Research on Audio-visual Emotion Fusion based on Superposition Response

Hua Zhang[1, a], Wei Jiang[2, b]

[1] Key Laboratory of Acoustic Visual Technology and Intelligent Control System, Ministry of Culture Beijing, China

[2] Key Laboratory of Acoustic Visual Technology and Intelligent Control System, Ministry of Culture Beijing, China

[a]email: zhanghua_cuc@126.com, [b]email: jw@cuc.edu.cn

Keywords: Audio-visual; Superposition Response; Emotion Fusion;

Abstract. The research tries to study humans' responses to superimposed costimulatory signals. The information, including music, natural images, may be perceived through sound and visual stimulus. The procedure is interested in the superposition state of superimposed vibrations, whereby two or more signals are perceived simultaneously, producing a perceptual impression that is considerably different than of each signal alone, owing to the interactions between perceived stimulus vibrations that induce a coordinated percept of a vibrational chord. We show that these temporal attributes and statistics are strongly related. At the same time, we reveal that this kind of superposition response is closely related to human's psychological mechanism.

Introduction

Information is the state and the way to know the things that are perceived by the subject or the things that are expressed. Real information transfer of audio-visual media is its own stimulus to the audience, rather than the content it conveys. Artificial emotion makes use of the information science method to simulate, recognize and understand human's emotion, and make the machine produce human emotion[1][2]. Researchers have put forward the affective computing, emotional engineering, artificial psychology [3-5].

Real time multimedia information plays an important role in the new human computer interaction [6]. The computer can form the computer vision and the sense of hearing by collecting the information of the image and music [7-8]. However, its research aims at simulating human intelligence [9], and it is far from enough to study the single perception process. The structural characteristics of the integration of meaning, visual and acoustic determine the social and natural attributes of audio-visual media. From the perspective of calculating how to see the emotional fusion of audio-visual media, How to study the emotion fusion of audio-visual media information, and what we care about the core issues.

In this paper, a core issue: the dynamic evolution law of audio-visual media fusion space is explored and utilized. First of all, the audio visual media information is described and defined in detail. And then, to study the human emotional response mechanism to the superimposed stimuli, complete subjective experiment and result evaluation of visual image. Then, we give the correlation of visual and audio on multi-dimension. The last section discusses the related problems of the superposition response.

Information Recognition

What is Audio-visual Media "Information"?

Information is the state and the way to know the things that are perceived by the subject or the thing that are expressed. Real information transfer of audio-visual media is its own stimulus to the audience, rather than the content it conveys. The content is only a carrier of information. Two kinds

of media information collision, the production of new media forms, the similarity of the two media allows us to stay on the edge of the two media. The superposition time is bound to form a new ratio, not only a variety of perception will form a new ratio, but also the interaction between them to form a new ratio.

Audio-visual Media "Information" Studies?

The pitch combinations used by music from different cultures may be unique, but the psychological basis of human perception of pitch is the same [10]. Nothing is more important than to study the response of the audience to the music stimulus. Study on the emotional response of music needs to consider the mechanism of emotion [11]. Most of the researches focus on the audience's perception of music emotion [12-13]. However, there is a lack of evidence for a perceptual component ─ the subjective awareness of people, which raises the question of how to accurately receive and evaluate musical stimuli. Emotional contagion is a process of "natural generation" which is fully involved in human's consciousness, because the audience receives the musical expression and is internalized as a mental representation, or in the presence of an emotional state, or more directly in the formation of visual imagery in the brain. People will have such a feeling: when listening to music, will be accompanied by the visual experience (such as a beautiful scene) of music related, which is the result of close interaction between music and image, which is the response to the superposition occurring in the different sensory stimuli.

Acoustic Visual Psychological Mechanisms

Music can make people fall into a reverie, how to explain the psychological mechanism of musical image. For example, the legal program, the voice gives a deep feeling. We think, "deep" is the height of the space, "heavy" is the weight of the object, we take these two words to describe the sound of hearing, which is in itself a joint sense phenomenon. Psychological phenomenon caused by a feeling of other feeling called synesthesia. The art of music, which can represent the image, scene, emotion, emotion, thought, philosophy and so on, is the core of the truth. In this sense, the composer chooses and organizes the voice to show what he wants, and our audience is in the same set of psychological reactions, in the influence of the composer to feel his performance. Whether music can cause people to have a clear vision of the image, depend on the continuous and stable joint sense of the corresponding relationship.

From all of the acoustic visual properties, Music in pitch, loudness, timbre, melody, rhythm and other elements can establish rich contraposition relationship with the characteristics of their visual perception of color, brightness, shape and volume. There is a corresponding relationship between the frequency of the music and the brightness of the color. The higher frequency sound can make people produce relatively bright visual image, and the lower frequency sound can make people produce darker visual image.

Responses to Superimposed Costimulatory Signals

Music-Based Representations

The audio signal has "character of short-time stationary, the variable", namely in a short period of time can approximate that its characteristic is stable, and beyond this period of time it is a kind of typical non stationary signals. It is found that these short time periods are called Frame Audio, which is the smallest element in audio processing. Frame processing is usually adopted in short time frame. The window function is Hamming window:

$$w(n)=\begin{cases}0.54-0.46\cos\left[2\pi n/(N-1)\right], 0\le n\le N-1\\ 0, \quad n=else\end{cases} \tag{1}$$

Chroma: the visual color is the result of the frequency distribution of music time series on the time axis, which forms the basis of the color spectrum. All kinds of musical instruments playing

music, not only contains a note of basic frequency, also contains a lot of homonym, homophone frequency and fundamental frequency between integral ratio relationship.

$$RTFI(n, w_m) = s(n) * I_R(n, w_k) \tag{2}$$

$$I_R(n, w) = (1\text{-}e^{\frac{-r(w_k)}{f_s}})e^{\frac{-r(w_k)+jw_k}{f_s}n} \tag{3}$$

In the above formula, I_R said the oscillation frequency of w_k first-order complex digital echo filter impulse response, $r(w)$ for the standardization of frequency response of incremental, attenuation factor r dependence on w and decision index window length and time resolution, also decided to band width.

Image-Based Representations

Color space has two important features: one is human changes in perceptions of independence the space for each color component; the second is color in the color space of triples (Hue, Saturation, Value) between the Euclidean distance and the feeling of human eyes to the corresponding color difference of cable shape, which is in accord with the characteristics of human visual perception of color model. Measurement based on HSV color space can better describe the feeling of the human eye.

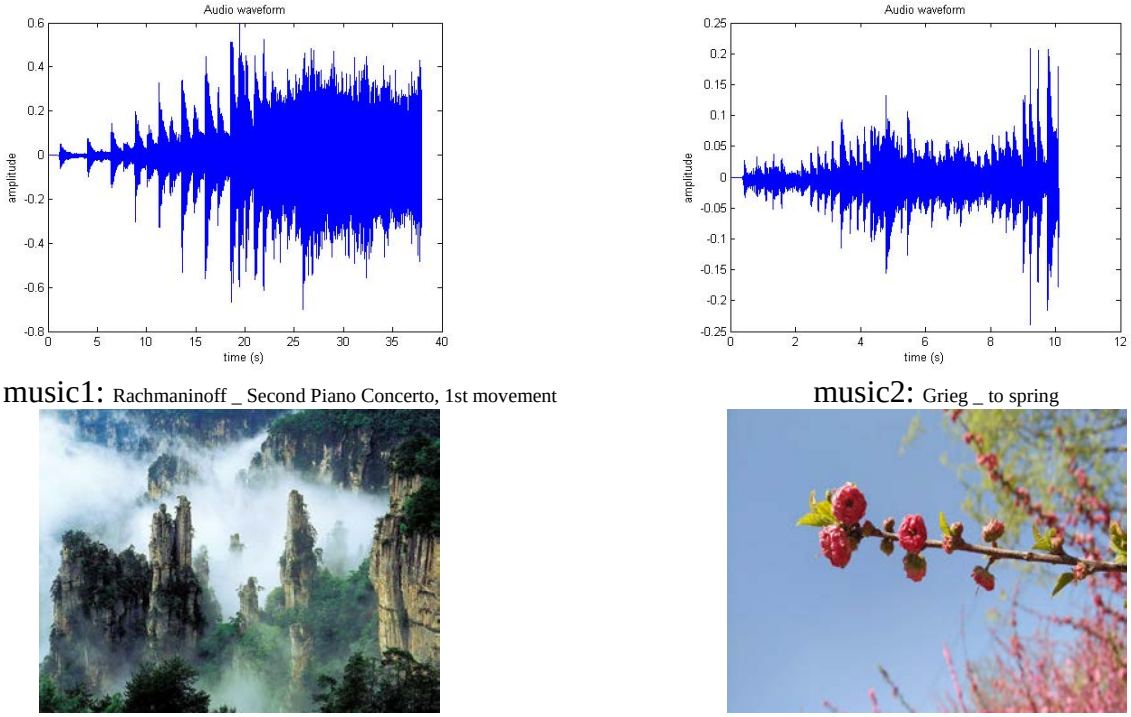
Result Analysis

(a) Table 1 shows one example of findings from studies with acoustic visual. Based on the musical acoustics and the psychological acoustics, the response relation of the visual stimulation is analyzed from the angle of the sound speed. Stimulus material (music and pictures in Figure 1) come from the subjective evaluation test.

Table1. *Corresponding relationship of acoustic visual psychological mechanisms*

*A*ttack time and speed	*Fast*	*Slow*
Tactile sensation	Hard	Soft
Personality	Stiff	Soft
Behavior	Decisive、resolute	Hesitant、retarded
Modal	Impatient	Moderated
Vision	Suddenly、blunt、straight	Moderate、soft
Survival relation	Evil、harmful、dreadful	Good、favourable、beautiful

(b)Music1 （Rachmaninoff _ second piano concerto, 1st movement）：First of all, the sound is very strong, give a person the feeling is very imposing manner; secondly, very low give a person feel very deep. Based on the subjective evaluation of the visual image of the music as shown in Figure 1，the steep mountains. Music2（Grieg _ to spring）：First of all, the sound is very relaxed, give a person feel very soft; secondly, very high give a person feel very light. Based on the subjective evaluation of the visual image of the music as shown in Figure 1， the warm spring.



music1: Rachmaninoff _ Second Piano Concerto, 1st movement

music2: Grieg _ to spring

Fig.1. Natural scene with different music and their corresponding representation

(c) Chord histogram is the percentage of time that the size of a chord is occupied in a song. Here's just a histogram of the size and size of the chord. But the chord in different position and the order of the song can cause similar emotions, if the number of songs in the middle and small chords accounted for a significant proportion of the songs tend to be dim and melancholy, and vice versa. Through shown in Figure 2, it can be seen that the music of the chord of frequency distribution and proportion. The music1 genre tend deep, thus chord histogram as a feature is able to explain the psychological mechanism of musical image.

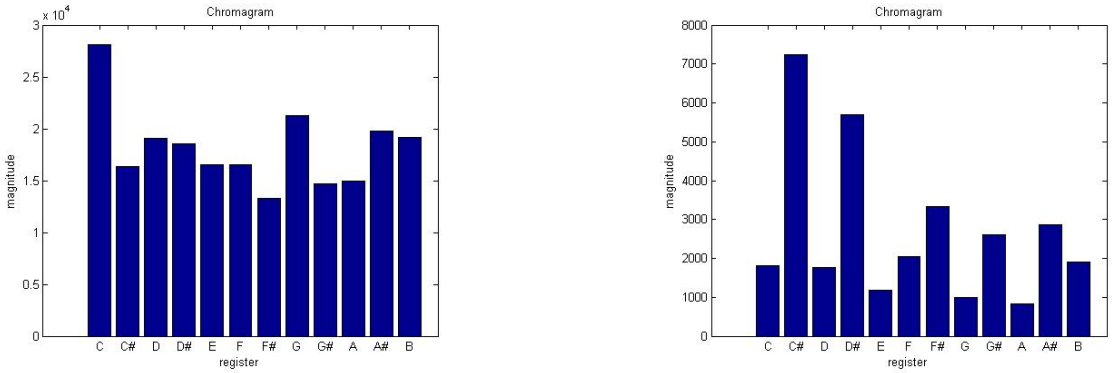


Fig.2. Frequency distribution and proportion of chords with different music

Conclusion

This paper introduces a kind of audio-visual media emotional superposition response, music and image is described as a set of perceptual features. These attributes are related to emotions, and it is meaningful to sense fusion. We show that these temporal attributes and statistics are strongly related. At the same time, we reveal that this kind of superposition response is closely related to human's psychological mechanism. However, the current research only considers the single emotion induced and ignores the potential correlation model. Further research will be helpful to the establishment of the theory and the integration of the subject.

Acknowledgement

In this paper, the research was sponsored by key cultivation project of engineering plan (Project No. 3132016XNG1603).

References

[1] Wilson, I, Simulation artificial emotion and personality. 2004. Stanford, CA, United States: American Association for Artificial Intelligence, Menlo Park, CA94025-3496, United States.

[2] Lucía Teijeiro-Mosquera;Joan-Isaac Biel; José Luis Alba-Castro;Daniel, Gatica-Perez What Your Face Vlogs About: Expressions of Emotion and Big-Five Traits Impressions in YouTube. IEEE Trans. Affective Computing, Vol. 6 no. 2, pp. 193-205, 2015

[3] R. Picard, E. Vyzas, and J. Healy, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175-1191, Oct. 2001.

[4] Nagasawa, S. Present state of Kansei engineering in Japan. 2004. The Hague, Netherlands: Institute of Electrical and Engineers Inc., New York, NY10016-5997, United States.

[5] W. James, The Principles of Psychology. Holt, 1890.

[6] Yang Yi, Nie Feiping, Xu Dong, Luo Jiebo, Zhuang Yueting, Pan Yunhe, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 723–742.

[7] Maria Teresa Riviello, Anna Esposito, and Mohamed Chetouani, Inferring Emotional Information from Vocal and Visual Cues: a Cross-Cultural Comparison. Cognitive Infocommunications (CogInfoCom), 2011 2nd International Conference on, 7-9 July 2011

[8] J. Holm and H. Siirtola, "A Comparison of Methods for Visualizing Musical Genres", in Information Visualisation (IV), 2012 16th International Conference on, pp. 636-645.

[9] Anne-Sylvie Crisinel and Charles Spence, As bitter as a trombone: Synesthetic correspondences in nonsynesthetes between tastes/flavors and musical notes. Attention, Perception, & Psychophysics 2010, 72 (7), 1994-2002

[10] A. Esposito, M. T. Riviello, N. Bourbakis: Cultural Specific Effects on the Recognition of Basic Emotions: A Study on Italian Subjects. In: Holzinger, A., Miesenberger, K. (eds.) USAB 2009. LNCS, vol. 5889, pp. 135–148. (2009)

[11] Eerola, T., & Vuoskoski, J. K. A comparison of the discrete and dimensional models of emotion in music. Psychology of Music, 39(1), 18-49. 2011

[12] Stein, N. L. & Levine, L. J. The early emergence of emotional understanding and appraisal: Implications for theories of development. In: Handbook of cognition and emotion, ed. T. Dalgleish & M. J. Power, pp. 383–408. Wiley. [SJH].1999

[13] Kejun Zhang, Shouqian Sun. Web music emotion recognition based on higher effective gene expression programming. Neuro computing, 105 (2013) 100-106.