A JOURNAL OF THE SOCIETY FOR MUSIC THEORY

Response to Trevor de Clercq (2025)

Nicholas Shea, Lindsey Reymore, Christopher Wm. White, Ben Duinker, Leigh VanHandel, Matthew Zeller, and Nicole Biamonte

REFERENCE: https://www.mtosmt.org/issues/mto.25.31.4/mto.25.31.4.declercq.php

DOI: 10.30535/mto.31.4.10

Volume 31, Number 4, December 2025
Copyright © 2025 Society for Music Theory

Introduction

[0.1] We appreciate the time and effort that Trevor de Clercq has dedicated to commenting on our article "Diversity in Music Corpus Studies." A number of productive points arise from his response, including the importance of distinguishing music-corpus type and a potential alternate method for relating demographic information to genre (see de Clercq 2025, [3.4]). However, we remain in fundamental disagreement on a number of points regarding both corpus construction and sampling methods.

[0.2] At the broadest level, de Clercq misrepresents our goals and, by extension, our methods, contending that our application of the Anti-Discriminatory Alignment System (ADAS) to the Timbre in Popular Song (TiPS) corpus creates a corpus of *music as imagined* in contrast to the more-typical *music as heard* or *music as produced* paradigms. As we will demonstrate, the TiPS corpus meets the standards for a "music as heard" corpus. Our response begins by reiterating our goals and methods in our use of the ADAS to construct the TiPS corpus. We focus on three topics: ways to operationalize a representative population, the role of differing methods in supporting converging evidence, and the myth of objectivity in corpus studies. We then close with a discussion of these concepts in practice, to again underscore that corpus-building tactics can still maintain statistical rigor even when moving beyond existing corpus-selection practices.

1. Ways to establish a statistical population outside of listenership

[1.1] As de Clercq rightly notes, corpora can be designed to capture the experiences of listening to or creating music, or what he calls "music as heard" and "music as produced." As he defines it, the statistical population of music as heard is "the complete set of encounters between a group of listeners and a song during a particular period of time, i.e., a set of listening experiences" [2.3], whereas the statistical population of music as produced is "all the songs written and recorded in some musical style during some window of time" [3.1]. He errs, however, when he portrays these two categories as having inflexible ideal implementations. Building a corpus requires operationalization; researchers must make decisions about how to capture listeners' experience, which artists to focus on, and so on.

[1.2] With the sample implementation of the Anti-Discriminatory Alignment System (ADAS) presented in our original article, we created corpora by selecting popular tracks in some genre, such that the final list is adjusted toward better representation of the overall demographics of the United States. The resulting corpus, then, contains tracks in a given musical genre that all meet the criteria of achieving popular success but that come from artists with a more diverse range of backgrounds and identities than lists based on critical acclaim or commercial success. As stated in our article, we agree with de Clercq that the resulting corpora are not precise models of which songs were most frequently heard and that they do not reflect the complete diversity of amateur and non-commercial artists working in those genres. The ADAS was not designed around either of these goals. However, the resulting corpus *does* allow analysts to study how artists with different identities and backgrounds achieve popular success by expressing a genre through a hit song or songs.

[1.3] When we adjust a heavy-metal corpus consisting primarily of white men to include more songs performed by women- and non-white-fronted bands, we are better characterizing what "heavy metal" means to a wider group of commercially successful musicians. Our method offers one way to balance measures of commercial popularity, which are often influenced by hidden biases, with the diversity of musicians participating in a genre at any point in time. As our review of genre scholarship in the initial article illustrates ([8.3–8.9]), genre is not just a top-down way of grouping music, as practiced by commercial entities; it is also a bottom-up amalgamation of the practices of its creators and listeners. Top-down commercial and social forces often marginalize artists whose identities do not align with corporate priorities. Our bottom-up methodology helps to counteract these forces.

The Statistical Population of the TiPS Corpus

[1.4] Our implementation of the ADAS sifts through the landscape of popular tracks to build a corpus that is both reflective of popularity and more representative of demographic diversity within a genre when compared to unaltered convenience samples. In parallel, our operationalization of popularity allows us to meet our goal of creating cohesive corpora for multiple genres, while simultaneously controlling for the innate variability, and ambiguity, of artist selection when using convenience samples such as "best of" lists to form a corpus. We begin with an initial convenience sample, adjust for demographic representation, and then sample the remainder of the child corpus via random selection. When de Clercq asks, "What, then, is the statistical population that Shea et al. are attempting to represent with the ADAS?" ([1.6]), his question is overly broad when considering the many ways researchers can choose to adopt the ADAS. However, we can provide a straightforward answer to the specific question of statistical population represented by the Timbre in Popular Song corpus (TiPS): tracks that are popular within a genre at a given moment in time—songs that constitute some of the most widely heard music in that style.

[1.5] By de Clercq's assessment, the ADAS disrupts both a purely random sample *and* one based on the very highest-rated or most-popular songs by inserting tracks from artists with underrepresented identities. However, any "most-popular," "most-played," or "greatest-of-all-time" list is a panoramic landscape of a population's listening practice. If a corpus substitutes, say, the fifth most-heard track with the twenty-fifth in order to include an artist from an underrepresented background, that list still reflects what is *popular* in a particular time and place. Likewise, if an analyst supplements a "best of heavy metal" list with selections from, say, "most-streamed women-led metal bands," that corpus continues to capture what is *popular* in heavy metal. As we argue in our article, mainstream commercial popularity and critical acclaim are not neutral—they are shaped by forces that systematically favor bands with certain identities and demographics. Yet, our operationalization of popularity still captures mainstream listening experiences particular to a genre and time period.

[1.6] Listenership is an alternative way to operationalize popularity, as applied by de Clercq in [2.3] of his commentary. However, this metric is also not without its issues: in arguing that Luminate data would provide an "accurate" source for sampling music as heard, de Clercq acknowledges neither the interpretive implications of his specification of the statistical population nor the implicit assumptions and limitations of this approach. A tally of total plays or streams treats all encounters as equal, ignoring qualitative differences, the number of listeners, and other key contextual factors. For example, say both Song A and Song B had 1000 plays, but song A was heard once by 1000 people and song B was heard 100 times by 10 people. Is there a functional difference in terms of "listenership"? Or popularity? Does it make a difference if song A's plays were the result of a playlist set on repeat in a big-box store, with each play heard—but not attended to—by dozens of people, or if Song B's plays were shared among friends at album-listening parties? There are no definitive answers, as each scenario speaks to some senses of the term "popular" but not others. Importantly, these questions are not meant to undermine the clear usefulness of plays or streams as a measure of popularity, but rather to illustrate that any operationalization is a necessary reduction for practical purposes, no matter how objective it may appear at first glance.(1)

[1.7] Parallel to our discussion of population, de Clercq also misconstrues our motivations for using, and thus our application of, demographic data. In [4.2], de Clercq expresses skepticism about "whether the authors in fact achieved their stated goal of making their corpora more representative of the overall population of the United States," noting that our adjustments only apply to marginalized groups. Here, de Clercq's reading diverges notably from the intent we explicitly articulate throughout the article; he chooses one remark of many on this topic. We concede that the sentence in [5.5] should have also included a specification about our focus on historically marginalized artists; yet de Clercq's criticism clearly overlooks the many times we establish that our goal is to increase representation of artists from historically marginalized identities, for example, as in [5.2], elsewhere in [5.5], and [5.6]. In [4.4] of his commentary, de Clercq writes that the "real goal appears to be to increase the proportion of female and BIHAP artists in a corpus, no matter what the original distribution, whenever it falls below a population target." We state this goal clearly in [5.7]: "In our current implementation, because one of our goals is greater representation of marginalized identities, we exceeded benchmarks in some contexts." Despite this, de Clercq's point draws attention to the fact that more discussion of this particular application is warranted. Doing so, however, is presently outside the scope of our response.

2. *Contrasting methods and converging evidence*

[2.1] Creating a corpus requires researchers to make many decisions. Some of these are made explicitly by the researcher, but some are made by other agents and are often not acknowledged. Although convenience samples can help mitigate *researcher* bias, they do not eliminate all types of bias. Rather, they typically shift bias from the researcher to some other source. For example, the content of the Rolling Stone 200 list, which has been used by multiple corpus studies to represent popular music, was shaped by the *Rolling Stone Magazine* editors who invited raters to participate, and by the raters themselves. Convenience samples may be practical and accessible, but they almost always remain subject to some type of bias, possibly including hidden prejudices such as racism or sexism.

[2.2] It is beneficial to step back and consider the role of any individual empirical study in the context of scholarship. No claim or theory is established by a single observation or analysis. Individual studies should *always* be interpreted with caution—specific operational definitions and sampling methods inevitably limit generalizability, and false positives and negatives are possible. Knowledge is more reliable when multiple approaches to testing an idea point to the same interpretation. Thus, the best support for a claim or theory comes from converging evidence. If different perspectives on the same trend support the same conclusion, then we can be more confident in the result. Converging evidence can come from different disciplines, cross-cultural research, or diverse methods, or as in the case of the ADAS, from varied operationalizations and sampling methods (Huron 2018).

[2.3] De Clercq expresses concern that:

> "a corpus created using the ADAS may distort or misrepresent a musical style, such that it is not clear whether research findings from a corpus created using the ADAS have much if any explanatory power beyond the limited scope of the corpus itself. In other words, the authors' recalibration of the corpus study may strip it of any value as a research tool" [1.3].

We address this concern directly and in detail in the original article (Shea et al. 2024, [4.1–4.5] and [8.1–8.9]) and will not reiterate those arguments here. However, it is pertinent to consider how we would interpret results from an ADAS-adjusted corpus that differed from those of a comparable, non-adjusted corpus from the standpoint of converging evidence.

[2.4] The ADAS-adjusted corpus provides detailed data on artist identity, allowing for the study of whether and how artist identity interacts with musical features. If the musical parameters of the songs represented in the ADAS-adjusted corpus do not significantly differ from a non-adjusted corpus, we would expect results of similar queries to converge. For example, if artists of all genders within a genre tend to use the same instrumentation, then there would be no interaction between gender and instrumentation, and the results from the ADAS-adjusted corpus would converge with those from other corpora.

[2.5] If results do *not* converge, it may be that artist identity is interacting with a musical feature in a way that is not evident in a non-adjusted corpus. Such cases merit further study, and previous theories may need to be refined. For instance, some interactions are historically plausible, as with the relationship between harmonic complexity and race discussed in Shea et al. 2024, [4.6–4.13]). It is therefore possible that an ADAS-adjusted corpus may result in selection bias in the statistical sense. However, if and where there are interactions between musical features and the detailed demographic information encoded in the ADAS-adjusted corpus, we can easily identify patterns through comparison of subgroups and with results from other studies. These patterns would then yield hypotheses that can be tested directly in a new corpus designed for such a purpose. This possibility presents rich opportunities for increasing our understanding of popular music in general, as well as how artist identity may influence musical features both within and across genres.

3. *The inherent subjectivity of corpus studies*

[3.1] The notion that the TiPS corpus may have limited explanatory power beyond the scope of the corpus itself (de Clercq 2025, [1.3]) requires further unpacking. Foremost, because every corpus is tethered to the cultural and social contexts of the creators and listeners (as we show; see also Covach 2022, Huron 2013, Laybourn 2018, and McDermott 2021), *every* corpus has limited explanatory power. Additionally, subjective choices are made at every stage of research, including decisions about corpus size and scope, data encoding annotation, analytical methods, and interpretation of the results. De Clercq himself acknowledges this elsewhere: "While statistical analyses ostensibly reflect an objective view of rock, it will be argued that significant subjectivity underlies the process" (de Clercq 2020, 150).

[3.2] In practice, de Clercq suggests an alternate methodology that targets a specific demographic before the sampling stage, when defining the population to study—for example, focusing a study on Black country artists.[(2)] This is certainly a viable approach. However, different studies may have contrasting goals and their sampling strategies should accordingly align with those goals. Speaking to this possibility, Alan Marsden notes that

> " . . . even when we have a clear idea of the population from which a sample should be drawn . . . we can see that there is no such thing as a 'representative sample' in the abstract. A sample is only representative by reference to the distribution of certain specified characteristics in the population, and it is usefully representative to the extent that those characteristics are the ones which relate to the phenomena of interest" (Marsden 2022).

In this respect, we would suggest that any time a corpus compiler makes a decision about sampling, it reflects some sort of goal or priority. No decision is, in of itself, purely *objective* and devoid of bias and predispositions, regardless of its position in the corpus-compilation process.[(3)]

Conclusion

[4.1] De Clercq's response seems to interpret our position as presenting a singular imperative method. However, in our article, we situate the ADAS as one of many possible approaches to addressing the problem of marginalization in corpus studies ([8.10]) and clarify that the ADAS is not an appropriate tool for all research questions ([8.11]). Yet, we *do* believe that research should consider and respond to the issues we raise. While the ADAS provides one option, other options might include intentional selection of convenience samples, some other method of adjustment, or, as suggested by de Clercq, the choice to study music from a particular demographic population of artists at the outset. Depending on the study and on the choice of sample, it may be most appropriate to sample without adjustment and to acknowledge the relevant issues, or to include a discussion of how social bias, as it relates to corpus construction, may or may not affect interpretation of the results. As we state, "by raising these issues, we are not accusing any individual corpus analysts of discrimination, racism, or misogyny. Our project is simply to raise awareness around particular latent issues in music-data analysis, and to suggest some ways to address these issues" ([1.7]). Our practical response, then, is to introduce the ADAS as a model that intentionally includes more artists with marginalized identities so as not to perpetuate existing biases. It is a non-prescriptive approach for our specific goals. We do, however, expect that others may share this goal and should feel free to adopt (and adapt) the ADAS as they see fit.

Nicholas Shea
Arizona State University
50 E Gammage Pkwy
Tempe, AZ 85257
njshea@asu.edu

Lindsey Reymore
Arizona State University
McGill University
50 E Gammage Pkwy
Tempe, AZ 85257
lreymore@asu.edu

Christopher Wm. White
UMass Amherst
151 Presidents Dr
Amherst, MA 01003
cwmwhite@umass.edu

Ben Duinker
Schulich School of Music
McGill University
555 Sherbrooke St. West
Montreal QC, Canada
H3A 1E3
benjamin.duinker@mail.mcgill.ca

Leigh VanHandel
University of British Columbia
6361 Memorial Rd
Vancouver, BC V6T 1Z2

leigh.vanhandel@ubc.ca

Matthew Zeller
Musical Instrument Museum
4725 E. Mayo Blvd.
Phoenix, Arizona 85050
matthew.zeller@mim.org

Nicole Biamonte
McGill University
Schulich School of Music
555 rue Sherbrooke O.
Montreal, QC H3A 1E3
nicole.biamonte@mcgill.ca

Works Cited

Burgoyne, John Ashley, Jonathan Wild, and Ichiro Fujinaga. 2011. "An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis." In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, ed. Anssi Klapuri and Colby Leider, 633–38.

Covach, John. 2022. "Corpus Formation in Popular Music Scholarship: History, Style, Form, and Data." In *The Oxford Handbook of Music Corpus Studies*, ed. Daniel Shanahan, John Ashley Burgoyne, and Ian Quinn. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190945442.013.29.

de Clercq, Trevor. 2020. "Computational Musicology in Rock." In *The Bloomsbury Handbook of Rock Music Research*, ed. Allan Moore and Paul Carr, 149–64. Bloomsbury. https://doi.org/10.5040/9781501330483.ch-010.

——————. 2023. "Representation in Corpus Studies of Music: Commentary on Shea's (2022) 'A Demographic Sampling Model and Database for Addressing Racial, Ethnic, and Gender Bias in Popular-music Empirical Research.'" *Empirical Musicology Review* 18 (2). https://doi.org/10.18061/emr.v18i2.9726.

——————. 2025. "What Does a Corpus of Music Represent? Commentary on 'Diversity in Music Corpus Studies.'" *Music Theory Online* 31 (4).

Fareed, Naleef, Christine M. Swoboda, John Lawrence, Tyler Griesenbrock, and Timothy Huerta. 2022. "Co-Establishing an Infrastructure for Routine Data Collection to Address Disparities in Infant Mortality: Planning and Implementation." *BMC Health Services Research* 22 (1): 4. https://doi.org/10.1186/s12913-021-07393-1.

Huron, David. 2013. "On the Virtuous and the Vexatious in an Age of Big Data." *Music Perception* 31 (1): 4–9. https://doi.org/10.1525/mp.2013.31.1.4.

——————. 2018. "Converging Evidence in Music Research." Empirical Musicology Workshops, Columbus, Ohio, May 17.

Laybourn, Wendy M. 2018. "The Cost of Being 'Real': Black Authenticity, Colourism, and Billboard Rap Chart Rankings." *Ethnic and Racial Studies* 41 (11): 2085–103. https://doi.org/10.1080/01419870.2017.1343484.

London, Justin. 2013. "Building a Representative Corpus of Classical Music." *Music Perception* 31 (1): 68–90. https://doi.org/10.1525/mp.2013.31.1.68.

Marsden, Alan. 2022. "Reliability and Validity of Research with Corpora of Music." In *The Oxford Handbook of Music and Corpus Studies*, ed. Daniel Shanahan, John Ashley Burgoyne, and Ian Quinn. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190945442.013.7.

McDermott, Courtney. 2021. “An Interdisciplinary Analysis of Popular Music Genres in the Recording Studio and Billboard Charts.” MA thesis, University of Lethbridge. https://hdl.handle.net/10133/6121.

Shea, Nicholas J. 2022. “A Demographic Sampling Model and Database for Addressing Racial, Ethnic, and Gender Bias in Popular-Music Empirical Research.” *Empirical Musicology Review* 17 (1): 49–58. https://doi.org/10.18061/emr.v17i1.8531.

Shea, Nicholas, Lindsey Reymore, Christopher Wm. White, Ben Duinker, Leigh VanHandel, Matthew Zeller, and Nicole Biamonte. 2024. “Diversity in Music Corpus Studies.” *Music Theory Online* 30 (1). https://doi.org/10.30535/mto.30.1.8.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. “Forecasting Elections with Non-Representative Polls.” *International Journal of Forecasting* 31 (3): 980–91. https://doi.org/10.1016/j.ijforecast.2014.06.001.

Footnotes

1. Tangential to our discussion of population.
Return to text

2. Here, de Clercq’s vision of corpus development disregards a variety of other viable sampling strategies and, perhaps more critically, forgoes the notion that different studies may have contrasting goals and their sampling strategies should accordingly align with these goals. Regarding the first point specifically, de Clercq implies that deciding demographic scope as a first step substantiates objectivity, but in real-world practice, a target population can be codified at any point along the corpus-development process, or even after a corpus has been developed. Some examples from other disciplines: Wang et al. (2015) adopt a top-down approach to sampling, where the researchers accurately predicted the results of the 2012 US presidential election from a demographically skewed sample of Xbox gamers, in a more accurate manner than traditional polling, which relies on unadjusted or “raw” sampling. Contrastingly, Fareed et al. (2022) demonstrate a bottom-up sampling approach to address infant mortality rates in Ohio by compiling, sorting, and synthesizing a variety of public databases consisting of survey responses from a demographically diverse population of mothers. They used this data to connect outreach organizations with Black mothers, whose infants are at a nearly three times greater risk of premature deaths than white infants.
Return to text

3. Our goals resonate with those outlined by Justin London (2013, 72): “while the inclusion of works by certain composers would seem to be obligatory, a rationale rather less arbitrary than choosing ‘the top 50’ is needed . . . Therefore, to produce a more general model, we need to more clearly define a list of obligatory composers, establish a principle for weighting their representation in the corpus . . . and then define an algorithm for selecting additional composers/pieces”, a sentiment that also recalls the sampling process used by Burgoyne, Wild, and Fujinaga (2011) to develop the McGill Billboard corpus. While they focused on song ranking, developing a similar model based on demographic characteristics may be a productive avenue for future studies.
Return to text

Copyright Statement

Copyright © 2025 by the Society for Music Theory. All rights reserved.

[1] Copyrights for individual items published in *Music Theory Online* (*MTO*) are held by their authors. Items appearing in *MTO* may be saved and stored in electronic or paper form, and may be shared among individuals for purposes of scholarly research or discussion, but may *not* be republished in any form, electronic or print, without prior, written permission from the author(s), and advance notification of the editors of *MTO.*

[2] Any redistributed form of items published in *MTO* must include the following information in a form appropriate to the medium in which the items are to appear:

> This item appeared in *Music Theory Online* in [VOLUME #, ISSUE #] on [DAY/MONTH/YEAR]. It was authored by [FULL NAME, EMAIL ADDRESS], with whose written permission it is reprinted here.

[3] Libraries may archive issues of *MTO* in electronic or paper form for public access so long as each issue is stored in its entirety, and no access fee is charged. Exceptions to these requirements must be approved in writing by the editors of *MTO,* who will act in accordance with the decisions of the Society for Music Theory.

This document and all portions thereof are protected by U.S. and international copyright laws. Material contained herein may be copied and/or distributed for research purposes only.

Prepared by NAME, TITLE

SMT
SOCIETY FOR MUSIC THEORY