Evidence of Task-Independent Person-Specific Signatures in EEG using Subspace Techniques

Mari Ganesh Kumar, Student Member, IEEE, Shrikanth Narayanan, Fellow, IEEE, Mriganka Sur, Member, IEEE and Hema A Murthy, Senior Member, IEEE

Abstract—Electroencephalography (EEG) signals are promising as a biometric owing to the increased protection they provide against spoofing. Previous studies have focused on capturing individual variability by analyzing task/condition-specific EEG. This work attempts to model biometric signatures independent of task/condition by normalizing the associated variance. Toward this goal, the paper extends ideas from subspace-based text-independent speaker recognition and proposes novel modification for modeling multi-channel EEG data. The proposed techniques assume that biometric information is present in entirety of the EEG signal. They accumulate statistics across time in a higher dimension space and then project it to a lower-dimensional space such that the biometric information is preserved. The embeddings obtained in the proposed approach are shown to encode task-independent biometric signatures by training and testing on different tasks or conditions. The best subspace system recognizes individuals with an equal error rate (EER) of 5.81% and 16.5% on datasets with 30 and 920 subjects using just nine EEG channels. The paper also provides insights into the scalability of the subspace model to unseen tasks and individuals during training and the number of channels needed for subspace modeling.

Index Terms—Biometric, Task-independent, EEG, i-vector, x-vector;

I. INTRODUCTION

Person recognition using EEG is an emerging technology. Previous studies in EEG-based person recognition have been constrained to a particular task or condition. Several elicitation protocols have been proposed for EEG-based person recognition. A detailed review of these different protocols and their performance can be found in [1–3].

Multiple factors suggest that the EEG can contain signatures [4, 5] that help to uniquely identify individuals irrespective of the task, condition, or state of the brain. These include genetic differences between individuals, compounded by neural plasticity due to environmental factors and learning [6], which help specify neuronal connections and brain activity that are reflected in the EEG. The focus of the present work is to identify individuals independent of task or condition across sessions. To this end, the systems proposed in this paper builds upon and extend existing state-of-the-art text-independent speaker recognition techniques, namely, the i-vector system [7] and the x-vector system [8]. These modified systems were initially proposed in [9]. This paper provides a consolidated analysis of these systems and shows that the proposed modifications are better than simple early and late fusion techniques used for modeling channel information.

For the work presented in this paper, a database was systematically collected with different elicitation protocols. A 128-channel EEG system was used for this purpose. EEG data were obtained from 30 healthy volunteers while they performed various tasks. In addition, we also use a large clinical dataset with 100 subjects, a subset of an openly available EEG dataset collected in a clinical environment [10]. This clinical dataset was recorded with clinical tasks without standard elicitation protocols. Using these two complementary datasets with multiple tasks and sessions, we show evidence for task-independent person-specific signatures in EEG.

EEG analysis requires adequate spatial sampling to capture the functionality of the brain. Our work suggests that person-specific information is observed in signals from all regions of the brain, so that high spatial-resolution may not be essential. Although previous works have explored different sets of channels for task-dependent EEG biometrics, a systematic study of the spatial resolution needed for task-independent EEG biometrics is still lacking. Using the 128 channels EEG dataset, this paper systematically compares the models built with various subsets of sensors. Further, different methods of spatial subsampling are examined to find the best set of channels necessary for person recognition.

The organization of the paper is as follows. Prior art on EEG biometrics and our contributions are summarized in the remainder of this section. The details of the baseline and proposed EEG person recognition systems are given in Section II. Section III discusses the different datasets used in this paper. The general experimental setup and the features used are outlined in Section IV. The experiments and results are presented in Section V, followed by a discussion in Section VI. Section VII concludes the paper.

A. Related Work

Multiple factors such as (i) sessions, (ii) tasks used to elicit subject-specific signatures, (iii) number of channels used and their location, and (iv) the choice of features and classifier have all been shown to influence the performance of EEG biometrics system. This section presents a review of related work addressing the above-mentioned factors.

1) Testing across sessions: Most previous studies on EEG biometrics have used datasets that have only a single acquisition session [19]. The exogenous conditions such as impedance...
between the electrodes and the scalp, minor displacement in electrode location, power supply artifacts, and other factors that vary from session to session can affect both inter- and intra-subject variability of the EEG recordings [27]. Consequently, many recent studies have shown that the performance of EEG biometrics system is significantly affected by cross-session testing. In [21], the EER for 60 subjects was observed to degrade from 5.34% to 22.0%. In [28], without cross-session testing, 100% accuracy was observed with 40 subjects, whereas while testing 15 subjects across sessions, the performance dropped to 86.8%. Without cross-session testing, the performance of the EEG biometric system can be heavily influenced by session-specific conditions. Owing to this, all results in this paper are reported only by evaluating data from sessions unseen during training. Table I gives a summary of previous works in the literature that have tested EEG-based person recognition across multiple acquisition sessions. The results in Table I shows immense potential for using EEG as a possible biometric. In [27], even with an intersession interval of 36 months, the biometric system is shown to work with an average equal error rate (EER) of about 6.6%.

2) Tasks used for EEG biometrics: From Table I, it is important to observe that different studies have employed different elicitation protocols. The primary interest involved in studying different elicitation protocols is that individuals can have a distinctive signature for a given task, which can be leveraged to identify them. However, in Table I, almost every elicitation protocol has demonstrated success in EEG person recognition across sessions. This suggests that person-specific signatures are present under all task conditions, and hence biometric systems need not be designed for a particular elicitation protocol. Consequently, EEG biometric systems have been shown to work on multiple tasks or conditions by training and testing on each task separately [23–25, 27]. Since these systems are trained only on a specific task, they may not scale to other protocols. Building a task-independent EEG biometric system can eliminate the constraint of using an elicitation protocol. Recent studies have explored the task-independent nature of EEG biometric using single-session data and a small set of tasks [26, 29–32]. This lack of cross-session testing is stated as a significant limitation [26, 30]. Cross session testing is important because the session-specific factors are known to influence the EEG biometrics (Section I-A1). In [26], the results of using different tasks for training and testing across two sessions with <= 10 subjects on two datasets are presented. The first dataset has a resting state and four mental subtasks data. The second dataset has two tasks, which are EEG captured with self and non-self images as stimuli. These subtasks are not known to influence the EEG significantly. This has been stated as a major limitation in [26]. In [28], we studied task-independent nature on 15 subjects using five different elicitation protocols but limited to the closed eye condition. The present paper provides a detailed analysis of

| Ref. | Sessions per-person | Inter-Session Interval | No. of Individuals | No. of Channels | EPS/ Tasks | Feature | Duration of EEG segment | Classifier | Channel Handling | Performance |
|------|---------------------|------------------------|--------------------|----------------|------------|---------|-------------------------|------------|-----------------|-------------|
| [11] | 3                   | 3 days                 | 9                  | 8              | MT         | PSD     | N.A                     | UBM-GMM    | FC              | HTER = 36.2% |
| [12] | 2                   | 12 - 15 months         | 9                  | 8              | REC        | PSD based| 290s (median)           | LR         | FC              | ACC = 88%    |
| [13] | 2                   | 12 - 15 months         | 9                  | 8              | REC        | PSD based| 189s (median)           | LR         | FC              | ACC = 88%    |
| [14] | 4                   | N.A                    | 6                  | 106            | IS         | AR      | N.A                     | SVM        | FC              | ACC = 78.6% to 99.8% |
| [15] | 2                   | 1 year                 | 9                  | 53             | WM         | AR      | 60s                     | MD         | FC              | ACC = 64.7% to 77.8% |
| [16] | 2                   | 1 week (min)           | 4                  | 128            | REC        | CWT     | 2s                      | k-NN       | FC              | ACC = 92.58% |
| [17] | 2                   | 5 - 40 days            | 15                 | 1              | N400 [18]  | ERP     | 1.1s x 100               | Cross Corr.| N.A             | ACC = 89%    |
|      | 3                   | 4 - 6 months           | 9                  | 1              |           |         |                         |            | N.A             | ACC = 93.0% |
| [19] | 2                   | 9 months (avg)         | 20                 | 26             | VEP        | ER       | N.A                     | Cross Corr.| Voting          | ACC = 100%  |
| [20] | 3                   | 25 - 49 days           | 50                 | 17             | VEP        | ER       | 600ms x 50              | CS         | SF              | EER = 10% to 15% |
| [21] | 2                   | 2 weeks                | 60                 | 27             | REC        | Multiple | 20.5s                   | Cross Corr.| FC              | EER = 22%    |
| [22] | 2                   | N.A                    | 8                  | 9              | SSVEP      | PSD based| 1s                      | CNN        | FC              | ACC ≈ 97%    |

| Tasks/Methodology                        | Channel Handling | Performance |
|------------------------------------------|------------------|-------------|
| Elicitation Protocol                     |                  |             |
| MI - Motor Imagery                       |                  |             |
| REC - Resting Eye Closed                 |                  |             |
| WM - Working Memory                      |                  |             |
| IS - Imagined Speech                     |                  |             |
| MM - Motor Movement                      |                  |             |

Abbreviations: EP - Elicitation Protocol, MI - Motor Imagery, REC - Resting Eye Closed, WM - Working Memory, IS - Imagined Speech, MM - Motor Movement, VEP - Visually Evoked Potential, REC - Resting Eye Open, MT - Mental Tasks, SSVEP - Steady State Visually Evoked Potential, AR - Auto-Regressive, PSD - Power Spectral Density, CWT - Continuous Wavelet Transform, ERP - Event-Related Potential, LR - Linear Regression, GMM - Gaussian Mixture Model, SVM - Support Vector Machine, MD - Mahalanobis Distance, CS - Cosine Similarity, NB - Gaussian Naive Bayes, UBM - Uniform Background Model, CNN - Convolutional Neural Network, HMM - Hidden Markov Model, FC - Feature Concatenation, SF - Score Fusion, HTER - Half Total Error Rate, ACC - Accuracy, EER - Equal Error Rate.
the task-independent nature of biometric signatures in EEG by using a dataset collected using 12 different elicitation protocol with both auditory and visual stimuli.

3) Channels used for EEG biometric: In prior work, multiple techniques have been used to reduce the number of channels needed for biometric recognition using EEG. This subsampling is essential because increasing the number of channels increases the computational complexity of the biometric system. Some works sample the channels according to the task, for instance, centro-parietal channels for a mental task [11] and parietal-occipital lobe for a visual task [22]. In [16], principal component analysis (PCA) was used to reduce the number of channels. However, the most widely used technique is to select the subset of channels based on performance of the biometric system [17, 19, 20, 23–25, 27].

In [24, 25], it is shown that the performance of the electrodes from the occipital lobe is better for the eyes-closed recording owing to the alpha activity in the visual cortex. However, for the recordings with the eyes-open condition, the performance was similar to the frontal, central lobe. In [12, 13], three electrodes were chosen such that they are spatially located far from each other. The location of the electrode also accounts for the spatial variation. The dominant frequencies at the frontal lobe are generally lower in other lobes such as parietal, occipital [33]. In this work, we initially sample 9 channels from the standard 10-20 EEG system such that they are spatially apart and cover different lobes of the brain. This selection is justified by empirically studying different configurations of channels in Section V-D.

4) Features and Classifiers used for EEG Biometrics: The most commonly used features include spectral analysis using discrete Fourier transform (DFT) [11–13, 24–26] or continuous wavelet transform (CWT) [16] and autoregressive (AR) coefficients [14, 23, 25, 27]. Besides, few studies have averaged the EEG signal across multiple trials and have used the event-related-potential (ERP) as features [17, 19, 20]. Using ERP is not feasible in task-independent EEG biometrics.

In the case of AR features, a small change in the estimated coefficients can change the location of the roots in the z-domain. This can affect the frequency spectrum of the EEG signal quite significantly. The raw power spectral density (PSD) estimated on short windows has been shown to identify subjects across sessions in [28]. Hence, in this paper, raw PSD estimated over short windows are used as features for recognizing individuals.

Table I shows that longer the duration of EEG signal used, better the performance of EEG biometrics. [23] achieved 100% accuracy on 9 individuals using 60s of EEG data. [20] used ERP averaged across multiple trials to achieve 100% recognition on 20 participants. The best performance obtained for short duration of EEG, such as 5s is EER of 6.6% for 45 subjects in [27]. A short duration of 15s and a long duration of 60s are both used to evaluate systems in this paper.

Most of the prior-work discussed in Table I have used relatively simple classification/verification methods like SVM [14], Bayes classification [26], scoring techniques like L1 distance, cosine similarity, Mahalanobis distance [17, 19–21, 24, 25] or a nearest neighbor classifier [16]. However, the challenge involved in task-independent EEG person recognition is minimizing the information about the task or state of the brain and the session related information present in EEG. This problem is similar to text-independent speaker recognition. In speaker recognition, the primary assumption is that the speaker information is present in the entirety of the signal in addition to that of phoneme and channel information. Consequently, subspace techniques such as i-vector [7] and x-vector [8] were proposed for speaker recognition. These models try to encode the speaker information present in the speech signal on to a compact vector representation. i-vector is an expectation-maximization based algorithm introduced in [7] based on distributional statistics of a data model (Gaussian Mixture Model). x-vector [8] is a recent deep neural network (DNN) based state-of-the-art technique that has outperformed i-vector in the speaker recognition task. Both methods assume speaker information is present in the entirety of the speech signal and estimate various statistics across time and project them on to a lower-dimensional space. This paper proposes modifications to both the i-vector system and the x-vector system to take advantage of parallel information available across multiple EEG channels.

B. Contributions

EEG signals are typically collected from multiple sensors, which are distributed across the scalp. Different sensors capture signals from sources located in various regions of the brain. The primary contribution of the proposed system is the novel way in which the data from multiple EEG channels are processed. The proposed techniques first convert the input data to a higher dimensional space by pooling data from all the channels. In the higher dimensional space, various statistics are estimated for each channel. These statistics are then concatenated and reduced to a single vector in a subspace that enhances person-specific information. The major contributions of this article are summarized below:

- The proposed modified version of the i-vector and the x-vector systems are shown to outperform the baseline systems on two large datasets with a simple cosine similarity backend. The preliminary results of the proposed systems using SVM backend has been reported by the authors in [9] on only one of the datasets.

- The UBM-GMM system and baseline versions of the subspace systems do not model the data from different channels explicitly. The modification proposed to the subspace system model the channel information explicitly by concatenating statistics in an intermediate level of processing. However, the channel information can also be modeled by either early concatenation of features or late fusion of scores from systems built using individual channels. Early and late fusion are the popularly used techniques to model channel information in EEG biometrics (refer Table I). We show that in the context of i-vector and x-vector subspace, the proposed modification is better than the simple early and late fusion techniques to model the data from different models.

- This paper uses 12 significantly different elicitation protocols to test the task-independence by mismatched testing.
This data includes EEG collected with oddball paradigm (with beeps and visual object), steady-state visually evoked potential, motor tasks, mental tasks, and imaging tasks. Further, we combine tasks into open and closed eye conditions such that they influence the EEG significantly and test the task-independence. By examining across sessions on a challenging set of tasks/conditions, this paper builds evidence for task-independent signatures addressing the limitations of previous works (Section I-A2).

- Using a 128-channel EEG system for data collection, channel sub-sampling techniques are proposed to achieve better performance with a task-independent setup.
- The scalability of the trained subspace models to unseen persons in the training data is evaluated.

II. PROPOSED AND BASELINE SYSTEMS

A. Baseline: Universal Background Model-Gaussian Mixture Model (UBM-GMM)

The UBM-GMM system proposed in [34] is a precursor to the i-vector system. A Gaussian mixture model (GMM) is trained using data pooled from the training sessions of all the individuals. This GMM is also called a universal background model (UBM) as it is estimated using multiple subjects and acts as a reference to the person-specific models. The UBM is then converted to person-specific models by maximum-a-posteriori (MAP) adaptation on person-specific data. While testing, the score is calculated as the log of likelihood ratio between the adapted person-specific model and UBM. A detailed description of the UBM-GMM system for speaker recognition task can be found in [34].

Both EEG and speech are essentially time series. Hence many studies in EEG biometrics literature have explored UBM-GMM based techniques [11, 28, 35–37]. We use this well-studied system as a baseline system for evaluating subspace systems. In this implementation, for building the UBM-GMM system, features were pooled from all the available channels. Hence, this system does not model the channels explicitly.

B. i-vector

i-vector is a powerful speech signal representation that has led to state-of-the-art speaker recognition systems [7] and has demonstrated the ability to model person-specific information in a lower-dimensional space. The i-vector space is a subspace of the UBM space defined as follows:

\[ \tilde{M} = \tilde{m} + \mathbf{T} \tilde{w} \]  

where \( \tilde{M} \) is the supervector representing an EEG segment, \( \tilde{m} \) is the UBM supervector, \( \mathbf{T} \) is the total variability matrix that defines the subspace, and \( \tilde{w} \) is the lower dimensional i-vector. The supervector is a vector of concatenated means from the UBM or adapted models. Hence, the dimension of the supervector is \( Kd \times 1 \), where \( K \) is the number of Gaussian mixtures, and \( d \) is the dimension of the input power spectral density (PSD) feature vector. The \( \mathbf{T} \)-matrix is of dimension \( Kd \times R \), where \( R \) is the dimension of the subspace. \( R \) is an empirically determined hyper-parameter of the i-vector system.

Let,

\[ X = \{ \bar{x}_n^c \mid n = 1 \text{ to } N \text{ & } c = 1 \text{ to } C \} \]  

denote an EEG segment with \( C \) EEG sensors/channels and \( N \) feature vectors per channel. A \( K \) mixture UBM is trained using EEG segments from multiple subjects. Using the UBM, the zeroth and first order statistics required for estimating the i-vector are calculated as given in Equations 3 and 4, respectively.

\[ N_k(X) = \sum_{c=1}^{C} \sum_{n=1}^{N} P(k \mid \bar{x}_n^c, \lambda) \]  
\[ \bar{F}_k(X) = \sum_{c=1}^{C} \sum_{n=1}^{N} P(k \mid \bar{x}_n^c, \lambda)(\bar{x}_n^c - \bar{m}_k) \]

where \( \lambda \) represents UBM parameters, \( k \) denotes the mixture ID, and \( P(k \mid \bar{x}_n^c, \lambda) \) corresponds to the posterior probability of the \( k \)-th mixture component given the feature vector \( \bar{x}_n^c \). \( \bar{m}_k \) is the mean of the \( k \)-th UBM component. Given the zeroth and first-order statistics, the i-vector is estimated as follows

\[ \bar{w} = \left( I + \mathbf{T} \Sigma^{-1} \Sigma(X) \mathbf{T}^{-1} \right)^{-1} \mathbf{T} \Sigma^{-1} \bar{F}(X) \]  

where \( \bar{F}(X) \) is a supervector obtained by concatenation of \( \bar{F}_k(X) \) for \( k = 1 \ldots K \) mixtures. Hence the dimension of supervector \( \bar{F}(X) \) is \( Kd \times 1 \). \( \Sigma \) is a \( Kd \times Kd \) block diagonal matrix with \( \Sigma_n \) (covariance matrix of k-th Gaussian) as blocks along the diagonal. \( N(X) \) is also a block diagonal matrix of dimension \( Kd \times Kd \) with \( N_k(X) \) as diagonal blocks. The expectation-maximization (EM) algorithm for estimating \( \Sigma \)-matrix in the case of EEG person recognition is the same as that for speech and has been detailed in [7, 38]. This system will be referred to as “baseline-i-vector” in the rest of the paper. This system has been adopted for EEG biometrics in [39, 40]. However, this standard approach is not adequate for multichannel EEG as this system does not model channel information explicitly. To integrate information from different channels in the i-vector framework, we proposed a novel way of finding the zeroth and first-order statistics as given in Equation 6 and 7, respectively.

\[ N_{kc}(X) = \sum_{n=1}^{N} P(k \mid \bar{x}_n^c, \lambda) \]  
\[ \bar{F}_{kc}(X) = \sum_{n=1}^{N} P(k \mid \bar{x}_n^c, \lambda)(\bar{x}_n^c - \bar{m}_k) \]

In this approach, the UBM is still trained by pooling data from all the channels. However, during statistics estimation, it is done for each channel individually and then concatenated before projecting to the lower dimensional i-vector space. Hence the supervector \( \bar{F}(X) \) (in Eq 5) is obtained by concatenating \( \bar{F}_{kc}(X) \) for all \( k = 1 \ldots K \) Gaussian mixtures and \( c = 1 \ldots C \) channels. Hence the super vector dimension increases to \( KCd \times 1 \). Consequently, the dimensions of matrices \( \Sigma, N(X), \Sigma(X) \)
and \( T \) (in Eq 5) increases to \( KCd \times KCd \), \( KCd \times KCd \), and \( KCd \times R \), respectively. This system is henceforth referred to as “modified-i-vector”. Since the dimension of the supervector is high, for effective estimation of \( T \)-matrix, we use a smaller number of mixtures in the UBM compared to the baseline model.

After estimating the \( i \)-vector, a linear transform is applied using linear discriminant analysis (LDA). LDA makes the subspace more discriminative for person-specific signatures and hence improves the performance of the person recognition system. During testing, cosine similarity classifier discussed in Section II-D is used on the LDA projected \( i \)-vectors.

### C. x-vector

\( x \)-vector is a recent state-of-the-art DNN based speech representation approach aimed towards speaker recognition [8]. The \( x \)-vector system initially operates at the frame level, estimates statistics, and then the final few layers operate at the segment level. This architecture is analogous to the \( i \)-vector system with UBM acting at the frame level and the \( T \)-matrix operating at the segment level. Similar to the “modified-i-vector” in Section II-B, we remodel the \( x \)-vector system to handle information from multiple EEG sensors. \( x \)-vector proposed for speech data uses time delay neural network (TDNN) for modeling temporal context. However, upon experimenting with EEG data, we did not find long term context information to be helpful. Hence, 1-D convolution is used in place of TDNN for \( x \)-vector based EEG person recognition systems in this paper.

Figure 1 gives an overview of the \( x \)-vector architecture modified for multi-channel EEG. Spectrograms from all the channels are provided as input to this model. The model has four hidden layers. The initial two layers are single frame 1-D convolution layers that transform every feature vector of the spectrogram into a higher dimensional space. The third layer is a statistics pooling layer, which estimates the mean and variance for each channel. These statistics are concatenated and reduced to a lower-dimensional representation using the fourth hidden layer. The final output layer is a feed-forward layer with softmax activation. The number of nodes in the output layer is the total number of subjects in the training data. Similar to [8], cross-entropy error is used to train the network using Adam optimization [41]. After training, the output of the fourth hidden layer is considered as a subspace representation for the EEG segment, also referred to as \( x \)-vector. This way of estimating \( x \)-vectors will be henceforth referred to as “modified-\( x \)-vector” system.

The \( x \)-vector system with a single statistics pooling across all channels in the third hidden layer is identical to \( x \)-vector proposed for speech and will be used as a baseline. Similar to the “baseline-\( i \)-vector”, this system does not take any explicit information about the channels and will be referred to as “baseline-\( x \)-vector” system. Testing is performed using a simple cosine similarity classifier after subjecting the \( x \)-vectors to LDA, as discussed in Section II-D.

### D. Back-end

After estimating the \( i \)-vectors or \( x \)-vectors, any classifier can be used for implementing person recognition. Since the focus of this paper is on the subspace technique, we use a simple cosine similarity based recognition system as the backend.

Let \( \bar{w}_i \) be the subspace vector obtained by projecting all the available training data of person \( i \) and \( \bar{w}_{test} \) be the subspace vector under test. The cosine similarity score for \( \bar{w}_{test} \) belonging to a person \( i \) is calculated as given in Equation 8.

\[
S_i = \frac{\bar{w}_{i}^T \bar{w}_{test}}{\|\bar{w}_{i}\| \|\bar{w}_{test}\|}
\]

(8)

In this work, all the systems are evaluated using accuracy and equal error rate (EER). For calculating accuracy, the person with maximum score \( S_i \) is chosen as the final class label. For computing EER, a threshold on the score \( S_i \) is used to determine if the EEG segment is from a target or non-target person.

### III. DATASETS

#### A. Dataset 1: 128-Channel Multi-Task EEG Dataset

This dataset was collected from 30 subjects performing multiple tasks. Multiple tasks/elicitation protocols were designed with both open and closed eye conditions to collect this data. Table II gives a summary of these tasks and protocols. It is to be noted that all the 30 subjects did not perform all the 12 tasks mentioned in Table II. Form each subject, EEG data was collected for at least 2 sessions and at most 5 sessions. During each session, at most of 4 tasks from Table II were performed.

This dataset was collected by the authors in a laboratory setting using a 128-channel dense-array EEG system manufactured by Electrical Geodesics, Inc (EGI) [42]. The Ethics Committee of the Indian Institute of Technology Madras approved this study. All the subjects were informed about the aim and scope of the experiment, and written consent was obtained to collect the data. The EEG data were recorded at a sampling rate of 250Hz with the central electrode \( Cz \) as the reference electrode. After collecting the dataset, the artifacts present
TABLE II: Data collection protocols for dataset 1.

| S. No. | Experiment Name                  | Brief Description of Experiment                                                                 | No. of Participants | Total Duration (in Minutes) |
|--------|---------------------------------|---------------------------------------------------------------------------------------------------|---------------------|----------------------------|
| 1      | Odd Ball Classic                | Participants were presented with frequent non-target stimuli and infrequent target stimuli. The target and non-target stimuli consist of audio beeps differing in frequency or duration. | 13                  | 5 hr 2 min                 |
| 2      | Odd Ball Stereo                 | Similar to S.No 1. The target and non-target stimuli consist of audio beeps played in left and right ear. | 12                  | 1 hr 55 min                |
| 3      | Imagining Binary Answers        | A set of binary questions were presented to participants. They were asked to first imagine the answer and then respond with a mouse click. | 7                   | 3 hr 21 min                |
| 4      | Semantically Opposite Words     | Semantically opposite words such as yes and no were played to the subject over multiple trials. Subject was instructed to respond with left and right mouse clicks depending on the semantics of the word being played | 4                   | 1 hr 36 min                |
| 5      | Familiar and Unfamiliar Words   | The subjects were presented with common words and uncommon words. They were expected to respond with a mouse click on hearing a familiar word. | 6                   | 1 hr 50 min                |
| 6      | Proper and Improper Sentences   | Regular and ill-formed sentences were played to subject. The subject was required to respond with mouse click on hearing ill-formed sentences. | 8                   | 1 hr 52 min                |
| 7      | Motor and Mental Imaginary      | Participants were asked to imagine motor-movements such as left and right fist rotation. For mental imaginary task, they were asked to count numbers in reverse. | 6                   | 3 hr 13 min                |
| 8      | Passive Audio                   | Participants were passively listening to a variety of audio stimuli such as music, sentences, stories, and sounds that trigger attention (for example sound of sirens). | 17                  | 3 hr 33 min                |
| 9      | Passive Audio Stereo            | Similar to S.No. 7. The auditory stimuli were always played through only one ear (either left/right) at time using headphones. | 11                  | 2 hr 46 min                |

**Experiments conducted with Open Eye Condition**

| S. No. | Experiment Name                  | Brief Description of Experiment                                                                 | No. of Participants | Total Duration (in Minutes) |
|--------|---------------------------------|---------------------------------------------------------------------------------------------------|---------------------|----------------------------|
| 10     | Odd Ball Visual                 | Similar to S.No 1. The target and non-target stimuli consist of visual objects varying in shape and color. | 6                   | 33 min                     |
| 11     | Steady State Visually Evoked Potential | Visual objects flickering at different frequencies were displayed to participants. At the end of each trial, a question about the shape or color of the object was asked. | 12                  | 3 hr 13 min                |
| 12     | Passive Audio-Visual            | Audio-visual clips were played to the participants. At the end of each clip, a question was asked based on the stimuli. | 12                  | 3 hr 2 min                 |

Total number of subjects: 30
Total duration of the dataset: 31 hours

**Experiments conducted with Closed Eye Condition**

| S. No. | Experiment Name                  | Brief Description of Experiment                                                                 | No. of Participants | Total Duration (in Minutes) |
|--------|---------------------------------|---------------------------------------------------------------------------------------------------|---------------------|----------------------------|
| 1      | Odd Ball Classic                | Participants were presented with frequent non-target stimuli and infrequent target stimuli. The target and non-target stimuli consist of audio beeps differing in frequency or duration. | 13                  | 5 hr 2 min                 |
| 2      | Odd Ball Stereo                 | Similar to S.No 1. The target and non-target stimuli consist of audio beeps played in left and right ear. | 12                  | 1 hr 55 min                |
| 3      | Imagining Binary Answers        | A set of binary questions were presented to participants. They were asked to first imagine the answer and then respond with a mouse click. | 7                   | 3 hr 21 min                |
| 4      | Semantically Opposite Words     | Semantically opposite words such as yes and no were played to the subject over multiple trials. Subject was instructed to respond with left and right mouse clicks depending on the semantics of the word being played | 4                   | 1 hr 36 min                |
| 5      | Familiar and Unfamiliar Words   | The subjects were presented with common words and uncommon words. They were expected to respond with a mouse click on hearing a familiar word. | 6                   | 1 hr 50 min                |
| 6      | Proper and Improper Sentences   | Regular and ill-formed sentences were played to subject. The subject was required to respond with mouse click on hearing ill-formed sentences. | 8                   | 1 hr 52 min                |
| 7      | Motor and Mental Imaginary      | Participants were asked to imagine motor-movements such as left and right fist rotation. For mental imaginary task, they were asked to count numbers in reverse. | 6                   | 3 hr 13 min                |
| 8      | Passive Audio                   | Participants were passively listening to a variety of audio stimuli such as music, sentences, stories, and sounds that trigger attention (for example sound of sirens). | 17                  | 3 hr 33 min                |
| 9      | Passive Audio Stereo            | Similar to S.No. 7. The auditory stimuli were always played through only one ear (either left/right) at time using headphones. | 11                  | 2 hr 46 min                |

Total number of subjects with closed eye recordings: 30
Total Number of subjects with open eye recordings : 14

Total number of subjects with both open and closed eye recordings on all sessions: 10

were removed using [43], and bad channels were replaced by spherical spline interpolation [44] (plugins available with EEG lab toolbox [45]). The total duration of this dataset is about 31 hours, with 3 sessions per person on average. Further statistics on number of sessions per individual, number of EEG segments and intersession intervals between train and test are given in Table III. This dataset has been made publicly available at [https://www.iitm.ac.in/donlab/cbr/eeg_person_id_dataset/](https://www.iitm.ac.in/donlab/cbr/eeg_person_id_dataset/).

**B. Dataset 2: Temple University Clinical EEG Dataset**

This dataset is a subset of the Temple University hospital EEG data corpus (TUH-EEG) [10]. TUH-EEG corpus is a massive dataset containing over 20,000 EEG recordings collected from about 14,000 clinical patients. We preprocess this dataset for evaluating the proposed systems, as described below.

The dataset has data from 7424 patients with average EEG as reference, and 6770 patients with linked-ears data as reference. Out of this, only 2152 and 1341 patients have multi-session data collected using average and linked-ears as reference, respectively. Since this dataset was collected from clinical patients, some of the EEGs were recorded with abnormal pathological conditions such as seizures. Since these abnormalities can affect subject recognition performance, the recordings that were annotated to have abnormal EEG were removed1. After removing the abnormal recordings, the dataset contained 1033 and 155 patients with at least two-sessions with average and linked-ears reference, respectively. For further analysis, only the 1033 subjects recorded with average reference were chosen owing to the higher number of subjects. To improve the signal-to-noise-ratio, we adopted the methods in [43] for removal of artifacts and bad-channels (channels with flat or noisy data). In this process, if one of the nine channels considered in our experimental setup (Section IV) turns out to be bad, the corresponding EEG recording was discarded. After these preprocessing steps, the number of subjects with at least two sessions reduced to 920 subjects with 2889 sessions. Further detailed statistics on number of sessions per individual, number of EEG segments and intersession intervals between train and test are given in Table III.

Clinical tasks such as hyperventilation, photic simulations, and sleep and wakefulness EEG were used to collect this data. Since these data were collected for clinical purposes, the elicitation protocol has not been standardized across acquisitions. Also, the dataset does not have any annotations regarding the tasks performed. The set of tasks and the clinical setting makes this dataset distinct from dataset 1. Given the clinical nature, this dataset is used only to show the scalability of proposed approaches over the baseline on a diverse dataset with large number of subjects.

**IV. GENERAL EXPERIMENTAL SETUP**

**A. Channels**

Dataset 1 was collected using a 128-channel EEG system, whereas dataset 2 was variably collected using 24 to 36 sensors. To make a common analysis, we choose 9 electrodes,
namely, Fz, F7, F8, C3, C4, P7, P8, O1, and O2 of the standard 10–20 system. A diagrammatic representation of these 9 channels with all the 128 channels as the background is shown in Figure 2. These 9 electrodes are chosen such that they cover different regions of the brain, namely, the Frontal, Central, Parietal, and Occipital lobe. This kind of selection covering the entire scalp is essential because different stimuli/tasks elicit different regions of the brain. In Section V-D, we further analyze the effect of different sets of sensors for good performance. This analysis is essential to build practical biometric systems using EEG.

B. Features

Power spectral density (PSD) spectrogram is used as the feature. PSD spectrograms are computed in the range of 3Hz–30Hz for every channel with a window size of 360ms and no overlap. This configuration of PSD features for EEG person recognition was fine-tuned using the UBM-GMM system in [9, 28].

C. Setup for task-independent person recognition

In dataset 1, the EEG signals obtained from various experiments (given in Table II) are divided into segments of 15 seconds length. This segmentation is irrespective of the experimental protocol such as, whether the person is in the resting state or watching/listening to a stimulus/instruction or doing a task. Hence, recognizing individuals from these segments are task independent. We use the same uniform segmentation for dataset 2, which was collected for clinical purposes, unlike in a laboratory with formal control of data collection conditions.

V. Experiments and Results

A. Performance of proposed systems vs. the baseline systems

The UBM-GMM system, which does not use any subspace representation, is used as the baseline system to compare the performance of i-vector and z-vector systems. As described in Section IV, 60% of sessions from each person was used to train all the systems. It is to be noted that all the systems discussed in this paper support variable lengths of EEG segments. Therefore, for the purpose of evaluation, we further divided the test data into EEG segments of duration 15s and 60s.

| Systems       | Accuracy 15s | EER 15s | Accuracy 60s | EER 60s |
|---------------|--------------|---------|---------------|---------|
| UBM-GMM       | 71.2         | 10.9    | 70.4          | 8.47    |
| baseline-i-vector | 70.5       | 10.6    | 84.5          | 6.51    |
| baseline-x-vector | 67.1       | 11.3    | 74.7          | 8.53    |
| modified-i-vector | 85.1       | 5.81    | 93.0          | 2.84    |
| modified-z-vector | 76.8       | 8.16    | 84.0          | 5.84    |
| Dataset 2     |              |         |               |         |
| UBM-GMM       | 6.02         | 44.0    | 5.5           | 45.8    |
| baseline-i-vector | 9.02       | 20.8    | 16.4          | 22.2    |
| baseline-x-vector | 3.64       | 32.8    | 5.01          | 29.7    |
| modified-i-vector | 30.0       | 16.5    | 42.8          | 13.3    |
| modified-z-vector | 27.2       | 14.0    | 36.4          | 16.6    |
TABLE V: Performance comparison of different ways of modeling EEG channels in i-vector framework

| Systems      | Channel Information | Accuracy | EER  |
|--------------|---------------------|---------|------|
|              |                     | 15s     | 60s  |
|              |                     | 15s     | 60s  |
| Dataset 1    |                     |         |      |
| baseline-i-vector |                   | 70.5    | 84.5 |
| Concatenation |                     | 10.6    | 6.51 |
| Score-Fusion |                     |         |      |
| Dataset 2    |                     |         |      |
| baseline-i-vector |                   | 9.02    | 16.4 |
| Concatenation |                     | 26.8    | 22.2 |
| Score-Fusion |                     |         |      |
| modified-i-vector |                  | 85.1    | 93.0 |
| Concatenation |                     | 5.81    | 2.84 |

60s. Results of all the baseline and modified systems discussed in Section II are compared in Table IV for both datasets 1 and 2. All hyper-parameters associated with different systems were fine tuned using validation data.

The modified-i-vector system has most consistently given the best performance for both the datasets 1 and 2, followed by the modified-x-vector system. With 15s EEG segments, the modified-i-vector system achieves an EER of 5.81% and an accuracy of 85.1% on 30 subjects collected using multiple elicitation protocols. Further, as the duration of the EEG segment is increased to 60s, the EER improves to 2.84%. These results are comparable to best-performing systems in the literature (Table I). The modified-i-vector with an EER of 16.5% is the best performing system on dataset 2 containing 920 subjects. On observing manually, it was found that EEG recordings from many sessions in dataset 2 were still noisy and abnormal. In the supplementary material, we show that combining the modified-i-vector and the modified-x-vector representations give even better performance.

B. Proposed modifications for the i-vector and the x-vector systems vs. naive early and late fusion techniques on baseline models

In Section V-A, it was observed that explicit modeling of the EEG sensors by the proposed approaches gave a significant improvement in performance. In the literature, information from multiple channels is handled by either concatenating the input feature vector or by performing voting/score fusion on channel-specific models (Table I). In this section, we explicitly model the EEG channels in baseline versions of the i-vector and the x-vector system by feature concatenation and score fusion. Table V compares performance of proposed i-vector systems with the early and late fusion version. Table VI presents a similar comparison for x-vector based systems. Hyper-parameters for all the systems reported in Tables V and VI were fine tuned using validation data.

From Tables V and VI, it can be seen that modified versions of the i-vector and the x-vector systems have outperformed the baseline versions that use explicit channel information through early concatenation or late fusion. In the case of i-vector, feature concatenation gives a better result than score fusion. Owing to discriminative training, the x-vector systems trained on individual channels are better than i-vector trained on a single channel. Hence the x-vector model gives a better score fusion result compared to the baseline and concatenated approach. However, for both the x-vector and i-vector system, the proposed method is observed to give the best performance by concatenating statistics from various channels at an intermediate level of processing.

C. Testing task-independent person-specific signatures in EEG

In this experiment, the task-independent nature of the EEG biometric-signatures are tested in two steps as given below:

1) The subspace embeddings are tested for their ability to normalize the variance across tasks seen during subspace training.
2) The same is tested for tasks unseen during training.

First, the modified version of i-vector and the x-vector systems trained in Section V-A using all the tasks (Table II) is used. Keeping EEG data of a particular task in Table II for testing, the reference subspace vector is computed using other tasks performed on different sessions. Hence this experiment tests the task-independent nature of the vector embedding for a known task during subspace training. Later, the subspace (i-vector and the x-vector) is trained again by leaving out a task from Table II and using them only for testing. The results of both these studies of testing left-out tasks are given in Table VII.

From the results in Table VII, it can be seen that the results are slightly different for different tasks. However, for all tasks, the results are significantly high (accuracy ≈≥ 75% and EER ≈≤ 10% for i-vector systems). This result was obtained with a simple cosine similarity classifier using reference vectors that used no EEG data from the task and session under test. The results of this experiment show that, when trained with all tasks, the proposed approach can account for task and session related variance in the EEG data. When the task under test is also excluded from subspace training, the performance reduces slightly for most of the tasks. However, the results are still high, showing that the proposed approach can extract person-specific signatures even for tasks and sessions not seen during training.
To further test the task-independence of proposed systems, the tasks in Table II were combined into data collected with the open and closed-eye conditions. It is well-known that the activation and inactivation of the visual cortex by open/closed eye conditions have a significant impact on the brain activation patterns [47]. Using EEG, the open/closed eye state can be classified with ≈97% accuracy [48]. We again repeat the previous experiment but only using two conditions (the open and closed-eye conditions) rather than all tasks from Table II. Out of all the 30 subjects from dataset 1 only 10 subjects have recordings with both open and closed-eye conditions during all the sessions. Only the data from those 10 subjects are used in this experiment. First, we use the subspace trained using all the tasks from Section V-A and extract the reference vector without using the data from the condition used for testing. Later, we repeat the experiment using a subspace that was trained without the condition reserved for testing. The result of these experiments testing task-independence across the open and closed-eye conditions are given in Table VIII.

The results in Table VIII show that even when the subspace representation obtained from open/closed-eye condition is tested against the reference vector formed using the opposite condition, the simple cosine similarity measure is able to recognize the individuals with an EER ≈ 7%. This suggests that the proposed subspace techniques have the ability to model the variations across major changes in the underlying circuit, such as activation and inactivation of the visual cortex. When the subspace trained without the test condition, the performance degrades. Especially when eye closed condition is retained only for test, the EER reduces to ≈ 30%. Although this result is significantly higher than chance, it clearly shows that the proposed subspace may not scale if the underlying circuit generating the EEG changes significantly. Further, the results also show that, if the same variability in tasks are seen in the training data, the proposed approach is able to extract task-independent features.

D. Channels needed for effective estimation on the subspace signatures

All the results reported in other sections of this paper has used only the 9 channels highlighted in Figure 2. This section explores various spatial subsampling methods to analyze the number of channels required for EEG person recognition using all 30 subjects of dataset 1 collected using 128 channels.

Given a particular number of sensors, we first explore different possible ways to sample them from the available 128 channels. Accordingly, in Figure 3, nine channels are sampled locally from different regions of the brain, namely, Frontal, Central, Parietal, Temporal, and Occipital lobes. In addition, three combinations of sensors are chosen such that they cover all the regions equally. The modified-i-vector system is observed to recognize subjects with much better EERs when sensors are sampled from all the regions rather than locally from a particular region. However, among the different areas of the brain, the Central region is observed to give better recognition, followed by the Parietal region. Nevertheless, by showing consistently good EERs for three different selections, Figure 3 shows that sampling sensors from across the entire scalp yields a better result for task-independent person recognition. In Figure 4, we analyze the number of channels required for EEG biometric using channel subsets of different sizes from the available 128 sensors. The sensors for systems with 16, 32, and 64 channels are sampled by incrementing the channel numbers (given in Figure 2) by 8, 4, and 2, respectively. This selection ensured that the entire scalp was covered. The EER of the systems using a larger number of sensors are compared with the 9 channel systems used in other sections of this paper. The results for the modified systems with channels ≥ 64 are not shown because higher the number of channels, greater the data needed to train the system owing to the concatenation of statistics across channels. In Figure 4, observe that the modified-i-vector system using just 9 channels achieves an EER of 5.81%. Whereas, using all the 128 channels, the baseline-i-vector system achieves a performance of 11% EER.
These results suggest that not all channels are required for subjective identification. Nevertheless, it is important to ensure that the sensors are spatially apart and cover the entire scalp. Further, using 32 channels, the performance of the modified-i-vector is observed to increase gracefully.

To study the performance of systems with fewer than 9 EEG sensors, we also analyze systems with 6 and 4 channels in Figure 4. While selecting sensors for the 6 and 4 channel system, higher importance was given to Central and Parietal regions as they were observed to provide better recognition in Figure 3. It is interesting that the system with just 4 channels is found to give slightly better recognition than baseline systems using all the 128 channels. Adding more channels need more data to train the systems. Further, as more channels are added, this improvement in performance degrades gradually. This is a significant result as it indicates that additional channels only lead to redundant information and increased dimensionality requiring more training data.

**E. The subspace defined by modified versions of the i-vector and the x-vector systems generalizes for unseen subjects**

Training of the entire subspace for every new user to be enrolled in a biometric system can be a tedious task. When trained using many subjects, the subspace should represent biometric signatures independent of specific individuals used during training. This hypothesis is well-established in the speaker recognition literature. The speakers used for evaluation are seldom used during subspace training. Taking this idea forward, we test the subspace system performance for subjects not seen during training.

A random 20% of the subjects from both datasets 1 and 2 was selected for evaluation in this experiment. The performance for this 20% of the random subjects is shown under two cases. In Case 1, all the subjects are used for training and evaluation is done only on the selected 20% of the subjects. In Case 2, we retain the selected 20% of the subjects for only testing. The subspace is trained with all the remaining subjects. The results for Case 1 and 2 are compared in Table IX for both datasets using EEG segments of 15 seconds duration. It can be observed that the results degrade when the subjects are not
TABLE IX: Accuracy (%) and EER (%) when subspace is trained with all subjects including the subjects under test (Case 1) vs when the subspace is trained without the subjects under test (Case 2).

| Dataset 1 | Case 1 | Case 2 | Dataset 2 | Case 1 | Case 2 |
|-----------|--------|--------|-----------|--------|--------|
|           | ACC    | EER    | ACC       | EER    | ACC    |
| Modified-\(i\)-vector | Split-1 92.3 6.85 70.3 14.9 | Split-2 95.4 4.97 90.3 13.6 | Split-1 89.4 10.3 80.5 14.5 | Split-2 92.4 7.37 80.4 14.3 |
| Modified-\(x\)-vector | Split-1 92.3 4.87 64.3 19.4 | Split-2 88.1 8.72 64.75 22.8 | Split-3 91.1 6.91 88.9 11.5 | Split-3 90.8 6.83 72.3 17.9 |
| Avg       | 90.8 6.83 72.3 17.9 | 89.4 10.3 80.5 14.5 | 91.1 6.91 88.9 11.5 | 90.8 6.83 72.3 17.9 |

The signal, we can also concatenate the spectrograms from specific signatures from the EEG signal independent of tasks. Of the proposed subspace systems over the baseline systems and EER in Table IV. This improvement in the performance of the proposed subspace systems is due to the modification that allows for the pooling of data to create a common high-dimensional space for all channels. In high-dimensional space, various statistics are estimated across time for each channel. These statistics are then concatenated across channels and reduced to a single vector in a subspace that enhances person-specific information. When this channel information is explicitly modeled by the modified version of the \(i\)-vector and the \(x\)-vector frameworks, a significant improvement in performance is observed in the respective systems. This observation is consistent with both accuracy and EER in Table IV. This improvement in the performance of the proposed subspace systems over the baseline systems demonstrates that the former can better model the person-specific signatures from the EEG signal independent of tasks.

Rather than concatenating the statistics computed across the signal, we can also concatenate the spectrograms from different channels to form a huge input spectrogram. The disadvantage of this early concatenation approach over the proposed modification is that it reduces the training data for initial frame level processing and also increases the number of parameters. An alternative approach is score fusion, wherein a subspace system is trained for each channel individually, and the final score is averaged across channels. The downside with the score fusion system is that the model does not take statistical advantage over the parallel information from different sensors. Moreover, since the score fusion system defines a subspace for each channel, this also increases the time complexity.

All these different techniques of incorporating the data from various channels have been evaluated against the proposed technique in Section V-B. From the results, it can be clearly observed that the proposed modification has better performance than the early concatenation and late fusion techniques.

The proposed approach has a drawback that the same set of channels should be present during training and testing as compared to the baseline subspace systems. In addition, owing to the concatenation of statistics, these models require more data when the number of channels is increased. Nevertheless, Section V-D shows that the modified-\(i\)-vector system with just 4 channels outperforms the baseline-\(i\)-vector model using all the 128 channels. This result is significant because using a larger number of channels is not feasible for building real-time biometric systems using EEG. In addition, we also show that sampling electrodes from across the entire scalp gives a better EER than choosing the sensors locally from a particular region of the brain.

Section V-E shows that the subspaces of the modified-\(i\)-vector and the modified-\(x\)-vector scales to the subjects that were not used to train the subspace. Therefore, the subspace technique discussed in this paper is shown to generalize across individuals. This result eliminates the need to retrain the subspace when new users are enrolled.

VI. DISCUSSION

A. Significance of the proposed modification for the subspace system

The most important observation from Section V-A is that the proposed approaches, namely, modified-\(i\)-vector and modified-\(x\)-vector system, give a significant improvement in performance over the baseline systems. Both the baseline-\(i\)-vector and the baseline-\(x\)-vector systems (Section II) assume that the biometric information is present in the entire EEG signal. Hence these systems do not perform any sequence modeling. These subspace systems accumulate statistics across time in a higher dimension space and then project to a lower-dimensional space such that the biometric information is preserved. The UBM-GMM system and baseline versions of the subspace systems do not model the data from different sensors explicitly.

The proposed modification suggests the pooling of data to create a common high-dimensional space for all channels. In the high-dimensional space, various statistics are estimated across time for each channel. These statistics are then concatenated across channels and reduced to a single vector in a subspace that enhances person-specific information. When this channel information is explicitly modeled by the modified version of the \(i\)-vector and the \(x\)-vector frameworks, a significant improvement in performance is observed in the respective systems. This observation is consistent with both accuracy and EER in Table IV. This improvement in the performance of the proposed subspace systems over the baseline systems demonstrates that the former can better model the person-specific signatures from the EEG signal independent of tasks.

Rather than concatenating the statistics computed across the signal, we can also concatenate the spectrograms from different channels to form a huge input spectrogram. The disadvantage of this early concatenation approach over the proposed modification is that it reduces the training data for initial frame level processing and also increases the number of parameters.

An alternative approach is score fusion, wherein a subspace system is trained for each channel individually, and the final score is averaged across channels. The downside with the score fusion system is that the model does not take statistical advantage over the parallel information from different sensors. Moreover, since the score fusion system defines a subspace for each channel, this also increases the time complexity.

All these different techniques of incorporating the data from various channels have been evaluated against the proposed technique in Section V-B. From the results, it can be clearly observed that the proposed modification has better performance than the early concatenation and late fusion techniques.

The proposed approach has a drawback that the same set of channels should be present during training and testing as compared to the baseline subspace systems. In addition, owing to the concatenation of statistics, these models require more data when the number of channels is increased. Nevertheless, Section V-D shows that the modified-\(i\)-vector system with just 4 channels outperforms the baseline-\(i\)-vector model using all the 128 channels. This result is significant because using a larger number of channels is not feasible for building real-time biometric systems using EEG. In addition, we also show that sampling electrodes from across the entire scalp gives a better EER than choosing the sensors locally from a particular region of the brain.

Section V-E shows that the subspaces of the modified-\(i\)-vector and the modified-\(x\)-vector scales to the subjects that were not used to train the subspace. Therefore, the subspace technique discussed in this paper is shown to generalize across individuals. This result eliminates the need to retrain the subspace when new users are enrolled.

B. Task-independent EEG biometrics

The person-specific signatures in EEG were tested for task-independence in Section V-C using mismatched tasks for train and test. First, the task-independence was tested with subspace trained using all the data and with a mismatched task/condition for the cosine similarity backend. Later, the subspace system was also retrained without using the task/condition reserved for the test. For both these cases, in Tables VII and VIII, all the results are significantly above chance accuracy. This result shows that biometric information is present in EEG irrespective of the task and condition. From Tables VII and VIII, it is evident that the subspace model generalizes better when the task/condition under test is also used for training the subspace.

In Table VIII, the worst performance of EER \(\approx 30\%\) was obtained when the subspace was trained using the closed eye condition and tested on the open eye condition. To further analyze this, the modified-\(i\)-vector space was reduced to a two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) [49]. t-SNE is a non-linear dimension...
reduction technique such that it preserves the distance between the two points in the original space. t-SNE is applied to reduce the EEG signatures in the i-vector space to a visualizable 2D space. In Figure 5, the t-SNE plots were made for all four conditions given in Table VIII.

In Figure 5, A and B, it can be seen that when the subspace is trained with all the data, the EEG segments from different conditions and sessions are located close to each other. The symbol “+” denotes the training condition from which the reference vector is formed, and “o” indicates the test condition. In Figure 5, A and B, subject 6 highlighted by a black circle, has two distinct clusters made of data from the train (“+”) and test (“o”) conditions close to each other. This shows the ability of the subspace to normalize the variance related to tasks and sessions. In Figure 5, C and D, the distance between train and test clusters increases when data from test conditions are not used to train the subspace. It is well known that alpha oscillations are present in the occipital lobe when the eyes are closed. The subspace trained only using open eye condition is unaware of the dominant alpha oscillations in closed eye condition. These alpha oscillations could be one of the reasons for poor task normalization in Figure 5. D (subject 6, for example). Hence, the subspace model cannot be expected to scale for a significant change in the underlying EEG (such as alpha wave during closed eye or change in EEG due to a brain injury). However, when these conditions are included while training the subspace, the model is able to generalise even across such conditions.

C. Limitations

Both the datasets 1 and 2 used in this paper have some limitations. The limitation of dataset 1 is that it does not have data from all the subjects performing all the tasks. Hence, it is not possible to analyze the performance from Table VII across tasks and compare them. However, dataset 1 has been collected using a wide range of tasks, making it a suitable candidate for task-independent EEG analysis. Dataset 2 has a limitation that it is obtained from clinical patients. Various traits of clinical conditions and the use of non-standardized tasks may contribute to the performance. Therefore, this dataset may have a bias towards the clinical conditions. However, in contrast to dataset 1, dataset 2 has larger number of subjects with longer intersession interval.

D. Future research direction

With different elicitation protocols being a primary focus in the literature, all the results discussed in this paper question the need for specific (constrained) elicitation protocols for studying biometric signatures. These results show that there can be significant person-specific signatures in any EEG being collected. These person-specific signatures can negatively affect generalization across individuals when EEG is being used for building task rich Brain-Computer Interfaces (BCI). In the speech processing literature, while training models for speech recognition, speaker information is suppressed by various speaker normalization techniques. Since biometric information is always present in EEG, similar methods need to be developed to scale BCIs across individuals. The subspace systems proposed in this paper gives a single vector representation of biometric information present in the signal. While building BCIs, the subject-specific vectors can also be used as features for normalizing variance across subjects. In addition, the results in this paper can also be used as a baseline for studying task-dependent EEG biometrics.

VII. Conclusion

The paper proposes adaptations of state-of-the-art task-independent speaker recognition techniques to EEG based biometric authentication systems. The proposed methods are shown to perform better than the baseline systems by testing on two large datasets across tasks and sessions. The proposed approach is shown to reliably encode the person-specific signatures into a single vector using just four channels and a simple cosine similarity scoring. This subspace is then used to build evidence for the presence of task-independent person-specific signatures in EEG using different tasks/conditions for training and testing at various levels. The results discussed in this paper question the use of elicitation protocols for EEG biometrics and suggest the task-independent setting to be used as a baseline for studying task-dependent EEG biometrics.
ACKNOWLEDGMENT
We thank the Center for Computational Brain Research (CCBR), IIT Madras for enabling the collaboration with Sur Lab, Massachusetts Institute of Technology and and DST VAJRA for enabling collaboration with University of Southern California.
EVIDENCE OF TASK-INDEPENDENT PERSON-SPECIFIC SIGNATURES IN EEG USING SUBSPACE TECHNIQUES: SUPPLEMENTARY MATERIAL

S1. Extensive analysis of Temple University dataset

After all preprocessing steps mentioned in Section III-B, the Temple University dataset contained 920 patients data. In the main article, all 920 subjects were used to compare various models. In this supplementary section, we use 500 and 100 best performing subjects using baseline UBM-GMM subjects. The results are given in Table S1. Detailed statistics of these datasets are given in Table S2.

Similar to the results given in the main article, the proposed systems have achieved a significant improvement over the baseline systems although the baseline system was used to select the subjects.

| Systems                        | Accuracy | EER |
|--------------------------------|----------|-----|
|                                |          | 15s | 60s | 15s | 60s |
| **500 Subjects**               |          |     |     |     |     |
| UBM-GMM                        | 10.1     | 11.4| 38.8| 42.4|
| modified-i-vector              | 33.1     | 47.1| 16.9| 13.9|
| modified-x-vector              | 20.0     | 27.1| 22.4| 19.5|
| **100 Subjects**               |          |     |     |     |     |
| UBM-GMM                        | 6.02     | 5.5 | 44.0| 45.8|
| modified-i-vector              | 27.6     | 42.0| 16.4| 13.5|
| modified-x-vector              | 13.9     | 20.5| 23.1| 19.6|

S2. Combining modified-i-vector and modified-x-vector representation for improved performance

In this section, we explore the advantage of combining both the i-vector and x-vector representation. The subspace representation of both the modified i-vector and x-vector systems are trained individually. This subspace representation is then concatenated to form a combined vector. This combined subspace representation is then used to recognize the individuals using cosine similarity back-end. The result of this combined model for datasets 1 and 2 are given in Table S3.

The concatenation of i-vector and x-vector is observed to give a significant improvement in performance, especially for dataset 2 (Table S3). The combined system achieves a relative improvement of 28% and 42% for EEG segments of length 15s and 60s in dataset 2, respectively. The result shows that the modified i-vector and x-vector subspace capture complementary information. Hence, the combined representation system can generalize better for the test data.

| Systems                                      | Accuracy | EER |
|----------------------------------------------|----------|-----|
|                                              | 15s      | 60s | 15s | 60s |
| **Dataset 1**                                |          |     |     |     |
| modified-i-vector                            | 85.1     | 93.0| 5.81| 2.84|
| modified-x-vector                            | 76.8     | 84.0| 8.16| 5.84|
| modified-i-vector + modified-x-vector        | 86.4     | 93.3| 5.02| 2.59|
| **Dataset 2**                                |          |     |     |     |
| modified-i-vector                            | 30.0     | 42.8| 16.5| 13.3|
| modified-x-vector                            | 27.2     | 36.4| 16.6| 14.0|
| modified-i-vector + modified-x-vector        | 35.9     | 47.0| 14.2| 11.7|
TABLE S4: Result of fixing the number of days between training to testing to 1-day and 1-week for 6 subjects in Dataset 1

| System        | Segment length | Intersession testing interval | 1 day | > 1 day and < 1 week |
|---------------|----------------|-------------------------------|-------|----------------------|
|               |                | Acc | EER | Acc | EER |
| modified-i-vector | 15s            | 82.3 | 6.22 | 81.8 | 6.83 |
| modified-x-vector | 15s            | 81  | 4.78 | 78.7 | 5.9  |

TABLE S5: Result of fixing the number of days between training to testing to 1-week and 1-month for 3 subjects in Dataset 1

| System        | Segment length | Intersession testing interval | < 1 week | > 1 week and < 1 month |
|---------------|----------------|-------------------------------|----------|------------------------|
|               |                | Acc | EER | Acc | EER |
| modified-i-vector | 15s            | 89.5 | 3.94 | 97 | 2.39 |
| modified-x-vector | 15s            | 82.4 | 6.56 | 84.8 | 2.76 |

S3. ANALYSIS ON TEMPORAL INTERVAL

In this experiment, the influence of the temporal distance between training and testing is shown at various intervals for the same set of subjects. Table S4, shows the result for 6 subjects with a temporal distance of 1 day and 1 week. Table S5, shows the same for 3 subjects with a temporal distance of 1 week and 1 month.

Table S4 shows that the results are almost stable between 1-day and 1-week testing with a slight degradation in performance. It is interesting to see that in Table S5, the results are slightly better for a longer temporal distance of 1 month as compared to 1 week.

S4. ANALYSIS OF WHY CHANNEL MODELING IS IMPORTANT FOR TASK-INDEPENDENT EEG BIOMETRICS

From the result of Sections V-A and V-B, it can be seen that the systems which model the information from different channel explicitly give better results than the baseline systems that pools data from all the channel. To further analyze why this channel modeling is important, in this section, we study the contribution of each sensor to the final subspace representation as follows:

Step 1: The subspace representation of modified-i-vector and modified-x-vector systems used in Section V-A are first found using data from all the 9 channels (given in Section IV).

Step 2: Channel specific subspace representation is found using the data from only that particular EEG sensor. Cosine similarity is calculated between this channel-specific subspace vector and the subspace vector obtained in Step 1. This step is repeated for all 9 channels, and the cosine similarity values obtained are plotted as a topographic plot using the locations of each channel. Examples of topographic plots so obtained from EEG segments of 60s are given in Figure S1 for the modified-i-vector system and in Figure S2 for the modified-x-vector system, respectively.

From the examples given in Figures S1 and S2, it can be seen that channels contribute differently to each subject. Further, when open or closed eye conditions influence the EEG, the channel contributions are observed to change. Especially for the closed eye condition, the channels form the occipital lobe have higher cosine similarity with the subspace vector obtained using all the channels. This experiment validates the importance of proposed systems and confirms that data from different channels are essential to model variations amongst individuals and tasks.
| Subject | Closed Eyes | Open Eyes |
|---------|-------------|-----------|
|         | Segment 1   | Segment 2 | Segment 3 | Segment 1 | Segment 2 | Segment 3 |
| 1       | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) |
| 2       | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) |
| 3       | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) |

Fig. S1: Visualization of contribution from different channel in the modified-\(i\)-vector subspace. Red represents high contribution, and blue represents low contribution.

| Subject | Closed Eyes | Open Eyes |
|---------|-------------|-----------|
|         | Segment 1   | Segment 2 | Segment 3 | Segment 1 | Segment 2 | Segment 3 |
| 1       | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) |
| 2       | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) |
| 3       | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) |

Fig. S2: Visualization of contribution from different channel in the modified-\(x\)-vector subspace. Red represents high contribution, and blue represents low contribution.