

FAULTS IN THE LATIN MUSIC DATABASE AND WITH ITS USE

Bob L. Sturm

Centre for Digital Music, Queen Mary University of London, U.K.

ABSTRACT

We find several significant problems in the Latin Music Database (LMD): 1) more than 6.8% of its tracks are replicated (exact or with minor changes to recording playback speed); 2) there are tracks that have a large amount of speech (e.g., live concert setting); 3) the spectral signatures of *Gaucha* appear distinct from those in other classes (leading to a possibility of confounding); 4) the use of LMD in the MIREX Audio Latin Genre Classification train/test task (ALGC) appears ambiguous and flawed.

1. SURVEY OF FAULTS

LMD [2], or portions of it, have been used in about 5% of all published research on music genre recognition [3], including each year 2008-15 of ALGC. Table 1 shows the numbers of tracks in each of the three folds used in ALGC (determined from raw evaluation results). Despite efforts taken to avoid replicas [2], we find at least 220 among its 3227 tracks.¹ We also find 336 tracks in *Tango* are by or of “Carlos Gardel.” Though the evaluation of ALGC is said to use “artist filtering,”² “Carlos Gardel” tracks must appear in both folds 1 and 3. There also appear to be artists across classes, e.g., Marc Anthony in *Bolero* and *Salsa*.

\ Fold Label \	1	2	3	Total	No. (%) replicas
<i>Axe</i>	257	14	42	313	16 (5%)
<i>Bachata</i>	1	131	181	313	53 (17%)
<i>Bolero</i>	68	172	75	315	14 (4%)
<i>Forro</i>	183	0	130	313	6 (2%)
<i>Gaucha</i>	0	126	186	312	6 (2%)
<i>Merengue</i>	224	80	11	315	30 (10%)
<i>Pagode</i>	60	246	0	306	10 (3%)
<i>Salsa</i>	75	217	19	311	47 (15%)
<i>Sertaneja</i>	0	272	49	321	16 (5%)
<i>Tango</i>	114	0	294	408	22 (5%)
Totals	982	1258	987	3227	220 (6.8%)

Table 1. Compositions of the three folds used in ALGC, and the number of replicas we find in the classes.

Table 1 also shows that there exists large differences in track proportions across folds. This lack of balance introduces significant variance in any point estimate com-

¹ List available at <http://www.eecs.qmul.ac.uk/~sturm>.

² “Evaluation” section, MIREX 2014 train/test website [http://www.music-ir.org/mirex/wiki/2014:Audio_Classification_\(Train/Test\)_Tasks](http://www.music-ir.org/mirex/wiki/2014:Audio_Classification_(Train/Test)_Tasks)



puted from the measurements [1] (c.f., beg. chapter 3). To illustrate this, consider the confusion table in Table 2. In this case, ALGC reports an “accuracy (normalised for class sizes)” of 51.77%, meaning all ten classes are considered. Had this system incorrectly labeled the single *Bachata* recording in fold 1, that value would be 41.77%. The significance of such imbalance turns on the explicit identification of the plots and treatments, and the hypothesis being tested [1] (c.f., beg. chapter 1). This imbalance also affects trained systems. To learn to identify *Axe* tracks in fold 1, a system will have been trained using only the 56 *Axe* tracks in folds 2 and 3. To learn to identify the single *Bachata* track, it will have been trained on 312 such tracks.

	<i>A</i>	<i>Ba</i>	<i>Bo</i>	<i>F</i>	<i>G</i>	<i>M</i>	<i>P</i>	<i>Sa</i>	<i>Se</i>	<i>T</i>
<i>A</i>	54	0	0	6	0	1	0	0	0	0
<i>Ba</i>	4	1	3	0	0	23	0	1	0	0
<i>Bo</i>	0	0	31	4	0	1	0	3	0	0
<i>F</i>	19	0	1	65	0	5	0	1	0	0
<i>G</i>	69	0	1	66	0	12	2	9	0	2
<i>M</i>	10	0	1	1	0	175	1	0	0	1
<i>P</i>	57	0	1	11	0	1	43	8	0	1
<i>Sa</i>	15	0	0	4	0	3	2	52	0	0
<i>Se</i>	27	0	28	18	0	0	12	1	0	0
<i>T</i>	2	0	2	8	0	3	0	0	0	110

Table 2. Confusion table of system AP1 in the fold 1 test of ALGC 2014. Column is ground truth.

Another problem is the ambiguity arising from the fact that fold 1 has no tracks from two labels. Table 2 shows the system has correctly identified all *Gaucha* and *Sertaneja* tracks, of which there are none. Should both these recalls then be 100%? In this case, the “normalised accuracy” would be 71.77%. If not, then why divide the sum of the 8 recalls by 10 and not 8?

Figure 1 shows the power spectra of all LMD tracks. We see that a good way to determine if a track is not *Gaucha* is to see if its cutoff frequency is not around 14 kHz. Clear as well is the significant impact of recording medium for most *Tango* tracks (which appear to date from between 1917-1935). Other problems include: *Merengue* “El Torito & Krissy & Tamarindo - Se A Loco” has a watermark sound (i.e., “EXCLUSIVO! Techno trafico ... ”); several tracks end prematurely (e.g., *Merengue* “Manikkomio - Merengue Mania 2003 - CD1 - 11 - Manikkomio”); *Gaucha* “Tch Barbaridade - 10 Anos Mais Fandanguero - 1 - Vinheta de Abertura.mp3” is 18 seconds long. The first 30 seconds of *Gaucha* “Grupo Rodeio - Festchê 1 - 6 - Gritos de liberdade” is speech, as is the first 72 seconds of *Pagode* “Sorriso Maroto - Ao vivo na Providncia - 9 - Por voc” is speech, and the last two minutes of *Salsa* “CELIA CRUZ - CELIA CRUZ & FRIENDS, A NIGHT OF SALSA - 3 - La Vida Es Un Carnaval.”

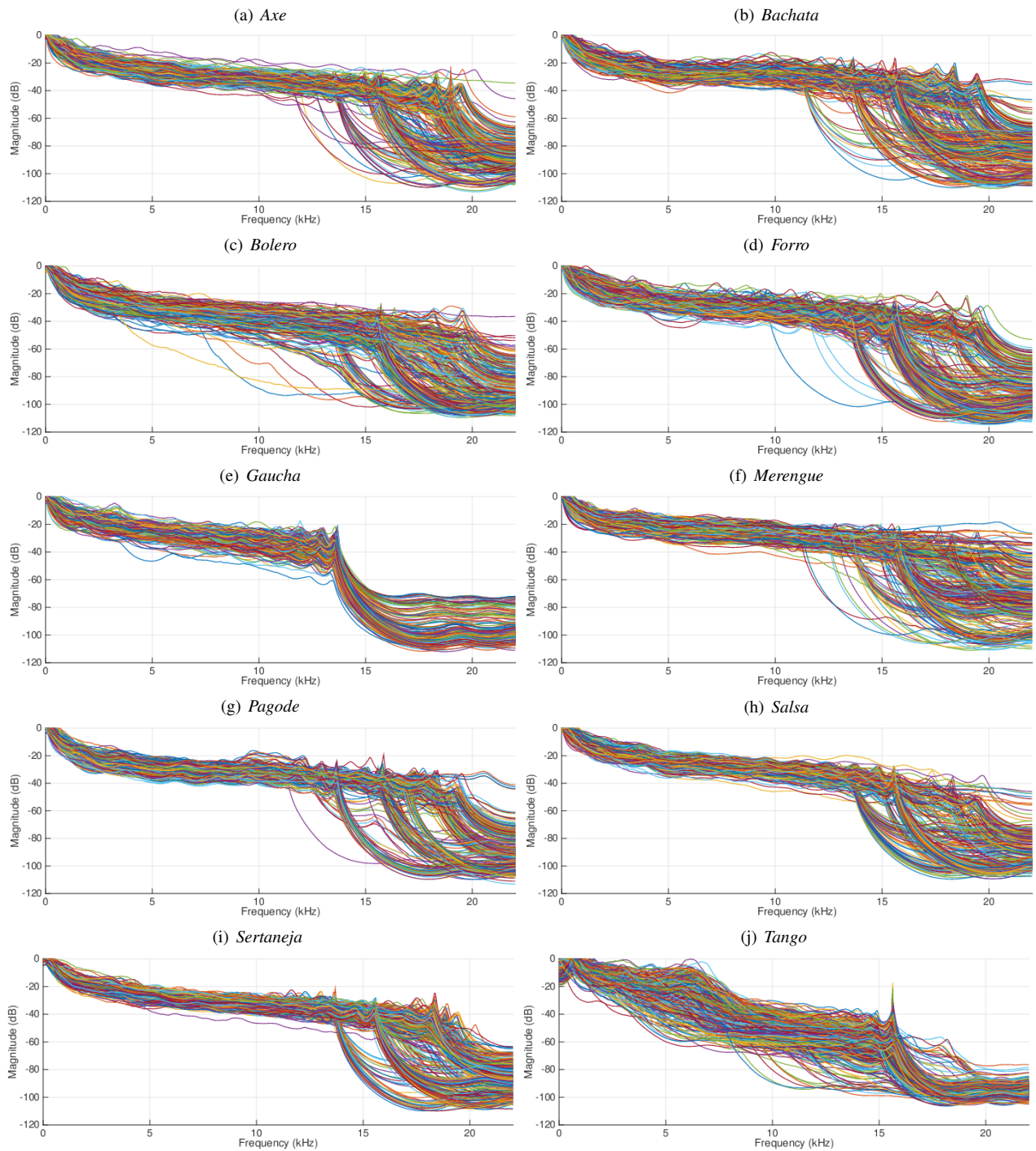


Figure 1. Power spectra of each *LMD* track computed using 25-order LPC analyses of all non-overlapping 2 second windows from either the entire track, or the first 120 seconds, whichever is shortest.

2. REFERENCES

- [1] R. A. Bailey. *Design of comparative experiments*. Cambridge University Press, 2008.
- [2] C. N. Silla, A. L. Koerich, and C. A. A. Kaestner. The Latin music database. In *Proc. ISMIR*, pages 451–456, 2008.
- [3] B. L. Sturm. A survey of evaluation in music genre recognition. In A. Nürnberger, S. Stober, B. Larsen, and M. Detyniecki, editors, *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation*, volume LNCS 8382, pages 29–66, Oct. 2014.