Audio Engineering Society

Convention Express Paper 369

Presented at the AES 159th Convention
2025 October 23–25, Long Beach, CA, USA

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the authors advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Kick Drum Microphone Taxonomy: Timbral Clustering of Controlled Kick-Out Recordings

Matthew Cheshire[1]

[1]*Birmingham City University*

Correspondence should be addressed to Matthew Cheshire (matthew.cheshire@bcu.ac.uk)

ABSTRACT

This study presents a kick drum microphone taxonomy of 46 distinct recording configurations from 32 different microphones, six of which featured varied voicing options. A range of spectral, temporal, and time–frequency features tailored to the transient and low-frequency nature of kick drums were extracted. These features were then standardised before dimensionality reduction via Principal Component Analysis (PCA). Microphones were then grouped using agglomerative hierarchical clustering with Ward's method, resulting in a dendrogram that visualises timbral similarity across both fixed- and switchable-voicing models. Cophenetic distance was used to rank microphone pairings by dissimilarity, revealing that voicing options, while measurable, typically result in subtler variations than changing between different models. A subset of microphones formed a maximally distinct cluster, suggesting strong timbral deviation from the rest of the dataset. The findings support practical use cases for microphone substitution, timbral matching, and educational demonstration. Rather than ranking microphones based on perceptual quality, this work provides a reproducible, analytical framework for understanding timbral relationships for kick drum recording.

1 Introduction

The kick drum remains a defining element of modern music production, providing both rhythmic grounding and genre-defining tonal weight across a wide spectrum of musical styles. Its centrality to the groove and energy of a mix is well documented, with both artistic and engineering communities emphasising its importance for perceived impact and musicality [1, 2]. Consequently, microphone selection has long been considered one of the most critical early-stage decisions available to engineers, enabling intentional timbral shaping even before the application of post-processing [3, 4].

Despite the abundance of microphone models marketed specifically for kick drum recording and the widespread anecdotal discourse surrounding their timbral signatures, there remains a lack of formal, data-driven research that characterises these microphones beyond perceptual preference tests. Pearce et al. [5] investigated perceptual differences more generally, using a range of musical sources recorded with studio and MEMS microphones, and found brightness, harshness, and clarity to be consistently prominent attributes. Prior work has explored subjective evaluations of snare drum

[6, 7] and vocal microphones [8], as well as studies on kick drum signal optimisation in live environments [9] and acoustical differences arising from changes to drum head configuration [10]. However, microphone shootouts typically focus on subjective listener ratings rather than providing a systematic, comparative framework for understanding microphone similarity from an analytical standpoint.

A key motivation for this work is not to rank microphones as "better" or "worse," but to compare their behaviour objectively and provide recording engineers with a relational framework for identifying timbral alternatives when preferred models are unavailable. In practical scenarios, access to specific microphones may be limited, yet engineers often aim to replicate familiar tonal profiles. By clustering microphones based on measurable audio features, this study highlights which models occupy similar regions of the feature space, offering guidance for informed substitution. Additionally, it examines how internal voicing options affect a microphone's position within that space, assessing whether such voicing yields perceptually and acoustically meaningful changes.

While previous online comparisons have informally evaluated subjective kick drum microphone differences [11, 12], the present study distinguishes itself by providing a reproducible methodology grounded in objective acoustic analysis. As such, this research contributes to the broader body of microphone evaluation literature by offering an analytical foundation upon which future perceptual or DSP-based studies can build.

This paper presents a taxonomy of kick drum microphones derived from spectral, temporal, and time–frequency characteristics, captured under tightly controlled recording conditions. Leveraging techniques from Music Information Retrieval (MIR) and audio signal processing, a comprehensive set of low-level features is extracted from each microphone's recording. These features are then standardised and subjected to Principal Component Analysis (PCA) to reduce dimensionality while preserving variance. The resulting feature space is used to compute similarity between microphones, which are visualised using hierarchical clustering and a dendrogram representations.



Fig. 1: Kick drum microphone position.

2 Methods

2.1 Recording

The dataset comprises recordings made with 32 different microphones, each capturing the same acoustic kick drum under tightly controlled conditions. All microphones were recorded sequentially at an identical position, directly in the centre of the port of the resonant head, shown in Figure 1. The capsule of each microphone was carefully angled to the same target point to maintain positional consistency across takes.

Recordings were conducted in a single session within the same acoustically treated studio environment, using the same performer, drum kit, tuning, beater, preamplifier (Focusrite ISA 828), and analogue-to-digital converter (RME Fireface 802). Gain was adjusted manually for each microphone to avoid clipping. All takes were captured at 44.1 kHz, 24-bit resolution. While a mechanical midi-controlled drum beater may have produce more consistent strikes, it was found that it lacked the force needed to excite the drum in a manner comparable to a human player, for this reason an experienced performer was chosen to deliver a real-world strike.

Each microphone was used to record four isolated kick drum strikes, performed at consistent levels. These were manually aligned in ProTools to account for minor timing variations introduced by the performer. All samples were normalised to -23 LUFS [13] prior to

feature extraction to enable meaningful comparisons across recordings. LUFS-based normalisation was chosen over peak normalisation to ensure perceptual loudness consistency across microphone recordings. This minimises feature distortion introduced by loudness differences and enables more meaningful comparisons of spectral and temporal content. No EQ, compression, or other signal processing was applied. While the broader dataset included multi-track recordings and 4-bar phrases, this study used only four isolated kick-out recordings per microphone. the term "*Kick-out*" refers to a microphone placed outside of the kick drum shell, most commonly on the resonant head. Although the microphone can be placed on the batter head for certain application, in this context, it exclusively refers to a microphone used on the resonant head. A microphone place directly inside the kick drum shell is referred to as "*Kick-in*", and will typically be position via the sound hole, if the resonant head is ported, or via an internal mounting system.

Table 1 shows the full list of microphones used. Microphones that had multiple polar patterns (i.e., AKG C414, Neumann U87, and Lauten FC-357) were all set to cardioid. Additionally, several microphones included various voicing options resulting in several recordings of these microphones being captured with different settings. While some microphones (i.e., AKG C414, Electro-voice RE20, Lauten FC-357, Neumann U47 FET, Neumann U87, Sennheiser MD421, Sennheinser MD441, and Shure SM7B) feature built-in high pass filters (HPF) these options were never used. Only frequency contouring/emphasis options were utilised. The list of the microphones that included these features is shown in Table 2. Additionally the Audio-Technica ATM250DE is a dual-capsule microphone, featuring both condenser and dynamic elements, these two channels were treated as separate recordings and analysed independently of one another. The addition of these extra recordings from the same microphones resulted in a total of 46 recordings being used for the taxonomy.

2.2 Feature Extraction

In order to construct a meaningful taxonomy of kick drum microphones, this study relies on a selection of low-level audio features that have demonstrated perceptual and acoustical relevance in capturing timbral characteristics. The objective extends beyond basic spectral description, aiming instead to model multidimensional characteristics that differentiate microphone responses to transient-rich, low-frequency audio sources.

The selection of features was informed by prior research in instrument identification, drum sound classification, perceptual timbre modelling, and statistical analyses of microphone recordings. Peeters' comprehensive taxonomy of low-level descriptors [14] served as a fundamental reference, particularly informing the choice of spectral descriptors including spectral centroid, spread, rolloff, flatness, kurtosis, skewness, and slope. These features have consistently demonstrated utility in applications such as musical genre classification [15], drum classification tasks [16], and perceptual audio content indexing [17].

Spectral descriptors were selected to capture spectral shape and harmonic structure. Features such as spectral centroid, spread, rolloff, and kurtosis are closely associated with perceptual qualities like brightness, noisiness, and harmonic richness [18, 17]. MFCCs provide a compact yet comprehensive representation of spectral shape, widely validated for timbre-based recognition and similarity tasks across musical instrument recognition and audio indexing applications [14, 17, 16, 19]. While the optimal number of MFCCs can vary depending on context, prior work [14, 16] retains the first 13 coefficients as they robustly characterise spectral envelope features of musical instruments and percussion sounds. This work employs the first 13 MFCCs, with each coefficient averaged across all audio frames, providing a stable, time-invariant representation that mitigates sensitivity to micro-timing fluctuations and supports comparison across varied recordings.

Sub-band RMS energy was calculated by dividing each kick drum recording into ten log-spaced frequency bands (from 0–50 Hz to 12.8–22 kHz) and computing the RMS within each sub-band, following the methodology proposed by Hove et al. [20]. This facilitates meaningful comparative analysis of spectral balance, revealing microphone-specific emphasis or attenuation within frequency bands. Spectral contrast, which quantifies the difference between spectral peaks and valleys within sub-bands, was also included as a descriptor of timbral texture. Initially proposed by [21] and further applied in genre classification by [22], spectral contrast has demonstrated considerable discriminative power in capturing harmonic richness and broadband content variations across both musical and percussive signals [19]. Given the transient and low-frequency characteristics of kick drum recordings, spectral contrast serves

Make	Model	Polar Pattern	Type
AKG	C414	Multi-pattern	Condenser
AKG	D112	Cardioid	Dynamic
AKG	D12 VR	Cardioid	Dynamic
AKG	P2	Cardioid	Dynamic
Audio-Technica	ATM250DE	Hypercardioid	Both
Audix	D6	Cardioid	Dynamic
Beyerdynamic	M160	Hypercardioid	Ribbon
Beyerdynamic	M201	Hypercardioid	Dynamic
Beyerdynamic	M88 TG	Hypercardioid	Dynamic
Beyerdynamic	M99	Hypercardioid	Dynamic
Beyerdynamic	Opus 99	Cardioid	Dynamic
DPA	4055	Cardioid	Condenser
DPA	4099	Supercardioid	Condenser
Earthworks	M30	Omnidirectional	Condenser
Electro-Voice	PL33	Supercardioid	Dynamic
Electro-Voice	RE20	Cardioid	Dynamic
Lauten	FC-357	Multi-pattern	Condenser
Neumann	U47 FET	Cardioid	Condenser
Neumann	U87	Multi-pattern	Condenser
Samson	Q-Kick	Supercardioid	Dynamic
sE Electronics	V Kick	Supercardioid	Dynamic
Sennheiser	e602	Cardioid	Dynamic
Sennheiser	e902	Cardioid	Dynamic
Sennheiser	MD421	Cardioid	Dynamic
Sennheiser	MD431	Supercardioid	Dynamic
Sennheiser	MD441	Supercardioid	Dynamic
Shure	Beta 52A	Supercardioid	Dynamic
Shure	Beta 57A	Supercardioid	Dynamic
Shure	SM57	Cardioid	Dynamic
Shure	SM58	Cardioid	Dynamic
Shure	SM7B	Cardioid	Dynamic
Telefunken	M82	Cardioid	Dynamic

Table 1: Specifications of selected microphones by Make, Model, Polar Pattern, and Type.

Microphone	Voicing Options	Number and Naming Used
AKG D12 VR	3 phantom-powered EQ presets	4 (Passive, SW1, SW2, SW3)
Beyerdynamic M99	EQ: Linear / Contour B / Contour C	3 (Lin, B, C)
sE V-Kick	Voicing (Classic/Modern) + HF (Classic/Modern)	4 (CC, MC, CM, MM)
Sennheiser MD441	Treble boost	2 (Flat, Treble)
Shure SM7B	Presence Boost	2 (Flat, Presence)
Telefunken M82	Kick EQ On/Off + High Boost On/Off	4 (OffOff, OnOff, OffOn, OnOn)

Table 2: Microphones voicing options, their functional switching modes, and the number of settings used.

as a perceptually relevant metric for microphone-driven shaping of spectral energy distributions.

Temporal features were chosen to capture the transient-dominant nature of kick drum recordings. Features such as log-attack time, temporal centroid, crest factor, and zero-crossing rate were included due to their proven efficacy in characterising percussive sounds and their time-domain envelopes [14, 16, 17]. A study on snare drums demonstrated how descriptors such as spectral centroid, rolloff, and decay-related metrics vary with microphone choice and playing dynamics [26]. Temporal centroid further quantifies the distribution of energy within signal envelopes, providing insight into how each microphone captures energy distribution across transient and sustain phases. Additionally, time–frequency descriptors such as spectral flux and onset strength were incorporated to quantify spectral evolution and rhythmic articulation. These descriptors have previously been validated in contexts ranging from emotion recognition tasks [24] to timbral analysis of audio for cochlear implants [18], and environmental sound classification [19]. Features were extracted individually for each of the four kick drum strikes from each microphone and the mean was taken.

Overall, the presented feature set integrates both theoretical relevance and practical interpretability. While recent literature has explored advanced learned descriptors such as scattering coefficients [27], the present study consciously prioritises established, interpretable features. This approach supports transparent clustering, facilitates intuitive interpretation by recording engineers, and links microphone choice to perceptually and technically meaningful distinctions in behaviour. The full list of features used can be seen in Table 3. All features were standardised across recordings using Z-score normalisation (mean-centred and scaled to unit variance) to enable meaningful cross-microphone comparison. This step ensured compatibility with downstream dimensionality reduction and clustering algorithms by preventing scale-related feature bias.

To ensure stable clustering and reduce the influence of correlated or low-variance features, PCA was applied prior to hierarchical clustering. This dimensionality reduction technique transforms the original feature space into a set of orthogonal components ranked by explained variance, thereby mitigating issues of multicollinearity and feature redundancy. The first 14 components were retained, capturing over 95% of the total variance in the dataset. These components formed the basis for computing distances between microphones, which were then used to construct the hierarchical dendrogram. Feature extraction and preprocessing were implemented in Python using librosa, scipy, numpy, and pyloudnorm. Dimensionality reduction and clustering were performed using scikit-learn and scipy.

2.3 Clustering

Following dimensionality reduction, the PCA-transformed feature space was used to compute pairwise similarity between microphones, quantifying how closely related each microphone was in terms of its multivariate timbral features. Hierarchical agglomerative clustering was employed to uncover structure within the dataset without requiring a predefined number of clusters. Specifically, Ward's minimum variance method [28] was used, whereby each microphone began as its own cluster, and iteratively merges the pair of clusters that results in the smallest increase in total within-cluster variance.

The resulting hierarchical structure was visualised as a dendrogram, providing an interpretable taxonomy of microphone similarity. This representation captures not only the relative proximity of microphones in feature space, but also their nested relationships, enabling analysis of both fine- and coarse-groupings.

3 Results

The clustering procedure described in Section 2.3 yielded a hierarchical dendrogram (Figure 2) that reflects microphone similarity within the PCA-reduced feature space. Each microphone is represented as a terminal leaf, and branches indicate the dissimilarity between clusters based on Ward's minimum variance criterion. The vertical axis represents the linkage distance at which clusters are merged, enabling examination of microphone relationships at multiple levels of similarity. No fixed number of clusters was imposed, allowing for flexible interpretation of the taxonomy. To aid visual clarity, the dendrogram uses colour to distinguish the major branches of the hierarchy (orange, green, red). These colours indicate primary sub-trees based on linkage distance but do not reflect predefined categories or manually assigned groupings.

Table 4 presents a ranked summary of pairwise microphone comparisons based on cophenetic distance,

Spectral Features		Temporal Features	
Harmonic Spectral Deviation	[14, 19]	Crest factor	[19, 23]
Harmonic-to-noise ratio	[14, 19]	Log-attack time	[14, 16, 18]
MFCCs (13)	[14, 15, 16, 19]	Temporal centroid	[14, 16, 18, 24]
Spectral centroid	[14, 16, 18, 19]	Temporal decrease	[14]
Spectral contrast	[21, 22]	Zero-crossing rate	[14, 18, 19]
Spectral entropy	[19, 25]		
Spectral flatness	[14, 19]		
Spectral kurtosis	[14]	**Time–Frequency Features**	
Spectral rolloff (85%)	[14, 19]	Log energy envelope	[19, 24]
Spectral skewness	[14]	Onset strength	[15, 24]
Spectral slope	[14, 19]	Spectral flux	[14, 18, 19]
Spectral spread	[14, 18, 19]		
Sub-band RMS energy	[20]		

Table 3: Audio features proposed for kick drum similarity clustering, and corresponding literature.

a metric derived from the hierarchical clustering process. For any two microphones, the cophenetic distance reflects the height in the dendrogram at which their respective branches first merge, representing the point at which they are grouped into the same cluster. This measure captures how dissimilar two microphones must be before being considered part of a common group under Ward's criterion. Unlike raw Euclidean distance in the PCA space, cophenetic distance incorporates the full structure of the dendrogram, making it a useful index for evaluating the relative proximity of microphones in the context of the overall clustering. Table 4 highlights the most similar (lowest cophenetic distance) microphone pairings, as well as the most similar distinct model pairs and most dissimilar intra-model pairs. It does not include the most dissimilar distinct model pairs, as it was found that several microphones all produced to the same maximum cophenetic distance.

In total, 2,070 pairwise comparisons were evaluated (excluding self-comparisons), of these, 408 pairs exhibited the same maximum cophenetic distance of 28.860. This value represents the greatest possible dissimilarity between microphone pairs. The 12 microphones that make up the orange branch of the dendrogram in Figure 2 each had the same maximum cophenetic distance when measured against the other 34 recordings of the data-set. These 12 microphones were never maximally dissimilar to each other, and instead formed a cohesive cluster within the orange branch of the dendrogram, which were all equally dissimilar to the microphones of the green and red clusters.

4 Discussion

The hierarchical dendrogram shown in Figure 2 offers a high-level visualisation of microphone similarity in the PCA-reduced feature space. Three primary clusters emerge (orange, green, and red) each representing a group of microphones with greater internal similarity and broader dissimilarity from those in other clusters. The orange branch predominantly contains a range of dynamic microphones and rather surprisingly one large diaphragm condenser. There are also other examples where the closest similarity of microphones pairs are instances of a condenser paired with a dynamic microphone, such as in the case of the DPA 4099 and Beyerdynamic M201, as well as the Lauten FC357 and EV PL33. Unsurprisingly, it is quite apparent that microphone with switched voicing options such as the sE VKick and Telefunken M82 are closely similar to themselves. The dendrogram reflects both close intra-family relationships and broader structural separations.

Interestingly, although the 12 microphones in the orange branch of Figure 2 were maximally dissimilar from all other 34 microphones, none of them were maximally dissimilar to one another. This indicates that while they are individually distinct from the rest of the dataset, they share sufficient acoustic similarity to form a coherent higher-level cluster. This reinforces the interpretation that the orange branch represents a group of microphones that is internally cohesive yet collectively distinct from the rest of the dataset

Table 4 revealed that the top five most similar microphone pairs were instances of the same model with

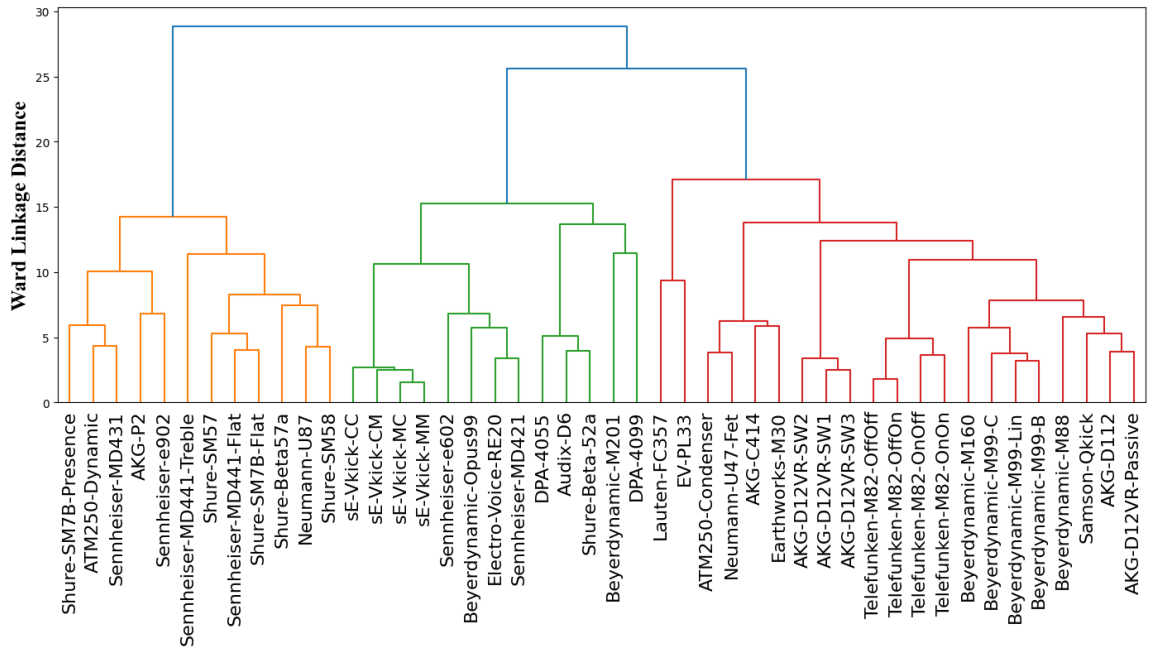


Fig. 2: Hierarchical clustering dendrogram showing pairwise similarity between kick drum microphones in PCA-reduced feature space. Microphones are joined based on Ward's method.

different voicing options. In fact, of the top 25 most similar microphone, only seven involved distinct models, while the remaining 18 pairs consisted of voicing variants of the same mic. The top five most similar distinct-model pairings are reported in Table 4, showing pairs that one might associate as being similar, such as the Audix D6 and Shure Beta52a, both marketed as dedicated kick drum microphones. These findings suggests that although voicing switches introduce measurable variations, the resulting changes are generally smaller in magnitude than the differences observed between entirely different microphone designs. In other words, switching voicing modes within these particular microphones may influence tone and frequency balance, but not to the extent that it fully mimics or replaces the characteristics of changing different models.

To the contrary of earlier observation, a small number of multi-voiced microphones demonstrated substantial variation across their different modes. As shown in Table 4. The SM7B Presence boost produced the highest observed cophenetic distance among any intra-model pairs compared to its Flat setting. Similarly, the passive mode of the AKG D12VR exhibited notable dissimilarity when compared with any of its three active EQ settings (SW1, SW2, SW3), and the Treble boost setting of the MD441 also caused a pronounced shift relative to its Flat configuration. By contrast, the voicing options offered by the sE V Kick, Telefunken M82, and Beyerdynamic M99 showed far smaller cophenetic distances between modes, suggesting these switches result in subtler spectral or dynamic changes. These findings indicate that the perceptual and acoustic effect of voicing options is highly microphone-dependent, with some designs exhibiting greater difference across voicings, and others maintaining closer timbral character.

While the recordings in this study were tightly controlled to isolate microphone-related differences, several factors inevitably influence drum timbre beyond the choice of transducer. While inconsistency of strikes were managed through an experienced performer with emphasis on consistent strike velocity as well as LUFS-based normalisation, subtle variations could still be introduced. Microphone placement is also known to affect the balance between attack and resonance, and alternative positions would likely alter the relative proximity of models within the taxonomy. Similarly, differences in shell material and size as well as changes

Mic 1	Mic 2	Cophenetic Distance
Most Similar Pairs		
sE-Vkick-MC	sE-Vkick-MM	1.52
Telefunken-M82-OffOff	Telefunken-M82-OffOn	1.83
sE-Vkick-CM	sE-Vkick-MM	2.49
sE-Vkick-MC	sE-Vkick-CM	2.49
AKG-D12VR-SW1	AKG-D12VR-SW3	2.51
Most Similar Distinct Model Pairs		
Electro-Voice-RE20	Sennheiser-MD421	3.39
ATM250-Condenser	Neumann-U47-Fet	3.83
AKG-D112	AKG-D12VR-Passive	3.91
Audix-D6	Shure-Beta-52a	3.97
Sennheiser-MD441-Flat	Shure-SM7B-Flat	4.00
Most Dissimilar Intra-Model Pairs		
Shure-SM7B-Flat	Shure-SM7B-Presence	14.23
AKG-D12VR-Passive	AKG-D12VR-SW3	12.39
AKG-D12VR-Passive	AKG-D12VR-SW2	12.39
AKG-D12VR-Passive	AKG-D12VR-SW1	12.39
Sennheiser-MD441-Flat	Sennheiser-MD441-Treble	11.42

Table 4: Microphone similarity and dissimilarity based on cophenetic distance. The table includes the top five: most similar pairs, most similar distinct model pairs, and the most dissimilar intra-model pairs. Lower cophenetic distances indicate greater similarity in the PCA-reduced feature space.

in head tension or internal damping would modify the source spectrum and decay profile, potentially shifting cluster boundaries. These considerations highlight that the present taxonomy should be understood as valid for the specific controlled conditions employed, with future work extending the approach across varied placements, tunings, and performance contexts.

5 Summary

This study presented a taxonomy of 46 recordings from 32 microphones based on carefully selected spectral, temporal, and time–frequency descriptors chosen for their relevance to the transient-rich, low-frequency characteristics of kick drum recording. Using dimensionality reduction and hierarchical clustering, microphones were grouped according to timbral similarity, providing a structured visualisation of relationships across both fixed- and switchable-voicing kick-out microphone recordings. The analysis revealed that while internal voicing options produce measurable changes, these are often subtler than the differences observed between distinct microphone models. A subset of microphones exhibited high levels of dissimilarity from the rest of the dataset, forming a cohesive cluster that remained globally distinct within the feature space.

The resulting taxonomy has several practical applications. It can serve as a reference tool for recording engineers seeking timbrally similar substitutes when preferred microphones are unavailable. It may also support educators in illustrating how different microphones cluster acoustically despite differences in manufacturer, type, or marketing, and can aid in demonstrating timbral contrast in studio and live settings. As all recordings were made under tightly controlled studio conditions, the findings are not intended to generalise across all kick drum sizes, tunings, styles, mic placements, or room environments. By making the dataset openly available, the aim is to encourage creative and novel applications in areas such as intelligent mic selection, spectral modelling, or virtual drum synthesis.

6 Distribution

The kick drum dataset can be downloaded from: `https://doi.org/10.5281/zenodo.17137497`
The audio files are released under a *Creative Commons Attribution–NonCommercial 4.0 International License* CC BY NC.

References

[1] Owsinski, B. and Moody, D., *The Drum Recording Handbook*, Hal Leonard, 2016.

[2] Toulson, R., *Drum Sound and Drum Tuning: Bridging Science and Creativity*, Audio Engineering Society Presents, Taylor & Francis, 2021.

[3] Corbett, I., *Mic It!: Microphones, Microphone Techniques, and Their Impact on the Final Mix*, Focal Press, New York, NY, 2nd edition, 2021.

[4] Brennan, B., "What You Need to Know About…Drum Mics," *https://www.moderndrummer.com/2015/01/need-know-drum-mics/*, 2015, accessed: May 11, 2025.

[5] Pearce, A., Brookes, T., Dewhirst, M., and Mason, R., "Eliciting the most prominent perceived differences between microphones," *The Journal of the Acoustical Society of America*, 139(5), pp. 2970–2981, 2016.

[6] Cheshire, M., Hockman, J., and Stables, R., "Microphone Comparison for Snare Drum Recording," *journal of the audio engineering society*, 2018.

[7] Cheshire, M., Stables, R., and Hockman, J., "Microphone Comparison: Spectral Feature Mapping for Snare Drum Recording," *journal of the audio engineering society*, 2019.

[8] Cheshire, M. and Jones O'shannessy, E., "Microphone Comparison for Female R&B Vocal Recording," *journal of the audio engineering society*, 2023.

[9] Hill, A. J., Hawksford, O. J. M., Rosenthal, A. P., and Gand, G., "kick-drum signal acquisition, isolation, and reinforcement optimization in live sound," *journal of the audio engineering society*, 2011.

[10] Herman, G. and Jasinski, C., "measurement of the particle velocity in front of kick drums in three states," *journal of the audio engineering society*, 2021.

[11] Fuston, L., "Kick Drum Mic Shootout – with Sound Samples," *https://www.sweetwater.com/insync/kick-drum-mic-shootout-with-sound-samples/*, 2024, accessed: June 24, 2025.

[12] Rodgers, J., "Kick Drum Shootout - The Results But Do They Really Matter?" *https://www.production-expert.com/production-expert-1/kick-drum-shootout-the-results-but-do-they-really-matter*, 2024, accessed: June 22, 2025.

[13] "Loudness Normalisation and Permitted Maximum Level of Audio Signals," Technical Report EBU R 128-2023, European Broadcasting Union, Geneva, 2023.

[14] Peeters, G., "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project," Technical report, IRCAM, 2004, technical Report.

[15] Tzanetakis, G. and Cook, P., "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, 10(5), pp. 293–302, 2002.

[16] Herrera, P., Yeterian, A., and Gouyon, F., "Automatic Classification of Drum Sounds: A Comparison of Feature Selection Methods and Classification Techniques," in *Lecture Notes in Computer Science*, 2002.

[17] Kim, H., Moreau, N., and Sikora, T., *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, Wiley, 2006.

[18] Kong, Y.-Y., Mullangi, A., Marozeau, J., and Epstein, M., "Temporal and Spectral Cues for Musical Timbre Perception in Electric Hearing," *Journal of Speech, Language, and Hearing Research*, 54(3), pp. 981–994, 2011.

[19] Alías, F., Socoró, J. C., and Sevillano, X., "A Review of Physical and Perceptual Feature Extraction Techniques for Speech, Music and Environmental Sounds," *Applied Sciences*, 6(5), p. 143, 2016.

[20] Hove, M. H., Athanasopoulos, G., VanHandel, L. M., and Todd, P. M., "Increased Levels of Bass in Popular Music Recordings 1955–2016," *Empirical Musicology Review*, 14(3-4), pp. 112–124, 2019.

[21] Jiang, D.-N., Lu, L., Zhang, H.-J., Tao, J.-H., and Cai, L.-H., "Music Type Classification by Spectral Contrast Feature," in *Proceedings of the*

IEEE International Conference on Multimedia and Expo (ICME), pp. 113–116, IEEE, Lausanne, Switzerland, 2002.

[22] West, K. and Cox, S., "Finding an Optimal Segmentation for Audio Genre Classification," in *ISMIR*, 2005.

[23] Wendl, M. and Lee, H., "The Effect of Dynamic Range Compression on Loudness and Quality Perception in Relation to Crest Factor," in *Proceedings of the 136th Audio Engineering Society Convention*, Audio Engineering Society, Berlin, Germany, 2014.

[24] Grekow, J., "Audio Features Dedicated to the Detection and Tracking of Arousal and Valence in Musical Compositions," *Journal of Information and Telecommunication*, 2(3), pp. 322–333, 2018.

[25] Misra, H., Ikbal, S., Bourlard, H., and Hermansky, H., "Spectral entropy based feature for robust ASR," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pp. I–193, 2004.

[26] Cheshire, M., Stables, R., and Hockman, J., "Investigating Timbral Differences of Varied Velocity Snare Drum Strikes," in *Audio Engineering Society Convention 148*, Online, 2020.

[27] Lostanlen, V., Hennequin, R., Lagrange, M., Andén, J., Andén, M., and Mallat, S., "Time–Frequency Scattering Accurately Models Auditory Similarities Between Instrumental Playing Techniques," 2020.

[28] Ward, J. H., "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 58(301), pp. 236–244, 1963.