

VOCAL TRILL AND GLISSANDO THRESHOLDS FOR INDIAN LISTENERS

Vishweshwara Rao, Preeti Rao
Electrical Engineering Department
Indian Institute of Technology Bombay
Email: {vishu, prao}@ee.iitb.ac.in

Abstract

This paper hypothesizes that listeners who have been exposed to pitch continuum traditions of music, in which microtonal variations are a common occurrence, will have greater sensitivity to pitch changes as compared to listeners who have not. To test this hypothesis, two perceptual experiments are designed and performed in order to study the validity of the trill and glissando thresholds, reported in the literature, for listeners who have been trained in or exposed to Indian classical music. Additionally, the dependence of these thresholds on classical music training, center frequency, rate of modulation and direction of modulation (for glissando threshold) is investigated. It was found that thresholds for such listeners were below the thresholds reported previously and that trained vocalists had lower perceptual thresholds than untrained listeners.

Keywords: Indian classical music, Music perception, Pitch perception, Vibrato, Glissando

1. Introduction

Vibrato is one of the most commonly used ornaments in western classical music. Sundberg [1] describes vibrato for singing as a periodic, sinusoidal modulation of phonation frequency. Vibrato is characterized by three parameters: the center frequency, the rate and the extent of vibrato. For western music, the center frequency is usually the note pitch as depicted on the score. The rate of vibrato is defined as the number of oscillations per second in Hz (cycles per second). The extent of vibrato is the frequency interval between the upper and lower limits of the frequency modulation, expressed in cents. Another commonly used ornament in western classical music is the trill, which is also described by a phonation frequency modulation. A trill consists of a quick alternation between two notes, a tone or a semi-tone apart [2].

With respect to the difference in production of a vocal vibrato and a vocal trill, Sundberg states that the extent for the former generally lies in between 1 and 2 semitones (ST), while that of the latter exceeds 2 ST. Castellengo [2] found that the rates of vibrato and trill for several western classical singers was about the same, ranging from 5.5 Hz to 7.5 Hz. She too states that the main difference between the production of vocal vibrato and vocal trill lies in the extent of modulation. Modulations with an extent below 200 cents were always vibrato, modulations with extents above 300 cents were always trills. However, in some cases, both vibrato and trill had extents in the same data interval of 200 – 300 cents, which is called the common area.

In experiments related to perception, Miller and Heise [3] and Shonle and Horan [4] attempted to find a threshold extent above which listeners would perceive a trill and below which they would perceive a vibrato. They called this threshold, the trill threshold. Further, perception of a vibrato i.e. a single vibrated note is termed as ‘fusion’, while perception of a trill i.e. two separate interrupted notes is termed as ‘separation’. The trill threshold extent value found by the former was close to 250 cents and was independent of center frequency. The trill threshold extents found by the latter, however, were smaller and dependent on center frequency; 217 cents at 250 Hz, 100 cents at 500 Hz and 50 cents at 1 kHz.

However, in both of the above experiments the stimuli presented to the listeners consisted of alternating notes generated by two *separate* sources, whose frequency difference was variable. This is consistent with an instrumental trill e.g. the piano, where a trill is produced by alternate fingering of two keys. This is not the case with the singing voice, in which a single source generates a sound that alternates between two pitches, so the validity of these results with respect to the singing voice is ambiguous. In an experiment related to the pitch perceived for short duration vibrato tones d'Allesandro and Castellengo [5] found that the pitch perceived for all stimuli with 200 cents extent was significantly higher or lower than the center frequency. This was then related to the perception of two alternating notes (separation) rather than vibrato notes (fusion).

All of the experiments described above used listeners who had no exposure to pitch continuum traditions such as Indian classical music. In Indian classical music microtonal deviations from the standard intonation are a common occurrence. Although these may only be used in oscillation and may not be sustained for long periods as a steady note, their occurrence is so frequent and widespread that certain musicians use the term 'sruthi' to indicate the subtle intervals produced as a result of this oscillation in pitch [6]. As a result, it may be that listeners who have been exposed to Indian classical music or even Indian film music, a lot of which is based on the former, for long periods of time have greater sensitivity to microtonal variations in music. This assumption directs us to investigate the trill threshold for such listeners, with the expectation that the extent thresholds might be lower than previously reported values.

Further, it is also possible that prolonged exposure to such microtonal deviations might have increased the listener's sensitivity to pitch change. Battey [7], who researched the perceptually equivalent simplification of pitch-time curves in Hindustani (north Indian) vocal music, feels that JND-discrimination between discrete-pitch traditions, such as most Western classical music, and pitch continuum traditions, such as Indian classical music, merits psycho-acoustical investigation. In this context, we are concerned with how rapidly, in a given interval of time, should the phonation frequency change in order to evoke a sensation of pitch change. This is also called the absolute threshold of pitch change or the glissando threshold. 'tHart et. al. [8] studied the distribution of glissando thresholds published in the literature and showed that the glissando thresholds were closely distributed around a curve G_{tr} , as shown in Eq. (1).

$$G_{tr} = 0.16/T^2, \quad (1)$$

where G_{tr} is the glissando threshold in ST/sec and T is the duration of the event.

In this work, we investigate the trill threshold and the glissando threshold for listeners who have been trained in / exposed to Indian classical and film music. The glissando threshold is investigated in the context of a particular gamakam (ornament) called the kampitham. The method of production of the kampitham, in Carnatic (south Indian) music, is described with respect to the veena (an Indian stringed instrument) in [9] as a simple shaking of the string with the finger. From the pitch contour of a kampitham extracted from the sound examples provided in [10] using a customized pitch tracker for the Indian classical music scenario [11] it was observed that the kampitham may be modeled as half a vibrato (sinusoidal) cycle with a low value of extent. This model of the kampitham, having either upward or downward excursions in pitch, is used in the determination of glissando thresholds.

The next section describes the individual experiments for each of the thresholds with respect to the stimuli, the listeners, the parameters under study and the experimental procedure. The results are presented in Section 3. Section 4 contains the conclusions from the study and directions for future work.

2. Experiments

Pitch perception has been shown not to have any strong dependence on the complexity of the signal waveform [4]. Since we are studying perception issues related to singing, all stimuli presented to the listeners for both experiments are synthetic vowels /a/ that have been generated from a formant

synthesizer. All stimuli are multiplied by an amplitude envelope which linearly increases from 0 to 1 for the first 150 ms, remains constant at 1, and linearly decreases from 1 to 0 for the last 150 ms. This is done to avoid the clicks that are otherwise perceived when sounds are abruptly started or ended.

Listeners are divided into two categories: four classically trained singers and four listeners who have had minimal or no classical training. All eight listeners have been listening to either classical Indian music or Indian film music or both since a long time.

2.1 The Trill Threshold

The parameters under study for this experiment are the center frequency, the rate of modulation and the listener category. The parameter values for center frequency and rate are 220, 440 Hz and 4, 6 Hz respectively.

Each listener is presented with four (2 center frequencies x 2 rates) sets of stimuli presented in random order. There are 20 stimuli per set. Each stimulus is a 2 second long synthetic vowel at the given center frequency with the given fixed rate of vibrato with an extent that is randomly selected from a set of extents ranging from 20 to 400 cents in steps of 20 cents. For each stimulus, the listener is instructed to choose whether the sound falls into one of two categories. Category A is labeled as 'Single vibrated note' and represents a vocal vibrato, while category B is labeled as 'Alternating in between two notes' and represents a vocal trill. Since both of these are alien concepts to Indian classical music, prior to the start of each set, the listeners are trained with a sound example of vibrato and trill. Both the sound examples are synthetic vowels /a/ that have the same center frequency and rate as the stimuli of that set. The example of vibrato has an extent of 50 cents while the example of a trill has an extent of 300 cents. These examples are available for listening only prior to each set and are unavailable during the actual classification experiment. For each stimulus, the listeners had no time limits to choose a particular category in the forced choice test. The average time taken to complete the entire experiment (all 4 sets of stimuli including the training stimuli) was about 15 minutes.

2.2 The Glissando Threshold

The parameters under study for this experiment are the base frequency, the rate of modulation, the listener category and the direction of the modulation. The listeners and the center frequencies used are the same as before. The rates of modulation used are 2, 4 and 6 Hz. The inclusion of a low rate is based upon the observation of the kampitham rate in [9], which is about 3 Hz. Since the kampitham is modeled as half a vibrato cycle, two directions of modulations (positive and negative) are separately investigated.

Each listener is presented with 12 stimuli (2 center frequencies x 3 rates x 2 directions of modulation). Each stimulus consists of 0.5 seconds of a synthetic vowel at a steady pitch with no modulation followed by the kampitham followed by another 0.5 seconds of a synthetic vowel at a steady pitch with no modulation. Thus the three events occur successively to form one continuous sound. For each stimulus, the listener is asked to adjust the value of the extent until he/she feels that they can just perceive the pitch oscillation. The extent can be adjusted by moving a sliding bar on the computer screen using the mouse pointer. The lower and upper limits of the sliding bar are 0 and 150 cents respectively. The resolution of the slider was about 0.6 cents. For each stimulus, the rates and directions of the kampitham are selected in random order but the low and high center frequencies are picked alternately. This is done to reduce the drawback of the visual memory, where a subject might try and adjust the slider to the same location for consecutive stimuli having the same base frequency. Again there is no time limit for each adjustment. The average time taken for the entire experiment was about 10 minutes.

3. Results and Discussion¹

3.1 Trill Threshold

For every listener, for each set of trials it was found that there was a relatively clear threshold, with a maximum ambiguity of upto 40 cents, below which they would perceive a vibrato and above which they would perceive a trill. In case of hazy regions (extents for which the perception of vibrato and trill overlap), the center of the region is considered as the trill threshold.

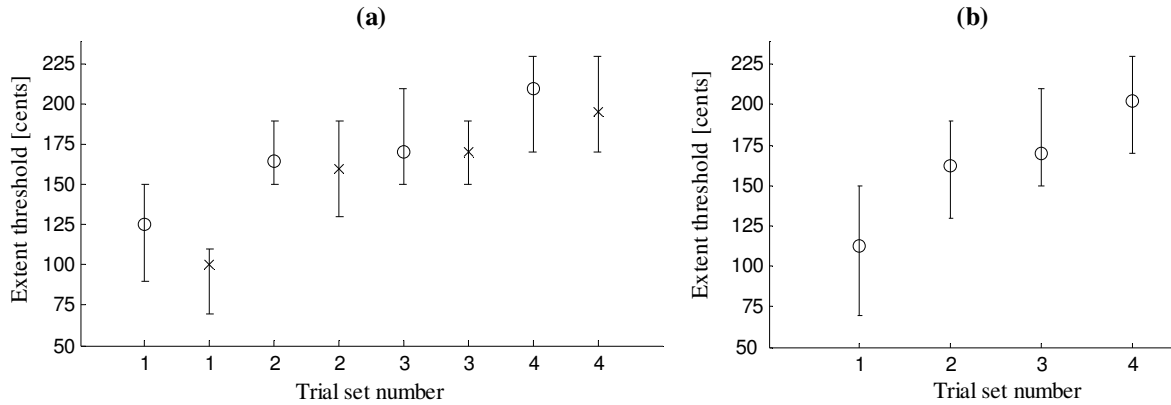


Figure 1: Mean and range plots of the trill threshold extents for each trial set for **a.** different categories of listeners (circles and crosses indicate untrained and trained listeners respectively and **b.** all listeners

Fig. 1a. shows the mean and upper and lower limits for the trill threshold for untrained and trained listeners, indicated by circles and crosses respectively for each of the trial sets. The center frequency and rate of modulation for each trial set is given in Table 1. For all the cases, the trill thresholds for all listeners are found to lie between 70 and 230 cents. While the mean trill threshold for trained listeners is always equal to or less than that for untrained listeners, the difference between the two never exceeds 25 cents. Also, the upper and lower limits on the trill threshold for trained listeners were always equal to or less than those of untrained listeners. This difference is greater at the lower value of center frequency. From Fig. 1b, we can see that the mean values of the trill threshold, computed over all listeners show a clear dependence on the center frequency as well as the rate of modulation. The average increase in the trill threshold for an octave increase in center frequency (from 220 to 440 Hz) and for an increase in rate (from 4 to 6 Hz) is 40 and 50 cents respectively or approximately half a ST.

Table 1. Parameter values for each trial set of stimuli for the trill threshold experiment

Parameters	Trial set 1	Trial set 2	Trial set 3	Trial set 4
Center frequency [Hz]	220	220	440	440
Rate of modulation [Hz]	4	6	4	6

3.2 Glissando Threshold

The parameter values for each stimulus are given in Table 2. From Fig. 2 we can see that for trained and untrained listeners, for all 12 stimuli, the mean extent thresholds above which a pitch modulation is perceived lies between 10 and 60 cents. The mean extent threshold for trained listeners always lies below those of the untrained listeners, indicative of greater sensitivity to pitch change for the former. The difference between mean extent threshold values for trained and untrained listeners is larger (upto 30 cents) at the lower center frequency (stimuli 1 to 6) than at the higher center frequency (stimuli 7

¹ Related audio examples, figures and experimental results are available at <http://www.ee.iitb.ac.in/uma/~daplab/PitchPerception/index.htm>

to 12) where the difference is much smaller (upto 12 cents). Further, for trained listeners we can clearly see that, for the same center frequency and rate, the thresholds for the downward modulation (even numbered stimuli) are always lower than those for the upward modulation (odd numbered stimuli). This trend is also observed for untrained listeners with the exception of stimuli 1 and 2. It is also evident that the difference in the thresholds for upward and downward modulations, for the same rate, increases with an increase in center frequency. Additionally, for the trained listeners, there is a negligible change in the extent threshold with rate, for the same center frequency and direction of modulation, while for untrained listeners the change in the threshold with the change in rate is slightly greater.

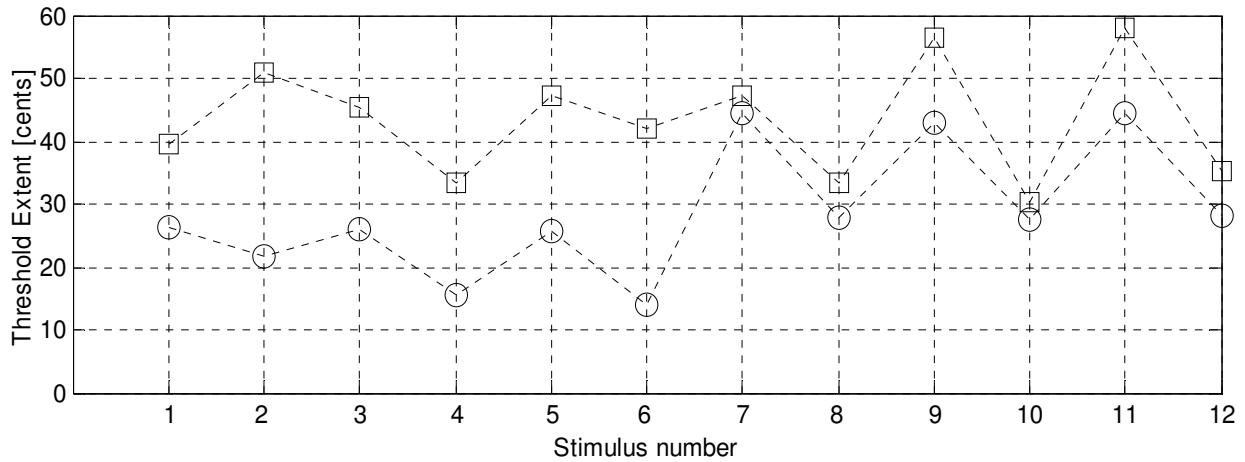


Figure 2: Average extent thresholds for trained (circles) and untrained (squares) listeners for each of the 12 stimuli

Table 2 also contains the glissando threshold values for each rate computed as per Eq.(1), and the average maximum rate of change of pitch (glissando rate) for each stimulus (computed as the mean threshold extent of the kampitham divided by its duration) for trained and untrained listeners. Clearly all the glissando rates for trained as well as untrained listeners are far below the respective glissando thresholds.

Table 2. Parameters, Glissando threshold, Glissando rate for trained and untrained listeners for each stimulus

Stimulus	Center freq. (Hz)	Rate (Hz)	Direction	G_{tr} (ST/s)	Glissando rate (ST/s) (untrained)	Glissando rate (ST/s) (trained)
1	220	2	Up	10.2	1.6	1.1
2	220	2	Down	10.2	2.0	0.9
3	220	4	Up	41.0	3.6	2.1
4	220	4	Down	41.0	2.7	1.2
5	220	6	Up	92.2	5.7	3.1
6	220	6	Down	92.2	5.0	1.7
7	440	2	Up	10.2	1.9	1.8
8	440	2	Down	10.2	1.3	1.1
9	440	4	Up	41.0	4.5	3.4
10	440	4	Down	41.0	2.4	2.2
11	440	6	Up	92.2	7.0	5.3
12	440	6	Down	92.2	4.2	3.4

4. Conclusion and Future Work

This paper investigates the validity of the trill and glissando thresholds reported in previous literature with respect to listeners from a pitch continuum tradition such as Indian classical music.

Individual perceptual experiments, specifically designed for the two thresholds, were carried out on two sets of listeners: untrained, and trained in Indian classical music. For the trill threshold, it was observed that a large number of the measured thresholds were well below 200 cents. Classical training was not found to be an important factor for the threshold values. However, with an octave increase in center frequency or an increase in the rate of modulation from 4 to 6 Hz, the trill threshold was found to increase by about half a semitone. For the glissando threshold, the mean glissando rates at threshold for trained and untrained listeners were found to be far below the glissando threshold values stated in the literature. Additionally, it was found that the glissando rates for untrained listeners were higher than those for trained listeners. Also, the glissando rates for downwards pitch oscillations were generally significantly lower than those for upward frequency oscillations. We intend to verify that these lower thresholds are indeed due to our hypothesis that listeners from pitch continuum traditions are likely to be more sensitive to pitch change and not due to the fact that new experimental procedure was used to measure the glissando threshold. To resolve this issue, we intend to perform the same experiment with listeners who have had no prior exposure to Indian or any other pitch continuum tradition of music at all.

The trill threshold has little significance as far as Indian classical music is concerned, since neither vibrato nor trill are explicitly defined ornaments in Indian music. A more relevant perceptual threshold would be the kampitham to vali threshold. The vali, in Carnatic music, is a gamakam that can be described as an exaggerated kampitham i.e. a kampitham with a slower rate and a much larger extent. The Hindusthani name for the vali is andolan, defined as a slow shake in [10]. Such a threshold can play an integral part in the process of modeling / transcribing Indian classical music. We intend to design and perform another perceptual experiment along the lines of the first experiment in this paper to investigate this threshold.

References

- [1] Sundberg, J. (1987) "A Rhapsody on Perception," *The Science of the Singing Voice*, Northern Illinois University Press.
- [2] Castellengo, M. (1993) "Fusion or separation: from vibrato to vocal trill," *Stockholm Music Acoustics Conference*, Stockholm, Sweden.
- [3] Miller, G. and Heise, G. (1950) "The trill threshold," *Journal of the Acoustical Society of America*, vol. 22, no. 5, pp. 637 – 638.
- [4] Shonle, J. and Horan, K. (1976) "Trill threshold revisited," *Journal of the Acoustical Society of America*, vol. 59, no. 2, pp. 469 – 471.
- [5] d'Alessandro, C. and Castellengo, M. (1994) "The pitch of short-duration vibrato tones," *Journal of the Acoustical Society of America*, vol. 95, no. 3, pp. 1617 – 1630.
- [6] Jairazbhoy, N. A. (1995) "The Rāgs of North Indian Music," Popular Prakashan Pvt. Ltd., Mumbai.
- [7] Battey, B. (2004) "Bézier spline modeling of pitch-continuous melodic expression and ornamentation," *Computer Music Journal*, vol. 28, no. 4, pp. 25 – 39.
- [8] tHart, J., Collier, R. and Cohen, A. (1990). "A perceptual study of intonation," Cambridge U.P., UK,
- [9] Dikshitar, S. (1904) "Sangita Sampradaya Pradarshini." English translation available at <http://www.ibiblio.org/guruguha/ssp.htm>
- [10] Jayalakshmi, R. S. (2002) "Gamakas explained in Sangita-sampradaya-pradarsini of Subbarama Diksitar," *Ph.D. Dissertation*, University of Chennai.
- [11] Bapat, A. and Rao, P. (2005) "Pitch tracking of voice in a tabla background by the Two-Way Mismatch Method," *Proc. Of the 13th Int. Conf. On Advanced Computing and Communications*, Coimbatore, India.
- [12] Subramanian, M. (2002) "Analysis of Gamakams in Carnatic music using the computer," *Sangeet natak*, vol. 37, no. 1, pp. 26 – 47.