ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV1194

Mel Frequency Cepstral Coefficients Properties Optimization Due to Ultrasonic Bands and Data Structure: Application to Acoustic Signals Identification

Bi Tra Jean Claude YOUAN[1]
Laboratory of Technology (Labtech),
Felix Houphouet-Boigny University
Abidjan Autonomous District, Côte d'Ivoire

N'tcho Assoukpou Jean GNAMELE[2]
Mathematics-Physics-Chemistry
University Peleforo Gon Coulibaly (UPGC)
Korhogo, Côte d'Ivoire

Digrais Moïse MAMBE
Computer Science and Telecommunications Research Laboratory,
Nangui Abrogoua University, Abidjan Autonomous District, Côte d'Ivoire

Abstract:- In the work presented in this article, we highlight the interest of choosing low-frequency ultrasound for the calculation of Mel cepstral coefficients combined with a singular restructuring of the study data. These coefficients were used as descriptors in the classification of sound samples produced by chainsaws in forest environments in order to combat the destruction of Ivorian fauna and flora. Three restructuring methods were compared, namely: the Time Domain Channel Fusion method, the cepstral Domain Channel Fusion method and the one channel method. To do this, we first calculated the MFCC on different frequency bands in the acoustic band [170 Hz-22000 Hz]. The different frequency bands selected range from 1 kHz to 21 kHz, increasing by 2 kHz at each new calculation phase. Low-frequency ultrasound produced better classification rates than the other acoustic bands. The best rate of 98.40% was obtained for the 3 kHz bandwidth on the acoustic band [21170 Hz-24170 Hz] combined with the 'Time Domain Channel Fusion' method. A study of the ultrasounds deduced from the central frequencies of the octave bands was then carried out. A comparative approach of the sample classification rates led to selecting the band [11313 Hz - 22627 Hz] deduced from the central frequency of the 16 KHz octave band as the best ultrasonic band for the calculation of the MFCCs.

Keywords:- *Ultrasound, Octave Band, KNN,, Data Structure.*

I. INTRODUCTION

The use of Mel's frequency cepstral coefficients in pattern recognition is a commonly used approach due to the better results obtained [1]. However, as the results obtained in the article [2] show, the effectiveness of a classification algorithm intrinsically depends on the appropriate choice of classifier and descriptor parameters. The results of this work show a very large variability in classification rates depending on the choice of these parameters.

In this sense, although offering better results, the choice of MFCCs should only be made in the light of the nature of the acoustic band defined to obtain them. Such a study is all the more important when the work concerns the classification of wide-band acoustic samples of great diversity, as is the case with forest environments.

In their work [2], the authors highlighted the interaction that existed between the choice of descriptors and that of classifiers with a view to defining an optimal algorithm for identifying the sounds emitted by chainsaws in a forest environment with a view to designing an effective system for combating the destruction of the forest cover in Cote d'Ivoire. This study was based on a comparative approach between KNN and SVM coupled with MEL frequency cepstral coefficients (MFCC).

The classification rates of the samples obtained were 40.51% and 53.16% for KNN and SVM respectively. The frequency bands used to obtain these rates were [8170 Hz-9170 Hz] and [7170 Hz-8170 Hz] and are located in the very high frequency range with a bandwidth of 1 kHz. Increasing the frequencies used to obtain MFCCs seems to provide a clear advantage in terms of sample classification. We are therefore interested in much higher frequencies, namely ultrasound.

Ultrasound is used in a variety of fields: In obstacle detection [3][4], in industry [5][6], and also for the study of communications in animals [7][8][9], to name but a few examples.

We will look at this in another new context, namely the extraction of relevant characteristics from the sound samples under study to facilitate their classification.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV1194

➢ *Three (03) Criteria will be of Particular Interest:*

- The acoustic band width,
- The nature of the acoustic bands, and more precisely the frequencies that define them
- Data structure for MFCC calculation

The objective of this study is to analyze the results obtained by studying the possibilities that could be offered by the calculation of MFCC by taking into account ultrasound and the restructuring of acoustic signals. The presentation of this work is as follows: Section 2 presents the methodology of our study. Section 3 is devoted to the results and their interpretation, followed by a discussion phase in 4 and then the conclusion in 5

II. METHODOLOGY

A. Description of Data from the Training and Test Phases

Since the object of study is acoustic signals, we need to build a database that will be used both to train our classifiers and to test them. Our study is based on two (2) labelling classes. The training phase consists of two classes: The class of chainsaw sounds that we refer to as the chainsaw class and the class of other sounds referred to as the forest class. The chainsaw class is made up of sounds recorded in the Armainvilliers Sunday forest (Gretz-Armainvilliers, France) and chainsaw sounds from an online database. The forest class is made up of the sounds of various animals and birds, vehicle sounds and meteorological sounds (rain). These sounds all come from online databases [1].

The test phase is as follows: The chainsaw class: this consists of sounds recorded at two forestry sites in Côte d'Ivoire: the Yapo-Abbe classified forest (Agboville, Côte d'Ivoire) and the national floristic center (Abidjan, Côte d'Ivoire).

The forest classroom: here we find a variety of animal and bird sounds, motorbike sounds and storm sounds. All these sounds come from online databases.

The recordings we made in the forestry centers were made using the DR-05 Dictaphone with a sampling rate of 44.1 *kHz*. The recording distances were between 10 *m* and 100 *m* between the transmission source and the Dictaphone.

We have a total of 2530 sound samples for the training phase and 1172 sound samples for the test phase. All the sound samples are stereo in WAV format and 5 seconds long. The distribution of sound samples by phase and class is given in Table 1.

Table 1 Breakdown of Sound Samples by Phase and Class

Training phase		**Test phase**
Forest class	**Chainsaw class**	**Chainsaw class**
2173	357	565

B. Description of the Calculation of Mel's Frequency Cepstral Coefficients

➢ *MEL Frequency Cepstral Coefficients*

MEL frequency cepstral coefficients (MFCC) are the most efficient descriptors used in speech-related applications [12]. The speech signal is represented as a short-term power spectrum of sound, based on the discrete cosine transform of a logarithmic power spectrum over a non-linear frequency scale of Mel. MEL frequency cepstral coefficients give better frequency resolution of low-frequency study samples. As a result, it can be applied to all types of signal and is not affected by noise. Mel's cepstral coefficients are obtained by the DCT (Discrete Cosine Transform) of Mel's frequency spectrum:

$$\boldsymbol{MFFC_k = \sum_{j=1}\left(\log_{10}(P_p)\frac{\cos(k(2j-1))\pi}{2M}\right)} \quad \textbf{(1)}$$

The frequency spectrum of Mel P_p is :

$$\boldsymbol{P_p = \sum_{v}^{\frac{N}{2}-1}\left(d_{(p,v)}F_v\right),\ \ p = 0, 1, \dots, M} \quad \textbf{(2)}$$

Where, N is the number of samples in a speech frame, $d_{(p,v)}$is the amplitude of the band-pass filter of index p for frequency v and F_v the Fourier transform of the signal [13]. The first thirteen (13) coefficients are retained to serve as relevant signal characteristics, as the lower order coefficients contain most of the information such as the average power of the input signal and the distribution of spectral energy between low and high frequencies. Selecting a larger number of cepstral coefficients leads to greater complexity in the [14] system.

➢ *MEL Frequency Cepstral Coefficients Calculation*

Three (03) methods of obtaining MFCCs depending on the structure of the signals are presented: Time Domain Channel Fusion (TDCF); cepstral Domain Channel Fusion (CDCF); One Channel

- *Time Domaine Channel Fusion'(TDCF)*

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:*

$$s_{resultant}(n) = \left[C_{L_1}; C_{R_1}; C_{L_2}; C_{R_2}; \dots; C_{L_{m-1}}; C_{R_{m-1}}; C_{L_m}; C_{R_m}\right] \quad (3)$$

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

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV1194

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 $\boldsymbol{s_{resutant_i}(n)}$).

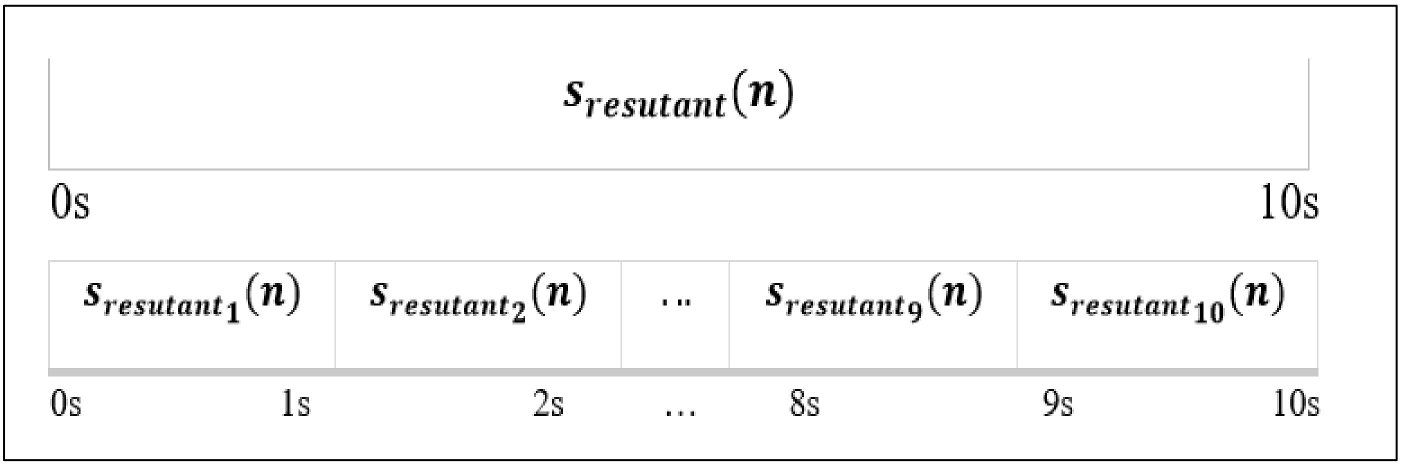

Fig 1 Segmentation of the Signal Resulting from Merging the Left and Right Channels

We compute the MFCCs for each of the 10 "subframes" $\boldsymbol{s_{resutant_i}(n)}$, 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 $1 \times 10N$.

- *Cepstral Domaine Channel Fusion(CDCF)*

In this approach, the calculation of cepstral coefficients is performed on the left channel and the right channel taken separately ((4) and (5).

$$MFCC_{channel\text{-}left} = [MFCC_{L1}; MFCC_{L2}; \dots; MFCC_{L64}; MFCC_{L65}] \quad (4)$$

$$MFCC_{right\text{-}channel} = [MFCC_{R1}; MFCC_{R2}; \dots; MFCC_{R64}; MFCC_{R65}] \quad (5)$$

And then the results are combined according to the approach defined in (1), this time in the cepstral domain. we obtain the values of the MFCC as defined in (6)

$$MFCC = [MFCC_{L1}; MFCC_{R1}; MFCC_{L2}; MFCC_{R2} \dots; MFCC_{L64} MFCC_{R64}; MFCC_{L65}; MFCC_{R65}] \quad (6)$$

- *One Channel*

The calculation of cepstral coefficients is carried out by considering a single channel: either the left or the right channel.

The steps of this calculation are given in (7) and (8)

$$\boldsymbol{recording_{channel-left}} = [\boldsymbol{L_1}; \boldsymbol{L_2}; \dots; \boldsymbol{L_{m-1}}; \boldsymbol{L_m}] \quad \textbf{(7)}$$

$$MFCC_{channel\text{-}left} = [MFCC_{L1}; MFCC_{L2}; \dots; MFCC_{L64}; MFCC_{L65}] \quad (8)$$

C. The Mel Frequency Cepstral Coefficients

Mel's frequency cepstral coefficients are calculated over the following frequency band: [170 *Hz*-21000 *Hz*]. The choice of 170*Hz* as the lower frequency is justified by the observation of a spike in the vicinity of this frequency in the amplitude spectrum of the acoustic signal from our chainsaw. And the upper limit of 22 *kHz* is taken to be equal to half the sampling rate of our DR-5 Dictaphone. The different bandwidths used range from 3 *kHz* to 21 *kHz*, increasing by 2 *kHz* for each new calculation phase. The number of Mel filters chosen is 35 [2] and the duration of the "sub-frames" $s_i(n)$ is fixed at 1*s*.

D. k- Neighborhoods Algorithm (KNN)

The objective of this study not being to assess the effectiveness of the classifiers, we made the choice of the KNN whose results were demonstrated to be satisfactory in [1].

KNN is a fundamental training algorithm that is often used for categorization, pattern recognition, and prediction [10]. The Distance measure between datasets is used by KNN to discover data neighbors. This approach is then used to address classification and regression problems. Furthermore, the k value (where k is an application variable) will detect all initial feature cases that are suitable for the new case and will surround all occurrences in order to obtain the incident for a similar place [11]. As a result, the Value of *k* is important and should be carefully planned, as a low *k* value could result in over-fitting. The value of *k* will be 3 [2].

III. RESULTS

A. Nature Et Largeur De La Bande De Frequence Acoustique

First, we consider only the TDCF restructuring model. Based on this approach, we study the parameters related to the nature of the acoustic frequency band and its width. Once these two parameters have been highlighted, they will serve as a basis for comparing the different signal restructuring approaches

- ➢ *Analysis of the Impact of the Width of the Acoustic Band on the Classification Rate*

The classification results for the chainsaw samples from the test phase are shown in Fig. 3 to 7.We can see that the classification rate depends on the width of the acoustic band: the best classification rates are obtained for acoustic

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV1194

bandwidths greater than 1 *kHz*. However, these classification rates do not necessarily increase with the increase in the width of the acoustic band used to calculate the MFCCs. The best classification rates 97.69% is obtained for a bandwidth of 3 *kHz* over the acoustic band [21170 *Hz*-24170 *Hz*].Arranging the 35 Mel filters over the entire study spectrum is not an approach that is likely to improve the classification rate: for the widest acoustic bandwidth (21 *kHz*) the best classification rates is 92.92%. This value rates is 4.76% lower.

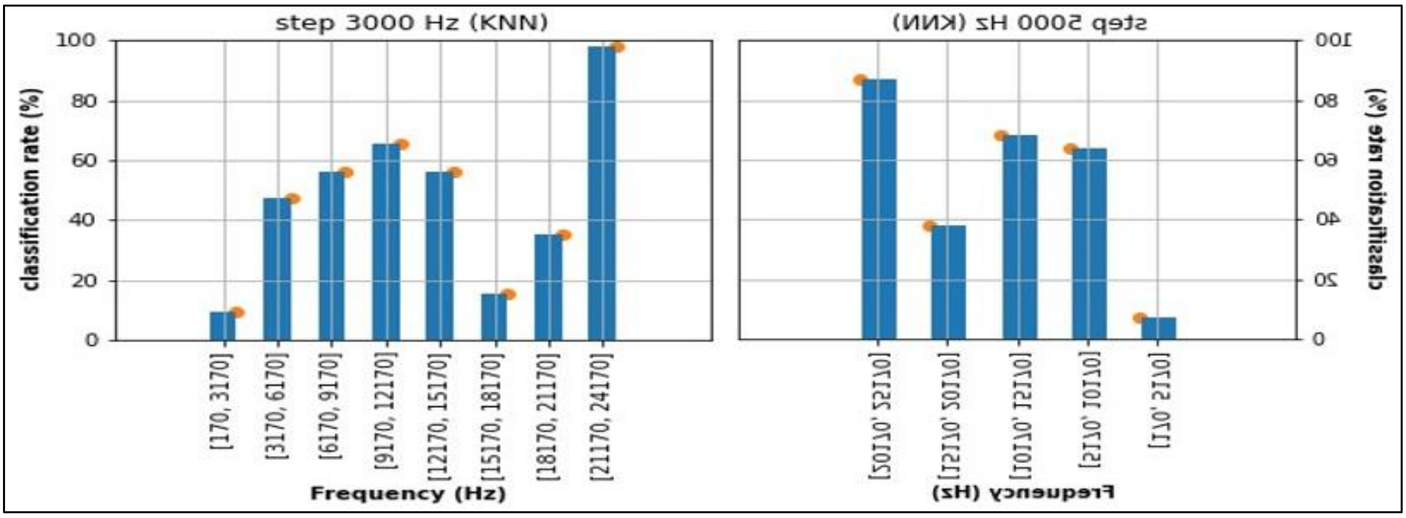

Fig 2 Classification Rates for Step 3000 Hz and Step 5000 Hz

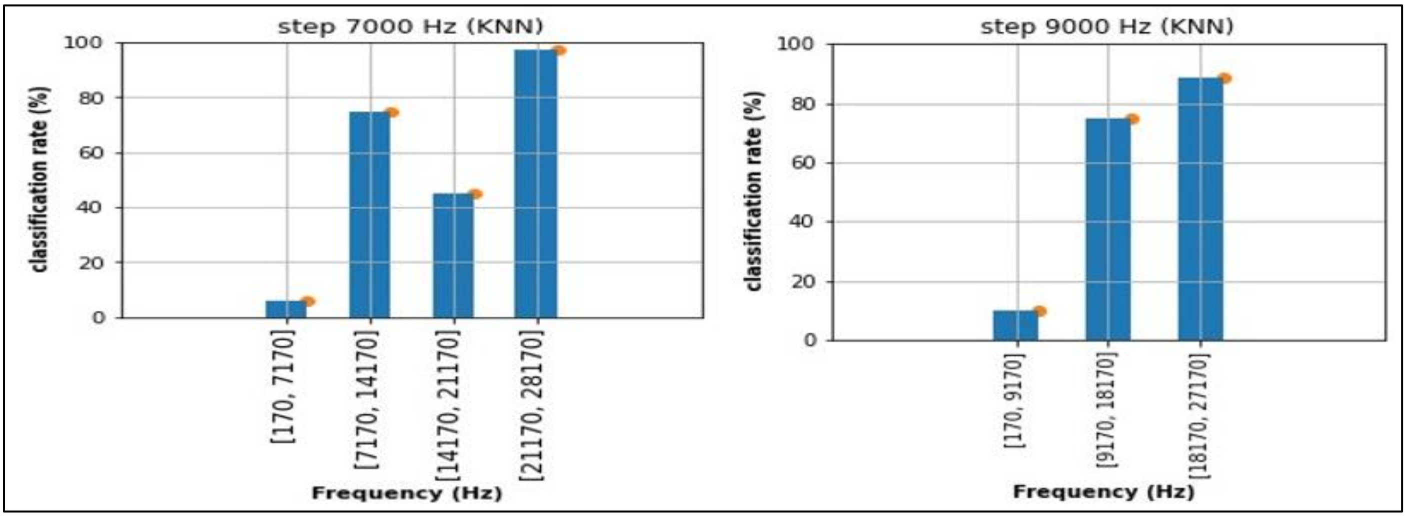

Fig 3 Classification Rates for Step 7000 Hz and Step 9000 Hz

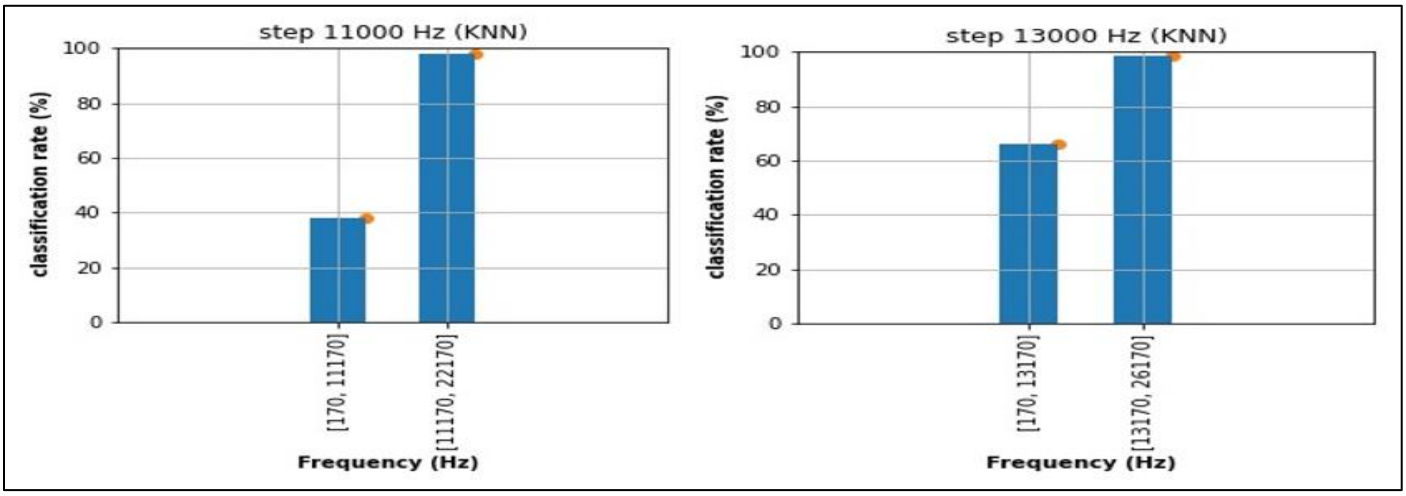

Fig 4 Classification Rates for Step 11000 Hz and Step 13000 Hz

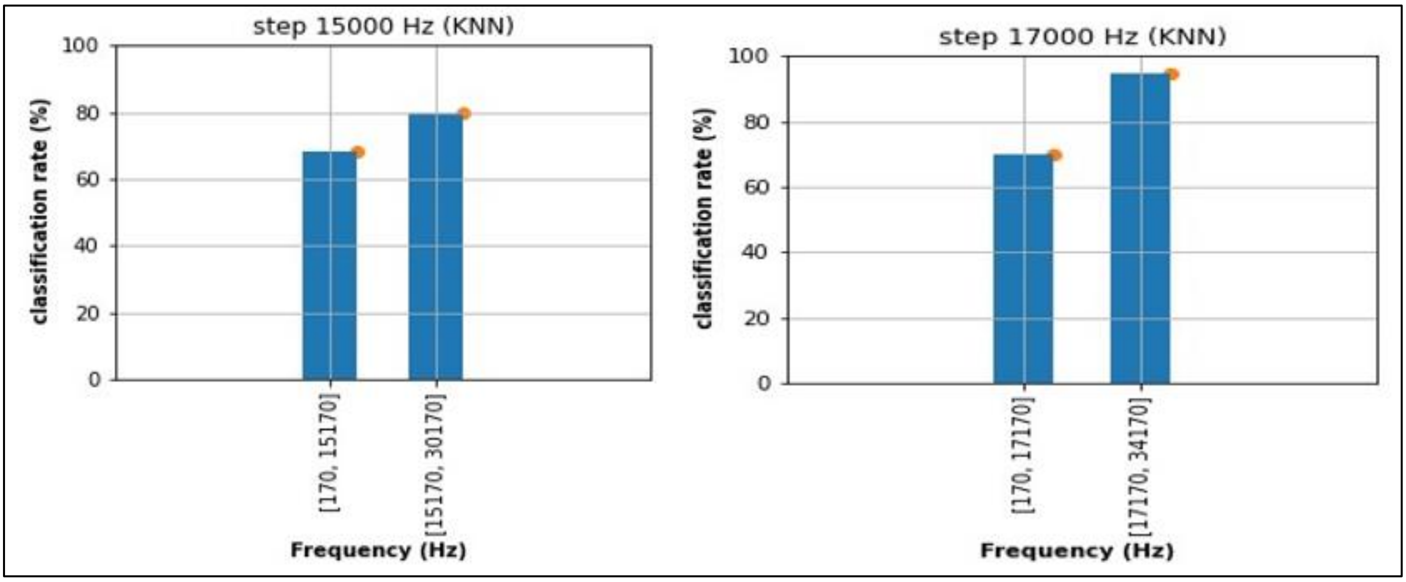

Fig 5 Classification Rates for Step 15000 Hz and Step 17000 Hz

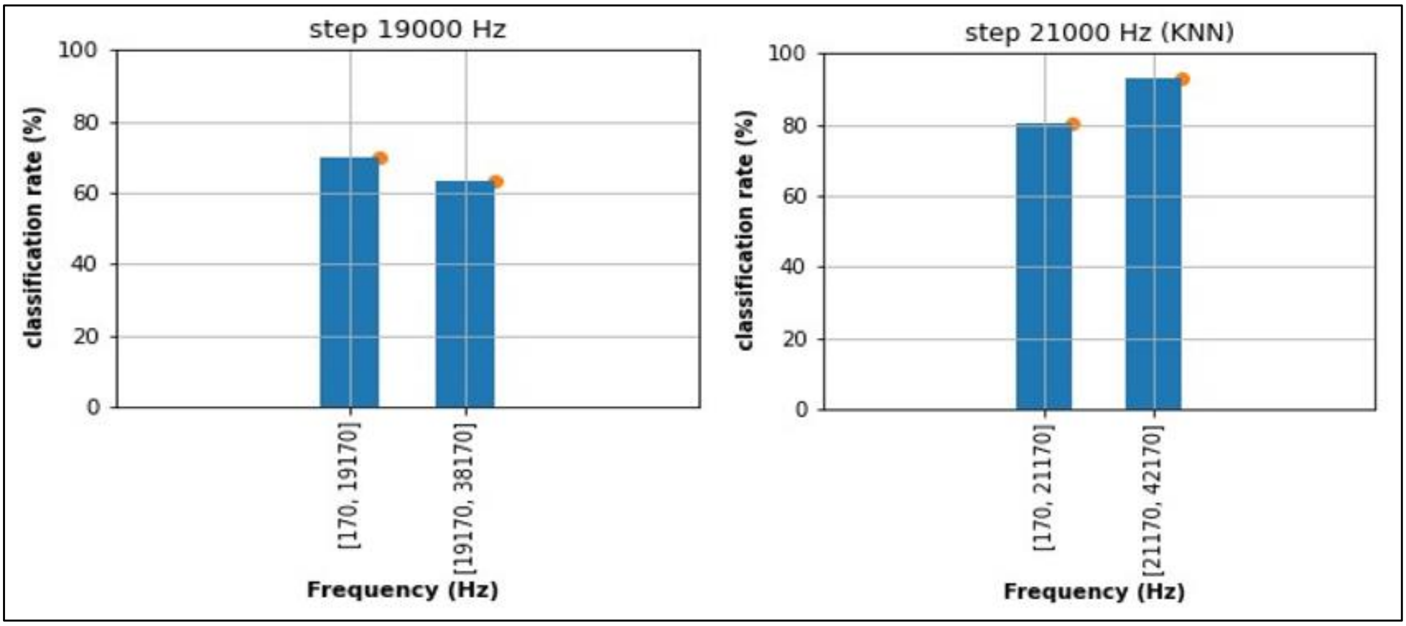

Fig 6 Classification Rates for Step 19000 Hz and Step 21000 Hz

- *Analysis of the Influence of the Nature of the Acoustic Band on the Classification Rate*

The acoustic bands over which the MFCCs are calculated range from low frequencies ([100 *Hz*-300 *Hz*]) to ultrasound (16 *kHz*).The best classification rates for all acoustic bandwidths are obtained either for ultrasound (Fig. 3 to 5 and Fig 7) or for bandwidths including ultrasound (Fig 6).The ultrasound concerned is power ultrasound ([16 *kHz*-1 *MHz*]), and more specifically low-frequency ultrasound ([16 *kHz*-100 *kHz*]); These results for this acoustic band can be explained by the fact that many animal species do not ’communicate’ over long periods of time using ultrasound. Fig. 8, where we see the spectrogram of a bird (the canary) and a frog, illustrates this fact. Studies on animal ”communications” on ultrasonic bands have been carried out by various authors [12] [13] [14]. On the other hand, motorized machines, because of their mechanical make-up, are able to maintain the same cadence, and therefore the same energy, over longer periods of time.

In the case of chainsaw sounds, we found an energy of over-50dB for ultrasound over longer recording times (12*s*), in contrast to the sounds of animal species. For some samples, constant values close to - 40dB over more than 5*s* of recording exist (Fig. 9).This energy distribution contrasts well with that of animal sounds. This property is all the more important as the calculation of MFCCs is based on the recovery of energy by Mel filter banks, which makes classification easier. On the basis of this observation, we understand the better classification results obtained for ultrasound. The classification rates for chainsaws should therefore be error-free, given this energy difference. However, this is not the case. It is important to consider the acoustic samples from motor vehicles and aircraft. These samples have an energy density in the ultrasound with certain similarities to the energy distribution observed for chainsaw sounds (Fig. 10).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV1194

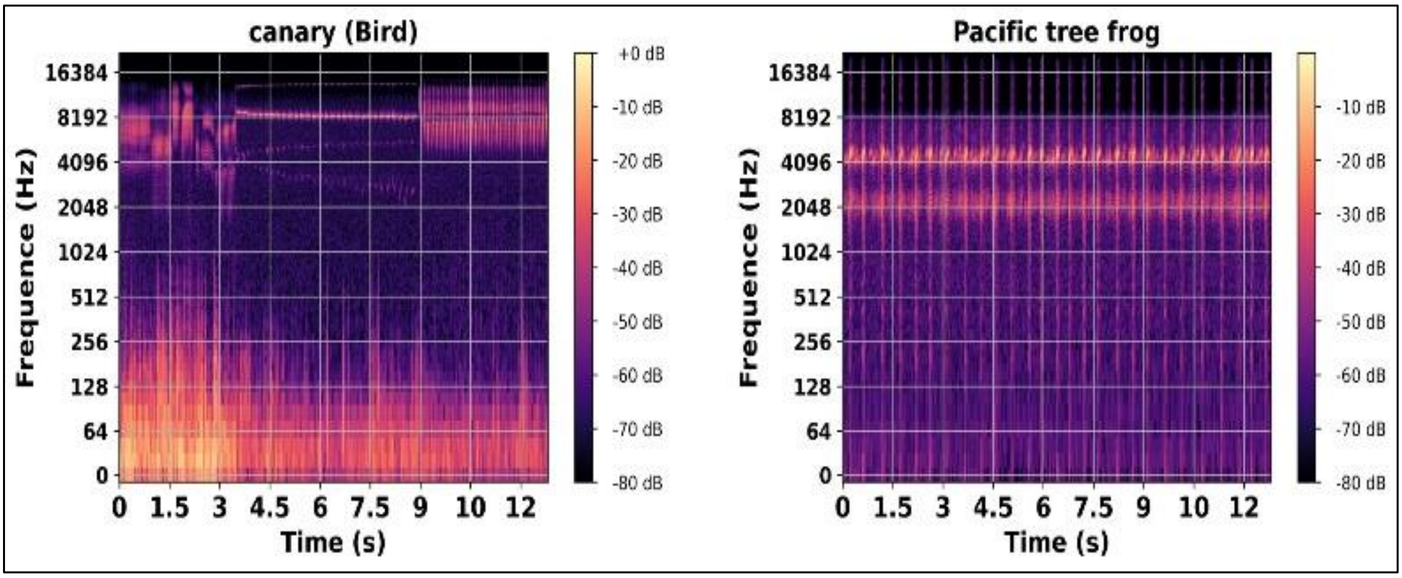

Fig 7 Spectrogram of Acoustic Samples of some Animal Species

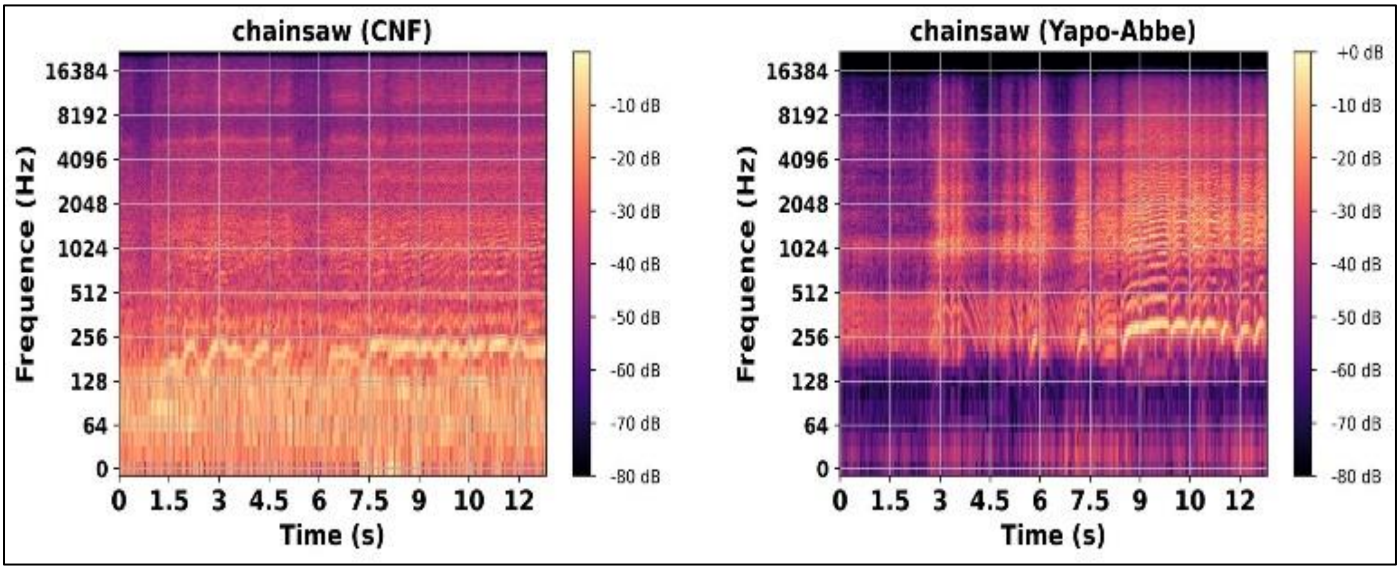

Fig 8 Spectrogram of Sound Samples from Chainsaws

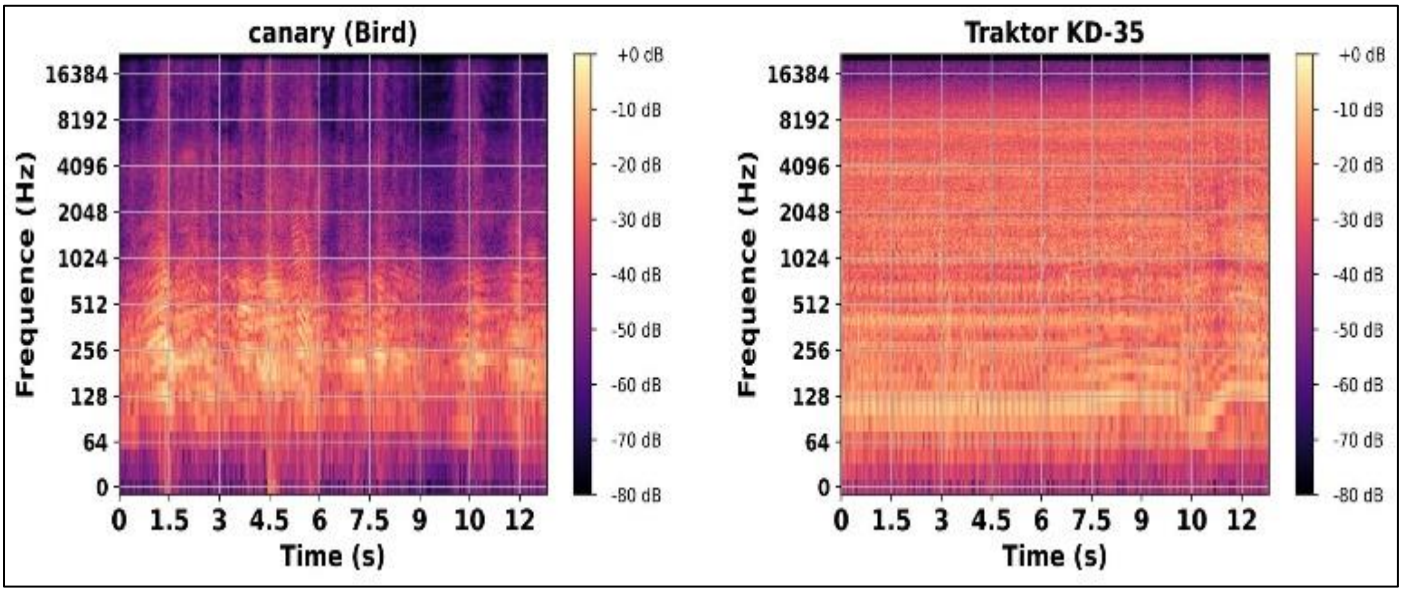

Fig 9 Spectrogram of Samples of a Tractor and a Motorbike

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24NOV1194

This similarity inevitably affects the classifiers' ability to carry out error-free labelling. Taking account of this energy distribution over the recording time is very important.

The calculation of Mel frequency cepstral coefficients is essentially based on the recovery of energy by Mel frequency banks. As a result, the energy recovered over the 5*s* will be extremely different depending on the spectrogram.

➤ *Study of the Calculation of MFCCs on Octave Band Center Frequencies*

The results obtained in the previous sections were based on a random approach to the choice of acoustic bandwidth. A more detailed study should be carried out on predefined bandwidths whose study properties relate to the acoustic perception that humans have of their environment, following the example of the MFCC calculation.

In this sense, we analyze octave band center frequencies. We therefore consider the entire range of octave band center frequencies from 31.5 *Hz* to 16 *kHz*.

Fig. 11 shows the classification rates of chainsaw sounds for each classifier as a function of the chosen octave band center frequency. The best classification rates for the two classifiers are obtained for the central frequency of 16 *kHz* for the acoustic band of [11313 *Hz*-22627 *Hz*]. The values of these identification rates is respectively 98.40% .

In order to fix the best acoustic band, either the [11313 *Hz*-22627 *Hz*] band (center frequency 16 *kHz*) or the [21170 *Hz*-24170 *Hz*] band (bandwidth 3 *kHz*), we study the classification rates of the forest class warning samples from the test phase.

Table 2 gives the best classification rates for the chainsaw class and forest class samples in the test phase. And table 3 highlights the overall classification rate of the test phase samples. This overall classification rate relates to the percentage of samples from the test phase correctly classified out of the 1172 sounds.

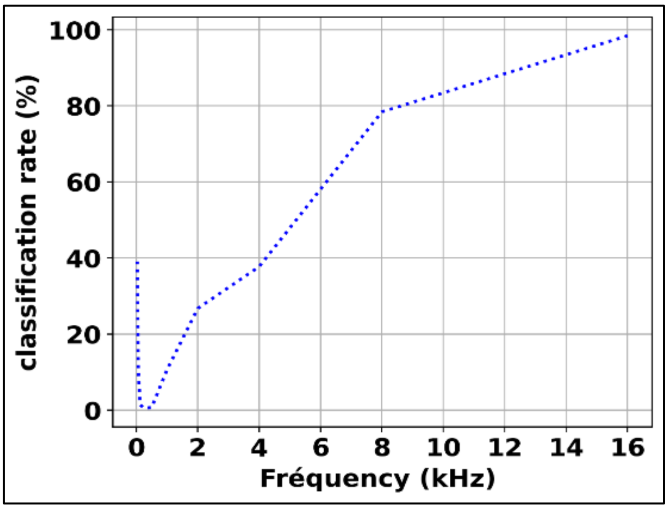

Fig 10 Classification Rate for Chainsaws in the Test

We retain as the best ultrasonic tape and best classifier, the ultrasonic tape and the classifier for which the classification rates are ≥ 98% whether we are talking about the overall classification rate. Once the above criteria have been met, this approach should have the highest classification rate for chainsaw samples.

The best approach is obtained over the ultrasonic band [11313 *Hz* - 22627 *Hz*] deduced from the use of the central octave band frequency of 16 *kHz*.

Table 2 Comparison of Classification Rates Obtained on Acoustic Bands

[11313 - 22627] *Hz*		**[21170 - 24170] *Hz***	
chainsaw class	**Forest class**	**chainsaw class**	**Forest class**
98.40%	98.68%	97.7%	88.13%

Table 3 Comparison of Acoustic Bands: Overall Classification Rate Values

[11313 - 22627] *Hz*	**[21170 - 24170] *Hz***
overall rate	overall rate
98.54%	92.74%

B. Change in Classification Rate due to Data Structure

We consider at preset the impact of the data restructuring mode on the classification rate. We keep the two central octave band frequencies for which the classification rates were the most important namely 8000Hz and 16000 Hz.

Fig. 12 and 13 show the evolution of the chainsaw classification rates according to the signal structuring method for obtaining Mel's frequency cepstral coefficients. For both study frequencies, the TDCF method offers better classification rates compared to the CDCF method and the one channel method. Of the three signal restoration methods, the one channel approach has the lowest classification rates.

➤ *Compared to the TDFC Method,*

- For the 16000Hz frequency: We have a decrease of 14.87% for the CDCF and 16.46% for the one channel.
- For the 8000 Hz frequency: We have a decrease of 43.89% for the CDCF and 50.09% for the one channel.

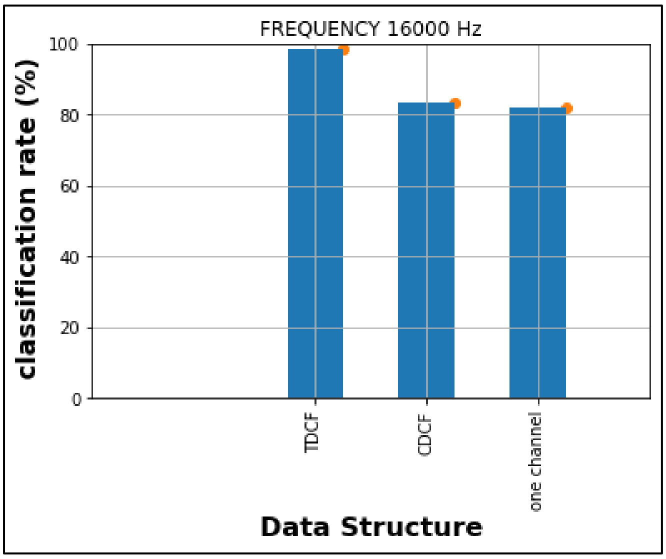

Fig 11 Classification Rate for 16000Hz

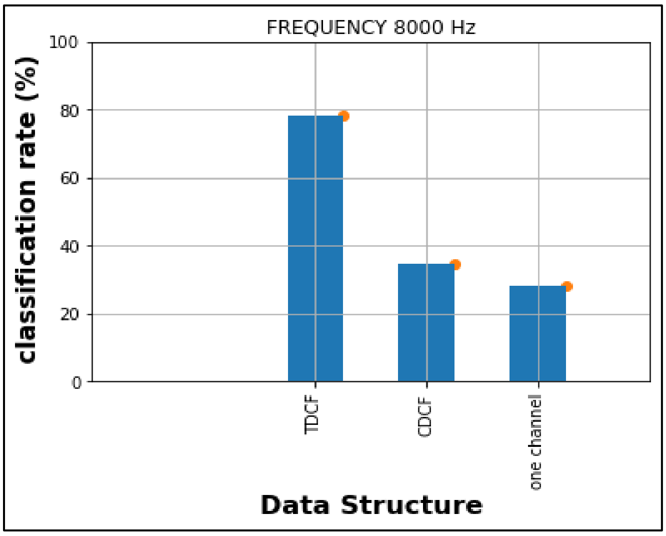

Fig 12 Classification Rate for 8000 Hz

IV. DISCUSSION

The nature and width of the acoustic frequency bands in the calculation of the Mel frequency cepstral coefficients is a determining factor for the optimization of the results obtained. The values presented in the articles [15] [16] [17] although indicating good classification rates in the use of MFCC, leave no information on these two notions. Also, the internal structure of the signals must be indicated, but in the articles [18][19], just as in the above-mentioned articles, this information does not appear.

Some results obtained in these articles that could be judged insufficient, could experience a revaluation with the optimal combination between nature and width of the acoustic frequency bandwidth and internal structure of the signal.

Therefore, the judgement of the effectiveness or not of an algorithm and/or the quality of the descriptors depends on a large number of parameters with regard to our results obtained and that of the articles [18] [1]

V. CONCLUSION

The study highlighted the interest of choosing low-frequency ultrasound as the band for obtaining MFCCs. This choice, combined with a restructuring of the acoustic signal, made it possible to better exploit the description properties of MFCCs for the classification of chainsaw noise samples. This judicious choice allowed us to obtain a classification rate of 98.40% for chainsaw noise samples.

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]. K. C. Shahira, S. Tripathy, et A. Lijiya, « Obstacle Detection, Depth Estimation And Warning System For Visually Impaired People », in TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India: IEEE, oct. 2019, p. 863-868. doi: 10.1109/TENCON.2019.8929334.
[4]. A. M. Joseph, A. Kian, et R. Begg, « State-of-the-Art Review on Wearable Obstacle Detection Systems Developed for Assistive Technologies and Footwear », Sensors, vol. 23, no 5, p. 2802, mars 2023, doi: 10.3390/s23052802.
[5]. J. Ye et N. Toyama, « Automatic defect detection for ultrasonic wave propagation imaging method using spatio-temporal convolution neural networks », Structural Health Monitoring, vol. 21, no 6, p. 2750-2767, nov. 2022, doi: 10.1177/14759217211073503.
[6]. M. Marchevsky, S. Prestemon, O. Lobkis, R. Roth, D. C. Van Der Laan, et J. D. Weiss, « Ultrasonic Waveguides for Quench Detection in HTS Magnets », IEEE Trans. Appl. Supercond., vol. 32, no 6, p. 1-5, sept. 2022, doi: 10.1109/TASC.2022.3164035.
[7]. J.-X. Shen et al., « Ultrasonic frogs show hyperacute phonotaxis to female courtship calls », Nature, vol. 453, no 7197, p. 914-916, juin 2008, doi: 10.1038/nature06719.
[8]. H. C. Gerhardt, M. A. Bee, et J. Christensen-Dalsgaard, « Neuroethology of sound localization in anurans », J Comp Physiol A, vol. 209, no 1, p. 115-129, janv. 2023, doi: 10.1007/s00359-022-01576-9.
[9]. N. Lee, A. Vélez, et M. Bee, « Behind the mask(ing): how frogs cope with noise », J Comp Physiol A, vol. 209, no 1, p. 47-66, janv. 2023, doi: 10.1007/s00359-022-01586-7.

[10]. N. A. Zaidan et M. S. Salam, « MFCC Global Features Selection in Improving Speech Emotion Recognition Rate », in Advances in Machine Learning and Signal Processing, vol. 387, P. J. Soh, W. L. Woo, H. A. Sulaiman, M. A. Othman, et M. S. Saat, Éd., in Lecture Notes in Electrical Engineering, vol. 387. , Cham: Springer International Publishing, 2016, p. 141-153. doi: 10.1007/978-3-319-32213-1_13.
[11]. M. A. Kassem, K. M. Hosny, R. Damaševičius, et M. M. Eltoukhy, « Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review », Diagnostics, vol. 11, no 8, p. 1390, juill. 2021, doi: 10.3390/diagnostics11081390.
[12]. X. Huang, A. Acero, F. Alleva, M. Hwang, L. Jiang, et M. Mahajan, « From Sphinx-II to Whisper — Making Speech Recognition Usable », in Automatic Speech and Speaker Recognition, vol. 355, C.-H. Lee, F. K. Soong, et K. K. Paliwal, Éd., in The Kluwer International Series in Engineering and Computer Science, vol. 355. , Boston, MA: Springer US, 1996, p. 481-508. doi: 10.1007/978-1-4613-1367-0_20.
[13]. R. M. Foratto, D. Llusia, L. F. Toledo, et L. R. Forti, « Treefrogs adjust their acoustic signals in response to harmonics structure of intruder calls », Behavioral Ecology, vol. 32, no 3, p. 416-427, juin 2021, doi: 10.1093/beheco/araa135.
[14]. L. R. Forti, M. R. De Melo Sampaio, C. R. Pires, J. K. Szabo, et L. F. Toledo, « Torrent frogs emit acoustic signals of a narrower spectral range in habitats with longer-lasting biotic background noise », Behavioural Processes, vol. 200, p. 104700, août 2022, doi: 10.1016/j.beproc.2022.104700.
[15]. 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.
[16]. S. Gupta et A. Cosgun, « Audio-Visual Traffic Light State Detection for Urban Robots », 2024, arXiv. doi: 10.48550/ARXIV.2404.19281.
[17]. 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.
[18]. 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.
[19]. 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.