VOCE

Voice Coach for Reduced Stress

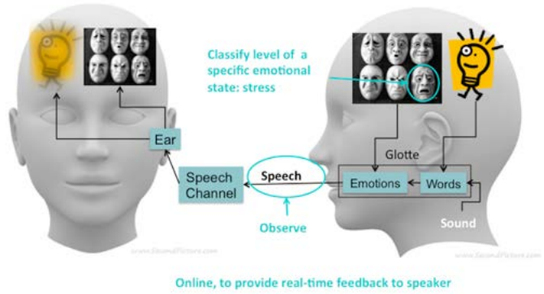

Fig. 1 Project overview.

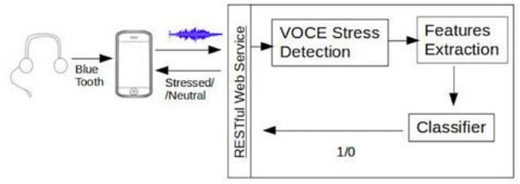

Fig. 2 Prototype overview.

VOCE developed methods and algorithms to enable the online classification of stress from live speech with the goal of providing feedback cues to the speaker. The work focused on building an annotated dataset, thereby developing a novel methodology for data collection and stress assessment, identifying relevant features for stress detection in speech, and building the corresponding classifier.

Main Project Team	
Ana Cristina Costa Aguiar	**NS-Po**
Jorge Miguel Barros da Silva	NS-Po
Mariana Dimas Julião	NS-Po
Mariana Henriques da Silva Kaiseler	NS-Po
Tânia Maria Pereira Lopes	NS-Po
Tiago Miguel Carmo Borba	NS-Po
Funding Agencies	
FCT - Fundação para a Ciência e a Tecnologia	**63,693€**
Start Date	01/03/2012
Ending Date	31/08/2015
Indicators	
Journal Papers	1
Conference papers	6
Concluded MSc:	4

Two Main Publications

A. Aguiar, M. Kaiseler, J. S. Silva, H. Meinedo, P. Almeida Almeida, M. D. Julião, **VOCE Corpus: Ecologically Collected Speech Annotated with Physiological and Psychological Stress Assessments,** ELRA International Conference on Language Resources and Evaluation - LREC, Reykjavik, Iceland, Vol. 978-2-9517408-8-4, pp. 1 - 6, May, 2014

M. D. Julião, J. S. Silva, A. Aguiar, H. R. Moniz, M. M. Batista, **Speech Features for Discriminating Stress Using Branch and Bound Wrapper Search**, Symposium on Languages, Applications and Technologies - SLATE, Madrid, Spain, Vol. 1, pp. 1 - 10, June, 2015

PROJECT WEBPAGE URL
http://paginas.fe.up.pt/~voce/

GENERAL MOTIVATION AND OBJECTIVES

Emotional stress is commonly experienced while speaking in public, producing changes to the various speech productions subsystems, affecting the speech signal in predictable ways and being easily conveyed to listeners. Speech stress indicators, however, are typically studied under laboratory settings, allowing little generalization to real life settings. VOCE's goal was to develop a system that would provide feedback cues based on stress detection to improve public speaking.

CHALLENGE

Collecting an annotated dataset proved to be the biggest challenge of all. Beyond collecting synchronised speech and physiologic signals of sufficient quality for automated processing outside of laboratory conditions, validating the method to detect stress from the collected physiologic signals proved to be extremely difficult and fuzzy. finally, time frames for speech processing and stress extraction are of different granularities, further increasing the difficulty of the task.

WORK DESCRIPTION AND ACHIEVEMENTS

We designed an interdisciplinary approach to assess speech stress during public speaking events, and implemented it on a platform that records speech simultaneously annotated with physiological and psychological measures. This approach enabled the collection of the VOCE corpus of annotated speech in ecological settings, i.e. in objectively stressing situations. We collected 43 samples, of which 32 are annotated with heart rate, 25 are annotated by experts, and 16 are annotated by RR sequences extracted from the electrocardiogram.

The collected recordings were sent to a backoffice that automated processing. We integrated a segmentation and feature extraction step made available by INESC-ID into the automatic backoffice processing for the automatic dataset generation, making all data available consistently to all project members at any time. The annotation of the collected speech with stress indicators, and specifically its validation, turned out to be the most challenging part of the work. We used various metrics extracted from physiologic sensors collected during the recordings, as well as annotations by experts, to perform annotation of the speech segmented into utterances, the chosen unit of work in the project. We carried out agreement and consistency tests among the different annotations, both for binary and 5-level stress annotation, and opted for binary annotation because the agreement among labelers is greater.

We carried out additional data gathering using out platform for ecologic momentary assessment during application of the Trier Social Stress Test (TSST) on 12 volunteers, with the goal of performing this validation for the heart measurement device used in out platform. We have been able to show that heart rate variability (HRV) measures can be obtained for short time windows of 50 s. Nevertheless, these short windows for HRV evaluation are still large for annotating speech, which uses processing time frames of a few seconds.

The classifier was chosen to be an SVM, as it has been identified in the literature as the best trade-off between robustness and required computing power. We used wrapper feature selection for searching for small feature subsets that are most useful to identify stress. Due to the large initial search space, we opted for parallelizing the wrapper, which is known to be a very computationally consuming task. The features that showed to perform better as stress discriminants were Audio Spectral, MFCC, PCM and TEO, and the size of the final feature sets was lower than 20 in all experiments.

Based on the results previously obtained, we created a demonstrator. A smartphone application collects the speech from microphone or Bluetooth headset, and sends it to a RESTful service that extracts features and classifies the speech segment, and returns the probability that it belongs to the "Stressed" class. This is used to provide feedback to the user in near real-time by varying the color of an indicator.