Machine Perception and Robotics for Understanding Interpersonal Synchrony

Mohamed CHETOUANI
Professor
Institut Systèmes Intelligents et Robotique (ISIR)
Université Pierre et Marie Curie - Sorbonne Universités

Mohamed.Chetouani@upmc.fr
Motivations

- Understanding social interaction is one of the big issue for the next decade: Characterization and interpretation of social interactions, Designing social interfaces

- Social signal processing and social intelligence (around 2005)

  *Inferring from signals social informations (emotions, role...)*

- **Main issues:** Multimodality, Real-life, Context
Social Signal Processing

Reality mining (Pentland, 2005)

Role recognition during meetings (Vinciarelli, 2007)

Building Autonomous Sensitive Artificial Listeners (Schröder et al., 2011)

Social learning (Meltzoff et al., 2009)
Human communication dynamics (Delaherche et al. 2012a):
- Computational models with explicit notion of social interaction
- From signal processing to interpretation of behaviours
- Inter-personal interaction: mutual and dynamic influence of partners
- Key concepts in psycho-pathology and robotics

Social signal processing

Still face experiments

Cerebral basis

- Imitation
- Synchrony
- Posture
- Body movements
- Non-verbal cues

Facial expressions
- Mutual gaze
- Prosody
- Speech
- Verbal cues

Social signal

Model

Imitator

Alpha-Mu 8-12Hz
Beta 13-30Hz
Gamma 31-48Hz

Physiology

Mean Duration, seconds

Touch
Social Reciprocity
Social Gaze
Object Manipulation

Father
Infant

Dumas et al., 2011

Weisman et al., 2012

Weisman et al., 2012
Human communication dynamics

Definition?

«The degree to which the behaviors in an interaction are non-random, patterned, or synchronized in both timing and form» (Bernieri et al., 1988)

Social resonance, mirroring, mimicking, matching, congruence, imitation, convergence, the chameleon effect... or interactional synchrony

Lack of synchrony

Trevarthen + Nadel

Micro-annotation of behaviours
Statistical modeling

Bernieri et al., 1991
Messinger et al. 2010

Mono-modal approaches
Few explicit cross-modal models

Ramseyer et al., 2006

Evaluation of Inter-personal synchrony

Controlled behaviors, meetings

Metrics: correlation, coherence, recurrent analysis

1. Role of social signals: motherese, motion, turn-taking and others...

2. Modeling synchrony: a focus on motor imitation

3. Synchrony and social intelligence for personal robots

4. Using social signal processing and developmental robotics for clinical investigations in autism
Role of social signals in synchrony

- Human communication dynamics
  - Nature of signals
  - Rhythm

Window analysis

P1
- voc.
- gaze
- smile

P2
- smile
- gest.
- voc.

Meta-signal
- voc.
- smile

High and low level informations

Multi-modal integration

Interactional synchrony characterization

Role of social signals in synchrony

- Toward a model of synchrony (Delaherche et al. 2012a)
  - General approach for characterization
  - From social perception to social interaction
  - Useful in various models

Some examples

- Developmental psychology: modeling parent-infant interaction (Saint-Georges et al. 2011)
- Cognitive robotics: social engagement (Al Moubayed et al. 2009)

Human communication dynamics

Using high-level information:

- **Real-life corpus**: Family home movies
- Manually annotated by psychologist:
  - **Infant behaviors**: vocalization, behaviors with objects, orienting toward people...
  - **Parent behaviors**: vocalization, touching....

C. Saint-georges et al. : Do parents recognize autistic deviant behavior long before diagnosis? taking into account interaction using computational methods. *PLOS one*, 2011

Meta-signal

\[ \text{voc.} \quad \text{smile} \]

Multi-modal integration

Interactional synchrony characterization

bi-gram  Non-negative Matrix
Human communication dynamics

Non-negative matrix modeling of interactive situations

One of the first initiatives to employ data mining methods for understanding social interactions

C. Saint-georges et al. : Do parents recognize autistic deviant behavior long before diagnosis? taking into account interaction using computational methods. *PLOS one*, 2011
Human communication dynamics

Non-negative matrix modeling of interactive situations:
Part-based representation

Data vector (nxm)

$\mathbf{V} \approx \mathbf{W}\mathbf{H}$

Expansion/activation coeff. (kxm)

Basis vectors «dictionary» (nxk)

Basis Interactive Behaviors vectors


Original

VQ

PCA

Original
Application to investigations on early signs of autism:

- Diagnostic > 36 month
- Developmental issues: Semesters S1 (0-6 months), S2 (6-12) and S3 (12-18)
- Comparisons of clusters obtained by NMF: Typical development, Intellectual disability, Autism

Normalized Mutual Information

«Deviant behaviors»:
✓ Reality mining method used by clinicians
✓ Coherent with qualitative impressions of clinicians
Role of social signals in synchrony

2A

Infant’s synchrone response and infant’s presence improve motherese

Motherese works in an interactive loop

Motherese improves infant’s affective engagement, reactivity and attractivity

2B

Decisive role of affect

Emotional charge is responsible for the prosodic features of motherese and triggers infant’s preference ⇒ motherese of depressed mothers is unable to promote infant’s learning

Infant’s preference helps infant’s engagement

Motherese helps affect transmission and sharing (accordage)

Socio-cognitive development

Motherese stimulates attention, joint attention, learning

Motherese improves language acquisition

1. Role of social signals: motherese, motion, turn-taking and others...

2. Modeling synchrony: a focus on motor imitation

3. Synchrony and social intelligence for personal robots

4. Using social signal processing and developmental robotics for clinical investigations in autism
Imitation characterization through social signal processing

- **Social learning**
  - Infant’s development
  - Learning in robotics

- **Problem:**
  - Modeling imitation during interaction

- Computational modeling of synchrony (Delaherche et al 2012b):
  - **Time** (rhythm of partners, delay between responses)
  - **Pattern** (similar gesture)
Unsupervised action recognition (Delaherche et al. 2012b)

Imitation characterization through social signal processing

- Unsupervised learning of a model A
- Unsupervised learning of a model B
- Evaluation of model B on data of A
- Evaluation of model A on data of B
- Metrics in the model space

Interpretation:
Novelty detection

\[
\begin{align*}
H_0 : P_A &= P_B \text{ (the gestures are identical)} \\
H_1 : P_A &\neq P_B \text{ (the gestures are different)}
\end{align*}
\]
Imitation characterization through social signal processing

- Unsupervised action recognition (Delaherche et al. 2012b)

![Image](image-url)

**Distance.**

denotes the size of the histogram.

**Table 1.**

| Frequency | Synchrony and asynchrony (Kadri et al., 2008) to define the distance in which a model learned on the histograms of a given threshold. The likelihood is determined by solving the 1-class SVM problem:

\[
    h_i = \begin{cases} 1 & \text{if } \exists P \in \mathbb{R}^n : \frac{1}{n} X_i \cdot P \geq \frac{1}{n} \sum_{i=1}^n X_i \cdot P \\
    0 & \text{ otherwise} \end{cases}
\]

- SVM

**Fig. 6.**

**Outliers**

Non-Imitation classifiers.

- 1-class SVM

**Feature space**

- 1-class SVM Model A

**Fig. 7.**

Imitation assessment in interaction.
Imitation characterization through social signal processing

- Unsupervised action recognition (Delaherche et al. 2012b)

The aim of 1-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:

\[
R_h = \{ h \in X \mid f(h) \geq 0 \}
\]

Decision function:

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

- \( h \) represents an histogram of codewords
- Intersection kernel:

\[
k(h_i, h_j) = \sum_{i=1}^{d} \min(h_i, h_j)
\]
Imitation characterization through social signal processing

Unsupervised action recognition (Delaherche et al. 2012b)

Unsupervised learning of a model A

Unsupervised learning of a model B

Evaluation of model B on data of A

Evaluation of model A on data of B

Distance:
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 $h_{B_i}$

$$s_{A_iB_j} = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \alpha^A_i k(h_{B_j}, h_{A_i}) \right) + \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \alpha^B_i k(h_{A_j}, h_{B_i}) \right)$$
Imitation characterization through social signal processing

- How to analyze the dynamics?
  - Delay between to gestures
  - Response time
  - ...

Interest point detection

Feature description

HOG/HOF

Codebook generation

K-means

Mapping to codebook

Histogram of visual words

Imitation classifier

1-class SVM Model A

1-class SVM Model B
Imitation characterization through social signal processing

- Recurrence analysis assesses the points in time that two systems visit similar states, called "recurrence points".
- They represent the points in time that the two systems show similar patterns of change or movement.

Imitation characterization through social signal processing

Consider, for example, two time series of numeric measurements.

First, time-delayed vectors $v$ of $m$ points are constructed from the time series, where $m$ represents the embedding dimension and $t$ the delay between sequential time-points.

Every vector from the first time-series is compared with every vector from the other time-series using a distance measure (e.g., Euclidean Distance).

A cross-recurrence matrix is created at this stage.

A threshold on the distance between vectors is fixed to decide whether two vectors are similar or not.

A time-point $(i, j)$ on the cross-recurrence matrix is set to 1 if the vectors $i$ and $j$ are similar and set to 0 otherwise.

The cross-recurrence plot is the two-dimensional representation of the cross-recurrence matrix.

Recurrence analysis assesses the points in time that two systems visit similar states, called "recurrence points". They represent the points in time that the two systems show similar patterns of change or movement.

FIGURE 6 | A basic sketch of how recurrence is constructed from one time series (top left). The time series is lagged (by 10), copied (3 times), and overlaid with itself (top right). If we use 3 dimensions (copies), then it is possible to visualize this reconstructed phase space (bottom left). By drawing a radius of a given size around parts of this reconstructed phase space (thick line, bottom left), one can determine when recurrence is taking place. The time indices of these recurrence points can be used to construct the recurrence plot (bottom right). Cross recurrence is done in almost exactly the same way, except two time series are used.

Imitation characterization through social signal processing

- Metrics:
- (%REC) is the percentage of recurrent points on the plot:
  - Ranging from 0% to 100%, it informs on the degree to which both systems tend to visit similar states.
- Diagonal structures represent periods in one time series that show a similar trajectory as another time series at a different time.
  - Stochastic behavior tends to produce very short diagonals whereas deterministic behavior produces longer diagonals.
  - The rate of recurrence points forming diagonal lines is informative of the determinism of the interaction between the two time series.
  - The average length of the diagonal line represents the duration that both systems stay attuned.
- By computing a histogram of the length of all diagonals from the plot, we can deduce the entropy of the cross-recurrence plot.
  - Entropy reflects the complexity of the deterministic structure in the system.
Back to our imitation detection problem:

- **1-Class SVMs provide a metric**
- **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.

- The recurrence matrix represents the points in time when the dyadic partners are in similar states.
Imitation characterization through social signal processing

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

Imitation characterization through social signal processing

- Using behaviors to analyze brain synchronization

Imitation characterization through social signal processing

- Using behaviors to analyze brain synchronisation
  (Delaherche et al., 2015)

![Diagram showing brain activity and behavior synchronization]

Linking behaviors and cerebral activities

![Brain activity visualization with T values]

(a) Manual indexing

(b) Automatic indexing

Imitation characterization through social signal processing

- Using EEG to analyze behaviors

Non-Negative Matrix Factorization A

Non-Negative Matrix Factorization B

Co-factorization

Metrics in the model space

\[ V_{f,t,i} = \sum_k W_{f,k} H_{t,k} Q_{i,k} \]

- \( f \) and \( t \) are signals in the model space.
- \( i \) is a temporal activation matrix.
- \( k \) is a pattern matrix.
Imitation characterization through social signal processing

- Using EEG to analyze behaviors
Outline

1. Role of social signals: motherese, motion, turn-taking and others...

2. Modeling synchrony: a focus on motor imitation

3. Synchrony and social intelligence for personal robots

4. Using social signal processing and developmental robotics for clinical investigations in autism
Social intelligence for personal robots

- Social intelligence influences (shared) task performance:
  - Second-perspective taking
  - Dynamics of interaction (synchrony)

- Experiment:
  - Object learning through human interaction

Ivaldi et al. : Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. *Frontiers in Neurorobotics* (2014)
Social intelligence for personal robots

Protocol:
- Robot initiates the phase
- Human initiates the phase

Teaching phase

RI
- I don’t know this object.
- What is its color?

HI
- Please, choose an object.
- What is its color?

Verification phase

- Please, choose an object.
- It is green!

Ivaldi et al.: Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. *Frontiers in Neurorobotics* (2014)
Some results:

Ivaldi et al.: Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. *Frontiers in Neurorobotics* (2014)
Subjects in the RI group react faster than the ones of the HI group, and the interaction with the robot has a higher rhythm

Table 2. Reaction time (seconds) in response to robot attention stimuli (utterances) during verification phase

<table>
<thead>
<tr>
<th>Group</th>
<th>mean</th>
<th>std</th>
<th>median</th>
<th>Wilcoxon’s test</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI</td>
<td>1.932</td>
<td>0.711</td>
<td>1.917</td>
<td>W=418, p-value=1.6e-5</td>
</tr>
<tr>
<td>RI</td>
<td>1.296</td>
<td>1.145</td>
<td>1.106</td>
<td>p-value=0.005</td>
</tr>
</tbody>
</table>

Table 3. Time interval (seconds) between consecutive robot attention stimuli (utterances) during verification phase

<table>
<thead>
<tr>
<th>Group</th>
<th>mean</th>
<th>std</th>
<th>median</th>
<th>Wilcoxon’s test</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI</td>
<td>9.524</td>
<td>1.515</td>
<td>8.588</td>
<td>W=447; p-value=0.128</td>
</tr>
<tr>
<td>RI</td>
<td>7.287</td>
<td>1.653</td>
<td>7.257</td>
<td>p-value=1.6e-5</td>
</tr>
</tbody>
</table>

Ivaldi et al.: Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement. *Frontiers in Neurorobotics* (2014)
Social intelligence for personal robots

- Combining social and physical interaction
- Engagement during Human Humanoid Interaction

1) acceptance test
2) teaching through physical interaction
3) functional acceptability
4) social acceptability
Outline

1. Role of social signals: motherese, motion, turn-taking and others...
2. Modeling synchrony: a focus on motor imitation
3. Synchrony and social intelligence for personal robots
4. Using social signal processing and developmental robotics for clinical investigations in autism
Role of social signals in synchrony

- Low-resolution brain scanning
  - Oxytocin modulates proximity (kind of motionese)
  - Infant’s OT reactivity positively correlated with father’s head acceleration

Role of social signals in synchrony

- Low-resolution brain scanning
  - Oxytocin modulates proximity (kind of motionese)
  - Infant’s OT reactivity positively correlated with father’s head acceleration

Extraction of social signatures during Human-Robot Interaction

- Case of Human-Human Interaction
  - Mutual influence of partners
  - Paradigm-shift Looking at partner A to analyze partner B!

Continuous space

Speech turns, Gestures, Interpersonal synchrony...

SVM regression

Correlation between speech turns and developmental age of child

Extraction of social signatures during Human-Robot Interaction

- Social learning (Boucenna et al. 2014)

  - Proposition: Evolution of computational model's parameters inform us about the human partner.

Boucenna et al.: Learning of social signatures through imitation game between a robot and a human partner.

*IEEE Transaction on Autonomous Mental Development (2014)*
Extraction of social signatures during Human-Robot Interaction

- Social learning (Boucenna et al. 2014)
  - Proposition: Evolution of computational model’s parameters inform us about the human partner

<table>
<thead>
<tr>
<th></th>
<th>ASD (N=15)</th>
<th>TD (N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (± SD), year</td>
<td>9.25 (± 1.82)</td>
<td>8.06 (± 2.49)</td>
</tr>
<tr>
<td>Male - Female</td>
<td>13-5</td>
<td>9-6</td>
</tr>
<tr>
<td>ADI-R, current, mean (± SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social impairment score</td>
<td>10.77 (± 5.3)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Verbal communication score</td>
<td>7.72 (± 4.22)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Non verbal communication score</td>
<td>4.3 (± 3.5)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Repetitive interest score</td>
<td>2.5 (± 1.88)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Developmental score</td>
<td>3.3 (± 1.5)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Total score</td>
<td>31.1 (± 5.46)</td>
<td></td>
</tr>
<tr>
<td>ADI-R, 4-5 years, mean (± SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social impairment score</td>
<td>17.33 (± 8.47)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Communication verb score</td>
<td>13.75 (± 5.72)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Communication non-verb score</td>
<td>8.08 (± 4.4)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Repetitive interest score</td>
<td>5.25 (± 3.52)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Developmental score</td>
<td>3.83 (± 1.47)</td>
<td>Not relevant</td>
</tr>
<tr>
<td>Total score</td>
<td>48.25 (± 7.34)</td>
<td></td>
</tr>
<tr>
<td>Developmental age</td>
<td>7.47 (± 2.9)</td>
<td>8.06 (± 2.49)</td>
</tr>
<tr>
<td>IQ*</td>
<td>73 (± 14)</td>
<td>All controls &gt; 80</td>
</tr>
<tr>
<td>GAF score</td>
<td>40.27 (± 9.44)</td>
<td>All controls &gt; 90</td>
</tr>
<tr>
<td>Imitation score / therapist**</td>
<td>18.0 (± 3.46)</td>
<td>19.66 (± 1.29)</td>
</tr>
<tr>
<td>Imitation score / Nao**</td>
<td>17.27 (± 5.24)</td>
<td>19.53 (± 1.81)</td>
</tr>
</tbody>
</table>

Recognizing human postures

Boucenna et al. : Learning of social signatures through imitation game between a robot and a human partner. IEEE Transaction on Autonomous Mental Development (2014)
Extraction of social signatures during Human-Robot Interaction

- Social learning (Boucenna et al. 2014)
  - Proposition: Evolution of computational model’s parameters inform us about the human partner

![Evolution of parameters for different groups during learning](image1)

- Changing group during learning

![Changing group during learning](image2)

Boucenna et al. : Learning of social signatures through imitation game between a robot and a human partner. *IEEE Transaction on Autonomous Mental Development (2014)*
Extraction of social signatures during Human-Robot Interaction

- Joint attention (Anzalone et al. 2014)
  - Proposition: How to extract social cues of joint attention during interaction?

Social Signal Processing for studying Parent-Infant Interaction

- Investigating lack of synchrony:
  - Severe emotional neglect
  - Coding Interactive Behavior

Social Signal Processing for studying Parent-Infant Interaction

- Investigating lack of synchrony:
  - Severe emotional neglect
  - Coding Interactive Behavior

![ROI_Maman](image1.jpg) ![ROI_Bebe](image2.jpg)

<table>
<thead>
<tr>
<th>Pathological dyad</th>
<th>Control dyad</th>
</tr>
</thead>
</table>
| Shoulder orientation | Mother mostly oriented toward table and bench  
Child focused almost exclusively on the table |
| Relative shoulder orientation | Dyad focused essentially on the task, just three periods when they are facing |
| Conclusion | Shoulder orientation of the child and parent in the pathological dyad is less mobile than the control dyad. That could be interpreted as a poorer ability to share attention while alternating the focus of attention. The control dyad illustrates a fluid alternation of attention |

![ROI_Maman](image3.jpg) ![ROI_Bebe](image4.jpg)

Table 3

<table>
<thead>
<tr>
<th>Attention to the task and the partner features – results analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathological dyad</td>
</tr>
</tbody>
</table>
| Shoulder orientation | Mother mostly oriented toward table and bench  
Child focused almost exclusively on the table |
| Relative shoulder orientation | Dyad focused essentially on the task, just three periods when they are facing |
| Conclusion | Shoulder orientation of the child and parent in the pathological dyad is less mobile than the control dyad. That could be interpreted as a poorer ability to share attention while alternating the focus of attention. The control dyad illustrates a fluid alternation of attention |

![Figure 6](image5.jpg)

**FIGURE 6**
Evolution of relative shoulder orientation during the interaction (Left: pathological dyad; Right: control dyad).

In this graph, we report the shoulder orientation according to the relative angle between the two partners' shoulders over time. When the angle is equal to 0°, the partners are facing. When to the angle is 45 to 90°, both shoulders are oriented in the direction of the table that is a point of interest in the given task. In the left graph, the pathological dyad is focused essentially on the task, as partners are facing only three times. In contrast, the control dyad had many face to face positions and showed clear turns between task focusing and other partner focusing. A blank or a cross line in figures indicates uncollected data.

Conclusions

‣ Modeling and characterizing human communication dynamics

‣ Models and methods for the evaluation of inter-personal synchrony

‣ Robot is employed as a tool for both stimulation and clinical investigation

‣ New ways to study social interactions: «Low-Brain Resolution Scanning» (Pentland)
Thank you for your attention
The MICHELANGELO project intends to bring the assessment and the therapy of autism out of the clinical environment and develop a patient-centric home-based intervention requiring minimal human involvement and therefore extremely cost-effective.

The research leading to these results has received funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n° #288241.

Questions?

©FP7 Michelangelo