Feature Switching in the i-vector Framework for Speaker Verification

Asha T [1], Saranya M S [1], Karthik Pandia D S [1], Srikanth Madikeri [2], Hema A Murthy [1]

[1] IIT Madras, India
[2] EPFL, Switzerland

asha@cse.iitm.ac.in, gnusaran@gmail.com, pandia@cse.iitm.ac.in
srikanth.madikeri@idiap.ch,hema@cse.iitm.ac.in

Abstract

Feature fusion is a paradigm that has found success in a number of speech related tasks. The primary objective in applying fusion is to leverage the complementary information present in the features. Conventionally, either early or late fusion is employed. Early fusion leads to large dimensional feature vectors. Further, the range of feature values for different streams require appropriate normalisation. Late fusion is carried out at score level, where the contribution from each type of feature is determined from the set of weights used. Feature switching is yet another paradigm that attempts to capture the diversity in the feature types used. Feature switching gains significance particularly in the context of speaker verification, where the feature type that best discriminates a speaker is used to verify the claims corresponding to that speaker. Earlier, feature switching was attempted in the conventional UBM-GMM framework. In this paper, the idea is extended to the Total Variability Space (TVS) framework. Two different feature types namely Modified Group Delay (MGD) and Mel-Frequency Cepstral Coefficients (MFCC) are explored in the proposed framework. Results are presented on NIST 2010 male database for the speaker verification task.

Index Terms: speaker verification, shared nearest neighbour, feature-switching

1. Introduction

Fusion of multiple systems is common for performance improvement in the state-of-the-art speaker verification systems. Improved performance was reported using weight-based score fusion[1] in one of the NIST SRE 2012 submissions. The feature switching approach proposed in this paper suggests another way of combining multiple systems. Feature switching in TVS framework computes similarity between target speaker's ivector and non-target speakers' ivectors in each feature domain. The feature domain that provides minimum inter-class similarity is selected as optimal feature domain for the target speaker. The procedure involves labelled train data and various similarity measures like cosine similarity, Shared Nearest Neighbour Similarity (SNNS) using cosine similarity and negative of Euclidean distance. A novel scoring method which uses SNNS scores, which was applied earlier in other domains, is used for speaker verification in this paper. In the paper [1], the seventeen systems fused varied on the basis of both features and models used. It is of interest to explore the factors contributing to the overall performance from each system. This study can also be categorised as an analysis of two systems based on feature variation and the contribution of each system is analysed speaker-wise.

Speaker-wise analysis is carried out because of the following reasons. Although the weight-based score fusion system is simple to implement, the weights chosen are same for all classes. Essentially, this scheme associates higher weights to constituent systems that perform better across all classes of the training data. On the other hand, the scheme proposed in this paper associates entire weight-age to the system that is hypothesised to discriminate the target speaker well for each speaker. This is based on the conjecture that a particular feature may be more than adequate to verify the claim instead of having to compute fused score after extracting all feature types[2]. We consider two different features for this study, namely, the standard MFCC and MGD features. These two features are chosen primarily because there is sufficient literature that establishes their complementarity for speaker verification [2, 3]. The choice of the feature that is relevant for each speaker is obtained using the training data and development data. The result might be a single feature or a combination of features for each given speaker in systems that are made of multiple number of constituent systems. This methodology is referred to as feature switching. Although the procedure was studied earlier in the UBM-GMM framework [4], it has to be newly devised for the TVS framework(also termed as i-vector framework) due to lack of extendability of the former procedure.

Feature switching method used in GMM-UBM framework could not be used in i-vector framework, because the information theoretic measures used are based on Mutual Information and Kullback-Leibler divergence computed on the distributions that characterise the target and the rest of the speakers. Therefore, a suitable metric has to be identified for the i-vector framework where each speaker is represented by a single vector. In this paper, apart from devising a feature selection strategy in i-vector framework, an attempt is made to study different distance measures to perform the feature selection. The feature switching method proposed in this paper allows us to combine any given number of systems that are based on feature type variation in the state-of-the-art TVS framework. Moreover, its applicability extends to pattern recognition tasks in other domains as well.

In Section 2, the complementary nature of the features namely, MFCC and MGD is established through both theoretical and experimental studies and the need for using complementary features in this methodology is emphasised. Section 3 provides a comprehensive discussion of feature switching methodology in the GMM-UBM framework vs i-vector framework. Further experimental details and results are presented in Section 4 followed by conclusion in Section 5.

2. Features for Feature Switching

Feature switching will be meaningful only if features capture diversity in speaker characteristics. In this paper, we explore MFCC and MGD features which have shown to capture diverse

  14 – 18 September 2014, Singapore

characteristics across speakers, sounds in the literature[5, 6, 7]. A speaker verification experiment is carried out on NIST 2003 female database that has data for 207 speakers and the results are plotted to represent the complementarity of features. Figure 1 shows plot of score by MFCC based system vs score by MGD based system where horizontal and vertical line correspond to MGD and MFCC system thresholds respectively. In the Figure 1, scores in north-west and south-east quadrants represent the trial scores that are correctly identified by either of the systems, but not both. The scores in these regions indicate complementary results. Table 2 shows an improvement in EER, when fixed weighting based score fusion was applied for NIST 2010 male database on MFCC and MGD feature based systems with weights 0.8 and 0.2 respectively. This further highlights the scalability of complementary information even on a different database namely, the male NIST 2010 database. MFCC features are well-known in speech domain and need no elaboration. In the subsequent section, the importance of phase based features for speaker verification is illustrated. MGD is shown to be performing well for speaker verification[4], and feature extraction procedure for MGD is detailed in the paper by Murthy et al.[8].

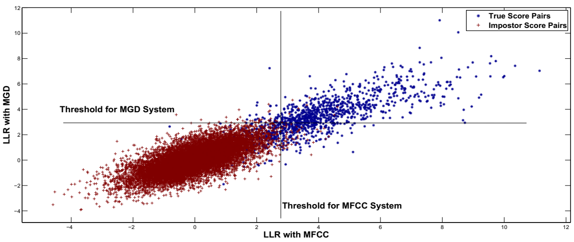

Figure 1: NIST 2003 Speaker Verifications Systems for MFCC and MGD : Plot depicting complementary results

2.1. Importance of Group Delay based Features

Magnitude spectrum based MFCC features are efficient in capturing the spectral envelope. While phase based MGD features are known for more accurate capture of the formant information, or essentially, the vocal tract shape[9]. Formant details are crucial for a reliable representation of a speaker. Figure 2 compares the capability of mel-filterbank energy contour and modified group delay spectrum in resolving formant locations (vocal-tract resonances) and vocal-tract anti-resonances. Contours in black colour are the original formants which are overlaid on MFCC (top) and MGD (bottom) spectrograms. It can be seen from the Figure 2 that MGD formants (red colour) offers a better match with the original formants compared to MFCC formants (yellow colour). Third and fourth formants are shown to be significant in distinguishing a speaker[10]. Also the study in the paper [11] suggest that first four formants are crucial in retaining sounds' characteristics and it is known that sounds' characteristics vary across speakers.

Recently, there has been increase in the use of group delay based features for a variety of pattern recognition problems. Feature fusion/concatenation of MGD and MFCC features is shown to be leading to a robust music instrument recognition system[12]. MGD features employed individually did not perform better than MFCC, while feature concatenation of both the features did. Similar results can be observed even with the speaker verification task, but using late fusion (Table 2). MGD features were applied for depression detection, which is a common psychiatric disorder, where speech by the patient is used as input[13]. Though, feature concatenation was not attempted in that paper, MGD was shown to perform comparably with MFCC features. MGD was helpful in efficient bispectrum estimation owing to its capacity to resolve frequencies better[14], and bispectrum estimation is useful for a wide variety of applications. All these applications reiterate the effectiveness of the deployment of MGD features for different tasks. Therefore, MGD features can be termed as *representative* of speaker information, because of their formant resolution capability. Also, they are *complementary* to MFCC which can be observed from the results mentioned before.

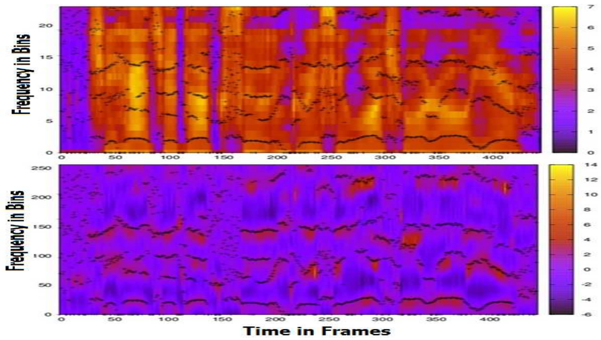

Figure 2: Comparison of Formant resolution capability of MFCC and MGD features

3. Feature Switching

System fusion aims at preserving the correctly identified uncommon results among the systems fused that are indicated in Figure 1. Score fusion is a typical solution to this problem[1]. A linear binary classifier[1] that gives a linear boundary separating target score vectors and non-target score vectors is trained on development data and applied on evaluation data. But this method might not preserve uncommon results to its entirety because of the non-linear nature of the boundary. Therefore an attempt is made to choose the feature that scores best with respect to a given speaker instead of a weighted score, with the aim of retaining all correctly identified uncommon results. So, the task of feature switching reduces to selection of the optimal feature for every target class/speaker. Such a framework based on *Mutual Information (MI)* and *Kullback - Leibler Divergence (KLD)* measure was proposed in GMM-UBM framework by Padmanabhan et al.[4]. In this paper, a method for feature switching in the i-vector framework is proposed. This method involves various distance and similarity measures namely *Euclidean Distance, Cosine Similarity, Shared Nearest Neighbour Similarity*[15].

3.1. Mutual information and Kullback - Leibler Divergence Measure

MI is a measure of dependence[16] between two random variables. Assuming that the random variable $\mathcal{X}$ takes a set of values $\mathcal{X} = \{x_1, x_2, , x_M\}$ with associated probabilities $\{p_{x_1}, p_{x_2}, , p_{x_M}\}$, and that the random variable $\mathcal{Y}$ takes a set of values $\mathcal{Y} = \{y_1, y_2, , y_N\}$ with associated probabilities $\{p_{y_1}, p_{y_2}, , p_{y_N}\}$. MI is calculated using the following formula.

$$MI(\mathcal{X}, \mathcal{Y}) = \sum_{x_i \epsilon X} \sum_{y_j \epsilon Y} p(x_i, y_j) log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (1)$$

KLD [17] is an asymmetric measure of the discrimination between any two probability distributions $p(x_i)$ and $p(x_j)$. KLD for two multivariate Gaussian distributions is given by the following formula[4]

$$D_{KL}(p(x_i)\|p(x_j)) = (1/2)[log\frac{|\mathbf{\Sigma_i}|}{|\mathbf{\Sigma_j}|} + Tr|\mathbf{\Sigma}_j^{-1}\mathbf{\Sigma}_i| - d$$
$$+ (\mu_i - \mu_j)^T {\mathbf{\Sigma}_j}^{-1}(\mu_i - \mu_j) \quad (2)$$

where $p(x_i)$ and $p(x_j)$ are given by $\mathcal{N}(\mu_i,\mathbf{\Sigma}_i)$ and $\mathcal{N}(\mu_j,\mathbf{\Sigma}_j)$ respectively.

3.2. Feature Switching in the GMM-UBM Framework

Feature selection[4] is carried out in the following way in the GMM-UBM system. KLD mentioned in Section 3.1 is defined for any two multivariate Gaussian distributions. But it is to be noted that we need to compute KLD between two Gaussian Mixture Models(GMM's) and is therefore modified (Eqn. (3)). KLD is computed between the speaker-specific model and universal background model. Therefore, KLD serves as a measure of discrimination with respect to the background model. Similarly, MI is computed between complex short-term Fourier transform and corresponding feature vector, for every frame and accumulated for entire recording. MI serves as a measure of representativeness of feature vector with respect to the Fourier transform of the original signal. For every speaker, a weighted sum of both MI and KLD is computed for every feature type in order to select the optimal feature. Feature which is more representative and discriminative is chosen as the optimal feature.

$$KLD(s, u) = \mathbf{\Sigma}_{ui}\pi_{ui}KLD(s_i, u_i) \quad (3)$$

where s is speaker specific GMM($\mathcal{N}(\hat{\pi},\mu_s,\mathbf{\Sigma})$) and u is the UBM ($\mathcal{N}(\hat{\pi},\mu_u,\mathbf{\Sigma})$), is the UBM.

Although, the improvement using feature switching in the UBM-GMM framework was not statistically significant [4] over that of joint features, the approach did show that features that are relevant for speaker verification need not be the same across all speakers. Since the results presented in [4] are encouraging, we explore the paradigm of feature switching in the i-vector framework. Since i-vector cannot be characterised by a distribution owing to the paucity of i-vectors, new measures must be designed for feature switching in the i-vector framework.

3.3. Cosine Similarity and Euclidean Distance

Cosine similarity is a similarity measure while Euclidean distance is a distance measure defined for any two arbitrary vectors as per the Eqn. 4 and 5 respectively. Assuming there are two i-vectors v_i and v_j.

$$cos(\vec{v_i}, \vec{v_j}) = \frac{\vec{v_i}.\vec{v_j}}{||\vec{v_i}||\,.\,||\vec{v_j}||} \quad (4)$$

$$euc(\vec{v_i}, \vec{v_j}) = ||\vec{v_i} - \vec{v_j}|| \quad (5)$$

3.4. Shared Nearest Neighbour Similarity

Shared Nearest Neighbour Similarity (SNNS) is a rank-based similarity measure found to be relevant for higher dimensional data[15]. It comes under the category of secondary measures whose computation depends on other primary similarity measures like cosine, Euclidean etc. Recently, this measure is also shown to combat the *hubness phenomenon and curse of dimensionality* observed while using higher dimensional data[18, 15]. Earlier research work related to the SNNS measure can be found in the domains of earth science, word clustering (text)[19]. Given n other vectors $D = \{d_1, d_2, ... d_n\}$, two sets of nearest neighbours $\mathcal{X}$ and $\mathcal{Y}$ of cardinality m are computed for each

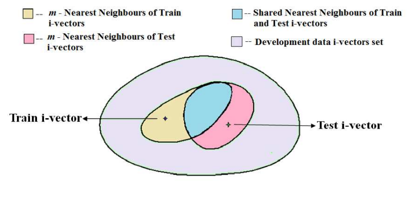

Figure 3: SNNS depiction with 2D vectors

of $\vec{v_i}$ and $\vec{v_j}$ from set D. The SNNS between the two i-vectors $\vec{v_i}$ and $\vec{v_j}$ is given by the equation6. The distance/similarity measure used to determine nearest neighbours could be any one among the Euclidean distance measure, cosine similarity measure or other similarity measures.

$$SNNS(\vec{v_i}, \vec{v_j}) = \frac{N(\mathcal{X} \cap \mathcal{Y})}{m} \quad (6)$$

3.5. Shared Nearest Neighbour Similarity Scoring with i-vectors

This scoring measure was initially experimented by the authors for the NIST i-vector challenge, 2012[20] to confirm its applicability and scrutinise its performance trends with i-vectors. A total of 12,582,004 trials are evaluated for a set of 1306 speakers. Five i-vectors for every speaker, 9634 test i-vectors and 36,572 development i-vectors are provided as part of the challenge. The i-vectors projected onto unit sphere were used in this experiment to highlight the effect of SNNS without subjecting them to any further discriminant analysis, dimensionality reduction and channel compensation techniques like Linear Discriminant Analysis (LDA), Within Class Covariance Normalisation (WCCN) and Probabilistic Linear Discriminant Analysis (PLDA). Results are shown in Table 1, and the relative gain is 16% (a drop of 0.062 in terms of minDCF) for SNNS with cosine similarity.

In the experiment using SNNS, say $\vec{v_i}$ and $\vec{v_j}$ correspond to train vector and test vector of a particular trial respectively. The development vectors set is chosen to comprise the set from which nearest m neighbours are identified i.e D, mentioned in Section 3.4. Set $\mathcal{X}$ and $\mathcal{Y}$ are determined using the chosen distance measure, and number of nearest neighbours is chosen to be m. Score is computed for every trail by applying Eqn. (6). SNNS with cosine similarity is observed to give the best results in this experiment. Cosine similarity can be considered as a meaningful measure for higher dimensional data as it captures angular similarity and, it is not affected by a single attribute change as much as in Euclidean distance. As can be observed from the nature of this distance measure, it can be used as an alternative to cosine similarity scoring without normalization, because we use cardinality of shared nearest neighbours and n (Total number of nearest neighbours) instead of using data (i-vectors) directly in computing the scores. But normalization was found to improve results further in the experiments. Generally, PLDA scoring is the last stage of current state-of-the-art speaker verification systems. It is also known to offer an alternative scoring method for cosine distance scoring without normalization[21].

3.6. Feature Switching in the i-vector Framework

It is to be noted that KLD (discriminative measure) and MI (representative measure) have to be computed between GMM's in GMM-UBM system. While in i-vector framework, all models reduce to a vector of dimension of order 10^2. The possibility

Table 1: Comparison of Results with Different Scoring Measures

Distance measure	Performance (minDCF)
Cosine Similarity(Baseline)	0. 386
SNNS using Euclidean Distance (m = 10000)	0. 371
SNNS using Cosine Similarity (m = 10000)	0. 324

for computation of representative measure is absent because of significant dimensionality reduction and inability to establish a correspondence to original signal. Therefore, feature selection is to be done on the basis of discriminative measures alone. The procedure for optimal feature selection in this framework is detailed below.

1. Assuming there are n speakers and m different i-vector systems, and i represents the index of target speaker for whom optimal feature has to be identified.

$$S_{ij} = \sum_{k\epsilon\{1...n\}-\{i\}} \phi(i,k) \tag{7}$$

   where S_{ij} refers to sum of inter-speaker similarity calculated in the feature domain j for a particular target speaker i against all non-target speakers $k\epsilon\{1...n\} - \{i\}$, and $\phi(i,k)$ refers to the chosen similarity measure between train i-vector of speaker i and train i-vector of speaker k. Similarity measures are chosen to be one amongst cosine similarity, shared nearest neighbour similarity and negative of Euclidean distance. In this paper, all these three measures are used individually to determine the feature that must be used for every speaker.

2. Compute ($S_{i1},\ S_{i2},...S_{im}$) $\forall\ i\ \epsilon$ (1,2,...n) i.e similarity between each speaker and corresponding non-target speakers in every feature domain.

3. Optimal feature index, $O(i)$ is chosen as that feature domain that offers minimum amount of similarity between a speaker and corresponding non-target speakers.

$$O(i) = \{l : l\epsilon\{1...m\}\&\min_{l}\left(S_{il}\right)\} \tag{8}$$

4. After subjecting i-vectors to LDA and WCCN, feature selection is carried out as per the procedure enlisted above followed by scoring for all the evaluation trails. For each claim of a particular target speaker, the score is computed using i-vector in selected feature domain and test i-vector in the same feature domain. The score computed in this study is chosen to be SNNS owing to the results in Section 3.5 and its characteristics.

4. Experimental Results

4.1. Experimental Setup

A gender dependent TVS system is built using NIST 1999, 2003, 2004, 2005, 2006 and 2008 databases as development dataset. Short-term features and corresponding delta features are concatenated to form 38-dimensional feature vector for every frame. These feature vectors are subjected to short-term gaussianisation. A 1024-mixture GMM-UBM is estimated from the development dataset and 500-dimension i-vectors are extracted. Before the application of feature switching, SNNS

Table 2: Comparison of MFCC system, MGD system and their fusion systems' performance(EER) on NIST 2010 male database

Feature/Condition	det5	det6	det8
MFCC with cosine(1)	4. 96	6. 05	1. 98
MGD with cosine(2)	6. 35	8. 06	2. 82
i-vector fusion(1,2)	5. 01	6. 79	3. 13
score fusion(1,2)	4. 57	5. 63	1. 68
MFCC with SNNS(3)	4. 36	7. 02	2. 17
MGD with SNNS(4)	5. 31	9. 15	2. 26
score fusion(3,4)	3. 90	7. 12	1. 66

Table 3: Comparison of Feature Switching System Results using different Distance/ Similarity metrics

$\phi(i,k)$/Condition	det5	det6	det8
Euclidean	4. 07	**5. 32**	**1. 43**
Cosine	**3. 64**	6. 93	1. 57
SNNS	5. 14	9. 31	3. 11

scoring and score normalisation, LDA and WCCN are applied to i-vectors in all chosen feature domains. T-Normalisation is used as the score normalisation strategy with a cohort set of size 150 from the same evaluation set. Telephone-telephone conditions are evaluated for NIST 2010 male database[22].

4.2. Results

Results are presented in terms of Equal Error Rate (EER) for NIST 2010 male database using MFCC-based, MGD-based and fusion systems in Table 2. Score fusion out performed i-vector fusion in the case of cosine similarity scoring. SNNS with cosine similarity was also used for the same conditions, and improved performance was obtained for two out of three conditions (ref Table 2). Therefore, SNNS was chosen for scoring for all the feature switching experiments carried out.

Results for the feature switching system with varying $\phi(i,k)$ mentioned in Section3.6 are presented in Table 3. Different measures used as $\phi(i,k)$ are detailed in Section 3.3, 3.4. It is important to verify the behaviour with different similarity measures. MGD is found to be the optimal feature for 1306, 1244 and 1097 speakers using SNNS, cosine and Euclidean based measures respectively. It can be inferred from the results that, feature switching carried out using cosine and Euclidean distance showed improved results compared to i-vector fusion. The results also indicate that MGD features are more efficient in discriminating majority of speakers when compared to MFCC which concurs with the formant resolution argument presented before.

5. Conclusion

This paper presents a simple method for feature switching in TVS framework and results are better compared to the fixed weight-based score fusion. SNNS scoring is shown to be performing better than cosine scoring. The feature switching method is important because of its applicability to all those high-data resource domains in which TVS modeling can be used. Further detailed analysis regarding the success of certain measures over others can be performed as part of future work.

6. References

[1] R. Saeidi, K. Lee, T. Kinnunen, T. Hasan, B. Fauve, P. Bousquet, E. Khoury, P. S. Martinez, J. Kua, C. You *et al.*, "I4u submission to nist sre 2012: A large-scale collaborative effort for noise-robust speaker verification," in *Proc. Interspeech*, 2013.

[2] R. Padmanabhan, S. H. K. Parthasarathi, and H. A. Murthy, "Robustness of phase based features for speaker recognition." in *INTERSPEECH*, 2009, pp. 2355–2358.

[3] T. Thiruvaran, E. Ambikairajah, and J. Epps, "Group delay features for speaker recognition," in *Information, Communications & Signal Processing, 2007 6th International Conference on*. IEEE, 2007, pp. 1–5.

[4] S. Dey, R. Rajan, R. Padmanabhan, and H. A. Murthy, "Feature diversity for emotion, language and speaker verification," in *Communications (NCC), 2011 National Conference on*. IEEE, 2011, pp. 1–5.

[5] J. Kumar, R. Janakiraman, and H. A. Murthy, "Kl divergence based feature switching in the linguistic search space for automatic speech recognition," in *Communications (NCC), 2010 National Conference on*. IEEE, 2010, pp. 1–5.

[6] R. Ramya, R. M. Hegde, and H. A. Murthy, "Significance of group delay based acoustic features in the linguistic search space for robust speech recognition." in *INTERSPEECH*, 2008, pp. 1537–1540.

[7] R. Padmanabhan and H. A. Murthy, "Acoustic feature diversity and speaker verification." in *INTERSPEECH*, 2010, pp. 2110–2113.

[8] H. A. Murthy and V. Gadde, "The modified group delay function and its application to phoneme recognition," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 1. IEEE, 2003, pp. I–68.

[9] H. A. Murthy and B. Yegnanarayana, "Group delay functions and its applications in speech technology," *Sadhana*, vol. 36, no. 5, pp. 745–782, 2011.

[10] P. Rose, *Forensic speaker identification*. CRC Press, 2003.

[11] H. Hermansky, "Perceptual linear predictive (plp) analysis of speech," *the Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738–1752, 1990.

[12] A. Diment, P. Rajan, T. Heittola, and T. Virtanen, "Modified group delay feature for musical instrument recognition," 2013.

[13] M. Guo, J. Wang, L. Dongxin, and L. Chang, "Depression detection using the derivative features of group delay and delta phase spectrum," in *Advance Computing Conference (IACC), 2013 IEEE 3rd International*. IEEE, 2013, pp. 1275–1278.

[14] S. Narasimhan, N. Basumallick, and R. Chaitanya, "Improved phase estimation based on complete bispectrum and modified group delay," *Signal, Image and Video Processing*, vol. 2, no. 3, pp. 261–274, 2008.

[15] M. E. Houle, H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Can shared-neighbor distances defeat the curse of dimensionality?" in *Scientific and Statistical Database Management*. Springer, 2010, pp. 482–500.

[16] W. Li, "Mutual information functions versus correlation functions," *Journal of statistical physics*, vol. 60, no. 5-6, pp. 823–837, 1990.

[17] S. Kullback, "The kullback-leibler distance," *American Statistician*, vol. 41, no. 4, pp. 340–340, 1987.

[18] A. Flexer and D. Schnitzer, "Can shared nearest neighbors reduce hubness in high-dimensional spaces?" in *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 460–467.

[19] L. Ertoz, M. Steinbach, and V. Kumar, "A new shared nearest neighbor clustering algorithm and its applications," in *Workshop on Clustering High Dimensional Data and its Applications at 2nd SIAM International Conference on Data Mining*, 2002, pp. 105–115.

[20] *NIST i-vector Challenge, 2014*, 2013, http://www.nist.gov/itl/iad/mig/ivec.cfm.

[21] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 788–798, 2011.

[22] *NIST SRE 2010 Evaluation Plan*, 2010, http://www.itl.nist.gov/iad/mig//tests/sre/2010/NIST_SRE10_evalplan.r6.pdf.