ACOUSTIC CLASSIFICATION OF PACKAGING WASTE BASED ON ULTRASONIC AND AUDIBLE SOUND WAVE EMISSIONS IN REVERBERANT ENCLOSURES

Marc Arnela, Ester Vidaña-Vila, Augusto Fantinelli, Sergi Rubio-Garcia, Alejandro Moñux-Bernal, Jesús Vaquerizo-Serrano, Joan Claudi Socoró[1]

[1] *Human-Environment Research (HER), Department of Engineering, La Salle, Universitat Ramon Llull, Barcelona, Catalonia, Spain*

Abstract

Reverse vending machines encourage recycling, but require automatic container classification. This is typically achieved by reading barcodes or using computer vision, which can fail due to unreadable codes or poor lighting. Recently, an active acoustic approach using sound waves has been proposed as an alternative, offering low computational cost and minimal maintenance. However, its performance depends on the acoustic environment. This study investigates how enclosure geometry and wall absorption influence classification accuracy. Acoustic impulse responses were measured using exponential sine sweeps, with an omnidirectional parametric loudspeaker generating ultrasonic and audible waves via the parametric acoustic array effect. The data trained machine learning and deep learning models to classify plastic, metal cans, glass, and cartboard-tetrabrick. The tests were performed in a mini-reverberation chamber and in a shoebox enclosure with and without wall-mounted absorbers. The highest accuracy was achieved in the mini-reverberation chamber (96%), highlighting the benefits of diffuse and reverberant environments.

Keywords: *Ultrasounds, parametric acoustic loudspeaker, acoustic classification, machine and deep learning, packaging waste.*

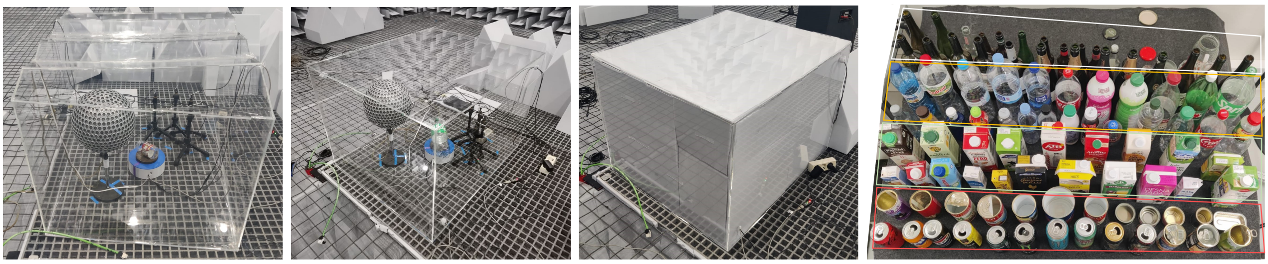
1. Introduction

Recycling at home often requires manual classification of products, a process that is not always followed correctly despite clear guidelines from public entities. Improving recycling rates is crucial for transitioning to a circular economy, as emphasized by the European Commission and the United Nations' Sustainable Development Goal 12 on sustainable production and consumption [1]. A device that ensures proper waste sorting at the source would significantly aid this goal.

Reverse vending machines (RVMs) can enhance waste sorting by collecting empty recyclable containers and offering rewards, thus encouraging recycling. These machines typically use various sensors, such as near-infrared, barcode, weight sensors, and cameras, to classify recyclable items [2]. However, barcode-based systems fail with damaged or unreadable barcodes, and computer vision systems face challenges including high computational demands and sensitivity to environmental conditions such as lighting [3]. Acoustic systems provide a promising alternative. They rely on microphones, which are easy to maintain, and processing sound requires less computational power than image-based systems. Most acoustic systems operate passively, recognizing the sound produced by the container when it is returned [4], [5], [6], [7]. A limitation of passive approaches, however, is their vulnerability to background noise.

This article is based on an active acoustic approach previously developed by the authors [8], [9]. The system uses a parametric acoustic loudspeaker (PAL) as a sound source, capable of simultaneously emitting ultrasonic and audible sound waves. While PALs are commonly used to reproduce highly focused sound beams [10], [11], [12], [13], here we use an omnidirectional parametric loudspeaker (OPL) [14], [15], [16], [17], which radiates ultrasonic and audible waves in all directions. These sound waves are emitted inside an enclosure containing the target container, ensuring a high signal-to-noise ratio. The acoustic impulse response with the container inside is measured using the Exponential Sine Sweep (ESS) technique [15], [16], [17]. These impulse responses capture the container's effect on the acoustic field, enabling classification of

Corresponding author: *marc.arnela@salle.url.edu.*



(a) (b) (c) (d)

Figure 1: Experimental setups with the three enclosures: (a) Mini-reverberant, (b) shoebox without absorption, and (c) shoebox with absorption, and (d) dataset of containers made of metal (red box), cardboard (green box), plastic (orange box) and glass (white box).

various types of packaging waste. Previous work validated this method in a downscaled reverberation chamber under near-ideal acoustic conditions. However, the geometry and acoustic properties of a typical RVM differ substantially from a reverberation chamber, and the physical size of the OPL (0.125 m radius, 750 transducers) poses integration challenges. The present study addresses some of these issues by investigating how enclosure characteristics—specifically geometry and wall absorption—affect classification accuracy. Three enclosures are considered (see Figure 1): (1) a mini-reverberation chamber as a reference (mini-reverberant), (2) a shoebox-style enclosure resembling the tunnel of an RVM (shoebox reverberant), and (3) the same shoebox enclosure lined with sound-absorbing material (shoebox absorbent). Experiments in these enclosures allow us to explore how sensitive the technology is to environmental changes, which in an RVM can arise from variations in tunnel geometry and wall materials. The findings may guide the design of RVMs optimized for the active acoustic technology under development.

The remainder of this work is organized as follows. Section 2 describes the experimental setups for the different enclosure configurations, outlines the procedure for generating the dataset of ultrasonic and audible impulse responses, and presents the classification methodology. Section 3 reports the classification results, providing performance metrics for each enclosure. Finally, Section 4 concludes the paper with a summary of the results and the main conclusions.

2. Methodology

2.1. Experimental setup

Three different enclosures were built to assess the sensitivity of the technology to environmental conditions (see Figure 1). First, the downscaled reverberation chamber described in [8] was used as a reference (mini-reverberant), but reconstructed with methacrylate instead of polycarbonate to increase sound reflection and improve mechanical stability (subfigure 1a). Second, a methacrylate shoebox-shaped enclosure with dimensions 0.98 m × 0.82 m × 0.67 m (shoebox reverberant) was constructed to evaluate sensitivity to geometric variations (subfigure 1b). Finally, the third enclosure (shoebox absorbent) incorporated sound-absorbing material into the shoebox design to investigate the impact of sound absorption on the technology's performance (subfigure 1c). The experiments were conducted in the anechoic chamber at La Salle – Universitat Ramon Llull to ensure minimal background noise.

An omnidirectional parametric loudspeaker (OPL) [14], [15], [16], [17], composed of 750 ultrasonic transducers arranged on a spherical surface, was employed to generate ultrasonic and audible sound waves inside each enclosure using the parametric acoustic array (PAA) effect [18]. The exponential sine sweep method developed in [16] was used to simultaneously measure the ultrasonic and audible impulse response of each container. A set of four free-field microphones (two GRAS 40BF and two GRAS 46BF-1) was used to measure the sound field inside the different enclosures. The four microphones were placed approximately 20 cm around the center of the measured package. A Brüel & Kjaer LAN-XI Type 3160-A-042 module with a UA-3102-042 front-end was used to generate the ESS and record the response measured by the microphones. The generated signal was amplified to 17 V peak using an Ecler XPA3000 amplifier and sent to the OPL. The container was placed on an Arduino-controlled rotating table, which rotated the items by 30° between successive sweeps.

2.2. Dataset

Three datasets of ultrasonic and audible impulse responses have been generated for the three different enclosures (mini-reverberant, shoebox without absorbent, and shoebox with absorbent). Each dataset includes 100 containers made of four different materials: metal, cardboard, plastic, and glass, with a balanced distribution of 25, 20, 25, and 30 examples, respectively, to ensure statistically unbiased results. The containers are depicted in Figure 1d. Each container was subjected to 96 measurements, comprising 12 angles, 2 orientations (horizontal and vertical), and 4 repetitions for redundancy. Each measurement was recorded by 4 differently positioned microphones, resulting in 384 measurements per container and 38400 samples per dataset. Since impulse responses are computed for both the ultrasonic and audible frequency ranges, the size of each dataset is doubled to 76800 samples.

2.3. Ultrasonic and audible impulse responses

The characterization of each packaging waste item begins with the measurement of its impulse response, for which Exponential Sine Sweeps (ESS) are utilized. The sweep signal is mathematically described by [19]

$$x(t) = \sin\left[K\left(e^{t/L} - 1\right)\right] K, \tag{1}$$

with the constants K and L defined as

$$K = \frac{2\pi f_1 T}{\ln f_2/f_1}, \quad L = \frac{T}{\ln f_2/f_1}. \tag{2}$$

The sweep spans from $f_1 = 20$ Hz to $f_2 = 14$ kHz and lasts $T = 10$ seconds. To ensure the full sweep is captured, a 2 second silence precedes the signal, while a 3 second silence follows it to allow residual energy to decay. To mitigate pre-ringing effects during deconvolution [20], fade-in and fade-out ramps are applied at the beginning and end of the sweep, respectively . These ramps span from 20 Hz to 80 Hz at the start and from 12 kHz to 14 kHz at the end, effectively restricting the usable frequency range to 80 Hz–12 kHz.

To enable ultrasonic playback through the OPL, the ESS is modulated using Upper Side Band Amplitude Modulation (USBAM) with a carrier frequency of $f_c = 41.4$ kHz [16]. Once emitted, the modulated signal generates both audible and ultrasonic components due to the PAA effect [18]. These are captured by the microphones described in Section 2.1, resulting in the recorded signal $y(t)$, which, after its digitalization using a sampling rate of $f_s = 131072$ Hz and 24 bits/sample, results in its discrete counterpart $y[n]$. Both discrete audible and ultrasound impulse responses $h_a[n]$ and $h_u[n]$ are then obtained via convolution with the corresponding inverse discrete filter, following [16]. $h_a[n]$ and $h_u[n]$ encapsulate both linear and nonlinear system behaviors within their respective frequency domains. To isolate the linear component, a 100 ms rectangular window is applied. Subsequently, both $h_a[n]$ and $h_u[n]$ are downsampled to $f_s'' = 44100$ Hz. Since $h_a[n]$ contains audible content, it remains unaffected by this operation. In contrast, $h_u[n]$, which contains ultrasonic components centered around f_c, must be frequency-shifted into the audible range to enable further analysis. This yields $h_{ua}[n]$, an audible version of $h_u[n]$ obtained through the process detailed in [8].

An example of a measurement signal $y[n]$, along with the corresponding audible impulse response $h_a[n]$ and ultrasonic impulse response $h_{ua}[n]$, is illustrated in the block diagram shown in Figure 2. The next step involves extracting features from each of the two impulse response modalities.

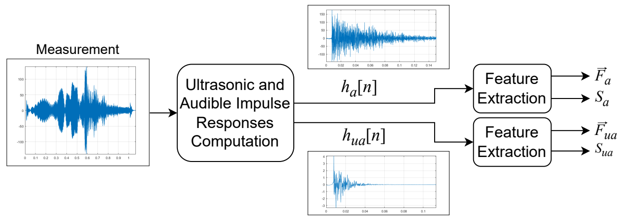


Figure 2: Computation of the audible impulse response $h_a[n]$ and ultrasonic impulse response $h_{ua}[n]$ from a measured signal $y[n]$, followed by the extraction of the vector of spectral features $\vec{F}$ and mel-spectrogram S.

2.4. Feature extraction

Two types of features are computed from $h_a[n]$ and $h_{ua}[n]$: short-term cepstral and spectral features, and mel-spectrograms. For both feature types, short-time analysis is applied using L Hanning windows of duration T_w and overlap O_v%. Two configurations are considered to evaluate the sensitivity of the window parameters : i) $C_1 = \{L = 5, T_w = 20 \text{ ms}, O_v = 50\%\}$, covering 60 ms of multipath delay, and ii) $C_2 = \{L = 20, T_w = 10 \text{ ms}, O_v = 75\%\}$, with 57.5 ms. Both configurations capture the most energetic multipath components of the audible and ultrasonic fields, with C_1 offering higher frequency resolution at the expense of lower temporal resolution, and C_2 providing the opposite trade-off.

From each k modality (a = audio, ua = ultrasonic audible) and analysis window, a vector $\vec{F}_k$ with 35 component features is extracted, comprising 13 Mel-Frequency Cepstral Coefficients (MFCCs), 13 Gammatone-Frequency Cepstral Coefficients (GFCCs), and a range of spectral descriptors: spectral spread, spectral flux, spectral flatness, spectral centroid, spectral rolloff point, spectral crest, spectral entropy, spectral kurtosis, and spectral skewness. Additionally, mel-spectrograms S_k are also computed using the same windowing parameters and N_B Mel subbands.

2.5. Machine learning and deep learning

Two different approaches have been evaluated for material classification. First, a Support Vector Machine (SVM), one of the best performing classical machine learning techniques (as shown in our previous study [9]), using as input the feature vectors $\vec{F}_k$. Second, a deep learning approach based on Ultralytics' YOLOv8 [21], which builds on the YOLO (You Only Look Once) architecture [22]. Originally designed for computer vision tasks, YOLOv8 was used in its pre-trained version on the ImageNet Dataset (https://www.image-net.org/) and subsequently adapted and fine-tuned with our database of spectrograms S_k for each enclosure.

Each training session employed 5-fold cross-validation, ensuring distinct waste packaging items as well as balanced number of all class items in each validation split. All impulse responses from a given package (includ-

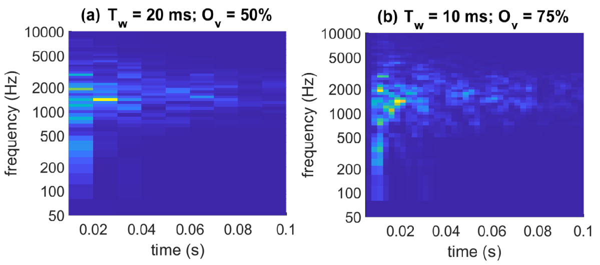


Figure 3: Mel-spectrograms of the audible impulse response $h_a[n]$ for a waste recycling item, obtained using windowing configurations (a) C_1 and (b) C_2.

ing positional variations, receiver microphones and measurement repetitions) were assigned exclusively to either training or validation within each fold, yielding realistic validation sets with entirely unseen examples. Both mean Accuracy and Macro F1 standard measures where computed to assess the quality of material classification for each experiment.

3. Results

Figure 3 shows an example of Mel-spectrogram analysis applied to an audible impulse response, using $N_B = 64$ Mel subbands and the windowing configurations C_1 and C_2. In C_2, the spectrogram offers a clear visualization of the dominant multipath components, which can be identified as bright yellow spots. In contrast, configuration C_1 uses fewer and larger windows, resulting in a noticeable reduction in time resolution and a less detailed representation. From a visual standpoint, C_2 thus offers a more accurate representation of the impulse response.

The classification results are presented in Table 1, showing the performances of SVM and YOLOv8 when using the windows C_1 and C_2 in the three enclosures under study. As can be observed, the enclosure plays a crucial role in determining the quality of the inputs for classification. The mini-reverberant chamber yields the highest accuracy and F1-scores with both the SVM and YOLOv8 classifiers, reaching a maximum F1 value of 96%. In contrast, performance is moderately reduced in the shoebox without absorbent material, with a maximum F1 score of 89%, and substantially degraded in the shoebox with absorbent material, where the maximum F1 score drops to 75%. These results suggest that reverberation is important for classification performance, indicating that not only the direct sound field but also the reflected components provide valuable information to the algorithms.

Focusing now on the algorithms, SVM offers similar or slightly better performance than YOLOv8. In terms of windowing, the higher temporal resolution of the inputs in C_2 produces slightly better results than C_1. In all three configurations, the improvement in C_2 is limited to a maximum of 5% in the shoebox with absorbent material and only 1% in the mini-reverberant chamber.

Table 1: Accuracy and Macro F1 (mean ± standard deviation) for the different enclosures and methods: SVM and YOLOv8 with window configurations C_1 and C_2.

Enclosure	Method	Accuracy	Macro F1
Mini-reverb	SVM + C1	94%±1%	94%±2%
	SVM + C2	95%±2%	96%±3%
	YOLOv8 + C1	95%±2%	95%±2%
	YOLOv8 + C2	96%±2%	96%±2%
Shoebox reverb	SVM + C1	87%±9%	86%±10%
	SVM + C2	90%±10%	89%±10%
	YOLOv8 + C1	86%±8%	86%±9%
	YOLOv8 + C2	89%±8%	88%±7%
Shoebox abs	SVM + C1	71%±5%	70%±5%
	SVM + C2	76%±7%	75%±8%
	YOLOv8 + C1	69%±6%	69%±5%
	YOLOv8 + C2	71%±5%	71%±5%

4. Conclusions

This study explored the use of active acoustic methods based on ultrasonic and audible sound wave emissions for packaging waste classification, with a focus on enclosure characteristics. Results show that reverberant and diffuse sound fields are most favorable, achieving 96% classification accuracy in a mini-reverberation chamber, while performance decreases to 90% in a shoebox-shaped enclosure and 75% with added absorption. These findings suggest that not only early reflections but also late reverberation carry valuable information, and that environments with non-absorbing, irregular walls—promoting diffusion and reverberation—may provide more robust classification. Among the tested algorithms, SVM slightly outperformed YOLOv8, while higher temporal resolution in feature extraction led to modest improvements. Overall, the study confirms the potential of active acoustic classification for container recognition. Future work will focus on integrating this approach into a reverse vending machine (RVM), which will include a reverberant acoustic enclosure, a single microphone to capture the sound waves, and a parametric acoustic loudspeaker (PAL) specifically designed to fit within the spatial constraints of the RVM.

5. Acknowledgments

This publication is part of the project CharPAL PID2025-175028NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU. The authors thank the Catalan Government (Departament de Recerca i Universitats) for the grant 2021 SGR 01396 awarded to the HER group.

References

[1] K. Kotyal, "Sustainable waste management in the circular economy: Challenges and opportunities," *Environmental Reports*, 2023.

[2] T. Yoo, S. Lee, and T. Kim, "Dual image-based CNN ensemble model for waste classification in reverse vending machine," *Applied Sciences*, vol. 11, no. 22, p. 11 051, 2021.

[3] S. Majchrowska et al., “Deep learning-based waste detection in natural and urban environments,” *Waste Management*, vol. 138, pp. 274–284, 2022.

[4] O. I. Funch, R. Marhaug, S. Kohtala, and M. Steinert, “Detecting glass and metal in consumer trash bags during waste collection using convolutional neural networks,” *Waste Management*, vol. 119, pp. 30–38, 2021.

[5] M. K. Korucu, Ö. Kaplan, O. Büyük, and M. K. Güllü, “An investigation of the usability of sound recognition for source separation of packaging wastes in reverse vending machines,” *Waste Management*, vol. 565, pp. 46–52, 2016.

[6] G. Lu, Y. Wang, H. Yang, and J. Zou, “One-dimensional convolutional neural networks for acoustic waste sorting,” *Journal of Cleaner Production*, vol. 271, p. 122 393, 2020.

[7] T. Munaka, F. Samiea, L. Bauer, and J. Henkel, “Improved feature extraction method for sound recognition applied to automatic sorting of recycling wastes,” *Journal of Information Processing*, vol. 28, pp. 658–665, 2020.

[8] M. Arnela, E. Vidaña-Vila, A. Fantinelli, A. Moñux-Bernal, J. Vaquerizo-Serrano, and J. C. Socoró, “Generation of ultrasonic and audible sound waves for the automatic classification of packaging waste in reverse vending machines,” *Waste Management*, vol. 204, p. 114 934, 2025.

[9] M. Arnela et al., “Automatic classification of packaging waste at source using acoustic technology,” in *Proc. of Forum Acusticum 2025*, Málaga, Spain, Jun. 2025.

[10] M. Alunno and A. Yarce Botero, “Directional landscapes: Using parametric loudspeakers for sound reproduction in art,” *Journal of new music research*, vol. 46, no. 2, pp. 201–211, 2017.

[11] C. Shi, H. Mu, and W.-S. Gan, “A psychoacoustical preprocessing technique for virtual bass enhancement of the parametric loudspeaker,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 31–35.

[12] M. Li, T. Zhuang, K. Chen, J.-X. Zhong, and J. Lu, “Deep learning-based approach for identification and compensation of nonlinear distortions in parametric array loudspeakers,” *IEEE Signal Processing Letters*, 2025.

[13] J.-X. Zhong, J. Ji, X. Xia, H. Heo, and Y. Jing, “Audible enclaves crafted by nonlinear self-bending ultrasonic beams,” *Proceedings of the National Academy of Sciences*, vol. 122, no. 12, e2408975122, 2025.

[14] M. Arnela, O. Guasch, P. Sánchez-Martín, J. Camps, R. M. Alsina-Pagès, and C. Martínez-Suquía, “Construction of an omnidirectional parametric loudspeaker consisting in a spherical distribution of ultrasound transducers,” *Sensors*, vol. 18, no. 12, p. 4317, 2018.

[15] M. Arnela, C. Martínez-Suquía, and O. Guasch, “Characterization of an omnidirectional parametric loudspeaker with exponential sine sweeps,” *Appl. Acoust.*, vol. 182, p. 108 268, 2021.

[16] M. Arnela, C. Martínez-Suquía, and O. Guasch, “Carrier frequency influence on the audible and ultrasonic fields generated by an omnidirectional parametric loudspeaker excited with exponential sine sweeps,” *Appl. Acoust.*, vol. 200, p. 109 073, 2022.

[17] M. Arnela, R. Burbano-Escolà, R. S. Ribeiro, and O. Guasch, “Reverberation time and random-incidence sound absorption measured in the audible and ultrasonic ranges with an omnidirectional parametric loudspeaker,” *Appl. Acoust.*, vol. 229, p. 110 414, 2025.

[18] F. J. J. Pompei, “Sound from ultrasound, The parametric array as an audible sound source,” Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

[19] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Audio Engineering Society Convention 108*, Feb. 2000.

[20] I. Bork, “Exponential sweep check using Hilbert transform,” *Acta Acust. united Ac.*, vol. 100, no. 4, pp. 659–666, 2014.

[21] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, “A review on yolov8 and its advancements,” in *International Conference on Data Intelligence and Cognitive Informatics*, Springer, 2024, pp. 529–545.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *You only look once: Unified, real-time object detection*, 2016. arXiv: `1506.02640 [cs.CV]`. [Online]. Available: `https://arxiv.org/abs/1506.02640`.