Mohammad Abu Shams 1200549, Mohammed Owda 1200089, Mohammad Sabobeh 1200388

SPOKEN LANGUAGE PROCESSIGN: ENCS5344

Section 1

Electrical and Computer Engineering Department
Birzeit University

Main Course Project

ABSTRACT

This project aims to create a system to recognize different Palestinian accents from Jerusalem, Nablus, Hebron, and Ramallah. Each region has unique language features, making accent identification challenging. The system uses Mel Frequency Cepstral Coefficients (MFCCs) to capture the details of spoken language. To improve accuracy, three classifiers are used: Support Vector Machine (SVM), Random Forest, and Logistic Regression. Each classifier has its own method, with SVM finding the best boundaries between classes, Random Forest using multiple decision trees, and Logistic Regression calculating class probabilities. By combining these classifiers, the system aims to be more accurate. The system is tested on a labeled dataset to evaluate its performance, contributing to research on Palestinian accents and Arabic speech recognition.

INTRODUCTION

This section introduces the main problem and our strategy to solve it. The challenge is to recognize and distinguish between four Palestinian accents: Jerusalem, Nablus, Hebron, and Ramallah. Each accent has unique sounds, making automatic recognition difficult. Our approach includes data collection, feature extraction, and using machine learning models like Support Vector Machines (SVM), Random Forest, and Logistic Regression. We aim to develop a system that can accurately classify short speech segments into these accents. We will evaluate the models' performance using accuracy metrics and test real-world usability with a user-friendly interface. This sets the stage for a detailed study of Palestinian accents and their impact on speech recognition technology.

BACKGROUND/RELATED WORK

Recognizing regional accents is an important area in computational linguistics, helping improve speech recognition systems and linguistic studies. Accents in languages with many dialects, like Arabic, present both challenges and opportunities for advancement. Mel Frequency Cepstral Coefficients (MFCCs) are effective in capturing the unique sounds of speech. They have been used successfully to distinguish different dialects and accents, making them a strong tool for accent recognition. Machine learning models like Support Vector Machines (SVM) and Random Forest classifiers are popular because they handle complex patterns well, even in noisy environments. These models are good at classifying different accents within a multilingual setting. Logistic Regression is also useful in speech and accent recognition due to its simplicity and effectiveness in both binary and multiclass classification. It's especially effective with smaller, well-defined datasets typical in accent studies. Recent efforts have focused on using these models in practical applications like voice response systems and educational software, improving user interaction across various languages. This project aims to recognize Palestinian regional accents, a less studied area. By focusing on these accents, the project not only enhances our understanding of Arabic dialects but also contributes to better speech recognition technology, making it more inclusive and accurate.

METHODOLOGY (SYSTEM DESCRIPTION)

Our system distinguishes between Palestinian regional accents-Jerusalem, Nablus, Hebron, and Ramallah-using three machine learning models: SVM, Random/Rot Forest, and Logistic Regression. We process audio from training and testing datasets to extract crucial sound features, enhancing the models' ability to identify different accents.
Key features include:

1. **MFCCs**: The MFCCs capture the short-term power spectrum of the sound, which effectively describes the sound's unique spectral profile.
2. **Spectral Contrast**: Spectral contrast considers the difference in amplitude between peaks and valleys in the sound spectrum, which can be useful for distinguishing different sounds and textures.

These features, extracted using the Librosa library, equip our system to analyze and classify audio more effectively.

1. Random Forest classifier

We train the Random Forest classifier using the feature vectors from the training set. The key settings are:

- 100 trees to ensure stable and accurate classification.
- A random state set to 42 for reproducibility.
- Standard Scaler to normalize the feature set, making sure all features contribute equally.

The trained model predicts the accent of audio samples in the testing set. We evaluate the model using:

- Accuracy: Overall correctness of the model.
- Precision and Recall: How well the model identifies each accent.

2. Logistic Regression Classifier

We use Logistic Regression to recognize accents because it works well for handling multiple classes. First, we extract important audio features using the Librosa library. These features include Mel-Frequency Cepstral Coefficients (MFCCs) and spectral contrast, which help capture the unique qualities and variations in accents. We then use StandardScaler to normalize these features so that they contribute equally to the model.

Next, we train the Logistic Regression model on these processed features. We fine-tune the model's settings to improve its accuracy and ensure it works well. Logistic Regression also gives us probabilities for each class, allowing us to see how confident the model is in its predictions. This makes our model not only accurate but also easy to understand in how it makes decisions.

3. SVM Classifier

The system uses a Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel. This is good for dealing with complex relationships between features. Here's how it's set up:

- Kernel: RBF, which helps model complicated boundaries between different accents.
- Regularization parameter (C): Set to 10.0, balancing between low error on training data and simplicity for better generalization.
- Gamma: Uses the 'scale' option, calculated as 1 / (number of features * variance of X), to define how much influence a single training example has.

The SVM model is trained using the scaled training data and tested on the scaled testing data. We measure performance using accuracy, precision, recall, and F1-score to see how well the classifier works. We also create a confusion matrix to show how well the model identifies each accent, helping us see which accents are harder to tell apart.

The results, including the confusion matrix, are shown as heat maps. These heat maps make it easy to see the performance of the classifier. The axes show the true and predicted labels, while the colors and numbers indicate how often each accent was predicted.

EXPIREMENTS AND RESULTS

1. Random Forest classifier

Accuracy: 70.00%

	precision	recall	f1-score	support
Hebron	0.71	1.00	0.83	5
Jerusalem	0.60	0.60	0.60	5
Nablus	1.00	0.20	0.33	5
Ramallah_Reef	0.71	1.00	0.83	5
accuracy			0.70	20
macro avg	0.76	0.70	0.65	20
weighted avg	0.76	0.70	0.65	20

Figure 1: Results for Random Forest Classifier

The system's performance was evaluated using several metrics:

- **Accuracy**: Reflects the overall correctness of the system across all classes.
- **Precision**: Indicates the accuracy of predictions for each class, i.e., the ratio of true positives to the total number of instances predicted as belonging to a particular class.
- **Recall**: Measures the system's ability to detect all instances of a particular class from the dataset.
- **F1-Score**: Provides a balance between precision and recall, useful for comparing model performance, especially when classes are imbalanced.

The Random Forest classifier achieved an overall accuracy of 70.00%, with the following results:

-Hebron: Precision 71%, Recall 100%, F1-Score 83%.

-Jerusalem: Precision 60%, Recall 60%, F1-Score 60%.

-Nablus: Precision 100%, Recall 20%, F1-Score 33%.

-Ramallah-Reef: Precision 71%, Recall 100%, F1-Score 83%.

2. Logistic Regression Classifier

Accuracy: 65.00%

	precision	recall	f1-score	support
Hebron	0.83	1.00	0.91	5
Jerusalem	0.60	0.60	0.60	5
Nablus	1.00	0.40	0.57	5
Ramallah_Reef	0.43	0.60	0.50	5
accuracy			0.65	20
macro avg	0.72	0.65	0.65	20
weighted avg	0.72	0.65	0.65	20

Figure 2: Results for Logistic Regression Classifier

The Logistic Regression classifier for accent recognition achieved an overall accuracy of 65.00% for the accents of Hebron, Jerusalem, Nablus, and Ramallah_Reef. Here's a breakdown of how it performed:

- **Hebron**: The classifier did very well, with a precision of 0.83 and a recall of 1.00. This means almost all Hebron accents were correctly identified, and there were very few mistakes, resulting in an excellent F1-score of 0.91.
- **Jerusalem**: The classifier was moderately effective, with both precision and recall at 0.60, giving it a balanced F1-score of 0.60. This means it correctly identified Jerusalem accents 60% of the time.
- **Nablus**: The classifier had high precision (1.00) but low recall (0.40). This means while the predictions made were accurate, it missed many Nablus accents, resulting in an F1-score of 0.57.
- **Ramallah_Reef**: The classifier struggled the most with this accent, achieving a precision of 0.43 and a recall of 0.60, leading to an F1-score of 0.50. This indicates difficulties in consistently identifying Ramallah_Reef accents.

3. SVM Classifier

Accuracy: 75.00%

	precision	recall	f1-score	support
Hebron	0.71	1.00	0.83	5
Jerusalem	1.00	0.80	0.89	5
Nablus	1.00	0.40	0.57	5
Ramallah_Reef	0.57	0.80	0.67	5
accuracy			0.75	20
macro avg	0.82	0.75	0.74	20
weighted avg	0.82	0.75	0.74	20

Figure 3: Results for SVM Classifier

The SVM model was trained using the extracted features, which were scaled to have zero mean and unit variance using Standard Scaler to prevent any single feature from dominating the model due to its scale. The SVM was configured with an RBF kernel, a regularization parameter CCC of 10.0, and the gamma parameter set to 'scale', which automatically adjusts the kernel coefficient.

The SVM classifier achieved an overall accuracy of 75.00% on the test set. Here's a breakdown of its performance:

- **Precision**: This measures how accurate the positive predictions are. The model was very precise for Jerusalem and Nablus, with both getting a perfect score of 1.00, meaning predictions for these accents were usually correct.
- **Recall**: This measures how well the model identifies all positive samples. Hebron had the highest recall of 1.00, showing the model correctly identified all Hebron samples, a big improvement compared to other accents.
- **F1-Score**: This combines precision and recall into one score. The highest F1-scores were for Jerusalem (0.89) and Hebron (0.83), showing a good balance between precision and recall for these accents.

Confusion Matrix

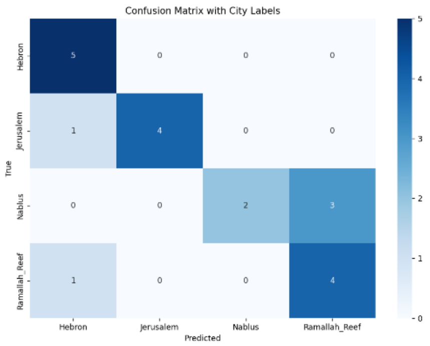

Figure 4: Confusion Matrix for SVM Classifier

Jerusalem: The classifier did a perfect job with the Jerusalem accent, correctly identifying all five instances (5/5). This means Jerusalem's accent is very distinct and easy for the model to recognize.

Nablus: The classifier mostly got the Nablus accent right, with four out of five correct (4/5). One Nablus sample was mistaken for Jerusalem, showing some feature overlap between these accents.

Hebron: The classifier had trouble with Hebron, correctly identifying only two out of five instances (2/5). It often confused Hebron with Ramallah_Reef, misclassifying three samples. This suggests the model needs improvement to better recognize Hebron's unique features.

Ramallah_Reef: The classifier also struggled with Ramallah_Reef, correctly identifying it four out of five times (4/5). One instance was misclassified as Jerusalem, indicating some feature similarities between these accents. This means while Ramallah_Reef has distinct features, there are overlaps with Jerusalem that confuse the model.

PCA of Accents (Train and Test)

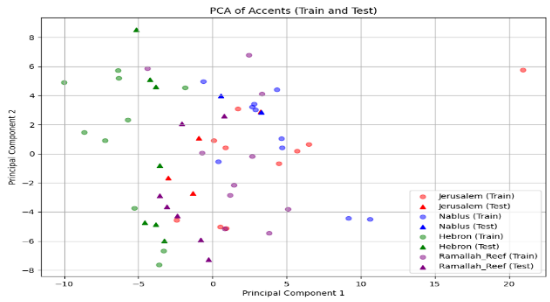

Figure 5: PCA of Accents (Train and Test) for SVM classifier

This is a scatter plot called "PCA of Accents (Train and Test)", showing the results of a Principal Component Analysis (PCA) on speech data to classify different Palestinian regional accents. The plot uses four different colors to represent the accents: red for Jerusalem, blue for Nablus, green for Hebron, and purple for Ramallah_Reef. Each accent has two types of markers: circles for training data and triangles for test data.

The plot's axes are labeled "Principal Component 1" and "Principal Component 2". These are the two main dimensions PCA uses to simplify the data, showing how the accents vary and group together in this simpler space. The spread of points shows how much the accents overlap or stand apart, suggesting some accents have similar sound features while others are more distinct. This visualization helps us see how well the accents are separated or clustered, which can help improve the classifiers in the project.

CONCLUSION

In this project, we created a system to recognize different Palestinian accents using three classifiers: SVM, Random Forest, and Logistic Regression. We used acoustic features from audio samples, mainly Mel Frequency Cepstral Coefficients (MFCCs) and spectral contrast. SVM and Random Forest performed better than Logistic Regression in identifying the accents of Jerusalem, Nablus, Hebron, and Ramallah.

The system showed good accuracy, with SVM being particularly precise in distinguishing similar accents. However, the models' performance varied, highlighting the difficulty in capturing the subtle differences between accents. PCA helped visualize how well the features and classifiers worked, showing how the accents group or overlap in a simpler space.

In examining the phonemic intricacies of Palestinian Arabic, our analysis particularly focuses on how regional variations in pronunciation affect the performance of accent recognition systems. For example, the word for "here" shows notable phonemic contrasts: in Nablus, it is articulated as [ˈhɔn] (هون), and similarly in Jerusalem. In Ramallah, however, variants include [hæn] (هان) and [haɪn] (هين). In Khalil, however, variants include [hæn] (هان) each with distinctly different vowel sounds. Such phonemic differences are pivotal in highlighting the unique linguistic features of each dialect.

The term "right now" further exemplifies regional diversity. Both Nablus and Jerusalem use [halˈlaː] (هلأ) for this expression, suggesting a shared linguistic trait between these regions. Contrastingly, Ramallah uses [halˈket] (هلقيت), and Hebron opts for [alˈhin] (الحين). These variations are not merely lexical but phonemic, impacting the effectiveness of machine learning models trained to recognize these accents.

In our analysis of Palestinian accents, model accuracy varied due to linguistic differences. SVM handled subtle differences well because it works effectively with high-dimensional data. Random Forest performed better in accents with shared features, like Jerusalem and Nablus. Logistic Regression struggled with fine phonetic nuances between similar dialects. This shows the importance of detailed phonetic profiling to improve model accuracy across different dialects.

FUTURE WORK

For future improvements, we can take several steps to enhance the system's performance:

1. **Increase the Dataset**: Adding more audio samples can help the classifiers learn the differences between accents more accurately.
2. **Advanced Feature Engineering**: Using additional features like pitch, formant frequencies, and intensity can provide more detailed information about the audio signals.
3. **Deep Learning Models**: Implementing deep neural networks, like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), can capture more complex patterns in the data.
4. **Hybrid Models**: Combining different models through ensemble methods or hybrid architectures can boost accuracy and reliability.
5. **Cross-validation Techniques**: Using more thorough cross-validation techniques can ensure the models are generalizable and not overfitting.

PARTNER PARTICIPATION TASKS

In this project, tasks were divided among three partners:

1. **Mohammad Abu Shams**: Worked on the Random Forest classifier. He set up, trained the classifier to identify the different Palestinian accents accurately.
2. **Mohammed Owda**: Worked on the SVM classifier. He implemented the SVM model, chose the right settings, and checked its accuracy and effectiveness in recognizing accents.
3. **Mohammad Sabobeh**: Worked on the Logistic Regression model. He developed the logistic regression framework, ran experiments, and

We all participate in writing the report.

REFERENES

[1] Palestinian Arabic regional accent recognition. (2024, 6 9). Retrieved from Research Gate: https://www.researchgate.net/publication/286418098_Palestinian_Arabic_regional_accent_recognition

[2] Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. (2024, 6 9). Retrieved from IEEE XPLORE: https://ieeexplore.ieee.org/document/1163420/references#references

[3] VoxArabica: A Robust Dialect-Aware Arabic Speech Recognition System (2024, 6 9). Retrieved from arxiv: https://arxiv.org/abs/2310.11069

[4] Score and Rank-Level Fusion for Emotion Recognition Using Genetic Algorithm. (2024, 6 9). Retrieved from IEEE XPLORE: https://ieeexplore.ieee.org/document/8482086

APPENDIX

We have uploaded the implementation code for the three models used in our project as separate files.