PRECISION OF SUNG NOTES IN CARNATIC MUSIC

Venkata Subramanian Viraraghavan[1,2] **Hema A Murthy**[3] **R Aravind**[2]
[1] TCS Research and Innovation, Embedded Systems and Robotics, Bangalore, India
[2] Department of Electrical Engineering, Indian Institute of Technology, Madras
[3] Department of Computer Science and Engineering, Indian Institute of Technology, Madras
venkatasubramanian.v@tcs.com, hema@cse.iitm.ac.in, aravind@ee.iitm.ac.in

ABSTRACT

Carnatic music is replete with continuous pitch movement called *gamaka*s and can be viewed as consisting of constant-pitch notes (CPNs) and transients. The stationary points (STAs) of transients – points where the pitch curve changes direction – also carry melody information. In this paper, the precision of sung notes in Carnatic music is studied in detail by treating CPNs and STAs separately. There is variation among the nineteen musicians considered, but on average, the precision of CPNs increases exponentially with duration and settles at about 10 cents for CPNs longer than 0.5 seconds. For analyzing STAs, in contrast to Western music, *rāga* (melody) information is found to be necessary, and errors in STAs show a significantly larger standard deviation of about 60 cents.

To corroborate these observations, the music was automatically transcribed and re-synthesized using CPN and STA information using two interpolation techniques. The results of perceptual tests clearly indicate that the grammar is highly flexible. We also show that the precision errors are not due to poor pitch tracking, singer deficiencies or delay in auditory feedback.

1. INTRODUCTION

The precision of sung notes in Western classical music has been well studied [3, 13, 19]. However, as far as we know, they have not been published for Indian classical music. Previous controlled precision studies were typically concerned with long constant-pitch notes (e.g. [3]), or vibratos [18]. This approach is not suitable for Carnatic Music (CM), where *gamaka*s are characterized by expansive pitch movements. Previous work on Indian music, such as [11], studied *gamaka*s by analyzing *svara*s. However, the term '*svara*' denotes both the note in the musical scale and the *gamaka*s that embellish it. Thus, separating the steady parts of the pitch from the continuous movement is beneficial [5] [17].

In this paper, we quantify the precision of constant-pitch notes (CPNs) and stationary points (STA) separately (see

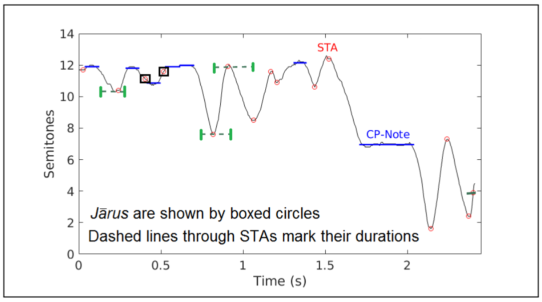


Figure 1. (From [21]) Blue lines are CPNs and red circles, STAs. Some STAs' neighbors and duration are shown.

Name	Sa	Ri		Ga		Ma		Pa	Da		Ni	
Carnatic	S	R1	R2	G2	G3	M1	M2	P	D1	D2	N2	N3
Western	C	C#	D	D#	E	F	F#	G	G#	A	A#	B

Table 1. *Svara*-names and positions of the 12 semitones/octave for Carnatic & Western music. C is the tonic; the correspondence is well-defined only for CPNs.

Figure 1 for examples). We adapt below their definitions from [21] and use the *svara* names given in Table 1.

1. *Silence*-segments (SIL) are identified by the pitch-tracking algorithm [15].

2. A *constant-pitch note* (CPN) is one whose pitch does not vary from its mean pitch by more than 0.3 semitones and lasts for at least 80 ms. Non-SIL and non-CPN regions are called *transients*. *Anchor note(s)* are CPN(s) that flank transient(s).

3. *Stationary points* (STAs) [4, 20], are pitch positions where a continuous pitch curve changes direction. In [4] STAs also occur in CPNs, but they are restricted to the transients in this paper. STAs carry melody information [12] and are useful analytically [4, 14].

4. The *duration* of CPNs and SILs is fairly straightforward. The duration of a STA, typically 100 ms, is defined in [21]. See Figure 1 for an illustration.

The rest of the paper is organized as follows. Section 2 describes a method to statistically analyze precision in CM. Section 2.2 then focuses on precision-errors in CPNs

(CPN-errors). In Section 2.3, we discuss the ambiguity inherent in measuring the precision-errors in STAs (STA-errors) and propose the use of *rāga*-specific information to overcome it. In Section 3, we observe that STAs have about half the precision of short CPNs, suggesting a flexible grammar. Section 4 describes two re-synthesis techniques with different interpolation schemes, which are used in a listening experiment that confirms this flexibility. Section 5 discusses the nature of this flexibility in the grammar.

2. PRECISION-ERROR MEASUREMENTS

2.1 Database, CPNs and STAs

For this work, the database comprised the same 84 concert pieces used in [21], which is a subset of [8]. These are in the *rāga*s *tōḍī*, *bhairavī*, *kharaharapriyā*, *kāmbhōji*, *śankarābharaṇam*, *varāḷī*, and *kalyāṇī*. The database has the dominant pitch (strictly 'fundamental frequency') tracked according to [15] and the tonic identified by the algorithm specified in [7]. Silence segments identified by the pitch tracker are ignored. For non-SIL segments in each piece in the database, we convert the fundamental frequency values to semitones (or equivalent cents) with reference to its tonic. Henceforth, the term 'pitch' in this paper implies measurement in semitones or cents.

Algorithm 1 of [21] is run hierarchically for the duration-threshold of CPNs set to 1000 ms, 300 ms, and 80 ms to get an initial set of CPNs (CPN-set-f). It is then run backwards (in time) to get CPN-set-b. Only CPNs in the intersection of these two sets are retained. Algorithm 2 of the same work is used to identify STAs. STAs adjacent to two CPNs of unequal pitch on either side, and having an intermediate pitch value (*jāru*s, see Figure 1) are ignored. Nineteen professional singers, whose renditions had sufficient data for analysis are chosen.

2.2 Precision-errors of CP-notes

We measure the statistics of the error of the *mean value* of a CPN compared to a target. Instead of assuming a musical scale, the target pitch-values of CPNs are obtained statistically as the mean-values of a pitch class [13, 19]. That is, the locations of the significant peaks ($> 0.01\times$ max value) in the histogram of CPN pitch values folded to one octave are chosen as the target pitch values. This step is repeated for each piece independently and only CPNs longer than 150 ms are considered in finding target CPN pitch-values. Two examples are shown in Figures 2(a) and 2(b). In Figure 2(a), which corresponds to a piece of length just under 49 minutes, the important notes of the *rāga* *śankarābharaṇam*, S, G3 and P are evident. Three other notes (M1, R2 and D2) that seldom occur in the *rāga* as CPNs or anchor notes, have very small peaks. There is no peak at N3, which reflects its rarity as a CPN in the *rāga*. In Figure 2(b), corresponding to a piece of length 47 minutes, the peak at R2 is not in the defined scale of *rāga* *tōḍī*, but Carnatic musicians are aware of its use as an anchor note.

A CPN-error is defined as the difference of a CPN's mean in semitones from the closest target CPN pitch-value. Qualitatively, it is expected that longer CPNs have better precision. To study this behavior, CPNs were grouped by duration according to Equation 1, where the bin-width, $B_w = 40$ ms. An additional bin was used for any duration over 440 ms.

$$\text{Bin}_i = [iB_w, (i+1)B_w), i \in \{2, 3, \ldots, 10\} \quad (1)$$

Figure 3(a) shows the histogram of CPN-errors for three duration bins. Figures 3(b) and 3(c) show the quantile plots [22] for two duration ranges. Note that the number of samples for the longer CPNs are smaller than for shorter ones and thus show more outliers. We also ran the Shapiro-Wilk parametric hypothesis test of composite normality [1], with the default confidence level, $\alpha = 0.05$. The test showed that CPN-errors are, in general, *not* normally distributed. In fact, less than a dozen duration-bins out of over 200 across all singers showed a normal distribution. Nevertheless, we focus on the first two orders of statistics – mean and standard deviation – of the CPN-errors and STA-errors. Further, we treat the standard deviation of the errors as a quantitative measure of the precision.

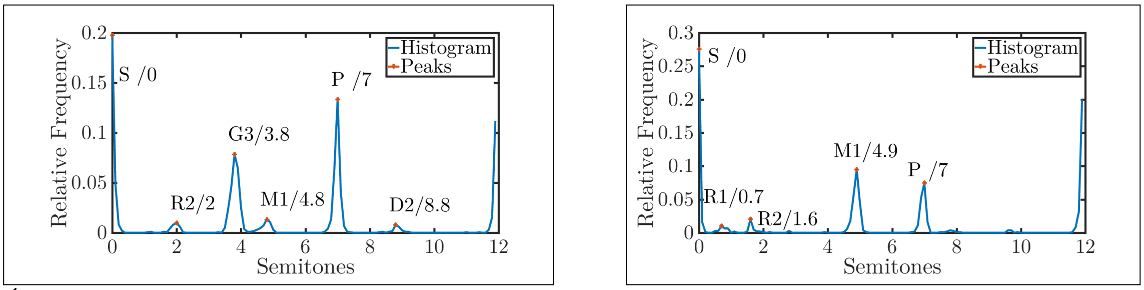
The means and standard deviations of CPN-errors for the 19 singers are shown in Figure 4 as a function of duration. The means are ± 3 cents for all duration-bins. While there is variation among singers, there is a trend of the standard deviation of CPN-errors decreasing with duration.

2.3 Statistics of Stationary Points

2.3.1 *Ambiguity in defining STA targets*

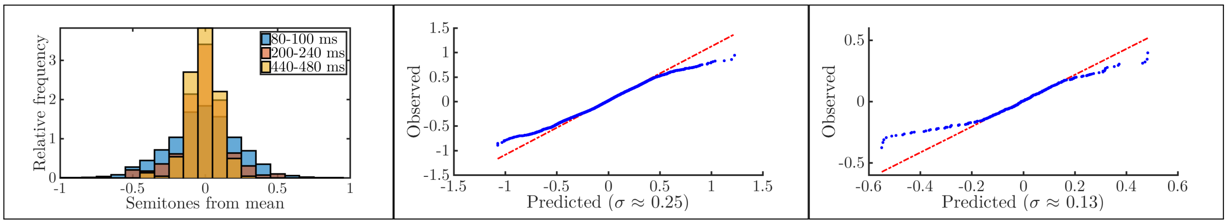
The peaks identified in Figure 2(a) correspond well with *rāga*-characterics even though explicit *rāga*-information is not used in identifying them. This result is encouraging, and the natural step is to adopt the same procedure for identifying STA target pitch-values. Figure 5(a) shows the histogram of STAs, with significant peaks identified exactly as for CPNs for the same piece that corresponds to Figure 2(a). Clearly, they do *not* cluster around scale notes. Further, where the peaks are visible, they are wider than in the case of CPNs. This suggests a larger tolerance for STA pitch errors and is worth verifying. Figure 6 shows a manually annotated spectrogram (using the method of reassignment [1, 10]) of an excerpt from a piece by a very famous singer, known for her exceptional tonal purity. Manual annotation removes the possibility of errors in fundamental frequency tracking. Sixteen of 37 STAs are at semitone values that are not expected in the *rāga*, but on listening to this sample, there is no hint of pitch errors. With STA-errors being of the order of a semitone, the simple histogram-based technique used for CPNs will not suffice. Thus, we propose the use of domain knowledge from CM to define target pitch-values for STAs.

[1] https://in.mathworks.com/matlabcentral/fileexchange/13964-shapiro-wilk-and-shapiro-francia-normality-tests?focused=3823443&tab=function



(a) *Śankarābharaṇam* (Singer 17) S, R2, G3, M1, P, D2, N3. (b) *Tōḍī* (Singer 18) S, R1, G2, M1, P, D1, N2.

Figure 2. Histogram of CPNs of two sample *rāga*s. The CPNs cluster around centers close to the notes of the just-tempered scale. Indian music note names and their values in semitones are marked per peak. Each *rāga*'s scale-sequence is also given.



(a) Normalized histogram (b) Duration-bin 80 ms to 120 ms (c) 440 ms to 480 ms

Figure 3. CPN-errors for Singer 04: Histogram and quantile-plots two duration-bins. Unmarked axes are in semitones.

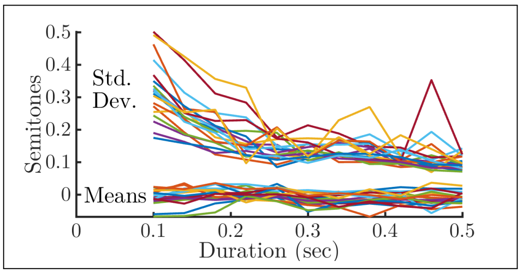


Figure 4. CPN-error statistics for 19 singers. Mean-values are ±3 cents, except for two singers. The standard deviations are in a wider band, but decrease with duration.

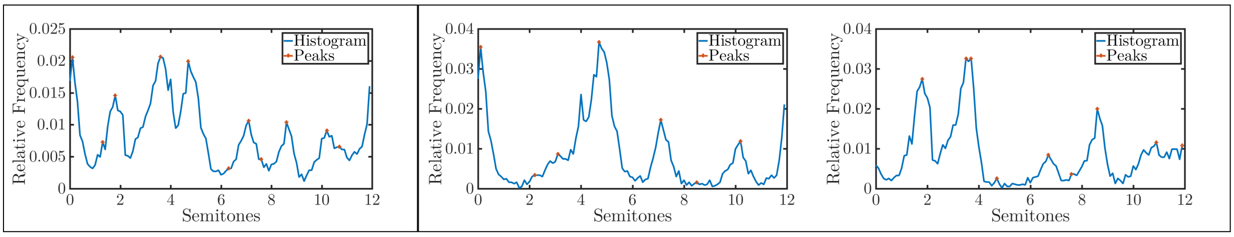
2.3.2 Restricting Measurements to Specific STAs

As explained above, when a sequence of adjacent STAs is encountered, their target pitch-values are not easy to define. To reduce ambiguity, we propose restricting the measurements to a specific type of STA: one that is adjacent to at least one CPN. This effectively pegs one side of the continuous pitch movement, thus providing a practically usable reference. We can then define the precision-error of such a STA with respect to its adjacent CPN. Let such a CPN have a mean pitch p_c in semitones. Then, the target scale-note of this STA is from one of $\mathbb{S} = \{[p_c \pm 1], [p_c \pm 2], [p_c \pm 3]\}$, where $[\cdot]$ denotes rounding the pitch to the nearest integer semitone. In the rare cases that a STA is adjacent to a CPN on both sides, $\mathbb{S}$ is a union of the sets formed by each adjacent CPN. Note that the elements of $\mathbb{S}$ are integer semitones. The mean errors will be affected by a few cents, but as we shall see later, the standard deviation of STA-errors is much larger than the differences between corresponding notes of different musical scales. Thus, the equal-tempered scale, or any other similar scale, can be used to define target pitch-values for STAs, with only a marginal effect on the measured precision. For consistency with Section 2.2, only CPNs and STAs that have a duration $\geq$ 150 ms are included in the measurement.

*Gamaka*s	Elements from $\mathbb{S}$	In $\mathbb{S}'$
To {R2, M1, P}	$\{p_c - 2, p_c + 1, p_c + 3\}$	Yes
To {R1, G2, M2}	$\{p_c - 3, p_c - 1, p_c + 2\}$	No

Table 2. Oscillatory *gamaka*s at G3 in *śankarābharaṇam*

For each CPN in a *rāga*, the valid STA pitch-targets are a subset $\mathbb{S}'$ of $\mathbb{S}$. For example, with the context being an anchor note, say p_c = G3 in the *rāga śankarābharaṇam*, the choices shown in Table 2 can be made. Such rules are not fully documented and are known more by practice. A (proprietary) synthesis algorithm that uses these rules was used to check and correct them in an iterative manner. Finally, overshoots and undershoots of STAs have been reported in the literature [17]. To account for them, STAs in ascending movements of pitch, i.e. where a STA is a local maximum, and in descending movements, where a STA is a local minimum, were measured separately. These subsets of STAs show histograms with sharp peaks in Figures 5(b) and 5(c). Corresponding to Table 2, the upward *gamaka*s from G3 (4 semitones) to M1 (5 semitones) and P (7 semitones) are visible in Figure 5(b) and the downward *gamaka* to R2 (2 semitones), in Figure 5(c). Further, the upward *gamaka* from D2 (9 semitones) to N2 (10 semitones) in Figure 5(b) matches CM practice, although N2 is not in the *rāga*'s scale.



(a) (b) (c)

Figure 5. (a) Histogram of STAs for the same piece as in Figure 2(a). Peaks identified automatically are not clustered around notes of any musical scale. However, visually, STAs adjacent to anchor notes show sharp peaks in both (b) upward and (c) downward movements. To avoid clutter, note names and exact peak-locations are not given.

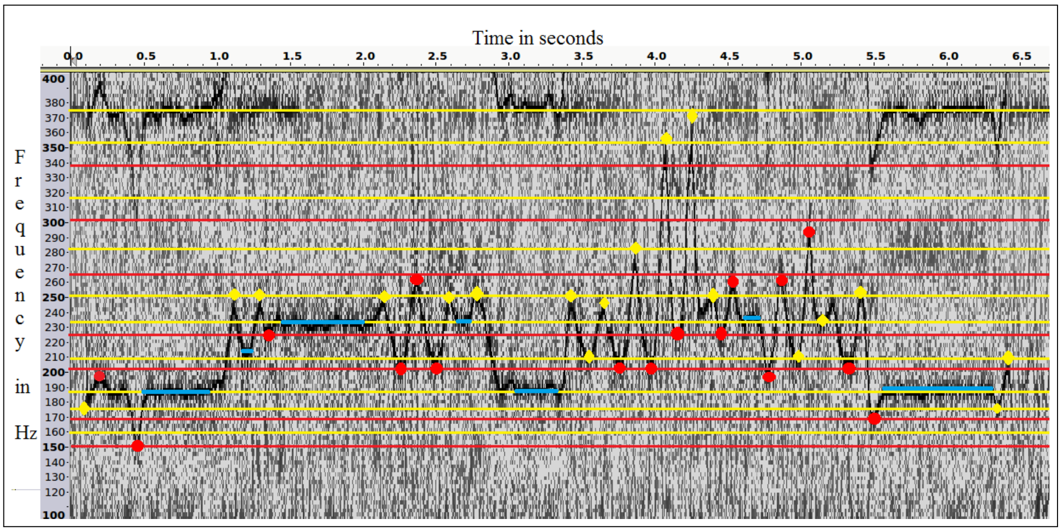
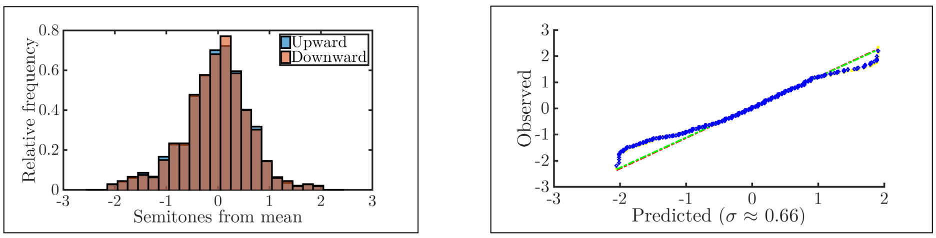
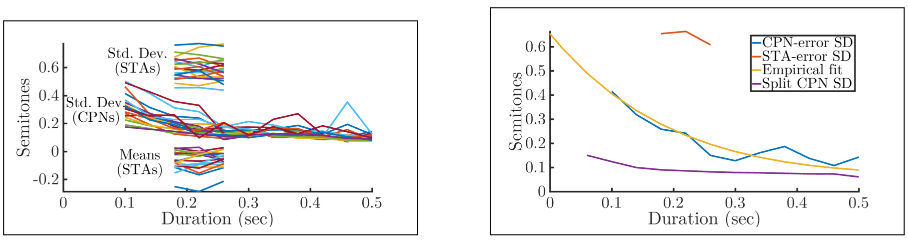


Figure 6. CPNs (blue lines) and STAs (circles and diamonds) in the spectrogram of an excerpt in the *rāga śankarābharaṇam*. The lower dark black curve is the pitch emphasized by reassignment. The tonic (Sa) is 188 Hz. Horizontal lines mark semitones from Ḍ1 to Ṡ. Yellow diamonds mark STAs at expected pitch values and red circles, at unexpected ones. This musician is famous for tonal purity and singer-errors can be discounted.



(a) Normalized histogram (b) Quantile plot for all duration-bins

Figure 7. STA-errors for Singer 04: Histogram and quantile-plot. Unmarked axes are in semitones.



(a) (b)

Figure 8. (a) Means and standard deviations (SDs) of precision errors in STAs for 19 singers. The SDs are mostly in a band of ±10 cents and is more or less constant, unlike that of CPNs (also shown for comparison). (b) SDs of CPN-errors and STA-errors for Singer 04, and for smaller CPNs 'split' from larger ones. The empirical fit for this singer is also shown.

The precision error, ϵ, of any STA chosen thus, with pitch value p semitones, is measured as:

$$I = \arg\min_i |p - p_i| \quad (2)$$

$$\epsilon = p - p_I \quad (3)$$

where i indexes $\mathbb{S}'$ *per* anchor note per *rāga*. Table 2 does not consider *gamaka*s traversing more than 3 semitones, which are rare in CM. Even so, STAs for which $|\epsilon| > 2$ semitones are not included in the measurement. With this more reliable definition of the precision error of these STAs, their statistics were collected per singer. The histogram and quantile plots of STA-errors (Singer 04) are presented in Figures 7(a) and 7(b), which have two virtually indistinguishable plots each: one for STAs in ascending movements, and another for downward movements. The STA-errors also do not follow a normal distribution.

3. ANALYSIS OF OBSERVATIONS

Figure 8(a) shows the means and standard deviations of the STA-errors for the 19 singers. Unlike that of CPN-errors, the standard deviation of STA-errors does *not* vary much with duration. The standard deviation of STA-errors is also about twice as large as that of CPN-errors for CPNs of duration around 100 ms. The slight negative bias of the means of STA-errors is not yet understood.

Modeling the observed precision can be useful in applications such as transcription. We propose a singer-dependent, composite empirical model to predict the standard deviations of *both* CPN-errors and STA-errors. In this model with two components, the first exponentially decreases with time, and can be expressed as:

$$\sigma_x(t) = \sigma_s e^{-t/T} \quad (4)$$

As most singers do not ever reach zero precision-error for even very long CPNs, we need to introduce a constant term σ_r. Thus, the overall standard deviation, as a function of time t, can be written as:

$$\sigma(t) = \sqrt{\sigma_x^2(t) + \sigma_r^2} \quad (5)$$

The forms of Equations 4 and 5 serve to emphasize the first component for low values of t (STAs and short CPNs) and the second, for larger t (long CPNs). For each singer, σ_r was set as the average of the CPN standard deviations for the last three duration-bins. The values of t were chosen as the mid-points of the duration-bins of Equation 1, i.e. $(i + 0.5)B_w$. We also propose that STAs can be viewed as 'point-CPNs'. For example, the short CPN around 1.4 seconds in Figure 1 can be shrunk to a point, which would make it a STA. Practically, a STA lasts at least as long as the shift in windowing algorithms. For the data presented in this paper, this shift is 4.44 ms. Consequently, we set the value of $\sigma(0)$ as the average value of the standard deviation of STA-errors.

The best values of T and σ_s were found by minimizing the mean squared error of the standard deviation predicted by Equations 4 and 5 over the following ranges of values: $T \in \{20, 21, \ldots, 300\}$ ms and $\sigma_s \in \{i\theta,\ i = 1, 2, \ldots, 20\}$, $\theta = 0.05$ semitones. An example for Singer 04 is given in Figure 8(b), which shows a good fit of the model with the observations. The quantitative measure of the fit (RMSE) and the values of T, σ_s and σ_r for the 19 singers are given in Table 3. The typical value of σ_r being in the range 0.08 to 0.15 semitones (one outlier at 0.19) is in good agreement with the precision range of 0.1 to 0.15 semitones reported for choir singers [19]. It remains to be seen if STAs of types other than in Section 2.3 also follow the same statistics.

Singer ID	σ_r	T	σ_s	**RMSE**
01	11	120	45	2.6
03	9	115	60	2.7
04	13	170	65	2.5
05	10	155	55	0.8
06	9	200	45	2.6
07	10	165	55	2.0
08	13	225	65	3.7
09	8	155	70	3.7
10	10	165	55	1.8
13	15	145	50	2.4
17	9	165	60	3.1
18	11	180	85	6.0
19	8	115	60	2.0
20	19	140	50	7.0
21	11	210	40	2.8
27	9	110	50	1.8
28	11	120	65	1.8
30	10	75	65	1.8
31	14	145	55	3.9

Table 3. Prediction parameters for the nineteen musicians aliased by Singer ID. The parameter T is in ms and σ_r, σ_s and the root mean squared error (RMSE) are in cents.

The cause(s) of the precision-error trend is (are) not fully clear, but we eliminate one possibility here. Two types of auditory feedback have been reported in the literature. The first is involuntary, and takes about 100 ms to take effect, and another, voluntary taking about 300 ms [9]. For the smaller duration bins, the voluntary mechanism does not have time to effect corrections. Even the involuntary mechanism does not seem to explain all of the variance. Specifically, the precision error of CPNs is not mirrored in successively longer initial segments of CPNs. That is, for each CPNs of duration $t \geq 300$ ms, and preceded by SIL, several CPNs of duration $iT_{\text{split}}, i \in \{1, 2, \ldots, \lfloor \frac{t}{T_{\text{split}}} \rfloor\}$, where $T_{\text{split}} = 20$ ms, were split from it and their precision for the duration-bins (Equation 1) were calculated. This result for Singer 04 is also shown in Figure 8(b). It is clear that, for durations around 100 ms, the standard deviation of precision error for such 'split notes' is far lower than that for CPNs found from the definition. Thus, it cannot explain the trend seen in CPN-errors. Further, this was seen to be true for all the singers.

Pair-set	U, cosine	R, cosine	R, linear
1	B	T	NA
2	NA	B	T

Table 4. Pair-sets *per rāga* for the quantization algorithms (U or R) and interpolation schemes (linear or cosine) of the synthesized samples in the pair-wise comparison test.

4. EXPERIMENTAL CONFIRMATION

A musicological view (e.g [12]) is that STAs should be precise, which is not consistent with the observations presented hitherto. In a previous experiment [2], the STAs at R1 and D1 in the *rāga pantuvarāḷī* were shifted to R2 and D2 respectively. The shift is not perceivable in the audio synthesized from manual notation [16]. While this experiment confirms the relatively large precision-error of the STAs, perceptual tests were not conducted. Independently, we designed an experiment [3] to confirm the large variability in the pitch-values of STAs. We concatenated one excerpt at slow speed, and another relatively fast one, both chosen from *ālāpana*s in four important, *gamaka*-heavy *rāga*s. These pieces of approximately one-minute duration were transcribed in two ways. In uniform quantization (U), the pitch in semitones of each CPN or STA, p, was set to $p' = [p]$, where $[\cdot]$ denotes rounding towards the nearest integer semitone. In this method, 13% to 19% of STAs were quantized to pitch values not in the *rāga*-specific list $\mathbb{R}$, i.e. Table 2 extended to all anchor notes and octaves. In the second method (R), with i indexing $\mathbb{R}$, and $e_i \in \mathbb{R}$, a CPN/STA pitch (p) was set to $p' = e_I$ where:

$$I = \arg\min(|p - e_i|) \tag{6}$$

The STAs and CPNs were then synthesized by constructing a pitch curve that was constant at CPN-locations. STAs and CPNs (and SILs) were connected to each other by using linear or cosine-interpolation. For the latter, the phase was set to $0(\pi)$ at a starting higher (lower) STA/CPN and to $\pi(0)$ at the ending lower (higher) STA/CPN. These pitch curves, sampled at 1 kHz, and the short-term energy of the original excerpts resampled to 1 kHz, were fed to a good-quality, 5-harmonics synthesis algorithm [6]. We asked listeners to rate pair-wise, the synthesis samples on the basis of adherence to the *rāga*. The *pair-sets* per *rāga* are given in Table 4. Twenty four participants (twelve experts) heard all *rāga*s of pair-set 1 in the order *kāmbhōji*, *śankarābharaṇam*, *tōḍī*, and *bhairavī*. Within a pair, the order was random. This was then repeated for pair-set 2.

Table 5 shows the results of the listening test. For each pair, the preference-percentages, the average rating across participants and *rāga*s, the average difference between ratings, and the average absolute-difference between the ratings are given. All measures indicate that there is no clear

Measure	U vs. R		Cosine vs. Linear	
	U	R	Cosine	Linear
Preference percentage	34 (33)	34 (35)	25 (33)	26 (21)
	Equal: 31 (31)		Equal: 49 (46)	
Avg. rating	3.5 (3.3)	3.4 (3.3)	3.6 (3.4)	3.6 (3.4)
Avg. diff.	0.0 (0.1); 0.7 (0.7)		0.0 (0.1); 0.7 (0.7)	

Table 5. Results of the pair-wise comparison test (expert-ratings in brackets). In the last row, average differences and average absolute-differences are separated by semicolons.

preference among the possibilities. This result can also explain why many interpolation schemes for *gamaka*s – Bezier curves [2], Hermite polynomials [6, 14], Gaayaka software [16], sine curves [17] etc. – all seem to work.

5. CONCLUSIONS

We presented the statistics of precision errors of CPNs and STAs, and measured their means and standard deviations as a function of duration. While the analysis was done separately, the precision-errors for both CPNs and STAs were empirically fitted in a single model. We also presented the results of a listening experiment using the outputs of two synthesis algorithms, both of which also treat CPNs and STAs separately. The key conclusions that can be drawn from this work are:

1. The standard deviation of the precision error in CPNs decreases with duration. A nominal value of 20 cents may be used for a duration of 200 ms, and 10 cents for long CPNs.

2. The standard deviation of the precision error in STAs is independent of duration (45 to 85 cents across singers). A nominal value of 60 cents may be used.

3. Even experts could not tell apart samples that had STAs quantized to notes within a *rāga*'s grammar and those that did not. Also, samples that used linear and cosine interpolation were not distinguishable.

Thus, it appears that there is a large tolerance for both the precision of STAs and the way they are connected, which implies a highly flexible grammar for CM. Point 3 suggests that a rich transcription for a CM piece need not be unique. It may also indicate that re-synthesis quality cannot be used to rate algorithms for rich transcription. However, given that Figures 5(b) and 5(c) show peaks only at expected locations, the existence of unique rich transcription with a large tolerance for STAs is likely.

Finally, it should be noted that the flexibility in its grammar does **not** mean that 'CM is imprecise' or that the precision of STAs is unimportant. Instead, this flexibility should be seen as natural in a form of music that employs continuous pitch movement in profusion.

[2] http://carnatic2000.tripod.com/maya.zip
[3] https://www.iitm.ac.in/donlab/pctestmusic/index.html?owner=venkat1&testid=test1&testcount=8

6. ACKNOWLEDGMENTS

This research was partly funded by the European Research Council under the European Unions Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583). We thank all participants of the listening test and Ms. Anju Leela Thomas and Mr. Krishnaraj Sekhar for helping set it up.

7. REFERENCES

[1] Audacity 2.1.2 [Audio editor] (2016). Latest version freely available from: http://www.audacityteam.org/download/. Accessed: 2017-04-16.

[2] Bret Battey. Bezier spline modeling of pitch-continuous melodic expression and ornamentation. *Computer Music Journal*, 28(4):25–39, 2004.

[3] Simone Dalla Bella, Jean-François Giguère, and Isabelle Peretz. Singing proficiency in the general population. *The journal of the Acoustical Society of America*, 121(2):1182–1189, 2007.

[4] Shrey Dutta and Hema A Murthy. A modified rough longest common subsequence algorithm for motif spotting in an alapana of carnatic music. In *Communications (NCC), 2014 Twentieth National Conference on*, pages 1–6. IEEE, 2014.

[5] Kaustuv Kanti Ganguli, Ashwin Lele, Saurabh Pinjani, Preeti Rao, Ajay Srinivasamurthy, and Sankalp Gulati. Melodic shape stylization for robust and efficient motif detection in hindustani vocal music. In *Proc. of the National Conference on Communication (NCC)*, 2017.

[6] Kaustuv Kanti Ganguli and Preeti Rao. Discrimination of melodic patterns in indian classical music. In *Communications (NCC), 2015 Twenty First National Conference on*, pages 1–6. IEEE, 2015.

[7] Sankalp Gulati, Ashwin Bellur, Justin Salamon, Vignesh Ishwar, Hema A Murthy, and Xavier Serra. Automatic tonic identification in indian art music: approaches and evaluation. *Journal of New Music Research*, 43(1):53–71, 2014.

[8] Sankalp Gulati, Joan Serrà, Kaustuv Kanti Ganguli, Sertan Sentürk, and Xavier Serra. Time-delayed melody surfaces for rāga recognition. In *ISMIR*, pages 751–757, 2016.

[9] Timothy C Hain, Theresa A Burnett, Swathi Kiran, Charles R Larson, Shajila Singh, and Mary K Kenney. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research*, 130(2):133–141, 2000.

[10] Kunihiko Kodera, Roger Gendrin, and C de Villedary. Analysis of time-varying signals with small bt values. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):64–76, 1978.

[11] Gopala Krishna Koduri, Joan Serrà Julià, and Xavier Serra. Characterization of intonation in carnatic music by parametrizing pitch histograms. In *Proceedings of the 13th International Society for Music Information Retrieval Conference; 2012 Oct 8-12;*, 2012.

[12] TM Krishna and Vignesh Ishwar. Carnatic music: Svara, gamaka, motif and raga identity. In *Serra X, Rao P, Murthy H, Bozkurt B, editors. Proceedings of the 2nd CompMusic Workshop; 2012 Jul 12-13; Istanbul, Turkey.* Universitat Pompeu Fabra, 2012.

[13] Peter Q Pfordresher, Steven Brown, Kimberly M Meier, Michel Belyk, and Mario Liotti. Imprecise singing is widespread. *The Journal of the Acoustical Society of America*, 128(4):2182–2190, 2010.

[14] HG Ranjani, Deepak Paramashivan, and Thippur V Sreenivas. Quantized melodic contours in indian art music perception: Application to transcription. In *Proceedings of the 18th International Society for Music Information Retrieval Conference*, pages 174–180, 2017.

[15] Justin Salamon and Emilia Gómez. Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics. *IEEE Transactions on Audio, Speech and Language Processing*, 20:1759–1770, 2012.

[16] M Subramanian. Carnatic music-automatic computer synthesis of gamakams. *Sangeet Natak*, 43(3):28–36, 2009.

[17] Srikumar K Subramanian, Lonce Wyse, and Kevin McGee. A two-component representation for modeling gamakas of carnatic music. In *Proc. of 2nd CompMusic Workshop*, 2012.

[18] Johan Sundberg. Acoustic and psychoacoustic aspects of vocal vibrato.

[19] Sten Ternström and Johan Sundberg. Intonation precision of choir singers. *The Journal of the Acoustical Society of America*, 84(1):59–69, 1988.

[20] Venkatasubramanian V, K R Ramakrishnan, and H V Sahasrabuddhe. Music information retrieval using continuity. In *Proceedings of the Symposium on the Frontiers of Research on Speech and Music (FRSM 2004)*, number 228, pages 74–83. Annamalai University, Chidambaram, 2004.

[21] Venkata S Viraraghavan, R Aravind, and Hema Murthy. A statistical analysis of gamakas in carnatic music. In *Proceedings of the 18th International Society for Music Information Retrieval Conference*, pages 243–249, 2017.

[22] Martin B Wilk and Ram Gnanadesikan. Probability plotting methods for the analysis for the analysis of data. *Biometrika*, 55(1):1–17, 1968.