

This paper reports a study that examines time estimation, judged sense of arousal, judged valence, and pupillometry of three types of stimuli: film (V), film + music (VA), and music (A). The within-groups design is well set to add to current empirical knowledge about the impact of music in film perception (or more specifically the impact of music on the sense of passing time, arousal level and valence, and pupillary responses) as no study has examined all of these dependent measures simultaneously before. The researchers were interested in determining whether the modality condition (A, V, or AV) would influence time estimation, and they predicted longer duration for the multimodal (AV) condition (based on Wöllner et al., 2018). They were also interested in determining whether time estimation would be predicted from arousal ratings and by pupillometry. The study led to several observations: time was underestimated for all conditions, but less so for the AV condition (contrary to the expectation). Arousal level as indexed by pupil size and ratings also predicted duration – the higher the arousal the longer the judged duration. The effect was stronger for the AV condition. An incidental finding was that the number of cuts in the film led to longer time estimates. The figures are for the most part quite helpful.

The authors interpret their results in terms of a biological model of time estimation of pulse counting which, if I understand correctly, postulates a different timing system for each modality, and more pulses to count in the AV condition than either A or V condition alone.

The topics of audiovisual integration, time perception, and the role of music in film perception, explored in the present study, would be of interest to readers of this journal. Yet there are several concerns about this paper. On page 16, it is mentioned that 2 of the 27 clips were removed from the analysis because they contained dialogue or written text. The authors treat the data for these clips as missing. One wonders if the data for these clips appeared as outliers and alerted the researchers to removal, or was it that somehow after running the entire experiment, the authors became aware of the problem without the data appearing as outliers.

Much space is directed to explaining the development of the various measures, and still some of them were not clear. I was wondering if some of these could have a simpler explanation in the Method section, but an expanded version in an appendix or online supplement including a worked example.

The design chosen by the authors of presenting a component of the same original clip to each participant in each of three groups (i.e., A, V, or AV) was ideal, and it seemed that it might have provided a way of performing a deeper analysis regarding how the information in A and V came together (linear or nonlinearly) to form AV. Perhaps their statistical analysis provide that information, but the expression of this was not clear to me.

Finally, the authors overlook the narrative aspect of film, and the implication of degree of narrative (or meaningfulness, or engagement) on time estimation. Visual clips alone subjectively (intuitively) are not as engaging as either the music or the video plus music. It

seems an oversight to overlook the storytelling aspect of film and not to consider that degree of engagement in a story might have an impact on time estimation. The biological theory of time estimation (pace-maker-counter device) deserves acknowledgement as the authors have done, but a more human-centred theory might explain that the AV clips are more engaging and more coherent and therefore lead to the particular pattern of time estimation that was obtained. The early work of Boltz (1995) seems to take this other approach, and see also Bangert et al (2020).

Bangert, A. S., Kurby, C.A., Hughes, A. S., & Carrasco, O. (2020). Crossing event boundaries changes prospective perceptions of temporal length and proximity. <https://doi.org/10.3758/s13414-019-01829-x> *AP&P*, 82, 1459-1472.

Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments . *Perception & Psychophysics*, 57, 1080-1096.

Below are some individual line by line comments.

p. 8 reference is made to perceived absence of film music, yet the “unheard melodies” of film as named and discussed by Claudia Gorbman (1987) *Unheard melodies: narrative film music*, Indiana University Press, is an equally important if not more important phenomenon of film. It may be presumptuous for the authors to suggest the conscious hearing of film music when psychologically the film music may not be fully in consciousness (attended). A complex question for attention research concerns the level of consciousness directed to film music. It seems that the “unheard” aspect of film music should be acknowledged.

Reference is made to Jones and Boltz (1989). This would seem to be a good segue to the paper by Boltz (1995) that actually does estimate time of a film. Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments . *Perception & Psychophysics*, 57, 1080-1096. The two papers seem to require a different theory than the pulse counting theory.

p. 9 Hypotheses are made based on Wöllner et al., but it is pointed out in the present paper, they did not run the uni-modal auditory condition. The rationale therefore is not clear for stating that the effect of arousal would be strongest in the multimodal condition (and incidentally. Incidentally, a recently published paper in *AP&P* by Wöllner and Hammerschmidt (2021) <https://doi.org/10.3758/s13414-020-02227-4> shows no effect of musical arousal on time estimation.

p. 10 – if the bone-anchored device is a cochlear implant, it would be reasonable to say so.

p. 10 – describing the films in the LIRIS-ACCEDÉ data base as “often little known” seems less informative than is warranted, when “Most of the 160 movies used for creating LIRISACCEDÉ come from the video platform VODO. This references best free-to-share feature films and short films that have been submitted on the website and makes them easily available to millions of people.” (Bavay et al., 2015, p. 5).

I am not familiar with some of the methods described and therefore can only assume the authors are on the right track (e.g., Satterthwaite's method and the Imer Test package; cluster-size correction permutation analysis). I am quite lost on p. 22, and would be happy with a worked example in an appendix, unless it is felt that the majority of the readers of the paper will be able to follow easily.

p. 23 refers to black lines on Fig. 2A. It is not clear what this refers to.

In summary, a lot of original and useful work has gone into this paper. Some further work seems needed with a focus on developing a more logical rationale (it does not need to be complex) for the study and following through with a simpler presentation of the data, discussion (with expressed limitations) and conclusion

Signed

Annabel Cohen