ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 7, Issue 3, September 2017

A Novel Hybrid Feature Set for Automatic Classification of Hindi-and Mandarin-Accented English

Anas J. Abumunshar and Enas J. Abumunshar

***Abstract*—Automatic accent recognition has attracted much attention in speech applications designed for security, business, and a plethora of other fields. In the US, a country with a growing number of immigrants, performance of Automatic Speech Recognizing (ASR) systems is unfortunately significantly degraded by foreign accents. Accordingly, comprehensive speech and accent analysis has been administered within the last decade focused mainly on acoustic waveform and physical speech production including prosody, phonetics, and human perception. It has been suggested that phonological features offer deep analysis in distinguishing between the accents, while breadth and depth analysis is performed on characterizing different articulatory aspects, such as vowel, roundness, duration, etc.**

In this study, a comprehensive analysis is presented to improve the classification accuracy among two major groups of accented speakers in the United States, Hindi- and Mandarin- accented English. The analysis is based on vowel articulatory features. A corpus of around six hundred /hVd/ speech utterances, for fifty-four male and female speakers, is achieved for robust modelling and analysis. Standard cepstral, 13-MFCC features are extracted and a sufficient support vector machine modeling is performed on the set date. Classification accuracy ranges between 72-80%. Later, novel, hybrid acoustical-phonological features, namely State Duration, Voicing Strength, Homogeneity, and Gradient, are extracted to characterize articulatory aspects and improve classification accuracy among the accented-groups. Furthermore, a cascaded linear discriminant model is developed. The new accuracy for our speaker- and phoneme-independent hybrid model with the novel feature set, is 86-91%, with an average improvement of 9% in accuracy among the standard MFCC/SVM model.

***Index Terms*—accent recognition, accented English, automatic speech recognition, and phonological features.**

I. INTRODUCTION

Increasing immigration to English-speaking countries, an international collaboration of ideas, and globalization in business have created an immediate need for well-performance accent recognizers for advanced ASR systems[1], pronunciation modeling and scoring, speaker recognition [2], and language learning [3][4][5]. In addition, familiarity with an accent has been noticed to affect the intelligibility and comprehensibility of any speech in a negative or positive way[6]. The United States immigrant population is growing exponentially. Mandarin and Indian accents are undoubtedly among the largest groups of immigrants in the US [8][9]. Current English speech recognition systems in the U.S are designed to deal with standard American English and are unable to perform as well in foreign-accented conversations, due mainly to ignoring the importance of 'imported' accents, or in other words, the effect of native language L1 on the spoken language L2.

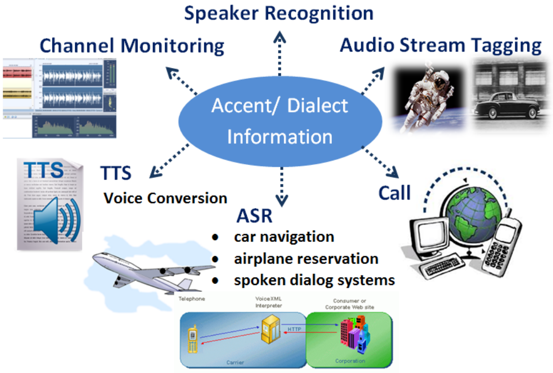

Fig. 1: Automatic accent recognition applications

Thus, accent becomes a crucial factor in speech technology in various areas of the United States, including services based on user-agent voice commands, targeted advertisements, guidance, language education, specifically 'interactive voice response' (IVR), systems entertainment, forensics, security, and areas of intelligence in general[7](Fig. **1**).Consequently, comprehensive research work focused on accent recognition has recently been performed.

Accent recognition based on the most discriminating characteristics including intonation pattern, phone duration, and formant behavior are employed in [10][11][12][13]. In [12]Hansen used voice onset time detection for unvoiced stops (/p/, /t/, /k/) in accent classification. Similarly, accent classification based on modelling techniques such as Gaussian mixture models (GMMs), support vector machines (SVMs), and Hidden Markov models (HMMs) are developed as well to learn accent characteristics and adopt good classification performance. Early work employs GMMs and HMMs techniques [7] [8] [9] [10]. After HMMs, GMMs have become predominant especially, in text independent applications [11] [13] [14]. Later on, the support vector machine (SVM) demonstrated high classification accuracy rates in gene expression classification and multi-class protein-fold recognition

DOI:10.17605/OSF.IO/VMB9N

[15]–[16]. Furthermore, feature extraction and signal processing such as cepstral features, prosodic features (phoneme segment duration, pitch and pitch slope), and categorical features about contextual information have been presented in accent recognition[21]. In [14][15], the authors used articulatory features (AFs) such as nasality, voice, and roundness to represent the acoustic signal in a compact manner. Also, Kirchhoff [14], showed that AF based ASR systems outperform HMM based ASR systems in certain adversative conditions.

On the other hand, some researchers have included the knowledge of grammar or vocabulary in order to simplify the procedures of phoneme recognition and to improve the accuracy of accent recognition [20].Intonation and rhythm are also used as an indication to a mother tongue that varies from the spoken L2. [16][17][18]. At the rhythm level, different parameters have been proposed to proof or deny the existence of rhythm classes [19][20][21]. Some of these measurements have resulted in varied success [22].In addition, training techniques are developed to build accent-specified acoustic models and adaptation strategies are designed to contain accent-related variations in pronunciation dictionaries [23][24][25][26]. The latter technique requires a linguistic background on foreign accents that might be a challenge. The recent NIST language recognition evaluation gathering contained accent and dialect verification by depending on a large amount of accented speech [27].

The total variability model or *i-vector* approach originally used for speaker recognition [28] has also been successfully applied to accent recognition [29][30]. These studies indicate that i-vector approach is promising for dialect and foreign accent recognition tasks. However, due to subtle linguistic variations, highly accurate i-vector is attributed to availability of a massive number of spoken utterances for training purposes. This challenge makes i-vector receive less attention in accent than language or speaker recognition. [31].

To this end, it is suggested that phonological features, that describe the articulatory characteristics of the speech, would offer deep analysis in distinguishing between the accents [32][9].Generally, accents lead to phonetic variations and overlapping. This overlapping yields fuzziness between the phoneme boundaries and classes [27][2]. Therefore, understanding these articulator variations would be beneficial in many research trials directed toward phonemes in recognizing the accent [33][34][35][36]. Most of them focused on vowels [33], since vowels are one of the easiest articulatory features to classify [37].

Encouraged by these findings, an in-depth feature extraction analysis that contributes to a better understanding of the differences between Hindi and Mandarin accented English is performed. Then, based on these new findings, a hybrid modelling approach is developed as well to improve classification accuracy. The remainder of this paper is organized as follows: In section II, the foreign- accented English speech corpus is described. In this study, data of Hindi and Mandarin accented speech utterances are used for development and analysis. In section III, a brief review of the novel feature sets with respect to speech technology is presented. The SVM modeling system is also presented. In section IV, the proposed accented model based on novel feature extraction and the cascaded hybrid Discriminant model is introduced. An in-depth comparative analysis of Hindi- vs Mandarin- accented English speech is also presented and discussed. Section V concludes this work.

II. HINDI- AND MANDARIN-ACCENTED ENGLISH SPEECH CORPUS

The foreign-accented English speech corpus of this study consists of utterances spoken by native speakers of Hindi and Mandarin. Participants of this study were students from the University of Bridgeport. Thirty-two Mandarin speakers (20 male and 12 female) and twenty two Hindi speakers (13 male and 9 female) participated in this study. They are all adult native speakers of Mandarin Chinese and Hindi, with no reported history of speech and/or hearing problems. All speakers are age matched with a range from 22–25 years, and their length of residence in the U.S. ranges between 2 months and 2 years. All participants filled out a questionnaire. Each participant was informed about the study and the procedures before recordings began.

The speech materials include eleven /hVd/ words in English: Heed, Hid, Had, Hawed, Hayed, Head, Hod, Hoed, Hood, Hud and Whod. Each /hVd/ word was inserted into a short phrase “Say /hVd/ again”. The sentences were randomized. Each volunteer repeated the sentences three times. Volunteers were asked to read the sentences into the mask filter using their normal voice. The recordings of all speech samples took place at the same laboratory for all subjects to maintain uniformity. A cakewalk professional microphone and accompanying Sonar 6 LE software was used. The microphone was placed at approximately 20 cm away from the speakers and a mask filter is placed in front of the microphone. All samples were recorded with a sampling frequency of 44.1 KHz.

For each speaker, each of the three repetitions for each word was examined for the least amount of noise and distortion. When this was determined, the /hVd/ word was cut from the original sentence and copied into another file for further analysis. [30]– [31].The speakers then stated the same sentences. The vowel word occurs in the same position in the sentence to cancel any tonal and intonation effects such as energy or duration that is induced by the intonation and tone.

III. ACCENT MODELING BASED ON LIKELIHOOD SCORE

Many different approaches were investigated for incorporating articulatory and phonetic features in speech and accent recognition. Zissman et. al. [38] combined phoneme recognition with dialect-dependent language modelling to discriminate dialects of Latin American Spanish. The linear dynamic model [39]and dynamic Bayesian network [40]are investigated and show high accuracies. Also, artificial neural networks (ANNs) have shown high accuracies for classifying AFs[15]. For smaller tasks, SVM offers favourable features like a high generalization of data with high-dimensional distribution. It has been applied to classification of AFs [41]and reported accuracies ranging from 79-95%. In addition, SVM has been used for the automatic classification of multi-level AF features and it produced successful results [42].

The major features that are widely used in depicting the different acoustic and physical traits of voices are LPCC (Linear Predictive Cepstral Coefficients), LPC, and MFCC (Mel Frequency Cepstral Coefficients). Among these three, MFCC offers the capability to fully capture the characteristics of the channel spectrum and simulate the human's auditory function, whose approximation of speech is linearly spaced in frequency scale [28].MFCC has been recently used for speech, speaker and accent recognition[21] [9]. Sangwan [9] used MFCC/HMM for Mandarin Accented speaker detection. He achieved 71% frame level accuracy for vowel hybrid features.

A. *Cepstral Features*

Our data set consists of594 recorded speech samples that are related to 54 Hindi and Mandarin accented speakers. The utterances provide sufficient samples for robust modelling and analysis. The utterances are divided into two groups based on their gender, male and female, respectively. The male group consists of 13Hindi-accented male (HM) speakers and 20 Mandarin-accented male (MM) speakers. The female group contains 9 Hindi-accented female (HF) speakers and 12 Mandarin-accented female (MF) speakers. Our extraction scheme is based on the SVM Framework. In our extraction system, all speech utterances are pre-emphasized. Afterthat, 13 MFCCs (melfrequencycepstralcoefficients) are extracted. The features are used to train the SVM (support vector machine) based classification system. The performance of the feature detection system on the Hindi- and Mandarin accented speech corpus (test only) in terms of frame level accuracy is shown inTable I Table 1.

From the achieved accuracies in Table 1, it is noted that classification accuracy of greater than 80% is achieved for mandarin accented speakers. This accuracy is larger when compared to the accuracy (71%) that is reported in [9] for vowel features except where HMM is used instead of SVM. This proves the sufficiency of SVM in folded multi-class recognition.

B. *Proposed Hybrid Features*

While the modeling techniques and features provide good classification accuracies, they do not offer a mature understanding of the major differences between the accents. Cepstral coefficients, for instance, are sensitive to auxiliary information inherent in the speech signal such as pitch, energy, and rate-of-speech. Emerging from the origin of the accents that are embedded in production differences, it would be beneficial to analyze and compare the major articulatory characteristics of accents. Therefore, it is suggested that a phonological features based framework will offer deep analysis in distinguishing between the accents[32][9].Furthermore, it is recommended to improve accent identification by following one of these; i) create a vector of hybrid features or combine phonological features with more standard approaches (such as cepstral features along with Gaussian mixture model and support vector machines); ii) specific acoustic models and pronunciation dictionaries can be designed for automatic speech recognition to reduce error rate on accented-speech; iii) benefit can be derived from speech synthesis. The hybrid features were firstly proposed in [43]. In [9] the author used the phonological features vowel, height, and frontness to describe vowel characteristics.

In this section, novel hybrid, phonological, and acoustical features are presented. These features tend to increase the classification accuracy among the Hindi- and Mandarin-accented English groups. The extraction of these features is based on physical traits and difference in pronunciation. Specifically, Indian-accented English tends to have more peculiarities, whereas, Mandarin accent is described as a homogenous, stress-timed accent and its syllables consist maximally of an initial consonant such as a glide, a vowel, a final, and tone [44][45].

TABLE I. CLASSIFICATION ACCURACY (%)

		MFCC/SVM
Male	**HM**	72.12
	MM	84
Female	**HF**	80.8
	MF	80.24

Before we move to discuss in more detail regarding the features, it is suggested by [46][47] that the representation of a speech carries advantages for certain applications. Accordingly, R.W. Schafer et.al.[46] proposed several representation techniques for various speech processing applications. M. Cannon [48] used

DOI:10.17605/OSF.IO/VMB9N

normalized autocorrelation functions for analysis. They are independent of speech amplitude and relatively insensitive to small fluctuations in speech waveforms, including peak clipping.

In this work, we represent the speech utterances through their autocorrelation. The proposed hybrid features consist of acoustical and phonetical features. The phonological features characterize *state duration*, *homogeneity*, inv*oicing strength* that are embedded in phonetics. The acoustical features are *gradient* and *cepstral features*. *State duration*, that reflects the duration spent in the dental place of the articulation, has been exploited by Sangwan et. al. [8][9] in accent analysis of Chinese speakers of English. Particularly, the author observed that native mandarin Chinese speakers gain higher proficiency in a) the duration feature of the articulation among the transitional feature, and b) vowel articulations among consonant articulation. Yet, the work has not been performed from the accent identification perspective.

In this study, state duration is utilized for accent classification between the Hindi-and Mandarin-accented English utterances. The *State Duration* feature is determined based on the spectral density of the autocorrelation sequence, and a single value feature is obtained for each utterance.

Voicing strength is modeled in [49], through hidden markov models (HMMs), as a sequence of stationary random regimes. Here, we followed a different approach in which the strength of the first major peak, after the center peak in the normalized autocorrelation sequence, is used as an indication of voicing strength. This approach is inspired by [50] in determination of voiced, non-voiced frames.

Also, *Homogeneity* of the speech is examined. Utterance or segment with less peculiarities is expected to carry a high level of homogeneity. Tuerk et.al [51], examined the homogeneity in feature vectors. He found that successive feature vectors are more dependent (correlated) for slow speech than for fast speech. Here, we applied the homogeneity metric on our data set. Each speech utterance is divided into three equal-duration frames. After that, the cross-similarity between the frames is determined, resulting in a vector of three values. In [9][52][53], the match between different spoken words, or model and listener data, is measured through correlation sequence. Similarly, the cross-correlation between the different portions of the spoken utterance is determined and utilized to indicate match/homogeneity.

Cepstral features are extracted as mentioned in the previous section. A cepstral features vector is extracted for each phoneme segment. The vector combines 12 MFCC features and log energy. Finally, the rate of change in the amplitude of the speech signal is utilized as an acoustical feature.

Finally, a *gradient* sequence is extracted for each speech utterance. The gradient maps the rate of change in the amplitude of the speech signal over time. To reduce the redundancy, the gradient sequence is represented by a single value. This value represents the highest rate of change among the sequence.

After all, the extracted features for each phoneme segment in the training and test utterances are gathered and combined into a single hybrid-features vector. The feature vector combines; 12 MFCC features and log energy, three *homogeneity* features, a *Voicing Strength* feature, a *State Duration* feature, and a *Gradient* feature. Then, the data is sent to the SVM model for classification. SVM is selected since it is found to be sufficient in class-dependent accent discrimination [54]. A flow chart clarifying hybrid-features/SVM model is shown in

Fig. **2**.

The classification accuracy of this mode is also determined and provided in Table. In addition, a comparison between the proposed model and the standard MFCC/SVM is also presented. As shown, the proposed hybrid features have improved the classification accuracy by; 8% MF, 7% HF, 6% MM and 1% HM, i.e. over 5% on average. This is a promising result especially when dealing with highly correlated accents that came from close geographical origins [55].

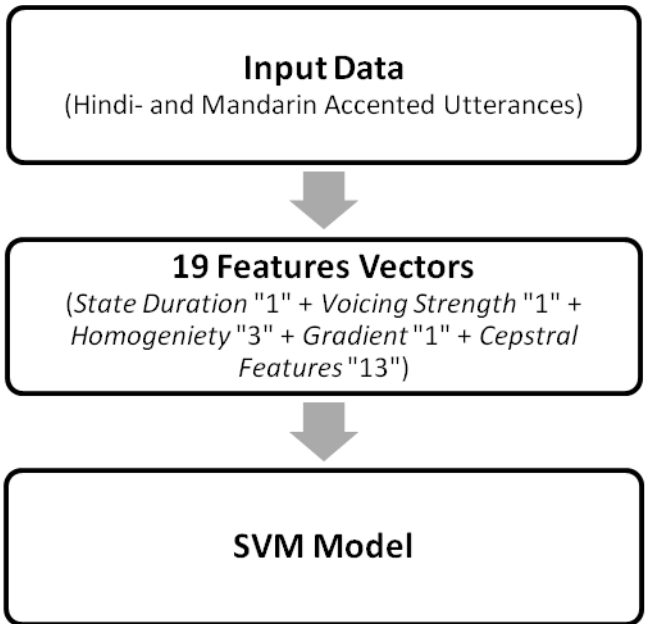
Input Data
(Hindi- and Mandarin Accented Utterances)
19 Features Vectors
(*State Duration* "1" + *Voicing Strength* "1" + *Homogeniety* "3" + *Gradient* "1" + *Cepstral Features* "13")
SVM Model

Fig. 2: Flow chart of the Hybrid features/SVM Model

IV. HYBRID CASCADED DISCRIMINANT MODEL

In this section, a hybrid classifier is developed to further improve the classification accuracy between

DOI:10.17605/OSF.IO/VMB9N

Hindi- and Mandarin- accented groups. The proposed model combines between SVM and linear discriminant function (LDA) in a cascaded multi-layer perception. Cascaded multi-layer perceptions (MLP) for speech segmentation and subsequent foreign speaker accent classification is proposed by Blackburn et. al. [56]. Similarly, Miller and Trischitta [57] applied LDA to the discrimination of different American English dialects. Kumpf et.al.[58]proposed a more flexible approach. This approach models the phoneme-dependent accent information, train phoneme class-dependent LDA models, processing and scoring utterances progressively.

The proposed hybrid Discriminant model is depicted in Fig. **3**. Multiple linear discriminant functions (LDAs) are cascaded. In each layer, a single feature vector is pooled across the tested data, and LDA is used to eliminate data with very low variance. An optimized feature sub-set is selected for the training of each LDA model to maximize accent discrimination performance. At the end of the sequence, SVM is used to classify the remaining ambiguous speech utterances.

The train and test groups for Mandarin male speakers consisted of 10 and 10 distinct speakers, respectively. The train and test groups for Hindi male speakers consist of 6 and 7 distinct speakers, respectively. Similarly, the train and test groups for Mandarin female speakers consisted of 6 and 6 distinct individuals, respectively. Moreover, for Hindi female speakers the train and test groups consist of 4 and 5 distinct speakers, respectively. In other words, the train and test sets consist of 286 and 308 samples of isolated word utterances. The proposed classifier is speaker-independent and phoneme-independent.

Input data are sent to train the model and find the optimum threshold levels of the LDAs. LDA1 is specified to classify speech utterances based on *State Duration* feature; more specifically, the classification of Mandarin utterances with timely-stressed tone. Since vowel is more sensitive to rate of speech (ROS) [59], the timely-stressed aspect appears clearly and a low class is given for Mandarin utterances.

TABLE II. CLASSIFICATION ACCURACY (%)

	MFCC/SVM	Hybrid Features/SVM
HM	72.12	73.54
MM	84	90.02
HF	80.8	87.83
MF	80.24	87.9

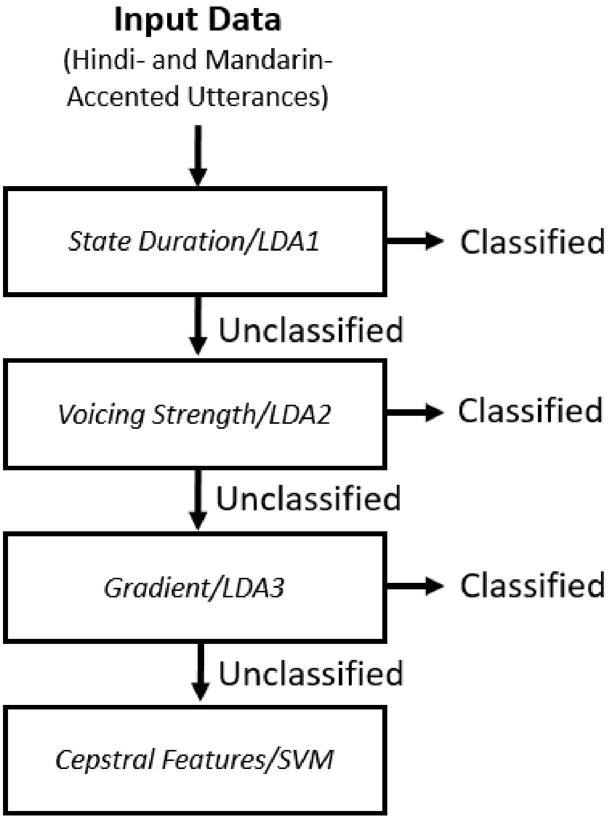

Fig. 3: Block diagram of the hybrid cascaded discriminant model

Unclassified speech utterances from LDA1proceed to the next LDA. In LDA2,it is found that low and high classes of the *Voicing Strength* are occupied by Hindi speakers. This confirms that Hindi speakers are inhomogeneous [44].As noted, the *Homogeneity* feature, that describes the homogeneity in the speech utterance itself, is eliminated from the hybrid model. This is simply because the feature-values were mixed and LDA performs poorly in this specific case.

After this, testing data are sent to the model. The performance of each sub-classifier in the hybrid model is determined based on the confusion matrixes, see Table and Table, for male and female speakers respectively. The confusion matrix in Table represents the number of classified and unclassified male utterances from both Hindi- and Mandarin-accented groups (HM and MM), whereas, in Table, female groups (HF and MF) are presented. The overall classification accuracy of the hybrid model is calculated based on the following formula:

$$\begin{aligned} Accuracy = {} & (number\ of\ true\ positives + number\ of\ true\ negatives)/(number\ of\ true\ positives \\ & + number\ of\ false\ positives + number\ of\ true\ negatives + number\ of\ false\ negatives) \end{aligned}$$

TABLE III.. HYBRID-MODEL CONFUSION MATRIX: MALE SPEAKERS

			Predicted Class	
			HM	**MM**
Actual Class	Cascade 1	**HM**	0	0
		MM	4	100
	Cascade 2	**HM**	11	0
		MM	0	0
	Cascade 3	**HM**	9	1
		MM	1	10
	Cascade 4	**HM**	93	25
		MM	18	91

HM: Hindi Male.
MM: Mandarin Male.

Finally, Table V aims to compare the classification accuracy of the hybrid model and the previous models. As shown, our hybrid model has improved the accuracy among the standard SVM in the previous, hybrid features/SVM model, and a classification accuracy of 85.6%-90.9% is achieved. In more details, the proposed cascaded linear discriminant model has improved the classification accuracy over the SVM model by; 1% in HM, 6% in MM, 7% in HF, and 8% in MF. Moreover, with the hybrid features, it outperforms the standard MFCC/SVM by; 13.5% in HM, 3% in MM, 7% in HF, and 10.5% in MF.

TABLE IV..HYBRID-MODEL CONFUSION MATRIX: FEMALE SPEAKERS

			Predicted Class	
			HF	**MF**
Actual Class	Cascade 1	**HF**	14	5
		MF	0	65
	Cascade 2	**HF**	0	0
		MF	0	6
	Cascade 3	**HF**	22	0
		MF	1	0
	Cascade 4	**HF**	51	7
		MF	11	49

HF: Hindi Female.
MF: Mandarin Female.

TABLE V. CLASSIFICATION ACCURACY (%)

	MFCC/SVM	Hybrid Features/SVM	Hybrid Model
HM	72.12	73.54	85.6
MM	84	90.02	87
HF	80.8	87.83	87.9
MF	80.24	87.9	90.9

HM: Hindi Male.
MM: Mandarin Male.
HF: Hindi Female.
MF: Mandarin Female.

V. CONCLUSION

Comprehensive studies concerning Hindi-and Mandarin-accented English in the United States is presented. Breadth and depth analysis in accent classification are proposed to enhance the classification accuracy between Hindi- and Mandarin- accented English. Sufficient data set that consists of around six hundred recorded /hVd/utterances is prepared for robust modeling and analysis. Cepstral, 13-MFCC, features with support vector machine modeling is implemented and classification accuracy ranging from 70-84% is achieved. Then, hybrid acoustical-phonological features, *State Duration*, *Voicing Strength*, *Homogeneity*, and *Gradient*, are developed and extracted. These features have improved the classification accuracy by an average of 7% in HF, HM, MM, and 1% in HM case. Moreover, a cascaded linear discriminant model is designed to further enhance performance. As a result, 85.6%-90.9% speaker- and phoneme-independent classification accuracy is achieved. With overall improvement of

DOI:10.17605/OSF.IO/VMB9N

13.5% in HM, 3% in MM, 7% in HF, and 10.5% in MF over the standard MFCC/SVM. The work can be further extended and the accuracy can be significantly increased by eliminating ambiguity due to the speaker or phoneme contribution. Furthermore, a dimensions-reduction stage can be added to the feature vector to remove the redundancy as proposed [60][61], and/or improved training procedures can be adapted.

ACKNOWLEDGMENT

The Author would like to thank Professor Buket D. Barkana from the University of Bridgeport, CT, USA, for her support and guidance throughout this research.

REFERENCES

[1] Y. Zheng, R. Sproat, L. Gu, I. Shafran, H. Zhou, Y. Su, D. Jurafsky, R. Starr and S.-Y. Yoon, "Accent detection and speech recognition for Shanghai-accented Mandarin," Interspeech-05, p. 217–220, 2005.

[2] S. Mangayyagari, T. Islam and R. Sankar, "Enhanced speaker recognition based on intra-modal fusion and accent modeling," in Internat. Conf. on Pattern Recognition., 2008.

[3] A. Neri, C. Cucchiarini and H. Strik, "ASR-based corrective feedback on pronunciations: does it really work ?," Interspeech., 2006.

[4] S. Wei, Q. Liu and R. Wang, "Automatic Mandarin pronunciation scoring for native learners with dialect accent," Interspeech, vol. 6, 2006.

[5] B. Mak, M. Siu, M. Ng, Y. Tam, Y. Chan, K. Leung, S. Ho, F. Chong, J. Wong and J. Lo, "Plaser: pronunciation learning via automatic speech recognition," in Human Language Technology Conf., 2003.

[6] A. Ikeno and J. H. and Hansen, "Perceptual Recognition Cues in Native English Accent Variation:" Listener Accent, Perceived Accent, and Comprehension," in IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, 2006.

[7] GAO, "Border Security: Fraud Risks Complicate States Ability to Manage Diversity Visa Program," DIANE Publishing, 2007.

[8] A. Sangwan and J. Hansen, "On the use of phonological features for automatic accent analysis," in Proc. Interspeech, Brighton, UK, pp. 172–175, 2009.

[9] A. Sangwan and J. H. and Hansen, "Automatic analysis of Mandarin accented English using phonological features," Speech Communication, vol. 54, no. 1, pp. 40-54., 2012.

[10] L. Arslan and J. Hansen, "Language accent classification in American English," Speech Comm. 18 (4), p. 353–367, 1996a.

[11] L. Arslan and J. Hansen, "A study of temporal features and frequency characteristics in American English foreign accent," J. Acoust. Soc. Amer. (JASA) 102, p. 28–40, 1996b.

[12] J. Hansen, S. Gray and W. Kim, "Automatic voice onset time detection for unvoiced stops (/p/, /t/, /k/) with application to accent classification.," Speech Comm. 52 (10), p. 777–789, 2010.

[13] P.-J. Ghesquiere and D. Van Compernolle, "Flemish accent identification based on formant and duration features," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2002.

[14] K. Kirchhoff, "Robust speech recognition using articulatory information," Ph.D thesis, University of Bielefield, 1999.

[15] M. Wester, "Syllable classification using articulatory-acoustic features," in Eurospeech, Geneva, Switzerland, pp. 233-236, 2003.

[16] M. Freland-Ricard, "Organisation temporelle et rythmique chez les apprenants étrangers. Étude multilingue," Rev. Phonét. Appl., vol. 118/119, p. 61–91, 1996.

[17] P. B. d. Mareüil and B. Vieru-Dimulescu, "The contribution of prosody to the perception of foreign accent," Phonetica, vol. 63, p. 247–267, 2006.

[18] M. Jilka, "The Contribution of Intonation to the Perception of Foreign Accent," Ph.D. Thesis, University of Stuttgart, Germany, 2000.

[19] T. Arai and S. Greenberg, "The temporal properties of spoken Japanese are similar to those of English," in Proc. European Conf. on Speech Communication and Technology (Eurospeech), Rhodes, Greece, pp. 1011–1014., 1997.

[20] F. Ramus, "Rythme des langues et acquisition du langage," Ph.D. Thesis, EHESS Paris, France., 1999.

[21] E. Grabe and F. Low, "Durational variability in speech and the rhythm class hypothesis," C. Gussenhoven, N. Warner (Eds.), Papers in Laboratory Phonology, Vol. VII, Mouton de Gruyter, The Hague, p. 515–546, 2002.

[22] A. Romano, "Speech rhythm and timing: structural properties and acoustic correlates," Vols. EDK Editore, Torriana, Torriana , La dimensione temporale del parlato, EDK Editore, 2010, p. 45–75.

[23] K. Livescu and J. Glass, "Lexical modeling of non-native speech for automatic speech recognition," in Proc. Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, pp. 1683–1686., 2000.

[24] G. Silke, R. Stefan and K. Ralf, "Generating non-native pronunciation variants for lexicon adaptation," Speech Comm., vol. 42, no. 1, p. 109–123, 2004.

[25] T. Cincarek, R. Gruhn and S. Nakamura, "Proc. Interspeech.," in Speech recognition for multiple non-native accent groups with speaker-group-dependent acoustic models., Jeju Island, Korea, pp. 1509–1512., 2004.

[26] G. Bouselmi, D. Fohr, I. Illina and J.-P. Haton, "Multilingual non-native speech recognition using phonetic confusion-based acoustic model modification and graphemic constraints," in Proc. Interspeech, Pittsburgh, USA, pp. 109–112., 2006.

[27] A. Martin and A. Le, "NIST 2007 language recognition evaluation," in Proc. Odyssey-The Speaker and Language Recognition Workshop, Stellenbosch, South Africa, paper 016., 2008.

[28] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel and P. Ouellet, "Front-end factor analysis for speaker verification.," IEEE Trans. Audio, Speech Lang. Process. , vol. 19, no. 4, p. 788–798., 2011a.

[29] A. DeMarco and S. Cox, "Iterative classification of regional British accents in I-vector space.," in Machine Learning in Speech and Language Processing (MLSLP),, Portland, OR, USA,, September 14–18, pp. 1–4..

[30] M. Bahari, R. Saeidi, H. hamme and D. Leeuwen, "Accent recognition using i-vector, Gaussian mean supervector and Gaussian posterior probability supervector for spontaneous telephone speech.," in IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26–31, 2013. pp.7344–7348..

[31] N. Chen, W. Shen and J. Campbell, "A linguistically-informative approach to dialect recognition using dialect-discriminating context dependent," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, USA, March 14–19, pp. 5014–5017..

[32] O. Scharenborg, V. Wan and R. Moore, "Towards capturing fine phonetic variation in speech using articulatory features," Speech Comm., vol. 49, no. 10-11, p. 811–826, 2007.

[33] J. E. Flege, C. Schirru and Ian R.A. MacKay, "Interaction between the native and second language phonetic subsystems," Speech Communication, vol. 40, no. 4, pp. 467-491, June 2003.

[34] R. Yamada, W. Strange, J. Magnuson, J. Pruitt and W. Clarke, "The intelligibility of Japanese speakers' productions of American English /r/, /l/ and /w/, as evaluated by native speakers of American English," in Proc. Internat. Conf. on Spoken Language Processing (ICSLP), Yokohama, Japan, pp. 2023–2026., 1994.

[35] J. Flege, "The detection of French accent by American listeners," The Journal of the Acoustical Society of America, vol. 76, no. 3, p. 692–707, 1984.

[36] J. Flege and R. Port, "Cross-language phonetic interference: Arabic to English," Lang. Speech, vol. 24, no. 2, p. 125–146, 1981.

[37] O. Scharenborg, V. Wan and a. R. K. Moore, "Towards capturing fine phonetic variation in speech using articulatory features," Speech Communication, vol. 49, no. 10, pp. 811-826, 2007.

[38] M. A. Zissman and e. al., "Automatic dialect identification of extemporaneous conversational, Latin American Spanish speech," Acoustics, Speech, and Signal Processing, vol. 2, 1996.

[39] J. Frankel, "Linear dynamic models for automatic speech recognition," Ph.D. thesis, The center for Speech Technology Research, Edinburgh University, 2003.

[40] K. Livescu, J. R. Glass and a. J. A. Bilmes., "Hidden feature models for speech recognition using dynamic Bayesian networks," INTERSPEECH, 2003.

[41] A. Juneja, "Speech recognition based on phonetic features and acoustic landmarks," Ph.D. thesis, University of Maryland., 2004.

[42] K. Saenko and e. al., "Visual speech recognition with loosely synchronized feature streams," in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. Vol. 2. IEEE, Beijing, China, 2005.

[43] J. Frankel, M. Magimai-Doss, S. King, K. Livescu and O. Cetin, "Articulatory feature classifiers trained on 2000 h of telephone speech," Interspeech., 2007a.

[44] P. Avery and S. and Ehrlich, Teaching American English pronunciation, Oxford: Oxford University Press, 1992.

[45] S. R. Ramsey, The languages of China, Princeton: Princeton University Press, 1987.

[46] R. W. Schafer and L. R. Rabiner, "Digital representations of speech signals," Proceedings of the IEEE, vol. 63, no. 4, pp. 662-677, April 1975.

[47] R. W. Schafer and L. R. and Rabiner, "Parametric representations of speech," Speech Recognition Invited Paper Presented at the 1974 IEEE Symposium, pp. 99-150, 1975.

[48] M. Cannon, "A method of analysis and recognition for voiced vowels," IEEE Transactions on Audio and Electroacoustics, vol. 16, no. 2, pp. 154-158, 1968.

[49] M. Benzeghiba and e. a. ". s. r. a. s. v. A, Speech Communication, vol. 49, no. 10, pp. 763-786, 2007.

[50] T. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval," IEEE Trans. Speech Audio Process., p. 309–319, 1979.

[51] A. Tuerk and S. Young, "Modeling speaking rate using a between frame distance metric," in Proceedings of Eurospeech, vol. 1., Budapest, Hungary, pp. 419–422., 1999.

[52] K. Y. Leung, M. W. Mak, M. H. Siu and S. Y. and Kung, "Adaptive articulatory feature-based conditional pronunciation modeling for speaker verification," Speech Communication, vol. 48, no. 1, pp. 71-84., 2006.

[53] K. J. Palomäki and G. J. and Brown, "A computational model of binaural speech recognition: Role of across-frequency vs. within-frequency processing and internal noise," Speech Communication, vol. 53, no. 6, pp. 924-940, 2011.

[54] K. Kumpf, "LDA Based Modelling of Foreign Accents in Continuous Speech," in Sixth Australian International Conference on Speech Science and Technology, 1996.

[55] E. Singer, P. A. Torres-Carrasquillo, D. A. Reynolds, A. McCree, F. Richardson, N. Dehak and D. E. and Sturim, "The MITLL NIST LRE 2011 language recognition system," Odyssey, pp. 209-215, June, 2012.

[56] C. S. Blackburn, J. Vonwiller and a. R. W. King, "Automatic accent classification using artificial neural networks," Eurospeech, pp. 1241-1244, 1993..

[57] D. R. Miller and J. Trischitta, "Statistical dialect classification based on mean phonetic features," in Fourth International Conference on Spoken Language, ICSLP 96, 1996.

[58] K. Kumpf and a. R. W. King, "Foreign speaker accent classification using phoneme-dependent accent discrimination models and comparisons with human perception benchmarks," Proc. EUROSPEECH., pp. 2323-2326, 1997.

[59] H. Kuwabara, " Acoustic and perceptual properties of phonemes in continuous speech as a function of speaking rate," in Proceedings of Eurospeech, Rhodes, Greece, pp. 1003–1006., 1997.

[60] R. Haeb-Umbach and H. Ney, "Linear discriminant analysis for improved large vocabulary continuous speech recognition," in Proceedings of ICASSP, San Francisco, CA, pp. 13–16, 1992.

[61] N. Kumar and A. Andreou, "Heteroscedastic discriminant analysis and reduced rank HMMs for

DOI:10.17605/OSF.IO/VMB9N

improved speech recognition," Speech Communication, vol. 26, no. 4, p. 283–297, 1998.

AUTHOR BIOGRAPHY

Anas J. Abumunshar was born in Hebron, Palestine in 1986. He received his B.Eng. degree in Telecommunications from Mu'tah University, Jordan, in 2008, his M.Sc. degree in electrical engineering from University of Bridgeport, CT, USA, in 2012, and his Ph.D. degree in electrical engineering from The Ohio State University, OH, USA, in 2017.

During his master, he was a student researcher at the signal processing research lab at University of Bridgeport, CT, USA. From 2013-2017, he was a Graduate Student Researcher with the Electro Science Laboratory at The Ohio State University. His research interest includes phase-array systems, ultra-wideband arrays, and microwave/millimeter-wave circuit designs. He is also interested in, speech signal processing for automatic accent detection and enhanced speech recognition, and the inter-relation between sound and magnetic waves.

Enas J. Abumunshar was born in Hebron, Palestine in 1988. She received her bachelor degree in applied mathematics from Palestine Polytechnic University, Palestine, in 2010. Her master degree in mathematics from Palestine Polytechnic University, Palestine, in 2014. Her research was in the field of computational geometry, she worked on developing a solid algorithm for efficient and timely manner spatial search and query.

DOI:10.17605/OSF.IO/VMB9N