SUPERVECTOR PRE-PROCESSING FOR PRSVM-BASED CHINESE AND ARABIC DIALECT IDENTIFICATION

Qian Zhang, Hynek Bořil, John H. L. Hansen[*]

Center for Robust Speech Systems (CRSS), Erik Jonsson School of Engineering,
University of Texas at Dallas, Richardson, Texas, U.S.A.
{qian.zhang,hynek,john.hansen}@utdallas.edu

ABSTRACT

Phonotactic modeling has become a widely used means for speaker, language, and dialect recognition. This paper explores variations to supervector pre-processing for phone recognition–support vector machines (PRSVM) based dialect identification. The aspects studied are: (i) normalization of supervector dimensions in the pre-squashing stage, (ii) impact of alternative squashing functions, and (iii) N-gram selection for supervector dimensionality reduction. In (i) and (ii), we find that several alternatives to commonly used approaches can provide moderate, yet consistent performance improvements. In (iii), a newly proposed dialect salience measure is applied in supervector dimension selection and compared to a common N-gram frequency based selection. The results show a strong correlation between dialect-salience and frequency of occurrence in N-grams. The evaluations in this study are conducted on a corpus of Chinese dialects, a Pan-Arabic corpus, and a set of Arabic CTS corpora.

***Index Terms*—** Dialect identification, phonotactic modeling, PRSVM, dialect-salience, squashing function

1. INTRODUCTION

State-of-the-art dialect identification (DID) systems share similar techniques with speaker and language recognition. Cepstral features with shifted delta cepstra (SDC) [1, 2], Gaussian mixture modeling with universal background models (GMM-UBM) and GMM supervectors [3, 4], phonotactic models realized by parallel phone recognizers and language modeling (PPRLM) [5–8], and phone recognizers combined with support vector machines (PRSVM) [9, 10] are highly popular in current systems as seen in NIST-SRE [11] and NIST-LRE [12] submissions.

The focus of our study is on PRSVM-based phonotactic modeling for DID. In PRSVM, speech signal is first decoded by a phone recognizer into a sequence of phones [9] or phone lattices [10]. The phone recognizer (PR) can be trained on a language or a mixture of languages that are not necessarily related to the dialects targeted in the DID task [4]. The idea is that when decoding an utterance, even from an unknown language, the PR will generate sequence of phones or phone lattices that reflect the PR's acoustic model states closest to the processed signal, and different dialects may generate different PR outputs. PR outputs are subsequently normalized, processed by a squashing function, stacked into supervectors, and passed to SVM classifiers.

In our recent study on PRSVM-based Arabic DID [13], we observed that replacing the traditional logarithmic squashing function in the PRSVM supervector pre-processing stage by alternative functions, sigmoid and hard limit, had a positive impact on the system performance. The goal of our current study is to verify whether the observed performance gains can transfer to a set of dialects drawn from another language (Chinese dialects), and investigate further options in supervector pre-processing.

Three aspects of the PRSVM supervector pre-processing are studied: (i) scaling and normalization of N-gram relative frequencies prior to applying squashing, (ii) efficiency of traditional versus alternative squashing functions, and (iii) selection of N-grams for supervector dimension reduction. To reduce the computational overhead, it is a common practice in literature to preserve only dimensions corresponding to the most frequently occurring N-grams [14] and drop the rest. However, to our best knowledge, there has not been a study on the correlation between the frequency of N-grams and their dialect salience. In theory, some frequent N-grams could appear with similar probability in all or a majority of the target dialect classes and be of a little use to their discrimination. To analyze this, we propose a so called *dialect-salience* measure that rank orders N-grams based on the non-uniformity of their occurrence across dialects. Frequency-based and dialect-salience based dimension reduction are compared side-by-side by utilizing the respective reduced supervectors in DID tasks, as well as by directly comparing the overlap of the N-gram sets selected by the two methods.

All approaches discussed in this study are evaluated on one Chinese and two Arabic DID data sets, each capturing four dialect classes. PRSVMs based on nine BUT phone recognizers [15] are used in the evaluations.

2. CORPORA

The two Arabic data sets used in this paper, a Pan-Arabic corpus and a set of Arabic CTS corpora, are identical with our previous study [13].

The CTS set comprises the following corpora: Iraqi Arabic CTS (IRQ; LDC2006S45), Gulf Arabic CTS (GLF; LDC2006S43), Arabic CTS Levantine Fisher Training Data Set 3 (LEV; LDC2005S07), and CALLHOME Egyptian Arabic Speech (EGY; LDC97S45 and supplement LDC2002S37). It is noted that these databases were acquired by LDC for the purpose of automatic speech recognition projects and LDC as such did not make any suggestions for their use in dialect identification. In [13], we found that each dialect data in these CTS collections capture unique and fairly distinctive long-term channel characteristics that are sufficient themselves for performing a successful DID. The reason for using the CTS set here is that a similar CTS set was previously used in Arabic DID studies in [14, 16, 17] (where LEV was represented by Levantine Arabic CTS, LDC2007S01 instead of Fisher) and hence, the performance of the systems here can be directly compared to the previous studies.

The Pan-Arabic corpus [18] consists of Arabic dialect data from five different regions, including United Arab Emirates (AE), Egypt (EGY), Iraq (IRQ), Palestine (PS), and Syria (SY). Each dialect set captures conversations of 100 speakers (genders balanced). In every session, two speakers complete four combined conversational recordings using lapel microphones. Four dialects – PS, IRQ, SY, EGY are used in the evaluations. The Chinese corpus [18] utilized in our study consists of four Chinese dialects (sub-languages): Mandarin (CMN),

[*] This project was funded by AFRL under contract FA8750-12-1-0188 and partially by the University of Texas at Dallas from the Distinguished University Chair in Telecommunications Engineering held by J.H.L. Hansen.

Cantonese (YU), Xiang (HSN), and Wu (WU). All data in this corpus capture spontaneous conversational noise-free speech.

For the purpose of PRSVM training and evaluation, all recordings were cut into approximately 10–12 second long segments. Training and test sets comprise non-overlapping speaker groups. The amount of training and evaluation data for the three DID sets is summarized in Table 1.

	Chinese				Arabic CTS				Pan-Arabic			
Dialect	CMN	HSN	WU	YU	GLF	IRQ	LEV	EGY	PS	IRQ	SY	EGY
Train (Hrs)	6.3	8.9	5.1	7.7	32.7	16.1	11.9	33.9	10.6	9.3	10.8	9.9
Test (Hrs)	2.2	2.9	1.7	2.6	2.0	2.3	1.6	10.1	2.8	2.7	2.5	2.6
Avg. Dur.	10.0 sec				11.3 sec				11.9 sec			

Table 1. Distribution of speech samples in Chinese and Arabic sets.

3. PRSVM SUPERVECTOR PRE-PROCESSING

In a typical PRSVM system, relative N-gram frequencies observed at the PR output for a training or test token are stacked into a supervector, each dimension being mapped to a unique N-gram. Subsequently, the dimensions are normalized by the inverse square root of the global frequencies of the corresponding N-grams [9] (*global frequency normalization, GFN*); the global frequency is estimated from the training tokens across classes. In the next step, a logarithmic *squashing function* $g(x) = \log(x) + 1$ is applied. The purpose of GFN combined with squashing is to equalize the typicality of N-grams across all classes and limit the probability that some supervector dimension would dominate the inner product in the SVM kernel [10].

To reduce the computational overhead, only frequent N-grams are usually included in the supervector and utilized in the subsequent SVM modeling [4] (*frequency-based supervector dimensionality reduction*). Finally, before entering SVM, the supervectors can go through an *adaptation* stage. Our study follows [19] where a so called universal N-gram language model is MAP adapted towards the token's supervector. The adaptation helps reduce sparseness of the supervectors caused by the limited number of N-grams occurring in short tokens [9]. In this text, the universal language model is called a universal background supervector (UBS) for the resemblance with the universal background model (UBM) in the GMM-UBM paradigm.

3.1. Pre-Squashing Normalizations

In the pre-squashing stage, the impact of the following normalizations is investigated.

Within-dimension mean/variance norm (WD MVN): for each N-gram, mean and variance of its relative frequency (no squashing) is estimated from training tokens across all classes. During the supervector extraction for SVM modeling and classification, the stored 'train' means and variances are used for dimension-wise MVN. This can be viewed as alternative or complementary norm to GFN as discussed above, only here rather the means and variance of across-class priors are equalized.

Across-dimension mean/variance norm (AD MVN): mean and variance are estimated across the elements of an individual supervector, and are applied in MVN of the supervector elements. AD MVN normalizes the frequency profile seen across the supervector dimensions. In AD MVN, the variance normalization is paired with a multiplicative constant α to control the dynamics of the N-gram normalized frequencies (NNF):

$$NNF_{t,m}^{ADMVN} = \left(NNF_{t,m} - \overline{NNF_t}\right) \Big/ \sqrt{\alpha \mathrm{var}\left(NNF_t\right)} \qquad (1)$$

where t is the token index and m is the supervector dimension index. As will be discussed in Sec. 3.2, when combined with a nonlinear squashing function, α can be effectively used to shape contours of the frequency distributions.

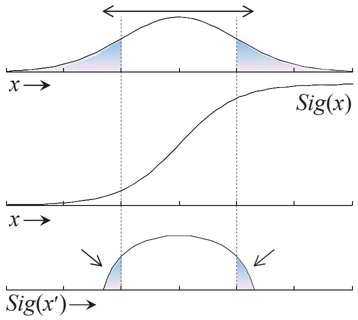

Fig. 1. Transformation of N-gram frequency distribution by sigmoid squashing function. The rate of expansion or compression (contour rounding) of distribution tails can be controlled through altering the pre-squash distribution variance by α in AD MVN. Increasing variance of the pre-squash distribution will results in stronger compression of the distribution tails due to sigmoid limiting properties.

GFN with uniform priors: instead of global N-gram frequencies extracted from the training set (*train priors*), uniform N-gram priors are substituted in the square root term. This uniform amplitude norm is introduced to allow for 'switching-off' the traditional GFN and yet numerically accommodate the log squashing function (see Sec. 4 for details).

3.2. Squashing Functions

Besides the traditional log function mentioned at the beginning of Sec. 3, the following two squashing functions are considered (i) hard limit, (ii) sigmoid. *Hard limit* was introduced in [13] based on the observation that with certain flavors of GFN applied only to non-zero supervector dimensions, the normalized frequencies tended to occupy two numerically coherent clusters (zeros vs. relatively narrow non-zero interval). This motivated the introduction of a hard limit function, where zero dimensions are kept zero and non-zero dimensions are replaced by a constant. This rather crude function merely detects the presence or absence of an N-gram in the token and ignores any prior knowledge and the actual N-gram counts in the token, yet in [13] it consistently outperformed the GFN–log setup in all PRSVMs on two Arabic DID sets. *Sigmoid*,

$$g(x) = \frac{1}{1 + e^{-x}} \qquad (2)$$

was also introduced in [13] as a squashing alternative and found to provide competitive results to log and hard limit. Combined with AD MVN, α in Eq. 1 can be used to control the variance of relative frequencies and together with Eq. 2 also their distributions (see Fig. 1). Expanding the variance will push the distribution tails into the saturated regions in the sigmoid, resulting in the compression of the tails. This may help equalize the impact of N-gram outliers (extremely frequent or extremely rare) on the subsequent SVM modeling.

3.3. Supervector Dimension Reduction

As an alternative to the traditional frequency-based dimensionality reduction (see Introduction and the overview at the beginning of Sec. 3), we propose a so called *dialect salience measure* for N-gram selection (see Table 2). For each dialect training set, a separate UBS is calculated. Here, GFN and squashing are not applied and the relative N-gram frequencies represent the estimation of dialect specific N-gram priors. In the second step, the dialect specific UBS' are compared dimension by

Dialect Salience Measure for N-gram Selection:

- Initialize dialect-specific universal background supervectors (UBS) by dialect-specific N-gram frequencies observed in the training data;
- Normalize UBS' dimensions by the total number of N-grams seen in the corresponding dialect training data (i.e., convert frequencies into prior probability estimates); the normalized dimensions are denoted *normalized N-gram frequencies* (NNF);
- In each dimension *m*, calculate cumulative distance (*CD*) between all dialect UBS pairs:

$$CD_m = \sum_{k=1}^{L}\sum_{l=1}^{L} d\left(NNF_{m,k}, NNF_{m,l}\right),\ k \neq l\ ,$$

  where *L* is the number of dialects, *m* is the UBS dimension index, *k* and *l* are dialect indices, and *d*() is the distance measure. We choose:

$$d\left(NNF_{m,k}, NNF_{m,l}\right) = \left|NNF_{m,k} - NNF_{m,l}\right|;$$

- Rank order UBS dimensions in descending order by their associated cumulative distances *CD*;
- Include only top ranking dimensions (representing most dialect-salient N-grams) in the dimension-reduced SVM supervector.

Table 2. Dialect salience measure and its application to supervector dimension selection.

dimension. A UBS dimension with high frequencies for some dialects and low frequencies for others is ranked high, a dimension with nearly uniform frequency distribution across dialects is ranked low. The ranking, a dialect salience measure, is calculated as a sum of relative frequency differences for all non-trivial dialect pairs, assuring that higher variability of frequencies across dialects will result in a higher ranking. Dimensions with higher rankings are considered more dialect salient as they represent N-grams that occur frequently in some dialects and rarely in others.

	Logarithm		Hard limit	Sigmoid			
Prior Norm	Train Priors	Uni-form	None	Train Priors	None		
MVN	Off	Off	Off	Off	Off	WD	AD
EN	18.4	18.4	17.5	16.7	17.4	15.4	14.7
CZ	23.5	23.8	22.7	22.9	22.8	22.1	21.6
HU	23.0	23.1	22.1	21.7	22.1	21.0	20.4
RU	21.9	21.9	22.0	21.4	22.2	20.0	19.6
GER	15.0	14.9	14.2	13.4	14.0	13.0	12.7
HIN	**12.8**	**13.0**	**12.0**	**11.7**	**12.1**	**11.4**	**10.5**
JAP	16.2	16.0	15.6	14.8	15.5	13.5	12.8
MAN	13.2	13.4	12.4	12.0	12.3	12.0	11.1
SPA	14.3	14.1	13.3	12.8	13.3	11.8	11.2
minEER	12.8	13.0	12.0	11.7	12.1	11.4	**10.5**
avgEER	17.6	17.6	16.9	16.4	16.9	15.6	**15.0**

Table 3. Chinese DID task; comparison of squashing and pre-squashing strategies.

4. EXPERIMENTAL SETUP AND EVALUATIONS

	Logarithm		Hard limit	Sigmoid			
Prior Norm	Train Priors	Uni-form	None	Train Priors	None		
MVN	Off	Off	Off	Off	Off	WD	AD
EN	**7.3**	**7.3**	**6.8**	**6.7**	**6.8**	**6.7**	**6.4**
CZ	16.3	16.2	15.7	15.8	15.9	15.3	15.0
HU	14.6	14.7	14.2	14.2	14.4	13.9	13.7
RU	14.8	14.9	14.5	14.3	14.5	14.1	13.8
GER	12.9	13.0	12.1	12.0	12.1	11.8	11.6
HIN	12.8	12.9	11.7	11.5	11.7	11.2	10.9
JAP	15.3	14.8	14.4	14.0	14.3	13.9	13.5
MAN	12.2	12.3	11.4	11.3	11.4	11.2	10.7
SPA	14.4	14.4	13.4	13.1	13.4	12.7	12.3
minEER	7.3	7.3	6.8	6.7	6.8	6.7	**6.4**
avgEER	13.4	13.4	12.7	12.6	12.7	12.3	**12.0**

Table 4. Arabic CTS DID task; comparison of squashing and pre-squashing strategies.

Nine BUT phone recognizers [15], English, Czech, Hungarian, Russian, German, Hindi, Japanese, Mandarin, and Spanish, are used to build dialect specific PRSVMs. All PRSVM setups utilize UBS MAP adaptation and (besides the dialect salience experiments), a frequency-based 70% dimension reduction. For a given DID dataset, a binary SVM classifier is trained for each target dialect. The training data from the remaining dialects are pooled together and used to represent the anti-class in the training procedure. In the evaluation, the task is to decide whether the token contains the target dialect or not (pick 1-vs-3). The PRSVM performance is reported by means of an equal error rate (EER) averaged per PR across four dialect-specific SVMs. Performance of individual PR-based systems is presented together with *minEER* and *avgEER* measures representing the best performance found in each category and the average performance across all PRSVMs.

Methods from Sec. 3.1 and 3.2 are evaluated for the three DID tasks in Tables 3, 4, and 5. The row *Prior Norm* denotes whether the standard GFN, *Train Priors*, GFN with uniform priors, *Uniform*, or no GFN, *None*, is applied. Note that log squashing is not presented in combination with *None* as it yielded a poor performance. Instead, *Uniform* that ignores N-gram priors and substitutes them for a uniform probability is used. Hard limit is not combined with pre-squashing norms as they would not benefit the N-gram detection. On the other hand, MAP adaptation is effective with hard limit as hard limit UBS will contain non-binary values and so will the adapted supervectors. Sigmoid is combined with GFN on/off and with within-dimension (WD) and across dimension (AD) MVN. WD/AD MVN was not combined with log squashing as MVN yields distributions centered to zero.

Among the three DID tasks, the Arabic CTS gives the most optimistic numbers, however as discussed in [13], the dialect-specific channel information is a significant contributor here. The Chinese DID task may be easier than the Pan-Arabic as the Chinese dialects here are in fact sub-languages with a shared grammar and written form, but completely different spoken form (up to the point of being unintelligible to the native speaker of one of the dialects).

Considering the baseline performance in the first column of the tables, the *avgEER* is absolutely reduced by 2.9 % for the Pan-Arabic corpus and by 3.8% for CTS compared to [13]. This can be attributed to the effects of MAP adaptation and frequency-based dimension reduction which were not employed in [13]. Compared to [14], where the best performance of a single PRSVM on the CTS corpora was 9.53 %, our best baseline PRSVM (*EN*) provides an absolute improvement of 2.23 % EER (Table 4).

The overall trends seen in the Tables can be summarized as follows. Standard GFN, *Train Priors* and GFN utilizing uniform priors, *Uniform*, combined with log squashing, provide almost identical *avgEER* on all three DID tasks. With one exception (*RU*, Chinese), hard limit al-

	Logarithm		Hard limit	Sigmoid			
Prior Norm	Train Priors	Uniform	None	Train Priors	None		
MVN	Off	Off	Off	Off	Off	WD	AD
EN	30.9	31.0	29.6	29.3	29.7	29.1	29.4
CZ	33.0	32.8	32.4	32.1	32.2	32.3	31.6
HU	33.4	33.2	33.0	33.0	33.0	32.5	32.3
RU	34.0	34.1	33.3	33.3	33.5	32.9	32.8
GER	30.0	29.9	29.3	28.8	29.4	28.1	27.8
HIN	**27.8**	**27.9**	**27.1**	**27.1**	**27.2**	**26.5**	**26.3**
JAP	31.5	31.4	31.0	30.8	30.9	29.7	29.6
MAN	28.4	28.2	27.9	27.9	28.0	26.8	26.7
SPA	30.1	29.8	28.9	28.4	28.7	27.5	27.7
minEER	27.8	27.9	27.1	27.1	27.2	26.5	**26.3**
avgEER	31.0	30.9	30.3	30.1	30.3	29.5	**29.3**

Table 5. Pan-Arabic DID task; comparison of squashing and pre-squashing strategies.

ways provides reduced EER compared to the log squashing setups. On average, GFN–sigmoid (*Train Priors*) further reduces the EER of hard limit in all three DID tasks. Sigmoid without GFN and MVN performs somewhat worse than with GFN employed. Combining WD MVD and sigmoid provides on average better EER than sigmoid without MVD or any non-sigmoid setup. The combination of AD MVD–sigmoid yields the lowest EERs out of all setups on all three DID tasks (α was experimentally set to 0.16 and kept fixed in all systems). The absolute *avgEER* reduction over the GFN–log baseline ranges from 2.6 % to 1.7 %, which is a moderate improvement, yet the size of the corpora utilized in the experiments, and the consistency of the trends across various PRSVMs suggest that this improvement is significant.

Since the *Uniform* norm is found successful in the evaluations, the authors assume that the log squashing function simply requires the relative N-gram frequencies appearing in a certain 'typical' range and that the non-uniformity of the N-gram priors is probably not exploited to such an extent as expected. Our preliminary results obtained close to the submission deadline suggest that also the combination *Uniform*–sigmoid yields lower EERs than *Train Priors*, even though the improvement does not reach the values of the sigmoid AD MVN setup.

Finally, the hypothesis that the most frequently observed N-grams in the pooled dialect training data are also the most discriminative ones is tested by confronting the frequency-based and dialect salience based N-gram selection. The results are demonstrated on the Chinese DID task in Figures 2 and 3. Figure 2 shows the average N-gram overlap for various rates of reduction. The trend is averaged across all nine phone recognition setups. It can be seen that the overlap decreases with the reduction rate, however, even for a 90 % dimension reduction, nearly 85 % of the N-grams are shared by the two methods. Also, it can be seen that this trend is quite independent of the phone recognizer used, in spite of the fact that different phone recognizers utilize completely different N-gram lexicons. Figure 3 compares the PRSVM performance as a function of the reduction rate for the two methods. It can be seen that the dialect-salient methods provides more discriminative supervectors for reduction rates over 60 %, and that at 70 %, the average EER is actually slightly improved compared to using the complete supervector. The results here suggest there is a high correlation between the N-gram frequency and its dialect salience.

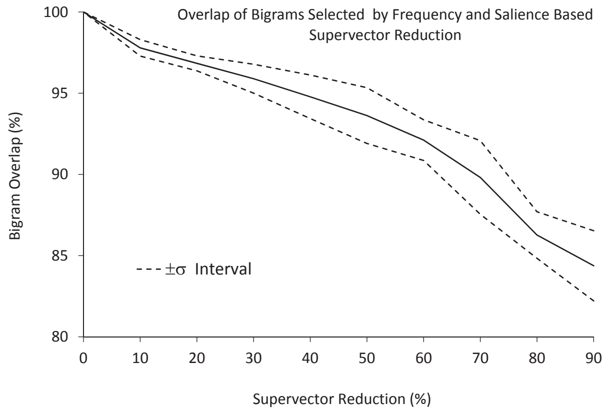

Fig. 2. Overlap between N-grams selected by frequency and dialect-salience based methods; Chinese corpus; solid line represents average N-gram overlap seen across 9 phone recognizer systems, dashed lines denote ±1 standard deviation interval.

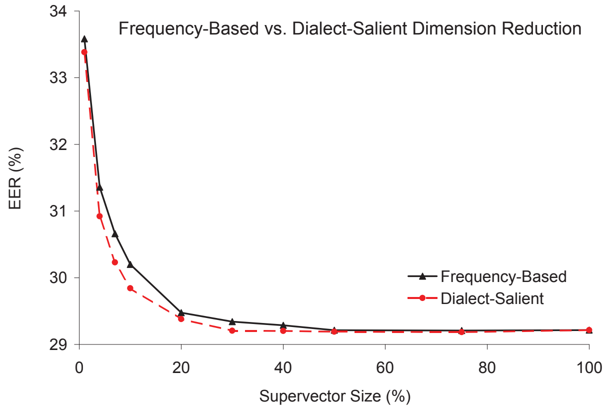

Fig. 3. Impact of dimension reduction on DID performance. Results averaged across 9 phone recognizer systems

5. RELATION TO PRIOR WORK

The study focused on the supervector pre-processing strategies for PRSVM based dialect identification. The concepts presented here built upon paradigms pioneered and practiced in [4, 9, 10, 14, 19]. Compared to those studies, our focus is on alternative frequency normalizations, squashing functions, and dimension selection. In the last case, we investigated the correlation between the frequency of N-gram occurrences and their dialect salience, which, to our best knowledge, has not been previously thoroughly considered. Alternative techniques to term selection and feature reduction can be found in [20]. Squashing functions inspired by other types of activation functions in neural networks (e.g., [21]) may further benefit PRSVM DID.

6. CONCLUSIONS

We have investigated the efficiency of within dimension and across dimension mean-variance normalization, global frequency normalization with uniform N-gram prior probabilities, alternative squashing functions hard limit and sigmoid, and dialect-salience based dimension reduction. Evaluations on three different DID tasks suggest that sigmoid squashing function combined with across dimension mean-variance normalization can provide consistent performance improvements over the traditional pre-processing strategy. In the second part of the study, a newly proposed dialect salience measure was applied in the analysis of the most frequent N-grams and has shown a strong correlation between N-gram frequency and dialect salience.

7. REFERENCES

[1] P. A. Torres-Carrasquillo, E. Singer, M. A. Kohler, R. J. Greene, D. A. Reynolds, and J. R. Deller Jr., “Approaches to language identification using Gaussian mixture models and shifted delta cepstral features,” in *INTERSPEECH'02*, Denver, Colorado, 2002, pp. 89–92.

[2] P.A. Torres-Carrasquillo, E. Singer, W. M. Campbell, T. Gleasonand A. McCree, D. A. Reynolds, F. Richardson, W. Shen, and D. E. Sturim, “The MITLL NIST LRE 2007 language recognition system,” in *INTERSPEECH'08*, Brisbane, Australia, 2008, pp. 719–722.

[3] E. Singer, P. Torres-Carrasquillo, D.A. Reynolds, A. McCree, F. Richardson, N. Dehak, and D. Sturim, “The MITLL NIST LRE 2011 language recognition system,” in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, June 2012, pp. 209–215.

[4] A. Stolcke, M. Akbacak, L. Ferrer, S. Kajarekar, C. Richey, N. Scheffer, and E. Shriberg, “Improving language recognition with multilingual phone recognition and speaker adaptation transforms,” in *Odyssey'2010*, Brno, Czech Republic, 2010.

[5] M.A. Zissman, “Comparison of four approaches to automatic language identification of telephone speech,” *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 1, pp. 31–44, Jan. 1996.

[6] M.A. Zissman, T.P. Gleason, D.M. Rekart, and B.L. Losiewicz, “Automatic dialect identification of extemporaneous conversational, Latin American Spanish speech,” in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, Atlanta, Georgia, May 1996, vol. 2, pp. 777 –780.

[7] H. Suo, M. Li, T. Liu, et al., “The design of backend classifiers in PPRLM system for language identification,” in *Proc. International Conference on Natural Computation*, Haikou, China, June 2007, p. 678682.

[8] W. Shen, W. Campbell, T. Gleason, D. Reynolds, and E. Singer, “Experiments with lattice-based PPRLM language identification,” in *IEEE Odyssey'06: Speaker and Language Recognition Workshop, 2006.*, San Juan, Puerto Rico, June 2006, pp. 1 –6.

[9] W. M. Campbell, J. P. Campbell, D. A. Reynolds, D. A. Jones, and T. R. Leek, “Phonetic speaker recognition with support vector machines,” in *Advances in Neural Information Processing Systems*. 2004, pp. 1377–1384, MIT Press.

[10] W.M. Campbell, F. Richardson, and D.A. Reynolds, “Language recognition with word lattices and support vector machines,” in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, april 2007, vol. 4, pp. IV–989 –IV–992.

[11] NIST, “Speaker recognition evaluation (SRE),” Dec. 2012.

[12] NIST, “Language recognition evaluation (LRE),” Dec. 2011.

[13] H. Bořil, A. Sangwan, and J. H. L. Hansen, “Arabic dialect identification – ‘Is the secret in the silence?’ and other observations,” in *INTERSPEECH 2012*, Portland, Oregon, 2012.

[14] M. Akbacak, D. Vergyri, A. Stolcke, N. Scheffer, , and A. Mandal, “Effective Arabic dialect classification using diverse phonotactic models,” in *INTERSPEECH'11*, Florence, Italy, 2011.

[15] P. Schwarz, *Phoneme recognition based on long temporal context*, Ph.D. thesis, Brno University of Technology, Czech Republic, 2009.

[16] F. Biadsy, J. Hirschberg, and N. Habash, “Spoken Arabic dialect identification using phonotactic modeling,” in *Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages*, Athens, Greece, 2009, pp. 53–61.

[17] F. Biadsy, J. Hirschberg, and D. P. W. Ellis, “Dialect and accent recognition using phonetic-segmentation supervectors,” in *INTERSPEECH'11*, Florence, Italy, 2011, pp. 745–748.

[18] Y. Lei and J.H.L. Hansen, “Dialect classification via text-independent training and testing for Arabic, Spanish, and Chinese,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 85 –96, jan. 2011.

[19] B. Xu, Y. Song, and L.R. Dai, “The adaptation schemes in PR-SVM based language recognition,” in *Chinese Spoken Language Processing, 2008. ISCSLP '08. 6th International Symposium on*, dec. 2008, pp. 1 –4.

[20] H. Li, B. Ma, and C.-H. Lee, “A vector space modeling approach to spoken language identification,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 1, pp. 271–284, Jan.

[21] W.E. Wong, Y. Shi, Y. Qi, and R. Golden, “Using an RBF neural network to locate program bugs,” in *Software Reliability Engineering, 2008. ISSRE 2008. 19th International Symposium on*, Nov., pp. 27–36.