

Variations of the perception of mood and tension of music excerpts depending on the visual context

Blas Payri

Universidad Politécnica de Valencia, Spain
bpayri@har.upv.es

ABSTRACT

We study the influence that image features may have on music tension and liveliness perception. 72 music excerpts from different genres and periods were selected, and 72 still shots were taken from different animation features little known to the subjects. 62 subjects rated the isolated images for tension and liveliness, 37 subjects rated the isolated music excerpts for tension and liveliness, and 153 subjects rated the music excerpts combined with the images for music tension and liveliness, and for music-image congruence. There is a significant variation of tension and liveliness of the music as a function of the tension and liveliness of the pairing image, showing a transfer of mood from image to music. The significance of ANOVA tests showed that 40% of music excerpts were image-sensitive for liveliness and 32% for tension. The transfer of mood was dependent on congruence: music excerpts with high congruence with the image had a higher correlation in tension and liveliness rating deviations with the image ratings. For low congruence, the liveliness correlation was not significant and the tension deviation was negatively correlated with the image tension. Feature transfer from image to music depends on the image-music congruence rated by each subject.

INTRODUCTION

Film music, and in general audiovisual music, has been studied from the perspective of film theory, music composition and music perception and cognition. The works of musicians and film theorists describe the interaction of image and music (Chion 1995, 2005) and the “secret influence” of music in the perception of the film by the audience (Nieto, 2003; Ondaatje, 2005) and how the music structures itself around the image to convey specific messages and emotions (Arcos, 2006).

In the field of music perception there is an important ongoing research on the influence that music has on film, and how it changes the perceived intention of characters, and the mood of the scenes (e.g. Boltz, 2001, 2004; Cohen, 2000, 2001, 2005; Tan, Spackman & Bezdek, 2007; Tan, Sapckman & Wakefield, 2008).

Both approaches, music perception and film music theory, tend to remain truthful to filmmaking process: first comes the screenplay and the image shooting, and once the image editing has been completed, the music is conceived to influence the general mood, add suspense or create narrative information. The non-diegetic music composer is

then someone in charge to analyze the intention of the filmmaker and put it into music, modifying the final perception of the audience.

Most of the perception studies focusing on music-film interaction use some visual scenes that have a certain mood or genre (usually neutral or ambiguous) and then they are combined with different musical excerpts that are previously rated for the features under study. The study questions how the music affects the perception of the scene.

A. Aim of the present study

Most experiments assume that music excerpts have certain characteristics, and that these characteristics will influence the film perception. What this study wishes to understand is if there is actually an interaction of image and music: that is to say, whether the image changes the perception of the music.

For that purpose, we are going to use similar material to previous studies (film image, music excerpts) and rate the perception of images and music independently. Our experimental questions refer to the perception of music tension and liveliness when heard accompanying an image, instead of the changes of the image with the music.

We chose music excerpts as well as images that varied greatly in liveliness and tension but the goal of this experiment is to understand the transfer of features from image to music, and we leave out of the scope of this article the explanation of the intrinsic ratings of music liveliness and tension.

I. EXPERIMENTAL DESIGN

A. Common audiovisual material

72 still shots from 14 animation works of different genres were selected. Animation was chosen as the genres, mood, colors and other visual codes are clearer than in real image. Images varied in mood and contents. 72 music excerpts from 26 musical works of different genres and periods were selected. Music excerpts were chosen to vary in mood.

B. Tasks and subjects

Rating the image alone. The 72 images were divided in 6 groups of 12 images. 61 subjects rated 86 image sets, with an average of 14,3 ratings per image question. Subjects were first year students of audiovisual communication at UPV. For each image, subjects rated the *tension* and *liveliness* from 1 to 5. They had the images they had to rate in an .rtf document along with the questions, and opening the document in a computer, they gave their answers directly on the document that was then collected.

Rating the music alone. The 72 music excerpts were divided in 6 groups of 12 excerpts. 37 first year students of audiovisual communication at UPV rated 57 sets for a mean of 9,5 ratings per excerpt. For each music excerpt, subjects rated the *tension* and *liveliness* of the music excerpt from 1 to 5. They answered a total of 12 questions about the genre, period and origin of the music excerpt. Subjects listened individually the music excerpts on a computer with headphones and answered directly on a .rtf document. There was no time limit.

Rating the music with the image. 6 sets of music-image combinations were made for a total of 360 rated combinations. The sets respect two conditions: no image and no music excerpt was repeated for a given set and the order of the 72 images was fixed for every set, and the music excerpts were permuted from set to set so that if music excerpt number m was associated with image i in one set, then music excerpt $m+1$ was associated with the same image in the following set.

Each of the sets of music-image was rated by a different group of subjects. The subjects participating in this part did not participate in the other tasks. The task was performed in a silent classroom, using a video projector to display the images and a couple of high quality loudspeakers to play the music. The subjects saw an image first in silence (1s) and when with a music excerpt (14s). Then there was no image and no sound (15s) and subjects had to rate from 1 to 5: 1- the congruence of the music as non-diegetic music for the image, 2- the tension of the music alone, 3- the cheerfulness of the music alone. A total of 153 raters participated in this task including first-year students of audiovisual communication and students of the master of music of UPV. Each pair music-image

was rated as a mean 23,5 times on tension, liveliness and congruence.

II. ANALYSIS OF THE RESULTS

A. Music alone and image alone rating means

The main focus of this article is to analyze the influence of the characteristics of the image on the music, when rating the music in the presence of an image. The main experimental task is then the third (rating the music with the image). We computed the means for the different answers for the music alone rating task and the image alone rating task. These means were then inserted as new variables in the table of results for the image with music rating task, and the following analyses focus on the music-image responses and how they can be explained by the music and image alone responses.

B. Influence of the image on music tension perception

The presence of an image had a very highly significant effect on music tension ratings. A one-way ANOVA showed that subjects rating the music alone perceived less tension than subjects that rated the music with the presence of an image ($p < .001$, $F=59.8$).

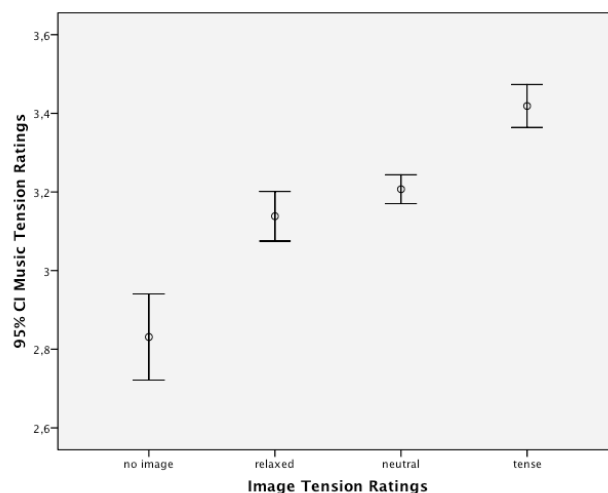


Figure 1. Mean values of the ratings of tension in the music excerpts with several conditions: no image present, presence of images with different rated tension.

To study further the influence of image tension on music tension, we divided the images in three groups according to their mean tension rating: we will refer to these groups as “relaxed” (lower tension ratings), “neutral” and “tense” (for the higher tension ratings). A one-way ANOVA

showed that the rated tension of the image had a very highly significant effect on music tension ratings ($p < .001$, $F = 27.7$). The music excerpts that were rated with “tense” images had higher tension ratings ($p < .01$) than the other groups. There were no significant differences between the “neutral” and “relaxed” image groups. A graphic summary of these findings can be seen in figure 1.

A bivariate correlation between image tension ratings and tension ratings for the music excerpts rated with these images showed that the correlation factor is small (0.077) but the correlation is significant at the 0.01 level. To study further the importance of each factor on tension ratings, we performed a stepwise multiple regression using as dependent variable the music-with-image tension ratings, and as factors the ratings for the music and image alone, as well as the congruency. The music-alone tension was entered first and explained 40,4% of music-with-image tension ratings ($F_{1,8993} = 6107$, $p < .001$). Image-alone tension ratings was entered last (fourth of four predictors) and explained a further 0,1% ($F_{1,8990} = 1540$, $p = .004$). When more variables describing the music alone were entered (orchestration, origin...) the influence of the image tension became negligible.

Consequently, the tension for the music is mainly rated on intrinsic factors, but a small and always significant portion is due to the characteristics of the image.

C. Influence of the image on music liveliness perception

The presence of an image had a very highly significant effect on music mood ratings. A one-way ANOVA showed that subjects rating the music alone perceived less liveliness than subjects that rated the music with the presence of an image ($p < .001$, $F = 95.2$).

To study further the influence of image mood on music mood, we divided the images in three groups according to their mean mood rating: we will refer to these groups as “sad” (lower liveliness ratings), “neutral” and “lively” (for the higher liveliness ratings). A one-way ANOVA showed that the rated mood of the image had a very highly significant effect on music mood ratings ($p < .001$, $F = 84$), the higher the image liveliness, the higher the music rated liveliness. Every group had mean differences with the other groups significant at the 0.01 level. A

graphic summary of these findings can be seen in figure 2.

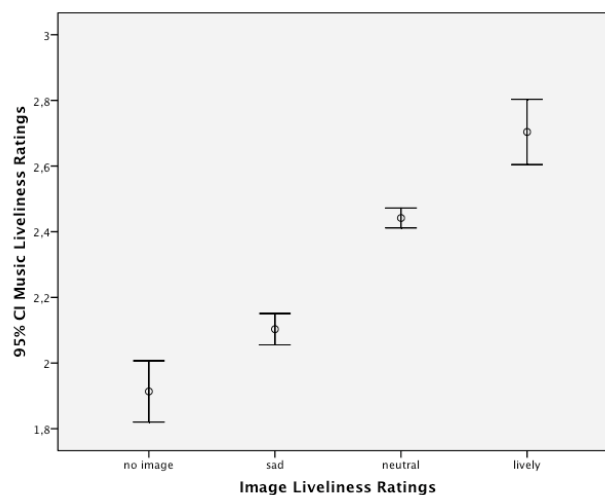


Figure 2. Mean values of the ratings of mood in the music excerpts with several conditions: no image present, presence of images with different rated mood.

A bivariate correlation between image and music mood ratings showed a small correlation factor (0.125) but significant at the 0.01 level. To study further the importance of each factor on liveliness ratings, we performed a stepwise multiple regression using as dependent variable the music-with-image liveliness ratings, and as factors the ratings for the music and image alone, as well as the congruency. The music-alone liveliness was entered first and explained 38,2% of music-with-image tension ratings ($F_{1,9004} = 5553$, $p < .001$). Image-alone liveliness ratings was entered second and explained a further 1% ($F_{1,9004} = 2897$, $p < .001$). Finally image-alone tension ratings was entered third and explained a further 0,2% ($F_{1,9004} = 1946$, $p < .001$). When more variables describing the music alone were used as factors the influence of image tension decreased but remained significant.

As with the tension, music liveliness is mainly rated on intrinsic factors, but a significant small portion is due to the characteristics of the image, and it is larger than for tension.

D. Influence tension and liveliness ratings on congruence

In a previous study (Payri, 2008) on part of the data used for this article, we described the process of congruence of image and music that was explained by factors such as rated tension and mood, but predominantly there were other factors like the information of place and period, as well as other codes of genre and specific situation that

explained why a music excerpt was considered congruent.

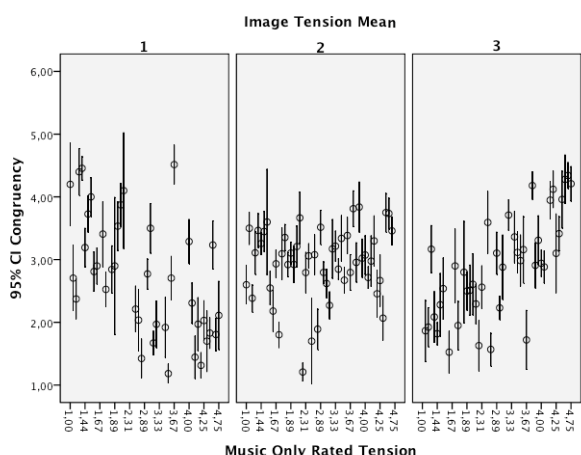


Figure 3. Mean values of the congruency ratings as a function of the music-alone rated tension, divided in three groups depending on the image tension mean.

The results of this experiment show that the congruence depends partly on the similarity of tension and liveliness of the image and the music. As can be seen in figure 3, the congruency has a different relation with the tension of the music for images with different tension, and yet a lot of the variation is not related to tension. Namely, the correlation between congruency and music-alone tension is negative for images with low tension ($R=-0.30$, $p<.001$) is not significantly correlated for images with a medium tension and is positively correlated for images with a higher tension ($R=0.38$, $p<.001$). Similarly, for images with a low liveliness, there was a negative correlation between the congruency and the music-alone liveliness ($R=-0.27$, $p<.001$), it was still negatively correlated for neutral images ($R=-0.07$, $p<.001$), and it was positively correlated with music-alone liveliness ratings for images with a higher liveliness ($R=0.48$, $p<.001$). It is interesting to note that the correlations are stronger between the congruency and the tension or liveliness for the music-alone condition than for the tension of liveliness of the music-with-image condition.

To take in account the non-linear relation between congruency and the ratings for liveliness and tension for the music, we computed a standardized score for the ratings of liveliness and tension for the image and the music, and computed the difference and the multiplication of the image-music tension and liveliness ratings. We obtained thus four new compound variables. We performed a stepwise multiple regression on the congruency

using as factors the different music and tension direct and compound values, and the total amount of explained variance was 18%, the compound image*music tension being the most important factor that only explained 6% of the congruency ($(F_{1,9031}=639, p<.001)$).

E. Influence of the congruence on tension and liveliness deviation

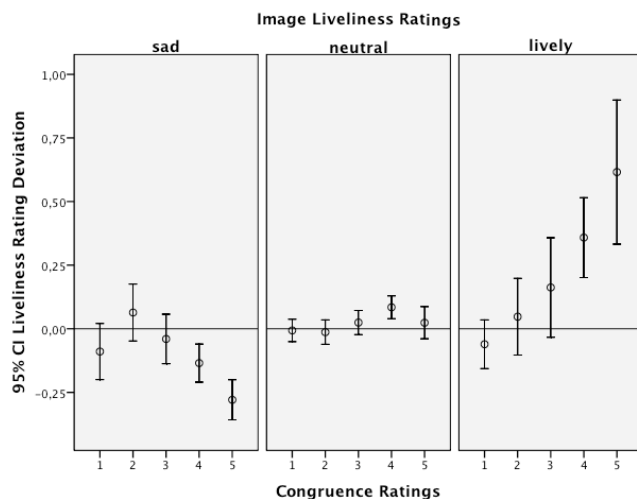


Figure 4. Graph of the music-with-image liveliness rating deviation from the mean of the music extract as a function of the image-music congruence, separated by the image liveliness rating means.

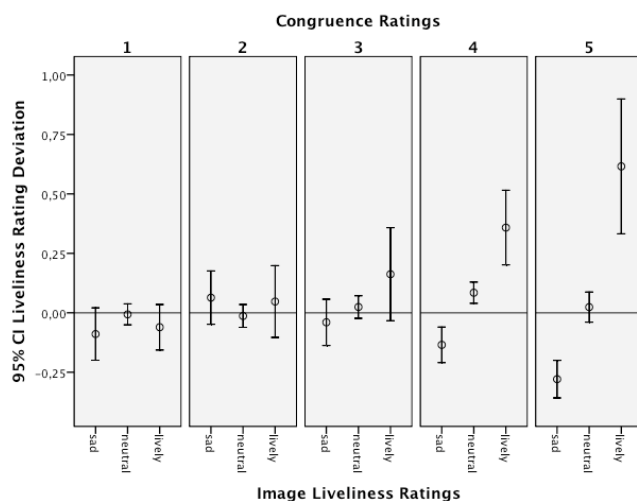


Figure 5. Graph of the music-with-image liveliness rating deviation from the mean of the music extract as a function of the image liveliness rating, separated by the image-music congruence.

In order to understand the process of transmission of characteristics from the image to the music, we proceeded to study if the congruence music-image had an influence on this variation.

As our interest is the variation of the ratings for music excerpts when in conjunction of a given image, we computed a new variable containing the deviation of the ratings from the mean of a music excerpt, that is first we computed the mean of the ratings for the music excerpt, and then the difference between an individual rating and the mean. This deviation variable was computed for both tension and liveliness ratings. We computed as well the absolute deviation, which is simply the absolute value of the deviation.

As can be seen in figure 4, the deviation of liveliness ratings for a given music excerpt depends both on the liveliness of the image and the congruence of the image and the music. Lively images will influence positively the liveliness ratings of music depending quite linearly on the congruence. Images with the lowest liveliness ratings will have the contrary effect depending as well on the congruence, although the congruence needs to be really strong (4 or 5) for the effect to be clear. Neutral images have little impact on the music. Figure 5 displays similar data: we can observe that the influence of image liveliness on music ratings happen for higher congruence ratings, while low congruence limit the impact.

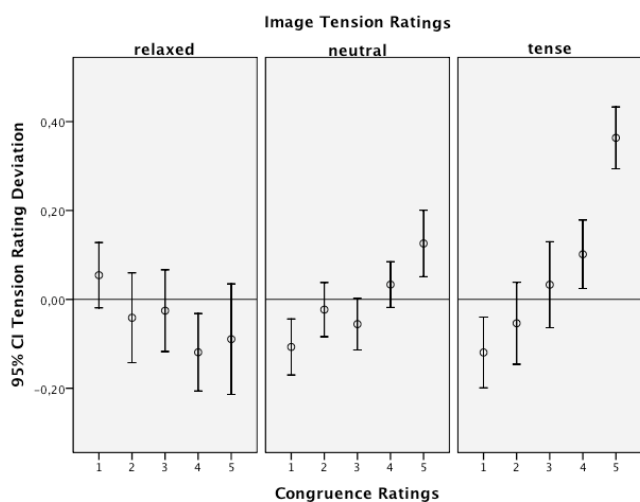


Figure 6. Graph of the music-with-image liveliness rating deviation from the mean of the music extract as a function of the image-music congruence, separated by the image liveliness rating means.

Figure 6 displays a similar pattern for music tension deviation. As with liveliness, the tension in music is linked to image tension depending on the congruence music-image.

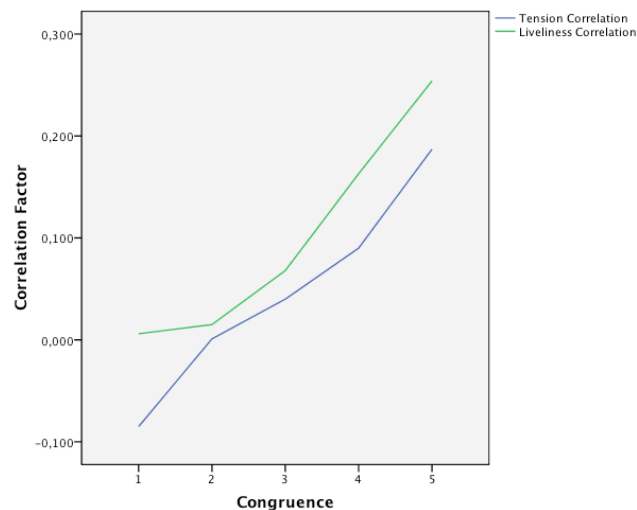


Figure 7. Evolution as a function of music-image congruence of the correlation factor between the tension of image and music and the liveliness of image and music.

We computed the correlation between the deviation of the music tension ratings and the congruence music-image for each image. There is a clear correlation between those correlation factors and the values of tension for the image ($R=0.6$, $p<0.001$).

We proceeded similarly with the liveliness ratings: for each of the 72 images we computed the correlations between the music liveliness ratings and the congruence music-image for all the music excerpts that were associated to that image. This resulted in 72 correlation-factor values that we correlated with the mean liveliness ratings for each of the 72 images. There is a correlation factor of 0,6, significant at the 0,001 level.

Finally, we computed the correlation factor between image and music tension and liveliness, for the music-image pairs that had the same congruence ratings, as can be seen in figure 7. The results show that clearly the correlation depends on the congruency and that it is generally higher for liveliness than for tension.

F. Characteristics of the music and sensitivity to the image

Not all music excerpts varied equally in liveliness and tension with the pairing images. We will define here the sensitivity to the image with the statistical measure of the variance (one-way ANOVA): the music excerpt is image-sensitive if the ANOVA shows a significant change ($p<0.05$) of the ratings of a same music excerpt for different images.

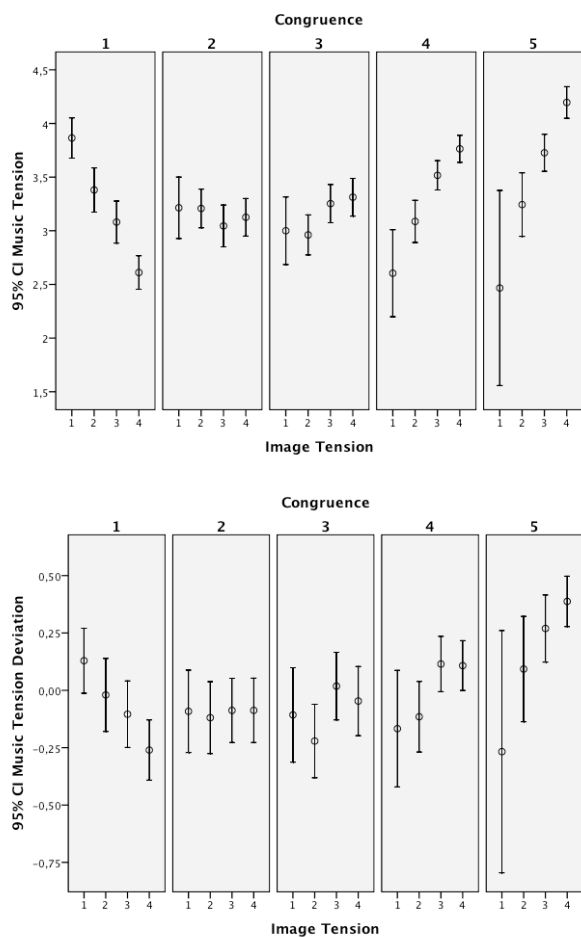


Figure 8. Mean deviation of the tension ratings for the music as a function of the tension of the corresponding image and divided by the congruence between image and music. Only music excerpts with significant variations of tension are selected.

For each music excerpt, we computed a one-way ANOVA for the differences of mean for the tension and liveliness ratings for each image. We obtained the F value and the significance of the ANOVA for each music excerpt. 32% of the music excerpts were image-sensitive for tension, and 40% for liveliness, for the samples of our study.

To understand what could explain that some music excerpts are more sensitive to the image, we computed correlations between the F values of liveliness and tension and all the different rating means for each music excerpt: there was no significant correlation. We stress that there was no correlation between the value F for tension or liveliness that measures the sensitivity to image features and the mean tension or liveliness of the music excerpts: this goes again the assumption that more tense or lively excerpts would be less sensitive to the environment than music excerpts that have no salient features. An ANOVA showed no significant differences either between groups of

different tension level and their sensitivity to the image.

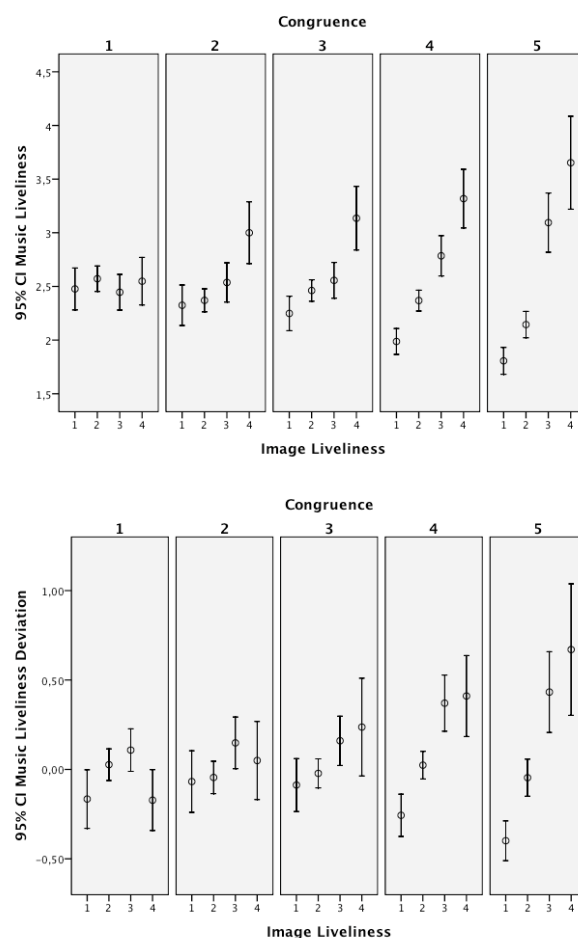


Figure 9. Mean music liveliness (above) and deviation of music liveliness (below) as a function of the liveliness of the corresponding image separated by image-music congruence. Only music excerpts with significant variations of tension are selected.

We conclude then that the variations explained by the regression values are due to the fact that listeners agree better on music excerpts with more extreme values, but that is the general standard deviation and not the sensitivity to the image environment.

The F values for liveliness and tension were not correlated between them and were not correlated with the F value for congruency, meaning that a music excerpt could have very different congruency levels with one image or another but that would not influence the sensitivity to the image tension or liveliness. It means also that the influence of tension and liveliness are separate matters.

A case per case analysis of the sensitivity of music excerpts did not prompt a clear hypothesis that would explain the difference in sensitivity.

G. Analysis of image-sensitive music excerpts

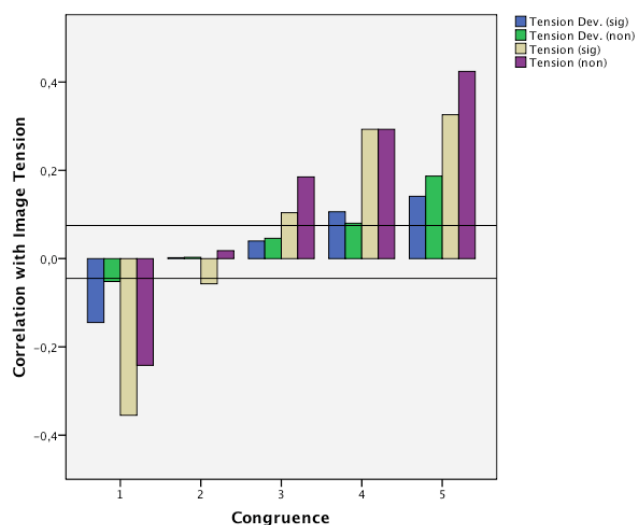


Figure 10. Correlation factor between image tension ratings and music tension ratings considering the tension ratings and tension deviation from the mean of the music excerpt and considering the music excerpts that had a significant variation of tension depending on the image (sig) versus those that had no significant variation of tension (non). The horizontal lines define the zone where the correlation music-image tension was non significant.

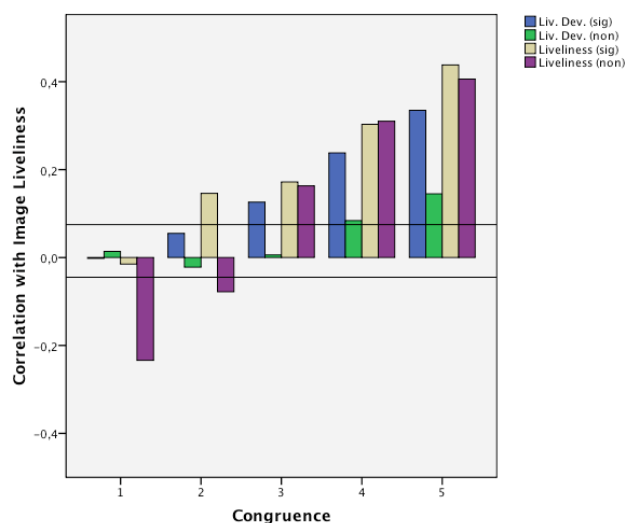


Figure 11. Correlation factor between liveliness ratings for music and image considering the tension ratings and tension deviation from the mean of the music excerpt and considering the music excerpts that had a significant variation of tension depending on the image (sig) versus those that had no significant variation of tension (non). The horizontal lines define the zone where the correlation music-image tension was non significant.

In order to analyze further the differences between the image-sensitive and non sensitive music excerpts, we computed the correlations between the image liveliness ratings and the music liveliness ratings and liveliness deviation from the excerpt mean (figure 11) and the equivalent for

image tension and music tension and tension deviation (figure 10) dividing in music excerpts that were respectively image-sensitive for liveliness and tension.

The results displayed in figures 10 and 11 call for several conclusions. First, the correlations are more important for the ratings of tension (respectively liveliness) than for tension deviation from the mean. This means on one hand that when there is a high congruence, the music tension partially correlates with the image tension and vice versa, with a very low congruence, the music tension will vary contrary to the image tension. The correlations are higher for music tension are higher than for tension deviation because on top of the transfer from image to music of tension (resp. liveliness) a small part of the congruence is due to the matching of the tension (resp. liveliness) of the image and the music. Thus congruent music-image pairs will tend to have music excerpts that have a similar intrinsic value in tension than the music. This difference is clearer for tension than for liveliness, meaning that for tension ratings the congruence depends on the intrinsic tension for image and music, and moreover, congruent music will inherit some of the tension of the image. On the other hand, the correlation music-image for liveliness is mainly due to the transfer of liveliness from image to music as there is a similar correlation factor for liveliness and liveliness deviation.

Finally, there are differences between the correlations of music excerpts that have significant differences in tension and liveliness depending on the images with which they are rated. We can formulate the hypothesis that music excerpts that are image-insensitive should not have any correlation with the image in tension or liveliness. This is not the case, as for tension deviations the correlation is basically the same for image-sensitive and insensitive, which means that when a given rater finds that the image and the music are congruent this rater will transfer the tension of the image to the music, even if the average rating is not affected as other raters will not find the image congruent. To a lesser extent, the same applies to liveliness.

CONCLUSION

The main result of our study is that there is a transfer of perceived characteristics from image to music, both for liveliness and tension, and that this

transfer is very dependent on the congruence of the image and the music. The higher the congruence, the higher the correlation of music and image features. In the case of tension ratings, a low congruency (meaning that the music was not at all fit for the image) resulted in a negative correlation: the subjects rate the music as less tense when the image is very tense but the music and image do not fit together.

40% of the music excerpts had significant variations of the mean ratings depending on the associated image for liveliness and 32% for tension. Even though a music excerpt had no significant difference of the mean, the correlation of image and music features was still significant and depending on the congruency. That means that each individual subject changed the ratings for the music depending on the subject's evaluated congruency.

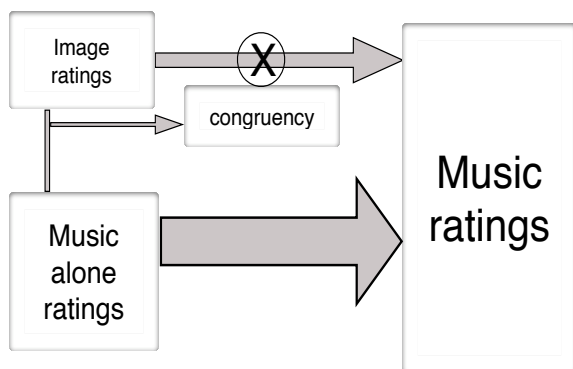


Figure 12. Model for the ratings of music features when in presence of an image, taking in account the congruency of the image and the music.

The graphic in figure 12 summarizes the model we propose to explain the results of this research. Music ratings for tension and liveliness, the two features under study, are explained mostly by their intrinsic values, that is, when the music features are rated in isolation. That is particularly true as we had an extremely wide range of music genres and styles, with a great diversity of expected tension and liveliness. If we had chosen more similar or expressionless music excerpts, we might expect that the ratings would depend less on intrinsic values of the music and more on the context.

A significant influence for music ratings comes from the features of the images that were associated, and that transfer is multiplied, or rather influenced, by the congruency of the music excerpt and the associated image. In turn, a small amount of this congruency is explained by the similarity of

features of music and image, here the liveliness and tension.

This model should be completed by the general findings in the literature that describe the variations of image perception depending on the associated music, which we have not studied here. There are several models that are proposed for the cognitive processing of film and image (e.g. Cohen, 2000). Also, the complex could have much more detail and complexity, as the congruence of image and music is derived from other features that liveliness and tension, and as the cognitive processing should take in account narrative codes and knowledge of the actual use of music styles in film to convey codified information of liveliness and tension.

REFERENCES

- Arcos, M. d. (2006) *Experimentalismo en la música cinematográfica*. Spain: Fondo de cultura económica de España.
- Bolivar, V., Cohen, A. J. & Fentress, J. (1994). Semantic and formal congruence and effects of film music. *Psychomusicology*, 13, 28-59
- Boltz, M. G. (2001). Musical soundtracks as a schematic influence on the cognitive processing of filmed events. *Music Perception*, 18, 427-454.
- Boltz, M.G. (2004). The cognitive processing of film and musical soundtracks. *Memory & Cognition*, 32, 1194-1205.
- Chion, M. (2005). *L'audio-vision : Son et image au cinéma*. France: Armand Colin.
- Chion, M. (1995). *La musique au cinéma*. France: Fayard.
- Cohen, A. J. (2000). Film music: Perspectives from cognitive psychology. In J. Buhler, C. Flinn & D. Neumeyer (Eds.). *Music and cinema* (pp. 360-377). Hanover: Wesleyan Univ. Press
- Cohen, A. J. (2001). Music as a source of emotion in film. In Juslin P. & Sloboda, J. (Eds.). *Music and emotion*. (pp.249-272). Oxford: Oxford University Press
- Cohen, A. J. (2005). How music influences the interpretation of film and Video: Approaches from experimental psychology. In R.A. Kendall & R. W. Savage (Eds.). *Selected Reports in Ethnomusicology: Perspectives in Systematic Musicology*, 12, 15-36
- Nieto, J. (2003). *Música para la imagen: la influencia secreta*. Spain: Iberautor Promociones Culturales.
- Ondaatje, M. (2005). *The conversations: Walter Murch and the art of editing film*. NY: Alfred Knopf
- Payri, B. (2008). Perceived Fitness of Music and Film Genre: Color, Light, Style and Period Combinations. *10th International Conference on Music Perception and Cognition*. Japan: Hokkaido University.
- Tan, S. L., Spackman, M. P., & Bezdek, M. A. (2007). Viewers' interpretations of film characters' emotions: Effects of presenting music before or after a character is shown. *Music Perception*, 25 (2), 135-152.
- Tan, S. L., & Spackman, M. P., & Wakefield, E. M. (2008). Effects of diegetic and non-diegetic presentation of film music on viewers' interpretation of film narrative. *10th International Conference on Music Perception and Cognition*. Japan: Hokkaido University