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

Nicklas Linz[1], Xenia Klinge[1], Johannes Tröger[1], Jan Alexandersson[1], Philippe Robert[2], Alexandra König[2]
[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

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 $*$.

	Male		Female	
	No	**Apathy**	**No**	**Apathy**
N	25	32	38	23
Age	77.6 (6.8)	78.8 (6.2)	78.8 (6.5)	79.1 (6.0)
MMSE	21.9 (4.3)	19.1 (3.7)	20.1 (3.8)	17.9 (5.0)
AI	1.68 (1.6)	5.5* (1.6)	1.61 (1.7)	5.0* (1.8)

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. [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 accoustic 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 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).

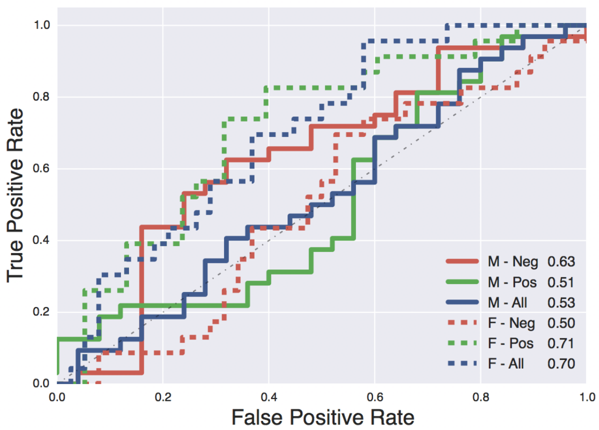

Figure 1: Receiver Operator Curves (ROC) of classification experiments. Color coding and AUC is reported in legend. 'M'=Male; 'F'=Female; 'Neg'=Features from negative story; 'Pos'=Features from positive story; 'All'=Features from both positive and negative story

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 patients' 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

[Aalten *et al.*, 2007] P. Aalten, F. R. Verhey, M. Boziki, R. Bullock, E. J. Byrne, V. Camus, M. Caputo, D. Collins, P. P. De Deyn, K. Elina, G. Frisoni, N. Girtler, C. Holmes, C. Hurt, A. Marriott, P. Mecocci, F. Nobili, P. J. Ousset, E. Reynish, E. Salmon, M. Tsolaki, B. Vellas, and P. H. Robert. Neuropsychiatric syndromes in dementia. Results from the European Alzheimer Disease Consortium: part I. *Dement Geriatr Cogn Disord*, 24(6):457–463, 2007.

[Alghowinem *et al.*, 2016] Sharifa Alghowinem, Roland Goecke, Julien Epps, Michael Wagner, and Jeffrey F Cohn. Cross-cultural depression recognition from vocal biomarkers. In *INTERSPEECH*, pages 1943–1947, 2016.

[Asgari *et al.*, 2014] Meysam Asgari, Izhak Shafran, and Lisa B Sheeber. Inferring clinical depression from speech and spoken utterances. In *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, pages 1–5. IEEE, 2014.

[Bayard *et al.*, 2014] Sophie Bayard, Jean-Pierre Jacus, Stéphane Raffard, and Marie-Christine Gely-Nargeot. Apathy and emotion-based decision-making in amnesic mild cognitive impairment and alzheimer's disease. *Behavioural neurology*, 2014, 2014.

[Boersma and Weenink, 2001] Paul Boersma and David Weenink. Praat, a system for doing phonetics by computer. *Glot international*, 5:341–345, 2001.

[Cummins *et al.*, 2015a] Nicholas Cummins, Stefan Scherer, Jarek Krajewski, Sebastian Schnieder, Julien Epps, and Thomas F Quatieri. A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, 71:10–49, 2015.

[Cummins *et al.*, 2015b] Nicholas Cummins, Vidhyasaharan Sethu, Julien Epps, Sebastian Schnieder, and Jarek Krajewski. Analysis of acoustic space variability in speech affected by depression. *Speech Communication*, 75:27–49, 2015.

[Cummins *et al.*, 2017] Nicholas Cummins, Bogdan Vlasenko, Hesam Sagha, and Björn Schuller. Enhancing speech-based depression detection through gender dependent vowel-level formant features. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 209–214. Springer, 2017.

[De Jong and Wempe, 2009] Nivja H De Jong and Ton Wempe. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, 41(2):385–390, 2009.

[Di Iulio *et al.*, 2010] Fulvia Di Iulio, Katie Palmer, Carlo Blundo, Anna Rosa Casini, Walter Gianni, Carlo Caltagirone, and Gianfranco Spalletta. Occurrence of neuropsychiatric symptoms and psychiatric disorders in mild alzheimer's disease and mild cognitive impairment subtypes. *International Psychogeriatrics*, 22(4):629–640, 2010.

[Eyben *et al.*, 2013] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 835–838, New York, NY, USA, 2013. ACM.

[Fitts *et al.*, 2016] Whitney Fitts, Lauren Massimo, Nicholas Lim, Murray Grossman, and Nabila Dahodwala. Computerized assessment of goal-directed behavior in parkinson's disease. *Journal of clinical and experimental neuropsychology*, 38(9):1015–1025, 2016.

[Folstein *et al.*, 1975] M. F. Folstein, S. E. Folstein, and P. R. McHugh. "Mini-Mental State". A Practical Method for Grading the Cognitive State of Patients for the Clinician. *J Psychiatr Res*, 12(3):189–198, 1975.

[Hurt *et al.*, 2008] Catherine Hurt, Sarmishtha Bhattacharyya, Alistair Burns, Vincent Camus, Rossella Liperoti, Anna Marriott, Flavio Nobili, Philippe Robert, Magda Tsolaki, Bruno Vellas, et al. Patient and caregiver perspectives of quality of life in dementia. *Dementia and geriatric cognitive disorders*, 26(2):138–146, 2008.

[Karakostas *et al.*, 2014] Anastasios Karakostas, Alexia Briassouli, Konstantinos Avgerinakis, Ioannis Kompatsiaris, and Magda Tsolaki. The Dem@Care Experiments and Datasets: a Technical Report. Technical report, Centre for Research and Technology Hellas (CERTH), 2014.

[König *et al.*, 2014] Alexandra König, Pauline Aalten, Frans Verhey, Gregory Bensadoun, Pierre-David Petit, Philippe Robert, and Renaud David. A review of current information and communication technologies: can they be used to assess apathy? *International journal of geriatric psychiatry*, 29(4):345–358, 2014.

[Kret and De Gelder, 2012] Mariska E Kret and Beatrice De Gelder. A review on sex differences in processing emotional signals. *Neuropsychologia*, 50(7):1211–1221, 2012.

[Levy *et al.*, 1998] Morgan L Levy, Jeffrey L Cummings, Lynn A Fairbanks, Donna Masterman, Bruce L Miller, Anne H Craig, Jane S Paulsen, and Irene Litvan. Apathy is not depression. *The Journal of neuropsychiatry and clinical neurosciences*, 10(3):314–319, 1998.

[Manera *et al.*, 2015] Valeria Manera, Pierre-David Petit, Alexandre Derreumaux, Ivan Orvieto, Matteo Romagnoli, Graham Lyttle, Renaud David, and Philippe H Robert. 'kitchen and cooking,'a serious game for mild cognitive impairment and alzheimer's disease: a pilot study. *Frontiers in Aging Neuroscience*, 7:24, 2015.

[Marin, 1991] R. S. Marin. Apathy: a neuropsychiatric syndrome. *J Neuropsychiatry Clin Neurosci*, 3(3):243–254, 1991.

[MATLAB, 2010] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[Mundt *et al.*, 2012] James C Mundt, Adam P Vogel, Douglas E Feltner, and William R Lenderking. Vocal acoustic biomarkers of depression severity and treatment response. *Biological psychiatry*, 72(7):580–587, 2012.

[Pedregosa *et al.*, 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Robert *et al.*, 2002] P. H. Robert, S. Clairet, M. Benoit, J. Koutaich, C. Bertogliati, O. Tible, H. Caci, M. Borg, P. Brocker, and P. Bedoucha. The apathy inventory: assessment of apathy and awareness in Alzheimer's disease, Parkinson's disease and mild cognitive impairment. *Int J Geriatr Psychiatry*, 17(12):1099–1105, Dec 2002.

[Robert *et al.*, 2005] Philippe H Robert, Frans RJ Verhey, E Jane Byrne, Catherine Hurt, Peter Paul De Deyn, Flavio Nobili, Roberta Riello, Guido Rodriguez, Giovanni B Frisoni, Magda Tsolaki, et al. Grouping for behavioral and psychological symptoms in dementia: clinical and biological aspects. consensus paper of the european alzheimer disease consortium. *European Psychiatry*, 20(7):490–496, 2005.

[Robert *et al.*, 2008] Philippe H Robert, Claudine Berr, Magali Volteau, Christelle Bertogliati-Fileau, Michel Benoit, Olivier Guerin, Marie Sarazin, Sylvie Legrain, and Bruno Dubois. Importance of lack of interest in patients with mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, 16(9):770–776, 2008.

[Sockeel *et al.*, 2006] P Sockeel, K Dujardin, D Devos, C Deneve, A Destée, and L Defebvre. The lille apathy rating scale (lars), a new instrument for detecting and quantifying apathy: validation in parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 77(5):579–584, 2006.

[Starkstein *et al.*, 1992] Sergio E Starkstein, Helen S Mayberg, ThomasJ Preziosi, Paula Andrezejewski, R Leiguarda, RG Robinson, et al. Reliability, validity, and clinical correlates of apathy in parkinson's disease. *J Neuropsychiatry Clin Neurosci*, 4(2):134–139, 1992.

[Stella *et al.*, 2015] Florindo Stella, Orestes Vicente Forlenza, Jerson Laks, Larissa Pires Andrade, João Castilho Cação, José Sílvio Govone, Kate Medeiros, and Constantine G Lyketsos. Caregiver report versus clinician impression: disagreements in rating neuropsychiatric symptoms in alzheimer's disease patients. *International journal of geriatric psychiatry*, 30(12):1230–1237, 2015.

[Tröger *et al.*, 2017] Johannes Tröger, Nicklas Linz, Jan Alexandersson, Alexandra König, and Philippe Robert. Automated Speech-based Screening for Alzheimer's Disease in a Care Service Scenario. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2017.

[Yeager and Hyer, 2008] Catherine A Yeager and LEE Hyer. Apathy in dementia: relations with depression, functional competence, and quality of life. *Psychological reports*, 102(3):718–722, 2008.