

ASSESSING THE TUNING OF SUNG INDIAN CLASSICAL MUSIC

Joan Serrà, Gopala K. Koduri, Marius Miron and Xavier Serra

Music Technology Group

Universitat Pompeu Fabra, Barcelona, Spain

joan.serraj@upf.edu, gopal.iiit@gmail.com, miron.marius@gmail.com, xavier.serra@upf.edu

ABSTRACT

The issue of tuning in Indian classical music has been, historically, a matter of theoretical debate. In this paper, we study its contemporary practice in sung performances of Carnatic and Hindustani music following an empiric and quantitative approach. To do so, we select stable fundamental frequencies, estimated via a standard algorithm, and construct interval histograms from a pool of recordings. We then compare such histograms against the ones obtained for different music sources and against the theoretical values derived from 12-note just intonation and equal temperament. Our results evidence that the tunings in Carnatic and Hindustani music differ, the former tending to a just intonation system and the latter having much equal-tempered influences. Carnatic music also presents signs of a more continuous distribution of pitches. Further subdivisions of the octave are partially investigated, finding no strong evidence of them.

1. INTRODUCTION

Pitch relationships are at the heart of composition and improvisation in the large majority of musical cultures [1]. The tuning system plays an important role in these relationships, deeply influencing many musical qualities such as mood, consonance or timbre [11, 17]. Traditionally, tuning has been a relevant subject of study from musicological, historical and theoretical perspectives (see [11] and references therein). Current technologies allow for a more empiric and quantitative analysis of the different tunings that enrich our musical experience, specially those used by non-Western cultures or the ones which substantially differ from equal temperament [6, 13, 17, 19].

In Indian classical music, both in Carnatic (South) and Hindustani (North) musical traditions, musicologists have comprehensively covered the issue of tuning (e.g. [15, 18]).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2011 International Society for Music Information Retrieval.

Typically, one finds seven notes, called swaras, which are denoted by Sa, Ri, Ga, Ma, Pa, Da and Ni¹. Except for the tonic and the fifth, all the other swaras have two variations each, which account for 12 notes in an octave, called swarasthanas. It is a well accepted notion that a *swarasthana* is a *region* rather than a point [18]. Therefore, in Indian musicology, a fixed tuning for each note is not as important as it is, say, in Western classical music. Ornamentations are essential, and part of the style.

Still, instruments are tuned using some method. For example, Shankar [18] first presents a set of 12 notes tuned in 5-limit just intonation [11] and later discusses the theory of the 22 shrutis², which fixes 22 unequal subdivisions of the octave. Sambamoorthy [15] directly advocates for this theory and states that “the number 22 represents the barest minimum of shrutis that has been actually used in Indian music from ancient times”, mentioning theories of 24, 27, 32, 48, 53 and 96 shrutis (see also [4] for a brief overview). Such further subdivisions of the octave are very debatable [8, 9, 14], even from a purely psychoacoustical perspective [1]. Therefore, today, the theory of 22 shrutis is under controversy, with some musicologists claiming that they are no more used, replaced by the basic set of 12 notes [14]. The tuning of these 12 notes is also debated, with authors claiming it to be either just intonation, equal-tempered or a mixture of both [9, 16].

Signal processing techniques and computational resources can shed light to the above discussion, providing empiric and quantitative evidence. Initial experiments along this line reported a high variability in the intonation of notes, both in Carnatic and Hindustani music (see [7–9] and references therein). In [7, 8], Krishnaswamy employed different pitch tracking methods (based on the Fourier transform, the autocorrelation function and source separation) to analyze several Carnatic music pieces with different instruments each and found only 12 distinctive intervals. Still, he did not provide any clear statement whether these intervals corresponded to equal temperament or just intonation. Mahendra et al. [12] used an autocorrelation-based method for pitch es-

¹ They are analogous to European solfège’s Do, Re, Mi, Fa, Sol, La, Si.

² Literally, “that which is heard”. A shruti “is the smallest audible difference of pitch. It is a fraction of a semitone”, with no constant value [15].

timation and found clear signs of equal temperament. However, their analysis was done with very few recordings and without providing much detail of them. Larger collections were analyzed by Datta et al. [2–4], who employed a phase-space analysis and a series of post-processing methods to study the shrutis in Hindustani music. Among other results, they report a slightly better fit for 12 interval scales compared to 22 shruti scales in [2]. However, no substantial differences were found between the considered scales (including the 12-interval equal-tempered one). In subsequent papers [3, 4], the same authors provided evidence for the existence of 22 shrutis, although they reported some contradictions.

In this paper we study the tuning of several Indian music recordings. In order to do so, it makes sense to focus on the singing voice, since it is not constrained in its pitch control, it is the reference to be followed by all the other instruments and it has many solo sections, easy to analyze, in all performances [15]. Using standard techniques, we estimate the fundamental frequencies of sung recordings and, based on a pool of these, build an interval histogram. Interval histograms for different data sources are then used to assess the plausibility of just intonation and equal temperament tunings, both in Carnatic and Hindustani music. We compare the interval histograms obtained from these sources against the ones obtained from both synthetic and real signals, the latter coming from commercial Western music recordings. The possibility of 22 shrutis and the issue of tuning variability in Indian classical music are also discussed.

The rest of the paper is organized as follows. First, we present our methodology, both for estimating fundamental frequencies and for computing interval histograms (Sec. 2). Next, we present and discuss our results for different sets of recordings (Sec. 3). Finally, we draw our main conclusions and highlight future research lines (Sec. 4).

2. METHODOLOGY

2.1 Fundamental frequency estimation

Strictly speaking, most Indian classical music should be considered heterophonic. However, for the current analysis, we carefully selected passages where the singer’s voice was very predominant and substantially louder than the rest of the usual accompaniment (Sec. 3.2). Therefore, the analyzed recordings can be roughly considered as monophonic. Under this situation, we choose to estimate their fundamental frequency with an implementation of De Cheveigné & Kawahara’s YIN algorithm [5].

First, we resample the audio to 44100 Hz, downmix it to mono and apply a low-pass filter with a cutoff frequency of 1200 Hz. Next, we estimate the fundamental frequency in a moving window of 4096 samples with a hop size of 133 (93 and 3 ms, respectively). To do so, the time lag

τ_{\min} that yields a minimum of the modified autocorrelation function $d'_n(\tau)$ for each window n is selected [5]. Such minimum $d'_n(\tau_{\min})$ corresponds to a value between 0 and 1, and provides a confidence value $c_n = 1 - d'_n(\tau_{\min})$ for the fundamental frequency. We set the threshold for the search of such minimum to $d'_{Th} = 0.15$ and discard lags below 40 or above 882 samples (above 1100 and below 50 Hz, respectively). For a further refinement of the fundamental frequency, a three-point parabola interpolation and a “best local estimate” within 20 ms is employed [5]. Finally, each fundamental frequency for each window is converted to cents, yielding what we call a *pitch contour*. Confidence values c_n are also kept.

Since we focus on tuning, and in order to mitigate potential errors of YIN, we consider only ‘stable’ regions of our pitch contour. Indeed, we do not need a complete transcription of the sung content, but only those parts of the contour where the tuning can be assumed to be more or less constant. Therefore, and following the ideas of Tidhar et al. [19], we can apply some principle of “conservative transcription”. In our case, we only do a kind of “steady state detection” [2] and keep the cent and confidence values of the windows centered at a *stable region* of 400 ms (we consider a region to be stable if the standard deviation of the pitch contour elements of such region is lower than 20 cents).

2.2 Interval histogram computation

For each recording, a *weighted histogram* with a resolution of one cent is built by rounding the stable cent values above. Confidence values c_n are used as weights. The value of the lowest/highest bin of the histogram corresponds to the minimum/maximum value found in the stable regions (minus/plus an arbitrary constant $\gg 1$). This weighted histogram is mean-smoothed by taking into account the 12 nearest magnitudes of each bin (the 6 immediately lower and upper ones) and linearly normalized between 0 and 1. Notice that, in contrast to Moelants et al. [13], we have not applied any octave equivalence at preliminary stages. Therefore, we are able to discriminate intervals larger than 600 cents.

In Indian classical music, the reference tuning frequency substantially varies within performers [2, 15]. An intuitive, natural and straightforward way to avoid the (usually not trivial, see e.g. [2]) estimation of this reference frequency is to employ intervals, a basic perceptual concept [1]. Therefore, we opt to build and study an *interval histogram*, which we denote as \mathbf{h} . To calculate \mathbf{h} , we select prominent peaks of the mean-smoothed weighted histogram and compute all the possible positive subtractions between peak bins (i.e. the intervals). A peak bin is defined such that it has a magnitude higher than the mean of all the histogram’s magnitudes and higher than the ones of the nearest 50 bins (the 25 lower and upper ones). Since in preliminary analysis we found clear octave equivalences for all intervals, we mapped them to

values between 0 and 1199 cents. For a better visualization, the interval histogram is mean-smoothed by taking into account the 6 nearest values of each bin (the 3 lower and upper bin magnitudes) and linearly normalized between 0 and 1.

3. RESULTS AND DISCUSSION

3.1 Synthetic data

To confirm the usefulness of our approach to detect tunings, we first test it with synthetic data. To generate this data we synthesize MIDI scores using a physically-modeled piano sound on Pianoteq³. We use the first 5 preludes and fugues of Bach’s Well-Tempered Clavier and record them using equal temperament and 5-limit just intonation (treating each MIDI channel as a different piece).

The interval histograms obtained for the synthetic signals show clear peaks at the theoretical interval positions of equal temperament (\mathbf{h}_E) and just intonation (\mathbf{h}_J) tuning systems (Figs. 1 and 2, respectively). These and subsequent figures show the positions of the theoretical interval values that would be obtained with equal temperament, just intonation and the 22 shrutis (vertical lines; see Fig. 1’s caption). The theoretical intervals obtained with just intonation overlap with the ones obtained with the 22 shrutis (c.f. [16, 18]). However, there are a few *idiosyncratic locations* where the shruti intervals do not overlap with the just intonation ones: at 21, 133, 337, 365, 835, 1067 and 1178 cents.

3.2 Carnatic and Hindustani intervals

We now analyze and comment the interval histograms obtained with Indian classical music. We employ two music collections composed of Carnatic (233) and Hindustani (133) recordings of 30 sec to 4 min duration. The recordings are of both male and female singers of various schools and singing styles (rachanas: geetams, varnas, keertanas and kruti). They comprise artists with an active period from 1930 to present such as Balamurali Krishna, Sudha Raghunathan, Maharajapuram Santhanam, John Higgins, Voleti Venkateswarlu, Pandit Ajoy Chakrabarty, Amir Khan, Bhimsen Joshi, Vidyadhar Vyas or Girija Devi. All recordings were selected such that the voice was very loud compared to the rest (primarily alap parts from khyal compositions). Voice was normally accompanied by the drone of a tambura and/or a sarangi, the percussion of a mridangam or the tabla and a violin⁴. In the case of Hindustani music we discarded recordings that contained a harmonium since this could be tuned to equal-temperament.

The interval histogram \mathbf{h}_C obtained for Carnatic music suggests that statements about the use of a just intonation

Location	Nearest JI	Diff. JI	Nearest ET	Diff. ET
187	182	5	200	-13
201	204	-3	200	1
206	204	2	200	6
281	275	6	300	-19
302	294	8	300	2
314	316	-2	300	14
374	386	-8	400	-26
397	406	-9	400	-3
427	427	0	400	27
496	498	-2	500	-4
703	702	1	700	3
766	773	-7	800	-34
799	792	7	800	-1
813	814	-1	800	13
891	884	7	900	-9
906	906	0	900	6
923	925	-2	900	23
985	977	8	1000	-15
996	996	0	1000	-4
1009	1018	-9	1000	9

Table 1. Exact location of prominent peaks of \mathbf{h}_C , the nearest locations of theoretical just intonation (JI) and equal-tempered (ET) profiles and their respective differences (in cents).

system could be true (Fig. 3). In general, prominent peaks are closer to the theoretical interval positions of a just intonation tuning (Table 1). In particular, clear peaks were identified at 314, 427, 496, 703, 813, 906, 923 and 996 cents. The correlation between the Carnatic (\mathbf{h}_C) and the just intonation (\mathbf{h}_J) histograms yielded a value of $r(\mathbf{h}_C, \mathbf{h}_J) = 0.552$, whereas the correlation between \mathbf{h}_C and the equal-tempered interval histogram \mathbf{h}_E yielded $r(\mathbf{h}_C, \mathbf{h}_E) = 0.448$. The difference between correlations $r(\mathbf{h}_C, \mathbf{h}_J)$ and $r(\mathbf{h}_C, \mathbf{h}_E)$ was found to be statistically significant at $p < 10^{-4}$, according to Lawley’s equicorrelation test [10].

The interval histogram \mathbf{h}_H obtained for Hindustani music, however, shows a clear tendency towards equal temperament (Fig. 4). Except for few values distributed near theoretical just intonation intervals (e.g. 92, 223, 792 and 884), the major part of the prominent peaks lie near theoretical equal-tempered intervals (e.g. 300, 400, 800 and 900). Prominent peaks for \mathbf{h}_H were found at 96, 202, 222, 299, 279, 395, 499, 597, 702, 794, 803, 897, 906, 1000, 1104 and 1108. All of them have an equal-tempered location as the nearest one except 222, 279, 702, 906 and 1108, which have a just intonation location as the nearest one. The correlation between \mathbf{h}_H and \mathbf{h}_E yielded a value of $r(\mathbf{h}_H, \mathbf{h}_E) = 0.723$, whereas the correlation against \mathbf{h}_J was $r(\mathbf{h}_H, \mathbf{h}_J) = 0.641$. The difference between $r(\mathbf{h}_H, \mathbf{h}_E)$ and $r(\mathbf{h}_H, \mathbf{h}_J)$ was found to be statistically significant at $p < 10^{-4}$.

³ <http://www.pianoteq.com>

⁴ A comprehensive summary of Indian instruments can be found at http://chandrakantha.com/articles/indian_music/instruments.html

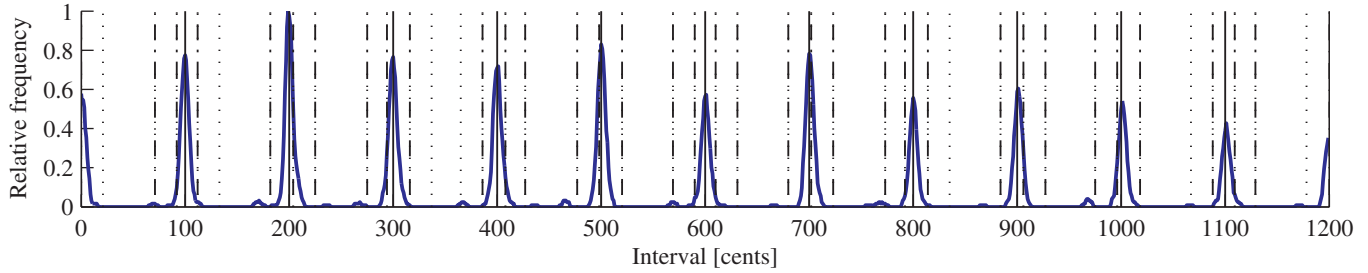


Figure 1. Interval histogram for synthetic equal-tempered data (\mathbf{h}_E ; bold line). Vertical black lines correspond to the theoretical interval values of equal temperament (solid lines), just intonation (dash-dotted lines) and the 22 shrutis (dotted lines). The last two overlap at many places (see text).

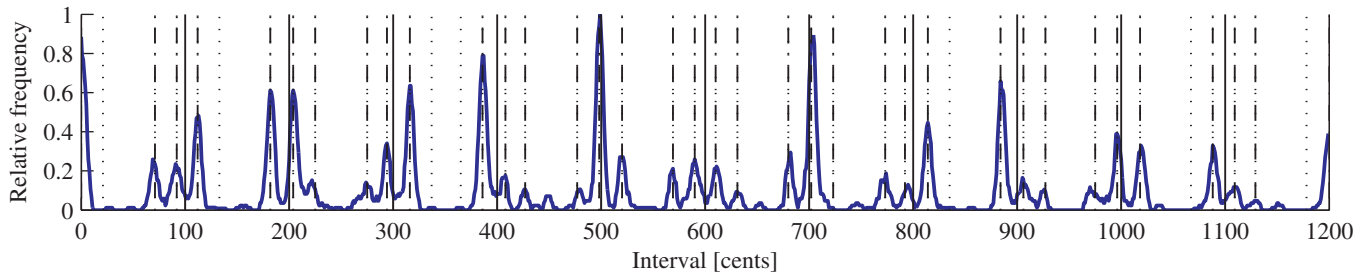


Figure 2. Interval histogram for synthetic just intonation data (\mathbf{h}_J ; bold line).

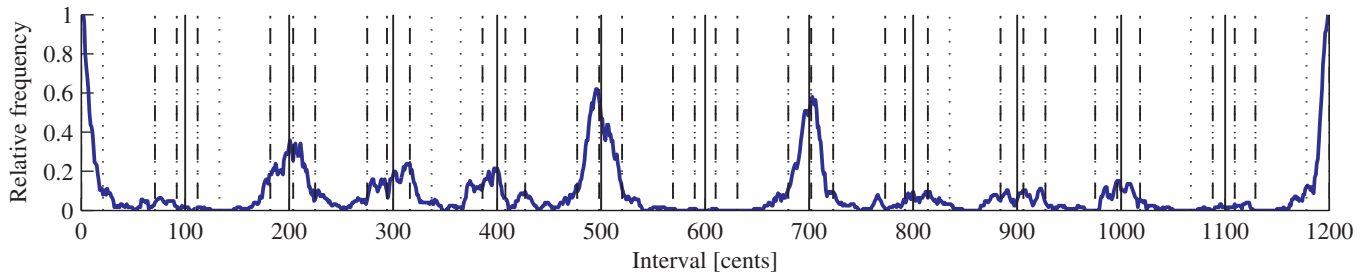


Figure 3. Interval histogram for Carnatic music (\mathbf{h}_C ; bold line).

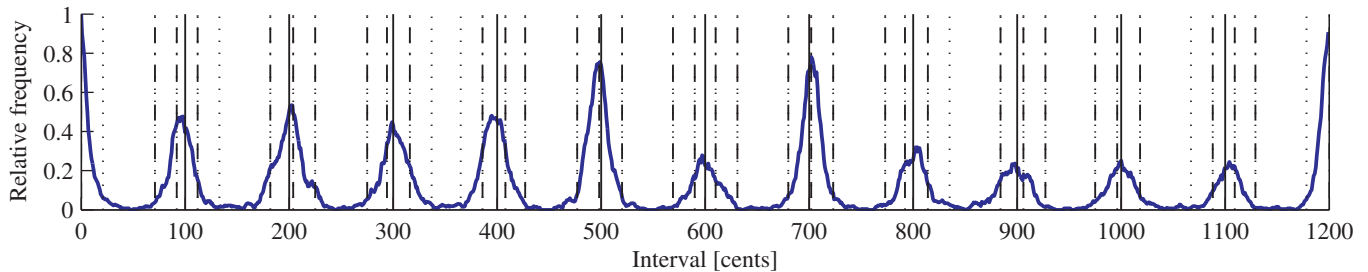


Figure 4. Interval histogram for Hindustani music (\mathbf{h}_H ; bold line).

This tendency towards equal temperament may be explained by the introduction of new instruments in Hindustani music in the last centuries. Hindustani music, unlike Carnatic, was more open to new influences, as the devotional aspect lost its importance. Most instruments used in Indian music have flexible tuning capabilities [15]. One major exception is the hand-pumped harmonium, a mostly equal-tempered instrument introduced in Hindustani music in the late 19th century which is used extensively to accom-

pany the singer soloist. Therefore, it is logical to think that such introduction has influenced the way the singers adjust their pitch. Such influence has yielded, according to some musicians, a “hybrid tuning system” [9], where most singers try to maintain the flat third, but only the purist ones try to also maintain the flat second and the flat sixth.

Going back to the interval histogram of Carnatic music (\mathbf{h}_C ; Fig. 3), we see that distributions around the theoretical interval locations are less peaky than the ones obtained

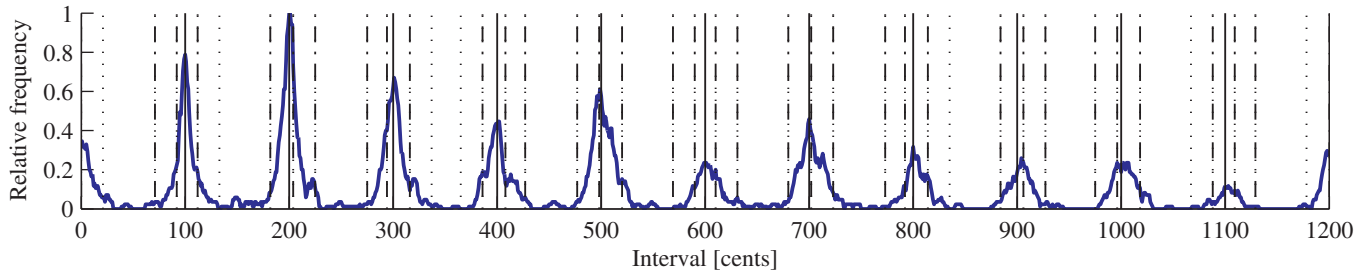


Figure 5. Interval histogram for the fretted electric bass tracks (\mathbf{h}_B ; bold line).

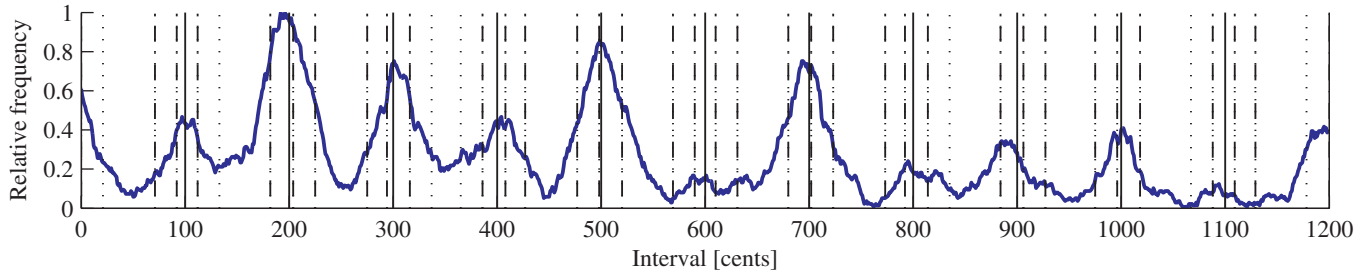


Figure 6. Interval histogram for the vocal Motown tracks (\mathbf{h}_M ; bold line).

with other sources (see below). They are flatter, showing less preference for specific locations, and extending up to the bounds of the just intonation ones. This partially supports the idea of swarasthana regions (Sec. 1). On the other hand, in the analyzed Hindustani music, we see considerably peaky distributions in \mathbf{h}_H (Fig. 4), which contrasts with the flatter distributions of \mathbf{h}_C . This suggests that in our Hindustani recordings the distribution of pitches is much less continuous than in the Carnatic ones.

As mentioned in Sec. 1, the actual use of the 22 shruti scale is controversial. In the light of the results presented here, we must conclude that there is no strong evidence for the existence of this scale in our recordings. For the Hindustani recordings, no peaks are observed at the idiosyncratic locations we mentioned in Sec. 3.1. Thus we cannot differentiate the intervals generated by the 22 shrutis from the ones generated by 12 note just intonation. As for the Carnatic recordings, we note that there is not much support for these idiosyncratic locations. If we look at, for example, 133, 835 or 1067 cents, we do not see any contribution to \mathbf{h}_C . However, small contributions seem to be made at 337, 365 and 1178 cents. Overall, claims that the 22 shrutis are perceivable but not actually played (e.g. [7–9]) acquire strength. However, this phenomenon needs to be further studied, also with behavioral and perceptual studies.

3.3 Comparison with Western practice

To conclude, we apply our methodology to a number of selected Western recordings. In particular, we select 121 tracks of electric (fretted) bass extracted from a collection of multitrack recordings of popular music pieces and 142 vocal

tracks extracted from a collection of multitrack Motown recordings. These recordings comprise different commercial artists whose active period was between 1960 and 2000. All tracks were monophonic and without any sound effect that could change their tuning.

The interval histogram of the electric bass recordings \mathbf{h}_B shows no surprises (Fig. 5). All prominent peaks are very close to (if not exactly at) the theoretical locations of equal temperament intervals. We can appreciate the peakiness of the distributions around, for example, 100, 200, 300, 400 or 500 cents. We find correlations $r(\mathbf{h}_B, \mathbf{h}_E) = 0.831$ and $r(\mathbf{h}_B, \mathbf{h}_J) = 0.478$, which have a statistically significant difference at $p < 10^{-4}$.

The interval histogram of the vocal Motown recordings \mathbf{h}_M presents some subtleties worth commenting (Fig. 6). Intuitively, since singers perform on top of an existing equal-tempered background mix, one would expect \mathbf{h}_M to have prominent peaks at the theoretical locations of equal temperament intervals, in the same vein as \mathbf{h}_B . Although this is true for many peaks (e.g. at 200, 300, 700 or 1000 cents), we can also see some other peaks closer to just intonation positions (e.g. at 92, 112, 406, 792 or 874 cents). Indeed, the correlations $r(\mathbf{h}_M, \mathbf{h}_E)$ and $r(\mathbf{h}_M, \mathbf{h}_J)$ are very similar (0.463 and 0.477, respectively), with a difference that is only just statistically significant ($p \approx 0.01$). The issue of whether there exist some traces of just intonation intervals in these recordings is left for further investigation.

The fact that even Western Motown recordings better approach the just intonation locations than Hindustani music reinforces our hypothesis that the latter has dramatically suffered from equal-tempered influences. In addition, we

should notice that the peakiness of \mathbf{h}_H seems to be slightly larger than the one of \mathbf{h}_M , which indicates that the notion of swarasthana regions may have been lost too.

4. CONCLUSIONS AND FUTURE WORK

The results in Sec. 3 shed light on some of the existing controversies in the tuning of sung Indian classical music. First, we demonstrate how a simplistic approach using standard techniques allows us to assess, in an empiric and quantitative manner, the usage of different tuning systems. Second, we provide evidence that Carnatic music does not make use of an equal-tempered tuning, showing that it presents a strong correlation with 5-limit just intonation. Furthermore, our findings support the notion that Carnatic music may be less confined to strict intervals than the other recordings we have analyzed (swarasthana regions). Hindustani music, on the other hand, seems to be explained by a mixture of equal-tempered tuning and 5-limit just intonation. In addition, we find prominent peaks in its interval histogram, showing a preference for stricter, more precise intervals. In the view of our analysis, the theory of the 22 shrutis lacks strong quantitative evidence. However, we cannot rule out this theory, since many of the intervals overlap with the just intonation ones.

As future work, to resolve this ambiguity, we could employ a method for the estimation of reference tunings [2, 6]. In addition, more recordings should be gathered and different categorizations than Carnatic and Hindustani should be studied (e.g. different epochs [13], different ragas [3] and different schools [2]). Furthermore, the fundamental frequency estimation procedure could be refined [7, 19], as well as the histograms' construction [6]. Unsupervised clustering techniques could be also introduced [3, 4]. Finally, a relaxation and the effect of our 'stability conditions' should be evaluated in depth.

5. ACKNOWLEDGMENTS

The authors wish to thank Enric Guaus and Perfecto Herrera for useful discussions and Justin Salamon for useful discussions and proofreading. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 267583 (CompMusic).

6. REFERENCES

- [1] E. M. Burns. Intervals, scales and tuning. In D. Deutsch, editor, *Psychology of Music, The*, chapter 7, pages 215–264. Academic Press, San Diego, USA, 2nd edition, 1999.
- [2] A. K. Datta, R. Sengupta, N. Dey, D. Nag, and A. Mukerjee. Study of shrutis in indian musical scales and relevance of schools in recent times using signal processing techniques. In *Proc. of the Int. Symp. on Frontiers of Research on Speech and Music (FRSM)*, 2003.
- [3] A. K. Datta, R. Sengupta, N. Dey, D. Nag, and A. Mukerjee. Srutis from objective analysis of the performances of hindustani music using clustering algorithm. In *Proc. of the Natl. Symp. on Acoustics (NSA)*, 2004.
- [4] A. K. Datta, R. Sengupta, N. Dey, D. Nag, and A. Mukerjee. Evaluation of shrutis in hindustani music from recorded performances. In *Proc. of the Int. Symp. on Frontiers of Research on Speech and Music (FRSM)*, 2005.
- [5] A. De Cheveigne and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.
- [6] A. C. Gedik and B. Bozkurt. Pitch-frequency histogram-based music information retrieval for turkish music. *Signal Processing*, 90:1049–1063, 2010.
- [7] A. Krishnaswamy. Application of pitch tracking to south indian classical music. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 557–560, 2003.
- [8] A. Krishnaswamy. Pitch measurements versus perception of south indian classical music. In *Proc. of the Stockholm Music Acoustics Conference (SMAC)*, pages 627–630, 2003.
- [9] A. Krishnaswamy. On the twelve basic intervals in south indian classical music. In *Proc. of the Conv. of the Audio Engineering Society (AES)*, page 5903, 2004.
- [10] D. N. Lawley. On testing a set of correlation coefficients for equality. *Annals of Mathematical Statistics*, 34:149–151, 1963.
- [11] M. Lindley. Temperaments. *Grove Music Online. Oxford Music Online*, 2011. Available online: <http://www.oxfordmusiconline.com/subscriber/article/grove/music/27643>.
- [12] S. R. Mahendra, H. A. Patil, and N. K. Shukla. Pitch estimation of notes in indian classical music. In *Proc. of the IEEE Indian Conf.*, pages 1–4, 2009.
- [13] D. Moelants, O. Cornelis, and M. Leman. Exploring african tone scales. In *Proc. of the Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 489–494, 2009.
- [14] N. Ramanathan. Shrutis according to ancient texts. *Journal of the Indian Musicological Society*, 12(3):31–37, 1981.
- [15] P. Sambamoorthy. *South Indian Music*. The Indian Music Publishing House, Chennai, India, 7th edition, 2006.
- [16] C. Schmidt-Jones. Indian classical music: Tuning and ragas. *Connexions*, 2011. Available online: <http://cnx.org/content/m12459/1.11/>.
- [17] W. A. Sethares. *Tuning, Timbre, Spectrum, Scale*. Springer, Berlin, Germany, 2004.
- [18] V. Shankar. *Art and Science of Carnatic Music, The*. Parampara, Chennai, India, 1999.
- [19] D. Tidhar, M. Mauch, and S. Dixon. High precision frequency estimation for harpsichord tuning classification. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 61–64, 2010.