Audio Engineering Society

Convention Paper 10240

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

This paper was peer-reviewed as a complete manuscript for presentation at this convention. 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.

Investigating Phrase and Vocalist Dependent Microphone Preferences for Male Hip-Hop Vocal Recording

Matthew Cheshire[1] and Owen Douglas-Brooks[1]

[1]*Birmingham City University*

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

ABSTRACT

This study investigates whether microphone preference for male Hip-Hop vocal recording varies with lyrical phrasing and vocalist. Four microphones (AKG C414, Shure SM7B, Neumann TLM103, and Neumann U87) were tested on four vocalists with distinct styles performing three 4-bar phrases: two common across all vocalists and one unique self-selected phrase. Listener preference was collected through a perceptual evaluation listening test and analysed using statistical measures. Results showed that microphone preference was not consistent across phrase or vocalist. The SM7B showed the greatest variability in both phrases and vocalists, while the TLM103 also showed some notable vocalist-dependent effects. In contrast, the C414 and U87 did not produce statistically significant differences between vocalists but did present some significant changes in preferences between phrases. These findings indicate that microphone selection in Hip-Hop recording should consider both the vocalist and lyrical content, as stylistic nuance can meaningfully affect listener perception.

1 Introduction

Since emerging in the late 1970s in New York's Bronx borough, Hip-Hop has become a leading force in popular music, evolving from its roots as a platform highlighting social injustice into a globally influential genre [1, 2]. Early pioneers such as Grandmaster Flash and groups like N.W.A introduced politically charged commentary and distinct vocal delivery styles, influencing subsequent generations of artists [3, 4]. The continued commercial success of Hip-Hop into the present day has seen it diversify significantly in stylistic approaches, yet one constant remains: the critical role of vocal delivery and lyrical performance [5, 6].

Within Hip-Hop, vocal performances differ notably in rhythmic complexity, articulation speed, dynamics, and tonal qualities—collectively referred to as 'flow' [7, 4]. For instance, classic performances such as Phife Dog's verse in *Can I Kick It?* by A Tribe Called Quest contrast starkly with Snoop Dogg's relaxed rhythmic delivery on *Gin and Juice*, exemplifying stylistic variance [8]. Consequently, the optimal capture of these stylistic nuances through appropriate microphone selection is essential for authentic Hip-Hop vocal recordings [9].

Microphone choice has historically leaned towards condenser microphones due to their dynamic range and superior transient response [10, 11, 12]. Conversely, dynamic microphones, such as the Shure SM7B, remain popular due to their durability and characteristic proximity effect, influencing the tonal quality desirable in specific Hip-Hop production contexts [9].

Previous studies suggest microphone preference varies

notably across individual vocalists and even across phrases within the same vocalist's performance. Cheshire and Jones O'Shannessy [13] demonstrated that microphone preference for female R&B vocalists was significantly vocalist-dependent, highlighting the absence of a universal 'best microphone'. Building upon these findings, further investigation explores whether similar vocalist-dependent preferences extend beyond the vocalist's timbral characteristics, explicitly examining the role of lyrical content and vocal delivery style. Understanding this relationship holds value both academically and practically: for researchers, it contributes to a clearer picture of how performance context shapes perceptual evaluation of recording tools; for engineers and producers, it highlights that microphone preference may shift not only with the artist's voice, but also with the rhythmic, dynamic, and articulatory qualities of specific phrases. Such insight can improve decision-making where selecting the most appropriate microphone can impact vocal clarity, artistic authenticity, and ultimately listener reception.

Moreover, understanding microphone preferences in Hip-Hop recordings requires consideration of human auditory perception. Critical aspects of vocal quality, such as intelligibility, richness, and clarity, are closely tied to the frequency content and delivery style, particularly in frequencies between 1 kHz and 5 kHz, to which the human ear is most sensitive [14, 15]. As vocalists use varying degrees of vocal force and distinct vocal registers, these differences can significantly impact subjective perceptions of microphone suitability [16, 17]. Given the broad stylistic range within Hip-Hop vocal performances, there is a clear practical benefit to understanding how phrase content and delivery style influence microphone preference. With home and project studio production becoming commonplace, providing precise, practical guidance for engineers and artists in microphone selection is increasingly valuable [18, 19, 20].

This paper investigates the relationship between vocal delivery style and microphone preference in male Hip-Hop vocal recordings. Four commonly used microphones are evaluated using recordings of male Hip-Hop vocalists. Through subjective listening tests involving trained audio professionals, this study aims to identify whether microphone preference varies significantly depending on the lyrical content and delivery style.

2 Methods

2.1 Recording

In order to assess phrase- and vocalist-dependent microphone preference for male Hip-Hop vocal recording, four commonly used microphones—the AKG C414, Shure SM7B, Neumann TLM103, and Neumann U87—were selected. These were evaluated using four male vocalists, each performed two identical phrases and one individually selected phrase suited to their personal vocal style. Specification of the four microphones are shown in Table 1.

Microphones were chosen based on previous research and industry practice indicating their suitability for Hip-Hop vocal recording, particularly due to their proven effectiveness capturing varied vocal delivery styles, ranging from aggressive rapping to melodic singing [9, 13, 11]. Variable polar patterns on the AKG C414 and Neumann U87 were set to cardioid, and all onboard filters and pads, were disengaged. Each microphone was commercially available at the time of this study. As the primary goal of this research was not to determine the single 'best' microphone for male Hip-Hop vocals but rather to explore phrase- and vocalist-dependent microphone preferences, an exhaustive selection of microphones was deemed unnecessary. While certain high-profile artists such as Dr. Dre, Snoop Dogg, Jay-Z, and Kendrick Lamar, have notably used the Sony C-800G [21, 22, 23], this particular microphone was unobtainable for inclusion in the study. The selected microphones nonetheless represent a robust cross-section of industry-standard tools frequently employed within contemporary Hip-Hop vocal recording practices.

Two 4-bar phrases were selected from well known Hip-Hop songs ('Life's A Bitch' by Nas and 'Elevators (Me & You)' by Outkast), providing controlled conditions to evaluate microphone preference across vocalists. Each song was chosen based on its commercial success, having over 96,050,000 and 64,600,000 streams respectively on Spotify at the time of this study. Additionally, each vocalist chose a personalised phrase, tailored to their own vocal style, to investigate how microphone preference might vary with artistic and stylistic choices. Three of the four vocalist chose phrases from their own material, whilst one vocalist chose to use a 4-bar phrase from 'Cleaning out my Closet' by Eminem due to it suiting his vocal delivery style, and working as a professional Eminem tribute act.

Make	Model	Type	Polar Pattern	Released	Colour	Frequency Response
AKG	C414	Condenser	Cardioid	1971	Dark Grey	20 Hz – 20 kHz
Shure	SM7B	Dynamic	Cardioid	2001	Charcoal	50 Hz – 20 kHz
Neumann	TLM103	Condenser	Cardioid	1997	Nickel	20 Hz – 20 kHz
Neumann	U87	Condenser	Cardioid	1967	Nickel	20 Hz – 20 kHz

Table 1: Microphones used in the listening test, including colour and frequency response specifications.

The four vocalists were asked to self-assess their delivery and vocal style, their responses can be seen in Table 2. Vocalists were selected based on their experience with studio recording, as well as to represent a diverse range of different delivery styles.

Although it has been recommended to record all microphones simultaneously [24], for a fairer comparison each microphone should be recorded in the exact same point in space [25]. While simultaneous recording would eliminate differences in performances across microphones, a serial approach, such as in [13] was employed to mitigate timbral variation associated with different microphones positions. This allowed each microphone to be positioned optimally and on-axis for the vocalist. The microphone position can be seen in Figure 1. The Vocalists were instructed to maintain a consistent distance from the pop filter, which was placed 3.5" from the grill of each microphone, based on recommendations in [10], and to limit front–back and lateral movement, even though greater physical expression may occur in real-world performances. Although optimal performance may vary with microphone-specific placement, a fixed distance was chosen to ensure consistency and comparability across all models. Determining the ideal distance for each microphone was outside the scope of this study. Each microphone was connected to an SSL AWS 924 δelta mixing console, all EQs, pads, and inserts were disengaged. While certain microphones may yield more pleasing results when paired with specific preamps, the SSL AWS 924 preamps were selected to provide a consistent front end across all microphones. Artists were permitted to change the level of the backing track as they deemed appropriate. Beyerdynamic DT 770 PRO closed-back headphones were used for monitoring.

Although vocalists were aware of which microphone was in use, measures were taken to minimise the influence of preconceptions on performance. the vocalist repeated the 3 phrases 5 times each, and the most consistent takes in regards to pacing, phrasing, and dynamics across the four microphones were selected independently of any input from the vocalist. Additionally, the recording order of each microphone and phrase was randomised for each vocalist, reducing any bias that vocal fatigue may introduce. Artists were also instructed to maintain consistent delivery across microphones as best they could. While subconscious variation cannot be entirely excluded, the use of performance-matched takes substantially reduces the likelihood that microphone reputation, appearance, or familiarity biased the results. while an acoustically transparent screen could have been deployed to obscure the microphones, this may have produced an unrealistic recording scenarios unfamiliar to the vocalist used to performing into a microphone that they can see.



Fig. 1: Example of microphone position shown with Neumann U87.

Vocalist	Age	Experience	Delivery and Vocal Style (Self-Assessed)
1	28	15	Most comfortable with melodic rapping styles, similar to Nafe Smallz. Proficient with more aggressive styles, with experience as a grime MC, similar to artists like Ghetts.
2	31	3	Punchy and aggressive, with a baritone vocal range. Emphasis of vocal grit combined with more intricate and technical delivery. Has experience as an Eminem tribute.
3	25	10	Measured and technical delivery, with an emphasis on articulation.
4	23	4	Smooth and laid back delivery. Less technical, gravitating towards more melodic styles.

Table 2: Summary of vocalists, including age, years of experience, and self-assessed vocal style.

2.2 Evaluation

To assess the subjective preference of the vocal recordings, a listening test was conducted using the Web Audio Evaluation Tool (WAET) [26] with the APE interface [27], seen in Figure 2. All stimuli were loudness normalised to -23 LUFS, and playback was carried out using Beyerdynamic DT 770 PRO headphones with a Focusrite Audio interface. Participants were not permitted to adjust the volume during the test. Although recordings were made using backing tracks to help maintain consistent timing, stimuli were evaluated a cappella, without any additional accompaniment.

Before the test, participants were provided the following instructions: "A scale from 0 to 1 will be presented on each page. Please ensure the full range of the scale is utilised (i.e. Least Preferred - leftmost, Most Preferred - rightmost), and the remaining stimuli relative to your least and most preferred". Enforcing the full scale ensured that all participants rated the stimuli in a consistent way, anchored by their relative most and least preferred within each set. Without this, each listener would effectively normalise against their own internal reference, making results less comparable across participants. By requiring full use of the scale, the relative distances reflect perceived differences more consistently, whereas normalising afterwards risks skewing these distances. They could not complete the test until every stimuli had been played at least once and all sliders had been moved from their initial positions. Each test page presented four stimuli corresponding to one recording from each microphone. The starting position, slider labels, and page order were randomised for every participant on each of the 12 test pages (four vocalists, three phrases). Rating the samples on separate perceptual dimensions such as clarity, impact, or tonal balance could have introduced additional ambiguity, as each participant may interpret these terms in different and personalised ways. Assessing only preference ensured the task was consistent across participants.

In total there were 15 participants aged 20—54, all with experience in an audio-related profession. Six were recording/mixing engineers, five were producers, and the remaining four were live sound engineers and musicians. Within their profession, five participants had less than 5 years of experience, four between 5-10 years, three between 12-14 years, and three had more than 30 years. Although the listening panel consisted exclusively of audio professionals, this was appropriate given the study design. The stimuli were unmixed, unprocessed recordings, which does not reflect how consumers typically encounter vocal tracks. As such, the demographic most directly responsible for microphone choice is the engineer, and including consumers would have been less relevant.

In addition to preference ratings, participants were ask to describe what features/qualities they evaluated for their ratings. The majority of responses highlighted perceptual attributes such as control of sibilance and plosives, tonal balance, and clarity as key decision-making factors. Several participants noted that microphones with excessive harshness in the high end or exaggerated plosive content were rated lower, while those that provided greater intelligibility tended to be favoured. A smaller number of responses suggested that preferences were also shaped by how well each microphone complemented the performer's voice, underscoring the context-dependent nature of microphone selection. Some of the other semantic terms used by participants were: dynamic, body, warmth, punch, intelligibility, depth, definition, and muddiness.

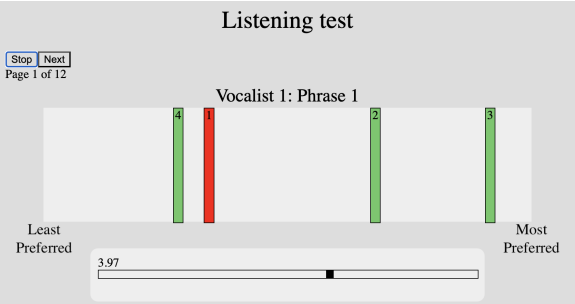


Fig. 2: Interfaced used for the subjective listening test.

3 Results

3.1 Statistical methods

To provide an initial visual overview of listener preferences, boxplots were used to summarise rating distributions for each microphone across different phrases and vocalists. This enables intuitive inspection of potential phrase- or vocalist-dependent trends prior to statistical analysis. In particular, visualising variation, consistency, and outliers supports the identification of microphones that may elicit more polarised listener preference, or whose suitability appears stable across conditions. Although formal statistical tests are required to confirm significance, the box plots in Figure 3 provide valuable insight into the structure and spread of listener ratings.

Before statistical analysis could be carried out, Shapiro-Wilk tests were conducted on the ratings to assess the distributional assumptions necessary for subsequent tests. The results indicated a mixture of normally and non-normally distributed responses across microphones, vocalists, and phrases. Out of the 48 individual microphone ratings tested (4 microphones 4 vocalists 3 phrases), 22 were found to be normally distributed ($p > 0.05$), while 26 were non-normally distributed ($p < 0.05$). Due to the prevalence of non-normality within the dataset, subsequent analyses employed non-parametric statistical methods to ensure robustness and methodological appropriateness.

Given the within-subject design of the listening tests—where each listener rated all four microphones for a given phrase—the first stage of statistical analysis used the Friedman test. The Friedman test is a non-parametric alternative to repeated-measures ANOVA, appropriate for non-normally distributed data collected across multiple conditions. Although originally developed for ranked data, it is suitable for continuous data where parametric assumptions such as normality cannot be met. It was selected to determine whether there were statistically significant differences in listener ratings among the four microphones within each individual listening test (i.e., each unique vocalist–phrase combination). The null hypothesis for each Friedman test states that all microphones are rated equivalently, with any differences in ratings arising from random chance. Rejection of the null hypothesis indicates that at least one microphone was rated significantly higher or lower than others within that specific test condition.

The Wilcoxon signed-rank test was then employed to assess pairwise differences in microphone ratings within each test page. This non-parametric test is specifically appropriate when the assumption of normality is violated, as determined by the Shapiro-Wilk tests. Although the Wilcoxon test was originally developed for ordinal or rank-transformed data, it remains valid and widely used for continuous data that are not normally distributed [28]. In the present case, participant ratings were captured using a continuous slider (0 to 1), producing data that did not adhere to a Gaussian distribution. The Wilcoxon signed-rank test evaluates whether the median of paired differences deviates significantly from zero without assuming any particular distribution. As a result, it provides a robust and distribution-free alternative to the paired t-test for within-subject comparisons of microphone preferences.

Table 3 shows the results of Friedman tests conducted independently for each listening test (i.e., each combination of vocalist and phrase), alongside post hoc Wilcoxon signed-rank tests (with Bonferroni correction) used to identify which microphones were responsible for any observed differences. A total of 8 out of 12 tests returned statistically significant Friedman results ($p < 0.05$), suggesting that microphone preference varied substantially across many of the conditions. The accompanying Wilcoxon results provide further insight by indicating which specific microphones were rated significantly higher or lower within those conditions.

The results also highlight cases where the Friedman and Wilcoxon tests did not fully align. For example, in Vocalist 4 – Phrase 1, the Friedman test did not indicate an overall significant difference, yet the Wilcoxon comparisons revealed significant differences between individual pairs. This suggests the presence of localised

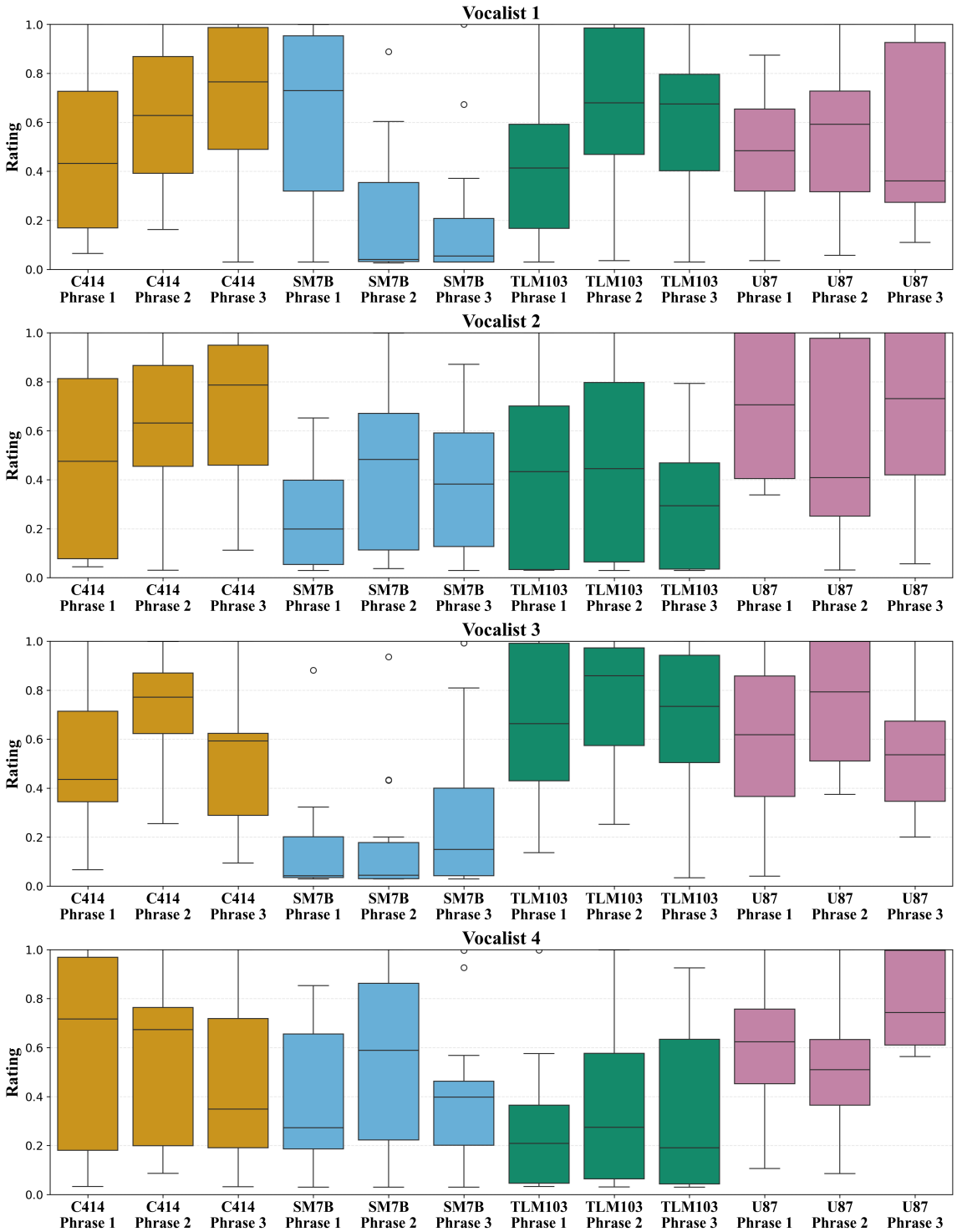


Fig. 3: Boxplots showing listener ratings for C414, SM7B, TLM103, and U87, across three phrases performed by each of the four vocalists. Each row corresponds to a single vocalist, with ratings grouped by microphone and phrase. Ratings were collected on a continuous scale from 0 (least preferred) to 1 (most preferred).

effects that were not strong enough to influence the overall ranking pattern. Conversely, in Vocalist 3 – Phrase 3, the Friedman test suggested an overall difference, but no individual Wilcoxon comparison remained significant after correction. This is likely explained by the combination of modest sample size and the conservative adjustments applied for multiple testing.

3.2 Phrase Dependence

In order to investigate phrase-dependent microphone preferences, statistical comparisons were conducted to assess whether the ratings of each microphone varied significantly across the three phrases performed by the same vocalist. This analysis aimed to determine whether the perceived performance of a microphone was influenced by the content of the vocal phrase. Although listeners did not directly compare the same microphone across all three phrases—since ratings were collected independently for each phrase—the within-subject design of the experiment allowed for meaningful statistical comparison of the rating distributions. As such, observed differences should be interpreted as indirect evidence that a microphone's preference may vary with phrase.

The Wilcoxon signed-rank test was used to compare listener ratings of the same microphone across different phrases. For each microphone, three pairwise comparisons were made: Phrase 1 vs Phrase 2, Phrase 1 vs Phrase 3, and Phrase 2 vs Phrase 3. Bonferroni correction was applied to adjust for multiple comparisons. Significant results ($p < 0.05$) are shown in Table 4 alongside the direction of effect based on median shifts between phrases.

3.3 Vocalist Dependence

In order to investigate vocalist-dependent microphone preferences, listener ratings were compared across vocalists for each microphone separately, using only Phrase 1 and Phrase 2 and excluding Phrase 3. These two phrases were common across all vocalists while phrase 3 was unique and could not be compared across vocalists. For each microphone and phrase combination, six pairwise comparisons were conducted between the four vocalists to determine whether any significant differences in listener preference emerged. This enabled the identification of microphones whose performance was particularly influenced by vocalist, as opposed to content or phrasing alone. As in previous sections, Wilcoxon signed-rank tests were used to assess these within-subject differences. Table 5 shows the results of pairwise Wilcoxon signed-rank tests for each microphone across different vocalists.

4 Discussion

From the combined results shown in Table 3, both Friedman tests and Wilcoxon signed-rank post hoc comparisons were used to evaluate microphone preferences across unique vocalist–phrase conditions. The Friedman test revealed significant differences in 8 out of 12 cases, with the strongest effects observed for Vocalist 3, where all three phrases yielded large test statistics (17.08, 20.52, and 9.00) and low p-values ($p < 0.05$). Post hoc Wilcoxon signed-rank tests (with Bonferroni correction) identified specific microphones responsible for these effects, most notably SM7B, which was consistently rated lower in multiple comparisons, particularly for Vocalist 3. Potentially as it was the only dynamic microphone used.

Table 4 shows that microphone ratings were not always consistent between phrases. Although most microphones' ratings remained stable, a few exhibited statistically significant changes. The SM7B showed the most variation, with three phrase-dependent shifts, while the C414, TLM103, and U87 each showed two. Changes were specific to certain vocalist-mic pairings, rather than following a broader pattern. These findings suggest that changes in lyrical content across phrases can influence listener preference for certain microphones, highlighting a potential need to consider phrasing when evaluating vocal mic suitability.

Table 5 showed several examples of vocalist-dependent microphone preference. The SM7B showed the highest degree of vocalist-dependent variability, with statistically significant differences in four out of six pairwise comparisons for Phrase 1, and three for Phrase 2, suggesting sensitivity to vocal identity across both passages. The TLM103 also exhibited some variation, particularly in Phrase 2, with four significant pairwise comparisons indicating inconsistent ratings across vocalists. In contrast, the C414 and U87 did not show any statistically significant differences between vocalists, implying more consistent performance irrespective of vocalist. This suggests that, while some microphones are more context-sensitive, others—such as the C414 and U87—may offer more consistent results across diverse vocal styles. Notably, no vocalist-dependent

Test	Friedman Stat	p-value	Higher Rated Mic(s)	Lower Rated Mic(s)
Vocalist 1 – Phrase 1	2.60	0.4575	—	—
Vocalist 1 – Phrase 2	11.96	0.0075	C414, TLM103	SM7B
Vocalist 1 – Phrase 3	13.16	0.0043	U87	SM7B
Vocalist 2 – Phrase 1	8.36	0.0391	U87	SM7B
Vocalist 2 – Phrase 2	1.96	0.5808	—	—
Vocalist 2 – Phrase 3	13.16	0.0043	C414	TLM103
Vocalist 3 – Phrase 1	17.08	0.0007	C414, TLM103, U87	SM7B
Vocalist 3 – Phrase 2	20.52	0.0001	C414, TLM103, U87	SM7B
Vocalist 3 – Phrase 3	9.00	0.0293	—	—
Vocalist 4 – Phrase 1	7.56	0.0560	U87	TLM103
Vocalist 4 – Phrase 2	3.08	0.3795	—	—
Vocalist 4 – Phrase 3	15.48	0.0014	U87	SM7B, TLM103

Table 3: Combined results of Friedman tests and Wilcoxon signed-rank post hoc analysis for each listening test (vocalist–phrase pair). The microphones with significantly ($p < 0.05$) different listener ratings based on the results of the Wilcoxon signed-rank are shown in the last two columns.

Vocalist	Mic	Phrase 1 vs 2	Phrase 1 vs 3	Phrase 2 vs 3
1	C414	ns	ns	ns
	SM7B	0.004 (↓)	0.004 (↓)	ns
	TLM103	0.030 (↑)	ns	ns
	U87	ns	ns	ns
2	C414	ns	ns	ns
	SM7B	0.030 (↑)	ns	ns
	TLM103	ns	ns	0.030 (↓)
	U87	ns	ns	ns
3	C414	0.041 (↑)	ns	0.010 (↓)
	SM7B	ns	ns	ns
	TLM103	ns	ns	ns
	U87	ns	ns	ns
4	C414	ns	ns	ns
	SM7B	ns	ns	ns
	TLM103	ns	ns	ns
	U87	ns	0.011 (↑)	0.009 (↑)

Table 4: Results of Wilcoxon signed-rank tests assessing phrase-dependent microphone preferences for each vocalist and microphone. Each row compares listener ratings of the same microphone across two different phrases for a given vocalist. Statistically significant differences ($p < 0.05$) are reported alongside an arrow indicating the direction of change: (↑) indicates the median rating increased in the second phrase pair, and (↓) indicates it decreased, ns denotes non-significant results.

Phrase	Mic	V1 vs V2	V1 vs V3	V1 vs V4	V2 vs V3	V2 vs V4	V3 vs V4
1	C414	ns	ns	ns	ns	ns	ns
1	SM7B	0.007 (↓)	0.004 (↓)	0.030 (↑)	ns	ns	0.013 (↑)
1	TLM103	ns	ns	ns	ns	ns	0.004 (↓)
1	U87	ns	ns	ns	ns	ns	ns
2	C414	ns	ns	ns	ns	ns	ns
2	SM7B	ns	ns	0.019 (↑)	0.035 (↓)	ns	0.009 (↑)
2	TLM103	0.022 (↓)	ns	0.018 (↓)	0.047 (↑)	ns	0.004 (↓)
2	U87	ns	ns	ns	ns	ns	ns

Table 5: Pairwise Wilcoxon signed-rank test results showing vocalist-dependent preferences for each microphone across Phrase 1 and Phrase 2. Statistically significant differences ($p < 0.05$) are reported alongside an arrow indicating the direction of change: (↑) indicates the median rating increased in the second vocalist pair, and (↓) indicates it decreased, ns denotes non-significant results.

scores for any of the microphones or phrase were found between Vocalist 2 and 4, despite them self-assessing their vocal style quite differently; punchy and aggressive vs smooth and laid back respectively. This could be due to several factors such as similar frequency content of the two vocalists despite drastically different style, although further investigation is required.

5 Summary

This study explored whether microphone preference for male Hip-Hop vocal recording varies depending on the lyrical phrase and vocalist. Using controlled studio recordings and a within-subject listening test, four microphones (AKG C414, Shure SM7B, Neumann TLM103, and Neumann U87) were evaluated across three vocal phrases and four vocalists. Statistical analyses revealed that microphone preference was not fixed, but instead varied according to both phrasing and the vocalist. The SM7B exhibited the greatest variability, showing both phrase- and vocalist-dependent shifts in listener preference. In contrast, the C414 and U87 demonstrated greater consistency across different vocalists, although both produced phrase-dependent ratings for vocalist 3 and 4 respectively. These findings suggest that in the context of Hip-Hop, certain microphones are more sensitive to the nuances of vocalist and lyrical content, supporting the notion that no single microphone is universally optimal. Engineers and artists should consider both the stylistic qualities of vocal performance and the specific lyrical content when choosing a microphone.

While the study offers meaningful insights, it is limited by the relatively small pool of vocalists and microphones, as well as the use of only three phrases. Listener preference is likely also influenced by other factors such as post-processing, context, or production style, which were not accounted for in this study. Future work could expand the dataset to include a broader range of microphones, vocalists, and musical genres, as well as incorporating processed recordings to better simulate real-world production conditions. Objective measurements such as sensitivity, dynamic range, directivity, and frequency response could also be integrated to provide a technical basis for subjective judgements, helping to identify which measurable qualities of microphones align with listener preferences for particular artists or phrases. In addition, future studies should consider blinding vocalists to the microphones in use to minimise potential performance bias, and explore multidimensional rating scales to capture perceptual attributes alongside overall preference.

References

[1] Land, R. R. and Stovall, D., *Hip hop and Social justice education: A brief introduction*, volume 42, Equity and Excellence in Education, 2009.

[2] Diaz, M., "The World IS Yours: A Brief History of Hip-Hop Education," https://www.academia.edu/1088920, 2011.

[3] Neal, M. A., *That's the Joint!: The Hip-Hop Studies Reader*, Routledge, New York, 2004.

[4] Davis, I. R., *A Rhythmic Analysis of Rap: What Can We Learn from 'Flow'?*, Master's thesis, Uni-

versity of Canterbury, Christchurch, New Zealand, 2017.

[5] Chartmetric, "Chartmetric A Year in Music," https://reports.chartmetric.com/2023/year-in-music/part-1, 2023.

[6] Ruusunen, R., "Mixing modern hip-hop music," https://www.theseus.fi/bitstream/handle/10024/804268/Ruusunen_Roope.pdf, 2023.

[7] Adams, K., "On the Metrical Techniques of Flow in Rap Music," *Music Theory Online*, 15, 2009.

[8] Komaniecki, R., *Analyzing the Parameters of Flow in Rap Music*, Ph.d. dissertation, Indiana University, Bloomington, IN, 2019, proQuest Dissertations Publishing, Publication No. 2299814298.

[9] Senior, M., *Recording Secrets for the Small Studio*, Routledge, 2023.

[10] Huber, D. M. and Runstein, R. E., *Modern Recording Techniques*, Routledge, sixth edition, 2005.

[11] Eargle, J., *The Microphone Book*, Routledge, second edition, 2005.

[12] Robjohns, H., "A brief history of microphones," https://web.archive.org/web/20101125131858/http://microphone-data.com/media/filestore/articles/History-10.pdf, 2001.

[13] Cheshire, M. and Jones O'Shannessy, E., "Microphone Comparison for Female R&B Vocal Recording," *Journal of the Audio Engineering Society*, (10679), 2023.

[14] Smith, S. W., *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Pub, 2011.

[15] Hermes, K., "Factors influencing the spectral clarity of vocals in music mixes," 2019.

[16] Childers, D. G. and Lee, C. K., "Vocal quality factors: Analysis, synthesis, and perception," *Journal of the Acoustical Society of America*, 90(5), pp. 2394–2410, 1991.

[17] Pickett, J. M., "Effects of Vocal Force on the Intelligibility of Speech Sounds," *Journal of the Acoustical Society of America*, 28(5), pp. 902–905, 1956.

[18] Hughes, R., "Home Studio Setup Costs Compared - 1980s And Now," 2024.

[19] Schonbrun, M., "Modern-Day Developments - Digital Home Recording," 2011.

[20] Stahl, G. and Percival, J. M., *The Bloomsbury Handbook of Popular Music, Space and Place*, Bloomsbury Publishing, 2022.

[21] Shaw, D., "The Sony C800G: the recording enigma that achieved immortality," https://happymag.tv/sony-c800g/, 2020.

[22] McDonough, M., "Sony C-800G: Hip-hop's (Not So) Secret Weapon," https://www.sweetwater.com/insync/sony-c-800g-hip-hops-not-so-secret-weapon/, 2023.

[23] Simion, M., "Which Are The Best Microphones for Recording Rap Vocals," https://songmixmaster.com/which-are-the-best-microphones-for-recording-rap-vocals, 2023.

[24] De Man, B. and Reiss, J. D., "A pairwise and multiple stimuli approach to perceptual evaluation of microphone types," *Journal of the Audio Engineering Society*, 2013.

[25] Pearce, A., Brookes, T., and Dewhirst, M., "validation of experimental methods to record stimuli for microphone comparisons," *journal of the audio engineering society*, (9385), 2015.

[26] Jillings, N., De Man, B., Moffat, D., and Reiss, J. D., "Web Audio Evaluation Tool: A Browser-based Listening Test Environment," in *12th Sound and Music Computing Conference*, 2015.

[27] De Man, B. and Reiss, J., "APE: Audio Perceptual Evaluation Toolbox for MATLAB," in *136th Convention of the Audio Engineering Society*, 2014.

[28] Conover, W., *Practical Nonparametric Statistics*, John Wiley & Sons, 3rd edition, 1999, ISBN 9780471160687.