AUTOREGRESSIVE FEATURES FOR A THOUGHT-TO-SPEECH CONVERTER

N. Nicolaou, J. Georgiou and M. Polycarpou
Department of Electrical and Computer Engineering, University of Cyprus, 75 Kallipoleos Street, Cyprus

Keywords: Brain-Computer Interface, electroencephalogram, Morse Code, thought communication, speech impairment.

Abstract: This paper presents our investigations towards a non-invasive custom-built thought-to-speech converter that decodes mental tasks into morse code, text and then speech. The proposed system is aimed primarily at people who have lost their ability to communicate via conventional means. The investigations presented here are part of our greater search for an appropriate set of features, classifiers and mental tasks that would maximise classification accuracy in such a system. Here Autoregressive (AR) coefficients and Power Spectral Density (PSD) features have been classified using a Support Vector Machine (SVM). The classification accuracy was higher with AR features compared to PSD. In addition, the use of an SVM to classify the AR coefficients increased the classification rate by up to 16.3% compared to that reported in different work, where other classifiers were used. It was also observed that the combination of mental tasks for which highest classification was obtained varied from subject to subject; hence the mental tasks to be used should be carefully chosen to match each subject.

1 INTRODUCTION

The development of techniques that offer alternative ways of communication by bypassing conventional means is an important and welcome advancement for improving quality of life. This is especially desirable in cases where the conventional means of communication, such as speech, is impaired. We envisage the development of a simple and wearable system that communicates by converting thoughts into speech via morse code and a text-to-speech converter.

In this paper we present preliminary investigations towards the development of such a system. The investigations form part of our search for features, classifiers and mental tasks that are appropriate for utilisation in our system. In particular, we compare the classification accuracy obtained between combinations of mental task pairs when (i) autoregressive (AR) coefficients and Power Spectral Density (PSD) values are utilised as features; and (ii) Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Neural Network (NN) are utilised as classifiers. Our investigations suggest that the combination of AR coefficients and SVM is more appropriate for our application, as an increase in classification accuracy ranging from 8.2-16.3% has been observed compared to classification of the same features using LDA and NN.

The paper is organised as follows. Section 2 provides a background into communication via thoughts and how morse code has been utilised for this purpose so far. This is followed by section 3 where a description of the system envisaged, the objectives that motivated these preliminary investigations and a description of the methods utilised are provided. The findings are presented in section 4 followed by a discussion towards how these could be interpreted and understood as part of the proposed system. The main conclusions and plans for future work emerging from these investigations are outlined in section 5.

2 BACKGROUND

A number of conditions, such as amyotrophic lateral sclerosis, strokes and speech impairment, affect the ability to communicate with the environment through speech. The problem becomes more severe when limb or muscle control is also affected, since other means of communication e.g. typing, are eliminated. An alternative method of communication

Nicolaou N., Georgiou J. and Polycarpou M. (2008).
AUTOREGRESSIVE FEATURES FOR A THOUGHT-TO-SPEECH CONVERTER.
In *Proceedings of the First International Conference on Biomedical Electronics and Devices*, pages 11-16
DOI: 10.5220/0001052900110016

is achieved by utilising brain activity as an input signal to a device for spelling purposes (brain-computer interface, BCI). A BCI is "a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles" (Wolpaw et al., 2000). This technology is primarily aimed at people who have lost conventional means of communication, but whose brain function remains intact.

Current BCI applications are limited by the trade-off between speed and accuracy. Thus, the most common application still remains 1-dimensional cursor movement on a computer screen, which offers the ability to communicate with the environment when teamed with a "virtual keyboard". Communication can be achieved by mentally controlling cursor movement on the screen for choosing letters on a "virtual keyboard" (Wolpaw et al, 2002) or to highlight the desired character from a scrolling list (Scherer et al., 2004). Different mental tasks are associated with left/right and/or up/down cursor movement, thus allowing the subject to pick characters and spell words. Despite the simplicity of these applications, current BCI systems are faced with som: (i) 25 bits/min is the maximum speed of communication reported (Vaughan et al., 2003). If we consider a character with 8 bit resolution this is equivalent to 3.13 chars/min, which is not acceptable for normal speech; and (ii) current systems are bulky and non portable. It is envisaged that the development of custom-built hardware as part of the proposed system will provide a solution to both these issues. In addition, these can be aided if the "virtual keyboard" is substituted by a simplified set of characters whose choice is directly associated with particular mental tasks, thus eliminating the intermediate step of cursor movement.

Such a potential simplification could be achieved via the use of Morse Code (MC), which has already been utilised for communication for disabled people. In MC transmission of information is based on short and long elements of sound (dots and dashes) and was originally created for telegraph communication. The elegance of MC lays in its simplicity and the high speech reception and transmission rates. A skilled MC operator can receive MC in excess of 40 words per minute (Coe, 2003). The world record for understanding MC was set in 1939 and still stands at 75 words per minute (French, 1993). Utilisation of MC for the disabled is commonly based on some form of muscle movement, such as operating a switch (Park et al, 1999) or a sip-puff straw (Levine et al., 1986). However, certain disabilities affect muscle movement, but even if not, then such systems are difficult to operate on a daily basis as they cause fatigue.

The use of MC for directly translating thoughts into words has been considered in very few BCI systems, mainly as an extension to traditional BCI communication methods. In (Palaniappan, 2005) the "virtual keyboard" was substituted with the two MC elements, "." and "-", and the user chose through mentally controlling cursor movement. Another MC-BCI system is described in (Altschuler and Dowla, 1998) based on the attenuation of power in the μ band (8-13Hz) during motor imagery, whose duration corresponds either to a "."or a "-" (shorter or longer motor imagery duration respectively). Spelling is achieved by interchanging motor imagery with baseline task (representing a "pause"). In addition, (Huan and Palaniappan, 2004) showed how communication in a BCI system could conceptually be achieved via a tri-state MC scheme and utilising a fuzzy ARTMAP as classifier. In such a system a ".", a "-" or a "space" would be represented by 3 mental tasks and the continuous EEG would be sampled every, e.g., 0.5s, for decision making. In (Huan and Palaniappan, 2002) it is stated that the conversion of a mental task into one of the 3 MC elements would take 6ms of computation time; however this heavily depends on a number of operating system factors.

The concept behind the latter two systems is closer to the concept of the proposed system, as the intermediate step of cursor movement is eliminated. The use of MC is advantageous as it simplifies the dictionary to 3 symbols, the choice of which will be achieved through 2 mental tasks. This reduces the system complexity and improves communication speed. Hence, we envisage the development of a portable, embedded, custom and wearable MC-based BCI system that could be used either as an assistive or as an enhancing communication aid.

3 PERFORMANCE OPTIMISATION

The proposed system is shown in figure 1 and consists of 4 parts: (1) EEG signals are recorded from a patient performing two mental tasks, each corresponding to either a "." and "-" (depending on the task duration) or a "pause". The patient is

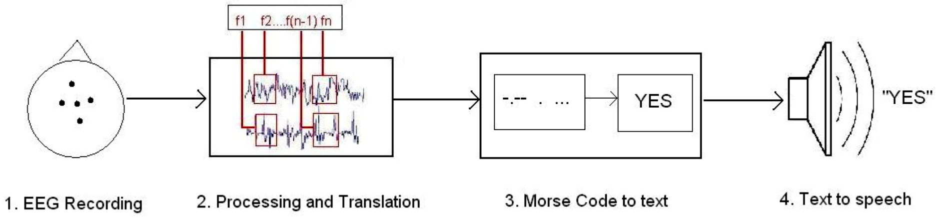

Figure 1: The proposed MC-BCI system.

mentally spelling letters and words in MC; (2) windows of specified duration of the recordings are processed and classified as ".", "-" or "pause"; (3) MC is then converted into text, which is in turn converted to speech via a text-to-speech converter (4). At this stage our priority is to maximise correct interpretation of EEG data. Computational efficiency is not a key consideration as we will be designing custom hardware tailored to the chosen processing methods. Therefore, it is imperative to firstly converge on a particular combination of signal processing methods that could be used reliably in the proposed system. The preliminary investigations presented in this paper are associated with part 2 of the proposed system and are part of our greater search for the optimal combination of features and classifiers.

3.1 Methods

3.1.1 Feature extraction

AR models are commonly utilised in EEG analysis (Wright et al., 1990). More specifically, the estimated AR coefficients have been shown to capture well the differences between various mental tasks, and as a result are frequently used as features in mental task classification and BCIs (Guger et al., 2000). Eq. 1,

$$x_t = \sum_{\tau=1}^{p} a_\tau x_{t-\tau} + \varepsilon_t \quad (1)$$

represents an AR(p) model where p is the model order, x_t is the time series to be modelled, a_τ, τ=1,…,p are the estimated coefficients of the p^{th}-order AR model and ε_t is zero-mean random noise (commonly Gaussian with unit variance). In EEG analysis an AR(p) is fitted to the data and the p^{th} dimensional vector of estimated coefficients represents the different mental tasks, as a variation of the coefficients depending on the mental task is observed. The AR model order used in EEG analysis ranges from 5 up to 13 (Lopes daSilva, 1998). For the specific dataset used here an order of 6 was chosen as suggested in (Keirn and Aunon, 1990). Estimation of the coefficients is possile via a number of ways – here we used the method of Least Squares.

The second set of features utilised is PSD values obtained via parametric spectral analysis. In particular an AR(p) model (here p=6) is first fitted on the data and the power spectrum is subsequently obtained from the estimated coefficients via

$$S(f) = \frac{\hat{\sigma}_p^2}{N \sum_{k=0}^{p} a_k e^{-j2\pi fk/N}} \quad (2)$$

where a_k, k=1,…,p are the estimated coefficients, f is a vector of chosen frequencies, $\hat{\sigma}_p^2$ is the estimated noise variance and N is the number of samples. The advantage of parametric methods for spectrum estimation is the ability to specify a set of frequencies of interest over which the spectrum is estimated.

3.1.2 Classification

The choice of the classifier should have little effect on the classification rate if the chosen features are good representations of the data to be classified. Given that the features capture the data characteristics well, then classification becomes an easier problem. However, the properties of the classifier must be well-matched to the feature dimensionality or separability (linear or non-linear). The problem of choosing a classifier is enhanced if the feature dimensionality is high, as this does not allow the visualisation of the features and, consequently, whether they are linearly separable or not.

SVMs offer a solution to this issue, as both linear and non-linear classification can be obtained simply by changing the "kernel" function utilised

(Burges, 1998). Due to the fairly new development of SVMs they are not commonly utilised in BCI systems (see (Gysels and Celka, 2004) for an example). Thus, their performance for mental task classification has not been widely assessed and their application in such systems can be considered novel.

SVMs belong to the family of kernel based classifiers. The main concept of SVMs is to implicitly map the data into the feature space where a hyperplane (decision boundary) separating the classes may exist. This implicit mapping is achieved via the use of Kernels, which are functions that return the scalar product in the feature space by performing calculations in the data space. The simplest case is a linear SVM trained to classify linearly separable data. After re-normalisation, the training data, $\{x_i, y_i\}$ for i=1, …, m and $y_i \in \{-1,1\}$, must satisfy the constraints

$$\mathbf{x}_i \mathbf{w} + b \geq +1 \quad \text{for } y_i = +1 \tag{3}$$

$$\mathbf{x}_i \mathbf{w} + b \leq -1 \quad \text{for } y_i = -1 \tag{4}$$

where $\boldsymbol{w}$ is a vector containing the hyperplane parameters and b is an offset. The points for which the equalities in the above equations hold have the smallest distance to the decision boundary and they are called the support vectors. The distance between the two parallel hyperplanes on which the support vectors for the respective classes lie is called the *margin*. Thus, the SVM finds a decision boundary that maximises the margin. Finding the decision boundary then becomes a constrained optimization problem amounting to minimisation of $\|w\|^2$ subject to the constraints in (3) and (4) and is solved using Lagrange optimisation framework. The general solution is given by

$$f(x) = \sum_i \alpha_i y_i \langle x_i, x \rangle \tag{5}$$

In the case of non-linear classification, Kernels (functions of varying shapes, e.g. polynomial or Radial Basis Function) are used to map the data into a higher dimensional feature space in which a linear separating hyperplane could be found. The general solution is then of the form:

$$f(x) = \sum_i \alpha_i y_i K \langle x_i, x \rangle \tag{6}$$

Depending on the choice of the Kernel function SVMs can provide both linear and non-linear classification, hence a direct comparison between the two can be made without having to resort to utilisation of different classifiers.

3.1.3 Data

At this stage we utilise EEG data that is available online. The dataset chosen is well-known and has been used in various BCI applications. It contains EEG signals recorded by Keirn and Aunon during 5 mental tasks and is available from (http://www.cs .colostate.edu/~anderson). Each mental task lasted 10s and subjects participated in recordings over 5 trials and a number of sessions (subjects 2 and 7 participated in 1 session, subject 5 in 3 and subjects 1, 3, 4 and 6 in 2). The data was recorded with a sampling rate of 250Hz from 6 EEG electrodes placed at locations C3, C4, P3, P4 and O1 (more details on the recording protocol can be found in (Keirn and Aunon, 1990)). The 5 mental tasks are: (1) Baseline: subjects are relaxed and should be thinking of nothing particular; (2) Multiplication: subjects are asked to perform non-trivial mental multiplication problems; it is highly likely that a solution was not arrived at by the end of the allocated recording time; (3) Rotation: a 3-dimensional geometric figure is shown on the screen for 30s, after which the subjects are asked to mentally rotate the figure about an axis; (4) Letter composition: subjects are asked to mentally compose a letter, continuing its composition from where it was left off at the end of each trial; and (5) Counting: subjects are asked to count sequentially by imagining the numbers being written on a blackboard and rubbed off before the next number is written. In each trial counting resumes from where it was left off in the previous trial.

This dataset has been chosen for two reasons. Firstly, it contains recordings from mental tasks that are traditionally associated with BCI systems. Secondly, it allows the investigation of a large combination of mental task pairs as it contains recordings from 5 different tasks – this will allow us to identify whether the choice of tasks depends on the subject and whether other non-traditional tasks should also be investigated. In addition, a third good reason is that it allows direct comparison with results from the literature.

4 RESULTS

To allow a direct comparison of the results with those presented in (Huan and Palaniappan, 2004), we used data from 2 sessions and 4 subjects

(subjects 1, 3, 5 and 6). The data was split in non-overlapping segments of 0.5s duration, resulting in 200 segments per task per subject, over 2 sessions. The SVM classification rate was averaged over 10 trials, where in each trial a randomly chosen set of 100 segments was used for training, with the remaining segments used for testing. All 10 pair combinations of the 5 mental tasks were classified and the pair of tasks with the maximum average classification rate for each subject was identified. The average classification rate was estimated as $(TP_1+TP_2)/2$, where TP_i (true positive) is the number of segments classified correctly for mental task *i*. The feature vectors describing each 0.5s segment are 36-dimensional in the AR(6) case and 300-dimensional in the PSD values case (6 AR coefficients and 50 PSD values per electrode; the final feature vectors consisted of the concatenated AR coefficients and PSD values for all electrodes respectively).

The classification results for the AR(6) features are presented in table 1. It can be seen that the choice of classifier had a positive effect on the classification accuracy. The use of an SVM increased the accuracy by up to nearly 13% compared to that obtained for the same features using LDA and by up to 16.3% using an NN (see table 2 for details), as presented in (Huan and Palaniappan, 2004). In theory, the choice of classifier has a smaller effect on the classification rate if the features utilised represent the data well. Nonetheless, the use of an SVM with RBF Kernel increases the classification rate by a large margin and, hence these results indicate that the use of an SVM is more appropriate for these features. In addition, the pair of tasks which provided the highest average classification was different than the equivalent pair from (Huan and Palaniappan, 2004). However, it was also observed that the task pair which gave highest average classification varied with each subject, in agreement with (Huan and Palaniappan, 2004). Hence a particular task pair for which optimal operation can be obtained should be identified for each subject. In addition, performance could be improved if the tasks utilised had a more intuitive connection with the way of thinking associated with MC.

The classification rates for the PSD features are presented in table 3. The rates obtained are much lower than the ones reported in (Palaniappan et al., 2002). This could be attributed to three reasons. Firstly, in this work classification between pairs of tasks was obtained as opposed to between 3 tasks as in (Palaniappan et al., 2002) hence a direct comparison is not appropriate. Secondly, the PSD features are already of high dimension (300-dimensional) and an SVM may not be appropriate for classification when the feature space is already of high dimension. Thirdly, the classification rates presented in (Palaniappan et al., 2002) were averaged for a single training set whose ordering of the training patterns was randomly varied 10 times, hence the high classification rate reported may have been a side-effect of the particular choice of training set. In addition, another issue with utilisation of PSD values as features is the partial spectrum overlarp of certain artefacts (such as eye movements) with EEG activity, which can potentially adversely affect the classification rate.

Table 1: Maximum average classification rate (%) for AR(6) features with SVM. Results presented are averaged over 10 trials.

Subj.	Class. Rate	Tasks	Kernel
1	88.4	Letter vs multiplication	RBF
3	87.9	Letter vs counting	RBF
5	83.9	Roration vs counting	RBF
6	92.4	Counting vs multiplication	Linear

Table 2: Maximum average classification rate (%) for AR(6) features. Column 2 presents our results, while columns 3 and 4 give the best rates presented in (Huan and Palaniappan, 2004) for LDA and NN.

Subj.	SVM	LDA	NN
1	88.4	80.2	78.9
3	87.9	73.6	73.9
5	83.9	71.4	67.6
6	92.4	84.3	77.6

Table 3: Maximum average classification rate (%) for power spectrum values with SVM. Results presented are averaged over 10 trials.

Subj.	Class. Rate	Tasks	Kernel
1	58.0	Letter vs multiplication	RBF
3	56.6	Letter vs counting	RBF
5	68.0	Roration vs counting	RBF
6	60.2	Counting vs multiplication	Polyno-mial

The feature vectors were created by concatenating the estimated AR coefficients from all 6 electrodes. However, the wearability and portability of an MC-based BCI is facilitated by employing a small number of electrodes –ideally two, or even a single, electrode(s). It may be possible to obtain higher classification rates by utilising a single electrode that is more relevant to the specific mental task rather than using a combination of electrodes, all of which are not as relevant to the task. This is also advantageous as it decreases the feature dimensionality.

5 CONCLUSIONS

This paper presents the results of initial investigations in the search for appropriate features and classifier towards the development of a thought-to-speech converter. The results indicate that the use of an SVM for the classification of AR coefficients is more appropriate than LDA and NN and will be utilised in the development of the proposed system.

The proposed system is promising as it offers the ability to communicate more efficiently via direct conversion of thoughts into speech. In order to ensure optimal operation other aspects of the system must also be investigated. Firstly, a more extensive set of features and classifiers will be examined such that the optimal combination in terms of maximising accuracy is determined – computational efficiency is not a consideration as the system will be customised and capable of parallel processing. Secondly, these investigations suggest that different combinations of mental tasks seem to be more appropriate for different subjects. We are going to look into finding a combination of tasks that are more intuitive and more closely related to the concept of MC, as this could improve classification accuracy and facilitate easier operation.

REFERENCES

Altschuler, E.L., and Dowla, F.U., 1998. Encephalolexianalyzer. United States Patent Number: 5,840,040. November 24.

Burges, C. J. C., 1998. A tutorial on Support Vector Machines for Pattern Recognition. In *Data Mining and Knowledge Discovery*, U. Fayyad, Ed. Boston: Kluwer Academic Publishers, pp. 121-167.

Coe, L., 2003. Telegraph: A History of Morse's invention and its predecessors in the United States. McFarland & Company.

French, T., 1993. McElroy, World's Champion Radio Telegrapher. Artifax Books.

Guger, C., Schlogl, A., Neuper, C., Walterspacher, D., Strein, T., and Pfurtscheller, G., 2000. Rapid Prototyping of an EEG-based Brain-Computer Interface (BCI). In *IEEE Trans. on Neural Systems and Rehab. Eng.,* Vol. 9, issue 1, pp.49-58, March.

Gysels, E., and Celka, P., 2004. Phase synchronisation for the recognition of mental tasks in a brain-computer interface. In *IEEE Trans. on Neural Systems and Rehab. Eng.*, Vol. 12, issue 4, pp. 406-415.

Huan, N.-J., and Palaniappan, R., 2004. Neural network classification of autoregressive features from electroencephalogram signals for brain-computer interface design. In *Journal of Neural Eng.*, Vol. 1, pp.142-150.

Keirn, Z.A., and Aunon, J.I., 1990. A new mode of communication between man and his surroundings. In *IEEE Trans. Biomed. Eng.,*, Vol. 37, pp.1209-1214.

Levine, S.P., Gauger, J.R.D., Bowers. L.D., and Khan, K.J., 1986. A comparison of Mouthstick and Morse code text inputs. In *Augmentative and Alternative Communication*, Vol. 2, issue 2, pp.51-55.

Lopes da Silva, F., 1998. EEG analysis: Theory and Practice. In *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, Ch. 6, pp.1135-1163.

Palaniappan, R., Paramesan, R., Nishida, S., and Saiwaki, N., 2002. A new brain-computer interface design using Fuzzy ARTMAP. In *IEEE Trans. On Neural Systems and Rehab. Eng.*, Vol. 10, issue 3, pp.140-148.

Palaniappan, R., 2005. Brain computer interface design using band powers extracted during mental tasks. In *Procs. of the 2nd International IEEE EMBS Conf. on Neural Eng.,* Arlington, Virginia. March 16-19.

Park, H.-J., Kwon, S.-H., Kim, H.-C., and Park, K.-S., 1999. Adaptive EMG-driven communication for the disability. In *Procs. of 1st Joint BMES/EMBS Conf. Serving Humanity, Advancing Technology.* Atlanta, USA, October 13-19. p.656.

Scherer, R., Muller, G. R., Neuper, C., Graimann, B., and Pfurtscheller, G., 2004. An Asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. In *IEEE Trans. on Biomed. Eng.*, Vol. 51, issue 6, pp.979-984.

Vaughan, T.M., et al., 2003. Brain-computer Interface Technology: a review of the second international meeting (Guest Editorial). In *IEEE Trans. on Neural Systems and Rehab. Eng.*, Vol. 11, issue 2, pp.94-109.

Wolpaw, J.R., et al., 2000. Brain-Computer Interface Technology: a review of the First International meeting. In *IEEE Trans. on Rehab. Eng.*, Vol. 8. issue 2, pp.164-173.

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., and Vaughan, T.M., 2002. Brain-Computer Interfaces for communication and control. In *Clinical Neurophysiology*, Vol. 113, pp.767-791.

Wright, J.J., Kydd, R.R., and Sergejew, A.A., 1990. Autoregression Models of EEG. In *Biological Cybernetics*, Vol. 62, pp.201-210.