CULTURA & SAÚDE

ISSN 2595-7546
Vol. 26 · No. 3 · August 2026
São Paulo, Brasil

ARTICLE · CULTURA & SAÚDE

Background music increases viewer engagement: a systematic review and meta-analysis

A meta-analysis of 26 effects separates the attentional benefit from the one known cost

Alexandre Campos Moraes Amato

PhD, University of São Paulo (USP). Vascular and endovascular surgeon, Amato – Institute of Advanced Medicine. Brazilian Lipedema Association (ABL). São Paulo, SP, Brazil

Received for publication in August 2026 · Section: Article · Language: English

ABSTRACT

The author synthesises, in a systematic review and meta-analysis, the full range of outcomes tested during exposure to video with and without background music. A total of 2,574 records were screened across six bibliographic databases and web sources; thirteen reports contributed 15 independent samples and 26 effect sizes, pooled in multilevel random-effects models. Music raised attentional engagement largely and consistently (Hedges g = 1.06), a result replicated across autonomic, ocular, electrophysiological and self-report measures, whose failure modes are unrelated to one another. No outcome family showed a statistically reliable cost. The memory penalty proved confined to tests competing with concurrent narration (g = -0.46), approaching zero without such competition (g = -0.13). The author concludes that background music is a well-evidenced tool for increasing engagement, immersion and enjoyment, with a single boundary condition, which is a matter of placement rather than avoidance.

KEYWORDS: *background music; attentional engagement; meta-analysis; instructional video; immersion; cognitive load*

RESUMO

A música de fundo acompanha quase toda produção audiovisual profissional, mas a literatura acadêmica tem sido lida como advertência, dominada pelo achado de que ela prejudica a aprendizagem de material instrucional. Essa leitura repousa sobre fatia estreita da evidência: engajamento, imersão e fruição, que motivam a decisão de produção, nunca foram sintetizados quantitativamente ao lado da literatura de memória. O objetivo foi avaliar toda a gama de desfechos já testados durante a exposição audiovisual. Foram triados 2.574 registros de seis bases bibliográficas e de fontes web em busca de estudos que contrastassem presença e ausência de música durante a visualização de vídeo, em três famílias pré-especificadas: engajamento atencional, memória do conteúdo e persistência comportamental. Os efeitos foram expressos em *g* de Hedges e agregados em modelos multiníveis de efeitos aleatórios. Treze relatos contribuíram com 15 amostras independentes e 26 tamanhos de efeito. A música aumentou o engajamento atencional de forma ampla e consistente (*g* = 1,06; IC 95% [0,76; 1,37]), replicado em quatro modalidades de medida com modos de falha não relacionados. Nenhuma família apresentou custo estatisticamente confiável. O prejuízo à memória restringiu-se a testes que competem com a narração simultânea (*g* = −0,46), sendo próximo de zero sem tal competição (*g* = −0,13). A música de fundo é ferramenta bem evidenciada para aumentar engajamento, imersão e fruição, com uma

única condição de contorno: música sobreposta à fala durante instrução prejudica a evocação verbal — questão de posicionamento, não de evitação.

Palavras-chave: Atenção. Imersão. Trilha sonora. Vídeo instrucional. Multimídia.

RESUMEN

La música de fondo acompaña casi toda producción audiovisual profesional, pero la literatura académica se ha leído como advertencia, dominada por el hallazgo de que perjudica el aprendizaje de material instruccional. Esa lectura se apoya en una porción estrecha de la evidencia: el compromiso, la inmersión y el disfrute, que motivan la decisión de producción, nunca se sintetizaron cuantitativamente junto a la literatura de memoria. El objetivo fue evaluar todo el rango de desenlaces ya probados. Se cribaron 2.574 registros de seis bases bibliográficas y fuentes web que contrastaran presencia y ausencia de música durante la visualización de vídeo, en tres familias preespecificadas: compromiso atencional, memoria del contenido y persistencia conductual. Los efectos se expresaron como *g* de Hedges y se agregaron en modelos multinivel de efectos aleatorios. Trece informes aportaron 15 muestras independientes y 26 tamaños de efecto. La música aumentó el compromiso atencional de forma amplia y consistente (g = 1,06; IC 95% [0,76; 1,37]), replicado en cuatro modalidades de medida con modos de fallo no relacionados. Ninguna familia mostró un costo estadísticamente confiable. El perjuicio a la memoria se restringió a pruebas que compiten con la narración simultánea (g = −0,46) y fue casi nulo sin tal competencia (g = −0,13). La música de fondo es herramienta bien evidenciada para aumentar el compromiso, la inmersión y el disfrute, con una única condición de contorno: la música superpuesta al habla durante la instrucción perjudica la evocación verbal — cuestión de ubicación, no de evitación.

Palabras clave: Atención. Inmersión. Banda sonora. Vídeo instruccional. Multimedia.

INTRODUCTION

Background music accompanies most professionally produced video, and the practitioner rationale is straightforward: music makes the viewing experience more absorbing. The academic literature, however, has often been summarised as advising against it. The coherence principle in the cognitive theory of multimedia learning holds that non-essential audio consumes limited processing capacity, and the finding that adding an instrumental music loop to a narrated animation reduced both recall and problem-solving transfer (MORENO; MAYER, 2000) has been widely generalised into a recommendation to strip music from instructional video.

That generalisation outruns its evidence in two ways. First, it treats a result obtained under one specific configuration — instrumental music playing concurrently with narration during a short animation — as a property of music in general. Second, it privileges a single outcome, verbal recall, over the outcomes that motivate the production choice in the first place: whether viewers are engaged, absorbed and enjoying what they watch. The gap, then, is not one of data but of synthesis: what is missing is a common metric on which the documented benefits can be weighed against the documented cost.

A parallel literature has been measuring exactly those outcomes, and its results are strikingly consistent. Music raises autonomic arousal across video genres (WÖLLNER; HAMMERSCHMIDT; ALBRECHT, 2018), increases pupil diameter (HAMMERSCHMIDT; WÖLLNER, 2018), produces greater alpha-rhythm suppression and larger P300 amplitudes than silence (KWON; LEE; LEE, 2022), and suppresses default-mode network activity associated with mind-wandering during free viewing (BEZDEK; WENZEL; SCHUMACHER, 2017). This is the state of the art on the engagement side, and it has never been confronted quantitatively with the memory literature.

This work provides that synthesis. The question is not whether music helps or harms in the abstract, but what it demonstrably does, how large those effects are, how consistently they replicate, and under precisely which conditions the one known cost appears.

MATERIAL AND METHODS

ELIGIBILITY

The review followed PRISMA reporting conventions. Eligible records involved human participants, an explicit contrast between a music-present and a music-absent condition, and a quantitative outcome in one of three prespecified families: attentional engagement (psychophysiological indices, eye movements, self-reported immersion, arousal or enjoyment); content memory (recall, recognition, comprehension, transfer); and behavioural persistence (automatically logged duration of voluntary exposure). Designs contrasting only different kinds of music, with no silence baseline, were recorded but excluded from pooling. No date, language or publication-status restrictions applied.

SEARCH AND SCREENING

Searches ran across OpenAlex, PubMed, arXiv, Crossref, bioRxiv/medRxiv and general web sources in three parallel tracks. Because the persistence literature proved sparse for video, that search was extended to adjacent domains where music's effect on time spent is the dependent variable, with domain recorded as a coded variable. After keyword searching, the three most relevant hits per track were expanded one step backward through their

reference lists and one step forward through their citing works. Every DOI was resolved against Crossref or OpenAlex, and all bibliographic metadata was taken from the resolved record.

EFFECT SIZES AND SYNTHESIS

All effects were expressed as Hedges *g*, signed so that positive values indicate music improved the outcome. Where per-condition means and standard deviations were reported, *g* was computed directly, pooling music-present cells across tempo or congruency levels. Otherwise *t* and *F* were converted to *d* and bias-corrected, and partial eta squared was converted via Cohen's *f*. Within-subject contrasts used the *d*av metric assuming $r = 0.5$. Each conversion is documented, effect by effect, in the supplementary dataset.

Two extraction decisions warrant note. In Kwon, Lee and Lee (2022), the dispersions reported around the EEG and event-related potential means are implausibly small relative to the reported *F* statistics and appear to be standard errors; effect sizes were therefore derived from the reported partial eta squared. In Moreno and Mayer (2000), the marginal means presented in the text are inconsistent in sign with the cell means in Table 2 of the original article; the cell means were used, as they match the reported direction.

Each family was pooled in a multilevel random-effects model (REML) with random intercepts at study and effect level. Pooled estimates are reported with 95% confidence and prediction intervals, τ^2 at both levels, Cochran's *Q*, and a multilevel I^2. Prespecified subgroup analyses split the memory family by test depth and the attention family by measurement type. Small-study effects were assessed by regressing effect size on standard error and by trim-and-fill on study-aggregated estimates. Analyses used the metafor 5.0-1 package in R.

RESULTS AND DISCUSSION

Of 2,574 screened records, 61 were assessed in full, 42 entered the qualitative synthesis, and 13 reports, contributing 15 independent samples and 26 effect sizes, were poolable (figure 1). The 26 individual effects, organised by outcome family, are presented in figure 4.

Figure 1 – PRISMA flow diagram of study selection.

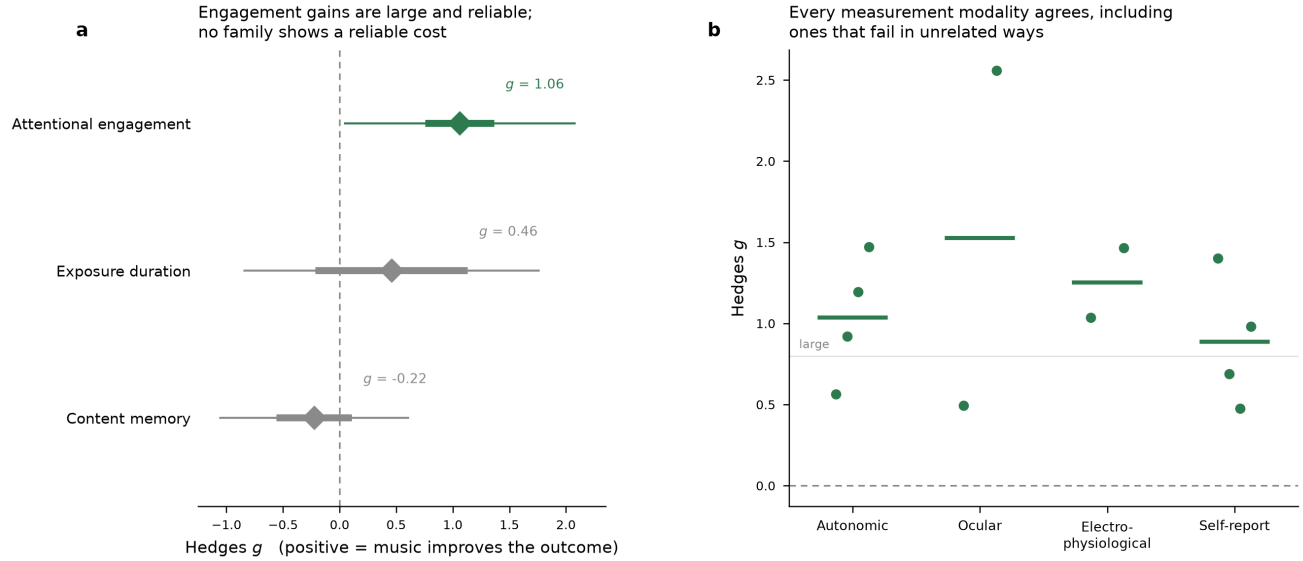
THE ENGAGEMENT BENEFIT IS LARGE, CONSISTENT AND MODALITY-INDEPENDENT

Background music increased attentional engagement by g = 1.06 (95% CI [0.76, 1.37]; $p < 0.0001$; k = 12 effects from six reports; figure 2a). Every included effect was positive. Heterogeneity was high (I^2 = 90%), but the prediction interval remained entirely above zero ([0.05, 2.08]), meaning a new study in this family would be expected to find a positive effect.

The strength of this result lies in its convergence (figure 2b). Autonomic measures (g = 1.04; for example, skin conductance during a suspense film, BENTE et al., 2022), ocular measures (g = 1.53), electrophysiological measures (g = 1.25) and self-report (g = 0.89) all detect the same benefit, and these methods have entirely unrelated failure modes: demand characteristics that could inflate a self-report cannot inflate skin conductance, and electrode artefacts that could distort an EEG signal cannot distort a questionnaire. Agreement across them is the strongest form of evidence available for a construct of this kind.

Physiological and experiential estimates were also close to one another (g = 1.18; 95% CI [0.76, 1.61] against g = 0.93; 95% CI [0.46, 1.40]; figure 3b). This matters for production: viewers do not merely register music subperceptually, they consciously experience the heightened engagement it produces. Muting the soundtrack of subtitled video reduced enjoyment (g = 0.69) and immersion (g = 0.48) in a within-subjects design with 161 participants (SZARKOWSKA et al., 2024), among the most precisely estimated effects in this review.

Figure 2 – The benefit profile of background music.



Caption: (a) Pooled random-effects estimates by outcome family; thick bars are 95% confidence intervals and thin lines 95% prediction intervals. (b) Individual attentional effects grouped by measurement modality, with horizontal bars showing modality means. Positive values of g indicate music improved the outcome. Families whose interval spans zero are drawn in grey.

Source: prepared by the author.

NO OUTCOME FAMILY SHOWS A RELIABLE COST

Eighteen of the 26 effects favoured music and 14 excluded zero. Neither of the other two families produced a statistically reliable effect in either direction: content memory pooled to g = −0.22 (95% CI [−0.56, 0.11]; p = 0.19) and exposure duration to g = 0.46 (95% CI [−0.21, 1.13]; p = 0.18).

The memory result deserves precision rather than reassurance. Its central tendency is negative, and the interval includes values that would represent a meaningful cost as well as values representing a benefit; under the fitted random-effects distribution, the probability that the true effect in a new study is trivial ($|d| < 0.20$) is only 0.34. What the data rule out is a reliable general memory penalty, not the possibility of a real one in specific configurations. Those configurations, as will be seen, are identifiable.

THE MEMORY COST IS A BOUNDARY CONDITION, NOT A GENERAL PROPERTY

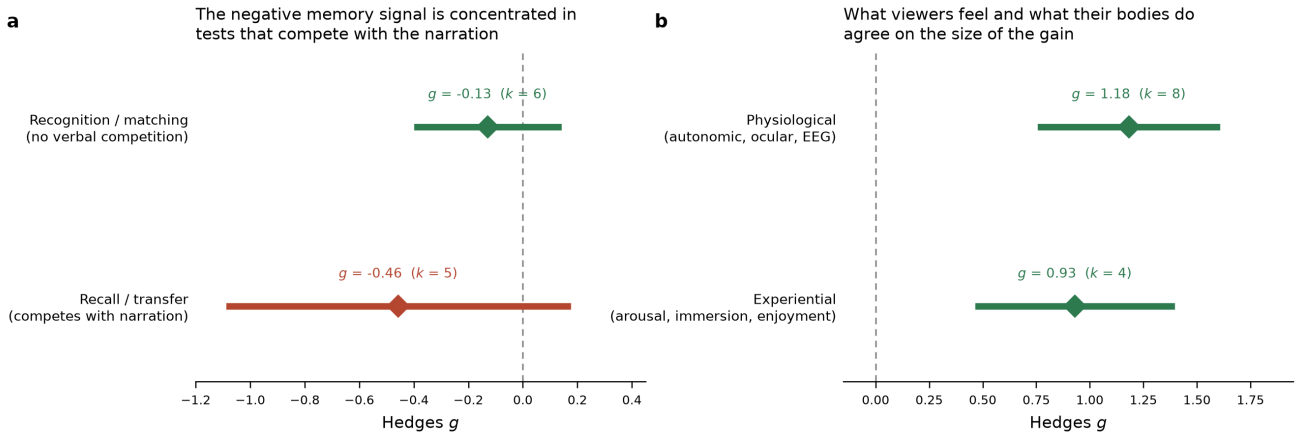
Splitting the memory family by test depth localises the negative signal precisely (figure 3a). Tests requiring effortful verbal reconstruction while narration is present — free recall and problem-solving transfer — pooled to g = −0.46 (k = 5). Tests that do not compete for the verbal channel — recognition and matching — pooled to g = −0.13 (95% CI [−0.40, 0.14]; k = 6), close to zero. The moderator contrast is significant (b = −0.44; p = 0.017).

Three findings independently confirm that concurrency with speech, rather than music itself, is the operative variable. Music placed in a pretraining segment, where it

did not compete with narration, improved retention ($F(1,88) = 8.01$; $p = 0.006$), while the same music inserted inside the narrated content produced no effect at all ($p = 0.53$) (PHAN, 2023). Music was detrimental for the easier of two instructional videos and null for the harder one, where the content itself already occupied attention (MEYERHOFF et al., 2022). And a high-powered replication of the television-news line of research found no recognition differences at all (KOPIEZ; PLATZ; WOLF, 2013).

A further caution against reading the memory literature as uniformly negative: in narrative film, music does not raise or lower accuracy so much as reshape *what* is encoded, through mood congruency (DAMJANOVIC; KAWALEC, 2021). For narrative and documentary work, that is a creative instrument rather than a cost.

Figure 3 – Where the observed effects come from.



Caption: (a) Memory effects split by whether the test competes with concurrent narration. (b) Attentional effects split by measurement type. Thick bars are 95% confidence intervals; *k* is the number of effect sizes. Subgroup estimates come from random-effects models fitted separately within each subgroup.

Source: prepared by the author.

MUSIC ALSO SHAPES AFFECTIVE INTERPRETATION

Beyond engagement, soundtrack choice systematically shifts how audiences interpret what they see. Music alters empathy toward on-screen characters (ANSANI et al., 2020) and anthropomorphic judgements about filmed animals (MANITZAS HILL et al., 2023). For documentary, advertising and narrative production, changing how an audience feels about the subject is frequently the objective, and these results establish that music does so reliably.

EXPOSURE DURATION: PROMISING BUT UNTESTED FOR VIDEO

Two of the three poolable duration effects favour music: stay in a virtual gallery tour averaged 74.7 s with music against 34.1 s without ($g = 1.16$; XU et al., 2023), and cafeteria meal duration was longer with sound ($g = 0.32$, converted from $F(1,244) = 4.67$; the source reports no condition means; MATHIESEN et al., 2022). Neither involves video, and the third effect, from a restaurant field experiment, is essentially null (MALCMAN et al., 2024). The pooled estimate is reported for completeness and is not interpreted: with $k = 3$ and $I^2 = 92\%$, it carries no evidential weight.

The honest statement of the position is that no retrievable experiment has measured video watch time under a music/no-music contrast. The one study reporting viewing duration under such a contrast (MORA VELASCO; HIRUMI; CHEN, 2021) is closed-access and reaches the literature only second-hand, reportedly finding no difference. The engagement benefit documented here is therefore established for how viewing *feels*, and remains an untested hypothesis for how long viewing *lasts*.

SMALL-STUDY EFFECTS

The memory funnel is symmetric (Egger-type $b = 1.16$; $p = 0.88$), with no studies imputed by trim-and-fill. The attention funnel is asymmetric ($b = 10.67$; $p < 0.0001$), with one study imputed and the estimate adjusted from $g = 1.04$ to 0.96 on study-aggregated data. With only six reports, this asymmetry plausibly reflects the genuine correlation between precision and effect magnitude in small-*n* laboratory physiology, but it should be read as indicating that $g = 1.06$ is an upper bound. The adjusted estimate remains large.

Figure 4 – Forest plot of the 26 effect sizes.

PRACTICAL IMPLICATIONS

For narrative, documentary and entertainment video, where no verbal learning objective competes for attention, the evidence supports the use of background music without qualification, and the affective-shaping results give the soundtrack a role beyond engagement.

For instructional video, the evidence supports music in segments that do not overlap narration: titles, transitions, pretraining and demonstration passages, as well as B-roll without voiceover. During narrated explanation of novel material, the recall cost is real and music should be reduced or removed. Difficulty moderates this effect: for content demanding enough to occupy attention on its own, the penalty was absent.

For advertising and marketing video, the engagement and affective results apply directly, and brand-recall objectives should be treated under the instructional rule when a spoken claim is the message.

One production variable deserves measurement rather than intuition. Music loudness relative to narration is almost never reported in this literature, although the pretraining result suggests it may be decisive. Producers can control it precisely; researchers have not yet told them where to set it.

LIMITATIONS

The pooled estimates rest on few reports, and the attention family draws 12 effects from only six reports, so the study-level variance component is estimated from little information. Within-subject conversions assume $r = 0.5$, which affects variance but not sign. Several attentional studies manipulate the soundtrack as a whole rather than isolated music, so part of the pooled engagement effect may reflect sound rather than music specifically; the test of this moderator was underpowered ($p = 0.635$) and cannot exclude that possibility. Funnel asymmetry in the attention family suggests the headline estimate is an upper bound. Music characterisation is poor across the field, which prevented the intended dose-response analysis. Coverage is anglophone and dominated by samples from Western, educated, industrialised, rich and democratic populations. Finally, the memory family's interval remains wide enough to accommodate a genuine cost, and the null result is presented as an absence of reliable evidence for a general penalty rather than as evidence of no effect.

THE EXPERIMENT THAT WOULD COMPLETE THE ARGUMENT

The engagement result creates a specific, testable prediction that no one has evaluated: if music increases attentional engagement and suppresses mind-wandering, viewers given the option to stop should watch for longer with music than without. One study design would settle it.

Design. Between-subjects, with three arms: no music, a low-arousal instrumental bed and a high-arousal instrumental bed. Participants are told they may stop watching at any time and move on. The option to stop must be real, unpenalised and incentive-neutral, since it is the entire point of the design.

Stimuli. Four to six videos of eight to twelve minutes, spanning instructional, documentary and narrative genres, each produced in all three audio conditions from a single visual master. Music must be isolated from dialogue and sound effects, and its level relative to speech measured in LUFS and reported. Existing stimuli run from 23 seconds to 5.5 minutes, far shorter than the horizon over which retention curves separate.

Primary endpoint. Proportion of the video watched before voluntary termination, analysed as time-to-event with completers right-censored; a Cox model stratified by video is the natural primary analysis. The retention curve, not a mean, is the object of interest.

Secondary endpoints. Recall and transfer scored relative to how much was actually watched; self-reported engagement and enjoyment; retrospective duration estimation, which tests the time-perception mechanism directly (APPELQVIST-DALTON et al., 2022); and continuous pupillometry where instrumentation permits, linking within-viewer arousal to the hazard of stopping.

Power. Detecting a hazard ratio of 0.75 for stopping, roughly a five-percentage-point difference in mean proportion viewed, requires approximately 380 events at 80% power and $\alpha = 0.05$, implying about 500 participants per arm. A platform field experiment randomising the music bed on real uploads would answer the same question with greater ecological validity and less control.

CONCLUSION

The evidence on background music in video is more favourable than its reputation. Music produces a large, consistent and cross-modally replicated increase in attentional engagement, raises immersion and enjoyment, and reliably shapes affective interpretation. Its one documented cost is specific and avoidable: instrumental music overlapping spoken narration impairs verbal recall of instructional content, an effect confined to tests that compete for the verbal channel and absent where no such competition exists.

Producers do not face a choice between engagement and comprehension; they face a placement decision that the evidence can now inform. Whether the engagement benefit extends to the viewer's decision to keep watching is the field's most consequential open question, and it is answerable with a single well-designed study.

DATA AVAILABILITY

The study-level dataset, with all extracted statistics, computed effect sizes and per-effect conversions, is available in the file study_level_effects.csv; the resolved reference list in references.csv; and the pooled model output in table1_pooled.csv. The files will be made available by the author upon request.

CONFLICT OF INTEREST DECLARATION

The author is the founder of Scora (https://scora.fm), a service that provides synchronised background music for video. The subject of this review — the effect of background music on viewer engagement — bears a direct commercial relationship to that activity, which constitutes a potential conflict of interest, declared here in full. No funding was received for this work, and Scora had no part in the design of the review, the selection of studies, the analysis of the data or the decision to publish.

So that the reader may weigh that conflict, the safeguards adopted are placed on record: the eligibility criteria, the three outcome families and the subgroup analyses were prespecified before pooling; the contrast of most direct commercial interest — exposure duration — is reported as not established for video, against the author's interest; and the complete study-level dataset is made available so that any reader may reproduce or contest the estimates reported here.

REFERENCES

ANSANI, A. et al. How soundtracks shape what we see: analyzing the influence of music on visual scenes through self-assessment, eye tracking, and pupillometry. *Frontiers in Psychology*, v.11, art. 2242, 2020. DOI: 10.3389/fpsyg.2020.02242

APPELQVIST-DALTON, M. et al. Time perception in film is modulated by sensory modality and arousal. *Attention, Perception, & Psychophysics*, v.84, n.3, p.926-942, 2022. DOI: 10.3758/s13414-022-02464-9

BENTE, G. et al. Building blocks of suspense: subjective and physiological effects of narrative content and film music. *Humanities and Social Sciences Communications*, v.9, n.1, art. 449, 2022. DOI: 10.1057/s41599-022-01461-5

BEZDEK, M.A.; WENZEL, W.G.; SCHUMACHER, E.H. The effect of visual and musical suspense on brain activation and memory during naturalistic viewing. *Biological Psychology*, v.129, p.73-81, 2017. DOI: 10.1016/j.biopsycho.2017.07.020

DAMJANOVIC, L.; KAWALEC, A. The role of music-induced emotions on recognition memory of filmed events. *Psychology of Music*, v.50, n.4, p.1136-1151, 2021. DOI: 10.1177/03057356211033344

HAMMERSCHMIDT, D.; WÖLLNER, C. The impact of music and stretched time on pupillary responses and eye movements in slow-motion film scenes. *Journal of Eye Movement Research*, v.11, n.2, 2018. DOI: 10.16910/jemr.11.2.10

KOPIEZ, R.; PLATZ, F.; WOLF, A. The overrated power of background music in television news magazines: a replication of Brosius' 1990 study. *Musicae Scientiae*, v. 17, n.3, p.309-331, 2013. DOI: 10.1177/1029864913489703

KWON, Y.-S.; LEE, J.; LEE, S. The impact of background music on film audience's attentional processes: electroencephalography alpha-rhythm and event-related potential analyses. *Frontiers in Psychology*, v.13, art. 933497, 2022. DOI: 10.3389/fpsyg.2022.933497

MALCMAN, M. et al. How does background music affect dining duration, tips and bill amounts in restaurants? A field experiment. *Behavioral Sciences*, v.14, n.12, art. 1188, 2024. DOI: 10.3390/bs14121188

MANITZAS HILL, H.M. et al. The influence of background music and narrative setting on anthropomorphic judgements of killer whale (Orcinus orca) emotional states and subsequent donation behavior. *PLOS ONE*, v.18, n.5, e0282075, 2023. DOI: 10.1371/journal.pone.0282075

MATHIESEN, S.L. et al. The sound of silence: presence and absence of sound affects meal duration and hedonic eating experience. *Appetite*, v.174, art. 106011, 2022. DOI: 10.1016/j.appet.2022.106011

MEYERHOFF, H.S. et al. Medical education videos as a tool for rehearsal: efficiency and the cases of background music and difficulty. *Instructional Science*, v.50, n.6, p. 879-901, 2022. DOI: 10.1007/s11251-022-09595-4

MORA VELASCO, E. de la; HIRUMI, A.; CHEN, B. Improving instructional videos with background music and sound effects: a design-based research approach. *Journal of Formative Design in Learning*, v.5, n.1, p.1-15, 2021. DOI: 10.1007/s41686-020-00052-4

MORENO, R.; MAYER, R.E. A coherence effect in multimedia learning: the case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, v.92, n.1, p.117-125, 2000. DOI: 10.1037/0022-0663.92.1.117

PHAN, T. Effects of background music in instructional videos on learners' retention. *The Journal of Educators Online*, v.20, n.3, 2023. DOI: 10.9743/jeo.2023.20.3.4

SZARKOWSKA, A. et al. Watching subtitled videos with the sound off affects viewers' comprehension, cognitive load, immersion, enjoyment, and gaze patterns: a mixed-methods eye-tracking study. *PLOS ONE*, v.19, n.10, e0306251, 2024. DOI: 10.1371/journal.pone.0306251

WÖLLNER, C.; HAMMERSCHMIDT, D.; ALBRECHT, H. Slow motion in films and video clips: music influences perceived duration and emotion, autonomic physiological activation and pupillary responses. *PLOS*

ONE, v.13, n.6, e0199161, 2018. DOI: 10.1371/journal.pone.0199161

XU, X. et al. Effect of music tempo on duration of stay in exhibition spaces. *Applied Acoustics*, v.207, art. 109353, 2023. DOI: 10.1016/j.apacoust.2023.109353

© Cultura & Saúde · ISSN 2595-7546 · culturaesaude.med.br · English edition