Towards Realtime Measurement of Connectedness in Human Movement

Michael Krzyzaniak[1], Rushil Anirudh[2,1], Vinay Venkataraman[2], Pavan Turaga[1,2] and Sha Xin Wei[1,3]
[1]School of Arts, Media + Engineering, Arizona State University
[2]School of Electrical, Computer, and Energy Engineering, Arizona State University
[3]Synthesis Center, Arizona State University
{mkrzyzan | ranirudh | vvenka18 | pturaga | xinwei.sha}@asu.edu

ABSTRACT
With the proliferation of wearable sensors, we have access to rich information regarding human movement that gives us insights into our daily activities like never before. In a sensor rich environment, it is desirable to build systems that are aware of human interactions by studying contextual information. In this paper, we attempt to quantify one such contextual cue - the connectedness of physical movement. Inspired by the Semblance of Typology Entrainments, we estimate the connectedness of trained dancers as observed from inertial sensors, using a diverse set of techniques such as quaternion correlation, approximate entropy, Fourier temporal pyramids, and discrete cosine transform. Preliminary experiments show that it is possible to robustly estimate connectedness that is invariant to frequency, amplitude, noise or time lag.

Author Keywords
Connectedness; Human Movement; Group Movement; Correlation; Cross Approximate Entropy; Discrete Cosine Transform; Fourier Temporal Pyramids; Automated Society; Automation; Group Intention; HCI; CHI; Wearable Sensing; Time Series Analysis; Social Signal Processing

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces; K.4.3 Computers and Society: Organizational ImpactsAutomation; J.5 Arts and Humanities: Performing Arts; H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
As daily life becomes increasingly automated, it is desirable for automated systems to be sensitive to human social contexts. If some people are chatting near an automatic door, without yet the intention of entering, the door should not continually open and close, nor should the interlocutors have to change their behavior to accommodate the door's socially agnostic behavior. The group's intention could be conveyed explicitly to the system via a button, wireless signal, or speech command. However, if life is to be naturalistic in a highly automated world, the systems should be capable of implicitly detecting group intention. We hypothesize that in many cases, a group's intentions can be inferred through the quantity and type of *connectedness* in their physical movement. We define 'connectedness' to mean, very broadly, any type of similarity in the movement of the individuals as compared to one-another. This notion was inspired by Adrian Freed's "Semblance Typology of Entrainments" [6], and includes but is not limited to temporal synchrony.

A growing body of research in social psychology has been establishing the link between sociality and coordinated movement [7], and this provides a foundation for our hypothesis. In general, three broad approaches have been taken in automatically measuring this interpersonal movement: cross-correlation, recurrence analysis, and spectral methods [4]. Here, we extend these approaches to include monolithic cross-correlation on three-dimensional sensor data; we use cross-approximate entropy which extends recurrence analysis as a numerical rather than graphical method, and additionally checks for repetitive patterns defined by a vectors rather than performing simple point-wise comparison; and we explore a variety of spectral techniques. We evaluate these methods in the hope that this will provide a signpost for future research addressing the larger questions presented here.

SIMPLIFYING ASSUMPTIONS
Connectedness may refer to a broad variety of phenomena, such as connectedness in time, effort, orientation, gesture, location, frequency, and so forth (simultaneous, isotropic, syntonic...) [6]. Computationally modeling these phenomena presents a challenge due to their ambiguous and disparate nature. The problem is further complicated by the fact that movements that are ostensibly "related" to the human observer, may manifest themselves as very different signals when observed through various sensors. Therefore, in the present study, which represents nascent research, we aim to assess the suitability of a variety of potential ways of measuring connectedness which are invariant to attributes which can be captured easily within the sensor data such as temporal shifts, amplitude and frequency changes. Furthermore, we focus on measuring connectedness of only two humans

MOCO'15, August 14–15, 2015, Vancouver, Canada.

ACM ISBN 978-1-4503-3457-0/15/08...$15.00
DOI: **http://dx.doi.org/10.1145/2790994.2791012**

at a time, although in the general case it will be desirable to extend these methods to include arbitrarily large numbers of people. We additionally choose trained dancers as our humans, because their training makes them suitable for use in a controlled movement study. In the future we plan to extend our techniques to untrained movers in quotidian settings. We measure each human's movement with a single inertial sensor (accelerometer, gyroscope, and magnetometer). The increasing prevalence of wearable sensors in fitness trackers, smart watches, and mobile phones provides a strong motivation for this approach.

MEASURING CONNECTEDNESS

In this work we aim to quantify the *connectedness* of human movements using the different techniques described next.

Hypercomplex Signal Correlation

Previous studies have used signal cross-correlation on gyroscope data to measure the similarity of movement of dancers [1] [2]. However, this technique only operates on one of each of the gyroscope's axes. This may be extended to a monolithic representation of all three axes by representing the data as pure quaternions with real part of zero. The hypercomplex result of such correlation encodes both the magnitude of correlation, time-lag, as well as the 'best-fit' rotation which achieves that magnitude. The standard definition of discrete correlation extends to quaternion signals thus[3]:

$$a \circ b = \sum_{t=0}^{N-1} a(\tau + t)\overline{b(t)}$$

Where $\circ$ indicates correlation, and τ is the time-lag between signals a and b. Because we are interested in realtime computation on signals of arbitrary length, we compute the correlation in windows of size N, and evaluate τ for $-N <= \tau <= N$. A direct implementation of this sliding sum has a time complexity of $\mathcal{O}(N^2)$ quaternion multiply-adds, which makes it unwieldy for realtime computation. As with real-valued correlation, the time-complexity can be reduced by making use of the well-known convolution property; the implementation described here is roughly $\mathcal{O}(6Nlog_2(N))$. This necessitates forward and inverse quaternion Fourier transforms (QFTs). Ell and Sanguine give definitions for these [5], which we reduce here from two spatial dimensions (for images) to one temporal dimension (for gyroscope data).

$$\text{QFT}(\omega) = \sum_{n=0}^{N-1} x(n) e^{-\mu\omega\frac{2\pi n}{N}}$$

$$\text{QFT}^{-1}(\omega) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{\mu\omega\frac{2\pi n}{N}}$$

Here, μ is any pure unit quaternion. If we select $\mu = i$ (where i is the imaginary unit) we may implement these each as two Complex Fourier Transforms, one on the real and i components of the quaternions, and one on the j and k components. For these, we may use the common, efficient FFT algorithm. This allows an efficient implementation of hypercomplex correlation using equation 117 in [8]. Because it is difficult to understand in the form given there, we expand it algebraically here to a form that can be directly implemented, again reduced to one dimension.

$$\text{QFT}(g \circ h)(\omega) =$$
$$(ac + bd + uy + vz) + (-ad + bc - uz + vy)i +$$
$$(-ae + bf + uw - vx)j + (-af - be + ux + vw)k;$$
$$a = G(\omega)_r; \quad b = G(\omega)_i; \quad c = H(\omega)_r; \quad d = H(\omega)_i;$$
$$e = H(-\omega)_j; \quad f = H(-\omega)_k; \quad u = G(\omega)_j; \quad v = G(\omega)_k;$$
$$w = H(-\omega)_r; \quad x = H(-\omega)_i; \quad y = H(\omega)_j; \quad z = H(\omega)_k;$$

Where G and H are the QFT of the signals g and h respectively, and the subscripts indicate the real, i , j or k part of the quaternion.

The result of this must be rotated by N samples so that $\tau = 0$ appears in the center. We furthermore convert the quaternion result to Euler angles, to obtain the magnitude and best-fit rotation for each value of τ. The result is a function of τ. We select the peak of this function to represent the overall magnitude and lag of correlation for this window. Because it would be meaningless for time-lag to jump discontinuously over time, in each window we only search for peaks nearby the previous peak.

Approximate Entropy

Approximate entropy is a probabilistic measure based on the log-likelihood of repetitions of patterns of length m being close within a defined tolerance window (or radius r) that will exhibit similar characteristics as patterns of length $(m + 1)$ [10]. It is defined using three parameters: embedding dimension (m), radius (r), and time delay (τ). Here, m represents the length of pattern (also called as embedding vector) in the data which is checked for repeatability, τ is selected so that the components of the embedding vector are sufficiently independent, and r is used for the estimation of local probabilities.

Cross Approximate Entropy (XApEn):

Cross approximate entropy is a bivariate measure and is defined as the amount of asynchrony between two time series data [9]. Let $\mathbf{u} = [u_1, u_2, \ldots, u_N]^T$ and $\mathbf{v} = [v_1, v_2, \ldots, v_N]^T$ denote two time series data of length N. The embedding vectors for given parameters m, τ, and r can be defined as

$$\mathbf{C}_i^m(r)(v||u) = \frac{1}{N-(m-1)\tau} \sum_{<j>} \Theta(r - d(\mathbf{x}_1(i), \mathbf{x}_2(j)))$$

The cross approximate entropy is then given by

$$XApEn(m, r, \tau) = \Phi^m(r)(v||u) - \Phi^{m+1}(r)(v||u)$$

where:

$$\Phi^m(r) = \frac{1}{N-(m-1)\tau} \sum_{i=1}^{N-(m-1)\tau} \ln \mathbf{C}_i^m(r)(v||u)$$

Transform based metrics

In the present framework, we observe data from the inertial sensors as a $3 \times T$ time series, where T is the number of observed samples. It is useful to think of connectedness as the generalization of a typical distance metric such as the ℓ_2

norm, because the connectedness between two completely different movements maybe low, but the *distance* between them will be large. While the ultimate goal is to develop measures that allow one to quantify connectedness over any two arbitrary movements, at present we focus on similar movement styles performed in different settings (such as speed, time lag, noise, intensity etc.), where connectedness and distance metrics can be approximated to be equal.

Fourier Temporal Pyramid

Accurately measuring connectedness involves obtaining measures that are robust to noise, sampling rates, and, more importantly, speed variations. While the former two are common for any kind of time series, the latter is an important problem in human movement analysis. Typically, we would like to have measures that are invariant to the speed of execution of various movements. Within time series analysis, a common trick is to use Dynamic Time Warping (DTW) which attempts to identify the minimum distance between two time series, for all possible speed variations between them. However, DTW is computationally very expensive with a complexity of $\mathcal{O}(mn)$ for two time series of lengths m and n respectively. Alternatively, recent methods have attempted to solve the mis-alignment problem by observing the time series in a different domain, such that the representation is inherently robust to noise and temporal alignment. Fourier Temporal Pyramid (FTP) [13] is one such technique originally introduced in the context of human skeletal action recognition. FTP performs a hierarchical short-time Fourier transform on the time series, and discards the high frequency components which are often associated with noise. Further, the representation takes care of misalignment because temporal translation results does not affect the magnitude of the Fourier coefficients.

Discrete Cosine Transform

Another recent work [11] proposed the use of the Discrete Cosine Transform (DCT) as an alternative way to obtain a metric that are robust to noisy measurements and temporal translations. Similar to the FTP, this metric transforms the time series, $q(t)$ into the frequency domain as, $Q(t) = Aq(j)$, where A is the DCT matrix that contains cosines of varying frequencies. We retain the top k coefficients as $\phi(t) = |Q(t)_{1:k}|$, which results in a k dimensional vector for a time series of arbitrary length. This also allows us to compare time series of different lengths easily, without the need for interpolation or truncation.

EVALUATION

In this study we focus on connectedness in gesture, time, and amplitude. In order to make a preliminary assessment of the performance of the above metrics, we gathered a dataset of humans moving as follows. Two dancers designed a short movement sequence approximately thirty seconds in duration. Inertial sensors were placed on their wrists. They then performed five variations on the basic movement sequence, where the variations were designed to exhibit different amounts and types of connectedness. Each variation was repeated five times in a row, resulting in twenty five instances of the basic thirty-second sequence which shall be referred to as sequence 001 - 025 throughout. The labels and descriptions of the variations are summarized in table 1.

Sequence	Label	Description
001-005	Identical	Both dancers perform the basic movement sequence identically
006-010	Dissimilar	One dancer performs the basic movement sequence while the other does something wholly unrelated (different in each instance)
011-015	Similar	One dancer performs the basic movement sequence while the other performs the basic movement sequence plus noise in the instrumented hands
016-020	Lag	Identical movement but one dancer leads the other by a few seconds
021-025	Amplitude	One dancer performs the basic movement sequence,while the other performs a 'small amplitude' version of the same

Table 1. Connectedness in Movement dataset

The labels for these variations are treated as ground-truth. In the overall research agenda we are interested in a continuous ground-truth, and do not view this as a classification problem. Instead, we treat the labels as continuous ground-truth with a coarse temporal granularity. The complete dataset is available from the authors upon request.

Parameters

In the following studies, we decimate the sensor stream to 100 Hz. This reduces computational load and does not have any significant effect on the results. We use a window-size of 512 sensor samples, for a maximum τ of about ± 5 seconds, and a hop-size of 5 samples for a temporal resolution of 50 ms.

For the quaternion correlation we search for peaks within ± 128 samples of the previous peak, which is chosen heuristically - which may sub-optimal in general, but worked well for our experiments. We normalize the contents of each window prior to correlation, by dividing each quaternion sample by the mean magnitude of that window. We use this over other possible definitions of normalization because it preserves the quaternions' angles [12].

Comparison of Metrics

If connectedness is taken as a one-dimensional phenomenon whose measurement must be robust to a variety of conditions, we may use Spearman coefficients to compare the performance of different measures discussed in this work. Once the connectedness scores are estimated, the empirical mean of the scores is used as a final measure per sequence. Then the Spearman correlation coefficient of the estimated measures is compared against an "ideal" score which is 1 when the movements are correlated and 0 when they are uncorrelated. In this labeling scheme, we mark the variations of 'amplitude', lag, and 'similar' as 1. As table 2 shows, the proposed measures perform significantly better than our baseline - the standard ℓ_2 distance metric.

Measure	Score
Hypercomplex Correlation	0.6933
Approximate Entropy	0.5269
Discrete Cosine Transform	0.5296
Fourier Temporal Pyramids	0.4715
ℓ_2-norm	0.2773

Table 2. The Spearman correlation coefficients for various measures discussed in this work. The estimated scores are compared to an ideal scoring rubric with correlated movement as 1 and uncorrelated movement as 0 (higher is better)

Connectedness in Gesture, Time or Amplitude

Next, we consider connectedness as a multidimensional phenomenon, and examine how quaternion correlation performs in greater detail. We first assess whether it is capable of ranking the degree the dancers are performing the same gestures. We accomplish this by examining the 'Identical', 'Dissimilar' and 'Similar' movement variations, which represent high, moderate, and no connectedness. Below is a plot of the magnitude of correlation as a function of time for each sequence in those categories.

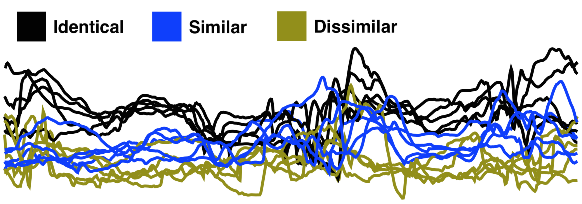

Figure 1. Magnitude of correlation as a function of time for 'Identical', 'Similar', and 'Dissimilar' movement sequences

Although the resolution of our ground-truth is too low to shed light upon the peaks and valleys in each movement sequence, it can be seen that on average the identical sequences have a greater magnitude of correlation than the similar ones, which are in turn greater than the dissimilar ones. Next, we are interested in identifying in whether or not the dancers are together in time. For this we analyze the 'Lag' and 'Identical' movement variations, which are the same except in the case of 'Lag', the dancers are separated in time. Figure 3 shows time-lag as a function of time, as given by hypercomplex correlation for these sequences.

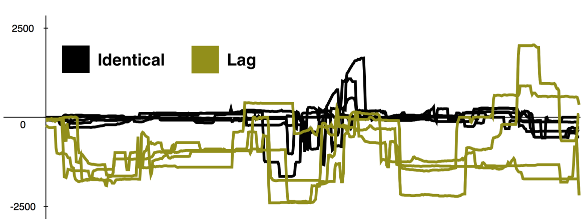

Figure 2. Correlation lag as a function of time for 'Identical' and 'Lag' movement sequences

Again, the ground-truth does not contain information about the individual peaks and valleys in these signals, but one may speculate that the movement sequence contains some self-similarity midway through. Nonetheless, It is clear that on average correlation correctly assesses connectedness in time.

CONCLUSION

This paper has introduced the notion of 'connectedness' and evaluated some methods of quantifying it. We hope that these methods will provide a foundation for the study of detecting group intention.

ACKNOWLEDGEMENTS

Thanks to Adrian Freed for suggesting hypercomplex signal correlation, and to the dancers Julie Akerly and Varsha Iyengar for their generosity and patience. The work of Vinay Venkataraman, Rushil Anirudh, and Pavan Turaga was supported by National Science Foundation grant number CCF-CIF 1320267.

REFERENCES

1. Aylward, R., Lovell, S. D., and Paradiso, J. A. A compact, wireless, wearable sensor network for interactive dance ensembles. In *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*, IEEE (2006), 4–pp.
2. Aylward, R., and Paradiso, J. A. Sensemble: a wireless, compact, multi-user sensor system for interactive dance. In *Proceedings of the 2006 conference on New interfaces for musical expression*, IRCAMCentre Pompidou (2006), 134–139.
3. Bahri, M. Relationships between convolution and correlation for fourier transform and quaternion fourier transform. *Int. Journal of Math. Analysis 7*, 43 (2013), 2101–2109.
4. Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., and Cohen, D. Interpersonal synchrony: A survey of evaluation methods across disciplines. *Affective Computing, IEEE Transactions on 3*, 3 (2012), 349–365.
5. Ell, T. A., and Sangwine, S. J. Decomposition of 2d hypercomplex fourier transforms into pairs of complex fourier transforms. In *Proc. Eusipco*, vol. 2 (2000), 1061–1064.
6. Freed, A. Semblance Typology of Entrainments. **http://adrianfreed.com/content/semblance-typology-entrainments**. Accessed: 2015-03-10.
7. Marsh, K. L., Richardson, M. J., and Schmidt, R. C. Social connection through joint action and interpersonal coordination. *Top Cogn Sci 1*, 2 (4 2009), 320–39.
8. Pei, S.-C. . C., Ding, J.-J. . J., and Chang, J.-H. . H. Efficient implementation of quaternion fourier transform, convolution, and correlation by 2-d complex fft. *Signal Processing, IEEE Transactions on 49*, 11 (2001), 2783–2797.
9. Pincus, S., and Singer, B. H. Randomness and degrees of irregularity. *Proceedings of the National Academy of Sciences 93*, 5 (1996), 2083–2088.
10. Pincus, S. M. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences 88*, 6 (1991), 2297–2301.
11. Pirsiavash, H., Vondrick, C., and Torralba, A. Assessing the quality of actions. In *13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI* (2014), 556–571.
12. SEXTON, B. A., and JONES, J. C. Means of complex numbers. *Texas College Mathematics Journal 1*, 1 (2005), 1–4.
13. Wang, J., Liu, Z., Wu, Y., and Yuan, J. Mining actionlet ensemble for action recognition with depth cameras. In *(CVPR), 2012* (June 2012), 1290–1297.