

Computerized home video detection for motherese may help to study impaired interaction between infants who become autistic and their parents

AMMAR MAHDHAOUI,¹ MOHAMED CHETOUANI,¹ RAQUEL S. CASSEL,^{2,3}
CATHERINE SAINT-GEORGES,^{2,3} ERIKA PARLATO,⁴ MARIE CHRISTINE LAZNIK,⁵ FABIO APICELLA,⁶
FILIPPO MURATORI,⁶ SANDRA MAESTRO⁶ & DAVID COHEN^{2,3}

1 Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie, Paris, France

2 Department of Child and Adolescent Psychiatry, AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Université Pierre et Marie Curie, Paris, France

3 Laboratoire Psychologie et Neurosciences Cognitives, Paris, France

4 Department of Linguistics, University of Campinas, Brazil

5 Department of Child and Adolescent Psychiatry, Association Santé Mentale du 13ème, Paris, France

6 Scientific Institute Stella Maris, University of Pisa, Italy

Key words

motherese, autism, computerized detector

Correspondence

David Cohen, Department of Child and Adolescent Psychiatry, Groupe Hospitalier Pitié-Salpêtrière, APHP, Université Pierre et Marie Curie, 47 bd de l'Hôpital, 75013, Paris, France.
Telephone (+33) 1-42-16-23-51
Fax (+33) 1-42-16-23-53
Email: david.cohen@psl.aphp.fr

Received 1 August 2008;
revised 15 June 2009;
accepted 22 July 2009

Abstract

Autism is a well-defined clinical syndrome after the second year of life, but information on autism in the first two years of life is still lacking. The study of home videos has described children with autism during the first year of life as not displaying the rigid pattern typical of later symptoms. Therefore, developmental/environmental factors are claimed in addition to genetic/biological ones to explain the onset of autism during maturation. Here we describe (1) a developmental hypothesis focusing on the possible implication of motherese impoverishment during the course of parent–infant interactions as a possible co-factor; (2) the methodological approach we used to develop a computerized algorithm to detect motherese in home videos; (3) the best configuration performance of the detector in extracting motherese from home video sequences (accuracy = 82% on speaker-independent versus 87.5% on speaker-dependent) that we should use to test this hypothesis. *Copyright © 2011 John Wiley & Sons, Ltd.*

Introduction

Autism is a severe psychiatric syndrome characterized by the presence of abnormalities in reciprocal social interactions,

abnormal patterns of communication, and a restricted, stereotyped, repetitive repertoire of behaviours, interests, and activities. Although it is a well-defined clinical syndrome after the second, and especially after the third,

year of life, information on autism in the first two years of life is still lacking (Short and Schopler, 1988; Stone *et al.*, 1994; Sullivan *et al.*, 1990). Home movies (that is, films recorded by parents during the first years of life, before diagnosis) and direct observation of infants at high risk due to having an autistic sibling are the two most important sources of information for overcoming this problem. They have both described children with autism disorder (AD) during the first year of life (and also in the first part of the second year) as not displaying the rigid patterns described in older children. In particular, the children can gaze at people, turn toward voices and express interests in communication as typically developing infants do. It is of seminal importance to have more insight into these social competencies and in which situations they preferentially emerge in infants who are developing autism. In this paper, we focus on a verbal stimulus called motherese (a special kind of speech that is directed towards infants) that has been shown to be of crucial importance in language acquisition (Kuhl, 2004), and question whether its course during parent/child interaction might be affected by the autistic child.

First clinical manifestations of autism

Long before a diagnosis is established, parents are often aware of the differences exhibited by their autistic children (Werner *et al.*, 2000). Parents' descriptions include extremes of temperament and behaviour (ranging from marked irritability to alarming passivity), poor eye contact, and lack of responsiveness to parents' voices or attempts to play and interact (e.g., Dahlgren and Gillberg, 1989; De Giacomo and Fombonne, 1998; Ohta *et al.*, 1987).

Data from prospective and longitudinal studies focusing on high-risk infants (siblings of AD children) have also yielded important insights into what constitutes the initial behavioural signs of autism. Studies have reported various behavioural markers that, in part, predict a subsequent diagnosis of autism in a sibling sample. These markers include atypicalities in eye contact, visual tracking, social smiling, and reactivity (Zwaigenbaum *et al.*, 2005), lack of orienting to the child's own name (Nadig *et al.*, 2007; Zwaigenbaum *et al.*, 2005), prolonged latency to disengage visual attention (Swettenham *et al.*, 1998; Bryson *et al.*, 2007; Zwaigenbaum *et al.*, 2005), deficits in joint attention skills (Mundy *et al.*, 1990; Baron-Cohen *et al.*, 1996; Sullivan *et al.*, 2007), lack of pointing to objects (Baron-Cohen *et al.*, 1996), lack of imitation (Baron-Cohen *et al.*, 1996; Zwaigenbaum *et al.*, 2005), delayed expressive and receptive language (Sullivan *et al.*, 2007; Landa and Garrett-Mayer,

2006), and a lack of interest or pleasure in self-initiated contact with others (Bryson *et al.*, 2007; Baron-Cohen *et al.*, 1996; Zwaigenbaum *et al.*, 2005).

Several studies focusing on early home videos have also revealed atypical developmental tendencies in infants later diagnosed with autism spectrum disorders (ASDs). In the first year, autistic infants tend to lack social interest, to interact poorly (Maestro *et al.*, 2002; Zakian *et al.*, 2000; Adrien *et al.*, 1993), to look less at others (Clifford and Dissanayake, 2007; Osterling and Dawson, 1994; Osterling *et al.*, 2002; Maestro *et al.*, 2002, 2005b, 2006) and to show less communicative skills (e.g., vocalizations *to people*). Compared with delayed infants, the characteristics of a reduced response to the infant's own name, a reduced quality of affect and a tendency to look less at others appear more specific to autistic children (Clifford and Dissanayake, 2007; Baranek, 1999; Osterling *et al.*, 2002). In the second year, pre-autistic signs (lack of social behaviour and joint attention) become more obvious and signs particular to autism more numerous: lower eye contact quality, gaze aversion, lower social peer interest, less positive affect, less nestling, less protodeclarative showing, less anticipatory postures and less conventional games (Clifford and Dissanayake, 2007; Clifford *et al.*, 2007); eye-contact quality and affect quality have been found to be the best predictors of a later diagnosis of autism (Clifford *et al.*, 2007; Maestro *et al.*, 2001). Whether these early signs impact the interactive process between an infant and his/her parents and, as a consequence, whether they influence the development of the infant himself, remain two challenging issues.

Early language acquisition and social interactions

To explore these points, we will now briefly review what we know about language learning, social interactions, and ways to integrate the disadvantages of autistic children in both domains within the course of development. Studies of language acquisition are too numerous for an extensive review; however, we wish to highlight some specific points regarding the developmental course of language acquisition in children developed by Kuhl (2007, 2004, 2003, 2000) and Goldstein *et al.*, (2003). (1) From birth, infants possess the ability to discriminate all languages at the phonetic level and to recognize their prosodic features. (2) They develop a strategy of learning based on input language signs and characteristics and explore language statistical properties, leading to "so-called" probabilistic learning (Milgram and Atlan, 1983). (3) The language experience will commit the perceptual system at the neural level, increasing native-language speech perception

and decreasing foreign-language speech perception with a “magnet effect” (Kuhl, 2000). However, simple exposure does not explain language learning: in both speech production and speech perception, the presence of a human being interacting with a child has a strong influence on learning (Goldstein *et al.*, 2003).

Regardless of the influence of the neurobiological and genetic factors on autism (Cohen *et al.*, 2005), one should keep in mind that infants’ survival and development depend on social interaction with a caregiver who serves the infant’s needs for an emotional attachment. In cases of early severe deprivation, the consequences for infant development are sometimes impressive. Rutter’s studies of children adopted into the United Kingdom following early severe deprivation in Romania showed that dramatic damage stemmed from institutional deprivation, and its heterogeneity in outcome includes cognitive deficit and attachment problems, but also quasi-autistic patterns (e.g., Rutter *et al.*, 2007a, 2007b; Croft *et al.*, 2007).

The quality of social interaction depends on a reciprocal process, an active dialogue between parent and child based on the infant’s early competencies and the mother’s (or father’s) capacity for tuning. Micro-analysis of film records examining rhythms and patterns in mother–infant face-to-face interaction in the neonatal period (Brazelton *et al.*, 1975) and early communication (Condon and Sander, 1974; Stern *et al.*, 1975) have emphasized the importance of synchrony and co-modality of these early interactions for infants’ development.

Combining early language acquisition and social interaction

Interestingly, researchers in language acquisition and researchers in early social interactions have encountered an important peculiarity that affects both the language and social development of infants: the way adults speak to infants. A special kind of speech that is directed towards infants, often called “motherese” (“parentese”), characterized by higher pitch, slower tempo, and exaggerated intonation contours (Fernald, 1985; Grieser and Kuhl, 1988), seems to play an important role in both social interaction and language development. Studies have revealed that this particular prosody may be responsible for attracting infants’ attention, conveying emotional affect, and conveying language-specific phonological information (Karzon, 1985; Hirsh-Pasek *et al.*, 1987; Kemler Nelson *et al.*, 1989; Fernald, 1985; Fernald and Kuhl, 1987). (1) When given a choice, infants show a preference for mothers’ infant-directed speech (so-called motherese) as opposed to adult-directed speech (Pegg *et al.*, 1992; Cooper and

Aslin, 1990; Werker and McLeod, 1989; Fernald, 1985; Glenn and Cunningham, 1983). This particular prosody helps to engage and maintain the limited attention of the baby. (2) Motherese contains particularly good phonetic exemplars – sounds that are clearer, longer, and more distinct from one another (e.g., vowel hyperarticulation) – when compared to adult-directed speech (Kuhl *et al.*, 1997; Burnham *et al.*, 2002).

In addition, the baby’s reactions amplify the contours of prosody curves in the mother’s voice (Burnham *et al.*, 2002). Mothers’ infant-directed speech was found to depend on the quality of infants’ responsiveness, which suggests that infants are actively involved in early interactions (Braarud and Stormark, 2008).

Is it possible to implicate motherese impoverishment in the pathogenesis of autism?

In our view, the well-known autistic impairments in language, cognition and social development, as well as the tendencies toward self-absorption, perseveration and self stimulation (Volkmar and Pauls, 2003), may be downstream effects of primary difficulties in the ability to engage in interactions involving emotional signals, motor gestures and communicative acts directed to others. This view is also supported by testimony of adult individuals with autism who expressed themselves (Chamak *et al.*, 2008). If learning depends on a normal social interest in people and the signals they produce, children with autism, who lack a social interest, may be at a cumulative disadvantage in language learning. Their poor response to parental engagement may impoverish both parental engagement and motherese production over time. As a consequence, this impoverishment will reinforce social withdrawal and language acquisition delay.

In a previous exploratory study, we observed home movie sequences in which a withdrawn infant who will later develop autism may suddenly appear joyful when the parent implements a vocal expression using motherese. During such interactions, infants and toddlers with autism exhibit social focal attention: their faces light up, unexpected interactive skills appear, and real protodialogues expand (Laznik *et al.*, 2005). However, we were unable to find any study comparing interest in motherese speech versus other human speech in children with autism. Even if we know that autistic children can process some aspects of human voices (Groen *et al.*, 2008), however, studies have shown that autistic individuals display no specific cortical activation in response to human voices (e.g., Zilbovicius *et al.*, 2006). Further, five-year-old autistic children do not show the expected preference for their mother’s speech when given it

as a choice with the noise of superimposed voices (Klin, 1991). Similarly, 32- to 52-month-old children with an ASD have a significant listening preference for electronic non-speech motherese analogue signals in comparison to motherese signals (Kuhl *et al.*, 2005). Given that home movies offer a unique opportunity to follow infant development and parent–infant interactions, we planned to use a multi-disciplinary approach with a child psychiatrist, developmental psychologist, psycholinguist and engineer to test the following research questions: (1) Are infants who later develop autism initially equipped to respond to motherese specifically? In other words, is this competency stable and correlated with positive interactions? (2) As motherese amplification is bidirectional, does the parental quantity of motherese decrease overtime?

Aims of the present paper

In this paper, we describe (1) a computerized algorithm to detect motherese developed for this research; (2) its characteristics in extracting motherese from home video sequences. Given that micro-analysis of home video is highly time consuming, we consider the production of this algorithm as the first step of our multi-disciplinary research project. Indeed, even if motherese is clearly defined in terms of acoustic properties, the modelling and detection is expected to be difficult, as is the case with the majority of emotional speech, because the characterization of spontaneous and affective speech in terms of features is still an open question since several parameters have been proposed in the literature (Schuller *et al.*, 2007). As a starting-point following the acoustic properties of motherese, we characterized verbal interactions by the extraction of supra-segmental features (prosody). Given that home movies are not recorded by professionals and often contain adverse conditions (e.g., with regard to noise, the camera, or microphones), however, acoustic segmentation of home movies shows that segmental features play a major role in robustness (Schuller *et al.*, 2007). Consequently, the utterances are characterized by both segmental (Mel Frequency Cepstrum Coefficients, MFCCs) and supra-segmental (e.g., statistics with regard to fundamental frequency, energy, and duration) features. These features, along with the known emotional labels of a training set of emotional utterances, are provided as input to a machine learning algorithm. Here, we compared the performance of two learning algorithms, GMM (Gaussian Mixture Model) and k-nn (k-nearest neighbours), on segmental and non-segmental features alone using the ROC (receiver operating characteristic) analysis method. To improve detection,

we also investigated several combinations or fusion schemes (segmental/supra-segmental; GMM/k-nn) to select the one that is optimal in comparison to manual segmentation of motherese sequences from home videos.

Method

Database

The speech corpus used in this experiment is a collection of natural and spontaneous interactions. This corpus contains expressions of non-linguistic communication (affective intent) conveyed by a parent to a preverbal child. The corpus consists of recordings of Italian mothers and fathers addressing their infants. We decided to focus on the analysis of home movies, as it enables a longitudinal study (months or years) and gives information about early behaviours of autistic infants long before the diagnosis is made by clinicians. All sequences were extracted from the Pisa home video database, which includes home movies from the first 18 months of life for three groups of children aged four to seven years (Maestro *et al.*, 2005a, 2005b). Children are matched for gender and socio-economic status. The first group was composed of 15 children (Male/Female 10:5) with a diagnosis of autism based on the Autism Diagnostic Interview-Revised (ADI-R) (Rutter *et al.*, 2003). All cases with Childhood Autism Rating Scale (CARS) total-scores (Schopler *et al.*, 1988) below 30 were excluded. The second group was composed of 12 children (Male/Female 7:5) with a diagnosis of Mental Retardation (MR) who had a non-autistic CARS total score under 25. Children with MR or AD secondary to neuropsychiatric syndromes (e.g., Fragile-X, Rett, or Down Syndrome) or evident neurological deficits or physical anomalies were excluded. Participants in the AD or MR group were administered the Griffiths Mental Developmental Scale or Wechsler Intelligent Scale in order to determine intellectual functioning. For both AD [mean IQ = 59.26; standard deviation (SD) = 8.49] and MR (mean IQ = 56.82; SD = 8.16) groups, the composite IQ score was below 70. A third group was composed of 15 typical developing (TD) children (Male/Female 9:6) recruited from children attending a local kindergarten.

The large size of this corpus, however, makes it inconvenient for human review. For the development of the first algorithm, we focused on one home video of superior acoustic quality that totalled three hours of the first year of life of an infant who later developed autism (at six years: ADI-R communication score = 9; ADI-R social interest score = 12; ADI-R repeated interest score = 5; CARS total score = 36; IQ = 71). The verbal interactions of the infant's mother were carefully annotated by two

blind psycholinguists into two categories: infant-directed speech (motherese) and adult-directed speech (normal) [Cohen's kappa = 0.82; 95% confidence interval (CI) = 0.75–0.90]. From this manual annotation, we extracted 100 utterances for each category. The utterances are typically between 0.5 seconds and 4 seconds in length. To conduct a secondary performance characteristic study, we collected an independent set of utterances (50 motherese and 50 normal speech) extracted from 10 randomly selected home movies (five from the AD group and five from the TD group).

System description

Automatic speech sounds segmentation requires several steps: feature extraction, classification and decision fusion are conducted sequentially. These steps can be divided into two major processing phases, namely the training phase and the test or classification phase. Figure 1 shows a schematic overview, which is described in more detail in the following paragraphs. The first step, namely feature extraction, shown in Figure 1 is needed in both training and test. The second step, called classification, aims at classifying the whole utterances. The last step, which is called decision fusion, is necessary for classification and combines the different streams to form one common output.

Feature extraction

After the segmentation of the database into utterances, the next step is to extract features, resulting in a representation of the speech signal as input for the motherese classification system. In this paper we evaluate two approaches, respectively termed segmental and supra-segmental features. The first is characterized by the MFCCs, while the second is characterized by statistical measures of both the fundamental frequency (F0) and the short-time energy. A 20 ms window is used, and the overlapping between adjacent frames is halved. A parameterized vector of order 16 was computed. The supra-segmental features are characterized by three statistics (mean, variance and range) on both F0 and

short-time energy, resulting in a six-dimensional vector. One should note that the duration of the acoustic events is not directly characterized as a feature but is taken into account during the classification process by a weighting factor. The feature vectors are normalized (zero mean, unit standard deviation).

Classification

In this study, two different classifiers – the k-nn classifier and the GMM-based classifier – were investigated. The k-nn is a distanced based classifier often used in pattern recognition. GMM is a statistical signal model trained by sequences of feature vectors that are representative of the input signal (Reynolds, 1995).

A posteriori probabilities estimation. The GMM is adopted to represent the distribution of the features. Under the assumption that the feature vector sequence $x = \{x_1, x_2, \dots, x_n\}$ is an independent identical distribution sequence, the estimated distribution of the d -dimensional feature vector x is a weighted sum of M component Gaussian densities $g(\mu, \Sigma)$, each parameterized by a mean vector μ_i and covariance matrix Σ_i ; the mixture density for the model C_m is defined as:

$$P(x|C_m) = \sum_{i=1}^M w_i g(\mu_i, \Sigma_i)(x) \quad (1)$$

Each component density is a d -variate Gaussian function:

$$g(\mu, \Sigma)(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} e^{-1/2(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (2)$$

The mixture weights w_i satisfy the following constraint:

$$\sum_{i=1}^M w_i = 1 \quad (3)$$

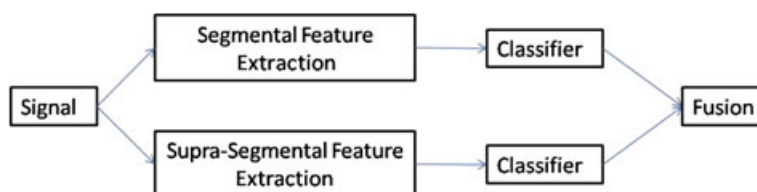


Figure 1 Motherese classification system used in the algorithm.

The feature vector x is then modelled by the following a posteriori probability:

$$P_{gmm}(C_m|x) = \frac{p(x|C_m)p(C_m)}{p(x)} \quad (4)$$

where $P(C_m)$ is the prior probability for class C_m , we assume equal prior probabilities. We use the expectation maximization (EM) algorithm for the mixtures to obtain maximum likelihood. The k-nn classifier is a non-parametric technique which classifies the input vector with the label of the majority k-nn (prototypes) (Duda *et al.*, 2000). In order to keep a common framework with the statistical classifier (GMM), we estimate the a posteriori probability that a given feature vector x belongs to class C_m using k-nn estimation (Duda *et al.*, 2000):

$$P_{knn}(x|C_m) = \frac{km}{k} \quad (5)$$

where k_m denotes the number of prototypes that belong to the class C_m among the k-nn.

Segmental and supra-segmental based characterization. Segmental features (i.e. MFCC) are extracted from all the frames of an utterance U_x independently to the voiced or unvoiced parts. One should note that the nature of the segments can also be exploited (vowels/consonants) (Ringeval and Chetouani, 2008). A posteriori probabilities are then estimated by both GMM and k-nn classifiers and are respectively termed $P_{gmm,seg}(C_m|U_x)$ and $P_{knn,seg}(C_m|U_x)$.

The classification of supra-segmental features follows the segment-based approach (SBA) (Shami and Verhelst, 2007). An utterance U_x is segmented into N voiced segments (F_{xi}) obtained by F0 extraction (see earlier). Local estimation of a posteriori probabilities is carried out for each segment. The utterance classification combines the N local estimations.

$$P(C_m|U_x) = \sum_{xi=1}^N P(C_m|F_{xi}) \times length(F_{xi}) \quad (6)$$

The duration of the segments is introduced as weights of the a posteriori probabilities: importance of the voiced segment ($length(F_{xi})$). The estimation is also carried out by the two classifiers, resulting in supra-segmental characterizations: $P_{gmm,supra}(C_m|U_x)$ and $P_{knn,supra}(C_m|U_x)$.

Fusion

The segmental and supra-segmental characterizations provide different temporal information, and combining them should improve the accuracy of the detector. Many decision techniques can be employed (Kuncheva, 2004), but we investigated a simple weighted sum of likelihoods from the different classifiers:

$$C_l = \lambda \cdot \log(P_{seg}(C_m|U_x)) + (1 - \lambda) \cdot \log(P_{supra}(C_m|U_x)) \quad (7)$$

with $l = 1$ (motherese) or 2 (normal directed speech); λ denotes the weighting coefficient. For the GMM classifier, the likelihoods can be easily computed from the a posteriori probabilities [$P_{gmm,seg}(C_m|U_x)$, $P_{gmm,supra}(C_m|U_x)$] (Reynolds, 1995). However, the k-nn estimation can produce a null a posteriori probability incompatible with the computation of the likelihood. We used a solution recently tested by Kim *et al.* (2007), in which the a posteriori probability is used instead of the log probability of the k-nn:

$$C_l = \lambda \cdot \log\left(e^{P_{knn,seg}(C_m|U_x)}\right) + (1 - \lambda) \cdot \log(P_{gmm,supra}(C_m|U_x)) \quad (8)$$

Consequently, for the k-nn classifier we used Equation 7, while for the GMM the likelihood is conventionally computed. We investigated cross combinations (Table 1).

Results

Classifier configuration

First, to find the optimal structure of our classifier, we had to adjust the parameters: the number of Gaussians (M) for the GMM classifier and the number of neighbours (K) for the k-nn classifier. We searched for the optimal configuration in terms of accuracy. Table 2 shows the best

Table 1 Table of combinations of classifiers (GMM/k-nn) × features (segmental/supra-segmental)

Combination 1	$P_{knn,seg}$	$P_{knn,supra}$
Combination 2	$P_{gmm,seg}$	$P_{gmm,supra}$
Combination 3	$P_{knn,seg}$	$P_{gmm,supra}$
Combination 4	$P_{gmm,seg}$	$P_{knn,supra}$
Combination 5	$P_{gmm,seg}$	$P_{knn,seg}$
Combination 6	$P_{gmm,supra}$	$P_{knn,supra}$

Note: GMM, Gaussian Mixture Model; k-nn, k-nearest neighbours; seg, segmental; supra, supra-segmental.

Table 2 Accuracy of optimal configurations for GMM and k-nn classifiers according to segmental and supra-segmental features

	Segmental	Supra-segmental
k-nn	72.5% ($K=11$)	61% ($K=7$)
GMM	78% ($M=15$)	82% ($M=16$)

Note: GMM, Gaussian Mixture Model; k-nn, k-nearest neighbours; M , number of Gaussians for the GMM classifier; K , number of neighbours for the k-nn classifier.

configuration of GMM and k-nn with segmental and supra-segmental features and further shows that the GMM classifier trained with prosody features outperformed the other classifiers in terms of accuracy.

Fusion of best system

In this study the performance of the classifier was quoted using class sensitivities, predictivities and overall accuracy, and the optimal classifiers were determined by employing ROC graphs to show the trade-off between the hit and false positive rates. A ROC curve represents the trade-off between the false acceptance rate (FAR) and the false rejection rate as the classifier output threshold value is varied. Two quantitative measures of verification performance, the equal error rate (EER) and the area under the ROC curve (AUC), were calculated. All calculations were conducted with Matlab (version 6). For best configuration data, results were given with the 95% CIs that were

estimated using Cornuéjols *et al.* (2002) method. It should be noted that while EER represents the performance of a classifier at only one operating threshold, the AUC represents the overall performance of the classifier over the entire range of thresholds. Hence, we employed AUC and not EER to compare the verification performance of two classifiers and their combination. However, the result shown in Table 2 motivated an investigation of the fusion of both features and classifiers following the statistical approach described in the previous section. Improvement by the combination of features and classifiers is known to be efficient (Kuncheva, 2004). However, caution should be used, because the fusion of best configurations does not always give better results; the efficiency will depend on the nature of the errors produced by the classifiers (independent versus dependent) (Kuncheva, 2004). Table 1 and the previous section show that six different fusion schemes could be investigated [Combination 1 (Comb₁) to Combination 6 (Comb₆)]. For each of them, we optimized parameter classifiers (K value for k-nn and M value for GMM, respectively) and weighting λ (Equation 6) parameters of the fusion. In Figure 2, we can see that for the k-nn classifier, the best scores (0.8113/0.812) were obtained with an important contribution of the segmental features ($\lambda=0.8$), which is in agreement with the results obtained without the fusion (Table 2). The best GMM results (0.932/0.932) are obtained with a weighting factor equal to 0.6, revealing a balance between the two features. Table 3 summarizes the best results in terms of accuracy as well as the positive predictive value (PPV) and negative predictive value (NPV) for each classifier fusion (top section) and cross-classifier fusion (bottom section).

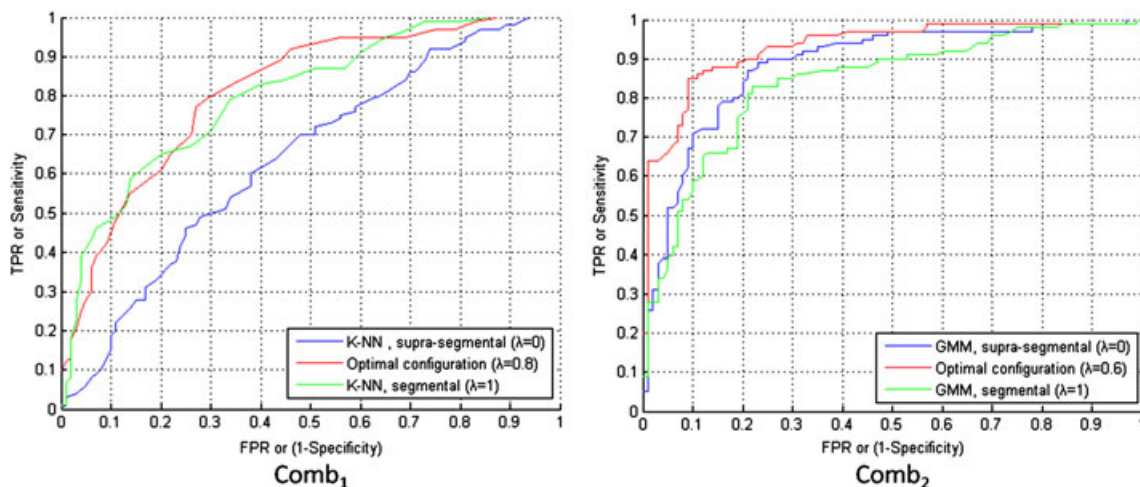


Figure 2 ROC curves for Combination 1 (Comb₁) and Combination 2 (Comb₂). Combination 1 = $P_{knn,seg} \times P_{knn,supra}$; Combination 2 = $P_{gmm,seg} \times P_{gmm,supra}$; λ = weighting coefficient used in the equation fusion for each combination.

Table 3 Accuracy, positive predictive value (PPV), and negative predictive value (NPV) of fusion configurations according to classifiers and segmental and supra-segmental feature parameters

		Accuracy	PPV	NPV	Λ
<i>Using only one classifier</i>					
Combination 1 k-nn, seg × k-nn, supra	k-nn, seg ($K=11$)	74%	77.91%	71.05%	0.9
	k-nn, supra ($K=1$)				
Combination 2 GMM, seg × GMM, supra	k-nn, seg ($K=7$)	72.5%	74.42%	71.84%	0.8
	k-nn, supra ($K=11$)				
	GMM, seg ($M=12$)	87.5%	88.47%	86.41%	0.4
	GMM, supra ($M=15$)				
	GMM, seg ($M=16$)	86.5%	84.85%	88.42%	0.5
	GMM, supra ($M=16$)				
<i>Combining k-nn and GMM classifiers</i>					
Combination 3 K-nn, seg × GMM, supra	k-nn, seg ($K=11$)	85.5%	89.41%	83.81%	0.7
	GMM, supra ($M=12$)				
	K-nn, seg ($K=5$)	85%	90.7%	82.69%	0.6
Combination 4 GMM, seg × k-nn, supra	GMM, supra ($M=12$)				
	GMM, seg ($M=15$)	80.5%	79.61%	81.44%	0.2
	k-nn, supra ($K=11$)				
	GMM, seg ($M=16$)	79.5%	77.57%	81.72	0.3
	k-nn, supra ($K=1$)				

Note: PPV (positive predictive value) = TruePositive/(TruePositive + FalsePositive). PPV evaluates the proportion of correctly detected motherese utterances out of all utterances labelled as motherese. PPV can be viewed as the reliability of positive predictions induced by the classifier. NPV (negative predictive value) gives similar information about the detection of normal speech.

Table 3 shows that different combinations of segmental and supra-segmental features can reach similar accuracies with different PPVs. With regard to the k-nn method, the best scores (accuracy = 74%; PPV = 77.91%) were obtained with an important contribution of the segmental features ($\lambda = 0.9$). The best GMM results (accuracy = 87.5%; PPV = 88.47%) were obtained with a weighting factor equal to 0.4 (i.e. a balance between the two features). We also investigated cross-classifier fusion (Table 3, bottom). The accuracy of this method can be lower than that for the single classifiers (Table 2). However, significant improvements (90.7%) in the PPV were reached for Combination 3: $P_{knn,seg} \times P_{gmm,supra}$ combination ($\lambda = 0.6$). This result reveals the importance of evaluation metrics for fusion, which is dependent on the task in this study (motherese detection).

In sum, the best fusion configuration used only the GMM classifier for both segmental and supra-segmental features ($M = 12$ and $M = 15$, respectively, and $\lambda = 0.4$). Performance of this fusion corresponds to Combination 2 from Table 3, and were as follow: accuracy = 87.5%

(95%CI = 82.91–92.08%); PPV = 88.47% (95%CI = 83.03–95.18%); NPV = 86.41% (95%CI = 79.4–92.88%). We also investigated the performance of the detector with these fusion parameters in detecting motherese versus normal speech within a second set of utterances extracted from 10 randomly-selected home videos with 12 independent speakers. Performances under speaker-independent conditions were as follows: accuracy = 82% (95%CI = 73.87–89.58%); PPV = 86.36% (95%CI = 66.52–89.48%); NPV = 77.55% (95%CI = 76.73–96.6%).

Discussion

Motherese detector validity

Our goal was to develop a motherese detector by investigating different features and classifiers. Using classification techniques that are often used in speech and speaker recognition (GMM and k-nn), we have developed a motherese detection system and tested it on mode dependent of speaker. The fusion of features and classifiers was also investigated. We obtained results from which we

can draw several interesting conclusions. First, our results show that segmental features alone contain much useful information for discrimination between motherese and adult direct speech, since they outperform supra-segmental features. Thus, we can conclude that segmental features can be used alone. However, according to our detection results, prosodic features are also very promising. Based on the previous two conclusions, we combined classifiers that use segmental features with classifiers that use supra-segmental features and found that this combination improves the performance of our motherese detector considerably.

Thus, we can conclude that a classifier based on segmental features alone can be used to discriminate between motherese and normal-directed-speech, but a significant improvement can be achieved in most cases when this classifier is fused with another that is based on prosodic features. Furthermore, fusing scores from k-nn and GMM is also very fruitful.

Limitations and strengths

For our motherese classification experiments, we used only speech excerpts that were already segmented (based on human transcription). In other words, detection of the onset and offset of motherese was not investigated in this study but can be addressed in a follow-up study. Detection of the onset and offset of motherese (motherese segmentation) can be seen as a separate problem that gives rise to other interesting questions, such as how to define the beginning and end of motherese and what kind of evaluation measures to use. These are problems that can typically be addressed within a Hidden Markov Model framework.

In its best configuration, the novel detector had positive and negative predictive values reaching 90%; this level of prediction is suitable for further studies of home videos. When we explored the motherese detector's performance on sequences blindly validated by two psycholinguists (speaker-independent), the performance of the detector remained very good (accuracy = 82%). As we hypothesize the detector to be language independent, however, we still need to explore the motherese detector's performance on sequences randomly extracted from different languages. As noted, although motherese is itself a highly variable and complex signal (Fernald and Kuhl, 1987), it would be interesting to investigate to what extent motherese is dependent on the speaker, culture, etc.

Finally, the aim of this study was to develop a motherese detector to enable emotion classification. Our plan for emotion recognition is based on the fact that emotion is expressed in speech by audible paralinguistic features or events, which can be seen as the “building

blocks” or “emotion features” of a particular emotion. Adding visual information could further help to improve emotion detection. Our plan is thus to perform emotion classification via the detection of audible paralinguistic events. In this study, we have developed a motherese detector to provide a first step in this plan towards a future emotion classifier.

Possible applications to the field of autism

Despite the limitations listed earlier, we are now systematically exploring in home videos from the Pisa database (Maestro *et al.*, 2005b) the natural courses of motherese and parent–infant interaction and their co-occurrence. The pattern of parent–infant interactions has been extensively explored in infants later diagnosed with autism (for reviews see Saint-Georges *et al.*, 2010; Palomo *et al.*, 2006). However, whether motherese and positive interactions co-occur has not been investigated. This will be tested on randomized sequences extracted from the database using group comparisons matched for age and sex: autism versus typically developing; autism versus children with intellectual disability; early onset autism versus late onset autism. These analyses are made possible by the availability of several cases matched for age and sex with normal and delayed developing controls (Maestro *et al.*, 2005b). We hypothesize that motherese should be correlated with positive interactions both in autism and non-autism, but to a lesser extent in autism.

The second analysis will focus on a longitudinal view of motherese using a subset of cases with several videos at different ages. We hypothesize that in early onset autism, motherese will decrease over time in terms of relative frequency due to lack of interactive feedback from the child (Muratori and Maestro, 2007). This deficit in stimulation would have developmental consequences, which would lead in particular to the specific social and language impairments found in older autistic children (Zilbovicius *et al.*, 2006). Indeed, motherese prosody is related to the mother's emotions and seems an authentic marker of the quality of parent–child interactions. In their infants' absence, mothers cannot produce “faked” motherese into a microphone (Fernald and Simon, 1984): the full range of prosodic modifications in mother's speech is evoked only in the presence of the infant. On the contrary, baby's reactions improve the amplitude of prosody contours in mother's speech (Burnham *et al.*, 2002). But when a baby seems indifferent, can his/her parents still continue to produce motherese? If motherese can no longer be produced by such parents, the deficit of motherese would appear to be one of the cofactors in the

development of an interactive vicious circle. Figure 3 shows how the detector may be useful in performing such a study. In a sequence of duration 12 seconds during which a positive interaction occurred (note the baby's smile), the detector can automatically analyse the sound signal, prosodic characteristics, and main speaker and determine whether motherese is being produced.

Possible applications outside the field of autism

If our previous hypotheses are confirmed (i.e. if a decrease of motherese appears as a cofactor in the pathogenic process), a possible application for infants at risk for autism could be to improve parental motherese by

training using a screen warning for motherese production (biofeedback). As motherese has been shown to improve responsiveness and learning in various developmentally disabled children (Santarcangelo and Dyer, 1988), such training could also be useful for nurses, teachers, or other professionals working in the field of infant mental health outside the field of autism (e.g., maternal depression).

As motherese is an important factor and a valuable tool to study both first interactions and language development, we can easily imagine possible applications of its automatic detection in various research projects performed with an experimental or naturalistic design (e.g., Braarud and Stormark, 2008). Similarly, motherese prosody reflects the

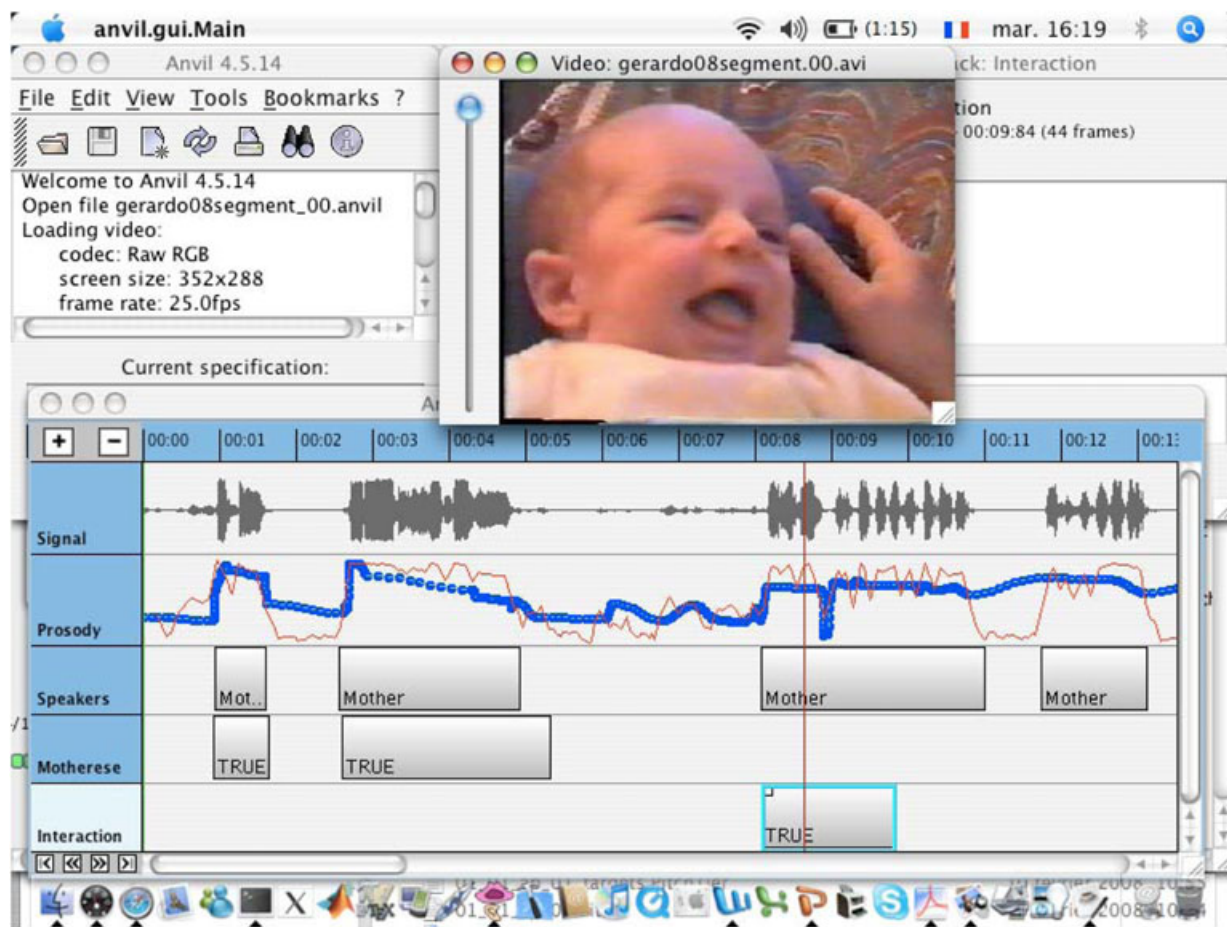


Figure 3 Use of the detector as a tool to study motherese in a home video. In a sequence of duration 12 seconds during which a positive interaction occurred (see baby's smile), the detector automatically analyses the acoustic signal, prosodic characteristics (frequency, in hertz, in blue; intensity, in dB, in red), main speaker, and whether motherese is produced. In parallel, existing positive interactions are computerized with the same temporal scale in the Pisa home video database. The red vertical line indicates the timing of the baby's picture shown (in this case, eight seconds after the sequence's starting point).

emotional affect conveyed in the voice (Trainor *et al.*, 2000). Because manual investigation are highly time consuming, automatic motherese detection could also be useful for various types of research on emotional speech (e.g., Beaucousin *et al.*, 2007).

Conclusion

By using different features, segmental and supra-segmental, we were able to automatically discriminate motherese segments from normal speech segments with low error rates. We conclude that the motherese detector is a powerful tool to study motherese in home videos, allowing in-depth

study of large samples of sequences and exhibiting good performance characteristics.

Acknowledgements

The Pisa “home video in autism” project was supported by grants PRIN 2003-2005 and PRIN 2005-2007 from the Italian Ministry of Instruction, University, and Research (MIUR). The current study was supported by a grant (n°2008005170) from the *Fondation de France* given to MC and DC.

Declaration of interest statement

The authors have no competing interests.

References

- Adrien J.-L., Lenoir P., Martineau J., Perrot A., Hameury L., Larmande C., Sauvage D. (1993) Blind ratings of early symptoms of autism based upon family home movies. *Journal of the American Academy of Child and Adolescent Psychiatry*, **32**, 617–626.
- Baranek G.T. (1999) Autism during infancy: a retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of Autism and Developmental Disorders*, **29**, 213–224, DOI: 10.1023/A:1023080005650
- Baron-Cohen S., Cox A., Baird G., Swettenham J., Nightingale N., Morgan K., Drew A., Charman T. (1996) Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry*, **168**, 158–163.
- Beaucousin V., Lacheret A., Turbelin M.-R., Morel M., Mazoyer B., Tzourio-Mazoyer N. (2007) FMRI study of emotional speech comprehension. *Cerebral Cortex*, **17**, 339–352, DOI: 10.1093/cercor/bhj151
- Braarud H.C., Stormark K.M. (2008) Prosodic modification and vocal adjustments in mothers’ speech during face-to-face interaction with their two- to four-month-old infants: a double video study. *Social Development*, **17**(4), 1074–1084, DOI: 10.1111/j.1467-9507.2007.00455.x
- Brazelton T.B., Tronick E., Adamson L., Als H., Wise S. (1975) Early mother–infant reciprocity. In Hofer M. (ed.) *Parent–Infant Interaction*, pp. 137–154, Amsterdam, Elsevier.
- Bryson S.E., Zwaigenbaum L., Brian J., Roberts W., Szatmari P., Rombough V., McDermott C. (2007) A prospective case series of high-risk infants who developed autism. *Journal of Autism and Developmental Disorders*, **37**, 11–24, DOI: 10.1007/s10803-006-0328-2
- Burnham C., Kitamura C., Vollmer-Conna U. (2002) What’s new pussycat: on talking to animals and babies. *Science*, **296**, 1435, DOI: 10.1126/science.1069587
- Chamak B., Bonniau B., Jaunay E., Cohen D. (2008) What can we learn about autism from autistic persons? *Psychotherapy and Psychosomatics*, **77**, 271–279, DOI: 10.1159/000140086
- Clifford S., Dissanayake C. (2007) The early development of joint attention in infants with autistic disorder using home video observations and parental interview. *Journal of Autism and Developmental Disorders*, **38**, 791–805, DOI: 10.1007/s10803-007-0444-7
- Clifford S., Young R., Williamson P. (2007) Assessing the early characteristics of autistic disorder using video analysis. *Journal of Autism and Developmental Disorders*, **37**, 301–313, DOI: 10.1007/s10803-006-0160-8
- Cohen D., Pichard N., Tordjman S., Baumann C., Burglen L., Excoffier S., Lazar G., Mazet P., Pinquier C., Verloes A., Heron D. (2005) Specific genetic disorders and autism: clinical contribution towards identification. *Journal of Autism and Developmental Disorder*, **35**, 103–116, DOI: 10.1007/s10803-004-1038-2
- Condon W.S., Sander L.W. (1974) Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science*, **183**, 99–101, DOI: 10.1126/science.183.4120.99
- Cooper R.P., Aslin R.N. (1990) Preference for infant-directed speech in the first month after birth. *Child Development*, **61**, 1584–1595, DOI: 10.1111/j.1467-8624.1990.tb02885.x
- Cornuéjols A., Miclet L., Kodratoff Y. (2002) *Apprentissage artificiel concepts et algorithmes*, Paris, Editions Eyrolles.
- Croft C., Beckett C., Rutter M., Castle J., Colvert E., Groothues C., Hawkins A., Kreppner J., Stevens S.E., Sonuga-Barke E.J.S. (2007) Early adolescent outcomes of institutionally-deprived and non-deprived adoptees. II: Language as a protective factor and a vulnerable outcome. *Journal of Child Psychology and Psychiatry*, **48**(1), 31–44, DOI: 10.1111/j.1469-7610.2006.01689.x
- Dahlgren S.O., Gillberg C. (1989) Symptoms in the first two years of life. A preliminary population study of infantile autism. *European Archives of Psychiatry and Neurological Sciences*, **238**, 169–174, DOI: 10.1007/BF00451006
- De Giacomo A., Fombonne E. (1998) Parental recognition of developmental abnormalities in autism. *European Child and Adolescent Psychiatry*, **7**, 131–136, DOI: 10.1007/s007870050058
- Duda R.O., Hart P., Stork D. (2000) *Pattern Classification*, 2nd edition, Chichester, John Wiley & Sons.
- Fernald A. (1985) Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, **8**, 181–195, DOI: 10.1016/S0163-6383(85)80005-9
- Fernald A., Kuhl P. (1987) Acoustic determinants of infant preference for Motherese speech. *Infant Behavior and Development*, **10**, 279–293, DOI: 10.1016/0163-6383(87)90017-8
- Fernald A., Simon T. (1984) Expanded intonation contours in mothers’ speech to newborns. *Developmental Psychology*, **20**, 104–113.
- Glenn S.M., Cunningham C.C. (1983) What do babies listen to most? A developmental study of auditory preferences in nonhandicapped infants and infants with Down’s syndrome. *Developmental Psychology*, **19**, 332–337.
- Goldstein M., King A., West M. (2003) Social interaction shapes babbling: testing parallels

- between birdsong and speech. *Proceedings of the National Academy of Sciences*, **100**, 8030–8035, DOI:10.1073/pnas.1332441100
- Grieser D.L., Kuhl P.K. (1988) Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Developmental Psychology*, **24**, 14–20, DOI: 10.1037/0012-1649.24.1.14
- Groen W.B., van Orsouw L., Zwiers M., Swinkels S., van derGaag R.J., Buitelaar J. K. (2008) Gender in voice perception in autism. *Journal of Autism and Developmental Disorders*, **38**, 1819–1826, DOI: 10.1007/s10803-008-0572-8
- Hirsh-Pasek K., Kemler Nelson D.G., Jusczyk P.W., Cassidy K.W., Druss B., Kennedy L. (1987) Clauses are perceptual units for young infants. *Cognition*, **26**, 269–286, DOI: 10.1016/S0010-0277(87)80002-1
- Karzon R.G. (1985) Discrimination of polysyllabic sequences by one- to four-month-old infants. *Journal of Experimental Child Psychology*, **39**, 326–342, DOI: 10.1016/0022-0965(85)90044-X
- Kim S., Georgiou P., Lee S., Narayanan S. (2007) Real-time emotion detection system using speech: multi-modal fusion of different time-scale features. *IEEE International Workshop on Multimedia Signal Processing*, October, DOI: 10.1109/MMSP.2007.4412815
- Kemler Nelson D.G., Hirsh-Pasek K., Jusczyk P.W., Cassidy K.W. (1989) How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, **16**(1), 55–68.
- Klin A. (1991) Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, **21**, 29–42.
- Kuhl P. (2000) A new view of language acquisition. *Proceedings of the National Academy of Sciences*, **97**(22), 11850–11857.
- Kuhl P. (2003) Human speech and birdsong: communication and the social brain. *Proceedings of the National Academy of Sciences*, **100**(17), 9645–9646, DOI: 10.1073/pnas.1733998100
- Kuhl P. (2004) Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, **5**, 831–843, DOI: 10.1038/nrn1533
- Kuhl P., Coffey-Corina S., Padden D., Dawson G. (2005) Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental Science*, **8**(1) F1–F12, DOI: 10.1111/j.1467-7687.2004.00384.x
- Kuhl P.K. (2007) Is speech learning 'gated' by the social brain? *Developmental Science*, **10**(1), 110–120, DOI: 10.1111/j.1467-7687.2007.00572.x
- Kuhl P.K., Andruski J.E., Chistovich I.A., Chistovich L.A., Kozhevnikova E.V., Ryskina V.L., Stolyarova E.I., Sundberg U., Lacerda F. (1997) Cross-language analysis of phonetic units in language addressed to infants. *Science*, **277**, 684–686, DOI: 10.1126/science.277.5326.684
- Kuncheva L.I. (2004) *Combining Pattern Classifiers: Methods & Algorithms*, Chichester, John Wiley & Sons.
- Landa R., Garrett-Mayer E. (2006) Development in infants with autism spectrum disorders: a prospective study. *Journal of Child Psychology and Psychiatry*, **47**, 629–638, DOI: 10.1111/j.1469-7610.2006.01531.x
- Laznik M.C., Maestro S., Muratori F., Parlato E. (2005) Les interactions sonores entre les bébés devenus autistes et leurs parents. In Castarede M.F., Konopczynski G. (eds). *Au commencement était la voix*, pp 171–181, Paris, Erès.
- Maestro S., Muratori F., Barbieri F., Casella C., Cattaneo V., Cavallaro M.C., Cesari A., Milone A., Rizzo L., Viglione V., Stern D., Palacio-Espasa F. (2001) Early behavioral development in autistic children: the first 2 years of life through home movies. *Psychopathology*, **34**, 147–152, DOI: 10.1159/000049298
- Maestro S., Muratori F., Cesari A., Cavallaro M.C., Paziente A., Pecini C., Grassi C., Manfredi A., Sommaro C. (2005a) Course of autism signs in the first year of life. *Psychopathology*, **38**, 26–31, DOI: 10.1159/000083967
- Maestro S., Muratori F., Cavallaro M.C., Pecini C., Cesari A., Paziente A., Stern D., Golse B., Palasio-Espasa F. (2005b) How young children treat objects and people: an empirical study of the first year of life in autism. *Child Psychiatry and Human Development*, **35**(4). DOI: 10.1007/s10578-005-2695-x
- Maestro S., Muratori F., Cavallaro M.C., Pei F., Stern D., Golse B., Palacio-Espasa F. (2002) Attentional skills during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**(10), 1239–1245.
- Maestro S., Muratori F., Cesari A., Pecini C., Grassi C., Apicella F., Stern D. (2006) A view to regressive autism through home movies. Is early development really normal? *Acta Psychiatrica Scandinavica*, **113**, 68–72, DOI:10.1111/j.1600-0447.2005.00695.x
- Milgram M., Atlan H. (1983) Probabilistic automata as a model for epigenesis of cellular networks. *Journal of Theoretical Biology*, **103**(4), 523–547, DOI: 10.1016/0022-5193(83)90281-3
- Mundy P., Sigman M., Kasari C. (1990) A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, **20**, 115–128, DOI: 10.1007/BF02206861
- Muratori F., Maestro S. (2007) Autism as a downstream effect of primary difficulties in intersubjectivity interacting with abnormal development of brain connectivity. *International Journal for Dialogical Science*, **2**(1), 93–118.
- Nadig S., Ozonoff S., Young G.S., Rozga A., Sigman M., Rogers S.J. (2007) A prospective study of response to name in infants at risk for autism. *Archives of Pediatrics & Adolescent Medicine*, **161**, 378–383.
- Ohta M., Nagai Y., Hara H., Sasaki M. (1987) Parental perception of behavioral symptoms in Japanese autistic children. *Journal of Autism and Developmental Disorders*, **17**, 549–563, DOI: 10.1007/BF01486970
- Osterling J., Dawson G. (1994) Early recognition of children with autism: a study of birthday home videotapes. *Journal of Autism and Developmental Disorders*, **24**, 247–258, DOI: 10.1007/BF02172225
- Osterling J.A., Dawson G., Munson J.A. (2002) Early recognition of 1-year old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology*, **14**, 239–251, DOI: 10.1017/S0954579402002031
- Palomo R., Belinchon M., Ozonoff S. (2006) Autism and family home movies: a comprehensive review. *Developmental and Behavioral Pediatrics*, **27**, S59–S68.
- Pegg J.E., Werker J.F., McLeod P.J. (1992) Preference for infant-directed over adult-directed speech: evidence from 7-week-old infants. *Infant Behavior and Development*, **15**, 325–345, DOI:10.1016/0163-6383(92)80003-D
- Reynolds D. (1995) Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication*, **17**, 91–108.
- Ringeval F., Chetouani M. (2008) Exploiting a Vowel Based Approach for Acted Emotion Recognition. *International Workshop on Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction*, Berlin, Springer.
- Rutter M., Colvert E., Kreppner J., Beckett C., Castle J., Groothues C., Hawkins A., Stevens S.E., Sonuga-Barke E.J.S. (2007a) Early adolescent outcomes for institutionally-deprived and non-deprived adoptees. I: Disinhibited attachment.

- Journal of Child Psychology and Psychiatry*, **48**(1), 17–30.
- Rutter M., Le Couteur A., Lord C. (2003) *ADI-R: The Autism Diagnostic Interview – Revised*, Los Angeles, CA, Western Psychological Services, DOI: 10.1111/j.1469-7610.2006.01688.x
- Rutter M., Kreppner J., Croft C., Murin M., Colvert E., Beckett C., Castle J., Sonuga-Barke E.J.S. (2007b) Early adolescent outcomes of institutionally deprived and non-deprived adoptees. III. Quasi-autism. *The Journal of Child Psychology and Psychiatry*, **48**(12), 1200–1207, DOI: 10.1111/j.1469-7610.2007.01792.x
- Saint-Georges C., Cassel R.S., Cohen D., Chetouani M., Laznik M.C., Maestro S., Muratori F. (2010) What the literature on family home movies can teach us about the infancy of autistic children: a review of literature. *Research in Autism Spectrum Disorders*, **4**, 355–366.
- Schopler E., Reichler R.J., Renner B.R. (1988) *The Childhood Autism Rating Scale*, Los Angeles, CA, Western Psychological Services.
- Santarcangelo S., Dyer K. (1988) Prosodic aspects of motherese: effects on gaze and responsiveness in developmentally disabled children. *Journal of Experimental Child Psychology*, **46**, 406–418.
- Schuller B., Batliner A., Seppi D., Steidl S., Vogt T., Wagner J., Devillers L., Vidrascu L., Amir N., Kessou L., Aharonson V. (2007) The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functionals. *Proceedings of InterSpeech*, 2253–2256.
- Shami M., Verhelst W. (2007) An evaluation of the robustness of existing supervised machine learning approaches to the classification of emotions. *Speech Communication*, **49**(3), 201–212, DOI: 10.1016/j.specom.2007.01.006
- Short A.B., Schopler E. (1988) Factors relating to age of onset in autism. *Journal of Autism and Developmental Disorders*, **18**(2), 207–216, DOI: 10.1007/BF02211947
- Stern D.N., Jaffe J., Beebe B., Bennett S.L. (1975) Vocalization in unison and alternation: two modes of communication within the mother–infant dyad. *Annals of the New York Academy of Science*, **263**, 89–100, DOI: 10.1111/j.1749-6632.1975.tb41574.x
- Stone W.L., Hoffman E.L., Lewis S.E., Ousley O.Y. (1994) Early recognition of autism: parental reports vs. clinical observation. *Archives of Paediatrics and Adolescent Medicine*, **148**, 174–179.
- Sullivan A., Kelso J., Stewart M. (1990) Mothers' views on the ages of onset for four childhood disorders. *Child Psychiatry Human Development*, **20**, 269–278, DOI: 10.1007/BF00706019
- Sullivan M., Finelli J., Marvin A., Garrett-Mayer E., Bauman M., Landa R. (2007) Response to joint attention in toddlers at risk for autism spectrum disorder: a prospective study. *Journal of Autism and Developmental Disorders*, **37**, 37–48, DOI: 10.1007/s10803-006-0335-3
- Swettenham J., Baron-Cohen S., Charman T., Cox A., Baird G., Drew A., Rees L., Wheelwright S. (1998) The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *The Journal of Child Psychology and Psychiatry*, **39**(5), 747–753, DOI: 10.1111/1469-7610.00373
- Trainor L.J., Austin C.M., Desjardins R.N. (2000) Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, **11**, 188–195.
- Volkmar F.R., Pauls D. (2003) Autism. *Lancet*, **362**, 1133–1141, DOI: 10.1016/S0140-6736(03)14471-6
- Werner E., Dawson G., Osterling J., Dinno N. (2000) Recognition of autism spectrum disorder before one year of age: a retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders*, **30**, 157–162, DOI: 10.1023/A:1005463707029
- Werker J.F., McLeod P.J. (1989) Infant preference for both male and female infant-directed talk: a developmental study of attentional affective responsiveness. *Canadian Journal of Psychology*, **43**, 230–246, DOI: 10.1037/h0084224
- Zakian A., Malvy J., Desombre H., Roux S., Lenoir P. (2000) Signes précoces de l'autisme et films familiaux: une nouvelle étude par cotuteurs informés et non informés du diagnostic. *L'Encéphale*, **26**, 38–44.
- Zilbovicius M., Meresse I., Chabane N., Brunelle F., Samson Y., Boddaert N. (2006) Autism, the superior temporal sulcus and social perception. *Trends in Neurosciences*, **29**(7), 359–366, DOI:10.1016/j.tins.2006.06.004
- Zwaigenbaum L., Bryzon S., Rogers T., Roberts W., Brian J., Szatmari P. (2005) Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, **23**, 143–152, DOI:10.1016/j.ijdevneu.2004.05.001