INTENSIF, Vol.9 No.1 February 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

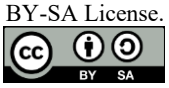
Environmental Acoustic Features Robustness Analysis: A Multi-Aspecs Study

Received:
25 September 2024
Accepted:
26 November 2024
Published:
23 February 2025

[1*]Andi Bahtiar Semma, [2]Kusrini, [3]Arif Setyanto, [4]Bruno da Silva, [5]An Braeken
[1-3]Informatics, Universitas AMIKOM Yogyakarta
[1,4,5]Industrial Engineering, Vrije Universiteit Brussels
[1]Informatics, Universitas Islam Negeri Salatiga
E-mail: [1]andisemma@uinsalatiga.ac.id, [2]kusrini@amikom.ac.id, [3]arief_s@amikom.ac.id, [4]bruno.da.silva@vub.be, [5]an.braeken@vub.be
*Corresponding Author

Abstract—**Background:** Acoustic signals are complex, with temporal, spectral, and amplitude variations. Their non-stationarity complicates analysis, as traditional methods often fail to capture their richness. Environmental factors like reflections, refractions, and noise further distort signals. While advanced techniques such as adaptive filtering and deep learning exist, comprehensive acoustic feature analysis remains limited. **Objective:** This study investigates which acoustic features maintain the highest robustness across diverse environments while preserving discriminative power. **Methods:** Audio samples were recorded in controlled environments (jungles, cafés, factories, streets) with varying noise levels. Standardized equipment captured 22050 Hz, 16-bit audio at multiple positions and distances. After amplitude standardization, various acoustic features were extracted and analyzed. **Results:** MFCCs demonstrated exceptional reliability, with correlation coefficients of 0.98819 and 0.98889 for closely positioned devices and a robustness score of 0.99. Across different acoustic scenes and sample lengths (1, 3, 5s), MFCCs maintained high correlation (≈0.978) and robustness (0.98), confirming their versatility. **Conclusion:** MFCCs proved highly effective for acoustic fingerprinting across settings. Despite limitations in tested environments (≤5m distance, ≤5s samples), their consistent performance validates the methodology. Future research should explore combining MFCCs with spectral features and expanding studies to broader environments and device types.

Keywords—Acoustic Fingerprinting; Signal Processing; MFCC; Environmental Sound



Corresponding Author:

Andi Bahtiar Semma,
Informatics,
Universitas AMIKOM Yogyakarta,
Email: andisemma@uinsalatiga.ac.id
Orchid ID: https://orcid.org/0000-0002-6487-1791

INTENSIF, Vol.9 No.1 February 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

I. INTRODUCTION

The complexity of acoustic signals encompasses multiple interrelated dimensions that challenge analysis and interpretation in the field of acoustics. These signals exhibit intricate frequency content, ranging from infrasonic to ultrasonic components. Amplitude modulations and phase relationships further complicate their characteristics. Spatial properties and stochastic elements add additional layers of complexity, necessitating advanced analytical techniques [1], [2]. Acoustic fingerprinting is a method that creates a compact digital summary, or "fingerprint," of an audio sample that can be used for various applications [3] [4]. These extracted features are then condensed into a compact digital signature, which is much smaller than the original audio file but still contains enough information to uniquely identify it [5]. Effective acoustic fingerprinting techniques are designed to be robust against various noise conditions [6]. In the realm of indoor positioning systems, researchers have developed hybrid approaches that combine acoustic ranging with Wi-Fi fingerprinting to achieve meter-level accuracy in non-line-of-sight environments [3]. This technology has also been applied to device authentication, with studies exploring the use of MEMS sensor readings in response to acoustic signals to identify devices [4]. In the context of drone security, acoustic noise fingerprinting has been proposed as an additional authentication factor for delivery drones, offering protection against impersonation attacks [5]. Furthermore, acoustic fingerprinting has been implemented in hardware architectures for audio ownership verification and content management [7]. The power consumption implications of acoustic sensing applications on smartphones have also been studied, providing insights into the feasibility of long-term acoustic monitoring tasks [8]. These diverse applications demonstrate the versatility and potential of acoustic fingerprinting in addressing challenges across multiple domains.

Real-world acoustic environments pose several challenges that impact the performance of acoustic technologies. Environmental factors play a significant role in distorting acoustic signals, often complicating the transmission and reception of sound in various settings [9], [10]. One primary environmental factor affecting acoustic signals is atmospheric conditions. Temperature inversions, for instance, can create acoustic shadows or enhance sound propagation over long distances. Wind also significantly influences sound transmission, potentially increasing or decreasing the effective range of acoustic signals depending on its direction relative to the sound source [11]. Physical obstacles in the environment present another major source of distortion. Objects such as buildings, trees, and terrain features can cause reflection, diffraction, and scattering of sound waves. These phenomena can lead to multipath propagation, where the same

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

signal arrives at the receiver via multiple paths with different time delays and amplitudes. This can result in interference patterns, echoes, and reverberation [12].

Background noise is a pervasive environmental factor that can interfere with acoustic signals. Urban environments, in particular, are characterized by a complex traffic, construction, and human activity soundscape, which can obscure signals of interest. Natural environments also contribute their own acoustic backgrounds, such as wind noise, water sounds, or animal vocalizations, which can interfere with signal detection and analysis [13], [14]. In the deployment of acoustic-based COVID-19 screening systems, the performance of classifiers can degrade due to variations in recording devices and noise contamination, as well as differences in the symptom status of individuals being tested. This variability can lead to inconsistent performance across different datasets [15]. In acoustic emotion recognition systems, the challenge lies in developing feature representations that can abstract away extraneous low-level variations while capturing relevant speaker characteristics [16]. Water bodies present unique challenges for acoustic signal propagation. In underwater environments, factors such as water temperature, salinity, and depth affect sound speed and create complex propagation paths [17]. Techniques such as adaptive filtering [18] and beamforming [19], [20] noise reduction are continually being developed to improve the robustness of acoustic systems in complex, real-world environments.

Acoustic signals are inherently complex, comprising many features that interact in intricate ways. These features include temporal characteristics (duration and rhythm), spectral components (frequency content and distribution), and amplitude variations. The complexity is further compounded by acoustic signals being often non-stationary, meaning their statistical properties change over time. This variability poses significant challenges for signal analysts, as traditional methods designed for stationary signals may fail to capture the full richness of acoustic data [21], [22], [23], [24]. One of the primary reasons for the complexity of acoustic signal features is the influence of environmental factors. Acoustic signals propagate through various media and are subject to reflections, refractions, and absorptions, which can significantly alter their characteristics. Moreover, background noise and interfering sources can mask or distort the signal of interest, making feature extraction and analysis even more challenging [25], [26], [27]. These environmental effects necessitate sophisticated signal processing techniques to isolate and accurately characterize the desired acoustic features. The novelty of this research is depicted in Figure 1.

However, existing research in acoustic signal analysis has primarily focused on applying pre-selected features to specific applications without systematically evaluating the robustness and effectiveness of different acoustic features across varying conditions. While previous studies [28], [29] have utilized commonly accepted acoustic features, they did not conduct comprehensive

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

comparative analyses to determine which features maintain their discriminative power under diverse environmental challenges. The difference between this research and previous research is its systematic approach to evaluating and comparing acoustic features' performance across different environmental conditions. Prior works have focused on applying specific features to solve particular problems, such as speech emotion recognition [30] or wildlife-vehicle collision prediction [31]. On the other hand, this study aims to provide a fundamental understanding of which acoustic features demonstrate superior robustness and discriminative power regardless of environmental variations. From that background, we formulate a research question: Which acoustic features demonstrate the highest robustness across diverse environmental conditions while maintaining discriminative power?

Ultimately, this research aims to comprehensively evaluate acoustic features to determine which ones maintain their effectiveness across diverse environmental conditions while preserving their discriminative capabilities. This fundamental analysis will contribute to developing more reliable and adaptable acoustic-based systems, particularly in challenging real-world environments where traditional approaches may fall short, like speech recognitions [32], [33].

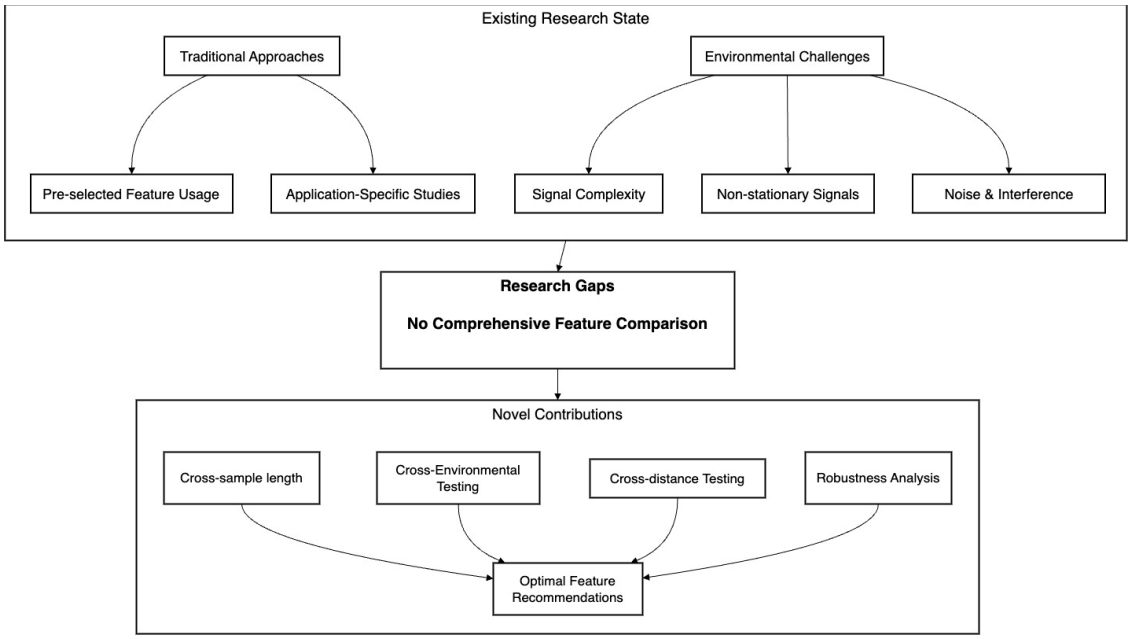


Fig 1. Novelty flowchart

II. RESEARCH METHOD

This study uses quantitative methodology with an experiment approach to analyse the most robust and invariant acoustic feature/s from diverse environmental scenes. This study is divided into several steps:

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

A. Data Acquisition

A diverse audio dataset was compiled by capturing audio samples across various controlled environmental conditions. Controlled recording locations encompassed a range of noise levels and ambient sounds, including jungle [34], café [35], machine [36], and street [37] settings. This approach aimed to establish a robust foundation for subsequent analysis. Recordings were captured from multiple device positions and distances within each controlled environment to characterize acoustic environments comprehensively.

Standardized recording equipment and procedures were implemented throughout the data collection to ensure consistency and comparability. These recordings will use 22050 Hz at 16-bit. Furthermore, this study will employ single and multiple sound sources. The sample length dividen into 1, 3, and 5 seconds with 50 chunks each. Figure 2 illustrates the recording setups. The microphone symbol is the recording device, and the speaker symbol is the sound source.

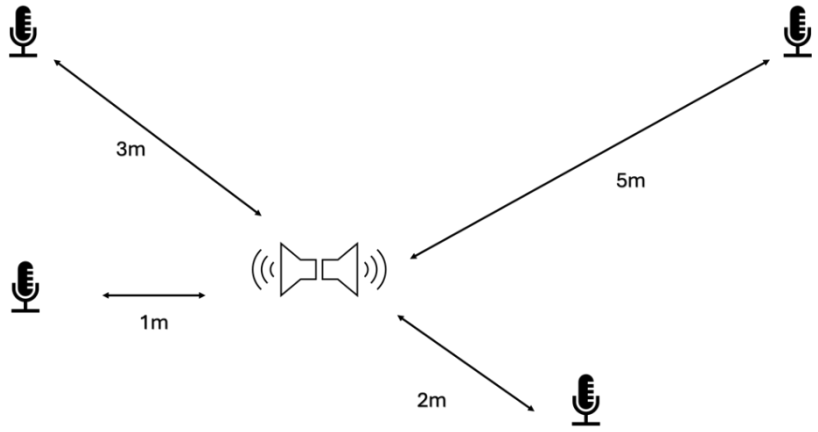


Fig 2. Single Source Setup

B. Pre-processing and Feature extraction

Before feature extraction, amplitude standardization was applied to the audio data to normalize amplitude. Furthermore, extracting a comprehensive set of features from the audio data will be conducted, including Mel-frequency cepstral coefficients (MFCCs), chroma_cens, chroma_cqt, chroma_stft, mel_spectrogram, onset_strength, spectral_bandwidth, Spectral Centroid, Spectral Contrast, Spectral Flatness, Spectral Rolloff, Tempo, Tonnetz and Zero Crossing Rate.

The spectral features form the backbone of audio fingerprinting. Mel-frequency cepstral coefficients (MFCCs) capture the timbral characteristics of sound, making them crucial for identifying unique audio signatures [38], [39]. Spectral Centroid, Bandwidth, Contrast, Flatness, and Rolloff collectively describe the distribution and quality of frequencies, helping distinguish between different audio samples with high precision [40]. The chroma variants (CENS, CQT,

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

STFT) are vital as they represent the harmonic and melodic content of audio. These features map the entire spectrum into 12 pitch classes, making them invariant to changes in octave while preserving the musical character of the audio, essential for robust fingerprinting [41]. Onset strength [42] and Tempo [43] features capture the rhythmic elements and temporal evolution of the audio signal. Zero Crossing Rate [44], [45] provides information about the signal's frequency content and noisiness. These temporal characteristics are crucial for distinguishing between similar-sounding audio segments. Mel spectrograms [46] provide a comprehensive frequency representation that aligns with human auditory perception. Tonnetz [46], [47] features capture harmonic relationships in a unique geometric space. This multi-dimensional representation makes fingerprints more unique and reliable for audio matching.

C. Robustness Analysis

A key aspect of robustness analysis is calculating Pearson's correlation. Pearson's correlation coefficient [48], [49] measures the linear relationship between two variables, ranging from -1 to 1. After calculating these metrics, results will be ranked to determine the most robust feature.

III. RESULT AND DISCUSSION

A. Pearson's correlation across scenes

MFCC consistently shows the highest correlation across all scenes, with values ranging from 0.97198 (machine) to 0.98344 (street). This suggests that MFCC is the most reliable feature in acoustic scene variations, maintaining a strong correlation regardless of the environment. The zero-crossing rate demonstrates a high correlation in most scenes, particularly in the jungle environment (0.95391). However, its performance varies across scenes, with lower correlations in street (0.73137) and machine (0.78992) environments.

Spectral features (bandwidth, centroid, rolloff, and flatness) generally show moderate to high correlations, but their performance varies depending on the scene. For example, the spectral centroid has a high correlation in the multi2 scene (0.84386) but performs poorly in the street scene (0.53809). Mel spectrogram exhibits consistently moderate correlations across all scenes, ranging from 0.54985 (multi2) to 0.84032 (jungle). Chroma-based features (chroma_cens, chroma_cqt, chroma_stft) and tonnetz show lower correlations than other features, with values typically below 0.7. However, chroma_stft performs exceptionally well in the jungle scene (0.89612), indicating its potential usefulness in specific environments. Spectral contrast consistently shows the lowest correlation across all scenes, with values ranging from 0.21764 (machine) to 0.28841 (cafe).

INTENSIF, Vol.9 No.1 February 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

The cafe scene shows high correlations for several features, including spectral bandwidth (0.80811) and spectral centroid (0.79405). The jungle scene demonstrates high correlations for multiple features, including zero-crossing rate (0.95391) and chroma_stft (0.89612), suggesting that these features could be valuable for identifying jungle environments. The machine scene shows relatively lower correlations for most features than other scenes, with only MFCC and zero-crossing rate performing well. This indicates that machine environments are more challenging. Figure 3 illustrates Pearson's correlation across scenes.

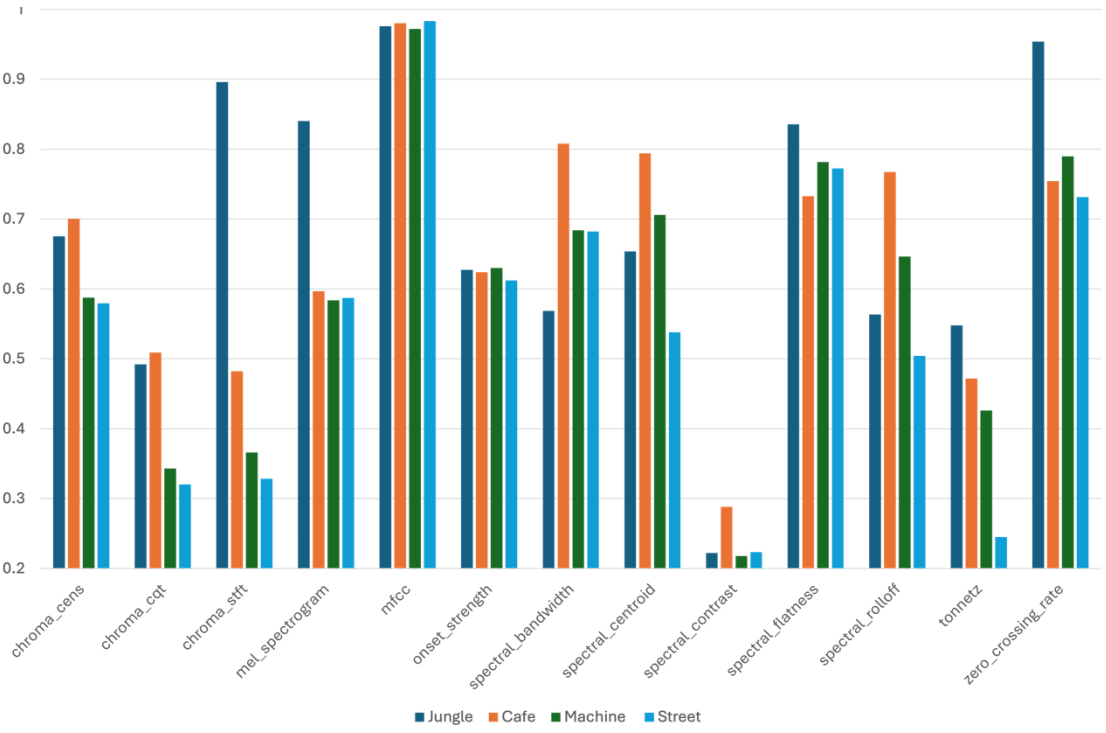


Fig 3. Pearson's correlation across scenes

B. Physical distance impact

MFCC consistently shows the highest correlation and robustness across distances. For pairs 1m and 3m, it achieves an impressive Pearson's correlation of 0.98819 and 0.98889. Even for 5m distance, MFCC maintains a high correlation of 0.95704. There is a general trend of decreasing correlation as we move from 1m to 3m and then to 5m. This indicates that increasing physical distances also increases differences in the acoustic characteristics captured. However, this also indicates that MFCC is the most reliable feature for acoustic fingerprinting in a physical distance context. Detailed pearson's correlation across distances is illustrated in Figure 4.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

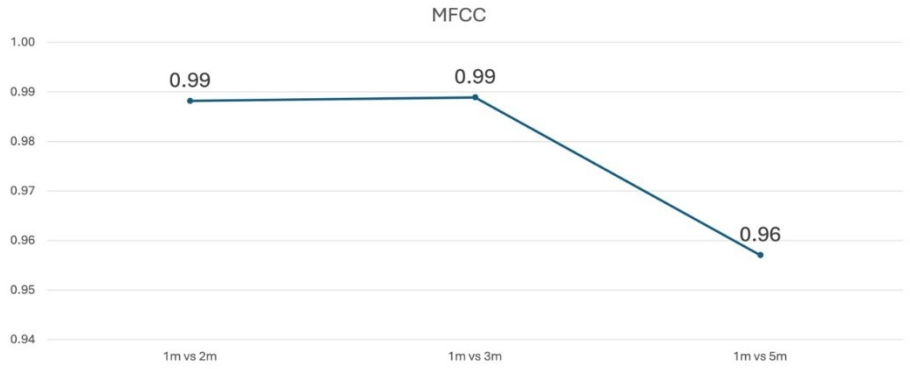


Fig 4. Pearson's correlation across distances

C. Sample lengths impact

Across all sample lengths (1, 3, and 5 seconds), MFCC maintains an exceptionally high Pearson's correlation of around 0.978. This remarkable stability indicates that MFCC is a highly reliable feature for acoustic fingerprinting, regardless of the sample duration. This indicates that while the linear relationship weakens with longer samples, the overall reliability of the feature remains consistent. Lastly, the result reveals that the spread of Pearson's correlation generally increases as sample length increases. This trend indicates that longer audio samples provide more variation correlations across different instances of the same audio, potentially leading to less reliable fingerprinting results. Figure 5 illustrates Pearson's correlation across sample length.

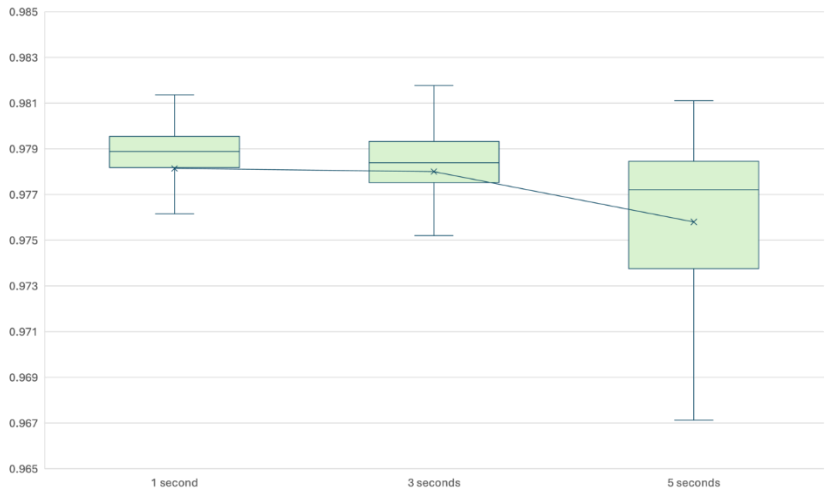


Fig 5. Pearson's correlation across sample length

The results of this research are focused on three key areas of acoustic analysis. In scene analysis, MFCC demonstrated the highest correlations, ranging from 0.97198 in machine environments to 0.98344 in street settings. The zero-crossing rate showed varying performance, with its strongest correlation in jungle environments (0.95391) but weaker correlations in street

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

(0.73137) and machine (0.78992) settings. In terms of distance analysis, MFCC maintained strong correlations across different distances. At 1m and 3m, it achieved impressive Pearson's correlations of 0.98819 and 0.98889, respectively, while still maintaining a high correlation of 0.95704 at 5m. This indicates a slight decrease in correlation as distance increases but overall robust performance. Finally, the sample length analysis revealed that MFCC maintained a consistently high Pearson's correlation of approximately 0.978 across all sample lengths (1, 3, and 5 seconds). However, the research noted that the correlation spread increased with longer sample lengths, suggesting more variation in correlations for longer audio samples. Overall, MFCC proved to be the most reliable feature across all testing conditions, demonstrating its effectiveness for acoustic fingerprinting.

The results of this research are in line with or supported by recent studies that utilize mfcc in various real-world applications. The biological relevance of MFCC is one of its key strengths, as it is designed to mimic human auditory perception [50]. This makes it particularly well-suited for speech and audio analysis, as evidenced by the remarkable 98% accuracy achieved by [51] in their heart sound classification study using MFCC features combined with deep learning. The bee queen presence detection system [52] demonstrated the ability of MFCC to effectively reduce the dimensionality of input signals while retaining essential information, which is another crucial factor contributing to its robustness. This characteristic is vital for efficient processing and classification tasks, especially in real-time applications. Using just 15 or 31 MFCC coefficients, their study achieved outstanding performance metrics of 0.99 across various measures. MFCC's resilience to background noise also became a significant advantage, demonstrated by [53] their speech recognition study, where they achieved an accuracy of 88.21%. More evidence was revealed by [54] successfully applying MFCC-based features for clinical depression recognition from speech, achieving an accuracy of 76.27% and outperforming state-of-the-art approaches. In ECG signal analysis [55], MFCC also proves its applicability beyond traditional audio processing tasks.

This study makes three key contributions: (1) We present a comprehensive evaluation framework for assessing acoustic feature robustness across diverse real-world environments including jungles, cafés, factories, and streets; (2) We provide empirical evidence establishing MFCCs as highly reliable acoustic features, demonstrating correlation coefficients above 0.988 and robustness scores of 0.99 in closely positioned devices; and (3) We validate MFCC performance across multiple temporal scales, maintaining correlation scores around 0.978 and robustness scores of 0.98 across different sample lengths (1-5 seconds).

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

IV. CONCLUSION

This study presents significant contributions to acoustic fingerprinting, demonstrating the exceptional effectiveness of MFCC. These findings have practical implications for device pairing and acoustic scene analysis, with potential applications in secure device communication and environmental monitoring. MFCC has demonstrated exceptional effectiveness and robustness as a feature for acoustic fingerprinting across various settings. The consistently high Pearson's correlation coefficients and robustness scores indicate MFCC's reliability in representing acoustic characteristics. For closely positioned devices, MFCC exhibited outstanding performance in device pairing, with correlation coefficients of 0.98819, 0.98889 and 0.96 for pairs 1-2, 1-3 and 1-5, respectively. MFCC's versatility is evident in its consistent performance across diverse acoustic scenes, with correlation values ranging from 0.97198 to 0.98344. This feature also maintains stability across different sample lengths (1, 3, and 5 seconds), with correlation scores around 0.978 and robustness scores of 0.98, regardless of duration.

While this study's strength lies in its comprehensive evaluation across various parameters, it may be limited by the specific acoustic environments tested, with a maximum of only a 5m distance and a maximum of 5s sample length. However, the consistent performance across different variables supports the methodology's validity and the reliability of the results. Future research directions could include exploring combined MFCC and spectral feature approaches, extending the study to a wider range of acoustic environments and device types, examining feature performance under various noise conditions, and assessing computational efficiency in resource-constrained devices. These avenues would further validate and expand upon the current findings, potentially leading to more robust and widely applicable acoustic fingerprinting techniques.

Author Contributions: *Andi Bahtiar Semma*: Conceptualization, Writing - Original Draft, Software, Investigation, Data Curation. *Kusrini, Arif Setyanto, An Braeken, Bruno da Silva*: Conceptualization, Methodology, Review, Supervision.

All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia number 107/E5/PG.02.00.PL/2024; 0609.20/LL5-INT/AL.04/2024; 025/KONTRAK-LPPM/AMIKOM/VI/2024.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The research data can be found here:

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i1.23723

ORCID:
Andi Bahtiar Semma: https://orcid.org/0000-0002-6487-1791
Kusrini: https://orcid.org/0000-0001-9573-3909
Arif Setyanto: https://orcid.org/0000-0003-0721-3941
Bruno da Silva: https://orcid.org/0000-0002-4877-9688
An Braeken: https://orcid.org/0000-0002-9965-915X

REFERENCES

[1] T. Heittola, A. Mesaros, and T. Virtanen, "Acoustic scene classification in DCASE 2020 Challenge: generalization across devices and low complexity solutions," Nov. 02, 2020, *arXiv*: arXiv:2005.14623. Accessed: Sep. 25, 2024. [Online]. Available: http://arxiv.org/abs/2005.14623

[2] S. Dröge *et al.*, "Listening to a changing landscape: Acoustic indices reflect bird species richness and plot-scale vegetation structure across different land-use types in north-eastern Madagascar," *Ecol. Indic.*, vol. 120, p. 106929, 2021, Accessed: Sep. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X20308682

[3] Z. Zhang, Y. Yu, L. Chen, and R. Chen, "Hybrid Indoor Positioning System Based on Acoustic Ranging and Wi-Fi Fingerprinting under NLOS Environments," *Remote Sens.*, vol. 15, no. 14, p. 3520, 2023.

[4] S. Ramesh, T. Pathier, and J. Han, "Sounduav: Towards delivery drone authentication via acoustic noise fingerprinting," presented at the Proceedings of the 5th Workshop on Micro Aerial Vehicle Networks, Systems, and Applications, 2019, pp. 27–32.

[5] B. Thoen, S. Wielandt, and L. De Strycker, "Fingerprinting Method for Acoustic Localization Using Low-Profile Microphone Arrays," presented at the 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2018, pp. 1–7.

[6] A. Berdich, B. Groza, R. Mayrhofer, E. Levy, A. Shabtai, and Y. Elovici, "Sweep-to-unlock: Fingerprinting smartphones based on loudspeaker roll-off characteristics," *IEEE Trans. Mob. Comput.*, vol. 22, no. 4, pp. 2417–2434, 2021.

[7] I. Algredo-Badillo, B. Sánchez-Juárez, K. A. Ramírez-Gutiérrez, C. Feregrino-Uribe, F. López-Huerta, and J. J. Estrada-López, "Analysis and hardware architecture on fpga of a robust audio fingerprinting method using ssm," *Technologies*, vol. 10, no. 4, p. 86, 2022.

[8] S. Zhidkov, A. Sychev, A. Zhidkov, and A. Petrov, "On smartphone power consumption in acoustic environment monitoring applications," *Appl. Syst. Innov.*, vol. 1, no. 1, p. 8, 2018.

[9] H. Nam, S.-H. Kim, and Y.-H. Park, "Filteraugment: An acoustic environmental data augmentation method," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 4308–4312. Accessed: Sep. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9747680/

[10] D. Li, S. Cao, S. I. Lee, and J. Xiong, "Experience: practical problems for acoustic sensing," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, Sydney NSW Australia: ACM, Oct. 2022, pp. 381–390. **doi:** 10.1145/3495243.3560527.

[11] G. W. Lyons, C. R. Hart, and R. Raspet, "As the wind blows: Turbulent noise on outdoor microphones," *Acoust. Today*, vol. 17, no. 4, p. 20, 2021, Accessed: Sep. 25, 2024. [Online]. Available: https://acousticstoday.org/wp-content/uploads/2021/11/As-the-Wind-Blows-Turbulent-Noise-on-Outdoor-Microphones-Gregory-W.-Lyons-Carl-R.-Hart-and-Richard-Raspet.pdf

[12] W. Lambert, L. A. Cobus, T. Frappart, M. Fink, and A. Aubry, "Distortion matrix approach for ultrasound imaging of random scattering media," *Proc. Natl. Acad. Sci.*, vol. 117, no. 26, pp. 14645–14656, Jun. 2020, **doi:** 10.1073/pnas.1921533117.

[13] A. Novak and P. Honzík, "Measurement of nonlinear distortion of MEMS microphones," *Appl. Acoust.*, vol. 175, p. 107802, 2021, Accessed: Sep. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X20309075

[14] N. Kamuni, S. Chintala, N. Kunchakuri, J. S. A. Narasimharaju, and V. Kumar, "Advancing Audio Fingerprinting Accuracy with AI and ML: Addressing Background Noise and Distortion Challenges," in *2024 IEEE 18th International Conference on Semantic Computing (ICSC)*, IEEE, 2024, pp. 341–345. Accessed: Sep. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10475656/

[15] D. Grant, I. McLane, V. Rennoll, and J. West, "Considerations and Challenges for Real-World Deployment of an Acoustic-Based COVID-19 Screening System," *Sensors*, vol. 22, no. 23, p. 9530, 2022.

[16] Z. Aldeneh and E. M. Provost, "You're Not You When You're Angry: Robust Emotion Features Emerge by Recognizing Speakers," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1351–1362, 2021.

[17] C. Abrahams, C. Desjonquères, and J. Greenhalgh, "Pond Acoustic Sampling Scheme: A draft protocol for rapid acoustic data collection in small waterbodies," *Ecol. Evol.*, vol. 11, no. 12, pp. 7532–7543, Jun. 2021, **doi:** 10.1002/ece3.7585.

[18] D. Shi, W.-S. Gan, B. Lam, and S. Wen, "Feedforward selective fixed-filter active noise control: Algorithm and implementation," *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 1479–1492, 2020, Accessed: Sep. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9086868/

[19] A. Li, W. Liu, C. Zheng, and X. Li, "Embedding and beamforming: All-neural causal beamformer for multichannel speech enhancement," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 6487–6491. Accessed: Sep. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9746432/

[20] G. Huang, J. Benesty, I. Cohen, and J. Chen, "A simple theory and new method of differential beamforming with uniform linear microphone arrays," *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 1079–1093, 2020, Accessed: Sep. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9037110/

[21] S. Cortesi, C. Vogt, E. Reinschmidt, and M. Magno, "Latency and Power Consumption in 2.4GHz IoT Wireless Mesh Nodes: An Experimental Evaluation of Bluetooth Mesh and Wirepas Mesh," *2023 19th Int. Conf. Wirel. Mob. Comput. Netw. Commun. WiMob*, pp. 200–205, 2023, **doi:** 10.1109/WiMob58348.2023.10187799.

[22] D. Uchida, Y. Yonezawa, and K. Akita, "Measurement-Based Latency Evaluation and the Theoretical Analysis for Massive IoT Applications Using Bluetooth Low Energy," *2023 IEEE 97th Veh. Technol. Conf. VTC2023-Spring*, pp. 1–5, 2023, **doi:** 10.1109/VTC2023-Spring57618.2023.10200332.

[23] C. Gomez, J. Oller, and J. Aspas, "Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology," *Sensors*, vol. 12, pp. 11734–11753, 2012, **doi:** 10.3390/s120911734.

[24] Y. Bai, L. Lu, J. Cheng, J. Liu, Y. Chen, and J. Yu, "Acoustic-based sensing and applications: A survey," *Comput. Netw.*, vol. 181, p. 107447, 2020.

[25] S. S. Sethi *et al.*, "Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set," *Proc. Natl. Acad. Sci.*, vol. 117, no. 29, pp. 17049–17055, Jul. 2020, **doi:** 10.1073/pnas.2004702117.

[26] S. R.-J. Ross, N. R. Friedman, M. Yoshimura, T. Yoshida, I. Donohue, and E. P. Economo, "Utility of acoustic indices for ecological monitoring in complex sonic environments," *Ecol. Indic.*, vol. 121, p. 107114, 2021, Accessed: Aug. 15, 2024.

[Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X20310530

[27] H. Nam, S.-H. Kim, and Y.-H. Park, "Filteraugment: An acoustic environmental data augmentation method," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 4308–4312. Accessed: Aug. 15, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9747680/

[28] Y. Diao, Y. Zhang, G. Zhao, and M. Khamis, "Drone authentication via acoustic fingerprint," in *Proceedings of the 38th Annual Computer Security Applications Conference*, 2022, pp. 658–668.

[29] S. Chamishka *et al.*, "A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling," *Multimed. Tools Appl.*, vol. 81, no. 24, pp. 35173–35194, Oct. 2022, **doi:** 10.1007/s11042-022-13363-4.

[30] X. Xu *et al.*, "Survey on discriminative feature selection for speech emotion recognition," presented at the The 9th International Symposium on Chinese Spoken Language Processing, IEEE, 2014, pp. 345–349.

[31] A. Bénard, T. Lengagne, and C. Bonenfant, "A biologically realistic model to predict wildlife-vehicle collision risks," *bioRxiv*, pp. 2023–02, 2023.

[32] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, "Automatic speech recognition: a survey," *Multimed. Tools Appl.*, vol. 80, no. 6, pp. 9411–9457, Mar. 2021, **doi:** 10.1007/s11042-020-10073-7.

[33] J. L. K. E. Fendji, D. C. M. Tala, B. O. Yenke, and M. Atemkeng, "Automatic Speech Recognition Using Limited Vocabulary: A Survey," *Appl. Artif. Intell.*, vol. 36, no. 1, p. 2095039, Dec. 2022, **doi:** 10.1080/08839514.2022.2095039.

[34] Trundlefly, "Amazon Jungle Morning | Royalty-free Music." Accessed: Sep. 10, 2024. [Online]. Available: https://pixabay.com/sound-effects/amazon-jungle-morning-24939/

[35] Hitrison, "Restaurant Sounds (Sunny Point Cafe)." Accessed: Sep. 10, 2024. [Online]. Available: https://pixabay.com/sound-effects/restaurant-sounds-sunny-point-cafe-25092/

[36] CSNmedia, "industrial sounds | Royalty-free Music." Accessed: Sep. 10, 2024. [Online]. Available: https://pixabay.com/sound-effects/industrial-sounds-25817/

[37] Aatreya_v, "Busy Street Ambience | Royalty-free Music." Accessed: Sep. 10, 2024. [Online]. Available: https://pixabay.com/sound-effects/busy-street-ambience-195884/

[38] D. Prabakaran and S. Sriuppili, "Speech processing: MFCC based feature extraction techniques-an investigation," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 012009. Accessed: Aug. 14, 2024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1717/1/012009/meta

[39] D. Á. Villafuerte-Lucio, "MFCC feature extraction for COPD detection," *J. Technol. Innov.*, 2023, Accessed: Aug. 22, 2024. [Online]. Available: https://www.semanticscholar.org/paper/MFCC-feature-extraction-for-COPD-detection-Villafuerte-Lucio/b07f35f25dee7c44de22fef153b70189655e28e1

[40] I. A. Thukroo, R. Bashir, and K. J. Giri, "A comparison of cepstral and spectral features using recurrent neural network for spoken language identification," *Comput. Artif. Intell.*, vol. 2, no. 1, Feb. 2024, **doi:** 10.59400/cai.v2i1.440.

[41] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Appl. Acoust.*, vol. 158, p. 107020, Jan. 2020, **doi:** 10.1016/j.apacoust.2019.107020.

[42] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx). Maynooth, Ireland (Sept 2013)*, Citeseer, 2013, p. 4. Accessed: Nov. 14, 2024. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b9588a8840f9f195c03847d9e3a95ac63d2bc5f2

[43] P. Grosche, M. Müller, and F. Kurth, "Cyclic tempogram—a mid-level tempo representation for musicsignals," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2010, pp. 5522–5525. Accessed: Nov. 14, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5495219/

[44] M. J. Carey, E. S. Parris, and H. Lloyd-Thomas, "A comparison of features for speech, music discrimination," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, Mar. 1999, pp. 149–152 vol.1. **doi:** 10.1109/ICASSP.1999.758084.

[45] K. El-Maleh, M. Klein, G. Petrucci, and P. Kabal, "Speech/music discrimination for multimedia applications," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, Jun. 2000, pp. 2445–2448 vol.4. **doi:** 10.1109/ICASSP.2000.859336.

[46] M. A. Aslam, M. Umer, M. Kashif, R. Talib, and U. Khalid, "Acoustic Classification using Deep Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, 2018, **doi:** 10.14569/IJACSA.2018.090820.

[47] R. Cohn, "Introduction to Neo-Riemannian Theory: A Survey and a Historical Perspective," *J. Music Theory*, vol. 42, no. 2, pp. 167–180, 1998, **doi:** 10.2307/843871.

[48] D. Weisburd, C. Britt, D. B. Wilson, and A. Wooditch, "Measuring Association for Scaled Data: Pearson's Correlation Coefficient," in *Basic Statistics in Criminology and Criminal Justice*, Cham: Springer International Publishing, 2020, pp. 479–530. **doi:** 10.1007/978-3-030-47967-1_14.

[49] R. Chattamvelli, "Pearson's Correlation," in *Correlation in Engineering and the Applied Sciences*, in Synthesis Lectures on Mathematics & Statistics. , Cham: Springer Nature Switzerland, 2024, pp. 55–76. **doi:** 10.1007/978-3-031-51015-1_2.

[50] D. Villafuerte-Lucio, "MFCC feature extraction for COPD detection," *J. Technol. Innov.*, pp. 1–7, Dec. 2023, **doi:** 10.35429/JTI.2023.27.10.1.7.

[51] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Netw.*, vol. 130, pp. 22–32, 2020, Accessed: Aug. 14, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608020302306

[52] B. S. Soares, J. S. Luz, V. F. de Macêdo, R. R. V. e Silva, F. H. D. de Araújo, and D. M. V. Magalhães, "MFCC-based descriptor for bee queen presence detection," *Expert Syst. Appl.*, vol. 201, p. 117104, 2022, Accessed: Aug. 14, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422005085

[53] A. Mahmood and U. Köse, "Speech recognition based on convolutional neural networks and MFCC algorithm," *Adv. Artif. Intell. Res.*, vol. 1, no. 1, pp. 6–12, 2021, Accessed: Aug. 14, 2024. [Online]. Available: https://dergipark.org.tr/en/pub/aair/issue/59650/768432

[54] E. Rejaibi, A. Komaty, F. Meriaudeau, S. Agrebi, and A. Othmani, "MFCC-based recurrent neural network for automatic clinical depression recognition and assessment from speech," *Biomed. Signal Process. Control*, vol. 71, p. 103107, 2022, Accessed: Aug. 14, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421007047

[55] Y. Arpitha, G. L. Madhumathi, and N. Balaji, "Spectrogram analysis of ECG signal and classification efficiency using MFCC feature extraction technique," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 2, pp. 757–767, Feb. 2022, **doi:** 10.1007/s12652-021-02926-2.