Title

Subliminal audio-visual temporal congruency in music videos enhances perceptual pleasure

Authors

Lin, Chenyang
Yeh, Maggie
Shams, Ladan

Publication Date

2022-05-01

DOI

10.1016/j.neulet.2022.136623

Peer reviewed

Neuroscience Letters 779 (2022) 136623

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Subliminal audio-visual temporal congruency in music videos enhances perceptual pleasure

Chenyang Lin [a,1], Maggie Yeh [b,1], Ladan Shams [a,b,c,*]

[a] Neuroscience Interdepartmental Program, University of California, Los Angeles, CA, USA
[b] Department of Psychology, University of California, Los Angeles, CA, USA
[c] Department of Bioengineering, University of California, Los Angeles, CA, USA

ARTICLE INFO

Keywords:
Multisensory
Pleasure
Temporal congruency
Audiovisual integration
Subliminal
Perception

ABSTRACT

Human perception is inherently multisensory, with cross-modal integration playing a critical role in generating a coherent perceptual experience. To understand the causes of pleasurable experiences, we must understand whether and how the relationship between separate sensory modalities influences our experience of pleasure. We investigated the effect of congruency between vision and audition in the form of temporal alignment between the cuts in a video and the beats in an accompanying soundtrack. Despite the subliminal nature of this manipulation, a higher perceptual pleasure was found for temporal congruency compared with incongruency. These results suggest that the temporal aspect of the interaction between the visual and auditory modalities plays a critical role in shaping our perceptual pleasure, even when such interaction is not accessible to conscious awareness.

1. Introduction

The positive affective response elicited by sensory stimuli is referred to as *perceptual pleasure*. Biederman and Vessel [1] associate perceptual pleasure from visual pathways to modulation of mu-opioid receptors located in gradients with increasing density from early to later visual processing stages. Utilizing naloxone antagonism, another experiment by Goldstein [2] also demonstrated an association between thrills in music perception and modulation of opioid receptors, potentially within the auditory pathways. Such findings are consistent with opioid receptor's involvement in reward processing, liking, behavior reinforcement, motivation, and pain regulation [3–7]. The field of neuroaesthetics has been speedily growing in the past decades, with accumulating findings on the sensory attributes and neural correlates of aesthetic judgments on a wide range of stimuli, visual and auditory primarily [8–14]. However, perceptual pleasure naturally depends on perceptual processing, and perception is a function of not only inputs from individual sensory modalities but also the interactions among them. Despite the fact that perception, in daily life, involves concurrent processing of inputs from multiple modalities, previous research on perceptual pleasure has been largely restricted to one sensory modality alone, resulting in a major limitation in this field of study. Therefore, how cross-modal interactions contribute to perceptual pleasure and principles governing such effects have yet to be investigated.

As vision and hearing are essential for one's ability to interact with the environment, the integration between the two senses has been extensively scrutinized in recent years. However, the effects of the integration on perceptual pleasure have remained largely unknown. A few studies have utilized audio-visual materials, such as music videos, to investigate emotional responses [15,16]. It has been proposed that observation of music performances, which activates both auditory and visual channels, enhances emotional responses owing to multiple converging sensory cues [17,18]. Along this line of research, Pan et al. [19] demonstrated that, compared to audio-only presentation, audio-visual presentation of music in which the two modalities are congruent in valence enhances one's emotional responses, with the integration effects larger for positive than for negative music. One study pairing affective pictures with musical stimuli found more accurate qualities of emotional responses, assessed by both psychometrical ratings and physiological measurements, in the audio-visual combined condition, compared with audio or visual only conditions [20]. Another study pairing dance movements with instrumental music also revealed shifts in affective responses to the movements through both subjective ratings and physiological measures [21]. The above findings on cross-

* Corresponding author at: Department of Psychology, University of California, Los Angeles, CA, USA.
E-mail address: lshams@psych.ucla.edu (L. Shams).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.neulet.2022.136623
Received 28 October 2021; Received in revised form 31 March 2022; Accepted 5 April 2022
Available online 8 April 2022

modal affective bias support the bidirectional relationship between the affect detection structures in the visual and auditory modalities [21,22].

In this study, instead of focusing on the audio-visual congruency in valence, we aimed to examine the effect of temporal congruency in cross-modal features on reported perceptual pleasure. In film studies, it has been suggested that audio-visual synchronization provides meanings and emphasis and points of synchronization have emotional values [23]. The influence of congruency in onset, duration, and temporal frequency of auditory and visual stimuli on multisensory integration has been studied extensively [24–27]. The temporal discrepancy between audio and visual in videos of talking heads has been reported to lead to experience of displeasure [28]. Therefore, it is reasonable to propose that temporal congruency in audio-visual stimuli enhances one's capability for cross-model integration, which may lead to a higher perceptual pleasure.

Here we examine a more subtle temporal feature, the alignment of the timing of cuts in a video with the timing of the beats in the accompanying music. Because both audio and visual streams are ongoing, the onset and duration of the auditory and visual stimuli are congruent, and yet depending on the rhythm and tempo of the music and the exact position of cuts in the video, the points of emphasis in the two streams may be aligned or misaligned. Because this is a subtle feature, it may not be available to the observer consciously, and in fact, our data revealed that almost none of the observers reported noticing the differences between the aligned and misaligned conditions.

Considering that this perceptual variable is largely subliminal, we asked whether it would influence the experience of pleasure. A change in the experience of pleasure despite the subliminal manipulation would demonstrate that cross-modal interactions play an important role in the experience of pleasure at an early sensory stage when the cross-modal relationship is not even consciously accessible.

2. Experiment 1

2.1. Materials and methods

2.1.1. Participants

A total of 143 subjects participated in this experiment. Participants were undergraduate students enrolled in psychology courses at the University of California, Los Angeles. They were compensated with course credits in exchange for participation. All participants provided written, informed consent, and this study was approved by the Institutional Review Board at the University of California, Los Angeles.

2.1.2. Stimuli

Six video clips were used in this experiment. All six video clips were 30–40 s long and were cut from the same original video, which consisted of time-lapses of Los Angeles at night. The original video was called PANO | LA – 10 K [29] and was selected due to its relatively homogeneous visual content and purely instrumental soundtrack. The soundtrack had clear rhythmic patterns with the first beat within each measure being the accented beat.

Each video-soundtrack pair was presented in one of two conditions: "congruent" or "incongruent". In the congruent version (which was mostly the same as the original video), the cuts in the video between scenes were aligned with the accented first beat of a measure in the soundtrack. In this manner, there was temporal congruency between the video cuts and the auditory accents of the stimulus. In the incongruent condition, the cuts in the video were displaced such that they fell a little less than halfway between the first beats of two consecutive measures in the soundtrack. This was accomplished by shifting the placement of the soundtrack relative to the video. In this manner, there was temporal incongruency between video cuts and auditory accents (See Fig. 1). The video cuts in the stimuli simply corresponded to the transitioning from a time-lapse of one location to a time-lapse of another location in LA. The cuts occurred as designed in the video, and thus did not occur every measure. Rather, each clip might contain around half a dozen cuts but many more measures of music.

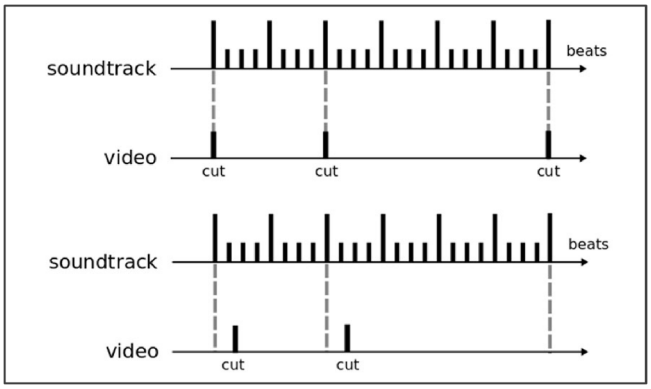

Fig 1. Schematic representation of congruent (top) and incongruent (bottom) conditions. In the soundtrack the first beat of each measure is accented (as is generally the case in most music genres).

The reason for placing the visual cuts a little less than halfway between the auditory accents in the incongruent condition was to avoid participants perceiving an implied congruency between the auditory and visual accents. Music is theoretically infinitely sub-dividable into smaller and smaller equally spaced beats, but a common theme in Western music theory is the division into groups of four beats per measure. Within these groups of four, the first and third beats are always implicitly accented, with slightly more emphasis on the first over the third. As such, if the incongruent condition consisted of visual cuts placed halfway between auditory accents, landing on the third beat of a measure, there is a possibility that participants would still perceive temporal congruency between the two modalities.

The visual stimuli were presented on a CRT monitor, and the auditory stimuli were presented from two speakers positioned symmetrically adjacent to the two sides of the monitor. The videos filled the screen of the monitor.

2.2. Design

Participants were presented with the 6 video-soundtrack clips across the 6 trials. All participants were presented with the same video (and soundtrack) clips, however, the order of the clips and the version (congruent vs. incongruent) varied across participants. For each participant, the order of the video clips was randomized, and the version of the clips was selected pseudo-randomly, with the constraint that an equal number of congruent and incongruent clips were presented (See Fig. 2).

2.3. Procedure

Participants sat 35–40 cm away from the screen where the visual stimuli were presented. Participants were instructed to rate how pleasant they found the videos to be on a continuous rating scale. The scale itself was marked, at the left end, the center, and the right end respectively, "Very Unpleasant", "Neutral", and "Very Pleasant," corresponding to values of −100, 0, and 100. After participants completed the experiment, they were asked if they had noticed anything unusual about the videos they had seen. They were also asked about the factors behind their ratings - for example, some participants cited lighting as influencing their rating choices.

2.4. Results

Because the two versions of each video were not both presented to each subject, to account for variabilities in the usage of the rating scale

Fig. 2. Stimuli and design of Experiment 1. a) A snapshot of the 6 different video clips used. b) Design of the experiment: The two rows represent schematically the order of trials for two participants. Each participant watched the same 6 videos in a random order, with the congruency condition of each video pseudo-randomized, such that each participant viewed a total of three congruent and three incongruent stimuli.

across participants, the scores of each participant were normalized and data analysis was performed using Z-scores. Analysis was also conducted taking into account which video clip was being rated, given that individual differences between the video clips may have also influenced pleasure ratings. A mixed two-way ANOVA with a within-subject factor of condition (with 2 levels: congruent and incongruent) and between-group factor of video ID (with 6 levels corresponding to the 6 video clips) was conducted on subject perceptual pleasure z scores and showed a significant main effect of video ($F = 32.6$, $p < .001$), and a significant effect of condition ($F = 5.32$, $p < .05$). A paired-samples *t*-test confirmed this Condition effect ($t = 2.07$, $p < .05$, Cohen's $d = 0.33$). Fig. 3 shows the violin and box plots for the congruent ($M = 0.075$, $SEM = 0.036$) and incongruent ($M = -0.071$, $SEM = 0.035$) conditions.

With one exception, none of the participants reported noticing a difference between videos in terms of alignment of music and video cuts. The one participant (out of 145) who was consciously aware of the temporal congruency manipulation had extensive experience in video editing. All other participants reported that the visual characteristics of the videos influenced their ratings.

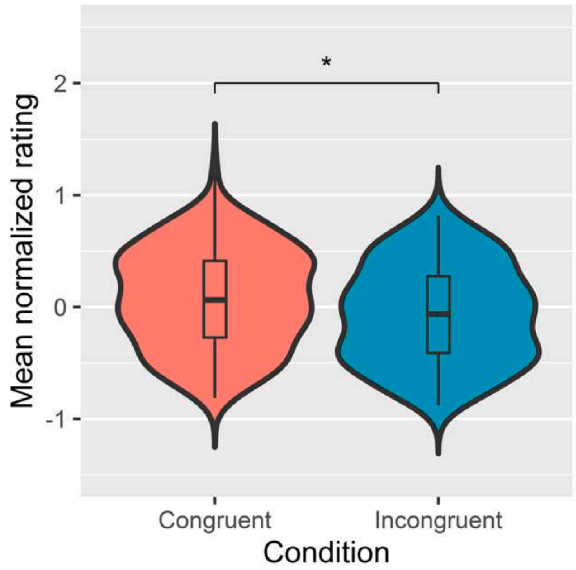

Fig. 3. Normalized scores (z-scores) for congruent and incongruent conditions. The left violin plot shows the distribution of the mean scores for congruent videos, while the right plot shows the distribution of mean scores for the incongruent videos across participants. Internal box plots depict the median and 1st and 3rd quartiles of each distribution. The asterisk indicates the significant difference between the mean ratings of the two conditions ($p < 0.05$).

2.5. Discussion

Experiment 1 examined the effects of temporal congruency between temporal features of concurrent auditory and visual stimuli upon the pleasure experienced by participants from observing the stimuli. The results indicate that congruency can in fact play a role in our preferences of video clips, specifically in terms of temporal congruency between the accents in the music and cuts in the video. Observers showed an overall preference for the congruent condition despite lack of verbal reporting of the alignment difference between the congruent and incongruent condition. It is possible that the participants did in fact notice the alignment difference but were simply not equipped with the expertise to report it. However, our results do suggest that the possibility of the subliminal effects of audiovisual temporal congruency bears further

investigation.

3. Experiment 2

In Experiment 1, the number of video clips presented per participant was relatively small, and all clips were segments of the same video, having the same visual theme, music, and editing style. Moreover, subjects in Experiment 1 viewed each video either in a congruent or incongruent condition, making a within-subject comparison of the two versions of the same clip not possible and reducing the experimental power. The goal of this experiment was to address the limitations of Experiment 1 and test for the generality of the effect of temporal congruency on the experienced pleasure of music videos.

Furthermore, to minimize the possibility of a syncopation effect, the relative timing of the video cuts in the asynchronous condition was slightly modified. A within-subject design was employed allowing direct comparison of the two versions of the same exact video clip for each participant. Importantly, a larger set of stimuli, with a significant variety of visual themes and music was used to probe the generality of the effect.

3.1. Materials and methods

3.1.1. Participants

A total of 30 subjects (12 males and 18 females ranging from 18 to 22 years of age, with an average age of 19.87 years) participated in this experiment. Subjects were undergraduate students enrolled in psychology courses at the University of California, Los Angeles. They were compensated with course credits in exchange for participation. All participants provided written informed consent and no participant reported a medical history of epilepsy, stroke, or head trauma prior to the experiment. This study was approved by the Institutional Review Board at the University of California, Los Angeles.

3.1.2. Stimuli

Thirty video clips were selected for this experiment. All clips were 30 $\pm$ 1 s long. The clips were cut from thirty videos collected online, covering a wide spectrum of contents, including landscapes, animals, daily routines, and sports. (see Fig. 4). The clips had a similar number of cuts (M = 13.33, SD = 2.14) leading to a time interval of 2–3 s between two consecutive cuts. The clips were then paired with purely instrumental soundtracks edited to match the lengths of the videos (30 $\pm$ 1 s). All soundtracks and videos are non-copyrighted materials sourced from YouTube. The selection of the original videos excluded those with a negative valence or a clear narrative. The soundtracks had varying tempos and were selected based on clear rhythmic patterns with the first beat within each measure being the accented beat and the meters being in multiples of two.

For each video clip, two versions were created: "congruent" or "incongruent." In the congruent version, the cuts in the video between scenes were aligned with the accented first beat of a measure in the soundtrack. In the incongruent version, to reduce the effects of potential syncopations, the video cuts were placed between the third beat and third offbeat of a measure in the soundtrack. To achieve this, we temporally shifted the placement of the video relative to the soundtrack 2–2.5 beats posterior to the congruent versions. In this manner, there was temporal incongruency between video cuts and auditory accents while the frequencies of the video cuts and the tempos of the soundtracks were separately maintained (See Fig. 5).

3.2. Design

Each subject participated in two sessions, with an approximately 24-hour gap in between. In both sessions, participants were presented with all the 30 video-soundtrack clips across the 30 trials, and the order of presentations of the videos was the same across the two sessions. The only difference between the two sessions was that the version of each video was alternated. The version (congruent vs. incongruent) of each video was selected with a random process in session 1, leading to an approximately equal number of congruent and incongruent clips presented in a randomly interleaved fashion. In session 2, the alternative version of each video was presented in the same order as in session 1 (see Fig. 6). The order of the presentation of videos in session 2 was identical to that of session 1 for the same subject. However, to increase the variation in the order of videos across individuals, two sequences of videos were used across participants. Half of the participants were assigned the first order, and half were assigned the second.

3.3. Procedure

Due to the occurrence of the COVID-19 pandemic during the time of the experiment, subjects in Experiment 2 participated in the study through the Qualtrics software [30], an online survey platform, under observance from a researcher over the Zoom conference [31]. Subjects were required to share their computer screens and turn on their cameras and microphones before the experiment, so the experimenter could remotely monitor the experimental sessions.

After watching the video in each trial, participants were asked to rate how pleasant they found the videos to be on a continuous rating scale.

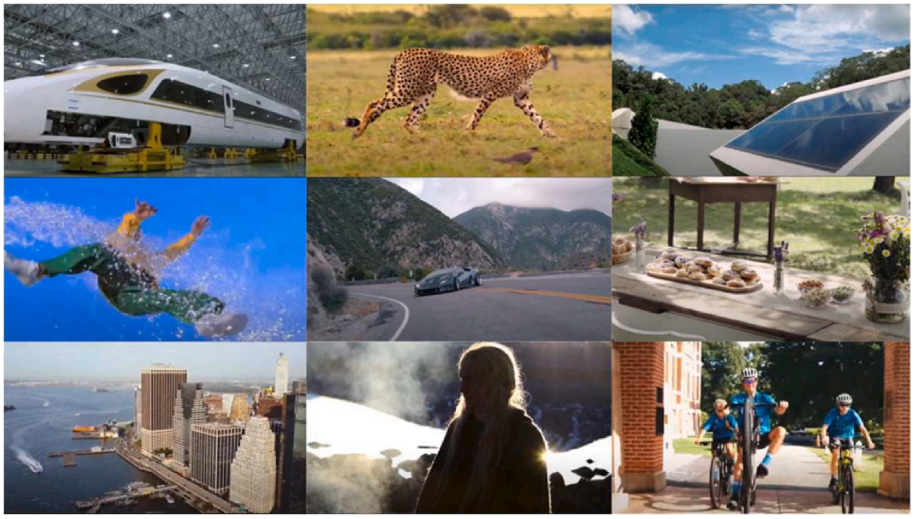

Fig. 4. Examples of frames in the video stimuli used in Experiment 2. The videos cover a wide range of contents.

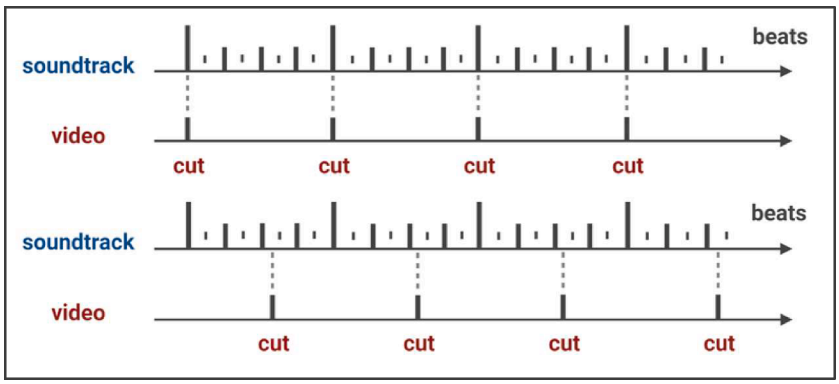

Fig. 5. Schematic representation of congruent (top) and incongruent (bottom) conditions. The longer ticks in the soundtrack timeline represent the accented beats of each measure.

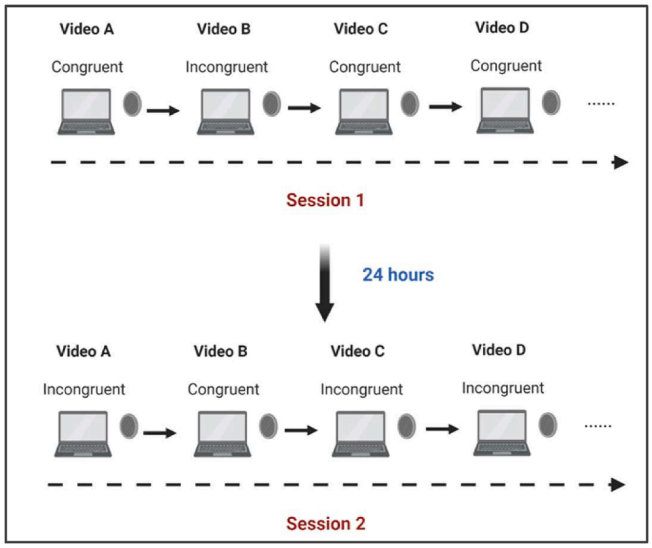

Fig. 6. Design of Experiment 2. Each subject participated in two sessions, with a ~ 24-hour gap in between. The congruency condition for each video in session 1 is randomized and is reversed in session 2.

The scale itself was marked, at the left end, the center, and the right end respectively, "Very Unpleasant", "Neutral", and "Very Pleasant," corresponding to values of −100, 0, and 100. At the end of the second session, participants were asked to complete a brief post-experiment questionnaire consisting of a series of demographic questions and questions regarding past audio-visual experiences, such as video gaming and video editing. They were also asked if they had noticed any systematic patterns or anything unusual in the ways the videos were presented.

3.4. Results

A mixed two-way ANOVA, with a within-subject factor of condition (congruent vs. incongruent) and between-group factor of video order, was conducted. No significant interaction between the two factors and no significant effect of the video order was found. However, the results showed a significant effect of condition ($F(29, 59) = 4.76$, $p < .05$). The ratings of congruent videos were significantly higher than the incongruent videos ($t(29) = 2.16$, $p = 0.019$, Cohen's $d_z = 0.395$) (See Fig. 7). Additionally, none of the participants reported noticing any differences between videos in terms of the alignment between music beats and video cuts.

4. General discussion

Experiment 1 used a between-subjects design and showed a preference for congruent music videos compared to incongruent music videos even though both the video and audio components were identical, and the difference in the relationship between the video and audio was not in an obvious temporal marker, such as onset or offset, but rather in a

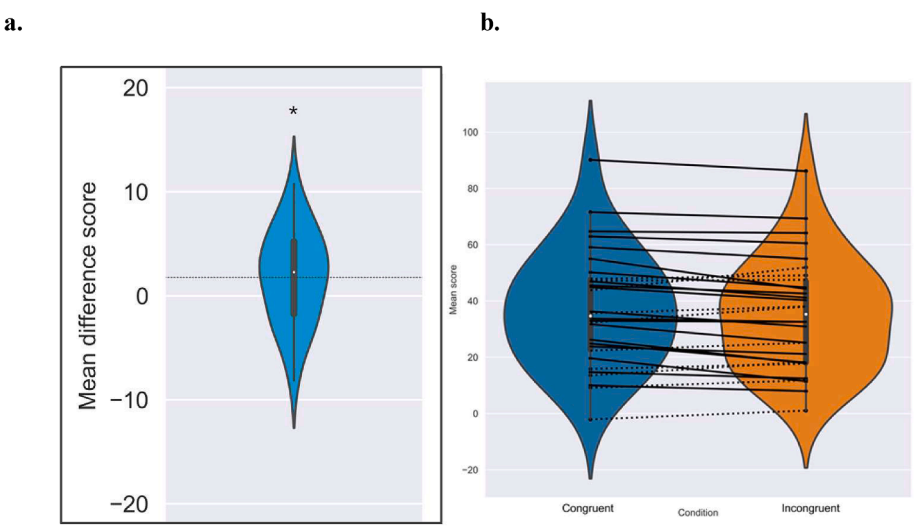

Fig. 7. (a) Mean difference scores (congruent - incongruent). The violin plot represents the distribution of mean difference scores across participants, and the dotted horizontal line represents the overall mean difference score. A significantly higher mean perceptual pleasure was found for congruency than for incongruency. **(b) Mean scores for congruent and incongruent conditions.** The left violin plot shows the distribution of the mean scores for the congruent condition, while the right plot shows the distribution of mean scores for the incongruent condition across participants. The mean scores for the two conditions were connected with lines for each participant. Solid lines show higher mean rating scores for the congruent condition, while the dotted lines show higher mean rating scores for the incongruent condition.

subtle temporal feature that the participants were not aware of. Experiment 2 aimed to investigate the robustness and generality of these findings. A within-subject experimental design and a larger number of trials were employed to improve the experimental power of the study. A larger set of stimuli, with greater variation in content across trials were used to examine the generality of the effect. Experiment 2 supported the results of Experiment 1 in that higher pleasure ratings were reported by observers for congruent versions of the video clips. This was despite the fact that observers, again, were not reporting any noticed difference regarding the manipulation of temporal congruency between the beats in the soundtracks and the cuts in the videos. In fact, the participants did not report noticing any manipulation of any relationship between the audio and video. The effect of congruency is not only robust but also relatively sizable. The effect sizes in both experiments were also quite consistent (0.33 stds in Exp. 1, and 0.39 stds in Exp. 2) despite the discrepancies in design and stimuli.

The experiments were planned in a way that the change in the temporal relationship between the audio and video is so subtle that it would not be consciously noticeable to the participants. Strictly speaking, we cannot entirely rule out the possibility that some of the participants were aware of the difference but decided not to report it. However, in the open-ended questions, in which we probed whether they noticed any pattern in the presentation of the stimuli or whether they noticed any manipulation of the videos, none of the participants reported any clue about any relationship between the audio and video, with the exception of one participant (the experienced video-editor in Experiment 1). Instead, they reported other suspicions and guesses. Therefore, these results strongly support a lack of participants' awareness of the temporal relationship.

According to Hekkert's Principle of "maximum effects for minimum means'' for design aesthetics [32], human brains are designed such that sensory stimuli that could be processed with a minimal amount of brain capacity and, in other words, are more perceptually fluent are preferred. Another study by Reber et al. [33] also supported that perceptual fluency enhances liking and positive affective judgment. Additionally, previous research has shown that temporal congruency plays an essential part in multisensory integration [24–27,34–40]. Therefore, it would be reasonable to propose that an enhanced cross-modal integration between the auditory and visual modalities may lead to better audio-visual perceptual fluency, which in turn increases perceptual pleasure. Such an interpretation may invite counter arguments, however, as the debate has long existed between aesthetic pleasure of novel versus familiar/fluent sensory stimuli [41]. Empirical evidence also revealed that visual complexity, instead of linearly reducing pleasure in perception, as the perceptual fluency theory might predict, actually associates with aesthetic pleasure following a Berlyne's U-shaped curve [42]. However, as Biederman & Vessel [1] suggested, the ultimate point is the subject's understanding of the sensory stimuli. Here, our interpretation of the results is grounded on the fact that the subjects were exposed to the stimuli for only twice across two days and were thus fairly unfamiliar with them. Therefore, while the general level of novelty remains comparable between the stimuli in the two sessions, a higher fluency in perception resulting from temporal congruency may facilitate the understanding of the stimuli and thus increase aesthetic pleasure. It still remains to be investigated whether, over a longer term of repetitive exposures, subjects would experience a higher degree of boredom towards congruent videos, and thus whether Biederman and Vessel's theory could be applied. Secondly, while the rhythmic auditory stimulation from the music results in an auditory entrainment, the temporal synchronization between the video cuts may formulate a type of subtle audio-visual entrainment, which would lead to a more pleasurable subjective experience. Previous research focusing on audio-visual entrainment has also found its association with different positive effects, including attention enhancement, calmness, and pain relief over a long term [43,44]. How such a technique could be used to induce positive affective response to sensory stimuli in a short term needs further investigation. Finally, the higher preference for temporal congruency may be the consequence of a better fit with predictive coding, in that the prediction errors of expecting the locations of video cuts in the continuum of the music is reduced when the cuts lie on the accented first beats in a measure, compared to when they lie on an unaccented place. As less weight would be put on the incoming stimuli in the congruent condition, less cognitive effort would be needed in correcting the prediction errors or updating the mental model derived from one's prior perception.

A limitation of the present study is that Experiment 2 was conducted online. Even though the participants were monitored by experimenters via video conference, environmental factors, such as the locations and devices, could not be controlled as strictly as in a lab setting and might, therefore, interfere with the effects of audio-visual temporal congruency we expected to see. Additionally, the measurements in the present study were restricted to subjective reports of pleasure rating. Future studies may consider using additional dependent variables (such as physiological recordings), including a wider variety of videos, and exploring the role of arousal in the ratings.

5. Conclusion

This study investigated the effects of a subliminal congruency between two subtle features of audio and video on perceptual pleasure, specifically in terms of the temporal congruency between the cuts in the video and the accents in the accompanying music. The results indicate that subliminal temporal congruency enhances perceptual pleasure.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to thank Bijan Mehdizadeh, Ashley Chen, Leah Eslamian, and Lakshita Vij for their help with data collection.

References

[1] I. Biederman, E.A. Vessel, Perceptual pleasure and the brain: A novel theory explains why the brain craves information and seeks it through the senses, Am. Sci. 94 (3) (2006) 247–253.
[2] A. Goldstein, Thrills in response to music and other stimuli, Physiol. Psychol. 8 (1) (1980) 126–129.
[3] J. Le Merrer, J.A. Becker, K. Befort, B.L. Kieffer, Reward processing by the opioid system in the brain, Physiol. Rev. (2009).
[4] V. Trezza, R. Damsteegt, E.M. Achterberg, L.J. Vanderschuren, Nucleus accumbens μ-opioid receptors mediate social reward, J. Neurosci. 31 (17) (2011) 6362–6370.
[5] S. Ghozland, H.W. Matthes, F. Simonin, D. Filliol, B.L. Kieffer, R. Maldonado, Motivational effects of cannabinoids are mediated by μ-opioid and κ-opioid receptors, J. Neurosci. 22 (3) (2002) 1146–1154.
[6] M. Funada, T. Suzuki, M. Narita, M. Misawa, H. Nagase, Blockade of morphine reward through the activation of κ-opioid receptors in mice, Neuropharmacology 32 (12) (1993) 1315–1323.
[7] J.K. Zubieta, Y.R. Smith, J.A. Bueller, Y. Xu, M.R. Kilbourn, D.M. Jewett, C. S. Stohler, Regional mu opioid receptor regulation of sensory and affective dimensions of pain, Science 293 (5528) (2001) 311–315.
[8] I. Aharon, N. Etcoff, D. Ariely, C.F. Chabris, E. O'connor, H.C. Breiter, Beautiful faces have variable reward value: fMRI and behavioral evidence, Neuron 32 (3) (2001) 537–551.
[9] T. Jacobsen, R.I. Schubotz, L. Höfel, D.Y.V. Cramon, Brain correlates of aesthetic judgment of beauty, Neuroimage 29 (1) (2006) 276–285.
[10] A. Ishai, S.L. Fairhall, R. Pepperell, Perception, memory and aesthetics of indeterminate art, Brain Res. Bull. 73 (4–6) (2007) 319–324.
[11] S. Koelsch, Towards a neural basis of music-evoked emotions, Trends in cognitive sciences 14 (3) (2010) 131–137.
[12] A. Chatterjee, O. Vartanian, Neuroscience of aesthetics, Ann. N. Y. Acad. Sci. 1369 (1) (2016) 172–194.
[13] A. Gallace, C. Spence, Tactile aesthetics: towards a definition of its characteristics and neural correlates, Social Semiotics 21 (4) (2011) 569–589.
[14] M. Reybrouck, P. Vuust, E. Brattico, Neural Correlates of Music Listening: Does the Music Matter? Brain Sciences 11 (12) (2021) 1553.

[15] I.E. Lee, C.F.V. Latchoumane, J. Jeong, Arousal rules: An empirical investigation into the aesthetic experience of cross-modal perception with emotional visual music, Front. Psychol. 8 (2017) 440.
[16] M. Van Elk, M.A. Arciniegas Gomez, W. van der Zwaag, H.T. Van Schie, D. Sauter, The neural correlates of the awe experience: Reduced default mode network activity during feelings of awe, Hum. Brain Mapp. 40 (12) (2019) 3561–3574.
[17] S.R. Livingstone, W.F. Thompson, The emergence of music from the Theory of Mind, Music Sci. 13 (2_suppl) (2009) 83–115.
[18] C. Chapados, D.J. Levitin, Cross-modal interactions in the experience of musical performances: physiological correlates, Cognition 108 (3) (2008) 639–651.
[19] F. Pan, L. Zhang, Y. Ou, X. Zhang, The audio-visual integration effect on music emotion: Behavioral and physiological evidence, PLoS ONE 14 (5) (2019), e0217040.
[20] T. Baumgartner, M. Esslen, L. Jäncke, From emotion perception to emotion experience: emotions evoked by pictures and classical music, Int. J. Psychophysiol. 60 (2006) 34–43.
[21] J.F. Christensen, S.B. Gaigg, A. Gomila, P. Oke, B. Calvo-Merino, Enhancing emotional experiences to dance through music: the role of valence and arousal in the cross-modal bias, Front. Hum. Neurosci. 8 (2014) 757.
[22] B. De Gelder, J. Vroomen, The perception of emotions by ear and by eye, Cogn Emot. 14 (3) (2000) 289–311.
[23] M. Chion, Audio-vision and sound, Sound. (2000) 201–221.
[24] R. Arrighi, D. Alais, D. Burr, Perceptual synchrony of audiovisual streams for natural and artificial motion sequences, J. Vision 6 (3) (2006) 6.
[25] G. Aschersleben, P. Bertelson, Temporal ventriloquism: Crossmodal interaction on the time dimension: 2. Evidence from sensorimotor synchronization, Int. J. Psychophysiol. 50 (1–2) (2003) 157–163.
[26] Y.H. Su, Content congruency and its interplay with temporal synchrony modulate integration between rhythmic audiovisual streams, Front. Integr. Neurosci. 8 (8) (2014 Dec) 92.
[27] R.B. Welch, L.D. DutionHurt, D.H. Warren, Contributions of audition and vision to temporal rate perception, Perception & psychophysics. 39 (4) (1986 Jul) 294–300.
[28] R. Steinmetz, Human perception of jitter and media synchronization, IEEE J. Sel. Areas Commun. 14 (1) (1996) 61–72.
[29] SCIENTIFANTASTIC. PANO | LA - 10K [Video]; 2016. Los Angeles. https://vimeo.com/188611665.
[30] The data or this paper was generated using Qualtrics software, Version: March 2021 of Qualtrics. Copyright © 2021 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com.
[31] A. Store Zoom Video Communications, Inc. ZOOM cloud meetings (Version 5.4.7) 2020 [Mobile app].
[32] P. Hekkert, Design aesthetics: principles of pleasure in design, Psychol. Sci. 48 (2) (2006) 157.
[33] R.W.P. Reber, N. Schwarz, Effects of perceptual fluency on affective judgments, Psychol. Sci. 9 (1) (1998).
[34] P. Maragos, P. Gros, A. Katsamanis, G. Papandreou, Cross-modal integration for performance improving in multimedia: A review, Multimodal Technol. Interact. (2008) 1–46.
[35] R.L. Miller, B.E. Stein, B.A. Rowland, Multisensory integration uses a real-time unisensory–multisensory transform, J. Neurosci. 37 (20) (2017) 5183–5194.
[36] L. Shams, R. Kim, Crossmodal influences on visual perception, Phys. Life Rev. 7 (3) (2010) 269–284.
[37] C. Spence, Audiovisual multisensory integration, Acoust. Sci. Technol. 28 (2) (2007) 61–70.
[38] B.E. Stein (Ed.), The new handbook of multisensory processing, MIT Press, 2012.
[39] R.A. Stevenson, S.H. Baum, J. Krueger, P.A. Newhouse, M.T. Wallace, Links between temporal acuity and multisensory integration across life span, J. Exp. Psychol. Hum. Percept. Perform. 44 (1) (2018) 106.
[40] A. Vatakis, C. Spence. Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. In Multisensory object perception in the primate brain 2010 (pp. 95-121). Springer, New York, NY.
[41] J. Song, Y. Kwak, C.Y. Kim, Familiarity and Novelty in Aesthetic Preference: The Effects of the Properties of the Artwork and the Beholder, Front. Psychol. 12 (2021).
[42] P. Chassy, T.A. Lindell, J.A. Jones, G.V. Paramei, A relationship between visual complexity and aesthetic appraisal of car front images: An eye-tracker study, Perception 44 (8–9) (2015) 1085–1097.
[43] F.J. Boersma, C. Gagnon, The use of repetitive audiovisual entrainment in the management of chronic pain, Medical Hypnoanalysis Journal. (1992). Sep.
[44] M. Joyce, D. Siever, Audio-visual entrainment program as a treatment for behavior disorders in a school setting, J. Neurother. 4 (2) (2000) 9–25.