

Automatic Motherese Detection for Parent-Infant Interaction

Ammar Mahdhaoui & Mohamed Chetouani

Université Pierre et Marie Curie Paris 6,
Institut des Systèmes Intelligents et de Robotique - ISIR
3 rue Galilée 94200 Ivry sur Seine
France

Ammar.Mahdhaoui@robot.jussieu.fr; Mohamed.Chetouani@upmc.fr

INTRODUCTION

During the first year of life, children with autism were distinguished from typical development children by poor social interaction (Zakian *et al.*, 2000), absence of social smiling, lack of appropriate facial expression (Adrien *et al.*, 1991), failure to orient to name. Autism manifestations are extremely variable, but infants inability to develop any sociability or to communicate with the entourage are in the core of this pathologic process.

In this study, we focus on verbal information which has been recently shown to be crucial for engaging interaction between the parent and infant and we investigated automatic motherese detection in home movies interaction. To complement this work, the engagement of interaction between the parent and the infant is analyzed. We manually identified and annotated sequences as engagement or not. The evaluation of the different annotations (two different psychologist) is done by using inter-coders agreement (kappa). One should note that this annotation is done independently from the motherese detection process. The key idea is to evaluate the number of sequences of interaction induced by the production of motherese.

Motherese (also termed infant-directed speech) and it is a simplified language/dialect/register (Fernald & Kuhl, 1987). From an acoustic point of view, motherese has a clear signature (high pitch, exaggerated intonation contours). The phonemes, and especially the vowels, are more clearly articulated. Motherese has been shown to be preferred by infants over adult-directed speech and might assist infants in learning speech sounds. The exaggerated patterns facilitate the discrimination between the phonemes or sounds. In addition, motherese plays a major role since it is a highly communicative and social event in parents child communication that can elicit emotional reactions. Even if motherese is clearly defined in terms of acoustic properties, the modeling and the detection is expected to be difficult which is the case of the majority of emotional speech. Indeed, the characterization of spontaneous and affective speech in terms of features is still an open question and several parameters have been proposed in the literature (Schuller *et al.*, 2007). As a starting-point and following the definition of motherese, we characterized the verbal interactions by the extraction of supra-segmental features (prosody). However, acoustic segmentation of home movies shows that segmental features play a major role for the robustness. Consequently, the utterances are characterized by both segmental (MFCC) and supra-segmental (statistics on fundamental frequency, energy and duration) features. These features are provided as input to a machine learning algorithm along with the known emotional labels of a training set of emotional utterances.

DATABASE

The output of the supervised learning phase is a classifier capable of distinguishing between the different emotional speech classes (motherese vs. non-motherese). The speech corpora used in this experiment is a collection of natural and spontaneous interactions. This corpora contain expressions of non-linguistic communication (affective intent) conveyed by a parent to a preverbal child. The corpus is a real interaction parents/child and consists of recordings in Italian of mother and father as they addressed their infants. In addition, the analysis of home movies makes it possible to set up a longitudinal study (months or years) and gives information about early behaviors of autistic infants, a long time before the diagnostic would be made by the clinicians. However, this large corpus makes it inconvenient for people to review it. Also, the recordings are not done by professionals resulting in adverse conditions (noise, camera, microphones...). We focus on one home video totaling 3 hours

during the first year of an infant. Verbal interactions of the mother have been carefully annotated by a psycholinguist on two categories: motherese and normal directed speech. From this manual annotation, we extracted 100 utterances for each class. The utterances are typically between 0.5s and 4s in length.

PRELIMINARY RESULTS

The performance of two learning algorithms GMM (Gaussian Mixture Models) and k-nn (k-nearest neighbors), is compared. To improve the detection, we investigate fusion schemes: segmental / supra-segmental, GMM / k-nn, the best approach yields an overall detection accuracy of 87.5%. Experimental result show that segmental features play a major role on the detection.

The second phase consists of the annotation of interaction in order to determine the relevance of this verbal information for the engagement of interaction. Obviously, the interaction may not be synchronized with the production of motherese. In addition, different verbal solicitations can produce an interaction. A sequence is annotated as interaction if the infant gives a feedback (vocalization, gaze, facial expression) following a verbal solicitation (motherese). The kappa's coefficient is estimated between two annotators among the previously defined sequences (47 sequences). We observed a good agreement ($\kappa=0.69$) for the characterization of the interaction. Approximately 78% of the segments of interaction have been induced by motherese. This preliminary result shows that the motherese seems to play a major role for the engagement of interaction between the parents and the infants who will become autistic. One should note that over verbal and/or non-verbal signals (gaze, facial expressions of the parents) may produce interactions and they should be considered for future works but the extraction and the characterization are more complicated.

REFERENCES

- Adrien, J. L., Faure, M., Perrot, A., Hameury, L., Garreau, B., Barthelemy, C. & Sauvage, D., 1991. Autism and family home movies: preliminary findings. *Journal of Autism and Developmental Disorders*, 21(1), 43-9.
- Fernald, A. & Kuhl, P., 1987. Acoustic determinants of infant preference for Motherese speech. *Infant Behavior and Development*, 10, 279-293.
- Schuller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., Kessous, 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*, pp. 2253-2256.
- 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*, XXVI, 38-44.