PARAMETRIC AND NON-PARAMETRIC SIGNAL ANALYSIS FOR MAPPING AIR FLOW IN THE EAR-CANAL TO TONGUE MOVEMENTS: A NEW STRATEGY FOR HANDS-FREE HUMAN-MACHINE INTERFACES

Ravi Vaidyanathan[1, 3], Hyunseok Kook[2], Lalit Gupta[2], & James West[3]
[1] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH, 44016, USA
[2] Department of Electrical and Computer Engineering, Southern Illinois University, Carbondale, IL 62901, USA
[3] Think-A-Move, Ltd., Beachwood, OH, 44122, USA

ABSTRACT

A complete signal processing strategy is presented to detect and precisely recognize tongue movement by monitoring changes in airflow that occur in the ear canal. Tongue movements within the human oral cavity create unique, subtle pressure signals in the ear that can be processed to produce commands signals in response to that movement. Once recognized, said movements can in turn be used in human-machine interface applications such as communicating with a computer and controlling mechanical devices. The processing strategy includes pressure signal acquisition using a microphone inserted into the ear-canal, PSD analysis to design bandpass filters to reject pressure changes due to sources other than tongue movements, start- and end-point detection in the waveforms through cross-correlation, signal estimation, and the design and evaluation of parametric and non-parametric signal classifiers. The non-parametric signal classifiers include non-linear alignment classifiers and matched filters, while the parametric classification involves a multivariate Gaussian classifier using AR model parameters. The complete strategy is tested on 4 tongue actions: touching the tongue to the left and right corners of the mouth, and to the top and bottom center of the mouth. Through extensive experiments, it is shown that the pressure signals due to tongue movements are distinct and can be detected with over 97% accuracy. It is thus concluded that the unique strategy will make hands-free control of devices using tongue movements a practical reality.

This work is supported by a NIH Phase I SBIR Research grant 1R43HD042367-01A1

1. INTRODUCTION

Although there is a well-recognized need in society for effective machine interface mechanisms that will enable the physically impaired to be more independent, much of the technology developed towards this goal still fails to meet their specific needs. At present, the majority of existing systems may be classified as mechanical-input devices; *i.e.* the user physically moves a part of a device in order to generate a control input signal. Examples of such systems include hand-operated joysticks and the use of head or chin movements to move a lever whose motion is translated into control commands. Systems of this nature require constant bodily movement that can be tiring and uncomfortable to the user, while regular use can also cause repetitive motion injuries and skin irritation. Furthermore, the single-lever design limits allowable commands that can be generated, and is extremely limiting for people with limited upper extremity function. Given that most patients with limited extremity function (such as victims of spinal cord injury (SCI) and arthritis) possess the ability to move their tongue and/or mouth effectively, the potential of the human oral cavity has been exploited as a source for machine control signals. Contemporary examples include inserting a track-ball, joystick, plastic palate, or "sip-and-puff" straw into the mouth of an individual with the tongue or lips providing control input. These devices, however, are extremely intrusive, irritate the mouth, impair verbal communication, present hygiene issues, and are also limited signal generation capacity. The goal of our work is to develop a patient-generated control strategy which can overcome the deficits of these existing systems. Specifically, *it is our intention to develop a non-intrusive tongue-movement based machine interface without the insertion of any device within the oral cavity*. We introduce a unique strategy for detecting tongue movement through the monitoring of air pressure changes within the ear canal.

The focus of this paper is a new means of detecting tongue movement in order to generate input signals that can be used for "hands-free" control of devices (such as a wheel chair) in human-machine interface applications. Our on-going investigations have shown that various movements within the oral cavity create unique traceable pressure changes in airflow within the ear canal that can be measured with a simple pressure sensor (e.g. microphone) placed in the ear. Individuals with limited upper extremity control can use the output of the microphone as an effective means to communicate with a computer and/or to control electro-mechanical assist devices (e.g. a power wheelchair). Patients suffering from spinal cord injuries (SCI), repetitive strain injuries (RSI), severe arthritis, loss of motion due to stroke, and central nervous system (CNS) disorders would all benefit greatly from this concept. The success of this strategy will clearly depend on the accurate classification of tongue movements based on air flow measured in the ear. The scope of the paper, therefore, is to demonstrate that pressure signals in the ear corresponding to tongue movement are distinct, and can be classified accurately. The detection of the tongue movement pressure signals in the ear canal is formulated as a M -class pattern classification problem in which the classes correspond to M distinct tongue movements. Pressure signals resulting from $M = 4$ tongue movements (left, right, up, and down) is used to demonstrate the effectiveness of the strategy.

2. SIGNAL ACQUISITION

Figure 1 illustrates a pressure sensor inserted partially into the ear of an individual (within the cavity defined by the pinna, if not deeper within the ear such as within the concha, at the opening of the ear canal). The sensor includes a shielding housing and an internal microphone. The internal microphone resides on the interior portion of the housing within the ear canal at a depth of 2.5 mm to 12.5 mm measured from the opening of the ear canal. Insertion of the microphone into the ear canal shields pressure signals from environmental noise. The external microphone (not used in the present study), will be used in future studies to monitor and exclude signals from external sources. Figure 2 shows examples of pressure signals in the ear (sampled at 2 KHz), when a subject was asked to touch the tongue lightly to the left, right, top, and bottom of the mouth respectively. Each movement was repeated 100 times, thus each figure has 100 superimposed signals corresponding to the same tongue movement.

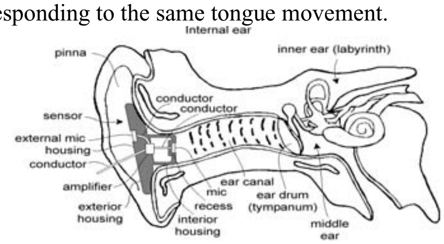

Fig 1 Pressure signal acquisition system

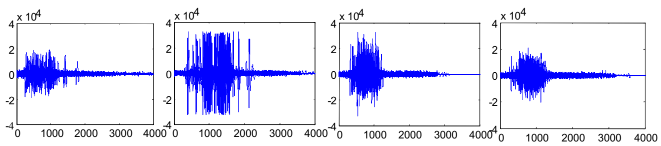

Fig 2. Pressure signals for the 4 different tongue movements.

3. SIGNAL ANALYSIS AND PROCESSING

Conventional signal processing techniques are generally inadequate to recognize the subtle pressure variations in the ear canal resulting from tongue movement. The ear canal itself is an interference-ridden, noise-amplified environment for acoustic recording. Furthermore, external noise (environmental sounds) can easily obscure the slight pressure deviations accompanying tongue movement. The following two sections enumerate the steps in our current processing and classification strategy.

Bandpass Filtering and Normalization

The first step in the analysis of the signals in Figure 2 is to identify the frequency range of interest in the signals. The averaged PSD of the signals are shown in Figure 3. It is observed that pressure signal activity is approximately in the band 10 to 50 Hz. Therefore, in the first step of processing, the signals are bandpass filtered using 10 and 50 as the lower and upper cutoff frequencies, respectively. By examining the signals in Figure 2, it is clear that the signals have amplitude differences within the same class and are not aligned in time. The signals can be easily amplitude normalized by dividing each sample of a signal by the standard deviation of the samples in the signal [1]. In the generalized formulation to follow, let

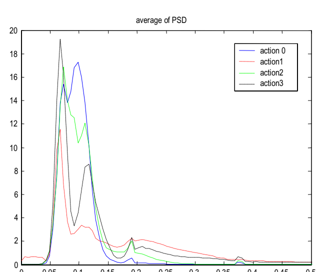

Fig 3. Averaged PSDs of the signals in Figure 2.

$$h_{m,i}(k),\ k = 1,2,..., N_m;\ i = 1,2,..., L$$

be the ith filtered and amplitude normalized signal of class

$$m,\ m = 1,2,..., M,$$

where, M is the number of signal classes, N_m is the number of samples, and L is the number signals in each class (assumed equal for convenience).

Signal Estimation

Signal averaging is one of the most frequently used operations to estimate signals from the outcomes of a random process [2,3] and can, therefore, be used to estimate the underlying signal of each pressure signal class from the amplitude normalized outcomes. However, directly averaging the signals $h_{m,i}(k),\ i = 1,2,...L,$ will result in a poor time-smeared estimate because the signals are not aligned in time. The accuracy of the estimate can be improved if the signals are first aligned in time with a template of each class and then averaged. The problem, however, is that the templates are not available because the true pressure signals are unknown. A pairwise cross-correlation based averaging procedure is introduced to first generate an initial signal template for each class and then use the initial template to align signals and estimate the signal of each class. If L is assumed to be an integer power of 2, the average $\bar{h}_{m,L}(k),\ m = 1,2,..., M$ of the L signals can be computed as:

$$\bar{h}_{m,L}(k) = (1/2)[\bar{h}_{m;1\to(L/2)}(k) + \bar{h}_{m;(L/2)+1\to L}(k)],$$

where,

$$\bar{h}_{m;1\to(L/2)}(k) = [1/(L/2)]\sum_{i=1}^{L/2} h_{m,i}(k),\ k = 1,2,...N_m$$

is the mean of the first half of the L signals and

$$\bar{h}_{m;(L/2)+1\to L}(k) = [1/(L/2)]\sum_{i=(L/2)+1}^{L/2} h_{m,i}(k),\ k = 1,2,...N_m$$

is the mean of the second half of the L signals. By further decomposing the first half and second half of the signals into equi-sized sets of size (L/4), the means can be computed as

$$\bar{h}_{m;1\to(L/2)}(k) = (1/2)[\bar{h}_{m;1\to(L/4)}(k) + \bar{h}_{m;(L/4)+1\to(L/2)}(k)]$$

$$\bar{h}_{m;(L/2)+1\to L}(k) = (1/2)[\bar{h}_{m;(L/2)+1\to(3L/4)}(k) + \bar{h}_{m;(3L/4)+1\to L}(k)]$$

The L signals can be decomposed into successively smaller sets until pairs of signals are left. The signals in each pair are averaged by aligning the sequences in the position of maximum cross-correlation. The means of the pairs are combined in a pairwise fashion according to the steps outlined above to determine $\bar{h}_{m,L}(k),\ m = 1,2,..., M$. The initial template for each class is formed by identifying the start- and end-points of the tongue action in $\bar{h}_{m,L}(k)$ and extracting the signal segment between these two points. If the start- and end-points in the initial template are denoted by a and b, respectively, each signal $h_{m,i}(k),\ i = 1,2,...L,$ is segmented by aligning it with the initial template in the maximum cross-correlation position and multiplying it with a rectangular window $R_{a,b}(k)$. That is, the segmented signals are given by

$$h_{m,i}(k)R_{a,b}(k),\ i = 1,2,..., L;\ m = 1,2,..., M.$$

If $N = (b - a + 1)$, then, the N samples of the segmented signals are re-ordered and represented by $v_{m,i}(k),\ k = 1,..., N$. The final estimate $\bar{h}_m(k),\ m = 1,2,..., M$, of the signal for each action class can be estimated by averaging the segmented signals

$v_{m,i}(k),\ k = 1,...,N$. Figure 4 shows estimates of the signals of the 4 action classes computed using $L = 64$.

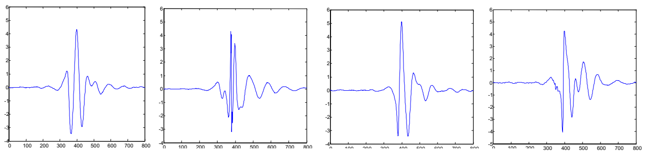

Fig 4. Estimates of the 4 pressure signals.

4. CLASSIFICATION STRATEGIES

Given the egmented signals belonging to the M classes, different classification methodologies can be applied to detect the classes of the signals. In this study, matched filtering, autoregressive modeling, and non-linear alignment methods are developed to determine the signal classes.

Matched Filter

A matched filter can be designed to detect a signal buried in noise under the conditions that the signal is known and the noise is stationary. The matched filter is designed to maximize the output signal-to-noise ratio at the time instant $k = N$. If it is assumed that the noise is white, then, the matched filter $h_m(k),\ m = 1,2,...,M$ is given by

$$h_m(k) = \bar{h}_m(N - k),\ k = 1,2,...,N$$

That is, the unit sample response is the signal reversed in time and delayed by N samples. The response of each matched filter to an input test signal represented by $T = t(k),\ k = 1,2,...,N$ is computed and the test signal is assigned to the class of the matched filter that yields the maximum value at time N. That is, if $y_m(N),\ m = 1,2,...,M,$ is the response of $h_m(k)$ to T at $k = N$, then, T is assigned to the class m^* given by

$$m^* = \arg \underset{m}{MAX}[y_m(N)]$$

Autoregressive (AR) Modeling

The underlying generation of the pressure signals of each action can be modeled by an AR process of the form

$$v_{m,i}(k) = \alpha_{m,i} + \sum_{j=1}^{p} \theta_{m,i,j} v_{m,i}(k - j) + \sqrt{\beta_{m,i}}\,\omega(i)$$

and the model parameters ($\theta_{m,i,j},\ \alpha_{m,i} / \sqrt{\beta_{m,i}}$) can be used as features for signal classification. If it is assumed that the class conditional density functions of the AR feature vector are Gaussian with mean vector μ_m and covariance matrix Ψ_m, the discriminant function of the resulting Gaussian classifier for class m, assuming equal prior probabilities, is given by

$$D_m(T) = -(1/2)\{\ln|\Psi_m| + (T - \mu_m)^T \Psi_m^{-1}(T - \mu_m)\} + \ln P(m)$$

where, $P(m)$ is the class prior probability. For this case, the test signal T is assigned to the class m^* given by

$$m^* = \arg \underset{m}{MAX}[D_m(T)]$$

Non-linear Alignment

Various alignment-based methods can also be formulated to determine the similarity of a test signal and a template of a signal [1,2]. Non-linear alignment, also called dynamic alignment, optimally aligns two signals to compensate for non-linear expansions and compressions in signal segments and also to compensate for duration differences. In the design of non-linear alignment classifiers, the goal is to determine a mapping W between the time-index p of a test signal $t(p)$ and the time-index q of a reference signal $\bar{h}_m(q)$ such that the best alignment between the two sequences is obtained. The mapping

$$W = [w(1), w(2),..., w(Z)]$$

where

$$w(z) = [i(z), j(z)];$$

$$p = i(z),\ z = 1,2,..., Z;\ q = j(z),\ z = 1,2,..., Z,$$

defines a piecewise linear alignment path in the (p, q) plane. Both time axes are transformed into a common time axis z of length Z. When there is no timing difference between the sequences, the warping path coincides with the diagonal line $(p = q)$. The best alignment path is given by determining W that minimizes

$$D = \sum_{z=1}^{Z} d[t(i(z)), \bar{h}_m(j(z))]$$

where D is the total accumulated distance between $t(p)$ and $\bar{h}_m(q)$ along W and $d[x, y]$ is the local distance between the samples x and y. Examples of local distance metrics include the absolute difference and the difference-squared norm. In order to restrict W in a meaningful manner in the (p, q) plane, end-point, continuity, and slope constraints are imposed on W [1,2]. If $D_m(T)$ is the aligned distance between a test sequence T and a reference sequence $\bar{h}_m(q),\ m = 1,2,..., M$, then, the test sequence is assigned to the class m^* given by

$$m^* = \arg \underset{m}{MIN}[D_m(T)]$$

5. EXPERIMENTS AND RESULTS

Pressure data corresponding to 4 tongue movements: touching the tongue lightly to the left, right, top, and bottom of the mouth were recorded to design and evaluate the strategy. Each movement was repeated 100 times; therefore, each tongue movement class had 100 pressure signals. Each signal was bandpass filtered and segmented (N =800) as described in Section 3. The signals were randomly partitioned into 2 mutually exclusive and equal-sized sets to generate a design set and a test set for each class. For each signal class, the signal estimated from the training set was used as the reference template for non-linear alignment and to determine the unit sample response of the matched filter. The AR model parameters for each signal class were determined from the signals in the respective training set using the Yule-Walker autocorrelation method. The model order $p = 10$ was determined empirically. The random resampling approach described in [1,3] was used to generate J design and test set pairs. Each pair is referred to as a trial and the classification accuracies were estimated over J =100 trials. Each trial consisted of testing 50 test signals from each class, therefore, the classification accuracy was estimated by testing (100x50x4) = 20,000 signals.

For convenience, the 4 pressure signal classes: left, right, up, and down, are represented by m=1, 2, 3, and 4, respectively. The matched filter, AR model, and non-linear alignment classification results, assuming equal prior probabilities, are presented in Tables 1, 2, and 3, respectively. The tables show confusion matrices as well as the classification accuracies. The confusion matrix part of the results can be interpreted by examining the first row of Table 2 which shows that out of the 5000 tests conducted with signals vectors drawn from class 1, 89.18% were classified correctly as belonging to class 1, 0.12% were misclassified as class 2, 9.98% were misclassified as class 3, and 0.71% were misclassified as class 4. The results show that an average classification accuracy of 90.86%, 85.06%, and 97.72% can be achieved by the matched filter, AR model classifier, and the non-linear alignment classifier, respectively.

The performance of the non-linear alignment classifier, which can compensate for non-linear variations, is superior to that of the matched filter which is essentially a cross-correlator. Cross-correlators are not capable of accommodating duration and non-linear variations. The performances of the non-linear alignment classifier and the matched filter are superior to that of the AR model classifier which compresses the 800 samples into a small set of 10 model parameters that are used as features.

TABLE 1 Matched Filter

m	1	2	3	4
1	100	0	0	0
2	6.06	64.45	28.28	1.01
3	0	0	100	0
4	1.01	0	0	98.99
Class. Accuracy = 90.86%				

TABLE 2 AR-Gaussian

m	1	2	3	4
1	89.18	0.12	9.98	0.71
2	0	77.24	0.08	22.67
3	17.51	1.94	78.33	2.22
4	0.20	0	4.33	95.47
Class. Accuracy = 85.06%				

TABLE 3 Non-linear Alignment

m	1	2	3	4
1	100	0	0	0
2	1.01	94.94	0	4.04
3	0	0	97.97	2.02
4	1.01	0	1.01	97.97
Class. Accuracy = 97.73%				

6. CONCLUSIONS

The goal of this paper was to develop a signal processing strategy to demonstrate that the pressure changes in air flow that occur in the ear canal due to tongue movement are distinct and that they can be detected accurately. PSD analysis was conducted to determine the frequency range of the pressure signals in order to design bandpass filters. A pairwise cross-correlation based averaging procedure was developed to obtain initial estimates of the pressure signals corresponding to the tongue movements. Start- and end-points in the initial template were identified and the signals were segmented between the end-points in the position of maximum cross-correlation with the initial template. A final estimate of the signal of each class was obtained by averaging the segmented signals. Three different classification methods were implemented to classify the signals. The matched filter and non-linear alignment classifier made use of the signal estimates for the unit sample responses and the reference templates, respectively. The parameters of the AR-Gaussian parametric classifier were estimated directly from the segmented signals in the training set. The results from experiments conducted on four tongue movements show that all three classifiers yield good results. The best results were obtained using non-linear alignment which yielded classification accuracies of over 97%.

7. FUTURE WORK

Current investigations are focused on the analyses and classification of a wider range of tongue actions and issues related to the practical application of this strategy. These issues include: (a) real-time detection of the onset of the tongue movement in the pressure signals, (b) filtering to isolate pressure signals from other bodily signals and external noise, and (c) determining the most suitable classification strategy, in terms of accuracy and speed, for real-time applications. We are presently targeting commercial applications for this technology including wheelchair control for quadriplegic users, [4], and robotic device control.

In summary, the results for the 4 tongue actions are highly encouraging. Based on these results as well as the results from our on-going investigations, it is concluded that the unique signal processing strategy developed for classifying air flow pressure signals in the ear canal will make hands-free control of devices using tongue movements a practical reality.

REFERENCES

1) L. Gupta and S. Ma, "*Gesture-based interaction and communication: automated classification of hand gesture contours*," IEEE Transactions on Sytems, Man, & Cybernetics – C, vol. 31, No. 1, 114-120, 2001.
2) L. Gupta, D. L. Molfese, R. Tammana, and P. G. Simos, *"Non-linear alignment and averaging for estimating the evoked potential,"* IEEE Transactions on Biomedical Engineering, vol. 43, No. 4, 348-356, l996.
3) L. Gupta, J. Phegley, and D.L. Molfese, "Parametric classification of multichannel averaged event-related potentials*,"* IEEE Transactions on Biomedical Engineering*,* Vol. 49, No. 8, 905-911, 2002.
4) G. Nemirovski, "System and method for detecting an action of the head and generating an output in response thereto"; U.S. Patent number 6,503,197, issued 2003