DOI: 10.61782/fa.2025.0385

FORUM ACUSTICUM
EURONOISE 2025

EXPLORING ACOUSTIC SIGNATURES OF DIFFERENT AIRCRAFT TYPES AND OPERATIONS USING ADVANCED DATA ANALYSIS

Irina Besnea[1*] **Alireza Amiri-Simkooei**[1] **Irene C. Dedoussi**[2] **Mirjam Snellen**[1]

[1] Operations & Environment, Faculty of Aerospace Engineering, Delft University of Technology The Netherlands
[2] Department of Engineering, University of Cambridge, United Kingdom

ABSTRACT

Understanding acoustic characteristics of aircraft is critical for designing optimal fleet compositions in terms of noise and improved airport operations. This study investigates acoustic signatures across different aircraft types, engine designs, and operational conditions. A dataset consisting of 457 field acoustic measurements of commercial turbofan aircraft landing and taking-off from Amsterdam Airport Schiphol was used. To unveil meaningful patterns, we focused on dimensionality reduction techniques—Principal Component Analysis (PCA) and t-distributed Stochastic Neighbour Embedding (t-SNE)—to analyse this high-dimensional acoustic data. These methods are complemented by clustering algorithms and supervised machine learning models, such as K-Means, random forests for feature importance, and multilayer perceptrons (MLP) to classify aircraft types, engine configurations, and operations.

Results reveal a strong loudness axis in the first principal component, overshadowing subtle spectral and time-based differences across aircraft families, especially for takeoffs. Nonetheless, focusing on higher-order components and alternative embeddings (t-SNE) highlights additional spectral and temporal markers. Operation classification (landing vs. takeoff) achieves 98% accuracy, but aircraft and engine family classification remain challenging, with accuracy capped below 50% using these feature sets. These findings suggest that advanced feature selection and dimensionality reduction while considering amplitude characteristics are essential for disentangling nuanced design-based acoustic traits.

Keywords: *PCA, K-Means, aircraft noise.*

1. INTRODUCTION

Aircraft noise remains a significant environmental issue, directly impacting community well-being, airport operations, and regulatory policies. With air travel demand continuing to grow, it has become essential to develop innovative approaches that effectively mitigate noise emissions without compromising operational efficiency. This study addresses this critical challenge by investigating the acoustic signatures of commercial turbofan aircraft through advanced, data-driven methodologies. Specifically, techniques such as dimensionality reduction, clustering, and machine learning-based classification are used to identify dominant noise mechanisms linked to various aircraft-engine configurations and operational scenarios. By analyzing real-world acoustic measurements, this research enhances the understanding of noise characteristics across different aircraft types and operational scenarios.

2. DATA COLLECTION AND PREPROCESSING

2.1 Dataset Description

The dataset used in this study consists of 457 acoustic measurements collected at Amsterdam Airport Schiphol in two different locations and two configurations of microphone arrays (see Figure 1). Most measurements were

**Corresponding author*: *i.besnea@tudelft.nl.*

taken at Location 1 between 2019 and 2023 north of runway 36C at approximately 600 m from the runway threshold. Location 2 is situated approximately 900 m from the threshold of runway 18L. All these measurements capture noise emissions from various commercial turbofan aircraft during 206 takeoff and 251 landing operations. Aircraft specifications and engine types were extracted from the ICAO database [1] by matching the ICAO identifiers of individual flyovers recorded during the measurements. While the dataset primarily consists of the new-generation Boeing 737-N aircraft family equipped with CFM56-7 engine family, the limited samples of other aircraft types may hinder accurate aircraft classification.

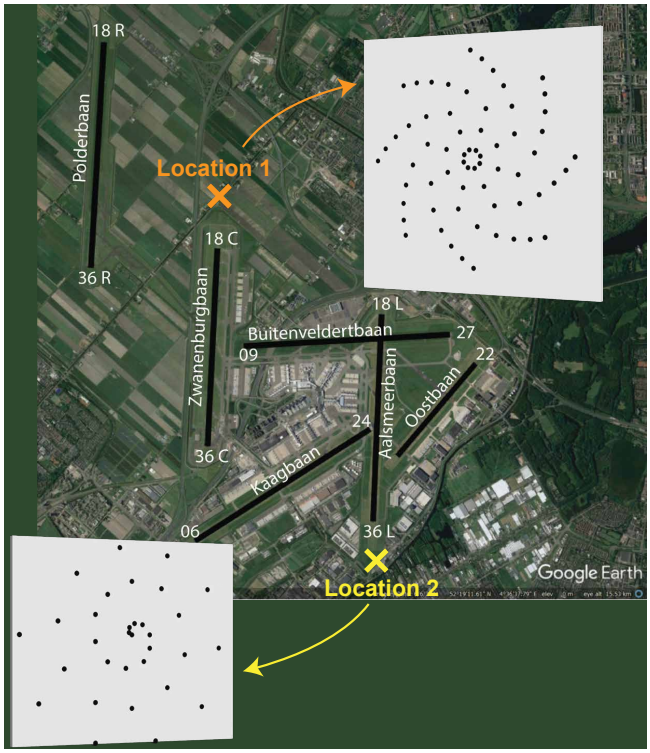

Figure 1. Measurement locations around Amsterdam Airport Schiphol. Location 1, marked in orange, used a 64-microphone array. Location 2, marked in yellow, used a 32-microphone array.

2.2 Data Preprocessing

To ensure the dataset was uniform and suitable for analysis, several pre-processing steps were carried out. First, the source levels were computed by correcting the pressure time series to include spherical spreading and atmospheric absorption. Since Automatic Dependent Surveillance-Broadcast (ADS-B) data can occasionally provide unreliable flight track information, for some flyovers the overhead height was estimated through beamforming (making sure the estimate is within realistic values) [2]. Following the amplitude normalization, feature extraction was then performed to identify 28 relevant spectral and temporal acoustic features, as follows: i) 8 time- and frequency–domain features, i.e. overall sound pressure level (OSPL), mean pressure, variance of the pressure signal, kurtosis of the pressure distribution, skewness of the signal distribution, peak pressure, zero-crossing rate (ZCR), peak frequency determined from FFT of the signal. ii) 11 OSPL-derived features from the smoothed OSPL curve computed over a 1-second window, i.e. maximum rate of change of OSPL curve, OSPL time above 60 dB, skewness, kurtosis and entropy of the OSPL distribution, the maximum, minimum, mean and range of the OSPL values observed, as well as OSPL variance and slope). iii) 5 spectral features from the FFT and power spectral density analysis, i.e. peak frequency determined from power spectrum, energy contained in 100–500 Hz band, 500–2000 Hz range, and 2000–8000 Hz range, as well as the sum of the spectral energy across frequencies). iv) 4 wavelet-based features (using a Daubechies-4 wavelet transform [3]), i.e. energy of the approximation (low-frequency) coefficients, energy of the detail (high-frequency) coefficients, entropy of the normalized approximation coefficients, and entropy of the normalized detail coefficients.

These features include fundamental frequency components, engine tone harmonics, and broadband noise characteristics. They are expected to play a crucial role in discriminating between different aircraft and operational procedures. Subsequently, the extracted features are standardized using Z-score normalization to ensure uniformity before further processing with dimensionality reduction techniques [4].

3. METHODOLOGY

3.1 Dimensionality Reduction Techniques

It is essential to employ dimensionality reduction methods to obtain meaningful patterns, simplify data representation, take the correlations among features into account, and facilitate an effective visual representation of these acoustic data.

One such method is Principal Component Analysis (PCA) [5]. PCA was applied to the complete dataset as well as separately to subsets defined by aircraft operations (takeoff, landing) and measurement location. Over 90% of the variance was captured by the first nine Principal Components (PCs) for the entire dataset, as shown in Figure 2(a). The loadings (i.e. correlation coefficients between the feature and PC) are presented in Figure 2(b). To calculate and interpret the feature importance using the PC loadings (unsupervised feature importance), we take the absolute value of the loading for each feature to give us the magnitude of the contribution. We then weigh them by the percentage variance of each PC and lastly, sum them over all PCs to get a total importance score per feature.

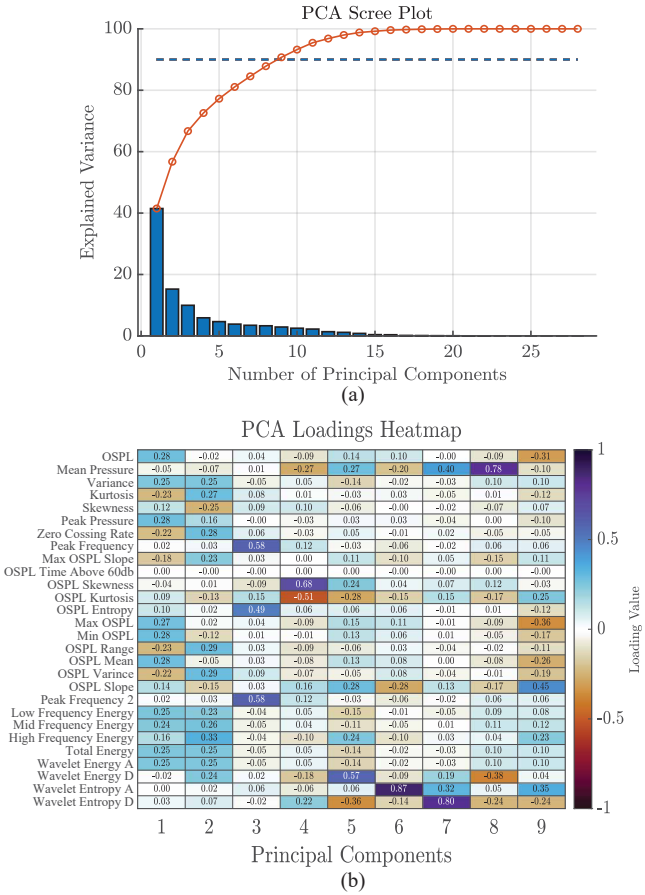

Figure 2. (a) Scree plot of cumulative variance explained by principal components (PCs) for the entire dataset. (b) Heatmap of first nine PCA loadings, highlighting significant features influencing each PC.

In addition to PCA, nonlinear dimensionality reduction methods, such as t-distributed Stochastic Neighbor Embedding (t-SNE) [6] and Uniform Manifold Approximation and Projection (UMAP) [7], were utilized to identify local structures and visualize the multi-dimensional acoustic data in two dimensions. These techniques allow to distinguish nuanced groupings of the dataset that are shown in Figure 3. After verifying the consistent distribution of operation labels between t-SNE and UMAP, only the PCA and t-SNE results will be presented for visualization purposes.

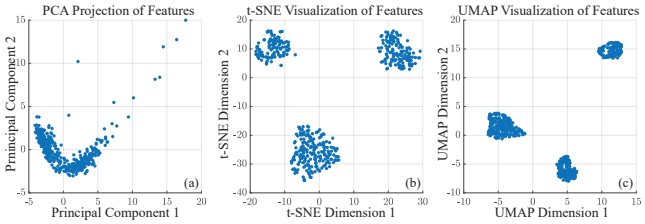

Figure 3. Dataset projection of (a) Principal Component Analysis, (b) t-SNE embedding, and (c) UMAP embedding of acoustic features.

3.2 Clustering Analysis

Since t-SNE discriminates three groups in the dataset, the next natural step was to apply clustering to the entire dataset. K-Means clustering was explored to segment acoustic signatures based on similarities in their dimensionality-reduced representations [8], [9]. K-Means was utilized primarily for the initial exploration and grouping of aircraft types based on PCs. To better understand the key features distinguishing operations, the number of clusters was set to two to confirm the natural clustering of the dataset per operation. However, for aircraft and engine family classification, the optimal number of clusters was determined using the elbow method for both PCA and t-SNE, yielding 4 or 5 clusters as optimum.

3.3 Machine Learning-Based Classification

After clustering, the complete dataset was divided into 60% training, 20% validation, and 20% testing sets to assess the performance of the classification methods. Additionally, a 10-fold cross-validation (CV) approach was used to ensure model robustness.

We tackled two classification tasks: (1) operation recognition (landing vs takeoff) and (2) aircraft/engine family classification. For operation classification, we

trained a Gaussian kernel Support Vector Machine (SVM) [10] on the t-SNE embeddings (over the complete feature matrix - including OSPL/amplitude features) of the entire dataset. This SVM was tuned via nested CV and maps each sample to either landing or takeoff categories. Its predictions then route incoming data down the correct branch (landing/takeoff) in our classification pipeline.

For aircraft and engine family prediction, we created reduced-feature embeddings (after removing amplitude-dominating features) and used PCA and t-SNE for dimensionality reduction within the operational category, as also mentioned in 3.1. We then fit three base models - Random Forest (RF), Gaussian SVM, and Multi-Layer Perceptron (MLP) [11] - on these new embeddings, each optimized (via nested CV) for hyperparameters such as tree depth (RF), kernel scale (SVM), or activation function (MLP). To address class imbalance, oversampling techniques were employed.

Finally, we adopt a hierarchical and stacked approach: the operation SVM first classifies a sample as landing or takeoff, after which the respective aircraft-family models (RF, SVM, MLP) generate predictions. A meta-classifier (linear SVM) then fuses these outputs into the final aircraft or engine label. Such stacking reduces generalization error by exploiting each model's complementary strengths—RF's ensemble partitioning, SVM's kernel-boundary, and MLP's learned representations—and often outperforms any single classifier by combining their complementary biases and structural assumptions, thus handling a broader range of patterns in the dataset [12].

4. RESULTS AND DISCUSSION

4.1 Dimensionality Reduction Insights

On the entire dataset, without separating into operational subsets, PCA revealed that the first nine components effectively captured the most variance ($\geq$90%), highlighting strong correlations among spectral and temporal acoustic features. Analysis of PCA loadings (Figure 2(b)) indicated dominant contributions from low-frequency tonal characteristics and broadband noise measurements.

Visual inspection of the t-SNE plots (see Figure 4) reveals overlapping patterns among aircraft and engine families under different operational conditions, suggesting acoustic similarities despite operational differences.

Due to these observed results, the above PCA was repeated onto the subsets defined by operation and location. The most variances in the datasets were captured by the

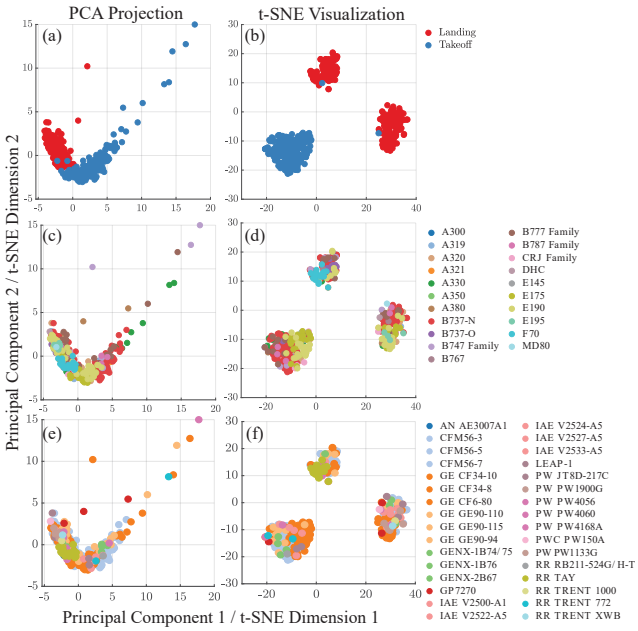

Figure 4. Dataset projection of (a, c, e) PCA and (b,d,f) t-SNE embedding color-coded per Operation (a, b), Aircraft family (c, d) and Engine family (e, f).

first seven PCs for takeoff operations and the top six PCs in the landing operations (regardless of measurement location). However, for consistency, only the first five PCs will be presented for all datasets as they are most representative.

4.1.1 Overall Observations from PCA Loadings

As described in 3.1, the feature importance can be derived from the PC loadings. The first five important features (in order) include: the OSPL variance, slope, the low-frequency energy content, pressure signal variance, the low-frequency coefficients.

It is rather difficult to draw concrete conclusions on the relation between the estimated PCs, extracted from acoustic features, with the operational parameters. We therefore present some observations from this dataset concerning the first four PCs. Interpreting the PCA loadings of Figure 2(b) over the entire dataset (i.e. the combined variance across both operations), reveals that **PC1** primarily reflects a global energy and amplitude scale. Higher scores correlate with louder, more energetic flyovers, while lower scores may indicate spiky or noisy shapes with less consistent energy. In the context of air-

craft noise, this dimension is crucial to distinguish high-energy aircraft from quieter ones. **PC2** reflects a balance between broadband noise and amplitude variability versus signal asymmetry. Therefore the high score of PC2 indicates the contribution of high-order moments of energy in aircraft flyovers. **PC3** may quantify whether a flyover is spectrally tonal (high peak frequency) or high entropy. This axis may help distinguish aircraft types with stable tonal components (turbofan harmonics) from broadband jet noise. **PC4** expresses the very high-order moments (skewness and kurtosis) of amplitude in operation classification. Our observations indicate that this PC is helpful in analyzing arrival vs. departure envelope characteristics, as departure profiles are more symmetric and less impulsive.

In terms of analysis per operations, the first seven PCs were investigated with the most contributing features (positive and negative) presented in Table 1. For landings, PCA was recomputed separately per location. Although classic amplitude metrics (like OSPL) were excluded, **PC1** for both locations still acts as a latent energy axis, now dominated by wavelet energies across multiple frequency bands. PCA helps isolate specific contrasts such as tone vs. broadband noise, burst vs. smooth, high vs. low-frequency behavior, and symmetric vs. asymmetric waveforms. Landings at locations 1 and 2 show similar structures in the first PCs, with some shifts in the order of features across PCs, expressing that the underlying dimensions are consistent but re-ordered.

Takeoff PCs also contain many of the same dominant features as landings (e.g., wavelet entropies, skewness, peak frequency), though their loadings appear more sharply clustered by component — potentially reflecting more distinct acoustic events. Just like in the case of landings, **PC1** still reflects a latent energy dimension, captured through multi-band energy and wavelet energies. The PC breakdown shows that takeoff noise contains rich micro-structural variation in both spectral peaks and time-domain characteristics. Removing amplitude features sharpens PCA's focus on complexity, asymmetry, and frequency content.

4.1.2 Overall Observation from t-SNE Embeddings

Since t-SNE does not produce linear 'loadings' as discussed previously for PCA, it does not inherently provide feature contributions or importance. Therefore, various methods have been proposed to approximate or infer feature importance in t-SNE embeddings, as detailed in [13]. One such approach involves training a predictive model to estimate t-SNE coordinates and analysing the model's feature importance. To implement this, we trained a random forest to predict the t-SNE embeddings from the feature matrix. Feature importance was then computed using MATLAB's built-in function, *predictorImportance*, which quantifies importance by summing changes in node risk due to splits on each predictor and normalizing by the total number of branch nodes [14]. Features with higher importance in the model contribute more significantly to the positioning of the t-SNE embeddings, as shown in Table 2.

4.2 Clustering Analysis and Classification Performance

As discussed in the previous section, we applied K-Means clustering to both principal components and embeddings to assess whether the clusters accurately capture the operational labels. Since t-SNE demonstrates more distinct groupings (see Figure 4(b)), we present the clustering results based on t-SNE in Figure 5. This figure shows boxplots of the two t-SNE embeddings, where cluster 1 corresponds to takeoff flyovers and cluster 2 corresponds to landing flyovers, respectively.

Regarding feature importance, t-SNE1 is driven primarily by OSPL Entropy, OSPL Range, Zero Crossing Rate, Kurtosis, and OSPL Variance, thus suggesting that amplitude fluctuation (range/variance), temporal distribution shape (kurtosis), and rapid oscillations (zero crossings) play major roles in clustering. In contrast, t-SNE2 is most influenced by Peak Frequency, OSPL Entropy, Zero Crossing Rate, OSPL Range, and OSPL Skewness, highlighting that both spectral characteristics (notably the peak frequency) and time-domain shape metrics (skewness/entropy) help distinguish between both operating procedures.

To summarize, the t-SNE plot from Figure 5 shows that landing and takeoff noise can be readily separated based on a combination of time-domain waveform shape (skewness, kurtosis, zero-crossing rate, entropy, variance) and frequency-domain features (peak frequency, spectral range). This suggests that these specific features effectively capture the fundamental acoustic differences between operational procedure events, allowing for a clear separation into two clusters in the embedded space.

Although not shown here, we also interpreted the results of the same boxplot analysis on the first three PCs, out of the nine most important. All in all, **PC1** offers a strong separation between Cluster 1 and Cluster 2, with Cluster 2 exhibiting higher values. This PC likely plays

Table 1. Feature contribution per Principal Component for takeoffs and landings

Component	Takeoffs Location 1		Landings Location 1		Landings Location 2	
Contribution	Positive	Negative	Positive	Negative	Positive	Negative
PC1	Global Energy	Skewness	Global Energy	-	Global Energy	-
PC2	Peak Frequecy; Kurtosis	-	Peak_Frequency; OSPL Kurtosis	OSPL Skewness; ZCR; Kurtosis	Kurtosis; Peak Frequency	OSPL Entorpy
PC3	OSPL Kurtosis; OSPL Entropy	Kurtosis; OSPL Skewness	OSPL Skewness; Peak Frequency; OSPL Entropy	OSPL Kurtosis	OSPL Skewness; OSPL Entropy	OSPL Kurtosis; ZCR
PC4	OSPL Skewness; OSPl Entorypy; Wavelet Entropy D	OSPL Kurtosis	OSPL Entropy	Kurtosis; Peak Frequency	OSPL Kurtosis	OSPL Skewness; ZCR
PC5	Skewness; Wavelet Entropy D	ZCR	Skewness; ZCR; Wavelet Entropy D	-	Wavelet Entropy D	Skewness
PC6	Wavelet Entropy A	Skewness	Wavelet Entropy A; Wavelet Entropy D	-	Wavelet Entropy A	Skewness
PC7	Wavelet Entropy D; ZCR	Skewness	Wavelet Entropy A; Skewness	-	Skewness; ZCR; Wavelet Entropy D	-

Table 2. Ranked features contributing to t-SNE embeddings for the entire dataset

t-SNE 1	t-SNE 2
OSPL Entropy	Peak Frequency
OSPL Range	OSPL Entropy
Zero Crossing Rate	Zero Crossing Rate
Kurtosis	OSPL Range
OSPL Variance	OSPL Skewness

a major role in distinguishing the clusters. **PC2** shows a moderate overlap, with differences in median and spread and although it is not as strong a separator as PC1, it still contributes. **PC3** shows considerable overlap between clusters and likely does not contribute much to the cluster separation. In terms of outliers, cluster 2 seems to have more outliers, especially in PC2 reflecting higher variability in that group.

Since t-SNE provided satisfactory discrimination between operations, the Gaussian kernel SVM model described in 3.3 was applied to the t-SNE Embeddings of the measurements. The model was applied over a 10-fold cross validation and finally yielded a 98% accuracy over the testing dataset.

Clustering, either with PC or t-SNE embeddings, is more challenging for aircraft types and engine types. The PC analysis from 4.1.1 per operation and location showed that energy-based features can significantly contribute to PC1. However it seems that these features cannot discriminate aircraft or engine types. Figure 6 illustrates the K-Means clustering for the case of Landings at Location 2 using the first seven PC components as input. There is no visible pattern in the distribution of either a specific aircraft type or engine type in a particular cluster. Therefore, the clusters are not representative of specific aircraft/engine families, and thus a low accuracy for subsequent classification is expected.

In the end, regardless of the considered dimensionality of the dataset or the considered domain (be it the feature domain, PCA, t-SNE, or UMAP domain), the overall testing accuracy of the models did not exceed 50% when trying to classify aircraft or engine families. This is likely due to the small and imbalanced nature of the dataset. For example the stacked model accuracy of classifying aircraft using the PCA data as input yielded an overall accuracy of 43.96% over the dataset (51% accuracy in takeoffs, 48% accuracy in landings at Location 1, and 21% accuracy for landings at Location 2).

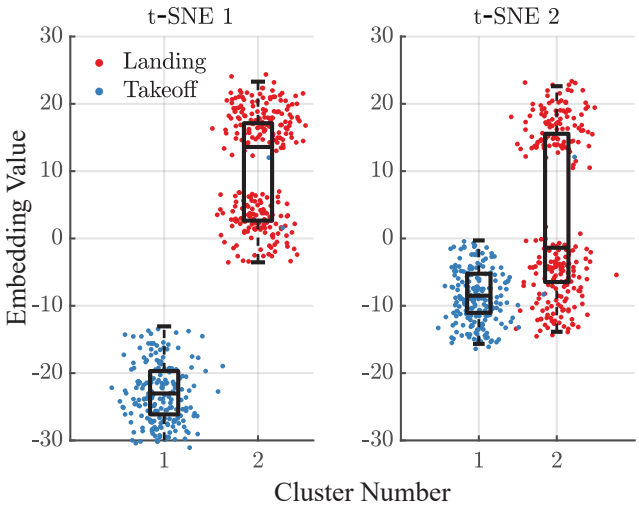

Figure 5. t-SNE Distributions by Operation cluster along with the data spread. The data points are color-coded based on the true operation label.

5. DISCUSSION AND CONCLUSION

In this work, we integrated dimensionality reduction (PCA, t-SNE), K-Means clustering, and machine learning (Random Forest, SVMs, multi-layer perceptrons) to investigate a set of 457 commercial turbofan noise measurements from both landing and takeoff operations.

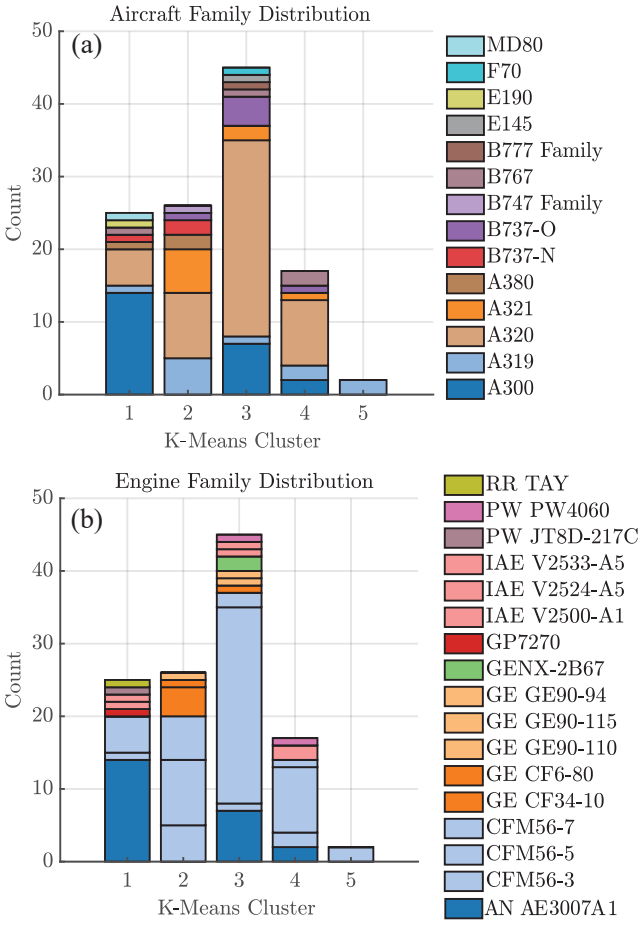

Figure 6. Histograms of (a) aircraft families and (b) engine families distributions per K-Means clusters, after applying PCA on landings at location 2.

Our analyses confirm that overall loudness—captured by amplitude-driven features such as peak pressure and OSPL—dominates the first principal component in all scenarios, overshadowing smaller but significant time-frequency variations. Nonetheless, higher-order PCA components and nonlinear embeddings expose modest but consistent differences in spectral peak distribution, wavelet-based entropy, and temporal shape parameters. These differences partly correlate with aircraft or engine families, although significant overlap remains, hindering subsequent classification problems.

Once the loudness and energy dimensions are factored out, time-domain shape (skewness, kurtosis, OSPL entropy) and spectral location (peak frequency) become important secondary and tertiary drivers of variance. They reflect how the noise changes over time (if the event is strongly peaked or more temporally spread out) and where the energy is centered in frequency. Clusters or patterns within takeoff vs. landing data can hinge on these PCs if certain aircraft or approach/departure profiles lead to distinct time-shape or spectral behaviors.

Operation classification is robust and achieves 98% accuracy, underscoring the value of amplitude-based metrics in distinguishing landing from takeoff. By contrast, family-level classification remains limited, typically below 50% accuracy, most probably due to shared engine technologies and standardized procedures among multiple aircraft. Consequently, advanced or domain-specific features, for instance finer harmonic analysis or amplitude-aware dimensionality reduction, may help to more clearly separate family-level differences.

In terms of location, when comparing landings based on the campaign location, both sets still have PC1 dominated by loudness, and then subsequent PCs revolving around OSPL shape (entropy, skewness, kurtosis) and spectral content (peak frequencies). The exact ordering (which PC is number 2 or number 3) can shift slightly, and the magnitudes of the loadings differ. This implies the relative importance of certain shape or wavelet features changes with array geometry, runway approach angles, or environmental factors (terrain, weather, etc.).

Moreover, our results highlight that removing or down-weighting amplitude-heavy features (e.g., maximum OSPL, Peak Pressure) can help reveal subtle time-frequency distinctions that might otherwise be masked. This approach, combined with a hierarchical classification strategy (identifying operations first, then specializing in the aircraft or engine separation), could show improved performance on a more extended and balanced dataset. Incorporating additional metadata—such as flight trajectory details, engine spool rates, or refined wavelet/harmonic indicators—could further boost class separability.

Overall, this study underscores loudness as the principal driver of acoustic variance while demonstrating that careful feature engineering and amplitude-aware dimensionality reduction are key to isolating more nuanced, design-based acoustic traits. Future work will investigate the adoption of genetic algorithms for hyperparameter tunning, as well as an extension of the current, quite small, and sparsely distributed dataset.

6. REFERENCES

[1] Human Environment and Transport Inspectorate, "Aircraft registrations," 2025. Accessed: 1-April-2025.

[2] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic, M. Wilhelm, and T. U. Kaiserslautern, "Bringing up OpenSky: A Large-scale ADS-B Sensor Network for Research," in *Proc. of the 13th international symposium on Information processing in sensor networks*, 2014.

[3] P. S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*. CRC Press, 1st ed., 2002.

[4] T. Tanaka, I. Nambu, Y. Maruyama, and Y. Wada, "Sliding-window normalization to improve the performance of machine-learning models for real-time motion prediction using electromyography," *Sensors*, vol. 22, no. 13, 2022.

[5] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987.

[6] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[7] L. McInnes, J. Healy, N. Saul, and L. Großberger, "Umap: Uniform manifold approximation and projection," *Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018.

[8] X. Jin and J. Han, *K-Means Clustering*, pp. 563–564. Boston, MA: Springer US, 2010.

[9] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining*, KDD'96, p. 226–231, AAAI Press, 1996.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, p. 273–297, Sept. 1995.

[11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[12] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[13] J. M. Klusowski and P. M. T. and, "Large scale prediction with decision trees," *Journal of the American Statistical Association*, vol. 119, no. 545, pp. 525–537, 2024.

[14] MathWorks, "predictorImportance - MATLAB Documentation," 2024. Accessed: 2025-04-07.