Inferring Perceptions of Control from Speech

Donal O’Sullivan¹, Rachel Msetfi², Pepijn Van de Ven¹, John Nelson¹

Electronic & Computer Engineering, University of Limerick, Psychology Department University of Limerick

Donal.osullivan@ul.ie

Keywords: Electronic and Computer Engineering; Biology & Life Sciences;

Abstract
Perceived Control or Perceptions of Control is defined as the belief that one can determine one’s own internal states and behavior, influence one’s environment, and/or bring about desired outcomes [1]. Individuals with a high Perceived Control are linked to positive mental health whereas individuals with a low Perceived Control are linked to depression and poor mental health. What we propose is to use a state-of-the-art Speech Emotion Recognition algorithm to infer Perceived Control from voice.

1. Introduction
Perceived control is a phenomenon whereby people think they have more control over events than they actually do [1]. This is linked to mental healthiness whereas realistic control perceptions are linked to depression. Perceived Control is usually tested using lab based tasks in which participants are repeatedly asked to perform some action and test whether or not it produces an outcome. The exact degree of control is of course programmed by the experimenter and can be compared to participant’s ratings of the same situation. Most experiments in this area are done in a controlled clinical environment where a subject is required to perform a task under supervision. However, results may not translate directly to a daily living scenario. Smartphones provide the ideal platform to present such experiments to the user in their natural environment as they are omnipresent and generally in close proximity to the subject throughout their day. The participants Perceived Control is measured via the participant providing a numerical rating of the amount of control they perceive they have over the experiment and the amount of control the environment has over the experiment. What we propose is to introduce a voice recorder to the experiment. This voice recorder will allow the gathering of voice data to determine if we can infer a participants Perceived Control from their voice signal.

2. Aim
The aim of this application is to provide a platform to gather voice data that will be used to determine if we can infer Perceived Control from a voice signal. A current Perceived Control experiment application will be used. The current application allows for the measuring of an individuals Perceived Control over an 8-hour period. A voice recorder will be added to the application to gather voice data periodically throughout the day. Once the data has been gathered a state-of-the-art machine learning algorithm will be used to model the speech signal. These models will then be used to classify the speech signals into High or Low Perceived Control.

3. Perception of Control App
To start an experiment, the experimenter will setup various parameters through a web interface. These include: how long the experiment runs for, the number of trials they would like to run during the experiment, the probabilities used in the experiment, the average time between each notification, and when to schedule the scoring system used by the participant to score how much control they felt they had. These data are stored on a server and synchronized with the participant’s phone. Trials are scheduled by the server and pushed to the user’s device; from here each trial is presented to the user as a notification. Once the notification appears, the participant has two minutes to start the trial. Initially the trial screen is blank for 3 seconds after which a button appears on the screen for 2 seconds. If the participant presses the button in time they are rewarded with a chime. However, the provision of rewards is provided with a certain probability set by the experimenter. For example, when the button is pressed it will generate a sound only 75% of the time. If the user does not press the button there is also a certain probability that the trial will generate a chime regardless. These trials will be performed throughout the participant’s day with the scheduling of the notifications being done via a variable interval schedule [2]. This allows for a varying time between each notification so the participant does not learn the pattern or time between each notification. After a specified number of trials, the participant is asked to rate how much control they felt they had over the trial generating a chime and how much control their environment had over the trial making a chime. These ratings will be numerical ratings of control, also the participant will provide a vocal rating of control by speaking into the phone.

4. Experiment Overview
The experiment is run over an 8-hour period from start to finish. The experiment is broken down into blocks. Within each block the participant is shown 8 trials and 1 judgement. Each trial involves the interaction with a button explained above. Each judgement involves the participant providing two numerical Judgement scores. Also the participant will be asked to speak into the phone during the judgement
5. System Overview

To infer Perceived Control from voice a statistical classifier will be employed. To ensure the correct methodology is used a current Speech Emotion Recognition System developed by [2] will be used. The classifier used in this system demonstrates current state-of-the-art in Speech Emotion Recognition. To use the system a number of steps are required. The first step involves feature extraction. The feature extraction step involves collecting data from the speech signals to create the Hidden Markov Models (HMM). The next step involves training a number of HMMs. The HMMs will be developed in Matlab using the PMTK3 tool kit. The figure 2 demonstrates the system overview.

6. Recording of Speech

The recording of the Perceived Control speech features will be done at five separate points during the experiment. The participant will be asked to speak into the phone as they are providing their control rating. The length of the audio recording will not exceed 10 seconds per speech occurrence. The audio recording will use a skip silence feature that will only record when the participant is speaking. The utterances will be recorded on an Android Smart phone using the WAV format with 16-bit PCM (Pulse Code Modulation) and a sampling frequency of 44.1 kHz. WAV (Waveform Audio File Format) is an audio format used for storing uncompressed raw audio.

7. Audio Signal Processing

Once the speech data has been gathered the initial stage involves extracting the features that will be used in the statistical classifier. Speech production in humans is generated by the source filter model. The source being the vocal cords. And the filter is the shape of the vocal tract which include the tongue and teeth. The shape of the vocal tract determines the sound that comes out. The speech feature we are using models the envelope of the short term power spectrum. These features that will be used are Mel Frequency Cepstrum Coefficients (MFCC) [3]. The MFCC are considered the state-of-art speech feature for use in Speech Recognition systems. The details of the MFCC extraction can be found in [3].

8. HMM Classification

A current HMM classifier has been implemented according to [2] with the same results. This works on speech emotion recognition but provides a suitable working methodology to follow. The HMMs are a popular classifier for speech recognition because of their ability to model time series data [4]. This is the reason for adopting them for speech emotion recognition and the reason we are adopting them for Perceived Control recognition. A HMM is a doubly embedded stochastic process where each observation is a probabilistic function of the state. The underlying stochastic process is hidden but can be observed through a set of stochastic processes that produce the sequence of observations. There are two main types of HMM that are used. The first is the Ergodic HMM where every state is reachable from another state. The second is the left-right HMM where the state sequence progresses in a left to right nature. Finally, there are two ways in which HMMs can represent their probability distributions, these are the discrete HMM and the continuous HMM. Discrete HMMs are used when the observations are a set of symbols chosen from a finite alphabet, the probability density function (PDF) is represented by a discrete set of variables. The continuous HMM are used when the observations are a set of continuous variables. The continuous PDF is usually represented by a Gaussian distribution. Normally the continuous HMM uses a mixture of Gaussians, which is defined as a linear combination of Gaussian distributions. A HMM will be created for each Perceived Control state. At the moment this is down to three different states. Perceived Control, No Control, and Preventative Control (meaning they are preventing something from happening). For the sake of the Perceived Control classification a 4 state HMM will be used with an Ergodic topology.

9. Conclusion

Once the data has been gathered we hope to be able to correlate the numerical ratings to the HMMs created for different Perceptions of Control thus demonstrating the use of current Speech Emotion Recognition algorithm can indeed infer Perceived Control from voice.

10. References
