F0 ANALYSIS OF GHANAIAN POP SINGING REVEALS PROGRESSIVE ALIGNMENT WITH EQUAL TEMPERAMENT OVER THE PAST THREE DECADES: A CASE STUDY

Iran R. Roman
New York University
`roman@nyu.edu`

Daniel Faronbi
New York University
`df2322@nyu.edu`

Isabelle Burger-Weiser
New York University
`ib2225@nyu.edu`

Leila Adu-Gilmore
New York University
`leila.adu-gilmore@nyu.edu`

ABSTRACT

Contemporary Ghanaian popular singing combines European and traditional Ghanaian influences. We hypothesize that access to technology embedded with equal temperament catalyzed a progressive alignment of Ghanaian singing with equal-tempered scales over time. To test this, we study the Ghanaian singer Daddy Lumba, whose work spans from the earliest Ghanaian electronic style in the late 1980s to the present. Studying a singular musician as a case study allows us to refine our analysis without over-interpreting the findings. We curated a collection of his songs, distributed between 1989 and 2016, to extract F0 values from isolated vocals. We used Gaussian mixture modeling (GMM) to approximate each song's scale and found that the pitch variance has been decreasing over time. We also determined whether the GMM components follow the arithmetic relationships observed in equal-tempered scales, and observed that Daddy Lumba's singing better aligns with equal temperament in recent years. Together, results reveal the impact of exposure to equal-tempered scales, resulting in lessened microtonal content in Daddy Lumba's singing. Our study highlights a potential vulnerability of Ghanaian musical scales and implies a need for research that maps and archives singing styles.

1. INTRODUCTION

Ghanaian popular music today is the result of a long history of European influences interacting with traditional styles, employing scales with microtonal variance outside equal temperament. Throughout the 20th century, Ghana featured hybrid music with complex harmony such as Palm-Wine, Highlife and more recently, Hiplife and A-frobeats [1,2]. These musics were influenced by traditional Ghanaian music, including European influences such as Christian hymns, sea shanties and brass band music [1].

Traditional Ghanaian music is full of features that set it apart from the Western European tradition. Ghanaian harmony is complex and compelling, yet there is a lot of harmonic information to be examined on scales, pitch sets and tuning. Kofi Agawu argues for transcriptions as a valid

means of studying African music [3] in keeping with our methods of music theory analysis. A notable in-depth study is Ampene's research on *nwomkoro*, traditional female Akan singing, with a leader's call and the chorus' harmonized response [4]. Adu-Gilmore (2017) [5] has researched Ghanaian harmony for over a decade, and although afrobeats has become a best-selling global phenomenon, the culture around African electronic music has been the focus of academic research but the music theory and harmony is still under-researched. Her research has built upon Ampene's showing parallel harmonies and step-wise chord progressions, creating tonal centers outside of the major and minor modes of the key, implying lesser-used modes in European styles such as Phrygian, Lydian and Mixolydian [5]. These harmonic tendencies continued to present-day popular music styles such as Ghanaian hip-hop, Highlife and Afrobeats, combining traditional Ghanaian scales with digital instruments and digital audio workstations (DAWs) [2], which are equal-tempered. Moreover, Moelants and colleagues claim that there is a move in African music towards "more elaborate, equally-tempered scales," even though African music traditionally has differed from the chromatic, equal-tempered twelve-tone scale [6].

The international dominance of Western equal-tempered tuning and tonality endangers traditional scales of the world. Internet music dissemination has increased audiences' music access globally. Ghanaians' consumption of music from overseas impacts Ghanaian music production and use of effects, including autotune. Although autotune is mainly a means to tune vocals towards equal-temperament, Cher's 1998 hit "Believe" [7] very audibly used the autotune effect for artistic purposes, producing an unnatural and robotic effect, later popularized by rappers, notably T-Pain, who even created a commercial microphone with an inbuilt autotune [8]. According to Ghanaian music producer, Kofi "IAmBeatmeance" Boachie-Ansah, Ghanaian tracks such as Sony Achiba's "Sony Maba" [9] used autotune early on. Boachie-Ansah states: "This was a few years after Cher's "Believe," when that song was dominating the charts. It had a heavy rotation even here, for years [. . .] Accra has always had a very vibrant nightlife culture driven by a lot of dance music so some of the sensibilities that informed those genres of music found a way of seeping into the popular music of the time" [10].

We hypothesize that the exposure to music technology and the internet, over the last three decades, has led to musical scales used in popular Ghanaian singing to be more

aligned with equal-tempered scales in artists such as Daddy Lumba. This paper presents a case study with quantitative evidence of a move over the years towards pitches that are more in tune with a discrete scale instead of traditional Ghanaian microtonal features, while also gradually trending towards equal temperament.

To test this hypothesis, we focus on the music of Daddy Lumba, who has been described by Ghanaian popular music scholar John Collins as "one of the most popular Burger Highlife artists" alongside his brothers, Nana Acheampong and Sarkodie [1]. Daddy Lumba's career spans from the earliest Ghanaian electronic style, Burger Highlife in the late 1980s, to similar present-day musics such as Ghanaian Hiplife [11]. Similar to most Ghanaian popular artists, Daddy Lumba is notoriously reclusive and biographical information in general is slim, mostly reliant on music entertainment websites (at times unreferenced, with no author name included) and *wikipedia*. Daddy Lumba's sings music in popular music genres that prominently feature the use of equal-tempered synthesizers and DAWs.

We curated a dataset of Daddy Lumba recordings to ask whether his singing reflects features that progressively abandon traditional Ghanaian scales with microtonal pitch variance in favor of equal temperament. We chose the dataset of a singular musician as a case study to refine our analysis method and to avoid over-interpreting the findings. Limiting the scope in this way allows us to ensure specificity in the results.

By presenting empirical evidence, we argue that Ghanaians' increased access to equal-tempered music through internet dissemination coupled with greater access to music technology is exponentially speeding up the threat to traditional Ghanaian music intonation and scale variance. That is, equal-tempered synthesizers and effects such as autotune enable a more pronounced tendency toward equal-temperament. Far from combating these phenomena, this case study aims to map this change and call for archiving and preserving traditional global musics and their practice.

2. RELATED WORK

Although music scales have been examined over centuries, their study using large datasets is made possible by two major advances in music information retrieval (MIR). The first consists of audio source separation algorithms, such as Wave-U-Net [12], recurrent neural networks [13], and spectrogram U-Nets [14], which can be used to automatically isolate vocal or single instrument tracks [1]. The second advancement consists of F0 estimation algorithms that operate directly on waveform signals, such as YIN [16,17] and CREPE [18]. Together, these advancements allow for the extraction of singing F0 values, and the subsequent approximation of musical scales via statistical modeling.

Previous research has used F0 estimation methods in non-Western music recordings. A recent study combined source-separation and F0 estimation methods to quantitatively analyze melodic mode switching in raga performances [19]. Additionally, the YIN algorithm has been used, on its own,

songs per year (1989-2016)	song dur (avg)	song dur (std dev)	total duration
3.54	5m42s	1m8s	9h23m23s

Table 1. Column 1: the average number of songs included in our dataset per year. Columns 2 and 3: the average song duration and standard deviation, respectively. Column 4: the total duration of our dataset.

F0 dur. : song dur. (avg.)	F0 dur. : song dur. (std. dev.)	total F0 duration
0.22	0.06	2h5m37s

Table 2. Columns 1 and 2: the average and standard deviation of the ratio between duration of F0 values extracted and total song duration. Column 3: the total duration of F0 values extracted from our dataset.

to extract the scales present in African and Turkish instrumental melodies [20,21], as well as Indian singing [22,23]. These studies sought to assess whether the scales found in digital musical recordings with real musical practice are consistent with music theories for each of these music traditions [24]. More recently, researchers used CREPE to compare the pitch content among human signing, speech, and birdsongs, finding that human singing, across cultures and countries, emphasizes integer ratio relationships between pitch content (i.e. perfect fifths), compared to speech or birdsongs [25]. Similar analysis (but using YIN instead of CREPE) on child and adult singing was used to study the possible sensorimotor origins of inter-ratio relationships commonly seen in sung pitch content [26].

These studies have been important in showing how modern MIR tools can be applied to analyze melodic pitch content in digital recordings of music. We have yet to find a study, however, that uses a data-driven methodology to analyze the impact of equal temperament on Ghanaian scales. We therefore believe that our study will be the first in the field to show such an effect.

3. METHOD TO ESTIMATE F0 VALUES IN DADDY LUMBA'S SINGING

Our goal is to study the changes in the scales and pitch content sung by Daddy Lumba over the course of twenty-seven years, as captured in his discography. To do this, we compiled a corpus of 99 songs released between 1989 and 2016 (stereo, mp3, 44.1 kHz sampling rate). Table 1 shows a summary of the most relevant metadata in our corpus.

Our corpus contains mastered songs, which means that vocals, instruments, and other effects are mixed in the tracks. We use CREPE [18] to do F0 estimation of the vocal tracks. However, CREPE was designed and tested to work on monophonic signals (i.e. tracks with only one instrument) [18]. To isolate the vocals from all other instruments and effects in a song, we used the pre-trained Wave-U-Net made available by its original authors [2]. Using the Wave-U-Net on

---

[1] these type of algorithms have recently been made available in dedicated software toolboxes like Spleeter [15]
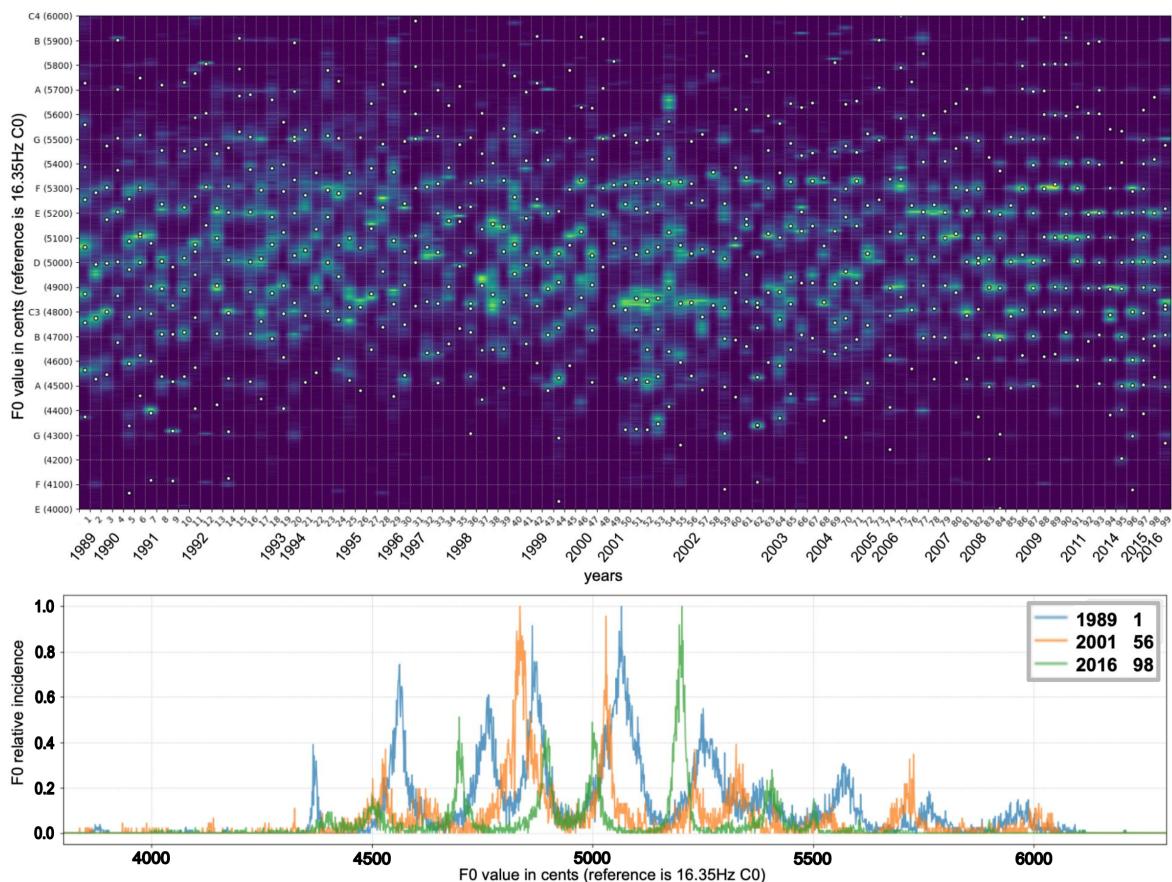
[2] https://github.com/f90/Wave-U-Net-Pytorch

28

Figure 1. Top panel: each song's F0 histogram, column-wise. Lighter colors indicate higher density. White dots represent the optimal GMM components found. Bottom panel: three sample F0 histograms (normalized so that the F0 with the highest incidence is equal to 1, for visualization purposes).

each song, we obtained tracks containing only the vocals sung by Daddy Lumba and vocal accompaniment.

Having the isolated vocal track, we can use CREPE to estimate the F0s in Daddy Lumba's songs. A pre-trained version of CREPE is made available by its original authors [3] . Given an audio signal, this version of CREPE estimates an F0 value, in Hz, every 10ms, and each value is accompanied by a confidence score between 0 and 1. We used CREPE version v0.0.7 to extract F0 values from each song's isolated vocal track. To ensure the quality of the F0 values we applied the following heuristic filtering method: 1) we discarded F0 values outside the range between 80Hz and 600Hz (the vocal range of male singing) 2) we also discarded F0 values that were given a confidence score below 0.8 by CREPE. Finally, we converted the F0 values from Hz to cents using Eq. (1)

$$F0_{cents} = 1200 \log_2\left(\frac{F0_{Hz}}{16.352_{Hz}}\right), \quad (1)$$

where 16.352 Hz corresponds to the C0 tone that we used as reference. Table 2 gives a summary of the amount of F0 values that we were able to obtain from the dataset using

our extraction and filtering method. In total, we were able to obtain over 2 hours of F0 values. On average, from each song we obtained 22% of its duration as F0 values.

We used these F0 values to carry out two analyses that quantify changes over Daddy Lumba's singing career. First, we used a Gaussian mixture model (GMM) to approximate the scale of each song and study the variance around pitches in the sung scale. Second, we used the GMM results to determine whether Daddy Lumba's singing scales reflect an equal-tempered structure.

4. ANALYSIS 1: ESTIMATING SONG SCALES AND VARIANCE AROUND THEM OVER YEARS

In this analysis we approximate each song's scale to look at the variations in pitch content. We observe whether this variation has changed over Daddy Lumba's career. Our specific research question is: has Daddy Lumba's singing reduced its pitch variations over the years?

4.1 Methodology

In general, a Gaussian mixture is a probabilistic unsupervised method that models data as a probability distribution

---
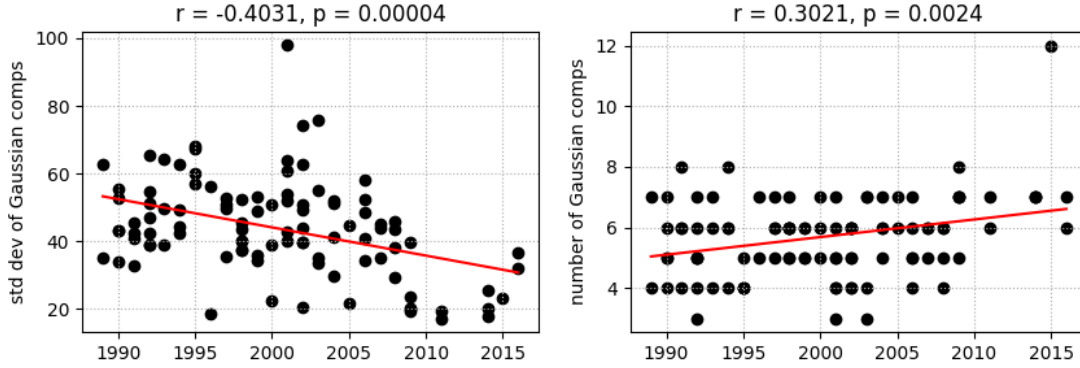
[3] https://github.com/marl/crepe

Figure 2. Left panel: a linear regression reveals a significant negative relationship for the F0 GMM variance over the years. Right panel: a linear regression reveals a significant positive relationship for the number of F0 GMM components over the years.

that is the sum of different weighted Gaussians. Eq. (2) describes the general equation

$$p(\theta) = \sum_{i=0}^{C-1} \phi_i \mathcal{N}(\mu_i, \sigma_i), \qquad (2)$$

where μ_i and σ_i are the i'th component's mean and variance, and each component is weighted by a scalar ϕ_i.

For each song's F0 values, we estimate the parameters of a Gaussian mixture with "tied" covariance [27]. For the specific case of "tied" covariance, the coefficient $\phi_i = 1$ is used and a constant σ is found for all components in the GMM [28]. In other words, the data is clustered by using the same distance criterion across the range of F0 values. We use a tied covariance because we are assuming that the distribution of F0 values sung around each pitch in a scale is the same (i.e. Daddy Lumba's singing has a characteristic variability around scale tones, independent of how high- or low-pitched a tone is). We limited the F0 values considered for the GMM to be an an octave around the most commonly-sung F0 value in each song.

We validate number of components C in a GMM via the highest euclidean silhouette score around each Gaussian component mean value. The silhouette score is a heuristics based method that yields a coefficient by

$$s = \frac{1}{N} \sum_{i=0}^{N-1} \frac{b_i - a_i}{\max(a_i, b_i)}, \qquad (3)$$

where a_i is the average distance between a point i and the other points in the same Gaussian component (i.e. pointwise inter-cluster distance), and b_i is the average distance between a point i and the points in the closest Gaussian component that i is not part of. The silhouette score ranges from -1 to 1, with values close to 1 indicating that datapoints, on average, belong to their assigned Gaussian component, while -1 indicates that datapoints, on average, do not belong to their assigned Gaussian component. Since our GMM components have a "tied" covariance, the silhouette score is a fair metric to validate components that should have similar density properties.

We tested integer values between 4 and 15 for the possible number of components, with 4 being the lowest number of pitch classes that we considered to be reasonable for a singing scale, and 15 the maximum within an octave. The GMM with the number of components that maximized the silhouette score was selected to model a song's scale.

4.2 Results: singing scales approximated via GMM

The top panel of figure 1 shows a matrix where columns are song histograms of F0 values, chronologically ordered from left to right. White dots are the mean of each component in the song's F0 GMM. On average, the silhouette score associated with each song's F0 GMM was 0.72 (lowest was 0.62 and highest was 0.87). The GMM components (i.e. the white dots) are an approximation of each song's scale based on our GMM approach. The bottom panel of Figure 1 shows three example F0 histograms, normalized so that the highest peak of each is equal to 1 (for visualization purposes). Note how the spread around peaks is visually smaller for the years 2016 and 2001 compared to 1989. With that said, we should be mindful that our dataset has a limited number of data points from 2015 onwards. We believe that less data starting at that point is not a big problem since, as shown in the top panel of Figure 1, Daddy Lumba's singing reflects the use of autotune.

4.3 Results: F0 GMM components over the years

The left panel of figure 2 shows the standard deviation associated with each song's F0 GMM components. A linear regression revealed a significant negative correlation (r = -0.3839; p < 0.05), indicating that, as years have progressed, the vocals have reduced their microtonal pitch diversity, suggesting a tendency to sing more "in tune" with the song's scale. The right panel of figure 2 shows the number of GMM components in each song's F0 GMM. A linear regression revealed a significant positive correlation (r = 0.2834; p < 0.05), indicating that, over the years, Daddy Lumba's sings scales with more tones per octave.

5. ANALYSIS 2: QUANTIFYING ALIGNMENT WITH EQUAL TEMPERAMENT OVER YEARS

Analysis 1 resulted in a GMM for each song, with a unique number of components and standard deviation around them, as well as a characteristic distance between components.

5.1 Methodology

In a song that perfectly follows the equal-tempered scale, the distance between any two notes n_i and n_j follows this rule (in units of cents):

$$|n_i - n_j| = N \times 100, \tag{4}$$

where 100 is the shortest possible distance between two consecutive notes in a scale, and $N \in \mathbf{Z}$ is the number of times such distance separates the notes n_i and n_j. 100 cents is the smallest possible distance between equal-tempered notes. However, no singer is perfect and musical performances have an error term associated, making Eq. (4):

$$|n_i - n_j| = N \times 100 + \epsilon, \tag{5}$$

where ϵ is a value between 0 and 50, reflecting an error from the perfect minimal distance of 100 cents. An ϵ closer to zero indicates better alignment with equal-temperament.

We want to quantify to what extent Daddy Lumba's singing reflects a scale that is consistent with equal temperament, and whether his singing has progressively become more equal-tempered over the years. Hence, we calculated Eq. (5) between the F0 GMM components found in each song to get the song's average ϵ_s.

5.2 Results: equal temperament alignment over years

A linear regression over the songs' ϵ_s values revealed a significant negative correlation (r = -0.2552; p < 0.05), indicating that, over the years, Daddy Lumba's singing has progressively reduced the ϵ error term in Eq. (5). Figure 3 shows this regression.

The code and F0 data needed to reproduce our results and generate the figures in this paper is openly available [4].

6. DISCUSSION

We are interested in examining the microtonal nuances in Ghanaian singing and its associated musical scales. This initial study analyzes Daddy Lumba's work over time because he has an influential catalog of work spanning from early Ghanaian electronic music called Burger Highlife in the 90s, until the present moment.

The first analysis used GMM to approximate the singing scales in Daddy Lumba's songs (independent of whether they are equal-tempered or not), and looked closely at the pitch variations. The results in the left panel of figure 2 showed that, over time, he has tended to sing closer to the discrete tones in his pitch sets (or modes), with less microtonal variance (i.e. he sings more "in tune" with the scale he uses in a song). The potential reasons for this could be manifold. For instance, his personal experience
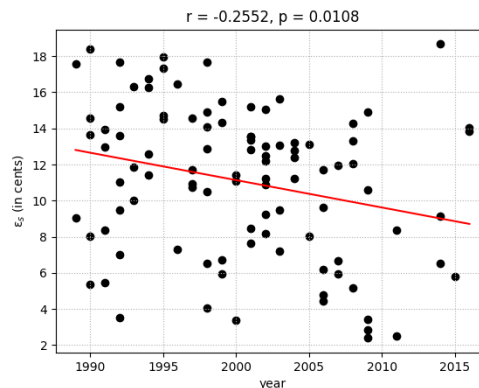
_____
[4] https://github.com/iranroman/AFRINUM

Figure 3. A linear regression of how error ϵ_s associated with deviations from equal temperament has been reducing over the years, indicating alignment with an equal-tempered scale

and developing artistic identity could be the driving force, as Daddy Lumba's career has spanned many years, and this experience could mean that, over time, the precision of his singing has been refined to target his chosen scale. However, over the same time-span, the pop music industry has seen an increased access to high-quality, budget alternatives to recording studios, such as DAWs that allow singers to iterate over potential recordings and select tracks where their voice reflects their preferred tuning precision and scale. Editing in DAWs is typically a lot faster, and different takes may be easier to record, and edit [29]. These changes in technology have implications everywhere.

We also analyzed how Daddy Lumba sings in relation to equal temperament, and exposed a gradual shift towards equal temperament over the years (see Figure 3). The reasons for this may also be linked to access to technology. First, traditional scales may be affected by fixed pitch equal-tem- pered instruments, such as piano, sound synthesizers, and wind instruments, as the string instruments and voices are tuned to them. Second, electronic instruments and software using "auto-tuned" MIDI are tuned to the equal-tempered system. Third, music made using electronic instruments and computers is pervasive. Fourth, the mass music distribution through the internet has increased the speed at which equal-tempered popular music is disseminated, consumed and absorbed. Therefore, digital instruments and effects embedded with tonal harmony of Western-European Christian scales and equal temperament may impact Ghanaian singing, as well as other music across the globe with traditional scales, tuning conventions and vocal inflections.

Why do we compare Daddy Lumba's singing against equal temperament? Because of its history as a British colony, to study Ghanaian popular music today, Daddy Lumba included, it is necessary to analyze it through the lens of both Western European tonal harmony and traditional Ghanaian harmony. Critical Sonic Practice (CSP) investigates global music by dissolving borders between music technology,

composition and theory, while including indigenous epistemologies. Western-European Christian tonal harmony is necessary in this discussion, not because this scale is "in tune," but because these scale systems are a cultural choice at the heart of the European colonizing force of music [30].

7. CONCLUSION

This study aims to catalyze a scientific discourse about how global access to music technology has influenced Ghanaian music, popular and traditional, as well as other global genres, including the global north, where traditional styles, like flamenco in Spain or Rímur in Iceland, feature microtonal inflections that cannot be defined by the equal-tempered system [31, 32]. Evolution in music formats, such as vinyl to tape dubbing, already increased access to broader genres; whereas internet dissemination increased access to equal-tempered music from abroad, and gives Ghanaians more access to global markets that picked equal temperament as an unofficial standard. Furthermore, recent advances in technology, such as cheaper (and free) DAWs, virtual studio technology (aka VST), and the prevalence of autotune increase the ubiquity of equal temperament. Changes in music scales and temperament are inevitable and MIR is an essential tool in mapping this change.

The findings in this study align with the mission of CSP, which includes decolonizing datasets, and departing from norms of tuning that are often overlooked, taken for granted, or even unknown. In doing so, we recognize that current algorithms are often trained with datasets that reflect majoritian musics. The specificity of this paper as a case study, which closely examines one artist, signifies that we can more easily understand the culturally-relevant nuances of a given artist's corpus. This allows us to foresee potential errors in cultural bias or expectation. Therefore, we are fine-tuning necessary steps in research design that are culturally sensitive to the dataset, before continuing to move forward toward our future work with larger datasets encompassing entire genres, music traditions, or geographical regions.

To reiterate, this paper demonstrates a research design for detecting microtonal variance over time in the music of one artist, rather than a comprehensive study of Ghanaian harmony. That said, the findings point towards more research into Ghanaian modes and tonalities, as well as the changes affected by technologies.

8. FUTURE RESEARCH

We will expand our work by analyzing other artists in the same genre. We also aim to research and expand the timeline to encompass releases before 1989, hoping to better map and describe traditional scales and influences. Future works using MIR could involve an in-depth study of Ghanaian scales and modes in the same dataset or others to expand the archive of Ghanaian popular and traditional musics. These could be cross-referenced with expert listening panels to verify the empirical data. This research will be augmented by interviews with artists, ensuring culturally specific context for the broader study in order to investigate the reason behind this phenomenon. The work

presented here lays the groundwork for the empirical, data-driven archiving of traditional and local modes and tonalities. With these goals in mind, this study has broader implications in the decolonization of music research requiring archiving of tonal systems from all over the world for cultural preservation.

9. REFERENCES

[1] J. Collins, "Contemporary ghanaian popular music," in *Hip Hop Africa: New African Music in a Globalizing World.*, E. Charry, Ed. Indiana UP, 2012, pp. 211–231.

[2] L. Adu-Gilmore, "Studio improv as compositional process through case studies of ghanaian hiplife and afrobeats," *Critical Studies in Improvisation/Études critiques en improvisation*, vol. 10, no. 2, 2015.

[3] K. Agawu, *Representing African music: Postcolonial notes, queries, positions*. Routledge, 2014.

[4] K. Ampene and A. Scharfenberger, *Female Song Tradition and the Akan of Ghana: The Creative Process in Nnwonkoro (SOAS Musicology Series)*. Aldershot: Ashgate, 2005.

[5] L. Adu-Gilmore *et al.*, "Music production in accra: Ghanaian hiplife, akan and highlife harmony and african-american minimalist hip-hop in two case studies," 2017.

[6] D. Moelants, O. Cornelis, and M. Leman, "Exploring african tone scales," in *10th International Society for Music Information Retrieval Conference (ISMIR-2009)*. International Society for music Information Retrieval, 2009, pp. 489–494.

[7] B. Higgins, S. McLennan, P. Barry, S. Torch, M. Gray, and T. Powell, "Believe [recorded by cher]," *On Believe [CD, track 1]. United States: Warner Brothers*, 1998.

[8] T. Dean and T. Brice, "How geophysics ruined music," *Preview*, vol. 2022, no. 217, pp. 43–47, 2022.

[9] S. Achiba, "Sony maba [recorded by achiba]," 2001.

[10] K. Boachie-Ansah, "personal communication with kofi boachie-ansah," 2023.

[11] E. Charry, "The birth of ghanaian hiplife: Urban style, black thought, proverbial speech," in *Hip Hop Africa: New African Music in a Globalizing World.*, E. Charry, Ed. Indiana UP, 2012, pp. 29–36.

[12] D. Stoller, S. Ewert, and S. Dixon, "Wave-u-net: A multi-scale neural network for end-to-end audio source separation," *arXiv preprint arXiv:1806.03185*, 2018.

[13] P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Singing-voice separation from monaural recordings using deep recurrent neural networks." in *ISMIR*, 2014, pp. 477–482.

[14] A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde, "Singing voice separation with deep u-net convolutional networks," in *18th International Society for Music Information Retrieval Conference*, 2017, pp. 23–27.

[15] R. Hennequin, A. Khlif, F. Voituret, and M. Moussallam, "Spleeter: a fast and efficient music source separation tool with pre-trained models," *Journal of Open Source Software*, vol. 5, no. 50, p. 2154, 2020.

[16] A. De Cheveigné and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.

[17] M. Mauch and S. Dixon, "pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 659–663.

[18] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, "Crepe: A convolutional representation for pitch estimation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 161–165.

[19] N. Shikarpur, A. Keskar, and P. Rao, "Computational analysis of melodic mode switching in raga performance," in *22th International Society for Music Information Retrieval Conference*, 2021, pp. 657–664.

[20] J. Six and O. Cornelis, "Tarsos: a platform to explore pitch scales in non-western and western music," in *12th International society for music information retrieval conference (ISMIR-2011)*. University of Miami, Frost School of Music, 2011, pp. 169–174.

[21] B. Bozkurt, R. Ayangil, and A. Holzapfel, "Computational analysis of turkish makam music: Review of state-of-the-art and challenges," *Journal of New Music Research*, vol. 43, no. 1, pp. 3–23, 2014.

[22] J. Serra, G. K. Koduri, M. Miron, and X. Serra, "Assessing the tuning of sung indian classical music." in *ISMIR*. Florida, 2011, pp. 157–162.

[23] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Rāga recognition based on pitch distribution methods," *Journal of New Music Research*, vol. 41, no. 4, pp. 337–350, 2012.

[24] G. Tzanetakis, "Computational ethnomusicology: a music information retrieval perspective." in *ICMC*, 2014, pp. 112–117.

[25] J. Kuroyanagi, S. Sato, M.-J. Ho, G. Chiba, J. Six, P. Pfordresher, A. Tierney, S. Fujii, and P. Savage, "Automatic comparison of human music, speech, and bird song suggests uniqueness of human scales," in *9th Folk Music Analysis Conference*, 2019, pp. 35–40.

[26] S. Sato, J. Six, P. Pfordresher, S. Fujii, and P. Savage, "Automatic comparison of global children's and adult songs supports a sensorimotor hypothesis for the origin of musical scales," in *9th Folk Music Analysis Conference*, 2019, pp. 41–46.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[28] D. A. Reynolds, "Gaussian mixture models." *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.

[29] A. P. Bell, *Dawn of the DAW: The studio as musical instrument*. Oxford University Press, 2018.

[30] L. Adu-Gilmore, "Embodied listening: Grassroots governance in electronic dance music venues in accra (ghana)," in *Electronic Cities*. Springer, 2021, pp. 243–260.

[31] E. Gómez and J. Bonada, "Automatic melodic transcription of flamenco singing," in *Fourth Conference on Interdisciplinary Musicology (CIM08)*, 2008.

[32] Á. H. Ingólfsson, *Jón Leifs and the musical invention of Iceland*. Indiana University Press, 2019.