Affective Speech for Alzheimer's Dementia Recognition.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
LREC: Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various
forms of cognitive/psychiatric/developmental impairments (RaPID)

Affective Speech for Alzheimer's Dementia Recognition

Fasih Haider, Sofia de la Fuente, Pierre Albert, Saturnino Luz

Usher Institute, Edinburgh Medical School
the University of Edinburgh
{Fasih.Haider, sofia.delafuente, pierre.albert, S.Luz}@ed.ac.uk

Abstract

Affective behaviour could provide an indicator of Alzheimer's disease and help develop clinical tools for automatically detecting and monitoring disease progression. In this paper, we present a study of the predictive value of emotional behaviour features automatically extracted from spontaneous speech using an affect recognition system for Alzheimer's dementia detection. The effectiveness of affective behaviour features for Alzheimer's Disease detection was assessed on a gender and age balanced subset of the *Pitt Corpus*, a spontaneous speech database from DementiaBank. The affect recognition system was trained using the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) and the Berlin database of emotional speech. The output of this system provides classification scores or class posterior probabilities of 6+1 emotions as an input for statistical analysis and Alzheimer's dementia detection. The statistical analysis shows that the non-AD subjects have higher mean value of classification scores for *anger* and *disgust*, along with a higher entropy of classification scores than AD subjects. The AD subjects have a higher classification scores for the *sad* emotional behaviour than non-AD. This paper also introduces a novel 'affective behaviour representation' feature vector for Alzheimer's dementia recognition. Results show that classification models based solely on affective behaviour attain 63.42% detection accuracy.

Keywords: Affective Computing, Social Signal Processing, Dementia, Alzheimer, Cognitive Decline Detection, Cognitive Impairment Detection

1. Introduction

Dementia is a a category of neurodegenerative diseases characterised by long-term and usually gradual decrease of cognitive functioning (American Psychiatric Association, 2000). Whilst memory loss is frequently considered the most prominent symptom of dementia, in particular of Dementia of the Alzheimer Type (DAT), speech and language alterations are also common (Kirshner, 2012). For instance, word-finding difficulties (i.e. anomia) are reported from early stages of cognitive impairment, when patients describe how they can see certain words "floating in front of them", although they do not manage to "catch" them in order to put them in a sentence. Literature also suggests that patients with DAT have difficulty accessing semantic information when they intend to do so (Bondi et al., 1996). Since successful communication is essential for meaningful social interaction, this takes a toll on patients' and their carers' wellbeing.

It is presumed that such difficulties increase the level of frustration and have an impact on the emotional life of these patients. The prevalence of apathy, dysphoria and depression in Alzheimer's Disease (AD) increases with the severity of the condition (Landes et al., 2005). In fact, the overlap between apathy and depression becomes particularly prominent in this clinical population (Starkstein et al., 2005). These comorbidities are relatively well established and have spurred research on differential diagnosis between dementia and depression (Leyhe et al., 2017).

One of the reasons suggested in an attempt to explain the comorbidity between dementia and depression is the role of emotions in memory encoding, since both conditions progress with an increased forgetfulness (Hart et al., 1987). Emotional abilities, amongst which are the expression of our own emotions as well as the recognition of those of others, are also decisive for social communication (Lopes et al., 2004). Emotional information can be conveyed in different ways, from explicit facial and verbal expression (e.g. smile, pout, happy statement) to more subtle non-verbal cues, such as intonation, modulation of vocal pitch and loudness of emotional expression. These non-verbal cues are generally referred to as emotional prosody. Both expression and recognition of emotional prosody seem to be impaired in DAT (Horley et al., 2010), though the latter has been more widely studied.

Research on computational speech technology to better characterise emotional prosody could shed a light on the expression of emotions in people with DAT. This study aims to apply a signal processing model for recognition of emotional prosody in DAT with two main objectives. First, we wish to determine whether certain emotions are predominant in DAT whilst others are subdued. With this purpose, we train an emotion recognition model on a high quality dataset of emotion expression, and then use this model to classify speech segments of a dataset containing speech of AD and non-AD participants into 6+1 emotional state labels. Second, once the distribution of emotions across each audio recording is established (i.e. classification scores or class posterior probabilities), this information will be used as an input for a classifier, aimed at automatic detection of DAT based on emotional prosody, as shown in Figure 1.

2. Related work

As far as research on emotions is concerned, the most common paradigms tend to rely on facial expression and image processing (e.g. (Seidl et al., 2012)), with less published work on prosody and other linguistic features. However, there is quantifiable evidence that acoustic analysis can give an account of emotional expression. For instance, sadness is associated with lower speech rate and lower mean fundamental frequency (F_0) than emotions such as happiness, fear or anger (Juslin and Laukka, 2003).

—Audio ▶ | Speech Segmentation | → | Volume Normalization | → | Feature Extraction | → | Affect recognition | → | AD/nonAD Recogntion | → AD/nonAD
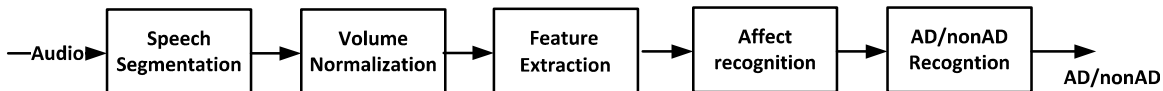
Figure 1: Pipeline for dementia recognition through Affective/Emotional Behaviour.

Previous research on DAT and emotional prosody has predominantly focused on recognition (*receptive emotional prosody*), as opposed to expression (*expressive emotional prosody*). Findings in both areas have yielded promising but as yet inconclusive results. For instance, research findings pointed at impaired emotional processing in DAT, though still relatively preserved in comparison to cognitive abilities, suggesting that impairment of emotional prosody might be secondary to the decline of another cognitive function (Bucks and Radford, 2004).

Receptive emotional prosody is generally evaluated as the accurate identification of certain emotional tones when someone speaks (Taler et al., 2008). By removing information based on words (i.e. filtering out the spectral energy above a certain frequency), promising results, based solely on prosodic features, have been reported. Not only there are signs of an impaired processing of emotional prosodic information in DAT, but there is also evidence suggesting that such impairment precedes the decline of other linguistic aspects (Testa et al., 2001).

The work presented in this paper focuses on expressive emotional prosody. There are three distinct ways to elicit data: a) prosodic modelling, which requires participants with DAT to repeat a sentence copying a previously heard emotional tone (Testa et al., 2001), b) commanded production, which requires participants to read semantically neutral sentences with a designated emotional tone (Roberts et al., 1996), and c) natural expression, whereby participants are required to describe an emotional experience (Testa et al., 2001). Testa et al. (2001) applied speech analysis to evaluate the quality of emotional expression from participants with DAT, based on prosodic information, and found that receptive prosody was impaired earlier than expressive prosody, but also that both were impaired early in the progression of the disease. However, they used a limited feature set, essentially analysing variability in fundamental frequency.

The same elicitation method, natural expression, was used by a more recent study where participants were asked to share an autobiographical memory (Han et al., 2014). They measure emotional prosody quality of memory retrieval, under the common assumption that traumatic events contain essential information for survival and hence benefit from superior encoding. They report an impaired ability to express emotions in early AD, regardless of whether the autobiographical memory is recent or remote - one of the first studies looking at emotional prosody instead of semantic content of those memories. More importantly, they found a correspondence between emotional expression and cognitive functioning. Another recent work develops a measure for emotional response as part of a comprehensive prosodic account, reporting gradual changes in spontaneous speech and emotional response as cognition declines (Lopez-de Ipiña et al., 2016). Even though the approach of these more recent studies extends previous research by using multiple acoustic measures, their acoustic feature set is still limited in both size and underlying rationale. While F_0 and its associated measures correlate acoustically to perceived pitch, we propose to use a standardised and theoretically motivated feature set to detect psychological changes in voice production, namely, eGeMAPS (Eyben et al., 2016).

Further research is clearly necessary to provide a solid account of the quality of emotional expression in the context of DAT. A computational approach to this task would lessen the problem of subjectivity and low inter-rater reliability, as well as contributing to a potentially automatic diagnostic support tool. We hypothesise that if the expression of emotions through speech is impaired in a person with AD, a classifier should have greater difficulty distinguishing emotions in the voice of a person with AD than in the voice of a person without AD (non-AD). Therefore, a measure of uncertainty in emotion classification, such as the Shannon entropy of posterior (emotion) class probabilities, might be a suitable feature for a classification model for DAT. Besides, there is controversy about the actual reliability of humans identifying other humans' emotions, with sadness and anger usually being the emotions with highest agreement. Research evidence shows that emotion recognition from voice samples is about 60% accurate (Johnstone and Scherer, 2000), which we will take as a baseline for our model.

3. Dataset Description

This section describes the Berlin Database of Emotional Speech, used for the training of our emotion recognition system, and the age and gender balanced subset of the Pitt dataset, used for DAT prediction based on emotional speech features.

3.1. Berlin Database of Emotional Speech (*EmoDB*)

The *EmoDB corpus* (Burkhardt et al., 2005) is a dataset commonly used in the automatic emotion recognition literature. It features 535 acted emotions in German, based on utterances carrying no emotional bias. The corpus was recorded in a controlled environment resulting in high quality recordings, but actors were allowed to move freely around the microphones, affecting absolute signal intensity. In addition to the emotion, each recording was labelled with phonetic transcription using the SAMPA phonetic alphabet, emotional characteristics of voice, segmentation of the syllables, and stress. The quality of the data set was evaluated by perception tests carried out by 20 human participants. In a first recognition test, subjects listened to a recording once before assigning one of the available category, achieving an

average recognition rate of 86%. A second naturalness test was performed. Documents achieving a recognition rate lower than 80% or a naturalness rate lower than 60% were discarded from the main corpus, reducing the corpus to 535 recordings from the original 800. We have normalized all the speech utterances' volume into the range [-1:+1] dBFS before acoustic feature extraction. The motivation behind this normalization is to make the model robust against different recording conditions such as distance between the microphone and the subject.

3.2. The *Pitt Corpus*

This study specifically uses the *Pitt Corpus*, gathered longitudinally between 1983 and 1988 on a yearly basis as part of the Alzheimer Research Program at the University of Pittsburgh (Corey Bloom and Fleisher, 2000). Participants are categorised into three groups: dementia, control (non-AD), and unknown status. All participants were required to be above 44 years of age, have at least seven years of education, have no history of nervous system disorders or be taking neuroleptic medication, have an initial MMSE score of 10 or more and be able to provide informed consents. Extensive neuropsychological and physical assessments conducted on the participants are also included; more detailed information of this cohort can be found in (Becker et al., 1994). This study selected only the dementia and control groups for a binary diagnosis of AD and non-AD.

The *Pitt Corpus* contains data elicited through the following tasks: the Cookie Theft stimulus picture description for AD and non-AD groups, and a word fluency task, a story recall task, and a sentence construction task for the AD group only. In this study, we specifically chose the Cookie Theft description task subset. Table 1 lists the data available in this set. Participants were shown the Cookie Theft picture and were asked to describe the picture in their own words.

Table 1: Statistics of the DementiaBank Pitt corpus

	non-AD	AD*
Number of patients	99	194
Number of visits (recordings)	242	307
with 1 visit	26	117
with 2 visits	28	53
with 3 visits	28	12
with 4 visits	9	9
with 5 visits	8	3

*One participant (ID:172) has changed the diagnosis from "Control" (in the first visit) to "Dementia" (in the remaining 3 visits).

The *Pitt Corpus* includes both the manual transcripts of the clinical sessions and the corresponding audio recordings for both participants (i.e. AD and non-AD) groups. The transcripts comprise both the speech of the Investigator (`INV`) and the Participant (`PAR`). Based on the information provided by DementiaBank, the AD and non-AD groups were not matched with age, gender or education. This study will thus create a subset matched for age and gender to eliminate bias.

3.3. Subset Selection from *Pitt Corpus*

The steps taken to select a balanced subset of the *Pitt Corpus*: Cookie Theft task, for our experiment are shown in Figure 2, and described in the remainder of this section.
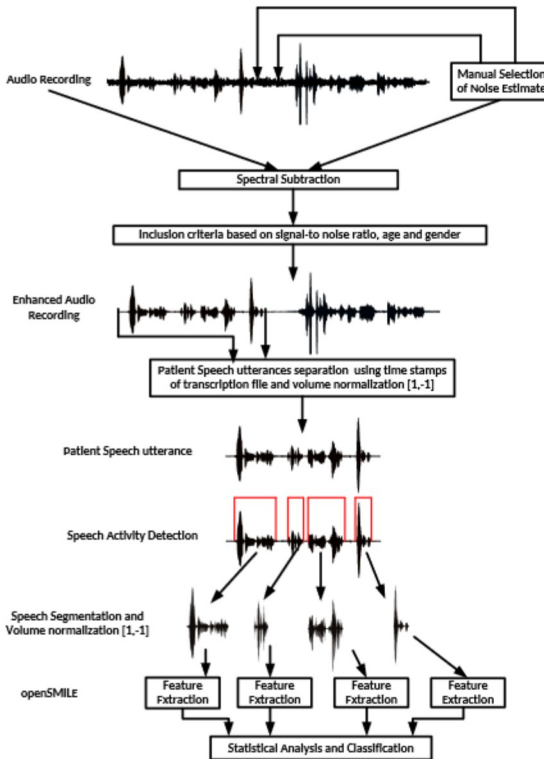

Figure 2: Data pre-processing steps.

3.3.1. Audio Enhancement and inclusion criteria

We manually selected a short interval from each audio recording which contained only the noise and applied spectral subtraction to eliminate that noise. Other non-target sounds such as background talk, ambulance sirens, door slamming, were minimised by selecting audio files with signal-to-noise ratio (SNR) greater than or equal to -17 dB. Where multiple audio files existed per participant only the most recent audio file of that participant was selected.

3.3.2. Matching the Data for Gender and Age

Age and gender are considered major risk factors for dementia (Dukart et al., 2011). Therefore, these variables are possible confounders between the AD and non-AD groups. To eliminate these confounders, we selected a subset of the *Pitt Corpus* in which the AD and non-AD groups are matched for age and gender. Along with the inclusion criteria defined in Section 3.3.1., matching gender and age for both AD and non-AD datasets ensured homogeneity of the sample population, reducing confounding and increasing the likelihood of finding a true association between exposure and outcomes. The age ranges were chosen empirically to optimise the number of recordings included in the final dataset. As a result, 164 participants matched the selection criteria to be included in the study. Of these, 82 were healthy and 82 were diagnosed with probable AD.

After testing the different ranges of the age intervals, the dataset was balanced and could produce the optimal number of recordings by using the age range from 45 to 80 years with the interval of 5 years. Table 2 presents the demographic data. Participants' age in each group ranged between 50 and 80 years old.

Table 2: Basic characteristics of the patients in each group (AD/non-AD)

| | AD | | non-AD | |
Age Interval	Male	Female	Male	Female
(50, 55)	2	1	2	1
(55, 60)	7	8	7	8
(60, 65)	4	9	4	9
(65, 70)	10	14	10	14
(70, 75)	9	11	9	11
(75, 80)	4	3	4	3
Total	36	46	36	46

3.3.3. Speech Segmentation

Speech segmentation was performed on the audio files that met the above described selection criteria. The study only focuses on the participants' speech; therefore, the investigators' speech were excluded from further processing.

First, we extracted the participants' speech utterances using the timestamps (start time and end time) from the DementiaBank transcripts. However, as the participants' speech exhibits long pauses and low volume, we normalised the volume to the range [-1:+1] dBFS and then used speech activity detection (with an energy threshold of 50 dB) [1] for speech segmentation (i.e. to separate speech from pauses). Volume normalization helps tackling different recording conditions, particularly variations in microphone placement in relation to the participant.

3.4. Feature Extraction

We used the openSMILE (Eyben et al., 2013) toolkit for the extraction of prosodic features using the eGeMAPS feature set, which is widely used for emotion recognition.

The **eGeMAPS** (Eyben et al., 2016) feature set contains the F_0 semitone, loudness, spectral flux, MFCC, jitter, shimmer, F_1, F_2, F_3, alpha ratio, Hammarberg index and slope V_0 features, including many statistical functions applied on these features, which results in a total of 88 features for every speech utterance. We removed features which are correlated ($|r| > 0.2$) with the duration of speech utterances. This left us with 75 remaining acoustic features for further processing.

4. Affect Recognition System

The Affect Recognition System was trained using Support Vector Machines (SVM) using the SMO solver with box constraint (k) of 0.75, and linear kernel function. We employed the MATLAB[2] implementation of this classifier, using the statistics and machine learning toolbox. A

leave-one-subject-out (LOSO) cross-validation procedure was adopted, where the training data do not contain any information on the validation subjects. The results are shown in Figure 3. The affect recognition system provides an accuracy of 69.72% with a Kappa of 0.638.

Figure 3: Emotion recognition results.

Once trained on *EmoDB*, our affect recognition system was used to identify emotions in the 4,076 speech segments in our dementia dataset. The results are shown in Figure 4. Noticeably, the AD subjects have more Sad (260 compared to 156) and Happy (616 compared to 580) instances than non-AD, and non-AD subjects have more Anger, Boredom, Disgust and Neutral instances.
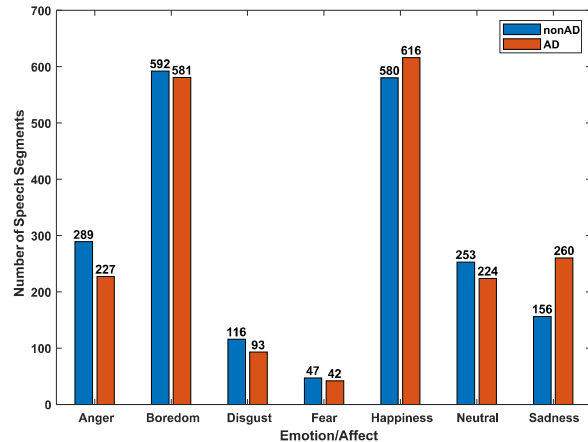
Figure 4: Emotion Recognition on subset of *Pitt Corpus* for AD and non-AD subjects.

5. Statistical Analysis

To find the relationship between emotions and AD, we used the matrix of scores (i.e. classification scores or class posterior probabilities) which indicates the likelihood that a speech segment expresses a particular emotion. As a results we have a vector of (1x7) for each speech segment representing likelihood of 6+1 emotions. We also calculated the entropy of the posterior probabilities per speech segment to measure the degree of the model's 'uncertainty' with regards to a classification.

The one-sample Kolmogorov-Smirnov test shows that the data (i.e. classification scores and entropy of scores) follows a normal distribution assumption with $p < 0.001$.

Table 3: Statistical Analysis: ANOVA test results; p-adj indicates p values adjusted for multiple comparison by controlling the false discovery rate.

	H_ϕ^{Anger}	H_ϕ^{Bore}	$H_\phi^{Disgust}$	H_ϕ^{Fear}	H_ϕ^{Happy}	H_ϕ^{Sad}	$H_\phi^{Neutral}$	$H_\phi^{Entropy}$
nonAD	0.1569	0.2178	0.1174	0.0446	0.1964	0.0815	0.1854	2.4750
AD	0.1285	0.2207	0.1026	0.0446	0.1996	0.1297	0.1742	2.4413
p-value	**0.0002**	0.7117	**0.0050**	0.9877	0.5469	**< 0.0001**	0.0642	**0.0216**
p-adj	**0.0011**	0.81	**0.0133**	0.9877	0.729	**< 0.0001**	0.1000	**0.0400**

We have set the following null hypotheses for the Anova test: the scores of emotion $x \in \{$Anger, Boredom, Disgust, Fear, Happiness, Sadness, Neutral$\}$ do not differ between the AD and non-AD groups (H_ϕ^x), and the Shannon entropy of the posterior probability distribution for an emotion given a speech segment ($H_\phi^{Entropy}$) does not differ between the AD and non-AD groups. The anova test results are shown in Table 3. The ANOVA test rejects the null hypothesis for Anger, Disgust, Sadness and entropy, when p values are corrected for multiple comparisons by using the Benjamini-Hochberg procedure to control the false discovery rate. However, no significant differences were found for Boredom, Fear, Happiness and Neutral emotion expressions.

6. Affective Behaviour Representation

To aggregate affective behaviour within an audio recording per subject for automatic classification of AD, we propose a novel Affective Behaviour Representation (ABR) feature vector. This consists of the following steps:

1. *Emotion Recognition* of segments: we used an emotion recognition model to recognise emotions within segments using audio features.

2. *Generation* of the Affective Behaviour Number ($nABR_{Ai}$) vector by calculating the number of segments in each emotion category for each audio (Ai) i.e. histogram representation of number of speech segments for 6+1 emotions for each audio recording.

3. *Normalisation of segments*: as the number of segments is different for each subject (i.e. the duration of all audio recordings is not constant), we normalise the ($nABR_{Ai}$) by dividing it by the total number of segments present in each audio recording (i.e. the L1 norm of $nABR_{Ai}$), as shown:

$$nABR_{A_{i_{norm}}} = \frac{nABR_{A_i}}{\|nABR_{A_i}\|_1} \quad (1)$$

4. *Generation* of the Affective Behaviour Score ($sABR_{Ai}$) vector by summing the score for each emotion category for each audio recording (Ai); that is, the histogram representation of scores for 6+1 emotions for each audio recording.

5. *Normalisation of score*: as the number of segments is different for each subject (i.e. the duration of all audio recordings is not constant), we normalise the ($sABR_{Ai}$) by dividing it by the sum of scores

of segments for each audio recording as we did for $nABR_{Ai}$:

$$snABR_{A_{i_{norm}}} = \frac{sABR_{A_i}}{\|sABR_{A_i}\|_1} \quad (2)$$

6. *Affective Behaviour Representation (ABR)*: we fused the $nABR_{A_i}$ and $snABR_{A_i}$ to generate the ABR, as shown in Equation 3

$$ABR_{A_{i_{norm}}} = [nABR_{A_{i_{norm}}}, snABR_{A_{i_{norm}}}] \quad (3)$$

6.1. AD Detection

We conducted three classification experiments to detect cognitive impairment due to AD, namely:

1. Segment Level (SL) classification: in this experiment we trained and tested our classifiers in a LOSO setting, with scores of emotions to predict whether the speech segments were uttered by a non-AD or AD patient;

2. Majority Vote (MV) classification: using the results of segment-level classification, we calculated the number of segments detected as AD and non-AD for each subject and then took a majority vote to assign an overall label to the subject; and

3. Affective Behaviour Representation: we generated the ABR using the score and labels of emotion recognition system as described in section 6., and then used $ABR_{A_{i_{norm}}}$ for classification as before.

6.2. Classification Methods

The classification experiments were performed using five different methods, namely decision trees (DT, with leaf size of 20), nearest neighbour (KNN with K=1), linear discriminant analysis (LDA), random forests (RF, with 50 trees and a leaf size of 20) and support vector machines (SVM, with a linear kernel with box constraint of 0.1, and sequential minimal optimisation solver). The classification methods were implemented in MATLAB using the statistics and machine learning toolbox. A leave-one-subject-out (LOSO) cross-validation setting was adopted.

6.3. Results

The AD recognition results for all three experiments (detailed in Section 6.1.) are shown in Table 4. It is noted that the ABR (59.76) provides better results than MV (58.54) and SL (52.70). The random forest classifier provides the best results for all three experiment. We have selected top three classifiers (57.93% for MV using KNN, 58.54% for

Figure 5: Fusion of Top Three Results

MV using RF and 59.76% for ABR using RF) and fused their label using the late fusion method. The fusion provides the best accuracy of 63.42%. The confusion matrix along with precision, recall and Kappa is shown in Figure 5.

Table 4: Classification Results using Affective Behaviour

Method	Blind	LDA	DT	1NN	SVM	RF
Segment Level	50.12	48.68	40.78	51.67	50.64	52.70
Majority Vote	50.00	50.61	40.24	*57.93*	54.27	**58.54**
ABR	50.00	54.27	50.61	54.27	54.27	*59.76*

7. Discussion

Statistical analysis showed that non-AD subjects' speech segments have a higher mean value (0.1569) of classification score for Anger than AD subjects (0.1285), the difference is statistically significant ($p < 0.01$). This suggests that the non-AD subjects expressed characteristics of Anger in their speech more than the AD subjects. Non-AD speech segments also have a higher mean value (0.1174) of classification score for Disgust than AD subjects (0.1026), similarly suggesting that non-AD subjects expressed this emotion in their speech more than AD subjects. This is likely due to the fact that the AD subjects usually have lower voice volume, speech rate and pitch than non-AD subjects, while Anger and Disgust emotions are associated with high voice volume, speech rate and pitch.

The non-AD subjects' speech segments have a significantly lower mean value (0.0815) of classification score for Sadness than AD subject (0.1297), suggesting that AD participants expressed speech with characteristics of Sadness more than non-AD subjects. For expression of Boredom, Fear and Happy, the differences in classifier scores are not statistically significant, which suggests that either AD and non-AD subjects can express those emotions equally in their speech, or that the model is not discriminating enough to detect those expressions. It is also noted that the emotion recognition system is more certain about the emotions of AD (mean entropy of 2.4413) than non-AD (mean entropy of 2.4750) participant, and this difference though quite small is statistically significant at the $p < 0.05$ level. As regards with the AD recognition results, using scores of the emotion recognition system and ABR as input features for the AD classifier we were able to detect AD speech with an accuracy of 63.42%. While this scores is below state-of-the-art AD classifiers (Haider et al., 2020), we note that the emotion recognition system evaluated on *emoDB* dataset

is also around 69.72% accurate. A speech dataset annotated for the emotions of elderly people and people with AD could conceivably improve the quality of the input features, and the performance of our emotion-based approach to AD recognition. It is also noted that the emotion recognition model was trained on a dataset (*emoDB*) recorded in a different language (German) to the one of the Pitt data (English). While the annotation quality of *emoDB* is higher than other datasets such as (Haq and Jackson, 2009; Costantini et al., 2014), it is possible that an affect recognition system trained directly on English data might improve the results.

8. Conclusion

In conclusion, we found that there are differences in (automatically) inferred affective behaviours regarding expressions of Sadness, Anger and Disgust among AD and non-AD subjects. Although these results need further study, they suggest, in agreement with the incipient literature on this topic, that AD speakers exhibit a deficit in the expression of those emotions reflected on voice volume, speech rate and pitch. The proposed Affective Behaviour Representation (ABR) and emotion classification scores are able to predict the AD with an accuracy of 63.42%. A limitation of this study which should be addressed in future work is the mismatch between the dataset used to generate the features for AD recognition (*emoDB*) and the *Pitt Corpus* (in which these features were used). This includes the facts that (unlike *emoDB*) the *Pitt Corpus* was not explicitly designed to elicit emotions, that the two datasets were recorded under different acoustic conditions and demographics, and that they are in different languages. In future work, we intend to manually annotate the Pitt corpus for emotions, and train an affect recognition system based on this augmented dataset to assess the effect of this model on AD recognition accuracy. The affect recognition system along with ABR script is made available to the research community through a git repository [3].

10. Bibliographical References

American Psychiatric Association. (2000). Delirium, dementia, and amnestic and other cognitive disorders. In American Psychiatric Association, editor, *Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR)*, chapter 2. Arlington, VA, fourth edi edition.

Becker, J. T., Boiler, F., Lopez, O. L., Saxton, J., and McGonigle, K. L. (1994). The Natural History of Alzheimer's Disease. *Archives of Neurology*, 51(6):585.

---

[3]https://git.ecdf.ed.ac.uk/fhaider/emotion2dementia.git

Bondi, M. W., Salmon, D. P., and Kaszniak, A. W. (1996). The neuropsychology of dementia. In *Neuropsychological assessment of neuropsychiatric disorders.*, pages 164–199. Oxford University Press, New York, NY, US, 2 edition.

Bucks, R. S. and Radford, S. A. (2004). Emotion processing in alzheimer's disease. *Aging & mental health*, 8(3):222–232.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., and Weiss, B. (2005). A database of german emotional speech. In *Proceedings of the ninth European Conference on Speech Communication and Technology*, pages 1516–1520.

Corey Bloom, J. and Fleisher, A. (2000). The natural history of alzheimer's disease. *Dementia*, 34:405–15.

Costantini, G., Iaderola, I., Paoloni, A., and Todisco, M. (2014). Emovo corpus: an italian emotional speech database. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, LREC 2014, pages 3501–3504. European Language Resources Association (ELRA).

Dukart, J., Schroeter, M. L., Mueller, K., Initiative, A. D. N., et al. (2011). Age correction in dementia–matching to a healthy brain. *PloS one*, 6(7):e22193.

Eyben, F., Weninger, F., Groß, F., and Schuller, B. (2013). Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 835–838. ACM, Association for Computing Machinery.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., et al. (2016). The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2):190–202.

Haider, F., de la Fuente, S., and Luz, S. (2020). An assessment of paralinguistic acoustic features for detection of alzheimer's dementia in spontaneous speech. *IEEE Journal of Selected Topics in Signal Processing*, 14(2):272–281.

Han, K.-H., Zaytseva, Y., Bao, Y., Pöppel, E., Chung, S. Y., Kim, J. W., and Kim, H. T. (2014). Impairment of vocal expression of negative emotions in patients with alzheimer's disease. *Frontiers in aging neuroscience*, 6:101.

Haq, S. and Jackson, P. (2009). Speaker-dependent audio-visual emotion recognition. In *Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP)*, pages 53–58, Sept.

Hart, R. P., Kwentus, J. A., Taylor, J. R., and Harkins, S. W. (1987). Rate of forgetting in dementia and depression. *Journal of Consulting and Clinical Psychology*, 55(1):101–105.

Horley, K., Reid, A., and Burnham, D. (2010). Emotional prosody perception and production in dementia of the alzheimer's type. *Journal of Speech, Language, and Hearing Research*, 53:1132–1146.

Johnstone, T. and Scherer, K. R. (2000). Vocal communication of emotion. *Handbook of emotions*, 2:220–235.

Juslin, P. N. and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, 129(5):770.

Kirshner, H. S. (2012). Primary Progressive Aphasia and Alzheimer's Disease: Brief History, Recent Evidence. *Current Neurology and Neuroscience Reports*, 12(6):709–714.

Landes, A. M., Sperry, S. D., and Strauss, M. E. (2005). Prevalence of apathy, dysphoria, and depression in relation to dementia severity in alzheimer's disease. *The Journal of neuropsychiatry and clinical neurosciences*, 17(3):342–349.

Leyhe, T., Reynolds III, C. F., Melcher, T., Linnemann, C., Klöppel, S., Blennow, K., Zetterberg, H., Dubois, B., Lista, S., and Hampel, H. (2017). A common challenge in older adults: Classification, overlap, and therapy of depression and dementia. *Alzheimer's & dementia*, 13(1):59–71.

Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schütz, A., Sellin, I., and Salovey, P. (2004). Emotional intelligence and social interaction. *Personality and social psychology bulletin*, 30(8):1018–1034.

Lopez-de Ipiña, K., Faundez-Zanuy, M., Solé-Casals, J., Zelarin, F., and Calvo, P. (2016). Multi-class Versus One-Class Classifier in Spontaneous Speech Analysis Oriented to Alzheimer Disease Diagnosis. In Esposito et Al., editor, *Recent Advances in Nonlinear Speech Processing*, volume 48, pages 63–72. Springer International Publishing.

Roberts, V. J., Ingram, S. M., Lamar, M., and Green, R. C. (1996). Prosody impairment and associated affective and behavioral disturbances in alzheimer's disease. *Neurology*, 47(6):1482–1488.

Seidl, U., Lueken, U., Thomann, P. A., Kruse, A., and Schröder, J. (2012). Facial expression in alzheimer's disease: impact of cognitive deficits and neuropsychiatric symptoms. *American Journal of Alzheimer's Disease & Other Dementias®*, 27(2):100–106.

Starkstein, S. E., Ingram, L., Garau, M., and Mizrahi, R. (2005). On the overlap between apathy and depression in dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(8):1070–1074.

Taler, V., Baum, S. R., Chertkow, H., and Saumier, D. (2008). Comprehension of grammatical and emotional prosody is impaired in alzheimer's disease. *Neuropsychology*, 22(2):188.

Testa, J., Beatty, W., Gleason, A., Orbelo, D., and Ross, E. (2001). Impaired affective prosody in ad: Relationship to aphasic deficits and emotional behaviors. *Neurology*, 57(8):1474–1481.