

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Motif discovery in speech: Application to monitoring Alzheimer's disease

Peter Garrard
*Cardiovascular and Cell Sciences Research Institute
St George's, University of London
London SW17 0RE
United Kingdom*

Vanda Nemes
*Institute of Physiology
University of Pécs Medical School
Pécs 7624
Hungary*

Dragana Nikolic and Anna Barney*
*Institute of Sound and Vibration Research
University of Southampton
Southampton SO17 1BJ
United Kingdom*

*Corresponding author

1

2 **Abstract**

3 Background: Perseveration – repetition of words, phrases or questions in speech – is
4 commonly described in Alzheimer’s disease (AD). Measuring perseveration is difficult, but
5 may index cognitive performance, aiding diagnosis and disease monitoring. Continuous
6 recording of speech would produce a large quantity of data requiring painstaking manual
7 analysis, and risk violating patients’ and others’ privacy. A secure record and an automated
8 approach to analysis are required.

9 Objectives: To record bone-conducted acoustic energy fluctuations from a subject’s vocal
10 apparatus using an accelerometer, to describe the recording and analysis stages in detail,
11 and demonstrate that the approach is feasible in AD.

12 Methods: Speech-related vibration was captured by an accelerometer, affixed above the
13 temporo-mandibular joint. Healthy subjects read a script with embedded repetitions.
14 Features were extracted from recorded signals and combined using Principal Component
15 Analysis to obtain a one-dimensional representation of the feature vector. Motif discovery
16 techniques were used to detect repeated segments. The equipment was tested in AD
17 patients to determine device acceptability and recording quality.

18 Results: Comparison with the known location of embedded motifs suggests that, with
19 appropriate parameter tuning, the motif discovery method can detect repetitions. The
20 device was acceptable to patients and produced adequate signal quality in their home
21 environments.

22 Conclusions: We established that continuously recording bone-conducted speech and
23 detecting perseverative patterns were both possible. In future studies we plan to associate
24 the frequency of verbal repetitions with stage, progression and type of dementia. It is
25 possible that the method could contribute to the assessment of disease-modifying
26 treatments.

27

28 **Keywords:** Alzheimer’s disease; perseveration; bone-conducted speech; motif discovery;
29 principal component analysis

30

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Introduction

The risk of dementia rises markedly with age (1), therefore an increased incidence of dementia and of its vast associated socio-economic costs (2) are inevitable consequences of improved life expectancy (3). Consequently, early, accurate detection and treatment of dementia have been identified as among the most important priorities for health research in the UK and other economically advanced societies (4). Impairment of cognition short of dementia (mild cognitive impairment (MCI)) is recognised as a potential precursor to future decline (5), yet symptom progression is by no means inevitable, and clinicians often rely on invasive, expensive biological measures such as structural and functional brain imaging and cerebrospinal fluid (CSF) analysis to verify the presence of a neurodegenerative pathology (6, 7). It is widely acknowledged that simpler, more widely available biomarkers of comparable sensitivity and specificity would have a transformative effect on the classification and clinical management of MCI (8).

As there is still no known method of slowing or arresting progression in AD, biomarker positive MCI patients are increasingly recruited to randomised controlled trials of candidate disease-modifying drugs. Outcomes are derived from global metrics of deterioration based on estimates made by clinicians or caregivers (9), abbreviated cognitive assessments such as the mini mental state examination (MMSE) (10) or the Addenbrooke’s cognitive examination (ACE) (11), and comprehensive batteries of formal neuropsychological tests. None of these methods is without shortcomings: global measures are based on subjective judgements and quantified using coarse-grained interval scales; the content of bedside cognitive batteries is fixed, and thus subject to learning or practice effects, which may mask

1 progression; and psychometric measures are time-consuming to complete, sensitive to
2 fluctuations in a patient's level of motivation and effort, as well as to learning effects, and
3 reflect performance in a laboratory, rather than 'real world' context (12). An outcome
4 measure based on a robust and naturalistic performance marker could radically simplify
5 dementia drug trial protocols, accelerating the rate at which efficacy of candidate
6 treatments could be either confirmed or eliminated.

7 There is therefore a compelling case for seeking new biomarkers, which correlate with
8 cognitive performance and can be measured with minimum invasion, to assist with
9 diagnosis or measurement of disease progression. We propose that perseveration – the
10 tendency to make the same statement, ask the same question, or carry out the same action
11 repeatedly over the course of the day (13) is a good candidate for such a marker.

12 Among the many abnormalities noticed by family members and caregivers of patients with
13 Alzheimer's disease (AD) – perseveration is among the most common. Perseverative
14 behaviour is likely to be related to a decline in day-to-day memory (the commonest first
15 symptom of AD): it is assumed that failure to store new information in episodic memory
16 results in incomplete updating of conscious experience, and hence to repetition. These
17 phenomena are widespread among patients with established dementia, probably increase
18 with progression of the condition (13), and are widely regarded as sensitive indicators of
19 disordered cognition (14).

20 To date, the frequency and severity of such errors have been quantified only in terms of
21 caregivers' estimates (13); more discriminating, quantitative measures (such as the mean
22 length of time between repetitions, the complexity of the repeated behaviour, and the
23 number of times the repetition is observed over the course of a day) have not been
24 acquired, despite the potential value of this information.

1 Lack of progress in this area is related to two important barriers to research - one ethical,
2 the other practical. The principal ethical consideration is that of privacy: continuous audio
3 and/or video recording of a patient with AD in the domestic environment may well capture
4 the appropriate data, but would be unacceptably intrusive, not only on the patient but on
5 all those with whom they interacted. Moreover, even if continuous voice recording were
6 deemed acceptable, the record would have to be subjected to considerable analysis.
7 Segmentation of the record to select only the subject's utterances, followed by the
8 identification and quantification of any repeated word strings, would be too time-
9 consuming to be feasible manually on a large scale and requires the development of
10 automated methods if it is to be practicable. In this paper we describe a technical solution
11 to both the privacy and the analysis problems.

12 To solve the privacy problem, we propose the use of an accelerometer to detect vibration
13 signals transmitted through the body during speech. Such a system, sometimes referred to
14 as a "contact microphone" differs from the airborne speech sound detected by a traditional
15 microphone and allows a continuous record of vocal activity to be acquired, which is robust
16 against sound sources external to the subject (15) including the sound of other speakers.

17 The signals so recorded represent a significantly low-pass filtered waveform compared to
18 recordings of the sound radiated from the mouth. Typically maximum speech frequencies
19 recorded by a microphone range up to about 5 kHz whereas in accelerometric speech, the
20 upper cut-off frequency is of the order of 1 kHz resulting in the loss of higher order formant
21 frequencies for vowels, frication and burst information for consonants and therefore
22 masking to some extent both the speaker's identity and the semantic content of their
23 speech (16). We hypothesised that repeated occurrences of identical phrases should

1 nevertheless be identifiable from the similarity of the recorded waveforms associated with
2 them.

3 In order to assess the practical application of such a device in a real world setting, we report
4 on a small scale test in patients with AD in their homes. The goal was to ascertain whether a
5 face-mounted recording device is acceptable to wearers in their home environment and
6 whether the signals so recorded are of sufficient quality to permit automatic analysis.

7 We then report the outcome of a study in healthy participants using scripted speech to test
8 the feasibility of automatically identifying repeated segments of speech in longer
9 conversations. We apply established algorithms for data mining that have been tested in
10 other contexts (music categorisation; astronomical data analysis, gene sequencing) to
11 identify repeated sub-sequences in the accelerometric speech. The technique of
12 identifying the existence and location of repeated sub-sequences in a data stream is known
13 as *motif discovery* where a *motif* can be described as a previously unknown but frequently
14 occurring pattern. The problem we consider here differs from the classical speech
15 recognition problem in that the linguistic and semantic contents of the utterance are not of
16 interest; instead we ask: Are any of the utterances in this sample of speech repeated more
17 than once? The possible repeated phrases are not identified *a priori*. Note that we are
18 interested in identifying phrases where the acoustic features match closely, suggesting
19 repetition of the identical utterance, not in phrases which are identical in meaning (so
20 “What is the time?” and “What time is it?” are not in our case a match). We note that motif
21 discovery algorithms have been applied to speech samples before (17-19) though we are
22 not aware of any studies applying these methods to accelerometric speech where the
23 information content is much reduced by the recording process.

1 Our two aims in this work are therefore:

- 2 • To determine the potential of accelerometric speech in terms of recording quality,
3 privacy and acceptability of device to a small group of AD users.
- 4 • To implement and test a data-mining protocol for identifying repeated phrases
5 embedded in accelerometric speech.

6 **Materials and methods**

7 We report in this paper two separate studies, one to test the device function and
8 acceptability in a small sample of AD patients in their home environment and one using only
9 healthy subjects to obtain speech samples for a proof-of-principle study to test an algorithm
10 for automatic detection of repeated speech phrases.

11 *Participants – device testing*

12 For this study 5 patients, all aged over 65, with a diagnosis of probable AD were recruited. In
13 each, the presence of perseverative verbal behaviour manifesting as repetitive speech
14 during everyday home life, had been reported. Patients who did not have a regular
15 conversational partner, those who suffered from separate speech and/or language disability
16 (e.g. due to previous stroke), those with significant hearing impairment, with dermatological
17 problems at or adjacent to the facial region where the accelerometer was to be fixed, or
18 recent ENT surgery, were not invited to participate.

19 All participants gave informed consent. The clinical study was approved after review by the
20 National Research Ethics Service (NRES) London South East Committee [REC reference
21 11/LO/1192].

1 *Participants – algorithm testing*

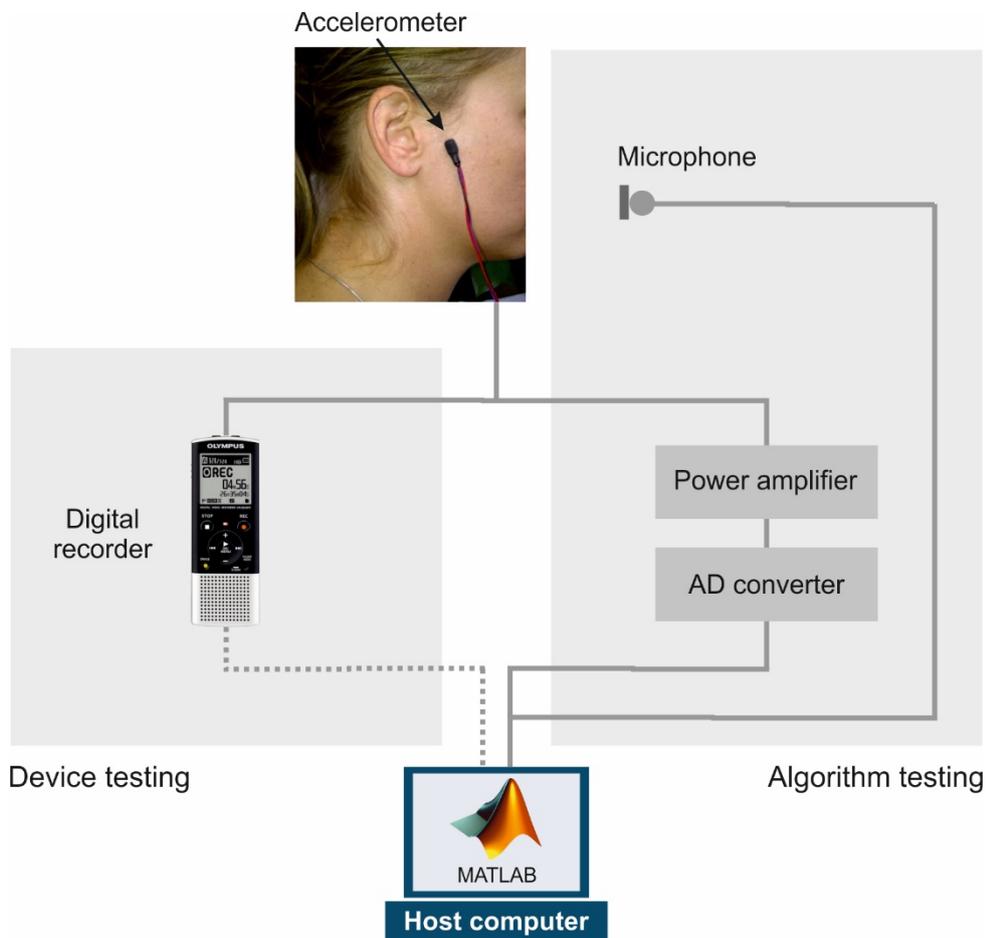
2 Recordings were obtained from a group of 5 healthy volunteers, who were undergraduate
3 or postgraduate students at the University of Southampton, aged between 20 and 30.
4 Volunteers were not invited to participate if they reported: a speech and/or language
5 disability (e.g. stammer), significant hearing impairment, dermatological problems at or
6 adjacent to the facial region where the accelerometer was to be fixed, or recent ENT
7 surgery. Ethical Approval for the study was obtained from the University's internal ethics
8 committee and all participants gave informed consent.

9 *Data acquisition*

10 In both studies, accelerometric speech was captured by recording bone-conducted
11 vibrations due to the vocal apparatus using an accelerometer. The data acquisition system
12 consisted of a small (approx. 10 mm³) accelerometer (Knowles BU 21771-000), which was
13 affixed to the skin over the temporo-mandibular joint (TMJ) using adhesive tape. The TMJ
14 was chosen specifically for locating the sensor because a previous study (20) found that the
15 speech intelligibility score for bone-conducted speech recorded via a contact sensor in the
16 region of the TMJ was as low as 65%, thus affording the greatest privacy to the patient.
17 Informal aural assessment of the recorded signal suggested it was largely devoid of linguistic
18 meaning.

19 For the device testing study, the accelerometer signal was routed to a lightweight, battery
20 powered digital recorder. For the algorithm testing study it was connected via an amplifier
21 and analogue to digital converter to a PC.

1 The data acquisition system for the algorithm testing study also included a headset
2 mounted boom microphone (Yoga Electronics, EM-174M, 20Hz-16kHz, -68 dB±3 dB, 680 Ω),
3 which used the traditional air-conducted route for recording speech, to give a reference
4 signal in which repeated phrases could be clearly identified and used to validate the
5 outcomes of the analysis of the accelerometric speech. Figure 1 shows the data acquisition
6 system for both data testing and algorithm testing studies.



7

8 *Figure 1:* Data acquisition system used for the device and algorithm testing studies. In the
9 device testing study with AD patients there was no microphone and the accelerometer
10 signal was recorded on a portable digital recorder and later transferred to the PC for
11 processing. Recorded signals from the accelerometer and the microphone were sampled at
12 a frequency of 16 kHz (12-bit resolution).

13

1 *Recorded speech data – device testing*

2 Upon recruitment, participants in the device testing study were given a trial speech
3 production task. This used the Cookie Theft picture from the Boston Diagnostic Aphasia
4 Examination (21), a picture frequently used to elicit free and fluent speech in linguistic
5 studies, and asked to: *Please tell me everything you see going on in the picture.* The aim of
6 this brief assessment was to ensure the participant's tolerance of the equipment while
7 speaking and to check for adequate quality of recorded data.

8 At a later date, the participant was visited at home by a member of the research team, and
9 the equipment reattached. The participant was instructed to avoid bringing the
10 accelerometer into contact with water, and to remove the device at the end of the day,
11 before going to bed, but otherwise to carry out their daily activities as normal. Speech
12 produced naturally by the patient during the course of the day was recorded continuously in
13 the home environment via the accelerometer and digital recorder until the patient chose to
14 remove the device. The following day a member of the research team visited the
15 participant's home to collect the recording equipment.

16 *Recorded speech data – algorithm testing*

17 For this study, the healthy participants were instructed to read aloud each of three pre-
18 prepared scripts, which contained, embedded in the speech, some short, repeated
19 questions and statements. The embedded speech segments used were selected based on
20 conversations with carers for people with AD as typical of the perseverated phrases they
21 encountered and consisted of: "What is the time?" (script 1); "I think I'm getting confused"
22 and "What day is it?" (script 2); and "Have you eaten?" and "Where did I leave my glasses?"

1 (script 3). Note that the choice of embedded sentences from among those suggested by the
2 carers is arbitrary as the proposed algorithm searches for any repeated phrases on the basis
3 of similarity of their acoustic features and has no underlying assumption about the semantic
4 attributes of the phrase or *a priori* model of what the phrase might be.

5 *Data analysis – device testing*

6 The purpose of the study with the AD patients was to test the feasibility of using the
7 accelerometer to capture speech-related data of sufficient quality in a patient's home
8 environment. Recorded signals were assessed for continuity and quality of recording and
9 the researcher also discussed with the patient after the trial whether they had any
10 difficulties with the device. Discussion with a family member or carer for the patient also
11 covered difficulties with the device and whether the patient was thought to have shown any
12 perseverative speech behaviour while wearing it for the trial.

13 *Data analysis – algorithm testing*

14 Data analysis for algorithm testing involved the application of established data-mining
15 methods to the accelerometric speech signal as detailed in this section.

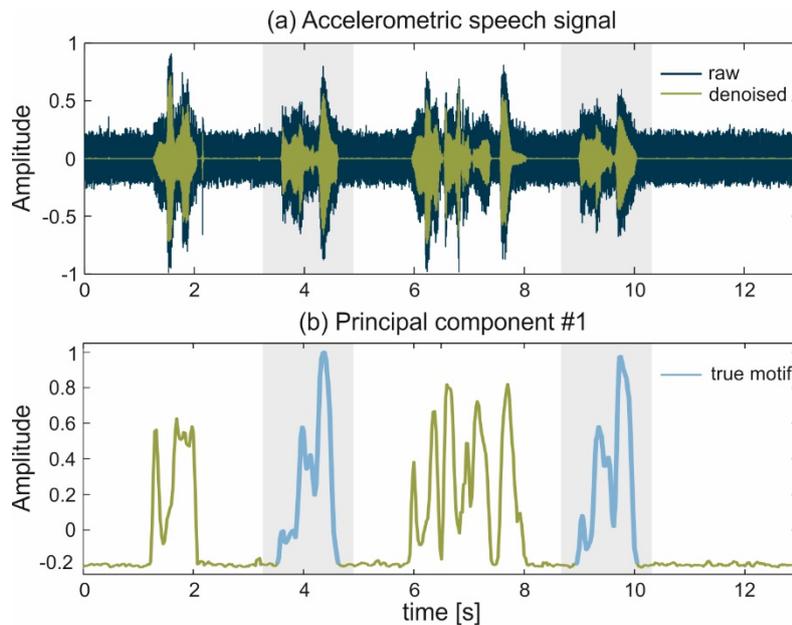
16 The accelerometer signal was pre-processed by applying a noise reduction scheme (22, 23)
17 and by the removal of any silent periods in order to reduce the size of the data stream
18 requiring processing. It was also high-pass filtered with a cut-off frequency of 20 Hz to
19 remove any motion artefacts related to articulatory movements of the TMJ.

20 The signal was then segmented into frames (frame length 50 ms, overlap between frames
21 60%) and a set of 95 features was calculated per frame: statistical measures (means,
22 extremes, zero and mean crossing rates, linear and quadratic regression coefficients);
23 spectral moments (centroid, spread, skewness, kurtosis), energy in defined frequency bands

1 (RMS energy, logarithmic energy), wavelet parameters, Mel-frequency and PLP cepstral
2 coefficients (to model aural and psychoacoustic perceptual aspects of hearing respectively)
3 and prosodic measures (fundamental frequency; voicing probability, pitch strength); using
4 the OpenSMILE toolbox (24, 25). OpenSMILE is a toolbox developed for extracting features
5 from speech signals that are typically used in speech recognition and emotion detection
6 applications. All the features selected have been shown to be useful in characterising air-
7 borne speech, though their applicability for characterising accelerometric speech has not
8 previously been tested.

9 Speech data characterised by these features contains a very large amount of information,
10 but equally, requires significant processing time to search for motifs. We wish to reduce the
11 dimensionality of the data without knowing prior to our analysis which features will be
12 useful in identifying repeated phrases and which may be discarded. We therefore carried
13 out a principal component analysis (PCA) on the feature vector and retained only the first PC
14 for further processing (26). The first PC for our data typically accounted for 25 to 30% of the
15 total variance of the feature vector signal. It is unlikely to contain elements related to all 95
16 features, but it contains the elements that contribute most to the variance of the signal.

17 More critically for this application, if we treat the component as a new time series it
18 provides a reduced information signal that nevertheless retains a similar shape in segments
19 that have similar acoustic features (i.e. repeated segments) as can be seen in Figure 2.



1

2 *Figure 2* (a) Accelerometric speech signal with two repeated segments (shaded background)
 3 identified for illustration and (b) first principal component extracted from the 95-D acoustic
 4 feature vector.

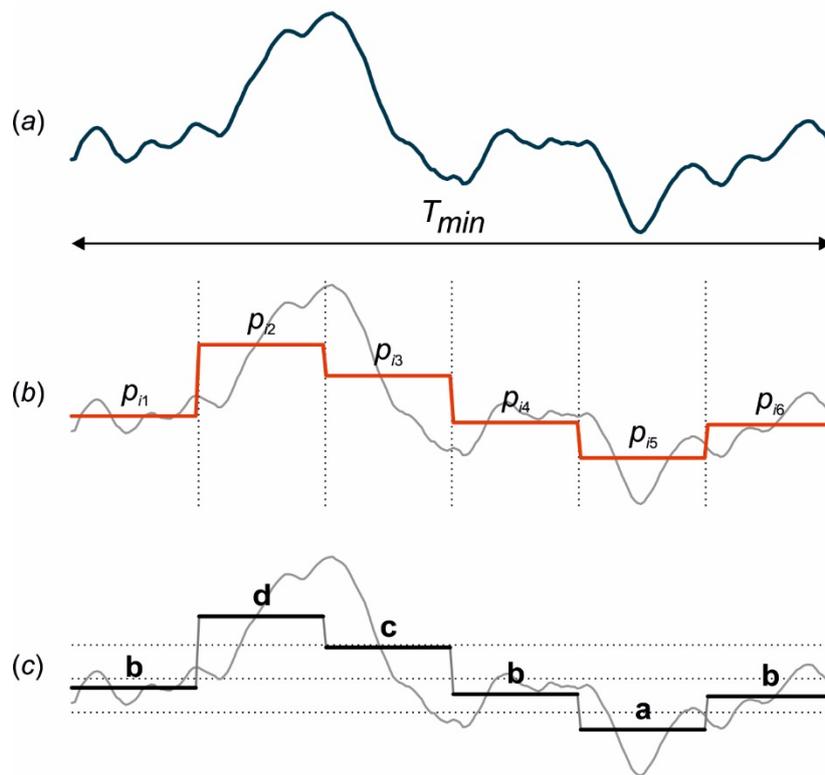
5

6 To search for motifs, it would now be possible to divide this new data stream into shorter
 7 sub-sequences and to compare each sub-sequence with every other. Repeated speech
 8 patterns represented in the data may not be absolutely identical, but are likely to be similar
 9 to each other. Similar sub-sequences can be defined as those that have a Euclidean Distance
 10 (or other suitable distance metric) between them of less than a pre-defined threshold value.
 11 Such a process is often referred to in the data-mining literature as the “brute force” search
 12 method. However, even with the reduction in dimensionality resulting from the use of only
 13 the first principle component, any given recording remains very large due to the still
 14 considerable amount of information it contains and will be extremely time consuming to
 15 analyse by such a method. A key focus in the development of the system was therefore the
 16 application of efficient search techniques that were able to reduce the computational
 17 burden of locating repeated utterances. The tactic adopted is to search first for the most

1 likely candidates for a motif and then to evaluate these in more detail to identify repetitions
2 that are sufficiently good matches under a given set of criteria.

3 The algorithms we used are established data mining techniques and their detailed
4 implementation can be found in the references given. We focus here more on the general
5 scheme of analysis and provide a largely qualitative description of the process we adopted.

6 The first step in the search process was to apply a coding to the data that further reduces
7 the information carried and which makes data segments easier to manipulate by
8 representing them in a symbolic format. This was achieved by the application of the
9 Piecewise Aggregate Approximation (PAA) (27) followed by the Symbolic Aggregate
10 Approximation (SAX) (28). PAA takes each sub-sequence of a pre-defined length T_{\min} within
11 the signal, z-score normalises it, divides it into sections of equal length and replaces the data
12 within each section with its mean value. The number of sections is known as the PAA word
13 size and we chose to work with a word size of 6 to represent each sub-sequence of data
14 with a step between the start of one sub-sequence and the next of one sample. SAX coding
15 then converts the PAA words into alphabetic symbols using, in our case a word size of 6 to
16 match the PAA word length and an alphabet size of 4. PAA sections are assigned an
17 alphabetic symbol based on the probability distribution of the subsequence. The process is
18 illustrated for a single subsequence in Figure 3.



1
2
3
4
5
6
7
8

Figure 3 a) Subsequence extracted from the first principal component of the 95-D feature vector with length T_{min} ; b) z-score normalised sequence divided into 6 sections with (red line) PAA coding applied by replacing the data in each section with its mean value p_{in} and c) division of the amplitude into 4 areas, based on equal probability, indicated by dotted lines and SAX coding of the PAA signal using an alphabet of 4 symbols according to which area it falls into such that the SAX code for this subsequence is: bdcbab.

9 Sub-sequences that are part of a motif in the data stream are defined as those with a
10 Euclidean difference between them of less than a pre-set threshold value. Sequences
11 associated with the same motif are likely to have the same, or at least similar, SAX codes.
12 The difference between SAX codes can be computed much more quickly than the Euclidean
13 difference between sub-sequences and is therefore a more efficient way to look for
14 similarity. Once candidates for motifs are identified, only those most likely to represent
15 motifs need be checked more rigorously by testing the Euclidean distance. The process can
16 be made yet more efficient by optimising the order in which SAX codes are searched so as to
17 start with the those most likely to result in finding a motif.

1 To search for candidate motifs using the SAX-coded sub-sequences, we selected the
2 Enumeration of Motifs through Matrix Approximation (EMMA) algorithm (29). In this
3 method, the hash-value for each SAX word is used to label a bucket which contains the
4 pointers to the starting position in the data stream of all SAX words with the same hash-
5 value. The buckets are then sorted according to the number of pointers they contain. The
6 bucket with the largest number of pointers in it is the most likely place to find a motif and is
7 known as the Most Promising Candidate (MPC). It is not certain that the MPC will contain a
8 motif, nor that all occurrences of a given motif will be found in the MPC. Therefore, prior to
9 search the MPC is given an associated neighbourhood of other buckets where the SAX word
10 differs from the word labelling the MPC by less than a predefined threshold value for a
11 distance metric known as the MinDist (30). Searching then starts with the MPC and its
12 neighbourhood and proceeds through subsequent buckets and their neighbourhoods in
13 order of bucket size until a pre-defined stop condition is met (e.g. N different motifs found
14 where N is a pre-defined value).

15 Even with the targeted search method defined by EMMA, the number of comparisons
16 between sub-sequences can be large and computationally costly. A further efficiency was
17 gained by using the Approximate Distance Map (ADM) algorithm (31) to first estimate the
18 upper and lower boundaries for the Euclidean distance between sub-sequences and then
19 only to calculate in full the Euclidean distances for those pairs not ruled out as too far apart
20 by ADM.

1 Results

2 *Device testing*

3 The device testing study indicated that the accelerometer was not found too intrusive by
4 wearers and that the recording device could normally be accommodated in a pocket.

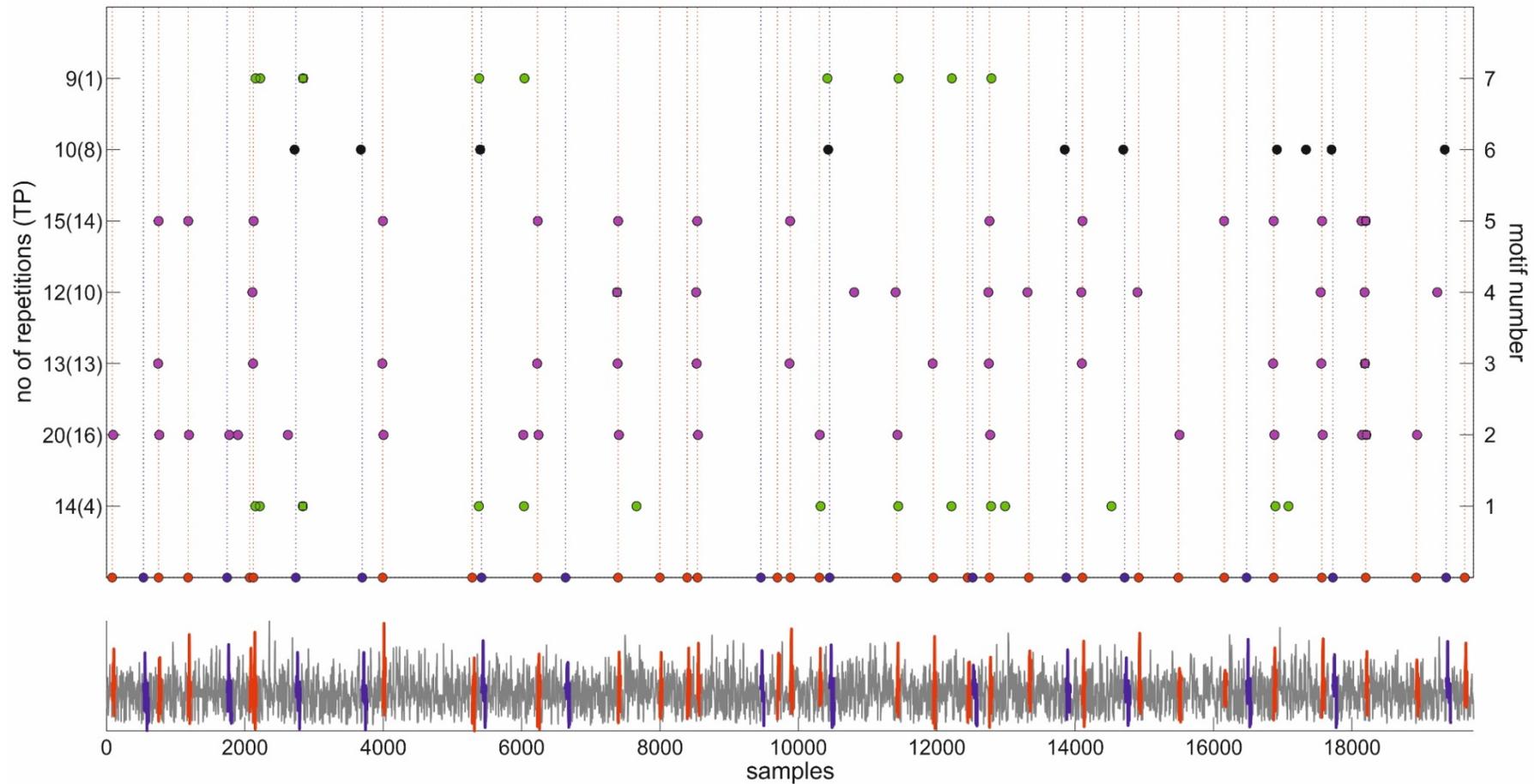
5 Recorded data were found to be of adequate quality for analysis with no breaks in recording
6 or excessive noise. Of the five subjects recorded, the carer of only one reported that they
7 had shown perseverative speech behaviour during the recording session. The data from this
8 subject is a current focus of further analysis.

9 *Algorithm testing*

10 In order to validate the results from the algorithm testing study, the location of the
11 embedded repeated phrases was identified in the signal recorded via the microphone and
12 the corresponding points in the accelerometric speech were manually labelled. We call
13 these embedded phrases *true motifs*.

14 An illustration of the detection results is shown in Figure 4. The bottom plot shows the first
15 principal component extracted from the accelerometric speech superimposed with matches
16 of the two deliberately embedded (true) motifs in this script; one coloured red and one
17 blue. The starting position of each true motif is denoted by a vertical line on the upper plot
18 emerging from a correspondingly-coloured dot on the horizontal axis. Each of the, in this
19 case 7, detected motifs, is shown as one row of dots on the upper plot. Each dot represents
20 starting position of a SAX word identified as forming part of the motif. The total number of
21 SAX words in a given motif is shown as the number of repetitions on the vertical axis with
22 the number of matches to a true (embedded) motif given in brackets. For instance, the third

- 1 motif consists of 13 matches and all these matches belong to one of the two motifs
- 2 embedded in the script.



1

2 *Figure 4:* Example of detected motifs (script 3, subject 5) showing (bottom plot) the first principal component extracted from the bone-
 3 conducted speech superimposed with two deliberately embedded (true) motifs; one red and one blue, and (top plot) the detected motifs each
 4 represented by a row of dots where each dot represents a SAX word within that motif. The starting position of each true motif is denoted by a
 5 vertical line on the upper plot emerging from a correspondingly-coloured dot on the horizontal axis. The total number of SAX words in a given
 6 motif is shown as the number of repetitions on the vertical axis with the number of matches to a true motif is given in brackets.

1 Where only some of the SAX words in a motif align with a true motif the non-aligned words
 2 nonetheless do indicate sections of the signal which are sufficiently similar to the true motif
 3 to be detected as such and thus they obey the rules for inclusion in the motif. Further there
 4 may be motifs where none of the SAX words align with true motifs, but these do
 5 presumably represent sufficiently similar sections of the signal to indicate repetitions even if
 6 not those repetitions that were deliberately embedded in the signals in this study. These
 7 might well represent other, naturally occurring, repetitions in the signal. The performance
 8 of the system as a detector of the true motifs deliberately embedded in the scripts is
 9 summarised in Table 1.

10 *Table 1: Average performance for the motif detection system over all subjects*

Script	Motif	Range of no. true motifs embedded	Average of true motifs detected (SD) [%]
1	#1 "What is the time?"	14	60.0 (13.9)
2	#1 "I think I'm getting confused"	4	70.0 (44.7)
2	#2 "What day is it?"	9	48.9 (44.9)
3	#1 "Have you eaten?"	18-29	75.1 (10.3)
3	#2 "Where did I leave my glasses?"	7-14	47.3 (27.7)

11
 12 Scripts 1 and 2 contained the same number of true motifs for all subjects and all reached the
 13 end of the script hence the number of motifs for each script and each subject is fixed. Script
 14 3 was long and not all subjects reached the end in the allotted recording time, hence the
 15 number of embedded true motifs in the recording for each speaker varied and is presented
 16 as a range.

17 In practice, for an implementation used with patients where the goal is to ensure privacy,
 18 one would not have access to the ground truth information about the location of any
 19 perseverative phrase within the recorded data, nor even necessarily any knowledge of

1 whether there was an unusual amount of phrasal repetition in a subject’s speech during the
 2 recording. It is therefore instructive to look at the difference in performance of the method
 3 using speech with and without the embedded repetitions. In order to achieve this, the true
 4 motifs were cut from the signals and the remaining portions of the signal were re-joined
 5 using the weighted overlap-add technique to ensure smoothness (32). Table 2 shows the
 6 average percentage reduction in the number of detected SAX words within the motifs for
 7 each script.

8 *Table 2: Average reduction in the percentage of detected Symbolic Aggregate*
 9 *Approximation (SAX) words between recordings with and without embedded true*
 10 *motifs.*

Script	Mean (SD) percentage reduction in number SAX words detected with and without true motifs embedded
1	41.7 (39.5)
2	32.2 (28.1)
3	47.2 (11.1)

11 It can be seen that, on average for each script, the number of detected SAX words does
 12 reduce when there are no embedded true motifs. Although the variance for this small
 13 sample of subjects is sufficiently large that statistical significance at the 95% level of the
 14 difference in the mean values of SAX words detected with and without embedded true
 15 motifs can only be shown for the difference in script 3 (mean no. with true motifs (SD) =
 16 110.6 (30.0); mean no. without true motifs (SD) = 59.2 (24.1); Mann Whitney U = 2.0, p (1-
 17 tailed) = 0.01) the effect sizes for the differences ($r = 0.3$ (small) for script 1; $r = 0.3$ (small)
 18 for script 2 and $r = 0.7$ (large) for script 3) suggest that there are generally a measurably
 19 larger number of SAX words found when true motifs are embedded in the scripts.

20

1 **Discussion**

2 *Device testing*

3 The device testing study was designed to give a preliminary understanding of whether the
4 recording system was likely to be acceptable to users and functional as a stand-alone system
5 in a patient's home environment. Our findings, based on this small scale study, were that
6 the device was generally acceptable. Patients did not appear to find it intrusive to have the
7 accelerometer fixed to the TMJ and were usually able to accommodate the digital recorder
8 within a pocket. Carers did not report any distress or unusual behaviour from patients when
9 wearing the system. The accelerometer remained securely attached to the TMJ for all
10 patients over the course of the test day and the recorder had sufficient memory and battery
11 capacity to cope with acquiring and storing the data for that duration. Recorded data had a
12 satisfactory signal to system-noise ratio and was of sufficient quality to permit automated
13 analysis. Aural assessment of the recorded signals suggested that the rejection of external
14 noise, including speech of conversational partners was robust and speech content was not
15 readily identifiable. We therefore suggest that the device meets the requirements we set
16 out for recording a data stream that represents speech behaviour but maintains due regard
17 for privacy of both the user and their interlocutors.

18 *Algorithm testing*

19 The motif discovery algorithm worked well on script 1, script 2 motif #1 and script 3 motif #1
20 detecting between 60 and 75% of all true motifs. Its performance was less good on script 2
21 motif #2 and script 3 motif #2 dropping to 49 and 47% respectively on these motifs. In order
22 to achieve good detection a number of parameters within the algorithms need to be chosen
23 *a priori*: the Euclidean distance threshold below which sub-sequences are considered to be

1 similar, the PAA word size, the SAX alphabet size and the minimum duration of a sub-
2 sequence, T_{\min} , that can be considered as a member of a motif. For the majority of these
3 values we were guided by values chosen in the literature for similar applications (26, 27).
4 For T_{\min} we used the empirically selected values of 108 ms for scripts 1 and 2 and 54 ms for
5 script 3. The good detection rates for the true motifs for script 1 and for the first embedded
6 motif in scripts 2 and 3 suggest that the chosen values were adequate. For detection in
7 scripts 2 and 3, a single value of T_{\min} was used despite there being two true motifs with
8 potentially rather different durations. T_{\min} was closer to the length for motif #1 in each case
9 and therefore did not perform so well at detecting examples of motif #2. It would be
10 possible to search for motifs in a given script with multiple values of T_{\min} (26), but the
11 resulting additional processing time could be significant. Further, accuracy of detection of
12 true motifs was much improved if data for one of the subjects, who spoke noticeably more
13 slowly than the others, were neglected; for example, for script 2, average accuracy
14 increased to 87.5% (25%) for motif #1 and 61.1% (41.1%) for motif #2 when this subject was
15 not included. Evidently some tuning of the parameters for individuals, whether in some way
16 algorithmic or manual, might be required in any real-life implementation.

17 Our primary concern was to show in principle that true motifs could be automatically
18 detected from the accelerometric speech signal using only the first principal component of
19 the feature vector and we did not therefore focus here on further optimisation of the
20 parameters of the data mining algorithms.

21 To estimate the likely specificity of the algorithm we compared the detection of similar data
22 sub-sequences in the time series from which the true motifs had been removed, with their
23 detection in unedited data sets. Naturally we expect there to be repeated phrases in healthy

1 speech (e.g. 'you know' or 'sort of' as well as words and phrases specific to a given topic of
2 conversation) and these may also be detected by the algorithm as motifs. We found a
3 measurable effect of more SAX words detected in the data containing embedded true
4 motifs than in the data where they had been edited out, and in one script (script 3) the
5 difference was statistically significant and the effect size large. The observed differences
6 suggest that with sufficient normative data we could define an upper threshold for the
7 number of detected SAX words in typical speech patterns in the absence of perseveration
8 and use this as a metric for determine the presence of abnormal repetition.

9 To improve the detection of repeated motifs and to apply it to unscripted speech samples
10 obtained from patients, a number of possible modifications may be considered to improve
11 the performance of the algorithm: Firstly, the features commonly used to characterize air-
12 conducted speech may not provide an optimal description of a data series obtained using an
13 accelerometer. Whether a better set of features can be specified for such data remains to
14 be determined. Secondly, the performance of the motif discovery algorithm was clearly
15 influenced by the choice of the sub-sequence length (T_{min}) and data-driven methods for
16 selecting an optimal value for T_{min} for any given recording should be developed. Thirdly, the
17 parameters of the motif discovery algorithm (word and alphabet size) and criteria of
18 similarity (thresholds for the Euclidean distance between a given pair of sub-sequences) that
19 work best will require empirical evaluation in larger and more complex sets of speech data.
20 Finally, metrics for normal and abnormal repetition must be developed to allow
21 discrimination between perseverative speakers and those engaged in normal conversation.

22

23 CONCLUSIONS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

In this preliminary study we have presented the technical details of a system that records a continuous signal related to the speech signal while preserving the privacy of the monitored individual and his/her conversational partners. We have shown that the proposed system can be used safely and is well tolerated by one group of patients in whom speech monitoring could give rise to clinically useful information – namely patients in the early stages of probable AD. In addition, we have described and implemented an algorithm for automatic processing of the recorded signal in order to identify any repeated patterns embedded within the signal and to count their occurrences, and estimated the accuracy of the algorithm in automatic detection of patterns corresponding to phrases that were deliberately embedded repeatedly within scripted speech samples. In order to develop this system further, a large scale study of normal and perseverative speech is required in order to set upper thresholds for expected amounts of repetition in the daily conversational speech of healthy participants and to set empirical values for the parameters in the motif discovery algorithm that optimise the detection of abnormal repetition.

Conflict of Interest

The authors whose names are listed as authors of this paper certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

1

2 **Acknowledgements**

3 The study reported in this paper was designed by Garrard (clinical lead) and Barney
4 (technical lead). Patient data collection and analysis was carried out by Nemes. Healthy
5 participant data was collected by Nemes and Nikolic. The motif discovery software was
6 implemented by Nikolic who also analysed the healthy participant data. The paper was
7 drafted jointly by Barney, Garrard and Nikolic. The work was supported by a Medical
8 Research Council Discipline Hopping award [Ref. G0902237] to PG and AB.

9

10 **References**

- 11 1. Matthews F, Brayne C. The Incidence of Dementia in England and Wales: Findings from the
12 Five Identical Sites of the MRC CFA Study. *PLoS Med.* 2005;2(8).
- 13 2. Comas-Herrera A, Wittenberg R, Pickard L, Knapp M. Cognitive impairment in older people:
14 future demand for long-term care services and the associated costs. *International Journal of*
15 *Geriatric Psychiatry.* 2007;22(10):1037-45.
- 16 3. Brayne C. Incidence of dementia in England and Wales - The MRC cognitive function and
17 ageing study. *Alzheimer Disease & Associated Disorders.* 2006;20(3):S47-S51.
- 18 4. Lakey L, Chandaria K, Quince C, Kane M, Suaunders T. *Dementia 2012: A national challenge.*
19 London: Alzheimer's Society, 2012.
- 20 5. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive
21 impairment: a concept in evolution. *Journal of internal medicine.* 2014;275(3):214-28.
- 22 6. Hoglund K, Fourier A, Perret-Liaudet A, Zetterberg H, Blennow K, Portelius E. Alzheimer's
23 disease - Recent biomarker developments in relation to updated diagnostic criteria. *Clinica chimica*
24 *acta; international journal of clinical chemistry.* 2015.
- 25 7. Hampel H, Burger K, Teipel SJ, Bokde AL, Zetterberg H, Blennow K. Core candidate
26 neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimer's & Dementia.*
27 2008;4(1):38-48.
- 28 8. Laske C, Sohrabi HR, Frost SM, Lopez-de-Ipina K, Garrard P, Buscema M, et al. Innovative
29 diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's & dementia : the journal of*
30 *the Alzheimer's Association.* 2015;11(5):561-78.
- 31 9. Schmidt R, Freidl W, Fazekas F, Reinhart B, Grieshofer P, Koch M, et al. The Mattis-Dementia-
32 Rating-Scale - Normative Data from 1,001 Healthy-Volunteers. *Neurology.* 1994;44(5):964-6.
- 33 10. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading
34 the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12(3):189-98.
- 35 11. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive
36 Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International*
37 *journal of geriatric psychiatry.* 2006;21(11):1078-85.
- 38 12. Cohen G. *Memory in the Real World.* Hove, E. Sussex: Psychology Press; 1996.
- 39 13. Pekkala S, Albert ML, Spiro A, Erkinjuntti T. Perseveration in Alzheimer's disease. *Dementia*
40 *and Geriatric Cognitive Disorders.* 2008;25(2):109-14.

- 1 14. Bayles KA, Tomoeda CK, McKnight PE, Helm-Estabrooks N, Hawley JN. Verbal perseveration
2 in individuals with Alzheimer's disease. *Seminars in Speech and Language*. 2004;25(4):335-47.
- 3 15. Nakayama M, Ishimitsu S, Nakagawa S. A Study of Making Clear Body-Conducted Speech
4 Using Differential Acceleration. *Ieee T Electr Electr*. 2011;6(2):144-50.
- 5 16. Monson BB, Hunter EJ, Lotto AJ, Story BH. The perceptual significance of high-frequency
6 energy in the human voice. *Front Psychol*. 2014;5.
- 7 17. Muscariello A, Gravier G, Bimbot F. Unsupervised Motif Acquisition in Speech via Seeded
8 Discovery and Template Matching Combination. *Ieee T Audio Speech*. 2012;20(7):2031-44.
- 9 18. Muscariello A, Gravier G, Bimbot F. Audio keyword extraction by unsupervised word
10 discovery. *Interspeech 2009: 10th Annual Conference of the International Speech Communication*
11 *Association 2009, Vols 1-5*. 2009:2811-4.
- 12 19. Brazma A, Jonassen I, Eidhammer I, Gilbert D. Approaches to the automatic discovery of
13 patterns in biosequences. *J Comput Biol*. 1998;5(2):279-305.
- 14 20. McBride M, Tran P, Letowski T, Patrick R. The effect of bone conduction microphone
15 locations on speech intelligibility and sound quality. *Appl Ergon*. 2011;42(3):495-502.
- 16 21. Goodglass H, Kaplan E, Barresi B. Innovations in aphasia testing: Preview of the new BDAE.
17 *Brain and language*. 1998;65(1):27-30.
- 18 22. Boll SF. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *Ieee T Acoust*
19 *Speech*. 1979;27(2):113-20.
- 20 23. Ephraim Y, Malah D. Speech Enhancement Using a Minimum Mean-Square Error Short-Time
21 Spectral Amplitude Estimator. *Ieee T Acoust Speech*. 1984;32(6):1109-21.
- 22 24. Eyben F, Wollmer M, Schuller B, editors. *OpenEAR – Introducing the Munich Open-Source*
23 *Emotion and Affect Recognition Toolkit. Affective Computing and Intelligent Interaction - ACII; 2009;*
24 *Amsterdam*.
- 25 25. Lartillot O, Toivainen P, Eerola T. A Matlab Toolbox for Music Information Retrieval. . In:
26 Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R, editors. *Data Analysis, Machine Learning and*
27 *Applications, Studies in Classification, Data Analysis, and Knowledge Organization*. Berlin: Springer-
28 *Verlag; 2008*.
- 29 26. Tanaka Y, Uehara K. Discover motifs in multi-dimensional time-series using the principal
30 component analysis and the MDL principle. *Lect Notes Artif Int*. 2003;2734:252-65.
- 31 27. Lin J, Vlachos M, Keogh E, Gunopulos D, Liu JW, Yu SJ, et al. A MPAA-based iterative
32 clustering algorithm augmented by nearest neighbors search for time-series data streams. *Advances*
33 *in Knowledge Discovery and Data Mining, Proceedings*. 2005;3518:333-42.
- 34 28. Lin J, Keogh E, Wei L, Lonardi S. Experiencing SAX: a novel symbolic representation of time
35 series. *Data Min Knowl Disc*. 2007;15(2):107-44.
- 36 29. Lin J, Keogh E, Lonardi S, Patel P, editors. *Finding motifs in time series. 2nd Workshop on*
37 *Temporal Data Mining; 2002; Edmonton, Canada*.
- 38 30. Chen JS, Moon YS, Yeung HW. Palmprint authentication using time series. *Audio and Video*
39 *Based Biometric Person Authentication, Proceedings*. 2005;3546:376-85.
- 40 31. Shasha D, Wang TL. New Techniques for Best-Match Retrieval. *Acm T Inform Syst*.
41 1990;8(2):140-58.
- 42 32. Smith, J. O., (2011). *Spectral Audio Signal Processing*,
43 <https://www.dsprelated.com/freebooks/sasp/>, online book, Section 9.6 WOLA, accessed
44 21/04/2016.

45

46

47 **Table legends**

48 *Table 1: Average performance for the motif detection system over all subjects*

1 *Table 2: Average reduction in the percentage of detected Symbolic Aggregate*
2 *Approximation (SAX) words between recordings with and without embedded true motifs.*

3

4 **Figure legends**

5 *Figure 1: Data acquisition system used for the device and algorithm testing studies. In the*
6 *device testing study with AD patients there was no microphone and the accelerometer*
7 *signal was recorded on a portable digital recorder and later transferred to the PC for*
8 *processing. Recorded signals from the accelerometer and the microphone were sampled at*
9 *a frequency of 16 kHz (12-bit resolution).*

10 *Figure 2: (a) Accelerometric speech signal with two repeated segments (shaded background)*
11 *identified for illustration and (b) first principal component extracted from the 95-D acoustic*
12 *feature vector.*

13

14 *Figure 3: a) Subsequence extracted from the first principal component of the 95-D feature*
15 *vector with length $T_{min} = 50$ ms; b) z-score normalised sequence divided into 6 sections with*
16 *(red line) PAA coding applied by replacing the data in each section with its mean value and*
17 *c) division of the amplitude into 4 areas, based on equal probability, indicated by dotted*
18 *lines and SAX coding of the PAA signal using an alphabet of 4 symbols according to which*
19 *division it falls into such that the SAX code for this subsequence is: bdcbab.*

20

21 *Figure 4: An illustration of detected motifs in a script with embedded true motifs (script 3,*
22 *subject 5).*