

Sensing Interpersonal Synchrony between Actors and Autistic Children in Theatre Using Wrist-worn Accelerometers

Jamie A Ward¹, Daniel Richardson¹, Guido Orgs², Kelly Hunter³, Antonia Hamilton¹

⁽¹⁾ University College London, ⁽²⁾ Goldsmiths, ⁽³⁾ Flute Theatre
London, UK. Correspondence: jamie@jamieward.net

ABSTRACT

We introduce a method of using wrist-worn accelerometers to measure non-verbal social coordination within a group that includes autistic children. Our goal was to record and chart the children's social engagement – measured using interpersonal movement synchrony – as they took part in a theatrical workshop that was specifically designed to enhance their social skills. Interpersonal synchrony, an important factor of social engagement that is known to be impaired in autism, is calculated using a cross-wavelet similarity comparison between participants' movement data. We evaluate the feasibility of the approach over 3 live performances, each lasting 2 hours, using 6 actors and a total of 10 autistic children. We show that by visualising each child's engagement over the course of a performance, it is possible to highlight subtle moments of social coordination that might otherwise be lost when reviewing video footage alone. This is important because it points the way to a new method for people who work with autistic children to be able to monitor the development of those in their care, and to adapt their therapeutic activities accordingly.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

wearable sensing; accelerometers; theatre; autism; interpersonal synchrony; cross wavelet

INTRODUCTION

Autism, or autism spectrum condition (ASC), refers to a range of developmental conditions that are characterised by difficulties with social interaction. People with ASC can struggle with non-verbal communication, including the use of gaze, imitation, and other social cues. These difficulties might arise during infancy, when ASC children are slower to grasp social signals, and fail to imitate or copy the movements of others in

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ISWC '18, October 8–12, 2018, Singapore, Singapore

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-5967-2/18/10...\$15.00

DOI: <https://doi.org/10.1145/3267242.3267263>

the same way that neurotypical children do [3]. This creates a barrier to both learning and social engagement.

An important measure of a person's social engagement is the degree to which they move in synchrony with others. Interpersonal synchrony – the non-random temporal coordination of two or more people [5] – is found to be less prevalent in ASC than in neurotypicals. Motion-capture technology was used to show this by measuring movement coordination (or lack of) between interactants [8]. Yet most studies of this kind are lab-based, and primarily focus on dyads rather than groups.

There are of course instances when even those with severe autism are able to engage and interact socially. Indeed many adults with ASC develop mechanisms to help them navigate social situations. It follows that moments of improved interpersonal synchrony do occur in ASC, however fleetingly. Capturing these moments might allow us to pinpoint instances of social engagement, both from the point of view of understanding the mechanisms at play, and as a way of charting an individual's development.

Wearable sensing provides an opportunity to take research on interpersonal synchrony and autism out of the lab and into "the wild". Body-worn accelerometers can be used to track ASC children – and everyone they typically interact with – as they go about their everyday life. The ecological richness of such data comes at a cost to privacy.

We propose an alternative approach: using wearables to study synchrony in autism during an interactive theatrical performance. Theatre can provide a snapshot of real-life, with many social situations and interactions built-in. It also has the research advantage of being in a constrained space, with a finite number of participants, and the opportunity to repeat the experience over several performances.

The work of Flute Theatre has been significant in transforming perceptions of ASC and what children with ASC are capable of. Flute's approach, developed over 20 years of working with autism, is to help autistic children learn how to synchronise and engage socially through a series of theatrical games [12]. These games draw on Shakespeare's plays, making use of the rhythm of the language, and exaggerated physical actions of the actors. Through these games, many children, some of whom are profoundly autistic with minimal language, learn to perform in front of an audience – sometimes to the surprise of

their parents and carers. Yet we do not know exactly why these games work, or what the neural and psychological mechanisms at play are. As a first investigation of this, we use wearable sensing to record the movements of actors and ASC children during a Flute performance, and use the data to chart their interpersonal synchrony.

In this paper we present an approach to collecting physical movement data in the challenging environment of a live theatrical performance involving autistic children. We describe a novel application of cross-wavelet analysis for exploring the interpersonal synchrony of up to 10 participants. We then demonstrate how this information can be used as a measure of social engagement to aid researchers studying autism, both as a way of charting a child’s development, and as a mechanism for automatically annotating videos of long-term interaction.

Related Work

Interpersonal synchrony measures the dynamics of interaction between people rather than the specific nature of their behaviours. It is more concerned with the temporal coordination and shared rhythm between interactants, rather than how they mirror, or imitate, one another [5]. It was originally studied by developmental psychologists, with early work attempting to quantify bodily synchrony by looking at stills of a movie [4]. Thanks to advances in sensing technology, interpersonal synchrony has become a topic of research in fields such as machine learning, robotics, and human-centered computing [5]. When people move together in synchrony, they tend to have greater rapport with one another [16]. Synchrony has been shown to be an important component in enhancing the success of joint goals [23]. And it has also been shown to effect affiliation in human-robot interactions [13].

In ASD, correlations were found between the ability to synchronise movement with others and sentence production [8]. The same work showed that ASD children are both less able to synchronise socially with others, and that their manner of movement when imitating is different. In another study autistic children who were sat on rocking chairs next to their caregivers were found to be less likely to rock in-phase than neurotypical children [20].

Wearables are a promising tool for researching interpersonal synchrony. The accelerometers built into Google Glass, for example, were used to measure dyadic synchrony during conversation [17]. And [24] used wrist-worn sensors (the same E4 devices used in the current work) to demonstrate how large groups of people moving in sync can enhance group affiliation.

There is also much potential for wearable applications that support and diagnose people with autism [6]. Google Glass has been explored as a tool to help ASC children with facial expression recognition [26]. Machine learning methods have been applied to wearable sensor data to automatically recognise stereotypical stimming behaviour in autistic participants [2, 28]. Similarly, accelerometer-based features were used to classify aggressive and self-harming behaviours in autism [18]. The focus of our current work, however, is not to automatically recognise specific behaviours, but rather to try and



Figure 1. Interacting with children, and actor-only performance.

uncover instances of interpersonal synchrony that might be hidden behind those behaviours.

Earlier work measured coordinations in dyadic body movement using Pearson’s cross-correlation (CC) [19]. CC, a measure of time-series similarity, is limited because it cannot capture synchrony across different frequencies.

Here we measure synchrony over time and frequency using a cross-wavelet analysis of participant’s wrist movement intensities [22]. Cross-wavelet analysis is a windowless method for comparing the similarity of two time-series in both time and frequency domain – and is well suited to analysing the co-synchrony of two movement signals. Historically, cross-wavelets were used to study co-variations in weather patterns [22, 15]. Recently, cross-wavelets were used to analyse movement features from dyads in conversation, and revealed how people tend to synchronise with one another at specific frequencies [9]. In a theatrical setting, cross-wavelet analysis of video data was also used to show how dancers entrain more effectively to one another than non-dancers [25].

Features based on the wavelet transform have long been used in wearable sensing (e.g. for ambulatory activity recognition [21, 14]). However, the work presented here is the first, that we are aware of, to apply cross-wavelet analysis to wrist-worn accelerometer data from multiple interacting people.

AUTISM AND THE DREAM

During 6 days in mid-March 2018, Flute theatre put on daily performances of Shakespeare’s “A Midsummer Night’s Dream” at the Bridge Theatre in London. Lasting just under 2 hours, each performance was a specially designed interactive workshop for ASC children. Six actors would work together with groups of between 3 and 18 children, all with autism and with a range of learning difficulties, within a round, floor-level stage. The audience, sat in an outer circle around the stage, was comprised of teachers, parents, researchers, and some members of the public.

The structure of the performance is a sequence of actor-only performances, interspersed with interactive re-enactments by the children. The interactive components encourage children to engage in aspects of social coordination with which they might otherwise struggle. Eye contact, for example, is difficult for many with ASC. One of the structured games is based around a moment in the play when Titania follows Bottom in the forest and makes eye-contact when she finds him. This moment is enacted by one actor using both hands to mimic

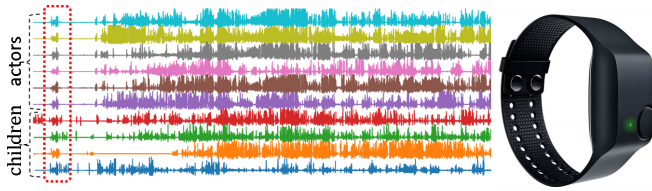


Figure 2. (Left) data from 6 actors and 4 children over ~2h Saturday performance. Synchronisation gestures highlighted. (Right) E4 watch.

cartoon-style eyes-on-stalks, to the exclamation, “Doyoyoying!... I love you!” At this, the person playing Bottom turns away in surprise, puts his hands up to mimic donkey ears, and makes a “Ee-aaw!” sound before walking on. The two actors will repeat this sequence 3 times. The children are then encouraged to step onto the stage, in ones or twos, and re-enact the moment. Sometimes they will be guided by the actors, and sometimes they will be able to do it on their own. A key part of the exercise is giving the children time to repeat the actions several times over, allowing them to connect the exaggerated physical movements with the intentions of the characters.

DATA COLLECTION

Data were collected using wrist-worn E4 sensor bands from Empatica [10]. The E4 can monitor skin temperature, heart rate, electrodermal activity (EDA), and 3-axis acceleration. Here we only use acceleration, sampled at 32 Hz. The E4 has a single, low-profile, button for marking annotations and, when held for 3 seconds, powering off. Because we used the devices as passive sensors, we prevented participants from accidentally switching off the devices by covering them in protective colour sweatbands.

Three days of performances (Thursday to Saturday) were recorded in front of live audiences of between 20 and 30 people. Data were recorded from the 6 actors and from those children for whom parental consent was granted. In summary, 3 children (out of 4) were recorded on Thursday, 3 (out of 12) on Friday, and 4 (out of 4) on Saturday. The Saturday was the only performance where data was obtained from everyone on stage. All participants wore E4 sensors on both wrists.

A camera and microphone, placed in the rigging directly above the stage, recorded each performance. Annotators kept a record of whenever each child stood up to perform, their role, and who they worked with. Usually, each child would have one actor with whom they were paired throughout. Annotations were also kept on whenever there was a shift between the two modes, ‘actors only’ (performing a scene), or ‘interaction’ (repeating scenes with the children). Performances lasted around 2 hours, producing a total dataset of 6 hours.

The 10 recorded children (6 female, 4 male) were aged between 8 and 16, and all have moderate to severe ASC. All but 2 have limited or no verbal ability. Two of the Thursday children used wheelchairs. Due to the conditions of our ethics approval, we are unable to report on specific diagnoses.

Synchronisation

Data from all devices were synchronised in 3 steps. First, the E4 internal clocks were updated each day by connecting the

devices, in sequence, to a reference laptop. (A limitation of the E4 software is that only one device can be connected at a time.) Second, before (and after) each performance all devices are powered on and enclosed in a padded box. The box is flipped sharply a few times to create distinctive spikes in the acceleration – see dotted region on the signals in Figure 2. These spike patterns are used to align all device data by cross-correlating against the pattern from a reference device, shifting signals accordingly to maximise the correlation. Third, to synchronise the video with sensor data, the actors performed a warm-up dance. The well-known Macarena was chosen because of the rhythmic hand gestures and jumps, which create distinctive signals in both the acceleration and video.

Managing Complexity

Comparing multivariate sensor data across many people is difficult – particularly when the coordinated activities we are interested in can be done in many different ways. A simple activity with an identical, but opposite, action, like one person pulling another towards her, will result in variations in the direction and orientation of the individual wrist x,y,&z acceleration signals (one person accelerates backwards, the other forwards). To complicate matters, sensors moved and shifted orientation on the participants’ wrists throughout the performance. To counter this we combine the acceleration axes into a single, orientation-invariant, measure of intensity ($|a| = \sqrt{x^2 + y^2 + z^2}$). At the cost of losing information on the spatial nature of movements (how people move), this measure highlights the temporal dynamics (when people move), which suits our ultimate purpose of tracking synchrony.

MEASURING INTERPERSONAL SYNCHRONY

Our primary measure of interpersonal synchrony is the average cross-wavelet (ACW), a method of summarising the time-frequency similarity of two signals. For comparison, we also calculate cross-correlation (CC) – a measure of how well two signals move together in time. CC, calculated using Pearson’s correlation, returns close to 1 if the signals move together, -1 if they diverge, and 0 if they are uncorrelated.

The cross-wavelet transform of two signals, s_1 and s_2 , is computed as the product of the continuous wavelet transform of one signal, W_1 with the complex conjugate of the other, W_2^* . That is: $W_{1,2} = W_1 * W_2^*$. The cross wavelet power (CW) is then calculated as, $CW = |W_{1,2}|$. The CW used in this work is calculated using the continuous wavelet transform function from the PyCWT module in Python (with Morlet base).¹

The average cross-wavelet power (ACW) provides a convenient measure of similarity between two signals, indicating whether they match at any frequency, and by how much. ACW is calculated by collapsing the CW across frequency: $ACW = \frac{1}{|F|} \sum_f CW^f$, where F are the wavelet frequency bands. F in this work covers the range between 0.025 and 2 Hz (40 to 0.5 second periods). The ACW measure we introduce is an adaptation for cross-wavelets based on the single-wavelet, scale-averaged wavelet power [22]. In the following sections, each ACW is normalized (to between 0 and 1).

¹<https://github.com/regeirk/pycwt>

THE INTERACTION MATRIX

With a single variable representing each person's wrist, $|a|$, we evaluate the similarity in movement between different combinations of pairings using CC and ACW. The process of calculating CC and ACW is highlighted for 30s of data in Figure 3. Acceleration data from two participants (actor *B* and child *k4*) is compared to generate a cross-wavelet transform in the time-frequency domain. This is averaged across frequencies (y-axis of Figure 3-iii) to give an indication of similarity at any frequency (ACW in Figure 3-iv). As a comparison, the cross-correlation between *B* and *k4* is also calculated (using a 5s window, rolled forward in increments of 1/32s). Both similarity measures, when averaged across the full 30s, and calculated for every combination of participant pairings, are then used to create the interaction matrices shown in 3-v.

Case Study: Stimming in Time

The example in Figure 3 allows us to examine interactions in detail. Here we notice at least three interesting findings. Trivially, the darker regions for both ACW and CC (v) show that five actors (*A* to *E*) are tightly coupled. This is to be expected because this section involves those actors moving together. One actor, *F*, and the children, watch from the side.

However, the ACW matrix shows that one of the children, *k4*, appears to be highly coupled in frequency with the actors. Yet he is not moving exactly as they are (as indicated by the low CC in v), and on first glance at the accompanying video he does not seem to be moving at all. Closer examination uncovers what is going on: he is gently stimming his hand at a frequency (~0.5 Hz) that is coupled to the rhythm of the actors. This can be seen in the band of low-frequency coherence with *B* shown in Figure 3-iii.

A third interaction seen in the data is the synchronicity of child *k3* with *k4* (and actors). Although slightly weaker than *k4*, none of the other children showed any synchronicity here. This pattern was spotted at various points throughout the recording, and might be explained, in part, because these two children are twins and have a strong awareness of one another. It should be noted that they were sitting at opposite ends of the stage.

This kind of data can be very valuable to a therapist, because without the accelerometer analysis, it is hard to spot that the behaviour of these children is socially relevant.

Note that the CC analysis fails to capture the clear engagement of *k4* in this example. CC is useful for finding specific temporally-aligned movements (e.g. when an actor would hold and guide a child to perform a specific movement). However, as might be expected from time-based correlation, it does not show the more general synchrony between participants that we are interested in here. For this reason, alongside space constraints, we focus on ACW for the remainder of this paper.

Case Study: Who's with Who

When applied across the entire dataset, interaction matrices can provide a valuable summary of a performance. The ACW matrices for all 3 days, shown in Figure 4, reveal two main findings: First, the actors tend to be strongly in sync with one another. Second, the highest ACW pairings for each child was

with the actor with whom they were paired (e.g. for Saturday: *k1-C*, *k2-D*, *k3-A*, and *k4-E*).

The ACW interaction matrices for the Thursday and Friday also summarize some interesting interactions on those days. On Thursday there were 4 children in total, with only 3 wearing sensors. *k8* was in a wheelchair, and mostly interacted with actors *C* and *D* (as is reflected in the highlighted areas in the lower right of Figure 4). *k9* was also in a wheelchair, but was attended mostly by one of the carers (who did not wear a sensor band), as well as with several of the actors – hence the more even spread of ACW power.

On the Friday, 3 children wore sensors (attended by *C*, *D* and *E*), but there were 9 other children without sensors. Consequently, the overall interactions between actors has lower power than on the other days - presumably because they had a bigger group to work with.

Let There Be Music

Music was another important influence in prompting synchrony. In the Thursday dataset, for example, *k8*, who was in a wheelchair, showed strong synchrony with actors *D* and *E*, yet they rarely interacted physically. It turned out that this connection was strongest when those actors were playing mandolin and drums – *k8* moved her arms in perfect synchrony to the rhythm of the music. Similarly, on the Saturday, a particularly active child, *k2*, who for the most part appeared to do her own thing and rarely engaged with the actors, would occasionally have moments where she synchronised perfectly with musicians – both in time (from the CC analysis), and in rhythm (from ACW). From surveying video footage alone, moments of remarkable synchrony like this are easily missed.

MOMENTS OF ENGAGEMENT

For a teacher or therapist, it is valuable to be able to measure how much and when an individual child is engaged in social interactions, but it is overwhelming to try to visualise every (pairwise) interaction that a child is involved in, particularly for large numbers of participants. Our solution is to combine all of the possible interactions of each person into a single measure that we call engagement. This is calculated by taking the maximum ACW output from each person in turn, to give an indication of his or her overall social engagement, $Eng(t)_i = \max_{k \neq i}^K ACW(t)_{i,k}$ for time t , person i , and K possible interaction partners.

Case Study: 'Doyoyoying!'

As an example of how engagement might be visualised, the scarf plot in Figure 5 shows a 2 minute extract from the scene, described in the introduction, where Titania proclaims her love to Bottom with a cartoon-style, eye-popping, 'Doyoyoying!' (Bottom, uncomfortable with this, turns away with an 'Ee-aaw!', while making donkey ears with his hands.) Each line of the plot represents the engagement for each actor (*A-F*) and child (1-4). The roles played by each person at each time, as recored by our annotators, are also highlighted. The first 20s shows a demonstration by *C*, as Titania, and *F*, as Bottom. This is followed by a sequence of repetitions of the scene involving the children (*b* through *c*).

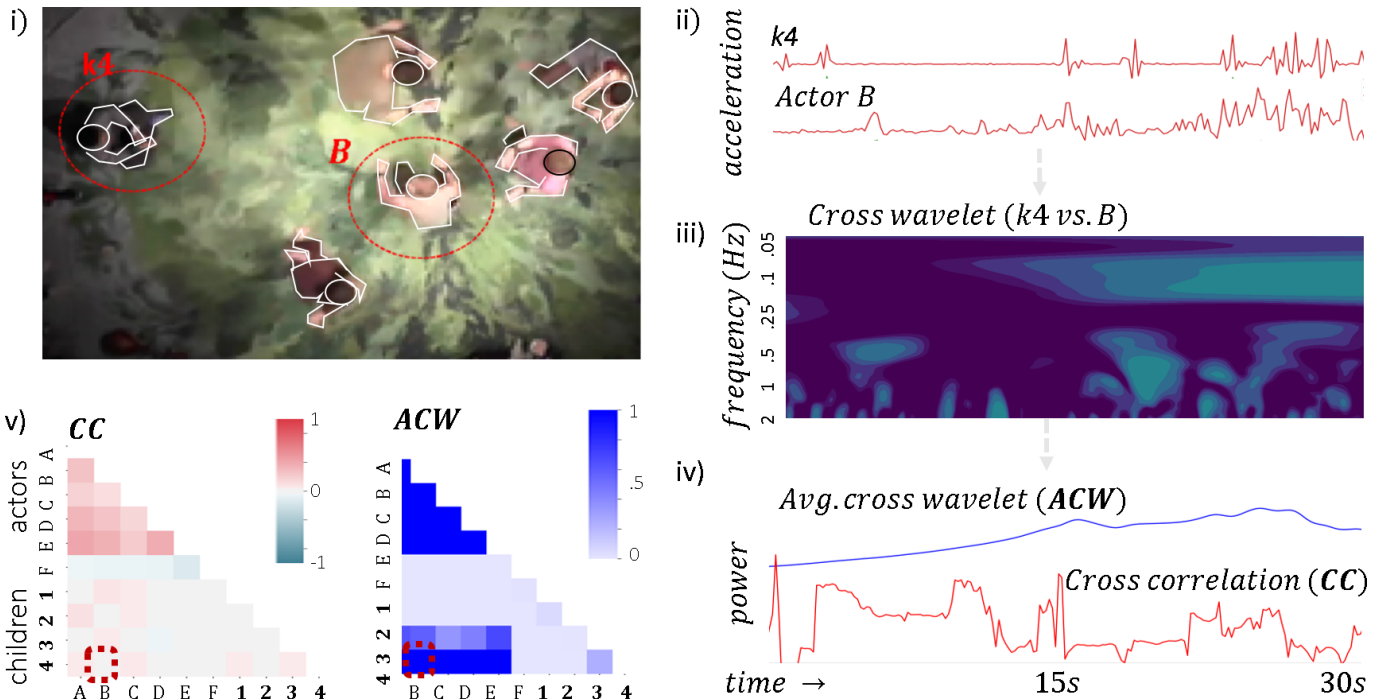


Figure 3. (i) Five actors perform while children (and actor F) watch. (ii) Acceleration from actor *B* and child *k4* over 30s. (iii) Cross-wavelet spectrogram of this data. (iv) Average cross wavelet power (ACW), plotted alongside cross-correlation (CC, calculated using 5s sliding window). (v) Interaction matrices for CC and ACW over 30s period. Note similarity in movement frequencies, but not in temporal correlation, between *B* and 4.

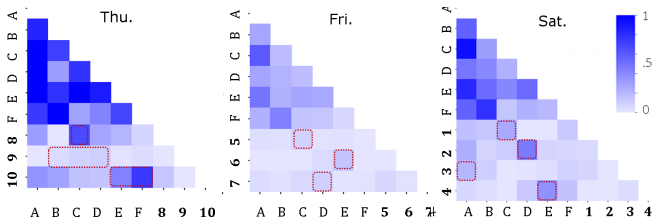


Figure 4. ACW interactions over 2h performances. Dotted squares show main actor-child pairings.

The engagement plot reveals much about the dynamics of this sequence. At time (a), for example, child 3 and A work together to play Titania. Actor *D* takes over as Bottom, and tries to get child 2 to join – but she is not interested, as shown by the low engagement values (lighter colouring) for that child in (a). After coaxing, she eventually joins in – shown by the solid engagement at (b). Finally, they all sit and actor *B* stands up with child 1 to play Titania. This child does not seem to engage with the sequence as a whole, but does manage a ‘Doyoyoying’ towards the middle of (c).

Figure 5 also shows the ACW interaction matrix computed from the entire 2 min sequence. Trivially this reveals supporting relationships (e.g. actor *D* helping child 2 play Bottom), but it also reveals more complex interactions, like connecting actor *D* with child 3, despite both playing opposite one another as different characters.

By combining the synchrony of a child with all possible interactants we can visualise whether or not a child is engaged in some way at any particular moment. Because the underlying

ACW measure does not require exact temporal correlation, and can capture complex interactions at varying time scales, engagement scores highly regardless of whether two people perform the same action, or simply perform an action that is in some way coordinated (e.g. donkey ears in response to Doyoyoying).

Event Spotting: Actors vs Children

Engagement can be used to automatically spot interesting interactions from long recordings. As an example, we build a classifier that can spot whenever the actors perform together, versus when they interact with the children. This is useful information because, when combined with our knowledge about the ordering of scenes (which alternate between actors-only and interactive), we can obtain the exact timings of specific scenes. This could be a useful aid for therapists wishing to survey and annotate behaviour from many hours of recordings.

Figure 6-i shows engagement outputs for the two groups (actors and children) across the entire duration of the Saturday performance. We summarise the engagement for each group by taking the maximum over the group members. Taking the difference of the two gives a measure that we then classify using a simple threshold (iii). This classification is compared to the ground truth from our hand annotation (actors only vs. interactive). Note that because the underlying ACW values are normalised, we choose an arbitrary decision threshold of 0.5 (this works well for this example, though more sophisticated classifiers and training regimes might be used in future).

We evaluate the frame-by-frame classifier outputs for each of the 3 days. Table 1 shows the results in terms of Precision

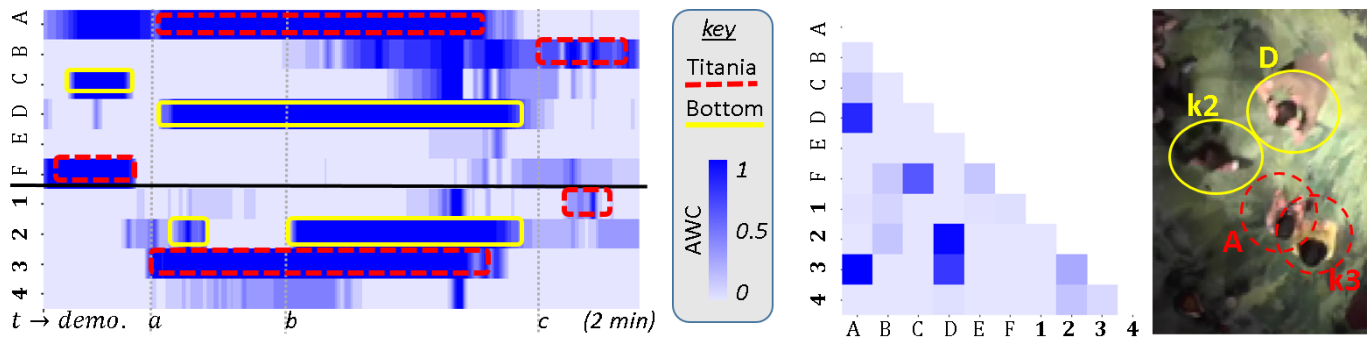


Figure 5. Engagement over 2 minute sequence: actors C (as Bottom) and F (as Titania) demonstrate the Doyoyoying sequence, a) actor A and child 3 then work as Titania, with D trying to coax child 3 to play Bottom, b) both children fully commit to their roles, c) B then helps child 1 take on the role of Titania (who manages a single Doyoyoying). (Center) ACW interaction matrix for the 2 minute scene. (Right) view from middle of the scene (b to c).

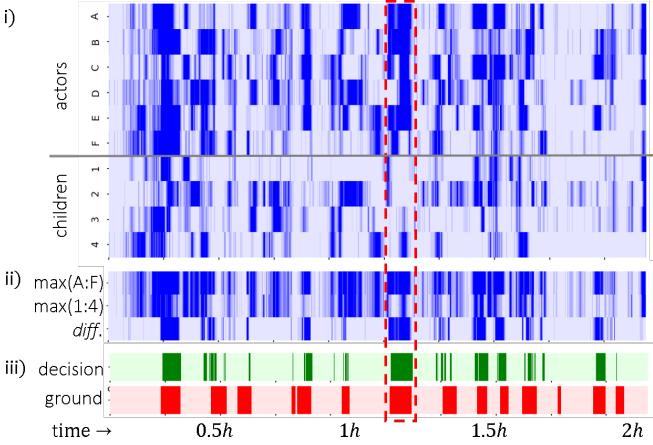


Figure 6. i) Engagement sequence of participants (Saturday). Dark areas indicate strong synchrony with at least one other person. ii) Maximum engagement for actors vs. children, and their difference. iii) Classifier decision (actors-only vs. interaction) compared to ground truth. Dotted area shows correct detection of an actors-only scene.

	Precision	Recall	AUC
Thursday	.72	.64	.80
Friday	.62	.65	.79
Saturday	.73	.67	.84

Table 1. Event spotting results

(proportion of returned frames which are correct) and recall (proportion of ground truth frames correctly detected). We also show area-under ROC curve (AUC), a threshold-independent measure of performance where 1 is perfect, 0.5 is random [7].

An AUC range of .79 to .84 indicates that a simple threshold-based classifier on engagement groupings can be sufficient to pick-out meaningful events from a long dataset.

DISCUSSION

Sensing in Practice

In a multi-person interaction like this, ideally everyone should be wearing sensors. Unfortunately, consent to record sensor data was only granted for 10 out of 20 ASC children. Despite this limitation, the majority of interactions for the sensed children were captured because their main interaction partners were actors (who all wore sensors).

We had several minutes with each child and their carers before the performance to introduce them to the technology, and to gently secure the devices in a way that was not too disturbing². Despite this, some children were irritated by the devices and attempted to discard them during the performance. This proved the main advantage of having sensors on both wrists - if one was removed, there would still be useable data from the other. The only instance where both sensors were lost was with k4 on the Saturday dataset. Thankfully this only happened during the final 20 minutes, as can be seen by the pale area for k4 in Figure 6-i. Considering the difficulty many autistic children experience with new things and situations, we recommend a longer period of adjustment to the devices in future.

Despite all 6 actors being recorded, occasionally some devices would break mid-performance (mostly due to sweat getting into the electronics). Again, the dual-wrist strategy ensured that there was at least one wrist recording for everyone. Choosing data from only one wrist per person (the dominant or non-broken one) does mean that some interactions are lost - e.g. if the other hand is mostly used when interacting. However, we found that ACW comparisons between left and right wrist data (where both exists), showed very high similarity, suggesting the drawback from using only one wrist is minimal.

Requiring everyone to wear a sensor, and to share their data, is an obvious limitation of our approach. As we argue (in Future work, below), this might be acceptable in certain environments, like workshops, therapy, or at home. The problem of getting and keeping the sensors on restless children, however, remains. Automated vision is an alternative to wearable sensing, and has been used previously to study the synchrony of dyadic body movement [5]. Software like OpenPose can perform multi-person tracking using only a single camera [27]. Occlusion can be a problem, however, and it is difficult for even the best algorithms to distinguish between several interacting children (especially when everyone is dressed in the same school uniform, as in our study).

Overall the parents and carers reported that their children had a positive experience. However, due to time constraints on the performance days, and the fact that many of our participants

²The E4 has a difficult to apply snap closing mechanism which did not work well with the children. Instead we attached each device individually using tape.

had limited or no verbal skills, we were unable to record the children's impressions. Future studies would ideally include a more thorough analysis of the children's point of view to help us better interpret the sensor-based measures. It would also be useful to record a performance using non-autistic children – this would provide a baseline for helping to understand any deviations in autism from typical interactions.

Interpersonal Synchrony

Average cross-wavelet is useful for finding times of co-synchrony between two movement signals. However, because the CW transform is biased towards matching high powered signals, some low powered moments of co-ordination might be missed. CW also returns a non-standardized output, which can be difficult to work with. (We get around this in our visualisations by globally normalising the output, although issues with this scaling might affect the generalisability of classifiers trained using this data.) An alternative is to use wavelet coherence, which is better at matching low power synchrony (and always returns a normalised output of between 0 and 1) [11]. In practice, the cross-wavelet approach is useful for this work because we are mostly interested in relatively high-power, hand and body movements.

Although we found time-based correlation (CC) to be of limited use here, the method can be informative as an additional descriptor of the nature of the synchrony taking place. For example, a high CC could indicate an instance of a child being physically guided by an actor, rather than it being a reciprocal engagement. An additional source of information would be on the interaction lag. This can be obtained by applying CC over a range of time shifts, or calculated from the CW phase response. Because ASC children are known to be slow to respond to social stimuli, information on phase and lag could be another marker of engagement [5].

Questions remain as to the precise mechanism revealed by our approach: is the child in our first case study (stimming in time) coordinating socially by watching the actors' movement, or is he responding to the rhythmic sound of that movement? In truth, both modalities play a role. Rhythm and music are used throughout Flute performances, and are an important component in engaging ASC children. But there are many instances where the coordination is primarily a physical one, such as in the 'Doyoyoying' case study. In future work it would be useful to distinguish between different modalities of synchronisation to provide richer feedback on the mechanisms of interpersonal coordination.

Future Work

Unexpected moments of social engagement might not be obvious from video or real-life interaction, but can be spotted using the accelerometer and cross-wavelet approach introduced here.

This work points to a new way for teachers and therapists to be able to measure the development of ASC children in their care. Although the present study is set within the context of theatre, the same sensor-based approach could be applied in a similarly structured environment, like the classroom, or behavioural workshop [6]. After each session, the engagement of each child could be visualised and used to highlight the

moments when they appear to be doing well – and then using this information to guide which activities to focus on in follow-on sessions.

Diagnosing autism in early life is important to ensure that the knowledge and support is in place to assist the individual and his or her carers. One future application of this methodology could be as a diagnosis aid, whereby groups of children wear sensors for a few hours (or longer) during their early play sessions to help identify those who might be struggling with certain social interactions [5].

Although it cannot be assumed that the wider population would wear sensors long-term, it is possible that carers and family of an ASC child might agree to wear them for monitoring and support purposes. Similar to the work of [26], where ASC children are given on-line assistance with interpreting people's faces, future wearables might be used to spot instances of low engagement and to provide on-line support either to the ASC child directly, or to those around her. Combined with automated classification of stereotypical (e.g. [2, 28]), information on interpersonal synchrony might also be used to highlight to others that a person's stimming behaviour is not random, but a response to a specific social situation.

As awareness grows of the importance of non-verbal synchrony in social situations, there is scope for wider deployment of wearable applications that monitor a person's engagement with others. Wearables are being developed that can monitor social rhythms in, for example, bipolar disorder [1]. Similar applications might also be developed to monitor interpersonal coordination patterns more generally, although as mentioned before, issues with privacy and acceptance remain.

Conclusion

This paper demonstrates that it is feasible to record data from autistic children in a social environment, like theatre, using wearable sensing. This data can be used to understand the social activities taking place, and reveals novel insights into behaviour. We show that sensors can be used to identify patterns of coordination by autistic children that are easily missed by human observers. We show how individual differences in social engagement might be quantified, providing a convenient way to chart a child's progress over time. And we show how an aggregate measure of engagement in groups (e.g. actors vs children) might be used to identify events within a recording, easing the process of annotating long video and data recordings. These findings point the way to using wearable sensors with children in classroom, theatre, and therapy contexts. Future work can test if this sensor data can predict how well children with limited verbal skills respond to different therapeutic activities. Such data will be valuable to children with autism, their teachers and families, and supports the development of new assistive technologies for autism.

ACKNOWLEDGMENTS

Many thanks to Roser Canigual, Mina Choi, the cast and crew of Flute theatre, as well as the front-of house and technical team at the Bridge Theatre. This work is supported by the UCL Grand Challenges *Adolescent Lives* scheme.

REFERENCES

1. S. Abdullah, M. Matthews, E. Frank, G. Doherty, G. Gay, and T. Choudhury. 2016. Automatic detection of social rhythms in bipolar disorder. *Jrnl of the American Medical Informatics Association* 23, 3 (2016), 538–543.
2. F. Albinali, M. S. Goodwin, and S. S. Intille. 2009. Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In *UbiComp '09*. ACM, 71–80.
3. T. Charman, J. Swettenham, S. Baron-Cohen, A. Cox, G. Baird, and A. Drew. 1997. Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. *Developmental psychology* 33, 5 (1997), 781.
4. W. S. Condon and W. D. Ogston. 1967. A segmentation of behavior. *Jrnl of psychiatric research* 5, 3 (1967), 221–235.
5. E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen. 2012. Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Trans. on Affective Computing* 3, 3 (July 2012), 349–365.
6. R. el Kaliouby, R. Picard, and S. Baron-Cohen. 2006. Affective computing and autism. *Annals of the New York Academy of Sciences* 1093 (December 2006), 228–248.
7. T. Fawcett. 2006. An introduction to ROC analysis. *Pattern recognition letters* 27, 8 (2006), 861–874.
8. P. Fitzpatrick, V. Romero, J. L. Amaral, A. Duncan, H. Barnard, M. J. Richardson, and R. Schmidt. 2017. Evaluating the importance of social motor synchronization and motor skill for understanding autism. *Autism Research* 10, 10 (2017), 1687–1699.
9. K. Fujiwara and I. Daibo. 2016. Evaluating Interpersonal Synchrony: Wavelet Transform Toward an Unstructured Conversation. *Frontiers in psychology* 7 (2016).
10. M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti. 2014. Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th Int'l Conf. on*. IEEE, 39–42.
11. A. Grinsted, J. C. Moore, and S. Jevrejeva. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics* 11, 5/6 (2004), 561–566.
12. K. Hunter. 2014. *Shakespeare's heartbeat: Drama games for children with autism*. Routledge.
13. M. Khoramshahi, S. Ashwini, R. Stéphane, B. Bardy, and B. Aude. 2016. Role of Gaze Cues in Interpersonal Motor Coordination: Towards Higher Affiliation in Human-Robot Interaction. *PLoS ONE* (08 2016).
14. J. Mantyjarvi, J. Himberg, and T. Seppanen. 2001. Recognizing human motion with multiple acceleration sensors. In *Proc. IEEE Int'l Conf. on Sys., Man, and Cybernetics*, Vol. 2. 747–752.
15. D. Maraun and J. Kurths. 2004. Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Processes in Geophysics* 11, 4 (2004), 505–514.
16. L. Miles, L. K. Nind, and C. Neil Macrae. 2009. The Rhythm of Rapport: Interpersonal Synchrony and Social Perception. *Jrnl of Experimental Social Psy.* 45 (05 2009), 585–589.
17. A. Paxton, K. Rodriguez, and R. Dale. 2015. PsyGlass: Capitalizing on Google Glass for naturalistic data collection. *Behavior Research Methods* 47, 3 (01 Sep 2015), 608–619.
18. T. Plötz, N. Y. Hammerla, A. Rozga, A. Reavis, N. Call, and G. D. Abowd. 2012. Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities. In *Proc. of the 2012 ACM Conf. on Ubiquitous Computing (UbiComp '12)*. ACM, 391–400.
19. F. Ramseyer and W. Tschacher. 2011. Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Jrnl of consulting and clinical psychology* 79, 3 (2011), 284.
20. M. J. Richardson, K. L. Marsh, R. W. Isenhower, J. R. Goodman, and R. Schmidt. 2007. Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science* 26, 6 (2007), 867 – 891.
21. M. Sekine, T. Tamura, M. Akay, T. Fujimoto, T. Togawa, and Y. Fukui. 2002. Discrimination of walking patterns using wavelet-based fractal analysis. *IEEE Trans. on neural systems and rehabilitation engineering* 10, 3 (2002), 188–196.
22. C. Torrence and G. P. Compo. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society* 79, 1 (1998), 61–78.
23. P. Valdesolo, J. Ouyang, and D. DeSteno. 2010. The rhythm of joint action: Synchrony promotes cooperative ability. *Jrnl of Exp. Social Psy.* 46 (07 2010), 693–695.
24. J. von Zimmermann, S. Vicary, M. Sperling, G. Orgs, and D. C. Richardson. 2018. The choreography of group affiliation. *Topics in Cog. Sc.* 10, 1 (Jan 2018), 80–94.
25. A. Washburn, M. DeMarco, S. de Vries, K. Ariyabuddhipongs, R. C. Schmidt, M. J. Richardson, and M. A. Riley. 2014. Dancers entrain more effectively than non-dancers to another actor's movements. *Frontiers in Human Neuroscience* 8 (2014), 800.
26. P. Washington, C. Voss, A. Kline, N. Haber, J. Daniels, A. Fazel, T. De, C. Feinstein, T. Winograd, and D. Wall. 2017. SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 112 (Sept. 2017), 22 pages.
27. S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. 2016. Convolutional pose machines. In *CVPR*.
28. T. Westeyn, K. Vadas, X. Bian, T. Starner, and G. D. Abowd. 2005. Recognizing mimicked autistic self-stimulatory behaviors using HMMs. In *Ninth IEEE Int'l Symp. on Wearable Computers (ISWC'05)*. 164–167.