Al-Rafidain Engineering Journal (AREJ) Vol. 29, No. 2, September 2024, pp. 148-156

Analyzing the MFCC and GFCC to Identify Reverberation Effects on The Sound

Abdalem A. Rasheed
alem12@uomosul.edu.iq

Electrical Engineering Department, College of Engineering, University of Mosul, Mosul, Iraq

Received: February 18th, 2024 Received in revised form: May 12th, 2024 Accepted: July 10th, 2024

ABSTRACT

The majority of acoustic signals contain additive reverberation noise, which degrades and distorts the reliability of the sound system and has detrimental effects on a variety of identification applications, including the speaker recognition field. This paper analyzed two techniques to mitigate and combat the impact of reverberation on sound and compared the performance of these methods. These techniques are Mel-Frequency Cepstral Coefficients (MFCC) and Gammatone Frequency Cepstral Coefficients (GFCC). The GFCC differs from the conventional MFCC in that it replaces the Mel filter bank with a Gamatone filter bank to increase durability.

To avoid the effects of environmental sounds and different features of the speaker voice duo to the variable situation of the speaker such as illness and emotion, a single tone of 1 KHz was applied to obtain a fair and impartial comparison between the GFCC and MFCC methods of sound signal recognition.

The comparison between the MFCC and GFCC features was accomplished by using PCA and corroborated by the normalized cross-correlation NCC. Reducing dimensions and removing correlation is the primary purpose of the PCA algorithm so that the features become orthogonalized. The PCA and NCC report that for both reverberant and non-reverberant single-tone recorded sound, there was a about 10% increase in the rate of detection and the variance increased by 11% for GFCC compared to MFCC features.

Then this work shows that method uses GFCC features is stronger and superior against the reverberation noise than classic MFCC features. Therefore, the GFCC mitigates the reverberation effect and presents a good candidate for functionality in actual recognition systems. In addition, this work examines the potential outcomes of joining the MFCC and GFCC as feature components to obtain a more robust speaker recognition system. The imrovment in the obtained variance is demonstrated by the results to be roughly 30% greater than in the case of GFCC feature coefficients variance.

Keywords:

MFCC; GFCC; Reverberation; Sound; PCA; recognition.

==

1. INTRODUCTION

The majority of sound recognition systems continue to be very perceptive of their auditory surroundings. Accurate recognition will rapidly decline due to various auditory contexts or test and training environments that are impacted by noise, channel distortion, and reverberation. As a result, improving the voice recognition system's resilience has emerged as a key concern.

Research on recognizing speakers has been conducted for at least thirty years in academic institutions worldwide[1].

Numerous reviews and instructional papers related to the broad range of investigations have been released since thereafter[2], [3], [4].

Studies concerning the voice thereby offer technological and economic value, as work remains continuing as a growing number of commercial applications appear.

The characteristics of the sound transmission path between a talker's mouth and the microphone are contaminated by reverberation and noise.

The direct sound signal is included in the room acoustics, which completely depicts the reverberation characteristics of any arrangement for the source-receiver location. The signal's timbre is mostly affected by the late reflections that create a fading noise [5][6].

Different techniques were proposed to lessen the effects of reverberation. Many approaches have been utilized to address this issue, such as using an array of microphone to reduce room reflections and enhance direct sound quality[7].

Mean subtraction of cepstral[8], and features normalization[9][10], are other techniques used to lessen the reverberation impact. It is recommended to combine the MFCC and GFCC feature components to increase a speaker recognition system's dependability [8], [11][12][13]

Additionally, multi-modal tactics were employed to address the reverberation issues but with limited success [14].

By utilizing a reference training model, the effects of reverberation were reduced and speaker recognition robustness was enhanced [15][16]. Initial echoes and their characteristics are crucial for an enclosure's acoustics[17][18]. Recently, has been utilizing from several techniques including Mel-frequency cepstral coefficients (MFCC)[19][20][19][21], Gaussian Mixture Model (GMM)[22], Support Vector (SVM)[23], and Gaussian Mixture Model–Universal Background Model (GMM–UBM)[24].

In the past work[10], the MFCC feature coefficients of discriminating for intonation in the brief text-dependent speech of the 1-Sec word utterance of ALLAH compared to the normal reading of the same word.

In this work, the comparison of MFCC and GFCC features for a single sound frequency of 1kHz short duration is achieved. Whereas the single tone confines the investigation solely on reverberation, avoiding the effects of other human voice variations.

2. MFCC FILTER

Figure 1 illustrates the components of the voice recognition system, which include sampling the input signal, Hamming window, FFT, Mel-scale filter, take the log, Discrete Cosine Transform (DCT), and MFCC extraction[25]. With 1024 samples overlapped 2048 samples for each hamming window, there are 86 windows along a 2-second sound wave. The sample frequency that is being used is 44.1kHz. The formula of the hamming window is[26]:

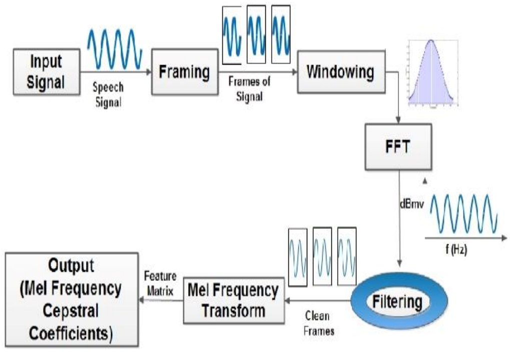

Fig.1 MFCC block diagram [21]

$$H(n) = \begin{cases} 0 \\ w(k) = 0.54 - 0.46\cos\left(\frac{2\pi k}{N-1}\right) \end{cases} \quad (1)$$

Where N represents the total number of samples per frame and k is the total amount of frames. After the windowing, the short forier transform is applied for each window to obtain the frequency response.

3. COEFFICIENTS OF MEL-SCALE

Elemets that more closely like human hearing, which focuses hearing on low frequency more than high frequency, are extracted using the Mel-scale. The following formula can be used to transform a frequency measurement into a Mel-scale

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (2)$$

Where to return the Mel-scale to frequency, as well:

$$f_{mel} = 700 \cdot (e^{M/1125} - 1) \quad (3)$$

The Mel-frequency bands, as shown in Fig.2, are equally spaced on the Mel-scale, which approximates the human auditory system's reaction more precisely than the linearly-spaced frequency bands employed in the standard spectrum, and hence offers a greater sense of speech.

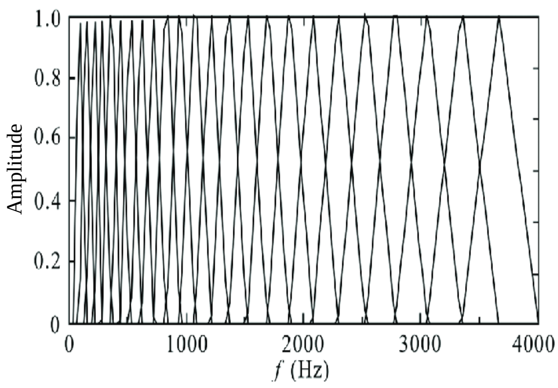

Fig.2 MFCC filter bank

4. SAMPLE FRAME

The applied sampling frequency is 44100Hz, where the frame duration is 23.2 msec and the overlapping is 11.6 msec. The duration of sound is 2 seconds which produces 86 windows (frames). Each frame contains 14 MFCC coefficients. The frequency band of 0Hz to 8KHz contains the majority of the characteristics of the human voice. In this work, 14 Mel-coefficients are employed since they are the most common number.

5. GAMMATIONE FILTER COEFFICIENTS

The gammatone filter represents a linear filter that results from the combination of a sinusoidal tone and a gamma distribution. The response of gammatone is provided by :

$$\boldsymbol{g(t) = a\, t^{n-1}. e^{-2\pi bt}.\cos(2\pi f t} + \phi) \qquad (4)$$

Where t is time, f is the center frequency, a is the amplitude, n is the filter's order, ϕ is the phase of the carrier, and b is the filter bandwidth [27].
The constant a controls the gain, the order filter is defined by the value n which is typically set to a value less than 4, and ϕ is the phase, which is usually set to zero. The factor b is defined as:

$$\boldsymbol{b = 25.17\left(\frac{4.37\, f_c}{1000} + 1\right)} \qquad (5)$$

The form of gammatone filter bands is shown in Fig.3.

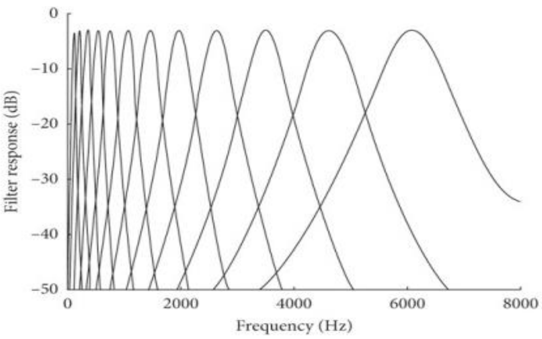

Fig.3 GFCC filter bank [24]

6. MFCC AND GFCC FEATURES ECTRACTION

The energy of the spectrum is subjected to Mel-filter. The resulting values are then logarithmically computed and utilized in the discrete cosine transformation (DCT), Which produces the MFCC or GFCC coefficients by Eq.6

$$\boldsymbol{S[m] = \log\left(\sum_{k=0}^{N-1} |X_a[k]|^2.\ H_m[k]\right)} \qquad (6)$$
$$\boldsymbol{0 \le m < M}$$

The DCT is used to depict the signal as MFCC or GFCC. The energy of the signal spectrum is represented by a vector, which consists of fourteen real numbers. The obtained features identify the frequencies that each individual considers to be most important, enable frame compression and a reduction in the amount of information processed[28], [29].
To summarize all of these processes, Fig.4 illustrates the sequence in the stages to obtain the MFCC, GFCC, or combine features extraction.

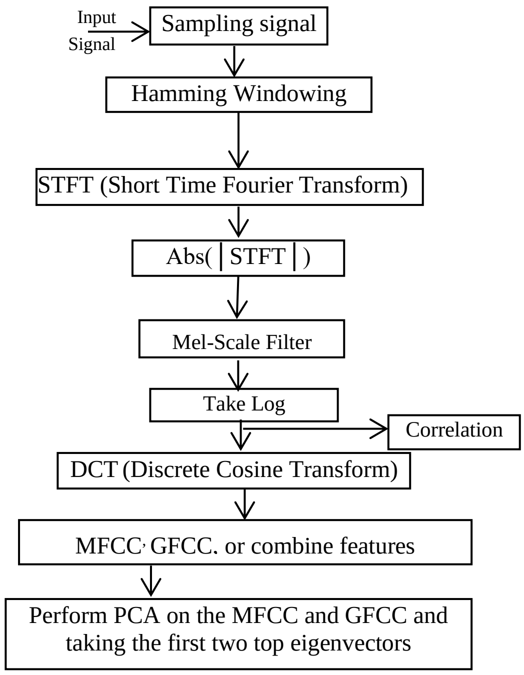

Fig. 4 Flow chart of sound recognition system

7. EXPERIMENTAL RESULTS

First, a limited by band audio signal recording is performed, with each recording file a .wav extension, between 0Hz and 8KHz. This is called "data acquisition" with a sampling frequency is 44.1KHz. The applied tone is about 0.2 Sec of 1KHz to avoid various influences that affect the characteristics of the tested human voice, such as feeling, illness, psychological state, and emotion. The room dimensions are 10m width, 12m length, and 3m height. The dimensions of another room are the same, but it has a lot of cloth with sponge material furniture that absorbs sound waves that are reflected back to it, resulting in minimal reverberating. The experiment was also conducted in another hall with dimensions of 12m, 15m, and 4.5m dimensions.
The aim was to gather information on the single frequency of tone signal of 1KHz. The inspection was accomplished in the absorbed room and no

absorbed room. The Excel file helps the Matlab software collect the findings of an experimental analysis technique for voice signal recognition.

The component that is most crucial for pattern recognition is the database. The purpose of this study was to provide a straightforward database-based system testing procedure. To this method, additional computing capacity and a database can be added.

8. TONE FFT RESPONSE

The spectrum analysis of a single tone that was applied at 1KHz for 0.2 second is displayed in Fig.5. It is evident from Fig.5a that the spectrum is concentrated at 1KHz center frequency in the non-reverberant environment. The spectrum of the same single tone in the reverberant room is displayed in Fig.5b. It is evident that reflected waves are picked up at various phases. Due to the difference in the reflected waves phase, the power spectrum is diminished in some places and enhanced in others as shown in Fig.5b. Fig.5a, also shows a low effect of reverberant signals where the test in the room is not the exactly non-reverberant environment.

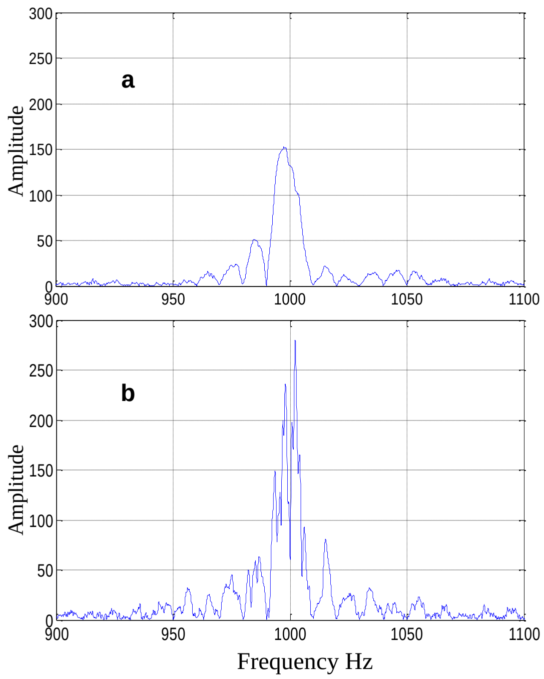

Fig.5 Spectrum analysis of 1KHz tone. a- In the non-reverberant room, b- In the reverberant room

9. SPETROGRAM

A spectrogram is a visual representation of a signal's strength, at different frequencies within a given waveform across time. Spectrograms are used to examine and show trends over extended periods of time. For instance, if there is a constant noisy signal, it will appear as a strait horizontal line someplace. The location and the frequency of the applied signal are displayed in Fig.6a. It is clear that the location of the low reverberation and signal duration extension is from 0.56 second to 0.8 second.

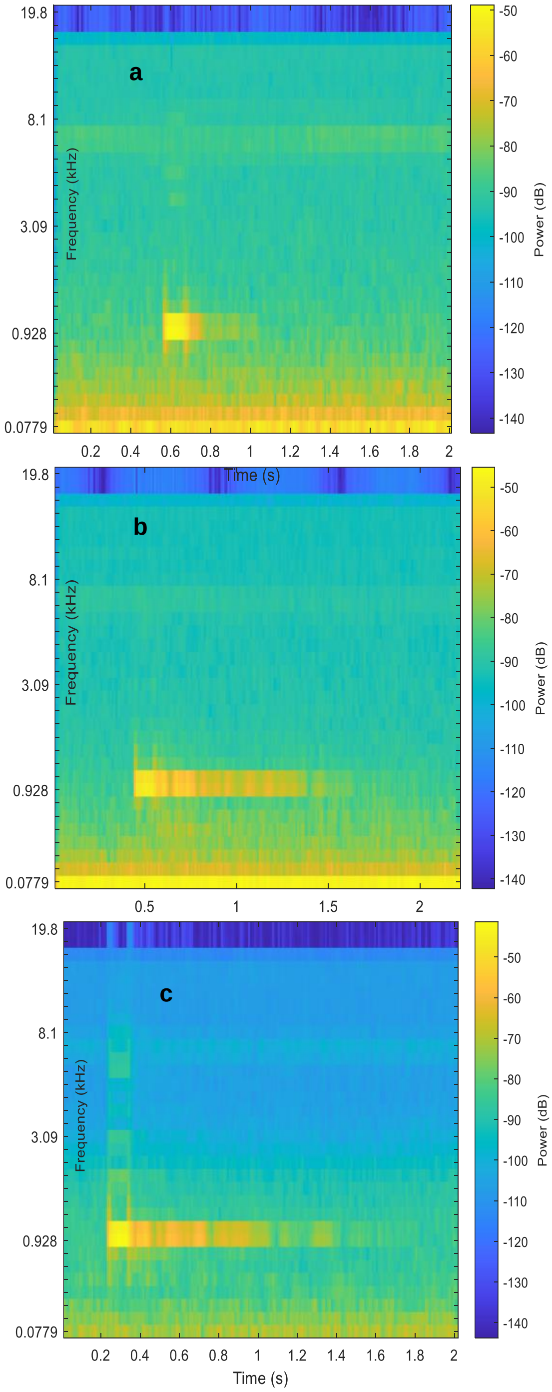

Fig.6 Spectrogram of the signal: a-without reverberation, b-wiwith reverberation, c-reverberation in wider room.

The applied signal is in Fig.6b with an extended reverberation that lasts for over 1.8 seconds and starts at 0.46 seconds. Fig.6c shows the spectrogram of reverberant sound that occurred when the room dimensions were 12m in width, 15m in length, and 4.5m in height. The reverberation took a longer time which is extended from 0.23seconds to 1.6 seconds.

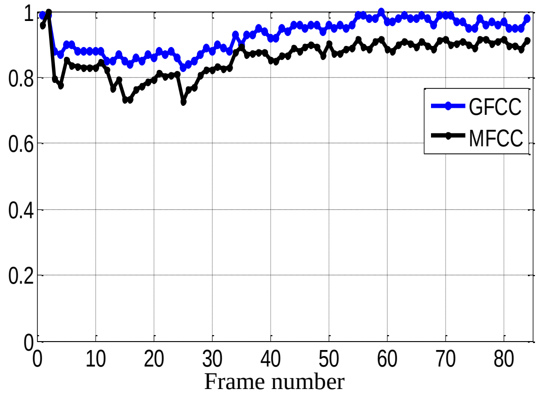

Fig.7 NCC in reverberant and non-reverberant for MFCC and GFCC techniques

As a result, the spectrogram can be used to clearly and easily distinguish between the two reverberant and non-reverberant room responses.

A robust speaker recognition system needs a method that can reliably and precisely represent the acoustic signals of a particular utterance.

Figure 7 represents the NCC between the reverberant and non-reverberant recorded sound for GFCC and MFCC feature extraction methods. Each frame shows how the GFCC maximizes the NCC for detection. The NCC for MFCC reaches a value of about 0.73, whereas the NCC for GFCC reaches a value of about 0.82. Therefore, applying the GFCC approach improved the correlation by a ratio of 10% between the recorded sound that was reverberant and non-reverberant across most of the framing windows.

10. PRINCIPAL COMPONENT ANALYSIS (PCA)

A straightforward technique for lowering a data set's dimensionality forms the basis of Principle component Analysis (PCA). It converts the spaces into a data space with fewer dimensions. Analyzing data and identifying trends more clearly is quite beneficial. The distribution's covariance matrix's eigenvectors as the foundation for primary components.

It is possible to minimize the feature vector without significantly losing information while dealing with redundant data, provided that the collection of features is assumed to be somewhat connected. The original feature vectors are transformed into a smaller space using PCA utilizing Eigen system decomposition, which produces an orthogonal transformation matrix[30][31]. The primary component with the highest variance is the first one, and the least variable is the last one, which in this instance can be discharged.

The Eigenvalue and eigenvector hold significant meaning in the area of PCA. A scalar called an eigenvalue indicates how much the data varies in the direction of the related eigenvector. Eigenvalues in PCA show how crucial the matching eigenvectors are for capturing the variability in the data.

11. MFCC AND GFCC FEATURES EXTRACTION

The MFCC and GFCC discriminations between two cases of reverberant and non-reverberant sound are more clearly depicted by the PCA algorithm, which compresses the 14-dimensional MFCC and GFCC features into a two-dimensional feature space. In Fig.8 and Fig.9, the two-dimensional PCA characteristic patterns of MFCC and GFCC respectively are shown. The PC1 and PC2 represent the first and second variances, respectively. For MFCC and GFCC, the first variance's eigenvalues are roughly 6.9. In MFCC, the second variance's eigenvalue is approximately 1.3, but in GFCC, it is approximately 2.9. Fig.8 shows that there is a clear overlap between the red and blue circles. This suggests that there is a low difference in the MFCC features space between the two tones (reverberant and non-reverberant). While there is less overlap between the red and blue circles in Fig.9's PCA differentiation of GFCC between the two cases of reverberant and non-reverberant. These findings support the above-mentioned NCC content contour. Thus, it is proposed that we should use the GFCC instead of the MFCC for recognition, as the GFCC approach outperforms and enhances the capacity for discrimination between reverberant and non-reverberant sounds.

The combination of MFCC and GFCC coefficients results in a more obvious distinction between reverberant and non-reverberant sound, as illustrated in Fig.10, wherein uses 28 coefficients rather than 14. Compared to the GFCC or MFCC alone, the distinction between reverberant and non-reverberant was improved by almost 30% where the discrimination exploited the accuracy of MFCC and noise overcome by GFCC. As a result, by merging multiple types of discrimination, this concept will reveal new viewpoints and aspects of discrimination.

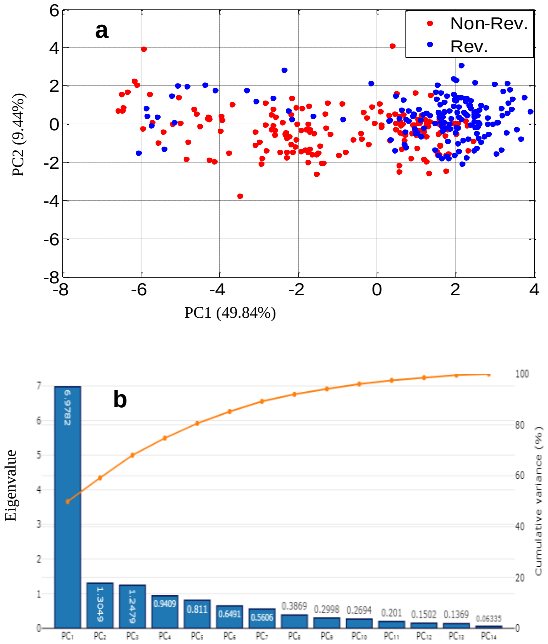

Fig.8 a- PCA characteristic patterns of MFCC. b- Eigenvalue of variances

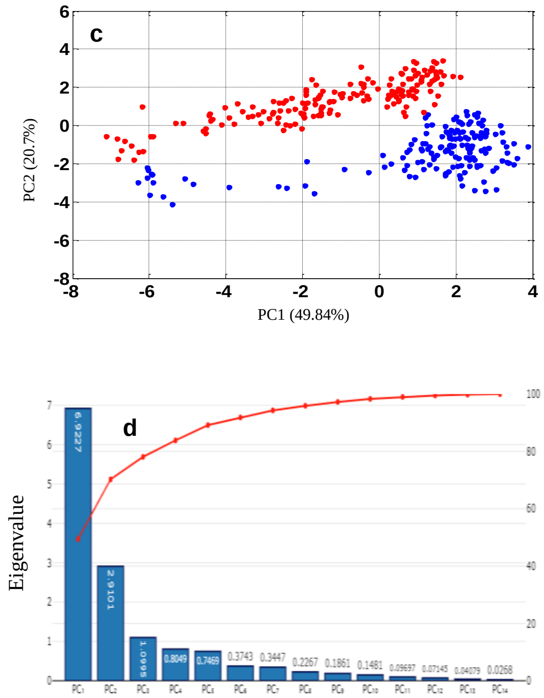

Fig.9 a- PCA characteristic patterns of GFCC. b- Eigenvalue of variances

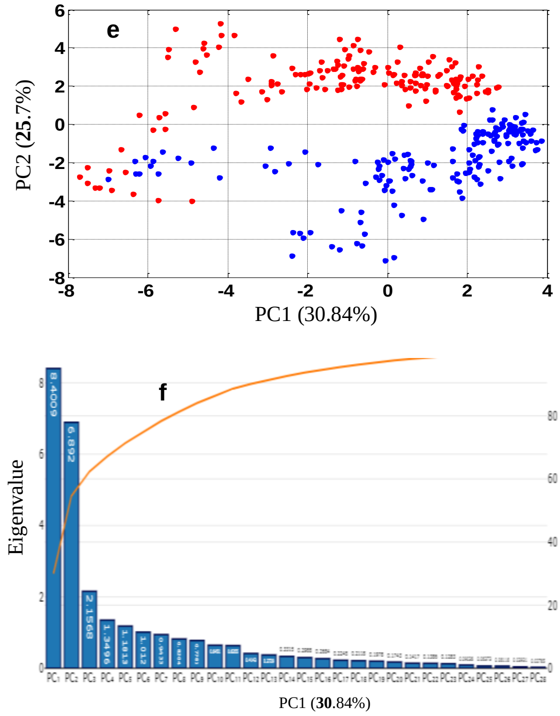

Fig.10 a- PCA characteristic patterns of GFCC. b- Eigenvalue of variances

To get a clear view of the amount of variation in the above three cases for PC1 and PC2 components, table I shows the variation amount in the three cases of MFCC, GFCC, and MFCC with GFCC combined. The obtained variance by the GFCC method is better than MFCC and the combining case has a higher variance than MFCC and GFCC alone, where the improvement by about 30% than GFCC. In the combining case, the rate of 54% represents the rate from 28 coefficients, but other rates were from 14 coefficients.

Table 1: Variance of PCA main components of used methods

Coeff. type	PC1	PC2	PC1 and PC2 Comulative Var.
MFCC Var.	49%	9%	58% from 14 coeff.
GFCC Var.	49%	20%	69% from 14 coeff.
MFCC with GFCC Combine Var.	30%	24%	54% from 28 coeff.

12. RELATIVE TO PUBLISHED WORKS

When assessing and describing speaker recognition skills, the MFCC and GFCC can be helpful in an effort to improve the SV's effectiveness, they can also be used to evaluate other aspects. Table II presents the caomparison between the proposed and published studies.

Table II Variance of PCA main components in another work

Coeff. Type [32]	PC1	PC2	PC1 and PC2 Comulative Variance
MFCC Var.	25%	15%	40% from 15 coeff.
LPCC Var.	45%	20%	64% from 15 coeff.
MFCC+GMM Var.	29%	28%	57% from 10 coeff.

13. CONCLUSION

The methods for MFCC and GFCC feature extraction have been discussed and evaluated. When compared to the known MFCC approach, the suggested system, which employs the GFCC feature methodology, has demonstrated adequate adaptability and resilience in reverberation settings. As a result, the findings demonstrate that using the GFCC to increase recognition rates at low SNR levels can be a practical approach. To explore the performance of GFCC over MFCC for feature extraction, the PCA algorithm that is used for dimensionality reduction is applied where each one has 14 features. The PCA appears a clear image of the difference between GFCC and MFCC for rev. and non-rev. tone detection of 1khz at different locations. The MFCC produces variance between PC1 and PC2 of about 58% whereas GFCC produces a 69% variance. To confirm these findings, the NCC algorithm, which yielded a 10% improvement in GFCC over the MFCC method, is also employed for discrimination.

In addition, the enhancing performance through the development of more effective features through combining techniques, such as merging MFCC and GFCC feature extraction methods, was accomplished, where 28 coefficients were used instead of 14 and 30% improvement was obtained. In the future work, can explore alternative feature sets and integrate other types of methods and techniques to achieve an adaptive system for the speaker and voice recognition field.

REFERENCES

[1] W. S. J. MOHN, “STATISTICAL FEATURE EVALUTION IN SPEAKER IDENTIFICATION.,” 1970.

[2] F. Bimbot et al., “A tutorial on text-independent speaker verification,” EURASIP J. Adv. Signal Process., vol. 2004, pp. 1–22, 2004. doi.org/10.1155/S1110865704310024

[3] J. P. Campbell, “Speaker recognition: A tutorial,” Proc. IEEE, vol. 85, no. 9, pp. 1437–1462, 1997. doi: 10.1109/5.628714

[4] T. Kinnunen and H. Li, “An overview of text-independent speaker recognition: From features to supervectors,” Speech Commun., vol. 52, no. 1, pp. 12–40, 2010. https://doi.org/10.1016/j.specom.2009.08.009

[5] T. Rossing, Springer handbook of acoustics. Springer Science \& Business Media, 2007.

[6] R. Petrick, K. Lohde, M. Wolff, and R. Hoffmann, “The harming part of room acoustics in automatic speech recognition.,” in INTERSPEECH, 2007, pp. 1094–1097. doi: 10.21437/Interspeech.2007-112

[7] K. A Al-Karawi, A. H Al-Noori, F. F. Li, and T. Ritchings, “Automatic speaker recognition system in adverse conditions implication of noise and reverberation on system performance,” Int. J. Inf. Electron. Eng., vol. 5, no. 6, 2015. doi.org/10.7763/IJIEE.2015.V5.571

[8] M. Mohammadamini, “Robustness of DNN-based speaker recognition systems against environmental variabilities,” 2023.

[9] DJordje GROZDIĆ, S. Jovičić, D. ŠUMARAC PAVLOVIĆ, J. Galić, and B. Marković, “Comparison of Cepstral Normalization Techniques in Whispered Speech Recognition.,” Adv. Electr. \& Comput. Eng., vol. 17, no. 1, 2017. doi: 10.4316/AECE.2017.01004.

[10] A. A. Rasheed, “Intonation speech for text-dependent speaker verification,” ICCSNIS’2024, no. Sousse, TUNISIA, 2024, [Online]. Available: https://fti-tn.net/publications

[11] D. Y. Mohammed, K. Al-Karawi, and A. Aljuboori, “Robust speaker verification by combining MFCC and entrocy in noisy conditions,” Bull. Electr. Eng. Informatics, vol. 10, no. 4, pp. 2310–2319, 2021. DOI: https://doi.org/10.11591/eei.v10i4.2957

[12] B. K. Swain, M. Z. Khan, C. L. Chowdhary, and A. Alsaeedi, “SRC: Superior Robustness of COVID-19 Detection from Noisy Cough Data Using GFCC.,” Comput. Syst. Sci. \& Eng., vol. 46, no. 2, 2023. DOI: 10.32604/csse.2023.036192

[13] K. A. Y. AL-Karawi, Robust speaker recognition in reverberant condition-toward greater biometric security. University of Salford (United Kingdom), 2018.

[14] A. H. Al-Noori, K. A. Al-Karawi, and F. F. Li, “Improving robustness of speaker recognition in noisy and reverberant conditions via training,” in 2015 European Intelligence and Security Informatics Conference, 2015, p. 180. DOI: 10.1109/EISIC.2015.20

[15] K. A. Al-Karawi and F. Li, “Robust speaker verification in reverberant conditions using estimated acoustic parameters: A maximum likelihood estimation and training on the fly approach,” in 2017 seventh international conference on innovative computing technology

(INTECH), 2017, pp. 52–57. DOI: 10.1109/INTECH.2017.8102427

[16] K. A. Al-Karawi, “Mitigate the reverberation effect on the speaker verification performance using different methods,” Int. J. Speech Technol., vol. 24, no. 1, pp. 143–153, 2021. https://doi.org/10.1007/s10772-020-09780-1

[17] K. A. Al-Karawi and D. Y. Mohammed, “Early reflection detection using autocorrelation to improve robustness of speaker verification in reverberant conditions,” Int. J. Speech Technol., vol. 22, no. 4, pp. 1077–1084, 2019. https://doi.org/10.1007/s10772-019-09648-z

[18] S. Huq, “Differentiation of Dry and Wet Cough Sounds using A Deep Learning Model and Data Augmentation,” Carleton University, 2023.

[19] S. Gergen, A. Nagathil, and R. Martin, “Reduction of reverberation effects in the MFCC modulation spectrum for improved classification of acoustic signals,” in Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[20] K. A. Al-Karawi and D. Y. Mohammed, “Using combined features to improve speaker verification in the face of limited reverberant data,” Int. J. Speech Technol., vol. 26, no. 3, pp. 789–799, 2023. https://doi.org/10.1007/s10772-023-10048-7

[21] X. Chen and S. A. Zahorian, “Improving speaker verification in reverberant environments,” in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 5854–5858. DOI: 10.1109/ICASSP39728.2021.9413731

[22] T. Sun, Y. Wen, X. Zhang, B. Jia, and M. Zhou, “Gaussian Mixture Model for Marine Reverberations,” Appl. Sci., vol. 13, no. 21, p. 12063, 2023.

[23] S. Ramoji, “Supervised Learning Approaches for Language and Speaker Recognition,” Indian Institute of Science Bangalore, 2023. doi.org/10.3390/app132112063

[24] H. Taherian, Z.-Q. Wang, and D. Wang, “Deep learning based multi-channel speaker recognition in noisy and reverberant environments,” in Interspeech, 2019. doi: 10.21437.

[25] P. M. Chauhan and N. P. Desai, “Mel frequency cepstral coefficients (MFCC) based speaker identification in noisy environment using wiener filter,” in 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), 2014, pp. 1–5. DOI: 10.1109/ICGCCEE.2014.6921394

[26] M. V. Sagvekar, M. Limkar, and B. R. Rao, “Speaker Identification Using MEL Frequency Cepstral Coefficients and Vector Quatization,” 2012.

[27] J. Qi, D. Wang, J. Xu, and J. Tejedor, “Bottleneck features based on gammatone frequency cepstral coefficients.,” in Interspeech, 2013, pp. 1751–1755.

[28] W. Burgos, “Gammatone and MFCC features in speaker recognition,” 2014. DOI: 10.13140/RG.2.2.25142.29768

[29] L. R. Rabiner and B.-H. Juang, Fundamentals of speech recognition. Tsinghua University Press, 1999.

[30] D. A. Reynolds, “An overview of automatic speaker recognition technology,” in 2002 IEEE international conference on acoustics, speech, and signal processing, 2002, pp. IV--4072. DOI:10.1109/ICASSP.2002.5745552

[31] M. Kim, E. Kim, C. Seo, and S. Jeon, “Speaker verification and identification using principal component analysis based on global eigenvector matrix,” in Hybrid Artificial Intelligence Systems: 5th International Conference, HAIS 2010, San Sebastián, Spain, June 23-25, 2010. Proceedings, Part I 5, 2010, pp. 278–285.

[32] A. I. Ahmed, J. P. Chiverton, D. L. Ndzi, and V. M. Becerra, “Speaker recognition using PCA-based feature transformation,” Speech Commun., vol. 110, pp. 33–46, 2019. https://doi.org/10.1016/j.specom.2019.04.001

تحليل MFCC و GFCC لتحديد تاثير الصدى على الصوت

عبدالعليم عبدالفتاح رشيد

قسم الهندسة الكهربائية, كلية الهندسة, جامعة الموصل, الموصل, العراق

alem12@uomosul.edu.iq

تاريخ الاستلام:18 فبراير2024   استلم بصيغته المنقحة: 12 مايو 2024   تاريخ القبول:10 يوليو 2024

الملخص

تحتوي غالبية الاشارات الصوتية المستلمة على اشارات الصدى مضافة الى الاشارة الاصلية, مما يؤدي الى تقليل وثوقية انظمة الصوت. وهذا له تأثيرات ضارة على مجموعة متنوعة من تطبيقات تحديد الهوية, بما في ذلك مجال تحديد والتعرف على المتحدثين.

تم في هذا البحث تحليل تقنيتين لتخفيف ومكافحة تأثير الصدى على الصوت ومقارنة اداء هذه الاساليب. هذه التقنيات ال MFCC *وتقنية* GFCC *. تختلف تقنية* GFCC *عن تقنية* MFCC *وذلك بأبدال مرشحات* Mel *بمرشحات* gammatone .

لتجنب تأثير الحالات المختلفة لصوت التحدث مثل المرض والمشاعر, تم تطبيق نغمة واحدة بتردد 1 كيلو هيرتز للحصول على مقارنة عادلة بين MFCC *و* GFCC *والتي تعتبر من أشهر طرق تمييز الصوت.*

تم اجراء المقارنة بين هاتين الطريقتين وذلك باستخدام تحليل المكونات الاساسية PCA *وتم تأكيدها بواسطة الارتباط المتبادل الاعتيادي* NCC *. تعتمد طريقة* PCA *على اختزال المعلومات وحذف الارتباط. أظهرت تحليلات* PCA *و* NCC *أنه بالنسبة للصوت المسجل بوجود وعدم وجود الصدى, كانت هناك زيادة بنسبة 10% تقريبا في معدل الكشف لطريقة* GFCC *مقارنة بطريقة* MFCC .

بذلك يوضح هذا البحث ان طريقة GFCC *تعتبر متفوقة على اسلوب* MFCC *التقليدي في التغلب على ضوضاء الصدى. وكنتيجة, فان اسلوب* GFCC *يخفف من تأثير الصدى ويقدم اختيارا افضل للعمل في انظمة التعرف على المتحدثين.*

الكلمات الداله :

MFCC، GFCC، الصوت، PCA، التمييز.