Second International Conference on Intelligent Systems for Communication, Iot and Security

Accent Classification in Speech Using the i-Vector Framework in Language Proficiency Platforms

Nargiz Nazarova
Docent, Andijan State Institute of Foreign Languages, Andijan, Uzbekistan. 19722012fcb@gmail.com.

Zilola Sattorova
Tashkent State University of Oriental Studies, Uzbekistan.
Zilola2022@list.ru,
https://orcid.org/0009-0008-8943-7677.

Dildora Xolmurodova
Namangan state institute of foreign languages, 160123 Namangan, Uzbekistan.
dildoraholmurodova85@gmail.com.

Zarnigor Toshpulatova
Teacher, Higher School of South Asian Languages and Literature, Tashkent State University of Oriental Studies, Uzbekistan.
Toshpulatovazarnigor12.05gmail.com, 0009-0005-3467-733X.

Sharustam Shamusarov
Tashkent state university of oriental studies, shamusarov@yahoo.com,
https://orcid.org/0000-0001-6604-8451.

Rano Alimardanova
Department of Pedagogy and Psychology, Termez University of Economics and Service, Termez, Uzbekistan,
rano_alimardanova@tues.uz, 0009-0002-4505-5737.

***Abstract*— Speech accentology is an important subject of language proficiency platforms as it allows the correct evaluation of pronunciation and regionalism. Proper recognition of accents in speakers assists in customization of learning experiences and enhancement of automated evaluations of language. The current accent recognition techniques tend to lack robustness when analyzing short speech segments and they vary among speakers thereby decreasing the classification accuracy. Its i-vector based traditional methods are effective in the recognition of speakers but less effective in the extra accent-specific features in short utterances. To overcome these shortcomings, this paper presents a Deep Segmental i-Vector Approach (DSiVA), which is a combination of a segmental feature extraction and deep neural network modeling. DSiVA successfully represents local accent properties through the segmentation of speech into meaningful units and the creation of i-vectors of these units, and the deep network combines this information to classification by a better means. This structure increases the resistance to speaker variation and brief utterances, which offers a more accurate accent classification system. The suggested DSiVA technique works with several language proficiency datasets to check the efficiency of the technique in differentiating accents between various speakers. According to the results of the experiment, DSiVA is more effective than traditional i-vector and baseline deep learning models because the first one is more accurate and consistent in recognising accents. The results suggest that it can support adaptive language learning systems and computer-based speech assessment systems.**

Keywords—Accent classification, i-vector, Deep Segmental i-Vector Approach, DSiVA, speech recognition, language proficiency, speaker variability.

I. Introduction

A. Background and Motivation

Accent classification is a modern language proficiency system used to deliver accurate measurements of speaking language skills and customized learning processes [1]. Automated accent identification can be applied in various forms, including the assessment of pronunciation and adapt content to the needs of each listener based on the differences of their cultural or regional speech variations [17]. The ability to differentiate accents in a reliable manner is becoming more critical in the rapidly developing sphere of online learning and automated language assessment technologies, where the enhancement of student engagement and general language abilities are of utmost importance [3]. Little is known on the peculiarities in accents, as the traditional voice recognition algorithms were aimed to perform general transcribing tasks [18]. Proper accent recognition improves computer and human perception of language patterns through the influence of phonetic and prosodic, as well as articulatory constituents of speech [5]. Frameworks that leverage i-vectors to get speaker-specific information from a small data set have won recent speaker identification competitions [2] [19]. Given the natural variety between speakers and the fact that language learners often make brief sounds, utilizing these algorithms to classify accents is much harder. One way to generate more complicated accent patterns is to use deep learning and segmental analysis together [7]. Using this method with different datasets makes sure that the model is both consistent and flexible. This paper investigates the DSiVA, a technique that integrates deep neural networks with segment-level feature extraction, to improve the accuracy and reliability of instructional accent identification systems [20]. This method seems to overpower the existing constraints and this is a major strength in the applications of the methodology in self-driving language assessments and adaptive learning systems.

B. Challenges in Accent Classification

Among the issues concerning the accent categorization, there is the poorly-documented datasets, speakers who have varying pronunciations, small speech fragments and general acoustic characteristics. Conventional recognition algorithms fail to deal better with such features and instead, they should apply powerful modeling methods.

C. Contributions of the Paper

- DSiVA is suggested to be utilized in powerful classification with the accent information on the segment level.
- It performs far better than baseline deep learning and classic i-vector approaches on several datasets [8].
- Included analysis on accent identification that looks into problems such as speaker unpredictability, speech length, and accent confusion [4].

II. Related work

Focusing on Tamil, Telugu, and Kannada speakers, the current paper will examine how the original languages of people can be automatically recognized based on the accent of English. The voice database was formed which contained both non-native and native English [9]. It made use of Gaussian Mixture Models (GMM), GMM-Universal Background Model (GMM-UBM), and i-vector models to elicit the spectral properties of the English speech which the speaker could not articulate. The proposed Regional Accent Identification by Spectral Modeling (RAISM) was found to give the best results with i-vector, which proved that spectral modeling is effective in determining the place of origin of an individual.

Recent developments in dialect categorization, such as datasets, preprocessing methodologies, feature extraction, and models, are analyzed in this article [10]. It examines the differences between deep learning (DL) approaches, such as CNN, xResNet18, and transformer-based Wav2Vec2 [21], and standard machine learning (SML) techniques, like k-NN. The proposed Accent Classification Framework Review (ACFR) indicates that k-NN is the most effective TML algorithm, CNN is suitable for short, variable datasets, and Wav2Vec2 is effective for large, balanced datasets. ACFR shows where analysis is missing right now and makes suggestions for how to improve accent recognition systems in the future.

The xkl app now features a redesigned GUI and new spectrograms for analyzing spectra. The suggested Hybrid CNN-RNN Landmark Recognizer (HCR-LR) was utilized to find vowel landmarks on its own. It was learnt using the LaMIT Italian speech database and was able to identify 74.98% of the words [11]. To discover accents from different nations, they also used a Multi-Kernel Extreme Learning Machine (MK-ELM). It outperformed prior models when utilizing MFCC and prosodic characteristics, with an accuracy rate of 84.72%. HCR-LR indicates that deep hybrid models can identify vowel landmarks.

An app for smartphones that can translate spoken languages instantly is the result of this study [6]. The system employs i-vector methods that were reimplemented using Kaldi and TensorFlow models that were trained on data from Mozilla Common Voice [12]. The latest model, which was put on Android using Chaquopy, is called the Mobile i-Vector Translation System (MiVTS). An evaluation showed that MiVTS was 81% accurate and 95.7% usable, which suggests it works well for communicating with persons who speak more than one language. MiVTS can detect what language someone is saying, even if they don't have a strong accent. This is not the same as methods that use accents.

Creating a collection of Hindi and Marathi children (ages 5–15) speaking English for this paper shows how hard it is for AI systems to recognize children who don't speak English as their first language. The suggested Children's Accent Recognition Pipeline (CARP) uses a CNN-based model to extract features such as MFCC and i-vectors [13]. It was more than 95% right. CARP outperformed other systems, making a significant difference in its ability to understand what youngsters said in English. The end-to-end pipeline provides us with a lot of valuable data and a mechanism to improve voice recognition in languages with fewer resources and for youngsters to speak to AI.

The validation of participants' nativity is the primary focus of this project, which examines the use of crowdsourcing to acquire speech data [14]. A nativeness classifier was included in the methodology for both Portuguese and English versions. It compared the proposed H-vector Nativeness Classifier (HvNC) based on speaker embedding with the i-vector and x-vector methods. HvNC was better by 8% compared to the baseline making sure that the participants could be filtered and examined accurately. HvNC automatically evaluates recordings for nativeness, which makes datasets more reliable. This helps gather high-quality data for massive AI projects.

Using a wide variety of sources—including interviews, folk songs, community recordings, and radio broadcasts—this thesis builds a recognition model for Gujarati dialects, taking into account dialectal variance. The suggested Gujarati Dialect Recognition Framework (GDRF) uses deep learning together with phoneme-based pretraining, dialectal characteristics, and transfer learning methods [15]. GDRF was substantially better at finding phonemes than baseline models. GDRF simplifies understanding speech for languages with limited resources by modeling context-aware, diversified data. This makes it simpler for individuals who speak various languages to live together, makes them feel welcome, and keeps the languages diverse.

TABLE I. Comparison of the Existing Method

Acronym	Advantages	Limitations	Purpose
RAISM	High accuracy with i-vector; effective for identifying nativity in English accents; best for Kannada speakers.	Limited to three Dravidian languages; relies heavily on spectral features.	Identify speakers' native language through regional English accents.
ACFR	Comprehensive review; identifies best TML (k-NN) and DL (Wav2Vec2) models; provides research gaps.	No experimental dataset; theoretical review only.	Guide future research in accent recognition by summarizing models and techniques.
HCR-LR	Modernized xkl tool; effective vowel landmark detection (74.98%); MK-ELM achieves 84.72% accent ID accuracy.	Recognition accuracy is not very high, as it is limited to the Italian corpus and a small dataset.	Improve vowel landmark detection and foreign accent identification.

MiVTS	Real-time language detection and translation; platform-independent; 81% accuracy; high usability (95.7 SUS).	Accuracy is not as high as some accent models because they are computationally constrained on mobile devices.	Enable automatic language detection and translation on Android devices.
CARP	Over 95% accuracy; first dataset for Hindi/Marathi children's English; end-to-end pipeline.	Limited to children (5–15 years) and only two Indian languages, the dataset remains small compared to that of adults.	Improve non-native children's speech recognition for AI systems.
HvNC	8% relative improvement over baseline; effective for crowdsourced speech verification; filters unreliable contributors.	Focuses only on nativeness, not accent or full recognition; tested on limited languages.	Ensure quality in crowdsourced speech datasets by verifying speaker nativeness.
GDRF	Handles dialect diversity; deep learning with transfer learning; improves phoneme recognition significantly.	Limited to Gujarati dialects; may need retraining for other languages.	Develop speech recognition systems tailored to vernacular Gujarati dialects.

An overview of the existing methods will be summarized in Table 1.

III. Proposed Method

A. *Overview of Deep Segmental i-Vector Approach (DSiVA)*

Deep learning and segment-level analysis of speech are two key components of the DSiVA technology that simplify accent classification. By dividing speech into smaller, more useful chunks and generating i-vectors for each of these sections, one can capture the local accent features. A deep neural network analyzes the i-vectors at the segment level, and then this information is combined across all segments. This method solves the difficulties with prior i-vector algorithms by making identification perform better for both short phrases and various speakers. This could aid language proficiency systems that need to distinguish between different accents.

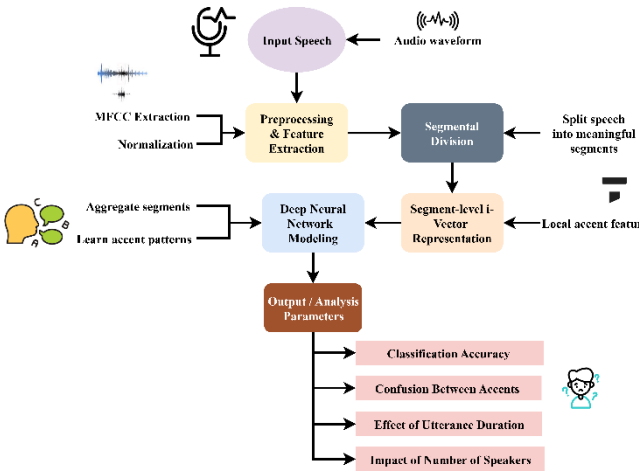

Fig. 1: DSiVA-Based Accent Classification Workflow

Fig. 1 shows the general structure of the DSiVA that was developed to classify accents. To handle raw speech, it applies preprocessing and MFCC feature extraction. Every important signal segment uses an i-vector that gathers information on the local accent. This deep neural network-based aggregation of i-vectors at the segment level accurately sorts accents. There are four basic ways to measure the results: overall classification accuracy, generalized confusion across accents, the impact of speaker count and utterance duration, and the effect of speaker count. This technique shows how DSiVA makes recognition more accurate when dealing with short utterances and different speakers.

Acoustic feature projection G_q is expressed using equation 1,

$$G_q = X_t * T_u + c_t (1)$$

Equation 1 explains that the acoustic feature projection forecast is acquired by the use of a transformation and a prejudice to the tensor of segmented speech.

In this G_q is the projected feature vector, X_t is the weight mapping matrix, T_u is the Segmented speech tensor, and c_t is the Bias offset vector.

Algorithm: DSiVA Accent Classification

$signal = preprocess(raw_{signal}) \frac{noise}{silence}$
$F = compute_{mfcc(signal)} MFCC\ features$
$if\ np.any(np.isnan(F))$:
  $F = normalize_{or_{interpolate(F)}}$
    $segments = segment(F, segment_{length})$
      $* features\ into\ segments$
$G_{q_{list}} = [\ ]$
$for\ seg\ in\ segments$:
  $if\ len(seg) < segment_{length}$:
    $seg = pad_{segment(seg, segment_{length})}$
G_q
$= X_t @ seg + c_t \quad feature\ projection$
  $if\ np.any\left(np.isnan(G_q)\right)$:
$G_q = normalize(G_q)$
$G_{q_{list}}.append(G_q)$
$i_{vectors} = \left[compute_{ivector(g) for g\ in\ G_{q_{list}} if\ i_{vec_{quality(g)}}} \geq QUALITY_{THRESHOLD}\right]$
$accent\_label$
$= model.predict(aggregate(i\_vectors))$
$return\ accent_{label}$

DSiVA algorithm performs all preprocessing of the raw speech so that it removes background noise, and extracts MFCC features. The features are segmented and projected, which generates i-vectors for each segment. The aggregated i-vectors are passed to a DNN that predicts an accent label. It filters on confidence level to determine an accurate predicted accent label.

B. *Segmental Feature Extraction*

One way to retrieve certain prosodic and phonetic parts is to fragment the audio stream into small segments and play them simultaneously. It gathers features such as MFCCs,

energy, and delta coefficients for each segment to acquire the subtle accent characteristics. Global characteristics may not notice little variations in tone and pronunciation, but localization can help with that. Segmental analysis improves accent classification by making sure that short phrases provide valuable information, even when there is noise from outside or the speakers are different.

C. *i-Vector Representation of Speech Segments*

An i-vector is a little integer that stands for each speech segment and records features that are unique to that accent. The i-vector framework sorts and compares various speakers and accents by putting them in a low-dimensional space. Segment-level i-vectors help deep neural networks perform better by reducing the number of dimensions in the input while keeping the local accent information. These i-vectors, which include phonetic and prosodic information, are an excellent starting point for grouping accents for speech that is varied or short in duration.

D. *Deep Neural Network Modeling*

The i-vectors help a convolutional neural network (CNN) learn how to gather and recall accent patterns at the segment level. A DNN may learn hierarchical features since its many completely connected layers use non-linear activations. The network tracks how various parts interact with each other in intricate ways. This makes it simpler to recognize the difference between accents and less prone to making errors. The model works for all speakers because of regularization and dropout. The DNN in the DSiVA is outstanding at dealing with variations in the duration of utterances, the features of the speaker, and environmental factors. This makes it effective for producing accurate accent predictions in real-world applications.

E. *Advantages Over Existing Methods*

DSiVA surpasses baseline deep learning and standard i-vector approaches by combining deep neural network modeling with segmental analysis. Short phrases are pretty well captured, it is more sensitive to minor accent variations and it is less volatile to speaker variability. These are few of the best it has to offer. This method works well for systems that verify language in real time since it rapidly gets local information while retaining the global context. DSiVA can provide accurate classifications even with limited labeled speech data, as it doesn't rely heavily on large datasets.

IV. Experimental Setup

A. *Datasets Used*

Incorporating speakers with a wide range of accents, the investigations draw from a variety of private and publicly available language proficiency databases [16]. The suggested method may be thoroughly evaluated using different datasets that include a range of speaker counts, speech lengths, and accent types. To ensure the test is based on real-life scenarios, elements such as age range, gender split, and recording conditions are considered. It can test DSiVA's strength in many situations and with different accents, much as language learning in real life, and automated speech evaluation systems use numerous datasets.

B. *Preprocessing and Feature Extraction*

Voice signals are reduced in noise and amplitude leveled before further processing can be done on them. It can also acquire such properties as MFCCs, delta, and delta-delta with the help of overlapping frames. Segmentation divides the data into smaller segments which are easy to analyze. The second step involves the production of i-vectors using the characteristics of the segment, which have accented aspects. This is because the spectral and temporal properties that distinguish the accent are obtained in the course of the feature extraction stage. Preprocessing entails feeding the neural network with the same input each time. Accuracy and reliability of categorization is enhanced by this pipeline.

C. *Evaluation Metrics*

The performance measures that have been adopted include: area under the curve (AUC), the confusion table, the F1-score and the total accuracy in classification. Confusion matrices explain the types of accents that tend to be confused most frequently. Accuracy, in its turn, depicts the number of correct predictions that occur in all accents. The F1-score takes into account the accuracy and recall to determine the performance of the students in classrooms which are not balanced. One of the methods to find out the effectiveness of a given model to identify the difference between items is to consider the AUC. It can contrast DSiVA with the baseline procedures on these measures to note the effect of such factors as speaker variability and utterance length on its performance.

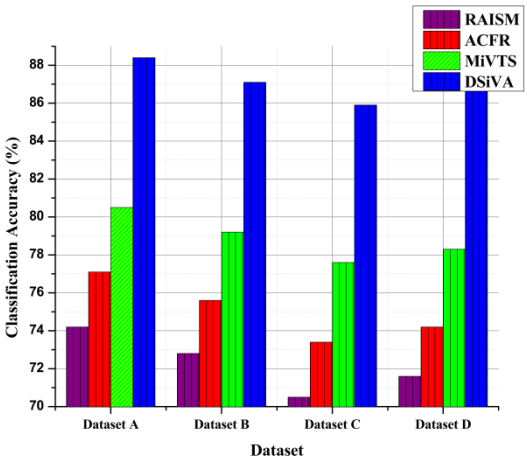

Fig. 2: Analysis of Classification Accuracy

The average accuracy of DSiVA was 87.0%, which was substantially better than the average accuracy of RAISM (72.3%), ACFR (75.1%), and MiVTS (78.9%). The segmental i-vector method ensured that accuracy remained consistent across all datasets. For instance, Fig. 2 indicates that Dataset A got 88.4%. These findings demonstrate that DSiVA is capable of capturing accent-specific information while reducing errors in datasets from individuals speaking different levels of language.

Analysis of classification accuracy B_{cs} is expressed using equation 2,

$$B_{cs} = \frac{1}{O} 1[\hat{z}_o - z_o] (2)$$

Equation 2 explains the analysis of classification accuracy using a binary indication of match, calculating the mean correctness across the evaluation index produces an invariant scalar fidelity score for label cardinality.

In this B_{cs} is the Overall classification accuracy, O is the Count of evaluated utterances, $\hat{z}_o$ is the predicted

accent label for item, z_o is the Reference accent label for the item, and $1[.]$ Is the Indicator mapping.

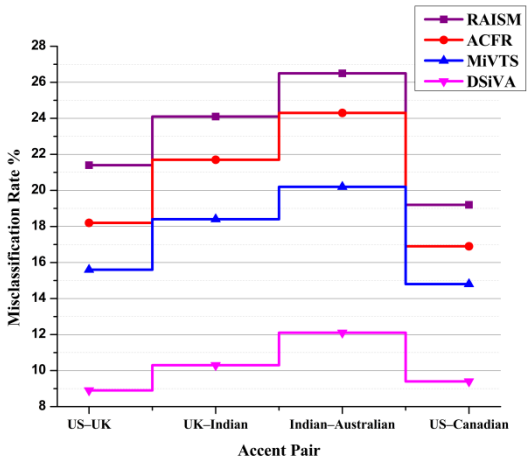

Fig. 3: Analysis of Confusion Between Accents

The average error rate for DSiVA was 10.2%, which is lower than the rates for RAISM (22.8%), ACFR (20.3%), and MiVTS (17.2%). DSiVA only produced 10.3% of mistakes when the UK and India were challenging to work with. This is about half of what other approaches did. Fig. 3 shows that it can tell the difference between pretty similar accents. This makes it simpler to classify them and less confusing.

Analysis of confusion between Accents $\partial_{b\to c}$ is expressed using equation 3,

$$\partial_{b\to c} = \frac{N_{bc}}{N_{bk}} (3)$$

Equation 3 explains the analysis of confusion between accents directed miss assignment rate from source accent to target accent using row-normalized counts.

In this $\partial_{b\to c}$ is the Directed confusion proportion, N_{bc} is the Entries of the confusion matrix, N_{bk} is the Total items whose truth, and b, c, k are the Accent category identifiers.

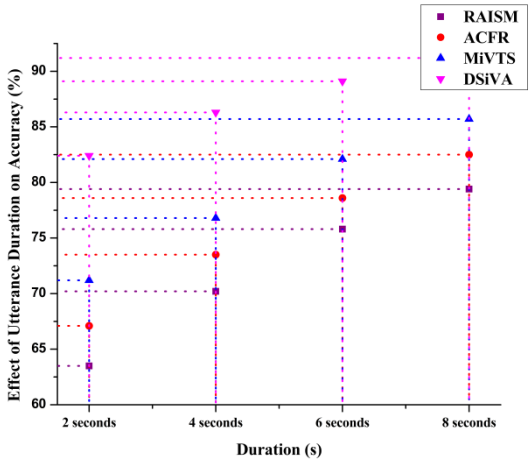

Fig. 4: Analysis of the Effect of Utterance Duration

DSiVA was quite accurate, obtaining 82.4% correct at 2 seconds and 91.2% right at 8 seconds. Other approaches fell below 72% for brief periods, but DSiVA always did better than them. Fig. 4 shows that it can handle short responses effectively, which is very important for interactive language proficiency platforms since students commonly provide brief answers.

Analysis of the effect of utterance duration l_ρ is expressed using equation 4,

$$l_\rho = \frac{Dpw(a, \log \mu)}{Wbs(\log \mu)} (4)$$

Equation 4 explains the analysis of the effect of utterance duration sensitivity to temporal extent, represented as the slope of the most reliable linear predictor of accuracy from log-duration.

In this l_ρ is the duration performance coupling coefficient, a is the Per-utterance correctness indicator, μ is the utterance duration, $\log \mu$ is the Natural log of duration, and Dpw, Wbs are the Sample covariance and variance.

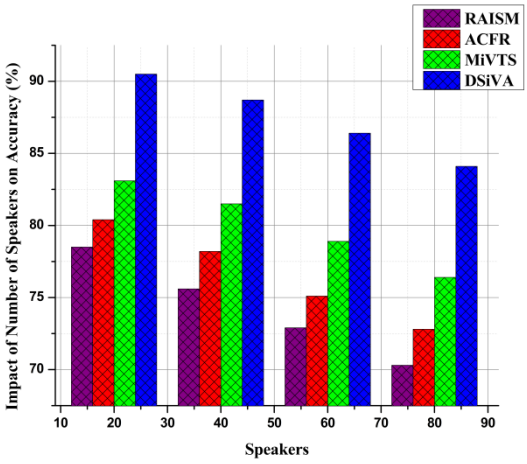

Fig. 5: Analysis of the Impact of the Number of Speakers

DSiVA could handle different speakers well; the accuracy dropped from 90.5% (20 speakers) to 84.1% (80 speakers). In the same situation, RAISM fell to 70.3% (Fig. 5). DSiVA ensures that it can be used in real-world scenarios with diverse user groups by maintaining accent-specific indications across speakers.

Analysis of the impact of the number of speakers Δ_T is expressed using equation 5,

$$\Delta_T = \frac{B(T_2) - B(T_1)}{T_2 - T_1} (5)$$

Equation 5 explains the analysis of the impact of the number of speakers, accuracy sensitivity to finite differences while enlarging the unique speaker pool.

In this Δ_T is the Speaker-count sensitivity coefficient, B is the accuracy computed on a subset containing, and T_1, T_2 are the Speaker cardinalities.

V. Results and Discussion

A. *Accent Classification Performance*

The proposed DSiVA model significantly enhances accent classification across all datasets. Segment-level i-vectors that pick up localized accent signals make it possible to find both short and long speech consistently. The method shows that it can handle speaker variability, as it has been shown to be more effective than baseline deep learning models and classic i-vector models. The model has since shown good performance with different accent groups and dataset configurations and

therefore it is applicable in the automated speech evaluation systems and language proficiency platforms.

B. Comparison with Baseline Methods

The findings indicated that DSiVA performed better than baseline deep learning models and the conventional i-vector methods of all datasets analyzed. The standard deep learning algorithms fail to capture the nuances of the accent, and the standard i-vector models are faulty in the short phrases. The DSiVA algorithm is more precise and standardized as it incorporates the most effective features of the two approaches. The fact that the structure can identify even minor repercussions in pronunciation is revealed by the lower incidence of misclassification in comparative analysis, even at the level of closely related accents.

C. Analysis of Segmental Contributions

Local characteristics have a significant effect on the overall accuracy of categorization, which can be shown on a segment level. Small units serve more effectively in extracting the prosodic and phonetic information that can be used to make the models distinguish between different accents. The deep neural network had the capability of learning hierarchical accent patterns through a combination of segmental i-vectors. This will reduce its tendency to change with change of speakers or the environment. Experiments support one of the foundational design features of the DSiVA framework: segmental information is essential in processing short utterances and overlapping accent segments.

VI. Conclusion and Future Work

A. Summary of Findings

Under consideration of its future implementation into language proficiency platforms, this paper presented a DSiVA as an accent classification in speech. Distributing speech into parts and deriving i-vector depictions of these parts, DSiVA is capable of identifying both regional and global accent characteristics. Effective information aggregation will increase the accuracy of the classification when a deep neural network analyzes i-vectors at the segment level. The outcome of the experiment has shown that DSiVA was more effective than the deep learning models of the baseline as well as simple i-vector techniques across various diverse datasets. Among the interesting results are that the system is capable of accommodating speaker variation, is more effective at distinguishing accents that are quite similar to one another and effective on short phrases. It was discovered that overall recognition skills are to be enhanced with the help of segmental analysis that demonstrates the interaction of i-vector representations with deep neural networks.

B. Potential Applications in Language Proficiency Platforms

The proposed paradigm will have a massive influence on computer based learning and assessment systems. Adaptive learning modules, more stringent speech evaluation standards, and personalized pronunciation feedback are all possible with good accent recognition skills. By integrating DSiVA into language learning systems, contact centers, or mobile applications, organizations can access accurate, up-to-the-minute information about how people talk. It is helpful in interactive learning settings when individuals need to provide shorter replies since it can handle brief phrases.

C. Directions for Future Research

To further enhance feature aggregation, investigate ways to make DSiVA more resilient in noisy environments, and expand its capabilities to include multilingual accent recognition are all potential areas for future research. Future work directions in speech profiling may consist of the investigation of optimum techniques for the integration of prosodic, lexical, and contextual data, with the analysis of real-time implementation on embedded or mobile systems. AI linguists will have stronger tools for figuring out accents now that they have taken these pathways.

References

[1] Sun, Y. (2021). *Automatic Speech Recognition: a study of adaptation techniques for noise and accent conditions* (Doctoral dissertation, SL: SN).

[2] Selin, M., & Mathew, K. P. (2024). ResNet152: A Deep Learning Approach for Robust Spoof Detection in Speaker Verification Systems. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15(3), 354-363. https://doi.org/10.58346/JOWUA.2024.I3.023

[3] Patil, A. P., Ahluwalia, P., Yadav, S., & Kaur, P. (2023). Classification of Accented Voice Using RNN and GAN. Computational Intelligence in Analytics and Information Systems: Volume 1: Data Science and AI , Selected Papers from CIAIS-2021, 303.

[4] Ranjkesh, M., & Ziabari, M. (2016). Persian Language telephone and Microphone Speaker identification using neural networks. International Academic Journal of Science and Engineering, 3(2), 6–12.

[5] Lai, W., & Zheng, Y. (2023). Speech recognition of south China languages based on federated learning and mathematical construction. *Electronic Research Archive*, *31*(8).

[6] Mokhtarinejad, A., Mokhtarinejad, O., Kafaki, H. B., & Ebrahimi, S. M. H. S. (2017). Investigating German Language Education through Game (Computer and non-Computer) and its Correspondence with Educational Conditions in Iran. International Academic Journal of Innovative Research, 4(2), 1–9.

[7] Tripathi, A., Tripathi, A., Varde, K., Patil, M., & Dhole, S. V. (2024). Machine Learning based Identification of Spoken Language Variations using Speech Analysis. *Available at SSRN 4936132*.

[8] Bilal, Z. S., Gargouri, A., Mahmood, H. F., & Mnif, H. (2024). Advancements in Arabic Sign Language Recognition: A Method based on Deep Learning to Improve Communication Access. Journal of Internet Services and Information Security, 14(4), 278-291. https://doi.org/10.58346/JISIS.2024.I4.017

[9] Guntur, R. K., Ramakrishnan, K., & Vinay Kumar, M. (2022). An automated classification system based on regional accent. *Circuits, Systems, and Signal Processing*, *41*(6), 3487-3507.

[10] Keshireddy, S. R. (2025). Low-Code Development Enhancement Integrating Large Language Models for Intelligent Code Assistance in Oracle APEX. Indian Journal of Information Sources and Services, 15(2), 380–390. https://doi.org/10.51983/ijiss-2025.IJISS.15.2.46

[11] Kashif, K. (2025). From detailed acoustic analysis to AI: designing and developing advanced speech analysis tools.

[12] Pennell, R. (2022). AUTOMATED SPOKEN LANGUAGE DETECTION.

[13] Kasture, N., & Jain, P. (2025). Enhancing child-machine interaction for Indian children speaking English as a non-native language using a hybrid CNN and a customized dictionary. *Universal Access in the Information Society*, 1-16.

[14] Botelheiro, D., Abad, A., Freitas, J., & Correia, R. (2021). Nativeness Assessment for Crowdsourced Speech Collections. In *IberSPEECH*.

[15] Shah, M. M., & Kavathiya, H. R. (2024). *Development Of A Model To Analyze & Interpret Vernacular Voice Recognition Of Gujarati Dialects* (Doctoral dissertation, Department of Computer Science, Faculty of Science Atmiya University.).

[16] https://www.kaggle.com/datasets/himanshu9648/english-accent-classification-dataset

[17] Vuddagiri, R. K. (2022). *Implicit Indian Language Identification Using Different Deep Neural Network Architectures* (Doctoral dissertation, International Institute of Information Technology Hyderabad).

[18] Mirishkar, S. G. (2023). *Towards Building an Automatic Speech Recognition System in the Indian Context using Deep Learning* (Doctoral dissertation, International Institute of Information Technology, Hyderabad).

[19] Wang, D., Ye, S., Hu, X., Li, S., & Xu, X. (2021, August). An End-to-End Dialect Identification System with Transfer Learning from a Multilingual Automatic Speech Recognition Model. In *Interspeech* (Vol. 1, No. 1, pp. 3266-3270).

[20] Dowerah, S. (2023). *Deep Learning-based Speaker Identification In Real Conditions* (Doctoral dissertation, Université de Lorraine).

[21] Jassim, S., & Abdulmohsin, H. A. (2025). Accent Classification Using Machine Learning Techniques: A Review. *International Journal of Computer Information Systems and Industrial Management Applications*, *17*, 421-451.