International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Proposed Gurmukhi-to-Braille Mapping and Preprocessing Rules for Grade 1 Braille Conversion of Punjabi Text

Charanjiv Singh Saroa[1a], Kawaljeet Singh[1]

Submitted: 05/10/2025 **Revised**: 20/11/2025 **Accepted**: 28/11/2025

Abstract: The conversion of Punjabi text written in Gurmukhi script into Braille is an important step to ensuring educational accessibility to the visually impaired person. Existing Punjabi-to-Braille converters are limited to accept only Unicode-encoded. This paper proposes a comprehensive Grade 1 Braille mapping for Gurmukhi script and introduces a set of preprocessing rules that resolve ambiguities arising from the shared use of Braille codes across multiple characters. The proposed mapping covers independent vowels, consonants, dependent vowel signs (matras), conjunct characters, numerals, punctuation marks, and special symbols. Key contributions include: (i) a modified mapping for the character ੲ to eliminate conflict with the numeric prefix code ⠼; (ii) context-based disambiguation rules for Braille codes shared by multiple characters; (iii) preprocessing rules for handling mixed numeric-alphabetic sequences, arithmetic operators, and conjunct consonants (Paar-ਹ); and (iv) an extended character encoding scheme for special symbols using a prefix convention. The paper also describes integration with an existing ASCII-to-Unicode font conversion module, enabling the system to process both Unicode and ASCII-based Punjabi text. The proposed framework provides a principled foundation for automated Grade 1 Braille conversion of Punjabi and serves as a basis for future Grade 2 converter development.

Keywords: *Braille conversion; Gurmukhi script; Punjabi accessibility; Bharati Braille; Grade 1 Braille; text preprocessing; non-Unicode font processing; visually impaired education*

Introduction

Braille is the primary medium through which visually impaired individuals access written content. For learners in linguistically diverse regions such as India, the availability of Braille materials in regional languages is a prerequisite for meaningful participation in education and public life. Despite the existence of Bharati Braille (India's national Braille standard), the availability of Punjabi-language Braille materials remains severely limited.

Text-to-Braille converters address this gap by automating the transformation of digital text into Braille-encoded output suitable for embossing or refreshable display. While converters for Roman-script languages are widely available, those supporting Indian regional languages, and Punjabi in particular, are sparse and technically limited. Most existing Punjabi text-to-Braille converters operate by performing a direct Unicode character-to-Braille mapping, without addressing the practical disambiguation rules, preprocessing requirements, and non-Unicode encoding challenges that characterize real-world Punjabi digital content.

[1]*Punjabi University, Patiala 147002, India*
[a]*Corresponding author: cjsinghpup@gmail.com |*
ORCID: 0000-0002-8940-9157

A further complication is that the majority of existing Punjabi digital content spanning educational materials, websites, and archival texts is encoded in ASCII-based fonts rather than standard Unicode [1]. These fonts assign different byte values to the same Gurmukhi glyphs, making them mutually incompatible and unprocessable by Unicode-dependent conversion systems. Any practical Punjabi-to-Braille system must therefore address the non-Unicode encoding problem as a preprocessing concern.

This paper makes the following specific contributions:

- Grade 1 Mapping: Comprehensive mapping between all characters of Gurmukhi script and Bharati Braille including all vowels, consonants, matras, digits, punctuation marks, and special symbols along with some modifications according to Braille Press Ludhiana.
- Mapping modification for ਞ: A specific modification to the default Bharati Braille mapping of the retroflex nasal consonant ਞ to solve its clash with numeric prefix indicator.
- Preprocessing Rules: Pre-processing steps defined to prepare the Punjabi text for conversion into Grade 1 Braille such as conversion of numeric-alphabetic sequence, arithmetic symbols, conjunct consonant treatment, quote marks, and expansion

of special symbols.

- Special symbol encoding: A systematic encoding method of special symbols in numeric prefix format to increase the number of characters from 63 standard Braille patterns.
- Non-Unicode Integration: An explanation of the integration of the proposed system with existing module for ASCII font detection and conversion into Unicode [1].

The remainder of the paper is organized as follows. The next section provides background on Gurmukhi script and Bharati Braille, followed by a review of related work, the proposed character mapping, the preprocessing and disambiguation rules, the extended character encoding scheme, integration with non-Unicode processing, a discussion, and a conclusion.

Background

Gurmukhi Script

Gurmukhi is the script used for writing the Punjabi language in India. It is an abugida a writing system in which consonants carry an inherent vowel sound that is suppressed or replaced by explicit vowel signs (matras). The script comprises 35 consonants, 10 independent vowels, a set of dependent matras, three carrier letters (ੳ, ਅ, ੲ), and several additional characters including nasalization marks (bindi ◌ਂ, tippi ◌ੰ), the addak ◌ੱ for geminate consonants, and the virama ◌੍ for forming conjunct clusters.

For Braille conversion purposes, the character set of Gurmukhi presents two principal challenges. First, the inventory of characters exceeds the 63 unique patterns available in the standard 6-dot Braille cell, necessitating shared codes and prefix-based extensions. Second, the context-dependent nature of several characters particularly the relationship between independent vowels and their corresponding matras, and the formation of conjunct consonants requires conversion rules that go beyond direct character-to-symbol substitution.

Bharati Braille

Bharati Braille is the national Braille standard for India, designed to represent the phonological systems of all major Indian scripts including Gurmukhi, Devanagari, Bengali, Tamil, Telugu, and others. A key feature of Bharati Braille is that phonetically equivalent characters across different Indian scripts share identical Braille codes. For example, the Gurmukhi consonant ਕ (ka), the Devanagari consonant क, and the Bengali consonant ক all map to the same Braille pattern ⠅ (dots 1-3). This cross-linguistic consistency facilitates literacy transfer for users of multiple Indian languages.

Bharati Braille is available in three grades. Grade 1 is a letter-by-letter transcription in which each print character has a direct Braille equivalent. Grade 2 employs contractions for common words and letter combinations. Grade 3 is a personal shorthand used for note-taking. The present work focuses exclusively on Grade 1, which forms the foundation for all Braille literacy instruction and is required for technical, scientific, and mathematical content.

Non-Unicode Punjabi Fonts

There have been no efforts towards the standardization of Punjabi digital content using Unicode until now. The vast amount of existing Punjabi textual data present on websites, e-books, and other archives still uses the legacy ASCII based fonts such as Asees, Akhar, Anmollipi, Satluj, Chatrik, Gurbani Akhar, and Joy. These fonts use different byte codes for the same Gurmukhi glyphs, thus making it impossible to develop a Punjabi-to-Braille converter which takes the input only in the form of Unicode strings [1].

Related Works

Kaur et al. developed one of the early systems for Gurmukhi to Braille translation which used letter-by-letter translation into Grade 1 Braille. They used the mapping of Unicode based Gurmukhi letters to their Braille equivalents and claimed an accuracy rate of 90%, 95% and 80% for characters, numerals and words respectively [2]. But this system is limited to Unicode inputs and does not solve the problem of conversion rules and transformations required for actual Braille printing.

Pushe and Kaur discussed a number of problems which are encountered in Punjabi to Braille translation such as character ambiguity, diacritical mark handling and compound character handling [3].Samota and Joshi conducted a survey of challenges and prospects in Punjabi-to-Bharati Braille translation [4]. Their study identified technical and linguistic hurdles including variations in Gurmukhi orthography, complexities in conjunct character representation, and the limited availability of validated Punjabi-Braille parallel corpora for evaluation.

Singh and Saroa addressed the non-Unicode encoding problem directly, developing a multilingual ASCII-to-Unicode conversion system for Indic scripts using n-gram frequency analysis [1]. Their work provides the Unicode conversion foundation on which the present system builds.

None of the existing Punjabi-to-Braille systems provides a complete treatment of all three requirements that practical Braille production demands:

comprehensive character coverage, systematic disambiguation of shared codes, and preprocessing rules grounded in operational guidelines from Braille production facilities.

Proposed Gurmukhi-to-Braille Character Mapping

Mapping Design Principles

The proposed mapping is grounded in the Bharati Braille standard and refined according to operational guidelines provided by Braille Press Ludhiana, an institution with established experience in producing Braille educational materials for schools serving blind students in Punjab. The mapping adheres to three principles: (i) phonetic consistency with Bharati Braille across Indian scripts; (ii) unambiguity, such that every Braille sequence produced has a unique interpretation for a trained Braille reader; and (iii) completeness, such that every character in the Gurmukhi Unicode block that occurs in standard Punjabi text has a defined Braille representation.

Complete Character Mapping Table

Table 1 (next page) presents the complete proposed mapping for Gurmukhi characters to Bharati Braille Grade 1 codes. The entry that differs from the existing mapping (ਣ) is discussed below.

Gurmukhi Character	Unicode	Existing Braille	Proposed Braille	Note
ੳ (oora)	U+0A73	⠥	⠥	
ਅ (airaa)	U+0A05	⠁	⠁	
ਆ (aa)	U+0A06	⠜	⠜	
ੲ (eeri)	U+0A72	⠕	⠕	Independent carrier
ਇ (i)	U+0A07	⠕	⠕	
ਈ (ii)	U+0A08	⠔	⠔	
ਉ (u)	U+0A09	⠥	⠥	
ਊ (uu)	U+0A0A	⠳	⠳	
ਏ (e)	U+0A0F	⠑	⠑	
ਐ (ai)	U+0A10	⠌	⠌	
ਓ (o)	U+0A13	⠕	⠕	
ਔ (au)	U+0A14	⠬	⠬	
ਅੰ (anusvara)	U+0A05+U+0A70	⠉	⠉	
ਆਂ (anusvaara)	U+0A06+U+0A02	⠄	⠄	
CONSONANTS				
ਸ (sa)	U+0A38	⠎	⠎	
ਹ (ha)	U+0A39	⠓	⠓	
ਕ (ka)	U+0A15	⠅	⠅	
ਖ (kha)	U+0A16	⠇	⠇	
ਗ (ga)	U+0A17	⠛	⠛	
ਘ (gha)	U+0A18	⠋	⠋	
ਙ (nga)	U+0A19	⠣	⠣	
ਚ (cha)	U+0A1A	⠉	⠉	
ਛ (chha)	U+0A1B	⠡	⠡	

ਜ (ja)	U+0A1C	⠒⠓	⠒⠓	
ਝ (jha)	U+0A1D	⠒⠓	⠒⠓	Context-disambiguated
ਞ (nya)	U+0A1E	⠒⠒	⠒⠒	Context-disambiguated
ਟ (tta)	U+0A1F	⠓⠓	⠓⠓	
ਠ (ttha)	U+0A20	⠒⠓	⠒⠓	
ਡ (dda)	U+0A21	⠒⠒	⠒⠒	
ਢ (ddha)	U+0A22	⠓⠓	⠓⠓	
ਣ (nna)	U+0A23	⠒⠓	⠒⠓⠓	MODIFIED — see below
ਤ (ta)	U+0A24	⠒⠓	⠒⠓	
ਥ (tha)	U+0A25	⠒⠓	⠒⠓	
ਦ (da)	U+0A26	⠒⠓	⠒⠓	
ਧ (dha)	U+0A27	⠒⠓	⠒⠓	
ਨ (na)	U+0A28	⠒⠓	⠒⠓	
ਪ (pa)	U+0A2A	⠒⠒	⠒⠒	
ਫ (pha)	U+0A2B	⠒⠒	⠒⠒	Context-disambiguated
ਬ (ba)	U+0A2C	⠒	⠒	
ਭ (bha)	U+0A2D	⠒	⠒	
ਮ (ma)	U+0A2E	⠒⠒	⠒⠒	
ਯ (ya)	U+0A2F	⠓⠓	⠓⠓	
ਰ (ra)	U+0A30	⠓⠒	⠓⠒	
ਲ (la)	U+0A32	⠓	⠓	
ਵ (va)	U+0A35	⠓⠓	⠓⠓	
ੜ (rra)	U+0A5C	⠓⠓	⠓⠓	
NUKTA (LOAN-SOUND) CONSONANTS				
ਸ਼ (sha)	U+0A38+U+0A3C	⠒⠒	⠒⠒	
ਖ਼ (kha)	U+0A16+U+0A3C	⠒⠓	⠒⠓	
ਗ਼ (ghain)	U+0A17+U+0A3C	⠲⠓	⠲⠓	
ਜ਼ (za)	U+0A1C+U+0A3C	⠒⠓	⠒⠓	
ਫ਼ (fa)	U+0A2B+U+0A3C	⠒⠒	⠒⠒	
ਲ਼ (lla)	U+0A32+U+0A3C	⠓	⠓	
DEPENDENT VOWEL SIGNS (MATRAS) & DIACRITICS				
◌ਂ (bindi)	U+0A02	⠲	⠲	
◌ੰ (tippi)	U+0A70	⠒	⠒	
◌ੱ (addak)	U+0A71	⠲	⠲	

◌ਾ (aa matra)	U+0A3E	⠨⠄	⠨⠄	
◌ਿ (i matra)	U+0A3F	⠈⠔	⠈⠔	
◌ੀ (ii matra)	U+0A40	⠈⠔	⠈⠔	
◌ੁ (u matra)	U+0A41	⠱⠱	⠱⠱	
◌ੂ (uu matra)	U+0A42	⠳⠳	⠳⠳	
◌ੇ (e matra)	U+0A47	⠢⠢	⠢⠢	
◌ੈ (ai matra)	U+0A48	⠨⠔	⠨⠔	
◌ੋ (o matra)	U+0A4B	⠱⠢	⠱⠢	
◌ੌ (au matra)	U+0A4C	⠈⠢	⠈⠢	
NUMERALS				
1	—	⠨⠂	⠨⠂	
2	—	⠨⠆	⠨⠆	
3	—	⠨⠒	⠨⠒	
4	—	⠨⠲	⠨⠲	
5	—	⠨⠢	⠨⠢	
6	—	⠨⠖	⠨⠖	
7	—	⠨⠶	⠨⠶	
8	—	⠨⠦	⠨⠦	
9	—	⠨⠔	⠨⠔	
0	—	⠨⠴	⠨⠴	
PUNCTUATION MARKS				
। (dandi / full stop)	U+0964	⠲⠲	⠲⠲	
‥ hyphen (–)	—	—	‥	Dot-36
・ comma (,)	—	—	・	Dot-2
⠒ colon / semicolon / ਞ	—	—	‥	Context-disambiguated
⠱⠱ opening quote / ?	—	—	⠱⠱	Context-disambiguated
⠨⠆ closing quote / ਝ	—	—	⠨⠆	Context-disambiguated
⠒⠒ ! / ਫ	—	—	⠒⠒	Context-disambiguated
( ) { } [ ]	—	—	⠶⠶	Unified: Dot-2356

Table 1: Proposed Gurmukhi-to-Bharati Braille Grade 1 mapping.

Proposed Modification: ਣ (Retroflex Nasal)

The standard Bharati Braille mapping assigns the code ⠨⠄ (dots 3-4-5-6) to the Gurmukhi retroflex nasal consonant ਣ. However, this is the same code used as the numeric prefix indicator in Braille the pattern that signals that subsequent symbols represent digits rather than letters. This creates a fundamental ambiguity whenever ਣ appears adjacent to, or within a word containing, a digit.

Ambiguity example: The word ਅਣਬਣ would be transcribed as ⠠ ⠨⠆ ⠨⠄. A reader encountering ⠨⠆ in this sequence could interpret it either as the numeric combination '2' (numeric prefix + digit b) or as the

consonant ਣ followed by ਬ. The ambiguity is irresolvable without semantic context.

Proposed solution: The double code ⠼⠼ is proposed to represent ਣ exclusively, reserving single ⠼ as the numeric prefix. Under this convention: single ⠼ always signals the start of a numeric sequence; double ⠼⠼ always represents the consonant ਣ; the word ਅਣਬਣ is unambiguously transcribed as ⠠ ⠼⠼⠒ ⠼⠼.

Preprocessing and Disambiguation Rules

Effective Grade 1 Braille conversion requires preprocessing transformations that resolve ambiguities before the character-to-Braille mapping stage. The rules described in this section are derived from the operational guidelines of Braille Press Ludhiana and are listed in the order in which they should be applied.

Numeric-Alphabetic Sequence Handling

When a word contains both digits and Gurmukhi alphabetic characters (e.g., 123ਚਲੋ or ਏ4), the numeric prefix ⠼ applied to the digit portion could be misinterpreted as extending into the alphabetic portion. Two preprocessing rules address this:

Rule 1: Hyphen insertion: Where a digit sequence is immediately followed by Gurmukhi alphabetic characters, a hyphen "-" is inserted at the boundary. For example, 123ਚਲੋ is preprocessed to 123-ਚਲੋ, which is then encoded as ⠼⠁⠃⠉⠤⠉⠇⠕. The hyphen code ⠤ (dot-3-6) marks the end of the numeric region.

Rule 2: Digit expansion within words: Where a digit is embedded within a Gurmukhi word containing ਣ, or where positional ambiguity cannot otherwise be resolved by hyphen insertion, the digit is expanded to its Punjabi word equivalent. For example, ਏ4 is preprocessed to ਏ-ਚਾਰ, where ਚਾਰ is the Punjabi word for 'four'. This eliminates any residual ambiguity between numeric codes and the ਣ modification described above.

Arithmetic Operator Expansion

Standard arithmetic operators are not represented directly in Bharati Braille for Punjabi educational materials. Instead, Braille Press Ludhiana replaces them with their full Punjabi word equivalents before conversion. Table 2 (next page) lists the standard expansions.

Symbol	Punjabi Word (Gurmukhi)	Meaning
+	ਪਲੱਸ	Plus / Addition
−	ਘਟਾਓ	Subtraction
×	ਗੁਣਾ	Multiplication
÷	ਵੰਡ	Division
=	ਬਰਾਬਰ	Equal

Table 2: Arithmetic operator expansion to Punjabi word equivalents.

Handling of Paar-ਰ (Subjoined Ra)

The subjoined ra, known in Punjabi as Paar-ਰ, forms conjunct consonants when the virama ੍ precedes ਰ within a word. Its Braille representation depends on its position within the word:

Case 1: Paar-ਰ at the first character position: When the conjunct ੍ਰ occurs with the first consonant of the word, the Braille sequence begins with ⠠ (dot-4) as a Paar marker, followed by the Braille code for the first consonant, then ⠺ for ਰ. For example, the word ਪ੍ਰਤੀ is encoded as ⠠⠏⠺⠞⠊, where ⠠ signals the Paar-ਰ, ⠏ encodes ਪ, ⠺ encodes ਰ, ⠞ encodes ਤ, and ⠊ encodes the matra ੀ.

Case 2: Paar-ਰ not at the first character position: When the conjunct ੍ਰ occurs at a non-initial position within the word, the ਰ is represented by its standard Braille code ⠺ without the ⠠ prefix.

Context-Based Disambiguation of Shared Codes

Several Braille codes in Bharati Braille represent more than one character in Punjabi text. The following disambiguation rules, applied in order during preprocessing, resolve these conflicts:

Rule 3: ⠨ (dots 3-5-6): This code represents both the consonant ਝ and the closing double quotation mark. If ⠨ is preceded by the opening quote code ⠦ (dots 2-3-6) within the same sentence, it is interpreted as the closing double quote. Otherwise, it is read as the consonant ਝ.

Rule 4: ⠦ (dots 2-3-6): This code represents both the opening double quotation mark and the question mark. If ⠦ appears at the end of a sentence (followed by a space, newline, or sentence-final boundary), it is interpreted as a question mark. Otherwise, it is read as the opening double quotation mark.

Rule 5: ⠖ (dots 2-3-5): This code represents both the consonant ਫ and the exclamation mark. If ⠖ appears at

the end of a sentence, it is interpreted as an exclamation mark. Otherwise, it is read as the consonant ਫ.

Rule 6: ⠒ (dots 2-5): This code represents the consonant ਏ, the colon, and the semicolon. Braille readers distinguish these three interpretations based on the grammatical and syntactic context of the surrounding text.

Table 3 (next page) summarizes the disambiguation rules.

Code	Characters Represented	Disambiguation Rule
⠨⠱ (dots 3-5-6)	ਝ / " (closing quote)	If preceded by ⠱⠂ in same sentence → closing quote; otherwise → ਝ
⠱⠂ (dots 2-3-6)	" (opening quote) / ?	If at sentence-final position → question mark; otherwise → opening quote
⠱⠒ (dots 2-3-5)	ਫ / !	If at sentence-final position → exclamation mark; otherwise → ਫ
⠒ (dots 2-5)	ਏ / : / ;	Determined by grammatical and syntactic context

Table 3: Context-based disambiguation rules for Braille codes shared by multiple characters.

Unified Code for Bracket Pairs

Given the 63-pattern limit of the 6-dot Braille cell, assigning separate codes to each bracket variant would consume a disproportionate number of codes. Braille Press Ludhiana uses the single code ⠶ (dots 2-3-5-6) for all opening and closing bracket characters: parentheses ( ), curly braces { }, and square brackets [ ]. The paired and nested structure of Punjabi prose provides sufficient context for readers to determine whether a given bracket is opening or closing.

Full-Word Expansion of Special Symbols

Symbols such as #, @, &, $, and ₹ are replaced with their full Punjabi equivalents before Braille conversion. This approach ensures complete comprehensibility for student Braille readers. Table 4 (next page) lists the standard expansions used by Braille Press Ludhiana.

Symbol	Punjabi Expansion (Gurmukhi)
#	ਹੈਸ਼
@	ਐਟਰੇਟ
&	ਐਂਡ
$	ਡਾਲਰ
₹	ਰੁਪਏ

Table 4: Full-word expansion of special symbols to Punjabi equivalents.

Single-Quote Handling

Single quotation marks in Punjabi text frequently appear as abbreviated forms of common words. In these cases the abbreviated form is expanded to the full word before conversion. For example, the abbreviated form 'ਚ (meaning 'in') is expanded to ਵਿੱਚ, and 'ਤੇ (meaning 'on') is expanded to ਉੱਤੇ. Where single quotes serve a purely punctuational role, they are replaced with double quotation marks.

Extended Character Encoding

The 63 unique patterns available in the 6-dot Braille system are insufficient to assign distinct codes to all characters required in Punjabi text, particularly special symbols, technical punctuation, and mathematical characters. To address this, the proposed system employs an extended encoding scheme using the numeric prefix ⠨⠼ followed by a character code. Under this scheme, ⠨⠼ signals that the following symbol should be interpreted as a special extended character rather than a letter or numeral.

Table 5 (next page) presents the complete extended character encoding scheme.

Symbol	Extended Braille Code	Symbol	Extended Braille Code
' (single quote)	⠨⠼⠒	! (exclamation)	⠨⠼⠖

Symbol	Extended Braille Code	Symbol	Extended Braille Code
((open paren)	⠲⠶	& (ampersand)	⠲⠯
) (close paren)	⠲⠾	# (hash)	⠲⠼
{ (open brace)	⠲⠷	$ (dollar)	⠲⠫
} (close brace)	⠲⠿	< (less-than)	⠲⠣
[(open bracket)	⠲⠻	> (greater-than)	⠲⠜
] (close bracket)	⠲⠽	, (comma)	⠲⠂
+ (plus)	⠲⠖	: (colon)	⠲⠒
* (asterisk)	⠲⠔	; (semicolon)	⠲⠆
@ (at)	⠲⠈	" (double quote)	⠲⠐
? (question mark)	⠲⠦		

Table 5: Extended character encoding using the numeric prefix ⠲ convention.

The prefix convention is extensible: any special character not listed in Table 5 can be accommodated by assigning a unique ⠲-prefixed code, following the same systematic pattern. This design ensures forward compatibility as new special characters emerge in digital Punjabi text.

Integration with Non-Unicode Text Processing

A significant proportion of existing Punjabi digital content is encoded in legacy ASCII-based fonts rather than Unicode. These fonts including Asees, Akhar, Anmollipi, Satluj, Chatrik, Gurbani Akhar, and Joy each assign different byte values to the same Gurmukhi glyphs, making them incompatible with Unicode-based processing tools. A Braille converter that accepts only Unicode input is therefore unable to process the majority of available Punjabi text.

The proposed Grade 1 mapping and preprocessing system is designed to accept Unicode Gurmukhi as its input. For non-Unicode Punjabi text, the system is integrated with an ASCII-to-Unicode conversion module [1] that performs two operations:

Font detection: The ASCII font type is identified using n-gram frequency analysis across seven supported fonts [1].

Character mapping: Each ASCII byte value is mapped to its corresponding Unicode Gurmukhi code point using a per-font lookup table. The conversion applies character reordering, virama normalization, and diacritical sequence canonicalization before producing NFC-normalized Unicode output [1].

Conversion module output is standard Unicode text in Gurmukhi script that then goes through the pre-processing module described above to follow by character to Braille mapping. Such integration allows complete end-to-end processing of text in Punjabi script regardless of initial encoding used.

Figure 1 describes the overall processing flow from raw input to Braille output.

Input: Punjabi text (Unicode or legacy ASCII font)
↓
[If non-Unicode] Font Detection → Character Mapping → NFC Unicode [1]
↓
Preprocessing Module (Rules 1–9)
↓
Character-to-Braille Mapping (Table 1)
↓
Output: Bharati Braille Grade 1

Figure 1: End-to-end processing pipeline from raw Punjabi text to Bharati Braille Grade 1 output

Discussion

Significance of the ਣ Modification

The most substantive departure from the standard Bharati Braille mapping proposed in this paper is the double-code representation of the retroflex nasal ਣ. This modification directly addresses a structural collision in the original standard that arises specifically in the Punjabi context, where ਣ is a phonologically active consonant appearing in common Punjabi words. The modification does not affect the encoding of any other character and preserves full backward compatibility for texts that do not contain ਣ.

The cost of this modification is a marginal increase in Braille cell consumption for words containing ਣ. This is negligible in practice given the relatively low frequency of ਣ in standard Punjabi text.

Grounding in Braille Press Ludhiana Guidelines

A distinguishing feature of the proposed mapping is its grounding in the operational guidelines of Braille Press Ludhiana an institution with direct experience producing Braille educational materials for schools serving visually impaired students in Punjab. As mentioned above, the rules of preprocessing, especially the arithmetic operator expansion rule, bracket unification, symbol to word expansion, and abbreviation expansion rules, follow conventions that have been proved valid by continued application.

This is different from previous research efforts aimed at converting Punjabi Braille into plain text, which focused on developing mappings only according to the Bharati Braille standard.

Limitations

There are several restrictions associated with the proposed system which need to be stated. Firstly, the ambiguity resolution process requires context such as sentence breaks or other codes, and in case of an ambiguous situation, manual inspection might be needed. Secondly, the proposed system is only applicable for the standard written form of Punjabi, and does not consider the archaic and dialectal words, very specialized technical information or bilingual documents using Punjabi and English/Hindi together. Thirdly, the extended characters representation system is a proposed convention and needs to be standardized first before its application.

Future research can focus on the formal standardization of the proposed modifications, empirical testing of the disambiguation rules on a valid corpus and development of grade 2 Punjabi contractions.

Conclusion

A complete framework for grade 1 Braille conversion of Punjabi text in Gurmukhi script has been proposed in this paper. The framework has four parts - a full mapping of Gurmukhi characters, a specific modification of Bharati Braille mapping scheme for character ਣ, a set of pre-processing and disambiguation rules based on the operation of Braille Press Ludhiana, and a new encoding scheme for special characters.

The framework proposed in this paper remedies the two major problems in current Punjabi to Braille converters – incomplete character mapping and lack of disambiguation rules for common Braille codes. In addition, the use of this framework along with existing ASCII to Unicode font converter will extend Braille conversion support to large amount of digital content in legacy ASCII based Punjabi fonts.

This research lays down the theoretical and practical foundation for automation of Punjabi Braille translation with direct applicability in conversion of educational materials for institutes that cater to visually challenged students of Punjab. In addition, the grade 1 Braille framework developed here serves as the foundation for development of grade 2 Punjabi Braille with contractions.

References

[1] R. Singh and C. S. Saroa, "Multilingual conversation ASCII to Unicode in Indic script," Computer Science & Information Technology (CSIT), vol. 7, no. 3, pp. 83–94, Feb. 2017, doi: 10.5121/csit.2017.70308.

[2] R. Kaur, Vandana, and N. Bhalla, "Gurmukhi script to Braille code," Int. J. Eng. Res. Appl. (IJERA), vol. 2, no. 4, pp. 1298–1302, 2012. Available: www.ijera.com

[3] V. Pushe and R. Kaur, "Issues in converting Punjabi to Braille," Asian J. Comput. Sci. Inf. Technol., vol. 1, no. 1, pp. 12–15, 2011. Available: http://www.innovativejournal.in/index.php/ajcsit

[4] H. Samota and N. Joshi, "Exploring the landscape of Punjabi to Bharati Braille translation: A comprehensive examination of challenges and prospects," Int. J. Modern Res. Eng. Technol., vol. 8, no. 10, pp. 1–8, 2023. Available: www.ijmret.org