Pattern Recognition Letters 000 (2014) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Short communication

Automatic measure of imitation during social interaction: A behavioral and hyperscanning-EEG benchmark ☆

Emilie Delaherche [a,1,*], Guillaume Dumas [b,c,d,e,1], Jacqueline Nadel [b,d,e], Mohamed Chetouani [a]

[a] Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie-Paris 6, CNRS UMR 7222, Paris, France
[b] Université Pierre et Marie Curie-Paris 6, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, UMR-S975, Hôpital de La Salpêtrière, Paris, France
[c] INSERM, U975, Paris, France
[d] CNRS, UMR 7225, Paris, France
[e] ICM, Paris, France

ARTICLE INFO

Article history:
Received 27 February 2014
Available online xxx

Keywords:
Imitation
Video indexing
Unsupervised learning
DTW
Hyperscanning

ABSTRACT

Social neuroscience shows a growing interest for the study of social interaction. Investigating its neural underpinnings has been greatly facilitated through the development of hyperscanning, a neuroimaging technique allowing to record simultaneously the brain activity of multiple humans engaged in a social exchange. However, the analysis of spontaneous social interaction requires the indexing of the ongoing behavior. Since spontaneous exchanges are intrinsically unconstrained, only a manual indexing by frame-by-frame analysis has been used so far. Here we present an automatic measure of imitation during spontaneous social interaction. Participants gestures are characterized with Bag of Words and 1-Class SVM models. Then a measure of imitation is derived from the likelihood ratio between these models. We apply this method to hyperscanning EEG recordings of spontaneous imitation of bimanual hand movements. The comparison with manual indexing validates the method at both behavioral and neural levels, demonstrating its ability to discriminate significantly the periods of imitation and non-imitation during social interaction.

1. Introduction

Social interaction is at the core of human behavior. While cognitive science have made tremendous progress in the understanding of cognition from isolated individuals, less is known about people engaged in an interactive context [1,2]. Developmental psychology pioneered in the study of reciprocal interaction, pointing out its key role during early life for the development of our sociocognitive abilities [3,4]. Recently, social psychology and social neuroscience also moved from the study of social perception in isolated individuals to the study of social interaction in pairs or group of humans. In social neuroscience, this has been facilitated by the development of hyperscanning, a neuroimaging technique allowing the simultaneous recording of brain activity in multiple participants [5–7]. Nevertheless, hyperscanning studies of social interaction need protocols where human interaction unfolds in a spontaneous manner thus leading to unconstrained dynamics at the social level [8]. To date, the behavioral analysis of the interaction—especially imitation—has been mostly made by hand, a long process of frame-by-frame video analysis. Here we propose an automatic indexing of imitative behavior during spontaneous interaction. We compare this technique with the traditional frame-by-frame approach and quantify how it impacts subsequent neurodynamical analyses at both intra- and inter-individual levels with hyperscanning-EEG.

1.1. Imitative behavior in spontaneous interaction

During social interaction, people spontaneously imitate their social interaction partners, including mimicry of his gestures [9], his facial expressions [10,11], his mannerism [9], and his posture [12,13]. Mimicry facilitates affiliation [14] and good understanding between individuals [12,15]. Many terms are associated with mimicry in the literature: *behavior matching* [16], *mirroring*, *congruence* and *the chameleon effect* [9]. This nonconscious form of imitation is notably different than conscious imitation which is commonly considered as a foundation for learning, socialization and communication [17,18]. In spontaneous exchanges, it becomes for instance a mean of communication [19]. While mimicry is still present, the behavior becomes more complex, giving rise to alternation between roles of imitator and driver. In neuroscience, this lack of control forced the study of imitation to be limited at the intra-individual level and induced

☆ This paper has been recommended for acceptance by G. Sanniti di Baja.
* Corresponding author: Tel.: +336 22348661.
E-mail address: emilie.delaherche@gmail.com (E. Delaherche).
[1] These authors contributed equally to this paper.

http://dx.doi.org/10.1016/j.patrec.2014.09.002
0167-8655/

context [20–22]. Hyperscanning studies however allowed to investigate spontaneous imitations, thus helping to identify the neurodynamic signatures of spontaneous human interactions at both intra- and inter-individual levels [23–25]. However, investigating spontaneous imitation requires a fine grained analysis of the ongoing interaction, only accessible by video manual indexing.

1.2. Automatic analysis of imitative behavior

To overcome the tedious task of frame-by-frame analysis, automatic methods have been proposed to assess imitative or coordinate behavior. Several studies propose to assess movement coordination in spontaneous interaction (meetings [26–28], music bands [29,30], psychotherapy sessions [31–34] etc.).

Among these studies, some focuses on head motion, assessed by motion capture [35–37] or image-based tracking methods [26–28,30,29,38]. Other studies assess movement of the participants globally, with image processing techniques like motion energy [39–41,33,34,42,31,43]. The main pitfall of these algorithms is that they capture the movement globally. They can assess if two participants have the same activation but cannot discriminate between two motions with the same dynamic.

Sun and colleagues proposed to combine a variant of motion energy image with quadtree decomposition to localize motion regions [44,45,43]. Then, kinematic features between the regions that contain motion are compared. Yet, none of these methods characterizes finely the shape of gestures or combine both shape and dynamic description of gestures to assess imitation.

The method we propose in this paper leverages refined gesture description, as proposed for action recognition, to improve automatic analysis of imitative behavior.

2. Material and methods

2.1. Material

2.1.1. Participants

Five adults participated in the study. All subjects had normal or corrected-to-normal vision. They were all right-handed. All were volunteers to participate to the study and had given their written informed consent. The project of this study was reviewed and approved by the local Ethical Committee for Biomedical Research (agreement No. 104-10). None of the participant reported a history of psychiatric or neurological disease.

2.1.2. Protocol

The experimental protocol was divided into three blocks separated by a 10 min rest. Each block comprised three runs of 2 min. A run was composed of three conditions: observation of a prerecorded library of 20 meaningless hand gestures, a spontaneous imitation and the imitation of a video. In this paper, we focus on the spontaneous imitation where the subjects were told that they could produce hand gestures of their own and imitate the other's hand gestures whenever they would like it. Imitation is thus produced at will and the social roles (i.e. imitator or driver) are not fixed by the experimenter but spontaneously emerge from the interaction between the two subjects. Each run started by a 30 s period of rest, and before each imitation conditions, the subjects were asked to produce a 30 s of meaningless hand gestures. At the end of the experiment, a short block of calibration comprised periods of blinks, jaws contraction, and head movements of 30 s each.

2.1.3. Dual-video acquisition

The experiment was conducted in three connected laboratory rooms, one for each participant and the third one for the computerized monitoring of the experiment (see Fig. 1). The participants were comfortably seated, their forearms resting on a small table in order to prevent arms and neck movements. They faced a 21-in. TV screen. Two synchronized digital video cameras filmed the hand gestures. An LED light controlled manually, via a switch, by an experimenter located in the recording room, signaled the session start. The output of the video records was transmitted to two TV monitors installed in the recording room allowing the experimenter to control that participants followed the requested instructions.

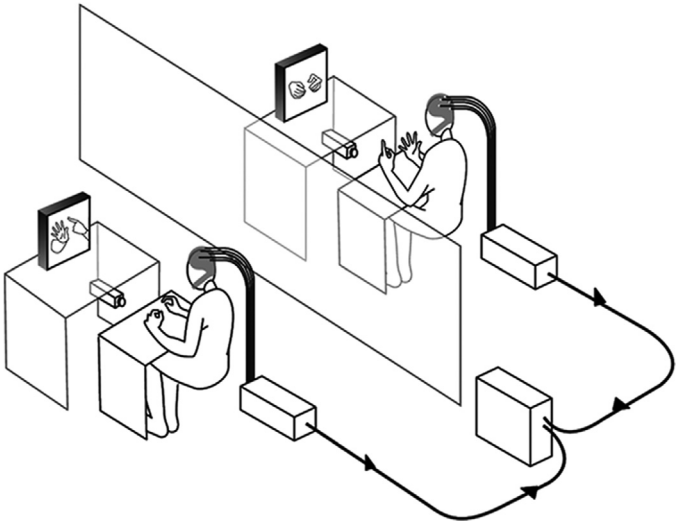


Fig. 1. Experimental setup.

2.1.4. Hyperscanning-EEG acquisition

The neural activities of the two participants were simultaneously recorded with a dual-EEG recording system. It was composed of two Acticap helmets with 64 active electrodes arranged according to the international 10/20 system. The helmets were aligned to nasion, inion and left and right pre-auricular points. A three-dimensional Polhemus digitizer was used to record the position of all electrodes and fiducial landmarks (nasion and pre-auricular points). The ground electrode was placed on the right shoulder of the subjects and the reference was fixed on the nasion. The impedances were maintained below 10 kΩ. Data acquisition was performed using two 64-channels Brainamp MR amplifiers from the Brain Products Company (Germany). Signals were analog filtered between 0.16 Hz and 250 Hz, amplified and digitalized at 500 Hz with a 16-bit vertical resolution in the range of ±3.2 mV.

2.2. Behavioral data analysis

The video records of hand movements during the free episodes of imitation of each other's hand movements were digitized. Then, the LED signals recorded on the two videos at the beginning of each session was used to synchronize the frames of the two partners. They were coded using a revised version of the ELAN program [46,47] that offers a simultaneous presentation of two frames from different sources on the ELAN window. This software allows an analysis of the behavioral frames on separate channels of the window and a recording of time (latency, duration) and occurrence of behavioral events.

Imitation was assessed when the hand movements of the two partners showed a similar morphology (describing a circle, waving, swinging etc.) and a similar direction (up, down, right, left etc.). We labeled respectively Im and NIm the periods with imitation and without imitation.

The reliability of our fine grained analysis was assessed using Cohen's kappa. Inter-observer agreement between two independent coders was performed on 25% of the recordings following previous

Please cite this article as: E. Delaherche et al., Automatic measure of imitation during social interaction: A behavioral and hyperscanning-EEG benchmark, Pattern Recognition Letters (2014), http://dx.doi.org/10.1016/j.patrec.2014.09.002

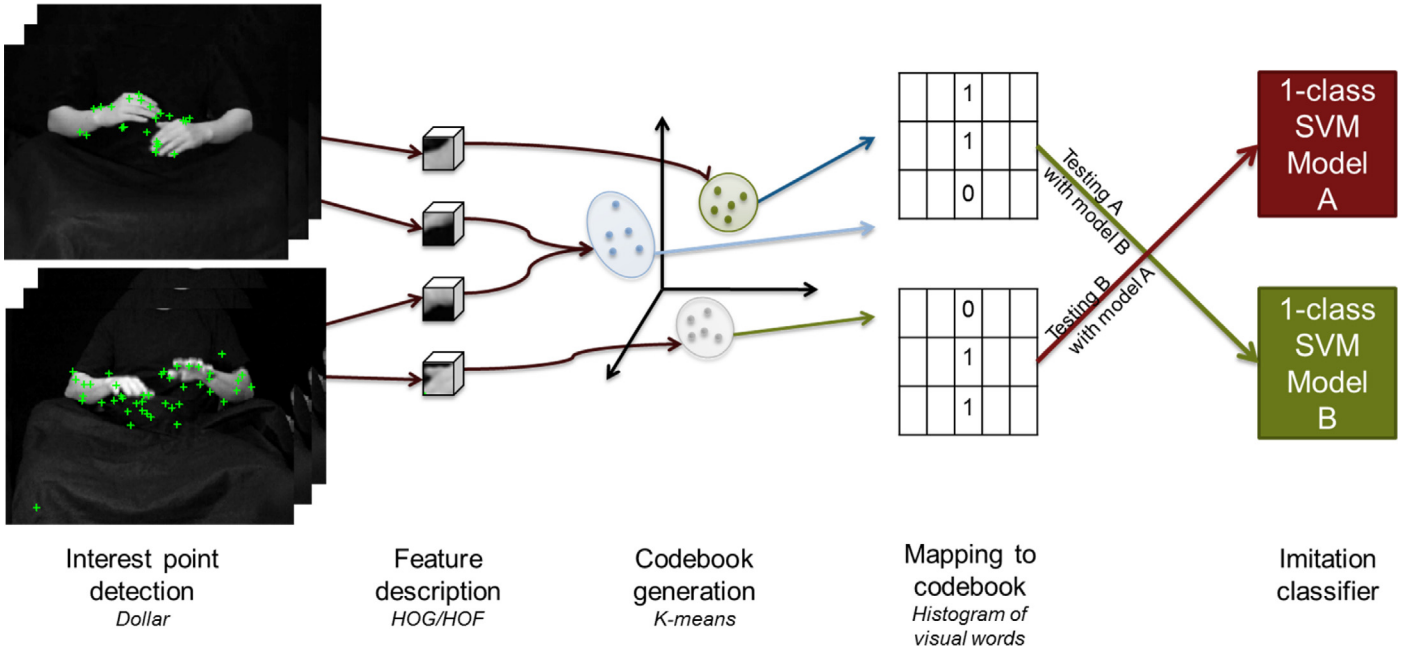


Fig. 2. Automatic indexing.

studies with the identical task [23,25]. The values of kappa coefficient was 0.83.

2.3. Automatic indexing

We proposed a binary classifier to automatically index the participants gestures as identical ("imitation") or different ("non-imitation"). The gestures are represented with histograms of visual words. Then a metric between gestures, based on 1-Class SVM is proposed and a threshold on the metric is learnt with a Leave One Out approach to differentiate imitation from non-imitation (see Fig. 2).

2.3.1. Gesture visual description

Bag of Words models have been successfully applied in computer vision to describe objects, gestures or actions [48–50]. The method is based on a dictionary modeling where each image contains some of the words of the dictionary. In computer vision, the words are features extracted from the image. Bag of Words models rely on four steps: interest point detection, interest point description, codebook generation, mapping to codebook.

Interest point detection. In this step, spatio-temporal interest points are extracted on the image. Several detectors exists in the literature like STIP [51] which is derived from the Harris Detector or Dollãr [52] etc. Dollãr detector was preferred to other detectors for its robustness and for the number of interest points detected was superior, leading to a better characterization of the gesture performed. Dollãr selects local maxima over space and time of a response function based on a spatial Gaussian convolved with a quadrature pair of 1D Gabor filters along the time axis ($\sigma = 1$ et $\tau = 4$).

Feature description. This step consists in describing the variations of image values in the spatio-temporal neighborhood of each interest point. This is the description of the interest points that will differentiate events from each other. Several descriptors exist in the literature: SIFT, SURF, Histogram Of Oriented Gradient (HOG) and Histogram Of Oriented Flow (HOF) etc. In this study, we used a combination of HOG and HOF, which characterizes both the shape and dynamics of gestures. The performance of these descriptors was demonstrated for gesture and action recognition tasks [49]. The size of the neighborhood is $19 \times 19 \times 11$ (nine pixels each side of the point in the spacial domain and five pixels in the temporal domain). These neighborhoods are then divided in $3 \times 3 \times 2$ cells. Each cell is described with a histogram (four orientations for HOG and four orientations for HOF plus an extra bin to code the absence of motion). The size of the feature vectors is 162 (72 bins for HOG and 90 bins for HOF).

Codebook generation. This step consists in generating a codebook from the feature vectors. A K-means clustering is applied to all the feature vectors. The number of clusters define the number of words in the codebook and the codewords are represented by the centers of the clusters.

Mapping to codebook. This last step consists in mapping the feature vector for each frame to the codewords of the codebook. Each frame of the video is represented by an histogram of the codewords for the frame. For each dyad, a new visual word codebook is learnt.

2.3.2. Metric between gestures

We proposed to derive an algorithm for novelty detection based on 1-Class SVM to estimate the similarity between two gestures A and B [53,54]. First, each gesture is modeled with a 1-Class SVM to estimate their probability density functions P_A and P_B. A distance is derived from the likelihood ratio between the following hypothesis:

$$\begin{cases} H_0 : P_A = P_B \text{ (the gestures are identical)} \\ H_1 : P_A \neq P_B \text{ (the gestures are different)} \end{cases}$$

Distribution estimation (1-Class SVM). 1-Class SVM was proposed to estimate the density of a unknown probability density function [55]. For $i = 1, 2, \ldots, n$, the training vectors h_i are assumed to be distributed according to a unknown probability density function $P(.)$. The aim of 1-Class SVM is to learn from the training set a function f such that most of the data in the training set belong to the set:

$$R_h = \{h \in X \setminus f(h) \geq 0\}$$

and the region R_h is minimal. The function f is estimated such that a vector drawn from $P(.)$ is likely to fall in R_h and a vector that does not fall in R_h is not likely to be drawn from $P(.)$. The decision function is:

$$f(h) = \sum_{i=1}^{n} \alpha_i k(h, h_i) - \rho$$

As in our case, h represents an histogram of codewords, we chose the histogram intersection kernel. The kernel $k(.,.)$ is defined over $X \times X$ by : $k(h_i, h_j) = \sum_{i=1}^{d} \min(h_i, h_j)$, where d denotes the size of the histogram.

Distance. Let $h_{A_i}, i = 1, \ldots, n$ and $h_{B_i}, i = 1, \ldots, n$ be the sequence of codewords histograms for a pair of gestures, n denotes the size of the window. Two gestures are similar if the likelihood ratio between H_0 and H_1 is inferior to a given threshold. The likelihood ratio can be interpreted as the similarity $s_{A_iB_j}$ between h_{A_i} and h_{B_i} (see [56] for further details):

$$s_{A_iB_j} = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \alpha_i^A k(h_{B_j}, h_{A_i}) \right) + \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \alpha_i^B k(h_{A_j}, h_{B_i}) \right)$$

where α_i^A (resp. α_i^B) is determined by solving the 1-Class SVM on h_{A_i} (resp. h_{B_i}). This distance can be interpreted as testing a model learned on h_{A_i} with the data from h_{B_i}. For robustness [54], we adopt the following distance in which the histograms of h_{A_i} and h_{B_i} are alternatively used for learning and for testing.

Imitation/non-imitation classifiers. Imitation is synchronous when the partners produce the same gesture at the same time. Imitation is deferred when the same gestures are produced with a slight delay between the partners.

To assess these two forms of imitation, the proposed metric is computed: (a) between simultaneous gestures, (b) between slightly delayed gestures. Thus, we obtain a recurrence matrix $R_{i,j}$ where point (i, j) corresponds to the similarity between the gesture produced at time i by participant A and the gesture produced at time j by participant B. Recurrence matrices represent the points in time when the dyadic partners are in similar states.

The main diagonal of this recurrence matrix corresponds to in-phase gestures. The similarity between slightly delayed gestures is represented in a neighborhood around this main diagonal. Points located below the main diagonal inform on time when partner A is leading and B is following. Points located above are informative of an opposite leading-following relationship.

The recurrence matrix is then quantified:

$$R_{i,j}^{\text{quantif}} = \Theta(\epsilon - s_{A_iB_j})$$

where Θ is the Heaviside function, ϵ a threshold on the similarity measure and $s_{A_iB_j}$ the similarity measure. The recurrence points ($R_{i,j}^{\text{quantif}} = 1$) represent points in time when the partners gestures are similar.

Based on this recurrence matrix, we proposed three classifiers to identify imitation phases:

- *Classifier 1:* Similarity from the main diagonal of the recurrence matrix. This first method only quantifies synchronous imitation. If $s_{1i} = R_{i,i}^{\text{quantif}} = 1$, the decision is "imitation" otherwise "non-imitation". The quantification threshold ϵ varies the decision "imitation/non-imitation".
- *Classifier 2:* Similarity from the number of recurrence points on the main diagonal. This second method also assess synchronous imitation. To ensure the metric is robust to slight variations, the decision is taken on a group of point located in the temporal neighborhood of the point under decision. It comes up to applying a mean filter on the main diagonal of the recurrence matrix.

$$s_{2i} = \sum_{l=i-k_1/2}^{i+k_1/2} R_{l,l}^{\text{quantif}}$$

For this method, the decision relies on the selection of two parameters: the quantification threshold ϵ and the size of the mean filter k_1.

- *Classifier 3:* Similarity from the number of recurrence points in the neighborhood of the main diagonal. To assess deferred imitation, we can consider a neighborhood of the main diagonal of the recurrence matrix. This way, we do not need to make assumption on who is the leader and who is the follower.

$$s_{3i} = \sum_{l=i-k_2/2}^{i+k_2/2} \sum_{m=i-k_2/2}^{i+k_2/2} R_{l,m}^{\text{quantif}}$$

For this method, the decision relies on the selection of two parameters: the quantification threshold ϵ and the size of the neighborhood: k_2.

The Fig. 3 represents for the same interaction the similarity measure from the three classifiers described above.

2.4. EEG data analysis

All EEG analyses were conducted with [57] software and utilized the built-in statistics and signal processing toolbox.

2.4.1. Artifacts correction

Blink, muscles and head movements artifacts were filtered by optimal projection (FOP) methodology [58].

EEG signals were then controlled visually another time. The few remaining artifacts (<0.1% of the data) were excluded from the analysis and we smoothed the joints by a convolution with a half-Hanning window of 400 ms in order to avoid border artifacts induced by the suppression.

2.4.2. Neurodynamical analyses

Following filtering corrections, EEG data were re-referenced to a common average reference (CAR) and transformed by discrete Hilbert methods for specific narrow frequency bands: theta (4–7 Hz), alpha-mu (8–12 Hz), beta (13–30 Hz) and gamma (31–48 Hz). Phases and amplitudes extracted using the Hilbert transform on all band passed signals met the reliability criteria defined in past studies [59].

The local activity was measured by the power. We averaged the square of the amplitude over windows of 400 ms. The connectivity at both intra- and inter-brain levels was analyzed using phase locking value (PLV) [60]. For each pair (i, k) of electrodes, this was done for each frequency band according to the relation:

$$\text{PLV}_{i,k} = \frac{1}{N} \left| \sum_{t=1}^{N} \exp^{j(\phi_i(t) - \phi_k(t))} \right|$$

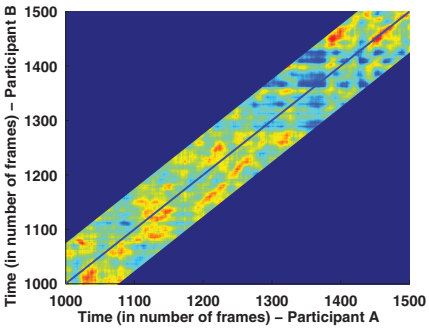
where N is the number of samples considered in each 800 ms window, ϕ is the phase and || the complex modulus. Thus, PLV measure equates 1 if the two signals are perfectly phase locked across the whole time window observed, and equates 0 if they are totally unsynchronized. Thus, PLV is equal to one minus the circular variance of phases' differences. Note that for the inter-brain PLV, so-called hyper-PLV or hPLV, electrode i and k are respectively for the helmets 1 and 2.

For all the EEG measures (i.e. power, PLV and hPLV), we calculated the two-tailed t-statistics across all dyads between the period of imitation and non-imitation. To test the validity of the manual and automatic indexing of behavior, we also calculated these statistics on surrogate data following [23]. This method reconstructs behavioral indexings by shuffling the manually indexed phases of imitation and non-imitation. By picking alternatively, and at random, the durations from the real imitative and non-imitative periods, the timing between EEG data and behavioral data are broken but the distributions of imitative and non-imitative duration stay identical.

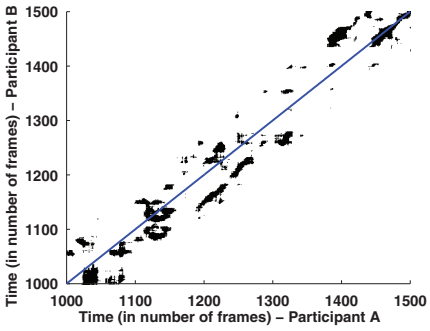
3. Results

3.1. Behavioral data: manual vs. automatic indexing

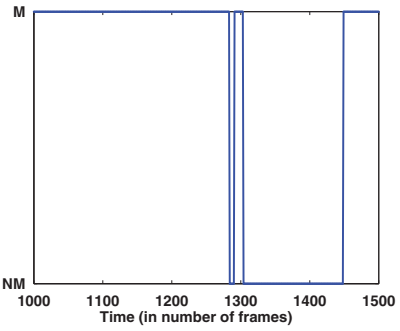
We evaluated the performance of the metric based on 1-Class SVM to detect phases of imitation in the BBC database. The system performance was evaluated in different configurations for the



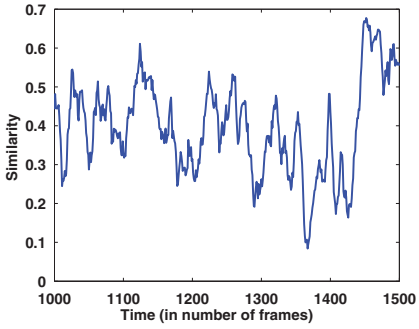
(a) Recurrence matrix



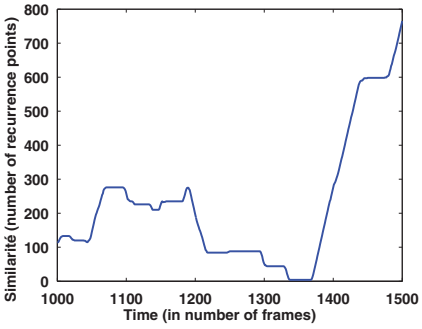
(b) Quantified recurrence matrix



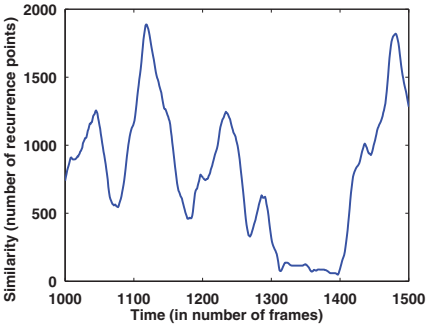
(c) Manual Indexing



(d) Similarity from the main diagonal of the recurrence matrix.



(e) Similarity from the number of recurrence points on the main diagonal.



(f) Similarity from the number of recurrence points in the neighbourhood of the main diagonal.

Fig. 3. Imitation/non-imitation classifiers.

gesture description (see Table 1) and the three imitation classifiers ($k_1 = 1\,s–4\,s$, $k_2 = 1\,s–7\,s$). Performance is obtained with "Leave One Out" cross-validation. The parameters are adjusted to optimize the system performance on $N - 1$ video sequences and the performances are evaluated on the video left out (with N, the number of video sequences in the database). For each configuration and each method, we trained binary classifier for discriminating conditions imitation/non-imitation. Three window sizes were used for training 1-Class SVM models $T = 0.6$, 0.8 and 1 s.

We used $F1$ measure to assess the performance of the classifier. The results are presented in Table 2. The best performance is obtained with method 3 (similarity based on the number of points of recurrence in a neighborhood of the main diagonal) for a dictionary of 256 visual words, the use of combined HOG and HOF descriptors and windows of 0.8 s ($F1 = 0.7848$). It is nevertheless noted that the combination of both types of descriptors only slightly improves the performance compared to configurations where HOG or HOF are used alone. In general, the HOG descriptors, which describe the shape of gestures, give better performance than the descriptors based on the optical flow (HOF), which rather characterize the dynamic gestures. A baseline measure of mimicry was assessed with motion energy image as gesture descriptors and correlation as the measure of similarity. Our classifiers based on HOG + HOF and SVM outperform this baseline measure. Moreover, the Classifier 3 consistently gives better performance than the Classifier 2, itself performing better than Classifier 1.

For the best classifier (Classifier 3, $T = 0.8\,s$), we compared the percentage of time windows correctly classified according to the

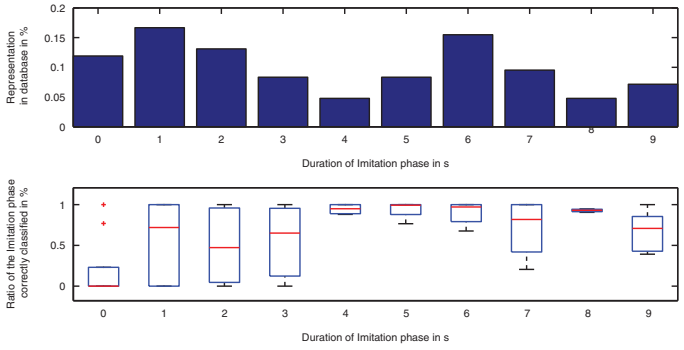
Table 1
Configurations for gesture description.

Codebook size	HOG	HOF	HOG + HOF
$k = 64$	x	x	x
$k = 128$	x	x	x
$k = 256$			x

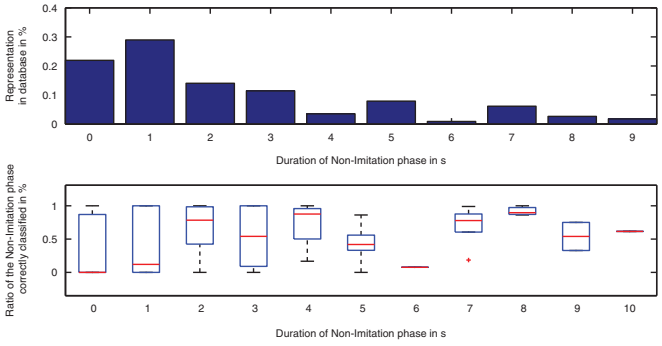
Table 2
Classification results imitation/non-imitation: $F1$ score.

Window size	0.6	0.8	1
Baseline motion energy	0.5348	0.5562	0.5946
Classifier 1			
HOG – $k = 64$	0.6825	0.6969	0.7083
HOG – $k = 128$	0.6998	0.7163	0.7281
HOF – $k = 64$	0.6400	0.6594	0.6872
HOF – $k = 128$	0.6387	0.6598	0.6901
HOG + HOF – $k = 64$	0.6581	0.6763	0.7064
HOG + HOF – $k = 128$	0.6711	0.6958	0.7223
HOG + HOF – $k = 256$	0.6946	0.7221	0.7474
Classifier 2			
HOG – $k = 64$	0.7277	0.7383	0.7367
HOG – $k = 128$	0.7491	0.7505	0.7469
HOF – $k = 64$	0.7181	0.7120	0.7325
HOF – $k = 128$	0.7180	0.7153	0.7234
HOG + HOF – $k = 64$	0.7185	0.7144	0.7277
HOG + HOF – $k = 128$	0.7419	0.7397	0.7471
HOG + HOF – $k = 256$	0.7368	0.7451	0.7672
Classifier 3			
HOG – $k = 64$	0.7320	0.7349	0.7355
HOG – $k = 128$	0.7653	0.7453	0.7459
HOF – $k = 64$	0.7384	0.7427	0.7475
HOF – $k = 128$	0.7392	0.7496	0.7443
HOG + HOF – $k = 64$	0.7476	0.7300	0.7206
HOG + HOF – $k = 128$	0.7668	0.7602	0.7360
HOG + HOF – $k = 256$	0.7786	0.7848	0.7766

Please cite this article as: E. Delaherche et al., Automatic measure of imitation during social interaction: A behavioral and hyperscanning-EEG benchmark, Pattern Recognition Letters (2014), http://dx.doi.org/10.1016/j.patrec.2014.09.002



(a) Imitation phase



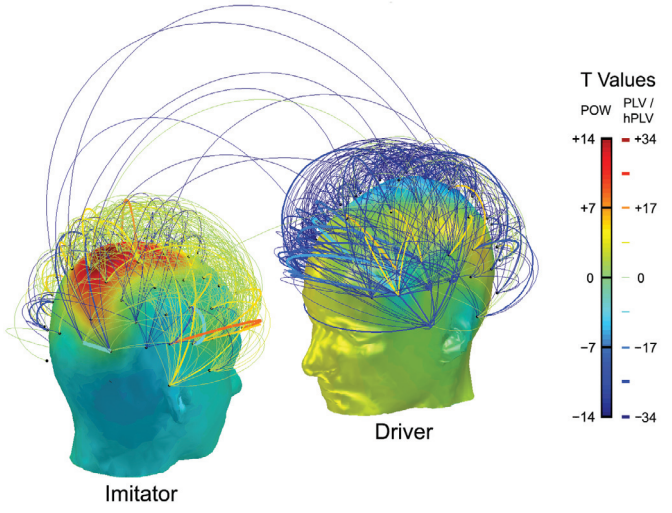
(b) Non-Imitation phase

Fig. 4. Duration of imitation and non-imitation phases and corresponding percentage of correctly classified time windows.

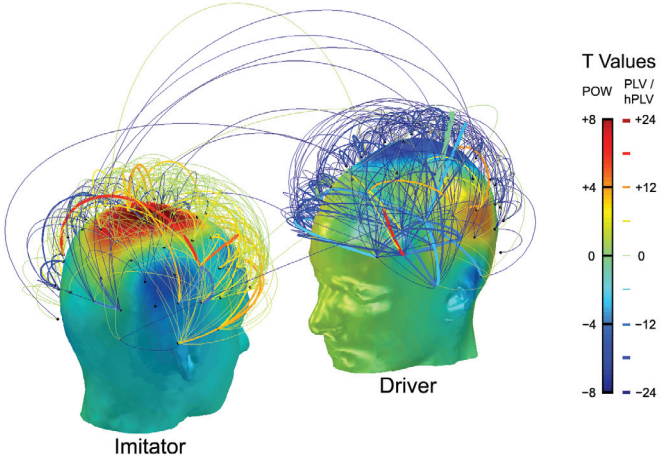
duration of the imitation or non-imitation phase. We can see in Fig. 4 that the classifier performance is higher when the duration of the phase is above 1 s for imitation phases and 3 s for non-imitation phases. These performances are also more stable when the duration of the phase is higher.

3.2. Neurophysiological data: impact of the indexing on common EEG measures

Automatic and manual indexing are compared for both local and distant EEG measures: power, phase-locking value (PLV), and hyper-phase-locking value (hPLV). Fig. 5 illustrates the statistical differences between periods of imitation and non-imitation for the theta frequency band. Colors indicate the t values. The power differences are mapped on the colors of the heads, and synchronization (PLV, hPLV) differences are indicated by links between related electrodes. Notice how the contrasts are similar for both automatic and manual segmentation: in both cases the imitator tends to have an increase of theta activity at both power and PLV levels while the opposite effect occurs for the driver. Fig. 6 summarizes the global results across all frequency bands. For each measure, positive and negative variations are indicated in the case of manual, automatic and scrambled indexing (i.e. null hypothesis). Statistically significant differences with scrambled indexing are indicated with an asterisk (*, $p < 0.05$). We can notice how the automatic indexing gives intermediate results between the scrambled and manual indexing. This demonstrates a less precise detection of neurophysiological relevant periods of imitation and non-imitation than with frame-by-frame analysis. Overall, the automatic method appear to better detect PLV differences in general, and power differences in the high frequency bands. Despite a tendency, there is no significant results for hPLV contrasts using the automatic indexing.



(a) Manual indexing



(b) Automatic indexing

Fig. 5. Illustration of the different EEG measures with the contrast in the theta frequency band (3–8Hz) of imitation and non-imitation periods.

4. Discussion

We showed how unsupervised indexing of imitation can be applied to spontaneous social interaction. The method is compared to the traditional manual and frame-by-frame indexing. Results show some differences at the behavioral level and measure how they impact subsequent hyperscanning-EEG analyses.

4.1. Interpretations

At the technical level, we showed that our unsupervised indexing of imitation outperforms traditional methods based on motion energy image and correlation. Taking into account a series of windows to make the decision (Classifier 2), rather than a single interaction window (Classifier 1) improves the performance of the automatic indexing. Classifier 3 in turn integrates a neighborhood around the diagonal for taking the decision, thus compensating small delays between participants. If partners are slightly offset in time and they perform periodic movements in opposite directions, the similarity measure will be low on the diagonal of the matrix of recurrence. In particular, if the partners are in opposite phase (one raises his arm while the

Please cite this article as: E. Delaherche et al., Automatic measure of imitation during social interaction: A behavioral and hyperscanning-EEG benchmark, Pattern Recognition Letters (2014), http://dx.doi.org/10.1016/j.patrec.2014.09.002

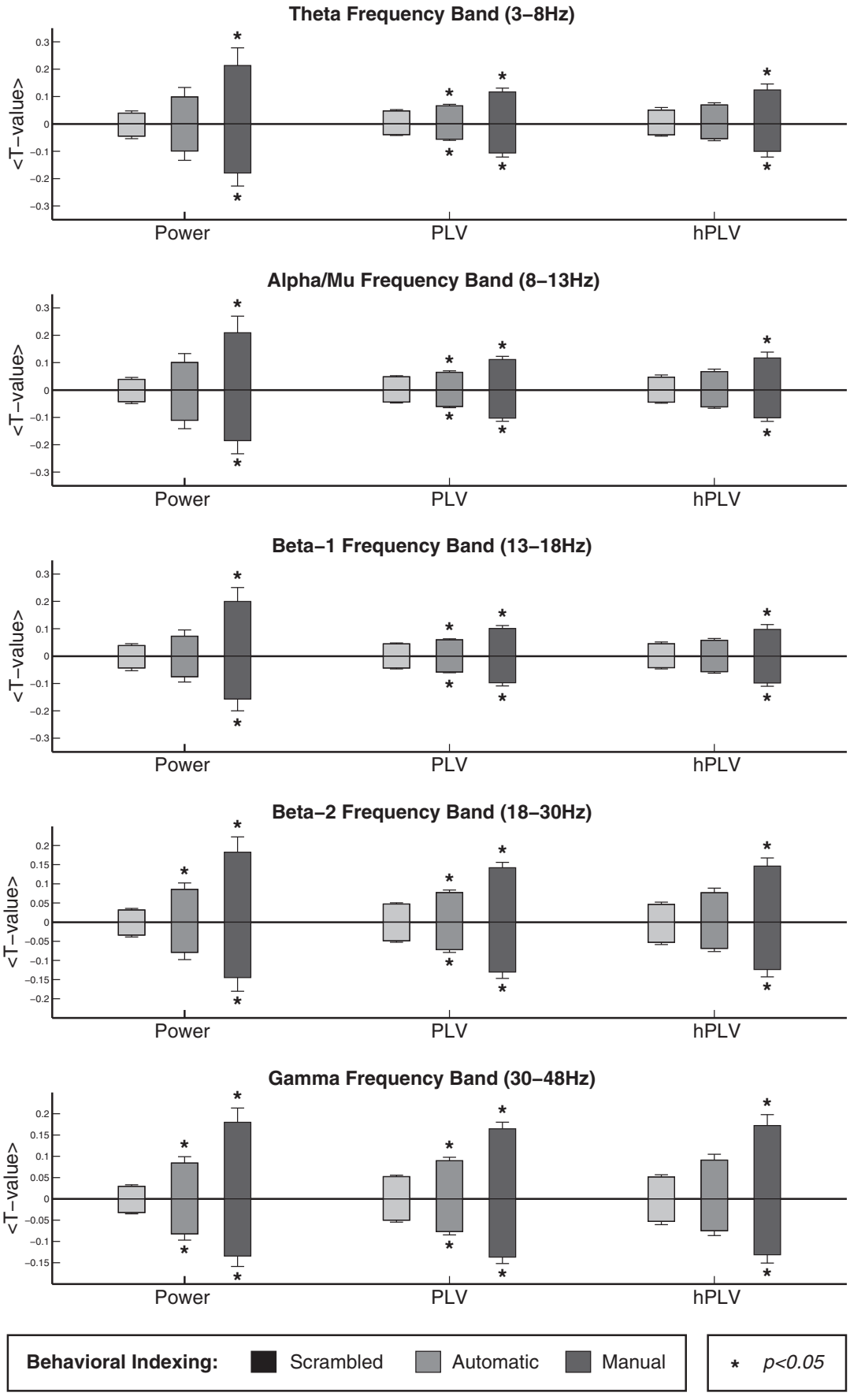


Fig. 6. Effects of the different indexing methods on the common EEG measures.

Please cite this article as: E. Delaherche et al., Automatic measure of imitation during social interaction: A behavioral and hyperscanning-EEG benchmark, Pattern Recognition Letters (2014), http://dx.doi.org/10.1016/j.patrec.2014.09.002

other down, for example). As against, if viewed in the neighborhood of the diagonal (for an offset equal to the period of the beat frequency), a high similarity measure is found. In turn, the automatic indexing is more accurate on long phases of imitation or non-imitation. This is a side effect of considering several seconds to take a decision.

At the neurophysiological level, the EEG analyses validated the ability of the automatic indexing to uncover biologically relevant periods of imitation and non-imitation. PLV results were the most statistically robust, with all frequency bands displaying a similar differences than those uncovered by manual indexing. The automatic indexing appearing better at detecting longer imitative periods, the absence of statistical differences for the hPLV measure may reflect the importance of short desynchronized and non-imitative phases in the interaction. In any case, the statistical differences at the neurophysiological level demonstrate that the automatic algorithm can capture parameters of biological relevance.

4.2. Perspectives

First, at the technical level, some simple enhancement could improve the characterization and comparison of gestures. For instance, HOG and HOF descriptors do not capture the spacial location of the interest points. Motion capture [61] would facilitate the matching between participants gestures. However, it is delicate to set up in domestic environments or with pathological population. Capture sensors like kinect would constitute a good trade-off, providing skeletal data with minimal invasiveness. Then, the current method provides more than a binary decision, but a continuous distance between gestures. Understanding whether human and automatic assessment of imitation evolves according to the same continuum may help to improve the method. Finally, automatic indexing of imitation could facilitate the understanding of pathologies, the comparison of pathological groups [56] and to tie together physiological, neurophysiological and behavioral levels [62].

4.3. Limitations

The current study is limited by the following factors. First, the method concerns only the detection of imitation—and no imitation—while interactional synchrony has been shown as a key part of inter-individual dynamics at both behavioral and neural scales. Interactional synchrony can be assessed with a low level characterization of gestures (correlation between the motion dynamics [34,63]) or based on the synchronicity of high level nonverbal cues, e.g. head nods [61]. A better integration of such key parameter of social interaction may further improve the classification performance. Second, the current method does not provide the directionality of the imitation and thus the detection of the social roles (i.e. leader/follower). This may be an interesting development for the future since turn-taking constitutes an important aspect of spontaneous interaction. Third, unsupervised characterization of gestures has the advantage of not requiring labeled gestures for training. While adequate in this setting (meaningless hand gestures), uncovering which nonverbal cues are most frequently imitated in natural interaction could be of interest. Finally, our analyses concerned only five dyads. This is enough to make the proof of concept and assess the reliability of the method at both behavioral and neural scales, but not to uncover precise anatomic effects.

5. Conclusion

In summary, we have presented a new automatic indexing of imitation during spontaneous social interaction in dyads. Thanks to hyperscanning-EEG recordings, we have also compared how this automatic indexing affect common EEG measures in comparison with the traditional frame-by-frame manual indexing. These experimental results show that our method can significantly discriminate periods of imitation and non-imitation at both behavioral and neural levels. Future works need to investigate how to integrate other behavioral parameters such as interactional synchrony for further improving the classification performance and also for better interpretation of the neurophysiological observations.

Acknowledgments

We thank Robert Soussignan and Emeline Mercier for their help in the manual indexing, Laurent Hugueville for his assistance in the setting of the hyperscanning system, and Florence Bouchet for her generous help in the EEG preparation. Guillaume Dumas was supported by a postdoctoral grant of the Orange Foundation for Autism Spectrum Disorders. This work was performed within the ANR SYNED-PSY (ANR-12-SAMA-06) of the program Santé Mentale et Addictions. This work was performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02.

References

[1] R. Hari, M.V. Kujala, Brain basis of human social interaction: from concepts to brain imaging, Physiol. Rev. 89 (2009) 453–479.
[2] L. Schilbach, B. Timmermans, V. Reddy, A. Costall, G. Bente, T. Schlicht, K. Vogeley, Toward a second-person neuroscience, Behav. Brain Sci. 36(4) (2012) 393–414.
[3] J. Nadel, L. Camaioni, New Perspectives in Early Communicative Development, Routledge, London, 1993.
[4] A. Fogel, Two principles of communication: co-regulation and framing, in: J. Nadel, L. Camaioni (Eds.), New Perspectives in Early Communicative Development, Routledge, London, 1993, pp. 9–22.
[5] G. Dumas, F. Lachat, J. Martinerie, J. Nadel, N. George, From social behaviour to brain synchronization: review and perspectives in hyperscanning, IRBM 32 (2011) 48–53.
[6] I. Konvalinka, A. Roepstorff, The two-brain approach: how can mutually interacting brains teach us something about social interaction? Front. Human Neurosci. 6 (2012) 215 doi: 10.3389/fnhum.2012.00215 See: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402900/.
[7] F. Babiloni, L. Astolfi, Social neuroscience and hyperscanning techniques: past, present and future, Neurosci. Biobehav. Rev. 44 (2012) 76–93.
[8] G. Dumas, Towards a two-body neuroscience, Commun. Integr. Biol. 4(3) (2011) 349–352.
[9] T.L. Chartrand, J.A. Bargh, The chameleon effect: the perception-behavior link and social interaction, J. Personal. Soc. Psychol. 76 (1999) 893–910.
[10] A.N. Meltzoff, M.K. Moore, Imitation of facial and manual gestures by human neonates, Science 198 (1977) 75–78.
[11] N.T. Termine, C.E. Izard, Infants' responses to their mothers' expressions of joy and sadness, Develop. Psychol. 24(2) (1988) 223–229.
[12] F.J. Bernieri, J. Reznick, R. Rosenthal, Synchrony, pseudo synchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions, J. Personal. Soc. Psychol. 54 (1988) 243–253.
[13] W.S. Condon, W.D. Ogston, A segmentation of behavior, J. Psychiat. Res. 5(3) (1967) 221–235.
[14] M.J. Hove, J.L. Risen, It's all in the timing: interpersonal synchrony increases affiliation, Soc. Cognition 27 (2009) 949–960.
[15] M. Lafrance, Posture Mirroring and Rapport, Human Sciences Press, New York, NY, 1982, pp. 279–298.
[16] F.J. Bernieri, R. Rosenthal, 11. interpersonal coordination: behavior matching and interactional synchrony in: R.S. Feldman, B. Rimé (Eds.). Fundamentals of Nonverbal Behavior (Studies in Emotion and Social Interaction), Cambridge University Press, Cambridge, 1991, pp. 401–432.
[17] G. Rizzolatti, M.A. Arbib, Language within our grasp, Trends Neurosci. 21 (1998) 188–194.
[18] A.N. Meltzoff, W. Prinz, The Imitative Mind: Development, Evolution and Brain Bases, vol. 6, Cambridge University Press, New York, 2002.
[19] J. Nadel, C. Guérini, A. Pezé, C. Rivet, The evolving nature of imitation as a format for communication, in: J. Nadel, G. Butterworth (Eds.). Imitation in Infancy Cambridge University Press, Cambridge, UK, 1999, pp. 209–234.
[20] M. Iacoboni, R.P. Woods, M. Brass, H. Bekkering, J.C. Mazziotta, G. Rizzolatti, Cortical mechanisms of human imitation, Science 286 (1999) 2526–2528.
[21] J. Decety, T. Chaminade, J. Grezes, A. Meltzoff, A pet exploration of the neural mechanisms involved in reciprocal imitation, Neuroimage 15 (2002) 265–272.
[22] P. Molenberghs, R. Cunnington, J.B. Mattingley, Is the mirror neuron system involved in imitation? a short review and meta-analysis, Neurosci. Biobehav. Rev. 33 (2009) 975–980.

[23] G. Dumas, J. Nadel, R. Soussignan, J. Martinerie, L. Garnero, Inter-brain synchronization during social interaction, PLoS ONE 5 (2010) e12166.
[24] G. Dumas, J. Martinerie, R. Soussignan, J. Nadel, Does the brain know who is at the origin of what in an imitative interaction? Front. Human Neurosci. 6 (2012) 128 doi: 10.3389/fnhum.2012.00128 See http://journal.frontiersin.org/Journal/10.3389/fnhum.2012.00128/abstract.
[25] S. Guionnet, J. Nadel, E. Bertasi, M. Sperduti, P. Delaveau, P. Fossati, Reciprocal imitation: toward a neural basis of social interaction, Cereb. Cortex 22 (2012) 971–978.
[26] N. Campbell, Multimodal processing of discourse information; the effect of synchrony, in: Second International Symposium on Universal Communication, 2008 (ISUC'08), IEEE, 2008, pp. 12–15.
[27] N. Campbell, Automatic detection of participant status and topic changes in natural spoken dialogues, 2008. Autumn Meeting of the Acoustical Society of Japan 2008 (ASJ'08).
[28] N. Campbell, An audio-visual approach to measuring discourse synchrony in multimodal conversation data, in: 10th Annual Conference of the International Speech Communication Association, INTERSPEECH 2009, Brighton, United Kingdom, September 6–10, 2009, ISCA, 2009, pp. 2159–2162.
[29] G. Varni, A. Camurri, P. Coletta, G. Volpe, Emotional entrainment in music performance, in: 8th IEEE International Conference on Automatic Face and Gesture Recognition, 2008 (FG'08), IEEE, 2008, pp. 1–5.
[30] G. Varni, G. Volpe, A. Camurri, A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media, IEEE Trans. Multimedia 12 (2010) 576–590.
[31] M. Komori, K. Maeda, C. Nagaoka, A video-based quantification method of body movement synchrony: an application for dialogue in counseling, Jpn. J. Interpersonal Soc. Psychol. 7 (2007) 41–48.
[32] F. Ramseyer, W. Tschacher, Synchrony: a core concept for a constructivist approach to psychotherapy, Construct. Human Sci. 11 (2006) 150–171.
[33] F. Ramseyer, W. Tschacher, Nonverbal synchrony or random coincidence? How to tell the difference, in: Development of Multimodal Interfaces: Active Listening and Synchrony, Springer, 2010, pp. 182–196.
[34] F. Ramseyer, W. Tschacher, Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome, J. Consult. Clin. Psychol. 79 (2011) 284–295.
[35] K.T. Ashenfelter, S.M. Boker, J.R. Waddell, N. Vitanov, E. Abadjieva, Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation, J. Exp. Psychol. Human 35 (2009) 1072.
[36] S.M. Boker, M. Xu, J.L. Rotondo, K. King, Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series, Psychol. Method 7 (2002) 338–355.
[37] R. Rienks, R. Poppe, D. Heylen, Differences in head orientation behavior for speakers and listeners: an experiment in a virtual environment, ACM Trans. Appl. Percept. 7 (2010) 2.
[38] A. Camurri, G. Varni, G. Volpe, Measuring entrainment in small groups of musicians, in: Third International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009 (ACII 2009), IEEE, 2009, pp. 1–4.
[39] G. Varni, A. Camurri, P. Coletta, G. Volpe, Toward a real-time automated measure of empathy and dominance, in: International Conference on Computational Science and Engineering, 2009 (CSE'09), vol. 4, IEEE, 2009, pp. 843–848.
[40] U. Altmann, Investigation of movement synchrony using windowed cross-lagged regression, in: Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues, Springer, 2011, pp. 335–345.
[41] E. Delaherche, M. Chetouani, Multimodal coordination: exploring relevant features and measures, in: Proceedings of the Second International Workshop on Social Signal Processing, ACM, 2010, pp. 47–52.
[42] C. Nagaoka, M. Komori, Body movement synchrony in psychotherapeutic counseling: a study using the video-based quantification method, IEICE Trans. Inform. Syst. 91 (2008) 1634–1640.
[43] X. Sun, K.P. Truong, M. Pantic, A. Nijholt, Towards visual and vocal mimicry recognition in human-human interactions, in: 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2011, pp. 367–373.
[44] X. Sun, K. Truong, A. Nijholt, M. Pantic, Automatic visual mimicry expression analysis in interpersonal interaction, in: 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2011, pp. 40–46.
[45] X. Sun, A. Nijholt, Multimodal embodied mimicry in interaction, in: Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues, Springer, 2011, pp. 147–153.
[46] A. Bickford, Using ELAN: a getting-started guide for use with sign languages, 2005. http://arts-sciences.und.edu/summer-institute-of-linguistics/teaching-linguistics/_files/docs/using-elan.pdf
[47] H. Lausberg, H. Sloetjes, Coding gestural behavior with the neuroges-elan system, Behav. Res. Method 41 (2009) 841–849.
[48] J. Zhang, M. Marszałek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: a comprehensive study, Int. J. Comput. Vision 73 (2007) 213–238.
[49] I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, Learning realistic human actions from movies, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008 (CVPR 2008), IEEE, 2008, pp. 1–8.
[50] L. Fei-Fei, P. Perona, A bayesian hierarchical model for learning natural scene categories, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005 (CVPR 2005), vol. 2, IEEE, 2005, pp. 524–531.
[51] I. Laptev, On space-time interest points, Int. J. Comput. Vision 64 (2005) 107–123.
[52] P. Dollár, V. Rabaud, G. Cottrell, S. Belongie, Behavior recognition via sparse spatio-temporal features, in: Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, IEEE, 2005, pp. 65–72.
[53] S. Canu, A. Smola, Kernel methods and the exponential family, Neurocomputing 69 (2006) 714–720.
[54] H. Kadri, M. Davy, A. Rabaoui, Z. Lachiri, N. Ellouze, Robust audio speaker segmentation using one class SVMs, in: European Signal Processing Conference (EUSIPCO-2008), 2007, Switzerland. http://www.eurasip.org/Proceedings/Eusipco/Eusipco2008/index.html <hal-00510423>.
[55] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, Neural Comput. 13 (2001) 1443–1471.
[56] E. Delaherche, S. Boucenna, K. Karp, S. Michelet, C. Achard, M. Chetouani, Social coordination assessment: distinguishing between shape and timing, in: Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction, Springer, 2013, pp. 9–18.
[57] MATLAB, version 7.14.0 (R2012a), The MathWorks Inc., Natick, Massachusetts, 2012.
[58] S. Boudet, L. Peyrodie, P. Gallois, C. Vasseur, Filtering by optimal projection and application to automatic artifact removal from eeg, Signal Process. 87 (2007) 1978–1992.
[59] M. Chavez, M. Besserve, C. Adam, J. Martinerie, Towards a proper estimation of phase synchronization from time series, J. Neurosci. Method 154 (2006) 149–160.
[60] J.P. Lachaux, E. Rodriguez, J. Martinerie, F.J. Varela, et al., Measuring phase synchrony in brain signals, Human Brain Map. 8 (1999) 194–208.
[61] S. Feese, B. Arnrich, G. Tröster, B. Meyer, K. Jonas, Detecting posture mirroring in social interactions with wearable sensors, in: Proceedings of the 15th Annual IEEE International Symposium on Wearable Computers (ISWC), 2011, pp. 119–120.
[62] O. Weisman, E. Delaherche, M. Rondeau, M. Chetouani, D. Cohen, R. Feldman, Oxytocin shapes parental motion characteristics during parent-infant interaction., Biol. Lett. 9 (2013) e20130828.
[63] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, D. Cohen, Interpersonal synchrony: a survey of evaluation methods across disciplines, IEEE Trans. Affect. Comput. 3 (2012) 349–365.

Please cite this article as: E. Delaherche et al., Automatic measure of imitation during social interaction: A behavioral and hyperscanning-EEG benchmark, Pattern Recognition Letters (2014), http://dx.doi.org/10.1016/j.patrec.2014.09.002