F-formations and Social Contact in Children: Exploring Dyadic Interactions and Modeling Groups

Master Thesis

to obtain the degree of Master in Computer Science with Specialization in Data Science and Technology at Delft University of Technology, to be publicly defended on 23rd October, 2024.

by

Yuan Tian

Socially Perceptive Computing Lab,
Faculty of Electrical Engineering, Mathematics and Computer Science (Faculteit Elektrotechniek, Wiskunde en Informatica), Delft University of Technology, Delft, The Netherlands.

Thesis Committee

Thesis advisor: prof. dr. Hayley Hung (Technische Universiteit Delft)
Core member: dr. Stephanie Tan (Technische Universiteit Delft)
Externel member: prof. dr. D. Messinger (University of Miami)
Externel member: dr. Lynn Perry (University of Miami)

Keywords: F-formation, Social Contact, Children Interactions, Homophily Effect

An electronic version of this thesis is available at
http://repository.tudelft.nl/.

Science is a wonderful thing
if one does not have to earn one's living at it.

Albert Einstein

Contents

Abstract

Understanding children's social interaction patterns is critical for their cognitive development; however, existing psychological studies often focus on dyadic interactions, overlooking the complexities of group dynamics. This study extends the concept of homophily—the tendency for individuals to interact with similar others—by exploring its implications within group settings. We introduce F-formations, a well-adopted notion in the computational field for detecting interaction groups in adults, which has yet to be extensively studied in children. By identifying this research gap, we aim to investigate how F-formations can be applied to children's social interactions.

Our case study was conducted in a preschool setting, specifically in the Starfish classroom, which includes children with hearing loss. Several key findings emerged from this study. First, we observed that F-formations can be detected in children's interactions, though some formations violate the definition of F-formations. Second, while the homophily effect was not evident in dyadic studies concerning children's hearing conditions, it was observed in group settings. We found that children prefer to form more homogeneous and smaller groups, with those who have hearing loss spending more time with peers who share similar characteristics.

Overall, our results suggest that integrating group interaction data into traditional dyadic analyses provides valuable insights into children's social behavior, highlighting the importance of studying group dynamics alongside dyadic interactions.

1

INTRODUCTION

Analyzing children's social interactions has proven to be a valuable area of research, as these peer interactions provide essential resources for children's development, supporting their cognitive, emotional, and social skills[3]. This became particularly important following the COVID-19 pandemic, which led to a decline in social interactions among children, especially those with neuroatypical conditions, such as Autism spectrum disorder (ASD)[21]. A central challenge in this research is identifying and measuring social interactions. Psychological studies have often focused on detecting dyadic (two-person) interactions. A significant advancement in this field came from Messinger et al. [13], who pioneered the use of automated data, such as continuous positional and movement data, instead of the manual encoding methods used in earlier studies. They introduced the concept of a radial distribution function, suggesting that children within a range of 0.2 m to 2.0 m of each other can be considered in social contact. Fasano et al. [7] extended this criterion by incorporating orientation, proposing that children who are oriented toward one another within a 45° angle are also considered to be in social contact. Building on these criteria, subsequent studies have investigated various social patterns, including the homophily effect [1] and reciprocal interaction patterns [15], as well as how additional factors, such as children's physical conditions and the activities they are engaged in, influence their social interactions.

While dyadic studies based on indexing social contact offer valuable insights into the interaction patterns among children, they fall short of capturing the complexity of social behavior within group settings. Group interactions involve more intricate dynamics that cannot be fully understood by examining pairs alone. For instance, an outgoing child may facilitate interactions between two other children, but by studying only dyads, we might miss the broader interactional patterns and how these interactions evolve in the presence of additional peers. Moreover, as acknowledged by Messinger et al. [13], mere co-location does not necessarily guarantee a genuine interaction. For example, children sitting around a table, whether face-to-face or side-by-side, engaged in activities like paper cutting, are not necessarily interacting with one another. Therefore, our aim is to

Codes of this thesis can be found at the github repository.

adopt a more comprehensive approach to modeling social interactions by extending the analysis to group-level interactions, thereby uncovering additional patterns in children's social behavior that are influenced by the broader social context.

To model children's group interactions, we also bring the psychological concept of f-formation, as many computational group interaction detection tasks from surveillance systems [9], computer vision [6][20][19], and Human-Robot Interaction [2][8] have done. An f-formation is a concept raised by Kendon[10]. An f-formation refers to a social and spatial arrangement in which individuals create and maintain a shared, enclosed space (known as the "o-space"), allowing all participants to have equal and direct interaction as seen in Figure 2.2. It is a useful tool for detecting interacting groups because, as noted by Kendon, f-formations can identify a specific interactional situation and provide a means of defining a social encounter as a unit for analysis. Although f-formation detection has already been extensively studied in adults [5, 19, 2, 4], its application to children remains largely unexplored. Children exhibit interactional patterns that are not typically found in adults. First, based on our animated children's movement data (seen in Figure 3.4), we found that children tend to be quite close to each other. Additionally, a study [16] points out that the interactions they form can be easily interrupted by a child's shift of attention from their partner to things such as other's approaching and toys, making their formations unstable and fluid. Hence, we first want to explore the applicability of f-formation analysis in the children's setting, leading us to the following research question:

RQ 1: *Do f-formations occur in children's group interactions?*

Due to a lack of annotated data in our study case (dataset details provided in Section 3), we cannot directly validate the extracted f-formations by comparing them with the annotations using standard metrics such as precision and recall. Additionally, manually annotating the data can be time-consuming and costly, and it introduces subjectivity [27]. Therefore, instead of relying on annotation, we adopt an indirect approach to answer this question: we hypothesize that children do form f-formations, and we employ an f-formation detection method to extract them. We then analyze the impact of these extracted f-formations as group context on one of the existing dyadic social interaction studies. If the conclusions of their study still hold or are further supported, it will bolster our hypothesis that f-formations do exist among children and enable us to explore additional social patterns with group information.

Specifically, we aim to extend the dyadic research on the homophily effect among children [1]—where children tend to associate with those who share similar physical conditions—to group settings. While the study primarily focused on children with ASD, our study case involves children with or without hearing loss. Therefore, we will first verify the homophily effect in children concerning their hearing conditions in dyads and then incorporate group information to examine the effects of grouping information. By comparing findings from these studies, we can determine how group context influences the dyadic conclusions drawn in [1]. This leads us to the following research questions:

RQ 2: *Does the homophily effect apply to children when considering their hearing conditions in dyadic interactions?*

RQ 3: *How does incorporating group context influence or further verify the homophily effect in children with respect to their hearing conditions?*

To address these questions, we adopt the Dominant Set (DS) clustering framework proposed by Hung et al. [9] to identify f-formations, followed by a series of experiments and analyses. The structure of this thesis is as follows: We first review related work in Chapter 2 to identify existing literature and the research gap that our study addresses. In the dataset section, we provide detailed information about the specific case study we are working on in Chapter 3. The methodology section, outlined in Chapter 4, describes the pipeline we use for f-formation detection and analysis. The experiments and results section in Chapter 5 presents our experimental findings and addresses the research questions. Finally, we conclude in Chapter 6 by summarizing our findings and discussing the limitations of our study.

2

Related Work

2.1. Children's Dyadic Interaction Study in the Psychological Field

Previous psychological studies on children's dyadic interactions primarily relied on manual encoding, which limited both the quality and quantity of observed interactions. To address this limitation, Messinger et al. [13] promoted the use of automated tracking to study children's locations and movements during social interactions. They modeled data obtained from 16 five-year-olds during three 1-hour classroom free play sessions and developed the radial distribution function (Figure 2.1) using a data-driven method. The radial distribution function serves as an index for determining when children are in social contact more than chance levels of colocation. They defined social contact as occurring when children were within 0.2 to 2 meters of each other.

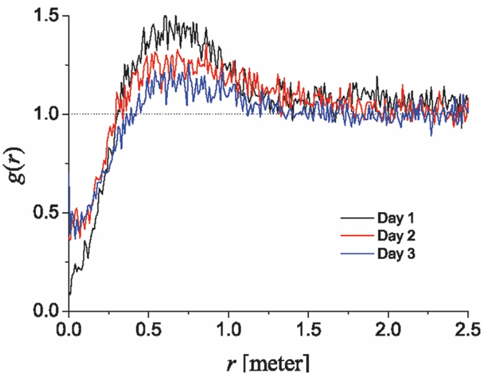


Figure 2.1: Radial Distribution Function. g(r) is the ratio of the observed distance between children to the expected distance by chance. When g(r) > 1, children are closer than expected by chance, indicating clustering. Social contact was defined as children being within a radius, r, of 1 m, where g(r) peaks.

Building on this, Fasano et al. [7] added the criterion of children being oriented towards one another within a 45° angle as another marker of social contact to improve the method's accuracy. Based on the interactions found by this updated criteria, their study made foundings that children with ASD were less central in the classroom and exhibited lower cohesion within the ASD group. Additionally, they found a reciprocal pattern that children were more likely to vocalize to peers who had previously vocalized to them.

Perry et al. [15] extended Fasano et al.'s reciprocal research to children with hearing loss, observing that vocalizations from a peer in one observation predicted a child's vocalizations to that peer in subsequent sessions. They also noted that social contact and vocalizations varied by activity type, with fewer interactions during free play than in structured activities.

Banarjee et al. [1] explored the homophily effect among children with developmental disabilities (DD) and typically developing (TD) peers. They employed mixed-effect analysis to investigate how the social contact time of a pair is influenced by their homophily status (as shown in Table 2.1). Their study revealed that children in concordant dyads (DD-DD/TD-TD) spent a greater proportion of time in social contact than discordant dyads (DD-TD, $p<0.001$). Additionally, DD-DD dyads were in social contact less than TD-TD dyads($p<0.001$). Children with DD did not differ from TD children in the overall time spent in social contact with other children. It is noteworthy that the DD children sampled are those with ASD, but the conclusions were generalized to encompass all cognitive disabilities.

Table 2.1: Results from the original homophily effect study [1]. Time in social contact is defined as the time the pair were in social contact divided by the total time both children were present in the classroom. The table presents the predictors affecting the time spent in social contact.

Predictors	**Time in Social Contact (B)**	**Standard Error (SE)**	**Confidence Interval (CI)**	**t-value**	**p-value**	**Effect Size (d)**
(Intercept)	0.04	0.00	0.04–0.05	13.22	< 0.001	-
[DD]	0.00	0.00	−0.00 to 0.00	0.02	0.987	0.00
Homophily [concordant]	0.01	0.00	0.01–0.01	6.63	**< 0.001**	0.24
[DD] x Homophily [concordant]	−0.01	0.00	−0.02 to −0.01	−4.29	**< 0.001**	−0.16
Random effects						
σ^2	0.00					
Child	0.00					
Classroom	0.00					
ICC	0.08					
Observations	3108					

While these studies provide valuable insights into dyadic interactions among children, they do not fully capture children's social behaviors in group settings. For instance, in a group of three, the presence of a third child C can either facilitate or hinder the interactions between children A and B. A and B's limited communication might be due to child C dominating the conversation, rather than a lack of closeness between A and B. Alternatively, C could also enhance A and B's interaction. Additionally, A and B may often appear together in group settings, but their interaction could be driven more by the social dynamics of the group than by direct dyadic engagement. Therefore, to better understand these interaction patterns, it is necessary to bring group perspective into existing research.

2.2. F-FORMATION

Detecting interactional groups has long been a popular topic of research, although a formal definition of these groups remains lacking. With the rise of social signaling literature, the social psychological notion of face-formation, or f-formation, has been widely adopted to address this gap [18]. According to Kendon [10],

"An F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access."

Such patterns are frequently observed in free-standing conversational groups, where individuals gather with the intention of conversing and exchanging information [9]. Kendon recognized the significance of studying f-formations because the f-formation system serves as a crucial means of maintaining the separate identity and integrity of an interactional situation. Furthermore, because of its boundary-defining function, the f-formation provides a useful way to define a social encounter as a unit of analysis. It also organizes the spatial structure of face-to-face interaction. Given these capabilities, the f-formation is an ideal focus for interactional group detection tasks.

In practice, an f-formation involves the organization of three social spaces: o-space, p-space, and r-space (see Figure 2.2)[6][11]). The o-space is a convex, empty area surrounded by participants engaged in social interaction, where each individual faces inward, and external people are excluded from this region. It is the core component of an f-formation. The p-space is a narrow band surrounding the o-space, occupied by the bodies of the interacting participants. Finally, the r-space extends beyond the p-space and encompasses the surrounding area.

Commonly observed f-formation arrangements are shown in Figure 2.3. For two participants, the most frequent formations include: the L-arrangement (Figure 2.3(a)), where two individuals stand perpendicularly to each other, forming the shape of the letter "L"; the vis-à-vis or face-to-face arrangement (Figure 2.3(b)), where two individuals face each other; and the side-by-side arrangement (Figure 2.3(c)), where two participants stand close together, facing the same direction. With more than two participants, the arrangements typically form semi-circular (d), circular (e), or linear (f) patterns [12].

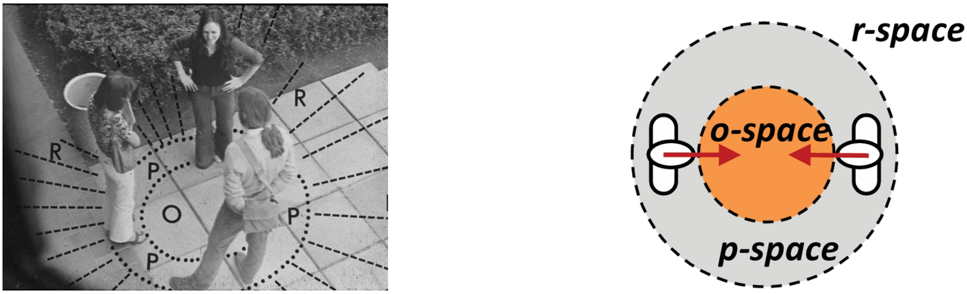


Figure 2.2: An example of a conversational group organized as an F-formation (left)[11] and a conceptual model (right)[6] showing the o-space, p-space, and r-space that constitute the F-formation.

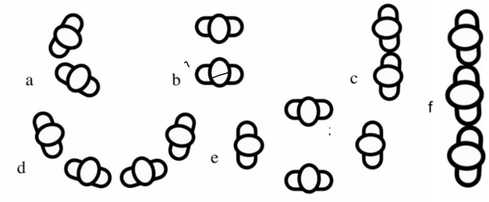


Figure 2.3: Common f-formation arrangements: (a) L-arrangement, (b) face-to-face, (c) side-by-side, (d) semi-circular, (e) circular, and (f) linear [12].

2.3. INTERACTIONAL GROUP DETECTION IN COMPUTATIONAL FIELDS

Considering an f-formation as a lose spatial arrangements of group, existing literature of detecting f-formations can be concluded as two lines of approaches[18]. They represent two perspectives of interpreting the F-formation. One is by direct mathematical modeling of spatial arrangements [26, 20, 6]. These model-based approaches tend to formalize the *transactional segments* of individuals, which is the space that extends forward from their lower body and that includes whatever they are currently engaged with[10]. Then, these methods find the intersection of transactional spaces in a scene, or o-spaces of the F-Formations.

The second approach adopts Graph Theory, where social scenes are represented as edge-weighted graphs. In these graphs, individuals are depicted as nodes, and their pairwise relationships as edges. The goal is to partition the graph into subgroups of nodes that correspond to interactional groups, as demonstrated in the works of [9, 25].

Setti et al. [18] provided a comparative study of two widely cited approaches from these categories: the Hough for f-formation (HFF) method developed by Cristani et al. [6] and the dominant set (DS) clustering method proposed by Hung and Kröse [9]. Their analysis extended the metrics for group detection by introducing a threshold to assess how accurately group members were identified. The study found that DS clustering was more accurate when only position data was available, while HFF was more robust to noisy data and performed better when head orientation information was included. However, we argue that DS clustering method offer the benefits that 1) it is more practical because, in many research scenarios, tracking devices primarily capture position data; 2) as a graph-based approach, DS clustering is inherently more flexible in modeling member relationships. This flexibility allows for the incorporation of additional factors by constructing various affinity matrices, creating significant potential for future enhancements in group detection methodologies.

In graph-based methods, the affinity matrix is typically constructed using position and orientation data. However, numerous techniques have since been employed to account for other factors that influence group interactions. To name a few, DANTE models dyadic and contextual interactions using relative positions and head/body orientations as input features in its deep learning framework[23]. Zhang et al. [28] incorporate environmental factors like furniture layout and crowdedness by modeling the geometric variations of potential f-formations in a space. Additionally, beyond the visually obtained features of position and orientation, other modalities have been incorporated to enhance group detection. For example, Thompson et al. [24] combine motion-based features with visual data to strengthen the estimation of pairwise affinities. Similarly, Ramirez et al. [17] proposed a synchrony-based method that refines clusters of individuals by integrating proxemics and the 2D field of view, highlighting the importance of multimodal data for robust group detection.

2.4. Research Gap

In summary, dyadic studies within the psychological field have primarily focused on indexing pairwise social contact based on two criteria: a distance of 0.2 to 2 meters and a relative orientation of 45° between individuals. However, these studies often overlook patterns that can only be understood within a broader group context. Therefore, we aim to adopt existing computational methods for detecting interactional groups.

While the framework for detecting f-formations has been well established, there has been no specific research on detecting and analyzing f-formations among children. Based on our analysis of children's movements in classroom settings, as discussed in Figure 3.4, and insights from developmental psychology literature[16], we find that children's spatial interactions are often intimate yet unstable, easily disrupted by external factors such as the presence of a third person or nearby objects. Thus, studying f-formations in children offers a unique perspective that we believe is a significant research gap worthy of exploration.

3
Dataset

3.1. Starfish

Our case study is on the Starfish Kindergarten inclusive classroom [22]. This dataset logs the activities of preschoolers in the classroom over 12 separate days from October 2022 to June 2023, with observations lasting approximately 3 hours each day (from 9:30 AM to 12:30 PM). The children engage in both unstructured and structured play activities, including breakfast time, storytime, circle time, and free play. The ongoing activities are logged, and each day follows a similar schedule. The classroom features four designated activity areas: book reading, dramatic play, circle time, and lunch/snack periods. Additionally, there is a terrace outside the classroom, as illustrated in Figure 3.1. The classroom dimensions are approximately 15 × 15 m.

Data were collected from 13 children, consisting of 9 girls and 4 boys, aged between 4 and 6 years. These children are labeled with IDs *DS_STARFISH_2223_27-33, 42-46*, with only the numbers used later for brevity. The children speak various languages, including English, Mandarin, Spanish, and Portuguese, with English as the dominant language. Among the participants, six children have hearing loss (HL), while the remaining seven have typical hearing (TH). Additionally, data were recorded for 4 teachers and 2 lab assistants, all of whom are female.

To collect data, each child wore a specially designed vest, as shown in Figure 3.2. The vest contains a LENA recorder that captures vocalization data, a binary value indicating whether the child is speaking (1) or not (0). Additionally, two Ubisense active tags (left and right) are sewn into the pockets of the vests, positioned around waist height. Utilizing RFID technology, sensors at the corners of the classroom track these active tags, as illustrated in Figure 3.3. The two tags provide tracked locations for each individual. The midpoint of the tags' XY coordinates indexes an individual's location, while their body orientation is estimated using these two coordinates.

The Ubisense system monitors children's positions at a rate of 1–4 Hz, recording at 1 Hz when static and up to 4 Hz when moving. After obtaining the raw Ubisense data, it was resampled to 0.1-second intervals. More details can be found in the data processing section 4.2. Following data preprocessing, we created an animation tracing the

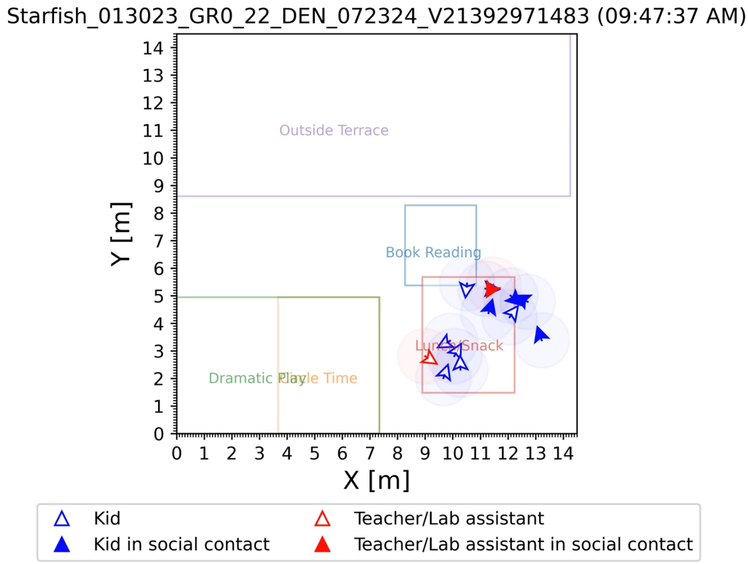


Figure 3.1: Snapshot of the children's activity animation. Date: 2023/01/30. Time: 09:38:46.300. The layout includes four main activity areas and a terrace. Each triangle represents a participant, with blue triangles indicating children and red triangles representing teachers/lab assistants. The triangles show the positions and body directions of each participant, with top angle indicating orientations. A filled arrow indicates that participants are in social contact. The origin of the XY plane is aligned with the left and bottom sides of the classroom boundary (top view).

movements of participants in the classroom; a snapshot of the animation can be seen in Figure 3.1. Observations from the animated video reveal that children tend to stand very close to one another, as illustrated in Figure 3.4.

As a dataset previously used for studies on dyadic interactions, it also includes logs of pairwise children's interactions when two subjects are within a defined social contact distance of 0.2 to 2.0 meters from each other or also within 45 degree relative orientations.

3.2. IDIAP POSTER DATA

The Idiap Poster Data[9] comprises 3 hours of real aerial video featuring over 50 people showing up in a scientific work during a poster session. From this video, 84 distinct images were selected and organized into 8 different sets. Each set of images was annotated by 3 different annotators. After getting proper definitions, the annotators were tasked with identifying F-formations and their associates from the static images. In total, 21 individuals contributed annotations for at least one set of images. Further details about the dataset can be found in the corresponding paper [9]. The dataset records the positions of individuals in the image plane, with the ordering reflecting the sequence in which the annotators clicked on the people in the scene to identify the F-formations.

Figure 3.2: Specially-designed Vest

Figure 3.3: Sensors in the classroom

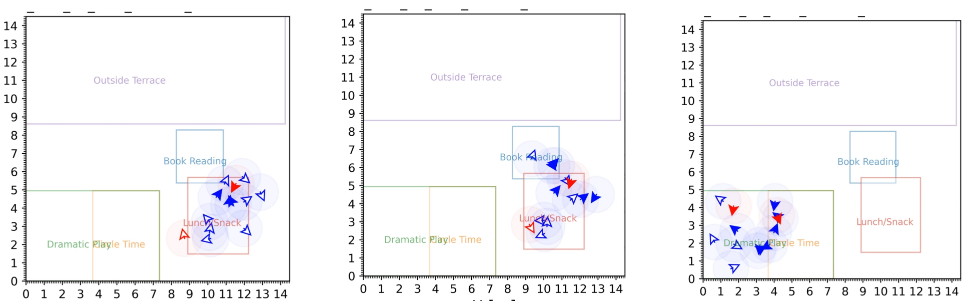


Figure 3.4: Snapshots from the classroom animation illustrating the close proximity of children during interactions.

4

Methodology

4.1. Overview

This chapter outlines our methodology to achieve the following objectives: 1) to examine f-formations in children, and 2) to investigate how the homophily effect influences dyadic and group interactions concerning hearing conditions. The fundamental task involves detecting f-formations among children's groups and extending the existing dyadic analysis using the detected f-formations.

After obtaining the raw data from the Ubisense devices, we undertook the following steps to accomplish our tasks. First, we performed preprocessing to remove unwanted data, interpolate missing values, and estimate individual positions and body orientations based on the two locations from the Ubisense devices.

Next, we employed Hung's Dominant Set (DS) clustering algorithm [9] in our f-formation extraction process. This involved constructing the interaction graph for each scene, building the affinity matrix, and calculating the dominant sets—i.e., the estimated f-formations in each scene. Each classroom scene at a given timestamp is represented as an interaction graph $G = \langle V, E, A \rangle$, where n vertices (V) represent the participants (children in our case). Each pair of nodes $i, j \in N$ is connected by an undirected edge $e \in E$, and the corresponding weight of this edge, w_{ij}, represents the pairwise affinity (i.e., $a_{ij} = w_{ij}$), indicating the likelihood that two subjects belong to the same group. In the DS extraction framework, the weights are organized into an affinity matrix, where each element denotes the affinity score between two nodes. We employed the Gaussian kernel to calculate pair affinities, as used in Hung's work [9]. Once the affinities for each pair of individuals in the social interaction graph are computed, the process continues with grouping individuals using the Dominant Sets (DS) extraction method by Hung and Kröse.

To assess the feasibility of detecting children's groups through f-formation extraction while considering the influence of group context, we extended the dyadic homophily studies conducted by Banerjee [1] to a group context. This was accomplished by integrating group properties derived from the previous f-formation extraction. Since the dataset lacks labels for direct performance measurement, we evaluated the f-formations by analyzing how the results of Banerjee's original works were influenced by group information through statistical analysis.

4.2. DATA PREPROCESSING

The original raw data were collected from two Ubisense devices located in the left and right pockets of each child's vest (at around waist height), with a frequency ranging from 1 to 4 Hz. The midpoint between the two devices is considered the child's location, while the line linking the two devices relative to the x-axis (as illustrated in Figure 3.1) is interpreted as the body orientation.

In our preprocessing steps, we generated three versions of the dataset for analysis:

1. **Linear and Kalman Interpolation:** We first applied linear interpolation to fill in any missing data points and resampled the data at 0.1-second intervals. To enhance interpolation accuracy, we incorporated Kalman interpolation for locations that were missing within a one-minute window.

2. **Manual Cropping of Overlapping Data:** During our analysis, we identified instances of overlapping triangular positions in the data. We manually cropped these entries.

3. **Outlier Removal**: We removed any data entries that were deemed erroneous based on a z-score threshold of $z/\text{height} > 1.25$.

The results of each version are detailed in the appendix, while subsequent experiments are conducted using version 2.

Following these preprocessing steps, we eliminated unrelated entries, including those associated with teachers, as well as times when children were not present in the classroom or were outside on the terrace (primarily during toilet breaks and playground time), according to the activity logs. Next, we constructed a coordinates dictionary for all participants visible in each scene at 0.1-second intervals. For each scene, we stored the x and y coordinates within the classroom, along with the body orientation for each child present.

4.3. AFFINITY MATRIX BUILDING

In this step, our goal is to build the $n \times n$ affinity matrix $A = a_{ij}$ for an interactional scene G using the provided features. As mentioned in Section 2.3, various techniques and feature choices exist for this step. While this is worth exploring in future research, it is not the primary focus here. In this study, we continue to utilize positions and orientations. When only positions are considered, the affinity calculation is defined by the following equation:

$$a_{ij} = \begin{cases} 0 & \text{if } i = j \\ e^{-\frac{d_{ij}}{2\sigma^2}} & \text{if } i \neq j \end{cases} \tag{4.1}$$

where $i, j \in V$ are the i-th and j-th children in an interactional scene. a_{ij} represents the affinity score between nodes i and j; d_{ij} is the Euclidean distance between i and j, calculated as $\|\text{positions}_i - \text{positions}_j\|_2$. Here, positions_i refers to the x, y coordinates of node i in scene G.

It is important to note that although orientation is a crucial aspect in the definition of f-formations, Setti et al.'s work [18] suggests that using position alone is sufficient to detect groups. Including head or body orientations may enhance detection accuracy, especially in crowded spaces, as shown in Figure 5.1. In this case, children 44 and 45 are grouped together using the position-only Dominant Set (DS) algorithm, yet their attention is not aligned.

When both position and orientation are utilized, a binary mask is applied to the kernel, ensuring that a_{ij} is non-zero only if the partner is within $-90°$ to $90°$ around the person's oriented direction. This can be represented as follows:

4

$$a_{ij} = \min\left(\begin{cases} e^{-\frac{d_{ij}}{2\sigma^2}} & \text{if } -\frac{\pi}{2} \leq \theta_i - \alpha_{ij} \leq \frac{\pi}{2}, \\ 0 & \text{otherwise} \end{cases}, \begin{cases} e^{-\frac{d_{ij}}{2\sigma^2}} & \text{if } -\frac{\pi}{2} \leq \theta_j - \alpha_{ji} \leq \frac{\pi}{2}, \\ 0 & \text{otherwise} \end{cases}\right) \tag{4.2}$$

where α_{ij} denotes the direction from i to j, and θ_i represents the body orientation of person i. In practice, this means that a child must be within their frontal space for their affinity score to be non-zero.

The choice of the parameter σ cannot be optimized through a standard loss function. Previous research by Messinger et al. [13] indicates that children within a distance of 0.2 to 2 m are considered to be in social contact. This aligns with the interpretation of σ, suggesting that children within a distance less than or equal to σ have a higher affinity for grouping. Therefore, we began searching for σ within the range of 0.2 to 2 m, observing the extracted f-formations with different values of σ for a random scene to determine the optimal choice.

4.4. DOMINANT SET CLUSTERING ALGORITHM

After Setti et al. [18] compared the Dominant Set (DS) clustering method [9] with the Hough for F-Formations(HFF) [6], they find that Hung's graph-based clustering approach is suitable for scenarios where only positional information is available. The HFF method exhibits greater robustness to noise and performs better when incorporating head orientation. In our study, we choose to adopt Hung's DS extraction approach for the reasons: 1) We have body orientation data estimated from devices attached to the children's waists, rather than head orientation as emphasized by Cristani et al.; 2) The Dominant Set framework is more flexible, allowing for the incorporation of various external factors through the construction of different affinity matrices, which may prove beneficial for future work.

Once the predicted affinities are organized into an affinity matrix, we proceed to extract F-formations using the DS extraction approach. This method is based on the principle that F-formations can be identified through graph clustering, framing the problem as one of finding dominant sets—a concept first introduced by Pavan and Pelillo [14].

In this context, a dominant set can be viewed as a maximal clique, generalized to edge-weighted graphs. It guarantees that the internal affinity among nodes in the set exceeds the affinity between any of the nodes in the set and those outside it. This property closely aligns with the definition of F-formations, where individuals share a mutual focus of attention, resulting in higher affinity among group members compared to those outside the group.

Given a graph $G = (V, E, w)$, where V denotes the set of vertices (individuals), E signifies the set of edges (connections between individuals), and w is the positive weight function representing the affinities between individuals, the DS clustering algorithm identifies a subset of vertices $S \subseteq V$ such that the internal connections within S (as measured by the weighted affinity matrix) are maximized relative to connections with vertices outside S.

The average weighted degree $k_S(i)$ of a vertex $i \in S$ with respect to the set S is defined as:

$$k_S(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij},$$

where a_{ij} represents the affinity between individuals i and j in the graph. This degree quantifies the average connection strength between node i and the remainder of set S. To assess the relative affinity between node $i \in S$ and node $j \notin S$, we define $\phi_S(i, j)$ as:

$$\phi_S(i, j) = a_{ij} - k_S(i),$$

which indicates the contribution of node j to the overall affinity of node i within the set.

The algorithm then recursively computes the weight $w_S(i)$ of node i concerning a subset S, which includes both the node and previously selected vertices:

$$w_S(i) = \begin{cases} 1 & \text{if } |S| = 1, \\ \sum_{j \in R} \phi_R(j, i)\, w_R(j) & \text{otherwise.} \end{cases}$$

Here, $R = S \setminus \{i\}$, and $w_S(i)$ measures the overall relative affinity between node i and the remaining members of set S. The algorithm continues iterating until a set S meets the conditions of a dominant set, defined by:

$$w_S(i) > 0, \quad \forall i \in S,$$

$$w_{S \cup \{i\}}(i) < 0, \quad \forall i \notin S.$$

These conditions ensure that members of the dominant set maintain stronger internal connections compared to connections with external nodes. The DS algorithm iteratively identifies such sets until the stopping criterion is satisfied.

However, a limitation of the standard peeling-off strategy for identifying dominant sets is that as more sets are removed, the remaining nodes often become singletons, which may not correspond to any meaningful F-formation. To overcome this issue, we modify the peeling-off approach proposed by Pavan and Pelillo [14] by introducing a principled stopping criterion. This criterion evaluates the global context of the complete graph by comparing the weight $w_{S \cup \{i\}}(i)$ for all $i \notin S$, ensuring that clustering remains meaningful within the broader context of the entire graph rather than just the local subgraph. This modification eliminates the reliance on an arbitrary threshold and facilitates more robust F-formation identification.

Finally, replicator dynamics [14] are employed to solve the underlying quadratic optimization problem, ensuring that the identified dominant sets are maximally cohesive. This approach enables the accurate identification of F-formations, which represent groups of individuals in a scene who have easy and equal access to a shared interaction space.

4.5. HOMOPHILY EFFECT WITH RESPECT TO HEARING CONDITIONS IN CHILDREN DYADS AND GROUPS

4

The original study on the homophily effect [1] analyzed a dataset of children with Autism Spectrum Disorder (ASD) and concluded that this phenomenon could extend to various developmental disabilities. To explore whether this generalization applies to children in relation to hearing conditions, we conducted a mixed-effect analysis for dyadic interactions, replicating the methodology used in the original study [1] with the Starfish dataset.

To further investigate how group contexts influence dyadic interactions in children with hearing conditions, we shifted our focus from dyads to individual children to assess how they are affected by their grouping. We extracted F-formations for each time frame (every millisecond). However, not all formations represent meaningful interactions; for example, brief encounters between children passing each other may not indicate engagement. Therefore, our goal is to identify key group properties across all extracted formations from the observation days to understand their influence on children's interactions.

Given that the original concept of social contact is inherently dyadic and cannot be directly applied to groups, we propose the following formula to calculate the social contact ratio in a group context. Notably, we do not adopt the definition of social contact from [13]; instead, we consider a child to be in social contact with a group if they are part of a detected F-formation.

$$\text{Social Contact Ratio} = \frac{\text{Total time a child is in a group}}{\text{Total time the child is in the classroom}}$$

Based on the findings from [1] and the results presented in Table 2.1, we recognize that children tend to spend time with peers who share similar characteristics. To investigate this homogeneity and extend it to group settings, we introduce the first group property known as the homophily degree, defined as the ratio of the majority condition count to the group size.

Homophily Degree: The proportion of individuals in the group who share a common characteristic (the majority condition), calculated as:

$$\text{Homophily Degree} = \frac{\text{Majority condition count}}{\text{Size of the group}}$$

Additionally, as shown in Table 2.1, children with developmental disabilities (DD) tend to spend more time with other children who also have DD. This observation suggests that children with hearing loss (HL) may similarly prefer to be in groups with a higher number of peers with HL. This leads us to investigate a second group property:

$$\text{HL Ratio} = \frac{\text{HL count}}{\text{Size of the group}}$$

Thirdly, based on the extracted F-formations, we noted that groups consisting primarily of sizes 2 and 3 appear most frequently. This observation indicates that children might prefer to spend time in smaller-size groups. Thus, we also analyze the impact of group size on the social contact ratio concerning the homophily effect.

In summary, we first examine the homophily effect in children w.r.t hearing conditions and then we examine how hearing conditions and three group properties, i.e. homophily degree, HL ratio, and group size, collectively affect the time children spend together.

5

Experiments and Results

5.1. Implementation and F-formation Extraction Results

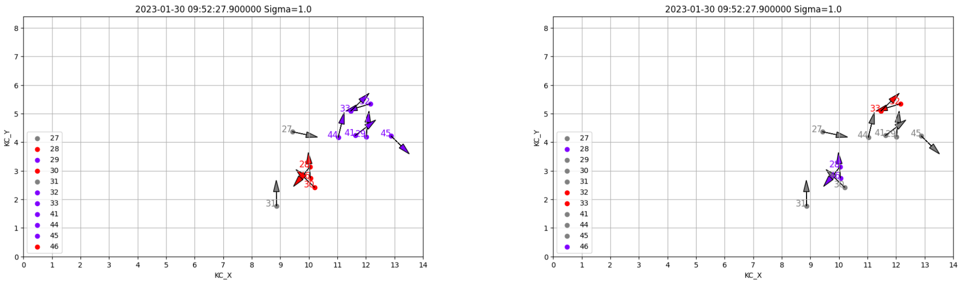
We implemented the Dominant Set (DS) algorithm following the reproduction of DANTE's work [23]. However, DANTE's approach to building the affinity matrix differs from ours, and they did not directly report their reproduction results with the DS algorithm. To validate our implementation of the affinity matrix building as described in [9] and DANTE's DS algorithm, we first applied it to the Idiap dataset, which contains only positional information. We assessed the results using the evaluation method proposed in Setti et al. [18], which establishes a threshold indicating the acceptable portion of correctly identified group members. A group is considered correctly identified if at least $\lceil (T \cdot |G|) \rceil$ of its members are detected by the grouping algorithm, with no more than $1 - \lceil (T \cdot |G|) \rceil$ false positives, where **G** represents the size of the annotated group and **T** is the threshold.

Once the group labeling was established, we applied standard metrics to quantify the accuracy of the detection method. We also compared our implementation results with those from Setti et al., who implemented the DS algorithm using the same affinity calculation equation with positional data. By setting the sigma value of the Gaussian Kernel in the affinity function to 40, we achieved optimal results comparable to those reported in their work [18] on the Idiap Poster Data. The performance metrics are shown in Table 5.1, confirming the correctness of our implementation of affinity matrix building and DANTE's DS algorithm.

Table 5.1: Comparison of precision, recall, and F1 scores for different thresholds using Setti et al.'s method and our proposed method.

	Threshold = 2/3			Threshold = 1		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Setti et al.'s	0.90	0.81	0.85	0.71	0.65	0.68
Our Implementation	0.88	0.80	0.84	0.74	0.64	0.68

After verifying the implementation of the DS clustering algorithm on Idiap Poster Data, we obtained F-formation results based on positional data alone from the Starfish dataset. For this dataset, the sigma value was set to 1.0, guided by the intuition from Figure 2.1, where the peak occurs at this value. We also performed the same extraction for features that included both position and orientations. The extraction results from one random scene are illustrated in Figure 5.1.



(a) Positions only

(b) Positions + orientations

Figure 5.1: Sample F-formation extraction results. The origin of the *KC_X-KC_Y* plane is aligned with the left and bottom sides of the classroom boundary (top view). Only children are shown, and they are labeled with numbers. Both figures represent the state at Time: 2023-01-30 09:52:27.900000, with $\sigma = 1$. Different colored arrows indicate membership in different F-formations, while grey arrows represent singletons. (a) Extraction based on positions only. (b) Extraction based on positions and orientations.

5.2. PARAMETER OPTIMIZATION

In our analysis, we set the sigma value appropriately when using the Gaussian kernel in the affinity function to achieve optimal extraction results for subsequent analysis. The sigma parameter represents the distance range within which individuals are considered to belong to the same group. Based on the findings by Messinger et al. [13], a distance range of 0.2 to 2 meters is indicative of social contact among children. Therefore, we adopted this range as our initial consideration for the sigma value.

Given the absence of annotations in our dataset, we could not optimize sigma using standard metrics and loss functions to evaluate the performance of the extracted F-formations. Instead, we plotted all F-formation extraction results for a random timestamp with sigma values ranging from 0.2 to 2.0 for both features: position only and position plus orientation. These results are presented in Figures 5.2 and 5.3.

From these figures, we observe that as sigma increases from 0.2 to 2.0, the clusters gradually expand. At lower sigma values (e.g., 0.2), the clusters are tight and compact, potentially constraining the variability within each group too much. As sigma increases, the clusters become more dispersed. A sigma value between 0.8 and 1.0 appears optimal, offering a balanced representation where the clusters are neither too rigid nor too dispersed, effectively capturing group formations. Coupled with the radial distribution function shown in Figure 2.1, which peaks at a radius of 1, we ultimately selected a sigma value of 1.0.

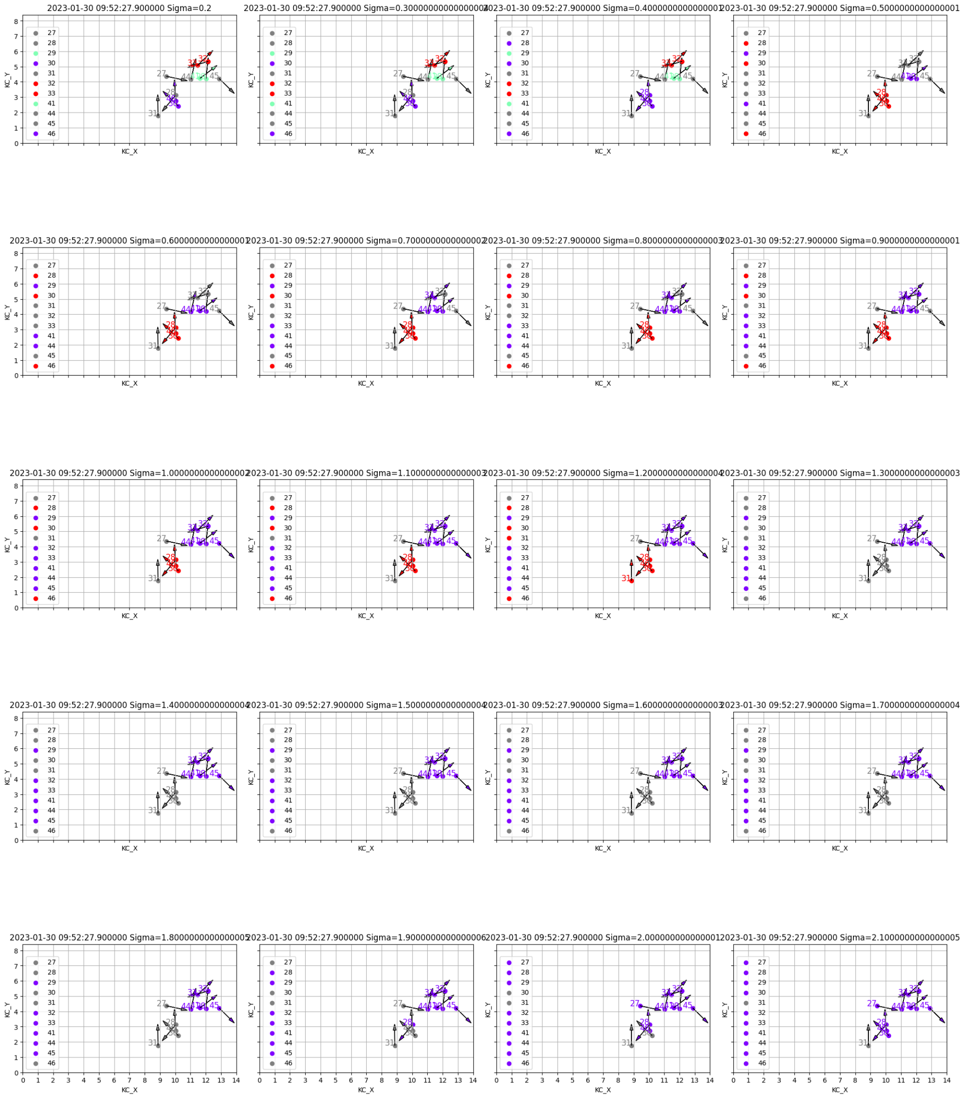


Figure 5.2: F-formations extracted at Time: 2023-01-30 09:52:27.900000, with sigma values ranging from 0.2 to 2.1, using position data only.

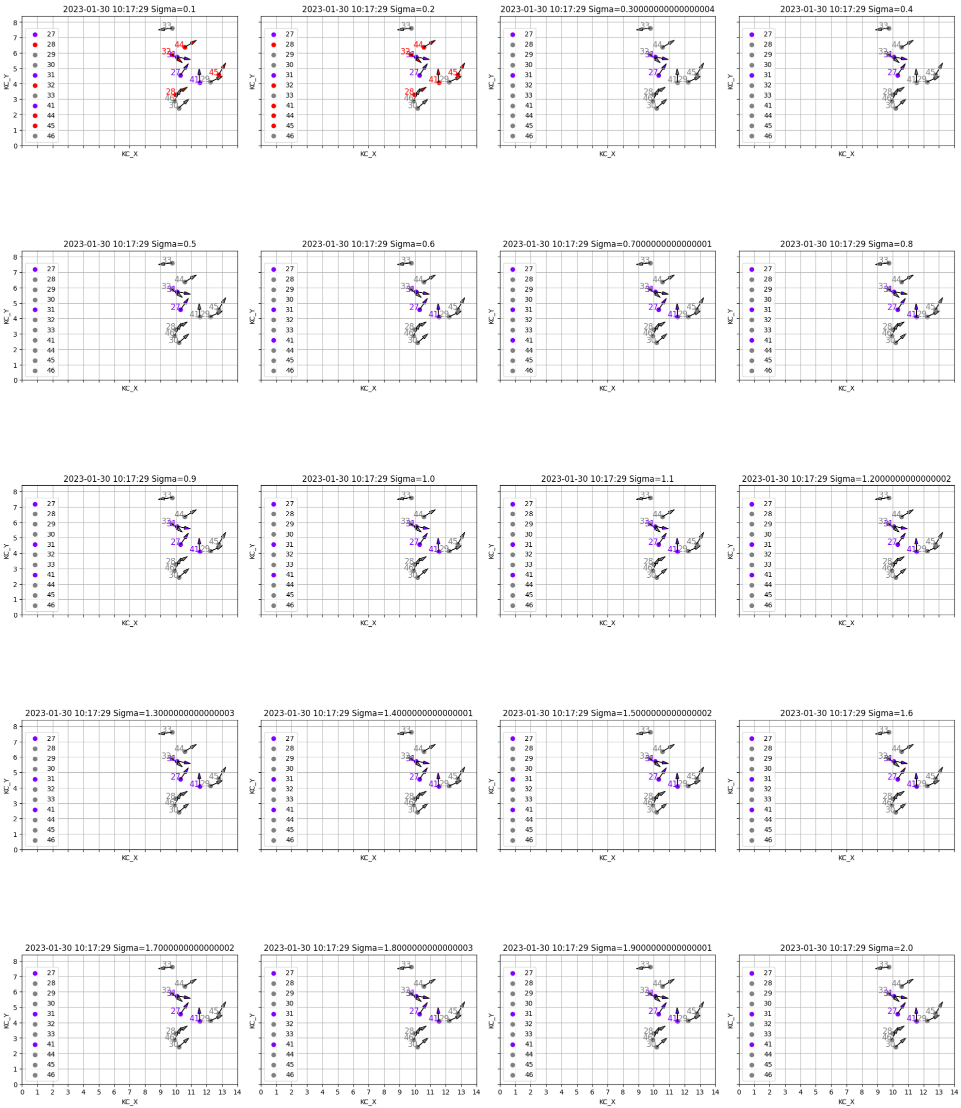


Figure 5.3: F-formations extracted at Time: 2023-01-30 09:52:27.900000, with sigma values ranging from 0.2 to 2.1, using both position and orientation data.

5.3. GROUP ANALYSIS

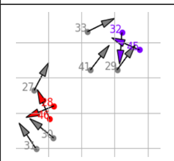
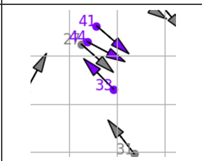
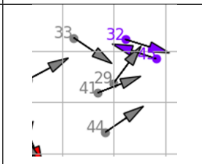
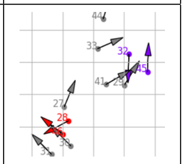
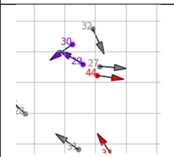
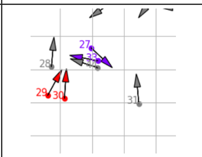
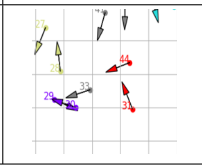
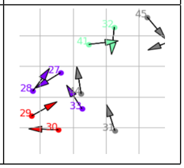
To address the proposed research questions, we conducted several experiments with the extracted F-formations per second and analyzed the group properties to investigate the influence of the homophily effect.

5.3.1. RQ1: DOES F-FORMATION OCCUR IN CHILDREN?

We explore this question by 1) presenting the clustered results and analyzing their representations, and 2) examining how the group analysis in the next subsection impacts the dyadic study. In this subsection, we first display the clustered results from three frequently occurring clusters extracted using positional and orientational data. During this process, we found out that frequent groups are all 2 or 3 people sized groups. This observation aligns with later studies indicating that children tend to spend more time in smaller groups.

From the results in Table 5.2, we observe that for 2-person groups, children do form typical F-formation arrangements such as L-shaped (a), side-by-side (b), and face-to-face (c) configurations. However, there are also instances (d) where the arrangement does not qualify as an F-formation due to a lack of common focused attention. For example, in Group 32,45 (column 3), their body orientations are parallel. Despite being detected by the Dominant Set (DS) clustering algorithm due to their spatial proximity within each individual's frontal area, it is unclear whether they are interacting or merely passing by each other. The absence of head orientation data over time limits our ability to accurately interpret these interactions.

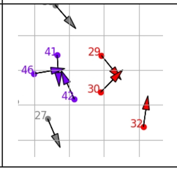
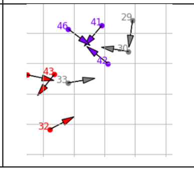
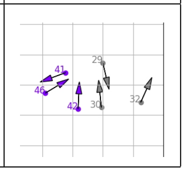
Table 5.2: Examples of two-sized most frequent clusters

Groups / F-formations				
Group	(a)L	(b)Side-by-Side	(c)Face-to-Face	(d)
32, 45				
29, 30				

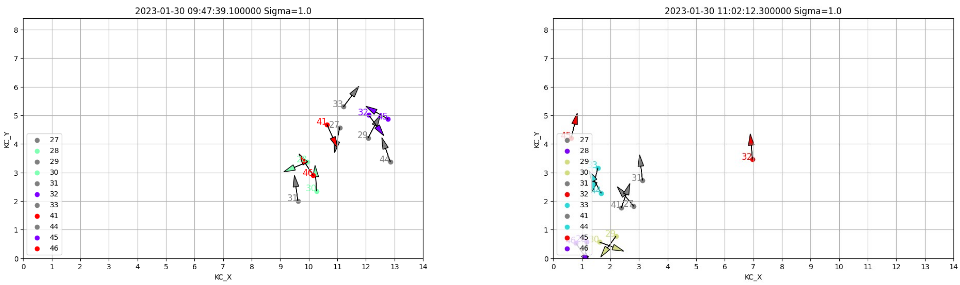
The observations for 3-person clusters, as shown in Table 5.3, reveal similar patterns. We can identify circular and semi-circular arrangements, yet in some cases (c), a common focus is not immediately evident.

Additionally, we encountered some irrationally clustered groups, illustrated in Figure 5.4. In case 5.4a, individuals 41 and 46 are grouped together, while 27, who is visibly closer to 41, is left out. Similarly, in case 5.4b, individuals 45 and 32 are grouped despite being quite far apart. We suspect these anomalies arise from the noise present in our

Table 5.3: Examples of three-sized most frequent clusters

Groups / F-formations			
Group	(a)Circle	(b)Semi-Circle	(c)Abnormal
41, 42, 46			

data, particularly in the orientation measurements, as the DS algorithm is known to be sensitive to such noise [18].



(a) Irrational Case 1

(b) Irrational Case 2

Figure 5.4: Examples of irrationally clustered results.

Overall, from the extracted clusters, we can identify five typical f-formation arrangements. However, some instances violate the definition of f-formation due to the lack of common focused attention. To partially answer RQ1, we conclude that f-formations can indeed be observed among children; however, our results also reveal irrationally extracted clusters. This may indicate two possibilities: 1) the Dominant Set (DS) algorithm's sensitivity to noisy orientation data, and 2) that children frequently break f-formations during interactions.

5.3.2. RQ2: Does the Homophily Effect in Dyads Still Hold for Children with Respect to Hearing Conditions?

The homophily effect has been observed in children with ASD, as reported in the work of Banerjee et al. [1]. However, their findings were generalized to all cognitive conditions. To further investigate this, we aimed to confirm the existence of the homophily effect in children concerning their hearing conditions by conducting a similar mixed-effects analysis on the Starfish dataset.

As shown in Table 5.4, our results indicate that the homophily effect cannot be confirmed in this context ($p = 0.086$). Specifically, there is insufficient evidence to conclude that children tend to spend more time with others who share the same hearing condition. Furthermore, no significant differences were observed in social contact patterns between pairs of children with hearing loss (HL-HL) compared to those with typical hearing (TH-TH) ($p = 0.322$). In contrast, the original study reported significant findings regarding ASD groups, with p-values less than 0.001, as shown in Table 2.1.

To answer our RQ2, we conclude that the homophily effect cannot be verified in the context of children's groups concerning their hearing conditions. We suspect that this lack of confirmation may be due to the limitations of dyadic analysis in capturing social patterns, which prompted us to extend our analysis to a group context.

Table 5.4: Homophily effect for children with respect to hearing loss conditions.

Predictors	**Estimates**	**std. Error**	**CI**	**Statistic**	**p**
(Intercept)	0.02	0.01	0.01 – 0.04	3.19	**0.001**
diagnosisPerson1 [HL]	0.00	0.01	-0.01 – 0.02	0.38	0.703
Homophily [same]	0.01	0.01	-0.00 – 0.03	1.72	0.086
diagnosisPerson1 [HL] × Homophily [same]	-0.01	0.01	-0.02 – 0.01	-0.99	0.322
Random Effects					
σ^2	0.00				
τ_{00} Subject	0.00				
ICC	0.01				
N Subject	12				
Observations	684				

5.3.3. RQ1 & RQ3: DOES ADDING GROUPING CONTEXT FURTHER VERIFY THE HOMOPHILY EFFECT?

HOMOPHILY DEGREE

Incorporating the first group property, homophily degree, within a group setting, we obtained the results presented in Table 5.5. From the table, it is clear that homophily degree significantly influences the social contact time of children within groups ($p < 0.001$). By comparing with the dyadic results from Table 5.4, this finding suggests that group-level dynamics capture more nuanced social patterns compared to dyadic analysis alone. Specifically, the results indicate that children tend to spend more time in groups that are more homogeneous, further supporting the homophily effect in children w.r.t. hearing conditions.

Table 5.5: Impact of Homophily Degree on Social Contact Ratio in Group Settings

Predictors	Estimates	std. Error	CI	Statistic	p
(Intercept)	-0.00	0.00	-0.00 – -0.00	-8.09	**<0.001**
diagnosisPerson [HL]	0.00	0.00	-0.00 – 0.00	0.20	0.843
Homophily degree	0.00	0.00	0.00 – 0.00	14.03	**<0.001**
diagnosisPerson [HL] × Homophily degree	0.00	0.00	-0.00 – 0.00	1.17	0.240
Random Effects					
σ^2	0.00				
τ_{00} group	0.00				
τ_{00} person	0.00				
ICC	0.98				
N person	13				
N group	4022				
Observations	19399				

HL RATIO

Next, we examined how an individual's hearing condition interacts with group composition by analyzing the HL ratio within groups (Table 5.6). The results indicate that when children with hearing loss are part of a group with a higher proportion of peers who also have hearing loss, they tend to spend more time in that group.

GROUP SIZE

In addition to homophily degree and HL ratio, we also explored the effect of group size on social contact patterns. Based on the extracted results, children were found to spend most of their time in smaller groups of two or three. This suggests that group size is another key factor in children's social interactions.

As shown in Table 5.7, increasing group size has a negative effect on the time children spend in the group. This trend highlights that smaller groups may facilitate more concentrated social contact.

ANSWERING RQ1 AND RQ3

In summary, by incorporating group-level properties (homophily degree, HL ratio, and group size), we can verify the existence of the homophily effect in children with respect

Table 5.6: Impact of HL Ratio on Social Contact Ratio in Group Settings

Predictors	**Estimates**	**std. Error**	**CI**	**Statistic**	**p**
(Intercept)	0.00	0.00	0.00 – 0.00	10.19	**<0.001**
diagnosisPerson [HL]	0.00	0.00	-0.00 – 0.00	0.25	0.804
HL ratio	-0.00	0.00	-0.00 – 0.00	-0.17	0.865
diagnosisPerson [HL] × HL ratio	0.00	0.00	0.00 – 0.00	2.34	**0.019**
Random Effects					
σ^2	0.00				
τ_{00} group	0.00				
τ_{00} person	0.00				
ICC	0.98				
N person	13				
N group	4022				
Observations	19399				

Table 5.7: Impact of Group Size and Hearing Conditions on Social Contact Ratio

Predictors	**Estimates**	**std. Error**	**CI**	**Statistic**	**p**
(Intercept)	0.01	0.00	0.01 – 0.01	30.09	**<0.001**
diagnosisPerson [HL]	0.00	0.00	0.00 – 0.00	5.65	**<0.001**
Group Size	-0.00	0.00	-0.00 – -0.00	-27.72	**<0.001**
diagnosisPerson [HL] × Group Size	-0.00	0.00	-0.00 – -0.00	-6.15	**<0.001**

to hearing conditions. While dyadic analysis alone did not capture these social patterns, the group context highlights the importance of homophily in shaping interactions.

These results demonstrate that:

- The homophily effect becomes more apparent in group settings, as homogeneous groups tend to have higher social contact times.
- The analysis indirectly supports the existence of f-formations in children.
- Group properties, homophily degree, HL ratio and group size, are essential for understanding children's social dynamics, which are often overlooked in simpler pairwise interactions.

Thus, the addition of group-level context significantly enhances our ability to detect and verify the homophily effect, providing a more comprehensive view of social behavior among children with hearing conditions.

5.3.4. RESULTS EXTRACTED USING BOTH POSITIONS AND ORIENTATIONS

The analysis presented earlier focused solely on the position feature, which, according to Setti et al. [18], is sufficient for extracting group formations, and DS clustering has been shown to perform well using only position data. However, since orientation is a key aspect of the f-formation definition, we conducted the same analysis by including orientation features. This allowed us to explore how homophily degree and the HL ratio interact with hearing conditions in relation to the time children spend in social contact.

The results presented in Table 5.8 show that when both position and orientation data are considered, the HL ratio no longer has a significant effect on social contact time in groups, nor does the interaction between HL ratio and hearing condition. This contrasts with the position-only results, where some significant relationships were observed.

In Table 5.9, we observe that homophily degree remains a significant predictor of social contact time, even when orientation data is included. Additionally, the interaction between homophily degree and hearing condition shows a marginally significant negative effect (p = 0.049), suggesting that for children with hearing loss, being in a homogeneous group might reduce the time spent in social contact. This nuanced result was not apparent when using position data alone.

5

Table 5.8: Impact of HL Ratio and Hearing Condition on Social Contact Ratio

Predictors	**Estimates**	**std. Error**	**CI**	**Statistic**	**p**
(Intercept)	0.00	0.00	0.00 – 0.00	6.07	**<0.001**
diagnosisPerson [HL]	0.00	0.00	-0.00 – 0.00	0.13	0.893
HL ratio	-0.00	0.00	-0.00 – 0.00	-0.32	0.746
diagnosisPerson [HL] × HL ratio	0.00	0.00	-0.00 – 0.00	1.00	0.318

Table 5.9: Impact of Homophily Degree and Hearing Condition on Social Contact Ratio

Predictors	**Estimates**	**std. Error**	**CI**	**Statistic**	**p**
(Intercept)	-0.00	0.00	-0.00 – -0.00	-2.52	**0.012**
diagnosisPerson [HL]	0.00	0.00	0.00 – 0.00	2.13	**0.033**
Homophily degree	0.00	0.00	0.00 – 0.00	5.38	**<0.001**
diagnosisPerson [HL] × Homophily degree	-0.00	0.00	-0.00 – -0.00	-1.97	**0.049**

The inconsistency in results between the analyses using position-only data and those incorporating both position and orientation suggests that the estimated orientations might introduce noise into the DS clustering process, reducing its accuracy in capturing children's social interactions. This reinforces the idea that while orientation is conceptually important in defining f-formations, in practice, the precision of orientation estimates may not be reliable enough to improve the detection of social groups in this dataset.

5.3.5. CONCLUSION

Based on the findings from both position-only and position-plus-orientation analyses, we conclude that children tend to spend the most time in smaller groups, particularly those with two or three members, and they generally prefer smaller groups over larger ones. Moreover, children with hearing loss show a preference for spending more time in groups with a higher proportion of peers who also have hearing loss, as indicated by the HL ratio. Additionally, children exhibit a clear preference for more homogeneous groups, where they share more similar characteristics with their peers. Finally, incorporating orientation data leads to less reliable results compared to using position data alone, likely due to noise in the estimated orientations. This suggests that, at least in this dataset, orientation may not provide additional value for detecting children's social interactions beyond what position data already captures.

These conclusions underscore the importance of group context in understanding children's social contact patterns. They also highlight that f-formation analysis, while conceptually based on both position and orientation, may face practical limitations when the orientation data is noisy or imprecise.

6

Conclusion

6.1. Summary of Observations and Analysis

In this thesis, we investigated children's social interactions in group contexts, with a specific focus on f-formations and the potential influence of the homophily effect, particularly regarding hearing conditions. Utilizing the Dominant Sets (DS) clustering algorithm, we extracted group formations from real-world datasets, the Starfish dataset, and analyzed their properties in terms of proximity, orientation, and group homogeneity.

Through our analyses, several key insights emerged:

Firstly, we demonstrated that f-formations, as defined in adult interactions, can indeed be observed among children. However, we also found instances where children's social behavior violated the classic definition, particularly in the lack of common focused attention. This suggests that while f-formations exist among children, they may be more fluid and subject to breaking more frequently than in adult interactions.

Secondly, the homophily effect, in terms of hearing conditions, could not be confirmed using dyadic analysis. However, our group-level analysis verified its existence. Children with hearing loss (HL) were found to prefer spending time in groups with a higher proportion of HL peers, and they also spent more time in more homogeneous groups. This effect was particularly strong when homophily degree was considered, confirming that group-level dynamics reveal more about children's social patterns than dyadic interactions alone.

Thirdly, the analysis showed that children predominantly prefer smaller groups, especially groups of size 2 or 3. As the group size increased, the time children spent in that group decreased, indicating that smaller groups might foster more interaction and closer social bonds.

Finally, while positions alone were effective for extracting f-formations and analyzing social contacts, incorporating orientations into the analysis introduced noise, leading to less accurate clustering results. This suggests that the available orientation data may not be reliable enough for precise interaction analysis in our dataset.

In conclusion, this study demonstrates that group-level analysis, particularly incorporating features such as homophily degree, provides a deeper understanding of children's social interactions, revealing patterns that dyadic analyses may overlook. Furthermore, it confirms the existence of f-formations in children while highlighting the dynamic and sometimes irregular nature of these formations. Our findings have implications for future research on children's group dynamics, suggesting that both group composition and homogeneity play crucial roles in shaping social interactions.

.

6.2. LIMITATIONS AND FUTURE WORK

While this study offers insights into children's group dynamics and social interaction patterns, several limitations need to be addressed. These limitations suggest potential areas for future research, which could further improve our understanding of social formations in children's groups.

One limitation is related to the scope of the participant group. Throughout the study, we use the terms "children" and "preschoolers" interchangeably. However, children at different developmental stages—such as toddlers, preschoolers, and teenagers—can exhibit vastly different social behaviors and group interaction patterns [16]. By focusing solely on preschoolers, our analysis does not fully capture the diversity of social behaviors across childhood and adolescence. Future studies could extend this research to include a wider range of age groups, allowing for a more comprehensive understanding of how social patterns evolve with age.

Another key limitation concerns the specific focus on F-formations. F-formations are typically used to describe conversational group structures, but children's social interactions often go beyond verbal communication. Physical activities like ball games, cooperative play, or other types of non-conversational interactions are common among children and are not adequately captured by F-formation detection alone. Moreover, as our results show, many detected groups do not conform strictly to the definition of an F-formation. This suggests that future work could expand the scope of analysis to include alternative group formations, which might provide deeper insights into children's social dynamics, especially in non-conversational settings.

A third limitation is the static nature of the F-formation analysis conducted in this study. Our current approach analyzes group formations at discrete timestamps, which aligns with previous dyadic mixed-effects models. However, this method overlooks the continuous, fluid nature of social interactions. Group dynamics, particularly among children, are highly dynamic and evolve over time. To gain a fuller understanding of these evolving social interactions, future work could incorporate temporal dynamics, tracking how F-formations emerge, dissolve, and change over time. Studies like Swofford et al. [23] provide examples of how temporal analysis could be employed to capture this richness.

Furthermore, since our graph-based approach provides flexibility in constructing the affinity matrix, it could be enhanced by integrating additional contextual factors. For instance, the layout of furniture, the ongoing activities, or the vocalization data captured in the environment could be incorporated into the model. These factors are known to influence how children form groups, and their inclusion could improve the accuracy of group detection.

In terms of methodology, we chose to use the Dominant Set (DS) algorithm for group detection, drawing from Setti's work [18]. However, as we observed in our study, the orientation data was noisy, reducing the accuracy of the extracted groups. To address this, future studies could introduce robustness measures to account for noisy data or compare different F-formation extraction methods. Additionally, Hung et al. [9] proposed an alternative approach to estimating orientations based on focused attention, which might offer a less noisy solution for future applications.

In conclusion, while this work lays the groundwork for understanding group dynamics in children's social interactions, expanding the age range, incorporating alternative group formations, addressing the fluidity of social interactions, and improving orientation estimation will be crucial for future research.

7

Statement

We ensured that all ethical guidelines were followed throughout this research, with parental consent obtained for tracking children's data. The children wore the tracking devices voluntarily.

From a responsible computer science perspective, we ensured that the algorithms and computational methods used in this research were applied in a manner that respects the integrity of the data and the participants. The dominant set clustering algorithm was chosen for its non-invasive nature, and we took care to validate its use in this novel context to avoid misinterpretations or bias in the results.

Furthermore, the AI tool,ChatGPT, was utilized solely for language refinement and grammar correction in the thesis writing process. These tools did not contribute to the generation of research findings, data analysis, or conceptual insights.

Additionally, I received significant support from my supervisor, Hayley Hung, and the psychological aspects of the study, as well as the dataset, were provided in collaboration with researchers, Daniel Messinger, Lynn Perry, Laura Vitale and Debasish Sarker from the University of Miami.

Acknowledgements

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Hayley Hung. Her unwavering support and encouragement have been instrumental throughout my research journey. She has always been open to the ideas and approaches I wanted to explore, providing me with the confidence and autonomy to pursue this study. Hayley is not only a brilliant researcher who offered insightful perspectives on my work but also a dedicated and caring mentor. Her patience, kindness, and commitment to my thesis process have been invaluable, and I cannot thank her enough for all the heart and energy she has put into guiding me.

I would also like to extend my heartfelt thanks to all the professors and faculty members at TU Delft who have provided me with invaluable advice both in my academic studies and in life. Their dedication has fostered a healthy and positive learning environment, deepening my passion for computer science.

A special thanks goes to my mother, Wang Xiuping, for her constant financial support and belief in me. Her help has made this journey possible.

Lastly, I want to thank the important friends and peers I have met during my master's program. Studying abroad for the first time was undoubtedly a challenging experience, but their companionship, collaboration, and friendship turned these past two years into the best journey of my life. I am especially grateful to Liu Xinyu, Gao Shuyi, and Zhu Caiwei, who have been my steadfast companions in both study and life during my second year. Your friendship has made me braver, more outgoing, and more willing to step out of my comfort zone. Thank you for pushing me and for making my life so much richer and more wonderful.

Bibliography

[1] Chitra Banarjee et al. "Objective quantification of homophily in children with and without disabilities in naturalistic contexts". In: *Scientific reports* 13.1 (2023), p. 903.

[2] Dan Bohus, Sean Andrist, and Eric Horvitz. "A study in scene shaping: Adjusting F-formations in the wild". In: *Proceedings of the 2017 AAAI Fall Symposium: Natural Communication for Human-Robot Collaboration*. 2017.

[3] Lindsey Cameron and Harriet R Tenenbaum. "Lessons from developmental science to mitigate the effects of the COVID-19 restrictions on social development". In: *Group Processes & Intergroup Relations* 24.2 (2021), pp. 231–236.

[4] Joe Connolly, Nathan Tsoi, and Marynel Vázquez. "Perceptions of Conversational Group Membership based on Robots' Spatial Positioning: Effects of Embodiment". In: *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 2021, pp. 372–376.

[5] Marco Cristani et al. "Human behavior analysis in video surveillance: A social signal processing perspective". In: *Neurocomputing* 100 (2013), pp. 86–97.

[6] Marco Cristani et al. "Social interaction discovery by statistical analysis of f-formations." In: *BMVC*. Vol. 2. 4. 2011, pp. 10–5244.

[7] Regina M Fasano et al. "A granular perspective on inclusion: Objectively measured interactions of preschoolers with and without autism". In: *Autism research* 14.8 (2021), pp. 1658–1669.

[8] Hooman Hedayati, Daniel Szafir, and Sean Andrist. "Recognizing f-formations in the open world". In: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2019, pp. 558–559.

[9] Hayley Hung and Ben Kröse. "Detecting f-formations as dominant sets". In: *Proceedings of the 13th international conference on multimodal interfaces*. 2011, pp. 231–238.

[10] Adam Kendon. *Conducting interaction: Patterns of behavior in focused encounters*. Vol. 7. CUP Archive, 1990.

[11] Adam Kendon. "Spacing and Orientation in Co-present Interaction". In: *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers*. Ed. by Anna Esposito et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–15. ISBN: 978-3-642-12397-9. DOI: 10.1007/978-3-642-12397-9_1. URL: https://doi.org/10.1007/978-3-642-12397-9_1.

[12] Paul Marshall, Yvonne Rogers, and Nadia Pantidi. "Using F-formations to analyse spatial patterns of interaction in physical environments". In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 2011, pp. 445–454.

[13] Daniel S Messinger et al. "Continuous measurement of dynamic classroom social interactions". In: *International Journal of Behavioral Development* 43.3 (2019), pp. 263–270.

[14] Massimiliano Pavan and Marcello Pelillo. "Dominant sets and pairwise clustering". In: *IEEE transactions on pattern analysis and machine intelligence* 29.1 (2006), pp. 167–172.

[15] Lynn K Perry et al. "Reciprocal patterns of peer speech in preschoolers with and without hearing loss". In: *Early childhood research quarterly* 60 (2022), pp. 201–213.

[16] Gloria Quiñones, Avis Ridgway, and Liang Li. "Holding hands: Toddlers' imaginary peer play". In: *Peer Play and Relationships in Early Childhood: International Research Perspectives* (2020), pp. 77–92.

[17] Omar Adair Islas Ramírez et al. "Modeling the dynamics of individual behaviors for group detection in crowds using low-level features". In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2016, pp. 1104–1111.

[18] Francesco Setti, Hayley Hung, and Marco Cristani. "Group detection in still images by F-formation modeling: A comparative study". In: *2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*. IEEE. 2013, pp. 1–4.

[19] Francesco Setti et al. "F-formation detection: Individuating free-standing conversational groups in images". In: *PloS one* 10.5 (2015), e0123783.

[20] Francesco Setti et al. "Multi-scale F-formation discovery for group detection". In: *2013 IEEE International Conference on Image Processing*. IEEE. 2013, pp. 3547–3551.

[21] Sonia Sultana and Tawhida Jahan. "Physiological Effects of Pandemic COVID-19 On Children With Autism Spectrum Disorder In Bangladesh". In: *Social Science Review* 38.1 (2021), pp. 203–216.

[22] Anchen Sun et al. *Who Said What? An Automated Approach to Analyzing Speech in Preschool Classrooms*. 2024. arXiv: 2401.07342 [eess.AS]. URL: https://arxiv.org/abs/2401.07342.

[23] Mason Swofford et al. "DANTE: deep affinity network for clustering conversational interactants". In: *arXiv preprint arXiv:1907.12910* (2019).

[24] Sydney Thompson et al. "Conversational group detection with graph neural networks". In: *Proceedings of the 2021 International Conference on Multimodal Interaction*. 2021, pp. 248–252.

[25] Sebastiano Vascon et al. "Detecting conversational groups in images and sequences: A robust game-theoretic approach". In: *Computer Vision and Image Understanding* 143 (2016), pp. 11–24.

[26] Marynel Vázquez, Aaron Steinfeld, and Scott E Hudson. “Parallel detection of conversational groups of free-standing people and tracking of their lower-body orientation”. In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2015, pp. 3010–3017.

[27] Lu Zhang and Hayley Hung. “Beyond f-formations: Determining social involvement in free standing conversing groups from static images”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 1086–1095.

[28] Lu Zhang and Hayley Hung. “On social involvement in mingling scenarios: Detecting associates of f-formations in still images”. In: *IEEE Transactions on Affective Computing* 12.1 (2018), pp. 165–176.