An Analysis of the GTZAN Music Genre Dataset

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

ABSTRACT

Most research in automatic music genre recognition has used the dataset assembled by Tzanetakis et al. in 2001. The composition and integrity of this dataset, however, 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, identifying nearly all excerpts. We also catalog numerous problems with its integrity, including replications, mislabelings, and distortion.

Categories and Subject Descriptors

H.3 [**Information Search and Retrieval**]: Content Analysis and Indexing; H.4 [**Information Systems Applications**]: Miscellaneous; J.5 [**Arts and Humanities**]: Music

General Terms

Machine learning, pattern recognition, evaluation, data

Keywords

Music genre classification, music similarity, dataset

1. INTRODUCTION

In their work in automatic music genre recognition, Tzanetakis et al. [21, 22] created a dataset (GTZAN) of 1000 music excerpts of 30 seconds duration with 100 examples in each of 10 different music genres: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Popular, Reggae, and Rock.[1] Its availability has made possible much work exploring the challenges of making machines recognize something as complex, abstract, and often argued arbitrary, as musical genre, e.g., [2–6, 8–15, 17–20]. However, neither the composition of GTZAN, nor its integrity (e.g., correct labels, duplicates, 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 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 show the contrary is of these observations is true. We highlight numerous other errors in the dataset. Finally, we create for the first time a machine-readable index listing the artists and song titles of almost all excerpts in GTZAN: `http://removed.edu`.

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, different recording), and 44 conspicuous and 64 contentious mislabelings. We also find significantly large sets of excerpts by the same artist, e.g., 35 excerpts labeled Reggae are of Bob Marley, 24 excerpts labeled Pop are of Britney Spears, and so on. There also exist distortion in several excerpts, and in one case this makes 80% of it digital garbage.

In the next section, we present a detailed description of our methodology for analyzing this dataset. The third section then presents the details of our analysis, summarized in Tables 1 and 2, and Figs. 1 and 2. We conclude with some discussion about the implication of our analysis on much of the decade of work conducted and reported using GTZAN.

2. METHODOLOGY

As GTZAN has 8 hours and twenty minutes of audio data, manual analyzing and validating its integrity are difficult; in the course of this work, however, we have listened to the entirety of the dataset multiple times, as well as used automatic methods where possible. In our study of its integrity, we consider three different types of problems: repetition, labeling, and distortions.

We consider the problem of *repetition* at a variety of specificities. From high to low specificity, these are: excerpts are exactly the same; excerpts come from same recording (displaced in time, time-stretched, pitch-shifted, etc.); excerpts are of the same song (versions or covers); excerpts are by the same artist. The most highly-specific repetition of these is exact, and can be found by a method having high specificity, e.g., fingerprinting [23]. When excerpts come from the same recording, they may overlap in time or not; or one may be an equalized or remastered version of the other. Versions or covers of songs are also repetitions, but in the sense of musical repetition and not digital repetition. Finally, we consider as repetitions excerpts featuring the same artist.

The second problem is *mislabeling*, which we consider in two categories: conspicuous and contentious. We consider a mislabeling *conspicuous* when there are clear musicological

[1] Available at: http://marsyas.info/download/data_sets

ACM Special Session '12 Nara, Japan

Label	FP IDed	by hand	in 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
Total	60.6	88.3	80.9	33814

Table 1: Percentages of GTZAN excerpts identified: in Echo Nest Musical Fingerprint database (FP IDed); with additional manual search (by hand); of songs tagged in last.fm database (in last.fm). The number of last.fm tags returned (tags) (July 3 2012).

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 really fit the musicological criteria of the label. One example is an excerpt that comes from a Hip hop song but the majority of it is a sample of a Cuban music. Another example is when the song and/or artist from which the excerpt comes can fit the given label, but a better label exists, either in the dataset or not.

The third problem we consider is *distortion*. Though Tzanetakis et al. [21, 22] purposely created the dataset to have a variety of fidelities, we find errors such as static, and digital clipping and skipping.

To find exact replicas, we use a simplified version of the fingerprinting method in [23]. This approach is so highly specific that it only finds excerpts from the same recording if they significantly overlap in time. It can neither find covers nor identify artists. In order to approach these three types of repetition, we first identify as many of the excerpts as possible using The Echo Nest Musical Fingerprinter (ENMFP) and application programmer interface.[2] With this we can extract and submit a fingerprint of each excerpt, and query 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., making "River Rat Jimmy (Album Version)" be "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 be "George Gershwin" and "Rhapsody in Blue." We also review every identification and correct four cases of misidentification: Country 15 is misidentified as Waylon Jennings when it is George Jones; Pop 65 is misidentified as being Mariah Carey when it is Prince; Disco 79 is misidentified as "Love Games" by Gazeebo when it is "Love Is Just The Game" by Peter Brown; and Metal 39 is misidentified as coming from a new age CD promoting deep sleep.

We then manually identify 277 more excerpts in numerous ways: by recognizing it ourselves; by querying song lyrics on Google and confirming using YouTube; finding track listings on Amazon (when it is clear excerpts come from an album), and confirming by listening to the excerpts provided; or finally, using friends or Shazam.[3] The third column of Table 1 shows that after manual search, we only miss information on 11.7% of the excerpts. With this record then, we are able to find versions and covers, and repeated artist.

With our index, we query last.fm[4] via the last.fm API to obtain the tags that users have applied 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"), and so on. There are no rules for these tags, but we often see that many tags applied to music 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.

3. COMPOSITION AND INTEGRITY

The index we create provides artist names and song titles for GTZAN: `http://removed.edu`. Figure 1 shows how specific artists compose six of the genres; and Fig. 2 shows "wordles" of the tags applied by users of last.fm to songs in four of the genres. (For lack of space, we do not show all genres in GTZAN.) A wordle is a pictorial representation of the frequency of specific words in a text. To create a wordle of the tags of a set of songs, we extract the tags (removing all spaces if a tag is composed of multiple words), their frequencies, and use `http://www.wordle.net/` to create the image. As for the integrity of GTZAN, in Table 2 we list all repetitions, mislabelings and distortions so far found. We now discuss in more detail specific problems for each label.

For the set of excerpts labeled Blues, Fig. 1(a) shows its composition in terms of artists. We find no wrong labels, but 24 excerpts by Clifton Chenier and Buckwheat Zydeco are more appropriately labeled Cajun and/or Zydeco. To see why, Fig. 2(a) shows the wordle of tags applied to all identified excerpts labeled Blues, and Fig. 2(b) shows those only for excerpts numbered 61-84. We see that users do not describe these 24 excerpts as Blues. Additionally, some of the 24 excerpts by Kelly Joe Phelps and Hot Toddy lack distinguishing characteristics of Blues [1]: vagueness between minor and major tonalities; use of flattened thirds, fifths, and sevenths; call and response structures in both lyrics and music, often grouped in twelve bars; strophic form; etc. Hot Toddy describes themselves[5] as, "Atlantic Canada's premier acoustic folk/blues ensemble;" and last.fm users describe Kelly Joe Phelps with the tags "blues, folk, Americana."

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 how the

[2] http://developer.echonest.com

[3] http://www.shazam.com

[4] http://last.fm

[5] http://www.myspace.com/hottoddytrio

Genre	Repetitions				Mislabelings		Distortions
	Exact	Record.	Version	Artist	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; etc.			static (49)
Country		(08,51) (52,60)	(46,47)	Willie Nelson: 18; Vince Gill: 16; Brad Paisley: 13; George Strait: 6; etc. (see Fig. 1(b))	RP "Tell Laura I Love Her" (20); BB "Raindrops Keep Falling on my Head" (21); KD "Love Me With All Your Heart" (22); (39); JP "Running Bear, Little White Dove" (48)	Willie Nelson "Georgia on My Mind" (67), "Blue Skies" (68); George Jones "White Lightning" (15); Vince Gill "I Can't Tell You Why" (63)	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; etc.	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 "High-Energy" (25), "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; etc. (see Fig. 1(c))	Aaliyah "Try again" (29); Pink "Can't Take Me Home" (31); unknown Jungle Dance (30)	Ice Cube "We be clubbin'" DMX Jungle remix (5); 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; etc.	Classical by 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; etc.	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; etc. (see Fig. 1(d))	Destiny's Child "Outro Amazing Grace" (53); Ladysmith Black Mambazo "Leaning On The Everlasting Arm" (81)	Diana Ross "Ain't No Mountain High Enough" (63)	
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; etc. (see Fig. 1(e))	unknown Dance (51); Pras "Ghetto Supastar (That Is What You Are)" (52); Funkstar Deluxe 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; (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.

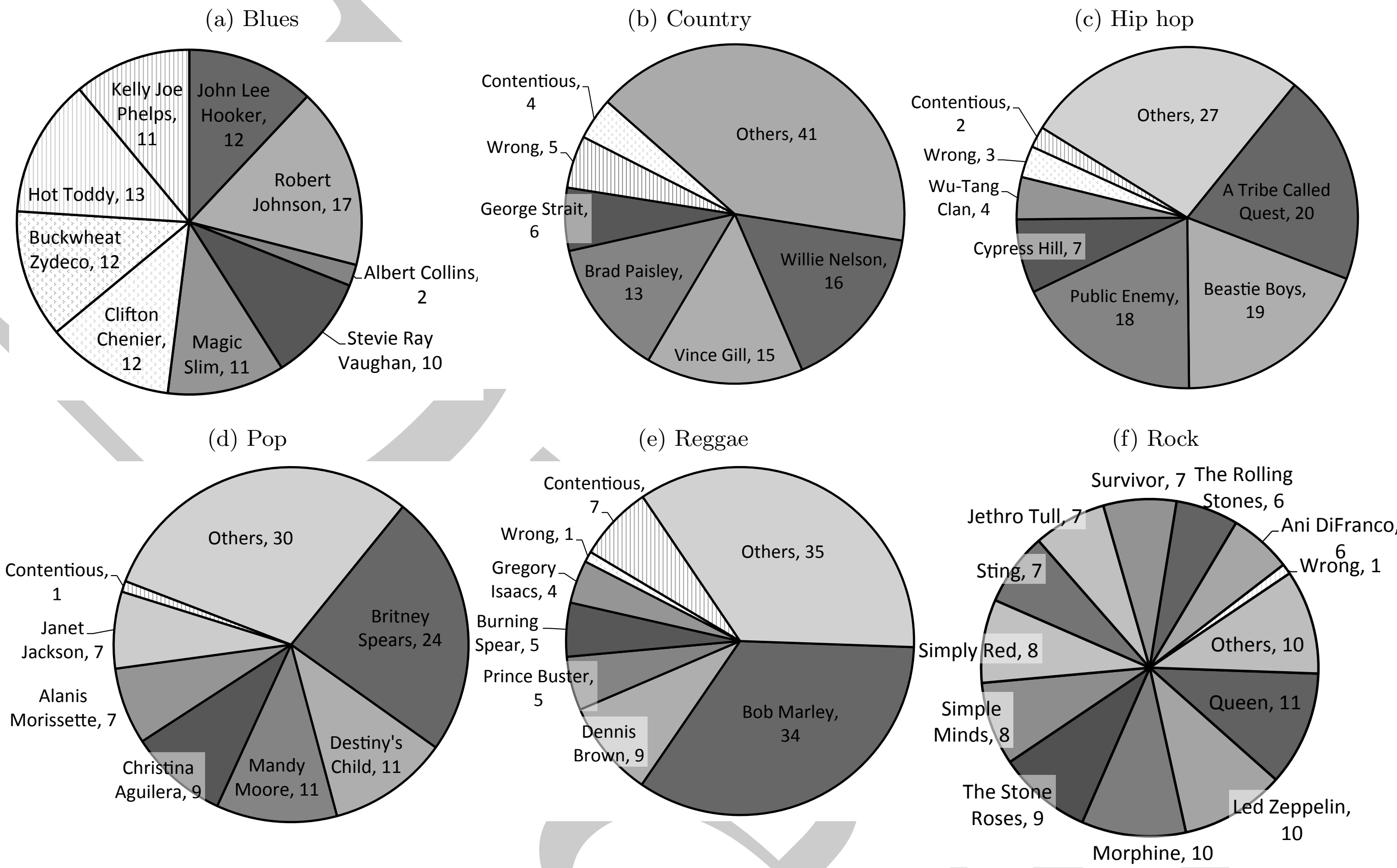

Figure 1: Make-up of GTZAN excerpts in 6 genres in terms of artists and no. Mislabeled excerpts patterned.

majority of these come from only 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; additionally, Country Rock combines Rock rhythms with fancy arrangements, modern lyrical subject matter, and a compressed produced sound. With respect to these characteristics, we find 5 excerpts conspicuously mislabeled Country. Furthermore, Willie Nelson’s rendition of Hoagy Carmichael’s “Georgia on My Mind” and Irving Berlin’s “Blue Skies,” and Vince Gill’s “I Can’t Tell You Why” are examples of Country musicians crossing-over into other genres, e.g., Soul in the case of Nelson, and Pop in the case of Gill.

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 heavy bass lines. We find seven conspicuous and four contentious mislabelings. First, the top last.fm tag applied to Clarence Carter’s “Patches” and Heatwave’s “Always and Forever” is “soul.” Music from 1989 by Latoya Jackson is quite unlike the Disco preceding it by 10 years. Finally, “Disco” is not in the top last.fm tags for The Sugar Hill Gang’s “Rapper’s Delight,” Tom Tom Club’s “Wordy Rappinghood,” and Bronski Beat’s “Why?” For the contentious labelings, excerpt 21 is a Pop version of Gloria Gaynor signing “Never Can Say Goodbye.” The genre of Evelyn Thomas’s two excerpts is closer to the post-Disco electronic dance music style Hi-NRG; and the few last.fm users who have applied tags to her post-Disco music use “hi-nrg.” Finally, excerpt 47 comes from Barbara Streisand and Donna Summer singing “No More Tears,” which in its entirety is Disco, but the portion of the recording from where the excerpt comes has no Disco characteristics.

The Hip hop-labeled excerpts of GTZAN contain many repetitions and mislabelings. Fig. 1(c) shows how the majority of excerpts come from only four artists. Aaliyah’s “Try again” is most often labeled “rnb” by last.fm users; and Hip hop is not among the tags applied to Pink’s “Can’t Take Me Home.” Though the material remixed in excerpt 5 is by Ice Cube, its Jungle dance characteristics are very strong. Finally, excerpt 44 contains such a long sample of Cuban musicians playing “Guantanamera” that the genre of the excerpt is arguably not Hip hop — even though sampling is a classic technique of Hip hop.

In the Jazz-labeled excerpts of GTZAN, we find 13 exact replicas. At least 65% of the excerpts are by five musicians and their groups. In addition, we find two excerpts of Leonard Bernstein’s symphonic work performed by a orchestra. In the Classical excerpts 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 these excerpts are better labeled Classical than Jazz.

Figure 2: Last.fm tag wordles of GTZAN excerpts.

Of the Metal excerpts, we find 8 exact replicas, and 2 versions. Twelve excerpts are by The New Bomb Turks, which last.fm users tag most often with "punk, punk rock, garage punk, garage rock." And 6 excerpts are by Living Colour and Rage Against the Machine, both of whom are most often tagged on last.fm as "rock." Thus, we argue that these 18 excerpts are conspicuously labeled as Metal. Figure 2(d) shows that the tags applied by last.fm users to the identified excerpts labeled Metal cover a variety of styles, including Rock and "classic rock." Queen's "Tie Your Mother Down" is replicated exactly in Rock (16) (where we find 11 other excerpts by Queen). Excerpt 85 is Ozzy Osbourne covering The Bee Gees' "Staying Alive," which is in excerpt 14 of Disco. We also find here two excerpts by Guns N' Roses (81,82), whereas one is in Rock (38). Finally, excerpt 87 is of Metallica performing "So What" by Anti Nowhere League, but in a way that sounds more Punk that Metal. Hence, we argue that this is contentiously labeled Metal.

Of all genres in GTZAN, those labeled Pop have the most repetitions. We see in Fig. 1(d) that five artists compose the majority of these excerpts. Labelle's "Lady Marmalade" covered by Christina Aguilera et al. appears four times, as does Britney Spear's "(You Drive Me) Crazy," and Destiny's Child's "Bootylicious." Excerpt 37 is from the same recording as three others, except it has ridiculous sound effects added. Figure 2(c) shows the wordle of all last.fm tags applied to the identified excerpts, where the bias toward "female vocalists" is clear. The excerpt of Ladysmith Black Mambazo is thus conspicuously mislabeled. Furthermore, we argue that the excerpt of Destiny's Child "Outro Amazing Grace" is more appropriately labeled Soul — the top tag applied by last.fm users.

Figure 1(e) shows that more than one third of the excerpts labeled Reggae are of 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 Dance, though the material of the latter is Bob Marley. To the excerpt by Pras, last.fm users apply most often the tag "hip-hop." And though Bounty Killer is known as a dancehall and reggae DJ, the two repeated excerpts of "Hip-Hopera" with The Fugees are better described as Hip hop. Excerpt 88 is "Electric Boogie," which last.fm users tag most often by "funk, dance." Excerpts 94 and 97 by Prince Buster sound much more like popular music from the late 1960's than Reggae; and to these songs the most applied tags by last.fm users are "law" and "ska," respectively. Finally, 80% of excerpt 86 is extreme digital noise.

As seen in Fig. 1(f), the majority of Rock excerpts come from six groups. Figure 2(e) shows the wordle of tags applied to all 100 Rock-labeled excerpts, which overlaps to a high degree the tags of Metal. Only two excerpts are conspicuously mislabeled: The Tokens' "The Lion Sleeps Tonight" and The Beach Boys' "Good Vibrations." One excerpt by Queen is repeated in Metal (58). We find here one excerpt by Guns N' Roses (38), while two are in Metal (81,82). Finally, to Sting's "Moon Over Bourbon Street" (63), last.fm users most frequently apply "jazz."

4. CONCLUSION

We have provided the first ever detailed analysis and catalogue of the contents and problems of the most used dataset for work in automatic music genre recognition. The most significant problems we find are that 7.2% of the excerpts come from the same recording (including 5% being exact duplicates); and 10.8% of the dataset is mislabeled. We have provided evidence for these claims from musicological indicators of genre, and by inspecting how last.fm users tag the songs and artists from which the excerpts come. We have also created a machine-readable index into all 1000 excerpts, with which people can apply artist filters to adjust for artificially inflated accuracies from the "producer effect" [16].

That so many researchers over the past decade have used

GTZAN to train, test, and report results of proposed systems for automatic genre recognition brings up the question of whether much of that work has been spent building systems that are not good at recognizing genre, but at finding replicas, recognizing extra-musical indicators such as compression, peculiarities in the labels and genre compositions, and sampling bias. Of course, the problems with GTZAN that we have listed here will have varying affects on the performance of a machine learning algorithm. Because they can be split across training and testing sets, exact replicas can artificially inflate the accuracy of some systems, e.g., k-nearest neighbor, or sparse representation classification of auditory temporal modulations [19]. These may not help other systems that build generalized models, such as Gaussian mixture models of feature vectors [21,22]. The extent to which the problems of GTZAN affect the results of a particular method is beyond the scope of this paper, but it presents an interesting problem we are currently investigating.

It is of course extremely difficult to create a dataset that satisfactorily embodies a set of music genres; and if a requirement is that only one genre label can be applied to one excerpt, then it may be impossible, especially when we have to reach a size large enough that current approaches to machine learning can benefit. Furthermore, as music genre is in no minor part socially and historically constructed [7,14], what is accepted 10 years ago as an essential characteristic of a particular genre may not be acceptable today. Hence, we should expect with time and the reflection provided by musicology, that particular examples of a genre become much less debatable. We are currently investigating to what extent the problems in GTZAN can be fixed by such considerations.

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 Information Retrieval*, pages 387–392, Amsterdam, The Netherlands, Aug. 2010.

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

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

[9] 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.

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

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

[12] 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.

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

[14] C. McKay and I. Fujinaga. Music genre classification: Is it work pursuing and how can it be improved? In *Proc. Int. Symp. Music Info. Retrieval*, 2006.

[15] 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.

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

[17] Y. Panagakis and C. Kotropoulos. Music genre classification via topology preserving non-negative tensor factorization and sparse representations. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pages 249–252, Dallas, TX, Mar. 2010.

[18] Y. Panagakis, C. Kotropoulos, and G. R. Arce. Music genre classification using locality preserving non-negative tensor factorization and sparse representations. In *Proc. Int. Symp. Music Info. Retrieval*, pages 249–254, Kobe, Japan, Oct. 2009.

[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] Y. Panagakis, C. Kotropoulos, and G. R. Arce. Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification. *IEEE Trans. Acoustics, Speech, Lang. Process.*, 18(3):576–588, Mar. 2010.

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

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

[23] A. Wang. An industrial strength audio search algorithm. In *Proc. Int. Conf. Music Info. Retrieval*, pages 1–4, Baltimore, Maryland, USA, Oct. 2003.