INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Out of Set Language Modelling in Hierarchical Language Identification

Saad Irtza[1,2], Vidhyasaharan Sethu[1], Sarith Fernando[1,2], Eliathamby Ambikairajah[1,2], and Haizhou Li[3]
[1]School of Electrical Engineering and Telecommunications, UNSW Australia
[2]ATP Research Laboratory, National ICT Australia (NICTA), Australia
[3]Institute for Infocomm Research, A*STAR, Singapore
s.irtza@unsw.edu.au

Abstract

This paper proposes a novel approach to the open set language identification task by introducing out of set (OOS) language modelling in a Hierarchical Language Identification (HLID) framework. Most recent language identification systems make use of data sources from other than target languages to model OOS languages. The proposed approach does not require such data to model OOS languages, instead it only uses data from target languages. Additionally, a diverse language selection method is incorporated to further improve OOS language modelling. This work also proposes the use of a new training data selection method to develop compact models in a hierarchical framework. Experiments are conducted on the recent NIST LRE 2015 data set. The overall results show relative improvements of 32.9% and 30.1% in terms of C_{avg} with and without the diverse language selection method respectively over the corresponding baseline systems, when using the proposed hierarchical OOS modelling.

Index Terms: Language identification, Hierarchical framework, out of set language modelling, i-vector, BNF

1. Introduction

The most widely adopted approaches to language identification use acoustic and phonotactic information [1-3]. Specifically, current systems employ the i-vector framework trained on both acoustic and phonotactic front-ends. MFCCs and Phone Log Likelihood Ratios (PLLRs) continue to be the most commonly utilised acoustic and phonotactic front-ends [4-7] and recently bottleneck features (BNF) have exhibited promising performances [8]. State-of-the-art language identification (LID) systems also make use of score level fusion to combine individual systems based on different speech cues [2]. However, these systems are all single level approaches where all language hypotheses are treated identically. Hierarchical LID frameworks have been proposed as an alternative to single level structures [9-11]. These are based on the observation that it is easier to distinguish between languages that are significantly dissimilar than those having a lot of similarities. Hierarchical systems utilise different speech cues that are more discriminative in different hierarchy levels (e.g., Prosodic cues are significantly better at distinguishing between a tonal and a non-tonal language than they are at distinguishing between two non-tonal languages) [1]. Preliminary work on the use of hierarchical structures have shown some promising performances [9-11].

Most recent research on language identification has been focused on closed set language identification where all test utterances correspond to one of a small set of target languages. However, in most practical scenarios where language identification systems may be employed, test utterances are not likely to be strictly limited to a small set of target languages but may also correspond to some unknown languages. This scenario is referred to as open set language identification. Some recent approaches to open set LID aim to capture out of set (OOS) language characteristics by using additional data from language that are not in the set of target languages [12]. This approach performs well if one has enough non-target language data to model OOS languages. However, given that the number of non-target languages is likely to be very large, it would be desirable to be able to detect OOS languages without using any additional non-target language data for training the system. In [13], the OOS language model was developed using training data from only the target languages. The analysis in [13] shows that this approach significantly reduces the false acceptance (FA) but also increases the false rejection (FR) rate.

The hierarchical framework for language identification [11] allows for multiple level of classification and consequently can include different OOS language models at different hierarchy level. This paper investigates: 1) the effectiveness of hierarchical frameworks in recognising OOS languages without using any additional non-target language data for OOS language model training; 2) the effectiveness of multiple OOS language models; and 3) the effectiveness of a diverse language selection method for developing OOS language models and to improve false rejection. This paper also investigates the effect of a new training data selection method in the context of hierarchical language identification systems.

2. Hierarchical Framework

The hierarchical framework for language identification was introduced with the aim of separating the problem of language identification into a top-down hierarchy of decisions, with initial high level decisions pertaining to identification of language groups followed by identification of specific languages and dialects from a smaller subset at the lower levels of the hierarchy. Hierarchical frameworks have been shown to significantly reduce the confusion among similar languages [11]. In the hierarchical language identification system employed in the experiments reported in this paper, the major language groups are identified in the first (topmost) level (Figure 1) of hierarchy to reflect the languages available in the NIST 2015 LRE database that is used in these experiments and harder problem of identifying dialects [14] is carried out at the lowest level. We then propose the introduction of out of set (OOS) language models at multiple levels of the hierarchy within this framework.

http://dx.doi.org/10.21437/Interspeech.2016-558

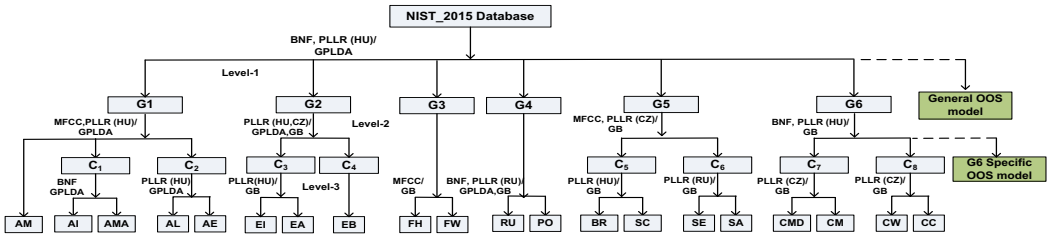

Figure 1: *Language Hierarchical Tree Structure estimated from NIST 2015. Each node is labelled as features/classifier [A*].*

2.1. Language Clustering

The hierarchical clustering algorithm [11] is used to determine the language clusters in each level. In the first layer of hierarchy, languages are grouped based on phonotactic content and linguistic information. The similarity, $S(\cdot)$, between language pair (L_a, L_b) is computed as per equation 1, where $a, b \in (1, N)$ and N is number of languages.

$$S(L_a, L_b) = \left(1 - K_s(L_a, L_b)\right) \times E(L_a, L_b) \quad (1)$$

Where, $K_s(\cdot)$ is the symmetric K-divergence between phoneme probability distribution of language L_a and L_b. K-Divergence is a measure of similarity between two probability distributions [15] and has been effectively used in clustering algorithms [16, 17]. It should be noted that using the symmetric KL-divergence in place of the K-divergence also results in identical language clustering. $E(\cdot)$ is language grouping information of language L_a and L_b according to Ethnologue linguistic community [18] and is given by:

$$E(L_a, L_b) = \begin{cases} 1, & L_a \; and \; L_b \in G \\ 0.5, & otherwise \end{cases} \quad (2)$$

Here, G is a language group in Ethnologue. Each language pair is given a prior probability of '1' if they belong to same language group in Ethnologue and 0.5 otherwise if the language group is unknown, they are given a prior probability of 0.5. These constant prior probability values were selected empirically.

The symmetric K-divergence between two languages, L_A and L_b, is defined as [15]:

$$K_s(L_a, L_b) = \sum_{i=1}^{N_P} \left[P(p_i|L_a) \ln\left(\frac{2P(p_i|L_a)}{P(p_i|L_a) + P(p_i|L_b)}\right) + P(p_i|L_b) \ln\left(\frac{2P(p_i|L_b)}{P(p_i|L_a) + P(p_i|L_b)}\right)\right] \quad (3)$$

Where N_p is the number of phonemes considered, and $P(p_i|L_j)$ is the posterior probability of the i^{th} phoneme, p_i, given language j.

In the system outlined in this paper, a Hungarian (HU) TRAPs/NN phone decoder [19] is used to estimate the phoneme probability distribution for each language from the training data. The set of 61 phonemes recognised by the chosen Hungarian phone decoder is reduced to a smaller set of 54 by discarding the phonemes that did not occur in the training data of any of the languages.

The agglomerative clustering algorithm previously used for determining the structure of a hierarchical language identification system [11] is used again to determine the language groups at the top of the hierarchy (G_1 to G_6 in Figure 1).

Following this, subsequent language groups in all lower levels of the hierarchy are determined using average cost performance (C_{avg}) [20] as a measure of similarity between languages. The C_{avg} is computed on the development set from the baseline system described in section 5.1. A different similarity metric is used for the lower levels of the hierarchy since it may be necessary to make a distinction between different dialects of the same language which share similar phonetic distributions and consequently a similarity measure based on phonetic distributions may not be suitable [14]. Figure 1 show the hierarchical structure obtained from the NIST 2015 LRE database.

3. Out of Set Language Modelling

In this work, we propose modelling out of set languages at multiple levels of the hierarchy where the out of set model at each node is a model of all languages not considered at that node. In addition a diverse language selection method is employed in order to produce a broad out of set model using only training data from target languages.

3.1. Proposed Hierarchical OOS Modelling

The hierarchical framework offers a range of possibilities for the inclusion of OOS language models. One possibility is the inclusion of a universal language model, trained from all target languages, at the topmost level of the hierarchy that distinguishes between broad language groups. This would serve as a background model (similar to a Universal Background Model in speaker verification) and it is expected that test utterances from an unknown language would match this background model better than any of the language group models. This OOS model is developed using all the target languages data also known as target independent (TI) OOS language model [13]. In this paper we propose that this idea be extended to other nodes in the hierarchical structure as well. It is expected that this extension is advantageous since the OOS models at each node provides a different background model that is specific to the languages considered at each node and are consequently the hierarchical structure on the whole can detect OOS languages more reliably. In the systems evaluated in this paper, OOS language models are incorporated into the nodes in the second layer of the hierarchy as well as the topmost (first) layer. These OOS models are developed separately for each language group identified at the topmost layer (Group Specific OOS). Group specific OOS models are trained on the data from target languages that are not present in that group e.g. OOS model for G1 is developed on training data for G2 to G6 (OOS language models are shown in green in Figure 1).

3.2. Diverse Language Selection for OOS Model

It is expected that the role of data selection in training of target independent OOS language models is similar to that of data selection in Universal Background Modelling (UBM) [13]. It has been shown that speaker selection can play an important role in UBM data selection [21]. It has further been shown that

using all available development data for UBM training can lead to clusters from different speakers having similar characteristics, which in turn leads to high rates of false rejection [21]. A similar effect can be expected in target independent OOS language modelling, particularly since only target data is used in training the model (as opposed to the use of development data from other background speakers in speaker verification). Therefore, in this work, we incorporate a diverse language selection method to develop target independent models. The selection of diverse languages is based on symmetric Kullback-Leibler (KL) divergence method between all pairs of target languages [15]. For each language i-vector L_j, where $j \in (1, N)$ and N is the total number of languages, a single Gaussian model is trained with shared covariance matrix (Σ) across all languages i-vector data [13]. KL divergence between single Gaussian models λ_i and λ_j is computed as:

$$D_{KL}^{s}(\lambda_i, \lambda_j) = 0.5(\overline{\mu_j} - \overline{\mu_i})\left(\frac{2}{\Sigma}\right)(\overline{\mu_j} - \overline{\mu_i}) \quad (4)$$

Where, μ_i and μ_j denote the means of the Gaussian models λ_i and λ_j respectively.

In order to measure how diverse a language L_i is from the other languages, a diversity factor D_i is computed from the KL divergence using equation (5):

$$D_i = \frac{1}{N}\sum_{j \in N, j \neq i} D_{KL}^{s}(\lambda_i, \lambda_j) \quad (5)$$

D_i is a measure of average divergence of model (λ_i) from all other models. Based on the average divergence of all N models, $\{D_i; 1 \leq i \leq N_D\}$, the models corresponding to the highest average divergences (N_D) are selected for OOS modelling.

4. Proposed Training Data Selection

In addition to speaker selection, it has been shown that training data selection of a model might improve performance by reducing the overlap between models [21]. Based on the assumption that this would hold for language identification systems as well, we introduce a training data selection method for training the models in the first two levels of the hierarchical structure. A cosine similarity factor, S_i^a, is computed for all training utterance (i-vectors) u_i^a corresponding to each language/language group, M^a, where $a \in (1, N_M)$ and N_M denotes the number of languages/language groups modelled in the first two levels of the hierarchy:

$$S_i^a = \frac{1}{U^a}\sum_{j \in U^a, j \neq i} s(u_i^a, u_j^a) \quad (6)$$

Where, U^a denotes the total number of training utterances available for language/language group M^a and $s(u_i^a, u_j^a)$ denotes the cosine similarity between utterances u_i^a and u_j^a and is given by:

$$s(u_i^a, u_j^a) = \frac{u_i^a \cdot u_j^a}{\|u_i^a\|\|u_j^a\|} \quad (6)$$

The training data selection is then implemented by discarding utterances that correspond to negative values of S_i^a.

5. Classification

The hierarchical framework allows for different front-ends and back-ends to be employed at different nodes for classification. In the system presented in this paper, both are chosen to provide the best classification at that node from a predetermined set of features and classifiers. Specifically, at each node the front-end is chosen as one of or a combination of the following feature sets – PLLRs from Hungarian, Czech, and Russian phone decodes, MFCCs, and bottleneck features (BNF) based on J-Measures [22] of these features and feature combinations estimated on a development dataset. The features employed at each node of the system presented here are denoted in Figure 1.

Also, the back-end at each node is chosen as either one of the two most commonly used probabilistic back-ends used in language identification or a score level fusion of both. Namely, Generative Gaussian Back-ends (GB) and Gaussian Probabilistic Linear Discriminant Analysis (GPLDA) [23, 24]. The choice of the back-end is based on performance as evaluated on a development set. Both of these back-ends operate on i-vectors estimated from the chosen front-end and Linear Discriminant Analysis (LDA) is applied on the i-vectors prior to using GPLDA (400D to 150D) but not the generative Gaussian back-end (preliminary experiments indicated that LDA degraded performance when used in conjunction with the generative Gaussian back-end).

At each node of the hierarchical framework, the log likelihoods (LLs) of all languages/language groups that are considered in that node are estimated as in [11]. Multiclass calibration models are then estimated from these LLs, using FoCal toolkit, on a development data set and applied to the test set. Finally these calibrated LLs are converted into log likelihood ratios (LLRs) as defined by NIST [20].

Average cost performance (C_{avg}) and log likelihood ratio cost (C_{llr}), as defined by NIST [20] are employed as performance metrics in this paper. For the closed set experiments, these measures are computed for each of the six language groups defined by the NIST 2015 evaluation plan [20] and averaged to compute overall system performance. In the open set experiments, C_{avg} and C_{llr} are defined as in [25].

5.1. Baseline System

The performance of the proposed HLID framework is compared to a baseline system that comprises of five sub-systems that are fused at a score level. The five sub-systems are all i-vector-GPLDA systems that each uses a different front-end. The five front-ends are: two acoustic front-ends (MFCC and BNF) and three phonotactic front-ends (PLLR using Hungarian, Russian and Czech phonemes) as described in section 6.2. The performance of the baseline system matches that of the fourth best performing system in the NIST 2015 LRE. For the open set experiments, the OOS language models (one for each subsystem) in the baseline system was estimated from the training data from all target languages.

6. Experimental Setup

6.1. Database

The closed set results are reported on the NIST 2015 LRE dataset as per the fixed test conditions given in [20] and involves 20 target languages. For development purposes, 10 conversations from each language were randomly chosen. The open set LID experiments required additional test data corresponding to out of set languages. This additional out of set test data of duration 30, 10 and 3 seconds from 17 different languages was selected from NIST 2007 and 2011 LRE datasets. These languages are Bengali, Czech, Dari, Farsi,

Thai, Urdu, Japanese, Vietnamese, Ukrainian, Hindi, Punjabi, Pashto, Tamil, Turkish, German, Korean and Lao.

6.2. Feature Extraction

The three sets of frame based PLLR features - Hungarian (HU), Russian (RU) and Czech (CZ), and MFCCs used in all the systems reported in this paper are augmented with SDCs and estimated as outlined in [11]. The Bottleneck features (BNF) of 42 dimensions are extracted using Deep Neural Networks implemented with the Kaldi toolkit [26]. The DNN was trained on 300 hours of Switchboard-I data and uses 13 dimensional MFCC's features as input [8]. The DNN comprises of 5 layers with 1024 nodes in each layer except the 4th layer which served as a 42 node bottleneck layer. All i-vectors based on these front ends were 400 dimensional [23] and estimated using Universal Background Models (UBMs) of 1024 mixture components.

7. Results

In this work, two set of experiments were conducted 1) closed set language detection and 2) open set language detection. The closed set experiments were conducted to a) quantify the performance of the hierarchical structure without OOS language modelling and b) investigate the effectiveness of data selection approaches in this hierarchical structure. The open set experiments were conducted to investigate a) the proposed hierarchical OOS modelling, b) the effect of having OOS models in different hierarchy level c) the use of diverse language selection for training OOS models.

7.1. Closed Set Detection Results

Table 1 reports the performances of the baseline system as well the hierarchical language identification system with and without the training data selection method outlined in section 4. From these results it can be seen that the hierarchical approach significantly outperforms the baseline system and the training data selection method further improves this.

Table 1: *Closed Set Detection Results on NIST 2015*

Language Groups	100* C_{avg} / C_{LLR}		
	Baseline	Hierarchical (using all data)	Hierarchical (data selection)
Arabic	20.3/0.61	18.5/ 0.57	**17.9/0.55**
Chinese	17.9/0.63	14.9/ 0.52	**14.2/0.50**
English	11.2/0.44	10.8/ 0.39	**9.8/0.38**
French	47.2/1.3	36.9/ 0.91	**33.6/0.87**
Slavic	3.85/0.27	3.49/ 0.15	**2.7/0.14**
Iberian	21.7/0.60	19.08/ 0.60	**18.9/0.60**
Overall	20.3/ 0.64	17.2/ 0.52	**16.2/ 0.50**

7.2. Open Set Detection Results

Table 2 shows the performance of baseline and hierarchical systems with open set LID. Compared to the closed set results (Table 1), it can be seen that the inclusion of out of set test utterances degrades performance even with explicit out of set language modelling in the baseline system. It can also be seen that the proposed hierarchical OOS modelling in either of the two top levels is significantly better than OOS modelling in a non-hierarchical framework and combined OOS modelling in both levels further improves the performance.

Table 2: Open Set Detection Results using Baseline Approach

Language Groups	100* C_{avg} / C_{LLR}			
	Baseline	Hierarchical (OOS level 1)	Hierarchical (OOS level 2)	Hierarchical (OOS level 1&2)
Arabic	29.2/0.82	21.3/ 0.64	20.9/0.64	**20.7/ 0.64**
Chinese	24.3/0.76	19.7/ 0.61	19.1/0.61	**18.1/ 0.59**
English	25.9/0.69	16.8/ 0.54	16.0/0.53	**15.5/ 0.52**
French	37.9/ 0.99	35.4/ 0.89	34.5/0.88	**34.0/ 0.88**
Slavic	27.8/ 1.2	8.1/ 0.27	7.6/0.25	**6.4/ 0.24**
Iberian	25.8/0.71	21.6/ 0.66	21.1/0.65	**20.3/ 0.64**
Overall	28.5/ 0.86	20.4/ 0.60	19.8/ 0.59	**19.1/ 0.58**

The results of the experiment conducted to determine the effectiveness of the diverse language selection method are reported in Table 3 and when compared to Table 2 it can be seen that the diverse language selection improves system performance. The results in Table 3 are obtained as a result of selecting 15 languages for each model based on best performance on development test set.

Table 3: *Open Set Detection Results using Proposed Approach*

Language Groups	100* C_{avg} / C_{LLR}			
	Baseline	Hierarchical (OOS level 1)	Hierarchical (OOS level 2)	Hierarchical (OOS level 1&2)
Arabic	27.4/0.75	20.7/ 0.63	19.5/0.60	**19.0/ 0.59**
Chinese	22.5/0.74	18.2/ 0.58	16.8/0.56	**15.8/ 0.54**
English	23.1/0.67	15.2/ 0.51	15.7/0.52	**15.0/ 0.51**
French	35.4/ 0.90	34.0/ 0.88	34.29/0.88	**33.6/ 0.87**
Slavic	25.9/ 0.75	6.6/ 0.24	7.2/0.26	**6.1/ 0.23**
Iberian	23.1/ 0.68	19.7/ 0.62	19.8/0.65	**19.1/ 0.63**
Overall	26.2/ 0.74	19.06/ 0.57	18.9/ 0.57	**18.1/ 0.56**

The performance of the diverse language selection is also evaluated in terms of false acceptance rates and false rejection rates by comparing systems (baseline and hierarchical) with and without the language selection in Table 4.

Table 4: *False Acceptance and Rejection Results*

Language Groups	False Acceptance rate / False Rejection rate			
	Using all languages		Using diverse languages	
	Baseline	Hierarchical	Baseline	Hierarchical
Arabic	59.8/23.6	13.9/ 20.5	53.8/ 19.3	**6.8/ 11.6**
Chinese	28.5/ 20.9	13.1/ 16.0	18.2/ 13.4	**2.4/ 10.7**
English	66.7/ 36.0	20.1/ 19.5	50.4/ 27.9	**13.8/ 9.0**
French	36.6/ 27.9	24.5/ 21.4	27.9/ 14.4	**19.9/ 10.3**
Slavic	58.8/ 52.3	15.5/ 10.9	51.5/ 33.0	**12.9/ 7.7**
Iberian	31.2/ 44.5	18.3/ 20.5	22.7/ 31.3	**10.2/ 10.8**
Overall	46.9/ 29.4	19.4/ 18.8	37.4/ 19.9	**11.0/ 10.5**

8. Conclusion

This paper has focused on addressing the problem of open set language identification and shows that the hierarchical framework is well suited for this task. The study indicates that the development of OOS language models in each level of a hierarchical framework is able to better reject unknown languages when compared to a non-hierarchical approach. Incorporation of diverse language selection into the system further improves performance and reduces both the false acceptance and rejection rates in hierarchical and non-hierarchical systems. Finally, the proposed training data selection is also shown to improve system performance in both open and closed set language identification tasks.

9. References

[1] E. Ambikairajah, H. Li, L. Wang, B. Yin, and V. Sethu, "Language identification: A tutorial," *Circuits and Systems Magazine, IEEE,* vol. 11, pp. 82-108, 2011.

[2] H. Li, B. Ma, and K. A. Lee, "Spoken language recognition: from fundamentals to practice," *Proceedings of the IEEE,* vol. 101, pp. 1136-1159, 2013.

[3] D. A. Reynolds, W. M. Campbell, W. Shen, and E. Singer, "Automatic language recognition via spectral and token based approaches," in *Springer Handbook of Speech Processing*, ed: Springer, 2008, pp. 811-824.

[4] N. Dehak, P. A. Torres-Carrasquillo, D. A. Reynolds, and R. Dehak, "Language Recognition via i-vectors and Dimensionality Reduction," in *INTERSPEECH*, 2011, pp. 857-860.

[5] M. Soufifar, M. Kockmann, L. Burget, O. Plchot, O. Glembek, and T. Svendsen, "iVector Approach to Phonotactic Language Recognition," in *INTERSPEECH*, 2011, pp. 2913-2916.

[6] M. Díez, A. Varona, M. Peñagarikano, L. J. Rodríguez-Fuentes, and G. Bordel, "On the use of phone log-likelihood ratios as features in spoken language recognition," in *SLT*, 2012, pp. 274-279.

[7] L. D'Haro, R. Cordoba, C. Salamea, and J. Echeverry, "Extended phone log-likelihood ratio features and acoustic-based i-vectors for language recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014, pp. 5342-5346.

[8] F. Richardson, D. Reynolds, and N. Dehak, "A Unified Deep Neural Network for Speaker and Language Recognition," *arXiv preprint arXiv:1504.00923,* 2015.

[9] B. Yin, E. Ambikairajah, and F. Chen, "Hierarchical language identification based on automatic language clustering," in *INTERSPEECH*, 2007, pp. 178-181.

[10] B. Yin, E. Ambikairajah, and F. Chen, "Improvements on hierarchical language identification based on automatic language clustering," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 4241-4244.

[11] S. Irtza, V. Sethu, E. Ambikairajah and H. Li " A Hierarchical Framework for Language Identification," published in the proceedings of 41st IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Shanghai, China, 2016.

[12] Q. Zhang and J. H. Hansen, "Training candidate selection for effective rejection in open-set language identification," in *Spoken Language Technology Workshop (SLT), 2014 IEEE*, 2014, pp. 384-389.

[13] M. F. BenZeghiba, J.-L. Gauvain, and L. Lamel, "Gaussian backend design for open-set language detection," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 4349-4352.

[14] Y. Lei and J. H. Hansen, "Dialect classification via text-independent training and testing for arabic, spanish, and chinese," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 19, pp. 85-96, 2011.

[15] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *City,* vol. 1, p. 1, 2007.

[16] M. Das Gupta, S. Srinivasa, and M. Antony, "KL Divergence based Agglomerative Clustering for Automated Vitiligo Grading," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2700-2709.

[17] K. Chaudhuri and A. McGregor, "Finding Metric Structure in Information Theoretic Clustering," in *COLT*, 2008, p. 10.

[18] M. P. Lewis, G. F. Simons, and C. D. Fennig, *Ethnologue: Languages of the world* vol. 16: SIL international Dallas, TX, 2009.

[19] P. Schwarz, "Phoneme recognition based on long temporal context," Ph.D. dissertation, Faculty of Information Technology, Brno University of Technology, http://www.fit.vutbr.cz/, Brno, Czech Republic, 2008.

[20] A. F. Martin, C. S. Greenberg, J. M. Howard, D. Bansé, G. R. Doddington, J. Hernández-Cordero*, et al.*, "NIST Language Recognition Evaluation—Plans for 2015," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[21] T. Hasan and J. H. Hansen, "A study on universal background model training in speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 19, pp. 1890-1899, 2011.

[22] S. Nicholson, B. P. Milner, and S. J. Cox, "Evaluating feature set performance using the f-ratio and j-measures," in *EUROSPEECH*, 1997, pp. 413-416.

[23] S. Irtza, V. Sethu, P. N. Le, E. Ambikairajah, and H. Li, "Phonemes Frequency Based PLLR Dimensionality Reduction for Language Recognition," in Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[24] S. Irtza, Haris Bavattichalil, V. Sethu and E. Ambikairajah, (2015) "Scalable I-vector Concatenation for PLDA based Language Identification System," In the proceedings of 7th Asia-Pacific Signal and Information Processing Association Conference (APSIPA), Hong Kong.

[25] A. F. Martin and C. S. Greenberg, "The 2009 NIST Language Recognition Evaluation," in *Odyssey*, 2010, p. 30.

[26] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel*, et al.*, "The Kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and understanding*, 2011.

[A*] AM=Arabic Modern, AI=Arabic Iraqi, AMA=Arabic Maghrebi, AL=Arabic Levantine, AE=Arabic Egyptian, EI=English Indian, EA=English American, EB=English British, FH=French Haitian, FW=French West African, RU=Russian, PO=Polish, BR=Brazilian, SC=Spanish Caribbean, SA=Spanish American, SE=Spanish European, CMD=Chinese Ming Dong, CM=Chinese Mandarin, CW=Chinese Wu, CC=Chinese Cantonese