Examining the Correlation Between Dance and Electroacoustic Music Phrases: A Pilot Study

Andreas Bergsland
andreas.bergsland@ntnu.no
Norwegian University for Science and Technology, NTNU
Trondheim, Norway

Sanket Rajeev Sabharwal*
sabharwal@edu.unige.it
Casa Paganini
Genova, Italy

ABSTRACT

In this paper, we will present a pilot study that explores the relationship between music and movement in dance phrases spontaneously choreographed to follow phrases of electroacoustic music. Motion capture recordings from the dance phrases were analyzed to get measurements of contraction-expansion and kinematic features, and the temporal location of the peaks of these measurements was subsequently compared with the peaks of a set of audio features analyzed from the musical phrases. The analyses suggest that the dancers variably accentuate their movements to the peaks or accents in the music. The paper discusses the findings in addition to possible improvements of the research design in further studies.

CCS CONCEPTS

• **Applied computing → Sound and music computing**.

KEYWORDS

motion capture, movement features, music-movement relationship, salience, accents

ACM Reference Format:
Andreas Bergsland and Sanket Rajeev Sabharwal. 2023. Examining the Correlation Between Dance and Electroacoustic Music Phrases: A Pilot Study. In *Audio Mostly 2023 (AM '23), August 30–September 01, 2023, Edinburgh, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3616195.3616218

1 INTRODUCTION

In this pilot study, we aim to investigate the degree of synchronization between dance and music phrases focusing on the relationship between accents in the music and spontaneously choreographed dance. Our objective is to identify movement and audio parameters that can capture accents and allow us to measure the degree of temporal alignment of these. We recorded motion capture data of two dancers spontaneously performing dance phrases to electroacoustic music rich with accents but without a salient rhythm. Initial results are reported on examining the relationship between audio and movement data by comparing the peaks in the two datasets and use this to reflect on further investigations of the material.

*Both authors contributed equally in the writing of the paper.

AM '23, August 30–September 01, 2023, Edinburgh, United Kingdom

ACM ISBN 979-8-4007-0818-3/23/08.
https://doi.org/10.1145/3616195.3616218

2 BACKGROUND

Dance usually engages a temporally synchronized interplay between movement and music, despite some exceptions in contemporary dance like Merce Cunningham, who famously choreographed independently of music [12]. The nature of the relationship between dance, or more generally, movement, and music has been an object of several empirical studies in recent years.

A body of research has investigated the rhythmic interplay between dance and music. Particularly, Naveda and colleagues have conducted extensive studies investigating the relationships between dance and music in popular genres like Samba and Charleston [11]. These studies often focus on how standardized and repetitive patterns are related to music. Similarly, Burger et al. [1] found correlations between rhythmical pulses and whole-body movements, as well as between spectral flux and percussiveness, and specific body parts.

There seem to be fewer studies involving music without periodic rhythmical patterns. Casciato and his team's work reported similarities in movement density and gestural qualities with musical features when dancers moved spontaneously to music without any salient rhythm [3]. Furthermore, they reported that quantity of motion (QoM) decreased along with spectral brightness and energy. Sound tracing studies have also shown how rising pitch and spectral centroid are correlated to upward as well as downward movement, and impulsive sounds would most often make the subjects produce an accentuated attack (=high acceleration value in the beginning of the sound), in other words, an accent [15]. This is also following Mora and Pellicer's definition of movement accents as the peak of the magnitude of the kinematic acceleration vector [13].

This brings us to the focus of our study, namely accents, which are temporal events of relatively short duration with a high degree of salience that occur both in music and dance [6, 8]. Thoresen notes that accents in music stand out from the context they appear in "most often by being louder; having a more ample mass, a brighter spectrum or a sharper onset quality; and/or being longer than other elements in their context" [18]. Similarly, Jordan remarks how accents in dance refer to "physical movements at these particular points [that] are particularly powerful (or, [...] 'salient') within their context" [8].

3 METHOD

We enlisted two female dancers from the Dance, Choreography MA Program at Zurich University of the Arts (ZHdK), selected via open call and rewarded with 30 CHF gift cards to a local ballet shop. Individual motion capture recordings were performed at at the Immersive Arts Space (ZHdK) with an OptiTrack system, featuring 40 Prime 17W cameras and Motive 2.2 software, complemented by Canon Legria HF G50 video footage. The dancers wore full-body suits with 37 markers adhering to Motive's Standard skeleton template. The selection of female dancers was due to availability, not to answer a gender-specific research question.

3.1 Tasks and selection of musical phrases

Prior to the test itself, the first author, an expert in electroacoustic music analysis, selected 15 phrases with a duration of 7 to 15 seconds based on criteria including manageable duration and complexity for choreography, and a range of different beginnings and endings. Following a pilot test with a single dancer, six phrases were removed due to excessive complexity and difficulties synchronizing with the music. To facilitate synchronization, the remaining 9 phrases were gently manipulated to impose an audible pulse in the music.

In the main test, each dancer's task was to individually create and perform short dance phrases that closely followed a set of electroacoustic music phrases. The dancer first had to listen to all musical phrases before the choreography phase. The remainder of the session was split into nine parts - one for each phrase. During each part, the dancer created and tested out choreographic ideas while listening to the phrase multiple times. Once ready, the mocap recording began after which the dancer performed the choreography four times. After each performance, the dancer completed an oral survey that included a question about their preferred phrase.

3.2 Data cleaning and preparation

Before further calculation of all movement features except mocapgrams (see below), mocap markers with gaps of over 20 subsequent frames (1/6 sec.) were identified and removed. To fill gaps of less than 20 missing markers, a shape-preserving piecewise cubic spline interpolation algorithm from the *fillmissing* function in MATLAB was used. Moreover, one mocap recording also had to be discarded because it had missing data during the initial clap that provides the synchronization point between mocap, audio playback, and video. A few occurrences of low-amplitude flicker noise also made it necessary to use a Savitsky-Golay smoothing filter. Here, we used a 15-frame window to achieve a good balance between rapid response and curve smoothness. The mocap recordings were subsequently divided into four separate sections (takes), each preceded and followed by a few seconds of stillness.

3.3 Movement feature analysis

Standard kinematic features (velocity, acceleration, and jerk) and their vector norms were calculated from gap-filled and smoothed mocap data, using Savitzky-Golay filtering. Mocapgrams, a mocap visualization technique, were plotted for all kinematic features, serving as a valuable tool for a qualitative inspection of the data (see Fig. 1)[14, 15]. After a qualitative evaluation, kinematic features were calculated separately for seven body regions (head, upper torso, right arm, left arm, lower torso, right leg, and left leg), and means for the kinematic vector norms.

By drawing upon a previous study [20] and carefully assessing the initial qualitative visual representation of the dancers' movements in the mocapgram, we decided to focus our initial studies on the contraction-expansion index (CEI). It evaluates how a dancer's body moves through space during a dance phrase by measuring

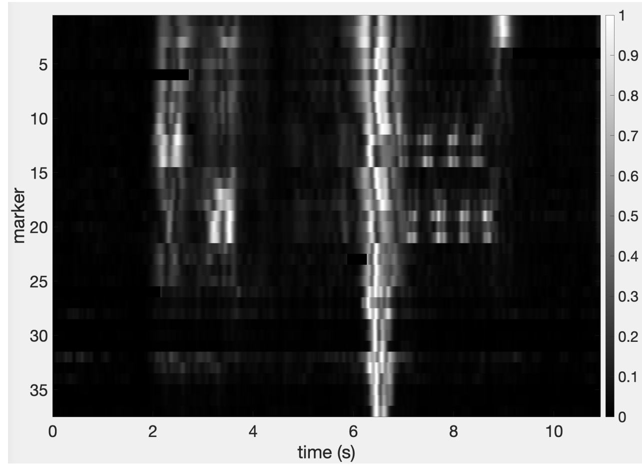

Figure 1: Mocapgram showing norm acceleration (m/s) for markers, normalized to 0-1 and shown as gray scale: Head 1-3; upper torso 4-9, 15-16; right arm 10-14; left arm 17-21, lower torso 22-25, right leg 26-31, left leg 32-37.

the extent to which their limbs and torso enlarge or reduce in size relative to their barycenter. The barycenter, or center of mass, is a geometrical point around which the mass of a body or system is distributed and is determined for each frame before the CEI was calculated. We used four dorsal markers, two of which were positioned on the hips and two underneath the shoulders, to compute the mean coordinate as the barycenter. The distance between the barycenter and five markers on the dancer's extremities including the top of the head, two for each wrist (Left and Right), and two for each foot (Left and Right) was then measured. Given a dataset D with barycenter coordinates $B_x(j)$, $B_y(j)$, $B_z(j)$ for each frame j and extremity markers set E = HeadTop, LWristOut, RWristOut, LToeOut, RToeOut, we computed the distance $d_e(j)$ between the barycenter and each extremity marker e, where $e \in E$:

$$d_e(j) = \sqrt{(x_e(j) - B_x(j))^2 + (y_e(j) - B_y(j))^2 + (z_e(j) - B_z(j))^2} \tag{1}$$

The raw contraction-expansion index $CEI(j)$ is the sum of the distances $d_e(j)$ for all extremity markers e:

$$CEI(j) = \sum_{e \in E} d_e(j) \tag{2}$$

This number was then normalized to the range from 0 to 1, yielding the final CEI where a value close to 0 signifies contraction and 1 indicates expansion. Since both can be included in the articulation of accents, both the peaks (expansion) and troughs (contraction) were identified.

3.4 Audio feature analysis

All audio analyses were performed using the MIRToolbox 1.8.1 in Matlab [10]. The extracted features were RMS, spectral flux, brightness (amount of energy above 1.5 kHz), and novelty.[1] We then calculated the most prominent peaks for all features. The time values were converted into corresponding mocap frames to analyze the coincidence of movement feature peaks with audio features.

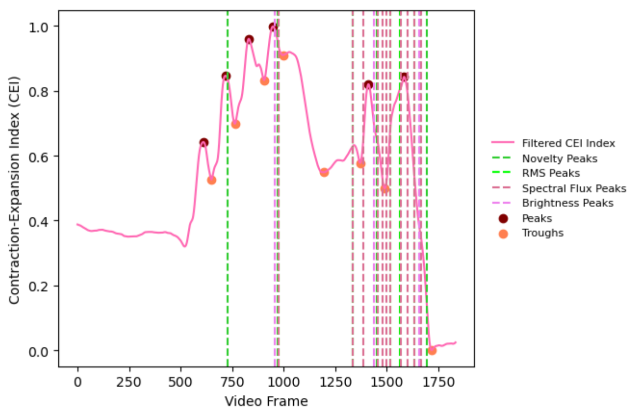

Figure 2: Plot indicating CEI with its peaks and troughs, and vertical lines for RMS, Novelty, Spectral Flux, and Brightness peaks for Phrase 3.

3.5 Movement-audio analysis

Since accents are brief events, a simple way to investigate the correspondence between music and dance is to compare the peak timings of the selected audio and movement features. Two peaks would likely be experienced as simultaneous if they fall within a so-called temporal binding window (TBW). Since the average adult TBWs lie around 160 ms for brief audio-visual stimuli like flashes and beeps [21], a simultaneity boundary of 166 ms (20 mocap frames) was chosen.[2] Since peak detection will affect the results directly, we have tuned our peak finding algorithms for both audio and movement features (*argrelextrema*, Scipy; *mirpeaks*, MIR Toolbox and *findpeaks* MATLAB) so that only the most prominent peaks will be identified [9, 19]. We analyzed a total of 14 takes in total across 5 different phrases since they fit our criteria of no missing data for the markers that were utilized to extract movement features.

4 RESULTS

To measure the degree of synchronization between accents in music and movement, we employed a quantitative measure known as the 'Jaccard Similarity Index', a method that calculates the degree of overlap between two sets. The index is determined by dividing the size of the intersection of the sets by the size of the union of the sets [16]. For our study, this methodology was applied to compute Jaccard Similarity percentages between the movement and audio feature peaks, where the criterion for similarity was that the peaks must reside within a TBW of 20 frames. Consequently, the percentages we derived represent the proportion of peaks in both music and dance that coincide within this defined TBW.

In Table 1, due to space constraints, we report a subset of these percentages which illustrate the varying degrees of coinciding movement features. This suggests a diverse distribution across different phrases and takes. Notably, certain features, such as Novelty and RMS, demonstrate higher alignment in certain scenarios, while others, like Brightness, exhibit lower synchronization. This implies the existence of underlying patterns and relationships between these features. Fig. 2 provides a visual illustration of the audio and movement features plotted together, where one can observe how the peaks and troughs follow, coincide with, or precede the audio features. This suggests a selective influence of the music on the dancers' movements, varying for different accents. Interestingly, for some phrases there is a larger difference, up to 50% between the minima and maxima for the peaks of kinematic features for separate body parts. For several phrases, this coincides with salient differences between the body parts from our qualitative evaluations.

We provide a summary statistics of the percentage of coinciding peaks and troughs (see Table 2) where the median values for coinciding acceleration peaks range from 4.55 to 26, while the mean values range from 8.86 to 25.38 with standard deviations between 7.16 and 10.41. The coinciding CEI peaks and troughs exhibit lower median and mean values, suggesting less alignment in these features across the dataset. These insights aid in understanding the relationships between music and movement among dancers, and provide a foundation for selecting the most suitable features for future research endeavors.

5 DISCUSSION, CONCLUSION AND FUTURE WORK

Our findings revealed a diverse distribution of alignment across different phrases and takes, as well as among different features. The results suggest that individual preferences play a significant role in choreography and the relationship between audio features and movement patterns is dynamic. In addition, they highlight the intricate interplay between auditory stimuli and human kinematics and provide some understanding of the factors influencing audio-visual spatial synchronization among dancers. These findings are essential for expanding our knowledge of the interplay between auditory stimuli and human kinematics, with the potential to impact fields such as dance pedagogy, performance analysis, and human-computer interaction in general.

While our initial analysis provides insights, these are only preliminary results which going forward will be augmented by broadening our dataset and performing further statistical tests to validate our findings. This could permit us to perform more sophisticated statistical tests such as autoregressive modeling for time-series [4, 5] and discretization of timestamps to correlate with expert annotations for salient moments in the dance phrases. Based on previous studies [2, 17] we also aim to investigate other movement features such as motion index, postural tension and symmetry, and intra-personal phase-synchrony. As we continue to explore these future research directions, we aim to progressively enhance our comprehension of the relationship between music and movements and gain a deeper

[1] Based on similarity matrix calculation [7] of Mel-frequency spectrogram. The novelty feature was calculated by cross-correlating local configurations along the diagonal of the similarity matrix with an ideal Gaussian checkerboard kernel with a kernel size of 400 samples.

[2] This value might be adjusted in future studies due to e.g. the asymmetry of visual- and audio-leading stimuli for the TBW.

Table 1: Jaccard similarity percentages between peaks in audio and movement analysis features

		% of coinciding accel. peaks				% of coinciding CE peaks				% of coinciding CE troughs			
Phrase & ID	Take	Novelty	RMS	SF	Brightness	Novelty	RMS	SF	Brightness	Novelty	RMS	SF	Brightness
1 ID2	1	9.1	23.1	30	9.1	0	0	0	0	0	0	0	50
2 ID1	1	20	27.3	25	0	33.33	0	20	0	0	25	0	0
3 ID1	1	0	0	13	0	40	0	0	33.33	0	0	9.09	0
6 ID1	3	20	37.5	37.5	12.5	20	33.33	33.33	33.33	20	0	0	0
9 ID2	2	17	16.7	25	20	0	11.11	20	0	0	0	0	0

Table 2: Summary statistics of Jaccard similarity percentages of coinciding peaks and troughs

	Acceleration peaks				CEI peaks				CEI troughs			
Particulars	Nov	RMS	SF	Brt	Nov	RMS	SF	Brt	Nov	RMS	SF	Brt
Median	18.50	21.55	26.00	4.55	0.00	0.00	0.00	0.00	0.00	0.00	9.09	0.00
Mean	16.44	22.12	25.38	8.86	10.36	1.02	8.48	14.29	10.00	9.27	7.40	7.14
Std. Deviation	7.16	10.41	8.22	10.17	17.37	3.82	11.26	20.52	15.57	12.28	7.73	26.73

understanding of perceived relevance and salience in music and dance.

ACKNOWLEDGMENTS

We thank Antonio Camurri, Gualtiero Volpe, Roberto Sagoleo and Eleonora Ceccaldi at Casa Paganini for their efforts, guidance, and support and Sébastien Schiesser for accomodating MoCap recordings in the at Zurich University of the Arts (ZHdK). This research has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101070278 (ReSilence STARTS project, ReSilence.eu) and Norwegian University of Science and Technology.

REFERENCES

[1] Birgitta Burger, Marc Thompson, Geoff Luck, Suvi Saarikallio, and Petri Toiviainen. 2013. Influences of Rhythm- and Timbre-Related Musical Features on Characteristics of Music-Induced Movement. *Frontiers in Psychology* 4 (2013), 1–10. https://doi.org/10.3389/fpsyg.2013.00183

[2] Antonio Camurri, Gualtiero Volpe, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Nicola Ferrari, and Corrado Canepa. 2016. The dancer in the eye: towards a multi-layered computational framework of qualities in movement. In *Proceedings of the 3rd International Symposium on Movement and Computing*. Association for Computing Machinery, ACM, Thessaloniki, Greece, 1–7.

[3] Carmine Casciato, Alexander Refsum Jensenius, and Marcelo M Wanderley. 2005. Studying free dance movement to music. In *ESCOM Performance Matters!*

[4] Roger Dean and Freya Bailes. 2010. Time Series Analysis as a Method to Examine Acoustical Influences on Real-time Perception of Music. *Empirical Musicology Review* 5 (10 2010). https://doi.org/10.18061/1811/48550

[5] Roger T. Dean and Freya Bailes. 2010. Time Series Analysis as a Method to Examine Acoustical Influences on Real-time Perception of Music. *Empirical Musicology Review* 5 (2010), 152–175.

[6] Carolyn Drake and Caroline Palmer. 1993. Accent structures in music performance. *Music perception* 10, 3 (1993), 343–378.

[7] Jonathan T Foote and Matthew L Cooper. 2003. Media segmentation using self-similarity decomposition. In *Storage and Retrieval for Media Databases 2003*, Vol. 5021. SPIE, 167–175.

[8] Stephanie Jordan. 2011. Choreomusical conversations: Facing a double challenge. *Dance Research Journal* 43, 1 (2011), 43–64.

[9] Olivier Lartillot and Petri Toiviainen. 2007. A Matlab toolbox for musical feature extraction from audio. In *International conference on digital audio effects*, Vol. 237. Bordeaux, 244.

[10] Olivier Lartillot, Petri Toiviainen, and Tuomas Eerola. 2008. *A Matlab Toolbox for Music Information Retrieval.*

[11] Marc Leman and Luiz Naveda. 2010. Basic Gestures as Spatiotemporal Reference Frames for Repetitive Dance/Music Patterns in Samba and Charleston. *Music Perception* 28, 1 (2010), 71–91. https://doi.org/10.1525/mp.2010.28.1.71

[12] Paul H Mason. 2012. Music, dance and the total art work: choreomusicology in theory and practice. *Research in Dance Education* 13, 1 (2012), 5–24. https://doi.org/10.1080/14647893.2011.651116

[13] Xavier Mora and Marta Pellicer. 2013. Understanding and measuring rhythmic quality in dance. What is a movement accent?. In *Proceedings of Bridges 2013: Mathematics, Music, Art, Architecture, Culture*, George Hart and Reza Sarhangi (Eds.). 111–118.

[14] Kristian Nymoen. 2012. *Methods and Technologies for Analysing Links Between Musical Sound and Body Motion.* Doctoral thesis. http://urn.nb.no/URN:NBN:no-33134

[15] Kristian Nymoen, Jim Torresen, Rolf Inge Godøy, and Alexander Refsum Jensenius. 2012. A Statistical Approach to Analyzing Sound Tracings. Springer Berlin Heidelberg, 120–145.

[16] Raimundo Real and Juan M. Vargas. 1996. The Probabilistic Basis of Jaccard's Index of Similarity. *Systematic Biology* 45, 3 (1996), 380–385. http://www.jstor.org/stable/2413572

[17] Sanket Rajeev Sabharwal, Manuel Varlet, Matthew Breaden, Gualtiero Volpe, Antonio Camurri, and Peter E. Keller. 2022. huSync - A Model and System for the Measure of Synchronization in Small Groups: A Case Study on Musical Joint Action. *IEEE Access* 10 (2022), 92357–92372. https://doi.org/10.1109/ACCESS.2022.3202959

[18] Lasse Thoresen. 2010. Form-Building Patterns and Metaphorical Meaning. *Organised Sound* 15, 02 (2010), 82–95. https://doi.org/10.1017/s1355771810000075

[19] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[20] Gualtiero Volpe. 2003. *Computational models of expressive gesture in multimedia systems.* Ph. D. Dissertation.

[21] Michael Zerr, Christina Freihorst, Helene Schütz, Christopher Sinke, Astrid Müller, Stefan Bleich, Thomas F. Münte, and Gregor R. Szycik. 2019. Brief Sensory Training Narrows the Temporal Binding Window and Enhances Long-Term Multimodal Speech Perception. *Frontiers in Psychology* 10 (2019). https://doi.org/10.3389/fpsyg.2019.02489