

# Dynamic Complexity in Audiovisual Aesthetics

Ana Clemente<sup>1, 2, 3, 4</sup>, Frances Board<sup>5</sup>, Marcus T. Pearce<sup>4, 6</sup>, and Guido Orgs<sup>5, 7</sup>

<sup>1</sup>Department of Cognition, Development and Educational Psychology, Institute of Neurosciences, University of Barcelona

<sup>2</sup>Cognition and Brain Plasticity Unit, Bellvitge Biomedical Research Institute, Barcelona, Spain

<sup>3</sup>Human Evolution and Cognition Research Group, University of the Balearic Islands

<sup>4</sup>School of Electronic Engineering & Computer Science, Queen Mary University of London

<sup>5</sup>Department of Psychology, Goldsmiths, University of London

<sup>6</sup>Centre for Music in the Brain, Department of Clinical Medicine, Aarhus University

<sup>7</sup>Institute of Cognitive Neuroscience, University College London

The appreciation of dance, film, and other temporal art forms relies on the continuous integration of auditory and visual streams. In this study, we investigate how bimodal audiovisual preferences arise from unimodal auditory and visual preferences. To this end, we created and validated the open-resource complexity in audiovisual aesthetics stimulus set (<https://osf.io/e5uh9/>), consisting of 120 short, dynamic, and abstract auditory, visual and audiovisual stimuli in which auditory and visual complexity corresponds to the number and variety of elements. In Experiment 1, 87 participants rated liking and perceived complexity for each stimulus, with visual, auditory, and audiovisual blocks fully randomized. In Experiment 2, 53 participants rated how much they liked each stimulus with the audiovisual block presented first to avoid potential bias arising from prior experience of unimodal stimuli and the simultaneous complexity judgements. Structural equation modeling and linear mixed-effects analysis show that liking for audiovisual stimuli can be explained by a weighted sum of liking for their auditory and visual components modulated by audio-visual congruence. Audiovisual preferences exhibit inverted-U-shaped relationships with auditory and visual complexity, the latter mediated by perceived complexity and modulated by congruence. Our findings provide a carefully controlled departure point for better understanding the role of prediction of sequential structure for the experience of dynamic audiovisual art forms such as dance or film.

**Keywords:** audiovisual, complexity, information content, liking, movement

**Supplemental materials:** <https://doi.org/10.1037/aca0000685.supp>


Most experiences in life are multimodal (Møller et al., 2021; Stein, 2012) and unfold in time (Spence & Squire, 2003; Stevenson & Wallace, 2013). This is especially true for the appreciation of temporal

arts such as film, dance, or theatre. These art forms involve integrating at least two—auditory and visual—sensory modalities. Film, media, and performance scholars have long argued that the appreciation of (e.g., liking for) these art forms relies on the congruence of their structural and semantic unimodal features (Chion, 1994; Cohen, 2013; Jordan, 2011; Tsay, 2013). However, research in audiovisual aesthetics using naturalistic stimuli like dance or film is difficult because their formal features (e.g., auditory and visual complexity) are not easily separated from their semantic features, including narrative or emotional expression through the human body. In the present study, we use carefully controlled nonrepresentational stimuli varying in auditory and visual complexity to address two critical questions: First, we test how liking for audiovisual stimuli relates to liking for their unimodal auditory and visual components. Second, we explore how liking for dynamic audiovisual stimuli relates to objective and subjective complexity measures, introducing new measures for dynamic visual complexity.

## Relationship Between Liking for Audiovisual Stimuli and for Their Unimodal Auditory and Visual Components

The Gestalt principle that the whole is greater than the sum of its parts refers to the perception of an object or phenomenon being different from the mere addition of its constituent elements (Köhler, 1971/1930; Wertheimer, 1938/1924). In the context of appreciation

Amy M. Belfi served as action editor.

Ana Clemente  <https://orcid.org/0000-0002-0460-6793>

This research was funded in whole, or in part, by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant 864420 - Neurolive). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Ana Clemente was supported by the Spanish Ministry of Universities, within the framework of the recovery, transformation, and resilience plan and funded by the European Union (NextGenerationEU) with the participation of the University of the Balearic Islands. Guido Orgs was supported by the European Research Council under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement 864420—Neurolive). All authors approved the final version of the article. The authors have no conflicts of interest. The data and materials are available at <https://doi.org/10.17605/OSF.IO/E5UH9>.

Correspondence concerning this article should be addressed to Ana Clemente, Department of Cognition, Development and Educational Psychology, Institute of Neurosciences, University of Barcelona, Passeig de la Vall d'Hebron, 171, 08035 Barcelona, Spain. Email: [ana.c.magan@gmail.com](mailto:ana.c.magan@gmail.com)

of audiovisual stimuli, it remains unclear how appreciation of bimodal (e.g., audiovisual) stimuli is related to appreciation of its unimodal components. We expect the Gestalt principle to apply, but there are several ways in which it could do so: First, liking for bimodal stimuli might reflect a weighted sum of liking for their unimodal components. Second, the unimodal influences may not be independent but interact. Third, combining auditory and visual components may produce new emergent properties influencing liking for the audiovisual whole.

Research on audiovisual integration provides clues as to how auditory and visual components combine to produce liking for audiovisual stimuli. Audiovisual integration is flexible and adaptive, varying with the relative relevance and reliability of the underlying perceptual components and expectations about the origin and causality of the signal (Meijer et al., 2019; Parise et al., 2012; Rohe & Noppeney, 2018). Spatial and temporal cues influence the relative weighting of visual and auditory information, resulting in the dominance (capture) of one kind of information over the other. Examples include visual capture—for example, McGurk (McGurk & MacDonald, 1976; Spence & Soto-Faraco, 2010, for a review) or ventriloquist effects (Alais & Burr, 2004)—and auditory capture—or temporal ventriloquism (Burr et al., 2009; Morein-Zamir et al., 2003). Research suggests preeminence of the auditory stream (auditory capture) when temporal properties are more salient and, conversely, preeminence of the visual stream (visual capture) when spatial properties are more salient. Integration is optimal (Holmes, 2007; Stanford et al., 2005; Stevenson & James, 2009) when the unimodal components share spatial (Meredith & Stein, 1986a, 1996) and temporal (Meredith et al., 1987; Miller & D'Esposito, 2005; Senkowski et al., 2007) sources—consistent with the common cause or unity assumption (see Chen & Spence, 2017 for a review)—and when they are similarly salient (Holmes, 2009; Kayser et al., 2005; Meredith & Stein, 1983, 1986b; Perrault et al., 2005). Therefore, an uneven weighting of liking for the unimodal components could point to attention directed to either unimodal component driving liking for the audiovisual composite.

Regarding emerging properties, we focus on congruence between the complexity of the auditory and visual components of audiovisual stimuli. In research on the perception of film music, Lipscomb and Kendall (1994; see also Lipscomb, 2005) proposed that attentional focus is maintained on the audiovisual composite rather than on either unimodal stream in isolation when semantic associations between auditory and visual streams are deemed appropriate based on previous experience and to the extent to which auditory and visual accent structures are consistent. Lipscomb (1999) found support for such structural consistency when using animations by Norman McLaren (corroborated with artificially controlled materials in Lipscomb, 2005) but not when using film extracts. This suggests that the effect may depend on the nature of the material and led the author to call for reliable, quantitative metrics of audiovisual complexity. These ideas have been substantially developed in Cohen's (2013) congruence-association model. According to it, semantic and structural relationships between music and visual components of film (also text, speech, and sound effects) are initially processed independently, allowing for structural congruence (as investigated here) to be assessed independently from semantic congruence before these components are integrated into a working narrative, which is in turn influenced by expectations derived from long-term memory.

## Relationship Between Complexity and Liking for Audiovisual Stimuli

Stimulus complexity influences the experience of both auditory and visual stimuli (Berlyne, 1970, 1971; Chmiel & Schubert, 2017; Nadal et al., 2010). It affects recognition memory (Halpern & Bartlett, 2010), learning (Flagg et al., 1976), physiological arousal (Potter & Choi, 2006), and attention, with auditory dominating visual complexity for attention maintenance (Alwitt et al., 1980; Wartella & Ettema, 1974). Moreover, complexity appears to influence appreciation across cultures, albeit with some elements of cross-cultural variation (Che et al., 2018). Liking has typically been reported to be maximal for stimuli with intermediate complexity (Berlyne, 1970, 1971; Berlyne & Boudewijns, 1971). However, many empirical results deviate from this general trend due to different definitions of complexity, experimental manipulations—including stimuli and measures (Che et al., 2018; Marin et al., 2016; Martindale et al., 1990; Nadal et al., 2010)—and analytical approaches—for example, whether considering individual or group-averaged responses (Clemente, 2022; Güçlütürk et al., 2016; Marin & Leder, 2018). Regarding dynamic auditory stimuli like music, genuine evidence for negative quadratic relationships between musical complexity and liking constitutes only a minority (26.3%) of the results reviewed by Chmiel and Schubert (2017). It is, therefore, unclear whether complexity and liking for dynamic audiovisual stimuli will display a linear or nonlinear (i.e., quadratic) relationship.

The influence of objective, feature-based complexity might be mediated by perceived complexity (e.g., Berlyne, 1971). However, most literature focuses on direct relationships between stimulus properties or their perceptual representations and liking, with scarce counterexamples (e.g., Clemente et al., 2023). Taking subjective ratings of liking and complexity for each stimulus enables a direct and systematic examination of whether perceptual representations of complexity (subjective complexity) mediate the impact of objective complexity on liking. Our study aims to delineate the relative influence of objective and subjective complexity for dynamic audiovisual stimuli.

## Measures of Auditory, Visual, and Audiovisual Complexity

Whereas research on complexity and liking for dynamic auditory or static visual stimuli is relatively widespread, liking for dynamic audiovisual complexity has yet to receive comparable attention. One of the main difficulties in operationalizing audiovisual complexity is the need to identify a measure of complexity that can be applied across both auditory and visual streams.

In music and other sound streams, complexity can vary as a result of the number of tones in a sequence (Mindus, 1968), chord structure (Berlyne et al., 1967), rhythm and syncopation (Heyduk, 1975), structural change (Mauch & Levy, 2011), variety in pitch, note durations, loudness, and timbre (Berlyne & Boudewijns, 1971). The expectancy-based model (Eerola & North, 2000; Eerola et al., 2006), or expectancy-violation model (EV; Eerola, 2016), and the MUSical STimulus complexity model (Clemente et al., 2020) consist of composite measures of weighted structural features that have also been found prominent in the visual modality (Nadal et al., 2010): number (event density) and variety and organization of elements (computed as different forms of entropy in the MUSical STimulus complexity and pitch proximity, tonal ambiguity, and rhythmic variation in the optimal EV models).

While complexity measures are relatively well defined for static images (e.g., Fernandez-Lozano et al., 2019; Machado et al., 2015), attempts to develop a formal measure of dynamic visual complexity are scarce. Existing tools are highly stimulus-specific and need to be sufficiently validated. For instance, Watt and Welch (1982) measured the visual complexity of children's television programs. In an alternative approach, Kearns and O'Connor (2004) applied Shannon entropy to calculate moving-image complexity. Finally, Orlandi et al. (2020) assessed the dynamic visual complexity of dance movements as entropy characterizations in the context of dance aesthetics. They showed that viewers prefer movements with variable yet predictable changes in speed and acceleration. However, none of these studies address how the complexity of dynamic visual and auditory components might interact.

One way to overcome the modality-specificity of existing complexity measures is to apply information-based theories and models (Kesner, 2014; Koelsch et al., 2019; Van de Cruys et al., 2017; Van de Cruys & Wagemans, 2011). However, to our knowledge, complexity measures based on information-based models of dynamic visual or audiovisual stimuli had yet to be available. The Information Dynamics Of Music (IDyOM; Pearce, 2005, 2018) is a well-established framework for investigating the impact of information-theoretic properties on the perception and appreciation of music (Clemente et al., 2020; Sauvé & Pearce, 2019). It is a system for constructing multiple-viewpoint, variable-order Markov models for predictive modeling of probabilistic structure in symbolic, sequential auditory domains like music. IDyOM acquires knowledge about a domain through statistical learning and generates conditional probability distributions representing the estimated likelihood of each event in a sequence, given the preceding context and incremental training on the current stimulus.<sup>1</sup> From the conditional probability estimates for each event, IDyOM computes Shannon entropy, which reflects the prospective predictive uncertainty of the model's probabilistic prediction given the context, and information content (IC), which reflects the contextual unpredictability of the event that actually follows from the model's perspective. Although no information-based complexity model existed for dynamic visual or audiovisual stimuli, IDyOM's architecture is flexible enough to enable the computation of information-theoretic measures of any sequence of discrete events. Here, we used IDyOM to quantify auditory complexity and, for the first time, dynamic visual complexity, providing comparable and generalizable measures of dynamic auditory and visual complexity.

## Aim and Hypotheses

Our overarching aim was to investigate liking for dynamic audiovisual displays in relation to the complexity of their auditory and visual components. Specifically, our first goal was to test the hypothesis that liking for the audiovisual whole is greater than the plain sum of liking for the unimodal components. Our second goal was to unveil the structure of relationships between stimulus complexity, perceived complexity, and liking for dynamically time-varying audiovisual stimuli.

To this end, we first created a novel stimulus set of dynamic unimodal and bimodal stimuli varying parametrically in auditory, visual, and audiovisual complexity, respectively. Then, we ran two experiments: Experiment 1 investigated links between

objective and subjective complexity measures and how they relate to liking. Experiment 2 replicated Experiment 1 in an in-person setting and established a direct link between objective complexity and liking, eliminating two potential confounds: First, bimodal stimulus blocks were always presented before unimodal blocks to avoid priming of bimodal preferences by the prior rating of unimodal preferences. Second, participants rated only liking to prevent influence of concurrently rating perceived complexity on liking ratings.

To address our two main goals, we ran two analyses: First, we tested the hypothesis conforming to the Gestalt principle that the whole is greater than the (plain) sum of its parts (Köhler, 1971/1930; Wertheimer, 1938/1924; Experiments 1 and 2) and examined the nature of the relationship between liking for audiovisual stimuli and liking for the unimodal components. We expected that liking for the auditory and visual components would differently influence liking for the audiovisual composite and a moderating role of audiovisual congruence. Second, we tested the hypothesis of a mediating role of perceived complexity on the impact of stimulus complexity on liking (Experiment 1). In addition, we expected negative quadratic effects of auditory and visual complexity on liking.

## Method

### Participants

In both experiments, native English speakers from the general population were unaware of the study's purpose and reported normal or corrected-to-normal vision and hearing and no cognitive impairments. Ethical approval was granted by the Psychology Department at Goldsmiths, University of London.

Previous research using mixed-effects models with random effects per participant and stimulus consistently employed sample sizes of around 40 participants (e.g., Clemente et al., 2021, 2022; Corradi et al., 2020). Following recommendations for online studies (Sauter et al., 2020; Stewart et al., 2017), we doubled this sample size. According to Judd et al.'s (2017) power calculator ([https://jakewestfall.shinyapps.io/two\\_factor\\_power/](https://jakewestfall.shinyapps.io/two_factor_power/)), both experiments—in which all participants rated 120 stimuli on each scale—would have a power of 1 for a moderate effect size of 0.5, which was expected by default given the lack of previous research with our experimental manipulations. In Experiment 1, 90 participants were recruited through Prolific (<https://www.prolific.co/>) with a minimum approval rate of 80% and compensated for participation following Prolific recommendations. Three participants with missing or invalid data were consequently excluded from the analyses. In Experiment 2, 53 UK residents took part and were compensated with £5.

We used the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014) general factor scale to characterize the sample emulating the distribution in the general population. The demographic data distributions in the final sample (Table 1)

<sup>1</sup> In the present research, IDyOM was configured to use a short-term model (STM) only, though it can also be configured to use a separate model given training prior to the stimulus (the long-term model or LTM), often from a large body of stimuli representing the outcome of long-term schematic statistical learning. The STM and LTM can also be combined in both configurations.

**Table 1**  
*Demographic Data*

Experiment	<i>N</i>	Age (range, <i>M</i> , <i>SD</i> )	Gender (w, m, other)	Gold-MSI (range, <i>M</i> , <i>SD</i> )
1	87	19–68, 34.15, 11.88	45, 40, 2	48–99, 70.24, 9.96
2	53	20–61, 29.64, 8.53	29, 24, 0	35–120, 77.51, 21.77

*Note.* Gold-MSI refers to the score in the general factor scale. w = women; m = men; Gold-MSI = Goldsmiths Musical Sophistication Index.

approach those in the general population (age median = 40.50, 50.57% women, 49.43% men, United Kingdom, 2022<sup>2</sup>).

## Stimuli

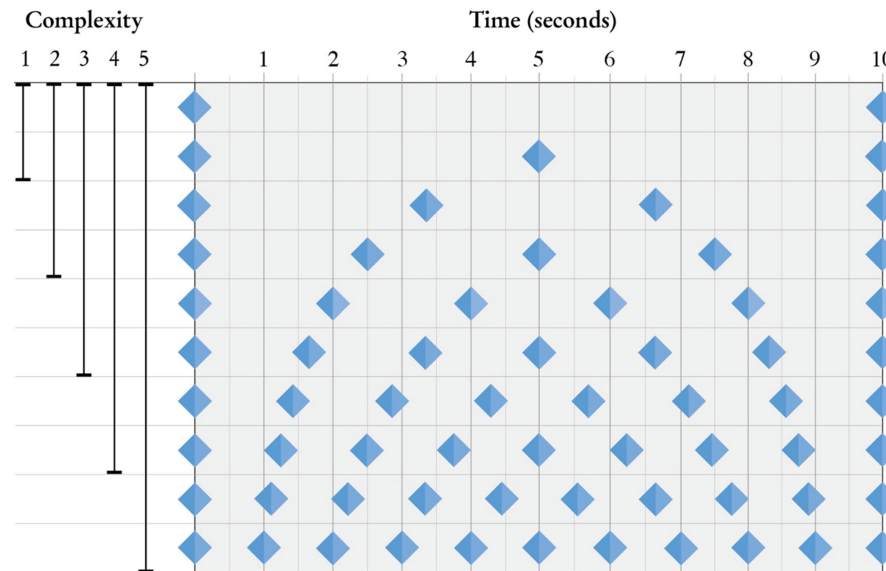
A novel set of 120 nonrepresentational 10-s stimuli was created: the complexity in audiovisual aesthetics (CAVA) stimulus set consists of auditory, visual, and audiovisual sequences varying in complexity. Complexity was manipulated by systematically increasing the number (event density) and variety of elements (modality-specific variability between events; Nadal et al., 2010) in the auditory and visual modalities separately, as described below. The CAVA set is publicly available as an open resource for research at <https://osf.io/e5uh9/> in MOV and MP4 formats.

The visual stimuli (V) consist of white horizontally moving vertical lines on a black background, inspired by McLaren and Lambart's abstract animation *Lignes Verticales*. The number of lines (event

density) and velocities (modality-specific variability) determine feature-based complexity in the visual modality. Five feature-based complexity levels were defined. The simplest visual stimulus contains two moving lines: The first line completes a horizontal journey from one side to the other for the duration of the video (0.1 Hz). The second line makes a return trip horizontally across the screen, thus traveling twice as fast (0.2 Hz). The most complex sequence contains 10 moving lines, the fastest traveling 10 times (1 Hz) across the screen. Therefore, each feature-based complexity level increases the number of lines by two, resulting in two, four, six, eight, or ten lines for each level (1–5). In addition, each level is presented in two movement directions: In left–right, lines move from left to right, whereas in right–left, they move from right to left. This results in 10 visual stimuli (Figure 1) created using Adobe After Effects CC2018.

<sup>2</sup> Source: <https://www.statista.com>.

**Figure 1**  
*Visual Sequence Structure*



*Note.* Each row represents a vertical line. Stimuli of a given complexity level include the corresponding vertical lines indicated on the left of the figure. Diamonds on the far left indicate a line leaving one side of the display at the beginning of the stimulus, while diamonds on the far right indicate a line arriving at the same side of the display at the end of the stimulus. Diamonds at intermediate time points indicate a line arriving at one end of the display and leaving in the opposite direction. For example, the first row indicates a vertical line that started on one side of the visual display and arrived at the other side 10 s later, while the second row indicates a vertical line that starts on one side of the visual display, arrives at the other side 5 s later, and returns to the original side after 10 s. A stimulus with a complexity level of 1 comprises both these lines presented simultaneously. See the online article for the color version of this figure.



The auditory stimuli (A) were designed to replicate the structure of the visual stimuli mapped into the auditory domain. They consist of 0.35-s sine tones belonging to the A-major triad in the range A3–A6. Each stimulus comprises between two and ten pitches, each of which sounds repeatedly with a specific period. Thus, the number of sounds (event density) and pitches (modality-specific variability) constitute feature-based complexity in the auditory modality. Five auditory feature-based complexity levels (1–5) were defined: The simplest sequence comprises two pitches and five events. The most complex sequence comprises 10 pitches distributed across 65 events. Each complexity level is presented in two pitch directions: In low–high, the pitches proceed from low to high, whereas in high–low, they proceed from high to low. This results in 10 auditory stimuli (Figure 2) created using Adobe Audition CC2018.

The 10 auditory and 10 visual sequences were orthogonally combined to produce 100 audiovisual stimuli in Adobe Premier Pro CC 2018. Therefore, each audiovisual stimulus is characterized by a specific degree of congruence between auditory and visual complexity: Minimal when the feature-based complexity levels are most dissimilar and maximal when these levels are the same in both sensory modalities. Furthermore, because the auditory and visual stimuli followed the same principles of rhythmic construction (cf., Figures 1 and 2), audiovisual congruence entails synchrony, in the sense that each visual event (a line reaching the frame of the screen) occurs at the same time as the corresponding auditory event (a sine tone) and vice versa.

The stimuli conform to, or leverage, the three principles of optimal multimodal integration (Chen & Spence, 2017; Holmes,

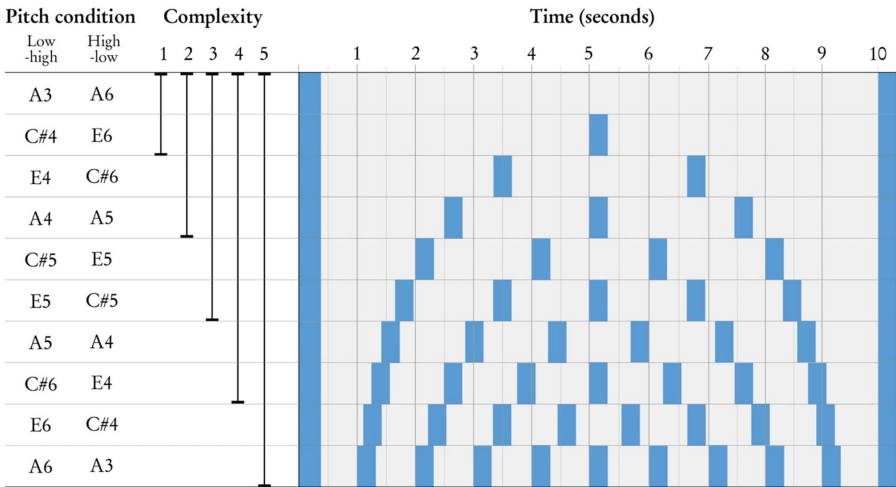
2007; Stanford et al., 2005; Stevenson & James, 2009): First, they originate from the same source, the laptop, complying with the spatial rule (Meredith & Stein, 1986a, 1996). Second, they involve perfect synchrony between the beginning and end of the auditory and visual streams when congruence is high, accommodating the temporal rule (Meredith et al., 1987; Miller & D’Esposito, 2005; Senkowski et al., 2007). Third, by combining visual and auditory components with greater or lesser degrees of congruence in complexity, they vary in compatibility with the inverse-effectiveness rule (Holmes, 2009; Kayser et al., 2005; Meredith & Stein, 1983, 1986b; Perrault et al., 2005).

Measures

We manipulated feature-based complexity as the number (event density) and variety of elements (modality-specific variability between events; Nadal et al., 2010) in each sensory modality over time. This was achieved by varying event density and event variability while orthogonally contrasting presentation modes (H–L vs. L–H pitch and R–L vs. L–R line movement direction). However, this feature-based operationalization of complexity is limited because it is, by definition, specific to the features manipulated—which may not be present in other stimuli. Furthermore, neither event density nor event variability reflects the complexity of the sequential configuration of the stimulus.

Therefore, we measured information-based complexity by developing IDyOM (Pearce, 2018) models of the auditory and visual streams. Information-based complexity measures are not

Figure 2  
Auditory Sequence Structure



Note. Each horizontal row is dedicated to a particular pitch sounding with a particular periodicity, as indicated by the blue (dark gray) rectangles, each of which represents the timing of a sine tone with the corresponding pitch class (A, C#, E) and octave designation (3–6). A stimulus of a given complexity level includes the corresponding pitches (rows) indicated on the left of the figure. Two versions of the stimuli are provided: the high–low condition in which higher pitches are more frequently repeated, and the low–high condition in which lower pitches are more frequently repeated—given the symmetric (mirror-like) structure of the stimuli. For example, in the low–high condition, the first row indicates an A3 that sounds at 0 and 10 s, while the second row indicates a C#4 that sounds at 0, 5, and 10 s. A stimulus with complexity Level 1 includes the pitches indicated on the first two rows of the figure presented simultaneously. See the online article for the color version of this figure.

only sensitive to the number and variety of elements but also the sequential structure of the stimuli. Therefore, they are more generalizable to other kinds of discrete sequential stimuli. In this context, the auditory stimuli can be represented as sequences of sound events differing in pitch, while the visual stimuli can be represented as sequences of visual events (line arrivals at one side of the display or the other) differing in movement velocity. Sequences of auditory and visual events with the same complexity level have identical timing.

Among the IDyOM information-based measures available, IC was preferred over entropy on theoretical grounds—it reflects the unpredictability of actual events rather than the model's prospective predictive uncertainty given the context—and on practical grounds—it has been used successfully as the main information-theoretic complexity measure in previous research (Cheung et al., 2019; Clemente et al., 2020; Gold et al., 2019; Sauvé & Pearce, 2019), which affords comparison across studies. The feature-based and information-based measures used in this research accompany the CAVA stimulus set.

To assess congruence, we subtracted from four the absolute value of the difference between the auditory and visual complexity levels and divided by four to normalize in the range from 0 (*least congruence*) to 1 (*perfect congruence*). Therefore, although the scale is continuous and comparable to the complexity measures, its levels are categorical and were treated as such in the analyses (Table 2).

The CAVA set in MOV and MP4 formats and the corresponding feature-based and information-based measures are publicly available as open resources for research at <https://osf.io/e5uh9/>. Detailed validation analysis is provided as online supplemental materials. The results show that our systematic manipulation captured participants' abilities to perceive and appreciate different levels of auditory complexity and show a nearly perfect correlation between information-based and feature-based complexity measures. Importantly, all three measures of complexity (feature-based, information-based, and perceived) predicted liking for audiovisual stimuli, and neither line movement direction (L–R or R–L) nor pitch movement direction (H–L or L–H) predicted liking.

## Procedure

In Experiment 1, participants completed the online experiment created and hosted using Gorilla Experiment Builder (<https://www.gorilla.sc>; Anwyl-Irvine et al., 2020). After providing informed consent, prospective participants performed a browser soundcheck, a built-in volume calibration and a headphone

**Table 2**  
*Audiovisual Congruence Levels*

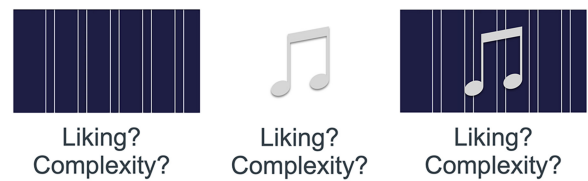
Levels	–	Auditory complexity				+
Visual complexity	0	.25	.50	.75	1	
	.25	.50	.75	1	.75	.75
	.50	.75	1	.75	.50	.50
	.75	1	.75	.50	.25	.25
+	1	.75	.50	.25	0	

*Note.* Auditory complexity increases from left (*very simple*) to right (*very complex*). Visual complexity increases from down (*very simple*) to top (*very complex*). Numbers denote audiovisual complexity congruence from 0 to 1.

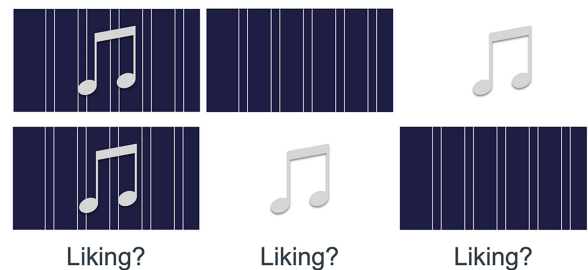
**Figure 3**

*Design of E1 and E2*

## E1 Counterbalanced block order



## E2 AV blocks before A&V blocks



*Note.* E1 = Experiment 1; E2 = Experiment 2; AV = audiovisual; A = auditory; V = visual. See the online article for the color version of this figure.

screening (Milne et al., 2021). Block (auditory, visual, and audiovisual) order was counterbalanced across participants, and stimulus order was individually randomized for each participant (Figure 3). A 1-s fixation cross preceded each visual and audiovisual stimulus. Immediately after stimulus presentation, participants rated their liking for each stimulus on a 7-point Likert scale anchored by *dislike a lot* (1) and *like a lot* (7) and how complex they perceived the stimulus to be on a 7-point Likert scale anchored by *very simple* (1) and *very complex* (7). They were instructed to make judgments quickly and intuitively and encouraged to use the full range of each rating scale. Following the rating blocks, the participants completed the Musical Sophistication Index general factor scale (Gold-MSI; Müllensiefen et al., 2014), an abridged version incorporating aspects from all five subscales. Finally, participants provided basic, customary demographic data (age and gender). Participants were allowed to take breaks between blocks but were required to complete the experiment in a single session, which took about 40 min.

In Experiment 2, participants were tested individually on site. The experiment was conducted on a 15-inch MacBook Pro using PsychoPy2 Version 1.85.4 (Peirce et al., 2019). All participants sat approximately 50 cm from the experimenter's laptop, were undisturbed during the experimental session, and listened to the audio and audiovisual stimuli through ATH-R70x headphones with the volume set at a comfortable level. The stimuli, questionnaires, measures, and experimental paradigm were identical to Experiment 1 except for two differences: First, we only collected liking ratings to prevent any influence of perceived complexity on liking. Second, the bimodal block always preceded the unimodal blocks to prevent a direct influence of unimodal on bimodal liking ratings. The order of auditory and visual blocks was counterbalanced across participants (Figure 1). Therefore, Experiment 2 was conducted in a more controlled setting, restricted to liking

ratings to avoid any potential confounding effects of taking simultaneous complexity ratings and prevented the experience of the audiovisual stimuli from being biased by prior experience of the unimodal stimuli.

## Data Analysis

### *Relationship Between Liking for Audiovisual Stimuli and for Their Unimodal Auditory and Visual Components*

We analyze the data in Experiments 1 and 2 separately using linear mixed-effects modeling (Hox et al., 2010; Snijders & Bosker, 2012), following Barr et al.'s (2013) recommendation to model the maximal random-effects structure justified by the experimental design. We compare four models: Model 1 reflects bimodal audiovisual preferences as the sum of unimodal auditory and visual preferences. Model 2 retains the summative analysis but allows for weighted contributions of auditory and visual preferences. Model 3 adds congruence to the weighted sum without interactions. Model 4 introduces interactions between congruence and liking for each component. Therefore, Models 1–2 test for simple or weighted additive linear effects, whereas Models 3–4 test for the contribution of an emerging property, congruence, in addition or interaction with such additive effects.

All models include intercepts and slopes per participant and intercepts per stimulus as random effects to account for the variability within and between participants and stimuli. To examine the relevance of such variability, we test whether removing fixed and random effects from each model in a stepwise model reduction improves the model fit through likelihood-ratio tests. Otherwise, we prefer the saturated model. For statistically significant differences ( $p < .05$ ), lower Akaike information criterion (AIC) and Bayesian information criterion (BIC) indicate a better fit to the data of one model over another. In cases of conflicting criteria, Vrieze (2012) and Yang (2005) recommend prioritizing AIC over BIC. Finally, we compare the best-fitting Models 1–4 following the same procedure.

All analyses are performed within the R environment for statistical computing, R Version 4.2.3 (R Core Team, 2023). For the mixed-effects models, we use the `lmer` function in the “lme4” package (Bates et al., 2015) and the “lmerTest” package (Kuznetsova et al., 2017) to estimate the  $p$  values for the  $t$  tests based on the Satterthwaite approximation for degrees of freedom, which produces acceptable Type-1 error rates (Luke, 2017). We report the model  $r^2$  for fixed effects only (marginal) and fixed plus random effects (conditional) regarding each response variable in the preferred model in each Experiment and interpret them according to Chin (1998). Effect sizes are calculated using the `effectsize` function in the “effectsize” package (Ben-Shachar et al., 2020) and are interpreted following Gignac and Szodorai's (2016) recommendations.

### *Relationship Between Complexity and Liking for Audiovisual Stimuli*

We analyze data in Experiment 1 using structural equation modeling (SEM) to test for a mediating effect of perceived complexity on the relationship between the stimulus properties and liking in the context of audiovisual stimuli varying in complexity (IC). Following the rationale above, the internal models making up the

SEM involve random effects per participant and stimulus, and the variables are not assumed to be normally distributed. Therefore, we apply piecewise SEM (or confirmatory path analysis; Shipley, 2009), which expands upon traditional SEM by introducing a flexible mathematical framework that can incorporate a variety of model structures, distributions, and assumptions, including interactions, non-Gaussian responses, random effects, hierarchical models, and alternate correlation structures.

We compare models with linear and quadratic terms to test for the inverted-U-shaped relationship between complexity and liking (Berlyne, 1971; Chmiel & Schubert, 2017; Nadal et al., 2010). Then, a stepwise model reduction is conducted for each configuration, starting with the following fixed-effects structure in R syntax—while the random effects have the form (1|stimulus) + ([fixed]|participant)—:

```
perceivedComplexity ~ auditoryIC + visualIC +
  auditoryIC:congruence + visualIC:congruence
liking ~ perceivedComplexity + auditoryIC +
  visualIC + congruence + auditoryIC:congruence +
  visualIC:congruence
```

Each SEM is adjusted according to three parameters: unsaturated model (nonsignificant paths to test the SEM fit:  $df > 0$ ), global goodness of fit (sufficiently low  $C: p > .05$ ), and absence of missing paths (tests of directed separation:  $p > .05$ ). We implement the SEM analysis using the `psem` and `plot` functions in the “piecewiseSEM” package (Lefcheck, 2016), Version 2.3.0. The `psem` output includes unstandardized and standardized estimates for each predictor (allowing comparisons within and between models), statistical significance, and coefficients of determination reported and interpreted as in the previous section.

## Results

### *Relationship Between Liking for Audiovisual Stimuli and for Their Unimodal Auditory and Visual Components*

Removing effects (whether predictors, interactions or random effects) significantly worsens the model fit (all  $ps < .01$ ). Consistently across experiments, the model of liking as a function of aggregated liking for the unimodal stimuli is outperformed by the model allowing for different weightings. The model fit is further improved when introducing congruence and is finally surpassed by the model considering interactions between congruence and liking for each unimodal component (Table 3 and Figure 4).

Across experiments, the best-fitting model reveals that liking for audiovisual stimuli is positively associated with liking for the corresponding unimodal stimuli. Audiovisual congruence shapes how much audiovisual liking is explained by liking for the unimodal components, as it boosts the effect of liking for auditory stimuli in Experiment 1 and that of liking for visual stimuli in Experiment 2 (Table 4).

### *Relationship Between Complexity and Liking for Audiovisual Stimuli*

SEM analyses are conducted to examine whether perceptual ratings of complexity mediate the relationship between stimulus

**Table 3***Models of Liking for Audiovisual Stimuli as a Function of Liking for Auditory and Visual Stimuli and Audiovisual Congruence*

E	M	Model configuration of fixed effects in R syntax	AIC	BIC	$\chi^2$	df	p
1	1	I (a.liking+v.liking)	26,538	26,587			
	2	a.liking+v.liking	25,985	26,063	560.52	4	<.01
	3	a.liking+v.liking+c	25,866	25,979	129.59	5	<.01
	4	<b>a.liking+v.liking+c+a.liking:c+v.liking:c</b>	<b>25,852</b>	<b>25,979</b>	<b>17.49</b>	<b>2</b>	<b>&lt;.01</b>
2	1	I (a.liking+v.liking)	18,934	18,980			
	2	a.liking+v.liking	18,684	18,756	257.98	4	<.01
	3	a.liking+v.liking+c	18,568	18,673	126.09	5	<.01
	4	<b>a.liking+v.liking+c+a.liking:c+v.liking:c</b>	<b>18,559</b>	<b>18,677</b>	<b>12.74</b>	<b>2</b>	<b>&lt;.01</b>

*Note.* ANOVA mixed-model likelihood ratio tests of different configurations of models of liking for the audiovisual stimuli as a function of liking for their auditory (a.liking) and visual (v.liking) streams and congruence in each experiment. The configurations represent the fixed effects of the model equation in R syntax. AIC and BIC for each model are reported together with the statistics of the multiple comparisons in each Experiment:  $\chi^2$ , *df*, and *p* value. Preferred models are highlighted in bold. E = experiment; M = model; AIC = Akaike information criteria; BIC = Bayesian information criteria; c = congruence; ANOVA = analysis of variance.

complexity measures and liking. The models with linear predictors fail to converge. The model with the following structure in R syntax yields the best fit to the data (AIC = 53,284.34,  $\chi^2 = 2.20$ ,  $p = .53$ ,  $df = 3$ ,  $C = 10.95$ ,  $p = .09$ ,  $df = 6$ ):

```
perceivedComplexity ~ visualIC2 + auditoryIC2:
  congruence + visualIC2:congruence
liking ~ perceivedComplexity + auditoryIC2 +
  congruence + visualIC2:congruence
```

Perceived complexity follows a negative quadratic relationship with visual IC. Liking is positively associated with perceived complexity, follows a negative quadratic relationship with auditory IC, increases with congruence, and follows a negative quadratic relationship with visual IC moderated by congruence. In other words, auditory IC and congruence directly affect liking, whereas the effect of visual complexity on liking is partially mediated by perceived complexity and moderated by congruence (Table 5 and Figure 5).

## Discussion

We created, validated, and made publicly available a new stimulus set and generalizable measures of auditory and visual dynamic complexity. These tools allowed us to investigate two important aspects of audiovisual aesthetics. First, we tested the Gestalt principle that the whole is greater than the sum of its parts in the context of liking for audiovisual stimuli varying in auditory and visual complexity. Second, we examined the relationship between stimulus properties, their perceptual representations, and liking for audiovisual stimuli varying in auditory and visual complexity.

The results of the validation analysis (see online supplemental materials) suggest that the CAVA set has good construct validity in systematically manipulating perceived auditory, visual, and audiovisual complexity, and support the use of information-based measures over feature-based measures. Extending IDyOM model to dynamic visual information provides a meaningful and generalizable quantification of information-based complexity to study the structural complexity of audiovisual stimuli.

The results of the model comparisons addressing our first research question suggest that liking for audiovisual stimuli can be explained as a weighted additive function of liking for both unimodal

components, distinctly moderated by audiovisual congruence: It enhanced the effect of liking for the auditory component when perceived complexity was overtly rated (Experiment 1), whereas it enhanced the effect of liking for the visual component when complexity was not explicitly rated (Experiment 2).

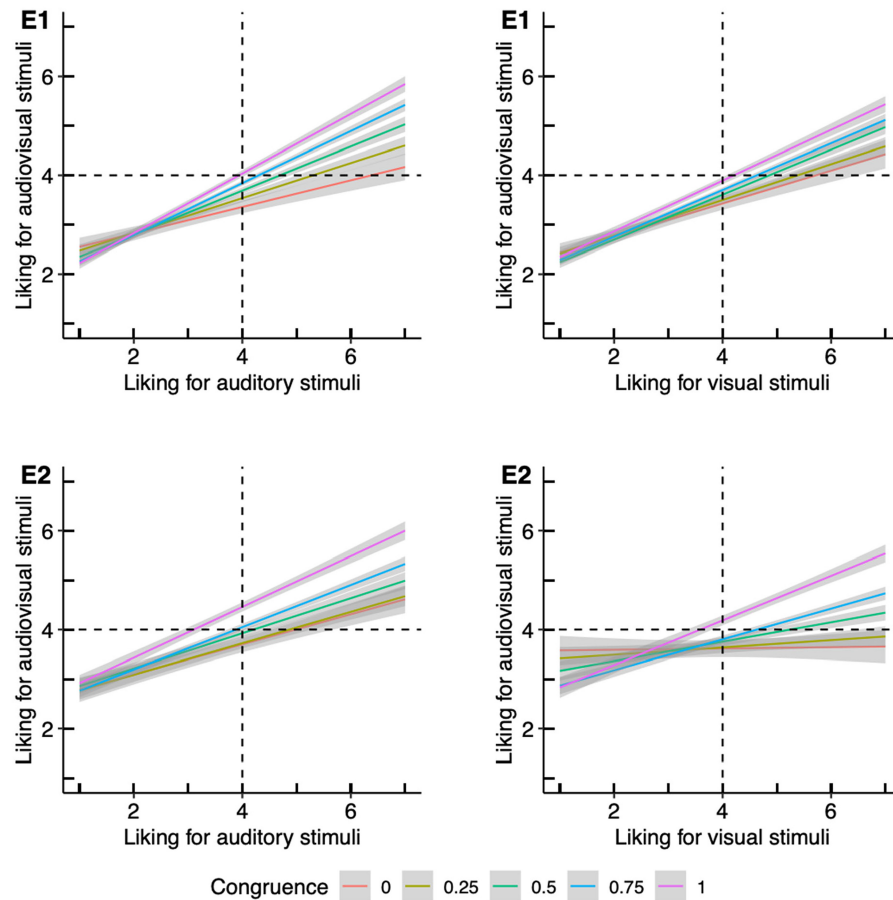
Several implications are worth considering here: First, the moderating role of congruence supports the Gestalt principle that the whole is greater than the (plain) sum of its parts (Köhler, 1971/1930; Wertheimer, 1938/1924) in that the contributions of each component are not identical and that an emergent property (congruence between auditory and visual complexity) exerts a moderating influence on such contributions sensitive to the context of evaluation. This entails shifts in the relative importance of each unimodal component, which aligns with capture effects (Spence & Squire, 2003; Stevenson & Wallace, 2013). Second, people do not usually experience auditory and visual components separately before experiencing the audiovisual whole, and smaller effects of liking for the unimodal components in Experiment 1 than in Experiment 2 suggest some mitigation by concurrent complexity ratings. In this sense, Experiment 2 more closely resembles liking for audiovisual stimuli in the real world. In this case, the results indicate a closer link between congruence and visual complexity or a greater malleability of the visual than the auditory component. Third, the differences between the impact of auditory and visual complexity on perceived complexity and liking and their relations with congruence align with distinct processing of auditory and visual streams (Griffin et al., 2002; Macaluso et al., 2004; Peretz & Coltheart, 2003; Peretz & Zatorre, 2005). Fourth, the susceptibility to context (i.e., presentation order and concurrence of evaluative judgements) concurs with current knowledge on sensory valuation, fruit from the interplay of subject, object, and context (Skov, 2019).

The results of the SEM addressing our second research question also suggest distinct effects of auditory and visual complexity and a moderating role of audiovisual congruence on visual complexity: Liking for audiovisual stimuli increased with perceived complexity—which in turn increased for intermediate visual complexity—intermediate auditory complexity, and congruence between auditory and visual complexity, which was enhanced for intermediate visual complexity. Thus, the results support an inverted-U-shaped relationship between auditory complexity and liking for audiovisual stimuli (Berlyne, 1970, 1971), in line with predictive processing theories of appreciation (e.g., Van de Cruys et al., 2022).



**Figure 4**

*Liking for Audiovisual Stimuli as a Function of Liking for Their Auditory (Left) and Visual (Right) Components in E1 (Top) and E2 (Bottom) for Each Congruence Level*



*Note.* Although the direction of the effects of congruence are consistent across experiments and unimodal stimuli, they only reach significance for auditory stimuli in Experiment 1 and for visual stimuli in Experiment 2. Dashed black lines represent neutral ratings in the Likert scale. Shaded areas represent 95% CI. E1 = Experiment 1; E2 = Experiment 2; CI = confidence interval. See the online article for the color version of this figure.

Together with the previous analysis, the results also suggest that visual complexity might be more directly relevant to perceptual evaluation of complexity, while auditory complexity might be more directly relevant to liking. The explanation for these observed modality differences is unclear. We might speculate that unpleasant auditory stimuli would be more difficult to ignore than unpleasant visual stimuli, leading to a greater effect of auditory complexity on liking. Furthermore, participants may be more familiar with processing dynamic auditory stimuli than dynamically time-varying visual stimuli in everyday life, so variations of complexity in dynamic visual components could have been more salient than variations of complexity in the auditory components and had a greater effect on complexity ratings. However, before such speculations can be tested, research is required to corroborate or contest these findings with other dynamic audiovisual stimuli.

Audiovisual congruence (Cohen, 2013; Lipscomb, 1999, 2005) is for the current stimuli equivalent to the simultaneous occurrence of lines and sounds, in keeping with the temporal (Meredith et al.,

1987; Miller & D'Esposito, 2005; Senkowski et al., 2007) and inverse-effectiveness rules (Holmes, 2009; Kayser et al., 2005; Meredith & Stein, 1983, 1986b; Perrault et al., 2005; Stanford et al., 2005) of audiovisual integration (Spence & Squire, 2003; Stevenson & Wallace, 2013). Correspondence between temporal structures allows inferring the environmental source causing the sensory data (Parise et al., 2012; Senkowski et al., 2007) so that perfectly or highly aligned visual and auditory streams contribute to a unitary, coherent, and fully integrated percept, facilitating processing fluency (Chenier & Winkielman, 2018; Reber, 2011; Reber et al., 2004; Schwarz et al., 2021). The results suggest that congruence amplifies the impact of liking for the unimodal components (Experiment 1) and visual complexity (Experiment 2) on liking for audiovisual stimuli. This finding advances our understanding of how perceptual and evaluative judgments are distinct albeit related cognitive processes (Jacobsen & Höfel, 2003).

The role of perceptual representations of stimulus properties on the relationship between those properties and appreciation has only

**Table 4**

*Best-Fitting Model of Liking for Audiovisual Stimuli as a Function of Liking for Auditory and Visual Stimuli and Audiovisual Congruence*

E	$r_m^2$	$r_c^2$	Fixed effects in R syntax	$\beta$	$df$	$t$	$p$	$d$ [95% CI]
1	.12	.58	<b>a. liking</b>	<b>.13</b>	<b>186.88</b>	<b>3.92</b>	<b>&lt;.01</b>	<b>0.20 [0.15, 0.25]</b>
			<b>v. liking</b>	<b>.18</b>	<b>141.03</b>	<b>5.17</b>	<b>&lt;.01</b>	<b>0.20 [0.14, 0.26]</b>
			congruence	-.12	381.32	-0.66	.51	0.08 [0.03, 0.12]
			<b>a. liking: congruence</b>	<b>.12</b>	<b>2,588.31</b>	<b>4.08</b>	<b>&lt;.01</b>	<b>0.04 [0.02, 0.05]</b>
			v. liking: congruence	.04	1,323.46	1.19	.23	0.01 [-0.01, 0.03]
2	.18	.37	<b>a. liking</b>	<b>.28</b>	<b>133.35</b>	<b>7.11</b>	<b>&lt;.01</b>	<b>0.33 [0.26, 0.40]</b>
			<b>v. liking</b>	<b>.08</b>	<b>155.86</b>	<b>2.20</b>	<b>&lt;.01</b>	<b>0.17 [0.12, 0.23]</b>
			congruence	-.10	502.94	-0.36	.72	0.12 [0.06, 0.17]
			a. liking: congruence	.04	2,639.75	1.02	.31	0.01 [-0.01, 0.04]
			<b>v. liking: congruence</b>	<b>.14</b>	<b>4,281.42</b>	<b>3.50</b>	<b>&lt;.01</b>	<b>0.04 [0.02, 0.07]</b>

*Note.* Estimated effects of liking for auditory (a.liking) and visual (v.liking) stimuli and audiovisual congruence. Significant effects are highlighted in bold. E = experiment;  $r_m^2$  = marginal coefficient of determination;  $r_c^2$  = conditional coefficient of determination;  $\beta$  = unstandardized estimate;  $d$  = effect size; CI = confidence interval.

started to be systematically examined (e.g., Clemente et al., 2023; Clemente, Board, et al., 2024; Clemente, Kaplan, & Pearce, 2024). The results of the *SEM* support a mediating role of perceived complexity on the effects of visual complexity on liking. This finding concurs with those regarding other domains and stimulus properties (e.g., visual contour in Clemente et al., 2023) and supports the role of personal and contextual factors in evaluative judgments (Skov, 2019). Perceived complexity was the main predictor of liking for audiovisual stimuli, showing preeminence of subjective over objective complexity in appreciation (Van Geert & Wagemans, 2020). However, it is plausible that the effects were somehow inflated as judgments of complexity and liking were given concurrently.

Across models, analyses, and experiments, the largest proportion of the variance was explained by the random effects and removing them worsened the fit to the data. This means that accounting for variability between and within participants and stimuli was essential for examining liking for audiovisual objects and the role of stimulus and perceived complexity in liking. This fact underscores the paramount relevance of individual differences in perceptual and evaluative judgments and the need to place them at the center of empirical investigation (Clemente, 2022; Corradi et al., 2020). Since the vast majority of studies on perception and appreciation typically disregard such variability as noise, reinterpretation of existing research might be advisable.

## Limitations and Future Work

Besides those already acknowledged, we highlight some methodological limitations here. First, our CAVA stimulus set manipulates

dynamic auditory and visual complexity using simple visual shapes and sine tones. Importantly, however, the CAVA stimulus set is based on genuine abstract audiovisual artworks and is thus relevant and ecologically valid as a model for nonrepresentational audiovisual art forms, including nonnarrative dance (Howlin et al., 2020). While we view this as a useful starting point, further research is needed to understand the role of dynamic visual and auditory complexity in the appreciation of naturalistic stimuli. A recent and relevant example in this vein is Frame et al.'s (2023) study investigating cross-modal influences on appreciation when presenting naturalistic auditory and visual stimuli simultaneously. This study differed from the present research in that the visual stimuli were static images (of artworks) and, therefore, did not have a dynamic temporal structure that could be aligned (or misaligned) with the auditory stimulus; the results showed independent additive contributions of auditory and visual pleasure ratings on pleasure for the audiovisual whole. Other studies have shown that music biases emotional responses to dance movements (Christensen et al., 2014; Reason et al., 2016), but hearing the sounds that dancers produce (e.g., footsteps and breathing) can reduce liking for the same dance movement despite meeting all criteria for optimal audiovisual integration (Jola et al., 2014; Orgs & Howlin, 2022). Conceivably, perceived meaningfulness (Martindale et al., 1990) of audiovisual congruence can outweigh the importance of structural complexity when people appreciate representational audiovisual artworks like film, theatre, or dance. Our new dynamic measure of visual complexity could be developed to distinguish the relative contribution of syntactic (dynamic audiovisual complexity) and semantic (narrative or storyline) features to further the understanding of the appreciation of these art forms.

**Table 5**

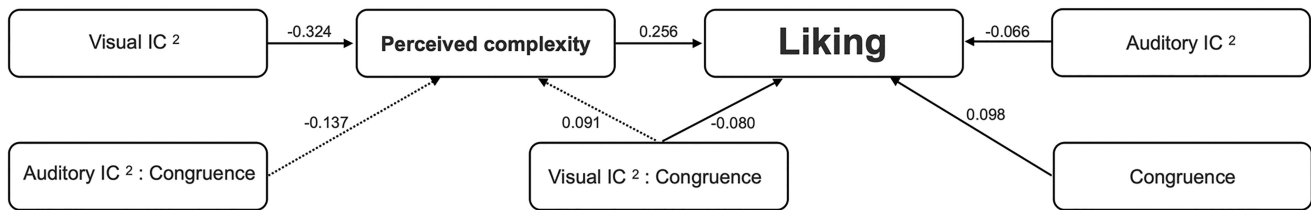
*Models in the Structural Equation Model of Liking for Audiovisual Stimuli*

Response	$r_m^2$	$r_c^2$	Predictor	$b$	$SE$	$df$	Critical value	$p$	$\beta$
Perceived complexity	.08	.65	<b>Visual IC<sup>2</sup></b>	<b>-.71</b>	<b>0.19</b>	<b>103.25</b>	<b>-3.64</b>	<b>&lt;.01</b>	<b>-.32</b>
			Auditory IC <sup>2</sup> : congruence	-.40	0.21	98.31	-1.90	.06	-.14
			Auditory IC <sup>2</sup> : congruence	.28	0.30	96.81	0.94	.35	.09
Liking	.10	.54	<b>Perceived complexity</b>	<b>.24</b>	<b>0.01</b>	<b>4,833.18</b>	<b>22.67</b>	<b>&lt;.01</b>	<b>.26</b>
			<b>Auditory IC<sup>2</sup></b>	<b>-.13</b>	<b>0.05</b>	<b>133.35</b>	<b>-2.72</b>	<b>.01</b>	<b>-.07</b>
			<b>Congruence</b>	<b>.53</b>	<b>0.14</b>	<b>122.51</b>	<b>3.78</b>	<b>&lt;.01</b>	<b>.10</b>
			<b>Visual IC<sup>2</sup>: congruence</b>	<b>-.23</b>	<b>0.08</b>	<b>141.97</b>	<b>-2.98</b>	<b>&lt;.01</b>	<b>-.08</b>

*Note.* Significant effects are highlighted in bold.  $r_m^2$  = marginal coefficient of determination;  $r_c^2$  = conditional coefficient of determination;  $b$  = unstandardized estimate;  $\beta$  = standardized estimate; IC = information content.

**Figure 5**

*SEM Representing the Structure of Relationships Between Auditory and Visual Quadratic IC<sup>2</sup>, Audiovisual Congruence, Perceived Complexity, and Liking for Audiovisual Stimuli*



*Note.* Numbers represent standardized estimates. Solid arrows denote a significant impact of one variable on another ( $p \leq .05$ ). Dashed arrows depict non-significant associations ( $p > .05$ ) necessarily included in the SEM—as removing them from the individual models would worsen the model fit, eliminating the degrees of freedom required to evaluate the global SEM. SEM = structural equation model; IC = information content.

Second, although these experiments systematically manipulated dynamic feature-based complexity across modalities and linked it with information-based complexity in the CAVA set, the relationship between feature-based complexity and IC and their contribution to liking may differ for other stimuli. Therefore, generalizing the present experimental approach to other bimodal stimuli is critical. In addition, auditory and visual complexity could be characterized in other ways. For instance, intermittent dots instead of moving lines would be more directly comparable to the pitches in the auditory modality. That said, we did not find an influence of pitch or line movement direction on liking for audiovisual displays, suggesting that specific low-level visual or auditory features are less important than the higher-level attributes of complexity or congruence.

Lastly, the samples were characterized in terms of musical but not visual sophistication, and we did not manipulate domain-specific expertise systematically. Nonetheless, domain-specific expertise is known to affect the role of complexity in liking (Hekkert & van Wieringen, 1996; Lahdelma & Eerola, 2020; Orr & Ohlsson, 2005; Popescu et al., 2019; Skov & Kirk, 2021). Therefore, future research should investigate the impact of visual and musical sophistication and other relevant traits on the relationship between complexity and liking for bimodal stimuli.

## Conclusion

The present research investigates the role of complexity and the relationships between stimulus complexity and perceived complexity in liking by systematically manipulating comparable auditory and visual dynamic complexity. In so doing, we contribute a novel stimulus set and generalizable computational measures of auditory and visual dynamic complexity, all of which are available as open resources for research at <https://osf.io/e5uh9/>. Liking for audiovisual stimuli can be explained as a weighted combination of liking for its auditory and visual components moderated by audiovisual congruence depending on the relative salience of perceived complexity. Furthermore, liking is maximal for intermediate levels of auditory and visual complexity, the latter mediated by perceived complexity and enhanced by congruence. Crucially, the most variance is explained by variability due to participants and stimuli, highlighting the importance of considering individual differences.

## References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>
- Alwitt, L. F., Anderson, D. R., Lorch, E. P., & Levin, S. R. (1980). Preschool children's visual attention to attributes of television. *Human Communication Research*, 7(1), 52–67. <https://doi.org/10.1111/j.1468-2958.1980.tb00550.x>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Ben-Shachar, M., Lüdtke, D., & Makowski, D. (2020). *Compute and interpret indices of effect size*. <https://github.com/easystats/effectsize>
- Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8(5), 279–286. <https://doi.org/10.3758/bf03212593>
- Berlyne, D. E. (1971). *Aesthetics and psychobiology*. Appleton-Century-Crofts.
- Berlyne, D. E., & Boudewijns, W. J. (1971). Hedonic effects of uniformity in variety. *Canadian Journal of Psychology*, 25(3), 195–206. <https://doi.org/10.1037/h0082381>
- Berlyne, D. E., McDonnell, P., Nicki, R. M., & Parham, L. C. C. (1967). Effects of auditory pitch and complexity on EEG desynchronization and on verbally expressed judgments. *Canadian Journal of Psychology*, 21(4), 346–367. <https://doi.org/10.1037/h0082987>
- Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, 198(1), 49–57. <https://doi.org/10.1007/s00221-009-1933-z>
- Che, J., Sun, X., Gallardo, V., & Nadal, M. (2018). Cross-cultural empirical aesthetics. *Progress in Brain Research*, 237, 77–103. <https://doi.org/10.1016/bs.pbr.2018.03.002>
- Chen, Y. C., & Spence, C. (2017). Assessing the role of the 'unity assumption' on multisensory integration: A review. *Frontiers in Psychology*, 8, Article 445. <https://doi.org/10.3389/fpsyg.2017.00445>
- Chenier, T., & Winkielman, P. (2018). The origins of aesthetic pleasure: Processing fluency and affect in judgment, body, and the brain. In M. Skov & O. Vartanian (Eds.), *Neuroaesthetics* (pp. 275–289). Routledge. <https://doi.org/10.4324/9781315224091-14>
- Cheung, V. K., Harrison, P. M., Meyer, L., Pearce, M. T., Haynes, J. D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical

- pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology*, 29(23), 4084–4092.e4. <https://doi.org/10.1016/j.cub.2019.09.067>
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, 295(2), 295–336.
- Chion, M. (1994). *Audio-vision: Sound on screen*. Columbia University Press.
- Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature. *Psychology of Music*, 45(6), 886–909. <https://doi.org/10.1177/0305735617697507>
- Christensen, J. F., Gaigg, S. B., Gomila, A., Oke, P., & Calvo-Merino, B. (2014). Enhancing emotional experiences to dance through music: The role of valence and arousal in the cross-modal bias. *Frontiers in Human Neuroscience*, 8, Article 359. <https://doi.org/10.3389/fnhum.2014.00757>
- Clemente, A. (2022). Aesthetic sensitivity: Origin and development. In M. Skov & M. Nadal (Eds.), *The Routledge international handbook of neuroaesthetics* (pp. 240–253). Routledge. <https://doi.org/10.4324/9781003008675-13>
- Clemente, A., Board, F., Pearce, M. T., & Orgs, G. (2024). *Dynamic complexity in audiovisual aesthetics, data and materials*. Open Science Framework. <https://doi.org/10.17605/OSF.IO/E5UH9>
- Clemente, A., Kaplan, T. M., & Pearce, M. T. (2024). Perceptual representations mediate effects of stimulus properties on liking for music. *Annals of the New York Academy of Sciences*, 1533(1), 169–180. <https://doi.org/10.1111/nyas.15106>
- Clemente, A., Pearce, M. T., & Nadal, M. (2022). Musical aesthetic sensitivity. *Psychology of Aesthetics, Creativity, and the Arts*, 16(1), 58–73. <https://doi.org/10.1037/aca0000381>
- Clemente, A., Pearce, M. T., Skov, M., & Nadal, M. (2021). Evaluative judgment across domains: Liking balance, contour, symmetry and complexity in melodies and visual designs. *Brain and Cognition*, 151, Article 105729. <https://doi.org/10.1016/j.bandc.2021.105729>
- Clemente, A., Penacchio, O., Vila-Vidal, M., Pepperell, R., & Ruta, N. (2023). Explaining the curvature effect: Perceptual and hedonic evaluations of visual contour. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. <https://doi.org/10.1037/aca0000561>
- Clemente, A., Vila-Vidal, M., Pearce, M. T., Aguiló, G., Corradi, G., & Nadal, M. (2020). A set of 200 musical stimuli varying in balance, contour, symmetry, and complexity: Behavioral and computational assessments. *Behavior Research Methods*, 52(4), 1491–1509. <https://doi.org/10.3758/s13428-019-01329-8>
- Cohen, A. J. (2013). Congruence-association model of music and multimedia: Origin and evolution. In A. J. Cohen, S. D. Lipscomb, & R. A. Kendal (Eds.), *The psychology of music in multimedia* (pp. 17–47). Oxford University Press.
- Corradi, G., Chuquichambi, E. G., Barrada, J. R., Clemente, A., & Nadal, M. (2020). A new conception of visual hedonic sensitivity. *British Journal of Psychology*, 111(4), 630–658. <https://doi.org/10.1111/bjop.12427>
- Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review*, 11(1), 2–17. <https://doi.org/10.18061/emr.v11i1.4836>
- Eerola, T., Himberg, T., Toivianen, P., & Louhivuori, J. (2006). Perceived complexity of Western and African folk melodies by Western and African listeners. *Psychology of Music*, 34(3), 337–371. <https://doi.org/10.1177/0305735606064842>
- Eerola, T., & North, A. C. (2000). *Expectancy-based model of melodic complexity* [Conference session]. Proceedings of the sixth international conference on music perception and cognition.
- Fernandez-Lozano, C., Carballal, A., Machado, P., Santos, A., & Romero, J. (2019). Visual complexity modelling based on image features fusion of multiple kernels. *PeerJ*, 7, Article e7075. <https://doi.org/10.7717/peerj.7075>
- Flagg, B. N., Allen, B. D., Geer, A. H., & Scinto, L. F. (1976). *Children's visual responses to Sesame Street: A formative research report*. Unpublished research report, Children's Television Workshop.
- Frame, J., Gugliano, M., Bai, E., Briellmann, A., & Belfi, A. M. (2023). Your ears don't change what your eyes like: People can independently report the pleasure of music and images. *Journal of Experimental Psychology: Human Perception and Performance*, 49(6), 774–785. <https://doi.org/10.1037/xhp0001118>
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: A reward for learning? *The Journal of Neuroscience*, 39(47), 9397–9409. <https://doi.org/10.1523/JNEUROSCI.0428-19.2019>
- Griffin, I. C., Miniussi, C., & Nobre, A. C. (2002). Multiple mechanisms of selective attention: Differential modulation of stimulus processing by attention to space or time. *Neuropsychologia*, 40(13), 2325–2340. [https://doi.org/10.1016/S0028-3932\(02\)00087-8](https://doi.org/10.1016/S0028-3932(02)00087-8)
- Güçlütürk, Y., Jacobs, R. H. A. H., & van Lier, R. (2016). Liking versus complexity: Decomposing the inverted U-curve. *Frontiers in Human Neuroscience*, 10, Article 112. <https://doi.org/10.3389/fnhum.2016.00112>
- Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In M. Riess Jones, R. Fay & A. Popper (Eds.), *Music perception* (pp. 233–258). Springer. [https://doi.org/10.1007/978-1-4419-6114-3\\_8](https://doi.org/10.1007/978-1-4419-6114-3_8)
- Hekkert, P., & van Wieringen, P. C. (1996). The impact of level of expertise on the evaluation of original and altered versions of post-impressionistic paintings. *Acta Psychologica*, 94(2), 117–131. [https://doi.org/10.1016/0001-6918\(95\)00055-0](https://doi.org/10.1016/0001-6918(95)00055-0)
- Heyduk, R. G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception & Psychophysics*, 17(1), 84–90. <https://doi.org/10.3758/BF03204003>
- Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia*, 45(14), 3340–3345. <https://doi.org/10.1016/j.neuropsychologia.2007.05.025>
- Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topography*, 21(3–4), 168–176. <https://doi.org/10.1007/s10548-009-0097-2>
- Howlin, C., Vicary, S., & Orgs, G. (2020). Audiovisual aesthetics of sound and movement in contemporary dance. *Empirical Studies of the Arts*, 38(2), 191–211. <https://doi.org/10.1177/0276237418818633>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications*. Routledge.
- Jacobsen, T., & Höfel, L. (2003). Descriptive and evaluative judgment processes: Behavioral and electrophysiological indices of processing symmetry and aesthetics. *Cognitive, Affective, & Behavioral Neuroscience*, 3(4), 289–299. <https://doi.org/10.3758/CABN.3.4.289>
- Jola, C., Pollick, F. E., & Calvo-Merino, B. (2014). “Some like it hot”: Spectators who score high on the personality trait openness enjoy the excitement of hearing dancers breathing without music. *Frontiers in Human Neuroscience*, 8, Article 718. <https://doi.org/10.3389/fnhum.2014.00718>
- Jordan, S. (2011). Choreomusical conversations: Facing a double challenge. *Dance Research Journal*, 43(1), 43–64. <https://doi.org/10.5406/danceresearchj.43.1.0043>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68(1), 601–625. <https://doi.org/10.1146/annurev-psych-122414-033702>
- Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, 48(2), 373–384. <https://doi.org/10.1016/j.neuron.2005.09.018>
- Kearns, J., & O'Connor, B. (2004). Dancing with entropy: Form attributes, children, and representation. *Journal of Documentation*, 60(2), 144–163. <https://doi.org/10.1108/00220410410522034>



- Kesner, L. (2014). The predictive mind and the experience of visual art work. *Frontiers in Psychology*, 5, Article 1417. <https://doi.org/10.3389/fpsyg.2014.01417>
- Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, 23(1), 63–77. <https://doi.org/10.1016/j.tics.2018.10.006>
- Köhler, W. (1971/1930). Human perception. In M. Henle (Ed.), *The selected papers of Wolfgang Köhler*. Liveright.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lahdelma, I., & Eerola, T. (2020). Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension. *Scientific Reports*, 10(1), Article 8693. <https://doi.org/10.1038/s41598-020-65615-8>
- Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, 7(5), 573–579. <https://doi.org/10.1111/2041-210X.12512>
- Lipscomb, S. D. (1999). Cross-modal integration: Synchronization of auditory and visual components in simple and complex media. *The Journal of the Acoustical Society of America*, 105(2), Article 1274. <https://doi.org/10.1121/1.426089>
- Lipscomb, S. D. (2005). The perception of audio-visual composites: Accent structure alignment of simple stimuli. *Selected Reports in Ethnomusicology*, 12, 37–67.
- Lipscomb, S. D., & Kendall, R. A. (1994). Perceptual judgment of the relationship between musical and visual components in film. *Psychomusicology: A Journal of Research in Music Cognition*, 13(1–2), 60–98. <https://doi.org/10.1037/h0094101>
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>
- Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: A PET study. *Neuroimage*, 21(2), 725–732. <https://doi.org/10.1016/j.neuroimage.2003.09.049>
- Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, 160, 43–57. <https://doi.org/10.1016/j.actpsy.2015.06.005>
- Marin, M. M., Lampatz, A., Wandl, M., & Leder, H. (2016). Berlyne revisited: Evidence for the multifaceted nature of hedonic tone in the appreciation of paintings and music. *Frontiers in Human Neuroscience*, 10, Article 536. <https://doi.org/10.3389/fnhum.2016.00536>
- Marin, M. M., & Leder, H. (2018). Exploring aesthetic experiences of females: Affect-related traits predict complexity and arousal responses to music and affective pictures. *Personality and Individual Differences*, 125, 80–90. <https://doi.org/10.1016/j.paid.2017.12.027>
- Martindale, C., Moore, K., & Borkum, J. (1990). Aesthetic preference: Anomalous findings for berlyne's psychobiological theory. *The American Journal of Psychology*, 103(1), 53–80. <https://doi.org/10.2307/1423259>
- Mauch, M., & Levy, M. (2011). *Structural change on multiple time scales as a correlate of musical complexity* [Conference session]. 12th International Society for Music Information Retrieval Conference (ISMIR 2011) (pp. 489–494). <https://ismir2011.ismir.net/papers/PS4-3.pdf>
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. <https://doi.org/10.1038/264746a0>
- Meijer, D., Veselic, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex*, 119, 74–88. <https://doi.org/10.1016/j.cortex.2019.03.026>
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. *The Journal of Neuroscience*, 7(10), 3215–3229. <https://doi.org/10.1523/JNEUROSCI.07-10-03215.1987>
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608), 389–391. <https://doi.org/10.1126/science.6867718>
- Meredith, M. A., & Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365(2), 350–354. [https://doi.org/10.1016/0006-8993\(86\)91648-3](https://doi.org/10.1016/0006-8993(86)91648-3)
- Meredith, M. A., & Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, 56(3), 640–662. <https://doi.org/10.1152/jn.1986.56.3.640>
- Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, 75(5), 1843–1857. <https://doi.org/10.1152/jn.1996.75.5.1843>
- Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *The Journal of Neuroscience*, 25(25), 5884–5893. <https://doi.org/10.1523/JNEUROSCI.0896-05.2005>
- Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, 53(4), 1551–1562. <https://doi.org/10.3758/s13428-020-01514-0>
- Mindus, L. A. (1968). *The role of redundancy and complexity in the perception of tonal patterns* [Doctoral dissertation]. Clark University.
- Møller, C., Garza-Villarreal, E. A., Hansen, N. C., Højlund, A., Bærentsen, K. B., Chakravarty, M. M., & Vuust, P. (2021). Audiovisual structural connectivity in musicians and non-musicians: A cortical thickness and diffusion tensor imaging study. *Scientific Reports*, 11(1), Article 4324. <https://doi.org/10.1038/s41598-021-83135-x>
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, 17(1), 154–163. [https://doi.org/10.1016/S0926-6410\(03\)00089-2](https://doi.org/10.1016/S0926-6410(03)00089-2)
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2), Article e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2), 173–191. <https://doi.org/10.2190/EM.28.2.d>
- Orgs, G., & Howlin, C. (2022). The audio-visual aesthetics of music and dance. In M. Nadal & O. Vartanian (Eds.), *The Oxford handbook of empirical aesthetics*. Oxford University Press. <https://doi.org/https://doi.org/10.1093/oxfordhb/9780198824350.013.29>
- Orlandi, A., Cross, E. S., & Orgs, G. (2020). Timing is everything: Dance aesthetics depend on the complexity of movement kinematics. *Cognition*, 205, Article 104446. <https://doi.org/10.1016/j.cognition.2020.104446>
- Orr, M. G., & Ohlsson, S. (2005). Relationship between complexity and liking as a function of expertise. *Music Perception*, 22(4), 583–611. <https://doi.org/10.1525/mp.2005.22.4.583>
- Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology*, 22(1), 46–49. <https://doi.org/10.1016/j.cub.2011.11.039>
- Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* [Doctoral dissertation]. City University London.
- Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, 1423(1), 378–395. <https://doi.org/10.1111/nyas.13654>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>

- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7), 688–691. <https://doi.org/10.1038/nn1083>
- Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, 56(1), 89–114. <https://doi.org/10.1146/annurev.psych.56.091103.070225>
- Perrault, T. J., Vaughan, J. W., Stein, B. E., & Wallace, M. T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *Journal of Neurophysiology*, 93(5), 2575–2586. <https://doi.org/10.1152/jn.00926.2004>
- Popescu, T., Neuser, M. P., Neuwirth, M., Bravo, F., Mende, W., Boneh, O., Santos, A. G., da Rocha, G. O., de Andrade, J. B., & Rohrmeier, M. (2019). The pleasantness of sensory dissonance is mediated by musical style and expertise. *Scientific Reports*, 9, Article 1070. <https://doi.org/10.1038/s41598-018-35873-8>
- Potter, R. F., & Choi, J. (2006). The effects of auditory structural complexity on attitudes, attention, arousal, and memory. *Media Psychology*, 8(4), 395–419. [https://doi.org/10.1207/s1532785xmep0804\\_4](https://doi.org/10.1207/s1532785xmep0804_4)
- R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.2.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>
- Reason, M., Jola, C., Kay, R., Reynolds, D., Kauppi, J.-P., Grobras, M.-H., Tohka, J., & Pollick, F. E. (2016). Spectators' aesthetic experience of sound and movement in dance performance: A transdisciplinary investigation. *Psychology of Aesthetics, Creativity, and the Arts*, 10(1), 42–55. <https://doi.org/10.1037/a0040032>
- Reber, R. (2011). Processing fluency, aesthetic pleasure, and culturally shared taste. In A. P. Shimamura & S. E. Palmer (Eds.), *Aesthetic science: Connecting mind, brain, and experience* (pp. 223–242). Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780199732142.003.0055>
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364–382. [https://doi.org/10.1207/s15327957pspr0804\\_3](https://doi.org/10.1207/s15327957pspr0804_3)
- Rohe, T., & Noppeney, U. (2018). Reliability-weighted integration of audiovisual signals can be modulated by top-down attention. *Eneuro*, 5(1). <https://doi.org/10.1523/ENEURO.0315-17.2018>
- Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, 10(4), 251. <https://doi.org/10.3390/brainsci10040251>
- Sauvé, S. A., & Pearce, M. T. (2019). Information-theoretic modeling of perceived musical complexity. *Music Perception*, 37(2), 165–178. <https://doi.org/10.1525/mp.2019.37.2.165>
- Schwarz, N., Jalbert, M., Noah, T., & Zhang, L. (2021). Metacognitive experiences as information: Processing fluency in consumer judgment and decision making. *Consumer Psychology Review*, 4(1), 4–25. <https://doi.org/10.1002/arcp.1067>
- Senkowski, D., Talsma, D., Grigutsch, M., Hermann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia*, 45(3), 561–571. <https://doi.org/10.1016/j.neuropsychologia.2006.01.013>
- Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. *Ecology*, 90(2), 363–368. <https://doi.org/10.1890/08-1034.1>
- Skov, M. (2019). The neurobiology of sensory valuation. In M. Nadal & O. Vartanian (Eds.), *The Oxford handbook of empirical aesthetics* (pp. 1–40). Oxford University Press.
- Skov, M., & Kirk, U. (2021). Expertise and aesthetic liking. In A. Chatterjee & E. R. Cardillo (Eds.), *Brain, beauty, and art: Essays bringing neuroaesthetics into focus* (Vol. 69, p. 70). Oxford University Press.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis. An introduction to basic and advanced multilevel modeling* (2nd ed.). SAGE Publications.
- Spence, C., & Soto-Faraco, S. (2010). Auditory perception: Interactions with vision. *The Oxford Handbook of Auditory Science: Hearing*, 3, 271–296. <https://doi.org/10.1093/oxfordhb/9780199233557.013.0012>
- Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, 13(13), R519–R521. [https://doi.org/10.1016/S0960-9822\(03\)00445-7](https://doi.org/10.1016/S0960-9822(03)00445-7)
- Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *The Journal of Neuroscience*, 25(28), 6499–6508. <https://doi.org/10.1523/JNEUROSCI.5095-04.2005>
- Stein, B. E. (2012). *The new handbook of multisensory processing*. MIT Press.
- Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*, 44(3), 1210–1223. <https://doi.org/10.1016/j.neuroimage.2008.09.034>
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: Task and stimulus dependencies. *Experimental Brain Research*, 227(2), 249–261. <https://doi.org/10.1007/s00221-013-3507-3>
- Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, 21(10), 736–748. <https://doi.org/10.1016/j.tics.2017.06.007>
- Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences*, 110(36), 14580–14585. <https://doi.org/10.1073/pnas.1221454110>
- Van de Cruys, S., Bervoets, J., & Moors, A. (2022). Preferences need inferences: Learning, valuation, and curiosity in aesthetic experience. In M. Skov & M. Nadal (Eds.), *The Routledge international handbook of neuroaesthetics* (pp. 475–506). Routledge. <https://doi.org/10.4324/9781003008675-13>
- Van de Cruys, S., Chamberlain, R., & Wagemans, J. (2017). Tuning in to art: A predictive processing account of negative emotion in art. *Behavioral and Brain Sciences*, 40, Article e377. <https://doi.org/10.1017/S0140525X17001868>
- Van de Cruys, S., & Wagemans, J. (2011). Putting reward in art: A tentative prediction error account of visual art. *i-Perception*, 2(9), 1035–1062. <https://doi.org/10.1068/i0466aap>
- Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, 14(2), 135–154. <https://doi.org/10.1037/aca0000224>
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17(2), 228–243. <https://doi.org/10.1037/a0027127>
- Wartella, E., & Ettema, J. S. (1974). A cognitive developmental study of children's attention to television commercials. *Communication Research*, 1(1), 69–88. <https://doi.org/10.1177/009365027400100104>
- Watt, J. H., & Welch, A. J. (1982). Effects of static and dynamic complexity on children's attention and recall of televised instruction. *Human Communication Research*, 8(2), 133–145. <https://doi.org/10.1111/j.1468-2958.1982.tb00660.x>
- Wertheimer, M. (1938/1924). Gestalt theory. In W. D. Ellis (Ed.), *A source book of gestalt psychology* (pp. 2–11). Routledge & Kegan P. <https://archive.org/details/in.ernet.dli.2015.198039/page/n13/mode/2up>
- Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika*, 92(4), 937–950. <https://doi.org/10.1093/biomet/92.4.937>

Received September 12, 2023

Revision received December 7, 2023

Accepted January 25, 2024 ■