Audiovisual aesthetics of sound and movement in contemporary dance

Claire Howlin1, Staci Vicary2, & Guido Orgs3

1. University College Dublin, Ireland
2. Psychological Sciences, Australian College of Applied Psychology, Australia
3. Department of Psychology, Goldsmiths University London, UK

Keywords; Dance; Aesthetics; Music; Audio-visual; Granger Causality, Multisensory, Time-series analysis

Introduction

Dance is often performed to music (Jordan, 2011) and the relationship between sound and movement has been a constant topic of interest for composers and choreographers alike (Cunningham, 1982; Hagendoorn, 2004, de Keersmaeker & Cvejić, 2014). If music and dance are combined, their relationship can be *congruent,* illustrating or amplifying structural and expressive similarities (Krumhansl & Schenk, 1997), or *incongruent* with no direct and sometimes even conflicting relationship between auditory and visual information. Cook (2000) refers to music-image relationships where each medium represents a different narrative structure that arises in a sense of collision or contradiction as in contest. In movies, music is considered empathetic if it resembles the rhythm, tone or phrasing of a scene, or unempathetic if it appears indifferent to what is displayed on screen (Chion, 1994). In this context, music is considered to be the composite mix of musical score alongside, ambient sound, dialogue, sound effects and silence, (Lipscomb & Tulchinsky, 2004). Compared to a long history of research in music perception from both humanities and empirical perspectives, the relationship between movement and sound from a dance perspective has only recently come into focus (Jordan, 2011). At the same time, embodied interactions with music play an increasingly important role in studies of music perception (Leman, Lesaffre, & Maes, 2017), including the cultural practice of embodiment itself (Van Dyck, Burger, & Orlandatou, 2017; Hinberg, 2017). In the present study, we examine the role of movement and sound congruency during aesthetic perception of contemporary dance and relate our findings to a prominent model of audio-visual integration in the arts, the congruence association model (CAM).

The CAM focuses on two analytic processes in audiovisual engagement, temporal structure on the one hand, and the corresponding associated meaning on the other hand (Marshall & Cohen, 1988; Cohen, 2013a; Cohen, 2013b). Temporal structure refers to how visual and auditory patterns change across time and space. Cohen (2016) argues that structural congruencies between the auditory and visual streams of movies direct attention to particular elements of the visual stream. These congruencies emphasize specific aspects of the narrative. Previous studies found that more congruent soundtracks enhance absorption in the movie (Cohen, MacMillan, & Drew, 2006), and engagement with the narrative (Cohen & Siau, 2008). Audiovisual congruency also influences the meaning of a movie on an aggressive-friendly dimension (Bolivar, Cohen, & Fentress, 1994). These studies show that structural and expressive congruency between audio and visual streams heightens spectators experience of movies, such influence however will also depend on meaningfulness of audio and visual streams, i.e. speech sounds vs. music (Cohen 2013a; Cohen 2013b).

Audiovisual congruency of movement and sound is central to dance and choreography. Krumhansl and Schenk (1997) demonstrated that people experience expressive similarities between music and dance, if dance movements are choreographed to a specific musical score. All three experimental conditions - music and dance, dance only and music only recordings of a choreography by George Balanchine to a Mozart Divertimento - were rated very similarly across a range of emotion dimensions. Another study demonstrated that audiences can detect a "match" between dance choreography and music that are intended go together, when asked to select from three pieces of music (Mitchell & Gallagher, 2001). Moreover, meter perception in dance choreography can be disrupted by a competing musical meter, and this effect is even stronger when a well-known choreography is used, such as the popular dance to gangnam style (Lee, Barrett, Kim, Lim, & Lee; 2015). Together, these studies show that audiovisual congruency indeed contributes to dance perception.

Congruent choreomusic parallels such as intertwining rhythms of movement and music are a common feature of contemporary dance and choreography (Jordan, 2011). Yet, post-dramatic forms of dance and choreography (Lehmann, 2006) also apply incongruent movement/sound relationships that appear unconnected and do not exhibit any obvious or intended choreomusical parallels, i.e. in the collaborations between choreographer Merce Cunningham and composer John Cage (Cunningham & Lesschaeve, 1985). Incongruence of musical and movement features arguably encourages to divide attention between visual and auditory streams of information (Fogelsanger & Afanador, 2006). Yet from an audience perspective, perceiving deliberate incongruencies or unconnectedness may be difficult to achieve. Intended or unintended, the mere presence of a soundtrack will likely influence how dance movements are perceived and interpreted. Accordingly, visual and auditory congruency may still occur incidentally, in line with gestalt principles of perception (Arnheim, 1974). In effect this can lead to a sense of aesthetic ‘capture’ or 'driving', where one element affects how the other element is perceived (Mitchell & Gallaher, 2001).

One of the first studies to examine the role of audiovisual congruency in aesthetic perception of live dance combined a qualitative approach with fMRI (Reason, Jola, Kay, Reynolds, Kauppi, Grobras, et. al., 2016). Reason and colleagues (2016) compared qualitative aesthetic reactions to a live dance performance, performed three times, in the presence of three different soundtracks; Bach’s concerto for Oboe and Violin in C, the sound of breathing and footfalls without music, and electronic music. While Reason et.al. (2016) expected each soundtrack to lead to a different interpretation by the audience, they did not find any differences between the two music conditions. However, they found that the novice audience members did not like the condition with the breathing soundtrack due to a heightened sense of physical presence that was perceived as unpleasant by some spectators. The authors interpreted this result as evidence for a stronger embodied response if the dancers breathing sounds were audible (Reason et. al, 2016, Jola, Pollick & Calvo Merino, 2014). In line with the concept of ’kinaesthetic empathy’, which suggests that increased bodily awareness operates as a mediator for aesthetic experience, due to a shift from primarily visual engagement to a proprioceptive body-to-body experience. However, completely different soundtracks were used for each condition, so it is possible that spectators preferred specific *soundtracks* rather than specific *relationships* between movement and soundtrack. To examine the role of audio-visual congruency in dance perception, it is essential to use identical soundtracks, and only manipulate the temporal relationship between movement and sound.

In this experiment we examine how congruency between sound and movement impacts on the aesthetic appreciation of contemporary dance videos. Importantly, we manipulate audiovisual congruency using a single soundtrack, thus matching both sound conditions for lower level sound features. In the congruent condition movement and sound were presented as recorded, in the incongruent condition, the soundtrack was played in reverse. Congruency of movement and sound might either enhance or reduce audience engagement with the dance video. Congruency may be preferred, as it emphasizes expressive structural features across visual and auditory domains (Krumhansl & Schenk, 1997). Alternatively, incongruency may lead to a sense of aesthetic capture, ambiguity and surprise that is potentially enjoyable (Mitchell & Gallaher, 2001; Fogelsanger & Afanador, 2006). Accordingly, the aesthetics of movement-sound congruency allow to empirically test the applicability of the CAM model to contemporary dance.

Method

Participants

Thirty-four participants, (9 male, 25 female), ranging in age from 18-51 years (*M* = 27.41; *SD* = 7.3), volunteered to take part. Musicality ranged from 48-100 (*M* = 74.71; *SD* = 17.85). on the Goldsmiths musical sophistication index v1.0 (GMSI) musicality scale. Musicality involves a wide range of musical behaviours and musical engagement such as understanding and evaluation (Müllensiefen, Gingras, Musil, & Stewart, 2014). The GMSI is a standardised instrument that provides a measure of musicality in the general population, which ranges from 18-126 with an average score of 81.58 (*SD* = 20.62) across the general population (Müllensiefen, Gingras, Musil, & Stewart, 2013).Dance experience was assessed using a custom-made questionnaire, which asked people how many years dance experience they had, and how often they watched recorded dance. Fifty-three percent of participants reported previous formal dance training, but none were professional dancers, see table 1 for a demographic breakdown by experimental group. Participants volunteered in response to a Goldsmiths participation page on Facebook, and all participants signed a written consent form before taking part in the experiment. Participants were signed up to experimental groups based on their availability; allocation of experimental group to condition was randomized using numbered sealed envelopes containing randomization cards for either the congruent, incongruent, or silent conditions. The groups included 10, 14 and 10 participants in the congruent, silent, and incongruent groups respectively. All participants were included in the physiological and qualitative analyses however the tablet data of 8 participants was lost due to technical problems, leaving only 8, 10 and 8 in the congruent, silent, and incongruent groups for the enjoyment analyses.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 1.  *Demographics* | | | |
|  | Congruent (N = 10) | Silent (N = 14) | Incongruent (N = 10) |
| Gender | F = 5 | F = 11 | F = 9 |
| Age | *M* = 29.6,  *SD* = 12.1  Range 18 - 51 | *M* = 26.0,  *SD* = 2.9  Range 22 - 31 | *M* = 26.2,  *SD* = 4.6  Range 23 - 38 |
| Gold MSI score | *M* = 78.2, *SD* = 13.1 | M = 73.9, SD = 8.5 | *M* = 80.3, *SD* = 16.2 |
| Dance Experience | Yes = 4 | Yes = 9 | Yes = 5 |
| Years Dance experience | *M* = 2.3, *SD* = 3.7 | *M* = 5.0 *SD* = 5.04 | *M* = 1.35, *SD* = 2.3 |
| Watch recorded dance at least once a month\* | 30.0% | 28.6% | 20.0% |
| Participants included in enjoyment analyses | 8 | 10 | 8 |
| Notes. *SD =* Standard Deviation, *M* = Mean, \*no distinction was made in terms of the type of dance watched, be it contemporary or otherwise | | | |

Materials

*Dance performance video*

A full length video recording of Matthias Sperling’s choreography ‘Group Study’, (<https://www.youtube.com/playlist?list=PLDAYdl_CLkZj2qWBDkxxVug4c5HQmCiB1>) with a running time of 33 minutes 54 seconds, was used as the basis for the three audiovisual conditions A full length presentation was used to enhance the ecological validity of the study in line with previous recommendations (Jola & Christensen, 2015). The perspective and point of view was identical in all three videos, there was no visual difference in the dance videos presented to the groups, only the soundtrack was altered to allow for a controlled comparison. This performative work for 10 performers, was initially developed in the context of research on aesthetic perception of movement synchrony in dance (Vicary, Sperling, von Zimmermann, Richardson, & Orgs, 2017). The choreography was based on a contemporary dance style that consisted of a series of rule-based improvised group movement tasks. Rather than specifying a fixed movement sequence, the choreographic score specifies dynamic relationships between performers and is based on a pedestrian movement vocabulary consisting of walking, running and falling movements, often associated with postmodern performing dance (Dempster, 2008). Familiarity with a movement vocabulary is an important predictor of aesthetic judgement (Kirsch, Drommelschmidt, & Cross, 2013; Orgs, Hagura, & Haggard, 2013). Using a pedestrian movement vocabulary therefore allows to control for movement vocabulary across spectators (Orgs, Dombrowski, Heil, & Jansen-Osmann, 2008). In addition to the rhythmical sounds associated with these movements, the score specified vocalisations by the dancers with a similar, rule-based structure. The dance performance did not include any music. In the congruent condition performer movement and sounds were presented as recorded. These sounds included breathing, humming, counting and running noises, which were rhythmic for some parts of the dance. The reversed soundtrack played the end of the soundtrack first, and all the corresponding sounds were played backwards in reverse order, to disrupt the temporal congruency of the music and dance presentation. Congruent and incongruent conditions were thus matched for lower level acoustic features, but only differed with respect to the correct alignment of performers sounds to performer movement. The silent condition was devoid of any sound silence with the original soundtrack muted.

*Continuous measures of dance movement, audience engagement and arousal*

To account for the dynamic experience of the performing arts (Orgs, Caspersen, Haggard, 2016; Vicary et al., 2017), we use continuous rather than discrete measures to capture audience enjoyment of the entire performance, rather than short excerpts which would potentially eliminate important dramaturgical cues. We used tablet computers and wrist sensors to collect these continuous measures, and applied agreement analysis and granger causality time-series analysis (Stevens, Schubert, Morris, Frear, Chen, Healey, Hansen, 2009; Schubert, Vincs, & Stevens, 2013; Dean & Bailes, 2010; Dean & Dunsmuir, 2015; Granger, 1969).

Aesthetic perception of the dance video was tracked with respect to three continuous variables, synchrony of the dancers, overall movement acceleration of the dancers, and the spectral power of the soundtrack. Performed synchrony was calculated by applying cross-recurrence quantification analysis (CRQA) to the non-windowed acceleration vectors to obtain a continuous measure of synchrony among performers across time using identical methodology to the procedure outlined by Vicary et al. (2017). Movement of the performers was based on the acceleration series extracted from the wrist sensors of each of the 10 performers. The magnitude of acceleration was then calculated from the 3-axis acceleration data, by taking the square root of the sum of squared x, y, and z values, leaving a single time series vector for each performer, with greater values indicating increased acceleration. The spectral power of the soundtrack at each time point was analysed using the miraudio toolbox for Matlab (Lartillot, Toiviainen, & Eerola, 2008). All measures were computed in non-overlapping time windows with a length of 2s.

ASUS tablets with a pixel display of 800 x 1280, running the android version of ‘OpenSesame’ experiment software based on Python script (Mathôt, Schreij, & Theeuwes, 2012) were used to collect continuous enjoyment ratings. This was done using a custom finger tracking application which was programmed to continuously collect the location of the participants index finger on the touch screen. The touch screen displayed a single axis ranging from 'enjoyed very little' to 'enjoyed very much', with white text on a black background (see figure 1c), as a measure of the spectator's enjoyment of the recorded dance video over time. The numerical range for the enjoyment data was reflected in the number of pixels on the screen, and ranged from +400 for 'liked very much' to -400 for 'liked very little', however these numerical values were never visible or made known to the spectators. Enjoyment data was collected at a rate of 25 samples per second.

|  |
| --- |
|  |
| Figure 1. Experimental Apparatus. Showing (1a) Empatica E4 wristband worn by wach member of the audience to measure Electrodermal Activity and Heart Rate while watching the dance performance; (1b) A still image from the contemporary dance performance 'Group Study'; (1c) Axis used to collect continuous enjoyment measures ranging from 'enjoyed very little' to 'enjoyed very much'. |

Ten Empatica E4 wristbands (Empatica, 2015; Garbarino, Lai, Bender, Picard, Tognetti, 2014; see figure 1a) were used to collect the physiological variables of heart rate (HR) and electrodermal activity (EDA) from the audience. HR data was collected at 1 Hz and EDA was collected at 4 Hz. Together HR and EDA can be used as direct measures of physiological arousal and are regularly interpreted as indirect measures of spectator engagement (Potter & Balls, 2012). Acceleration was also collected by the sensors at a rate of 32 Hz but was only used to align the start and end times of the datasets, from the sensors and was not used in further analyses.

*Qualitative analysis of spectator engagement*

Using the same tablet computers, we also collected qualitative data at the end of the experiment. The open question asked participants to 'Please use the space below to leave any comments that you have about the performance (e.g. what you liked/disliked the most, something strange you noticed etc) or your experience during the session (e.g. if you couldn’t hear the performance or you were distracted)'; and a blank box was provided to allow participants to write whatever they wished. The spectators comments were analysed using summative content analysis (Hsieh and Shannon, 2005).

Experimental Procedure

All three dance videos were shown in a large dimly lit lecture theatre with tiered seating in order to approximate the context of a live performance in a theatre setting (Jola & Christensen, 2015). All members of each experimental group watched the recorded dance performance together, and each group only watched one dance performance. The three conditions consisted of the dance choreography described above with either the congruent soundtrack in which the sounds and movements were not altered, the incongruent soundtrack or no soundtrack. As participants arrived they were shown to a seat and asked to complete the written consent form, the Gold-MSI (Müllensiefen et.al., 2014) and the dance experience questionnaire. Participants were seated adjacent to each other in rows of 3 or 4 participants. Once seated, participants were equipped with a wrist sensor on their left wrist and instructed to keep that arm as still as possible to minimise movement artifacts (Picard, Fedor, & Ayzenberg, 2015). Next participants were each given an ASUS tablet running the finger-tracking application (Open sesame) to record continuous aesthetic ratings. Participants were instructed to move their finger towards the top of the screen if they were enjoying the performance, and to move their finger towards the bottom of the screen if they were not enjoying their performance. Participants were given an opportunity to practice using the continuous rating scale in a demo version of the experiment, directly before the performance started. Following a short practice session using the app, participants were asked to focus on the performance and avoid looking at the tablet. Then, participants were asked to read a piece of information about the performance and choreographic style, written by the choreographer before watching the video stimuli to increase their understanding of the piece. The information about the dance film outlined that performers decided what movements to make themselves according to a set of instructions given by the choreographer, producing a unique combination of movements each time. Participants then watched the entire performance video without further interruption. As soon as the performance ended all participants answered a written open-ended question assessing their summative aesthetic judgement. The open question asked participants to 'Please use the space below to leave any comments that you have about the performance (e.g. what you liked/disliked the most, something strange you noticed etc) or your experience during the session (e.g. if you couldn’t hear the performance or you were distracted)'; and a blank box was provided to allow participants to write whatever they wished**.** Participants were debriefed at the end of the experimental session.

*Time series analyses*

**Temporal alignment:** From each participant, we recorded three time series, continuous enjoyment ratings, HR, and EDA. To align time series data across participants, all sensors were shaken at the beginning of the experiment. The resulting acceleration signal was then used as the reference signal. MATLAB was then used to count the number of UNIX time points between the reference signal and the timestamp that indicated the start of the performance, and each participants data was shifted accordingly. The same process was applied to HR, EDA, and enjoyment data. The number of points was adjusted to take into account the sampling rate of each variable, since HR data was collected at 1 sample per second and EDA was collected at 4 samples per second, EDA was shifted by four times more points than HR.

**Downsampling**: Granger causality analysis requires stationary time series with identical sampling rates. All time series were therefore resampled at 2 Hz, assuming that aesthetic judgements and physiological responses respond within seconds. The two second time-frame is consistent with information sampling periods in the context of aesthetic appreciation of music (Lindsen, Moonga, Shimojo, & Bhattacharya, 2011). The collected time-points were sorted into 1018 bins for each variable and the average of the data points in each bin was used for each time-point, (see Appendix for corresponding MATLAB script). Finally, the first five minutes of the performance were removed to allow for the audience to settle, and the last 5 minutes were removed as this is when the performers started leaving the stage, preventing a computation of synchrony across all performers (Vicary et al., 2017). This corresponded to the first and last 150 data points. In the end 718 data points for each variable, for each participant were included, corresponding to 24 minutes of dance performance at a rate of one sample every 2 seconds, for the three dependent variables of enjoyment, EDA and HR.

**Granger Causality analyses:** A variable X Granger causes Y if past values of X help to predict Y, above and beyond the information contained in Y alone (Seth, 2007, Granger, 1969). Unlike standard correlations or regression analysis, granger causality is more robust as it controls for the risk of auto-correlation, and allows to assess the temporal order of observed effects, that is whether changes in Y precede or follow changes in X. Granger causality was calculated using a bivariate Granger causality R software package, which mathematically accounts for potential auto correlation (Wessa, 2013). An ARX model was used with features of the video treated as the exogenous predictors, indicating that the features of the videos could influence physiological arousal or aesthetic judgement in a one-way interaction. A one-way ARX model was considered appropriate in this context because only one interaction was possible; the features of the video can impact audience responses, but audience responses cannot impact video features (Vicary et al., 2017). In order to meet the necessary assumption of stationarity and linearity, each dataset was plotted to check for stationarity of the mean. Where an upward or downward trend was observed, datasets were transformed using Box-Cox transformations based on Box-Cox normality plots for each dataset (Wessa, 2016). Following this, the stationarity of the mean level was checked using the autocorrelation function (ACF), using 1.96 standard deviations from the correlation co-efficient as the threshold for stationarity (Dean & Dunsmuir, 2016). Where data were observed to be stationary no differencing was performed. Where a linear trend was observed, first order differencing was implemented, and where a quadratic trend was observed, second order differencing was implemented according to the differencing equation by Box and Jenkins (1976). In line with Vicary and colleagues (2017), we computed uni-directional granger causality at temporal lags between 2 – 10s. For each condition, bivariate granger causality analyses were conducted to see if performed synchrony, visual change predicted HR, EDA, or Enjoyment. Additionally, for the congruent and incongruent sound conditions, granger causality analysis was used to see if the spectral centroid of the soundtrack predicted HR, EDA or enjoyment.

Results

We report analyses of overall aesthetic judgements and spectator agreement, followed by granger causality analyses and qualitative analyses of the open-ended questionnaire.

**Enjoyment**

On average, people disliked the congruent and silent conditions which had mean enjoyment scores of -79.27 (SD = 46.75) and -75.45 (SD = 46.93), respectively. However, aesthetic judgements were on average neutral in the incongruent condition which had a mean enjoyment score of -4.6 (SD = 1.18), with participants spending more time liking the video with a reversed soundtrack. Level of enjoyment was demonstrated using one sample *t*-tests, which indicated that group enjoyment scores were significantly different from 0 for the Congruent sound t(717) = -45.44; *p* < .001, 95% CI [-79.27, -82.69] ; Incongruent *t*(717) = -3.91; *p* < .001, 95% CI [-4.91, - 6.91], and Silent conditions t(717) = -43.08; *p* < .001, 95% CI [-75.45, -78.89]. To assess overall differences between performances, the mean enjoyment ratings of all three performances were compared collapsed across time. A one-way ANOVA showed a main effect of sound condition on enjoyment, *F* (2, 2151) = 707.8, *p* < 0.001, *η2* = .397). Post-hoc Games-Howell tests revealed that enjoyment ratings were significantly higher for the Incongruent condition compared to the Congruent condition (*p* < .001, 95% CI [-79, -69]), and the silent condition (*p* < .001, 95% CI [65.42, 76.09]) No difference found between the congruent and silent conditions (*p* = .202).

**Agreement analysis**

An agreement analysis was performed on the enjoyment data as outlined by Schubert, Vincs, and Stevens (2013). This involved computing the mean (M1) and standard deviations (SD1) of group enjoyment ratings for each time-point. Then the mean and standard deviation of SD1 was calculated which resulted in SD2. Higher SD2 values indicate less agreement in the group. The cutoff value was adjusted to 0.5\*SD2 above M1, and values below this level were considered points of good agreement. This analysis demonstrated that 67%-72% of all data points occurred within 0.5 standard deviations above the mean of all enjoyment values (Congruent: 71.71%; incongruent 67.41%; silent: 68.52%, see figure 2). Overall these results show that there was good agreement between audience members in each sound condition, and that the incongruent sound condition was rated as least unpleasant compared to congruent and silent conditions.

|  |
| --- |
|  |
| Figure 2. Results of Agreement analysis. Mean enjoyment shown in red. Good agreement shown in blue; good agreement defined in text. The blue line represents 67.4% of time points in the Incongruent condition, 71.1% of time points in the Congruent condition and 68.52% of time points in the Silent condition. It also indicates the specific time periods in each condition that the audience had good agreement. The overall mean enjoyment for the incongruent condition was -4.6 (SD = 1.18), the congruent condition was -79.27 (SD = 46.75) and the silent condition was (SD = 46.93). |

**Granger Causality Analysis**

As our hypothesis only focused on the impact of video features on immediate audience enjoyment responses, we examined lags of +10 seconds. Additionally, the presence of spurious relationships in the data was controlled for by calculating granger causality between performed synchrony and spectator time series of the same performance and comparing these to GC relationships derived from a mismatched performance. The mismatched performance data was derived from a previous study (Vicary et. al. 2017; synchrony of performance 1 or P1) which included three additional performances to the recorded performance displayed in the current study, and accordingly corresponding time series datasets. As in our previous study, GC results are only interpreted if they are ‘*performance-specific*’. GC analysis indicated that performer synchrony granger caused EDA, for the incongruent sound condition (F = 4.37, *p* = 0.0369; see figure 4), but not in the silent or congruent conditions. This relationship was both unidirectional and performance specific (all *p* > .05). No other significant granger causal relationships were found between any the video features, movement of the performer, or spectral power of the soundtrack and enjoyment, EDA or HR. Thus, the only significant effect found in the Granger causality analysis, was that performed synchrony granger-caused EDA in the incongruent movement and sound condition.

|  |
| --- |
| A screenshot of a cell phone  Description automatically generated |
| Figure 4. Granger causality results for performed synchrony in each soundtrack condition. Demonstrating significant p < 0.05 for the relationship between performed synchrony and electrodermal activity (EDA) in the incongruent condition at lag 1. Also showing the non-significant p > 0.05 for heart rate in each condition and the relationship between synchrony and EDA in the congruent and silent conditions. The dashed line indicates the significance level of p = 0.05. |

**Open-ended questionnaire**

After watching the dance performance participants were given an optional written questionnaire that asked to ‘please make a comment’, and a majority or participants decided to do so. Although comments tended to be brief, a summative content analysis was performed (see figure 3) in line with recommendations from Hseih and Shannon (2005). Several categories were repeated across comments including;

**Repetitive** e.g. P2: ‘Boring. I already have a low attention span. I would never be entertained by something that wasn’t more stimulating.’

**Searching for meaning**, e.g. P3: ‘I really enjoyed trying to figure out who was the leader/ who was following who / what the ‘rules’ were.’

**Physical movement**, e.g. P9: ‘Liked it the most when there was a lot of movement & when the movement contained a pattern I could decipher. flying guy was funny. really disliked the arm clapping making it seems their arms were broken.

**Sound**, e.g. P13: ‘Not hearing even the movements of the dancers was sort of distorting the ‘experience’ of watching this performance: hearing bodies moving, even if there is no music is part of dance and rhythm and I really missed this part.’  
P25: ‘The sounds were kind of unsettling, but occasionally made the performance more interesting, especially when it was congruent with the dancer’s movements.’ \*Notably, this participant was in the incongruent group, where the soundtrack and visual stimuli were intended to be incongruent throughout the entire performance.

**Patterns**, e.g. P8: ‘Spent time looking for patterns or beat couldn’t find either.

**Enjoyable**, e.g. P19: ‘I liked when the performers were synchronising, but when they weren’t I felt confused.’

**Unpleasant,** e.g. P21: ‘The noise was unpleasant and distracting. The performance at the end was more interesting. The middle bit, when the noise bits were increased was unbearable!’

|  |
| --- |
| A screenshot of a cell phone  Description automatically generated |
| Figure 3. Qualitative themes. Themes identified in participant’s responses to open-ended question, for each experimental condition of congruent, silent, and incongruent |

Discussion

The aim of this study was to explore how changes in the temporal congruency between movement and sound impact aesthetic judgements of recorded contemporary dance, in relation to the congruence association model (CAM; Cohen & Siau, 2008; Cohen, 2013a; 2013b). A secondary aim was to examine whether any change in aesthetic judgement was reflected in spectator arousal. Audience enjoyment was least negative for the incongruent condition, the only experimental conditions that also elicited an EDA response to the synchrony of the performer’s movements. Importantly, sound itself was *not* predictive of spectator arousal, implying that its influence on arousal was indirect, emphasizing or drawing attention to specific aspects of the dance. In contrast, the congruent and silent conditions which were judged more negatively did not induce any changes in spectator arousal. Thus, group enjoyment increased when the sounds were no longer coupled in time to the actions that produced these sounds. In line with these quantitative findings, qualitative participant reports suggest that participants in the incongruent condition found the presentation less boring than participants in the congruent and silent conditions. It is also worth noting that overall high levels of agreement were found for each audience group, with slightly higher agreement observed in the congruent condition than in the incongruent condition. This indicates that decoupling sound from movement was not only perceived as more enjoyable but might suggest that audience responses became more varied as the unequivocal causal link between sound and movement was replaced by the individual spectator’s idiosyncratic associations between auditory and visual input. Variability of aesthetic judgement increased with greater ambiguity.

Spectators did not enjoy the congruent condition, in line with Reason et al. (2016) who found that participants did not enjoy the sounds of the performer’s movements in the absence of music and even reported a sense of relief when this section of the performance ended. Our findings show that discomfort does not necessarily result from the sounds per se, but rather from the sounds being a direct consequence of the performers’ movements. In our study, when the same sounds were temporally decoupled from their movement origin, enjoyment ratings improved. Perhaps, removing the causal, embodied relationship between the performer’s actions and their auditory consequences disrupted the role of ‘kinaesthetic empathy’ on aesthetic judgements. This would suggest that the auditory channel was prioritized over the 'kinaesthetic channel' as mentioned in recent versions of the CAM, in this particular study. This might suggest hierarchical layers between each channel of audiovisual engagement, rather than parallel processing (Cohen, 2016). By removing the causal link between the performer’s actions, the corresponding sounds for each movement were turned into an independent soundtrack that invited spectators to actively search for correspondences between movement and sound. Similar to dancing to music, the performers movements do not produce the movements, which seem to align with the incongruent condition. Our findings suggest that non-expert spectators of dance typically expect musical accompaniment, or at least an arbitrary relationship between movement and sound that allows them to search for interesting correspondences and conflicts between auditory and visual content and structure, structuring their attention within the visual continuous stream of dance movement (Cook, 2000). It would be interesting to test whether our findings translate to Tap or Irish dance, where the rhythmical sound of the dancer’s steps are an intrinsic feature of the dance style. Further research in these different contexts may help to resolve these questions.

This study provides tentative evidence that in the context of contemporary dance, perfect congruency between sound and movement may not be desirable, in line with predictions by Fogelsanger and Afanador (2006). However, in line with original models of the CAM, it does seem that this change in aesthetic judgement was not simply a direct response to the soundtrack but occurred as a result of a different viewing pattern. This corresponds with the finding that musical meter influence or disrupt the perceived dance meter (Lee, et. al., 2015). The less negative aesthetic judgements in the incongruent condition coincided with an EDA response that was not predicted by the soundtrack, but by the visual movements of the performers, despite the fact that the visual movements were identical in all three performances. Therefore, this study lends support to the role of 'aesthetic capture' in contemporary dance, where the music drives visual attention to different physical gestures (Fogelsanger & Afanador, 2006; Mitchell & Gallaher, 2001). Additionally, this finding suggests that potentially conflicting media, that do not share the same underlying narrative can be perceived as complementary in a contemporary dance performance, rather than being perceived as conflicting, contradicting or in contest (Fogelsanger & Afanador, 2006; Cook, 2000). Conceivably, perfect audiovisual congruence conveys a relationship between sound and movement that is too simple, highlighting the importance of ambiguity and ‘remoteness from the habitual’ in contemporary dance appreciation (Kreitler & Kreitler, 1972) compared to other audiovisual contexts. In this particular study, reversing audio-visual input may have resulted in analogous, accidental audiovisual gestalts; paradoxically heightening the experience of congruency. It seems that in practice, multi-modal incongruence is difficult to achieve because each instance of coincident movement and sound can be perceived as a perceptual unity, in line with gestalt principles of perception (Sloboda, 1985). In fact, one participant explicitly reported a sense of congruency in the incongruent condition, and a similar outcome has been reported in previous studies (Bolivar, Cohen & Fentress 1994; Iwamiya, 1994; Lipscomb & Kendall 1994). Perceived matching between music and dances, not intended to go together (Mitchell & Gallagher, 2001) suggests that discovering patterns is an important feature of aesthetic appreciation (Muth & Carbon, 2013). Dance and music may benefit from being relative abstract, allowing people to project their own interpretation or meaning onto the performance.

The main strength of this study is that it examines the empirical evidence for extension of the CAM to contemporary dance settings, which until now was not empirically investigated (Cohen, 2016). This study also successfully employed advanced methodological approaches to the study of aesthetic responses to dance, using continuous enjoyment ratings as well as self-report accounts which gives a more complete picture of enjoyment ratings (Schubert et. al., 2013), compared to previous studies that tend to solely rely on post-hoc accounts (Kirsch, Drommelschmidt, & Cross, 2013; Reason et.al., 2016;). Additionally, this study maximized ecological validity by using a full-length dance designed by a choreographer and scientific rigour, by using a recorded version which guaranteed that all participants were exposed to identical dance movements (Jola & Christensen, 2015).

However, despite these strengths, the small sample size limits the generalizability of our findings. While the sample in our study was comparable in size to those used in the published literature (Cross, Kirsch, Ticini, & Schütz-Bosbach, 2011; Reason et.al. 2016;) a replication of our study with a larger sample size is necessary to confirm that structural incongruence between sound and movements is preferred to structural congruency. Another point to note in our results is that aesthetic appreciation was more often negative than positive. This may result from the lack of an immersive live performance context which has been shown to reduce aesthetic responsiveness to watching dance (Jola & Grosbras, 2013), or indeed it may have arisen from the experimental nature of the performance which does not reflect mainstream styles of dance. Perhaps future studies should include a wider range of spectator experiences, for example the Sense of Presence Inventory questionnaire (Lessiter, Freeman, Keogh, & Davidoff, 2001), to deduce the audience level of immersion and how this impacts on aesthetic engagement. Finally, we deliberately focused our analyses on the performance in its entirety rather than specific sections of the performance. Future studies should explicitly manipulate the relationship between audiovisual congruency, specific movement vocabulary and dramaturgy of the dance performance.

To conclude, the current study provides tentative evidence that cross-modal incongruence can be perceived as aesthetically pleasant, increasing ambiguity and thereby idiosyncratic associations. Our findings are consistent with the idea of aesthetic capture in the audiovisual context of contemporary dance, where the music drives visual attention to different physical gestures (Fogelsanger & Afanador, 2006; Mitchell & Gallaher, 2001). However, replication and extension of our findings to larger samples and using a variety of dance and music styles would be necessary before firmer conclusions can be drawn.

References

Arnheim, R. (1974). *Art and visual perception*. University of California Press.

Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology: A Journal of Research in Music Cognition*, *13*(1-2), 28.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control, revised ed*. Holden-Day.

Chion, M. (1994). *Audio-Vision: Sound on Screen,* ed. and trans. Claudia Gorbman. New York: Columbia University Press.

Cohen, A. J. (2013a).  Congruence-association model of music and multimedia: Origin and evolution. In Tan, Cohen, Lipscomb, & Kendal (Eds.), The Psychology of music in multimedia. Oxford, UK: Oxford University Press.

Cohen, A. J.  (2013b). Film music and the unfolding narrative.  In M. A. Arbib (Ed.).  Language, music and the brain: Strungmann Forum Reports (pp. 173-201). Cambridge, MA: Cambridge

Cohen, A. J. (2016). Music in performance arts: film, theatre, and dance. In S. Hallam, I. Cross, M. Thaut (Eds.), *The Oxford handbook of music psychology 2nd Ed*. (pp. 725-743). Oxford: Oxford University Press.

Cohen, A. J., MacMillan, K., & Drew, R. (2006). The role of music, sound effects & speech on absorption in a film: The congruence-associationist model of media cognition. *Canadian acoustics*, *34*(3), 40-41.

Cohen, A. J., & Siau, Y. M. (2008, August). The Narrative Role of Music in Multimedia Presentations: The Congruence-Association Model (CAM) of Music and Multimedia. In *Proceedings of the 10th International Conference on Music Perception & Cognition (ICMPC10)*.

Cook, N. (2000).  Analysing musical multimedia. Oxford, UK: Oxford University Press.

Cross, E. S., Kirsch, L., Ticini, L. F., Schütz-Bosbach, S. (2011). The impact of aesthetic evaluation and physical ability on dance perception. *Frontiers in human neuroscience*, *5*, 102.

Cunningham, M. (1982). A collaborative process between music and dance+ a selection of works by Cage, John and Cunningham, Merce. *TriQuarterly*, *54*, 173-186.

Cunningham, M., & Lesschaeve, J. (1985). *The dancer and the dance* (Vol. 1). Marion Boyars Publishers: London

Dean, R. T., & Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review*, *5*(4), 152–175. Retrieved from https://kb.osu.edu/dspace/handle/1811/48550

Dean, R. T., & Dunsmuir, W. T. M. (2015). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods*, *48*(2), 783–802. https://doi.org/10.3758/s13428-015-0611-2

De Keersmaeker, A. T., & Cvejic, B. (2014). *A Choreographer's Score: Drumming and Rain*. Mercatorfonds.

Dempster, E. (2008). The Choreography of the Pedestrian. *Performance Research*, *13*(1), 23–28. https://doi.org/10.1080/13528160802465458

Empatica (2015). *E4 wristband user’s manual*. <https://empatica.app.box.com/v/e4-user-manual>. doi: 10.4108/icst.mobihealth.2014.257418

Fogelsanger, A., Afanador, K. (2006). Parameters of perception: Vision, audition, and twentieth-century music and dance. *Paper presented at the 38th Congress on Research in Dance*. Tempe, Arizona.

Garbarino, M., Lai, M., Bender, D., Picard, RW., Tognetti, S. (2014). Empatica E3 – A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. *International Conference on Wireless Communication and Healthcare,* 39-42

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424. https://doi.org/10.2307/1912791

Hagendoorn, I. (2004). Some speculative hypotheses about the nature and perception of dance and choreography. *Journal of Consciousness Studies*, *11*(3-4), 79-110.

Hinberg, T. (2017). Group Flow. In M. Lesafffre, P.J. Maes, M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 133-140). Routledge: Cornwall.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, *15*(9), 1277-1288.

Iwamiya, S. (1994). *Interactions between Auditory and Visual Processing when Listening to Music in an Audio Visual Context: 1. Matching, 2. Audio Quality.* *Psychomusicology 13*, 133-154

Jola, C., Christensen, J. F. (2015). Towards ecological validity in the research on cognitive and neural processes involved in dance appreciation. In M. Nadal, J.P. Huston, L., Agnati, F. Mora, C.J. Cela-Conde (Eds.), *Art, aesthetics and the brain* (pp. 223-262). New York: Oxford University Press.

Jola, C., & Grosbras, M. H. (2013). In the here and now: Enhanced motor corticospinal excitability in novices when watching live compared to video recorded dance. *Cognitive Neuroscience, 4*(2), 90-98.

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*, 718.

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

Kirsch, L., Drommelschmidt, K. A., & Cross, E. S. (2013). The impact of sensorimotor experience on affective evaluation of dance. *Frontiers in human neuroscience*, *7*, 521, 1–10. https://doi.org/10.3389/fnhum.2013.00521

Kreitler, H., & Kreitler, S. (1972). *Psychology of the arts*. Durham, N.C: Duke University Press

Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae*, *1*(1), 63-85.

Lartillot, O., Toiviainen, P., Eerola, T. (2008). A Matlab Toolbox for Music Information Retrieval, in C. Preisach, H. Burkhardt, L. Schmidt- Thieme, R. Decker (Eds.), *Data Analysis, Machine Learning and Applications: Studies in Classification, Data Analysis, and Knowledge Organization*. Springer-Verlag.

Lee, K. M., Barrett, K. C., Kim, Y., Lim, Y., Lee, K. (2015). Dance and music in “gangnam style”: how dance observation affects meter perception. *PloS one*, *10*(8), e0134725.

Lehmann, H. T. (2006). *Postdramatic theatre*. Routledge.

Leman, M., Lesaffre, M., Maes, P.J. (2017). What is embodied music interaction? In M. Lesafffre, P.J. Maes, M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 1-10). Routledge: Cornwall.

Lessiter, J., Freeman, J., Keogh, E., Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence*, *10*(3), 282-297.

Lindsen, J. P., Moonga, G., Shimojo, S., & Bhattacharya, J. (2011). Swayed by the music: Sampling bias towards musical preference distinguishes like from dislike decisions. *Consciousness and Cognition*, *20*(4), 1781–1786. https://doi.org/10.1016/j.concog.2011.01.008

Lipscomb, S. D., Kendall, R. A. (1994). Perceptual Judgement of the Relationship between Musical and Visual Components in Film. *Psychomusicology 13*, 60-98.

Lipscomb, S., & Tulchinsky, D. (2004). The role of music communication in cinema.  In Miell, MacDonald & Hargreaves (Eds.)  Music Communication.  Oxford, UK: OUP.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. http://dx.doi.org/10.3758/s13428-011-0168-7

Mitchell, R. W., Gallagher, M. C. (2001). Embodying music: matching music and dance in memory. *Music Perception, 19*(1), 65-68.

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), e89642.

Muth, C., Carbon, C-C. (2013). The aesthetic aha: on the pleasure of having insights into Gestalt. *Acta Psychologica, 144,* 25-30.

Orgs, G., Caspersen, D., Haggard, P. (2016). You move, I watch, it matters: aesthetic communication in dance. In S. S. Obhi & E. S. Cross, (Eds). *Shared representations: sensorimotor foundations of social life.* Cambridge University Press : Cambridge.

Orgs, G., Dombrowski, J.-H., Heil, M., & Jansen-Osmann, P. (2008). Expertise in dance modulates alphabeta event-related desynchronization during action observation. *European Journal of Neuroscience*, *27*(12), 3380–3384. https://doi.org/10.1111/j.1460-9568.2008.06271.x

Orgs, G, Hagura, N., & Haggard, P. (2013). Learning to like it: Aesthetic perception of bodies, movements and choreographic structure. *Consciousness and Cognition*, *22*(2), 603–612. https://doi.org/10.1016/j.concog.2013.03.010

Picard, R. W., Fedor, S., Ayzenberg, Y. (2015). Multiple arousal theory and daily life elecrodermal activity asymmetry. *Emotion Review, 8*(1), 62-75.

Potter, R. F., & Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge.

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.

Schubert, E., Vincs, K., & Stevens, C. J. (2013). Identifying regions of good agreement among responders in engagement with a piece of live dance. *Empirical Studies of the Arts*, *31*(1), 1-20.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford : Oxford University Press.

Stevens, C. J., Schubert, E., Morris, R. H., Frear, M., Chen, J., Healey, S., Hansen, S. (2009). Cognition and the temporal arts: Investigating audience response to dance using PDAs that record continuous data during live performance. *International Journal of Human-Computer Studies*, *67*(9), 800-813.

Van Dyck, E., Burger, B., Orlandatou, K.(2017). The communication of emotion in dance. In M. Lesafffre, P.J. Maes, M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 122-130). Routledge : Cornwall.

Vicary, S., Sperling, M., von Zimmerman, J., Richardson, D. C., Orgs, G. (2017). Joint Action Aesthetics. *PLoS ONE, 12*(7): e0180101. https://doi.org/10.1371/journal.pone.0180101

Wessa P. (2013). Bivariate Granger Causality (v1.0.3) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL <http://www.wessa.net/rwasp_grangercausality.wasp/>