EFFECTS OF PITCH SHIFT AND TEMPO CHANGE ON LISTENER PREFERENCES FOR POPULAR MUSIC

Tomoyasu NAKANO (t.nakano@aist.go.jp) [1], **Sayaka TATSUMI** (contact@soc-research.org) [2], **Masataka GOTO** (m.goto@aist.go.jp) [1], and **Yoshinori HIJIKATA** (contact@soc-research.org) [2,3]

[1]**National Institute of Advanced Industrial Science and Technology (AIST)**, Japan
[2]**School of Business Administration, Kwansei Gakuin University**, Japan
[3]**Graduate School of Information Science, University of Hyogo**, Japan

ABSTRACT

A psychological experiment was conducted to investigate the effects of changing the pitch and tempo of a song of popular music by signal processing, *i.e.*, pitch shift and time stretch, on the listeners' preference ratings. Participants listened to songs (referred to as "original songs," which they were hearing for the first time), as well as to the same songs with their pitch and tempo altered. Participants gave their preference ratings for each on a 7-point Likert scale. The results confirmed that the changes in pitch and tempo had a significant effect on the preference rating values. The degree of variation in the ratings across nine different pitch and tempo changes (including the original song) was not correlated with participants' musical sophistication (Gold-MSI), but was significantly correlated with one musical factor, danceability, which is important for judging preference.

1. INTRODUCTION

Research on music preferences has been conducted in informatics and also in various fields such as psychology and psychophysiology. Music can produce emotional expressions that listeners within a given culture can agree upon [1], and features such as pitch and tempo in speech, music, and singing have been shown to be related to emotion [1,2]. Furthermore, different songs evoke varying listening impressions (e.g., "sadness" and "happiness") and physiological responses of listeners [3]. For example, songs with higher sadness ratings increase heart rate intervals and blood pressure, while songs with higher happiness ratings increase respiratory rate. Moreover, music preference has been found to significantly correlate with heart rate acceleration and facial electromyographic (EMG) activity [4].

It is also known that the melody and tempo of a song can influence its evaluation and the judgment of preference for a song on first listen [5]. Many other studies have also been conducted on the characteristics of songs appearing on music hit-charts, which have been discussed in terms of pitch and singing style [6], timbre and harmony [7], and chord progression [8]. Therefore, investigating how differences in pitch and tempo affect song preferences is expected not only to enhance our understanding of music's effects from a psychological perspective, but also to contribute to applications such as music recommendation from an informatics perspective.

The pitch and tempo of a song are not always the same as the original. In recent years, many cover performances of original songs, such as "me singing" versions, arranged renditions, and other derivative works, have been posted on online sharing platforms like NicoNico and YouTube. In such cases, the key or tempo of the original song may be changed, with pitch shifts or tempo changes sometimes applied. Even for mixed audio where instrumental sounds and vocals are combined, signal processing enables pitch and tempo to be changed independently [9]. As such technologies become more widespread in the future, listeners can enjoy songs with a different pitch and tempo from the original — just as adjusting sound quality with a graphic equalizer to suit personal preferences has become common. When a song is pitch-shifted to a higher (or lower) pitch by using the linear transformation of the frequency axis (*e.g.* [10]), especially in the case of vocals, the vocal timbre can give a significantly different impression due to the expansion or contraction of the spectral envelope caused. However, this change may also be part of the enjoyment.

The aim of this paper is to clarify the effects of changes in the pitch and tempo of a piece of popular music on listeners' preference. Since the evaluation of music preferences is influenced not only by the content of the music itself but also by various social factors, such as rankings [11] and the opinions of friends, this paper focuses on songs that listeners are hearing for the first time, ensuring that they have no prior information and that such external influences are minimized. The listener's knowledge and experience of music (referred to as *musical sophistication* [12]) may influence the listeners' evaluation of a song on first listen. Additionally, when judging preferences for a song, the elements that make up the song, such as melody, rhythm, and harmony, can have an influence [5]. Therefore, which musical factors each listener considers important during their preference evaluation may impact the results of the assessment.

Taking these factors into account, this paper aims to address the following research questions (RQs):

ISBN 978-3-200-10642-0

RQ1: Does changing the pitch or tempo affect the listener's preference evaluation?

RQ2: Is it possible for the preference to be higher after changing the pitch or tempo compared to the original song?

RQ3: Does the difference in musical sophistication influence the tendency of preference evaluations when pitch or tempo is changed?

RQ4: Do differences in the musical factors that listeners consider important lead to different trends in preference evaluation when pitch or tempo is changed?

2. RELATED WORK

This section describes previous works on effects of pitch, tempo, and musical factors and the use of pitch shifts and tempo changes.

2.1 Effects of pitch and tempo on emotions and attitudes

Much research has been conducted on the effects of musical pitch and tempo on people's emotions and attitudes. Scherer and Zentner [13] reviewed evidence to formalize the processes by which music produces emotional effects on listeners and proposed a method for systematically investigating such processes. They attempted to formalize production rules using various musical features as input variables, such as pitch (a segmental feature) and tempo and rhythm (suprasegmental features), and emotions as output variables. Scherer [2] reviewed studies on the expression of emotion and listeners' responses in voice and music. Taken together, these studies show, for instance, high pitch is associated with anger, low pitch with pleasantness, fast tempo with potency and happiness, and slow tempo with sadness and boredom. Holbrook and Anand [14] investigated the relationship between tempo and perceived activity and between tempo and preference. They reported a linear relationship between the logarithm of tempo and a perceived activity index and a nonlinear relationship in which moderate tempo was preferred. The study also revealed that participants with high arousal tended to prefer a faster tempo. Kamenetsky *et al.* [15] investigated whether expressive variations in tempo and dynamics applied to 30-second excerpts could influence the emotional expression and likeability of the same song. In the four musical excerpts they used, variations in dynamics were found to influence listeners' perception of emotion in music and their musical preferences, whereas variations in tempo did not have an effect.

These studies show that the effects of pitch and tempo on emotion are diverse and important for understanding how music elicits emotional responses, including preferences.

2.2 Effects of musical factors on emotion and attitude

In addition to pitch and tempo, many studies have explored the impact of various musical factors on emotions and attitudes. Trost *et al.* [16] reviewed studies on rhythmic entrainment in emotional induction, such as how moving the body to preferred music helps regulate mood and emotion, and discussed its theoretical implications and the empirical findings. Peretz and Coltheart [17] proposed a model of the functional architecture of music processing in the brain. This model represents that music-related processing, such as tonal encoding, plays a different role than other cognitive functions. Brattico *et al.* [18] investigated the experience of happiness in music (both happy and sad music) and the importance of lyrics in emotion by analyzing brain responses. They focused on songs selected by the experiment participants themselves, including popular and classical music. Eerola *et al.* [19] investigated how musical cues (mode, tempo, dynamics, articulation, timbre, and register) operate on emotional expression. Their importance ranking was established by multiple regression analyses, suggesting that their cue levels contribute to emotion in a linear fashion, that there is no significant interaction between cues, and that the cues operate primarily in an additive fashion. Keller and Schubert [20] showed that unpredictable rhythms due to musical syncopation have a positive impact on emotional valence.

These studies have shown that musical factors such as rhythm, melody, timbre, and syncopation also influence emotions and attitudes. In contrast, this paper focuses on changing pitch and tempo and explores how such changes influence listeners' preferences.

2.3 Applications using pitch shifts and tempo changes

Changing the pitch and tempo of an existing piece of music is easy in MIDI performances and is also possible with mixed sound acoustic signals through signal processing [9], so various studies have used these approaches. Fukumoto et al. [21] reported that listening to the piano piece "Ave Maria" at different tempi with a MIDI sound source changed impressions — feelings held towards the piece such as relaxed, energetic, strong, or noisy. Wager *et al.* [22] proposed a network architecture for pitch correction of singing voices with high accuracy using deep learning. The feature of this method is that it achieves singing pitch correction independent of musical scores. Rosenzweig *et al.* [23] proposed an auto-adaptive pitch-shifting method for pitch adjustment of a cappella recordings. The method can adjust the pitch of individual singers and compensate for intonation drift. Davies *et al.* [24] proposed a method to automatically adjust the pitch and tempo of different songs and combine multiple songs to create new musical works. This method allows for the natural superimposition of songs on each other by adjusting pitch and tempo to create new content. Nakano *et al.* [10] proposed a method to retrieve similar vocal timbre between male and female singers through the use of vocal timbre conversion by pitch shifting.

However, it is still unclear how listeners' preferences change when the pitch and tempo of a musical acoustic signal are changed, including vocals in popular music.

2.4 Novelty of this study

Previous studies investigating the effects of music on emotions and attitudes have often investigated effects of pitch and tempo on emotions and attitudes separately [1, 2, 16].

ISBN 978-3-200-10642-0

Therefore, this paper provides a new perspective by using signal processing techniques that enable changes in both pitch and tempo, thus revealing the combined effects of simultaneous pitch and tempo modification on listeners' preference ratings.

In addition, most previous studies have mainly focused on either simple note sequences or classical music [3, 15, 19, 20], and to our knowledge there have been no studies on musical acoustic signals containing vocals in popular music. Popular music is experienced differently from classical music, as its reception is influenced by factors such as vocal presence, lyrics, and accompaniment. In this paper, the effects of changes in pitch and tempo on preference evaluation are analyzed based on the results of a psychological experiment on popular music with lyrics.

3. EXPERIMENTAL SETTING

In this paper, songs whose pitch and tempo have been changed from the original song are referred to as "stimulus songs". These songs were presented in a random order to the participants, who were asked to rate their degree of preference for each. Participants were also asked questions to assess their level of musical sophistication and which musical factors they consider important. The influence of these on their preference ratings will also be analyzed.

3.1 Songs used in this experiment

A total of four songs were used in this experiment: two songs by male singers and two songs by female singers. These were all solo songs with one singer and were newly created for research purposes. Therefore, the participants had never listened to them. In this paper, the male songs are referred to as "song M1" and "song M2" and the female songs as "song F1" and "song F2".

In the experiment, 30-s excerpts were extracted from the chorus of each song and used. Their tempi were 113 BPM (beats per minute) for song M1, 185 BPM for song M2, 160 BPM for song F1, and 122 BPM for song F2.

3.2 Degree of change in pitch and tempo

The pitch and tempo of the songs were changed using SoX. Considerations when creating the stimulus songs were that too large a change in pitch and tempo would cause a significant deterioration in sound quality and that too many stimulus songs with small changes (finer granularity of change) would increase the listening burden on the participants. Therefore, as a preliminary experiment, preferences were evaluated using 25 patterns of 5×5 stimulus songs with pitch shifts in semitone increments of -3 to $+3$ and tempo changes in 0.1 increments of $\times 0.8$ to $\times 1.2$. In some cases, sound quality deteriorated and participants had difficulty perceiving differences. In addition, it was found necessary to evaluate each stimulus song multiple times, because when the participants evaluated the same stimulus song again, the evaluation results sometimes differed from the first listening.

Therefore, another preliminary experiment was conducted with a pitch shift of -2 to $+2$ in semitone increments and a tempo change of $\times 0.8$ to $\times 1.2$ in 0.2 increments, and as no participants found the sound quality or singing voice to be extremely unnatural, the conditions for creating stimulus songs were determined as follows.

1. Pitch：-2, ± 0, $+2$ (semitones)
2. Tempo：$\times 0.8$, $\times 1.0$, $\times 1.2$ (times)

A total of nine stimulus songs were thus created for a single piece of music from three different pitch shifts and three different tempo changes. Note that the stimulus song with a pitch shift of ± 0 and a tempo change of $\times 1.0$ is the original song.

3.3 Dummy stimuli

To prevent the participants from guessing the original song among the nine stimulus songs with the pitch shift and tempo changes, three dummy stimulus songs ("dummy stimuli") were added, resulting in a total of twelve stimulus songs. Two of these were selected at random from the nine stimulus songs, including the original song, and used as they were. The remaining one was a signal-processed sound in which the original song was modified in pitch and tempo in one of the following three ways and then modified to be closer to the original (these are referred to as "dummy stimuli X, Y, and Z").

X The pitch was changed to $+2$ and the tempo to $\times 1.2$, then the pitch was changed to -2 and the tempo to $\times 0.83$ to be closer to the original pitch and tempo.

Y The pitch was changed to $+2$, then changed to -2 to be closer to the original pitch.

Z The tempo was changed to $\times 1.2$, then changed to $\times 0.83$ to be closer to the original tempo.

Each of X, Y, and Z was presented once to each participant.

3.4 Randomization of the order of listening

In this experiment, nine stimulus songs per original song were presented in random order, followed by one of the dummy stimuli X, Y, and Z. Then, two dummy stimuli, randomly selected from the nine stimuli, were presented. In this way, participants listened to a total of twelve stimulus songs per original song. This was done for the four original songs, so each participant listened to and evaluated 48 stimulus songs (=4 songs $\times 12$ stimuli), which is referred to as "one set". As preference ratings can change when the same stimulus song is listened to again, three sets were conducted and the results of three preference ratings were obtained for each original song.

Here, because it was expensive to implement the experimental design to completely randomize the order of listening for each participant, three random patterns were predefined and used. Participants were randomly assigned one of these three patterns.

4. PREFERENCE EVALUATION EXPERIMENT

Participants were 12 males and 48 females (total of 60) aged 18 years or older who listen to music daily and who

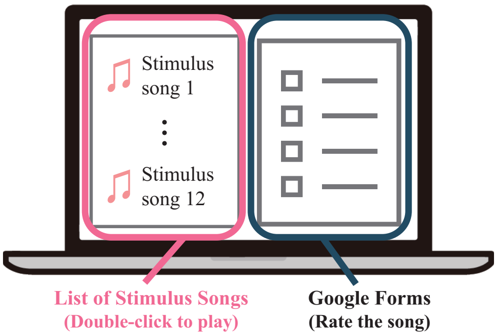

Figure 1. Experiment screen: The list of stimulus songs was displayed on the left side of the laptop screen, and the evaluation screen was shown on the right side, allowing participants to listen while making their evaluations.

gave their consent to participate in the experiment. Participants were recruited through convenience sampling. As a reward for their participation in the experiment, they received Amazon gift vouchers worth 2500 JPY.

The following equipment was used in the experiment:

- Laptop: Lenovo IdeaPad Gaming 315ARH7 (15.6-inch screen, 1920×1080 resolution)
- Headphones: Audio-Technica ATH-MSR7b

The volume of the headphones was adjusted by the participants themselves to a level they felt comfortable with.

This experiment was approved by the Research Ethics Review Committee of the University of Hyogo.

4.1 Experimental procedure

As mentioned in Sections 3.3 and 3.4, participants listened to twelve stimulus songs per original song, and they evaluated each stimulus song after each listening. This was repeated for the four original songs to make one set. Once one set was completed, a 10-minute break was provided to reduce the burden on the participants.

Figure 1 shows an overview of the experimental screen. On the left side of the laptop screen, a list of twelve stimulus songs for one original song was displayed, which the participants played by double-clicking them from the top. On the right side of the screen, a Google Forms window was displayed to enter the rating results for each song.

The specific procedure is outlined below.

1. Participants received an explanation of the procedure, instructions for use, and duration of the experiment on the laptop screen.
2. In each of the three sets, participants listened to twelve stimulus songs on the left side of the screen. Immediately after listening to each stimulus song, they rated their degree of preference for it on a 7-point Likert scale using Google Forms on the right side of the screen (1. Strongly dislike, 2. Dislike, 3. Slightly dislike, 4. Neutral, 5. Slightly like, 6. Like, 7. Strongly like).
3. After the completion of the three sets, the participants were asked questions based on the Music Sophistication Index (Gold-MSI) [12] with the aim of measuring their musical experience and musical competence. Then, on a 5-point Likert scale based on Tsukuda *et al.* [5], the participants rated the importance of ten musical factors (melody, singing voice, rhythm, lyrics, mood, tempo, harmony, sentiment, instruments, and danceability) in judging preferences for music.

4.2 Analysis

When evaluating preferences, there is a possibility that the song listened to immediately before may influence the evaluation of the next song or that the evaluation results may fluctuate depending on the participant's state or mood at the time of evaluation. In order to suppress these effects, participants listened to the same stimulus songs three times (three sets) in different orders as described in Section 3.4. The average of the three evaluation results for each of the nine stimulus songs, excluding the dummy stimuli, was calculated and analyzed as the "preference score".

In addition, the correlation between the participants' musical sophistication (Gold-MSI) and their preference scores for the stimulus songs was calculated. The correlation between the participants' preference scores for the stimulus songs and the rated importance of the ten musical factors in their preference scores was also analyzed.

5. EXPERIMENTAL RESULTS

To answer the research questions (RQs) described in Section 1, the preference scores were analyzed.

5.1 RQ1: Does changing the pitch or tempo affect the listener's preference evaluation?

To investigate the effect of pitch and tempo changes on preference scores, a heat map was created and analyzed for each of the four songs based on the preference scores from all 60 participants (the average of three ratings for each stimulus song). The results are shown in Figure 2. Each heat map visually represents the preference scores for the original song in the center and the eight stimulus songs surrounding it, which had their pitch and tempo changed. The horizontal axis represents the degree of tempo change, and the vertical axis represents the degree of pitch change. The color of the cells reflects whether the average is high or low, making it easy to intuitively grasp the trend of the preference scores.

In addition, to check statistics such as median and quartiles, a boxplot display is shown separately for each song in Figure 3. The notation of the change pattern of the stimulus song in the figure, such as (p+2, t0.8), indicates that the pitch was raised by $+2$ semitones and the tempo was reduced to 0.8 times the original. In Figure 3, the quartile ranges and medians of the preference scores for the 60 participants are different for each change pattern.

To further investigate the impact of pitch and tempo changes on preference scores in more detail, Wilcoxon signed-rank tests were conducted on preference scores for the original song (without change of pitch and tempo) and for each stimulus song. The test was conducted for each

ISBN 978-3-200-10642-0

Table 1. Results of Wilcoxon signed-rank tests for preference evaluations of the four original songs (M1, M2, F1, and F2) and each stimulus song (after Bonferroni correction: * denotes $p < 0.05$)

M1						M2					
Stimuli	Mean	S.D.	Median	V	p	Stimuli	Mean	S.D.	Median	V	p
(p+2, t0.8)	2.057	0.892	2.0	1823.5	1.76e-10*	(p+2, t0.8)	2.092	1.021	2.0	1824.5	1.68e-10*
(p+2, t1.0)	4.217	1.017	4.3	963.5	4.43e-05*	(p+2, t1.0)	4.153	1.203	4.3	1162.0	3.07e-06*
(p+2, t1.2)	4.222	0.977	4.3	1233.5	9.61e-03*	(p+2, t1.2)	2.867	1.070	3.0	1648.5	5.10e-10*
(p0, t0.8)	2.820	1.005	2.7	1820.5	2.03e-10*	(p0, t0.8)	3.050	1.142	2.7	1755.0	3.98e-10*
(p0, t1.0)	4.788	1.109	4.7	—	—	(p0, t1.0)	5.042	1.116	5.0	—	—
(p0, t1.2)	4.365	1.001	4.3	1142.5	0.0948	(p0, t1.2)	3.837	1.189	4.0	1318.5	7.38e-08*
(p-2, t0.8)	2.655	1.062	2.7	1748.5	5.55e-11*	(p-2, t0.8)	2.893	1.104	3.0	1824.5	1.67e-10*
(p-2, t1.0)	4.672	1.190	4.7	847.5	1.181	(p-2, t1.0)	4.860	1.216	5.0	720.0	0.386
(p-2, t1.2)	4.253	1.098	4.3	1176.0	0.00518*	(p-2, t1.2)	3.508	1.217	3.7	1575.5	1.71e-09*
F1						**F2**					
Stimuli	Mean	S.D.	Median	V	p	Stimuli	Mean	S.D.	Median	V	p
(p+2, t0.8)	2.902	1.306	2.5	1702.0	4.38e-10*	(p+2, t0.8)	2.948	1.191	2.7	1760.0	3.10e-10*
(p+2, t1.0)	4.782	1.368	4.7	1329.0	0.000497*	(p+2, t1.0)	5.580	0.934	5.7	526.0	4.227
(p+2, t1.2)	4.185	1.495	4.3	1474.0	2.09e-06*	(p+2, t1.2)	4.713	1.155	5.0	1182.0	0.001208*
(p0, t0.8)	4.257	1.247	4.3	1549.5	6.47e-09*	(p0, t0.8)	3.183	1.179	3.0	1770.0	1.85e-10*
(p0, t1.0)	5.533	1.143	5.7	—	—	(p0, t1.0)	5.552	0.950	5.7	—	—
(p0, t1.2)	4.523	1.106	4.7	1401.5	6.57e-06*	(p0, t1.2)	4.763	0.973	5.0	1131.5	0.00044*
(p-2, t0.8)	2.947	1.138	2.7	1711.0	2.74e-10*	(p-2, t0.8)	3.348	1.301	3.3	1748.5	5.60e-10*
(p-2, t1.0)	4.368	1.267	4.3	1336.0	3.08e-07*	(p-2, t1.0)	5.208	1.115	5.3	869.5	0.07651
(p-2, t1.2)	3.897	1.064	4.0	1651.0	5.64e-09*	(p-2, t1.2)	4.582	1.042	4.5	1477.5	1.13e-05*

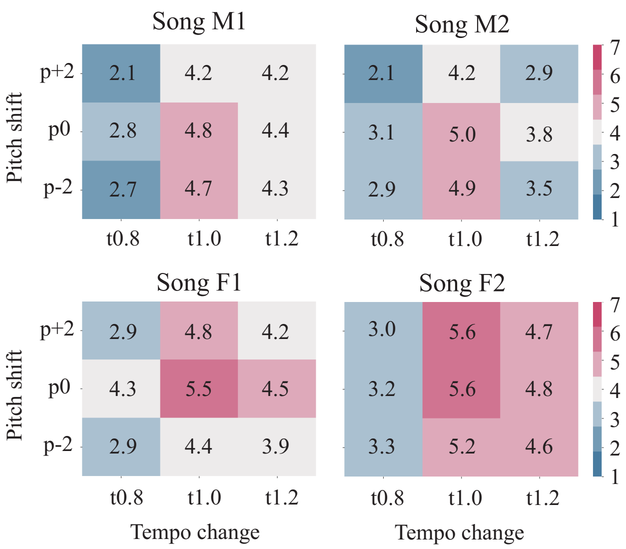

Figure 2. Heatmaps showing the average of the 60 participants' preference scores for each stimulus song.

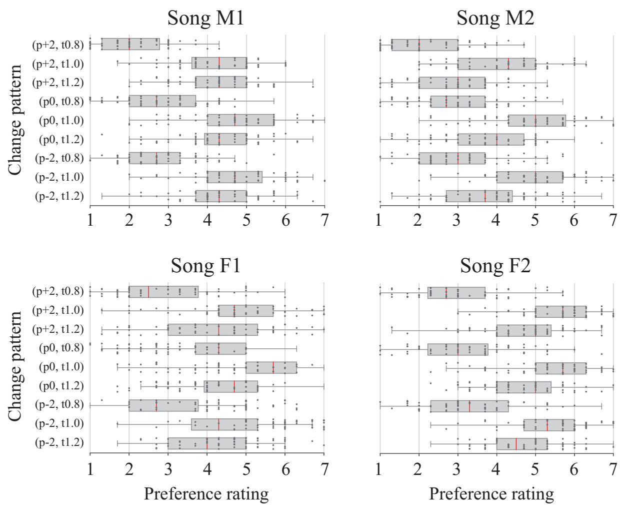

Figure 3. Distribution of preference scores of the 60 participants for each stimulus song.

original song and the preference score for the corresponding stimulus song. In all tests, the Bonferroni correction was applied and p values were adjusted ($\times 8$) to set the significance level at $p < 0.05$. This correction reduced the risk of false detection due to multiple tests.

The results are shown in Table 1. For song M1, significant differences were found between the scores of the original and stimulus songs in six conditions of (p+2, t0.8), (p+2, t1.0), (p+2, t1.2), (p0, t0.8), (p-2, t0.8), and (p-2, t1.2). For song M2, significant differences were also observed in seven conditions: (p+2, t0.8), (p+2, t1.0), (p+2, t1.2), (p0, t0.8), (p0, t1.2), (p-2, t0.8), and (p-2, t1.2). For song F1, significant differences were found in all conditions. For song F2, significant differences were also observed in six conditions: (p+2, t0.8), (p+2, t1.2), (p0, t0.8), (p0, t1.2), (p-2, t0.8), and (p-2, t1.2).

These results not only show that changing the pitch and tempo from the original song affected the preference scores but also that the change in preference scores varied depending on the four original songs. For example, as can be seen from the heatmap in Figure 2, lowering the pitch was preferred over raising it for M1 and M2, while raising the pitch was preferred over lowering it for F1 and F2. In other words, male singing voices tended to be preferred when pitched lower, while female singing voices tended to be preferred when pitched higher. Regarding tempo changes, there was a common tendency for the preference scores to decrease when the tempo was slowed down to t0.8. A faster tempo was generally preferred over a slower one for M1 and F2, whose original tempi were 113 and 122 BPM, respectively — slower than those of the other two songs (185 BPM for M2 and 160 BPM for F1).

ISBN 978-3-200-10642-0

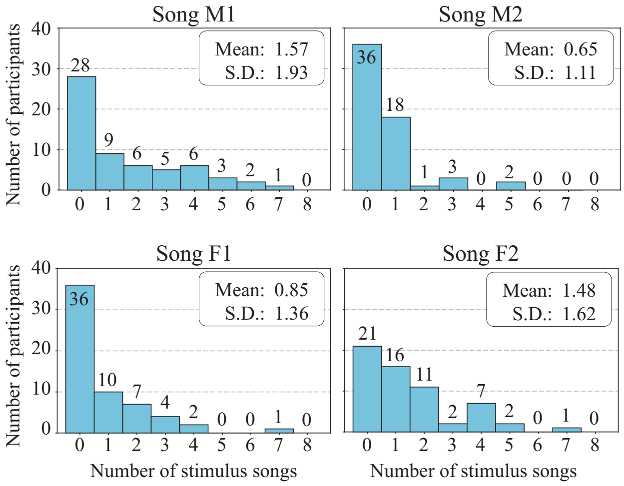

Figure 4. Aggregation of the number of stimulus songs with a higher preference score than the original song.

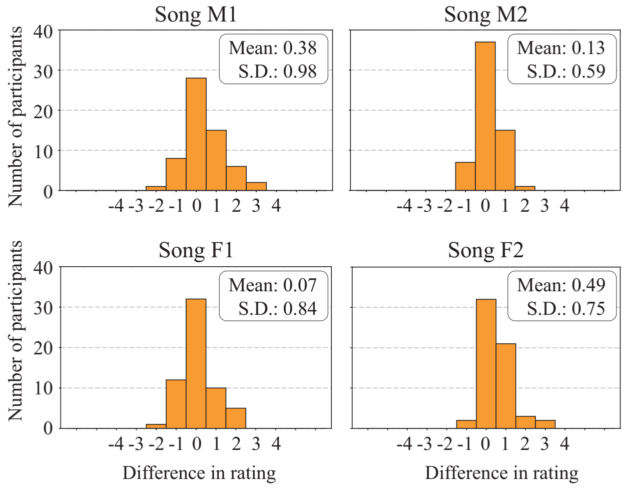

Figure 5. Aggregation of the difference between the highest preference score in the stimulus song and the preference score of the original song.

5.2 RQ2: Is it possible for the preference to be higher after changing the pitch or tempo compared to the original song?

In order to investigate whether changing the pitch and tempo can result in higher preference ratings than for the original song, the number of stimulus songs with higher preference scores than the original song was counted for each participant. Histograms for the 60 participants are shown in Figure 4. The horizontal axis represents the number of stimulus songs that had a higher preference score than the original song, while the vertical axis represents the number of corresponding participants. The largest number of participants gave the highest score to the original song: 28 of the 60 participants gave the highest score to the original song in song M1, 36 in M2, 36 in F1, and 21 in F2. This means that as many as 24–39 participants had at least one stimulus song with a changed pitch or tempo that they liked better than the original song.

Therefore, the difference in preference score between the original song and the most preferred stimulus song (other than the original) was calculated for each participant and each of the four songs. The histograms for the 60 participants are shown in Figure 5. The horizontal axis represents the difference between the highest score of the stimulus song minus the score of the original song, while the vertical axis represents the number of corresponding participants. The horizontal axis also includes negative values, as some participants gave the highest score to the original song. The horizontal axis at a difference of 0 includes cases where the stimulus song has a slightly higher preference score than the original song, as well as cases where it is slightly lower. The results in Figure 5 show that, for all four songs, the largest number of participants corresponds to a difference of 0. However, as seen in Figure 4, it was also found that some participants preferred the stimulus songs with modified pitch and tempo over the original songs, giving them a score 1 to 3 points higher on a 7-point scale.

5.3 RQ3: Does the difference in musical sophistication influence the tendency of preference evaluations when pitch or tempo is changed?

We hypothesized that participants with higher levels of musical sophistication would be more sensitive to changes in pitch and tempo, leading to differences in their tendency to rate their preferences for the pitch/tempo-changed stimulus songs. Therefore, for each participant, the correlation between the degree of variation in preference scores and the participant's music sophistication (Gold-MSI) was calculated.

First, the degree of variation of the preference scores for each participant in the experiment was obtained in two ways (called A and B), using the following procedure.

Step A1 Calculate the standard deviation of the preference scores for the nine stimulus songs.

Step A2 Calculate the average of the four standard deviations (*i.e.*, songs M1, M2, F1, and F2) calculated in step A1 for each of the four songs.

Step B1 Calculate the average from all three sets of four songs (twelve preference ratings) for the nine stimulus patterns.

Step B2 Calculate the variance of the nine values obtained in step B1.

This yields two values of the degree of variation for each participant in the experiment.

Next, each participant was asked to respond to the questions in the Gold-MSI on a 7-point scale, and the sum of the responses was obtained. The following 16 items, which are closely related to the present work, were used as specific question items ("*" means a reverse-scored item).

Factor 1: Active Engagement Nos. 8, 15, 24, 28, and 38
Factor 2: Perceptual Abilities Nos. 18, 22, and 23*
Factor 3: Musical Training Nos. 32, 35, and 36
Factor 4: Singing Abilities Nos. 10, 17*, and 29
Factor 5: Emotions Nos. 9* and 20

As the degree of variation of the preference scores and the respective Gold-MSI values were obtained for each

ISBN 978-3-200-10642-0

participant, a correlation analysis was conducted by calculating Spearman's rank correlation coefficient from the values for all participants. For the degree of variations A and B, the results showed that Spearman's rank correlation coefficients were 0.0467 and 0.0948 and the p values were 0.7229 and 0.4714, respectively, indicating no significant correlation at the 5% level. This indicates that there is no significant correlation between the degree of variation in preference scores and the level of musical sophistication.

5.4 RQ4: Do differences in the musical factors that listeners consider important lead to different trends in preference evaluation when pitch or tempo is changed?

We investigate whether the importance that participants place on the ten musical factors (melody, singing voice, rhythm, lyrics, mood, tempo, harmony, sentiment, instruments, and danceability) in their preference evaluation affects the degree of variation in their preference scores. We asked participants, 'How important is "melody" in deciding whether you like or dislike a song when you listen to music?' with the same question for each musical factor. The participants rated each of the factors on a 5-point Likert scale (1. Not important, 2. Mostly unimportant, 3. Somewhat important, 4. Important, 5. Very important). Furthermore, the sum of the importance ratings of those ten factors was treated as the "total". The degree of variation of the preference scores was the same as in RQ3.

For each participant, we obtained the ten importance ratings, their total, and the degree of variation of the preference scores. Using these values from all participants, we calculated Spearman's rank correlation coefficient and the p value for the test of no correlation. The results are shown in Table 2. The correlation coefficients were 0.0631 and 0.0195 and the p values were 0.6322 and 0.8827, respectively, for "total" in Table 2, which were not significant at the 5% level. This result indicates that there is no significant correlation between the total of the ten importance ratings and the degree of variation of the preference scores.

Next, when looking at the correlations between each of the musical factors in Table 2, correlation coefficients of 0.2230 and 0.2455 and p values of 0.0868 and 0.0587, respectively, were obtained for "singing voice," which were not found to be significant at the 5% level, but a tendency towards significance was observed. In contrast, for "danceability," correlation coefficients of -0.2550 and -0.3109 and p values of 0.0493 and 0.0156 were obtained, which were significant negative correlations at the 5% level. On the other hand, no significant correlations were confirmed for any of the other musical factors. These results indicate that the degree of variation in preference scores differs when the musical factors that listeners consider important during preference evaluation are "singing voice" or "danceability."

Table 2. Spearman's rank correlation coefficient and p value of the no-correlation test between the importance ratings of musical factors and the degree of variation in preference scores for all participants (* denotes $p < 0.05$).

Musical Factor	**Degree of Variation** (A) Coef.	(A) p	(B) Coef.	(B) p
Total	0.0631	0.6322	0.0195	0.8827
Melody	0.1864	0.1538	0.1441	0.2720
Singing voice	0.2230	0.0868	0.2455	0.0587
Rhythm	0.1518	0.2470	0.0763	0.5623
Lyrics	−0.1286	0.3274	−0.0812	0.5374
Mood	0.0620	0.6378	0.0402	0.7602
Tempo	0.0994	0.4497	0.0605	0.6459
Harmony	0.1408	0.2834	0.1312	0.3178
Sentiment	0.1289	0.3262	0.0675	0.6083
Instruments	0.1413	0.2815	0.1202	0.3603
Danceability	−0.2550	0.0493*	−0.3109	0.0156*

6. DISCUSSION

The results for RQ1 and RQ2 confirmed that the listeners' preference scores changed when the pitch and tempo were changed and that the preference scores were sometimes higher than for the original song. Since there are individual differences in musical preferences, it was found that the preference scores varied greatly among the participants even when the same changes were made and that the tendency of what changes are preferred is also different for each song. It was also confirmed that the original song tended to receive the highest scores. However, depending on the song, it is noteworthy that more than half of the participants liked at least one of the eight stimulus songs with pitch and tempo changes from the original song. This suggests that it is possible that *"listeners can enjoy songs with a different pitch and tempo from the original"* as described in Section 1.

The results for RQ3 showed that there was no significant correlation between the degree of variation in preference scores and the level of musical sophistication for each participant in the experiment. It can be inferred that other influences, such as differences in musical taste, were stronger in how much preference scores changed when pitch or tempo was changed.

The results of RQ4 are interesting in that participants who consider the musical factor "singing voice" to be important during preference evaluation tend to show a greater degree of variation in preference scores. Since changing the pitch causes a change in the impression of the singing voice, it is possible that the participants who considered the singing voice important changed their preference scores due to this change. On the other hand, when the musical factor "danceability" was considered important, conversely, the degree of variation in the preference scores tended to be smaller. From the perspective of whether a song is danceable, it is possible that the preference evaluation was less affected because changes in pitch and tempo within the range examined in this study had little impact.

7. CONCLUSION

This study investigated the effects of pitch and tempo changes in popular music on listeners' preference ratings. A key finding is that listeners' preference scores did change when pitch and tempo were changed. In some

ISBN 978-3-200-10642-0

cases, for more than half of the listeners, the preference scores were higher for the modified versions than for the original song. The tendency of what kind of change is preferred differed from song to song. It was also found that, for listeners who consider "singing voice" and "danceability" important when judging their preferences, the extent to which their ratings changed with alterations in pitch and tempo varied. One limitation is the small number of songs ($n = 4$), which may have influenced preference evaluations due to their unique vocals and accompaniment.

Future studies should include more songs while considering participant burden. We hope this study will serve as a starting point for further research on the preferences and attributes of various listeners, with a more detailed exploration of the relationship between pitch and tempo changes and preference scores across different types of music.

Acknowledgments

This work was supported in part by JST CREST Grant Number JPMJCR20D4, Japan.

8. REFERENCES

[1] P. N. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *JNMR*, vol. 33, no. 3, pp. 217–238, 2004.

[2] K. R. Scherer, "Expression of emotion in voice and music," *J. Voice*, vol. 9, no. 3, pp. 235–248, 1995.

[3] C. L. Krumhansl, "An exploratory study of musical emotions and psychophysiology," *Can. J. Exp. Psychol.*, vol. 51, no. 4, pp. 336–353, 1997.

[4] N. Fuentes-Sánchez, R. Pastor, T. Eerola, M. A. Escrig, and M. C. Pastor, "Musical preference but not familiarity influences subjective ratings and psychophysiological correlates of music-induced emotions," *Personality and Individual Differences*, vol. 198, pp. 1–6, 2022.

[5] K. Tsukuda, T. Nakano, M. Hamasaki, and M. Goto, "Unveiling the impact of musical factors in judging a song on first listen: Insights from a user survey," in *Proc. ISMIR 2023*, 2023, pp. 561–570.

[6] K. Shohei and K. Haruhiro, "Analysis of transition in singing styles in J-POP: Perspectives on pitch elevation and vocalization in male voice singing," *IPSJ Journal (in Japanese)*, vol. 64, no. 11, pp. 1463–1473, 2023.

[7] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi, "The evolution of popular music: USA 1960–2010," *Royal Society Open Science*, vol. 2, no. 5, pp. 1–10, 2015.

[8] E. Nakamura, "Computational analysis of selection and mutation probabilities in the evolution of chord progressions," in *Proc. CMMR 2023*, 2023, pp. 462–473.

[9] U. Zölzer, *DAFX: Digital Audio Effects*, 2nd ed., 2011.

[10] T. Nakano, K. Yoshii, and M. Goto, "Vocal timbre analysis using latent dirichlet allocation and cross-gender vocal timbre similarity," in *Proc. IEEE ICASSP 2014*, 2014, pp. 5239–5243.

[11] M. J. Salganik, P. S. Dodds, and D. J. Watts, "Experimental study of inequality and unpredictability in an artificial cultural market," *Science*, vol. 311, no. 5762, pp. 854–856, 2006.

[12] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart, "The musicality of non-musicians: An index for assessing musical sophistication in the general population," *PLOS ONE*, vol. 9, no. 2, 2014.

[13] K. R. Scherer and M. R. Zentner, "Emotional effects of music: Production rules," in *Music and Emotion: Theory and Research*, 2001, pp. 361–392.

[14] M. B. Holbrook and P. Anand, "Effects of tempo and situational arousal on the listener's perceptual and affective responses to music," *Psychol. Music*, vol. 18, no. 2, pp. 150–162, 1990.

[15] S. B. Kamenetsky, D. S. Hill, and S. E. Trehub, "Effects of tempo and dynamics on the perception of emotion in music," *Psychol. Music*, vol. 25, pp. 149–160, 1997.

[16] W. J. Trost, C. Labbé, and D. Grandjean, "Rhythmic entrainment as a musical affect induction mechanism," *Neuropsychologia*, vol. 96, pp. 96–110, 2017.

[17] I. Peretz and M. Coltheart, "Modularity of music processing," *Nat. Neurosci.*, vol. 6, no. 7, pp. 688–691, 2003.

[18] E. Brattico, V. Alluri, B. Bogert, T. Jacobsen, N. Vartiainen, S. Nieminen, and M. Tervaniemi, "A functional MRI study of happy and sad emotions in music with and without lyrics," *Frontiers in Psychology*, vol. 2, no. 308, pp. 1–16, 2011.

[19] T. Eerola, A. Friberg, and R. Bresin, "Emotional expression in music: Contribution, linearity, and additivity of primary musical cues," *Frontiers in Psychology*, vol. 4, no. 487, pp. 1–12, 2013.

[20] P. E. Keller and E. Schubert, "Cognitive and affective judgements of syncopated musical themes," *Advances in Cognitive Psychology*, vol. 7, pp. 142–156, 2011.

[21] M. Fukumoto, K. Okamatsu, and K. Matsuo, "Investigation of method for changing impression of musical piece by changing its tempo," *IEEJ Trans. EIS (in Japanese)*, vol. 127, no. 11, pp. 1953–1954, 2007.

[22] S. Wager, G. Tzanetakis, C. i Wang, and M. Kim, "Deep Autotuner: A pitch correcting network for singing performances," in *Proc. IEEE ICASSP 2020*, 2020, pp. 246–250.

[23] S. Rosenzweig, S. J. Schwär, J. Driedger, and M. Müller, "Adaptive pitch-shifting with applications to intonation adjustment in a cappella recordings," in *Proc. DAFx 2021*, 2021, pp. 121–128.

[24] M. E. P. Davies, P. Hamel, K. Yoshii, and M. Goto, "AutoMashUpper: Automatic creation of multi-song music mashups," *IEEE/ACM TASLP*, vol. 22, no. 12, pp. 1726–1737, 2014.

ISBN 978-3-200-10642-0