33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Ensemble classification for robust discrimination of multi-channel, multi-class tongue-movement ear pressure signals

Michael Mace, Khondaker Abdullah-Al-Mamun, Shouyan Wang, Lalit Gupta and Ravi Vaidyanathan

Abstract— **In this paper we introduce a robust classification framework for tongue-movement ear pressure signals based around an ensemble voting methodology. The ensemble members are comprised of different combinations of sensor inputs i.e. two in-ear microphones and an acoustic gel sensor positioned under the chin of the individual and classification using three different base models. It is shown that by using all nine ensemble members when compared to the individual (base) models, the average misclassification rate can be reduced from 23% to 2.8% when using the majority voting strategy. The correct classification rate is improved from 76% to 92.4% when utilizing either the borda count or condorcet methods. This is achieved through a combination of rejection based on ambiguity in the ensemble and diversity in the misclassified instances across the ensemble members.**

I. INTRODUCTION

Afflictions of the sensory-motor system both physical and neurological can profoundly inhibit human movement. The extremities tend to be at higher risk due to their inherent distance from the brain and body, thus increasing the potential for severing of the peripheral nervous system and/or limbs. Upper extremity motor loss can be induced by spinal cord injury (SCI), paraplegia, congenital limb deformities and stroke to name a few. Over the last few decades, a multitude of research has been conducted towards providing novel solutions that replace or compensate these degraded pathways. One particular area of interest is providing an individual with new ways for communicating with assistive technologies. This involves thinking of new and creative methods by which a user can express their intention and thereby control peripheral devices. The use of the head, tongue, eyes and brain in providing these new communication pathways, have been employed by researchers, due to the robust functionality of these craniofacial regions under said conditions.

Recently a non-invasive tongue based communication system has been developed, based around tongue-movement ear pressure (TMEP) signals [1]. The sensory unit is centered on a microphone positioned within the user's external acoustic meatus with user intention expressed through prescribed flicks of the tongue. These impulsive motions create unique low frequency (0 - 100 Hz) bio-acoustic pressure signals within the auditory cavity, allowing for inter-action discrimination and also discrimination from naturally occurring acoustic signals. Currently four actions have been defined and involve placement of the tip of the tongue at the base of the central incisor, left or right first molar and flicking the tongue up (bottom/left/right action) and placing the tip of the tongue against the top of the palate and flicking down (top action). This action set was chosen due to the tongue motions not normally occurring in daily activity, yet the actions themselves feel natural whilst executing, ensuring repeatability. Previous work has shown inter-class classification results of various algorithms using four a-priori collected data-sets [1], discrimination between controlled and non-controlled movements based on the signal frequency content extracted using a wavelet packet transform [2] and initial real-time classification across three of the actions [3]. All this work has concentrated on mono-channel classification using individual classifiers.

As an extension, an augmented bio-acoustic system based on a multi-channel ensemble classification framework is proposed. As opposed to obtaining data from a single microphone, a three channel system is implemented, consisting of a microphone placed within each ear and an acoustic gel sensor secured to the underside of the chin [4]. The acoustic gel sensor, although capturing a similar type of signal, provides additional information as the acoustic wave is propagating through a different facial region, thus providing different signal characteristics and therefore additional information. Fig. 1 gives an overview of the system (top-left), example waveforms associated with each channel from a single action (top-right) and a block representation of the ensemble process utilized (bottom). A multi-class system naturally allows for classifier rejection when there are conflicting channel outputs, preventing misclassification when there is ensemble disparities. This is vital when there is increased intra-class variance due to testing in non-controlled environments. In this paper the ensemble multi-channel framework is outlined, with its effectiveness for inter-class classification of TMEP signals highlighted through comparison to individual classification baselines.

II. METHODOLOGY

Combining of multiple channels for classification naturally lends itself to be formulated within an ensemble classification framework. An ensemble classifier methodology can be

This research was supported by the UK Engineering and Physical Sciences Research Council (EPSRC), grant EP/F01869X

Michael Mace is with the Department of Mechanical Engineering at Imperial College London, London, UK, SW7 2AZ (email: m.mace11@imperial.ac.uk)

Khondaker Abdullah-Al-Mamun and Shouyan Wang are with the Institute of Sound and Vibration Research (ISVR), University of Southampton, UK, SO17 1BJ (email: [kam1e08,sy.wang]@soton.ac.uk)

Lalit Gupta is with the Department of Electrical and Computer Engineering, Southern Illinois University, IL, USA, 62901 (email: lgupta@siu.edu)

Ravi Vaidyanathan is with the Department of Mechanical Engineering at Imperial College London, London, UK, SW7 2AZ and the US Naval Postgraduate School, Monterey, CA, USA, 93943 (email: rxv@case.edu)

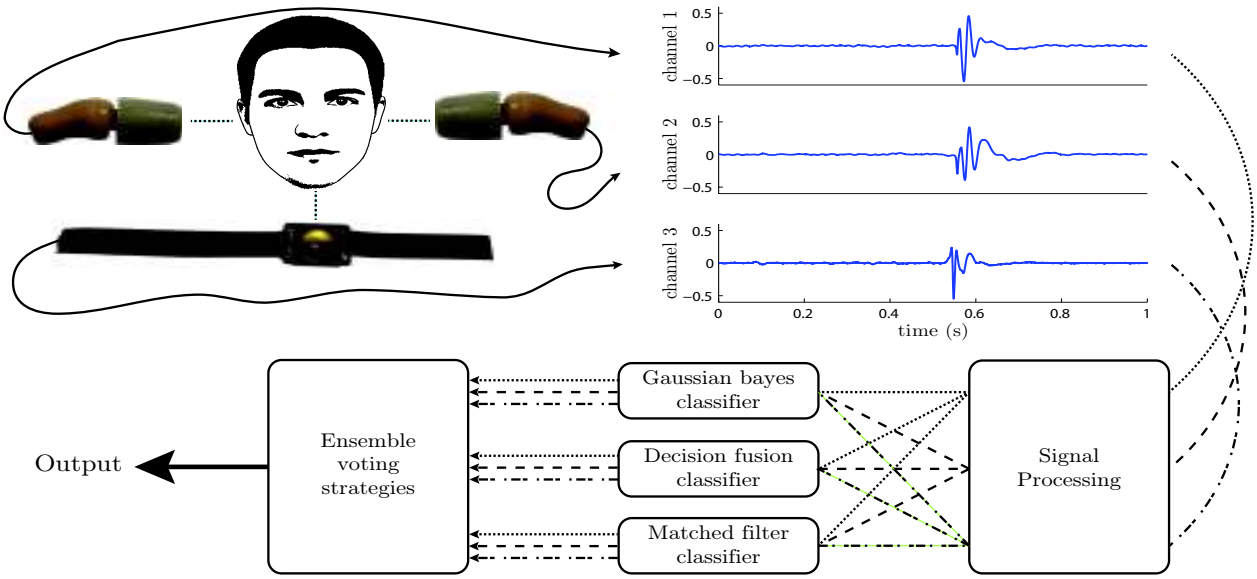

Fig. 1. Overview of the 3 channel system, examples of the signals for one action instance and the associated ensemble classification methodology

described as the weighted combination of several individual pattern classifiers to produce an output that outperforms every one of them. It mimics standard human nature of actively seeking several opinions before making a crucial decision [5]. A vast majority of ensemble methods rely on a single base model, with the ensemble created by training on different bootstraps of the data e.g. bagging techniques [6]. However if only a small training set is available this can tend not to provide the diversity required to produce a good ensemble. Random feature subspace methods are another technique for obtaining an ensemble of classifiers, however, when directly used with contiguous time series data it seems un-intuitive and requires tuning of additional parameters. In this paper we propose to combine the outputs of different sources (channels), classified through various base models outlined below. The basis for using different models and the selection of the models themselves is that each one extracts different information from each source channel based on frequency domain, time domain, and correlation (template matching) features.

A. Base classifiers

1) ***Gaussian Bayes classifier (GBC-DCT)****:* Bayes' decision theory provides a fundamental methodology for solving statistical classification problems. Under the assumption of a naive Gaussian distribution of the features, problems associated with dimensionality can be circumvented and a quadratic classifier is retained. The test instance is assigned to the class which maximizes the log posterior where all the sample statistics are estimated from the training set. For feature extraction, a frequency transform called the discrete cosine transform (DCT) is employed. The DCT is related to a real-valued DFT but allows more of the signals energy to be concentrated on fewer coefficients. This feature extraction can be represented as a linear matrix operation, $\mathbf{s} = \Phi\mathbf{x}$, where Φ represents the DCT basis matrix. Prior to classification the DCT coefficient vector $\mathbf{s}$ is reduced using a magnitude ranking and cross-validation procedure so as to reduce computation and over-fitting to the training set [3].

2) ***Decision fusion classifier (DFC)****:* The decision fusion classifier (DFC) is based on multinomial probabilities at each feature point that are then fused to give a single decision [1]. The individual distributions are estimated from the training set and are based on the likelihood of an instance being most similar to one of C-class templates. The similarity metric used is the Euclidean norm nearest mean discriminant function. The multinomial distribution of an instance at sample n, being classified to one of the C-classes, given its true class, can then be estimated from the training set based on the maximization of this discriminant function. The N-sample classifiers are then ranked based on the average correct classification across the C-classes. A test signal is then classified by reordering the vector according to this a-priori ranking and obtaining N-classifications at each sample using the discriminant function. The corresponding likelihood probabilities of each of these individual classification outputs for correct and misclassification to one of the '$C-1$' remaining actions, as found in the training phase, are concatenated at each sample to form 'C' probability vectors. The final classification is then given by the class which maximizes the sum of these vectors.

3) ***Matched filter classifier (MFC)****:* The matched filter classifier (MFC) is used to extract a known signal in stationary noise. It is designed to maximize the signal-to-noise ratio at $n = N$ and is equivalent to maximization of the cross-correlation between the test instance and the C-templates. The output is assigned to the class which gives the maximum value after convolution between the test instance and the C-templates.

B. Voting strategies

Use of rank-based (preferential) strategies allows for fusion and synergistic classification between base classifiers which may or may not exhibit varying types of output. A brief overview of some non-trainable voting schemes are outlined below and in [7].

1) ***Plurality voting****:* Plurality is a voting system in which the class with the most votes across the classifier and channel combinations is reconciled as the overall ensemble classification output. Rejection of the classification can only occur if there is a tie for the top place.

2) ***Majority voting****:* A majority vote is considered to give a slightly more democratic output, in the sense that a winner is only selected if a majority verdict is reached. If there is less than 50% of votes cast in favor of the winning class, then the classification is rejected.

3) ***Alternative voting****:* If a winner isn't assigned after a majority vote, the class with the minority is removed from the classification set and a majority vote takes place again. This is repeated until either a majority winner is eventually found or if the two remaining classes have the same number of votes then the classification is rejected.

4) ***Borda count voting****:* The individual outputs from each base classifier is ranked from $1-C$ with these converted directly to an associated score ranging from $C-1, C-2, ...0$. The scores are then tallied over the entire ensemble and the class with the biggest tally is selected as the winner, if a draw for top place is obtained then the classification is rejected.

5) ***Condorcet voting****:* This method involves pairwise head-to-head comparisons between each class, with a point tallied for each winner. This means that $C!/(2(C-2)!)$ comparisons are made for each voter with the class who wins the most head-to-head pairings declared overall ensemble winner. Due to the transitive nature of this method, rejection of the classification can occur, in the case of 2 or more classes being equal top scorers.

C. Experimental protocol

Data was collected from four healthy individuals aged between 21 and 28 and included collection of at least one hundred signals of each of the four actions outlined previously. Only subject 1 was familiar with making the actions, with the other subjects instructed over the course of 2-4 sessions (lasting approximately 30 minutes per session). The setup of the system was as follows; seating the subject comfortably in front of the laptop running the data collection software, placement of a generic earpiece in each ear of the individual (channels 1 and 2) and placement of an acoustic gel sensor under the chin of the subject, secured using a velcro strap around the head (channel 3). The sensors were connected to the laptop through pre-amplifiers, anti-alias filters set to 4 kHz and commercially available USB soundcards. Detection parameters were set by the experimental assistant and the software was allowed to autonomously detect and save the signals. The subjects were directed to follow a stimulus on screen with the program paused whenever a subject required a break. Any anomalous signals that the software segmented and saved but did not originate from a tongue motion were indicated by the subject and removed at the time of occurrence. The data was recorded at a sample rate of 8 kHz with the experimental protocol approved by the local ethics committee.

III. RESULTS

Results were run for the four subjects, using a 10-fold cross-validation procedure to estimate the generalization error, correct classification and rejected classification rates. The signals were downsampled to 2 kHz and segmented to 0.256 seconds (512 samples) and further decomposed using a wavelet packet transform, so that the 0 - 500 Hz sub-bands could be selected (thus reducing the number of samples and therefore computation time) [2]. Each fold was used as the test set (10 instances x 4 actions) in turn, with a reduced training set (32 instances x 4 actions) randomly selected from the remaining ninety instances. Thus within each cross-validated run the entire data set was used for testing. This was repeated a hundred times with the data set randomly shuffled each run. For each run the three base classifiers were trained on the three channels of data, indicating a maximum of nine voters in the ensemble. The five voting strategies were tested against different ensemble sizes and combinations. The total number of combinations for each ensemble size, $1-9$ being $\{9, 36, 84, 126, 126, 84, 36, 9, 1\}$ respectively.

Fig. 2 shows the error, correct and rejected classification accuracies for the five voting strategies versus different ensemble sizes. The results are averaged over the various combinations and subjects. The first result at ensemble size 1 gives the individual classification results averaged across all the base classifier and channel combinations and thus acts as the baseline for the five voting strategies with 0% rejection and 23.2% generalization error. There is a general trend of increased performance (reduced error and increased correct classification) as the ensemble size is increased, with this tailing off around the 6/7 ensemble size point. Alternative voting consistently performs the worst and produces no rejection at the odd prime ensemble sizes of 5 & 7. The majority vote generally performs the best, achieving the lowest generalization error of 2.8% when the full ensemble set is used. This is at the expense of a lower correct classification rate (86.2%) due to a higher rejection rate (11%). A similar error rate is achieved with an ensemble size of 2 for both the plurality and majority strategies but this is at the expense of a significantly increased rejection rate which reduces the correct classification to below the baseline. It should be noted that the borda count and condorcet methods are equal in all cases, giving their best correct classification rate of 92.4% (error 5.8%, rejection 1.8%) for an ensemble of size 9. They generally give lower rejection rates than the plurality and majority methods across all ensemble sizes leading to slightly increased correct classification rates but in conjunction with increased misclassification rates.

Further to this Fig. 3 shows error and rejection rates for different combinations of channels and base classifiers using the majority voting strategy. The majority voting strategy is

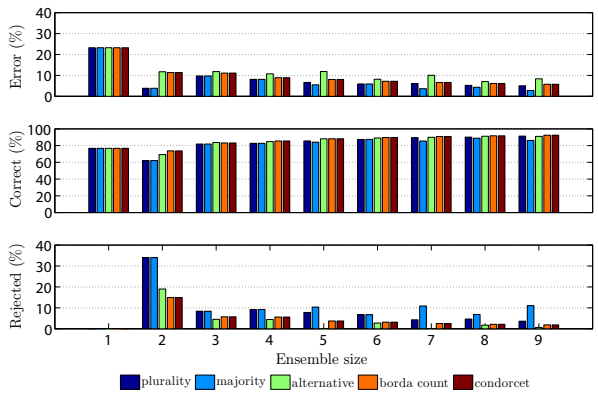

Fig. 2. Ensemble classification results for various combinations of the 3 channel, 3 classifier system

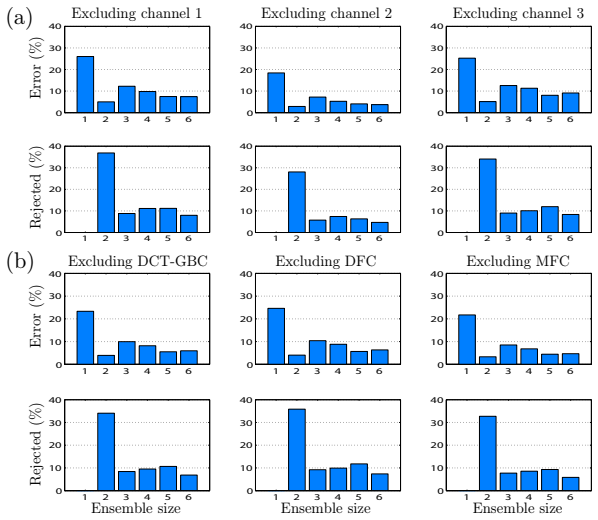

Fig. 3. Ensemble classification results for various combinations with exclusion of one channel or classifier only. (a) shows results pertaining to removal of each channel in turn while (b) gives results for removal of each base classifier in turn.

only considered as this gave the best results in terms of misclassification error. Fig. 3(a) gives results pertaining to each channel being excluded from the ensemble combinations, whilst Fig. 3(b) shows the removal of each base classifier in turn. This reduces the maximum ensemble size to 6, with the total number of combinations for ensemble sizes $1-6$ being $\{6, 15, 20, 20, 15, 1\}$ respectively. The lowest error from the classifier exclusions is when the MFC is removed (error 4.5%, rejection 9.3%, correct 86.2%, ensemble size 5) with the results generally quite uniform across the classifier removals, indicating relatively equal contributions from each classifier. The MFC has been shown to perform worst on individual channel classification, with these results indicative of this [1]. The lowest error from the channel exclusions is when channel 2 is removed (error 3.7%, rejection 4.7%, correct 91.6%, ensemble size 6) and even though a lower error is achieved whilst utilizing all channels and classifiers it is at the expense of a higher rejection rate. However, it cannot be elucidated from this result that channel 2 should be disregarded from the system entirely, as between subjects, channel 1 and 2 cannot be differentiated. This is because the placement and performance of each channel within each ear is completely subject specific. These results do however indicate the usefulness of the acoustic gel sensor but also show that if a two channel system was considered i.e. channels 1 and 2 only, as this is the simplest setup in terms of donning, obtrusion and cost, there is still a relative increase in performance (error 8.0%, rejection 12.0%, correct 80.0%, ensemble size 5).

IV. CONCLUSIONS

It has been shown that the use of additional channels/sensors combined with ensemble voting techniques can significantly increase the classification performance of tongue-movement ear pressure signals. This is in part due to the framework allowing for the rejection of classification when there is significant disagreement of the class of a particular test signal and thus rather then potentially misclassifying it, the system classifies the action as unknown. This increase in correct classification can also be affiliated with the diversity in classification across the ensemble set, implying that certain ensemble members which would give incorrect outputs are alleviated by the rest of the ensemble. Hence the philosophy of seeking several opinions before a crucial decision is made is an extremely useful methodology to implement. The robustness of the system will be fully tested in the future by analyzing its ability to specifically reject low frequency interfering signals that would otherwise be classified as one of the prescribed tongue actions.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge Think-A-Move, Ltd. of Cleveland, OH, USA for their commercial research in this area. This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC).

REFERENCES

[1] R. Vaidyanathan, B. Chung, L. Gupta, H. Kook, S. Kota, and J. D. West, "Tongue-Movement Communication and Control Concept for Hands-Free Human-Machine Interfaces," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 4, pp. 533–546, 2007.

[2] K. A. Mamun, M. Mace, S. Wang, R. Vaidyanathan, and M. E. Lutmen, "Multivariate Bayesian Classification of Tongue Movement Ear Pressure Signals Based on the Wavelet Packet Transform," in *In Proc. of the 2010 IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, (Kittila), pp. 208–213, IEEE, 2010.

[3] M. Mace, K. A. Mamun, R. Vaidyanathan, S. Wang, and L. A. Gupta, "Real-time Implementation of a Non-invasive Tongue-based Human-Robot Interface," in *Proc. of the 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, (Taipei), pp. 5486–5491, IEEE, 2010.

[4] M. V. Scanlon, "Acoustic sensors in the helmet detect voice and physiology," *Proceedings of SPIE*, vol. 5071, pp. 41–51, 2003.

[5] L. Rokach, *Pattern classification using ensemble methods*. World Scientific Publishing, 2010.

[6] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits And Systems*, vol. 6, no. 3, pp. 21–45, 2006.

[7] K. T. Leung and D. S. Parker, "Empirical comparisons of various voting methods in bagging," in *Proceedings of the 9th ACM int. conf. on Knowledge discovery and data mining - SIGKDD '03*, (Washington DC), pp. 595–600, ACM Press, 2003.