Stage de M2 (MIR / Reconnaissance de formes audio):
Audio data mining of a large corpus of human infant cries,
with applications to the developmental psychology of language.

Période: Février à Juillet 2017

Supervisors :

Jean-Julien Aucouturier ("Perception and Sound Design Team", IRCAM UMR9912), in collaboration with Kazuo Okanoya & Yulri Nonaka (University of Tokyo, Japan)

Context :

The internship is part of the project CREAM ("Cracking the Emotional Code of Music", http://cream.ircam.fr), supported by a Starting Grant from the European Research Council. The student will be integrated in the CREAM project, based within the "Perception and Sound Design" team, IRCAM (1 Place Stravinsky, 75004 Paris).

Project description :

During the first year of life, infants gradually acquire language by transitioning from gestures and prelinguistic vocalizations to referential speech. Baby cries even in the neonatal period are signals of intriguing acoustical complexity. Their communication bases are still poorly understood. On the one hand, the cry appears to be a reflexive graded signal that reflects certain states of the infant such as need or pain. On the other hand, increasing evidence shows that, as infants get older, the cry becomes a volitional categorical signal that conveys particular types of needs, such as hunger or loneliness (Okanoya, 2007)

This internship proposes an application of machine learning techniques to generate automatic descriptions from a large corpus of audio recordings of human baby cries, collected by our Japanese collaborators in University of Tokyo. Crying episodes were recorded daily from more than 20 infants, in their first 12 months of age, and annotated by the caregivers for their most likely context (hungry, sleepy, etc.).

The goal of the internship is to generate accurate acoustic metadata for each of the individual cry in the corpus (including whether the cry is an expiration or an inspiration, what is its loudness and pitch contour, as well as timbre descriptors such as roughness) and see how these acoustic characteristics can be used to predict annotated context for the cries. Following the internship, the corpus and its annotation will be published on a public directory to allow other researchers to test further scientific hypotheses about the development of the babies' linguistic abilities (see Goldstone & Lupyan, 2016).

Student role in the project :

The student will help develop audio machine learning algorithms to extract important elements of the cries:

- implement automatic expiration/inspiration segmentation (based on Aucouturier et al, 2011) as a public toolbox (matlab or python), and apply it to the whole database
- implement a pitch tracking algorithm adapted to the dataset (for instance, Camacho & Harris, 2008), and apply it to the whole database
- implement an automatic algorithm to cluster/recognize types of pitch contours (raising, falling, etc.), and apply it to the whole database
- format all the audio data as well as metadata (expiration, pitch, pitch contour) in a convenient format to share as a public dataset (as well as release the toolbox to do the analysis)
- help write a journal article describing the dataset (co-authored with the Japanese collaborators)

Student's profile :

We are looking for a Master's student with strong audio signal processing/machine learning/music information retrieval skills, as well as an interest in human communication and cognitive science. Programming proficiency with Matlab or (preferably) Python is required, as well as a solid understanding of audio pattern recognition and automatic sound description algorithms.

How to apply :

Send a CV and detailed cover letter by email to Jean-Julien AUCOUTURIER,
`aucouturier@gmail.com`

References :

Aucouturier, J. J., Nonaka, Y., Katahira, K., & Okanoya, K. (2011). Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. The Journal of the Acoustical Society of America, 130(5), 2969-2977.
A. Camacho & J. Harris, "A sawtooth-waveform inspired pitch estimator for speech and music," J. Acoust. Soc. Am., vol. 124, pp. 1638–1652, 2008.
Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring data sets. Topics in cognitive science, 8(3), 548-568.
K. Okanoya, "Language evolution as an emergent property," Curr. Opin. Neurobiol. 17, 271–276 (2007).