2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust

# Exploring Multimodal Social-Emotional Behaviors in Autism Spectrum Disorders

## An interface between social signal processing and psychopathology

Laurence Chaby[1,2,3], Mohamed Chetouani[1], Monique Plaza[1,3], David Cohen[1,3]

[1]Institute of Intelligent Systems and Robotics, ISIR, CNRS UMR 7222, Paris, France
[2]University Paris Descartes, Sorbonne Paris City, Paris, France
[3]Department of Child and Adolescent Psychiatry, Hôpital de la Pitié-Salpêtrière, Paris, France
*laurence.chaby@parisdescartes.fr, mohamed.chetouani@upmc.fr, {monique.plaza, david.cohen}@psl.aphp.fr*

*Abstract*—**The purpose of this paper is to present our original and multidisciplinary approach to study multimodal social-emotional behaviors in children with autism spectrum disorders. Our goal is to conduct fundamental and applied research regarding the reception and production of social signals involved in human interactions. To achieve this aim, we try to understand and model cognitive and multimodal emotional integration (e.g., auditory, visual, postural) during infancy and to analyze dysfunctions in pathologies that affect the dynamics of social interactions such as autism spectrum disorders. More specifically, we study the characterization of multimodal social-emotional signals (speech, prosody, faces, postures) and the dynamics of communication (e.g., synchrony, engagement). The fields of application covered are the improvement of differential diagnosis, interactive robotics, assisting people with autism spectrum disorders and their caregivers, and objectification in psychopathology.**

**Keywords: social signal processing; social cognition; multimodal processing; emotion; prosody; social interaction; interpersonal synchrony; autism spectrum disorders; child development**

## I. Introduction

Multimodal social-emotional interactions play a critical role in child development, and this role is emphasized in autism spectrum disorders. In typically developing children, the ability to correctly *identify*, *interpret* and *produce* social behaviors (Figure 1) is a key aspect for communication and is the basis of social cognition [1]. This process helps children to understand that other people have intentions, thoughts, and emotions, and act as triggers of empathy [2]. Social cognition includes the child's ability to spontaneously and correctly interpret verbal and nonverbal social and emotional cues (e.g., speech, facial and vocal expressions, posture and body movements, etc.); the ability to produce social and emotional informations (e.g. initiating social contact or conversation); the ability to continuously adjust and synchronize behavior to others (i.e., parent, caregivers, peers); and the ability to make an adequate attribution about other's mental state (i.e., "theory of mind") [3]. Autism spectrum disorder (ASD) is a group of behaviorally defined disorders with abnormalities or impaired development in three areas: verbal and nonverbal communication, social interaction, and rigid patterns of behavior, present in early childhood [4]. Several subtypes have been defined including autistic disorder (AD), pervasive developmental disorder not otherwise specified (PDD-NOS), Asperger syndrome and Rett syndrome. In this paper we focus more specifically on three domains of impairments: i) language, ii) emotion and iii) synchrony in social interactions.

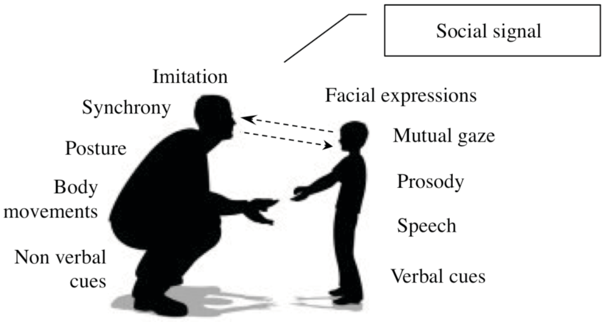

Figure 1. Reception and production of social signals. Multimodal verbal (speech, prosody) and non-verbal cues (facial expression, vocal expressions, mutual gaze, posture, imitation, synchrony, etc.) merge to produce social signal.

### A. Language

Language impairment is a common feature in autism spectrum disorders that is characterized by a core pragmatic disorder, abnormal prosody and impairments regarding semantics skills [5]. However, language functioning in ASD is variable. At one end, there are children with ASD whose vocabulary, grammatical knowledge, pragmatics, and prosody skills are within the normal range of functioning (e.g. Asperger syndrome), while at the other end a significant

978-0-7695-4848-7/12 $26.00 © 2012 IEEE
DOI 10.1109/SocialCom-PASSAT.2012.111

proportion of the population remains essentially non-verbal (e.g. AD with intellectual disability). In a recent clinical work [6], we tried to find differential language markers of pathology in autistic disorder without intellectual disability (AD), pervasive developmental disorder not otherwise specified (PDD-NOS) compared to specific language impairment (SLI) and to typically developing children (TD). Our findings suggest that expressive syntax, pragmatic skills and some intonation features could be considered as language differential markers of pathology. The AD group is the most deficient, presenting difficulties at the lexical, syntactic, pragmatic and prosodic levels; the PDD-NOS group performed better than AD in pragmatic and prosodic skills but was still impaired in lexical and syntactic skills.

### *B. Emotion*

Interpersonal communication involves the processing of multimodal emotional cues, which could be perceived and expressed through visual, auditory and bodily modalities. Autism spectrum disorder is characterized by problems in recognizing emotions that affect day-to-day life [7]. Research on emotion recognition abilities in ASD has been limited by over-focus to the visual modality, specifically the recognition of facial expressions. In addition, emotion production remains a neglected area. However, understanding emotional states in real life involves identifying, interpreting and producing a variety of cues that include non-verbal vocalizations (e.g. laughter, crying), speech prosody, body movements and posture. In a preliminary work [8], we recently studied neutral and emotional (facial, vocal) processing in children with pervasive developmental disorder not otherwise specified (PDD-NOS) that represent around two-thirds of autism spectrum disorders. Our results suggest that children with PDD-NOS present global emotional human stimuli processing difficulties (in both unimodal facial and vocal conditions), which dramatically contrast with their ability to process neutral human stimuli. However impairments in unimodal emotional processing are partially compensated using multimodal processing. Nevertheless, it is still not yet clear how children with ASD perceive and produce multimodal emotion depending of ASD subtypes (i.e., autism, PDD-NOS, high functioning autism, etc.) and stimulus domains (e.g. visual, auditory, etc.).

### *C. Synchrony in social interaction*

Synchrony in social interaction is a complex phenomenon that requires the perception and production of social and communicative signals (speech, linguistic cues, prosody, emotion, gesture, etc.) and also a continuous adaptation to other. In adulthood, interactional synchrony has been shown to act as a facilitator to high quality interpersonal relationships and smooths social interactions [9]. The role of synchrony during child development is not well known, but seems to provide the children a secure base from which they can explore their environment, regulate their affective states, and develop language and cognitive skills [10].

## II. Background And Objectives

### *A. Background*

In France, autism and autistic spectrum disorders (ASD) have become a national priority (see "national autism plan"). One child out of 150 suffers from a trouble belonging to ASD with a trend of an increasing prevalence. ASD prevalence was estimated about 20/1000 at the beginning of 2000 [11]. Although there have been many important advances in understanding autism spectrum disorders over the past twenty years, it still remains a serious disabling condition. Thus, recent studies have pointed out that early diagnosis, and early and intensive interventions were key issues of outcome [12]. Considering the importance of earlier diagnosis, interests have been focusing on elementary tools, which can be used by doctors and/or general health personal to identify children at risk of having social-communication disorders. The only instrument which succeeded in validation studies at large scale was the Checklist for Autism in Toddler (CHAT, [13]). However, according to the prospective studies, CHAT seemed to be quite specific but not enough sensitive to promote early identification of autism.

### *B. Objectives*

The study of social-emotional interactions in early pathological development is crucial, but currently, no commonly accepted method exists for detecting and assessing multimodal (verbal and nonverbal) behaviors or synchrony between interactive partners. A need has arisen for non-invasive tools for assessing and quantifying early-emerging developmental abnormalities. Social Signal Processing (SSP) [14] is an emerging research and technological domain that aims at providing computers with the ability to understand human social signals. SSP can also help to address some of the issues related to the study of early interaction. SSP can be used for several purposes such as modeling, assessing synchrony between partners and characterizing specific cues that participate to interpersonal exchanges. SSP may also be of interest for developing specific tools with human-like abilities to stimulate social behaviors in a controlled context.

The current multidisciplinary project is at the intersection of developmental psychology, psychopathology, social signal processing and computational neurosciences. It aims to develop engineering tools to detect and assess language and prosodic skills, multimodal emotional processing, social interactive synchrony during early pathological development.

## III. Multimodal Social-Emotional Paradigms

### *A. Recruitement and clinical evalution of participants*

We started to recruit children aged 6-12 years with oral language who meet criteria for Autism Spectrum Disorder (ASD); all of them were diagnosed and followed-up at the Pitié-Salpétrière Hospital (Paris, France). A control group, matched in developmental age and gender is composed by typically developing children.

This work was supported in part by the Fondation de France, the UPMC "EMERGENCE 2009" program, the French ANR grant SAMENTA "SYNED-PSY" and the European Project "FP7 MICHELANGELO".

All the children are administered the Autism Diagnostic Interview-Revised (ADI-R [15]) and the Autism Diagnostic Observation Schedule (ADOS [16]) to document the diagnosis of ASD. Severity is assessed with the Children Autism Rating Scale (CARS) [17]. In addition an expert clinician observe all the children to confirm that they meet DSM-IV criteria for ASD and define the clinical subtype.

All children are also assessed for cognitive level (WISC3/WPPSI, WISC4), oral language (ELO Battery [18]), motor skills/praxis, and basic visual/auditory functions.

### B. *Experimental design*

Collecting multi-modal data during clinical situations is challenging. A dedicated platform (Figure 2) is designed for the analysis of socio-emotional behavior and interactive abilities of children. The interactions are recorded using several cameras (including Microsoft Kinect) and sensors (e.g., Tobii T120 eye-tracker, omnidirectional microphones). One of the main objectives is to be able to collect social signals such as facial expressions, body movements, postures and gestures, speech behaviors and visual search strategies.

Gaze behavior is measured at 120 Hz via an integrated T120 eye tracker (Tobii Systems). The eye-tracking device is built into the screen and did not require fixing children's heads. The device tracks both eyes separately using corneal reflection. Audio recordings are collected at 48 kHz and the video at 25 or 30 fps depending on the used camera. Multimodal data are annotated with ANVIL [19].

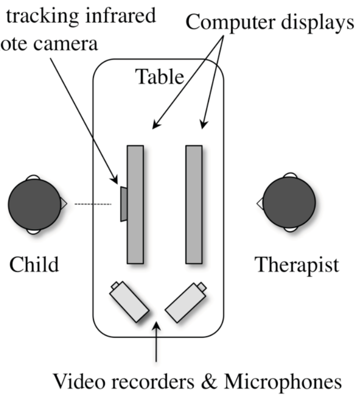
Eye tracking infrared remote camera
Computer displays
Table
Child
Therapist
Video recorders & Microphones

Figure 2. Layout of the experimental-clinical room, showing the location of the children (and if necessary the therapist) and the placement of the devices (computer displays, Tobii T120 eye-tracker, video recorders and microphones).

### C. *Socio-emotional based protocols*

Three socio-emotional dimensions are investigated through the data collected with the platform: language and prosodic skills, multimodal emotional processing and social interactive synchrony.

#### 1) *Language and prosodic skills:*

In studying language and prosodic skills we are interested in re-examining language profiles of children with autism disorders or pervasive developmental disorder not otherwise specified. Data collection is performed through two tasks: intonation imitation and affective speech production.

The 'imitation' task was initially developed to compare children's abilities to reproduce different types of intonation contours [20]. To avoid cognitive demand, the sentences were phonetically easy and relatively short. According to French prosody, 26 sentences representing different modalities (e.g., declarative, exclamatory, imperative) and four types of intonations (e.g., descending, falling, floating and rising) were defined for the imitation task (e.g., "I'm happy", "Anna will come with you"). Children were asked to repeat exactly the sentences they just heard even if they did not catch one or several words. After sentences segmentation and rejection of all disrupted sentences (e.g., noise, repetition, false-start), 2772 sentences equivalent to 1 hour of speech in total from thirty-five monolingual French-speaking subjects aged 6 to 18 years old were kept for analysis up to now (see section IV.A).

The 'production' task refers to a less constraint situation designed for investigation on emotional speech production in a more naturalistic setting. The task was based on a story telling of a pictured book ("Frog where are you?" [21]), wherein a little boy tries to find his escaped frog during the night (Figure 3a). The child is supposed to produce prosodic cues during the story telling that are correlated to the levels of the emotional valence in each picture. We categorized each image by its emotional valence in three categories: negative/neutral/positive. In total, the pictured book included 7 emotionally negative, 12 emotionally neutral and 5 emotionally positive pictures. Twenty children have already performed this task where behavioral performance, gaze behavior, and audio and videos data were recorded and are being analyzed.

#### 2) *Multimodal emotional processing:*

Many studies of emotional processing focus strictly on the visual or auditory modality; this means that far less is known about the processing of multimodal emotional information in both reception and production. In studying multimodal emotional processing we are interested in understanding factors mediating changes during unimodal and audiovisual integration of in children with ASD.

To characterize emotional versus non-emotional processing in children, we contrasted neutral with emotional human stimuli (e.g., happy, angry, sad and neutral) across different perceptual modalities (visual, auditory and multimodal). In an initial paradigm (see [8]) we required children to recognize targeted faces, first names and face/first name pairs among distracters, while recording behavioral performance. In a second paradigm (see layout Figure 2), we required children to recognize facial emotion, vocal emotion and congruent facial and vocal emotion (Figure 3b). A multiple forced choice paradigm was used with mouse clicks on one of the 6 labeled buttons what appear at the bottom of the computer screen; accuracies, eye-tracking data and child videos were recorded and are being analyzed.

### 3) *Social interactive synchrony:*

The lack of automatic tools for studying synchrony has limited the exploration of interactive abilities in autism spectrum disorders. In studying synchrony we were interested in a dynamical and clinical features from verbal and nonverbal exchange, attesting specific early relationship (Figure 3c). Two situations are proposed to investigate social coordination: face-to-face and computer-mediated communications.

The first situation aims at investigating the notion of coordination between dyadic partners: children-therapist. The task consisted in building a clown with 7 polystyrene elements (2 hands, 2 legs, body, head and hat). The child sat across from the therapist. The same polystyrene elements were arranged on a table in front of them. The children were asked to perform three different tasks: the "Imitation" task, the "Child Follows Instructions" task and the "Child Gives Instructions" task. Tasks were of increased level of difficulty regarding communication but realization per se in terms of motor and cognitive abilities was easy. The main differences between the tasks are 1) the perception of the actions of the partner and 2) the leader of the action (see Figure 4).

The computer-mediated communication exploits the platform described in Figure 2. A similar concept is proposed, the children and therapist had to coordinate for the realization of a puzzle. The main difference for the computer-mediated situation is that we employed a serious game: electronic puzzle (see Figure 3c). In addition to audio-visual data, we also collect mouse movements (e.g., manipulation of virtual objects) and visual strategies through the eye tracker. Twenty children have already performed the first face-to-face situation and initial analyses have been realized (see section IV.B).

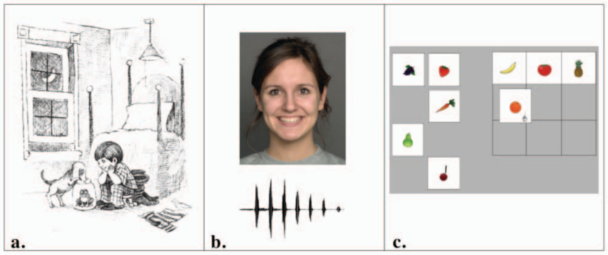

Figure 3. Example of stimuli used in the three modules: a. example of a picture used in the linguistic module *("Frog, where are you?" [21])*; b. example of facial expression (*from FACES [22]*) and vocal expression (*from "Emotional Affectives Voices [23]*) used in the emotional module; c. example of a puzzle used in the social coordination protocol.

## IV. Initial Analyses & Future work

### A. *Automatic intonation recognition and prosodic skills.*

In [20], we designed a system that automatically assesses a child's grammatical prosodic skills through an intonation contours imitation task. The key idea of the system is to propose computational modeling of prosody by employing static (k-nn) and dynamic (HMM) classifiers. The intonation recognition scores of typically developing (TD) children and language-impaired children (LIC) are compared. We also compared duration of the reproduced sentences in the two groups. The results showed that all LIC have difficulties in reproducing intonation contours because: i) sentence duration was lengthened by around 30 % compared to TD children; ii) they achieved significantly lower recognition scores than TD children on almost all studied intonations (respectively 56% and 70% for LIC and TD; p <.05; see Figure 4).

| Intonation | TD | AD | PDD-NOS | SLI |
|---|---|---|---|---|
| Descending | 64 | 64 | 70 | 63 |
| Falling | 55 | $35^{*T}$ | $45^{*T}$ | $39^{*T}$ |
| Floating | 72 | $48^{*T}$ | $40^{*T}$ | $31^{*T}$ |
| Rising | **95** | $57^{*T,S}$ | $48^{*T,S}$ | **81**$^{*T,A,P}$ |
| All | 70 | $56^{*T}$ | $53^{*T}$ | $58^{*T}$ |

Figure 4. Fusion intonation recognition performances: performances are given as percentage of recognition (%); * = p<.05: alternative hypothesis is true when comparing data from child groups, i.e., T, A, P and S; TD (T): typically developing; ASD (A): autism spectrum disorder; PDD-NOS (P): pervasive developmental disorders not-otherwise specified; SLI (S): specific language impairment.

The automatic approach used in this study to assess LIC's prosodic skills confirms the clinical descriptions of the subjects' communication impairments [6]. Combined with traditional clinical evaluations, these results also suggest that pragmatic skills and some intonation features could be considered as language differential markers of pathology (e.g. LIC vs. ASD), but also within LIC (e.g. AD vs. PDD-NOS vs. SLI).

### B. *Assessement of social interactive synchrony*

Investigations on social interactive synchrony require simultaneous multimodal processing of not only the children but also the partner (e.g., the therapist). We recently proposed to exploit speech and gestural turn-taking cues, dialog acts and synchronized motion cues for the analysis of social coordination (Figure 5). The children and therapist were performing the social coordination task described in III.B.3: cooperation for the completion of a puzzle (e.g., clown). In addition, questionnaires filled by judges were collected to evaluate the perceived coordination of dyads. The multimodal modeling gives insights on how a given partner tries to adapt and synchronize to the other partner. More interestingly, the leader's variations of rhythm and difference of rhythm between the partners were badly perceived by the judges [24].

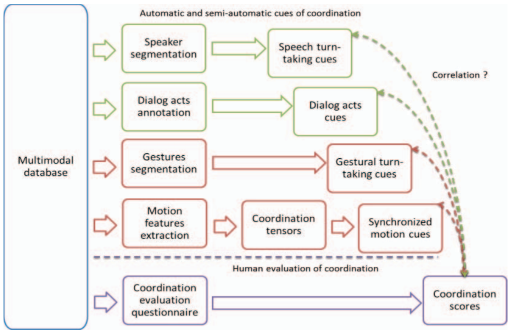

Figure 5. Synopsis of the automatic and semi-automatic audio-visual cues extracted on the child-therapist dyad.

### C. *Future work*

Our current works are firstly devoted to the analysis of the data collected through the experimental scenarios. One point that should be highlighted is that some children have performed all the scenarios. Consequently, one of our objectives is to propose a global view of multimodal, emotional and social processing in ASD.

Secondly, we are currently introducting social robots as tools for investigating interaction. Due to their agentivity, social robots open new opportunities for the convergence of social signal processing and clinical research as mentioned by Meltzoff et al. [25].

## V. Conclusions

To create experimental protocols and databases that will contribute the research in social signal processing for autism, interdisciplinary approaches and teams are required. By gathering researchers on psychopathology, neuroscience, engineering and robotics, we may efficiently address these challenges. By providing an automatic, detailed and objective measure of multimodal socio-emotional behaviors, we thought that our method would become a valuable tool for examining language, emotional and social interactions in clinical populations like autism spectrum disorders. We hope that our investigations, interface between social signal processing and psychopathology, will a useful aid to identify disorder-specific characteristics, improved early identification, and informed treatment.

## Acknowledgment

The authors would like to thank Dr. David Cohen's staff, from the Department of Child and Adolescent Psychiatry, AP-HP, Groupe Hospitalier Pitié-Salpétrière, Paris, France: Dr. Jean Xavier for participating in patient's diagnosis and recruitment; Dominique Chauvin, Julie Demouy and Elodie Tilmont for participating in tasks construction and clinical feedbacks; Nicolas Bodeau for his help in statistical analysis. We would also like to thank Emilie Delaherche and Fabien Ringeval from ISIR CNRS UMR 7222, for participating in tasks construction and for their helps in initial analyses.

## References

[1] J. Carpendale and C. Lewis, How Children Develop Social Understanding. Blackwell, London, 2006.

[2] J. Decety, and P.L. Jackson, "The functional architecture of human empathy". Behavioral and Cognitive Neuroscience Reviews, vol. 3, pp. 71–100, 2004.

[3] D.A. Baldwin and L.J. Moses, "The Ontogeny of Social Information Gathering". Child Development, vol. 67, No. 5, pp. 1915-1939, 1996.

[4] APA. DSM-IV, Diagnostic and statistic manual of mental disorders, Fourth Edition, Washington DC, American Psychiatric Association, 1994.

[5] H. Tager-Flusberg, "Language and understanding minds: connections in autism, in Understanding Other Minds 2nd edition, S. Baron-Cohen, H. Tager-Flusberg and D.J. Cohen Eds. New York: Oxford University Press, pp. 124-149, 2000.

[6] J. Demouy, M. Plaza, J. Xavier, F. Ringeval, M. Chetouani, D. Périsse, D. Chauvin, S. Viaux, B. Golse, D. Cohen, L. Robel. "Differential language markers of pathology in Autism, Pervasive Developmental Disorder Not Otherwise Specified and Specific Language Impairment", Research in Autism Spectrum Disorders, vol. 5, pp. 1402–1412, 2011.

[7] B. Chamak, B. Bonniau, E. Jaunay, D. Cohen. "What can we lean about autism from autistic persons?", Psychotherapy and Psychosomatics, vol. 77, pp. 271-279, 2008.

[8] L. Vannetzel, L. Chaby, F. Cautru, D. Cohen, and M. Plaza. "Neutral versus emotional human stimuli processing in children with pervasive developmental disorders not otherwise specified", Research in Autism Spectrum Disorders, vol 5(2), pp. 775-783, 2011.

[9] A. Kendon, "Movement coordination in social interaction: some examples described," Acta Psychologica, vol. 32, pp. 100–125, 1970

[10] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, D. Cohen. "Interpersonal Synchrony : A Survey Of Evaluation Methods Across Disciplines". IEEE Transactions on Affective Computing, to appear.

[11] E. Fombonne, "Epidemiology of autistic disorder and other pervasive developmental disorders". Journal of Clinical Psychiatry, vol. 66, pp. 3-8, 2005.

[12] S.J. Rogers and G. Dawson G. Early Start Denver Model for Young Children with Autism: Promoting Language, Learning, and Engagement, New York, Guilford Press, 2010.

[13] S. Baron-Cohen, S. Wheelwright, A. Cox, G. Baird, T. Charman, J. Swettenham, A. Drew, and P. Doehring, "The early identification of autism: the Checklist for Autism in Toddlers (CHAT)". Journal of the Royal Society of Medecine, vol. 93, pp. 521-525, 2000.

[14] A.Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," Image and Vision Computing, vol. 27, no. 12, pp. 1743–1759, 2009.

[15] C. Lord, M. Rutter and A. Le Couteur, A. "Autism diagnostic interview-revised: A revision version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders". Journal of Autism and Developmental Disorders, vol. 24(5), pp. 659–685, 1994.

[16] C. Lord, M. Rutter, S. Goode, J. Heemsbergen, H. Jordan, L. Mawhood, and E. Schopler. "Autism Diagnostic Observa- tion Schedule: A standardized observation of communicative and social behavior", Journal of Autism and Developmental Disorders, vol. 19, pp. 185–212, 1989.

[17] E. Schopler, R.J. Reichler, R.F. Devellis and K. Dally. "Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS)", Journal of Autism and Developmental Disorders, vol. 10(1), pp. 91–103, 1980.

[18] A. Khomsi, A. Evaluation du Langage Oral. Paris, ECPA, 2001.

[19] M. Kipp. "Multimedia Annotation, Querying and Analysis in ANVIL". In: M. Maybury (ed.) Multimedia Information Extraction, Chapter 19, IEEE Computer Society Press, to appear.

[20] F. Ringeval, J. Demouy, G. Szaszák, M. Chetouani, L. Robel, J. Xavier, D. Cohen, M. Plaza, M. "Automatic intonation recognition for the prosodic assessment of language impaired children". IEEE Transactions on Audio, Speech and Language Processing, vol. 19, no 5 pp. 1328-1342, 2011.

[21] M. Mayer. Frog, where are you? New York, Dial Press, 1969.

[22] N. Ebner, M. Riediger and U. Lindenberger. "FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation", Behavior Research Methods, vol. 42, pp. 351-362, 2010.

[23] P. Belin, S. Fillion-Bilodeau, F. Gosselin. "The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing", Behav Res Methods, vol. 40, pp. 531-539, 2008.

[24] E. Delaherche and M. Chetouani, "Characterization of coordination in an imitation task : human evaluation and automatically computable cues," in 13th International Conference on Multimodal Interaction, 2011.

[25] A.N. Meltzoff, P.K. Kuhl, J. Movellan, and T.J. Sejnowski, "Foundations for a new science of learning," Science, vol. 325, no. 5938, pp. 284–288, 2009.