International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 45 No. 2 Apr. 2025, pp. 439-447

http://www.ijias.issr-journals.org/

MFCC number limit for automatic sound recognition: Application to the chainsaw sound identification

N'Tcho Assoukpou Jean GNAMELE[1], Bi Tra Jean Claude YOUAN[2], and Konan Fernand GBAMELE[3]

[1]Department of Mathematics-Physics-Chemistry at University Peleforo Gon Coulibaly (UPGC), BP 1328 Korhogo, Côte d'Ivoire

[2]Technology Laboratory, Félix Houphouët-Boigny University (UFHB), 01 BP V34 Abidjan 01 Cocody, Côte d'Ivoire

[3]Optical Networks and Telecommunications Laboratory of the Joint Research and Innovation Unit in Engineering Science and Technology (UMRI-STI), National Polytechnic Institute Félix Houphouët-Boigny (INPHB), Bp 1093 Yamoussoukro, Côte d'Ivoire

ABSTRACT: This paper investigates the relevance of using a large number of Mel Frequency Cepstral Coefficients (MFCC) as descriptors of acoustic signals, and the interaction between these and the nature of the frequency band in which the Mel filters are arranged.
This study forms part of the wider field of automatic recognition of acoustic signals, with a particular focus on those that are not speech-related. We evaluated a series of MFCCs, spanning a range from 1 to 50, utilising the central octave band frequencies (31.5 Hz-16000 Hz) as the MFCC calculation frequencies. An application was made to the identification of chainsaw sounds among a plurality of signals from the forest environment.
The results revealed a threshold value for the number of MFCCs (LVMFCC) above which classification rates remain constant. The LVMFCC=39 was common to all frequencies, although specifically the LVMFCC for each centre frequency was between 5 and 39 MFCCs. We observed that the notion of an optimal value for the number of MFCCs could appear subjective. The best classification rate of 98.41% obtained with the 16000 Hz centre frequency corresponds to a number of MFCCs between 5 and 50. These results also reveal the need to restructure the.

KEYWORDS: acoustic, automatic recognition, KNN, MFCC, octave band.

1 INTRODUCTION

In the context of automatic recognition of speech and other signals not related to human language, several studies have demonstrated that 13 Mel frequency cepstral coefficients are an optimal choice for signal identification [1], [2]. However, numerous recent studies employ a comparatively large number of MFCCs in acoustic signal processing.

The rationale behind the selection of this extensive set of MFCCs remains opaque, with numerous scholars failing to elucidate the frequency band utilized in their derivation.

Indeed, in reference [3], the authors employ a convolutional deep learning model for the classification of respiratory diseases based on patients' respiratory sound signals with MFFCs as descriptors. There are 32 MFFCs in total. In this paper [4], the authors present a neural sensor fusion framework for drone detection based on audio and video data. With regard to the audio stream, the authors evaluate the Long Term Memory (LSTM) and Convolutional Recurrent Neural Network (CRNN) models and demonstrate the superiority of the CRNN model, which is based on Mel Frequency Cepstral Coefficients (MFCC) features, of which there are 40. In [5], the authors put forth a multimodal approach to traffic light status detection, employing both visual and auditory cues, as seen from the perspective of a quadruped robot traversing an urban setting. A cross-comparison between N MFCCs (with N ranging from 10, 12, 14, 16, 18, 20, 24 and 28) revealed that 24 MFCCs were optimal.

In their work [6], the authors examine the resilience of fully supervised convolutional neural networks (CNNs) and cutting-edge transfer learning methodologies within the domain of automatic speech recognition. The authors assess the influence of the number of Mel Frequency Cepstral Coefficients (MFCCs) on the word error rate (WER) by considering a range of MFCCs, from 10 to 24.

This study [7] presents an optimal method for the classification of cardiac noise based on machine learning technologies, with the objective of predicting cardiovascular diseases. In this study, the number of MFCC descriptors employed was 13, 25 and 42.

In another study [8], a novel approach to oestrus sound recognition was proposed, utilizing a fusion of two representative features as inputs and convolutional neural networks (CNNs) as the training model. The descriptors employed were MFCCs, with a total of 30, 40 and 50 coefficients.

In reference [9], a model of sound quality within high-speed trains (HSTs) is proposed. The model combines MFCC with convolutional neural networks (CNN) for the assessment of sound quality in high-speed trains (HSTs). The number of frequency cepstral coefficients is 26.

In [10], the authors assess the efficacy of the convolutional neural network in discerning the diverse spectral attributes of acoustic signals, utilizing two widely-used open databases: The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and the Berlin Database of Emotional Speech (EmoDB) were employed in this study. Five spectral representations are extracted from the original audio files, comprising 20 MFCC.

In the aforementioned literature, the authors do not explicitly indicate the reasons justifying the choice of MFCCs and/or do not mention the frequency bands used to obtain these coefficients. However, two key characteristics emerge in the calculation of MFCCs:

The frequency at which the signals are studied, namely the frequency at which Mel's filter banks are arranged.

The number of MFCCs.

In light of this work, two questions arise:

It would be beneficial to ascertain whether there is a dependency between the number of Mel frequency cepstral coefficients and the frequency band of the Mel filter arrangement for a given study.

Does the number of Mel frequency cepstral coefficients reach a point where it is no longer a determining factor in the recognition of study samples?

In this study, we address these questions by applying the aforementioned methodology to the identification of chainsaw samples among a number of environmental signals likely to be perceived in a forest environment. The following section presents the methodology employed in this study.

Section 2 outlines the methodology employed in this study. Section 3 presents the results, while Section 4 offers a discussion of these findings. The paper concludes in Section 5.

2 METHODS

2.1 MATERIALS

Given that the object of study is acoustic signals, it is necessary to construct a database that will be employed both for training our classifiers and for evaluating their performance. The study is based on two distinct labeling classes. The training phase comprises two classes. The chainsaw class comprises the acoustic signals produced by chainsaws, while the forest class encompasses all other sounds. The chainsaw class comprises audio recordings of chainsaws made in the Armainvilliers Sunday forest (Gretz-Armainvilliers, France) and chainsaw sounds sourced from an online database. The forest class is constituted by the sounds of diverse fauna and avifauna, as well as vehicle and meteorological sounds, including the auditory signature of precipitation. The sounds were sourced from online databases.

The test phase comprises chainsaw sounds, recorded at two forestry sites in Côte d'Ivoire: the Yapo-Abbe classified forest (Agboville, Côte d'Ivoire) and the National Floristic Centre (Abidjan, Côte d'Ivoire).

The recordings made at the forestry centers were made using a DR-05 dictaphone with a sampling rate of $44.1\ kHz$. The distance between the transmission source and the dictaphone was between 10 and 100 meters.

A total of 1400 sound samples were utilized for the training phase, while 565 sound samples were employed for the test phase. All of the sound samples are in the WAV format and are 5 seconds in duration, with a stereo recording. The distribution of sound samples by phase and class is presented in Table 1.

Table 1. Breakdown of sound samples by phase and class

Training phase		***Test phase***
Forest class	*Chainsaw class*	*Chainsaw class*
1043	*357*	*565*

2.2 Tasks And Methods

2.2.1 Mel Frequency Cepstral Coefficients

The Mel-frequency cepstral coefficient (MFCC) is a widely utilised feature that has been employed in numerous applications, particularly in sound signal processing [11]. The MFCC can be calculated by conducting five consecutive processes. These are: signal framing, computing of the power spectrum, application of a Mel filter bank to the obtained power spectra, calculation of the logarithm values of all filter banks, and finally, application of the DCT (Discrete Cosine Transform) to Mel's frequency spectrum.

2.2.2 Signal Structure For MFCC Calculation

The acoustic signal employed to obtain Mel's frequency cepstral coefficients is the result of merging the two channels of the stereo signal obtained after recording. This approach is analogous to that developed in [1]. This method of restructuring the study signal will be compared with the case where a single channel is used to obtain the MFCCs. This comparison will enable us to assess its impact on the results obtained.

The resulting signal is as follows:

$$s_{resultant}(n) = \left[C_{L_1}; C_{R_1}; C_{L_2}; C_{R_2}; \ldots; C_{L_{m-1}}; C_{R_{m-1}}; C_{L_m}; C_{R_m}\right] \quad (1)$$

- C_{L_i} left channel component
- C_{R_i}: left channel component
- m: number of samples in the signal frame.

This restructuring of the acoustic signal is applicable to both the test phase and the training phase samples.

The resulting signal, designated as $s_{resultant}(n)$, is then segmented into a one-second signal, as illustrated in Fig. 1, with no overlap between the sub-frames $s_{resutant_i}(n)$).

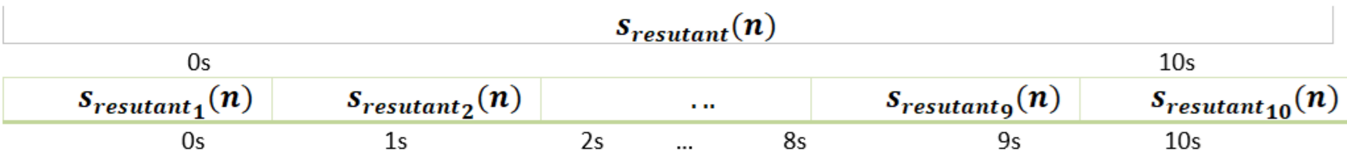

Fig. 1. Segmentation of the signal resulting from merging the left and right channels

We compute the MFCCs for each of the 10 "subframes" $s_{resutant_i}(n)$, for which we keep the first N MFCCs. We use the $10 \times N$ coefficients obtained to construct the descriptor vector of the signal. This vector descriptor is of type $1 \times 10N$.

2.2.3 The Frequency Band For The Mel Filter Bank

The configuration of the Mel filter banks is based on the center frequencies of the octave bands, which range from $31.5\ Hz$ to $16\ kHz$. The Mel filter banks will be evaluated by assessing the center frequencies of each octave band. Octave band analysis is a valuable tool in sound measurement as it offers an accurate representation of the human auditory response. This is an intriguing aspect as it aligns with the fundamental premise behind MFCC. The selected number of Mel filters is 35 [1].

2.2.4 K- NEIGHBOURHOOD ALGORITHM (KNN)

K-Nearest Neighbors (KNN) is a relatively straightforward machine learning algorithm that is commonly employed for classification purposes [12]. The algorithm is based on the minimum distance between the test point and all training points, which may be calculated using a variety of distance metrics, such as the Euclidean distance. Subsequently, the class of the test point is determined by the most frequent class of the k nearest neighbors of the test point. The most commonly utilized distances include: Consequently, the value of k is of significant importance and should be meticulously planned, as a low k value could potentially result in overfitting. The value of k is three (03) [1].

3 RESULTS

3.1 STUDY OF THE INFLUENCE OF THE NUMBER OF THE MFCC ON THE OCTAVE FREQUENCY BAND

Figs. 1 to 3 illustrate the impact of augmenting the number of Mel's frequency cepstral coefficients on the classification rate for each octave band center frequency. The curves in these figures are organized not on the basis of the central frequencies, but rather on the basis of closely related classification rates.

These figures facilitate an understanding of the MFCC values for which the dependency between the number of MFCCs and the classification rate is disrupted.

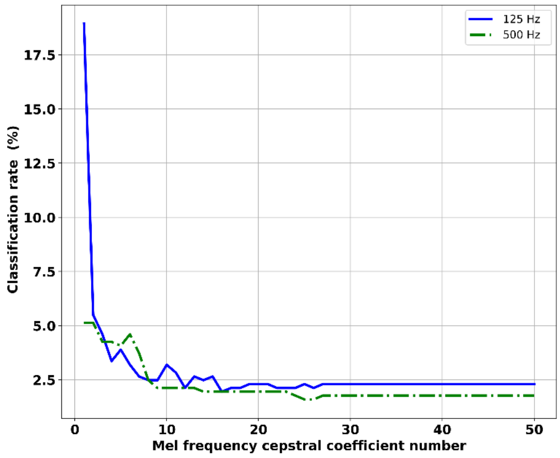

Fig. 2. Classification rate according to MFCC number for 125 Hz, 500 Hz

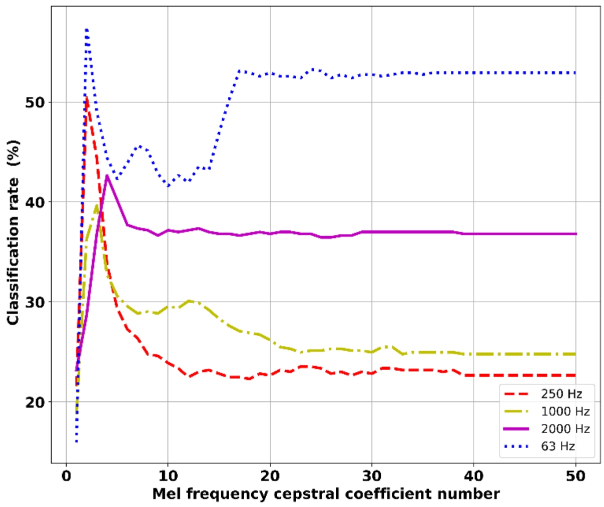

Fig. 3. Classification rate according to MFCC number for 63 Hz, 250 Hz, 1000 Hz, and 2000 Hz

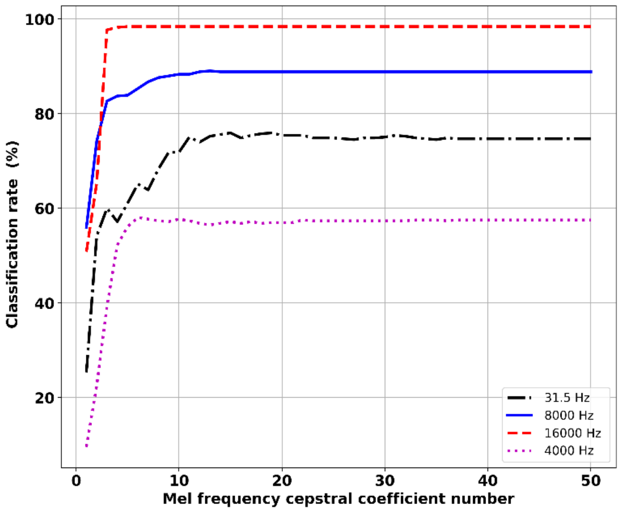

Fig. 4. Classification rate according to MFCC number for 31.5 Hz, 4000Hz, 8000 Hz, and 16000 Hz

The evolution of the classification rate as a function of the number of MFCCs can be divided into two phases: a "***constant***" phase and a ***"quasi-constant"*** phase.

The "***constant***" phase is characterized by a point on the curve at which there is no further change in the value of the classification rate from the given value of the number of MFCCs. This value of the MFCCs number is designated the limit value of the number of MFCCs (LVMFCC). Table 2 illustrates the LVMFCC for each octave band center frequency.

Table 2. "constant" phase

octave band centre frequency (Hz)	31.5	63	125	250	500	1000	2 000	4 000	8 000	16 000
range of cepstral coefficient number	≥ 37	≥ 36	≥ 27	≥ 39	≥ 27	≥ 39	≥ 39	≥ 37	≥ 14	≥ 5
LVMFCC classification rate(%)	74.69	52.92	2.3	22.65	1.77	24.78	36.81	57.52	88.85	98.41

The classification rates presented in Table 2 do not necessarily correspond to the highest classification rates for the specified frequency bands.

The ***"quasi-constant"*** phase results in the appearance of the LVMFCC, which is then followed by a variation in the classification rate up to the ***constant*** phase. In the quasi-constancy phase, multiple LVMFCC values may be observed. The selected value is that for which the rate of change is less than 1%. This variation in the rate is attributable to a small number of values for the number of cepstral coefficients, which exhibit a change in the classification rate. Table 3 illustrates the values for each central frequency, as well as the discrepancies between the peak of this fluctuation and the minimum value.

The values of the intervals vary according to the center frequency under consideration, encompassing a range of 11 to 40 Mel frequency cepstral coefficients.

It would appear that, contingent on the nature of the frequency band (see Table 2), there is a clearly defined LVMFCC for each octave band center frequency.

For a number of coefficients equal to or greater than, the classification rates are contingent upon the nature of the frequency band in which the Mel frequency cepstral coefficients are obtained: it is found to remain constant. We observe that the classification rates are not identical for all study frequencies. Indeed, the highest classification rate of is achieved for the central frequency of.

Table 3. Fluctuation in the classification rate: "quasi-constant" phase

octave band centre frequency (Hz)	*31.5*	*63*	*125*	*250*	*500*	*1000*	*2000*	*4000*	*8000*	*16000*
range of cepstral coefficient number	≥ 26	≥ 18	≥ 19	≥ 20	≥ 24	≥ 33	≥ 15	≥ 22	≥ 12	≥ 5
fluctuation in classification rates (%)	*0.89*	*0.88*	*0.18*	*0.89*	*0.18*	*0.18*	*0.53*	*0.18*	*0.18*	*0*

3.2 The Advantages Of Signal Restructuring: The Combination Of The Left And Right Audio Channels

The results displayed in Fig. 2 to 4 were derived by integrating Channels 1 and 2, as illustrated in (1). This approach has an impact on the quality of the information extracted from the signal under study, which in turn affects the Mel frequency cepstral coefficients obtained. Fig. 6 illustrates the classification rates obtained and the LVMFCC when the channels are not merged, but a single channel is utilised for the calculation of the MFCCs.

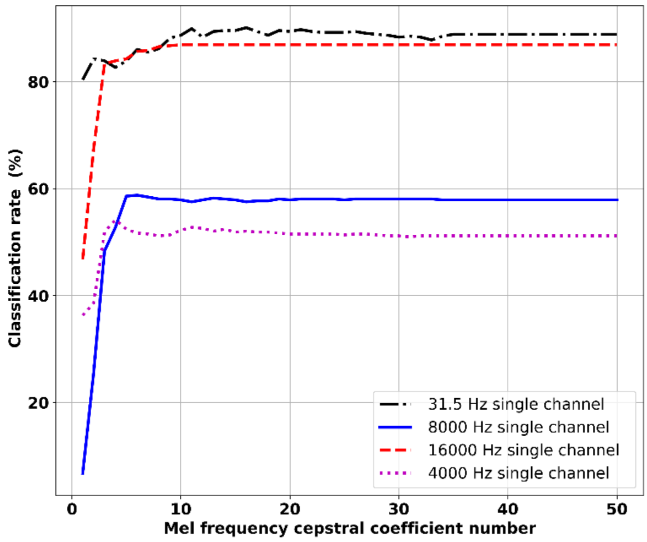

Fig. 5. Classification rate according to MFCC number for 31.5 Hz, 8000 Hz, and 16000 Hz single channel

The signal then assumes the following form:

$$s_{one-channel}(n) = \left[C_{L_1}; C_{L_2}; \dots; C_{L_{m-1}}; C_{L_m}\right] \quad (2)$$

The signal, $s_{one-channel}(n)$ is then divided into a one-second signal, as demonstrated in Fig. 6, with no overlap between the $s_{one-channel_i}$ subframes.

The MFCCs are computed for each of the 5 "subframes" $s_{one-channel_i}$, with the first N MFCCs retained for each.

The $5 \times N$ coefficients are then employed in the construction of the signal descriptor vector. This vector descriptor is of type 1 ×5N. In this phase, only the frequencies illustrated in Fig. 3 are considered, given their high classification rate.

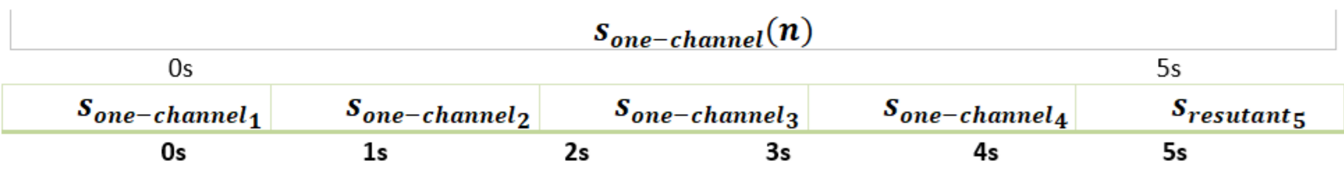

Fig. 6. Segmentation of one channel

We observe a change in the LVMFCC. The classification rate for the central frequency of $31.5\ Hz$ exhibits a notable increase, reaching values of ((14.16% at the maximum) and (14.16% at LVMFCC)), in comparison with the classification rates observed for the other two frequencies.

The classification rates for $8000\ Hz$ and $16000\ Hz$ exhibit a notable decline and Are respectively (30.27% (maximum value) and 30.97% (upper LVMFCC)) and (11.51% (maximum value) and 11.51% (upper LVMFCC)).

Additionally, an increase in the LVMFCC is observed for $8000\ Hz$ and $16000\ Hz$, while this value decrease for $31.5\ Hz$. Table 4 presents a comparison of the LVMFCC values obtained through the application of Eqn. (1) and Eqn. (2).

Table 4. comparative analysis of the values for channel fudging and for a single channel

octave band centre frequency (Hz)	*Merging of channels 1 & 2*			*One channel*		
	LVMFCC classification rate (%)	*The maximum classification rate (%)*	*LVMFCC*	*LVMFCC classification rate (%)*	*The maximum classification rate (%)*	*LVMFCC*
31.5	*74.69*	*75.93*	*37*	*88.85*	*90.09*	*35*
8000	*88.85*	*89.03*	*14*	*57.88*	*58.76*	*34*
16000	*98.41*	*98.41*	*5*	*86.9*	*86.9*	*10*

The LVMFCC is applicable at frequencies of $16000\ Hz$ and $8000\ Hz$, with the former being approximately twice as high as the latter. It is therefore evident that the structure of the processed signal represents a significant factor in the calculation of relevant characteristics for MFCCs.

4 DISCUSSION

It is important to indicate the specific frequency band employed to obtain the MFCCs, as illustrated in Figs. 1 to 3, during the calculation process. This value has a substantial impact on the quality of information extracted from the signal and, consequently, on the precision and reliability of the obtained results. However, the articles [1-9] do not address this factor. Information is provided regarding the potential filters for noise reduction [13], [14], the frequency band containing the signal information [15], and the sampling rate [13], [16], [17]. While this information is undoubtedly valuable, it does not provide a definitive indication of the precise frequency at which Mel's filter banks should be positioned. Nevertheless, the optimal value for the number of Mel frequency cepstral coefficients can be determined based on the specific characteristics of the frequency band in question.

The objective of the descriptor-classifier combination in the context of pattern recognition is twofold: firstly, to optimize the expected results and secondly, to reduce the information processing time.

The results demonstrate that there is a threshold beyond which the number of MFCCs is no longer indicative of relevant information. This results in more complex calculations, which consequently reduces the time taken to process the information.

Discussing the optimal number of MFCCs for achieving favorable outcomes becomes inherently subjective when we extend beyond the VLMFCC. This aspect does not provide sufficient justification for the selection of the number of MFCCs, particularly in light of the findings presented by the aforementioned authors. The number of MFCCs is a value intrinsic to the choice of study frequency and to signal processing linked to the structure of the signal (fusion of data, one channel, etc.). Therefore, a generalization of its value will not be appropriate without well-defined information.

We note that the concept of an optimal value for the number of MFCCs [5], may appear to be overly subjective in certain configurations. With regard to the center frequency of $16\ kHz$, the highest classification rate (98.41%) is observed for a number of MFCCs between 5 and 50. The number of MFCCs equal to 5 is not *the optimum* value, but rather *a value that optimises the classification rate*. This is particularly evident when considering that all the MFCC values greater than 5 give the same optimum classification rate. This observation was made for the $31.5\ Hz$ frequency, where the number of coefficients 15 and 19 yielded the same value of 75.93%. The same is true for the 500 Hz frequency, where the maximum classification rate of 5.13% corresponds to the number of MFCCs 1 and 2.

5 CONCLUSION

The selection of the number of Mel cepstral coefficients appears to be a parameter contingent upon the specific frequency band that has been selected. The values of the classification rates are inherently contingent upon these variables and necessitate the communication of their intrinsic nature. In the field of automatic sound recognition, particularly in the context of chainsaw signals, there exists a value for the number of Mel Cepstral Coefficients (MFCCs) that exceeds a threshold beyond which the interaction between the number of MFCCs and the nature of the frequency band is no longer linear (LVMFCC). A study of octave band center frequencies ($31.5\ Hz - 16\ kHz$) has revealed limit values for the number of MFCCs in the range $[5 - 50]$. However, for an LVMFCC of 39, the classification rate remains constant for all frequencies studied. Thus, determining the limit value for a given study frequency avoids selecting an excessively large value for the number of MFCCs, which would have no impact on classification. The sole consequence of this large number is that the calculation will become more complex, necessitating a greater number of iterations to run the identification algorithms.

The structure of the signal is contingent upon the nature of the frequency band and the selection of the number of Mel-frequency cepstral coefficients (MFCCs). To illustrate, the restructuring of an audio signal by combining the left and right channels into a single signal has been demonstrated to markedly enhance the classification rate in comparison to the utilization of a single channel of the audio acquisition for information extraction.

In consideration of these factors, the center frequency of $16\ kHz$ yields the optimal classification rate of 98.41% for a number of MFCCs of 5. Notably, this rate is consistent across all values of the number of MFCCs exceeding or equal to 5 for the same octave band centre frequency.

REFERENCES

[1] N. A. J. Gnamele, B. T. J. C. Youan, et A. M. L. Famien, « Improvement of chainsaw sounds identification in the forest environment using maximum ratio combining and classification algorithme«, *Eureka: PE,* nº 3, p. 3-16, mai 2024. doi: 10.21303/2461-4262.2024.003107.

[2] N. A. J. Gnamele, Y. Berenger, T. Arsene, G. Baudoin, et J.-M. Laheurte, « KNN and SVM Classification for Chainsaw Sound Identification in the Forest Areas«, *IJACSA,* vol. 10, nº 12, 2019, doi: 10.14569/IJACSA.2019.0101270.

[3] N. Babu, J. Kumari, J. Mathew, U. Satija, et A. Mondal, « Multiclass Categorisation of Respiratory Sound Signals Using Neural Network«, in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS),* Taipei, Taiwan: IEEE, oct. 2022, p. 228-232. doi: 10.1109/BioCAS54905.2022.9948675.

[4] I. Alla, H. B. Olou, V. Loscri, et M. Levorato, « From Sound to Sight: Audio-Visual Fusion and Deep Learning for Drone Detection«, in *Proceedings of the 17th ACM Conference on Security and Privacy in Wireless and Mobile Networks,* Seoul Republic of Korea: ACM, mai 2024, p. 123-133. doi: 10.1145/3643833.3656133.

[5] S. Gupta et A. Cosgun, « Audio-Visual Traffic Light State Detection for Urban Robots », 2024, *arXiv*. doi: 10.48550/ARXIV.2404.19281.

[6] A. M. Samin *et al.,* « BanSpeech: A Multi-Domain Bangla Speech Recognition Benchmark Toward Robust Performance in Challenging Conditions«, *IEEE Access,* vol. 12, p. 34527-34538, 2024, doi: 10.1109/ACCESS.2024.3371478.

[7] Y. N. Fuadah, M. A. Pramudito, et K. M. Lim, « An Optimal Approach for Heart Sound Classification Using Grid Search in Hyperparameter Optimization of Machine Learning«, *Bioengineering,* vol. 10, nº 1, p. 45, déc. 2022. doi: 10.3390/bioengineering10010045.

[8] P. Chen, D. Yin, B. Yang, et W. Tang, « A Fusion Feature for the Oestrous Sow Sound Identification Based on Convolutional Neural Networks », *J. Phys.: Conf. Ser.,* vol. 2203, nº 1, p. 012049, févr. 2022, doi: 10.1088/1742-6596/2203/1/012049.

[9] P. Ruan, X. Zheng, Y. Qiu, et Z. Hao, « A Binaural MFCC-CNN Sound Quality Model of High-Speed Train«, *Applied Sciences,* vol. 12, nº 23, p. 12151, nov. 2022, doi: 10.3390/app122312151.

[10] M. H. Pham, F. M. Noori, et J. Torresen, « Emotion Recognition using Speech Data with Convolutional Neural Network«, in *2021 IEEE 2nd International Conference on Signal, Control and Communication (SCC),* Tunis, Tunisia: IEEE, déc. 2021, p. 182-187. doi: 10.1109/SCC53769.2021.9768372.

[11] Z. Kh. Abdul et A. K. Al-Talabani, « Mel Frequency Cepstral Coefficient and its Applications: A Review », *IEEE Access,* vol. 10, p. 122136-122158, 2022, doi: 10.1109/ACCESS.2022.3223444.

[12] Y. Alsouda, S. Pllana, et A. Kurti, « IoT-based Urban Noise Identification Using Machine Learning: Performance of SVM, KNN, Bagging, and Random Forest », in *Proceedings of the International Conference on Omni-Layer Intelligent Systems,* Crete Greece: ACM, mai 2019, p. 62-67. doi: 10.1145/3312614.3312631.

[13] F. Li, Z. Zhang, L. Wang, et W. Liu, « Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning«, *Front. Physiol.,* vol. 13, p. 1084420, déc. 2022, doi: 10.3389/fphys.2022.1084420.

[14] B. Al-Naami, H. Fraihat, J. Al-Nabulsi, N. Y. Gharaibeh, P. Visconti, et A.-R. Al-Hinnawi, « Assessment of Dual-Tree Complex Wavelet Transform to Improve SNR in Collaboration with Neuro-Fuzzy System for Heart-Sound Identification«, *Electronics,* vol. 11, nº 6, p. 938, mars 2022, doi: 10.3390/electronics11060938.
[15] W. Hu, S. Feng, B. Zhang, Y. Gao, X. Xiao, et Qi-Huang, « Hybrid feature extraction method of MFCC+GFCC helicopter noise based on wavelet decomposition«, *J. Phys.: Conf. Ser.,* vol. 2478, nº 12, p. 122008, juin 2023. doi: 10.1088/1742-6596/2478/12/122008.
[16] H. Yan, H. Bai, X. Zhan, Z. Wu, L. Wen, et X. Jia, « Combination of VMD Mapping MFCC and LSTM: A New Acoustic Fault Diagnosis Method of Diesel Engine », *Sensors,* vol. 22, nº 21, p. 8325, oct. 2022, doi: 10.3390/s22218325.
[17] L. P. Sahu, G. Pradhan, et J. P. Singh, « Modeling Sub-Band Information Through Discrete Wavelet Transform to Improve Intelligibility Assessment of Dysarthric Speech », *IJIMAI,* vol. 7, nº 7, p. 56, 2022, doi: 10.9781/ijimai.2022.10.003.