# Addition of Indian Classical Music symbols used in Gurmukhi Script to the Unicode Character Set.

Hardeep Singh Jawanda[1], Dr. K.S Mann[2]

[1]Head Computer Engineering Department, Guru Nanak Dev Polytechnic, Ludhiana, Affiliated to Punjab Technical Education Board, Chandigar

[2]Head Information Technology Department, Guru Nanak Dev Engg. College, Ludhiana.Affiliated to IKG Punjab Technical University, Kapurthala.

Abstract: This Paper highlights the need of new Gurmukhi unicode characters to be included in the unicode character set. Indian classical music notation in Gurmukhi script is not present in the unicode character set. The work of translitration from Gurmukhi to Shahmukhi carried out previously by the author found that there are no unicode characters to represent Indian classical music notation written in Gurmukhi script. The base of transliteration carried out is the unicode character set, which acts as a common platform for both the scripts. Indian classical music books or literature available in Gurmukhi scripts are difficult to transliterarte due to non availability of Gurmukhi Indian classical notations in the unicode character set. This paper is an effort to highlight this problem and to find out a possible solution to this problem. An effort to include Indian classical notation written in Gurmukhi script to the standard Unicode Character Set is also discussed.

## I.    INTRODUCTION

Unicode is vast character base that promises to include and represent all popular languages of the world. Slowly the base is increasing as new languages, characters, and symbols are being included on the regular basis. The work of transliteration from Gurmukhi script to Shahmukhi script and vice versa is a mammoth task. Important literature in both the scripts needs to be transliterated for the spread of good knowledge. The two scripts are akin is Punjabi speaking, but still, these scripts are far away from each other in writing. The consonants, half-consonants, and salient words are some of the problems faced in transliteration. Unicode acts is a bridge between these two scripts. Unicode is only useful in transliteration if the consonants of the script are present in Unicode. The problem here is that while transliteration the Indian classical music literature notations are not present in the unicode set. Therefore the process of transliteration stops. This is threfore necessary to add the new unicode characters representing Gurmukhi script Indian classical musical notations into the unicode character set.

The complete paper is divided into various sections. The first section talks about the basic of Indian classical music. The second section talks about the Indian Classical Music Notations and Gurmukhi Script. The third section discusses the unicode character set and the regional language usage. The forth section explores the related work on the expansion and inclusion of regional charaters in the unicode character set, Indian classical musical notations in particular. The fifth section highlights the effort to include new Gurmukhi consonant of Indian classical musical notations into Unicode character set.

### A.    Section I - Basics of Indian Classical Music

Indian classical music is one of the oldest forms of music in the world. It has its roots in diverse areas such as the ancient religious vedic hymns, tribal chants, devotional temple music, and folk music. Indian music is melodic in nature, as opposed to Western music which is harmonic. Although Indian music is now divided into the two major classes of Hindustani (Northern Indian) and Carnatic (Southern Indian), the origins and fundamental concepts of both the types of music are the same.

The fundamental concepts that have to be understood at the outset are those of swara (musical note), raga (a melodic concept, or scale of notes) and tala (beats of timing or rhythm).

Swara:- Swara is a note of definite pitch which conveys an expression of charm and sweetness to the mind. Shruti is "the measure and the swara is the thing measured." There are seven basic swaras. These remain same but there appearance can be different based on the language used. It is as follows:-

SHADJA        -        Sa
RISHABH       -        Re
GANDHAAR      -        ga

| MADHYAM | - | Ma |
|---------|---|-----|
| PANCHAM | - | Pa |
| DHAIVAT | - | Dha |
| NISHAAD | - | Ni |

Types of swaras(notes) :-

Komal Swara:- From the above 7 swaras Rishabh, Gandhar, Dhaivat, Nishad can be moved below there original place on the scale. These types of swaras are called as komal (Soft or Flat).

Teevra Swara:- From the above 7 swaras only the Madhyam swara can be made Sharp by placing the vertical line above the symbol.

In Indian Classical music 3 Saptaks (Octaves) are usually utilized:-

Saptak: When the set of seven notes is played in the order it is called a Saptak (i.e. Sa , Re, Ga, Ma, Pa, Dha, Ni)

Maddhya Saptak: The normal tone of human voice, which is neither high nor low. It is called Maddhya Saptak (Middle Octave). This has got no symbol in the notation system.

Taar Saptak: The one higher than Maddhya Saptak is Taar Saptak (High). The notes are high and sharp. This shown by a dot above the note. Two dots above the note imply a note of an octave higher than the Taar Saptak i.e. Ati Taar Saptak.

Mandra Saptak: The one octave below the Maddhya Saptak is called Mandra Saptak (Low). Notes of this octave are sung or played in a low deep tone. This comprises of the saptak which is below the lower Sa of the Maddhya Saptak. Notes of this saptak are indicated by a dot below the note.

B.  *Section II - Indian Classical Music Notations and Gurmukhi Script.*

Indian Classical Music Notation has a different version, based on the language (and scripts) of different regions of India. But Fundamental Elements of Indian Classical Music Notation are same for all versions and have same rules. Following are the Characters that are common in all version.

1)  Teevar
2)  Komal
3)  Mandar
4)  Taar
5)  Kan Teevar
6)  Kan Komal
7)  Kan Mandar
8)  Kan Taar
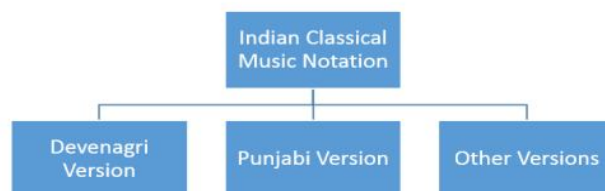


In Gurmukhi Script ਸ ਰੇ ਗ ਮ ਪ ਧ ਨੀ are the primary swar(notes) used in Indian Classical Music. Below are swar (notes) with their Unicode Counterpart in Hexadecimal.

ਸ = 0A38

ਰੇ = 0A30 + 0A47

ਗ = 0A17

ਮ = 0A2E

ਪ = 0A2A

ਧ = 0A27

ਨੀ = 0A28 + 0A40

**Name : Komal** — This symbol is same as 'b' symbol used in Sheet Music Notation. in Indian Classical Music, it is only used with the following swar(notes): ਰੇ ਗਾ ਧ ਨੀ and Result will be look like :

**in Gurmukhi Block :** There is similar character available in Unicode Standard in Gurmukhi Block: which has Unicode (Hex) value '0A41' But Unicode Encoding doesn't support combination like ਨੀ ਰੇ which is correct according to Punjabi Language. But it is required in Indian Classical Music Notation.

**Name: Teevar** — This symbol has the same usage as '#' in Sheet Music Notation. But in Indian Classical Music ( Gurmukhi Script) it is used only with: ਮ ( Unicode Hex '0A2E') The result will look like :

There is similar Unicode character available which has Unicode (Hex) Value: 030D But it has positioning problem like:

**Name: Mandar** — This swar ਮ symbol is used to lower the octave in Indian Classical Music. It is used with every (note). Result will look like: ਸ ਰੇ ਗਾ ਮ ਪ ਧ ਨੀ · ਰੇ ਗਾ ਮ ਧ ਨੀ

**Name: Taar** — This symbol is used to change to octave one up. E.g Octave 4 to Octave 5. In Indian Classical Music (Gurmukhi Script) it is used with every swar (note). The result will look like: ਸ ਰੇ ਗਾ ਮ ਪ ਧ ਨੀ · ਰੇ ਗਾ ਮ ਧ ਨੀ. As you can check that if there is another symbol with music note than this symbol is shifted to right of symbol, else it remains in center. Unicode has a similar symbol in this block which has Unicode (Hex) Value : 0A02. But it has some problems.

Some time it is required to write the whole music character in a manner similar to superscript way Like: ਸ ਨੀ

In Indian Classical Music, it is called as Kan.

For this we required the superscript versions of followings:

ਸ = 0A38

ਰੇ = 0A30 + 0A47

ਗਾ = 0A17

ਮ = 0A2E

ਪ = 0A2A

ਧ = 0A27

ਨੀ = 0A28 + 0A40

and all previous mentioned new Symbols.



Gurmukhi Music Notation Characters

| | | | | |
|---|---|---|---|---|
| ਸ Sa | ਰੇ Re | ਗਾ Ga | ਮ Ma | ਪ Pa |
| ਧ Dha | ਨੀ Ni | Teevar | Komal | Mandar |
| Taar | ਸ Kan Sa | ਰੇ Kan Re | ਗਾ Kan Ga | ਮ Kan Ma |
| ਪ Kan Pe | ਧ Kan Dha | ਨੀ Kan Ni | Kan Teevar | Kan Komal |
| Kan Mandar | Kan Taar | | | |

Devenagri Music Notation Characters

## C. Section II – Unicode and Regional Language Usage.

A language can be written with the help of a script and character set available in it. To access regional languages on a computer a standard universal encoding scheme is required. Most of the data stored and used in computers are in the ASCII standard. ASCII has 128 charcaters and at the most 256 in case of extended ASCII which is very less to accommodate all regional languages of the world. To use Indian regional languages an ASCII based font needs to be installed. Different fonts for the same language give conflicting results as there is no standard. The problem is sorted out by using a standard unicode scheme. The Unicode Character Set is one such effort that brings all the regional or local languages under one ecoding scheme. A computer can use any regional language if unicode encoding scheme is implemented. Unicode comes in two varints, UTF-8 bit and a UTF-16 bit. Unicode is accepted by a large number of IT companies and it also accepted by ISO under ISO/IEC 10646. The plus point of this encoding scheme is that there is always a space to accommodate new or missing script consonants. To include new or missing consonants of a script in unicode there are set of rules already defined in it. To produce softwares that are internationally acceptable, a software must break the barrier of language. Microsoft also from the beginning of NT based operating systems adapted unicode as its native text encoding scheme.

## D. Section III – Related work.

Unicode is still evolving and there are many characters which are yet to be included in the unicode encoding scheme. During this research, no literature was found to discuss the inclusion of Indian classical music notes written in Gurmukhi. Some literature on the inclusion of missing Gurmukhi character is discussed in this section. Sarabveer Singh of New Jersey, United States highlighted a problem with the current Gurmukhi Unicode consonatns. The problem is with representing half consonants. A consonant with adition of another half consonant must be represented as a new consonant. He proposed inclusion of new Gurmukhi consonants as in the table below.
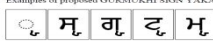
| Character Properties Graphical Appearance | Character Sequence | Character Name |
|---|---|---|
|  | 0A75 | GURMUKHI SIGN YAKASH |
|  | 0A31 (Proposed) | GURMUKHI LETTER YYA (Proposed) |
| Examples of proposed GURMUKHI SIGN YAKA  Expected Result | | |

Table showing work proposed by Sarabveer Singh

*E.  Section IV – Proposal for Inclusion of Gurmukhi consonant*

This proposal is to include the Indian Classical Music symbols with Unicode Character Set. The new symbols are 15, namely, Komal, Teevar, Mandar, Taarand also the super script versions of these symbols and some of the existing symbolsto the Gurmukhi block of the Unicode. First of all just have look at the basics of Indian Classical Music so we will get proper knowledge of required and existing symbols.

TABLE 1: Showing all the Swaras(notes) used in the Indian Classical Music Notation (in Gurmukhi script)

| NAME | SYMBOL | UNICODE | REMARKS |
|---|---|---|---|
| GURMUKHI SHADAJ (Sa) | ਸ | 0A38 | This is the first swara(note) in any Indian Classical Music script. This is the shuddha (pure) swara. |
| GURMUKHI RISHABHA (Re) | ਰੇ | 0A30 + 0A47 | This is the second swara(note) in any Indian Classical Music script. |
| GURMUKHI GANDHAR (Ga) | ਗ | 0A17 | This is the third swara(note) in any Indian Classical Music script. |
| GURMUKHI MADHYAM (Ma) | ਮ | 0A2E | This is the fourth swara(note) in any Indian Classical Music script. |
| GURMUKHI PANCHAM (Pa) | ਪ | 0A2A | This is the fifth swara(note) in any Indian Classical Music script. This is also the shuddha(pure) swara(note) as Sa. |
| GURMUKHI DHAIVAT (Dha) | ਧ | 0A27 | This is the sixth swara(note) in any Indian Classical Music script. |
| GURMUKHI NISHAD (Ni) | ਨੀ | 0A28 + 0A40 | This is the last swara(note) in any Indian Classical Music script. |
| KAN SWAR GURMUKHI SHADAJ (Sa) | ◌ ਸ | NOT IN UNICODE | Same as Shadja(Sa) swara(note) but used as superscript. Here the dotted circle will represent the other swara(note) in music notation. |
| KAN SWAR GURMUKHI RISHABH (Re) | ◌ □□ | NOT IN UNICODE | Same as Rishabha(Re) swara(note) but used as superscript. Here the dotted circle will represent the other swara(note) in music notation. |
| KAN SWAR GURMUKHI GANDHAR (Ga) | ◌ ਗ | NOT IN UNICODE | Same as Gandhar(Ga) swara(note) but used as superscript. Here the dotted circle will represent the other swara(note) in music notation. |
| KAN SWAR GURMUKHI MADHYAM (Ma) | ◌ ਮ | NOT IN UNICODE | Same as Madhyam(Ma) swara(note) but used as superscript. Here the dotted circle will represent the other swara(note) in music notation. |
| KAN SWAR GURMUKHI PANCHAM (Pa) | ◌ ਪ | NOT IN UNICODE | Same as Pancham(Pa) swara(note) but used as superscript. Here the dotted circle will represent the other swara(note) in music notation. |
| KAN SWAR GURMUKHI DHAIVAT (Dha) | ◌ ਧ | NOT IN UNICODE | Same as Dhaivat(Dha) swara(note) but used as superscript. Here the dotted circle will represent the other swara(note) in music notation. |
| KAN SWAR GURMUKHI NISHAD (Ni) | ◌ □□ | NOT IN UNICODE | Same as Nishad(Ni) swara(note) but used as superscript. Here the dotted circle will represent the other swara(note) in music notation. |

TABLE 2: Proposed Diacritical Marks used with the above Musical Notation Base Characters and are common for most of the Indian Classical Music Scripts.

| REMARK | NAME | SYMBOL | UNICODE |
|---|---|---|---|
| This is a symbol which is used to represent higher octave. | TAAR SAPTAK | | NOT IN UNICODE |
| This is the symbol which is used to represent the lower octave. | MANDRA SAPTAK | | NOT IN UNICODE |
| This is a symbol which makes the swara(note) flat. | KOMAL | | NOT IN UNICODE |
| This is the symbol which makes the swara(note) sharp. | | | NOT IN UNICODE |
| Same as symbol 1 but used as superscript. It is used with kan swaras. | SUPERSCRIPTED TAAR SAPTAK | | NOT IN UNICODE |
| Same as symbol 2 but used as superscript. It is used with kan swaras. | SUPERSCRIPTED MANDRA SAPTAK | | NOT IN UNICODE |
| Same as symbol 3 but used as superscript. It is used with kan swaras. | SUPERSCRIPTED KOMAL | | NOT IN UNICODE |
| Same as symbol 4 but used as superscript. It is used with kan swaras. | SUPERSCRIPTED TEEVAR | | NOT IN UNICODE |
| This is combining swara mark which can be used to combine 2 swaras(notes) to n swaras(notes). It has the formatting which is similar to as underline. | COMBINING SWAR MARK (TO REPRESENT COLLECTION OF SWARS) | Examples of usage: (i) (ii) (iii) | NOT IN UNICODE |

## II.   JUSTIFICATION OF EACH SYMBOL

There are altogether 23 symbols in the above tables. Among these 23 symbols, 7 symbols are already defined in the Unicode chart. Now there is certain symbols indicated in Table 1 which are not in the Unicode but they are just superscript of certain symbols of Gurmat Sangeet (Indian classical music gurmukhi version) as descripted in Table 1. So, there are a total of 16 (9 Symbols + 7 grace notes) need to be encoded.

The justification of each symbol (marked 'not in Unicode' in above tables) to encode it in UCS is given below:–

1) *Mandra saptak* symbol is required in Gurmat sangeet( rooted in Indian Classical Music) for lowering the octave as described earlier. It is used with all the swaras (notes + grace note too) even with other symbols such as of komal and teevra( descripted below). It is similar to gurmukhi sign nukta having Unicode '0A3C' but there is positioning problem with the swaras (notes). In actual we require output which is shown below in examples.

2) *Taar saptak* symbol is similar to the symbol gujrati sign anusvara having Unicode '0A82' but unfortunately it can only be used in the Gujrati language. The swaras(notes) having this symbol are sung in high and sharp voice. In Indian classical music, it is used with every swar (note) and we will have the results in the following manner:-

As we can see that this symbol is placed in the center of any swar (note ) till there is no other symbol placed with it and if any other symbol is placed it is shifted to the right side. We require a similar symbol in gurmukhi block of unicode character set so that we can get the required output as shown in figure-1.

3) *Komal* is same as 'b' symbol used in Sheet Music Notation. In Indian Classical Music, it is only used with the following swar(notes):   ਰੇ ਗ ਧ ਨੀ and Result will be look like :

There is similar character available in Unicode Standard in Gurmukhi Block which has Unicode (Hex) value '0A41' But Unicode Encoding doesn't support combination like:

Because it is not correct according to the Punjabi Language. But it is must and is required in Indian classical music example section given later in this proposal.

4) *Teevra* is similar to the '#' symbol in the sheet music notation. But in Indian Classical Music notation, it is used only with the swara (note): ਸ ( Unicode Hex '0A2E') and the results will look like this In the present time, in Unicode, we have a similar symbol to it but the result is not valid. It shows certain position problem like this:-

5) *Kan swara* is just the superscript version of all the main 7 swars in Indian Classical music. It s actually the grace note (of Western music).Although it is used less frequently, but without it Indian classical music can't be completed. It has appearance like this:-In the above example, the swaras (notes) Rishabh (ਰੇ) and Gandhar (ਗਾ) are represented as Kan swaras. In Indian Classical Music these are important for the proper rendition and essential to create the beauty of a raga. Some notes are linked with its preceding and succeeding note; these linked notes are called Kan-swars (= *grace notes*). Kan-swars deal with so called *touch notes*.

A. *Unicode Character Properties proposed.*

| |
|---|
| 0A52;Superscript GurmukhiShadja(Sa);Lm;0;L;<super> 0A38;;;;N;;;;; |
| 0A53;Superscript GurmukhiRishabh(Re);Lm;0;L;<super> 0A30 0A47;;;;N;;;;; |
| 0A54;SuperscriptGurmukhiGandhar (ga);Lm;0;L;<super> 0A17;;;;N;;;;; |
| 0A55;SuperscriptGurmukhiMadhyam(Ma);Lm;0;L;<super> 0A2E;;;;N;;;;; |
| 0A56;SuperscriptGurmukhiPancham (Pa);Lm;0;L;<super> 0A2A;;;;N;;;;; |
| 0A57;SuperscriptGurmukhiDhaivat(Dha);Lm;0;L;<super> 0A27;;;;N;;;;; |
| 0A58;SuperscriptGurmukhiNishad(Ni);Lm;0;L;<super> 0A28 0A40;;;;N;;;;; |
| 0A76;Komal Swara;Mn;220;NSM;;;;;N;;;;; |
| 0A77;Teevra Swara;Mn;232;NSM;;;;;N;;;;; |
| 0A78;Mandra;Mn;220;NSM;;;;;N;;;;; |
| 0A79;Taar;Mn;230;NSM;;;;;N;;;;; |
| 0A7A;Superscripted Komal Swara;Mn;220;NSM;<super> 0A76;;;;N;;;;; |
| 0A7B;SuperscriptedTeevar Swara;Mn;232;NSM;<super> 0A77;;;;N;;;;; |
| 0A7C;Superscripted Mandar;Mn;220;NSM;<super> 0A78;;;;N;;;;; |
| 0A7D;SuperscriptedTaar;Mn;230;NSM;<super> 0A79;;;;N;;;;; |
| 0A7E;Combining Swara;Mn;233;NSM;;;;;N;;;;; |

Table 3 Proposed Draft of Unicodes to be included.

| NAME | PROPOSED UNICODE |
|---|---|
| KAN SWAR GURMUKHI SHADJA (Sa) | 0A52 |
| KAN SWAR GURMUKHI RISHABH (Re) | 0A53 |
| KAN SWAR GURMUKHI GANDHAR (Ga) | 0A54 |
| KAN SWAR GURMUKHI MADHYAM (Ma) | 0A55 |
| KAN SWAR GURMUKHI PANCHAM (Pa) | 0A56 |
| KAN SWAR GURMUKHI DHAIVAT (Dha) | 0A57 |
| KAN SWAR GURMUKHI NISHAD (Ni) | 0A58 |
| KOMAL | 0A76 |
| TEEVRA | 0A77 |
| MANDRA | 0A78 |
| TAAR | 0A79 |

| SUPERSCRIPTED KOMAL | 0A7A |
| SUPERSCRIPTED TEEVRA | 0A7B |
| SUPERSCRIPTED MANDRA SAPTAK | 0A7C |
| SUPERSCRIPTED TAAR SAPTAK | 0A7D |
| COMBINING SWAR MARK (TO REPRESENT COLLECTION OF SWARS) | 0A7E |

*A. Examples*

Green marked areas in figure-1 shows the usage of Taar symbol in the Raga Majh of Swar samud book. Here the red marked areas show the wrong usage of Komal symbol in Gandhar(Ga) swara(note).
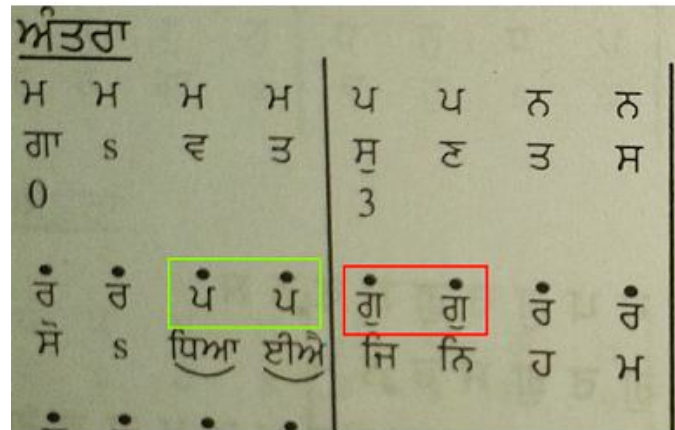


Figure -1: Usage of Taal and Komal symbol

Green marked area ( circled area ) in figure-2 shows the use of combing swar mark where it combines 2 swaras in one area and 4 swaras in other in a raga. As a compositon from Gurbani sangeet book (part -1,2) by sh. Gian Singh ji Ebtabaad.



Figure -2: Combing swar mark

In figure-3 Red marked ( squared ) areas show the wrong use of teevra swara Madhyam (Ma) due to lack of Unicode. So there is almost printing problems in every book printed.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
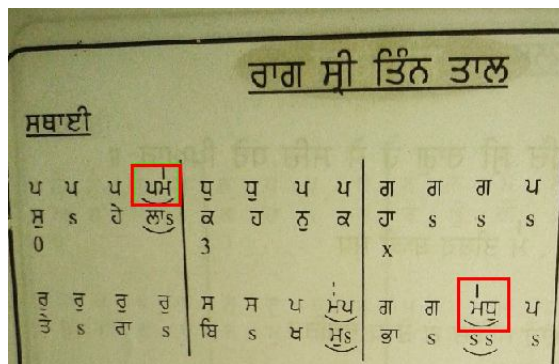Volume 5 Issue XII December 2017- Available at www.ijraset.com

Figure-3: Wrong use of teevra swara Madhyam

In figure-4 third vibhaag (part), under the second cross sign, of this composition the correct use of mandra saptak is shown which indicates that the respective swaras (notes) are of lowest octave (mandra saptak).
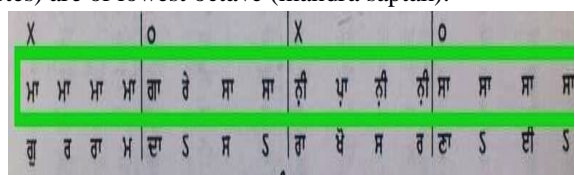


Figure-4: Correct use of mandra saptak

In figure-5 green marked areas ( squared first line ) shows the correct usage and symbol of komal swar (note) in the composition of Raga Bhairo. Red marked areas (squared fourth line ) shows the positioning problem for taar saptak (high octave) symbol on the shadja (Sa) swar (note) and rishabh (Re) swar (note)
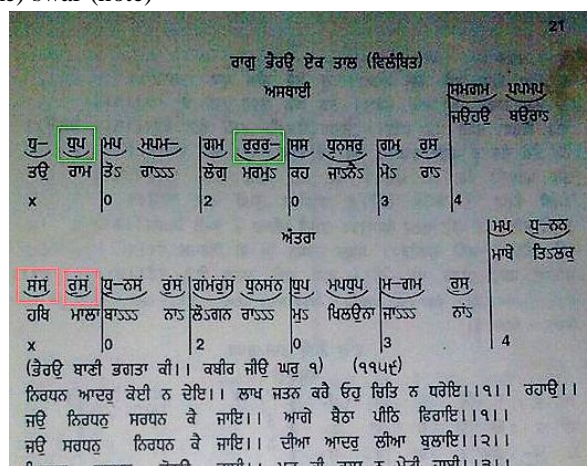


Figure-5: Correct usage and symbol of komal swar (note) and  positioning problem for taar saptak (high octave)

## BIBLIOGRAPHY

[1]  Swar samud (Shabad Reet Ratnawaali) -Book by prof. Paramjot Singh published by Sikh Missionary College, Ludhiana.
[2]  Gurbani Sangeet (part -1,2) -Book by sh. Gian Singh ji Ebtabaad published by Shromani Gurudwara Prabhandhak Committee (SGPC),
[3]  Sh. Amritsar Saral Sikh Sangeet Reet -Book by prof. Kulwant Singh Chandan published by Chattarsingh and Jeevansingh
[4]  Preetpal Kaur Buttar and Jaswinder Kaur, "A SURVEY OF LANGUAGE-DETECTION, FONT DETECTION AND FONT-CONVERSION SYSTEM FOR INDIAN LANGUAGES," International Journal on Cybernetics & Informatics (IJCI) , vol. 4, no. 3, pp. 15 - 23, 2015.
[5]  Jürgen Bettels, F. Avery Bishop, "Unicode: A Universal Character Code," D igital Technical Journal, vol. 5, no. 3, pp. 21 - 31, 1993.
[6]  S. Singh, "Changes to Gurmukhi 10," 27 Oct. 2016. [Online]. Available: http://unicode.org/L2/L2016/16294-gurmukhi-chg.pdf. [Accessed 12 Nov. 2017].
[7]  Gurpreet Singh and Manoj Sachan, "Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition," in 2014 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2014.

# INTERNATIONAL JOURNAL
# FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊘ (24*7 Support on Whatsapp)