applied sciences

MDPI

Editorial

New Advances in Audio Signal Processing

Giovanni Costantini *, Daniele Casali and Valerio Cesarini

Department of Electronic Engineering, University of Rome Tor Vergata, 00133 Rome, Italy; daniele.casali@uniroma2.it (D.C.); valerio.cesarini@uniroma2.it (V.C.)
* Correspondence: costantini@uniroma2.it

1. Introduction

The growth in computing capabilities has significantly transformed the realm of data analysis and processing, most notably through the widespread adoption of artificial intelligence (AI) and deep learning technologies [1]. Audio signals, rich in content, have benefitted immensely from enhanced analytical and interpretive methodologies. This advancement is evident in the rapid progress in traditionally challenging areas, such as speaker recognition [2] and sound categorization, while also paving the way for novel areas of study in audio analysis powered by machine learning, like emotional computing [3].

Consequently, contemporary approaches to audio signal processing are increasingly centered around AI methods that either analyse audio for tasks like classification and recognition or enhance audio quality through techniques like noise reduction and balancing. More broadly, the emphasis, in both the commercial and academic spheres, on signal-as-data involves automation. This prioritizes the development of swift, automated methods for dissecting, segmenting, annotating, and processing audio information as a dynamic area of interest. Furthermore, while AI relies on data, incorporating specialized knowledge within the domain can significantly refine these solutions, making signal processing a critical component of stages such as data enhancement and preparation.

Moreover, traditional signal processing strategies that delve into new or underexplored areas remain crucial, considering the extensive reliance on acoustic characteristics across various sectors, including AI.

Thus, the new advancements in audio signal processing inherently embed various fields, which can be summarized as follows:

- New techniques for processing and analyzing sound signals, such as acoustic imaging [4] or pitch detection algorithms [5].
- Models and methodologies used for the analysis of audio data to derive new information, especially those directed towards AI and Machine Learning-based sound analysis. With the recent growth of deep learning (DL), advanced AI techniques are nowadays employed for tasks such as the preliminary detection of diseases in human voice, sound event detection, speaker recognition or sound classification [6,7].
- The assessment of acoustic effects and/or relevant signals as indicators of the performance of another system (e.g., sound-based fault detection).

This Editorial provides an overview of the most recent and advanced applications of audio signal technology, whether used for their processing or analysis. As shown in the articles presented, deep learning represents a common topic of interest in audio data classification, whereas the analysis of acoustically relevant features such as pitch or reverberation will only become increasingly crucial as the importance of clean and quick data analysis grows.

2. Overview of Published Articles

Strianese et al. (Contribution 1) developed an experimental audio-based approach to evaluate the effects of noise from gas discharges, comparing different nozzle types under

check for updates

Citation: Costantini, G.; Casali, D.; Cesarini, V. New Advances in Audio Signal Processing. *Appl. Sci.* **2024**, *14*, 2321. https://doi.org/10.3390/app14062321

Received: 5 March 2024
Accepted: 7 March 2024
Published: 10 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

a variety of conditions and linking these effects to design features. The use of nozzles in gas-based fire suppression systems within data centers can produce sound levels that significantly degrade hard disk performance; to mitigate this issue, silent nozzles are utilized. Both standard and silent nozzles are analyzed during the release of inert gases and halocarbon compounds. The effectiveness of silent nozzles in maintaining noise levels below the critical 110 dB threshold is confirmed. The findings suggest a correlation between the Reynolds number and the peak noise levels [8], indicating that noise increases with higher flow rates. Slower hard drives were particularly vulnerable. A spectral analysis revealed that higher frequency noises could impair performance, even when below the set threshold. Additionally, the study found that the type of fire suppressant used did not affect noise levels relative to the volumetric flow rate; however, denser gases produced less noise for the same mass flow rate.

Contribution 2 by Kim et al. underlines another aspect of audio signal analysis by focusing on the deep learning-based detection of voice pathology; in this case, it is COVID-19. The authors propose a novel set of features tailored for COVID detection, comprising MFCC, Δ2-MFCC, Δ-MFCC [9], and spectral contrast, and integrate these features into a CNN architecture fusing a ResNet-50 [10] to a custom DNN. The proposed model + feature pairing, trained using selected cough samples sourced from public databases including Cambridge [11], Coswara [12], and COUGHVID [13], results in a 96% accuracy rate.

Tamulionis et al. (Contribution 3) explore a novel approach for estimating the impact of room acoustics, especially reverberation, on the intelligibility and information content of sound signals. This is achieved by training recurrent neural networks using both echo-free and echo-rich audio samples to build an estimator of the room impulse response. The models are trained on a newly created synthetic dataset carefully adjusted to match the characteristics of actual recorded impulse responses. Given the complexity and practical challenges involved in simulating especially long room impulse responses at 44.1 kHz using traditional time-domain methods, they propose a frequency domain-based strategy enhanced by a logarithmic band re-weighing of the auditory spectrum.

Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU) are compared to evaluate the model's ability to accurately predict the spectral characteristics of reverberated sound across adjacent frequency bands. The paper proves how different frequency bands have relevant impacts on the effect of reverberation, all the while determining the BiLSTM to be the most flexible network for estimating the room impulse response.

Atamer et al. (Contribution 4) also considered the analysis of sound for evaluating the performance of another system, aiming to assess the variations in both the signal and the psychoacoustic properties of canister-type vacuum cleaners. Fifteen vacuum cleaners were selected to evenly cover a broad range of reported sound power levels, and their differences in acoustic and psychoacoustic parameters were analysed to identify common noise characteristics, aiming to finding a prototypical noise for canister applications.

The second objective of this study is to learn how the noise from vacuum cleaners is perceived as annoying by listeners. This is achieved via two experiments, the first of which aimed to identify which factors contribute to the perception of annoyance caused by vacuum cleaner noise. Loudness, sharpness, and the presence of tonal components at both low and high frequencies are the primary factors influencing annoyance ratings. The subsequent experiment is designed to examine the potential interaction between loudness and sharpness in affecting annoyance, using listening tests with specifically chosen values of loudness and sharpness that reflect the ranges identified in the first experiment. Although no significant interaction between loudness and sharpness is detected, each factor individually shows a strong positive correlation with annoyance.

Contribution 5 by Lee et al. proposes a more efficient DL-based method for incomplete sound event detection (SED), specifically designed for situations where the training data only include the types of events, without precise timing. The authors propose an alternative weakly supervised architecture based on the U-Net [14], introducing a variant with con-

strained upsampling (LUU-Net) and global threshold average pooling (GTAP), aiming to diminish both the model's size and computational demands. This approach notably limits the increase in the frequency dimension in the U-Net's decoding phase, thereby shrinking the size of the output maps by almost 40% without causing a decrease in performances, as confirmed by experiments on a composite dataset from the DCASE 2018 Challenges 1 and 2 [15], showing how LUU-Net + GTAP trains 23% faster and scores higher in F1 metrics, achieving 0.644 for audio tagging and 0.531 for weakly supervised SED tasks, compared to the traditional approach. The principal advantage of the proposed LUU-Net lies in its efficiency, further enhanced by GTAP accelerating training and offering audio adaptability through the adjustment of a single hyperparameter.

Nanni et al. in Contribution 6 tackle the task of analyzing dolphin whistles through advanced deep learning technologies for monitoring marine environments through echo-bioacoustics [16]. In their experimentations, they also explore the effects of applying data augmentation on the test dataset rather than the training dataset, augmenting test spectrograms with random shift with black or wrap and symmetric alternating diagonal shift. They employed an ensemble of 10 ResNet50 networks merged with the sum rule on a well-known public benchmark, obtaining notable improvements in performance across various established metrics, including a state-of-the-art accuracy rate of 94.9%. While further studies are essential to verify the effectiveness of these methods in diverse marine settings and for different species, deep learning is assessed as a very promising technique for creating and implementing cost-effective, automatic marine monitoring systems.

Pitch detection is the task of identifying a sound's fundamental frequency (F0) and is applicable in many industrial fields, including music and fault detection, with different characteristics being sought from the algorithms, especially in terms of speed/real-time vs. accuracy. Contribution 7 by Coccoluto et al. proposes a novel, high-speed pitch detection algorithm called OBP (One Bit Pitch) to realize the fastest possible pitch detection, suitable for applications needing real-time tracking. It is implemented using a bitwise variant of the autocorrelation method used for estimating F0 [17] and operated on a synthetic dataset of signals with a known pitch. OBP is compared to the state-of-the-art algorithms for all other applications, including a deep learning-based one, and demonstrates higher speed than all of them, being nine times faster than the previous fastest one, with a mean elapsed time of 0.046 $\times$ real time. Despite being less accurate for high-precision landmarks and noisy signals, its performances are still within the realm of acceptability (<2%) and comparable to standards like YIN [18].

Danilo Greco's Contribution 8 deals with acoustic imaging for spatially mapping sounds: it is based on heavy "acoustic cameras" comprising spatially localized microphone arrays, and the study explores techniques for shrinking their size and enhancing their portability. A starting prototype employing a 128-microphone array spanning 50 by 50 cm^2, along with beamforming techniques, for real-time sound visualization is explored. Through simulations, to reduce the array size without compromising noise and distortion, it is shown that a strategically placed array of 32 microphones, approximately 20 by 20 cm^2, can surpass the larger prototype's performance in terms of directionality and noise cancellation for frequencies under 4 kHz, achieving a fourfold reduction in area with manageable compromises. The portability of these downsized acoustic imaging devices will broaden their utility in vehicle surveillance, urban soundscapes, and various industrial applications that are currently constrained by the larger sizes of traditional systems.

Staying within the realm of sound field analysis, Contribution 9 delves into enhancing personal audio systems with sound field control technology, with specific attention paid to the physical properties of sound such as particle velocity. The authors Wang and Zhang introduce a novel sound field control strategy that aims to minimize reconstruction error in the bright zone, reduce loudspeaker array exertion, and manage particle velocity and sound pressure in the dark zone. Utilizing a setup of five irregularly placed loudspeakers for computer simulations, they benchmark their method against established techniques

such as pressure matching (PM), the eigen decomposition pseudoinverse method (EDPM), and acoustic contrast control (ACC), proving that their method lowers the loudspeaker array's effort and surpasses those of PM and EDPM in the bright zone's acoustic contrast index and outperforms ACC in reducing reconstruction errors. Specifically, the average effort of their array is substantially lower, achieving reductions of 9.4790 dB, 8.0712 dB, and 4.8176 dB against the ACC, PM, and EDPM methods, respectively, indicating that their approach provides the most stable reconstruction system, especially with non-uniform loudspeaker placements.

This efficiency underscores the potential of their approach to significantly improve personal audio systems by ensuring more stable and efficient sound field reconstruction in setups where loudspeakers are unevenly distributed.

Finally, Scarpiniti et al. in Contribution 10 propose a CNN classifier based on scalograms for audio classification in construction sites. CNNs are a common solution for audio classification, and they are usually compared to more traditional machine learning solutions [19]. The scalogram is defined as the square modulus of the Continuous Wavelet Transform (CWT) [20], and here it is compared with the spectrogram, the most common solution for CNN-based audio analysis. Experimental results obtained using real-world sounds recorded at construction sites demonstrate state-of-the-art performances for their CNN + scalogram proposal, with a net based on AlexNet, compared with common architectures that employ spectrograms, demonstrating the effectiveness and completeness of the scalogram as a source of information for CNN-based sound analysis.

3. Conclusions

New technologies, the advancement of AI and its applications, and the availability of higher computational power are driving the evolution of audio processing and analysis techniques. Conversely, these fast-paced advancements make it possible to discover the informative power of audio signals, which are being ever-more utilized for automatic analysis.

New advancements in audio processing techniques include the deep learning-based analysis of sound signals, most notably speech and sound events, leading to the rise in CNN-based approaches. State-of-the-art results have been achieved and constantly revised for applications like COVID-19 monitoring and marine and construction sound event detection, and novel ways of enhancing CNNs have been explored, such as the usage of scalograms for challenging the current paradigm of using spectrograms or data augmentation on the test set.

On the other hand, pre-existing solutions such as pitch detection or sound field modeling are pushed to their limits with faster, more efficient or accurate algorithms, whose development is made possible by modern equipment and whose application is driven by current digital and real-time industrial needs.

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

List of Contributions

1. Strianese, M.; Torricelli, N.; Tarozzi, L.; Santangelo, P. Experimental Assessment of the Acoustic Performance of Nozzles Designed for Clean Agent Fire Suppression. *Appl. Sci.* **2023**, *13*, 186. https://doi.org/10.3390/app13010186.
2. Kim, S.; Baek, J.; Lee, S. COVID-19 Detection Model with Acoustic Features from Cough Sound and Its Application. *Appl. Sci.* **2023**, *13*, 2378. https://doi.org/10.3390/app13042378.
3. Tamulionis, M.; Sledevič, T.; Serackis, A. Investigation of Machine Learning Model Flexibility for Automatic Application of Reverberation Effect on Audio Signal. *Appl. Sci.* **2023**, *13*, 5604. https://doi.org/10.3390/app13095604.
4. Atamer, S.; Altinsoy, M. Vacuum Cleaner Noise Annoyance: An Investigation of Psychoacoustic Parameters, Effect of Test Methodology, and Interaction Effect between Loudness and Sharpness. *Appl. Sci.* **2023**, *13*, 6136. https://doi.org/10.3390/app13106136.

5. Lee, S.; Kim, H.; Jang, G. Weakly Supervised U-Net with Limited Upsampling for Sound Event Detection. *Appl. Sci.* **2023**, *13*, 6822. https://doi.org/10.3390/app13116822.
6. Nanni, L.; Cuza, D.; Brahnam, S. Building Ensemble of Resnet for Dolphin Whistle Detection. *Appl. Sci.* **2023**, *13*, 8029. https://doi.org/10.3390/app13148029.
7. Coccoluto, D.; Cesarini, V.; Costantini, G. OneBitPitch (OBP): Ultra-High-Speed Pitch Detection Algorithm Based on One-Bit Quantization and Modified Autocorrelation. *Appl. Sci.* **2023**, *13*, 8191. https://doi.org/10.3390/app13148191.
8. Greco, D. A Feasibility Study for a Hand-Held Acoustic Imaging Camera. *Appl. Sci.* **2023**, *13*, 11110. https://doi.org/10.3390/app131911110.
9. Wang, S.; Zhang, C. A Stable Sound Field Control Method for a Personal Audio System. *Appl. Sci.* **2023**, *13*, 12209. https://doi.org/10.3390/app132212209.
10. Scarpiniti, M.; Parisi, R.; Lee, Y. A Scalogram-Based CNN Approach for Audio Classification in Construction Sites. *Appl. Sci.* **2024**, *14*, 90. https://doi.org/10.3390/app14010090.

References

1. Gourishetti, S.; Grollmisch, S.; Abeßer, J.; Liebetrau, J. Potentials and Challenges of AI-based Audio Analysis in Industrial Sound Analysis. In Proceedings of the Conference: 48. Deutsche Jahrestagung für Akustik (DAGA), Stuttgart, Germany, 21–24 March 2022.
2. Faundez-Zanuy, M.; Monte-Moreno, E. State-of-the-art in speaker recognition. *IEEE Aerosp. Electron. Syst. Mag.* **2005**, *20*, 7–12. [CrossRef]
3. Wang, Y.; Song, W.; Tao, W.; Liotta, A.; Yang, D.; Li, X.; Gao, S.; Sun, Y.; Ge, W.; Zhang, W.; et al. A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances. *Inf. Fusion* **2022**, *83–84*, 19–52. [CrossRef]
4. Merino-Martínez, R.; Sijtsma, P.; Snellen, M.; Ahlefeldt, T.; Antoni, J.; Bahr, C.J.; Blacodon, D.; Ernst, D.; Finez, A.; Funke, S. A review of acoustic imaging methods using phased microphone arrays. *CEAS Aeronaut. J.* **2019**, *10*, 197–230. [CrossRef]
5. Ruslan, N.; Mamat, M.; Porle, R.; Parimon, N. A Comparative Study of Pitch Detection Algorithms for Microcontroller Based Voice Pitch Detector. *Adv. Sci. Lett.* **2017**, *23*, 11521–11524. [CrossRef]
6. Deep Learning for Audio Signal Processing | IEEE Journals & Magazine. Available online: https://ieeexplore.ieee.org/document/8678825 (accessed on 20 February 2024).
7. Costantini, G.; Di Leo, P.; Asci, F.; Zarezadeh, Z.; Marsili, L.; Errico, V.; Suppa, A.; Saggio, G. Machine learning based voice analysis in spasmodic dysphonia: An investigation of most relevant features from specific vocal tasks. In Proceedings of the 14th International Conference on Bio-Inspired Systems and Signal Processing—BIOSIGNALS, 2021, Vienna, Austria, 11–13 February 2021.
8. Bogey, C.; Marsden, O.; Bailly, C. Influence of initial turbulence level on the flow and sound fields of a subsonic jet at a diameter-based Reynolds number of 105. *J. Fluid Mech.* **2012**, *701*, 352–385. [CrossRef]
9. Bogert, B.P. The quefrency alanysis of time series for echoes; Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. *Time Ser. Anal.* **1963**, 209–243.
10. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
11. Xia, T.; Spathis, D.; Brown, C.; Chauhan, J.; Grammenos, A.; Han, J.; Hasthanasombat, A.; Bondareva, E.; Dang, T.; Floto, A.; et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. In Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Virtual, 6–14 December 2021.
12. Bhattacharya, D.; Sharma, N.K.; Dutta, D.; Chetupalli, S.R.; Mote, P.; Ganapathy, S.; Chandrakiran, C.; Nori, S.; Suhail, K.K.; Gonuguntla, S.; et al. Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection. *Sci. Data* **2023**, *10*, 397. [CrossRef] [PubMed]
13. Orlandic, L.; Teijeiro, T.; Atienza, D. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Sci. Data* **2021**, *8*, 156. [CrossRef] [PubMed]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. In *Lecture Notes in Computer Science*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Swizherland, 2015; pp. 234–241. [CrossRef]
15. Mesaros, A.; Heittola, T.; Virtanen, T. A multi-device dataset for urban acoustic scene classification. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018, Surrey, UK, 19–20 November 2018.
16. Stowell, D. Computational bioacoustics with deep learning: A review and roadmap. *PeerJ* **2022**, *10*, e13152. [CrossRef] [PubMed]
17. Staudacher, M.; Steixner, V.; Griessner, A.; Zierhofer, C. Fast fundamental frequency determination via adaptive autocorrelation. *EURASIP J. Audio Speech Music Process.* **2016**, *2016*, 17. [CrossRef]
18. de Cheveigne, A.; Kawahara, H. YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* **2002**, *111*, 1917–1930. [CrossRef] [PubMed]

19. Costantini, G.; Cesarini, V.; Leo, P.D.; Amato, F.; Suppa, A.; Asci, F.; Pisani, A.; Calculli, A.; Saggio, G. Artificial Intelligence-Based Voice Assessment of Patients with Parkinson's Disease Off and On Treatment: Machine vs. Deep-Learning Comparison. *Sensors* **2023**, *23*, 2293. [CrossRef] [PubMed]
20. Salles, R.S.; Ribeiro, P.F. The use of deep learning and 2-D wavelet scalograms for power quality disturbances classification. *Electr. Power Syst. Res.* **2023**, *214*, 108834. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.