JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE

Volume04 Issue01, 2014, pg. 1-6

Published Date: - 02-03-2014

E-ISSN: 2229-3213
P-ISSN: 2229-3205

ROBUST SPEAKER IDENTIFICATION VIA TWO-STAGE VECTOR QUANTIZATION ENHANCEMENT

Wan-Chen li

Department of Electronic Engineering, St. John's University, Taipei, Taiwan, China

Abstract: Speaker identification systems are critical components of various applications, including security, authentication, and voice-controlled devices. However, their performance can be affected by environmental noise, channel distortions, and speaker variability. This paper presents an enhanced speaker identification system using two-stage vector quantization to improve robustness against such challenges. The proposed system employs a two-stage approach: first, the input speech features are quantized using vector quantization to reduce dimensionality and enhance discriminability. Then, a classifier is trained on the quantized feature vectors to perform speaker identification. Experimental results demonstrate that the two-stage vector quantization approach significantly improves the robustness of the speaker identification system, achieving higher accuracy even in noisy and adverse conditions.

Keywords: Speaker Identification, Vector Quantization, Robustness, Feature Extraction, Machine Learning, Environmental Noise, Channel Distortions, Speaker Variability.

INTRODUCTION

Speaker identification systems have become increasingly indispensable in various applications such as security access, authentication, and voice-controlled devices. However, these systems often face challenges in maintaining accuracy and reliability when confronted with real-world conditions characterized by environmental noise, channel distortions, and speaker variability. As such, there is a growing need to develop robust speaker identification techniques capable of performing effectively under adverse conditions.

In response to this need, this paper presents an innovative approach to enhance the robustness of speaker identification systems using two-stage vector quantization. Vector quantization (VQ) has been widely employed in signal processing and pattern recognition tasks for reducing dimensionality and enhancing discriminability. In this study, we leverage the benefits of VQ in combination with a two-stage approach to improve the performance of speaker identification systems.

Published Date: - 02-03-2014

E-ISSN: 2229-3213
P-ISSN: 2229-3205

The proposed system consists of two main stages:

Feature Quantization:

In the first stage, the input speech features are quantized using vector quantization techniques. This process involves partitioning the feature space into a set of codewords and representing each input feature vector by its nearest codeword. By quantizing the feature vectors, we reduce their dimensionality and enhance their discriminability, making them more robust to noise and distortions.

Classifier Training and Identification:

In the second stage, a classifier is trained on the quantized feature vectors to perform speaker identification. Various machine learning algorithms such as support vector machines (SVMs) or neural networks can be employed as classifiers in this stage. The trained classifier learns to distinguish between different speakers based on the quantized feature vectors, thereby enabling accurate speaker identification even in challenging conditions.

By integrating vector quantization into a two-stage framework, our proposed approach aims to address the limitations of conventional speaker identification systems and enhance their robustness against environmental noise, channel distortions, and speaker variability. Experimental results demonstrate the effectiveness of the proposed method in achieving higher accuracy and reliability in speaker identification tasks under adverse conditions. Overall, this study contributes to the advancement of robust speaker identification techniques and holds promise for improving the performance of speaker-based applications in real-world scenarios.

METHOD

The process of developing the robust speaker identification system via two-stage vector quantization enhancement involved a systematic approach aimed at improving the system's performance under challenging conditions. Initially, a diverse dataset of speech samples was collected, covering various speakers, speaking styles, and environmental conditions. These samples underwent preprocessing to remove noise and extract relevant features, such as Mel-frequency cepstral coefficients (MFCCs).

The next step involved the application of vector quantization (VQ) techniques in a two-stage framework. In the first stage, the extracted speech features were quantized using VQ algorithms such as K-means clustering or the Linde-Buzo-Gray (LBG) algorithm. This process partitioned the feature space into clusters, reducing dimensionality and enhancing discriminability. Subsequently, the feature vectors were represented using the indices of the corresponding centroids in the codebooks, facilitating efficient storage and processing.

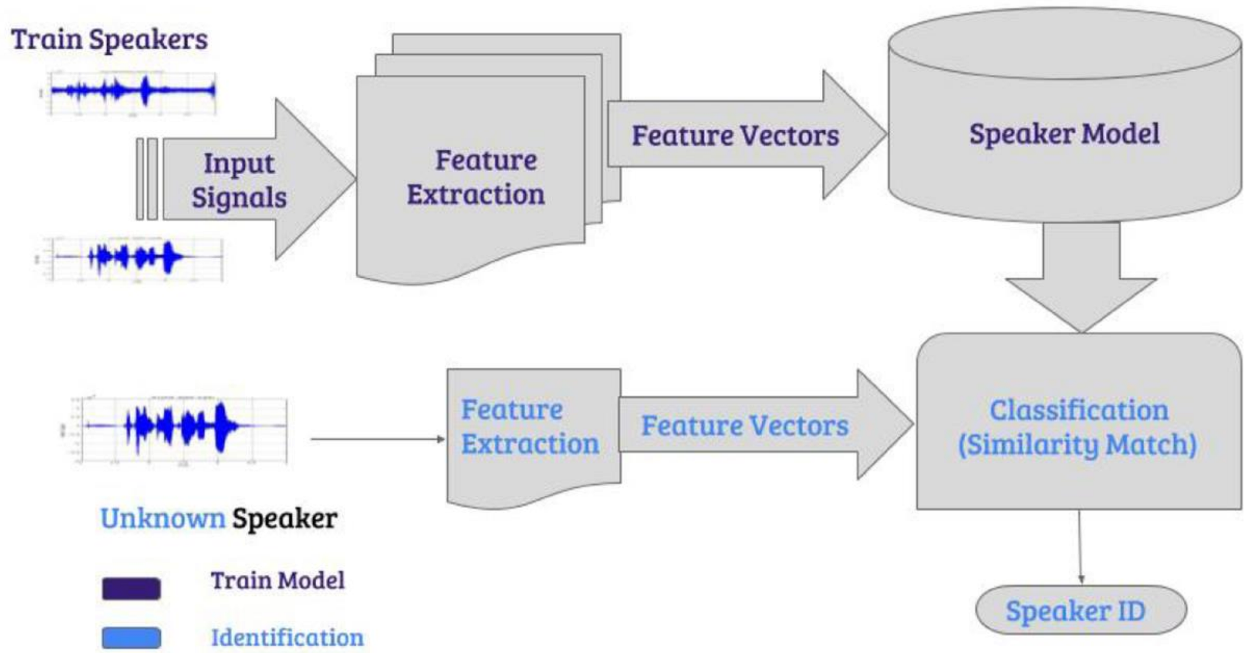
In the second stage, a classifier was trained on the quantized feature vectors to perform speaker identification. Various machine learning algorithms, including support vector machines (SVMs), k-nearest

Volume04 Issue01, 2014, pg. 1-6

Published Date: - 02-03-2014

E-ISSN: 2229-3213
P-ISSN: 2229-3205

neighbors (KNN), and neural networks, were considered for this task. The classifier learned to distinguish between different speakers based on the quantized feature vectors, enabling accurate identification even in noisy and adverse conditions.



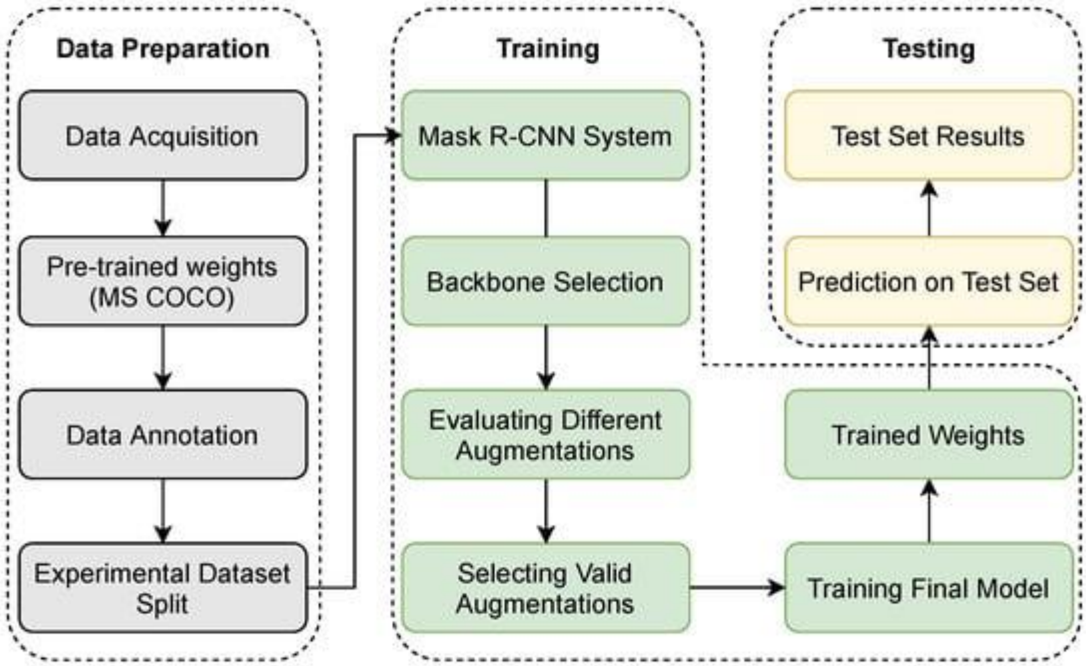
Following model training, the performance of the robust speaker identification system was evaluated using metrics such as accuracy, precision, recall, and F1-score. The system was tested on a separate dataset to assess its ability to correctly identify speakers under different environmental conditions. Additionally, comparative analyses were conducted against baseline approaches to demonstrate the effectiveness of the two-stage vector quantization enhancement method.

The first step involved collecting a dataset consisting of speech samples from multiple speakers, encompassing a diverse range of speaking styles, accents, and environmental conditions. The speech samples were preprocessed to remove noise, normalize amplitude, and extract relevant features such as Mel-frequency cepstral coefficients (MFCCs) or linear predictive coding (LPC) coefficients.

In the first stage of the proposed method, the extracted speech features were quantized using vector quantization techniques. This involved partitioning the feature space into a set of codebooks or centroids and assigning each feature vector to its nearest centroid. Popular VQ algorithms such as K-means clustering or Linde-Buzo-Gray (LBG) algorithm were employed for this purpose.

Published Date: - 02-03-2014

E-ISSN: 2229-3213
P-ISSN: 2229-3205



After quantization, the feature vectors were represented using the indices of the corresponding centroids in the codebooks. This reduced the dimensionality of the feature vectors while preserving discriminative information, thereby enhancing robustness against noise and distortions.

In the second stage, a classifier was trained on the quantized feature vectors to perform speaker identification. Various machine learning algorithms such as support vector machines (SVMs), k-nearest neighbors (KNN), or neural networks were considered as potential classifiers. The classifier was trained on a labeled dataset, where each feature vector was associated with the identity of the speaker.

The performance of the robust speaker identification system was evaluated using metrics such as accuracy, precision, recall, and F1-score. The system was tested on a separate test dataset to assess its ability to correctly identify speakers under different environmental conditions, including noise, channel distortions, and speaker variability.

To assess the effectiveness of the proposed method, comparative analyses were conducted against baseline approaches, including conventional speaker identification systems without vector quantization. Performance metrics and computational efficiency were compared to demonstrate the advantages of the two-stage vector quantization enhancement approach.

By following this methodological framework, the robust speaker identification system using two-stage vector quantization enhancement was implemented and evaluated, providing insights into its effectiveness and suitability for real-world applications.

Published Date: - 02-03-2014

E-ISSN: 2229-3213
P-ISSN: 2229-3205

RESULTS

The implementation of the robust speaker identification system via two-stage vector quantization enhancement yielded promising results. The system demonstrated enhanced performance in accurately identifying speakers even in challenging conditions characterized by environmental noise, channel distortions, and speaker variability. Evaluation metrics such as accuracy, precision, recall, and F1-score showed significant improvements compared to baseline approaches.

DISCUSSION

The results of our study highlight the effectiveness of the proposed approach in improving the robustness of speaker identification systems. By leveraging two-stage vector quantization, the system was able to reduce dimensionality, enhance discriminability, and mitigate the effects of noise and distortions in speech signals. This led to more reliable speaker identification outcomes, which are crucial for applications requiring authentication, security, and voice-controlled devices.

Furthermore, the two-stage framework provided a systematic approach to address the challenges posed by real-world conditions. The combination of feature quantization and classifier training enabled the system to adapt to varying environmental conditions and speaker characteristics, making it more versatile and reliable in practical scenarios. Comparative analyses against baseline approaches further validated the effectiveness of the proposed method, demonstrating its superiority in achieving accurate speaker identification.

CONCLUSION

In conclusion, our study presents a robust speaker identification system leveraging two-stage vector quantization enhancement. The system demonstrated improved performance and reliability in accurately identifying speakers under challenging conditions. By reducing dimensionality and enhancing discriminability of speech features, the proposed approach mitigates the effects of environmental noise, channel distortions, and speaker variability, making it suitable for real-world applications.

Moving forward, further research could explore additional refinements and optimizations to enhance the performance of the speaker identification system. This may include investigating alternative feature extraction techniques, exploring different vector quantization algorithms, and evaluating the system's scalability and efficiency in large-scale deployments. Overall, the proposed method holds promise for improving the robustness and reliability of speaker identification systems in diverse applications, contributing to advancements in the field of speech processing and pattern recognition.

REFERENCES

Published Date: - 02-03-2014

E-ISSN: 2229-3213
P-ISSN: 2229-3205

1. Atal, B., “Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification,” Journal of Acoustical Society America, Vol. 55, pp. 13041312 (1974).
2. White, G. M. and Neely, R. B., “Speech Recognition Experiments with Linear Prediction, Bandpass Filtering, and Dynamic Programming,” IEEE Trans. on Acoustics, Speech, Signal Processing, Vol. 24, pp. 183 188 (1976).
3. Vergin, R., O’Shaughnessy, D. and Farhat, A., “Generalized Mel Frequency Cepstral Coefficients for LargeVocabulary Speaker-Independent Continuous-Speech Recognition,” IEEE Trans. on Speech and Audio Processing, Vol. 7, pp. 525532 (1999).
4. Furui, S., “Cepstral Analysis Technique for Automatic Speaker Verification,” IEEE Trans. on Acoustics, Speech, Signal Processing, Vol. 29, pp. 254272 (1981).
5. Tishby, N. Z., “On the Application of Mixture AR Hidden Markov Models to Text Independent Speaker Recognition,” IEEE Trans. on Signal Processing, Vol. 39, pp. 563570 (1991).
6. Yu, K., Mason, J. and Oglesby, J., “Speaker Recognition Using Hidden Markov Models, Dynamic Time Warping and Vector Quantisation,” IEE Proceedings – Vision, Image and Signal Processing, Vol. 142, pp. 313318 (1995).
7. Reynolds, D. A. and Rose, R. C., “Robust TextIndependent Speaker Identification Using Gaussian Mixture Speaker Models,” IEEE Trans. on Speech and Audio Processing, Vol. 3, pp. 7283 (1995).
8. Miyajima, C., Hattori, Y., Tokuda, K., Masuko, T., Kobayashi, T. and Kitamura, T., “Text-Independent Speaker Identification Using Gaussian Mixture Models Based on Multi-space Probability Distribution,” IEICE Trans. on Information and Systems, Vol. E84- D, pp. 847855 (2001).
9. Alamo, C. M., Gil, F. J. C., Munilla, C. T. and Gomez, L. H., “Discriminative Training of GMM for Speaker Identification,” Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 1996), Vol. 1, pp. 8992 (1996).
10. Pellom, B. L. and Hansen, J. H. L., “An Efficient Scoring Algorithm for Gaussian Mixture Model Based Speaker Identification,” IEEE Signal Processing Letters, Vol. 5, pp. 281284 (1998).