COLLABORATIVE SONG DATASET (COSOD): AN ANNOTATED DATASET OF MULTI-ARTIST COLLABORATIONS IN POPULAR MUSIC

Michèle Duguay[1]   **Kate Mancey**[1]   **Johanna Devaney**[2]

[1] Department of Music, Harvard University
[2] Brooklyn College and Graduate Center, City University of New York
mduguay@fas.harvard.edu, johanna.devaney@brooklyn.cuny.edu

ABSTRACT

The Collaborative Song Dataset (CoSoD) is a corpus of 331 multi-artist collaborations from the 2010–2019 *Billboard* "Hot 100" year-end charts. The corpus is annotated with formal sections, aspects of vocal production (including reverberation, layering, panning, and gender of the performers), and relevant metadata. CoSoD complements other popular music datasets by focusing exclusively on musical collaborations between independent acts. In addition to facilitating the study of song form and vocal production, CoSoD allows for the in-depth study of gender as it relates to various timbral, pitch, and formal parameters in musical collaborations. In this paper, we detail the contents of the dataset and outline the annotation process. We also present an experiment using CoSoD that examines how the use of reverberation, layering, and panning are related to the gender of the artist. In this experiment, we find that men's voices are on average treated with less reverberation and occupy a more narrow position in the stereo mix than women's voices.

1. INTRODUCTION

As far back as the 1960s, *Billboard* charts have featured collaborations between independent acts. In recent years, however, the number of songs featuring a collaboration between artists has skyrocketed [1]. Part of this is due to the rising popularity of hip-hop in the 1980s, in which collaboration between different artists is a fixture. The 1986 version of "Walk This Way" by Aerosmith and Run DMC is an oft-cited example of such a collaboration. As Rose notes, the success of a collaboration between a hip-hop group (Run DMC) and a rock group (Aerosmith) "brought [hip-hop's] strategies of intertextuality into the commercial spotlight" [2, p. 51–52]. The 1990 success of "She Ain't Worth It" by Glenn Medeiros ft. Bobby Brown marked the first time a sung and rapped collaboration reached #1 on

Billboard's "Hot 100." Molanphy notes that during this period, multi-artist collaborations crystallized into two different frameworks: the "featured bridge rapper," and the "featured hook singer" [3]. Subsequently, tracks with one or more guest artist(s) have become a mainstay on the charts.

By 2021, over a third (39%) of the songs in *Billboard*'s "Hot 100" year-end chart credited more than one artist. Consider for instance "Save Your Tears," by singers The Weeknd & Ariana Grande, which occupied second place on the chart. A solo version of the song originally appeared on The Weeknd's album *After Hours* (2020). While this version achieved commercial success, the remix with Ariana Grande became a #1 single on the *Billboard* "Top 100" in May 2021 and became the longest-charting collaboration in *Billboard* "Hot 100" history. In the remix, Grande performs approximately half of the vocals, transforming the solo song into a dialogue between two characters. The collaboration between the two artists is responsible for the popularity of the remix, inviting both Grande's and The Weeknd's fans to stream, buy, and otherwise engage with the song. Several musicological studies have examined this relationship between collaborative songs and commercial success [4–6]. Other work has provided in-depth explorations of the musical characteristics of collaborative songs, with a particular focus on hip-hop [7–9].

Given the popularity of multi-artist collaborations, a more systematic exploration of their musical features is warranted. In this paper, we introduce the Collaborative Song Dataset (CoSoD), an annotated dataset that facilitates the study of various musical features in multi-artist collaborations. CoSoD provides metadata and analytical data for 331 multi-artist collaborations appearing on the *Billboard* "Hot 100" year-end charts between 2010 and 2019. The dataset also provides timed annotations on the song's formal structure, artists' gender, vocal delivery and pitch, and vocal production (reverberation, panning, and layering). As detailed in Section 2, the range of features included in the dataset makes it more broadly applicable for MIR research tasks. These include structural segmentation, vocal mixing, automatic music production, and examinations of gender in popular music. After outlining the contents of the dataset and the annotation methodology in Section 3, we present an experiment in Section 4 that examines the relationship between vocal production parameters and the gender of the performer in a subset of CoSoD.

2. RELATED WORK

CoSoD complements the growing list of annotated datasets that provide information on song structure in various popular music genres, e.g., [10–17], and is the first dataset to exclusively contain data on collaborative songs between independent acts. It can thus be used for training and evaluating structural segmentation tasks and for studying the specific structural characteristics of collaborative songs. CoSoD also complements existing datasets for multi-track mixing/analysis [18–23] and vocal analysis [24–26] by providing analytical annotations on the treatment of the voice in a mix.

In recent years, several studies have proposed tools and methods to automate the mixing of multi-track recordings [27, 28]. Such automatic production methods have various artistic and creative applications. One framework has been suggested to remix early jazz recordings, which are pre-processed using source separation then remixed with automatic production tools [29]. [30] proposes a prototype for an automatic DJ mixing system allowing for cross-fading via beat and tempo adjustment between songs. Studies on automatic mixing can be enhanced by knowledge of common mixing practices for specific instruments or sound sources. For instance, one study uses mixing practices that are consistent between mixing engineers to create a model that automatically mixes multiple drum tracks [31]. By focusing on vocals, which are a salient component of the mix in popular music [32], CoSoD provides a complementary approach to these studies on automated production. By providing annotations based on close listening of specific vocal mixing parameters in the different formal sections of a song, the dataset allows for the identification of trends in panning, layering, and use of artificial reverberation as they are applied to vocals in commercially successful post-2010 popular music. It enables the direct comparison of how various mixing parameters are applied to individual artists' voices within and across songs. In addition to facilitating the modeling of voice mixing, CoSoD also allows musicologists to ask questions about the way different voice types and individuals are mixed.

Finally, CoSoD facilitates the study of the relationship between gender and popular music. A number of previous studies have examined music programming and streaming services, exploring for instance how listeners tend to stream male artists more than women and mixed-gender groups [33]. Watson discusses gender inequality and low programming of women's music in country music radio [34]. Other work addresses how a listener's declared gender impacts automatic music recommendation [35] and musical preferences [36]. Additionally, various studies have addressed race and gender, along with sexist and racist discourses and practices, as they impact the music industry in general and the *Billboard* charts in particular [37–43]. By providing data on musical features, gender, and the role of these parameters within the formal structure of a song, CoSoD offers a new and complementary angle for the study of gender as it directly relates to the musical content of post-2010 popular collaborations.

3. COLLABORATIVE SONG DATASET (COSOD)

CoSoD[1] consists of metadata and analytical data of a 331-song corpus comprising all multi-artist collaborations on the *Billboard* "Hot 100" year-end charts published between 2010 and 2019. Each song in the dataset is associated with two CSV files: one for metadata and one for analytical data. We assembled the corpus by identifying every song on the charts that featured collaborations between two or more artists who usually perform independently from one another.

3.1 Annotation of Musical Features

The following analytical data is provided for each song in the dataset:

1. **Index number:** 1 to 33

2. **Time stamps:** In seconds (start of new section)

3. **Formal section label:** *Introduction, Verse, Pre-chorus, Chorus, Hook, Dance Chorus [44], Link, Post-chorus, Bridge, Outro, Refrain* or *Other*

4. **Name of artist(s):** Full name of the artist performing in each section. If all artists credited on the *Billboard* listing perform in a section, the label *both* or *all* is used.

Songs were assigned at random to one of two annotators, who generated time stamps at the onset of each formal section with Sonic Visualiser.[2] The annotators provided formal labels according to their analysis of the song. In case of ambiguity in the formal sections, both annotators discussed the analysis and agreed upon an interpretation.

For each formal section performed by *one artist only*, the following analytical data on the voice is provided:

1. **Gender of artist:** *M* (Man), *W* (Woman), *NB* (Non-binary)

2. **Function of artist:** *Feat* (Featured artist), *Main* (Main artist), *Neither, Uncredited*

3. **Style of vocal delivery:** *R* (Rapped vocals), *S* (Sung vocals), *Spoken*

4. **Minimum pitch value:** In Hz

5. **First quartile pitch value:** In Hz

6. **Median pitch value:** In Hz

7. **Third quartile pitch value:** In Hz

8. **Maximum pitch value:** In Hz

9. **Environment value:** On a scale of E1 to E5

10. **Layering value:** On a scale of L1 to L5

11. **Width (panning) value:** On a scale of W1 to W5

The annotators determined the name of the artist(s) performing in each section by ear, and using song lyric website Genius.com to validate their hearing. In cases where an artist only provides minimal background vocals (a few words) in a particular formal section, their name is not included. One annotator then provided analytical data on each formal section performed by one artist only. Data on gender was gathered from media interviews and social media statements from the artists, and matches the artist's gender identity at the time of the dataset creation. This methodology yielded three categories: man, non-binary, and woman. We understand these labels as umbrella terms that encompass a variety of lived experiences that intersect with race, sexuality, and other power structures. The style of vocal delivery was determined by ear. The distinction between rapping and singing is porous, with many vocalists adopting ambiguous modes of vocal delivery. We consider any formal section containing a melodic line performed with sustained pitches as sung.

The pitch data was obtained by first isolating the vocals from the full mix using Open-Unmix [45] and then running the pYIN Smoothed Pitch Track transform [46] on the isolated vocal file. The minimum, first quartile, median, third quartile, and maximum pitch points in each formal section were calculated and recorded in the dataset. [3]

The Environment, Layering, and Width values were determined by the first annotator to ensure consistency. Rather than attempting to reconstruct the mixing process itself, the annotations for these parameters represent the way a listener might perceive the final mix upon listening to it on stereo speakers. The Environment of a voice is the space in which the voice reverberates. Environment values were determined via an aural analysis of the full track by using the following scale [4] :

E1: The voice's environment sounds flat. There might be minimal ambiance added to the voice, but there is no audible echo or reverberation.

E2: The last word or syllable of most musical phrases is repeated through an echo or reverberation effect.

E3: The vocal line is repeated in one clear layer of echo. This added layer may be dry or slightly reverberant and has a lower amplitude than the main voice.

E4: The main voice is accompanied by a noticeable amount of reverberation. There is no clear echo layer, but rather a sense that the main voice is being reverberated across a large space.

E5: The main voice is accompanied by two or more layers of echo. The echo layers may be noticeably reverberant, similar in amplitude to the main voice, and difficult to differentiate from one another.

The Layering of a voice refers to the additional vocal tracks that are dubbed over a single voice. Layering values were determined via an aural analysis of the full track by using the following scale:

L1: The voice is presented as solo. Occasionally, a few words may be doubled with another vocal track for emphasis. Double-tracking is often used in the mixing process to create a fuller sound, with a final result sounding like a single vocal layer. Such cases fall into this category.

L2: The voice is presented as solo, but additional vocal layers are added at the end of musical phrases for emphasis.

L3: The main voice is accompanied by one or two layers. Layers might provide minimal harmonies or double the main voice. The layers have a noticeably lower amplitude than the main voice.

L4: The main voice is accompanied by two or more layers. These layers are close copies of the main voice, sharing the same pitch and similar amplitude.

L5: The main voice is accompanied by two or more layers. These layers add harmonies to the main voice, creating a thick and multi-voiced texture.

The Width of a voice refers to the breadth it occupies on the stereo stage. The Width was analyzed aurally with the aid of panning visualisation tool MarPanning [47]. The annotator simultaneously listened to the isolated vocal audio and observed the MarPanning visualization generated from the isolated vocals to determine the Width value. Since Open-Unmix occasionally omits reverberated components of the voice from the isolated file, the analyst then listened to the full track to confirm the Width value. Width values were determined according to the following scale:

W1: The voice occupies a narrow position in the center of the stereo stage.

W2: The voice occupies a slightly more diffuse position in the center of the stereo stage.

W3: The main voice occupies a narrow position in the center of the stereo stage, but some of its components (echo, reverberation, and/or additional vocal tracks) are panned toward the sides. These wider components have a lower amplitude than the main voice.

W4: The main voice occupies a slightly more diffuse position in the center of the stereo stage, and some of its components (echo, reverberation, and/or additional vocal tracks) are panned toward the sides. These wider components have a lower amplitude than the main voice.

W5: The main voice and its associated components (echo, reverberation, and/or additional vocal tracks) are panned across the stereo stage. All components have a similar amplitude.

---

[3] The accuracy of the F_0 estimates used to calculate this feature is impacted by the quality of the vocal source separation. A more accurate isolated vocal file would allow for more precise pitch data. Additionally, since pYIN Smoothed Pitch Track can only track a single melodic line, the accuracy of the pitch data is lessened in sections that feature multiple vocal layers with different pitch content.

[4] The scales were initially published in [9].

3.2 Metadata

The following metadata is provided for each song in the dataset:

1. **Index number:** From 1 to 331

2. **Year of first appearance on** *Billboard* **"Hot 100" year-end charts**

3. **Chart position:** As it appears on the *Billboard* "Hot 100" year-end charts

4. **Song title:** As it appears on the *Billboard* "Hot 100" year-end charts

5. **Name of artists:** As it appears on the *Billboard* "Hot 100" year-end charts

6. **Collaboration type:**

   - *Lead/featured:* Collab. with lead artist(s) and featured artist(s)
   - *No lead/featured:* Collab. with no determined lead
   - *DJ/vocals:* Collab. between a DJ and vocalist(s)

7. **Gender of artists:**

   - *Men:* Collab. between two or more men
   - *Women:* Collab. between two or more women
   - *Mixed:* Collab. between two or more artists of different genders

8. **Collaboration type + gender:**

   - *Collab M:* Collab. between men, no determined lead
   - *Collab M and W:* Collab. between men and women, no determined lead
   - *Collab NB and W:* Collab. betwen women and non-binary artists, no determined lead
   - *Collab W:* Collab. between women, no determined lead
   - *DJ with M:* Collab. between male DJ and male vocalist
   - *DJ with Mix:* Collab. between male DJ and mixed-gender vocalists
   - *DJ with NB:* Collab. between male DJ and non-binary vocalist
   - *DJ with W:* Collab. between male DJ and female vocalist
   - *M ft. M:* Men featuring men
   - *M ft. W:* Men featuring non-binary artist(s)
   - *W ft. M:* Women featuring men
   - *W ft. W:* Women featuring women

9. **MusicBrainz URL:** Link to the song on open music encyclopedia MusicBrainz

Each song in the dataset is labeled with an index number from 1 to 331. Songs are numbered in reverse chronological order, beginning with the 2019 charts and ending with 2010. One annotator obtained the metadata on year, chart
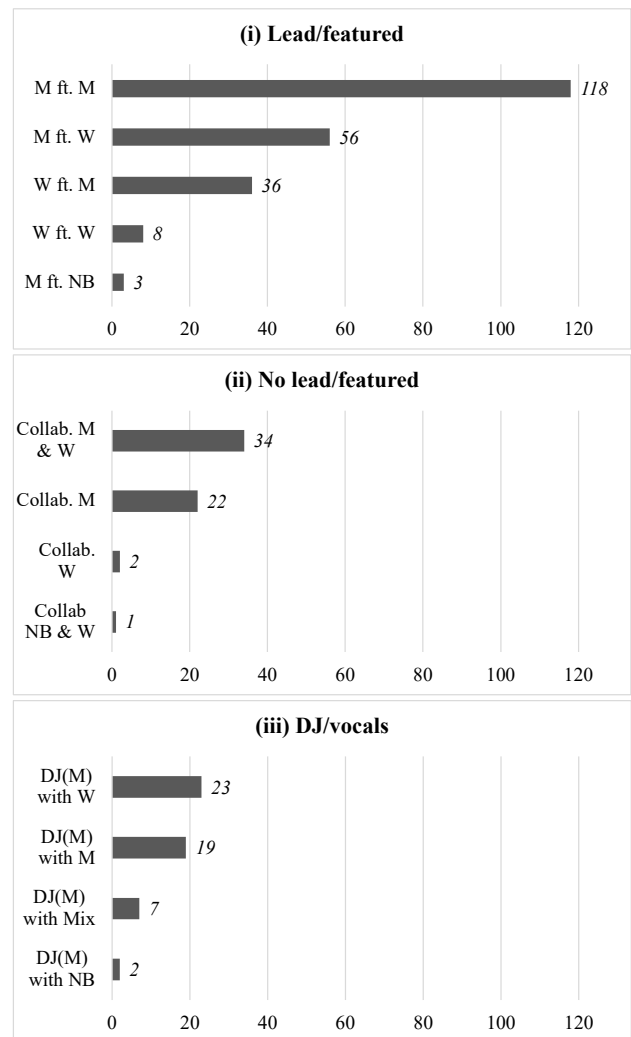

Figure 1. Summary of the gender distribution across different types of multi-artist collaborations. Subplot (i) shows gender counts for collaborations with lead and featured artists, subplot (ii) shows collaborations with no determined lead or featured artist, and subplot (iii) shows collaborations between DJs and vocalist(s).

position, title, and artists from the information available on the *Billboard* charts. Within years, songs are organized according to their position on the chart, from highest to lowest. Some songs appear on the charts two years in a row. In such cases, we only include the data for the earliest appearance.

3.3 Corpus Statistics

The dataset can be divided into three categories (shown in Figure 1): (i) collaborations between the lead artist(s) and featured artist(s), which account for 221, or 66.7% of the tracks, (ii) collaborations with no determined lead or featured artist, which account for 59, or 17.8%, of the tracks, and (iii) collaborations between a DJ and a vocalist, which account for 51, or 15.4% of the tracks. In category (i), the lead artist usually performs the majority of vocals. For example, in "No Limit" (2018) by G-Eazy ft. A\$AP Rocky & Cardi B, G-Eazy performs most of the vocals. A\$AP
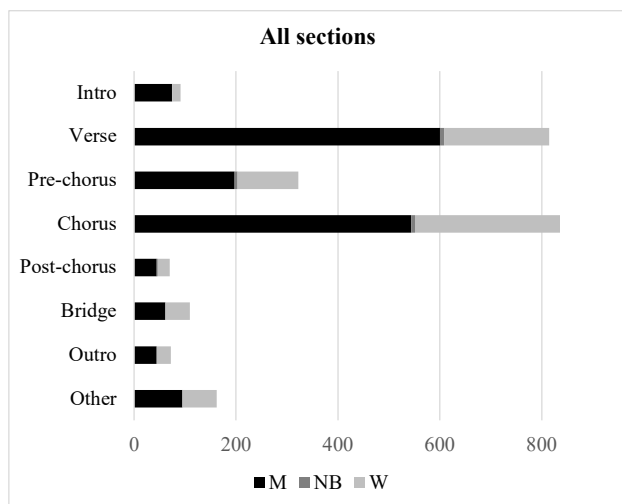
Figure 2. Number of formal sections performed by a single artist (main, featured, or neither), categorized according to formal section type

Figure 3. Number of formal sections performed by featured artists, categorized according to formal section type

Rocky accompanies him in the chorus and Cardi B raps the second verse. In category (ii), the performance of the vocals is often more equally distributed. Such collaborations are often billed as "duets," and the artists' names are separated by a "+", a "&", or a comma on the *Billboard* charts. For example, "Something' Bad" (2014) is labeled as a "Miranda Lambert Duet With Carrie Underwood." Both vocalists perform approximately equal portions of the song. In category (iii), the DJ does not provide vocals. In "Sweet Nothing" (2012), for instance, only the featured Florence Welch sings. The voice of DJ Calvin Harris is not heard.

Mixed-gender collaborations (including any combination of non-binary, women, and men artists) frequently appear on the Billboard charts and account for 162, or 49%, of the tracks in the dataset. Collaborations between two or more men account for 159 tracks, or 48% of the dataset. Finally, collaborations between women account for 10, or 3%, of the tracks. In six of the ten years under study–2011, 2012, 2015, 2017, 2018, and 2019–no collaborations between women reached the *Billboard* "Hot 100" year-end chart. Conversely, songs with two or more male vocalists were a consistent fixture on the charts. Mixed-gender collaborations, with any combination of men, women, and non-binary artists within the same track, also frequently appear on the charts.

Figure 2 shows the number and type of sections performed by individual artists in the corpus, categorized according to gender. This figure includes identical sections (such as choruses) that are repeated within a song. Sections in which more than one artist performs are not included. More sections are performed by men than by women and non-binary artists, which is to be expected given the over-representation of men in the dataset as a whole (Figure 1). Figure 3 displays the number and type of sections performed by *featured* artists only.
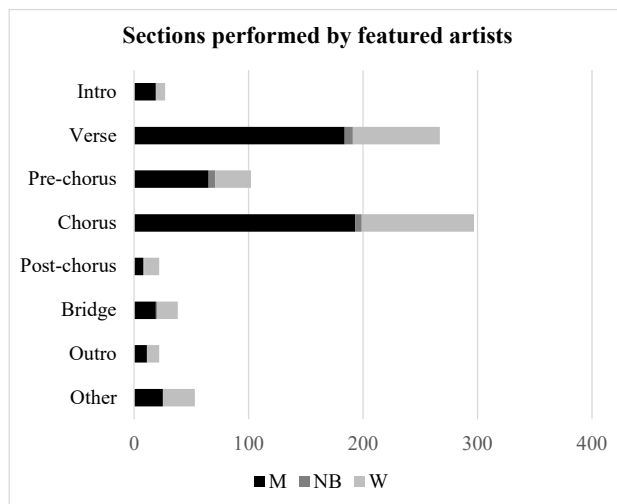
4. EXPERIMENT: VOCAL PRODUCTION FEATURES AND GENDER

This section examines the relationship between the gender of an artist and the treatment of their voice, as characterized by three of the annotated musical features in the dataset: Environment, Layering, and Width. For the purposes of statistical power in the experiment, only songs with men and/or women artists were included. We only included tracks that contained verse and chorus sections to remove section types that occur in only a few tracks. In order to avoid over-representations of tracks with repeated sections (i.e., several instances of the same chorus), we sampled the first verse and chorus performed by a single artist from each track.[5] This method resulted in the inclusion of two sections from 287 of the 331 dataset tracks in the experiment.

We analyzed the data with three separate logistic regressions–one for each feature–using the statsmodels package in Python. We encoded the different levels of the parameter scales (defined in Section 3.1) with one-hot encoding in order to allow us to examine whether there is a correspondence between specific parameter scale levels and gender.

Of the three logistic regressions, Environment ($R^2_{McFadden}$ (4, $N = 574$) = 0.028, $p < 0.0001$) and Width ($R^2_{McFadden}$ (4, $N = 574$) = 0.035, $p < 0.0001$) were statistically significant, while Layering ($R^2_{McFadden}$ (4, $N = 574$) = 0.0036, $p = 0.64$) was not. The McFadden R^2 values for both Environment and Width were very low. This was not surprising since we did not anticipate that these features, particularly in isolation, would be explanatory. We were instead interested in exploring whether there is a significance between these features with respect to the man/woman gender binary in these collaborations.

For Environment, there were significant effects ($p <$

---

[5] If the first verse of a song was performed by two artists simultaneously, while the second verse was only performed by one, we sampled the second verse.

0.0001) for E1 (β=-1.18, 95%CI [-1.49, -0.87]), E2 (β=-1.12, 95%CI [-1.56, -0.69]), and E3 (β=-0.78, 95%CI [-1.14, -0.42]). There was a significant negative effect for the lower/mid-level environment values and gender, meaning that men's voices are more likely to be set in less reverberant spaces than women's voices. For Width, there were significant effects at all of the levels: W1 (β=-1.84, 95%CI [-2.50, -1.17]), W2 (β=-1.58, 95%CI [-2.39, -0.77]), W3 (β=-1.13, 95%CI [-1.51, -0.75]), W4 (β=-0.47, 95%CI [-0.77, -0.17]), and W5 (β=-0.60, 95%CI [-0.95, -0.25]).

The Width results are harder to interpret than the Environment ones because the coefficient values are smaller and all negative. This is likely due to the imbalance between men and women in featured artist roles, both in the dataset (see Figure 1) overall and in the sample used in this experiment (404 of the included sections featured men while only of 170 featured women). However, the overall trend is similar to the one in the Environment experiment: lower-level values are more common for men than women. Men's voices are more likely to occupy a narrow, centered position on the stereo stage, while women's voices are more likely to occupy a wider space. These results were expected given that high Environment values tend to be associated with high Width values, as the reverberated components of a voice are generally panned across the stereo stage.

The lack of significant results for Layering indicates that there are no differences in the ways in which this parameter is applied to men's and women's voices. Since textural variation (such as the addition of vocal layers) is a standard feature of verse-chorus form, it is possible that Layering is linked to the type of formal section rather than to the gender of the vocalist. The significant results for the Environment and Width parameters can be interpreted in light of Brøvig-Hanssen's and Danielsen's work on technological mediation [48]. The authors establish a distinction between transparent and opaque technological mediation in recorded music. Transparent mediation, on one hand, is meant to create a recorded product that sounds natural and unaltered. Low Environment and Width values, for instance, are closer to transparent mediation because they sound closer to a real-life performance that is unmediated with artificial reverb or panning. Opaque mediation, on the other hand, highlights the use of technology by making it obvious to the listener. High Width and Environment values, with their clearly audible artificial reverberation and wide panning, are examples of opaque mediation. The results of the experiment therefore suggest that men's voices are more likely to be mixed to sound "transparent" and natural while women's voices are more likely to be mixed to sound "opaque" and technologically mediated.

Overall, this experiment demonstrates that within verse and chorus sections in CoSoD, there is a significant difference between the treatment of men's and women's vocals in terms of Environment and Width. This suggests that some mixing parameters contribute to the sonic differentiation of men's and women's voices in popular music.

5. CONCLUSION

CoSoD is a 331-song corpus of all multi-artist collaborations for faciliating appearing on the 2010–2019 *Billboard* "Hot 100" charts. Each song in the dataset is annotated with metadata, formal sections, and aspects of vocal production (including reverberation, layering, panning, and gender of the artists). As outlined in Section 2, CoSoD has several implications for MIR research. It provides annotated data for structural segmentation tasks and a listener-centered perspective on vocal mixing that could be useful for automatic music mixing tasks. The dataset could also be used to determine how these parameters interact with song form. Further study could also examine the relationship between the vocal range of an artist in a given section, their type of vocal delivery (rapped, spoken, or sung), and mixing parameters. Finally, the dataset also allows for the examination of the ways in which Environment, Layering, and Width values tend to be grouped together to create specific vocal production effects.

The dataset also facilitates musicological study of multi-artist collaborations post-2010 and gender norms. The experiment in Section 4 demonstrates this, as its results suggest that, for the chorus and verse data sampled from 287 songs in the dataset, men's voices are more likely to be narrow and less reverberated than women's. Opportunities for future research include examining whether there is a significant difference in the way Environment, Width, Layering, or other parameters are applied to women's and men's voices *within* collaborations that feature mixed- and same-gender vocalists. In other future work, we plan on expanding the annotations in the dataset with time-aligned lyrics, harmonic analyses, and additional performance data for the voice extracted using AMPACT [49, 50]. These annotations will include both spectral features and semantic descriptors, and the data will be encoded in relation to vocal-line transcriptions, where possible [51]. We also plan on providing annotations on vocal production parameters in sections performed by multiple artists and examining how vocal production parameters correlate with mixing parameters such as panning.

Finally, while our dataset focuses on gender, we are also interested in encoding other aspects of identity, such as race, in order to provide an intersectional perspective on artists' identities. However, categorizing artists according to race proves to be more problematic than gender. Matthew D. Morrison writes that "white (and other non-black) people freely express themselves through the consumption and performance of commodified black aesthetics without carrying the burden of being black under white supremacist structures" [52, p. 791]. In other words, white and non-Black artists–such as rappers Iggy Azalea and G-Eazy, or singer Bruno Mars–often assume particular sonic characteristics that implicitly associate them with commodified notion of Blackness. By categorizing all white artists together, for instance, we would ignore this phenomenon and the way it is sonically realized. Further work needs to be done to understand how to best expand on CoSoD, or datasets in general, to account for this dynamic.

6. REFERENCES

[1] Anonymous, "In popular music, collaborations rock," *The Economist*, February 2018. [Online]. Available: https://www.economist.com/business/2018/02/03/in-popular-music-collaborations-rock

[2] T. Rose, *Black Noise: Rap Music and Black Culture in Contemporary America*. Hanover, NH: Wesleyan University Press, 1994.

[3] C. Molanphy, "Feat. don't fail me now: The rise of the featured rapper in pop music," *Slate*, July 2015.

[4] A. Ordanini, J. C. Nunes, and A. Nanni, "The featuring phenomenon in music: How combining artists of different genres increases a song's popularity," *Marketing Letters*, vol. 29, no. 4, pp. 485–499, Dec 2018.

[5] M. Silva and M. Moro, "Causality analysis between collaboration profiles and musical success," in *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, 10 2019, pp. 369–376.

[6] G. P. Oliveira, M. O. Silva, D. B. Seufitelli, A. Lacerda, and M. M. Moro, "Detecting collaboration profiles in success-based music genre networks," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 2020, pp. 726–732.

[7] R. Komaniecki, "Analyzing collaborative flow in rap music," *Music Theory Online*, vol. 23, no. 4, 12 2017.

[8] B. Duinker, "Song form and the mainstreaming of hip-hop music," *Current Musicology*, vol. 107, pp. 93–135, 1 2020.

[9] M. Duguay, "Analyzing vocal placement in recorded virtual space," *Music Theory Online*, vol. 28, no. 4, 2022.

[10] A. Berenzweig, B. Logan, D. Ellis, B. Whitman, and C. A, "A large-scale evaluation of acoustic and subjective music similarity measures," *Computer Music Journal*, vol. 28, 11 2003.

[11] C. Harte, "Towards automatic extraction of harmony information from music signals," Ph.D. dissertation, Queen Mary, University of London, august 2010.

[12] J. Burgoyne, J. Wild, and I. Fujinaga, "An expert ground truth set for audio chord recognition and music analysis." 01 2011, pp. 633–638.

[13] J. Smith, J. Burgoyne, I. Fujinaga, D. De Roure, and J. Downie, "Design and creation of a large-scale database of structural annotations," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, Miami, FL, USA, 01 2011, pp. 555–560.

[14] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset," in *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, 2011.

[15] T. de Clerq and D. Temperley, "A corpus analysis of rock harmony," *Popular Music*, vol. 30, no. 1, pp. 47–70, 2011.

[16] F. Bimbot, G. Sargent, E. Deruty, C. Guichaoua, and E. Vincent, "Semiotic description of music structure: an introduction to the quaero/metiss structural annotations," in *Proceedings of the AES International Conference*, London, UK, 01 2014.

[17] O. Nieto, M. McCallum, M. Davies, A. Robertson, A. Stark, and E. Egozy, "The Harmonix Set: Beats, Downbeats, and Functional Segment Annotations of Western Popular Music," in *Proceedings of the 20th International Society for Music Information Retrieval Conference*, Delft, The Netherlands, 2019, pp. 565–572.

[18] C.-L. Hsu and J.-S. R. Jang, "On the improvement of singing voice separation for monaural recordings using the mir-1k dataset," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 310–19, 2010.

[19] R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, and J. Bello, "Medleydb: A multitrack dataset for annotation-intensive mir research," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, Taipei, Taiwan, 10 2014.

[20] T.-S. Chan, T.-C. Yeh, Z.-C. Fan, H.-W. Chen, L. Su, Y.-H. Yang, and R. Jang, "Vocal activity informed singing voice separation with the ikala dataset," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, 2015, pp. 718–722.

[21] R. Bittner, J. Wilkins, H. Yip, and J. Bello, "Medleydb 2.0: New data and a system for sustainable data collection," in *Extended abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference*, New York, NY, USA, 2016.

[22] B. D. Man and J. D. Reiss, "The mix evaluation dataset," in *Proceedings of the 20th International Conference on Digital Audio Effects*, Edinburgh, UK, 9 2017, pp. 436–442.

[23] Z. Rafii, A. Liutkus, F.-R. Stöter, S. I. Mimilakis, and R. Bittner, "The MUSDB18 corpus for music separation," Dec. 2017.

[24] R. Gong, R. C. Repetto, and X. Serra, "Creating an a cappella singing audio dataset for automatic jingju singing evaluation research," in *Proceedings of the 4th International Workshop on Digital Libraries for Musicology*, Shanghai, China, 2017, pp. 37–40.

[25] C.-i. Wang and G. Tzanetakis, "Singing style investigation by residual siamese convolutional neural networks," in *2018 IEEE International Conference on*

Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, 2018, pp. 116–120.

[26] J. Wilkins, P. Seetharaman, A. Wahl, and B. Pardo, "Vocalset: A singing voice dataset," in *Proceedings of the 19th International Society for Music Information Retrieval Conference*, Paris, France, 2018, pp. 468–474.

[27] B. D. Man and J. D. Reiss, "Ten years of automatic mixing," in *Proceedings of the 3rd Workshop on Intelligent Music Production*, Salford, UK, 2017.

[28] M. A. Martínez-Ramírez, W.-H. Liao, G. Fabbro, S. Uhlich, C. Nagashima, and Y. Mitsufuji, "Automatic music mixing with deep learning and out-of-domain data," in *Proceedings of the 23rd International Society for Music Information Retrieval Conference*, 2022.

[29] D. Matz, E. Cano, and J. Abeßer, "New sonorities for early jazz recordings using sound source separation and automatic mixing tools," in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, Málaga, Spain, 2015, pp. 749–755.

[30] H. Ishizaki, K. Hoashi, and Y. Takishima, "Full-automatic dj mixing system with optimal tempo adjustment based on measurement function of user discomfort." in *Proceedings of the 10th International Society for Music Information Retrieval Conference*, Kobe, Japan, 01 2009, pp. 135–140.

[31] J. J. Scott and Y. E. Kim, "Instrument identification informed multi-track mixing," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, Curitiba, Brazil, 2013, pp. 305–310.

[32] A. L. Knoll and K. Siedenburg, "The optimal mix? presentation order affects preference ratings of vocal amplitude levels in popular music," *Music & Science*, vol. 5, pp. 1–12, 12 2022.

[33] A. Epps-Darling, H. Cramer, and R. Takeo Bouyer, "Artist gender representation in music streaming," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 10 2020, pp. 248–254.

[34] J. Watson, "Programming inequality: Gender representation on canadian country radio (2005-2019)," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 2020, pp. 392–399.

[35] G. Vigliensoni and I. Fujinaga, "Automatic music recommendation systems: Do demographic, profiling, and contextual features improve their performance?" in *Proceedings of the 17th International Society for Music Information Retrieval Conference*, New York, NY, USA, 2016, pp. 94–100.

[36] A. Laplante, "Improving music recommender systems: What can we learn from research on music tastes?" in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, Taipei, Taiwan, 2014, pp. 451–456.

[37] C. L. Keyes, "Empowering self, making choices, creating spaces: Black female identity via rap music performance," *The Journal of American Folklore*, vol. 113, no. 449, pp. 255–69, 2000.

[38] T. J. Dowd and M. Blyler, "Charting race: The success of black performers in the mainstream recording market, 1940 to 1990," *Poetics*, vol. 30, pp. 87–110, 2002.

[39] R. N. Bradley, *Barbz and Kings: Explorations of Gender and Sexuality in Hip Hop*. Cambridge, England: Cambridge University Press, 2015, pp. 181–191.

[40] M. Lafrance, C. Scheibling, L. Burns, and J. Durr, "Race, gender, and the billboard top 40 charts between 1997 and 2007," *Popular Music and Society*, vol. 41, no. 5, pp. 522–538, 2018.

[41] K. J. Lieb, *Gender, Branding, and the Modern Music Industry: The Social Construction of Female Popular Music Stars*, 2nd ed. New York, NY: Routledge, 2018.

[42] J. E. Watson, "Gender on the billboard hot country songs chart, 1996–2016," *Popular Music and Society*, vol. 42, no. 5, pp. 538–60, 2002.

[43] C. Bauer and J. Devaney, "Constructing gender in audio: exploring how the curation of the voice in music and speech influences our conception of gender identity," in *Mediale Stimmentwürfe: perspectives of media voice designs*, ser. Schriftenreihe zur digitalen Gesellschaft NRW, M. Erbe, A. Riffi, and W. Zielinski, Eds. Munich, Germany: kopaed Verlag, 2022, vol. 7, pp. 83–100.

[44] A. Barna, "The dance chorus in recent top-40 music," *SMT-V*, vol. 6, no. 4, June 2020. [Online]. Available: http://doi.org/10.30535/smtv.6.4

[45] F.-R. Stöter, S. Uhlich, A. Liutkus, and Y. Mitsufuji, "Open-unmix - a reference implementation for music source separation," *Journal of Open Source Software*, 2019.

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

[47] K. McNally, G. Tzanetakis, and S. R. Ness, "New tools for use in the musicology of record production," *Unpublished Paper, University of Victoria.*, 2009.

[48] R. Brøvig-Hanssen and A. Danielsen, *The Impact of Digitization on Popular Music Sound*. Cambridge, MA, USA: The MIT Press, 2016.

[49] J. Devaney, M. I. Mandel, and I. Fujinaga, "A study of intonation in three-part singing using the automatic music performance analysis and comparison toolkit (ampact)." in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 2012, pp. 511–516.

[50] J. Devaney and M. Mandel, "Score-informed estimation of performance parameters from polyphonic audio using ampact," in *Extended abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference*, 2016.

[51] J. Devaney, "Using note-level music encodings to facilitate interdisciplinary research on human engagement with music," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, 2020.

[52] M. D. Morrison, "Race, blacksound, and the (re)making of musicological discourse," *Journal of the American Musicological Society*, vol. 72, no. 3, pp. 781–823, 12 2019.