

Investigating Grouping Behaviour of Dancers in a Silent Disco Using Overhead Video Capture

Nelson Mooren^{*1}, John Ashley Burgoyne^{#2}, Henkjan Honing^{#3}

^{*}*Institute for Interdisciplinary Studies, University of Amsterdam, The Netherlands*

[#]*Music Cognition Group, Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands*

¹mooren.nelson@gmail.com, ²j.a.burgoyne@uva.nl, ³honing@uva.nl

ABSTRACT

In this study, we present a method for the processing and analysis of overhead video recordings of dancers in a silent disco. Silent-disco events present experimenters with an ecologically valid environment in which to study the interactions between music and the social environment in large groups of participants compared to laboratory settings. The toolbox developed so far allows us to investigate the grouping dynamics of dancers in a silent disco, and how these dynamics change over time. Specifically, we found that participants listening to earworms are more grouped than those listening to matched controls. Further development will allow for more detailed investigation of grouping, individual movement and switching behavior under various conditions.

I. INTRODUCTION

Making music and moving to music is very much a social experience. Evolutionarily, one theory by Freeman (2000) suggests that music as a social experience fosters group cohesion and social bonding through jointly synchronized movement. For example, it has been shown that people synchronize their movements both with rhythms generated by musicians or in coordination with other observers. Music education also plays an important role in the development of social skills, such as empathy (Kalliopuska & Ruókönen, 1986; Rabinowitch, Cross, & Burnard, 2012). For example, joint music making in children has been found to increase cooperative behaviour (Kirschner & Tomasello, 2010), and drumming with a social partner improves synchronisation accuracy in children (Kirschner & Tomasello, 2009). Taken together, these findings show that music fulfils a role in the development of social behaviour; Freeman (2000) considers music and dance to serve as ‘a technology of social bonding’ (Freeman, 2000, p. 411).

There also seems to be an intrinsic relationship between movement and music, as infants are able to induce rhythmic regularities from birth (Winkler, Háden, Ladinig, Sziller, & Honing, 2009), as well as move in time to such induced rhythms (Zentner & Eerola, 2010). The spontaneous movement to rhythms is also seen in adults (Madison, 2006; Zatorre, Chen, & Penhune, 2007). Entrainment plays an important role in linking the musical, rhythmic, movement to social behaviour. That is, moving in synchrony with a rhythmic pulse (temporal) or with other people (affective). The latter plays a role in-group dancing and is shown to induce pro-social feelings and behaviour (Phillips-Silver & Keller, 2012).

Research on the interaction of music and social behaviour so far has mainly consisted of laboratory experiments and (self-report) questionnaires (Kirschner & Tomasello, 2009; Hove, Spivey, & Krumhansl, 2010). The silent-disco set-up

has also been used in the laboratory research: Hadley, Tidhar, and Woolhouse (2012) showed that participants listening to the same channel are more engaged with each other than with participants listening to other channels. Similarly, Leman, Demey, Lesaffre, Noorden, and Moelants (2009) recreated a dance-club setting in a motion capture laboratory in order to investigate music-driven social interaction. While these studies allow experimenters to control the conditions of the environment carefully, they are limited in the number of participants interacting at once. Additionally, laboratory participants are faced with *demand characteristics* that are related to their awareness of the study’s true nature and to attitudes towards the experimenter (Nichols & Maner, 2008).

The current study provides an ecologically valid setting to study music-driven social interaction through the use of silent-disco events, with overhead video being recorded. This minimises the interaction between experimenter and participant, which minimises demand characteristics. For the participants, these events should primarily function as a fun night out with their peers. In this paper, we present a method to process and analyse the data acquired using this overhead video recording set up.

The first dataset was recorded at an event organised as part of the ESCOM 2015 conference, where participants were divided into one of two conditions: songs known to be earworms and matched controls. Earworms, defined by the Oxford English Dictionary as ‘*a catchy song or tune that runs continually through a person’s mind*’, are a widely experienced example of involuntary cognition (Beaman & Williams, 2010). Most research on earworms has concentrated on the phenomenology of the experiences through interviews, diary study and questionnaires (Beaman & Williams, 2010; Williamson & Jilka, 2014). The present study places these earworms in a social environment and investigates their influence on grouping behaviour. We hypothesise that the earworm group would show more grouping behaviour compared to the control, as measured through increased clustering coefficients. One possible rationale for this hypothesis is that these songs, by virtue of being earworms, could be considered more memorable and thereby facilitating participants’ forming social groups with other listeners.

We use a second dataset to test the pipeline, which consists of a standalone silent-disco event where participants were able to switch between channels. The channels in this event consisted of pop music from three time periods from the 1940s to present. We hypothesise grouping behaviour to increase with recency. One rationale based on personal observations is that this is an effect of age, as many participants appeared to be in their twenties and thirties. However, it is important to note that no demographic data has been obtained for the current events.

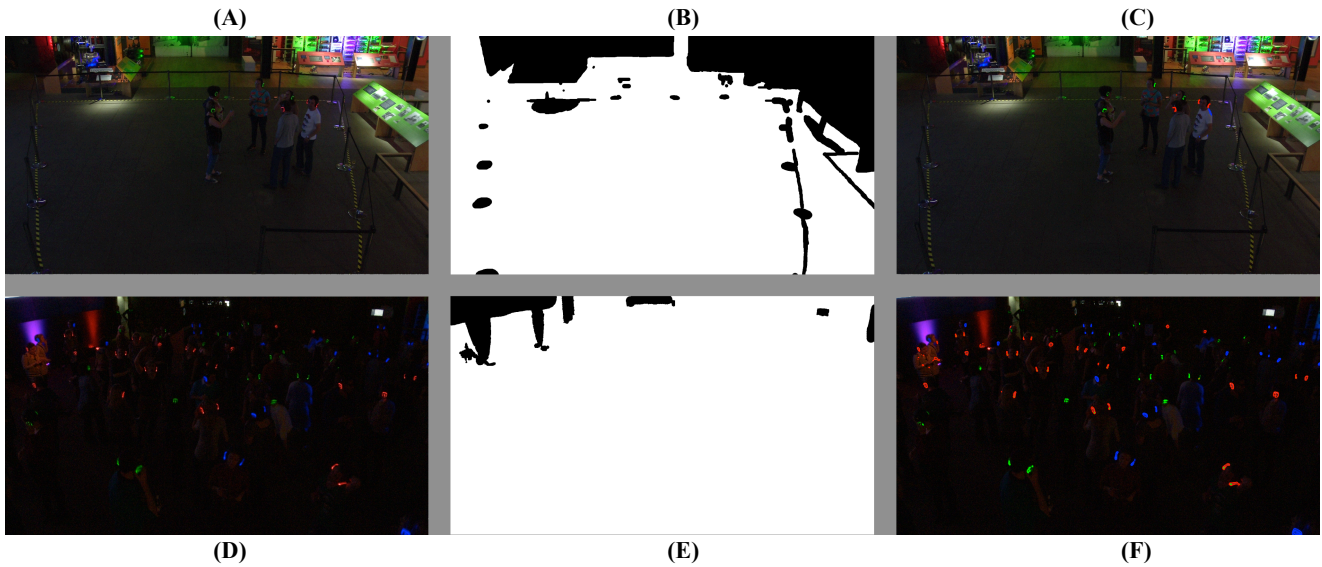


Figure 1. Example data for ESCOM (top row) and MOSI (bottom row) datasets; (A) and (D) example still of the dance floor, (B) and (E) dance floor mask with environment light blacked out, (C) and (F) example still with contours drawn.

II. METHODS

A. Data Acquisition

Data were acquired on two separate events, both being held at the Museum of Science and Industry (MOSI) in Manchester. Video was recorded using a camera suspended above the dance floor at an angle.

1) *ESCOM Dataset.* This event was organized as part of the evening programme of the ESCOM 2015 conference. This event consisted of one session, four hours in length. There were two channels: earworms (red) and a matched control (green). Participants were unable to switch between channels. Care was taken to make sure the dance floor was dark enough and a model headphone with large lights on the earpieces was used to make them easily distinguishable. Around the periphery, some light was present from exhibition pieces, but applying a mask before processing filtered this out.

2) *MOSI Dataset.* This event was a standalone silent disco, consisting of one session, approximately three hours in length. Here people could freely switch between three themed playlists: 1940s, 1950s, 1960s (blue); 1980s and 1990s (red); and post- 2000 (green). The same model of headphones was used, and extra care was taken with respect to ambient lighting. Again, there was some peripheral lighting present that required masking. The first half of this dataset had to be discarded because of camera movement (zooming and panning) during the first half of the evening, yielding one hour and 24 minutes of useful data.

B. Data Processing

Data processing consists of the following stages: mask creation, modeling of contours and centers, creating (undirected) graphs, extracting graph measures, statistical analysis. Figure 1 shows example stills of the dance floor and the first two processing steps. Image processing was done using the OpenCV library (Bradski, 2000), graph creation and processing is done using graph-tool module (Peixoto, 2014), both for Python 2.7.

1) *Mask Creation.* For each video, a mask is created to reduce noise from environmental features, such as spotlights, such that the masked regions will be ignored in further processing. Mask creation is done manually, with the following steps. First, for a number of frames, selected semi-randomly, the contours are modelled (following section), drawn on the frame. These frames are saved individually as images. Then, the contours that belong to environmental features are blacked out manually using a graphics editor, by tracing the outline of each contour and filling it in. The remainder of the canvas is left white. Then, while applying the mask, for a new set of frames the contours are modeled and the mask is updated. This process is repeated a number of times until no environmental features show up.

2) *Contour Modelling.* In order to extract the relevant data from each frame, contours describing the outline of visual features are modelled; in this case, these features consist of the lights on the headphones. Contours are modelled for each colour layer individually. Given that the video is shot in RGB and the headphone colours are red, green and blue, each colour layer can be processed separately. Video is sampled at a rate of one in every 10 frames (corresponding to 2.5 frames per second). First, the sampled images are binarized using Otsu's method for thresholding (Otsu, 1979) with Gaussian smoothing (kernel size 5). The mask is applied to these binarized images, and contours are extracted from them. Finally, the centres of the contours are determined; for each centre, the x -axis and y -axis coordinates, colour, and frame number are saved.

3) *Graph Creation.* Undirected graphs are created based on the centers of the modeled contours. These graphs provide snapshots of the dance floor at a time and describe how connected participants are. Per frame, a number of undirected graphs are created: one graph describing the entire dance-floor, thus containing all centers present in that frame; and one graph for each separate color present, thus with all nodes of a given color. The vertices are drawn based on the x -axis and y -axis location of the centers. Edges are drawn based on the

Euclidean distance between the vertices, using a threshold to set an upper limit. In other words, two nodes will be connected if their distance is less than the predefined threshold.

At first, graphs are created for a period of approximately 30 minutes in the middle of the event for a wide range of thresholds, between 50 and 900 points. For these data the clustering coefficients, average vector degree and number of vectors (following section) are calculated and plotted. Based on visual inspection of these plots three thresholds are selected for further analysis, namely 150, 200 and 250 points. Criteria used to determine the respective thresholds are the amount of noise as indicated by the size of standard deviations and the relative stability of the signal over time. For the ESCOM dataset, 150 points horizontally correspond to 60–120 cm at the near and far ends of the dance-floor, respectively; 150 points vertically correspond to roughly 120 cm at the near end of the dance-floor. For the MOSI dataset, the same distance thresholds are used, and these correspond to roughly the same distances.

4) *Graph Statistics*. For each graph the following statistics are determined: local clustering, global clustering, and average vertex degree. The local clustering coefficient is defined as follows by Watts and Strogatz (1998).

$$c_i = \frac{\text{existing connections between neighbours of } v_i}{\text{possible connections between neighbours of } v_i}$$

Thus, the local clustering coefficient c_i increases when the neighbours of vertex v_i are more interconnected, with completely interconnected neighbours of v_i resulting in $c_i = 1$. Then, this value is averaged over vertices:

$$c = \frac{1}{n} \sum_i c_i$$

Where local clustering takes the mean of a ratio, global clustering takes the reverse approach. Newman (2003) defines the global clustering coefficient as follows.

$$c = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}}$$

Here, a connected triple consists of a vertex connected to an (unordered) pair of other vertices, when the third edge is filled in a triangle is formed. Thus, a triangle consists of three triples, hence the multiplication factor.

Both clustering coefficients measure the density of triangles in the graph. However, local clustering gives a higher weight to low-degree vertices (Newman, 2003).

C. Statistical Analysis

A linear model is fitted for each of the three measures respectively, with group and time segment as independent variables. For the ESCOM data, *group* is a factor with two levels, corresponding to the green and red channels respectively, and *time segment* is a factor with 30 levels, corresponding to approximately eight minutes per segment. For the MOSI data, *group* is a factor with three levels, corresponding to the red, green and blue channels respectively, and *time segment* is a factor with 15 levels, corresponding to approximately six minutes per segment.

III. RESULTS

A. Video Processing

We created a toolbox for the processing and analysis of overhead video data of dancers in a silent disco. The toolbox yields consistent measures for both datasets; however, the measures are sensitive to the number of people on the dance-floor. When fewer than 10 participants per colour are present, the data seem to become less reliable since the connections between participants are less stable; this can be observed in figures 2A and 2C and seems to be unrelated to the threshold (not shown). The first hour of ESCOM data is unreliable because of the low number of participants present (for example see figures 2A and 2C). For the MOSI data, the end is particularly problematic, because in the last minutes there is a sudden jump in the number of vertices present (see figures 2B and 2D).

There is also a clear influence of the number of vertices detected and the measures themselves. Average vertex degree is especially susceptible to this effect; however, both local and global clustering measures are also related to the number of detected vertices.

B. Grouping Behaviour (ESCOM)

For the ESCOM dataset we created a linear model with a two-level factor for group and a 30-level factor for time segments; then we performed an ANOVA on and calculated partial η^2 for the model. This was done for three distance thresholds (150, 200, 250) and for each of three graph measures, respectively (local clustering, global clustering and average vertex degree). We follow the guidelines of Cohen (1988) for effect sizes, namely small ($0.01 < \eta^2 < 0.06$), medium ($0.06 < \eta^2 < 0.14$) and large ($0.14 < \eta^2$). The effects discussed in this and the following section will be based on the partial η^2 using these guidelines; because of the amount of data and the number of comparisons we consider significance to be less informative of any potential effects.

Results for local clustering are summarised in table 1. For each threshold, there is a main effect of group, a main effect of time, and an interaction effect of group \times time. The effect size for group is inversely related to distance threshold, with thresholds 150 ($\eta^2 = 0.12$) and 200 ($\eta^2 = 0.08$) showing a medium effect and threshold 250 showing a small effect ($\eta^2 = 0.05$). The effect size for time does not seem affected by the distance thresholds, showing a large effect for each threshold: 150 ($\eta^2 = 0.26$), 200 ($\eta^2 = 0.27$), 250 ($\eta^2 = 0.25$). The effect size for the group \times time interaction increases with distance thresholds, showing a small effect for thresholds 150 ($\eta^2 = 0.04$) and 200 ($\eta^2 = 0.05$) and a medium effect for threshold 250 ($\eta^2 = 0.06$).

Results for global clustering are summarized in table 2. For each threshold, there is a main effect of time and an interaction effect of group \times time. While there is no main effect of group, the effect size seems to be inversely related to the distance threshold as seen for local clustering. The effect size for time is similar between thresholds 150 ($\eta^2 = 0.08$) and 200 ($\eta^2 = 0.09$), showing a medium effect, but markedly higher for threshold 250 ($\eta^2 = 0.16$) showing a large effect. While all group \times time interactions are small, as with local clustering they seem to increase with the distance threshold: 150 ($\eta^2 = 0.02$), 200 ($\eta^2 = 0.03$), 250 ($\eta^2 = 0.04$).

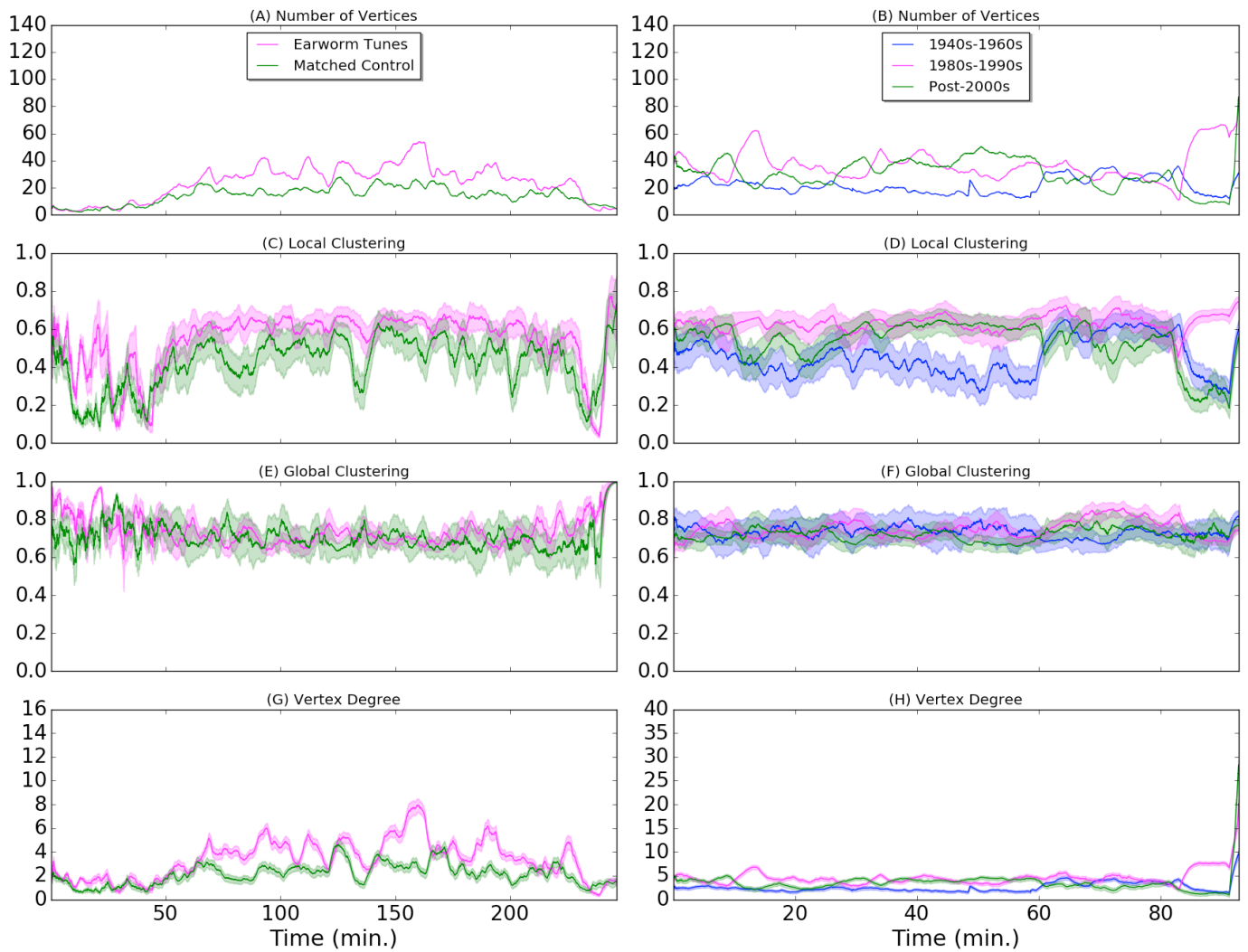


Figure 2. Statistical measures for ESCOM (left) and MOSI (right) datasets with threshold = 200; solid lines show the mean, with standard deviation represented by shading for local clustering (C, D), global clustering (E, F) and vertex degree (G, H).

Results for average vertex degree are summarised in table 3. For each threshold there is a main effect of group, a main effect of time, and an interaction effect of group \times time. While all effects are large, the effect of time is most pronounced, and all reported effects do seem to increase with distance thresholds. Group, 150 ($\eta^2 = 0.28$), 200 ($\eta^2 = 0.29$), 250 ($\eta^2 = 0.30$); time, 150 ($\eta^2 = 0.47$), 200 ($\eta^2 = 0.55$), 250 ($\eta^2 = 0.59$); group \times time, 150 ($\eta^2 = 0.19$), 200 ($\eta^2 = 0.25$), 250 ($\eta^2 = 0.27$).

C. Grouping Behaviour (MOSI)

For the MOSI dataset we created a linear model with a three-level factor for group and a 15-level factor for time segments; then we performed an ANOVA on and calculated partial η^2 for the model. This was done for three distance thresholds (150, 200, 250) and for each of three graph measures, respectively (local clustering, global clustering and average vertex degree).

Results for local clustering are summarized in table 4. The effect size for group is inversely related to distance threshold, with thresholds 150 ($\eta^2 = 0.16$) and 200 ($\eta^2 = 0.15$) showing a large effect, and threshold 250 ($\eta^2 = 0.12$) showing a medium

effect. The effect size for time increases with distance thresholds, showing a small effect for threshold: 150 ($\eta^2 = 0.04$), and a medium effect for thresholds 200 ($\eta^2 = 0.06$), 250 ($\eta^2 = 0.09$). The effect size for the group \times time interaction does not seem affected by distance thresholds, showing a medium effect for each threshold: 150 ($\eta^2 = 0.11$), 200 ($\eta^2 = 0.11$), 250 ($\eta^2 = 0.09$).

Results for global clustering are summarized in table 5. There is a small effect for all thresholds in both the group and the time main effects, as well as for the group \times time interaction. Group: 150 ($\eta^2 = 0.01$), 200 ($\eta^2 = 0.01$), 250 ($\eta^2 = 0.02$). Time: 150 ($\eta^2 = 0.01$), 200 ($\eta^2 = 0.01$), 250 ($\eta^2 = 0.02$). Group \times Time: 150 ($\eta^2 = 0.03$), 200 ($\eta^2 = 0.03$), 250 ($\eta^2 = 0.04$).

Results for average vertex degree are summarized in table 6. For all thresholds there is a small main effect of group: 150 ($\eta^2 = 0.04$), 200 ($\eta^2 = 0.04$), 250 ($\eta^2 = 0.05$). For all thresholds there also is a medium main effect of time: 150 ($\eta^2 = 0.12$), 200 ($\eta^2 = 0.12$), 250 ($\eta^2 = 0.11$). The group \times time interaction effect is small for all thresholds: 150 ($\eta^2 = 0.04$), 200 ($\eta^2 = 0.05$), 250 ($\eta^2 = 0.06$).

Table 1: Local Clustering results for ESCOM with distance thresholds 150, 200 and 250. All effects significant ($p < 0.001$).

Category	DF	F-Value	Partial η^2
150			
Group	1	10010.19	0.12
Time	29	901.86	0.26
Group \times Time	29	112.56	0.04
200			
Group	1	6667.95	0.08
Time	29	961.73	0.27
Group \times Time	29	135.32	0.05
250			
Group	1	3747.04	0.05
Time	29	867.41	0.25
Group \times Time	29	162.31	0.06

Table 2: Global Clustering results for ESCOM with distance thresholds 150, 200 and 250. All effects significant ($p < 0.001$).

Category	DF	F-Value	Partial η^2
150			
Group	1	479.48	0.01
Time	29	186.53	0.08
Group \times Time	29	36.33	0.02
200			
Group	1	366.42	0.01
Time	29	212.12	0.09
Group \times Time	29	67.95	0.03
250			
Group	1	212.66	0.00
Time	29	435.41	0.16
Group \times Time	29	92.74	0.04

Table 3: Average Vertex Degree for ESCOM with distance thresholds 150, 200 and 250. All effects significant ($p < 0.001$).

Category	DF	F-Value	Partial η^2
150			
Group	1	28479.93	0.28
Time	29	2241.63	0.47
Group \times Time	29	609.31	0.19
200			
Group	1	29744.35	0.28
Time	29	3050.55	0.55
Group \times Time	29	836.07	0.25
250			
Group	1	30926.36	0.30
Time	29	3586.98	0.59
Group \times Time	29	948.56	0.27

IV. DISCUSSION

The clustering coefficients are related to the number of participants present, that is: the number of vertices detected. Based on visual inspection of the plots this effect is strongest for the average vertex degree. This relation makes intuitive sense: as the number of vertices increases it becomes more likely for a vertex to be connected to others, especially for higher distance thresholds, thus leading to a higher vertex degree and increased clustering coefficients.

Of the graph measures used, it seems that local clustering is the best candidate for further investigation and interpretation. It shows stability over time compared to the global clustering coefficient (compare figure 2C, 2E and 2D, 2F). While it looks related to the number of vertices present

Table 4: Local Clustering results for MOSI with distance thresholds 150, 200 and 250. All effects significant ($p < 0.001$).

Category	DF	F-Value	Partial η^2
150			
Group	2	4093.11	0.16
Time	14	121.55	0.04
Group \times Time	28	190.64	0.11
200			
Group	2	3757.34	0.15
Time	14	184.64	0.06
Group \times Time	28	184.41	0.11
250			
Group	2	2859.04	0.12
Time	14	216.20	0.07
Group \times Time	28	147.73	0.09

Table 5: Global Clustering results for MOSI with distance thresholds 150, 200 and 250. All effects significant ($p < 0.001$).

Category	DF	F-Value	Partial η^2
150			
Group	2	100.39	0.01
Time	14	20.82	0.01
Group \times Time	28	37.57	0.03
200			
Group	2	190.36	0.01
Time	14	31.09	0.01
Group \times Time	28	50.38	0.03
250			
Group	2	315.64	0.02
Time	14	53.26	0.02
Group \times Time	28	63.81	0.04

Table 6: Average Vertex Degree for MOSI with distance thresholds 150, 200 and 250. All effects significant ($p < 0.001$).

Category	DF	F-Value	Partial η^2
150			
Group	2	837.93	0.04
Time	14	414.80	0.12
Group \times Time	28	59.71	0.04
200			
Group	2	936.32	0.04
Time	14	388.13	0.12
Group \times Time	28	72.37	0.05
250			
Group	2	1036.51	0.05
Time	14	373.02	0.11
Group \times Time	28	87.94	0.06

(see figure 2), the transformation that happens is not as direct as is the case for average vertex degree. In addition, local clustering has been used to describe social networks (Newman, 2003) and it is closely related to the definition of small-worldness given by Watts and Strogatz (1998) and elaborated upon by Newman (2003).

A. Interpretation of Results

1) *ESCOM*. In this dataset, judging from figure 2A, more red than green sources are picked up over the course of the night. Given that half the headphones were tuned to each channel, this suggests that the green light sources were detected less reliably than red, there is more red noise present, or there is a difference in participant behaviour.

The red channel, featuring earworms, shows higher clustering compared to the control. One possible explanation is that earworms are more memorable, by virtue of running through people's minds. This memorability, in turn, might lead to easier synchronization and tighter grouping between participants. Following this line of argument, the greater number of red sources detected could be attributed to more participants listening to earworms being on the dance-floor, while the control group was more prone to wandering about.

2) *MOSI*. This event allowed people to switch freely between channels, figure 2B shows aggregates of people switching between channels. The red (1980s and 1990s) and green (post-2000) channels seem most popular overall, while blue (1940s- 1960s) picks up later during the night. It would be of interest to compare songs between the playlists at times where a large number of participants switch, specifically focusing on popularity and their musical features.

B. Future Directions

Investigating (meta-) musical features of music used in this experiment was beyond the scope of this research project. The songs we used in this experiment all came from the Hooked on Music database (Burgoyne, Bountouridis, Van Balen, & Honing, 2013). Hooked on Music uses a recognition and verification task, combined with musical features, to investigate musical memorability and catchiness. These data can be combined with the data from these and future silent disco experiments. This adds a social dimension to the data, informs the analysis and interpretation of current silent disco data. Furthermore, in future silent discos conditions can be based on the Hooked on Music results.

While integration of the Hooked on Music data with silent disco data could yield richer descriptions, it could also increase the complexity given the sheer amounts data present. Because of this, we would suggest defining times or sections of interest. Suitable targets for further investigation are the transitions between songs, and transitions between sections. Regions of interest can also be defined around changes in musical features, such as tempo changes, as these might have an effect on grouping behavior and other measures that can be investigated in future silent discos.

Aside from investigating clustering and other measures that consider snapshots of the dance-floor at specific times, there is the possibility to examine participants on an individual basis. One way this can be done is to track individual movement over time, for which (off-the-shelf) methods used for surveillance data can be adapted (Ali & Dailey, 2012; Fradi & Dugelay, 2015; Hu, Bouma, & Worring, 2012). This might reveal how movement over time differs per listening conditions and to which extent participants synchronize their movements. Aside from increasing the sampling rate and tracking individual participants, there is the possibility to use a smartphone app that records motion, borrowing techniques from fitness apps that count steps and measure user activity.

More sophisticated models could be developed to better determine a participant's place in the room. Adding an additional camera would allow for more accurate location through triangulation, and could potentially take care of environmental noise and situations someone is partially obscured. Modeling gaze direction would be an ideal addition

because this allows connections to be drawn more informed than when simply using distance as a measure. For example, to participants dancing with their backs to one another can be very close in physical space, while being part of two different social groups existing on the dance floor.

C. Methodological Improvements

Between the three datasets discussed in this paper, there have been many improvements regarding the experimental setup. However, there are a number of further methodological improvements that should be explored. Ambient lighting is a clear source of noise in the data. Examples are (spot) lights shining on exhibition pieces and walls, as well as reflections on the clothes of participants. The masking procedure described in section 2.2.1 takes care of the static light sources, which account for most of the noise. Moving sources of noise, such as reflections on participants, are more difficult to target. So far, both ESCOM and MOSI datasets have been processed and analyzed using the same parameter settings unless mentioned otherwise. The parameters for binarization and smoothing can be fine-tuned per dataset to further reduce noise. Additionally, in setting up a silent disco, steps can be taken to reduce the amount of light on the dance floor while maintaining a pleasant environment for the participants.

The MOSI dataset suffered from camera movement during approximately the first half of the evening. To prevent such mishaps, all parties should be properly briefed on the experimental procedure and setup. This briefing can also encompass lighting, as detailed above.

Channel selection and switching poses a number of potential problems. One issue is that the number of participants per condition is not static, and cannot be controlled directly. While this is an obvious issue, it is also a feature, thus it should be controlled for in the analysis through normalization. Another issue stems from the location of the camera in relation to the dance floor. Since the camera is suspended at an angle, occlusions occur when participants move around and in front of each other. Using a second camera that records simultaneously, if it can be located properly, could alleviate these problems as well as noise from ambient lighting. However, this would require the images to be translated to a common space. This translation is technically possible, however given the amounts of data generated by these experiments; it could come with a significant extension in processing time.

1) *Image Processing*. The current image processing algorithm uses RGB space, this comes with an ease of use since the video is recorded using RGB channels and the data source, i.e. the headphone lights, use RGB LEDs. However, it might be worthwhile to explore other color spaces, such as HSV, in which the RGB geometry is rearranged to be more perceptually intuitive. Using HSV color space might reduce noise, with the trade-off of being potentially more finicky in parameter setting and adjustments to account for changes in lighting conditions.

The current experiment samples one in every ten video frames for further processing. Increasing the sampling rate will allow individual motion to be tracked. In addition, one possibility is to combine nearby frames to inform the contour

modeling, thereby boosting signal strength and taking advantage of the video's original frame rate.

2) *Contour Modeling*. The current process of contour modelling is relatively straightforward, as contours are directly estimated from the binarized images. Lower and upper bounds for contour area have been defined based on the ESCOM data, however these should be fine-tuned per dataset. Adding cycles of dilations and erosions will result in contours with smoother edges and result in fewer disjointed contours coming from the same source. Performance of the algorithm can be assessed by comparison with human scoring of contours for the same frames, in addition this can indicate sources of noise not considered before.

3) *Graph Creation*. In graph creation, one major improvement would be to use distance as a measure for edge strength. Currently, for each distance threshold, separate graphs need to be created. After processing video and extracting the centres this is the most time-consuming process. Using a measure for edge strength, each time point can do with only one graph. The formula used for encoding edge strength as a measure of distance can then be used to determine which connections to use during analysis, and which to discard. Additionally, this offers a more detailed view of grouping behaviour since groups can now be distinguished, not only by their size and being a group or not, but also by how close-knit the group is in physical space.

V. CONCLUSION

In this study, we developed a toolbox for the processing and analysis of overhead video recordings of dancers in a silent disco. Silent disco events present experimenters with an ecologically valid environment to study the interactions between music and the social environment in large groups of participants compared to laboratory settings. The toolbox developed so far allowed us to investigate the grouping dynamics of dancers in a silent disco, and how these dynamics change over time. Both datasets we investigated yielded consistent data. Specifically, we found that participants listening to earworms are more grouped than those listening to matched controls. Further development will allow for more detailed investigation of grouping, individual movement and switching behaviour under various conditions.

ACKNOWLEDGMENT

We would like to thank everyone at the Manchester Museum of Science and Industry for hosting the events and assisting with data acquisition. HH is supported by a Horizon grant (317-70-10) of the Netherlands Organization for Scientific Research (NWO), JAB is supported by a Talent Grant from the Amsterdam Brain and Cognition centre.

REFERENCES

- Ali, I., & Dailey, M. N. (2012). Multiple Human Tracking in High-density Crowds. *Image and Vision Computing*, 30(12), 966-977.
- Beaman, C. P., & Williams, T. I. (2010). Earworms (stuck song syndrome): Towards a natural history of intrusive thoughts. *British Journal of Psychology*, 101(4), 637-653.
- Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal of Software Tools. Retrieved from <http://www.drdobbs.com/open-source/the-opencv-library/184404319>
- Burgoyne, J. A., Bountouridis, D., Van Balen, J., & Honing, H. (2013). Hooked: A game for discovering what makes music catchy. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 245-250.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York, NY: Routledge.
- Fradi, H., & Dugelay, J. L. (2015). Towards Crowd Density-aware Video Surveillance Applications. *Information Fusion*, 24, 3-15.
- Freeman, W. (2000). A Neurobiological Role of Music in Social Bonding. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 411-424). Cambridge, MA: MIT Press.
- Hadley, L., Tidhar, D., & Woolhouse, M. (2012). Effects of observed music-gesture synchronicity on gaze and memory. In *Proceedings of the 12th International Conference in Music Perception and Cognition*, 384-388. Retrieved from http://icmpc-escom2012.web.auth.gr/sites/default/files/papers/384{}_Proc.pdf
- Hove, M. J., Spivey, M. J., & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, 36(6), 1525-1534.
- Hu, N., Bouma, H., & Worring, M. (2012). Tracking individuals in surveillance video of a high-density crowd. In *Proceedings of SPIE, The International Society for Optical Engineering*.
- Kalliopuska, M., & Ruokonen, I. (1986). Effects of music education on development of holistic Empathy. *Perceptual and Motor Skills*, 62(1), 187-191.
- Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, 102(3), 299-314.
- Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31(5), 354-364.
- Leman, M., Demey, M., Lesaffre, M., Noorden, L. V., & Moelants, D. (2009). Concepts, technology, and assessment of the social music game 'Sync-in-Team'. In *Proceedings of the 12th IEEE International Conference on Computational Science and Engineering*, 4, 837-842.
- Madison, G. (2006). Experiencing groove induced by music: Consistency and phenomenology. *Music Perception*, 24(2), 201-208.
- Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167-256.
- Nichols, A. L., & Maner, J. K. (2008). The Good-Subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, 135(2), 151-165.
- Otsu, N. (1979, jan). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.
- Peixoto, T. P. (2014). The graph-tool Python library. Figshare. doi: 10.6084/m9.figshare.1164194
- Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*, 6, 1-11.
- Rabinowitch, T.-C., Cross, I., & Burnard, P. (2012). Long-term musical group interaction has a positive influence on empathy in children. *Psychology of Music*, 41(4), 484-498.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440-442.
- Williamson, V. J., & Jilka, S. R. (2014). Experiencing earworms: An interview study of involuntary musical imagery. *Psychology of Music*, 42(5), 653-670.
- Winkler, I., Häden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *PNAS*, 106(7), 2468-2471.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception

and production. *Nature Reviews Neuroscience Neuroscience*,
8(7), 547-558.

Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music
in infancy. *PNAS*, 107(13), 5768-5773.