Automatic Detection of Apathy using Acoustic Markers extracted from Free Emotional Speech

Nicklas Linz¹, Xenia Klinge¹, Johannes Tröger¹, Jan Alexandersson¹, Philippe Robert², Alexandra König²

¹ German Research Center for Artificial Intelligence (DFKI)
² Memory Center, CoBTeK - IA CHU Université Côte d’Azur
nicklas.linz@dfki.de

Abstract

Apathy is a frequent neuropsychiatric syndrome in people with dementia. It leads to diminished motivation for physical, cognitive and emotional activity. Apathy is highly underdiagnosed since its criteria have been only recently established and rely heavily on the subjective evaluation of human observers. In this paper we analyse speech samples from demented people with and without apathy. Speech was provoked by asking patients two emotional questions. Acoustic features were extracted and used in a classification task. The resulting models show performances of $AUC = 0.71$ and $AUC = 0.63$. This is a decent first step into the direction of automatic detection of apathy from speech. Usefulness of stimuli to elicit free speech is found to depend on patients gender.

1 Introduction

Apathy is a neuropsychiatric syndrome that expresses itself in multiple domains: loss of interest, emotional blunting and diminished goal directed behaviour [Marin, 1991]. It is associated with a variety of neurodegenerative diseases, such as Alzheimer’s disease (AD), Parkinson’s disease or even Mild Cognitive Impairment (MCI) [Di Iulio et al., 2010]. Apathy is present in nearly 65% of dementia cases [Aalten et al., 2007; Robert et al., 2005] and has a negative predictive role for disease course [Stella et al., 2015], as well as a strong impact on the quality of life of patients and their caregivers [Hurt et al., 2008].

Diagnosis of apathy is usually conducted through clinical interviews and rating scales [Robert et al., 2002; Sockeel et al., 2006; Starkstein et al., 1992], which are limited because of their dependency on human observers. Apathy is often misdiagnosed, since characteristics (i.e. diminished interest and psychomotor retardation) overlap with those of other neuropsychiatric syndromes, such as depression [Yeager and Hyer, 2008]. Albeit, correct and early diagnosis of apathy is important, as e.g., in patients with MCI, a possible predecessor of AD, apathy’s ‘lack of interest’ domain has been shown to be the strongest predictor of conversion to AD [Robert et al., 2008]. Consequently, additional systematic and objective assessment tools are needed [König et al., 2014].

Automatic speech analysis (ASA) in combination with machine learning (ML) have been shown to effectively predict people with other neuropsychiatric syndromes, such as depression [Cummins et al., 2015b; Asgari et al., 2014]. Markers automatically computed from speech are objective and can be collected unobtrusively, rendering it a potential diagnostic tool.

This paper investigates the possibility of using acoustic markers extracted from free emotional speech to automatically classify people as having apathy. A short introduction to related speech processing research is given (Section 2), the experiment set-up is described (Section 3) and results and their implications are discussed (Section 4).

2 Related Work

Little to no information and communication technologies (ICT) have been previously applied in the assessment of apathy—speech has never been used. [König et al., 2014] performed a review of ICT for the assessment of apathy and concluded that no one had previously used ICT specifically in this context, but that techniques seemed promising. Since then, [Manera et al., 2015] evaluated a serious game with dementia patients showing signs of apathy. They found that patients with apathy played longer than non-apathetic patients, while they found no difference in the number of scenarios played. Since apathy seems to affect emotion-based decision making, other attempts to measure it have been made, such as with the Iowa gambling task [Bayard et al., 2014] or the Philadelphia Apathy Computerized Task (PACT) [Fitts et al., 2016] detecting impairments in goal-directed behavior including initiation, planning, and motivation.

2.1 Speech Analysis in Depression

A large body of research validates the use of speech in the assessment of depression. As a symptom, apathy has an association with depression in the context of neurodegenerative diseases [Levy et al., 1998]. Depression however, is rather expressed as negative affect, whereas, apathy is observed as emotional neutrality, where neither positive nor negative
emotions are observed. Deficits in ‘auto-activation’ and the cognitive domain seem common in both and therefore results from previous ASA studies on depression may be generalisable to apathy.

Previously, [Cummins et al., 2015b] investigated the effects of depression in speech manifesting as a reduction in the spread of phonetic variability in acoustic space. They analyse Average Weighted Variance (AWV), Acoustic Movement (AM) and Acoustic Volume (AV) and conclude that depressed people show significant reductions in all. [Asgari et al., 2014] used speech features—including jitter and shimmer, harmonic to noise ratio (HNR) and mel frequency cepstral coefficients (MFCC)—and language features extracted from natural conversation to detect depression. Speech features alone performed better than only language features. The best performance of 74% accuracy was reached with a combination of speech and language features alone performed better than only language features. The spread of phonetic variability in acoustic space. They analyse Average Weighted Variance (AWV), Acoustic Movement (AM) and Acoustic Volume (AV) and conclude that depressed people show significant reductions in all. [Asgari et al., 2014] used speech features—including jitter and shimmer, harmonic to noise ratio (HNR) and mel frequency cepstral coefficients (MFCC)—and language features extracted from natural conversation to detect depression. Speech features alone performed better than only language features. The best performance of 74% accuracy was reached with a combination of speech and language features. [Alghowinem et al., 2016] examined German and English speech data of depressed patients from three different corpora. They extracted vocal markers, such as fundamental frequencies (F0), energy, intensity, loudness, jitter, shimmer, HNR and MFCCs, and built classifiers to evaluate single resources and their combinations. They achieve 97% accuracy for one and 82% for two other corpora. [Mundt et al., 2012] elicited speech from 105 adults with major depression in a free speech, counting, reading and a sustained vowel task. They extracted fundamental frequencies (F0), first and second formants and features relating to the duration and proportion of silences and vocalisations. All features relating to silences, pauses and vocalisations were significantly different between the groups.

In general, speech analysis has found great applicability to either screen for or to compute robust and objective metrics for depression. We hypothesise that due to the above mentioned similarities some of the same features will show merit.

3 Methods

To provide evidence for the potential of ASA in apathy assessment, we recorded demented patients with and without apathy, extracted acoustic features from the speech signal and built, as well as evaluated, ML classifiers.

3.1 Data

Speech recordings from both the Dem@Care [Karakostas et al., 2014] and the ELEMENT [Tröger et al., 2017] projects were used. All participants were aged 65 or older and were recruited through the Memory Clinic located at the Institute Claude Pompidou in the Nice University Hospital. Speech recordings were collected using an automated recording app on a tablet computer.

To elicit free emotional speech, people were asked to perform two tasks: (1) talk about a positive event in their live and (2) to talk about a negative event in their live. Instructions were prerecorded to guarantee a standardised assessment.

Participants also completed a battery of cognitive tests, the MMSE [Folstein et al., 1975] and the Apathy Inventory (AI) [Robert et al., 2002]. Participants were excluded if they had any major auditory or language problems, history of head trauma, loss of consciousness, or psychotic or aberrant motor behaviour. Following the clinical assessment, patients were grouped into three categories in accordance with the DSM-V diagnostic guide: patients without any impairment, minor impairment or major impairment. In this study we only look at patients with either minor or major impairments, to prevent confounding of group differences by cognitive state. Males and females are treated separately to account for differences in acoustic features and anticipate differences in effects of apathy [Cummins et al., 2017]. Patients are split into groups according to their AI score (≥ 4) and groups are matched for MMSE. Demographic data and clinical test results by diagnostic groups are reported in Table 1.

3.2 Features

Multiple features were extracted, some due to their previous success in detection of depression [Cummins et al., 2015a] and the overlap of symptoms in free speech between both disorders, others encode task specific performance relating to diminished goal directed behaviour as examined in apathy.

We extract statistics relating to lengths of silence and sounding segments, determined based on intensity, calculated from the bandpass filtered sound signal, statistics relating to the audible pitch, in the form of fundamental frequency (F0), speech tempo, approximated using syllable nuclei [De Jong and Wempe, 2009], as provided by the Praat software [Boersma and Weenink, 2001]. Micro level variations in amplitude and period—jitter and shimmer—were determined using the openSmile software [Eyben et al., 2013]. A Matlab [MATLAB, 2010] script was used to compute Harmonic-to-Noise-Ratio (HNR) and statistics over the first three formants.

3.3 Classification

We construct ML models to verify the predictive power of the extracted features to classify between people with and without apathy. All features were normalised using z-standardisation. As classifiers, Support Vector Machines (SVMs) implemented in the scikit-learn framework [Pedregosa et al., 2011] were used. To evaluate the performance of the model on such a small dataset we rely on Leave-One-Out cross validation. As a performance metric we report Area Under the Curve (AUC).

Table 1: Demographic data of patients used in experiments. Statistically significant group differences from the control group inside a gender, based on a Mann-Witney-U test (p < 0.01) are indicated by ∗.

<table>
<thead>
<tr>
<th>Male</th>
<th>No Apathy</th>
<th>Female</th>
<th>No Apathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>25</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>Age</td>
<td>77.6 (6.8)</td>
<td>78.8 (6.2)</td>
<td>78.8 (6.5)</td>
</tr>
<tr>
<td>MMSE</td>
<td>21.9 (4.3)</td>
<td>19.1 (3.7)</td>
<td>20.1 (3.8)</td>
</tr>
<tr>
<td>AI</td>
<td>1.68 (1.6)</td>
<td>5.5∗ (1.6)</td>
<td>1.61 (1.7)</td>
</tr>
</tbody>
</table>
4 Results and Discussion

Classification results are reported in Figure 1. Results differ depending on the origin of used features. In the male population, classification results improve significantly from an AUC of 0.51 to 0.63 when using features from the negative story in contrast to the positive one. The female population shows the opposite behaviour with an increase in AUC from 0.50 to 0.71 switching from the negative to the positive task. When using features from both positive and negative stories, both male and female populations show worse performance compared to their baselines, with an AUC of 0.70 and 0.53 respectively.

The classification results are a promising first step showing that speech features clearly contain information relating to apathy and could therefore be used in its assessment. As anticipated, different patterns for males and females emerge. Classifiers trained on features from the negative story show superior performance for the male population, classifiers built on features from the positive one for the female population. We are unaware of any work on gender dependent symptoms of apathy that could explain this pattern. Parts of this effect could be explained by the fact that men from this generation are in general less likely to talk enthusiastically about a positive event and show greater responses to threatening cues [Kret and De Gelder, 2012]. Sex differences in emotional processing and memory retrieval could be another reason and should be further investigated, since current literature mostly focuses on exploring age as a variable. We conclude that ASA has the potential to be useful in the assessment of apathy, that the type of stimulus speech is being provoked with might play a major role and might have to be adapted depending on a patient’s gender.

Further work should examine what features in particular are predictive for apathy, how they relate to depression and how the two could be discriminated. Since patient data is always hard to acquire, our sample is relatively small and future studies should strive to draw more conclusive evidence from larger datasets.

References


