forum acusticum 2026
8-12 September · Graz, Austria

AUTOMATIC TIMBRE CHARACTERIZATION OF PLAYING TECHNIQUES IN NON-WESTERN PERCUSSION INSTRUMENTS: A CASE STUDY ON THE COLOMBIAN ALEGRE DRUM

Cristhiam Fidel Martínez Orellanos[1]

[1] *Institute for Systematic Musicology, University of Hamburg, Neue Rabenstr. 13, 20354 Hamburg*

Abstract

The underrepresentation of non-Western musical traditions in open-source datasets poses a persistent challenge for computational psychoacoustic analysis, particularly due to insufficient sample sizes, often the case in these collections. This study examines the automatic timbre characterization of non-Western percussion instruments under conditions of small sample size employing the Colombian Alegre drum as a case study. Three playing techniques (Abierto, Bajo, and Quemado) were analyzed. Recordings were obtained in an anechoic chamber using an artificial head system. Frequency spectra as well as temporal and spectral descriptors relevant to the perception of percussive sounds were extracted. To address the limited dataset size (N = 17), bootstrapping and Hedges' g were applied to generate 95% confidence intervals and assess effect sizes, respectively. Descriptors with the smallest differences in effect size among techniques were disregarded to maximize technique differentiation. The resulting perceptual profiles per technique are summarized in radar charts, displaying a multidimensional timbre assessment. These findings suggest that reliable computational timbral characterization is achievable with small datasets, supporting automatic data labeling pipelines for the digital preservation of non-Western musical heritage.

Keywords: *non-Western percussion, data science, musical heritage, computational musicology*

1. Introduction

The increasing use of generative artificial intelligence has brought to the forefront the need for annotated datasets of musical instruments and recordings. However, many open-source datasets such as [1] and [2] include only Western instruments and genres, which may cause models to underperform outside this musical tradition.

Developing more datasets is the obvious solution, yet high-quality recordings require conditions and equipment that are not always available, often resulting in compromised or small datasets. This study addresses the latter case by developing a data pipeline for the timbre characterization of percussive musical instruments under conditions of small sample size, aiming to characterize different playing techniques and evaluate the effectiveness of the chosen descriptors in discriminating among them. In Colombian traditional music, limitations related to the lack of datasets have been documented in several ML-based classification initiatives ([3]–[5]), therefore, the Colombian Alegre drum is analyzed as a proof of concept. This percussion instrument is a single-headed conical membranophone with a goatskin membrane, central to several rhythms on the Caribbean coast of Colombia [6]. Three playing techniques are characterized: Abierto, a mid pitched open tone produced by striking the membrane edge with a flat hand; Bajo, a deep bass tone obtained when hitting the membrane off-center with the palm; and Quemado, a short sound produced by striking between center and edge with a cupped hand. The following sections describe the procedure applied, present the results and discuss their implications regarding automatic timbre characterization.

2. Method

To computationally describe the sound characteristics of the three techniques previously portrayed, a dataset composed by timbral descriptors extracted from recordings performed under controlled conditions was evaluated. For this purpose, a three-step procedure was followed: first, the spectral contours

Corresponding author: *crismartinez@t-online.de.*

of each technique were compared; second, five temporal and spectral descriptors were extracted from each sample; and third, bootstrapping and Hedges' g were applied to increase the statistical reliability of the results. Finally, the resulting timbre profiles were visualized through radar charts. Each step is described below.

2.1. Spectral contour analysis

Samples from each technique were recorded in the anechoic room of the musical acoustics laboratory at the Institute for Systematic Musicology at the University of Hamburg using an artificial head HSU III.2 and the acoustic measurements system SQuadriga III. A Colombian percussionist was commissioned to perform single shots from each technique. From these recordings, a dataset with $N = 17$ sounds (Quemado: 6, Bajo: 6, Abierto: 5) was extracted. Over these samples two analyses were carried out. First, the spectral contours from the first 200 ms of each sample were analyzed via Short-Time Fourier Transformation (STFT) with a window length of 42.6 ms and a hop length of 10,6 ms at a sample rate of 48 kHz. The beginning of the analyzed time window was set at the highest amplitude peak in the time domain. All amplitude coefficients were normalized. Second, a qualitative analysis between strokes was performed by comparing the spectra from chosen examples of each technique across different frequency ranges along the spectrum.

Subsequently, five temporal and spectral descriptors were extracted from each sample: spectral centroid [7], spectral bandwidth [8], spectral flux [9], roughness [10] and crest factor, which quantifies the waveform's shape as a ratio of peak to average energy calculated as the root mean squared [11]. Each of these descriptors corresponds to at least one perceptual dimension relevant to the aural perception of percussive timbre [12]: centroid is related to brightness; flux to the dry/dead dimension, bandwidth and roughness to the noisiness or purity of the sound; and the crest factor is related to perceived sharpness.

2.2. Bootstrapping

The variability inherent to musical performance produces high within-technique variance which, combined with the limited sample size, reduces the precision of descriptor estimates. Bootstrapping [13] was therefore applied to generate 95% confidence intervals (CI) without assuming normality, enabling comparison of descriptor distributions across techniques.

The core idea is that one can estimate the uncertainty around a statistical measure, in this case, the mean. To do so, each sample from a dataset with J observations $\{x_1, x_2, \ldots, x_J\}$ is treated as a proxy for the true population. Then, B bootstrap samples are drawn with replacement from the original dataset, and the mean $\bar{x}$ is calculated for each resampled batch. The resulting distribution $\{\bar{x}_1, \ldots, \bar{x}_B\}$ approaches the sampling distribution of $\bar{x}$, from which the 2.5th and 97.5th percentiles are taken as the 95% CI bounds. In the present study, a value of $B = 2000$ as well as a fixed random seed were employed.

2.3. Standardized effect size - Hedges' g

Given the small and unequal sample size for each technique, comparing descriptor means alone does not constitute a reliable timbre characterization, because it does not account for within-technique variability or the uncertainty of the estimates. Consequently, the standardized effect size magnitude was computed to assess the differences across techniques independently of descriptor scale [14]. Hedges' g [15] was employed, as it applies a bias correction that reduces the overestimation of effect sizes characteristic of small samples, and correctly handles unequal group sizes. It is defined as

$$g = \frac{\bar{x}_{ai} - \bar{x}_{bi}}{s_p} J(\nu), \tag{1}$$

where $\bar{x}_a$ and $\bar{x}_b$ are the means for the i descriptor of techniques a and b respectively, s_p is the pooled standard deviation

$$s_p = \sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{n_a + n_b - 2}} \tag{2}$$

with n_a and n_b the sample sizes, s_a^2 and s_b^2 the unbiased sample variances of each group, and $J(\nu)$ the small-sample bias correction factor [16]

$$J(\nu) \approx 1 - \frac{3}{4\nu - 1}, \tag{3}$$

where $\nu = n_a + n_b - 2$ are the degrees of freedom. As ν increases, $J(\nu) \to 1$ and the correction becomes negligible. For small samples, $J(\nu) < 1$, shrinking the effect size estimate downward to correct for the tendency of small samples to overestimate the true population effect size.

The Hedges' g values for each descriptor and technique pair were visualized in a heatmap to facilitate comparison of effect size magnitudes across descriptors. Descriptors showing consistently small effect size differences across all technique pairs were considered insufficiently discriminative and discarded. The remaining descriptors were normalized to a $[0, 1]$ scale using global min–max across all hits and techniques, and their means per technique were plotted as radar charts to visualize the timbral profile of each technique. Results are presented in the following section.

3. Results

3.1. Spectral contour analysis

Fig. 1 shows the spectral contours from each playing technique. The main plot depicts the frequency spectrum only until 10 kHz for the sake of readability —frequency bins up to 15 kHz were included in the analysis—, and the inset illustrates in detail the spectra within the first 1 kHz. A relation between the technique's execution, its corresponding spectra and the obtained descriptor values is observed. The

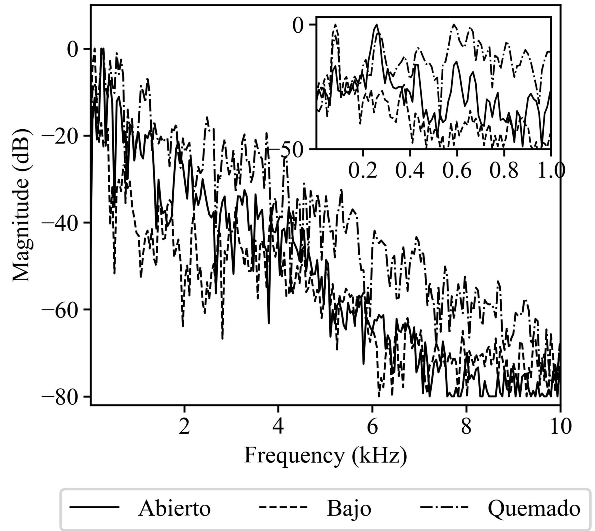

Figure 1: Spectra of the Abierto, Bajo, and Quemado techniques. STFT from the first 200 ms of each technique, sample rate = 48 kHz, 42.6 ms window and hop length 10,6 ms. $t = 0$ was set at the peak amplitude in the time-domain.The main plot displays the spectrum up to 10 kHz and the inset details the sub-1 kHz region. The palm-to-center strike of the Bajo technique excites the lowest mode of the system ($f_0 = 82$ Hz), accounting for the lowest centroid mean. The Abierto's highest peak, found at at approximately 240 Hz, is about three times f_0, followed by two resonant frequencies under 1kHz well above -10dB, which may explain the observed spectral flux values. Lastly, high-frequency density and sustained partials above -60 dB (notably 5–7 kHz) are observed for the Quemado technique; added to its short duration, illustrates its impulse-like sound.

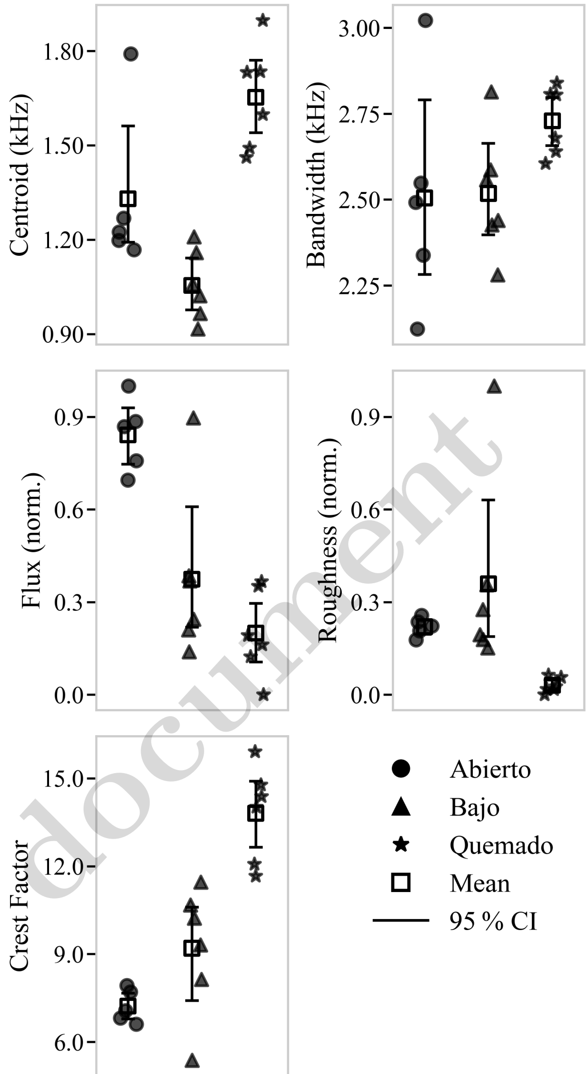

Figure 2: Bootstrap 95 % Confidence Intervals (CIs) per Timbre Descriptor and Playing Technique. Although narrow CIs are observed in some descriptors for each technique, the natural variability of musical performances leads to large CIs in others; thus, some features are not consistently produced by some techniques, e.g., the Abierto's bandwidth (top row, right) or the Bajo's roughness (middle row, right). In this regard, bootstrapping helps to decrease the bias produced by such variability, while still presenting honest estimates of the reliability of the mean.

Bajo technique injects most of its energy to the lowest mode of the system by hitting the membrane close to the center, which causes a considerable amplitude difference with regard to higher modes, lowering the centroid. The Abierto distributes its energy between higher modes: the frequency with the highest amplitude (f=240Hz) exhibits a ratio of approximately 3:1 with respect to f_0, which is approximately 10dB quieter; moreover, higher modes in the sub 1kHz range are well above -15dB during the first 200 ms. This behavior increases the perceived complexity/liveliness of the sound, as measured by the spectral flux. Lastly, the impulse-like behavior of the Quemado is explained by high energy in the middle and high frequency ranges, starting at approximately 2.5 kHz and the presence of partials between 5 and 7 kHz well above -60 dB during the analyzed time window, all of this summarized by the highest centroid and bandwidth.

3.2. Bootstrapping

The results are shown in Fig. 2. The per descriptor mean values show clear differences between techniques. The Bajo has the lowest centroid and highest roughness, while the Quemado shows the inverse tendency, namely highest centroid and lowest roughness. The crest factor mirrors the relation between the technique's execution and the corresponding sound decay, e.g. a very high coefficient represents the impulse-like behavior of the Quemado as opposed to the low coefficient of the Bajo. Nonetheless, large CIs are observed in four out of five descriptors (centroid, bandwidth, flux and roughness). These are caused by samples whose execution somehow differs from the rest of the data corresponding to the same technique; an intrinsic characteristic of musical data, specially in non-

Western traditional music where the use of samples and error-correcting techniques in the mixing stage is reduced. To reduce the impact of such variability in the technique classification task, the difference in the magnitude of the effect sizes is calculated.

3.3. Standardized effect size - Hedges' g

The Hedges' g analysis is shown in Fig. 3. Negative values result from the mean of the first technique being lower than that of the second technique for the evaluated descriptor. The largest effect size between Abierto and Quemado is observed in roughness; between Abierto and Bajo in flux; and between Bajo and Quemado in the centroid. Spectral bandwidth shows the lowest differences in effect size magnitude across all technique pairs, suggesting that this feature is not as relevant for distinguishing between techniques; therefore, it was disregarded in the timbre characterization of the playing techniques. The remaining descriptors were used to generate timbre profiles plotted as radar charts for each technique as a tool to visually inspect how different the resulting descriptions are (see Fig. 4).

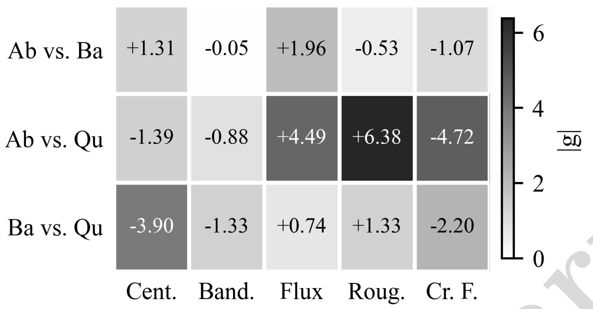

Figure 3: Hedges' g effect sizes per technique pair and descriptor. (Ab=Abierto, Ba=Bajo, Qu=Quemado) Negative values indicate the first technique has a lower mean than the second. Roughness, flux, and centroid show the largest effects for Abierto–Quemado, Abierto–Bajo, and Bajo–Quemado respectively. Bandwidth shows the lowest effect size magnitude differences across all pairs and is therefore disregarded in the timbral characterization.

4. Discussion

The results show a clear relation between the frequency spectra and the obtained profiles. Yet, large confidence intervals resulting from bootstrapping point towards a high variability in the data for some descriptors. Such variability obeys inherent characteristics of musical expression—the same musician may produce different sounds when performing the same technique multiple times—on one hand, and to the small sample size, where outliers strongly impact the mean, on the other. To alleviate the influence of the former case, the data may also be observed in context; that is, if a sample is classified as Bajo but differs strongly with other samples from this technique, including in the analysis when in time it was executed, e.g. on the first

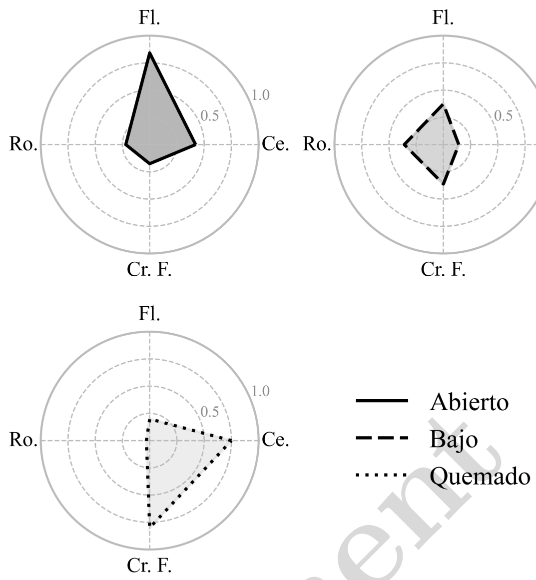

Figure 4: Timbral profiles per playing technique. (R.=Roughness, Fl.=Flux, Ce.=Centroid and Cr. F=Crest Factor) Descriptors normalized to $[0, 1]$ using global min–max across all hits; radar axes show the mean per descriptor per technique.

beat of a 4/4 metric, may improve the classification task, although this alternative was not explored and remains for future studies. The latter case, a usual constraint found in non-Western music datasets, may render the mean, as well as the corresponding descriptor, unreliable if left unaddressed. In this regard, calculating the magnitude of the effect size decreases the bias caused by a limited sample size, and its standardization facilitates comparison across techniques. This, in turn, transforms effect size magnitude into an index of the relevance of the chosen descriptors for the classification task. In the present study, since spectral bandwidth showed the smallest difference across all technique pair comparison, dismissing this feature may lead to a more accurate classification. Further studies are required to test this hypothesis. Moreover, since roughness also relates to the noisiness/purity perceptual dimension, it is possible that the bandwidth descriptor is redundant in this context. In other words, measuring different dimensions through different descriptors can also improve descriptor choices.

As a result, the reliability of the obtained profiles depends on observing (or including) the CIs from the bootstrap and the effect size analyses from Hedges's g. Including them as an optimization step that tests different descriptors, narrows their corresponding CIs, disregards those descriptors with small differences in effect size, and includes these coefficients as annotations of the data may improve automatic timbre classification pipelines working with small datasets.

5. Conclusion

This study demonstrates that reliable computational timbral characterization of non-Western percussion instruments is achievable under conditions of small sample size, provided that appropriate statistical methods are applied. By combining bootstrapping and Hedges' g, the proposed pipeline generates robust perceptual profiles that capture the most relevant (computational) perceptual descriptors for each playing technique of the Colombian Alegre drum. The resulting radar charts reveal technique-specific timbral signatures—lowest centroid for Bajo, highest crest factor for Quemado, and highest spectral flux for Abierto—which constitute a foundation for automatic data labeling and digital preservation of non-Western musical heritage. Future work should explore larger datasets, additional descriptors, and the optimization of descriptor selection based on bootstrapped confidence intervals and effect size magnitude differences to further improve classification accuracy and generalizability across instruments and traditions.

6. Acknowledgments

This research was partly funded by the Colombian Ministry of the Cultures, Arts, and Knowledges through the research grant *Beca de investigación de las colecciones de la Biblioteca Nacional de Colombia*. The author thanks the Institute for Systematic Musicology at the University of Hamburg for providing access to its anechoic chamber and recording equipment.

References

[1] J. Engel, C. Resnick, A. Roberts, *et al.*, "Neural audio synthesis of musical notes with wavenet autoencoders," in *Proc. 34th Int. Conf. on Machine Learning (ICML)*, vol. 70, JMLR.org, 2017, pp. 1068–1077. DOI: 10.5555/3305381.3305492.

[2] C. Hawthorne, A. Stasyuk, A. Roberts, *et al.*, "Enabling factorized piano music modeling and generation with the maestro dataset," in *Int. Conf. on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=r1lYRjC9F7.

[3] D. A. Cruz, C. C. Cristancho, and J. E. Camargo, "Automatic identification of traditional colombian music genres based on audio content analysis and machine learning techniques," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2019*, I. Nyström, Y. H. Heredia, and V. M. Núñez, Eds. Springer, Cham, 2019, vol. 11896, pp. 646–655. DOI: 10.1007/978-3-030-33904-3_61.

[4] I. T. Gonzalez and J. C. Osorio, "Identificación de instrumentos musicales de cuerdas pulsadas de la región andina colombiana en solo, mediante técnicas de aprendizaje de máquina," *Revista EIA*, vol. 15, pp. 177–193, 30 Dec. 2018. DOI: 10.24050/reia.v15i30.1245.

[5] D. Pico-Ripoll, W. S. Percybrooks, and C. Viloria-Núñez, "Machine learning assisted composition of traditional music from the colombian caribbean region," in *Proc. 2022 IEEE ANDESCON*, 2022, pp. 1–5. DOI: 10.1109/ANDESCON56260.2022.9989762.

[6] A. Duque, "A mano limpia," Ph.D. dissertation, Conservatorio Antonio María Valencia, 2016. [Online]. Available: https://repository.bellasartes.edu.co/handle/123456789/12.

[7] R. Bader, *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology*. Springer Berlin Heidelberg, 2013, pp. 57–108, ISBN: 978-3-642-36097-8. DOI: 10.1007/978-3-642-36098-5.

[8] A. Lazaro, R. Sarno, R. J. Andre, and M. N. Mahardika, "Music tempo classification using audio spectrum centroid, audio spectrum flatness, and audio spectrum spread based on mpeg-7 audio features," in *Proc. 3rd Int. Conf. on Science in Information Technology (ICSITech)*, IEEE, Oct. 2017, pp. 41–46, ISBN: 978-1-5090-5864-8. DOI: 10.1109/ICSITech.2017.8257083.

[9] P. Masri, "Computer modeling of sound for transformation and synthesis of musical signals," Ph.D. dissertation, University of Bristol, 1996.

[10] R. Bader, M. Blaß, and J. Franke, *Computational timbre and tonal system similarity analysis of the music of northern myanmar-based kachin compared to xinjiang-based uyghur ethnic groups*, 2021. [Online]. Available: https://arxiv.org/abs/2103.08203.

[11] W. M. Hartmann, *Signals, Sound, and Sensation*. Springer-Verlag New York, 1998, ISBN: 978-1-56396-283-7. DOI: 10.1007/978-1-4612-0575-3.

[12] W. Brent, "Physical and perceptual aspects of percussive timbre," Ph.D. dissertation, UC San Diego, 2010. [Online]. Available: https://escholarship.org/uc/item/5bx4j1fj.

[13] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1993.

[14] P. D. Ellis, *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press, 2010, ISBN: 9780521194235. DOI: 10.1017/CBO9780511761676.

[15] L. V. Hedges, "Distribution theory for glass's estimator of effect size and related estimators," *Journal of Educational Statistics*, vol. 6, pp. 107–128, 2 Jun. 1981, ISSN: 0362-9791. DOI: 10.3102/10769986006002107.

[16] L. V. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis*. Academic Press, 1985, ISBN: 9780080570655.