An Analysis of the GTZAN Music Genre Dataset

Bob L. Sturm
Department of Architecture, Design and Media Technology
Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450 Copenhagen SV, Denmark
bst@create.aau.dk

ABSTRACT

A significant amount of work in automatic music genre recognition has used a dataset whose composition and integrity has never been formally analyzed. For the first time, we provide an analysis of its composition, and create a machine-readable index of artist and song titles. We also catalog numerous problems with its integrity, such as replications, mislabelings, and distortions.

Categories and Subject Descriptors

H.3.1 [**Information Search and Retrieval**]: Content Analysis and Indexing; J.5 [**Arts and Humanities**]: Music

General Terms

Machine learning, pattern recognition, evaluation, data

Keywords

Music genre recognition, exemplary music datasets

1. INTRODUCTION

In their work on automatic music genre recognition, and more generally testing the assumption that features of audio signals are discriminative,[1] Tzanetakis and Cook [20,21] created a dataset (GTZAN) of 1000 music excerpts of 30 seconds duration with 100 examples in each of 10 different categories: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Popular, Reggae, and Rock.[2] Tzanetakis neither anticipated nor intended for the dataset to become a benchmark for genre recognition,[3] but its availability has facilitated much work in this area, e.g., [2–6,10–14,16,17,19–21].

Though it has and continues to be widely used for research addressing the challenges of making machines recognize the complex, abstract, and often argued arbitrary, genre of music, neither the composition of GTZAN, nor its integrity (e.g., correctness of labels, absence of duplicates and distortions, etc.), has ever been analyzed. We only find a few articles where it is reported that someone has listened to at least some of its contents. One of these rare examples is [15], where the authors manually create a ground truth of the key of the 1000 excerpts. Another is in [4]: "To our ears, the examples are well-labeled ... Although the artist names are not associated with the songs, our impression from listening to the music is that no artist appears twice."

In this paper, we catalog the numerous replicas, mislabelings, and distortions in GTZAN, and create for the first time a machine-readable index of the artists and song titles.[4] From our analysis of the 1000 excerpts in GTZAN, we find: 50 exact replicas (including one that is in two classes), 22 excerpts from the same recording, 13 versions (same music but different recordings), and 43 conspicuous and 63 contentious mislabelings (defined below). We also find significantly large sets of excerpts by the same artists, e.g., 35 excerpts labeled Reggae are Bob Marley, 24 excerpts labeled Pop are Britney Spears, and so on. There also exist distortion in several excerpts, in one case making useless all but 5 seconds.

In the next section, we present a detailed description of our methodology for analyzing this dataset. The third section presents the details of our analysis, summarized in Tables 1 and 2, and Figs. 1 and 2. We conclude by discussing the implications of this analysis on the decade of genre recognition research conducted using GTZAN.

2. DELIMITATIONS AND METHODS

We consider three different types of problems with respect to the machine learning of music genre from an exemplary dataset: *repetition, mislabeling*, and *distortion*. These are problematic for a variety of reasons, the discussion of which we save for the conclusion. We now delimit our problems of data integrity, and present the methods we use to find them.

We consider the problem of repetition at four overlapping specificities. From high to low specificity, these are: excerpts are exactly the same; excerpts come from same recording; excerpts are of the same song (versions or covers); excerpts are by the same artist. When excerpts come from the same recording, they may overlap in time or not, and could be time-stretched and/or pitch-shifted, or one may be an equalized or remastered version of the other. Versions or covers are repetitions in the sense of musical repetition and not digital repetition, e.g., a live performance, or one done by a different artist. Finally, artist repetition is self-explanatory.

We consider the problem of mislabeling in two categories:

[1]Personal communication with Tzanetakis.
[2]Available at: `http://marsyas.info/download/data_sets`
[3]Personal communication with Tzanetakis.
[4]Available at `http://imi.aau.dk/~bst`

MIRUM'12, November 2, 2012, Nara, Japan.

Label	ENMFP	manual	last.fm	# tags
Blues	63	96	96	1549
Classical	63	80	20	352
Country	54	93	90	1486
Disco	52	80	79	4191
Hip hop	64	94	93	5370
Jazz	65	80	76	914
Metal	65	82	81	4798
Pop	59	96	96	6379
Reggae	54	82	78	3300
Rock	67	100	100	5475
All	60.6	88.3	80.9	33814

Table 1: Percentages of GTZAN: identified with Echo Nest Musical Fingerprint (ENMFP); identified after manual search (manual); tagged in last.fm database (in last.fm); number of last.fm tags having "count" larger than 0 (tags) (July 3 2012).

conspicuous and contentious. We consider a mislabeling conspicuous when there are clear musicological criteria and sociological evidence to argue against it. Musicological indicators of genre are those characteristics specific to a kind of music that establish it as one or more kinds of music, and that distinguish it from other kinds. Examples include: composition, instrumentation, meter, rhythm, tempo, harmony and melody, playing style, lyrical structure, subject material, etc. Sociological indicators of genre are how music listeners identify the music, e.g., through tags applied to their music collections. We consider a mislabeling contentious when the sound material of the excerpt it describes does not strongly fit the musicological criteria of the label. One example is an excerpt of a Hip hop song, but the majority of it is a sample of Cuban music. Another example is when the song (not recording) and/or artist from which the excerpt comes can fit the given label, but a better label exists, either in the dataset or not.

Though Tzanetakis and Cook purposely created the dataset to have a variety of fidelities [20, 21], the third problem we consider is distortions, such as significant static, digital clipping and skipping. In only one case do we find such distortion rendering an excerpt useless.

As GTZAN has 8 hours and twenty minutes of audio data, the manual analysis of its contents and validation of its integrity is nothing short of fatiguing. In the course of this work, we have listened to the entire dataset multiple times, but when possible have employed automatic methods. To find exact replicas, we use a fingerprinting method [22]. This is so highly specific that it only finds excerpts from the same recording when they significantly overlap in time. It can find neither song nor artist repetitions. In order to approach the other three types of repetition, we first identify as many of the excerpts as possible using The Echo Nest Musical Fingerprinter (ENMFP),[5] which queries a database of about 30,000,000 songs. Table 1 shows that this approach appears to identify 60.6% of the excerpts.

For each match, ENMFP returns an artist and title of the original work. In many cases, these are inaccurate, especially for classical music, and songs on compilations. We thus correct titles and artists, e.g., changing "River Rat Jimmy (Album Version)" to "River Rat Jimmy"; reducing "Bach - The #1 Bach Album (Disc 2) - 13 - Ich steh mit einem Fuss im Grabe, BWV 156 Sinfonia" to "Ich steh mit einem Fuss im Grabe, BWV 156 Sinfonia;" and correcting "Leonard Bernstein [Piano], Rhapsody in Blue" to "George Gershwin" and "Rhapsody in Blue." We review all identifications and find four misidentifications: Country 15 is misidentified as Waylon Jennings (it is George Jones); Pop 65 is misidentified as Mariah Carey (it is Prince); Disco 79 is misidentified as "Love Games" by Gazeebo (it is "Love Is Just The Game" by Peter Brown); and Metal 39 is identified as a track on a CD for improving sleep (its true identity is currently unknown).

We then manually identify 277 more excerpts by either: our own recognition capacity (or that of friends); querying song lyrics on Google and confirming using YouTube; finding track listings on Amazon (when it is clear the excerpts are ripped from an album), and confirming by listening to the on-line snippets; or Shazam.[6] The third column of Table 1 shows that after manual search, we only miss information on 11.7% of the excerpts. With this index, we can easily find versions and covers, and repeated artists. Table 2 lists all repetitions, mislabelings and distortions we find.

Using our index, we query last.fm[7] via their API to obtain the tags that users apply to each song. A tag is a word or phrase a person applies to a song or artist to, e.g., describe the style ("Blues"), its content ("female vocalists"), its affect ("happy"), note their use of the music ("exercise"), organize a collection ("favorite song of all time"), and so on. There are no rules for these tags, but we often see that they are genre-descriptive. With each tag, last.fm also provides a "count," which is a normalized quantity: 100 means the tag is applied by most users, and 0 means the tag is applied by the fewest. We keep only tags having counts greater than 0.

For six of the categories in GTZAN,[8] Fig. 1 shows the percentages of the excerpts coming from specific artists; and for four of the categories, Fig. 2 shows "wordles" of the tags applied by users of last.fm to the songs, along with the *weights* of the most frequent tags. A wordle is a pictorial representation of the frequency of specific words in a text. To create each wordle, we sum the count of each tag (removing all spaces if a tag has multiple words), and use `http://www.wordle.net/` to create the image. The weight of a tag in, e.g., "Blues" is the ratio of the sum of its last.fm counts in the "Blues" excerpts, and the total sum of counts for all tags applied to "Blues."

3. COMPOSITION AND INTEGRITY

We now discuss in more detail specific problems for each label. Each mention of "tags" refers to those applied by last.fm users. For the 100 excerpts labeled Blues, Fig. 1(a) shows they come from only nine artists. We find no conspicuous mislabelings, but 24 excerpts by Clifton Chenier and Buckwheat Zydeco are more appropriately labeled Zydeco. Figure 2(a) shows the tag wordle for all excerpts labeled Blues, and Fig. 2(b) the tag wordle for these particular excerpts. We see that last.fm users do not tag them with "blues," and that "zydeco" and "cajun" together have 55% of the weight. Additionally, some of the 24 excerpts by Kelly Joe Phelps and Hot Toddy lack distinguishing characteristics of Blues [1]: a vagueness between minor and major tonalities from the use of flattened thirds, fifths, and sev-

[5] `http://developer.echonest.com`

[6] `http://www.shazam.com`

[7] `http://last.fm` is an online music service collecting information on listening habits. A tag is something a user of last.fm creates to describe a music group or song in their music collection.

[8] We do not show all categories for lack of space.

GTZAN	Repetitions				Mislabelings		Distortions
Category	Exact	Record.	Version	Artist (# excerpts)	Conspicuous	Contentious	
Blues				JLH: 12; RJ: 17; KJP: 11; SRV: 10; MS: 11; CC: 12; BZ: 12; HT: 13; AC: 2 (see Fig. 1(a))		Cajun and/or Zydeco by CC (61-72) and BZ (73-84); some excerpts of KJP (29-39) and HT (85-97)	
Classical		(42,53)	(44,48)	Mozart: 19; Vivaldi: 11; Haydn: 9; and others			static (49)
Country		(08,51) (52,60)	(46,47)	Willie Nelson: 18; Vince Gill: 16; Brad Paisley: 13; George Strait: 6; and others (see Fig. 1(b))	RP “Tell Laura I Love Her” (20); BB “Raindrops Keep Falling on my Head” (21); Zydecajun & Wayne Toups (39); JP “Running Bear” (48)	GJ “White Lightning” (15); VG “I Can’t Tell You Why” (63); WN “Georgia on My Mind” (67), “Blue Skies” (68)	static distortion (2)
Disco	(50,51, 70)(55, 60,89) (71,74) (98,99)	(38,78)	(66,69)	KC & The Sunshine Band: 7; Gloria Gaynor: 4; Ottawan; 4; ABBA: 3; The Gibson Brothers: 3; Boney M.: 3; and others	CC “Patches” (20); LJ “Playboy” (23), “(Baby) Do The Salsa” (26); TSG “Rapper’s Delight” (27); Heatwave “Always and Forever” (41); TTC “Wordy Rappinghood” (85); BB “Why?” (94)	G. Gaynor “Never Can Say Goodbye” (21); E. Thomas “Heartless” (29); B. Streisand and D. Summer “No More Tears (Enough is Enough)” (47)	clipping distortion (63)
Hip hop	(39,45) (76,78)	(01,42) (46,65) (47,67) (48,68) (49,69) (50,72)	(02,32)	A Tribe Called Quest: 20; Beastie Boys: 19; Public Enemy: 18; Cypress Hill: 7; and others (see Fig. 1(c))	Aaliyah “Try again” (29); Pink “Can’t Take Me Home” (31)	Ice Cube “We be clubbin’” DMX Jungle remix (5); unknown Drum and Bass (30); Wyclef Jean “Guantanamera” (44)	clipping distortion (3,5); skips at start (38)
Jazz	(33,51) (34,53) (35,55) (36,58) (37,60) (38,62) (39,65) (40,67) (42,68) (43,69) (44,70) (45,71) (46,72)			Coleman Hawkins: 28+; Joe Lovano: 14; James Carter: 9; Branford Marsalis Trio: 8; Miles Davis: 6; and others	Leonard Bernstein “On the Town: Three Dance Episodes, Mvt. 1” (00) and “Symphonic dances from West Side Story, Prologue” (01)		clipping distortion (52,54,66)
Metal	(04,13) (34,94) (40,61) (41,62) (42,63) (43,64) (44,65) (45,66)		(33,74)	The New Bomb Turks: 12; Metallica: 7; Iron Maiden: 6; Rage Against the Machine: 5; Queen: 3; and others	Rock by Living Colour “Glamour Boys” (29); Punk by The New Bomb Turks (46-57); Alternative Rock by Rage Against the Machine (96-99)	Queen “Tie Your Mother Down” (58) appears in Rock as (16); Metallica “So What” (87)	clipping distortion (33,73,84)
Pop	(15,22) (30,31) (45,46) (47,80) (52,57) (54,60) (56,59) (67,71) (87,90)	(68,73) (15,21, 22,37) (47,48, 51,80) (52,54, 57,60)	(10,14) (16,17) (74,77) (75,82) (88,89) (93,94)	Britney Spears: 24; Destiny’s Child: 11; Mandy Moore: 11; Christina Aguilera: 9; Alanis Morissette: 7; Janet Jackson: 7; and others (see Fig. 1(d))	Destiny’s Child “Outro Amazing Grace” (53); Diana Ross “Ain’t No Mountain High Enough” (63); Ladysmith Black Mambazo “Leaning On The Everlasting Arm” (81)		Strange sounds added to 37
Reggae	(03,54) (05,56) (08,57) (10,60) (13,58) (41,69) (73,74) (80,81, 82)(75, 91,92)	(07,59) (33,44)	(23,55) (85,96)	Bob Marley: 35; Dennis Brown: 9; Prince Buster: 7; Burning Spear: 5; Gregory Isaacs: 4; and others (see Fig. 1(e))	unknown Dance (51); Pras “Ghetto Supastar (That Is What You Are)” (52); Funkstar Deluxe Dance remix of Bob Marley “Sun Is Shining” (55); Bounty Killer “Hip-Hopera” (73,74); Marcia Griffiths “Electric Boogie” (88)	Prince Buster “Ten Commandments” (94) and “Here Comes The Bride” (97)	last 25 seconds are useless (86)
Rock				Q: 11; LZ: 10; M: 10; TSR: 9; SM: 8; SR: 8; S: 7; JT: 7; and others (see Fig. 1(f))	TBB “Good Vibrations” (27); TT “The Lion Sleeps Tonight” (90)	Queen “Tie Your Mother Down” (16) in Metal (58); Sting “Moon Over Bourbon Street” (63)	jitter (27)

Table 2: Repetitions, mislabelings and distortions in GTZAN excerpts. Excerpt numbers are in parentheses.

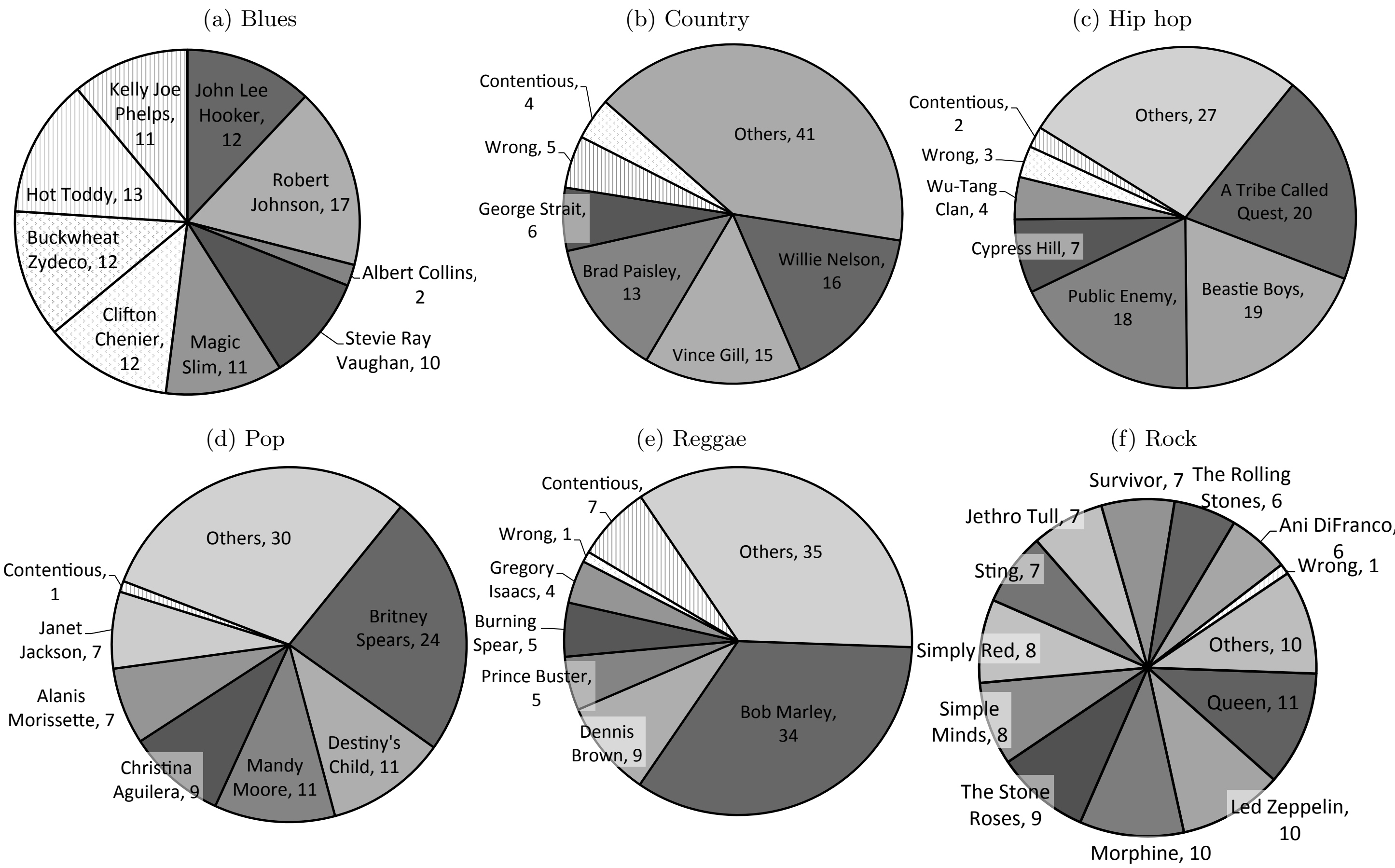


Figure 1: Number of excerpts by specific artists in 6 categories of GTZAN. Mislabeled excerpts are patterned.

enths; twelve bar structure with call and response in lyrics and music; etc. Hot Toddy describes itself as, "[an] acoustic folk/blues ensemble";[9] and last.fm users tag Kelly Joe Phelps most often with "blues, folk, Americana." We thus argue the labels of these 48 excerpts are contentious.

In the Classical-labeled excerpts, we find one pair of excerpts from the same recording, and one pair that comes from different recordings. Excerpt 49 has significant static distortion. Only one excerpt comes from an opera (54).

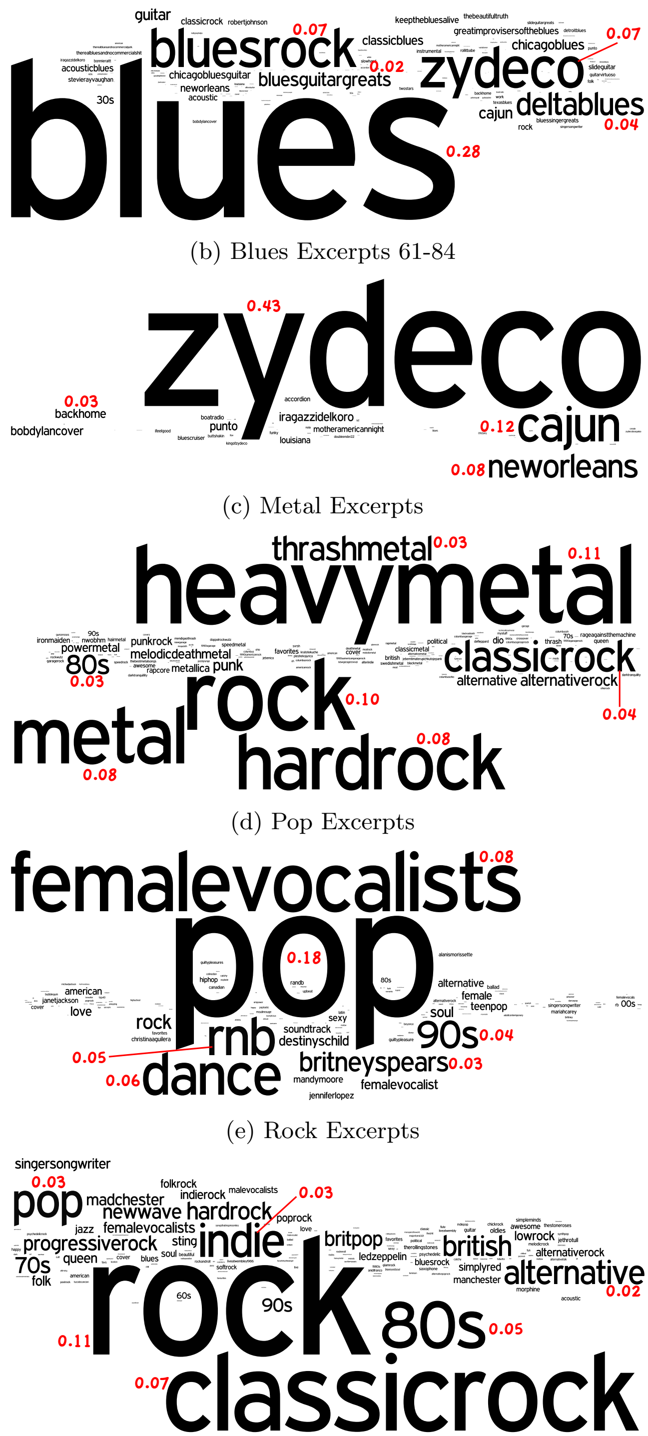
For the Country-labeled excerpts, Fig. 1(b) shows half of them are from four artists. Distinguishing characteristics of Country include [1]: the use of stringed instruments such as guitar, mandolin, banjo, and upright bass; emphasized "twang" in playing and singing; lyrics about patriotism, hard work and hard times. With respect to these characteristics, we find 4 excerpts conspicuously mislabeled Country: Ray Peterson's "Tell Laura I Love Her" (never tagged "country"); Burt Bacharach's "Raindrops Keep Falling on my Head" (never tagged "country"); an excerpt of Cajun music by Zydecajun & Wayne Toups; and Johnny Preston's "Running Bear" (most often tagged "oldies" and "rock n roll"). Contentiously labeled excerpts — all of which have yet to be tagged — are George Jone's "White Lightening," Vince Gill's cover of "I Can't Tell You Why," and Willie Nelson's covers of "Georgia on My Mind" and "Blue Skies." These, we argue, are of genre-specific artists crossing over into other genres.

In the Disco-labeled excerpts we find several repetitions and mislabelings. Distinguishing characteristics of Disco include [1]: 4/4 meter at around 120 beats per minute with emphases of the off-beats by an open hi-hat; female vocalists, piano and synthesizers; orchestral textures from strings and horns; and amplified and often bouncy bass lines. We find seven conspicuous and three contentious mislabelings. First, the top tag for Clarence Carter's "Patches" and Heatwave's "Always and Forever" is "soul." Music from 1991 by Latoya Jackson is quite unlike the Disco preceding it by a decade. Finally, "disco" is not among the top seven tags for The Sugar Hill Gang's "Rapper's Delight," Tom Tom Club's "Wordy Rappinghood," and Bronski Beat's "Why?" For contentious labelings, we find: a modern Pop version of Gloria Gaynor signing "Never Can Say Goodbye;" Evelyn Thomas's "Heartless" (never tagged "disco"); and an excerpt of Barbra Streisand and Donna Summer singing "No More Tears." While this song in its entirety is exemplary Disco, the portion in the excerpt has few Disco characteristics, i.e., no strong beat, bass line, or hi-hats.

The Hip hop category contains many repetitions and mislabelings. Fig. 1(c) shows that 65% of the excerpts come from four artists. Aaliyah's "Try again" is most often tagged "rnb," and "hip hop" is never applied to Pink's "Can't Take Me Home." Though the material in excerpts 5 and 30 are originally Rap or Hip hop, they are remixed in a Drum and Bass, or Jungle, style. Finally, though sampling is a Hip hop technique, excerpt 44 has such a long sample of musicians playing "Guantanamera" that it is contentiously Hip hop.

In the Jazz category of GTZAN, we find 13 exact repli-

[9] `http://www.myspace.com/hottoddytrio`



(a) All Blues Excerpts

(b) Blues Excerpts 61-84

(c) Metal Excerpts

(d) Pop Excerpts

(e) Rock Excerpts

Figure 2: last.fm tag wordles of GTZAN excerpts and weightings of most significant tags.

cas. At least 65% of the excerpts are by five artists. In addition, we find two orchestral excerpts by Leonard Bernstein. In the Classical category of GTZAN, we find four excerpts by Leonard Bernstein (47, 52, 55, 57), all of which come from the same works as the two excerpts labeled Jazz. Of course, the influence of Jazz on Bernstein is known, as it is on Gershwin (44 and 48 in Classical); but with respect to the single-label nature of GTZAN we argue that these excerpts are better categorized Classical.

Of the Metal excerpts, we find 8 exact replicas and 2 versions. Twelve excerpts are by The New Bomb Turks, which are tagged most often "punk, punk rock, garage punk, garage rock." Six excerpts are by Living Colour and Rage Against the Machine, both of whom are most often tagged as "rock." Thus, we argue these 18 excerpts are conspicuously labeled. Figure 2(d) shows that the tags applied to the identified excerpts in this category cover a variety of styles, including Rock, "hard rock" and "classic rock." The excerpt of Queen's "Tie Your Mother Down" is replicated exactly in Rock (16) — where we also find 11 others by Queen. We also find here two excerpts by Guns N' Roses (81, 82), whereas another of theirs is in Rock (38). Finally, excerpt 87 is of Metallica performing "So What" by Anti Nowhere League (tagged "punk"), but in a way that sounds to us more Punk than Metal. Hence, we argue it is contentiously labeled Metal.

Of all categories in GTZAN, we find the most repetitions in Pop. We see in Fig. 1(d) that 69% of the excerpts come from seven artists. Christina Aguilera's cover of Disco-great Labelle's "Lady Marmalade," Britney Spear's "(You Drive Me) Crazy," and Destiny's Child's "Bootylicious" all appear four times each. Excerpt 37 is from the same recording as three others, except it has had strange sounds added. The wordle of tags, Fig. 2(c), shows a strong bias toward music of "female vocalists." Conspicuously mislabeled are the excerpts of: Ladysmith Black Mambazo (group never tagged "pop"); Diana Ross's "Ain't No Mountain High Enough" (most often tagged "motown," "soul"); and Destiny's Child "Outro Amazing Grace" (most often tagged "gospel").

Figure 1(e) shows more than one third of the Reggae category comes from Bob Marley. We find 11 exact replicas, 4 excerpts coming from the same recording, and two excerpts that are versions of two others. Excerpts 51 and 55 are clearly Dance (e.g., a strong common time rhythm with electronic drums and cymbals on the offbeats, synth pads passed through sweeping filters), though the material of 55 is Bob Marley. The excerpt by Pras is most often tagged "hip-hop." And though Bounty Killer is known as a dancehall and reggae DJ, the two repeated excerpts of his "Hip-Hopera" (yet to be tagged) with The Fugees (most often tagged "hip-hop") are Hip hop. Finally, we find "Electric Boogie" is tagged most often "funk" and "dance." To us, excerpts 94 and 97 by Prince Buster sound much more like popular music from the late 1960s than Reggae; and to these songs the most applied tags are "law" and "ska," respectively. Finally, 25 seconds of excerpt 86 is digital noise.

As seen in Fig. 1(f), 56% of the Rock category comes from six groups. The wordle of the tags of these excerpts, Fig. 2(e), shows a significant amount of overlap with that of Metal. Only two excerpts are conspicuously mislabeled: The Beach Boys' "Good Vibrations" and The Tokens' "The Lion Sleeps Tonight," both of which are most often labeled "oldies." One excerpt by Queen is exactly replicated in Metal (58); and while one excerpt here is by Guns N' Roses (38), there are two in Metal (81,82). Finally, Sting's "Moon Over Bourbon Street" (63), is most often tagged "jazz."

4. DISCUSSION

Overall, we find in GTZAN: 7.2% of the excerpts come from the same recording (including 5% duplicated exactly); 10.6% of the dataset is mislabeled; and distortion significantly degrades only one excerpt. We provide evidence for these claims using listening and fingerprinting methods, musicological indicators of genre, and sociological indicators, i.e., by how songs and artists are tagged by users of last.fm.

That so much work in the past decade has used GTZAN to train and test music genre recognition systems raises the question of the extent to which we should believe any conclusions drawn from the results. Of course, since all this work has had to face the same problems in GTZAN, it can be argued their results are still comparable. This, however, makes the false assumption that all machine learning approaches so far used are affected in the same ways by these problems. Because they can be split across training and testing data, exact replicas will, in general, artificially inflate the accuracy of some systems, e.g., k-nearest neighbors [10], or sparse representation classification [6, 13, 19]; and when they are in training data, they can artificially decrease the performance of systems that build models of the data, e.g., Gaussian mixture models [20, 21], and boosted trees [4].

Measuring the extents to which the problems of GTZAN affect the results of particular methods is beyond the scope of this paper, as is any recommendation of how to "correct" its problems to produce, e.g., a "GTZAN2.0" dataset, or whether genre recognition is a well-defined problem. As music genre is in no minor part socially and historically constructed [7, 9, 16], what was accepted 10 years ago as an essential characteristic of a particular genre may not be acceptable today. It thus remains to be seen whether constructing an exemplary music genre dataset of high-integrity is even possible in the first place. We have, however, created a machine-readable index into GTZAN, with which others can apply artist filters to adjust for artificially inflated accuracies from the "producer effect" [8, 18].

Acknowledgments

Thanks to the identification contributions of Mads G. Christensen, Nick Collins, and Carla Sturm. I especially thank George Tzanetakis for his support for and contributions to this process. This work supported in part by: Independent Postdoc Grant 11-105218 from Det Frie Forskningsråd; the Danish Council for Strategic Research of the Danish Agency for Science Technology and Innovation in project CoSound, case no. 11-115328; EPSRC Platform Grant EP/E045235/1 at the Centre for Digital Music of Queen Mary University of London.

5. REFERENCES

[1] C. Ammer. *Dictionary of Music*. The Facts on File, Inc., New York, NY, USA, 4 edition, 2004.

[2] J. Andén and S. Mallat. Scattering representations of modulation sounds. In *Proc. Int. Conf. Digital Audio Effects*, York, UK, Sep. 2012.

[3] E. Benetos and C. Kotropoulos. A tensor-based approach for automatic music genre classification. In *Proc. European Signal Process. Conf.*, Lausanne, Switzerland, Aug. 2008.

[4] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl. Aggregate features and AdaBoost for music classification. *Machine Learning*, 65(2-3):473–484, June 2006.

[5] D. Bogdanov, J. Serra, N. Wack, P. Herrera, and X. Serra. Unifying low-level and high-level music similarity measures. *IEEE Trans. Multimedia*, 13(4):687–701, Aug. 2011.

[6] K. Chang, J.-S. R. Jang, and C. S. Iliopoulos. Music genre classification via compressive sampling. In *Proc. Int. Soc. Music Info. Retrieval*, pages 387–392, Amsterdam, The Netherlands, Aug. 2010.

[7] F. Fabbri. A theory of musical genres: Two applications. In *Proc. Int. Conf. Popular Music Studies*, Amsterdam, The Netherlands, 1980.

[8] A. Flexer. A closer look on artist filters for musical genre classification. In *Proc. Int. Soc. Music Info. Retrieval*, Vienna, Austria, Sep. 2007.

[9] J. Frow. *Genre*. Routledge, New York, NY, USA, 2005.

[10] M. Genussov and I. Cohen. Musical genre classification of audio signals using geometric methods. In *Proc. European Signal Process. Conf.*, pages 497–501, Aalborg, Denmark, Aug. 2010.

[11] E. Guaus. *Audio content processing for automatic music genre classification: descriptors, databases, and classifiers*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2009.

[12] M. Henaff, K. Jarrett, K. Kavukcuoglu, and Y. LeCun. Unsupervised learning of sparse features for scalable audio classification. In *Proc. Int. Soc. Music Info. Retrieval*, Miami, FL, Oct. 2011.

[13] C. Kotropoulos, G. R. Arce, and Y. Panagakis. Ensemble discriminant sparse projections applied to music genre classification. In *Proc. Int. Conf. Pattern Recog.*, pages 823–825, Aug. 2010.

[14] C. Lee, J. Shih, K. Yu, and H. Lin. Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features. *IEEE Trans. Multimedia*, 11(4):670–682, June 2009.

[15] T. Li and A. Chan. Genre classification and the invariance of MFCC features to key and tempo. In *Proc. Int. Conf. MultiMedia Modeling*, Taipei, China, Jan. 2011.

[16] C. McKay and I. Fujinaga. Music genre classification: Is it worth pursuing and how can it be improved? In *Proc. Int. Soc. Music Info. Retrieval*, Victoria, Canada, Oct. 2006.

[17] A. Nagathil, P. Göttel, and R. Martin. Hierarchical audio classification using cepstral modulation ratio regressions based on legendre polynomials. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pages 2216–2219, Prague, Czech Republic, July 2011.

[18] E. Pampalk, A. Flexer, and G. Widmer. Improvements of audio-based music similarity and genre classification. In *Proc. Int. Soc. Music Info. Retrieval*, pages 628–233, London, U.K., Sep. 2005.

[19] Y. Panagakis, C. Kotropoulos, and G. R. Arce. Music genre classification via sparse representations of auditory temporal modulations. In *Proc. European Signal Process. Conf.*, Glasgow, Scotland, Aug. 2009.

[20] G. Tzanetakis. *Manipulation, Analysis and Retrieval Systems for Audio Signals*. PhD thesis, Princeton University, June 2002.

[21] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.*, 10(5):293–302, July 2002.

[22] A. Wang. An industrial strength audio search algorithm. In *Proc. Int. Soc. Music Info. Retrieval*, Baltimore, Maryland, USA, Oct. 2003.