

Affective computing based on acoustic-prosodic cues

Introduction

- Describe the work developed by the L2F/INESC-ID Lisbon group on affective computing models based on acoustic-prosodic features for:
 - entrainment
 - emotions
 - personality traits
- Show contributions towards building bridges between speech and natural language processing, human-computer interaction, and linguistics

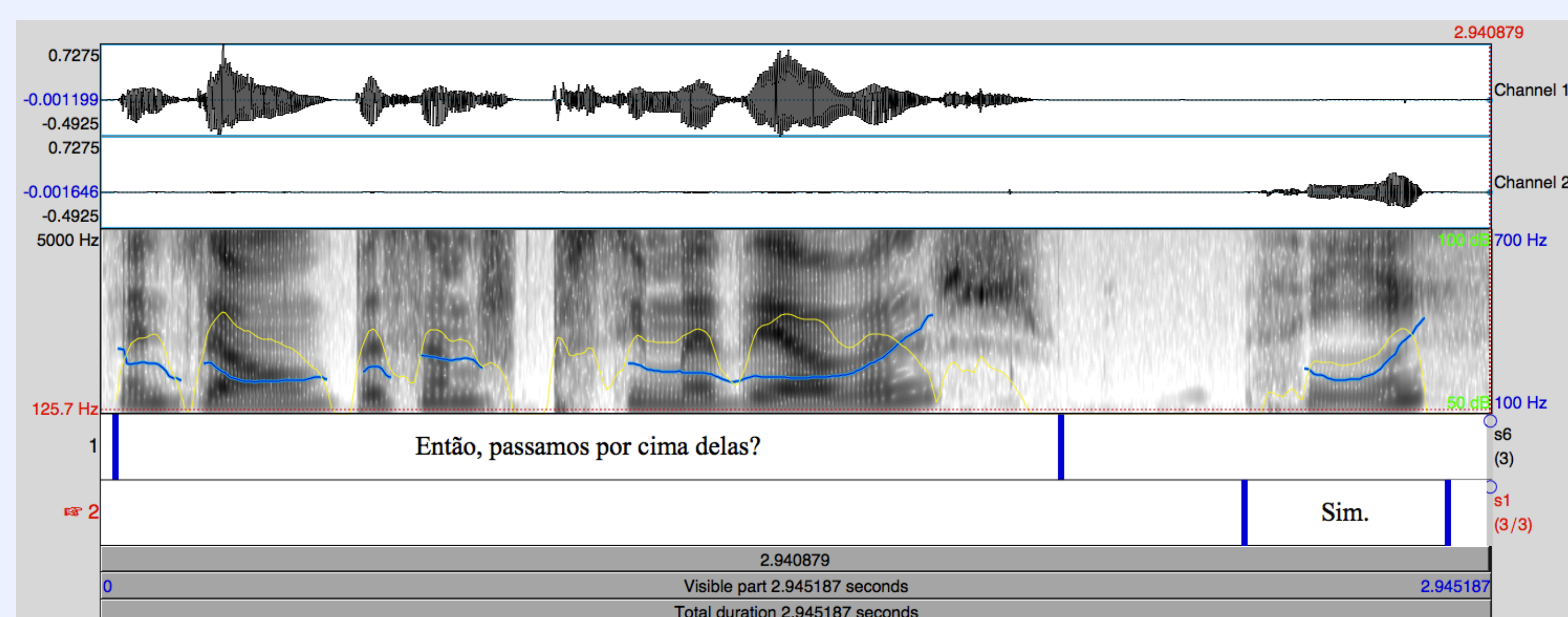
Entrainment

Goals

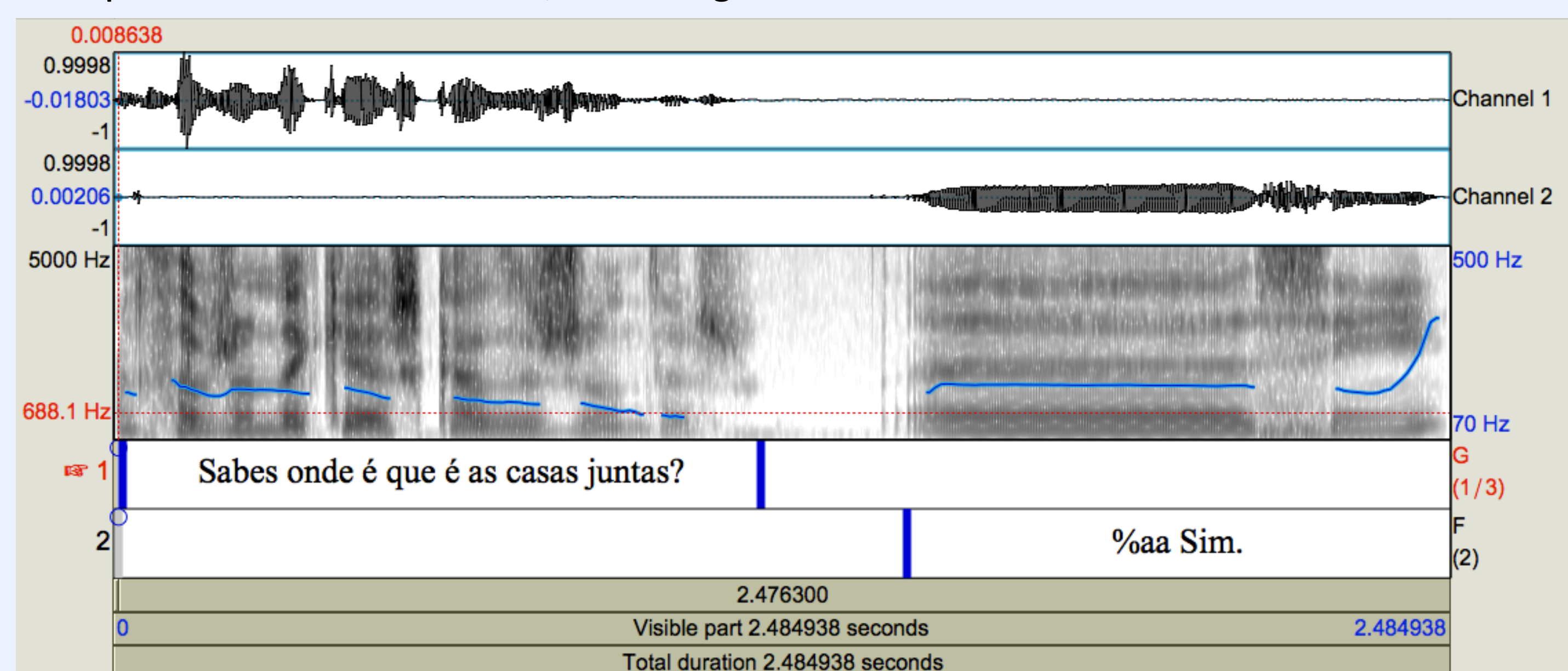
- Analyze acoustic-prosodic entrainment in spontaneous speech turn-by-turn
- Study entrainment with structural metadata events: sentence-like units, disfluencies, discourse markers, and affirmative cue words

Local entrainment

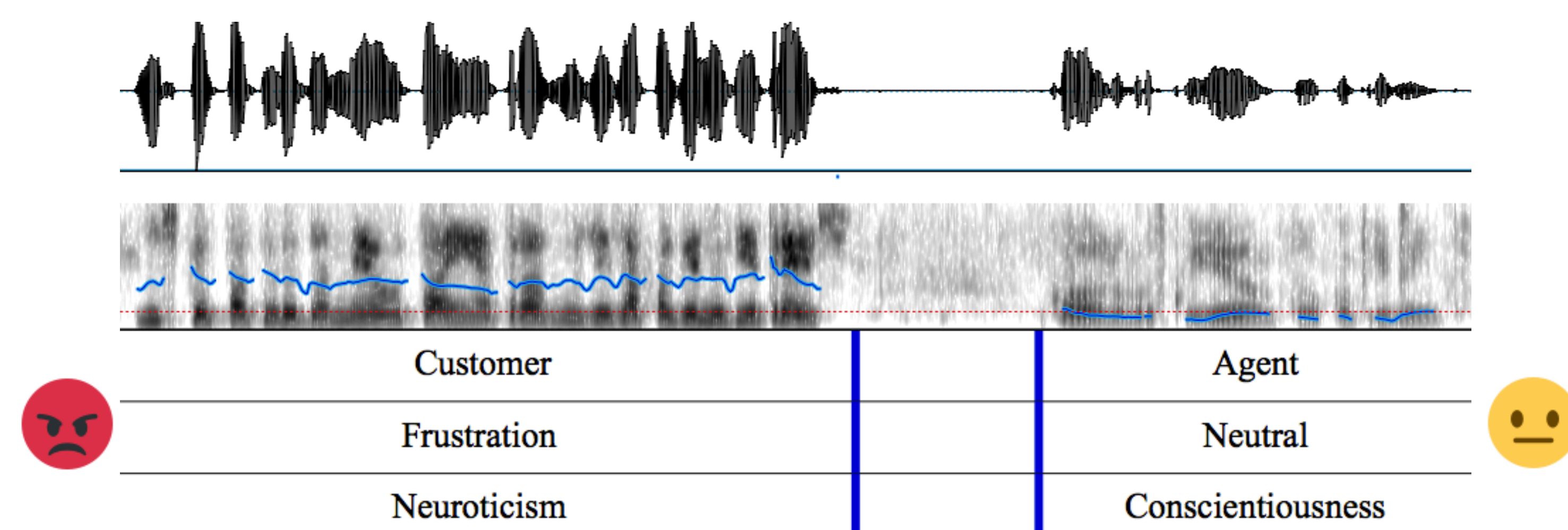
- Speakers match their interlocutors at turn-exchanges in pitch, energy, duration, and voice quality features
- Results not in line with English or Mandarin (Levitan, 2014): more entrainment in intensity features (mean, maxima), and HNR, but not in pitch



Example of entrainment: “So, we will go over them? Yes”



Example of non entrainment: “Do you know where the joint houses are? %aa Yes”



Future directions

- Enrich the acoustic-prosodic affective computing models with multimedia cues to model social behavior in virtual agents and robots
- Tackle the analysis of affective computing models in eHealth applications, an active area in our group
- Collect large in-the-wild multimodal datasets (vlogs) in which the speech of the subject is affected by certain medical conditions

Emotions

Goals

- Perform robust emotion classification in cross domains (acoustic environment and language)
- Assess emotional states of speakers during call center conversations

Approach

- eGeMAPS feature set (especially developed for affective computing)
- Different classifiers (neural networks) using different databases:
 - EMODB (Burkhardt et al., 2005) - German acted emotional speech (anger, boredom, disgust, anxiety/fear, happiness, sadness);
 - LetsGo (Schmitt et al., 2012) - English spontaneous speech (anger and neutral)
- Classifiers applied to Portuguese call center data with customer satisfaction evaluation

Results

- Comparable to the state-of-the-art (Accuracy of 73% on the test set)
- Success of cross-language and cross-domain knowledge transfer of emotional states

Personality

Goal

- Robust models for automatic perception of personality traits from speech

Approach

- Interpretable machine learning models
- Leverage available (very small and language specific) data sets:
 - Heterogeneous (age and language mismatch) → transfer learning
 - Partially non-labeled → semi-supervised training
- Knowledge-based acoustic-prosodic features
 - Personality-linked clues shared by speakers with different ages and languages
 - Pitch and energy ranges and dynamics, silence and phonation rates (IPU level)

Areas of research

- Perception of personality traits in Portuguese children/adults
 - Classmates dyadic interactions (with GAIPS/INESC-ID)
 - Autistic children-robot interactions (project INSIDE)
 - Affective analysis of customer service calls

