16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

EARLY MORNING ACTIVITY DETECTION USING ACOUSTICS AND WEARABLE WIRELESS SENSORS

Cheol-Hong Min, Nuri F. Ince, and Ahmed H. Tewfik

Department of Electrical and Computer Engineering, University of Minnesota
200 Union St. SE, 55455, Minneapolis, USA
phone: + (1) 612-625-5006, fax: + (1) 612-625-4583, email: cmin@umn.edu

ABSTRACT

In this paper, we study the classification of the early morning activities of daily living to assist those with cognitive impairments due to traumatic brain injuries. The system can be used to help therapists in hospitals or could be deployed in one's home. We briefly describe the infrastructure of our cost-effective system which uses fixed and wearable wireless sensors and show results related to the detection of activities executed in the morning. We focus on several early morning activities such as washing face, brushing teeth, shaving face and especially on the detection of shaving and brushing activities executed with electric devices. Features from accelerometer and acoustic sensors were extracted in time and frequency domain then used for classification using Gaussian mixture models, followed by a sequential classifier. We show promising classification results obtained from 7 subjects especially when electric devices are used to execute these early morning activities.

1. INTRODUCTION

The Center for Disease Control (CDC) states that Traumatic Brain Injury (TBI) to be one of the leading causes to death and permanent disabilities in US. Statistics show that over 5 million people in US alone, which is approximately 2% of the population, are diagnosed with TBI. They suffer from long term disabilities and are in need of assistance with their Activities of Daily Living (ADL). About 75% of these incidents are "mild" TBI which exhibit short term amnesia or unconsciousness while others are categorized as "severe" [1].

TBI is usually caused by sudden impact to the head that damages cognitive functions of the brain. Therefore, TBI patients have difficulties remembering, concentrating or making decisions [1, 2]. Alzheimer's disease patients or elderly population also have similar cognitive disabilities [1].

Most patients who are diagnosed with TBI receive initial treatment at hospitals with therapist and learn new memories but when they return to their homes, they cannot receive the same treatment due to the high cost of service. Therefore, the cost-effective assistive system we propose shows potential to decrease the cost and allow TBI and cognitively impaired patients to live an independent life.

To assist the people with cognitive disabilities, several scheduler or reminder systems that give instructions were developed [3]. But, due to its intrusive nature, some people do not prefer this kind of technology. Therefore, we have proposed to work with set of sensors which are non-intrusive that can monitor the user's progress with the movement patterns recoded with accelerometers and motion sensors etc. [4]. By detecting and monitoring the activities, the system can give intelligent instructions to the user only when necessary [11, 12]. Static sensors detect the location and coarse level of activity, while the wearable sensors can detect activities at a fine level. Finally the data collected from these sensor platforms are fused together and appropriate sensor data can be extracted and processed by intelligent algorithms enabling the reminder system to interact with the subject only when necessary. Such a feedback can range from reminding the patient to take their medicine, continue shaving, initiate lunch preparation or make emergency calls.

With the increase of the electric device usage in our daily lives, people have begun to use them around the house. It ranges from kitchen appliances to blow dryers, electric toothbrush and electric shavers in the bathroom. Since the electric devices produce different sounds when activated, we explore the potential for sound analysis in detecting ADLs around the house. Since different activities may generate different sounds, we believe acoustic data may give us distinguishing patterns. Therefore, in this study, we focus on fusing the audio data with other sensor data to detect several early morning activities such as face washing, tooth brushing and face shaving.

Detecting the early morning activities which involve electric devices can be a problem since those activities do not require much movement of the arms. Therefore, our previous system which relies on accelerometer alone would not be sufficient. Thus, we simultaneously analyze the sounds produced by the devices during the execution of activities.

There have been studies that use RFID tags on objects around the house to infer activities based on the touch of a device [5]. Although, it is simple and widely researched, there are several reasons why this might not work in our case. Firstly, a glove is widely used as an RFID reader but washing activity is not possible without getting it wet. Secondly, processing many sensor data around the house is complex and costly. Thus, we would like to minimize the number of sensors distributed in the house. Therefore, it is critical that we pickup motion information from the wearable sensors to detect execution of the activity. Others have used sound data from bathroom to detect if the subject was toileting or taking a shower or washing hands [6] but in our study, we focus more on using sound information to augment motion information.

The main purpose of this paper is to extend our previous studies in [11] by exploring audio data fusion to detect early morning activities including electric brush and shaver. We first describe our system architecture and data collection methodology and discuss the problems encountered during the study. We further explore the problems and present a solution to improve classification performance. As the number of subjects increase, more variation in data and feature overlap across the activities are observed. Finally, we discuss about other problems encountered during the study such as the presence of the background noise and future plans. In Section 2, we give a brief overview of our system. In section 3, we discuss classification approach and in section 4, we show initial experimental results for fusing the accelerometer and audio data features which we collected and analyzed from healthy subjects. Finally, we close the paper with conclusion and few remarks.

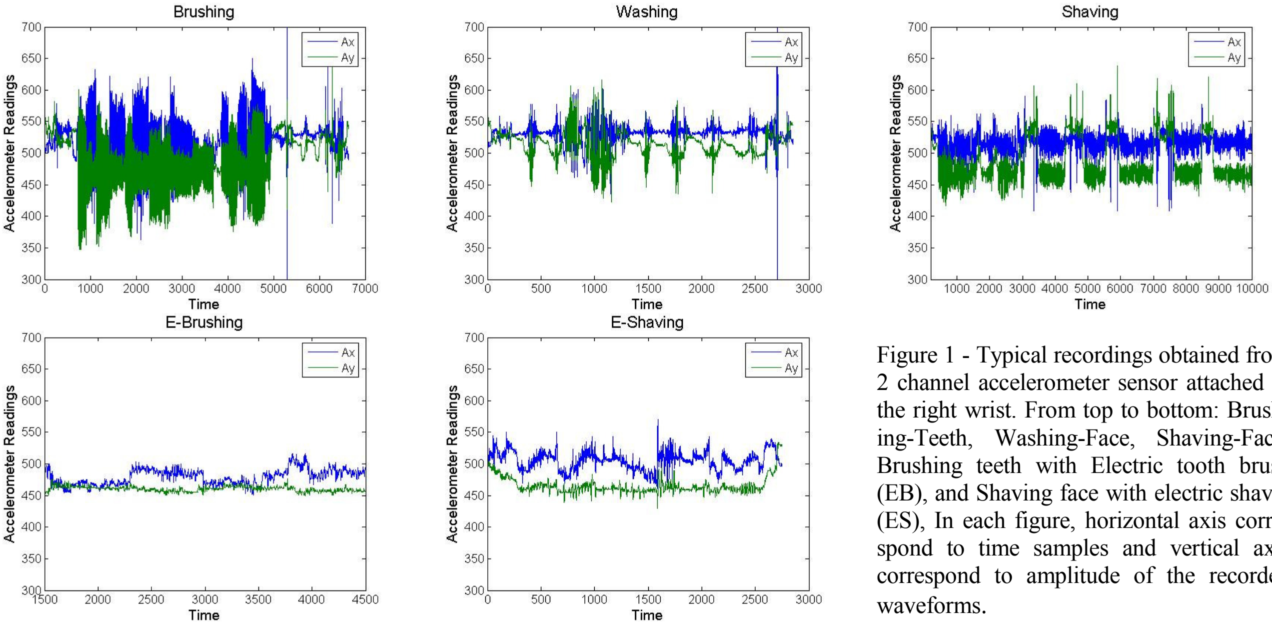


Figure 1 - Typical recordings obtained from 2 channel accelerometer sensor attached to the right wrist. From top to bottom: Brushing-Teeth, Washing-Face, Shaving-Face, Brushing teeth with Electric tooth brush (EB), and Shaving face with electric shaver (ES), In each figure, horizontal axis correspond to time samples and vertical axis correspond to amplitude of the recorded waveforms.

2. SYSTEM INFRASTRUCTURE AND DATA COLLECTION

To detect activities of daily living using wearable wireless sensors, we have developed a Data Acquisition (DAQ) platform which combines three sensor systems [4]. The first sensor system is a collection of static wireless sensors while the second sensor system relies on wearable sensors that provide detailed arm movements to complement the static sensor system data. The third is the audio visual data acquisition. Although our system is capable of recording audio and video simultaneously, we did not record video since some subjects did not feel comfortable with video recording. We recorded the activities on a "single trial" basis, which means each subject executed one of those activities we study, independently from other activities and trials.

For the fixed sensor system, we used the eNeighbor^{TM} system (eN) that was developed by RedWing Technologies which is now called the Healthsense^{TM} (www.healthsense.com). The system is equipped with several sensors such as a motion, bed, chair, door, flow and contact sensors which enables to track most of the daily activities in coarse level.

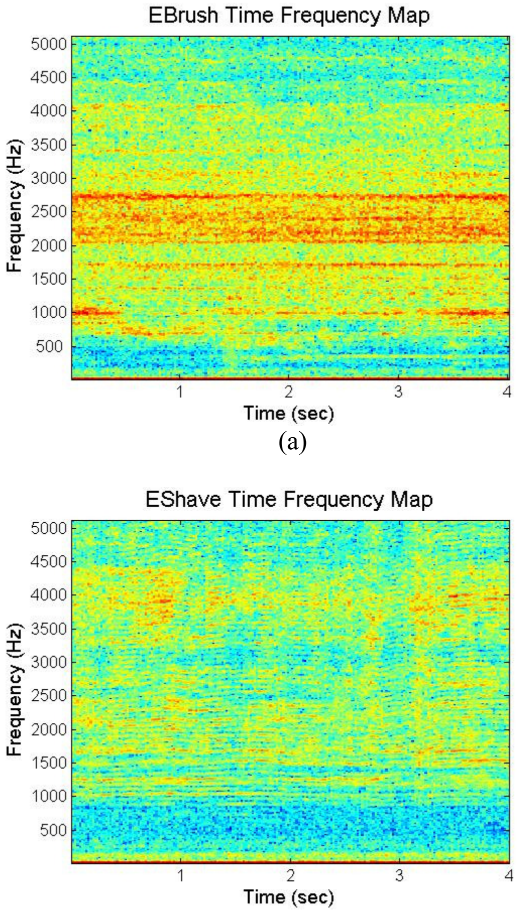
To obtain detailed information about the activity of the subject, we used wearable 2-axis accelerometer which is attached to the both wrists. We used the MICAZ mote kits and MTS-310 sensor kits to collect and wirelessly transmit accelerometer data developed by Crossbow Technology Inc. (www.xbow.com). Accelerometer data was sampled at 50Hz, digitized with 10-bit on-board A/D, transmitted to the gateway and transferred to the PC from MIB-510 gateway (base station) in real-time. The mote kits were attached to wrists to record movements. We recorded the acoustic data simultaneously with a microphone at 22 kHz with 8 bits.

Authors in [7] classified ADLs into 3 different categories such that personal hygiene (bathing, toileting, etc) and personal nutrition (eating) were classified as basic ADL. Each independently living person must be capable of performing these activities. Thus, we focused in the classification of several ADLs for personal hygiene such as washing face, brushing teeth and shaving face. Other activities we targeted were the use of electric devices such as electric toothbrush and electric shaver which are already widely used. We are currently waiting for the approval to apply the technology to the TBI patients; therefore, we continue our study which involves healthy subjects. We have 7 subjects that participated in the study and provided many recordings of washing face, conventional and electric tooth brushing, and both electrical and razor based face shaving.

We have collected 44 conventional tooth brushing activities, 46 face washing activities, 26 face shaving activity trials, 75 electrical tooth brushing (EB) and 73 electrical shaving (ES) activities which were recorded to generate a classifier model. Typical accelerometer recordings of the signals are shown in Figure 1, which shows the different patterns across the signals in time. We exploit these differences to distinguish one activity from the others. In addition to the 5 distinct tasks, we also recorded other activities that were not related to the activities of interests. These can be for example changing a towel, combing, drying hair, applying lotion to one's face, flushing toilet or arranging items on the sink etc. 27 trials were recorded and are categorized under Other-Activity (OAct). Associated sounds were also recorded with all trials mentioned above and data was recorded in subject's home and in a lab in which activities were freely executed.

3. DETECTION, CLASSIFICATION AND MONITORING OF ACTIVITIES

In our previous studies, we used the 2 step method to detect, classify and monitor the ADL. First, we localized the subject to a specific location in a home using the motion sensors. Then, we used the wearable accelerometer sensors to detect the arm motions while the system analyzes, classify and monitor the progress of activities in execution. But, in this study we include audio signals to detect events that generate specific sounds to enhance the classification rate. Figure 2 shows the reasons behind this approach. As we can see from Figures 2-(a) and 2-(b) electric tooth brushing (EB) and electric shaving (ES) can be distinguished in the frequency domain. Researchers have been using accelerometer to detect human activities [8, 9]. There have been studies that incorporate audio and video recording for labelling data purposes for activity detection [10]. They have also used audio and wearable sensors to distinguish whether the subject was in a noisy environment or not and inferred



(a)

(b)

Figure 2 – Time-Frequency map for Electric tooth brushing and Electric shaving. Above figures show the differences in time-frequency domain which can be used as features to distinguish EBrush (EB) and EShave (ES) activity.

if the subject was inside or outside [10]. But, in this study, we use both motion and audio data for activity detection and classification.

Similar to our previous studies, we used Gaussian mixture models (GMM), to model the 6 different classes discussed in section 2. For the 2-axis accelerometer, we extracted 5 frequency domain (FD) features from each axis and 3 time domain (TD) features such as mean, root mean square and the number of zero crossings for each accelerometer axis. Thus 10 FD features and 6 TD for the accelerometer data are extracted [12]. For audio, we extracted 12 FD features for each activity and combined them in the GMMs.

For each accelerometer data, we obtained FD and TD features by calculating them for each 64 sample segment (≅1.2secons), we calculated the energy in dyadic frequency bands as indicated in [12]. In addition to the 16 features mentioned previously, we performed analysis on the audio recordings which was recorded at 22kHz rate with 8 bit resolution and extracted 12 more FD features. Since the sampling rate was much higher for the audio compared to the accelerometer data, we sampled audio data to match the feature size of the accelerometer data. Therefore, the audio data is sampled across several time points by multiples of 1024 samples. Thus, the total duration of audio data used is only several seconds. Therefore, speech cannot be reconstructed intelligibly to invade privacy of the user. From the sampled audio data, we used a fixed window size of 1024 samples and shifted along the time with 50% overlap across

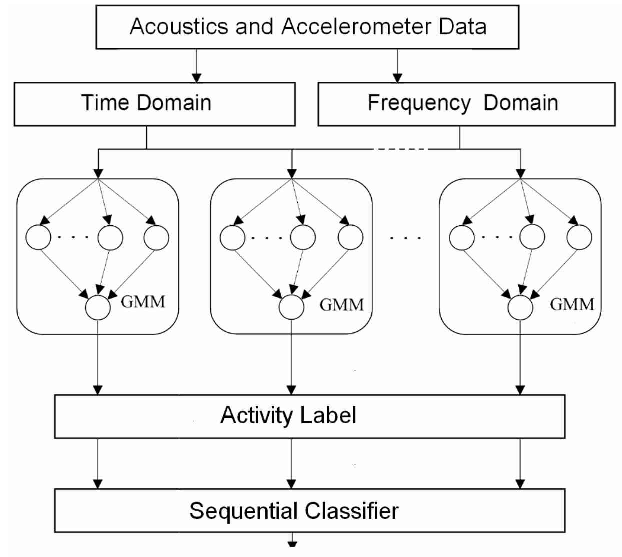


Figure 3 - A schematic diagram of the classification system. Activity Label module outputs discrete labels which is input to the Sequential Classifier for final decision output.

the sampled signal. FD features were extracted from the energies in the frequency domain.

In order to classify the extracted features Gaussian Mixture models were used. A GMM probability density function (pdf), is defined as a weighted combination of N Gaussians.

$$p(x/\lambda_k) = \sum_{c=1}^{N} w_c \eta(x/\mu_c, \Sigma_c) \qquad k = 1, .., K \quad (1)$$

where λ_k is the model, x is the feature vector, η is the D dimensional Gaussian pdf component

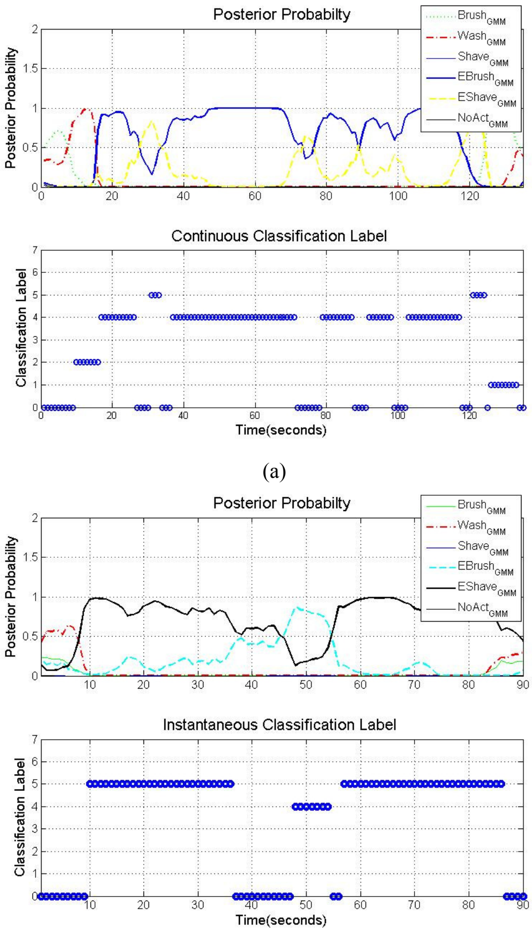
$$\eta(\mu_c, \Sigma_c) = \frac{1}{(2\pi)^{D/2} |\Sigma_c|^{1/2}} \exp(-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1} (x-\mu_c)) \quad (2)$$

with mean vector μ and covariance matrix Σ. The w_c is the weight of each component and satisfies

$$\sum_{c=1}^{N} w_c = 1. \quad (3)$$

Therefore, inputs to a GMM with a set of feature vector will be assigned a label with respect to the posterior probabilities of each GMM.

As we show in Figure 3, the TD and FD features are extracted from accelerometer and audio signals and fed to the system. Extracted FD features from acoustic sensor are converted to log scale and combined with accelerometer data features at the feature fusion stage. This resulting feature vector now has a dimension of 28. Separate GMMs were generated for each activity with 2 mixtures for each activity. Then, the outputs of all GMM were processed by Sequential Activity Label module and discrete activity label is assigned for each time point. Finally, Finite State Machine module takes into account the duration of the discrete classification labels and state transitions to generate an activity label of the current task in process. The reader is referred to [11] for details.



(a)

(b)

Figure 4 - (a) The input of the finite state machine module related to EB activity. (b) The input of sequential activity label module related to the ES activity. Note that several local FPs exists during the activity.

4. RESULTS

In this section, we summarize our findings in 2 subsections and discuss the subject and device generalization capability of a wearable early morning activity detection system.

4.1 2 Hand Data Analysis

Some activities in the bathroom require two hands to be used to execute them while for other activities only one hand is used whether it is left or right. For the subjects that participated in our study, all of them were right-handed. Thus, we present results obtained from this data. Using the data obtained from 2-hand motion, we observe the synchronous hand movements especially during the face washing activity. But, for other activities, such as brushing teeth, shaving face, electric brushing and shaving, each subject showed different behavioural patterns. Left hands were mostly placed on the waist, on top of the sink or they lifted it and placed the hand in front of the chest to balance the body. This is similar to the behavioural patterns observed during face washing activities [11], in which, different groups of people demonstrated different behavioural patterns. Small number of additional recordings from 4 other subjects showed that left hand patterns could be grouped. But, subjects also moved their left hands, touching several objects around the bathroom, opening the cabinet/drawer, etc. Therefore, it generated lot of variation in the feature space for left hand and

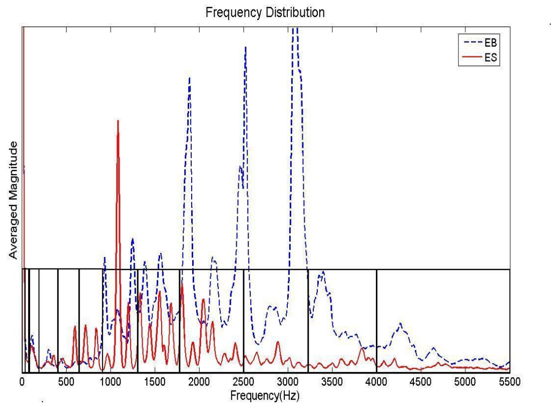


Figure 5 – Frequency division applied to real EB and ES data taken from few subjects. It is divided into 12 discrete bands. Much overlap is observed in the frequency domain but the band powers show differences between the two activities.

caused confusion for the system and system failed to classify correctly. Overall classification accuracy using 2 hand accelerometer data for brushing, washing and shaving showed 98% but for tasks such as EB and ES, classification accuracy was around 65%. Using 4 or more mixtures showed similar results compared to using 2 mixtures for 1 hand data. Therefore, we sacrificed complexity of the system for no performance gain. We therefore included audio sensor to record sounds generated during the execution of activities. Thus, in the following subsection, data from accelerometer and audio sensors will be discussed with its results.

4.2 Audio Data Analysis

Recall that, in our system, the finite state machine module is like a counter which checks the duration of sub activities. If the duration is shorter than predefined limit, the counter will neglect local FPs. Thus, even when we see many local FPs as in Figure 4, the sequential activity label module will neglect local FPs and output a correct label for figures 4(a) and 4(b) [12]. Note that this is the stage where the continuous probabilities are converted to discrete label assignments. Different classes are assigned with a unique numbers so that, 0: No Activity (NoAct), 1: Brushing, 2: Washing, 3: Shaving, 4: Electric Brushing (EB), 5: Electric Shaving (ES), 6: Other Activity (OAct). Here, NoAct means that the classifier cannot make a decision, since no posterior probabilities is greater than a threshold at a given point.

For classification, Leave One Subject Out (LOSO) analysis was performed across subjects, such that each subject's data was left out during training stages and was only used for testing the system. Our initial observation for LOSO subject training without audio showed that the correct classification rates for EB and ES were much lower compared to the results shown in the previous section. This is due to the fact when electric devices are used, not enough hand/arm motion is necessary to complete the activity. Also, motion observed for electrical tooth brushing (EB) and electrical face shaving (ES) are quite similar. Therefore the system was only able to achieve 54% classification rate for these 2 activities.

We note that the subjects used different electrical devices for EB and ES resulting differences in the sound patterns. We incorporated the sound data to distinguish EB from ES activity. The audio data was synchronously recorded with the accelerometer data using the system architecture.

The energy distribution across the frequency band showed a possible separation between the pure tone sounds of the electrical brush and electrical shaver. Thus, the frequency bands were di-

Table 1 – Confusion matrix for different activities for mixture 2.

Tasks	Brush	Wash	Shave	EB	ES	OAct
Brush	43	0	0	1	0	0
Wash	1	44	1	0	0	0
Shave	0	0	26	0	0	0
EB	4	0	0	64	7	0
ES	1	0	0	7	62	3
OAct	0	1	0	1	0	25

vided into 6 discrete bands to calculate energy in each band. But, when the activities were initiated the energy distribution changed greatly, especially for the ES activity. This was due to the fact that subjects had different amount of hair, different hair growth rate and some subjects shave every other or 3rd day, which caused variation in ES sounds. Thus using 6 bands caused lots of overlap in the frequency domain with EB activities that produced True Negatives (TN) and were classified as EB activities. Further analysis in the frequency distribution showed that 12 frequency bands were needed to better distinguish the EB from ES activities. Figure 5 shows the band split along the frequency axis and how they are position in a real EB and ES frequency data. Graph is obtained from data averaged across few different subjects. We observed the frequency feature to be very similar at frequencies higher than 5500Hz. Therefore, for the purpose of illustration, up to 5500 Hz is displayed.

Since the frequency components overlap during the execution stage of an activity we prefer to find a near pure tone of the device. In order to obtain a pure tone or a near pure tone for the electric devices, we observed the behavioural patterns of the subjects. For EB and ES activities, we noticed that people usually engage in the activity immediately after they turn on the device. But, especially for the ES activities, as the activity progresses more hair is cut off, we can record near pure device tones. Thus, we extract the audio feature towards the end of the activity.

When 6 frequency bands were used along with the accelerometer data, the classification accuracy was 73% for EB and 70% for ES activities. But, increasing the audio features to use 12 frequency bands, we were able to achieve 85% classification accuracy rate for both EB and ES activities. We observed that the error was mostly due to background noise. When the sampled audio portion included fan or speech sounds, the system misclassified the activities. This is an important factor since the background noise could be different for each subject since the activities are executed in their own home environment. The usual background noise found in the bathroom is the sound of the ventilation fan, sound of the water flushing down the toilet and speech that could occur during the execution of the activities. Also, the fan and water sound for each home could be different. Due to these factors, we also noticed that EB and ES activities with background noise are almost always misclassified. We also observed that ES are classified as Shaving and EB are classified as Brushing. In practice, these do not do any harm. Sometimes, the classifier could not make decision (NoAct) because the uncertainty was high. If this is the case, the system can issue a reminder to execute which user can simply ignore. If we assume that classifying EB to be brushing and ES to be shaving is tolerable, we obtain about 90% accuracy for these activities. But, our future work is to continue to explore different factors such as environment noise to enhance the performance of the classifier.

Finally, we show the confusion matrix in Table 1, which shows that for Brushing, Washing and Shaving activities, the classifier is able to make good detection but due to the reasons discussed above, higher misclassification is observed for EB and ES tasks. We observed that for EB, 4 instances were misclassified as the Brush and 7 cases were misclassified as ES. For ES, 7 trials were misclassified as EB and 1 trial misclassified to Brush. Therefore, the goal of the future studies would be to reduce the number of FPs for EB and ES while increasing TPs in classification.

5. CONCLUSION

In this paper, we described a system which is intended to assist people with cognitive impairments due to TBI. In particular, we focused on the problems of detecting, classifying and monitoring the progress of activities of daily living at home. We explored the potential of using audio sensors for activity detection and discussed several issues that arise during the detection of the activities. We plan to continue to explore the effects of environmental noise and noise removal to enhance the classification accuracy for system generalization. We showed experimental results from 7 subjects while completing face washing, face shaving and tooth brushing, electric tooth brushing and electric shaving activities. Our preliminary results using accelerometers and audio sensors are quite promising. We plan to explore the capability of the system to TBI patients upon receiving the approval and explore other important activities within a home setting to assist TBI patients. For the audio and accelerometer data analysis, we plan to compare our algorithm with other classification algorithms to better understand the efficiency of the proposed approach. We believe that integration of proposed classification system with a PDA like planner device would allow us to develop an intelligent reminder system. Such system could assist TBI and other cognitive impaired patients by allowing them to live an independent life.

REFERENCES

[1] National Center for Injury Control and Prevention, http://www.cdc.gov/ncipc/tbi/TBI.htm. 2006

[2] “Traumatic Brain Injury”, Department of Veterans Affairs Employee Education System, 2004.

[3] M. E. Pollack, “Planning technology for intelligent cognitive orthotics”, *in Proc. of the 6th International Conference on AI Planning and Scheduling*, Toulouse, France, Apr. 23-27, 2002

[4] N. F. Ince, C.-H. Min, A. H. Tewfik, “Integration of wearable wireless sensors and static home sensors to monitor the activities of daily living,” in 3rd *IEEE MDBS Symposium*, MIT Boston, 2006.

[5] M. Philipose, K.P. Fishkin, M. Perkowitz, D.J. Patterson, D. Fox, H. Kautz, and D. Hahnel. “Inferring Activities from Interaction with Objects,” in *Pervasive Computing*, vol. 3, no. 4, pp. 50-57. 2004

[6] J. Chen, A.H. Kam, J. Zhang, N. Liu and L. Shue, “Bathroom Activity Monitoring Based on Sound,” *Pervasive Computing LNCS*, pp. 41-67, Springer-Verlag, 2005

[7] E. D. Mynatt, I. Essa and W. Rogers, “Increasing the Opportunities for Aging in Place”, *Proceedings of ACM Conference on Universal Usability*, Arlington, VA, USA, Nov. 16-17, 2000.

[8] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Petola and I. Korhonen, “Activity Classification Using Realistic Data From Wearable Sensors”, *IEEE Trans. on Info. Tech. in Biomed.*, vol. 10, no. 1, pp. 119-128, Jan. 2006.

[9] L. Bao and S. Intille,“Activity Recognition from User-Anotated Acceleration Data”, in *Pervasive 2004 LCNS*, Springer-Verlag, pp. 47-61 2004.

[10] J. Lester, T. Choudhury, N. Kern, G. Borriello and B. Hannaford, “A Hybrid Discriminative/Generative Approach for Modelling Human Activities”, IJCAI 2005, pp. 766-772, 2005

[11] C.-H. Min, N. F. Ince and A. Tewfik, “Generalization Capability of a Wearable Early Morning Activity Detection System”, in *Proc. EUSIPCO 2007*, Poznan, Poland, Sep. 3-7.2007.

[12] N.F. Ince and C.-H. Min and A. Tewfik, “Detection of Early Morning Daily Activities with Static Home and Wearable Wireless Sensors”, EURASIP, Journal on Advances in Signal Processing, vol. 2008.