

# Modelling and testing the effects of a maternal mismatch of face and voice on 6-month-olds' interactions

Ken Prepin\*<sup>&\*\*\*</sup>, Maud Simon\*, Anne-Sophie Mahé, Arnaud Revel\*\*<sup>\*</sup>, & Jacqueline Nadel\*

\*UMR CNRS 7593, University P&M Curie, Paris \*\*UMR CNRS 8051, University of Cergy-Pontoise  
[prepin@ext.jussieu.fr](mailto:prepin@ext.jussieu.fr); [jnadel@ext.jussieu.fr](mailto:jnadel@ext.jussieu.fr)

## Abstract

*This paper emphasises the complementary of social interaction and multimodality and how they facilitate each other. In a conceptual model of the intermodal social interaction, agents are seen as dynamical systems which exchange energy via different channels. To make the interaction be the attractor of this two agents system, each agent must be both interamodal-synchrony and contingency dependent. A second model, of the social agent, aims to account for the previous conceptual model and to allow easy implementation on robotic systems. Dynamic Fields enable to benefit from both intermodal-synchrony and contingency. Two experiments with 6-month-old infants and mother test these two models. Results show that infants prefer a bimodal interaction to a unimodal one, but that they prefer a unimodal interaction rather than receiving two disconnected social sensory signals, one interactive and the other not. These results meet the prediction of our models and this shows that interaction is a main way to attribute coherence to sensory messages and that coherent sensory messages are a main way to convey social interaction messages.*

## Introduction

The recent goal to design robots that are able to cooperate with humans in every day life leads to equip them with social capabilities. Then two interdependent challenges arise: robots have to be accepted readily by humans and to be capable to select in their environment the information relevant to a social interaction with humans.

To be spontaneously accepted by humans, the robot should not monopolize the attention and competences of its interlocutor, the interaction should be intuitive, friendly and resistant to limited perturbations. A response emerges from Human Machine Interfaces (HMI) studies: "Multimodal exchange supports more transparent, flexible efficient and powerfully expressive means of human-machines interactions" (Oviatt, 2002). Intermodal IHM studies stress the robustness as well as the attractiveness of multimodal interactions. Multimodal interfaces are robust insofar as they enable adaptation to a continuously changing environment during visual exploration or shifts of attention (Oviatt 2000, Holzman, 1999).

Multimodal interfaces are attractive in as much as they allow the interacting agents to select and control over how they interact (Fell, et al. 1994, Karshmer&Blattner, 1998, Hauptmann, 1989, Oviatt 1997). If we intend to extend robots' capacities to natural interactions with naïve human agents, multimodality is imperatively needed. The reason why multimodal messages are so useful for interaction, we will claim, is because in a given individual each sensory channel shares with the others amodal parameters such as tempo. These amodal parameters are the ones that humans exchange during interactions. The person sends sensory messages that are in temporal harmony one with the other. When receiving multimodal messages from one source, the recipient perceives their temporal harmony and gives corresponding multimodal interactive responses. This means that not only are the sensory channels synchronized in an individual but that they are synchronized between individuals during multimodal interaction. The demonstration that humans have an intuitive and very early sensitivity to interactive multimodal synchrony is provided by developmental psychology in two series of research.

A first series of research focuses on social perception of intermodality. With classical experimental procedures using static displays and speech sounds, young infants have been shown to be sensitive to relationships between facial and vocal features of an adult (Bahrack, 2000). For instance, 4 month-olds can match a vowel sound with a facial pattern miming the vowel speech (Kuhl and Meltzoff, 1982). Shifting from static to dynamic displays, Walker-Andrews has presented two pre-recorded faces expressing different emotions accompanied by a voice matching the emotion of one of the two faces (Walker-Andrews, 1997). Infants around 6 months looked longer to the emotional face matching the voice, showing that they expect visual and auditory emotional signals to present a multimodal coherence.

A second series of research focuses on sensitivity to synchronic (i.e. contingent) interaction. Face-to-face interaction procedures with TV displays allow experimental manipulations of image and voice. One of these manipulations consists in presenting to the infant a non-contingent episode of mother's communication (i.e. replayed good sample of previous interaction) (Bigelow & DeCoste, 2003;

Muir & Hains, 1999; Murray & Trevarthen, 1985). Young infants of different ages including 2-month-olds (Nadel *et al.*, 1999; Nadel *et al.*, 2005; Soussignan *et al.*, 2006), detect a non-contingent communication of the mother: they are sensitive to the temporal delay between what they signal and their mother’s response.

What the two series of research teach us is that young infants detect and expect multimodal messages coming from a person, and that they detect and expect sensory signals that respond contingently to their own messages. We do not know much however about the link between the capacity to detect multimodality and the sensitivity to contingent interaction. The aim of this paper will be to investigate this question.

Considering the above mentioned results concerning human development, we will propose a model of intermodal interactions as a background for experiments with young infants. Note that our model does not account for the problem of attention focus during interaction. We will consider that the two agents interact face to face in a dyadic environment, which may solve or at least lessen the attentional problem.

## I. Conceptual model

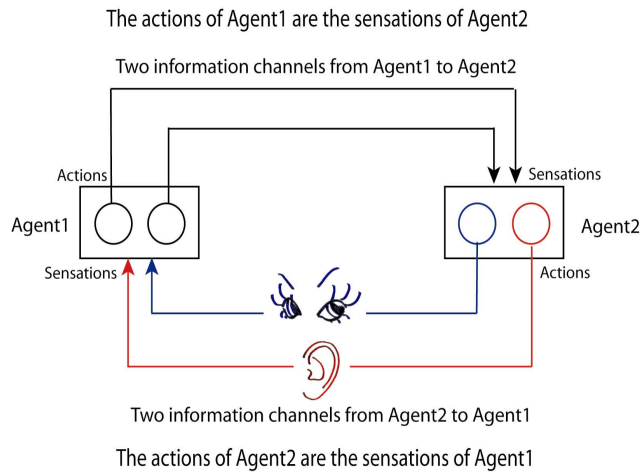


Figure 1: From a dynamical point of view, two multimodal social agents are two dynamical systems exchanging energy through different information channels in an interaction loop.

We propose that two agents, either robots or humans, that are interacting can be considered as two dynamic systems exchanging energy. The two agents exchange energy when visual, auditory or other sensory flows of information are outputs of one and inputs for the other (see figure 1).

Let us consider the dynamic system constituted of the two agents as a whole. Our model aims to make

the social interaction be the stable attractor of the “two agents” system. As the attractor of this system depends on each agent dynamic, we build our model from the agent point of view: the model of the agent is energetic and the lowest energy cost is when it multimodally interacts with another agent. If both agents respect this model then the *minimal loss* of energy of the whole system is when the *exchange* of energy between the two agents is *maximal*.

This “minimal loss-maximal exchange of energy” principle induces two assumptions for our model: the agent is both intermodal-synchrony (1) and contingency (2) dependent.

(1) When the agents interact using several modalities, they exchange energy through corresponding energy channels. As a multimodal entity, each system can receive simultaneously several sensory inputs coming from different channels, as well as deliver several outputs on different sensory modalities (see figure 1). The different sensory modalities may relate one to the other, or may not. If they are related, they do not only co-occur but also co-vary, insofar as they share amodal properties such as rhythm or speed: in short, they are synchronised. That is the case when the different modalities come from a unique agent: we talk about “intermodal-synchrony”.

(2) The second aspect of the processing of sensory inputs is their contingency, i.e. the fact that they are involved bilaterally in the interaction. If an input of a *system A* participates in the interaction, the information it contains is related to this *system A*. The input is involved in the closed loop that couples the two agents. In this case, from the *system A* point of view, the actions related to its inputs are closed to its previous actions: the energy consumption within the *system A* to decide of its actions is low and the exchange of energy between the two agents is high. Conversely, if the input of the *system A* does not participate in the interaction, actions related to its inputs are far from its previous actions: the energy used within the *system A* to decide of its actions is high, the energy exchange between the two agents is lessened. It must be noticed that this agent sensitivity and dependency to contingency has already been demonstrated in previews Double-Video and Still-face experiments (Nadel 1999, Tronick 1978).

The two assumptions of the model, contingency and intermodal-synchrony dependency of the agent, enable us to make predictions concerning interactive situations which may test this model:

- If the agent faces two contingent and intermodally-synchronous modalities, its internal energy cost will be minimal if it interacts. The attractor of the two-agents system dynamic is the social interaction which forces the agent to interact and synchronise with the incoming modalities.

- If *Agent1* faces only one contingent modality (other modalities are not available), the attractor of the whole “two agents” system is the interaction but this attractor may be weaker and less stable than when there are two contingent modalities: the energy exchange is less important with one modality than with two but the agent may interact and synchronise with the available incoming modality.
- If *Agent1* faces one contingent modality and one non contingent, the attractor of the “two agents” system induced by the *Agent1*'s dynamic is the interaction using the contingent modality. Yet, this attractor is not stable since its energy cost is high, the exchange of energy is poor (only one contingent modality) and the energy loss within the agent is high (non intermodally-synchronous modalities). In this case our model predicts that the agent should either interact and synchronise with the contingent modality or not manage to stabilise: the attractor of the whole system is not well determined.

Notice that in the trivial case where the *Agent1* faces two non contingent modalities, there is no energy exchange, the *Agent1* dynamic is not sufficient to build an attractor of the whole system dynamic, no interaction is possible.

We will use this model as a background to test the relationships between the early sensitivity to contingent interaction and the early capacity to link two sensory messages as coming from a multimodal agent. For that purpose, we have revisited the Double Video device previously designed to test contingent interaction so that it allows us now to present interactive or non-interactive samples of face and voice coming from the same person or coming from two different persons.

## II. Experiments

In our TV experiments, the exchange of energy is processed via two channels: the visual channel where facial and bodily motor outputs of one system are visual inputs for the other(s), and the auditory channel, where vocal and verbal outputs of one system are auditory inputs for the other(s).

### Experimental tests of model assumptions

The first assumption of our model of social interactions, contingency dependency, has already been tested (Nadel 1999). Contingency between partners is a prerequisite of the interaction.

The second assumption of our model, dependency to intermodal-synchrony, is the one which will be tested by our human development experiments. To test this assumption, we compared the infant's

response to a both contingent and intermodally-synchronous partner and to a bimodal mismatch of the sensory inputs (mother's interactive voice coupled to recorded mother's face). We also compared interaction with mother voice in live only to multimodal live interaction and to a bimodal mismatch of the sensory inputs.

### Population

Thirty-four volunteer dyads of French mothers and their 6-month-olds participated in the study. Eighteen dyads were involved in the first experiment, and sixteen in the second study. All mothers have had uneventful pregnancies followed by normal deliveries. The infants were all full-term. They were all aged 6 to 7 months (mean age 6M 2W).

The mothers had previously given their written informed consent concerning the experimental procedure. This means that they knew from start that the infants will not always be responding to the live of their facial and vocal communicative display.

### Experimental design

The experimental design was the upgraded device already described in Nadel et al. (1999). As depicted in Figure 2, three independent rooms in a quiet part of our lab in the hospital were equipped so as to allow to manipulate independently what mother and infant hear and see.

The infant's room was equipped with a baby seat facing a two-way mirror at a 50cms distance. The mirror reflected the mother's image from a TV monitor, obscuring the video camera that recorded the infant, at eye-level. The mother's room was equipped with an arm-chair and a similar set-up.

The recording room was equipped with four video recorders (VR B, VR M, VR&D M, VR BS), a mixer and various amplifiers and control TV screens. VR B recorded the infant while VR M recorded the mother. VR&D\_M, Video Recorder and Diffuser of the Mother, is a computer able to record and diffuse at will the mother's face and voice. VR BS, Video- Recorder of the Baby Screen, video-recorded the manipulated video that the baby sees (live, replay or mismatch).

As depicted on Figure 5, sound and image were combined in two different ways on the mixing table : contingent, with live sound and live image, or partially non-contingent, with live sound but replayed image from VR&DM computer.

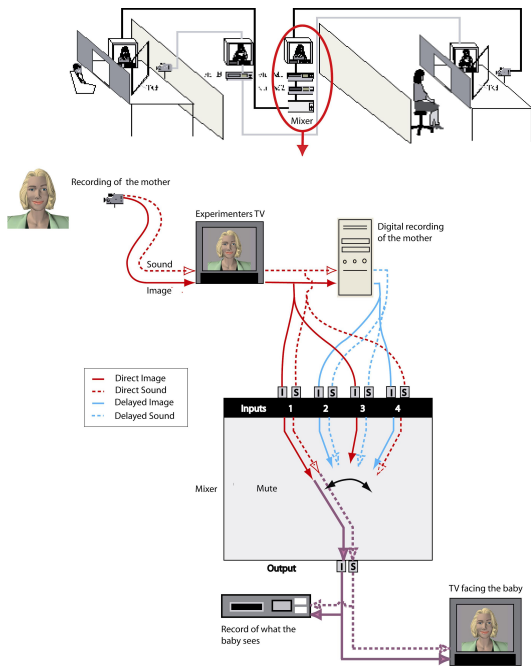


Figure 2: Experimental design with the 3 independent room and a zoom on the mixer. The mixer allows us first a seamless shift from live to recorded video, and second a decoupling of visual and auditory channels.

### Experimental procedure

The first experiment consisted of three uninterrupted 30-second face-to-face TV interaction episodes. During the first and third episodes, infants could see and hear their interactive mother. The second episode was manipulated in order to present to the infant the bimodal by-product of a mismatch between what she hears and what she sees: infants could hear their interactive mother but saw her recorded face. Table 1 summarizes the sequences presented according to this experiment

	Episode 1 30 sec	Episode 2 30 sec	Episode 3 30 sec
Exp 1	Mother's face + voice interactive	Mother's voice interactive + face recorded	Mother's face + voice interactive

Table 1- Episodes according to the experiments

In the second experiment, infants could hear their interactive mother but did not see her face.

### Data recording

In the recording room (depicted in Figure 2), the engineer had two tasks to fulfill. He first continuously records the mother (on VR M), the

baby (on VR B) and what the baby sees (VR BS) during all three conditions. A LED visual signal indicates the start of the two records. The engineer's second task is to record (on VR&D M) a 30 seconds communicative display of the mother, as soon as a good contact between the mother and infant is established. The criterion for a good contact was the mother talking to her infant and eye-to-eye contact between infant and mother. A judge present in the room signals when a good contact begins and immediately the engineer starts recording the exchange. The VR&D M, using a computer to manipulate video, allows the engineer to use the record immediately. As soon as the first episode of live interaction is 30 seconds long, the engineer switches and presents the coupling of live mother's (or stranger's) voice and recorded mother's (or stranger's) image on the infant's screen. After 30seconds of this mixed period the switch is pressed again and a bimodal interactive episode of mother (or stranger)'s communication is presented to the infant.

This technique of data recording allows us to offer to infant a continuous video presentation of the partner's behavior under both bimodal and decoupled sensory conditions.

### Coding system

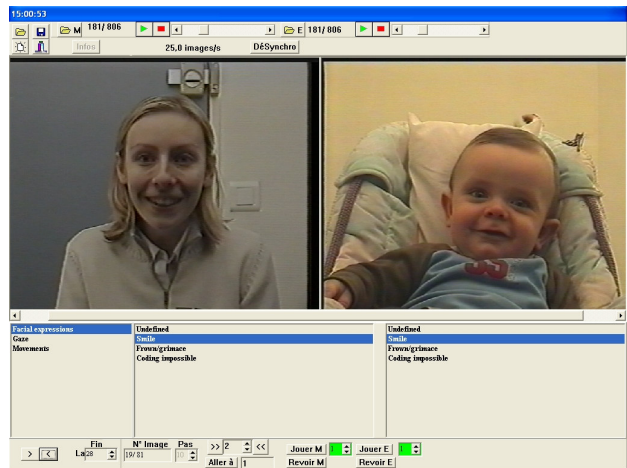


Figure 3: Coding software.

Video-computer interfacing software already created and used successfully in precedent studies (Kervella & Nadel, 1999) allows us to present simultaneously on the monitor the infant's and the mother's digitized single frames along with the two coding grids (see fig.3 and table 2).

The coding grid was composed so as to allow the coder to depict the frame for each category of behaviour by an item and an item only (Nadel et al., 1999).

Categories	Items
Facial expressions	Neutral Smile Frown/grimace Other
Gaze	Gaze at the person Gaze toward voice (speakers) Gaze away Other Impossible to code
Movements	Self-centred movements Relaxed movements Other

Table 2- Coding grid

The frames were synchronized according to the LED visual signal. For each 400/1000 of a second, a stable frame was automatically presented to the coder who chose for each category the item corresponding to the behaviour. As illustrated in Figure 7, for the category of “facial expression” both mother and her infant were judged to smile.

### Dependent measures

For the three experiments, the Analyses of Variances (ANOVAs) were conducted with four dependent variables: Gaze to mother, Smile, Frown/Grimace, Self-centred movements. Self-centred movements are indices of stress in young infants.

Two independent coders coded 25% of the frames with a Kappa agreement of .90 for gaze, .95 for self-centred movements, .84 for frowning and .80 for smile.

### Results

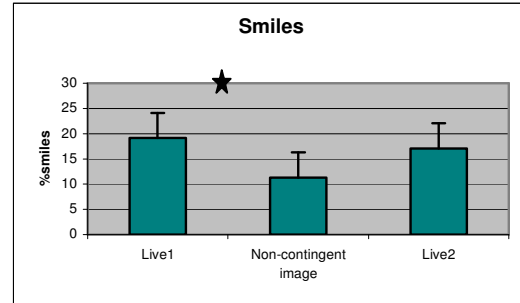
We compared the effects of a bimodal social environment where the two modalities are interactive (exp 1, episode 1) to a bimodal social environment where one modality only is interactive (exp 1, episode 2), and to a unimodal social environment where the unique modality in play was interactive (exp 2).

#### 1. Bimodal social environment: two versus one modality interacting.

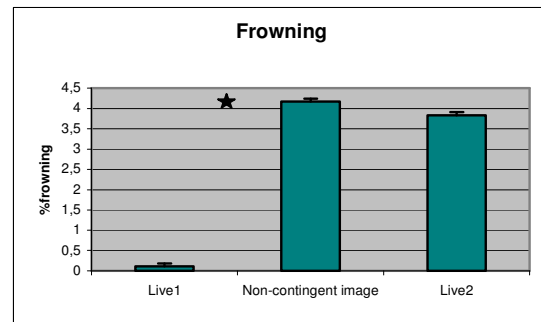
Analyses of variance comparing the positive and negative responses of infants in episodes 1 and 2 of experiment 1 all showed that infants were upset when their mothers appeared to present a mismatch between her interactive voice and her recorded face, compared to when their mother’s face and

voice were interactive: smile decreased [ $F(2, 34) = 4.5, p = .02$ ], self-centred movements increased [ $F(2, 34) = 4.0, p = .04$ ], and frowning increased [ $F(2, 34) = 5.2, p = .02$ ]. Obviously, they were able to detect that the two sensory signals were not co-varying during episode 2. The interaction was disrupted as an effect of the detection of bimodal mismatch (see figure 4,a,b,c).

a)



b)



c)

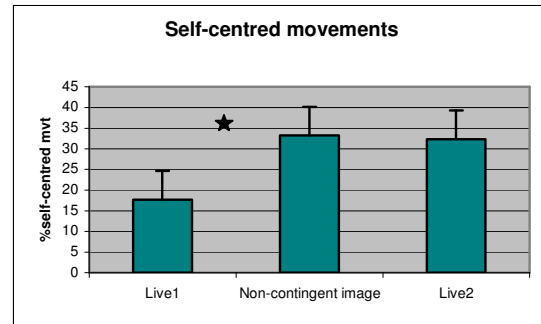


Figure 4- Bimodal mismatch generates the decrease of positive responses (a) and the increase of negative responses (b,c) to the mother

#### 2. Bimodal social interaction versus unimodal social interaction.

Comparing with ANOVAS the positive and negative responses of infants in episode 1 of experiment 1 and in experiment 2, we found a

significant difference for frowning/grimacing [ $F(1,32)=7.09$ ,  $p=.01$ ] which accounts for the infants being less positively engaged in the unimodal condition (fig 5).

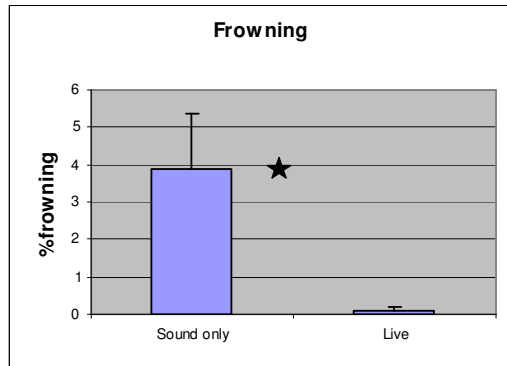


Figure 5- Interaction in an unimodal social environment is less rewarding than interaction in a bimodal social environment where the two modalities are interactive

### 3. One modality interacting in a bimodal vs. unimodal social environment

When we compared the effect of an interaction with one modality only according to the fact that the social environment was unimodal or bimodal, the analyses of variance showed a significant difference in self-centered movements in the bimodal environment [ $F(1,32)=4.93$ ,  $p=0.03$ ]. Infants were thus more stressed in the condition of a mismatch between face and voice with voice interacting only than in the condition where voice interacts and face is not visible (fig 6).

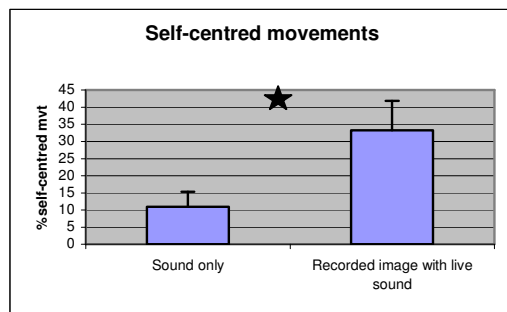


Figure 6- Only one modality interacting is less disturbing in an unimodal social environment than in a bimodal social environment

Our experimentations have shown two things which validate our conceptual model of the social interaction. The first is that the infant is able to interact using only one modality even if this interaction is not as easy as when she can use several modalities. The second thing is that when a modality is hetero-synchronous, i.e. it is used by both agents to interact together, if a second

modality, asynchronous with the first, is added, then the interaction cannot go on. It seems it generates instability in the infant.

### III. A model that exploits both multimodality and social interaction

In section I we proposed a coarse conceptual model focused on social interactions. To be much more realistic and to allow easy implementation on robotic systems, we detail here, a model of the social agent which may be involved in these interactions. According to our conceptual model, social interactions consist in the coupling of agents which are dynamic systems exchanging energy. To fit this dynamical model, we referred to “a general class of models that express behavioral decisions in the dynamics of continuously evolving activation fields” (Schöner, Thelen 2006, p.277), the Dynamic Field Theories (Amari 1977). (a) the variables are represented in a continuous metric space; all signals, within the agent, are assigned frequency variables which are continuous, metric variables that indicate their relative distance from one another. Additionally, each of these variables can assume different activations strengths, with an understanding that only variables that reach a particular threshold enter into behavioral decision. Moreover, within a field, interaction between different activation variables can also be specified. (b) the decision process evolves gradually over time: the field models assume that perceptual variables have dynamics, that is, they evolve gradually over time according to some dynamic regime, which may be nonlinear. This assumption of continuous time means that the preceding and the current states of the system are critical: the nearest of the current action are the action decision, the stronger they are into the behavioral decision.

Our social agent is made of three constitutive parts. The first part is the “features extraction” in which the different flows of information are processed so as to extract salient features. The features we are interested in are the frequency and the phase which can be extracted from the signal (see the bottom of fig.7, *Extraction of frequency spectri*). This extraction leads to a dynamic field, the *Input field* (center of fig.7), which combines the input coming from each modality.

The second constitutive part of our model and the most important one is the *Motor planning field* (see the 3D diagram, fig.7). That is the part which implements the “minimal loss-maximal exchange of energy principle”. The *Motor planning field* is a dynamic field as described by Amari which receives two inputs, the *Memory input* and the *Input field* (fig.7 upper part). The *Memory input* is a dynamic field which reintroduces in the *Motor planning field* the previous motor decision.

The third part of our model is the motor control which processes the *Motor planning field* outputs so as to perform the action .

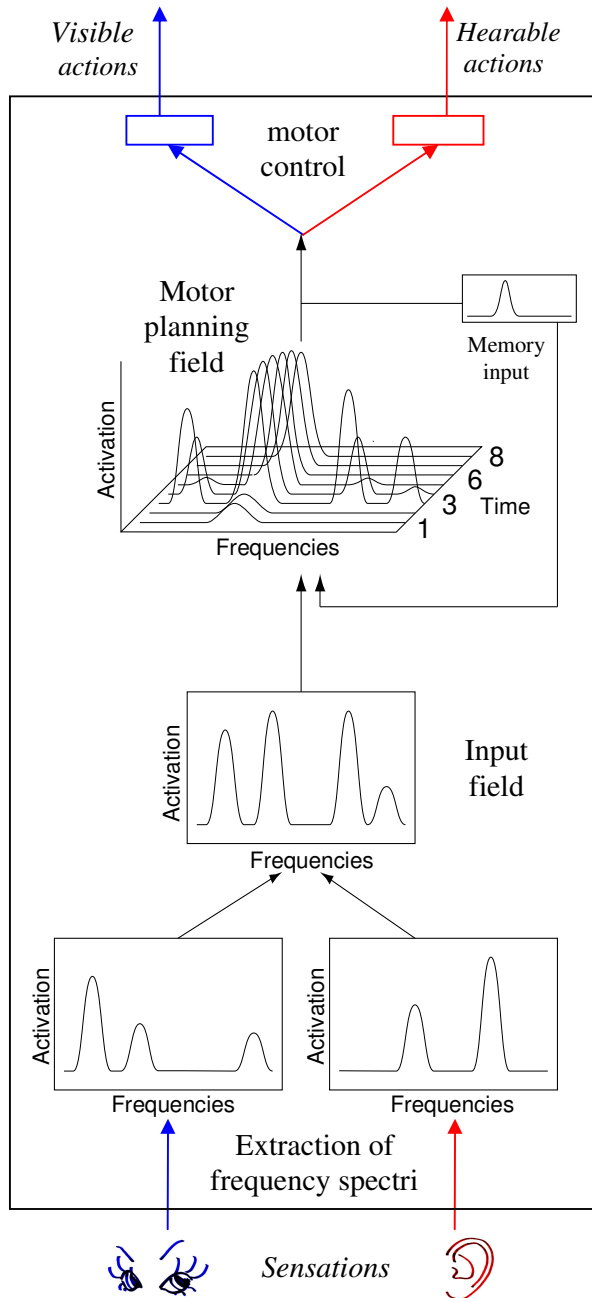


Figure 7- Dynamical model of a bimodal social agent: the “Motor planning field” makes the agent depend to both contingency and inter-modal synchrony.

The dynamic properties of our system make it fits the two assumptions of the conceptual model and thus feed our “minimal loss-maximal exchange of energy” principle:

- On one hand the *Motor planning field* emphasises the contingent energy flow. The agent’s actions are shaped by what it has most recently done. Either that consumes energy if its input are not

related to its recent action. Or that saves energy if the agent’s inputs are related to its recent actions, that means if its inputs come from a socially interactive second agent. The agent is contingency dependent.

- On the other hand, our model of the social agent best processes intermodally synchronous flows. At the level of the *Input field* the combination of the stimuli coming from different modalities may emphasises the properties shared by the different modalities and attenuate the modality-specific properties. Thus in the case of intermodally coherent stimuli, the *Input field* should emphasise specifically properties such as rhythm or synchrony. These properties are the crucial features of the interaction, the ones which enable to define a distance between the current action and the previous ones. The more salient these interaction features are, the more the agent can enter in interaction with its contingent partner. The agent abilities to interact is intermodal-synchrony dependent.

## Discussion

Do this latter model of the social agent accounts for our experimental results?

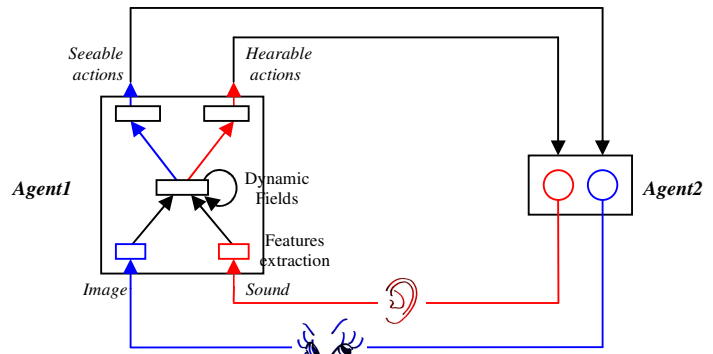


Figure 8- *Agent1* and *Agent2* are interacting: Multimodal stimuli accentuate interaction features detection and thus contingency dependency. The attractor of this two agents system is the interaction.

Let us consider our model of the social agent embedded in the context of an interaction (fig.8). The prediction of our model are that the different interactive conditions will be ranged in the following order:

- The condition favouring at best the sensitivity to interaction would be when two modalities are coherent and both interactive. The interaction features are emphasized by the *Input field* and at the level of the *Motor planning field* there is no loss since the action decision is near (in a spatial dynamic field representation) the ongoing action.
- The medium condition would be when only one modality is available to performan interaction. The energetic cost at the level of the *Motor*

*planning field* is not null: the *Input field* does not emphasise interaction features and lessen other features, thus the *Motor planning field* has to select within the signal the features of the interaction. The energy cost is higher.

- The worse condition would be when two modalities are available in the social environment but only one is involved in the interaction. In this case the energy cost at the level of the *Motor control field* is maximal. Both because the interaction features are not specifically selected by the *Input field* and because the action decision is perturbed by the second modality.

Our experimental results meet the prediction of our interrelated models of the social interaction and the social agent. Indeed, taken together, our results suggest a hierarchy of disturbances linked to the interaction between modality and contingency. The most rewarding condition is when the two modalities provided by the social environment respond contingently to the infant's behaviour, and the most disturbing is when only one of two modalities responds contingently to the infant's behaviour. To sum up, one contingent modality out of 2 < one modality that is contingent < two contingent modalities.

We are aware of the fact that a lot more has to be done before we can more precisely describe the process by which bimodality within one system generates the conditions to detect and achieve contingency between two systems, a step toward an understanding of intersubjective matching.

*Acknowledgements- The experiments were part of the European Project ADAPT IST- 2001-37173*

## References

- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. In *Biological Cybernetics*, Vol. 27:77—87.
- Bahrick, L. (2000). Increasing specificity in the development of intermodal perception. In D. Muir & A. Slater (Eds.), *Infant development: the essential readings* (pp.119-136). Oxford: Blackwell.
- Bigelow, A.E., & DeCoste, C. (2003). Sensitivity to social contingency from mothers and strangers in 2-4-, and 6—month-old infants. *Infancy*, 4 (1), 111-140.
- Fell, H., Delta, H., Peterson, R., Ferrier, L, Mooraj, Z., & Valteau, M. (1994). Using the babybabble-blanket for infants with motor problems. *Proceedings of the Conference on Assistive Technologies (ASSETS'94)*, 77-84. Marina del Rey, CA.
- Hains, S., & Muir, D. (1996). Effects of stimulus contingency in infant-adult interactions. *Infant Behavior and Development*, 19, 1, 49-61.
- Hauptmann, A. G. (1989). Speech and gestures for graphic image manipulation. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'89)*, Vol. 1 241-245. New-York: ACM Press.
- Holzman, T. G. (1999). Computer-human interface solutions for emergency medical care. *Interactions*, 6(3), 13-24.
- Kuhl, P.K., & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138-1141.
- Muir, D. (2004). Infant perception and production of emotions during face-to-face interactions with live and 'virtual' adults. In J. Nadel & D. Muir (Eds.), *Emotional development* (pp.207-233). Oxford: Oxford University Press.
- Muir, D.W., & Hains, S. (1999). Young infants' perception of adult intentionality: Adult contingency and eye direction. In P. Rochat (Ed.) *Early social cognition: Understanding others in the first months of life* (pp. 155-184). Mahwah, NJ: Erlbaum.
- Muir, D.W., & Nadel, J. (1998). Infant social perception. In A. Slater (Ed.), *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 247-285). London: Psychology Press.
- Murray, L., & Trevarthen, C. (1985). Emotional regulation of interactions between two-month-olds and their mothers. In T. M. Field & N. Fox (Eds.), *Social perception in infants* (pp. 101-125). Norwood, NJ: Ablex.
- Nadel, J. (2002). When do infants expect? *Infant Behavior and Development*, 25, 30-33.
- Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, 2, 2, 164-173.
- Nadel, J., Soussignan, R., Canet, P., Libert, G., & Gérardin P. (2005). Two-month-old infants of depressed mothers show mild, delayed and persistent change in emotional state after non-contingent interaction. *Infant Behavior & Development*, 28, 418-425.
- Oviatt, S. (2002). Multimodal Interfaces. In J. Jacko & A. Sears (Eds.), *Handbook of Human-Computer Interaction*, IV.A.14. Lawrence Erlbaum: New Jersey.
- Oviatt, S. L. (2000c). Multimodal signal processing in naturalistic noisy environments. In B. Yuan, T. Huang & X. Tang (Eds.), *Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000)*, Vol. 2, (pp. 696-699). Beijing, China: Chinese Friendship Publishers.
- Oviatt, S. L. (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction [Special issue on Multimodal Interfaces]*, 12, 93-129.

Schöner, G., Thelen, E. (2006). Using Dynamic Field Theory to Rethink Infant Habituation. In *Psychological Review* 2006, Vol. 113, No. 2, 273–299

Soussignan, R., Nadel, J., Canet, P., & Gérardin, P. (2006). Sensitivity to social contingency and positive emotion in 2-month-olds. *Infancy*.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of American Academy of Child Psychiatry*, 17, 1-13.

Walker-Andrews, A. (1997). Infants' perception of expressive behaviors: differentiation of multimodal information. *Psychological Bulletin*, 121, 437-456.