Feasibility of Identity Vectors for use as Subject Verification and Cohort Retrieval of Electroencephalograms

Christian Ward[1] and Dr. Iyad Obeid[1]

***Abstract*—The success of Identity Vectors in speech recognition as a tool for subject verification, language detection, word recognition, and accent/dialect classification suggests the technique is a robust method of unsupervised learning on high dimensional data, such as electroencephalograms. Tests run on the PhysioNet EEG Motor Movement/Imagery corpus concerning the matching of subject specific trials showed an average verification of 99% for the 109 subject-trial tests. Further tests on the ability to cluster repeated subject-trials produced at least one matching subject-trial for 60% of the subjects. The driving component of the Identity Vector process is the creation of Universal Background Models derived from single dimension Gaussian mixtures of user defined sizes operating on cepstrum feature coefficients. Taken as a whole, the results of this work indicate that Identity Vectors can be effective at distinguishing between subjects and show promise when asked to generate cohorts of related data.**

I. Introduction

Development of an EEG indexing system for research datasets is critical to the development of tools designed to process patient EEG records. Neurologists' ability to annotate records is limited by their time and their exposure to various EEG phenomena [1]. Annotated training sets lead to robust detection and categorization algorithms using the existing domain knowledge of neurologists. When faced with EEGs deviating from known phenomena neurologists struggle, as do the algorithms they helped develop [2]. The ability to link known recordings to previously unlabeled data will boost the development of detection algorithms and overall EEG diagnostic power.

Research from the speech community shows it is possible to discern speakers, environments, and forms of communication from speech recordings using unlabeled data [3]. The initial work in this area showed the effectiveness of I-Vectors as not only subject verification tools, but as a framework to resolve relationships between subject recordings [4]. As the EEG community struggles with similar problems the use Universal Background Models (UBMs), Joint Factor Analysis (JFA), and I-Vectors may prove beneficial to the advancement of unsupervised EEG annotation and diagnosis. However, the increased data complexity of EEGs with respect to speech data means the aforementioned techniques must be developed and verified to assure such techniques can be translated to the new field[5].

[1]Department of Electrical and Computer Engineering at Temple University in Philadelphia, Pennsylvania. *christian.ward@temple.edu is the corresponding author.*

The main aim of this work is to setup a framework for cohort retrieval across the NEDC EEG Corpus to aid clinicians and researchers in finding pertinent data from within the database for their work [6]. Implicit within this goal is reducing the need for professional neurologists to annotate EEG recordings. When professional annotation is needed the records neurologists review contain relevant features and events culled from this unsupervised learning process. At the same time, EEG records can be linked based upon their found features to enhance clinical searches otherwise carried out only on medical reports. The development of these tools will hopefully lead to a reduction in the need to manually review data will improve the diagnostics concerning EEG tests.

II. Materials & Methods

A. Data Source: 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 shown in figure 1 [7]. 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). There were two additional recordings per subject, resting eyes open (REO) and resting eyes closed (REC), which serve as calibration trials for the original experiment.

Data from the calibration trials were used to test the effectiveness of subject verification under near ideal conditions; No exterior sensory input to the subject. Each set contains a single trial spanning all subjects with the REO trials labeled as set R01 and the REC trials labeled as set R02. The non-calibration trials were grouped into similar sets labeled R03 through R14 with results aggregated over all subjects/trials to show error trends and cohort retrieval probability. In all cases the data for building the UBM, the training I-Vectors, and testing I-Vectors are identical for a given test. The full data set contained R01 through R14 and the motion data set contained R03 through R14.

B. Data Treatment

A one-second 90% overlapping sliding window was used to build the 26 cepstral coefficients used as the baseline features

for the UBM algorithm [8]. The generated UBMs spanned $\{2, 4, 8, 16, 32, 64, 128, 256, 512\}$ Gaussian mixtures to adequately capture the variance in the data given the number of subjects (109), channels (22), and features (26).

C. Universal Background Models and Joint Factor Analysis

The UBM represents subject-independent characteristics as a set of *n* Gaussian mixtures. Each mixture contains *m* independent Gaussian distributions matching the number of source features[9]. In this work, the Gaussian mixtures are built from the 26 cepstrum coefficients of the processed EEG signals. These UBMs provide a base over which models can be developed through I-Vector based recognition.

D. I-Vectors

Creation of an I-Vector depends upon a total-variability matrix $\{T\}$, the UBM supervector $\{m\}$, the targeted I-Vector $\{M\}$, and the subject specific I-Vector $\{w\}$ shown in Eq. 1. This process is an adaption of Joint Factor Analysis shown in equation 2 with the speaker/session factor vectors $\{y, x, z\}$ and variability matrices $\{V, U, D\}$.

$$M = m + Tw \tag{1}$$

$$M = m + Vy + Ux + Dz \tag{2}$$

The m vector is a supervector of the means and variances gathered from the UBM. The training data is used to generate I-Vector, s, which serves as an optimization target for w. As m contains both mu and sigma, w has rows equal to twice the number of features. The columns of w, and thus the size of the I-Vectors, is capped at the min of 100 or $n - 1$ where n is the number of subjects. The upper limit $\{100\}$ is adjusted based upon the dimensionality of the data and the needs of the user.

Training T, or its JFA equivalents, is reliant on the covariances of the UBM and Baum-Welch statistics generated from the adjusted means of the UBM means and test data. This process is identical for I-Vectors and joint factors, but is beyond the scope of the paper. A proper discussion of this technique is covered by Kenny *et al* in [10].

The dependent factors vector $\{w\}$ becomes the resultant I-Vector for each M once T is resolved. This differs from JFA where V, U, and D enable specific solutions based upon channel and session variability [11]. I-Vectors combine channel and session discrimination into one vector further reducing the dimensionality reduction seen in JFA.

E. Joint Factor Analysis

In an effort to catalog the influences {speaker, channel, and noise} present when a speaker produces an utterance, the speech community developed tools to split a speech super-vector {UBM m} into three parts {eigenvoice matrix V, eigenchannel matrix U, residual matrix D} shown in Eq. (2). This represents the fundamental application of JFA through the separation of the varied components as they are

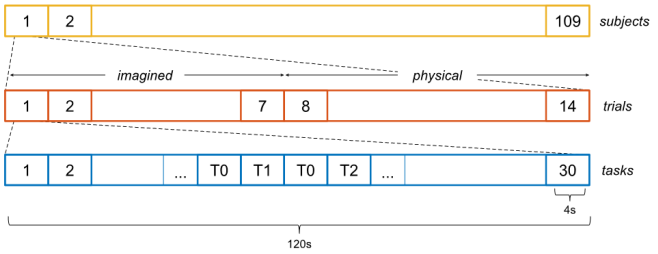

Fig. 1. Graphical representation of the PhysioNet data. Each trial contains 30 tasks which can be either event T1 (fists/left), T2 (feet/right), or T0 (rest)

sequential solved for in an iterative process [4]. The vector categorizing all of these influences is M which used the UBM and weighted components of each matrix to produce an ideal speaker supervector.

Solving for the next factor in Eq. (2) requires all previously solved matrices before updating the Baum-Welch statistics. After which the new estimations undergo the same mathematical process to create the missing matrix. A deeper treatment of this approach comes from Kenny *et al* in [12] where the idea was introduced.

F. Evaluation and Scoring

Both sets of models, I-Vectors and UBMs, are evaluated against the feature data to produce an Equal Error Rate (EER). EER indicates the intersection of the false negative rate and false positive rate for a given subject's model. The I-Vector score is produced by using a Gaussian Probabilistic Linear Discriminant Analysis (GLPDA) over the raw feature set. For the UBM mixtures, scores are generated using the loglikelihood result of the models compared to the raw feature set.

The UBM and I-Vector models are matching against each channel in each trial making their evaluation channel agnostic. Thus a match occurs on the channel level which requires 22 matches for perfect trial verification.

In both instances, the scores represent the model's ability to discriminant against the feature data instead of other subject models. This enables the ability to discern strength of match over a given subject's data set providing the ability to rank matches. Cohort scoring relies on these rankings not for the their raw score, but the order of the scores being in the top 3 or 5 as needed by the experiment.

G. Software: MSR Identity Toolbox

Development of software tools relied on importing and modifying the freely available Microsoft Research Identity Toolbox for MATLAB [13]. 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

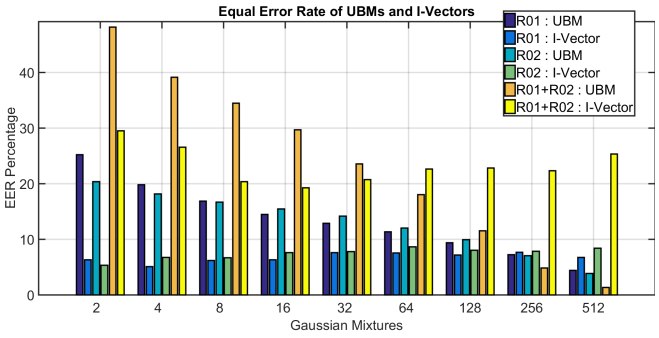

Fig. 2. Equal Error Rate on the eyes closed/eyes open calibration data shown as UBM mixture size increases.

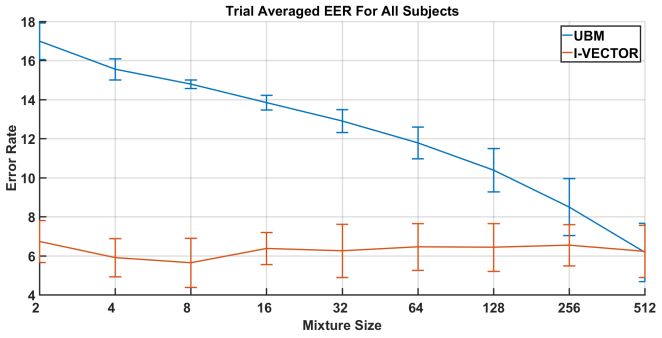

Fig. 3. Equal Error Rate of each motion trial averaged across each subject shown as UBM mixture size increases.

(EER) from detection error rate trade off plots. There are two confusion scoring matrices, Gaussian Mixture Models (GMMs) for the UBMs and Gaussian Probabilistic Linear Discriminant Analysis (GPLDA) trials (I-Vectors) which produce likelihoods of the models match to the data.

H. Experiments

The techniques proposed are new to the field of EEG requiring a baseline measurement to prove functionality and validity of the techniques. Testing on the channel and trial level is necessary to address the increase in dimensionality from speech data to EEG data.

1) Experiment 0: I-Vector Viability: Verification of a subject's trial specific channel, channel-trial, against all other channel-trials provides the lowest level of verification possible: matching channels to their common trial. An EER based on matching channels within a subject's resting state trials is used to determine an ideal minimum error rate. This experiment generates a baseline trend for accuracy with respect to the number of available Gaussian mixtures.

2) Experiment 1: Trial Verification: Verification at the trial level, that a subject's trial matches other subject's trials, constrains all channel data to be compacted into one I-Vector per subject trial. Two data sets were generated: (1) a set of all of a subject's trial data and (2) a set of only a subject's motion trials. These two sets provide overlapping testing environments to highlight the influence of the resting state trials. Evaluating these data sets for their trial specific EER showcases an abstraction of the feature data into a higher dimensional space than the channels.

3) Experiment 2: Cohort Retrieval: From the ranked results of Experiment 1, the top subset of matches is compared against similar trials within a subject. The PhysioNet experiment repeated four trials three times, the basic sets are trials {3, 7, 11}, {4, 8, 12}, {5, 9, 13}, and {6, 10, 14}. When the full data set is run the two resting state trials {1,2} are added to make each set five trials. As the scores are ranked, groups can be made out of the top 3 or 5 matches to search for similar subject trials from the previously sets.

III. Results

A. Experiment 0: I-Vector Viability

Figure 2 shows error rates on the resting trials, resting eyes open and closed, which are known to be distinct brain states given the presence of alpha waves during eyes closed [14]. These are a control because the subjects were not responding to stimulus during calibration as they were required to during motion trials. The results are aggregated across all subjects to highlight errors in matching channels to models as a function of Gaussian mixture size.

Error rates for the all subjects over the twelve motion trials data set {R03...R14} is shown in figure 3. Each subject's data was run as an individual analysis and then averaged over the common mixture size. The error bars for each mixture show $\pm$ 1 standard deviation across subjects.

B. Experiment 1: Trial Verification

Twelve subjects did not produce perfect trial verification within the full data set, but only one subject failed to produce perfect trial verification with the motion data set. The full data set subject trial match of 0.9908% is less than the rate seen for the motion data set of 0.9977%. The minimum verification for a subject's trials in the full data set is 0.8571% and 0.7500% for the motion data set. Figures 4 and 6 present individual results from Subject 001 using Gaussian mixtures of size 4 and 512 which exemplify the results being aggregated overall subject and mixture combinations.

C. Experiment 2: Cohort Retrieval

Evaluation of the ranked results based upon the clustering of common trials is seen in figures 7 and 5. Matching sets differ between the two data sets because figure 5 does not include the resting trials, producing matches based on a set of three as opposed to figure 7 with the resting trials having set of five. The first match is almost always the native trial, following from the result of Experiment 1, making the additional matches others within the trial set.

Within the motioin data set a second match is found for roughly 30% of the subjects and a third match is a non-zero percentage for all trials. Shifting to the full data set shows the strongest match cases are for two and three matches, over 60% of all trial matches, with a decrease in single matches.

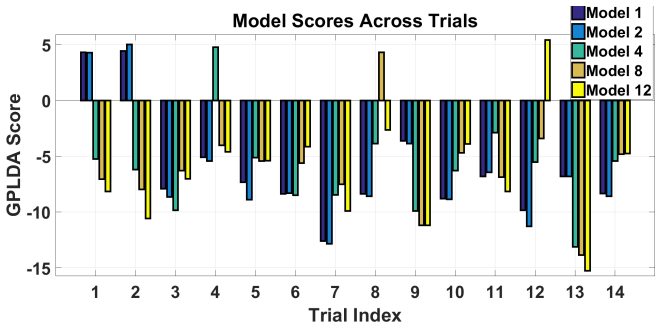

Fig. 4. Trial matches for subject 001 using 4 Gaussian mixtures, the trial set highlighted is {4, 8, 12}.

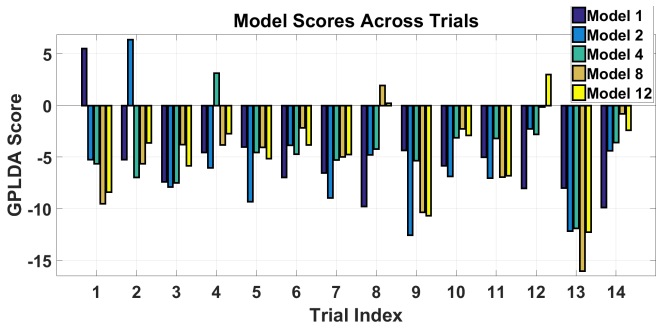

Fig. 6. Trial matches for subject 001 using 512 Gaussian mixtures, the trial set highlighted is {4, 8, 12}.

IV. Conclusion

A. Experiment 0

In figure 2 the EERs of the UBM and I-Vector evaluations track each other as mixture size is increased. The I-Vector EER results for the R01+R02 data set fail to show equivalent improvement as their related UBM EERs. This suggests a failure in the calculation of the I-Vectors and was investigated as such. Review of the individual results showed UBMs built with two Gaussian mixtures produced EERs over 50%, which is why figure 4 shows a mixture size of four instead of two.

Both figures, 2 & 3, trend toward an EER of 5% for the UBM and I-Vector models. The combine set {R01+R02} shows the EERs increases initially before the falling into the trend seen in the individual trials. Parallel behavior is seen for the motion trial data set in figure 3 indicating larger mixtures generate stronger matches across all the subject's trials.

Missing a channel match is a penalty of 4.5% which is near where the models converge for their minimum EER. This suggests for the datasets tested UBMs and I-Vectors are tenable for EEG data with an ideal channel based trial verification of 95%.

B. Experiment 1

Given the performance of trial verification throughout the datasets the cohort retrieval appears possible within a subject's trials. Figures 4 and 6 show the scores as a function of Gaussian mixtures. These plots illustrate the natural hierarchy formed by the I-Vector evaluations. Targets 8 and 12 in figure 6 both show a strong within set {4, 8, 12} preference for each other. Set {6, 10, 14} contains trials of a different imagery task, but appear as the next strongest matches after the native trial set.

Using only two Gaussian mixtures makes it difficult to properly track the two resting state trials {1, 2} as their scores mimic each other in figure 4. As the mixture count is increased to 512 Gaussian mixtures in figure 6, the discrimination of the resting states is improved. Additionally, the resultant scores are tightened up to the point that positive scores are reported for within set matches.

C. Experiment 2

Over the 109 subjects an I-Vector is always able to match to its training data trial and a secondary match within the top 3 scoring I-Vectors show in figure 5. When expanded to the top 5 results, figure 7, the matches exceed 60% for finding at least two within set trials for each subject. The presence of five matches in figure 7 shows that a full set matches do occur. These complete five matches appear less frequently than their complete three matches suggesting the full data set many not be producing improved scores.

As the placement of the matches was not tracked (outside of the best match), certainty of the cause of the improved

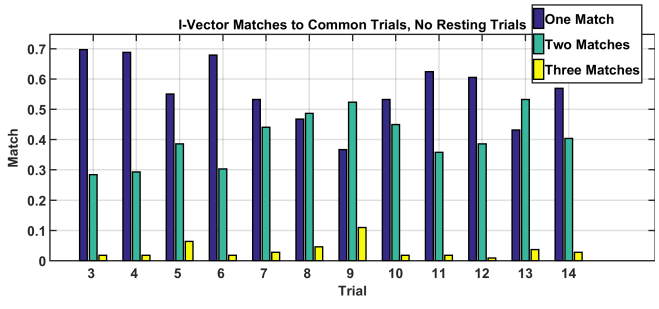

Fig. 5. Match percentage of I-Vectors based on four Gaussian mixtures over trials where sets are { (3, 7, 11), (4, 8, 12), (5, 9, 13) and (6, 10, 14)}. Trials 1 and 2 are not shown as they are resting eye trials and contain no attempts at motion.

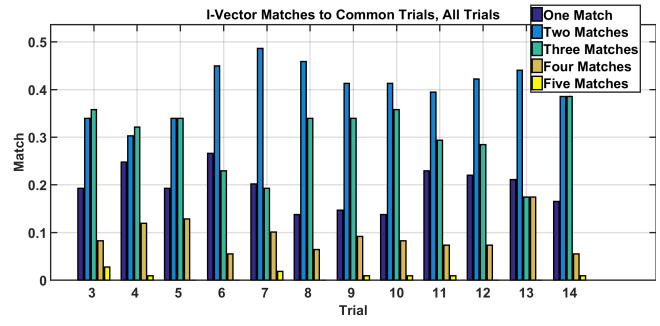

Fig. 7. Match percentage of I-Vectors based on four Gaussian mixtures over trials where sets are { (3, 7, 11, 1, 2), (4, 8, 12, 1, 2), (5, 9, 13, 1, 2) and (6, 10, 14, 1, 2)}. Trials 1 and 2 are not shown as they are resting eye trials and contain no attempts at motion.

matching is unclear. Figure 6 highlights this as the resting states, Trials 1 and 2, match poorly when compared to other motion trials. As such Trials 5 and 9, which exhibit the strongest three matches in figure 5, do not perform as well with five matches in figure 7.

An overall reduction in single matches coupled with an increase in two matches and three matches suggests the top ranking I-Vectors are indeed within the subject set. However, the weak performance of four and five matches suggests interplay between the sets may be weakening within set verification. This needs to be investigated further to determine if the highest matches are being mitigated by other trials that share properties {artifacts, imagery vs real motion} not accounted.

The approach detailed in this work suggests channel agnostic subject verification of EEGs is feasible when using UBMs and/or I-Vectors to model EEG data. Increases to the UBM Gaussian mixtures shows strong improvement of the EER across all subjects during all trials. Each Gaussian mixture size increase directly improves the UBM EER, but I-Vector EERs improve only in the early mixture sizes. With further development I-Vectors could be capable of producing reliable rankings when compared directly to the feature data and even against other I-Vectors.

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] Halford, J. J., Shiau, D., Desrochers, J. A., Kolls, B. J., Dean, B. C., Waters, C. G., ... & Sinha, S. R. (2015). Inter-rater agreement on identification of electrographic seizures and periodic discharges in ICU EEG recordings. *Clinical Neurophysiology*, 126(9), 1661-1669.

[2] Halford, J. J., Schalkoff, R. J., Zhou, J., Benbadis, S. R., Tatum, W. O., Turner, R. P., ... & Kutluay, E. (2013). Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis.' *Journal of neuroscience methods*, 212(2), 308-316.

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

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

[5] Del Pozo-Banos, M., B. Alonso, J., R. Ticay-Rivas, J., & M. Travieso, C. (2014). Electroencephalogram subject identification: A review. *Expert Systems with Applications* 41(15): 6537-6554

[6] 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.

[7] 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

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

[9] Reynolds, D. (2015). Universal background models. *Encyclopedia of Biometrics*, 1547-1550.

[10] Kenny, P., Boulianne, G., & Dumouchel, P. (2005). Eigenvoice modeling with sparse training data. *IEEE Transactions on Speech and Audio Processing*, 13(3), 345-354.

[11] Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., & Ouellet, P. (2011). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4), 788-798.

[12] 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.

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

[14] Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., & Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12), 2765-2773.