p<0.01$; '*' $p<0.05$.

Condition Comparison	Dependent Variable	*t*-value	*p*-value	Holm-adjusted *p*-value
T1L0 vs. T2L0	Energy	5.309	0.0000	0.0000***
T2L0 vs. T1L1	Energy	-4.614	0.0001	0.0075**
T1L0 vs. T2L2	Energy	4.494	0.0001	0.0075**
T1L0 vs. T2L0	Mood	4.281	0.0002	0.0146*
T2L0 vs. T0L2	Preference	-4.132	0.0003	0.0216*
T2L0 vs. T1L1	Mood	-4.127	0.0003	0.0216*
T1L1 vs. T2L2	Energy	3.911	0.0006	0.0420*

4 Discussion

This study investigated how phrasing patterns affect perceived emotion in music. Four hypotheses were tested, focusing on their impact on perceived energy (arousal), mood (valence), expressiveness, and listener preference. Below, each hypothesis and its relation to the statistical findings and relevant literature were discussed.

The first hypothesis, which predicts that forward phrasing patterns will lead to higher arousal and reverse phrasing patterns will lead to lower arousal, was well supported by the data. The two-way ANOVA showed a significant main effect of tempo and of loudness on energy ratings, which confirms that both features influenced energy. But the interaction of tempo and loudness showed no significance, which suggest that tempo and loudness contributed independently to perceived energy. Critically, the forward phrasing (T1L1) produced the highest energy rating among all conditions, while the reverse phrasing (T2L2) produced the second lowest energy ratings, as shown in Figure 4. Post hoc paired *t*-tests with Holm adjustment (Table 9) further confirmed this pattern – forward tempo with neutral loudness (T1L0) was rated significantly higher in energy than reverse tempo with neutral loudness (T2L0), and forward phrasing (T1L1) was rated significantly higher than reverse phrasing (T2L2). This evidence strongly supports the notion that forward phrasing builds perceived energy (arousal), consistent with Fernández-Sotos et al. (2016) and Patty (2009).

The second hypothesis predicting that reverse phrasing would be perceived as more negative while forward phrasing would be perceived as more positive was also supported by the data. The two-way ANOVA revealed a significant effect of tempo and of loudness on mood ratings, which suggest that both features independently influenced perceived valence. Additionally, the interaction between tempo and loudness was significant, suggesting that the emotional effect of one feature may depend on the direction of the other. Notably, the reverse phrasing (T2L2) resulted in the second lowest mood ratings, while forward phrasing (T1L1) led to the highest, as illustrated in Figure 5. This further supports Hypothesis 2. Moreover, post hoc paired *t*-tests with Holm correction further confirmed this finding, as forward tempo with neutral loudness (T1L0) was rated significantly higher than reverse tempo with neutral loudness (T2L0) in mood, and forward phrasing (T1L1) was rated significantly more positively than reverse tempo with neutral loudness (T2L0). Though the direct comparison between ratings of mood on forward phrasing and reverse phrasing was not significant after Holm adjustment, the interaction plot and the two-way ANOVA result still give us support for this hypothesis.

In contrast, the third hypothesis, which stated that the effects of tempo and loudness will exhibit an additive relationship, was only partially supported by the data. The two-way ANOVA revealed a significant interaction effect between tempo and loudness on mood ratings and emotional expressiveness. But no significant interaction can be confirmed for preference or energy. This could indicate that additive or canceling effects are more relevant for mood and emotional expression than for energy (preference is not considered as an emotional effect). In general, the forward phrasing

(T1L1) produced the highest energy rating, and also performed strongly in mood. Conversely, the reverse phrasing (T2L2) consistently produced lower scores across energy and mood. This trend suggests that when tempo and loudness changes are aligned, they amplify their combined emotional impact. But notice T1L1 and T2L2 does not show significant difference in emotional expressiveness.

Furthermore, the opposing direction conditions did not uniformly "cancel each other out." For example, forward tempo with reverse loudness (T1L2) produced the highest rating for emotional expressiveness. Similarly, reverse tempo with forward loudness (T2L1) also achieved high ratings for expressiveness and mood. These findings suggest that contrasting phrasing patterns may enhance emotional complexity or salience, rather than simply neutralizing each other. This challenges the "canceling out" aspect of hypothesis 3, but instead, may support the aspect of emotional ambiguity. These patterns may reflect how listeners perceive unexpected or shifting emotional cues – a combination of forward tempo and reverse loudness (or vice versa) may be interpreted as dynamic or emotionally layered, rather than neutral. This aligns with the idea that listeners are attuned not just to the direction of feature changes but also to their structural and expressive interaction within a musical phrase.

The findings related to the fourth hypothesis expand earlier research on the relative influence of tempo and loudness. The results for expressiveness were consistent with Kamenetsky et al. (1997) as loudness was found to have a bigger sum of squares compared to tempo. While Kamenetsky et al. concluded that variations in loudness but not tempo influence listeners' perception of emotion and musical preferences, their suggestion that this effect generalized broadly to listeners' perception of emotion may have been premature. Both statistical significance and variance in the previous section suggest that tempo has a stronger and more consistent influence than loudness on energy and mood. This conclusion is supported by the sum of squares values from the two-way ANOVA, which indicate how much variance each factor explains. For energy, tempo showed a bigger sum of squares than loudness. And for mood, tempo also showed a bigger sum of squares than loudness.

Additionally, although preference is not considered an emotional effect, loudness also played a more important role than tempo in participants' ratings. This finding suggests that loudness may be more influential when it comes to subjective judgment, but not necessarily emotional interpretation.

Finally, while all reverse tempo conditions (T2) were generally associated with lower ratings in emotional perception, a consistent and noteworthy difference emerged between reverse tempo with neutral loudness (T2L0) and reverse tempo with reverse loudness (T2L2). Across all three emotional dimensions, T2L2 was rated higher than T2L0, including emotional expressiveness, energy, and mood. While reverse tempo alone (as in T2L0) led to the lowest energy and mood ratings, introducing reverse loudness (as in T2L2) significantly elevated ratings for both energy and mood. This observation indicates that the interaction between reverse tempo and reverse loudness produces a different emotional response than reverse tempo alone – higher in arousal and more positive in valence. This pattern reflects a potential interaction effect, where two features that are independently associated with lower emotional intensity result in increased affective ratings when combined. The explanation may lie in the temporal structure of reverse loudness, which features a crescendo following a decrescendo. This gradual intensification in loudness toward the end of the phrase may serve as a recovery mechanism, as it offsets the emotional inertia introduced by the reverse tempo.

5 Conclusion

Our findings provide substantial support for the first two hypotheses: forward phrasing led to higher participants' ratings of arousal and valence, while reverse phrasing elicited lower ratings, aligning more with subdued or negative emotional perceptions. These results reinforce prior research on the emotional trajectory of musical climaxes and the intensifying effect of increasing tempo and loudness (Fernández-Sotos et al., 2016; Patty, 2009).

While we found partial support for the additive effect of aligned phrasing patterns, unexpected findings emerged. Specifically, contrasting phrasing combinations (e.g., forward tempo with reverse loudness) did not simply neutralize emotional effects, but instead produced high expressiveness ratings. This suggests listeners may interpret divergent cues as emotionally complex rather than ambiguous. Moreover, the prediction that loudness would have a stronger effect over tempo in shaping emotional perception, was only partially supported. While loudness does have a stronger effect on emotional expressiveness, tempo consistently explained more variance in energy and mood ratings than loudness. This observation suggests though loudness may have a stronger effect on shaping perceived emotional expressiveness, tempo may play a more central role in driving perceived energy and mood in short musical excerpts.

Several limitations must be acknowledged. First, the musical material was limited to a single composed excerpt, which may restrict generalizability. Emotional responses could vary with different genres, timbres, harmonic content, or performance styles. Second, the use of synthetic audio may not fully capture the expressive nuances of human performance, which potentially influences how listeners interpreted phrasing patterns. Third, while the participant sample was musically diverse, it was recruited solely from a university population, which may limit generalization. Lastly, all ratings were collected via self-report on Likert scales, which, while informative, may not capture subtle or unconscious emotional reactions.

Future studies could benefit from including a wider range of musical stimuli and incorporating physiological measures (e.g., heart rate, galvanic skin response) to complement self-report data and to improve the robustness and validity of results. Further, incorporating real-time continuous rating tools would add nuance to participants' perception in real time.

6 Acknowledgments

The author thanks Dr. Ève Poudrier for her supervision and guidance throughout the project, and Dr. Christopher Mole for his support as the COGS 402 course instructor. The author also thanks Dr. Keith Hamel and Dr. Xin Sun for allowing recruitment in their classes. Appreciation is extended to the COGS 401 and COGS 402 classes for their helpful feedback.

7 References

Baird, A., Parada-Cabaleiro, E., Fraser, C., Hantke, S., & Schuller, B. (2018). The perceived emotion of isolated synthetic audio: The EmoSynth dataset and results. *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*, 1–8. https://doi.org/10.1145/3243274.3243277

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex, 47*(9), 1068–1081. https://doi.org/10.1016/j.cortex.2011.05.009

Chen, S. Y., Feng, Z., & Yi, X. (2017). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease, 9*(6), 1725–1729. https://doi.org/10.21037/jtd.2017.05.34

Fernández-Sotos, A., Fernández-Caballero, A., & Latorre, J. M. (2016). Influence of tempo and rhythmic unit in musical emotion regulation. *Frontiers in Computational Neuroscience, 10*, 80. https://doi.org/10.3389/fncom.2016.00080

Hunter, P. G., & Schellenberg, E. G. (2010). Music and emotion. In M. Riess Jones, R. Fay, & A. Popper (Eds.), *Music perception* (Vol. 36, pp. 129–164). Springer. https://doi.org/10.1007/978-1-4419-6114-3_5

Jäncke, L. (2008). Music, memory and emotion. *Journal of Biology, 7*, 21. https://doi.org/10.1186/jbiol82

Kamenetsky, S. B., Hill, D. S., & Trehub, S. E. (1997). Effect of tempo and dynamics on the perception of emotion in music. *Psychology of Music, 25*(2), 149–160. https://doi.org/10.1177/0305735697252005

Panda, R., Malheiro, R., & Paiva, R. P. (2023). Audio features for music emotion recognition: A survey. *IEEE Transactions on Affective Computing, 14*(1), 68–88. https://doi.org/10.1109/TAFFC.2020.3032373

Patty, A. T. (2009). Pacing scenarios: How harmonic rhythm and melodic pacing influence our experience of musical climax. *Music Theory Spectrum, 31*(2), 325–367. https://doi.org/10.1525/mts.2009.31.2.325

Poudrier, È., Bell, B. J., Lee, J. Y. H., & Sapp, C. S. (2023). Listeners' perceived emotions in human vs. synthetic performance of rhythmically complex musical excerpts. *16th International Symposium on Computer Music Multidisciplinary Research (CMMR 2023)*. https://doi.org/10.5281/zenodo.10076248