Applications of UBMs and I-Vectors in EEG Subject Verification

Christian Ward
Department of Electrical and Computer Engineering
Temple University
Philadelphia, Pennsylvania
Email: christian.ward@temple.edu

Dr. Iyad Obeid
Department of Electrical and Computer Engineering
Temple University
Phildelphia, Pennsylvania
Email: iobeid@temple.edu

Dr. Joseph Picone
Department of Electrical and Computer Engineering
Temple University
Philadelphia, Pennsylvania
Email: picone@temple.edu

***Abstract*—The processing of electroencephalograms (EEGs) is a growing field where mature speech processing techniques are able to rapidly progress development and understanding of the associated neuroscience. I-vectors and Joint Factor Analysis (JFA), along with their foundational universal background models (UBMs) have progressed to a level of understanding that makes them prime for transition to the EEG community. To prove the capability of these techniques they are tested against two contrasting EEG data sets, PhysioNet's EEG Motor Movement/Imagery Dataset and the Temple University Hospital EEG Corpus, to highlight the effectiveness of the techniques with minimal domain knowledge modifications. The initial results, presented as equal error rates as low as 20%, support the development of these techniques as a viable approach to addressing subject verification within and across subjects.**

I. Introduction

Many techniques for analyzing EEGs come from the speech community [1], but require modification due to the increase in complexity of EEG recordings compared to speech recordings. EEG records are not standardized across subjects, hospitals, or research groups, as evidenced by our own TUH EEG Corpus [2], nor is there currently consensus on subject recognition techniques. Sorting recordings without their associated labels is a difficult and often unsuccessful process even with the aid of clinical neurologists [3].

The use of joint factor analysis (JFA), along with the resultant identity vectors (i-vectors), described in [4] progressed in addressing speaker recognition problems for speech signals. Since 2007 this work has shown merit as a technique to address speech related identification issues: subject identification, channel modeling, and phoneme detection [4][5][6][7]. Given the success of this tool when applied to speech signals, this paper applies the basic technique against electroencephalograms (EEGs) of the human brain.

By applying the tools of JFA and i-vectors to two unique EEG data sets, the PhysioNet EEG Motor Movement/Imagery Dataset [8] and the TUH EEG Corpus [2], the feasibility of these techniques can be determined with respect to EEG data. Both techniques rely on universal background models (UBMs) generated for each subject which makes them dependent on initial feature selection and data quality. Working with two unique data sets provides a robust testbed to validate feature selection and resultant UBMs that can be used to develop i-vectors. This is a preliminary step in developing standards for these techniques in address subject verification in EEGs, with the eventual goal of improving medical diagnosis.

II. Materials & Methods

A. Universal Background Models & I-Vectors

Universal Background Models capture speaker independent features from training data, turning it into decision surface for speaker verification. The UBM can be developed to determine speaker specific models via a *maximum a priori* scheme. This relies on a linear interpolation of the models that capture channel (mobile phone, land line, in-person), noise (traffic, dog barks), and speaker features (frequency, tone) [4]. Such an approach is not ideal given that the model conflates speaker and channel information.

A speaker's data M arises from a speaker supervector s and a channel supervector c, in eq. 1. In speech a channel refers to recording medium (land-line, mobile phone, in person, etc), whereas with EEGs a channel implies a specific electrode.

$$M = s + c \tag{1}$$

The speaker supervector can be broken down to highlight common features universal to speech, m, while capturing eigenvoices, V, and speaker factors, y with residual matrix D and associated factors, z in eq. 2. This approach is followed with the channel supervector, producing eigenchannels U and channel factors x in eq. 3.

$$s = m + Vy + Dz \tag{2}$$

$$c = Ux \tag{3}$$

Universal Background Models are capable of modeling features common across the training data, but with JFA they become more powerful once decomposed into their foundational factors. The underlying assumption is that channel and speaker factors are independent normal Gaussian distributions with means of zero [5]. This technique turns a Gaussian Mixture Model (GMM) supervector M into eigenvoices V,

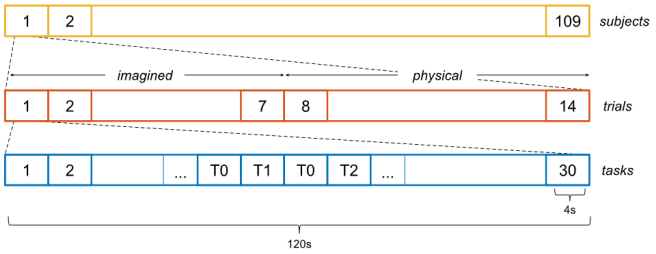

Fig. 1. Graphical representation of the PhysioNet data. Each trial contains 30 tasks which can be either event T1 (fists/left), T2 (feet/right), or T0 (rest)

eigenchannel U, and residual D matrices capable of tracking speaker and channel independence/dependence in eq. 4.

$$M = m + Vy + Ux + Dz \tag{4}$$

Starting with the eigenvoice matrix, each matrix is solved by an initial randomization coupled with a training data based estimation of the associated factor vector (y, x, z). The initial random matrix undergoes approximately 20 iterations of expectation maximization with respect to reducing the error of the 0^{th}, 1^{st}, and 2^{nd} order statistics of the training feature set. As each successive matrix is solved it is applied to the next solution for the remaining undetermined matrices thereby reducing the uncertainty in the supervector M.

$$s = m + Tw \tag{5}$$

The i-vector approach uses the independence of the three matrices to find a low-dimensional variability matrix to capture channel, voice and residual characteristics. Shown in eq. 5, this relies entirely on the speaker independent supervector m generated from the UBM. The i-vector weight matrix T is trained in the same manner as the V matrix from before, but there are now no channel subsets forcing the i-vector to capture channel, noise and subject characteristics. This generates an i-vector, w, that treats channel variations as minor features to the driving factors seen in the eigenvoices matrix. These variations on JFA reduce the computational need of the process and lessens the impact of under represented channels during training.

1) Software: MSR Identity Toolbox: Development of software tools relied on importing and modifying the freely available Microsoft Research Identity Toolbox for MATLAB [9]. The packaged software implements a Gaussian Mixture Model - Universal Background Model speaker-recognition and an i-vector Probabilistic Linear Discriminant Analysis speaker recognition. This quickly allowed for a baseline system to be tested without needing to adjust specific algorithm parameters.

In addition to processing the data, the toolbox supports evaluation by providing tools to present the equal error rate (EER) from detection error rate trade off plots. There are two confusion scoring matrices, one for GMM trials (UBMs) and one for Gaussian Probabilistic Linear Discriminant Analysis (GPLDA) trials (i-vectors), that are produced along with the EER. The results from these two evaluations form the basis of the results.

B. Data Sources & Selection

1) PhysioNet - EEG Motor Movement/Imagery Dataset: The PhysioNet EEG Motor Movement/Imagery Dataset contains recordings of 109 subjects at 160Hz from 64 electrodes placed in the standard 10-20 configuration. Each recording captures a single trial, with 14 unique trials per subject, each containing 30 tasks [8]. Half the trials require physical movement and half require imagined movement. The tasks are divided into contrasting actions: opening/closing fists (event T1) versus feet (event T2), opening/closing the left (T1) versus the right (T2) fist, and a rest state (T0). Figure 1 shows a trial sequence broken down into task order for the specific case of Trial 4. Each task's duration is 4 seconds. There were two additional recordings per subject, resting eyes open (REO) state and resting eyes closed (REC) state, which serve as calibration files and were not used in this work.

Data from the first three trials of subjects 1 through 5 were used to test the effectiveness of a subject specific UBM across each trial. The first three trials, labeled {R03, R04, R05}, were used with the UBMs training over over all trials for a given subject's results. This is later broken down into results from specific trials when compared against the full subject UBMs.

2) Neural Engineering Data Consortium - Temple University Hospital EEG Corpus: The data set is a collection

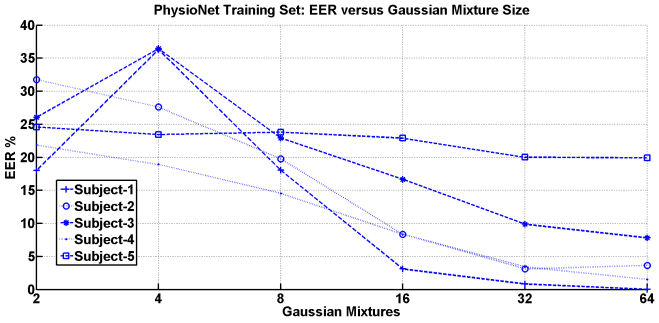

Fig. 2. Equal Error Rate plot for the PhysioNet subjects relative to number of Gaussian mixtures when comparing UBMs to training data.

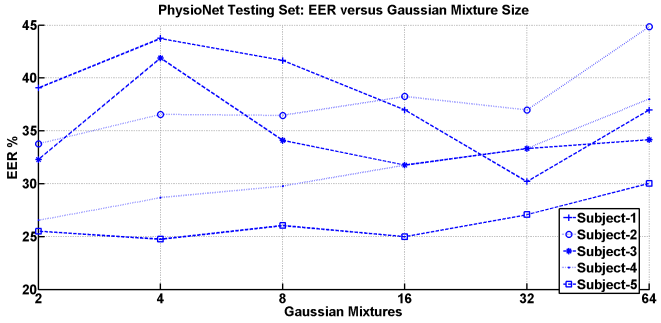

Fig. 3. Equal Error Rate plot for the PhysioNet subjects relative to the number of Gaussian mixtures when comparing UBMs to testing data.

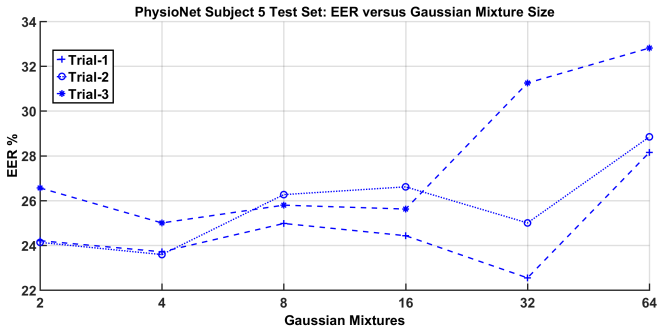

Fig. 4. Equal Error Rate plot for the PhysioNet subjects relative to the number of Gaussian mixtures when comparing UBMs to each trial of subject 5.

of patient recordings spanning 2002 to 2014 provided by Temple University Hospital Philadelphia, Pennsylvania. The data contains 247 sessions accounting for over 150 hours of EEG data. Recordings come from patients of various ages and both genders over the course of their treatment which often spans years. Each patient presents with a different electrode configuration, based upon treatment protocol, and a higher sampling rate than the PhysioNet data. In some cases the subjects were under photic stimulation to trigger responses as indicated in the associated header files. Multiple recordings are logged as sessions in the data, but will be referred to as trials in this paper to keep the language congruent between the two data sets.

Five subjects were chosen at random from the publicly available data set, subjects $\{1, 3, 5, 6, 9\}$ from folder 001 referred to here forward as subjects $\{1, 2, 3, 4, 5\}$. As this set lacks specific trials, subjects with multiple sessions were found and UBMs were trained that spanned two of each subject's trials. A single TUH EEG trial was longer than the aggregated three PhysioNet trials, so only two trials were taken for the TUH EEG data sets. These trials spanned anywhere from years later to hours later on the same day as the recordings were dictated by medical necessity.

C. Data Treatment

All of the files were processed in the same fashion regardless of data source. A one-second 50% sliding window was used to build the seven Mel Frequency Cepstral Coefficients which acted as seed features for the UBM algorithm [1]. These features were used to build UBMs that had $\{2, 4, 8, 16, 32, 64\}$ Gaussian mixtures over 10 iterations for each mixture for the PhysioNet data, while the TUH EEG data stopped at 32 mixtures due to less electrode channels. All data was partitioned into test-training sets of 10% test, 90% training.

Training T for the i-vectors was capped at 5 iterations given the limited size caused ill-conditioned matrices to be formed at higher iterations. The initial i-vector produced contained 100 rows, but was row reduced to one dimension less than the number of channels via Linear Discriminant analysis. The refined i-vector is modeled against all the trials to produce a general model for the subject that is then evaluated against the training set.

III. Results

A. PhysioNet Subjects

The training sets used to build the UBMs for each subject were initially tested against themselves to provide a baseline accuracy for the testing set. Figure 2 shows the results for each subject's UBM versus the GMM mixture count with increasing error reduction as the number of mixtures increases. The optimal mixture count of 64 was anticipated as that matches one-to-one with the number of electrodes. However, the equal error rates (EER) seen for mixtures of size 16 and 32 were not anticipated with most cases showing minimal error reduction after only 16 mixtures.

The testing data results in Figure 3 lacked an overall trend, but subject specific relationships emerged. Subject 5 performed poorly in terms of training accuracy, but presented with the most stable EER plots. For both sets of data, Figures 2 & 3, their EER is the within a similar range of 20%-30% error with minimal variance compared to the other subjects. Looking closer at subject five's individual trial training results in Figure 4 shows that individual trial UBMs performed no better than the UBM trained across all three trials in Figure 3.

B. NEDC Subjects

The TUH data training set baseline EERs in Figure 5 indicates a strong preference towards smaller mixture sizes

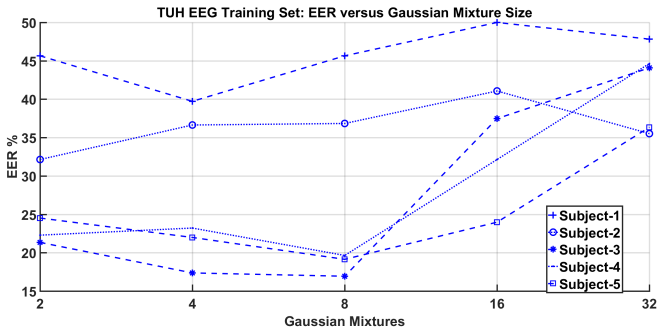

Fig. 5. Equal Error Rate plot for the TUH EEG Corpus subjects relative to the number of Gaussian mixtures when comparing UBMs to test data.

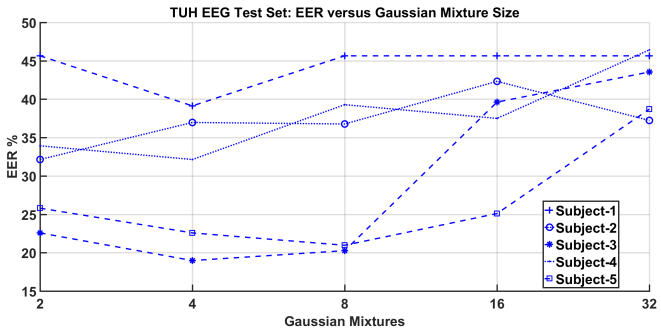

Fig. 6. Equal Error Rate plot for the TUH EEG Corpus subjects relative to the number of Gaussian mixtures when comparing UBMs to training data.

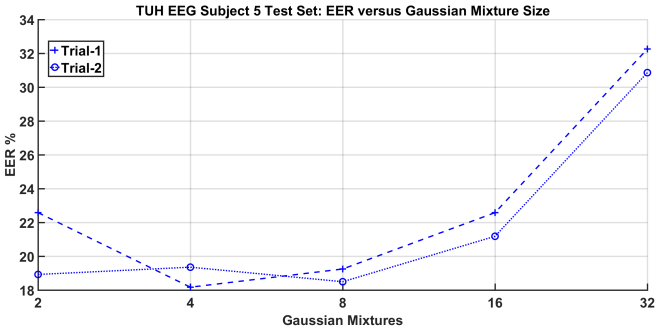

Fig. 7. Equal Error Rate plot of each trial(session) from subject 5 from the TUH EEG Corpus relative to the number of Gaussian mixtures.

for the UBMs. Subjects 3, 4, and 5 peak at 8 mixtures, while subject 5 peaks at 4 mixtures and subject 4 speakers at 16 mixtures. This differs from the trend seen in the PhysioNet data where they peaked at their number of electrodes, 64. Each TUH data set tested had between 16 and 32 active channels so an ideal mixture count was not achieved like it was in the PhysioNet testing.

With the training data EERs deteriorating as the number of mixtures is increased, the test data exhibits a similar response from the increased mixtures in Figure 6. Subject 4's increase in error contrasts with the consistency of the other low error training Subjects 3 & 5. The high error Subjects 1 & 2 remain consistent across test and training sets. This error is uniform across individual trial based UBMs, Figure 7, and shows at most a 5% improvement over being trained on all trials.

IV. Conclusion

The two datasets explored are divergent in terms of channels recorded, type of subject, and testing conditions. PhysioNet contains healthy subjects responding to visual cues captured with 64 electrodes while the TUH data contains medical patients being recorded to diagnosis brain conditions with variable electrode configurations. In both cases, generated UBMs from training datasets performed consistently across different mixture levels to identify the subject specific channels. Random channel selection of 64 electrodes yields an error of 98.44% while with 23 channels, the smallest TUH data subject tested, yields an error of 95.65%. The worst case EER never exceeds 50% and for most is below 40% highlighting

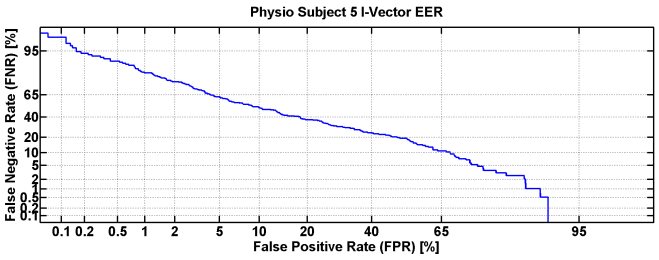

Fig. 8. Equal Error Rate curve of i-vector on PhysioNet Subject 5 training data using a 16 mixture UBM with a resultant EER of 29.7%.

the strength of UBMs to discern specific user channels on both data sets.

A disparity between the data is highlighted in Figure 8 showing the i-vector EER plot of the PhysioNet Subject 5 training set evaluation. Despite training across three unique trials, the optimal EER is 29.7% which is in range of error rates found for trial specific UBMs, Figure 4. The TUH subjects return i-vectors with EERs at or slightly below 50% which exceeds the EERs seen for the associated UBM training and testing data. Even when the results for both data sets on test data falls within an EER of 20% to 45%, the PhysioNet data results exceed the TUH data when tested on the training data.

The results suggest that the feature density of the two sets differs exposing complications when shifting from speech to EEG data. The PhysioNet data is hundreds of seconds long with multiple structured events while the TUH data is thousands of seconds long with minimal unstructured events. A lack of structure and increased variance from the TUH recordings increases the complexity of the UBMs when compared to the PhysioNet data. The development of robust i-vectors is feasible, given the PhysioNet results, but the technique must mature to address the feature complexity, cross-channel features and dynamic feature-to-noise ratios, present in clinical EEG data like the TUH data.

Acknowledgment

C. Ward and I. Obeid are supported by the Office of the Assistant Secretary of Defense for Health Affairs through the Neurosensory and Rehabilitation Research Award Program under Award No. W81XWH-15-1-0045.

References

[1] A. Harati, M. Golmohammadi, S. Lopez, I. Obeid and J. Picone. (2015). "Improved EEG Event Classification Using Differential Engery". *2015 IEEE SPMB Symposium*.

[2] Harati, A., Lopez, S., Obeid, I., Jacobson, M., Tobochnik, S., & Picone, J. (2014). The TUH EEG Corpus: A Big Data Resource for Automated EEG Interpretation. *In Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (pp. 15). Philadelphia, Pennsylvania, USA.

[3] Marcos Del Pozo-Banos, Jess B. Alonso, Jaime R. Ticay-Rivas, Carlos M. Travieso. (2014). Electroencephalogram subject identification: A review. *Expert Systems with Applications* 41(15): 6537-6554

[4] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel. (2007). "Joint Factor Analysis Versus Eigenchannels in Speaker Recognition". *IEEE Trans. on Audio, Speech, and Langauge Processing*, Vol. 15, No. 4.

[5] Kenny, P., Ouellet, P., Dehak, N., Gupta, V., & Dumouchel, P. (2008). A study of interspeaker variability in speaker verification. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(5), 980-988.

[6] C. Greenberg et al. (2014). "The NIST 2014 Speaker Recognition i-Vector Machine Learning Challenge". *Odyseey: The Speaker and Lanuage Recognition Workshop*.

[7] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet. (2011). "Front-End Factor Analysis for Speaker Verification". *IEEE Trans. on Audio, Speech, and Langauge Procressing*, Vol. 19, No. 4.

[8] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220

[9] S. O. Sadjadi, M. Slaney, and L. Heck. (2013). "MSR Identity Toolbox v1.0: A MATLAB Toolbox for Speaker-Recongition Research". *Speech and Language Processing Techincal Committee Newsletter*.