國立臺灣大學理學院心理學研究所

碩士論文

Graduate Institute of Psychology

College of Science

National Taiwan University

Master Thesis

音樂與演唱影像的整合對於情緒判斷的影響

Audio-visual Integration Modifies Emotional Judgment in Music

蘇紳源

Shen-Yuan Su

指導教授：葉素玲博士

Advisor: Su-Ling Yeh, Ph.D.

中華民國 98 年 6 月

June, 2009

誌謝

這份論文的完成，意謂著下個人生階段的開始。碩班的日子很單純也過得很快，在短短兩年間卻讓自己對未來有了新的想法。感謝這段時間陪伴的朋友、學長姐們幫忙與協助，另外更要感謝葉老師對於我的包容與體諒。音樂是我的根，像宗教般的信念植入我的靈魂中，音樂與心理的探討也是我進入碩士班的動機－唯一的動機。很感謝葉老師給我這個機會，做我想做的，念我想念的，再次感謝葉老師對於這篇論文的協助，沒有了她的幫助，這一切將不會發生。

很多人問我：音樂跟心理有甚麼關係？你以後要當音樂治療師嗎？剛開始我會慢慢地說，慢慢的解釋，後來的我笑而不答。有些事情是需要體認的，言語這時反而是種累贅。美可以是藝術的一部份，而美感更是人性的一部份，對於美感與藝術兩者的關聯我們所知甚少，相關科學研究更是冷門，更別說在台灣，根本沒甚麼人在意這件事。不同的人用不同的角度看美，用不同的語言去闡述它的意義，但我們有因此更了解什麼是快感什麼是美嗎？這篇論文指是個起頭，談的不多也不夠深刻，不過只要目標明確就有往下走的可能性。

音樂與審美從古希臘時代就是哲學中一個很重要的議題，東方文化中也有「聲無哀樂論」在探討音樂與情緒的關聯性。在哲學的思潮裡音樂的角色一直在改變，審美經驗的內涵卻只能被文字捕捉到部分精髓。情感在音樂中重不重要？音樂結構是否為美感唯一來源？我相信透過實徵研究我們可以更了解音樂與美，哪怕只能窺見冰山一角，還是有助於釐清這個大問題－何謂美感經驗？最後，謝謝大家的幫助，感謝周遭的每一個朋友與夥伴。

摘要

音樂經驗向來被視為是純聽覺的感受，導致視覺影像對音樂經驗的影響常被忽略。本文操弄音樂的大小調，探討影像是否會影響聆聽音樂時的正負向情緒強度。實驗一播放速度與節奏相同的大、小調音樂給參與者聆聽，確立台灣的參與者如同西方的參與者，會將大調樂曲評定為傳達正向情緒而小調樂曲傳達負向情緒。實驗二操弄音樂與影像在大小調的一致性與否，發現影音在大小調上一致時，對於音樂所感受到的正負情緒較強烈（一致性效果）。實驗三的結果進一步顯示，欣賞情緒一致的影音片段比單獨聽音樂還能感受到更強烈的情緒訊息（增強效果）。這些結果顯示表演者的視覺影像可以調節對音樂情緒的感知;當影音兩者的情緒意涵一致時可以增強所知覺到的情緒強度。Overy 與 Molnar-Szakacs (2009)所提出的 SAME 模型可適度解釋本研究的結果。

關鍵詞：視聽整合、多感官、歌唱、音樂、情緒。

Audio-visual Integration Modifies Emotional Judgment in Music

Shen-Yuan Su

Abstract

The conventional view that perceived emotion in music is derived mainly from auditory signals has led to neglect of the contribution of the visual image. In this study, we manipulated mode (major vs. minor) and examined the influence of a video image on the music. Melodies in either major or minor mode were controlled for tempo and rhythm and played to the participants. We found that Taiwanese participants, like Westerners, judged major melodies as expressing positive, and minor melodies negative, emotions (Experiment 1). The major or minor melodies were then paired with video images of the singers, which were either emotionally congruent or incongruent with their modes. Results showed that participants perceived stronger positive or negative emotions with congruent audio-visual stimuli (i.e., the congruency effect, Experiment 2). Compared to listening to music alone, stronger emotions were perceived when an emotionally congruent video image was added (i.e., the enhancement effect) and weaker emotions

were perceived when an incongruent image was added (i.e., the attenuation effect) (Experiment 3). These results can be explained by the SAME model of Overy and Molnar-Szakacs (2009).

Keywords: audio-visual integration, multisensory, singing, music, emotion.

CONTENTS

Chapter 1 Introduction

Music has long been considered a temporal form of art belonging to the *auditory* domain only. Recently, however, researchers have found that performers' body movements, in addition to auditory signals from music, can affect the perceiver's musical experience. In other words, visual images of the performer can also affect the musical experience of the perceiver (Schutz, 2008; Schutz & Lipscomb, 2007).

One of the essential ingredients of human musical experience is the emotion expressed or induced by music. As is well known, music has great power to convey specific emotional meanings and influence our feelings (Gabrielsson & Juslin, 2002; Gabrielsson & Lindstrom, 2001; Krumhansl, 2002; Sloboda & Juslin, 2001). For example, singing with rich emotional expression is typical in opera. Vocalists express their feelings through voice, which can be perceived by the listeners (Buelow, 1983; Juslin & Laukka, 2003).

The neglect of visual images in musical performance so far has resulted in there being few investigations into the relationship between images of performance and perceived emotions in music. In fact, studies have shown that participants could discriminate different emotions expressed by musicians even without any auditory input.

Body movements, for example, can convey emotions even more accurately than the music *per se* (Davidson, 1993). Recent studies have shown that specific emotions can be conveyed not only by the acoustic properties of music, but also by the performers' facial expressions, body movements, and gestures (Livingstone, Thompson, & Russo, 2009; Thompson, Russo, & Quinto, 2008; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2005; Vines, Krumhansl, Wanderley, & Levitin, 2006). Few studies, however, investigate whether music and visual image can be integrated to create a combined emotional effect.

Because of the close link between music and video images in entertainment industries, audio-visual integration in emotional perception has become a hot topic in studies of film and music (Shevy, 2007; Thompson, Russo, & Quinto, 2006; Vines et al., 2005; Vines et al., 2006). For example, researchers found that happy music could make a visual target in a movie appear emotionally more positive, even though it was judged as negative when presented alone (Shevy, 2007). Vines et al. (2006) used a dynamic continuing judgment task while playing the music to the participants, and asked them to compare the dynamic emotional difference (tension) between listening to the music alone (audio-only) and listening to music while watching the video (audio-visual). Their results showed that audio-visual combination did not have an enhancement effect on tension compared to music alone, and it was rated as having a weaker strength of

tension in some conditions. The results seemed to suggest audio dominance in perceived tension for musical performance. In their earlier study (Vines et al., 2005), they manipulated different performing styles (immobile, standard, and exaggerated) and asked participants to rate the perceived emotion in the conditions of music-alone, video-alone, or both. Factor analysis was used to transfer the emotional rating data into four emotional dimensions: active-positive, active-negative, passive-positive and passive-negative. They found that, compared to the music-alone condition, stronger emotional magnitude was found in the audio-visual condition with the exaggerated performing style in the active-positive dimension. Together, these two studies indicate that, in general, adding a visual image to music does not lead to a robust enhancement effect in perceived emotion.

Because they used modern music with irregularity in mode and rhythm, the lack of emotional enhancement effect might have resulted from the irregular musical components. It is possible that some critical musical components, such as specific mode and regular rhythm, are modulators of emotional audio-visual integration in music which could enhance the emotional effect compared to listening to music alone. Little is known, however, about the role of mode on audio-visual integration in music.

In this article, we focus on one characteristic of musical structure, the mode (i.e.,

major or minor), and examine the following issues in three experiments: (1) whether the mode of the music can have an effect on the perceived emotional valence (positive or negative) of music; (2) whether the congruency of mode between music and video is critical for the perceived negative or positive emotion; and (3) whether the combination of music and video has a greater effect on perceived emotion than listening to music alone.

Chapter 2 Experiment 1

Gabrielsson and Lindstrom (2001) reviewed studies on the relationship between musical components (e.g., mode, rhythm, harmony, etc.) and perceived emotions. Mode was found to be one of the components that can convey different emotional content; major mode conveys happiness and minor mode conveys sadness. Rhythm is another. Different complexity of rhythm leads to different levels of arousal; complex and varied rhythms lead to higher arousal than flowing ones. However, these findings came from studies using western music as the stimulus, and thus it is still a question whether these results can be generalized to Taiwanese participants.

The issue of whether culture affects perceived emotion in music in terms of mode manipulation has not yet reached consensus. The same musical form was found to convey different emotional meanings in different cultural contexts (Gregory & Varney, 1996), indicating that the emotions expressed by music could be modified by the listener's knowledge and life experience, and thus be perceived differently. However, other researchers found that mode is an efficient musical cue that conveys specific universal emotional connotations and is perceived similarly by people across different cultures (Balkwill & Thompson, 1999; Hoshino, 1996).

Experiment 1 examined whether mode affects emotional judgment in music, for Taiwanese participants, such that the major mode conveys positive emotion and the minor mode negative. The purpose was two-fold. One was to test whether the relationship between mode and emotional valence can be universally held, for both western and Taiwanese participants. If this held true, the other purpose was to provide the basis for testing audio-visual integration in emotional-valence judgment on musical performance (i.e., singing) in the next two experiments.

In order to investigate the link between mode and valence, we chose to control another musical component, that is, rhythm, since both mode and rhythm are shown to affect emotional judgment in music (Gabrielsson & Lindstrom, 2001). This manipulation let us explore the relation between mode and valence in music perception without interference from rhythmic variety, and allowed us to generate appropriate stimuli that could be used in the following experiments.

2.1 Method

Participants

Seventeen Taiwanese participants (eight males and nine females with mean ages of 21.7, ranged from 18-26 years old) participated in this experiment. The mean years of

musical training were 1.91 (range = 0-7 years) in the sample. Each participant reported normal or corrected to normal vision and hearing.

Stimuli

We selected six major-scale and six minor-scale tunes in the pool of stimuli from a pilot study[1] and then generated a new set of 12 major and 12 minor tonic melodies with equal rhythms of each pair that underwent relative major-minor or minor-major modulations. For example, a C major tune (labeled **C**) was modulated to become a relative minor tune (labeled **a'**), and a minor tune (labeled **a**) was modulated to become a relative major tune (labeled **C'**). All tunes were composed with the same length in measures (12 bars) and tempo (90 beats per minute) and synthesized with a sine wave sound and played in stereo channels. The 24 melodies were presented in a random order.

Procedure

The participant sat in a quiet dimly lit chamber and the music was played through an earphone. In the practice session, two Taiwanese folk songs were played. In the experiment session, each participant listened to 24 melodies and was asked, after each

[1] In a pilot study, we composed 12 major tonic and 12 minor tonic melodies that had the same measures and tempo. Fourteen Taiwanese participated in the experiment and each one was asked to judge the valence expressed by the music in a 7-point Likert's scale. The results indicated that Taiwanese participants evaluated major tunes as positive (mean=4.941, SD=.383; $F(1,13) = 359.027$, $p < .001$) and minor tunes as negative (mean = 3.042, SD = .454; $F(1,13) = 62.442$, $p < .001$). In this pilot study, rhythm was not controlled.

tune ended, to evaluate the emotional valence conveyed by the music in a 7-point Likert's scale (1 to 7, 1 means very negative, 4 means neutral and 7 means very positive). The experiment took 20-30 minutes to complete.

2.2 Results and Discussion

The relation between musical mode and perceived emotion was analyzed by a one-way ANOVA with a within-subject factor of mode (major, minor). The main effect of mode ($F(1,16) = 76.737$, $p < .001$) indicates that the emotional ratings of major and minor tunes were different. The mean rating value was 4.85 (SD = .516) for major and 3.03 (SD = .652) for minor tunes, and both were rated as significantly different from neutral ($F(1,16) = 43.034$, $p < .001$ in major mode; $F(1,16) = 31.641$, $p < .001$ in minor mode).

The 7-point rating scale was transformed by subtracting the rated value from 4 (i.e., the neutral point), thus making a positive rating value correspond to a positive valence, and a negative one to a negative valence. The results were plotted and shown in Figure 1.

With rhythm as a well-controlled factor, and by manipulating the mode of the music, we found in this experiment the linkage between mode and perceived emotional

valence for Taiwanese participants. The participants gave higher rating values to major tunes and lower values to minor tunes, indicating that they indeed perceived positive and negative emotion when they heard major and minor musical modes, respectively. This result is consistent with previous studies (Balkwill & Thompson, 1999; Hoshino, 1996). More importantly, the major-positive and minor-negative links derived from our musical stimuli provided the material base for investigating audio-visual integration in the following experiments.

Chapter 3 Experiment 2

After establishing the correspondence of mode and perceived emotion in music, we proceeded to test whether congruence between visual images of the singer and the mode of the music played affects the perceived emotional valence differently from an emotionally incongruent audio-visual combination. We selected two major tonic and two minor tonic tunes to be the main audio stimuli and video images (two singers performing these tunes were recorded to create the visual stimuli). We manipulated the congruence in mode of the bimodal stimuli by combining the visual image (a silent video clip in which a singer sang a major or a minor tune) and the music played (major or minor).

3.1 Method

Participants

Sixteen Taiwanese (seven males and nine females with mean ages of 22.4, ranged from age 19-25) participated in the experiment. The mean years of musical training were 1.8 (range = 0-6 years) in the sample. Each participant reported normal or corrected to normal sight and normal hearing ability.

Stimuli and conditions

Four melodies, including **C**, **G**, **a**, **e,** were selected from our MIDI music pool and eight videos were prepared, sung by two singers (labeled $\mathbf{C_1}$, $\mathbf{C_2}$, $\mathbf{G_1}$, $\mathbf{G_2}$, $\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{e_1}$ and $\mathbf{e_2}$; the number indicates different vocalists). The two singers were both female and trained in opera for at last two years. They were asked to sing the tune and express the emotion that was appropriate for the music they heard. All images presented the singer's full facial expression and body movements from the shoulders up. The videos were post-produced by iMovie HD and the soundtrack was deleted to form a silent version (labeled as $\mathbf{sC_1}$, $\mathbf{sC_2}$, $\mathbf{sG_1}$, $\mathbf{sG_2}$, $\mathbf{sa_1}$, $\mathbf{sa_2}$, $\mathbf{se_1}$ and $\mathbf{se_2}$; where the letter **s** stands for "silent"). The resolution of the video images was 640 x 480 pixels.

The bimodal stimuli in which mode and emotion were congruent were created by combining the silent videos ($\mathbf{sC_1}$, $\mathbf{sC_2}$, $\mathbf{sG_1}$, $\mathbf{sG_2}$, $\mathbf{sa_1}$, $\mathbf{sa_2}$, $\mathbf{se_1}$ and $\mathbf{se_2}$) and the sung melodies (**C**, **G**, **a** and **e**), which were synthesized by a virtual vocal sound. The visual movements of the singer in the video were synchronized with the melody in each pair.

The incongruent bimodal stimuli were made with the same process mentioned above but the four tunes **C'**, **G'**, **a'** and **e'** which had been modulated from **a**, **e**, **C** and **G** were used instead. We generated 8 incongruent pairs, including $\mathbf{sC_1}$ + **a'**, $\mathbf{sC_2}$ + **a'**, $\mathbf{sG_1}$ + **e'**, $\mathbf{sG_2}$ + **e'**, $\mathbf{sa_1}$ + **C'**, $\mathbf{sa_2}$ + **C'**, $\mathbf{se_1}$ + **G'**, and $\mathbf{se_2}$ + **G'**. Each trial contained either a

congruent or an incongruent bimodal pair that lasted 36 seconds.

Two within-subject factors were manipulated and four conditions were formed: mode (major/minor) x congruency (congruent/incongruent). Each condition included four trials, two for judgment of music and two for judgment of video. The 16 trials were presented in a random order in the experimental session.

Procedure

In the practice session, participants watched and listened to two videos played with music of two Taiwanese folk songs to familiarize themselves with the response procedure. In the experimental session, participants were asked to pay attention to both audio and visual stimuli and judge the valence expressed by either the video or the music. When the trial ended, a question was presented on the monitor: “What is the emotion expressed by the music?” or “What is the emotion expressed by the video?” Participants needed to enter an appropriate number reflecting the perceived valence based on a 7-point Likert’s scale. Other details in the experimental procedure were equal to Experiment 1.

3.2 Results and Discussion

Experiment 2 aimed to examine the mode congruence effect to see whether an

incongruent audio-visual pair could modify the pattern of perceived emotional valence in both major and minor mode, compared to a congruent pair. Accordingly, the main effects were less useful in drawing meaningful conclusions. In this and the following experiments, we will thus focus on the interaction effect, which reflects the pattern change in the emotional judgment, and the simple main effects, which reflect the emotional difference between congruent and incongruent pairs for major and minor mode respectively.

Emotional judgment on music

A two-way repeated-measures ANOVA (mode x congruency) was used to analyze the data. The main effect of mode was significant ($F(1,15) = 5.713$, $p < .05$), but the main effect of congruency was not significant ($F(1,15) = .008$, $p > .05$). The interaction effect was significant ($F(1,15) = 9.732$, $p < .01$). Post-hoc analysis indicates that the emotional rating of congruent stimuli (both in judgment of major and minor tunes) reflected stronger magnitude than the incongruent ones ($F(1,30) = 7.130$, $p < .05$ in major mode; $F(1,30) = 7.613$, $p < .01$ in minor mode). See Figure 2 for the results.

Emotional judgment on video image

As displayed in Figure 3, the main effects of mode and congruency were significant ($F(1,15) = 23.661$, $p < .001$; $F(1,15) = 10.113$, $p < .01$). These significant

main effects indicate that the video images of a performer singing a major tune were rated with higher scores than those in which she sang a minor tune, regardless of the actual mode of the music; and incongruent music caused video images to be perceived as more negative than congruent music. Planned comparison showed that for video images of performance of major tunes, incongruent stimuli were rated as having a weaker emotional magnitude than congruent ones ($F(1,15) = 6.713$, $p < .05$).

Hence, the mode congruence effect was demonstrated in judging major and minor tunes, and in judging video images of someone singing major tunes. What is more important is the effect of the video image on the music, because perceived emotion in music is the main concern here. Also, the absence of the congruence effect in the judgment of the video image in which the minor tune was sung might be due to our manipulation of the correspondence in musical mode and video image, and emotional valence was indirectly inferred from this correspondence. In Experiment 3 we test more directly the relationship between mode and valence.

In summary, we have successfully demonstrated the congruency effect of video images on music in Experiment 2, by comparing incongruent pairs with congruent ones. The emotional difference between congruent and incongruent stimuli in music may result from (1) emotional *attenuation* by mode *incongruent* stimuli, or (2) emotional

enhancement by mode *congruent* stimuli. In the next experiment, we test whether video image affects emotional judgment of music through *attenuation* or *enhancement*, by comparing congruent or incongruent pairs with a unimodal stimulus.

Chapter 4 Experiment 3

In Experiment 2, we manipulated the audio-visual congruency between modes, and asked the participants to judge, based on the bimodal stimuli, the emotional valence of either the music or the video image. When we recorded the video images, the singers listened to the major and minor music and were asked to “express emotion appropriately based on the music heard.” The performer relied on her understanding of music to express emotional cues that could be decoded by the listeners (Juslin, 2001). The mode congruence effect between music and video images found in Experiment 2 might be due to a correspondence in the musical structure, with emotional valence being indirectly inferred from this correspondence. To directly link the mode in which the music was performed (an audio-visual event) with emotion, which is more relevant to the musical experience of most people, the listeners’ emotional judgment of the music and video image *per se* are necessary to confirm their emotional valence. In Experiment 3A the participants were asked to rate the emotional valence of each unimodal stimulus (music alone or image alone) and we chose the stimuli that were rated as positive or negative by these participants to form the stimulus set in Experiment 3B and 3C.

Emotionally congruent pairs were compared with unimodal stimuli (music only or

video only) to test the enhancement effect in Experiment 3B. Emotionally incongruent stimuli were compared with unimodal stimuli in Experiment 3C, for the following two purposes: (1) to examine whether the emotional enhancement effect in Experiment 3B, if it exists, results from audio-visual emotional congruency, or whether any added video images would induce the enhancement effect on music; and (2) to test whether incongruent pairs would cause an attenuation effect on perceived emotion in music.

4.1 Experiment 3A

To confirm that our unimodal stimuli did indeed convey different emotional valences as intended and were perceived as such, we asked the participants to rate the emotional valence of each virtual song and silent video. Based on the subjective ratings of emotional valence of each unimodal stimuli, we could select stimuli that were rated as significantly positive or negative in emotional valence in order to test the enhancement or the attenuation effect of perceived emotion in music in the next two experiments.

4.1.1 Method

Participants

Sixteen Taiwanese (eight males and eight females with mean ages of 20.9, ranged from age18-24) participated in Experiment 3A. Their mean years of musical training were 4.06 (range = 0-15 years) in the sample. Each participant reported normal or corrected to normal sight and normal hearing ability.

Stimuli

All stimuli were selected from Experiment 2. Two types of block were designed: audio-only and visual-only. The audio-only block included eight virtual songs (**C**, **G**, **a**, **e, C'**, **G'**, **a'** and **e'**) and the visual-only block included eight silent videos (**sC$_1$**, **sC$_2$**, **sG$_1$**, **sG$_2$**, **sa$_1$**, **sa$_2$**, **se$_1$** and **se$_2$**), and they were randomized within each block in the experiment session. The presentation order of the audio-only and visual-only blocks was counterbalanced between participants.

Procedure

In the practice session, one familiar Taiwanese folk song and one silent video image with the singer performing another folk song were used to let the participants familiarize themselves with the response procedure. In the experiment session, participants were asked to pay attention to the music in the audio-only block, and pay

attention to the video images in the visual-only block. They then judged the emotional valence expressed by the stimuli at the end of the trial. Other details in experimental procedure were the same as in Experiment 2.

4.1.2 Results

Table 1 displays the mean rating values of emotional judgment of stimuli and the t scores (the mean rating values of stimulus compared to the neutral value of emotional valence). Results showed that in emotional judgment of music, major tunes (**C**, **G** and **C'**) were perceived as emotionally positive and minor tunes (**a**, **e**, **a'** and **e'**) as emotionally negative, but **G'** was perceived as neutral. The silent videos $\mathbf{sG_1}$ and $\mathbf{sa_2}$ were rated as neutral and other silent videos displayed either major-positive or minor-negative relations.

4.2 Experiment 3B

In Experiment 3B, we selected appropriate stimuli from the results of Experiment 3A to test whether an emotional enhancement effect exists when participants both see and listen to an emotionally congruent clip, as compared to the unimodal stimuli.

4.2.1 Method

Participants

Twenty Taiwanese participants (nine males and eleven females with mean ages of 22.7, ranged from age 16-27) participated in Experiment 3B. The mean years of musical training were 2.9 (range = 0-17 years) in the sample. Each participant reported normal or corrected to normal sight and normal hearing ability.

Stimuli

According to the results of Experiment 3A, we selected **G** and $\mathbf{sG_2}$ as positive unimodal stimuli and **a** and$\mathbf{sa_1}$ as negative unimodal ones. The emotionally congruent pairs were created by using the pairs (**G**, $\mathbf{sG_2}$) and (**a**, $\mathbf{sa_1}$). Unimodal stimuli were repeated twice and eight trials were generated that were presented in a random order in the unimodal block. Emotionally congruent stimuli were repeated four times, two judged on the basis of the music and two on the basis of the video, and the eight emotionally congruent trials were also ordered randomly within the bimodal block.

Procedure

In the practice session, two familiar Taiwanese folk videos were used to familiarize the participants with the response procedure. In the experiment session, participants were asked to pay attention to either the music or the video in the unimodal block, and

then to judge the emotional valence expressed by each unimodal stimulus. In the bimodal block, participants were asked to pay attention to both music and video images, and then to judge the emotional valence expressed by *either* the music *or* the video images. We presented the two blocks sequentially (always unimodal first) to avoid potential confounding from the other modality if bimodal stimuli were presented first. Other details were the same as in Experiment 2.

4.2.2 Results

Emotional judgment on music

A two-way repeated measure ANOVA with the factors of valence (positive, negative) and modality (unimodal, bimodal) was conducted. As Figure 4 displays, the main effects were significant ($F(1,19) = 121.435$, $p < .001$, for the factor of valence; $F(1,19) = 5.429$, $p < .05$, for the factor of modality). The significant interaction effect ($F(1,19) = 5.074$, $p < .05$) and significant simple main effect for positive valence ($F(1,38) = 10.096$, $p < .005$) indicate that emotionally congruent stimuli were rated higher, namely, more positive, than audio-only stimuli for positive valence. However, for negative valence the emotional rating was identical between congruent and audio-only stimuli ($F(1,38) = 0.280$, $p > .05$).

Emotional judgment on video image

As shown in Figure 5, in the emotional judgment of the video image, the main effects were significant ($F(1,19) = 111.865$, $p < .001$, for the factor of valence; $F(1,19) = 6.867$, $p < .05$, for the factor of modality). This indicates that the positive video was rated as positive, and the negative video was rated as negative, for both unimodal and bimodal stimuli; and the silent video was rated at a higher value than congruent stimuli regardless of emotional valence. The interaction effect was not significant ($F(1,19) = 0.375$, $p > .05$). It remains to be seen whether video images combined with emotionally congruent music could enhance perceived emotion in emotional judgment of video images.

4.3 Experiment 3C

In Experiment 3C, we compared emotionally incongruent stimuli with unimodal stimuli to investigate whether incongruent stimuli could attenuate the magnitude of emotional valence, and whether any audio-visual combination could enhance the perceived emotional magnitude compared to unimodal stimuli.

4.3.1 Method

Participants

Twenty Taiwanese participants (eight males and twelve females with mean ages of 22.9, ranged from age 19-29) participated in Experiment 3C. The mean years of musical training were 3.2 (range = 0-11 years) in the sample. Each participant reported normal or corrected to normal sight and normal hearing ability.

Stimuli and procedure

In Experiment 3C, the unimodal block included **e'**, **C'**, $\mathbf{sG_2}$ and $\mathbf{sa_1}$ and the bimodal block included emotionally incongruent stimuli created by using the pairs (**e'**, $\mathbf{sG_2}$) and (**C'**, $\mathbf{sa_1}$). Other details were the same as in Experiment 3B.

4.3.2 Results

Emotional judgment on music

The main effect of valence was significant ($F(1,19) = 36.578$, $p < .001$), but not the main effect of modality ($F(1,19) = 0.905$, $p > .05$). The significant interaction effect ($F(1,19) = 17.336$, $p < .001$) and the significant simple main effects for positive valence ($F(1,38) = 6.281$, $p < .05$) and for negative valence ($F(1,38) = 14.132$, $p < .001$) indicate that emotionally incongruent stimuli attenuated the emotional magnitude compared to

music played alone, for both positive and negative emotions (see Figure 6).

Emotional judgment on video images

Similarly, the main effects of valence and modality were significant ($F(1,19) = 49.237$, $p < .001$, for valence; $F(1,19) = 0.059$, $p > .05$, for modality). The significant interaction ($F(1,19) = 17.318$, $p < .001$) and the significant simple main effect ($F(1,38) = 10.885$, $p < .01$, for positive valence; $F(1,38) = 8.880$, $p < .01$, for negative valence) indicate that emotionally incongruent stimuli were rated weaker in emotional strength than seeing the silent video alone (see Figure 7).

4.4 Discussion of Experiment 3

Experiment 3A shows that, except for the silent videos $\mathbf{sG_1}$ and $\mathbf{sa_2}$ which were perceived as neutral, the mode of the music was still sufficient to provide emotional valence information for performers to express emotion that was consistent with the heard music through facial expressions and body movements. Without instructing the singers to express specific emotions in their performance, 75% of the silent videos could still convey emotional information in a correspondent manner (i.e., when the singer sang a major tune, her video image was perceived to be emotionally positive, and vice versa). Emotional intention in music is conveyed by performers based on their

facial expression, body movements, and dynamic acoustic cues from voicing. These visual and audio cues could be decoded by their audience in the process of perceiving the emotion in music (Juslin, 2001). The mode, one structural component of the music, could supply the emotional intention in order for performers to perform accurately, even without other instructions regarding specific emotion. The information derived from mode thus seems important in communicating the emotional connotation of music to both performer and audience. More importantly, this provides the basis to directly link emotion to mode for the unimodal stimuli to be used in comparison with the bimodal stimuli in the next two experiments.

In Experiment 3B and 3C, while the participants were asked to judge the perceived emotion in music, congruent stimuli enhanced the perceived strength of positive emotion compared to listening to music alone (Experiment 3B), and incongruent stimuli attenuated it for both positive and negative emotions (Experiment 3C). Generally speaking, the pattern of results suggests that more intense emotion can be perceived with the addition of a visual image than that obtained from listening to music alone.

Spreckelmeyer, Kutas, Urbach, Altenmüller, and Münte (2006) used event related potentials (ERPs) to investigate audio-visual integration in emotional valence. They found that happy pictures paired with happy voices could evoke stronger P2 potential

than happy-neutral pairs, but the difference of P2 potential between sad pictures paired with sad voices and those paired with neutral ones could not be displayed. The P2 potential indicated early audio-visual integration in emotion, and the existence of P2 difference in positive congruent pairs rather than negative congruent ones in their research is consistent with our results. Although the enhancement effect of negative emotion in the congruent pair did not reach statistical significance, there is a trend toward that direction, as can be seen from Figure 4. In fact, this is consistent with what has usually been observed in the literature: more robust incongruent effects than congruent effects (e.g., de Gerder & Vroomen, 2000). As we tried to use stimuli that are more similar to natural ones rather than exaggerated forms (cf. Vines et al., 2005), our finding of an emotional congruence effect on positive emotions is valuable and informative.

In the emotional judgment of the video images, the effect of music on the video was a bit more complicated. The insignificant interaction in Experiment 3B implies that the emotional enhancement effect might be absent in visual judgment. Nevertheless, the attenuation effect demonstrated in Experiment 3C indicates that when compared to the video image, emotionally incongruent music would eliminate the emotional magnitude of the perceived visual expression.

Chapter 5 General Discussion

Several important findings were obtained in this study. First, information relating to the mode of the music is a strong factor affecting our perceived emotion when listening to music. Experiment 1 first established this point by using music in major or minor mode and showed robust relationships between positive emotions and major mode, and between negative emotions and minor mode. Although this mode-emotion relationship has already been shown for western participants (Gabrielsson & Juslin, 2002; Gabrielsson & Lindstrom, 2001) and seems to be a commonly held view in western society, no empirical evidence so far has been provided for Taiwanese. Our finding of the mode-emotion relationship in Taiwan not only provides evidence consistent with the literature in western societies, but also supports the view that even with the influence of different cultural backgrounds, the mode of music conveys emotional valence, perhaps across different cultures (Balkwill & Thompson, 1999; Hoshino, 1996).

Furthermore, when we asked the participants to directly evaluate the emotional valence of each unimodal stimulus in Experiment 3A, although there were some stimuli rated as neutral in the silent video, the mode information of the music could still help the performer to express appropriate emotional intention which could also be detected

by the participant. The mode of the music was implicitly detected by listeners, but it still has power to influence perceived emotion and form a medium supplying the performer with the ability to communicate the emotional intention of the music to the audience through visual performance, even without acoustic emotional cues.

The second important finding is that congruency in mode between the video image and music mediates audio-visual integration in music. Based on the results of Experiment 1, Experiment 2 investigated whether an incongruent pairing of music and video (vs. a congruent pair) could modify our emotional magnitude (i.e. the mode congruence effect) in perceiving musical performance. The mode congruence effect demonstrated in Experiment 2 indicates that emotional information processed from one modality can be affected by information processed from the other modality, and that information from both modalities integrates to modify our emotional responses (de Gelder & Bertelson, 2003; Shams & Seitz, 2008).

Comparing our results with other studies, the mode congruence effect reflects that the musical component, mode, is not only important in musical perception but is also a medium that conveys emotional connotation visually. A musical piece which lacks mode information is thus likely to lose that medium with which to convey emotional intention. Accordingly, the result of audio dominance in perceived tension in Vines et al. (2006)

might have resulted from the fact that the musical stimuli they used did not have clear mode information.

The third important finding is that the combination of music and video gives us stronger perceived emotion than listening to music alone. The results in Experiment 3B and 3C indicate that emotionally congruent visual information could enhance the perceived positive emotion of music, whereas emotionally incongruent visual information could attenuate both positive and negative emotions. Judging from the results of the emotional judgments of both music and video, the possibility of visual dominance of perceived emotion can be excluded. Had the emotional enhancement effect resulted from the visual aspect being dominant, how could the attenuation of emotionally incongruent stimuli in visual judgment of music be explained? The emotional modulation effect from audio-visual emotional integration has also been found in other cross-modal studies on music (Baumgartner, Lutz, Schmidt, & Jancke, 2006; Shevy, 2007; Spreckelmeyer et al., 2006; Thompson et al., 2008; Vines et al., 2005). However, most studies explore the effect by examining the influence of the combination of music with a picture (or movie clip) but not videos of musical performance (Baumgartner et al., 2006; Shevy, 2007; Spreckelmeyer et al., 2006). Although Vines et al. (2005) found that, compared to listening to music alone, images of

exaggerated visual performance could enhance the perceived emotion, the effect might have resulted from the artificial manipulation of visual cues which were unrelated to the musical performance itself. In our study, we asked the performer to sing appropriately with regard to the music heard, without any exaggerated or unrelated expression. Our finding that an emotionally congruent video image could still enhance the perceived emotional magnitude points to a strong connection between mode and emotion.

Thompson et al. (2008) investigated whether the emotional congruence between a vocalist's dynamic facial expression and vocal sound in singing a major third or minor third interval could affect the emotional judgment of music. Their results indicated that the congruent pairs had the extreme scores, with the incongruent pairs in between. The results seemed to suggest that emotional judgment of music could be modified by visual information. Despite the fact that they found an emotional congruence effect, the question remains open whether audio-visual integration enhances emotional strength (i.e., cross-modal enhancement) or attenuates it (i.e., audio-dominance), since they did not compare participants' judgments of the congruent conditions with a music-alone condition, as we have done here. Also, the audio-visual stimuli in their experiments are more similar to binding pictures and simple vocal sounds than to real musical performance (their audio signals contained just two notes). This makes their results

difficult to generalize to real–life situations such as a musical concert. We have made an effort in this direction by using stimuli more similar to a musical performance in this study, and the emotional congruence effects found in our Experiments 2 and 3 indicate the robust mode-emotion relationship even across different modalities.

Music has been said to be analogous to motion, because of the dynamic sound flow in music that is associated with the motion generated in music production (Eitan & Granot, 2006). Some researchers have proposed that musical experience is derived from cross-modal processing between audition and visual motion. Through the correspondence of the musical dynamic change and the intention of expressing motion behind the auditory code, we can understand what connotations are conveyed by music (Livingstone & Thompson, 2009; Molnar-Szakacs & Overy, 2006; Overy & Molnar-Szakacs, 2009). Molnar-Szakacs and Overy (2006) reviewed many studies about the relation between music and motion, and they came to the conclusion that the acoustic components of music, such as amplitude variation, rhythm, and contour of melody, were systematically synchronized with performers' motions. Musical experience might thus be generated from the corepresentation of motor programming between audio signal and motion production. Facial expression and body movement express visual emotional cues that are decoded in order to understand the emotional

intention behind the visual image (Thompson et al., 2008). Brain imaging studies show that the functions of detecting emotion in music and understanding what others are thinking are highly associated with the mirror neuron system, which is considered a part of the motor system. As well as music, facial expression and body movements also activate the mirror neuron system (Livingstone & Thompson, 2009; Livingstone et al., 2009). Through sensory-motor transformation by the mirror neuron system, we understand the intention behind the sensory inputs, including emotional connotation.

Accordingly, we suggest that the audio-visual integration we found in this study might result from cross-modal information reorganized in the mirror neuron system. Ovary and Molnar-Szakacs (2009) propose the Shared Affective Motion Experience (SAME) model to explain the mechanism of perceived emotion in music. According to this model, musical signals enter the fronto-parietal mirror neuron system from the temporal and occipital cortex to be decoded and generate motor programming. The information flows to and is modified by anterior insula and is then transported to the limbic system where emotional information is processed. Finally, musical information processed by the neural network form the musical experience and emotional perception. Recent neuroimaging studies support this hypothesis by showing that musical processing activates the mirror neuron system and limbic regions (Blood & Zatorre,

2001; Green, Baerentsen, Stodkilde-Jorgensen, Wallentin, Roepstorff, & Vuust, 2008; Hasegawa et al., 2004; Molnar-Szakacs & Overy, 2006; Peretz & Zatorre, 2005).

Considering visual images as part of musical inputs as shown in this study, the SAME model can explain the cross-modal effect in perceiving the emotion in music performance. However, most brain imaging studies only focus on “music” perception and neglect the close relationship between the music itself and the image of the performance. Future studies can use fMRI and ERP techniques to investigate if the emotional enhancement effect of bimodal stimuli, compared to unimodal ones, reflects a difference in brain activities in the neural network as predicted by the SAME modal. Brain regions such as the frontal-parietal lobe, anterior insula, and limbic system may play an important role in audio-visual integration of emotional perception, an important part of musical experience.

Chapter 6 Conclusion

In this study we have demonstrated that *mode* is important in order for the performer to convey, and for the audience to perceive, emotional valence in music. In addition, the performer's facial expression and body movements are also important because they modify the magnitude of perceived emotion in music. Treating musical art as a purely auditory event might lose the enhanced emotional strength perceived in music, since going to a concert (an audio-visual event) leads to stronger perceived emotion than playing the CD (audition alone) at home.

Chapter 7 References

Balkwill, L. L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17,* 43-64.

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jancke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. *Brain Research, 1075,* 151-164.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 11818-11823.

Buelow, G. J. (1983). Johann Mattheson and the invention of the Affektenlehre. In G. J. Buelow & H. J. Marx (Eds.), *New Mattheson studies* (pp. 393-407). Cambridge: Cambridge University.

Davidson, J. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music, 21,* 103–113.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7*, 460-467.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion, 14*, 289-311.

Eitan, Z., & Granot, R. Y. (2006). How music moves. *Music Perception, 23*, 221-248.

Gabrielsson, A., & Juslin, P. N. (2002). Emotional expression in music. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 503-534). New York: Oxford University Press.

Gabrielsson, A., & Lindstrom, E. (2001). The influence of musical structure on emotional expression. In J. A. Sloboda & P. N. Juslin (Eds.), *Music and Emotion: Theory and Research* (pp. 223-248). New York: Oxford University Press.

Green, A. C., Baerentsen, K. B., Stodkilde-Jorgensen, H., Wallentin, M., Roepstorff, A., & Vuust, P. (2008). Music in minor activates limbic structures: A relationship with dissonance? *Neuroreport, 19*, 711-715.

Gregory, A. H., & Varney, N. (1996). Cross-cultural comparisons in the affective response to music. *Psychology of Music, 24*, 47-52.

Hasegawa, T., Matsukib, K.-I., Uenoc, T., Maedad, Y., Matsued, Y., Konishia, Y., et al. (2004). Learned audio-visual cross-modal associations in observed piano

playing activate the left planum temporale. An fMRI study. *Cognitive Brain Research, 20*, 510-518.

Hoshino, E. (1996). The feeling of musical mode and its emotional character in a melody. *Psychology of Music, 24*, 29-46.

Juslin, P. N. (2001). Communicating emotion in musical performance: A review and a theoretical framework. In J. A. Sloboda & P. N. Juslin (Eds.), *Music and Emotion: Theory and Research* (pp. 309-337). New York: Oxford University Press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770-814.

Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science, 11*, 45-50.

Livingstone, S. R., & Thompson, W. F. (2009). The emergence of music from the Theory of Mind. *Musicae Scientiae, Special Issue*, 83-115.

Livingstone, S. R., Thompson, W. F., & Russo, F. A. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception, 26*, 475-488.

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to 'e'motion. *Social Cognitive and Affective Neuroscience, 1*, 235-241.

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror Neuron System. *Music Perception, 26*, 489-504.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology, 56*, 89-114.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review, 3*, 84-108.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36*, 888-897.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*, 411-417.

Shevy, M. (2007). The mood of rock music affects evaluation of video elements differing in valence and dominance. *Psychomusicology, 19*, 57-78.

Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In J. A. Sloboda & P. N. Juslin (Eds.), *Music And Emotion: Theory And Research* (pp. 71-104). New York: Oxford University Press.

Spreckelmeyer, K. N., Kutas, M., Urbach, T. P., Altenmüller, E., & Münte, T. F. (2006).

Combined perception of emotion in pictures and musical sounds. *Brain Research, 1070*, 160-170.

Thompson, W. F., Russo, F. A., & Quinto, L. (2006). Preattentive integration of visual and auditory dimensions of music. *Proceedings of the Second International Conference on Music and Gesture* (pp. 217-221). Manchester, UK.

Thompson, W. F., Russo, F., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition & Emotion, 22*, 1457-1470.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2005). Dimensions of emotion in expressive musical performance. *Neurosciences and Music II: From Perception to Performance* (Vol. 1060, pp. 462-466).

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*, 80-113.

Chapter 8 Tables

Table 1. Mean rating values of emotional valence for the unimodal stimuli in Experiment 3A.

Stimulus	Mean	SE	t-score	*p*
(A) Music				
A	2.813	0.262	-4.538	.000
e'	2.875	0.301	-3.737	.002
a'	3.000	0.354	-2.828	.013
E	3.188	0.262	-3.105	.007
G'	4.438	0.329	1.331	.203
C	4.875	0.328	2.671	.017
C'	4.938	0.347	2.700	.016
G	5.063	0.295	3.597	.003
(B) Silent video				
sa_1	2.125	0.180	-10.434	.000
se_2	2.437	0.302	-5.168	.000
se_1	3.062	0.193	-4.858	.000
sa_2	3.562	0.302	-1.447	.168
sG_1	4.063	0.295	0.212	.835
sG_2	5.750	0.250	7.000	.000
sC_2	6.000	0.158	12.649	.000
sC_1	6.125	0.202	10.543	.000

Note: Stimuli in (A) the music group and (B) the silent-video group were arranged according to the emotional-rating values (from low to high). Gray marks indicate that the stimuli were rated as emotionally neutral.

Chapter 9 Figures

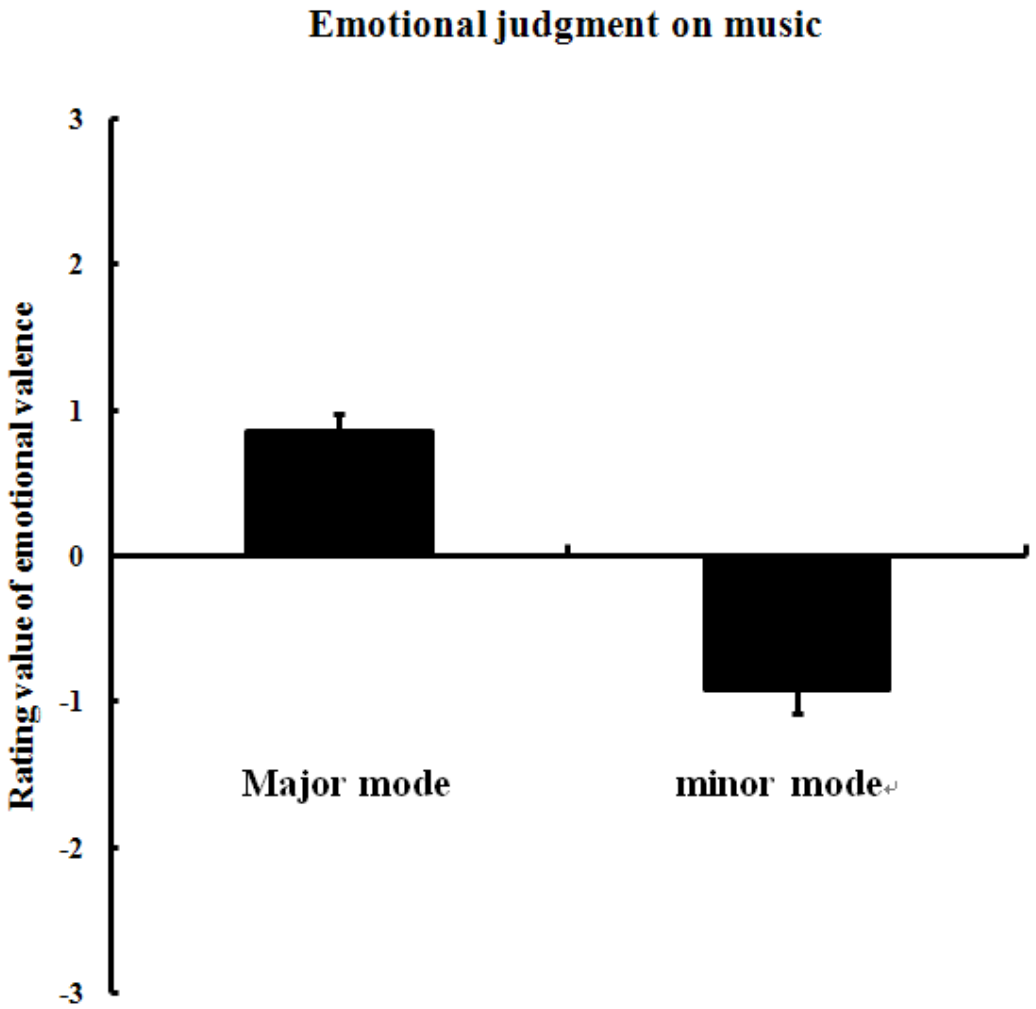

Figure 1. Mean rating values of emotional valence across 12 tonic tunes used in Experiment 1. The error bar indicates one standard error in all the figures.

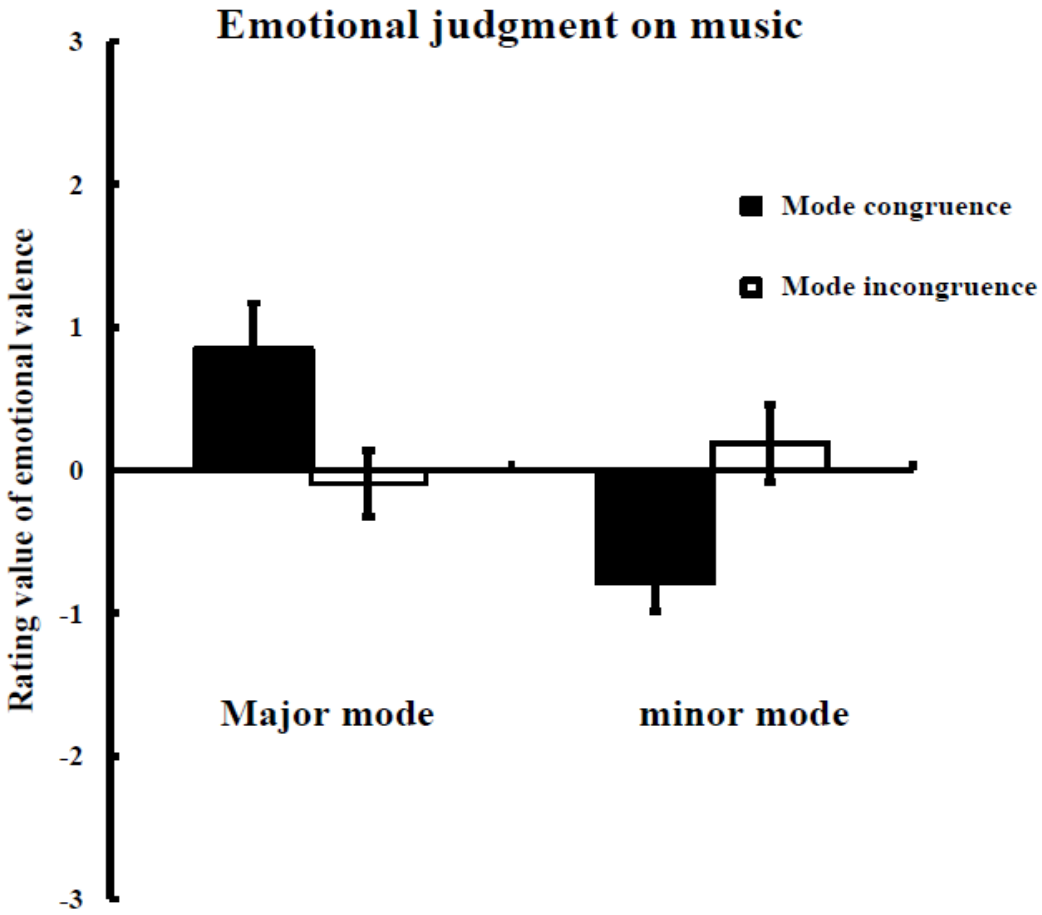

Figure 2. Mean rating values of emotional valence in judgment on music in Experiment 2. Mode congruence: the music and the video image were of the same (major or minor) mode; Mode incongruence: the music and the video image were generated from relative mode.

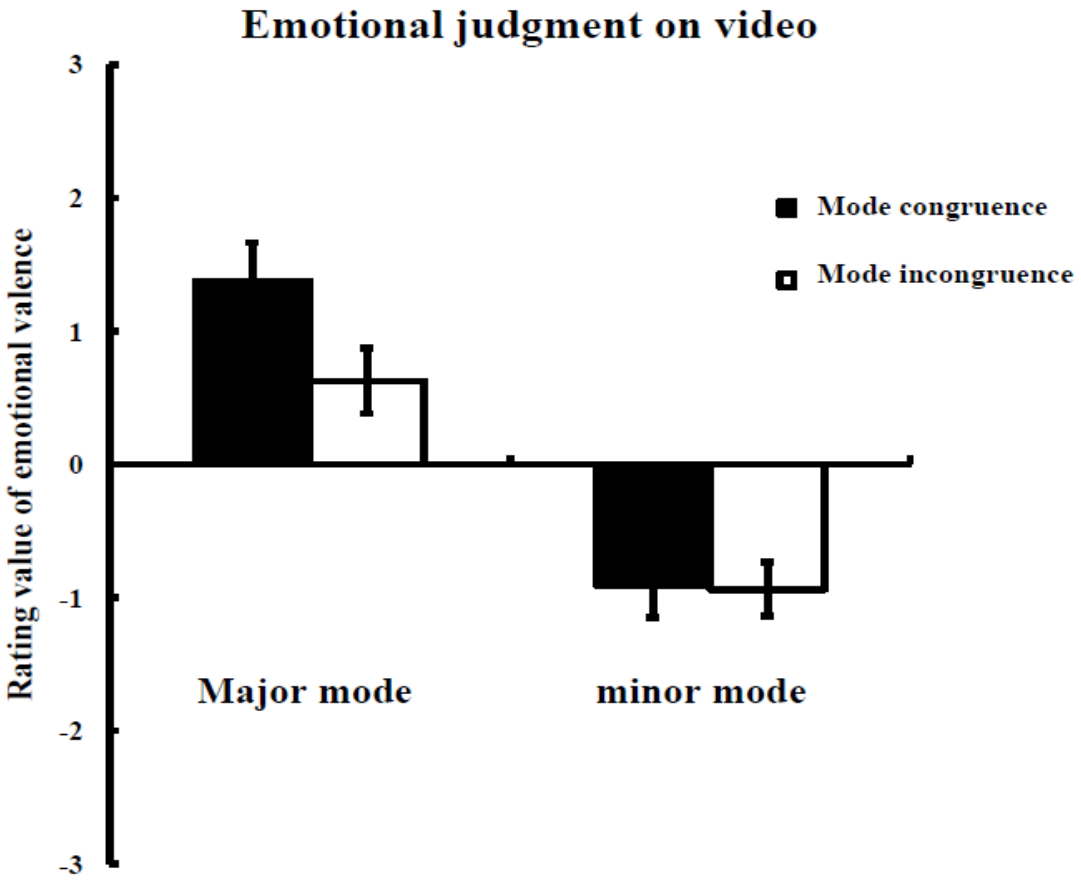

Figure 3. Mean rating values of emotional valence in judgment on video in Experiment 2. Mode congruence: the music and the video image were of the same (major or minor) mode; Mode incongruence: the music and the video image were generated from relative mode.

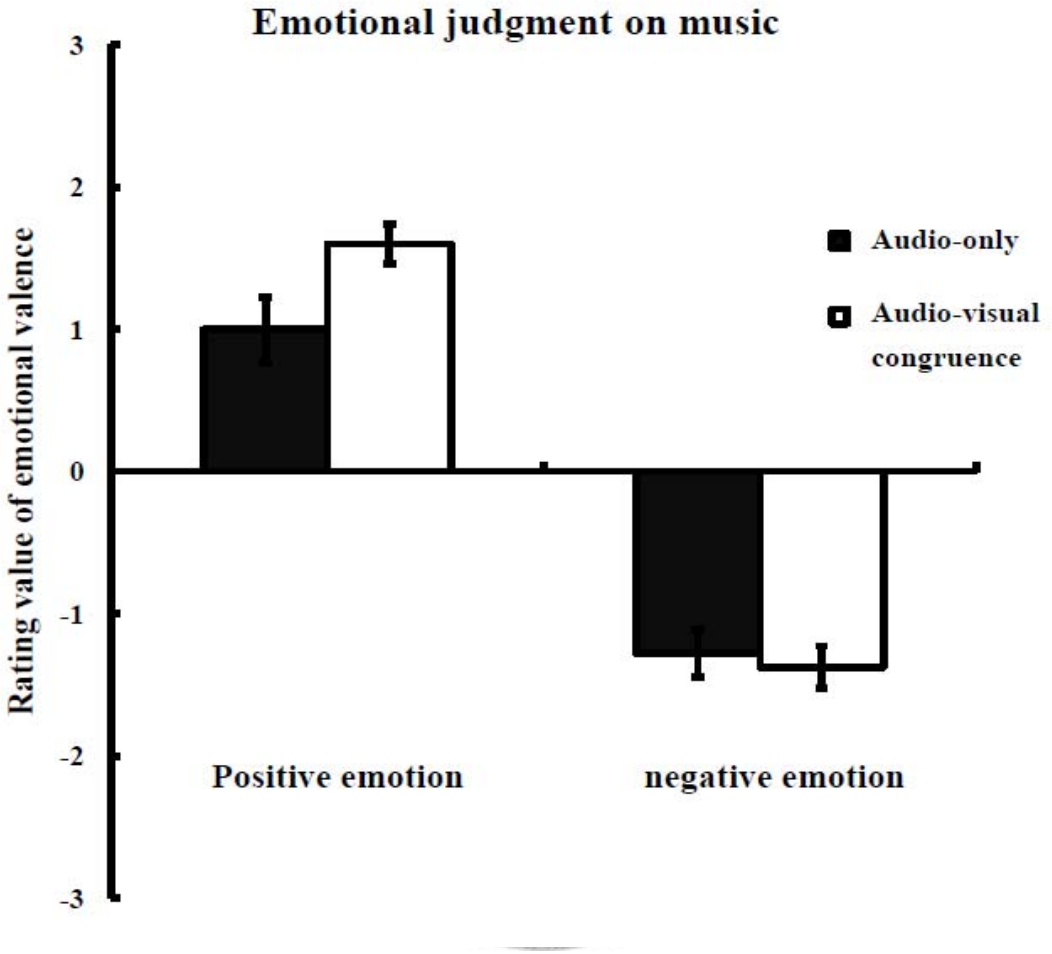

Figure 4. Mean rating values of emotional valence in judgment on music in Experiment 3B. Audio-only: only music was played; Audio-visual congruence: the music and the video image were perceived as the same emotional valence.

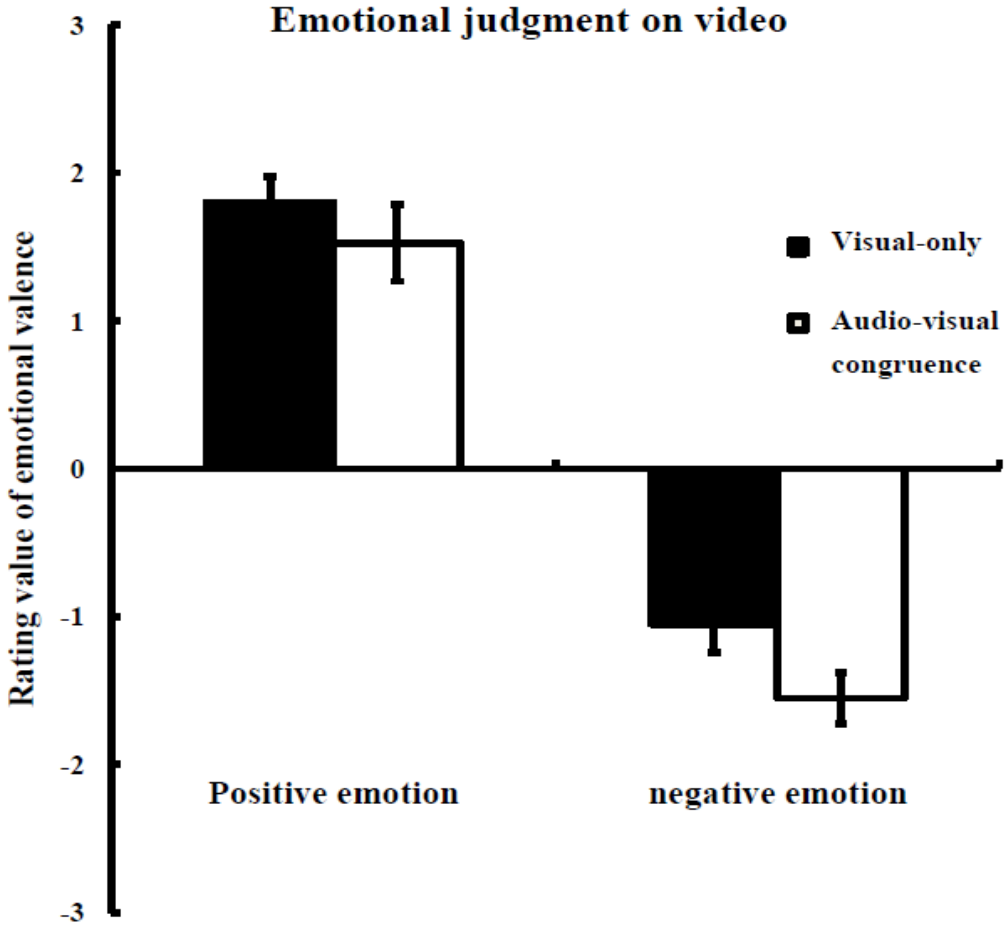

Figure 5. Mean rating values of emotional valence in judgment on video image in Experiment 3B. Audio-only: only music was played; Audio-visual congruence: the music and the video were perceived as the same emotional valence.

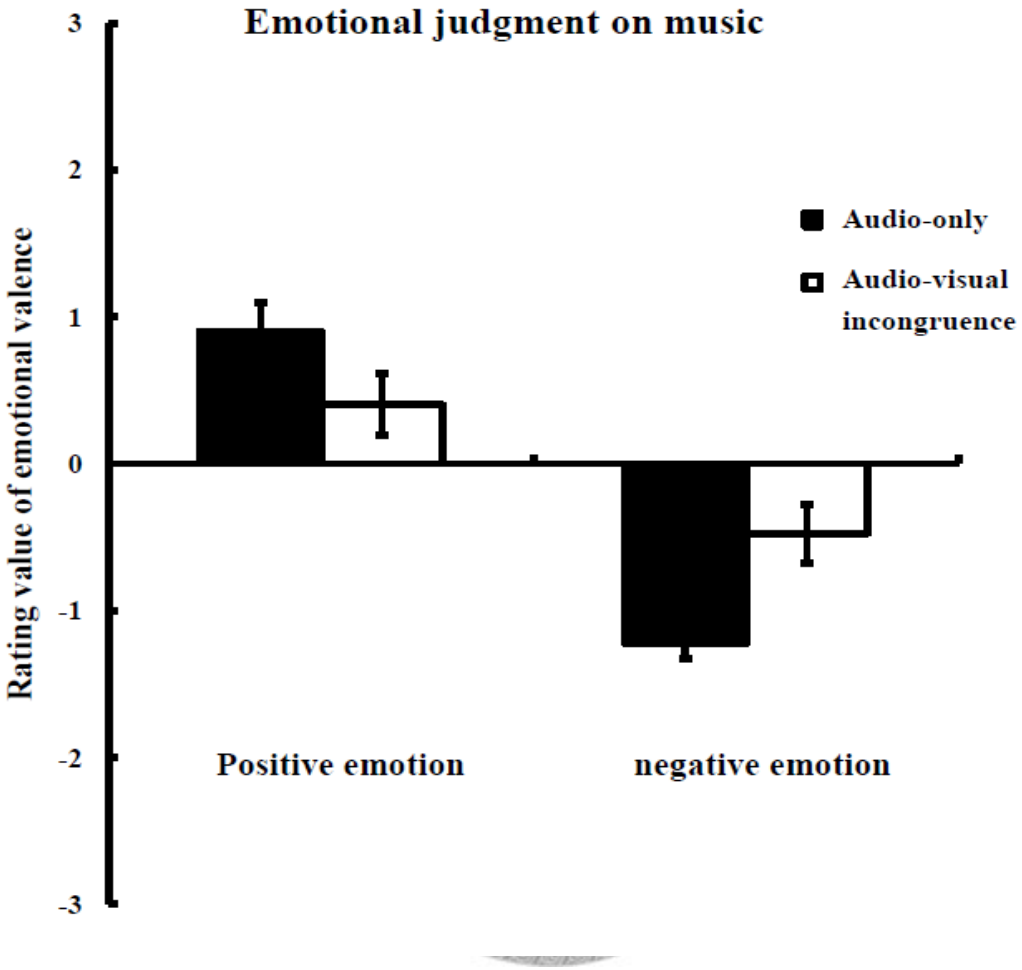

Figure 6. Mean rating values of emotional valence in judgment on music in Experiment 3C. Audio-only: only music was played; Audio-visual incongruence: the music and the video were perceived as different emotional valence.

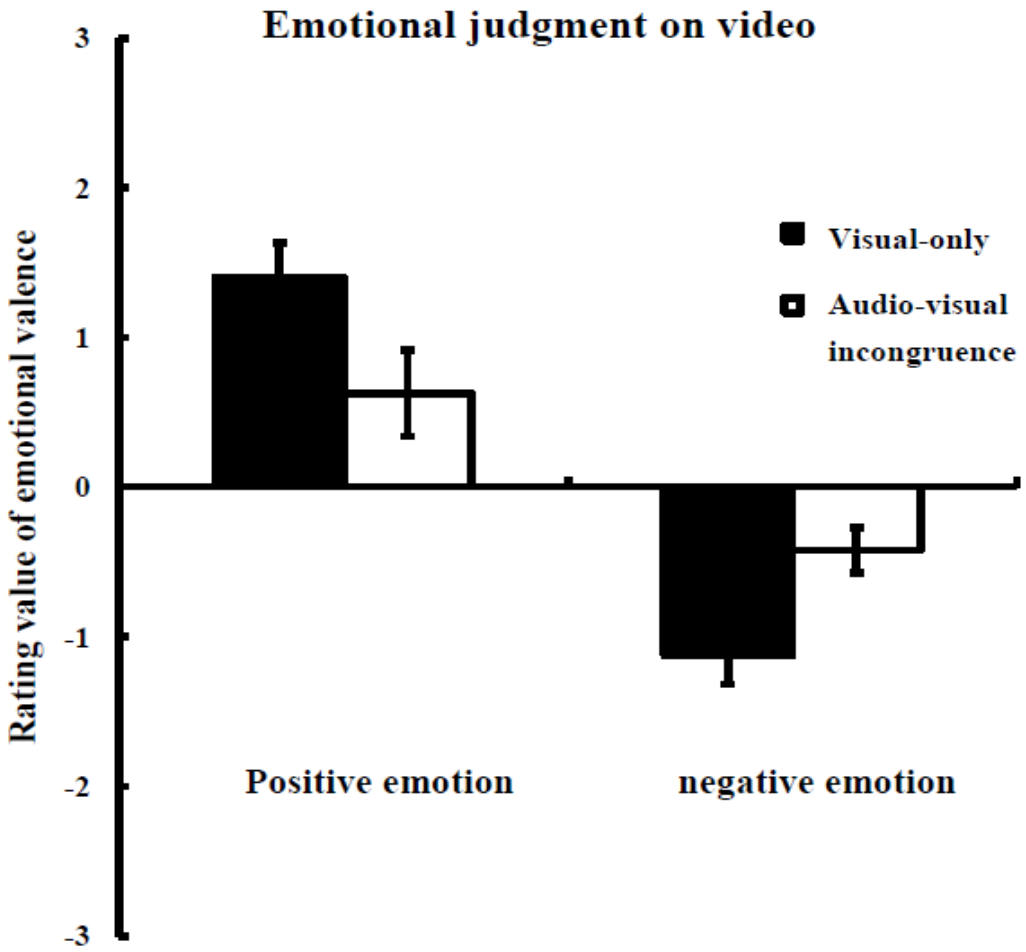

Figure 7. Mean rating values of emotional valence in judgment on video in Experiment 3C. Audio-only: only music was played; Audio-visual incongruence: the music and the video were perceived as different emotional valence.