

# Multimodal Coordination: Exploring Relevant Features and Measures

Emilie Delaherche  
emilie.delaherche@etu.upmc.fr

Mohamed Chetouani  
mohamed.chetouani@upmc.fr

Institut des Systèmes Intelligents et de Robotique  
CNRS UMR 7222  
Université Pierre et Marie Curie - Paris 6  
4 Place Jussieu  
Paris, France

## ABSTRACT

The decisive role of interpersonal coordination in social interactions is well established. As human beings, we are “experts” in decoding and producing social signals, learning since our birth. Though, equipping a machine with the same abilities represents a great challenge to design socially-accepted interfaces. Assessing automatically this coordination requires selecting the most relevant features. In this paper, various audio and visual features are investigated to characterize interactional synchrony in a cooperative task. A hierarchical representation of the similarity between extracted features is proposed. The properties of time-domain (correlation coefficients) and frequency-domain (coherence) synchrony measures are confronted. The presented results support the evidence that synchrony is a multimodal phenomenon. Moreover, it can be assessed both on local (short-length windows of interaction) or global scales (entire interaction).

## Categories and Subject Descriptors

I.2.11 [Distributed Artificial Intelligence]: Coherence and coordination

## General Terms

Algorithms, Measurement.

## Keywords

Interactional synchrony, cooperation, multimodality.

## 1. INTRODUCTION

The ability to understand, exploit and give in return social signals to a partner defines our social skills and our ability to engage in a successful interaction. The analysis and synthesis of those social signals present a great interest in order to design intelligent social interfaces.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SSPW'10, October 29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-4503-0174-9/10/10...\$10.00.

The adoption of an appropriate behaviour in response to one's solicitation requires a proper knowledge of the “rules” of coordination between two partners.

These “rules” are not static; they depend on the context (are you attending a wedding or a funeral?), on your relationship with your partner (is he a perfect stranger, your boss, your friend?), on your own personal background... and most of all on your partner's behaviour (Does he speak slowly, loudly? Does he gesture a lot? Does he smile a lot?).

Many studies have shown how partners tend to adopt unconsciously a part of each other's behaviour during a social interaction: convergence of inter-speaker silence duration, gestures congruence, rhythm convergence [7]... Various terms are used in the literature to define this coordination process: mirroring, mimicking, matching, congruence, imitation, convergence, the chameleon effect, social resonance... or interactional synchrony. Bernieri et al. define interactional synchrony as “the degree to which the behaviours in an interaction are non-random, patterned, or synchronized in both timing and form” [10]. We may distinguish interactional synchrony from intrapersonal synchrony, which refers to the use of several modalities to convey the same information, in order to draw someone's attention, for instance.

The role of interactional synchrony in social interactions has also been studied. It has been demonstrated that coordinated behaviours often give a good insight into the level of rapport between social partners [13], the engagement into a social interaction or the smoothness of an interaction [8].

Moreover, our perception of rapport between two partners is closely related to their effective coordination: in [5], Miles et al. explored the perceived degree of rapport between a pair of walkers whether their strides were temporally coordinated or not. The highest levels of rapport were perceived when their movements were highly coordinated (in-phase or anti-phase). In [9], Lakens established a link between our perception of groups forming an entity, a social unit and movement synchrony.

And occurrence of interactional synchrony deeply relies on the social skills of the partners. People with impaired social abilities like autism show asynchronous social behaviours with their partner [6] : absence of answer to a partner's elicitation, few eye contact, few anticipation...

However, if the link between quality of interaction and interactional synchrony seems clearly established, the way to measure and make use of this social cue to design social interfaces needs further investigation. Indeed, synchrony is not continuously observed in social interactions. As the saying

“Monkey see, Monkey do” refers to the learning of a process without understanding its rules, an interface proposing perfect imitation of its partner's behaviour would be disconcerting. We all know children's (successful) attempts to annoy a partner, repeating and mimicking each and every one of his words and gestures.

This suggests that synchrony is a much larger and subtler process than simple imitation.

According to various parameters (context, partner, personal preferences...), we may coordinate with some preferred channels or modalities. And, we may adjust to one modality with another (e.g. listener's head movements coordinated with speaker's speech). At last, synchrony occurs at certain special moments during interaction, which need to be investigated.

Thus a deeper analysis of the occurrence of interactional synchrony and its preferred channels represents a great concern in order to design intelligent social interfaces. This study addresses the question of interactional synchrony measurement. Our purpose is to explore several audio and video features to evaluate synchrony and try new types of measurements.

We present our experimental set-up in section 2. Then we describe our model to assess synchrony in section 3. At last, we present and discuss our results in sections 4 and 5.

## 2. EXPERIMENTAL SETUP

We set up a cooperative task involving two partners, a “demonstrator” and an “experimenter” to investigate interactional synchrony.

Originally, the proposed task was designed by a language therapist to evaluate the ability of young children (8-11 years old) to give instructions to a third party.

### 2.1 Participants

Sixteen students from ISIR Laboratory participated in this study. The « demonstrator » and « experimenter » roles were randomly assigned to each participant.

### 2.2 Procedure

The task consisted of putting together a clown with polystyrene elements. Parasites elements (which do not belong to the clown) were added to the original elements to complicate the task. Participants were sitting on both sides of a table, in front of each other, having the same polystyrene elements in front of them. A folding screen prevented the partners to see each other gestures. Thus, they could still gaze and see each other faces.



Figure 1. Clown puzzle setup

Each participant read the instructions before entering the room dedicated to the experiment. The purpose of the study (assess

interactional synchrony) was intentionally concealed from the participants to not influence on their behavior. The demonstrator was asked to guide his partner in putting together a clown with the polystyrene elements. The experimenter was asked to follow the instructions of the demonstrator and to question him, if his instructions were not accurate enough. At last, the experimenter was not informed that he was putting together a clown, to prevent him from completing the task without his partner's help.

The task ended when the experimenter had fulfilled his assembly work.

Figure 1 presents our experimental setup. Synchrony investigated here is related to *operating* synchrony (i.e. “how both partners coordinate to accomplish a given task?”). We intentionally proposed a task where coordination between partners is high by design, to focus on the study of relevant features and measures to assess this operating synchrony.

### 2.3 Material

The interactions were recorded using a Canon MV800 camera, put up above the participants. The audio recordings were collected at 32 kHz and the video recordings at 25 fps.

Both partners were equipped with colour bracelets to track easily their hands.

Audio data were annotated in Anvil annotation tool [16] to delimit both partners' speech turns.

One video was finally discarded because one of the participants did not understand the instructions properly. The total duration was approx. 35 min with a mean duration 4min57s and a standard deviation 4min17s.

## 3. COMPUTATIONAL MODEL

Previous attempts have been made to assess synchrony. Rolf et al. proposed a computational model of intra-personal synchrony based on *mutual information* [12]. They chose *audio energy* and *pixel intensity* to characterize respectively auditory and visual channels. Ramseyer et al. investigated interactional synchrony between psycho-therapists and their patient's movement [13]. Their model was based on *cross-correlation* between *motion energy image* features.

The audio and video features we chose to investigate and the synchrony measures we implemented are presented below. Figure 2 gives an overview of our computational model.

### 3.1 Audiovisual features

#### 3.1.1 Sampling rates

As audio and video were not sampled at the same frequencies, we set the window size for audio features extraction to be the same as the duration  $T$  between two video frames, with an overlap of duration  $T/2$ .

The video features were extracted between two consecutive frames of the video. A linearly interpolated sample was added between two video features samples.

#### 3.1.2 Audio features

Audio features were extracted every 20ms on 40ms windows of signals.

**-Pitch** is extracted on the voiced part of speech and represents its relative highness or lowness. We selected pitch as its contour or its emotional state.

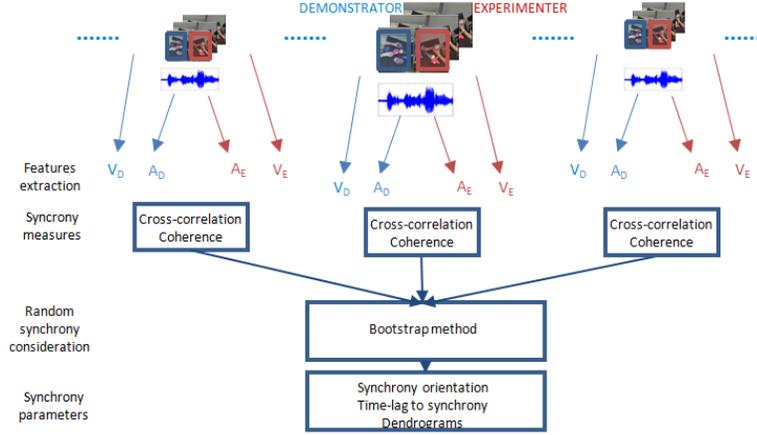


Figure 2. Overview of our computational model.

**-Energy** accounts for the loudness of the sound. It is extracted in dBs. Energy is computed as the averaged square value of the samples in a window signal. It has been previously used for synchrony assessment [12], intention or emotion recognition. *Pitch* and *energy* were extracted with Praat Tool.

**-Pause:** Pitch and energy features carry no information in the absence of speech. Nevertheless, pauses contain cues on the interactional state (next turn planning, information processing and communication breakdown). Though, we decided to design a binary feature to account for these pauses; it takes the value 1 when none of the partners was speaking and 0 otherwise. An audio window was considered as silent if the audio energy was below a predefined threshold and its duration was larger than 500ms. As filled pauses (“uh”, “uhm”...) contain comparatively equivalent audio energy as informative speech, note this feature only accounts for empty pauses.

**-Vocalic energy** was obtained by applying a set of perceptual filters to the audio signal. The method is fully described in [2]. It was originally designed by Cummins to detect the p-centers in speech. The resulting signal presents the evolution of the vocalic energy of the speaker. The peaks of this signal tend to be the most salient moments in speech. We intend to investigate if those salient points coincide with some behavioural response in the listener's attitude.

### 3.1.3 Video features

For each speaker, a region of interest was selected in the video. We extracted the following features for each region of interest:

**-Motion energy (ME)** is defined as the number of pixels in movement between the current video frame and a reference image. We computed the difference between those two images. For each pixel, if the difference was above a predefined threshold, the pixel was regarded as “in movement”. The reference image was updated after each iteration as a weighting sum of the current frame (weight  $\alpha = 0.1$ ) and the previous reference image (weight  $1-\alpha$ ). Precisiously studied to assess interactional synchrony [13], motion energy informs about a shared dynamics between the two partners movement or a listener's coordinating his movement with the speaker's speech.

**-Motion history image (MHI)** [1] is defined as the number of pixels in movement across N video frames (N=25). This feature is

derived from *Motion Energy*. A weight (between 0 and 1) is assigned to each pixel according to the recentness of its movement. Thus, the pixels “in movement” in the current frame of the video are given an important weight, the ones in movement in the previous frame a smaller weight, etc. This feature gives an interesting idea of the trajectory of movement during N frames and consequently on complete gestures. We may detect with this feature a loose synchrony between the two partner's movements.

**-Mean velocity** [3]. Motion energy measures the amount of movement between two video frames but can't discriminate between a large region moving slowly or a small region moving quickly. So, we computed an additional feature to measure how fast a person is moving. The velocity (direction and magnitude) of each pixel was computed by the Lucas and Kanade optical flow method [11]. Then mean velocity was obtained by averaging the velocity magnitude of all pixels in the region of interest.

**-Hands' trajectory.** When we compute motion energy or velocity, we consider the overall movement in the video: gesture, leaning, head nods... We decided to add a last feature to focus on hands' gestures. Hands' tracking was computed with the coupled Camshift (Continuously adaptative meanshift) algorithm [4]. At last, we compressed  $x$  and  $y$  Cartesian coordinates to the polar coordinate  $r = \sqrt{x^2 + y^2}$ .

Note that audio and video features were rescaled (minimum value subtraction and range division) to make all the elements lie between 0 and 1.

We obtained 14 features for each window of interaction (4 audio features belonging to one partner or the other according to the section of the video, 5 visual features for demonstrator and 5 visual features for experimenter).

## 3.2 Measures of synchrony

### 3.2.1 Correlation coefficient

Correlation coefficient estimates the linear relationship between two random variables  $x$  and  $y$ .

$$r_{xy} = \frac{\sum_n (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_n (x_n - \bar{x})^2} \sqrt{\sum_n (y_n - \bar{y})^2}}$$

It is normalized between [-1; 1]. The sign of the correlation coefficient indicates the direction of association between  $x$  and  $y$ . Though, it is positive when both variables increases at the same time and negative when one variable increases while the other

decreases. A correlation of 0 indicates that there is no dependency between the random variables. Accounting for positive or negative association, we considered absolute magnitude of correlation coefficients.

### 3.2.2 Magnitude coherence

Magnitude coherence is a function of frequency; it assesses the linear correlation between two signals  $x$  and  $y$  in the frequency domain.

$$\Gamma_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}$$

where  $P_{xy}(f)$  is the cross power spectral density of  $x$  and  $y$  and  $P_{xx}(f)$  is the power spectral density of  $x$ .

Magnitude coherence is particularly useful looking for synchronisation in a particular frequency band. Consequently, it is frequently assessed to study EEG signals [14].

Moreover, working in the frequency domain, we evade the question of time-lag between partners. Thus, magnitude coherence is much more related to syntony (i.e. the tuning between the time period of two oscillators) than to synchrony.

### 3.3 Synchrony scores

A synchrony matrix was built, performing synchrony measures across each possible pair of features described in section 3.1, for the entire videos. Synchrony matrix dimensions were 14x14. Each coefficient of the matrix represents the correlation or coherence between a pair of feature. Synchrony matrix is symmetric and the main diagonal is all ones.

Assuming that features involved in synchrony could vary across time, we also performed those measures on smaller units (one-second window with half-second overlap).

To account for a smooth synchronization between partners, we also computed these measures for +/- 1-4 seconds lagged windows.

### 3.4 Random synchrony consideration

Computing audiovisual synchrony, we needed to consider the chance that two features were correlated by coincidence. We adopted Ramseyer et al. bootstrap method to compare genuine synchrony with pseudo-interactions synchrony [13]. Genuine synchrony score was assessed on original features while 100 pseudo-interactions synchrony scores were computed on time-shuffled features. Genuine synchrony score was regarded as

significantly different from chance if it was two standard deviations above or below pseudo-interactions synchrony scores.

In the present experiment, we are investigating synchrony on a highly repetitive task. Though, shuffling one-second length windows from the videos would have led to equivalent synchrony scores on genuine interactions or pseudo-interactions. So, to break this repetitive structure, we build pseudo-interactions permuting every single point of data

This method was applied for both global (on the entire interactions) and local (on one-second length windows) measures.

### 3.5 Parameters of synchrony

We characterized synchrony between partners with three parameters.

**-Orientation of synchrony** which answers to the question who is driving the interaction. It is measured on the synchronous windows of the video, according to the time-lag between windows. A positive lag between partner 1's features and partner 2's features accounts for "Partner 1 is leading the interaction", a negative lag between partner 1's features and partner 2's features accounts for "Partner 2 leading the interaction". A zero lag between each partner's features accounts for mutual synchrony.

**-Time-lag to synchrony** is the time in seconds between a change in one of the partner's behaviour driving a corresponding change in the second partner's behaviour. It is assessed as the mean time-lag for all the synchronous frames in the video. A positive time-lag indicates a global dominance of the partner 1 while a negative lag indicates a global lead of the partner 2.

Orientation of synchrony and time-lag to synchrony were inspired by Feldman's work on parent-infant synchrony [15].

**-Hierarchy in features similarity:** Wondering how synchrony was established between features, we built dendrograms to represent the hierarchy in features similarity. Dendrograms are tree diagrams illustrating the hierarchical relationship between data. It is often used to represent a taxonomy or a hierarchical clustering in biology. U-shaped lines connect features according to their similarity. The height of each U-shaped line represents the distance between the two features being connected. The construction of a dendrogram relies on a dissimilarity symmetric matrix  $d_{xy}$ , containing the distance between every possible pairs of features. Synchrony matrices were transformed in dissimilarity matrices by computing  $d_{xy} = 1 - r_{xy}$  or  $d_{xy} = 1 - \Gamma_{xy}$ .

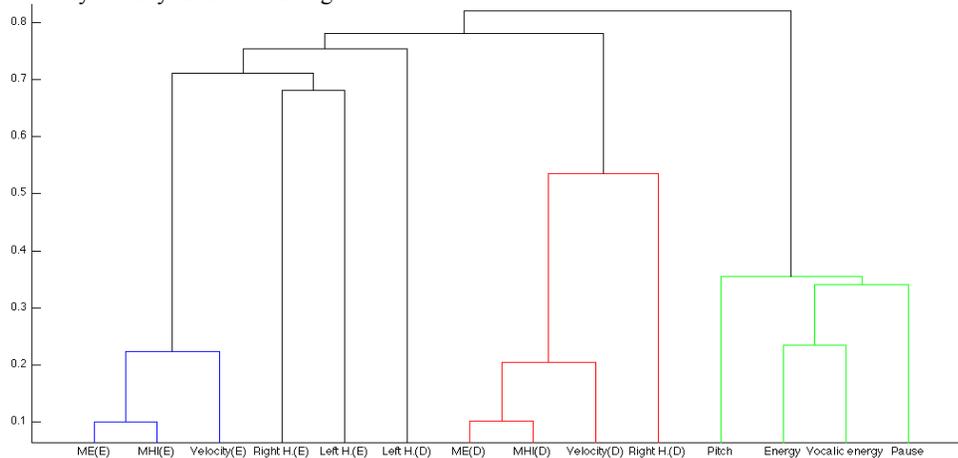


Figure 3. Dendrogram: a hierarchical representation of features similarity

## 4. RESULTS

### 4.1 Hierarchical relationship between features

Figure 3 presents a dendrogram representation of the relationship between features for one of the video we investigated. Low-level visual features belonging to each partner are highly correlated with one another (Motion Energy, Motion History Image and Velocity).

Their correlation with hands trajectory is less strong as low-level visual features consider the overall movement in the region of interest, and not hands gestures only. Audio features are also strongly related with each other, especially energy and vocalic energy. Then, a U-shape line connects the visual features of partners; this was predictable as the partners are performing the same task. At last, we found a weaker synchronization between audio and video features.

Looking for a shared structure in all the interactions, we compute Pearson's correlation coefficient between each dendrogram's distance coefficients. A significant relationship was found between the structure of all dendrograms ( $r > 0.743$ ,  $p < 0.0001$  for correlation,  $r > 0.874$ ,  $p$ -value  $< 0.0001$  for coherence).

### 4.2 Local synchronization

Looking for a shared pattern at a smaller scale, correlation and coherence were applied to one-second length windows of the signal. To investigate whether synchronization between demonstrator's and experimenter's features was significantly different from chance or not, we applied Ramseyer et al. normalization procedure to each pair of features. For each pair of features, we assessed the percentage of one-second length windows where synchrony measures were significantly different from chance across all interaction (two-sided Z-test). Tables 1.a and 1.b present these results. As demonstrator's speech turns cover a major part of the interactions, we focused on experimenter's visual features attunement to demonstrator's audio and video features.

Windows corresponding to each speaker turn were identified thanks to audio data annotation.

The letter in brackets refers to E the "Experimenter" or D the "Demonstrator".

**Table 1.a. Percentage of one-second windows presenting significant correlations across all interactions**

	Motion energy (E)	Motion history image (E)	Velocity (E)	Right Hand (E)	Left Hand (E)
Pitch (D)	28,93	31,82	20,81	19,71	14,18
Energy (D)	54,55	56,92	39,51	38,3	26,05
Pause (D)	38,7	41,3	28,7	30,23	20,46
Vocalic energy (D)	49,57	53,17	35,33	36,02	23,66
Motion energy (D)	73,23	78,96	56,86	53,66	37,69
Motion history image (D)	78,9	83,34	60	58,56	40,23
Velocity (D)	54,41	59,45	70,2	39,6	30,12
Right Hand (D)	61,5	71,21	49,02	65,82	53,29
Left Hand (D)	54,18	65,27	46,05	68,93	62,8

**Table 1.b. Percentage of one-second windows presenting significant coherence across all interactions**

	Motion energy (E)	Motion history image (E)	Velocity (E)	Right Hand (E)	Left Hand (E)
Pitch (D)	9,16	6,2	6,31	2,45	2,13
Energy (D)	46,46	56,89	36,57	54,44	37,2
Pause (D)	20,23	20,06	14,35	14,47	10,92
Vocalic energy (D)	22,45	19,51	14,64	6,22	3,98
Motion energy (D)	41,38	44,7	29,14	14,24	6,89
Motion history image (D)	49,63	61,3	37,03	37,78	20,12
Velocity (D)	37,64	39,05	28,39	9,25	4,03
Right Hand (D)	54,93	73,2	42,68	80,66	60,58
Left Hand (D)	54,64	73,63	42,51	83,43	63,43

The percentages of one-second windows presenting significant synchrony clearly depend on the pair of features studied. For both correlation and coherence, the lowest percentage was obtained for pitch and pause. An explanation could be that those features only bear information during the voiced part and the spoken part of the video. Experimenter's Motion History Image was the feature that fitted best demonstrator's features. This feature accounts for a loose synchrony between features and is certainly less sensitive to the onsets of event.

### 4.3 Parameters of synchrony

Parameters of synchrony presented in Section 3.5 where assessed for all interactions. We focused on experimenter's adjustment to demonstrator's behavior. Table 2 presents the percentage of interaction's duration where experimenter perfectly matches to its partner's behavior (mutual synchrony) and where there is a time-delay between a change in the demonstrator's behavior leading to a change in his partner's. It focuses on the pair of feature Audio-Energy (D) / Motion-Energy (E). Time-lag denotes the averaged delay in seconds for adjustment from experimenter's to demonstrator. We only considered time-windows where correlation was significantly different from chance (z-test,  $p < 0.05$ ).

**Table 2. Parameters of synchrony (Audio-Energy (D) / Motion-Energy (E))**

	Mutual synchrony (in %)	Experimenter adjusts to his partner (in %)	Time-lag (in s)
1	66,1	29,0	0,454
2	55,0	35,7	0,681
3	55,5	32,6	0,640
4	53,7	33,5	0,715
5	54,6	35,1	0,625
6	44,2	33,4	0,770
7	60,7	29,3	0,509

These figures illustrate an aspect of interpersonal coordination : movement adaptation to vocal activity. This pair of feature characterizes how experimenter's reacts to his partner's instructions. It informs of experimenter's attention and involvement in the task. For instance, Dyad 1's experimenter appeared really attuned to his partner during the task and this dyad was the quickest to finish the assembly work. Using Audio Energy and Motion Energy only, we assessed a part of the adjustment between partners. Considering other pair of features, we could measure other aspect of this coordination process : convergence of speed, gesture adjustment...

## 5. DISCUSSION AND CONCLUSION

In the present experiment, we are investigating synchrony on a cooperative task where both partners have to coordinate in order to build the same assembly work. So our main question here is not “Is there operating synchrony between partners?”. We already know there is. But “What are the best features and measures to characterise this coordination process?”.

Examining synchrony at a local scale we found synchronous short-length windows for all pair of visual or audio features. These first results give credit to the idea that synchrony needs to be investigated at a local scale and not on the overall interaction. Moreover, it appears that it is not a unique modality that convey information on interactional synchrony and that both partners choose the appropriate modality to adapt to each other. We found a low percentage of windows showing significant synchrony for pitch and speech pauses. One can argue that our approach is only quantitative (counting a percentage of window) and that pitch and speech pauses are by definition not informative during the entire interaction. Thus, these features should be used in conjunction with other audio features like (energy) and should be looked at only on the voiced part (pitch) or silent part of the video. To characterise the visual field, Motion History Image was particularly adapted to our task as it is less sensitive to the time-lag between partners.

Another issue, dealing with more than two features, was to extract and present information on the link between all those features. Dendrograms give a good insight on the hierarchical similarity between features investigated to assess synchrony. The structure of dendrograms (for entire interactions) was close for all the interactions we investigated. Moreover the obtained structure was predictable knowing the kind of task that was investigated.

Coherence provided complementary information on synchrony. It is particularly useful for the type of task we are dealing with. Both partners accomplish the same actions with some undefined delay between them. While correlation requires to be computed for different time-lags, coherence assesses the shared frequency of actions independently of the time-lag between partners.

Future work will focus on examining the appropriate window size to evaluate local synchrony. For the present work, one-second length windows were arbitrarily chosen. A trade-off needs to be found between sufficient window duration to assess a reliable measure of synchrony and a small window to detect quick occurrence of synchronization. At last, this work addresses local synchronization from a quantitative point of view (what percentage of windows present significant correlation or coherence?). A qualitative study of interactional synchrony should also be addressed: how does synchrony evolve across time? When does arousal of synchrony occur and why? How do we choose the adequate modality to adjust? The answer to those complex questions should lead to a better understanding of social interactions and help in the design of intelligent interfaces.

## 6. REFERENCES

- [1] A. Bobick and J. Davis. Motion detection, labelling, data association and tracking in real-time on risc computer. In Proceedings of the 3rd Workshop on Applications of Computer Vision, 39-42.
- [2] S. Tilsen and K. Johnson. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, 124:2, EL34-39.
- [3] S. Yildirim and S. Narayanan. 2008. Automatic Detection of Disfluency Boundaries in Spontaneous Speech of Children Using Audio-Visual Information. In *IEEE Transactions on Speech, Audio and Language Processing*, 2-12.
- [4] G.R. Bradski. 1998. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, Q2 :15.
- [5] L. K. Miles, L. K. Nind and C. Neil Macrae. 2009. The rhythm of rapport: Interpersonal synchrony and social perception. In *Journal of Experimental Social Psychology*, 45:3, 585-589.
- [6] C. Trevarthen and S. Daniel. 2005. Disorganized rhythm and synchrony: Early signs of autism and Rett syndrome. In *Brain and Development*, 27:1, S25-S34.
- [7] M. J. Richardson, K. L. Marsh, R.W. Isenhour, J. R. L. Goodman and R.C. Schmidt. 2007. Rocking together: Dynamics of intentional and unintentional interpersonal coordination. In *Human Movement Science*, 26:6, 867-891.
- [8] T.L. Chartrand and J.A. Bargh. 1999. The chameleon effect: the perception-behavior link and social interaction. In *Journal of Personality and Social Psychology*, 76:6, 893-910.
- [9] D. Lakens. 2010. Movement synchrony and perceived entativity. In *Journal of Experimental Social Psychology*, in press.
- [10] F. J. Bernieri and R. Rosenthal. 1991. Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior*. 401-432
- [11] B. D. Lucas and T. Kanade. 1981. An iterative image registration technique with an application to stereo vision. In *Proceedings of Imaging understanding workshop*, 121-130.
- [12] M. Rolf, M. Hanheide, K. J. Rohlfing. 2009. Attention via synchrony: Making Use of Multimodal Cues in Social Learning. In *IEEE Transactions on Autonomous Mental Development*, 1:1, 55-67.
- [13] F. Ramseyer and W. Tschacher. 2006. Synchrony : A Core Concept for a Constructivist Approach to Psychotherapy. In *Constructivism in the Human Sciences* 11, 150-171.
- [14] J. Dauwels, F. Vialatte, T. Musha, A. Cichocki. 2010. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. In *NeuroImage*. 49. 668-693.
- [15] R. Feldman. 2003. Infant-mother and infant-father synchrony : the coregulation of positive arousal. In *Infant Mental Health Journal*, 24(1), 1-23.
- [16] M. Kipp. 2010. Multimedia Annotation, Querying and Analysis in ANVIL. In M. Maybury (ed.) *Multimedia Information Extraction MIT Press*, 19.