International Journal of Research in Management & Social Science
Volume 14, Issue 2: April - June 2026
ISSN 2322 - 0899

DEVELOPMENT OF PUNJABI TEXT TO SPEECH SYSTEM WITH PROSODY ENHANCEMENT FOR INTERROGATIVE SENTENCES

Charanjiv Singh Saroa[1] and Kawaljeet Singh[2]
[1],[2]Punjabi University, Patiala 147002 India
[1]cjsinghpup@gmail.com, ORCID: 0000-0002-8940-9157

ABSTRACT
This paper presents the development of a comprehensive Punjabi Text-to-Braille and Text-to-Speech (TTS) system that addresses accessibility needs of visually impaired Punjabi speakers. The system introduces automatic conversion of non-Unicode Gurmukhi script data to Unicode, followed by Braille transcription and prosody-enhanced speech synthesis. A key contribution is the application of prosody rules for interrogative Punjabi sentences, distinguishing between yes/no (ki) questions with rising intonation and WH-questions with falling intonation. The system employs SSML with the Google Cloud Text-to-Speech API for online deployment, and a recorded-sound fallback for offline use. Validation with 30 human evaluators showed that 86.6% rated prosody-enhanced audio as more natural than baseline TTS output. Spectrogram analysis confirmed the acoustic effectiveness of the proposed intonation adjustments.

Keywords: *Punjabi TTS, Gurmukhi script, Braille conversion, prosody, intonation, SSML, accessibility, interrogative sentences, Unicode*

1. INTRODUCTION
Accessibility technologies for regional languages remain underserved, particularly for visually impaired communities in South Asia. Punjabi, one of the 22 officially recognized languages of the Indian Constitution and spoken by over 100 million people worldwide, is written in the Gurmukhi script—a script with significant structural complexity, including tonal markers, vowel diacritics, and conjunct characters.

Visually impaired individuals rely on Braille and audio-based assistive technologies for information access. However, existing TTS and Braille systems largely target major world languages, leaving Punjabi speakers with limited or inadequate tools. A dual-mode system that converts Gurmukhi text to both Braille and natural-sounding speech would substantially improve access to education, public information, and digital content for this community.

A particularly challenging aspect of TTS for Punjabi is prosody—the patterns of stress and intonation that convey meaning and emotional tone in spoken language. Interrogative sentences in Punjabi follow specific intonation patterns: yes/no questions rise in pitch toward the end, while WH-questions (who, what, where, when, why, how) begin at a higher pitch and fall toward the end . Without explicit modeling of these patterns, synthesized Punjabi speech sounds unnatural and robotic, reducing its utility for end users.

This paper makes the following contributions: (1) a unified Punjabi Text-to-Braille and TTS pipeline with automatic non-Unicode to Unicode conversion; (2) a prosody algorithm that identifies and applies intonation rules to interrogative sentences using SSML tags; (3) validation of the system through human perception evaluation and spectrogram analysis.

2. BACKGROUND AND RELATED WORK

2.1 Punjabi Language and Gurmukhi Script
Punjabi is an Indo-Aryan language with three lexical tones—level, rising, and falling—encoded in the Gurmukhi script through specialized tone markers called

Punjabi is an Indo-Aryan language with three lexical tones—level, rising, and falling—encoded in the Gurmukhi script through specialized tone markers. The script comprises 35 base characters, 10 vowel diacritics (*lāgā mātrā*), and several sub-script characters . Tonal and intonational features are closely intertwined, making prosody modeling a non-trivial task.

Legacy Punjabi digital content often uses ASCII-based proprietary fonts such as Asees, Anmollipi, or Gurbani Akhar, rather than Unicode-compliant Gurmukhi encoding (Unicode range U+0A00–U+0A7F). Interoperability between these formats requires automatic font detection and transliteration to Unicode .

2.2 Braille Systems for Indian Languages
Bharati Braille is a unified Braille code designed for Indian languages, mapping Devanagari-derived characters across scripts including Gurmukhi. Prior work has addressed Braille conversion for Hindi, Gujarati ,Tamil,

ISSN 2322 - 0899

Kannada, Bangla, Odia, and Arabic. For Punjabi specifically, Vandana et al. proposed a Gurmukhi-to-Braille implementation, and Pushe and Kaur documented structural challenges. Samota and Joshi provided a comprehensive examination of prospects for Punjabi-to-Bharati Braille translation. Despite these efforts, a production-ready system integrating Unicode normalization, Braille output, and natural speech remains absent from the literature.

2.3 Text-to-Speech for Indian Languages

Several TTS systems have been developed for Indian languages, including Dhvani, Shruti, and Vani for Hindi, and Singh and Lehal's system for Punjabi. Prosody modeling—pitch, duration, and energy—remains the most active research frontier in neural TTS. SSML has emerged as a standard markup layer for controlling prosodic features in commercial TTS APIs. However intonation-specific modeling for Punjabi interrogatives has not been systematically addressed in prior work.

3. SYSTEM DESIGN

3.1 System Overview

The proposed system operates as a pipeline with the following stages: (1) input acquisition and font detection, (2) Unicode normalization, (3) sentence segmentation, (4) interrogative sentence classification and flagging, (5) prosody tuning via SSML annotation, (6) audio synthesis (online via Google Cloud TTS API, or offline via recorded phoneme concatenation), and (7) audio output. Simultaneously, the Unicode text is passed through a Braille transcription module that maps Gurmukhi characters to their corresponding Bharati Braille dot patterns.

3.2 Unicode Normalization

The system first detects the encoding of the input text by analyzing character code ranges and known proprietary font signatures. Non-Unicode Gurmukhi text (e.g., Asees, Anmollipi) is mapped to its Unicode equivalent using lookup tables validated against the Punjabi Unicode standard. This step ensures downstream processing—Braille mapping and speech synthesis—operates on a canonical representation.

3.3 Speech Synthesis Markup Language (SSML)

SSML provides XML-based markup for controlling TTS attributes including pitch (

SSML provides XML-based markup for controlling TTS attributes including pitch (<*prosody pitch*>), rate (<*prosody rate*>), volume, pauses (<*break*>), and emphasis (<*emphasis*>). These controls enable developers to encode linguistically motivated prosodic variations directly in the text passed to a TTS engine. In this work, SSML is used to inject intonation patterns appropriate for Punjabi interrogative sentences before synthesis.

Google Cloud Text-to-Speech API was selected for online synthesis due to its WaveNet-based voice models, broad language support, configurable audio formats, and integration with existing cloud infrastructure. For offline or resource-constrained deployments, the system falls back to concatenating pre-recorded phoneme sounds with pitch/duration manipulation applied programmatically.

4. PROSODY ALGORITHM FOR INTERROGATIVE SENTENCES

4.1 Intonation in Punjabi Questions

Intonation—the variation in pitch across an utterance—is a primary cue for sentence type in Punjabi, as in most languages . Two question types require distinct intonation treatment:

Yes/No Questions (*Kī* Questions): These typically begin with the particle □□ and end with a rising pitch contour (L*H%), signaling that the speaker expects a binary affirmative or negative response. Example: □□ □□□□□ □□□□ □□ □□□*?* (Have you eaten?)

WH-Questions (Interrogative Pronoun Questions): These begin with interrogative pronouns such as □□□□ (where), □□□ (who), □□ (what), □□□□ (why), and follow a high-onset falling contour (H*L%), indicating the speaker expects detailed information. Example: □□□□ □□□□ □□ □□□ □□*?* (Where are you going?)

4.2 Algorithm Description

Algorithm 1 presents the complete prosody pipeline for Punjabi TTS. The algorithm processes an input Punjabi text string and produces an audio file with linguistically appropriate intonation.

Algorithm 1: Punjabi TTS Prosody Algorithm for Interrogative Sentences

Input: Punjabi text T (Gurmukhi script, Unicode or non-Unicode)

Output: Audio file with prosody-tuned speech

ISSN 2322 - 0899

1. T_raw ← Input(textPunjabi)
2. f ← DetectFont(T_raw)

T ← ConvertToUnicode(T_raw, f)

3. S ← SplitSentences(T)
4. JSON ← StoreInJSON(S, flag=0)
5. for each s_i in S:

if IsInterrogative(s_i):

JSON[i][flag] ← 1

6. for each s_i in S:

if JSON[i][flag] = 1:

type ← IdentifyInterrogativeType(s_i)

words ← SplitLastWords(s_i, type)

tuned_words ← ApplyProsodyRules(words)

s_i ← ReconstructSentence(s_i, tuned_words)

7. if online_mode:

audio ← GenerateAudio_API(S)

else:

audio ← ConcatenateRecordedSounds(S)

8. SaveAudio(audio)

Algorithm 1: Prosody-enhanced TTS pipeline for Punjabi interrogative sentences.

4.3 Prosody Rules

The ApplyProsodyRules function encodes the following intonation patterns using SSML markup:

Yes/No Questions: The terminal word and its constituent characters are annotated with *<prosody pitch="+20%">* SSML tags to produce a rising intonation contour. The penultimate syllable receives a moderate pitch increase and the final syllable a larger increase, approximating the L*H% tonal target.

WH-Questions: The interrogative pronoun at the sentence onset is annotated with *<prosody pitch="+15%">*, and the terminal word receives *<prosody pitch="-15%">*, implementing the H*L% falling contour characteristic of information-seeking questions.

Character-level splitting is employed to enable fine-grained pitch control over individual syllables where word-level markup alone is insufficient for naturalness.

5. VALIDATION AND RESULTS

5.1 Human Perception Evaluation

To evaluate perceptual quality, 30 native Punjabi speakers were recruited as evaluators. Each evaluator was presented with paired audio samples: one synthesized without prosody tuning (baseline) and one synthesized with the proposed algorithm applied, for both yes/no and WH-question types. Evaluators were asked to judge which sample sounded more natural and question-like.

Results showed that 26 out of 30 participants (86.6%) rated the prosody-enhanced audio as more natural than the baseline. This finding demonstrates that the proposed intonation rules produce a perceptible and preferred improvement in synthesized Punjabi speech for interrogative sentences.

Table 1: Human Perception Evaluation Results

Metric	Count	Percentage

ISSN 2322 - 0899

Metric	Count	Percentage
Total evaluators	30	100%
Preferred prosody-enhanced	26	**86.6%**
Preferred baseline	4	13.4%

5.2 Spectrogram Analysis

Spectrograms of the synthesized audio were examined to verify acoustic correlates of the applied prosody rules. For the closed question

Spectrograms were examined to verify acoustic correlates of the applied prosody rules. For the closed question □□□ □□□□□□ □□ □□□□ □□**?** (Are you going to Patiala?), the spectrogram without prosody tuning showed a flat or slightly declining fundamental frequency (F0) at the utterance end, whereas the prosody-tuned version showed a clear upward F0 modulation in the terminal region, consistent with a rising intonation contour.

For the open WH-question □□□ □□□□□ □□ □□□□ □□**?** (Where are you going?), the baseline spectrogram showed a nearly uniform pitch profile, while the tuned version exhibited elevated F0 at the interrogative pronoun onset and a descending contour toward the sentence boundary. These acoustic signatures confirm that the SSML annotations successfully shape the prosodic contour of the synthesized output in line with the intended phonological targets .

The implemented method generates audio files that undergo validation with the help of a group of people (30). The response from this group is positive, with around 26 (86.6%) out of the 30 individuals involved finding the generated sound is better than the audio produced without implementing intonations. These results are also validated with the help of spectrograms of the audio files

Input Text: ਤੁੰ ਪਟਿਆਲਾ ਜਾ ਰਿਹਾ ਹੈ ?

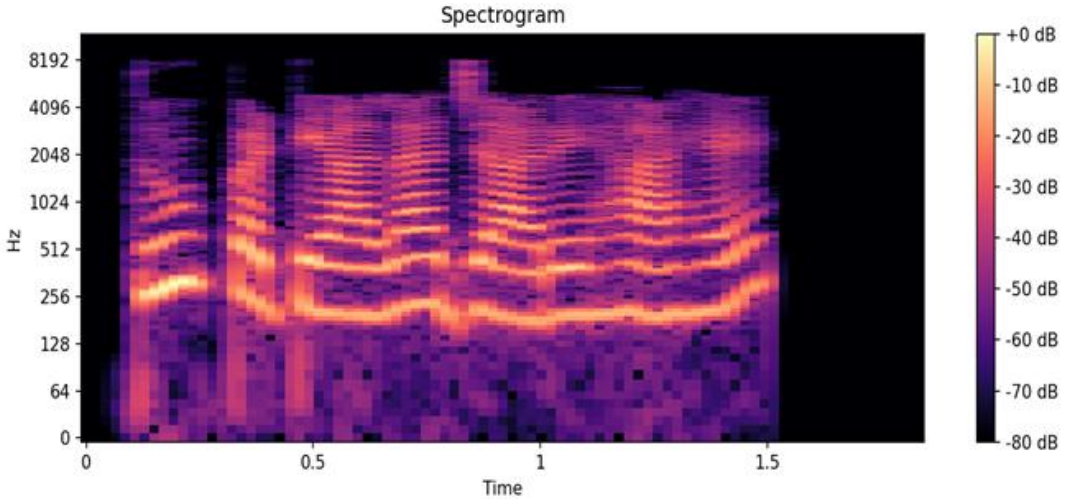

Figure: Spectrogram of Closed Question without Prosody Tuning of Text

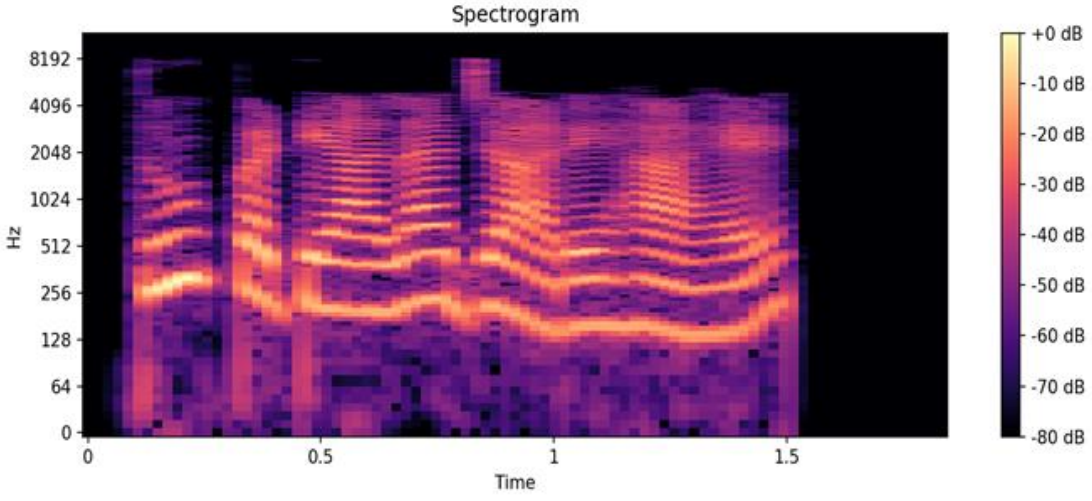

Figure: Spectrogram of Closed Question with Prosody Tuning of Text

Input Text: ਤੂੰ ਕਿੱਥੇ ਜਾ ਰਿਹਾ ਹੈ ?

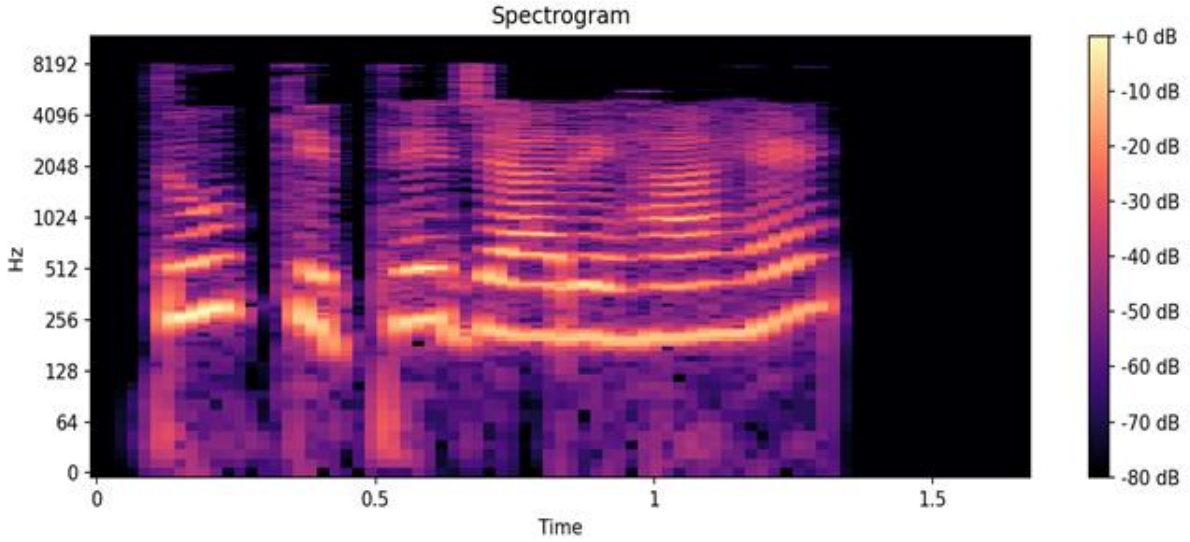

Figure: Spectrogram of Open Question without Prosody Tuning of Text

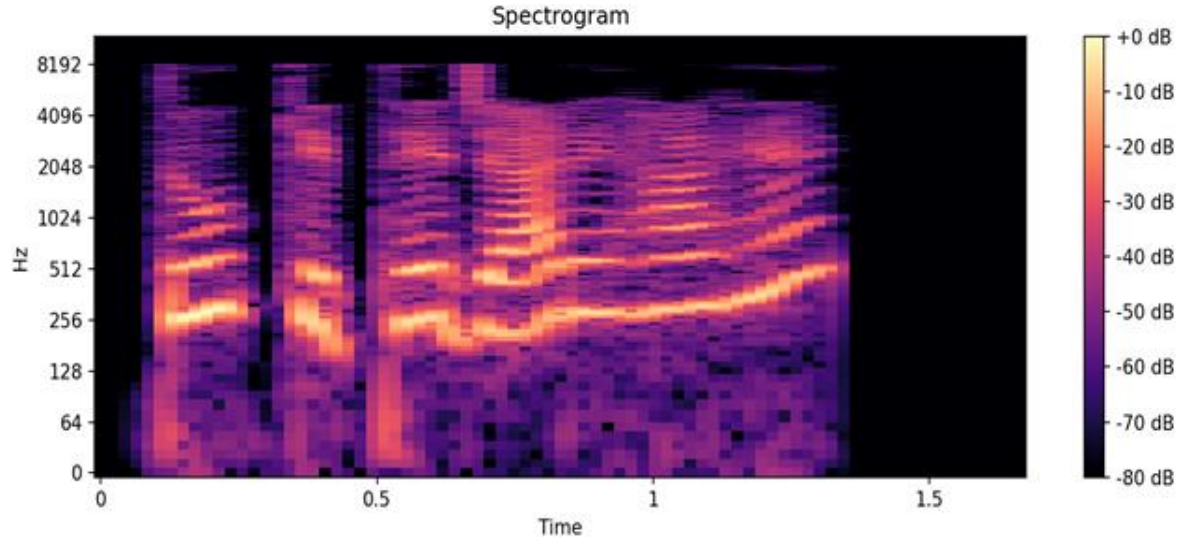

Figure: Spectrogram of Open Question with Prosody Tuning of Text

Above spectrograms shows the slight modulation in sound in the end of the text.

6. DISCUSSION

The proposed system addresses two gaps in the accessibility technology landscape for Punjabi: the absence of a production-ready Braille converter for Gurmukhi text, and the lack of prosody modeling in Punjabi TTS. The 86.6% preference rate in the human evaluation is a strong indicator of practical utility, though it is worth noting that the evaluation was conducted with a relatively small sample (n = 30) and future work should include larger-scale mean opinion score (MOS) studies.

The dual-mode deployment (online API + offline concatenation) is a deliberate design choice motivated by connectivity conditions in rural Punjab, where internet access may be intermittent. The offline mode sacrifices some audio quality but ensures the system remains functional in low-resource settings.

A limitation of the current prosody approach is that SSML pitch controls are applied at the word and character level based on simple rules (e.g., final word of a yes/no question receives a fixed +20% pitch increase). A data-driven approach using a Punjabi prosody corpus, or a neural end-to-end TTS model trained on Punjabi speech, would likely produce more nuanced and natural-sounding output. However, the absence of large annotated Punjabi speech corpora makes this a subject for future work.

Regional dialect variation in Punjabi (Majhi, Doabi, Malwai, etc.) also affects intonation patterns, and the current system does not differentiate between dialects. Extending the system to support dialect-specific prosody models is a natural next step.

7. CONCLUSION

This paper has presented a Punjabi Text-to-Braille and Text-to-Speech system that integrates automatic Unicode normalization, Braille transcription, and prosody-enhanced speech synthesis for interrogative sentences. The system applies linguistically motivated intonation rules—rising pitch for yes/no questions and falling pitch for WH-questions—using SSML markup delivered to the Google Cloud TTS API, with a recorded-sound fallback for offline deployment.

Validation with 30 native Punjabi-speaking evaluators demonstrated that 86.6% preferred the prosody-enhanced audio over the baseline, and spectrogram analysis confirmed the acoustic effectiveness of the applied intonation patterns. The system represents a meaningful step toward accessible, natural-sounding assistive technology for the Punjabi-speaking visually impaired community.

Future work will focus on: building a labeled Punjabi prosody corpus for data-driven prosody modeling; extending dialect-specific intonation support; integrating the system into mobile and web applications for broader accessibility; and conducting larger-scale MOS evaluation studies.

REFERENCES

[1] P. Singh and G. S. Lehal, "Text-to-speech synthesis system for Punjabi language," in Communications in Computer and Information Science, vol. 139, Springer Verlag, 2011, pp. 302–303.

[2] S. Dhanjal and S. S. Bhatia, "Development of a standard text and speech corpus for the Punjabi language."

[3] K. S. Fuglerud, "The Barriers to and Benefits of Use of ICT for People with Visual Impairment," 2011.

[4] I. Eligi and K. Mwantimwa, "ICT accessibility and usability to support learning of visually-impaired students in Tanzania," International Journal of Education and Development using ICT, vol. 13, no. 2, 2017.

[5] R. Singh and C. S. Saroa, "Multilingual Conversation ASCII to Unicode in Indic Script," AIRCC, Feb. 2017, pp. 83–94.

[6] V. Pushe and R. Kaur, "Issues in Converting Punjabi to Braille," 2011.

[7] H. Samota and N. Joshi, "Exploring the Landscape of Punjabi to Bharati Braille Translation," 2023.

[8] Vandana, R. K. Kaur, and N. B. Bhalla, "Implementation of Gurmukhi to Braille," IJCST, vol. 3, no. 2, pp. 596–599, 2012.

[9] D. Hirst, A. François, and D. Cristo, "A survey of intonation systems," 1998.

[10] J. Kanwal and A. Ritchart, "An Experimental Investigation of Tonogenesis in Punjabi."

[11] M. Singh and P. Bhatia, "Automated Conversion of English and Hindi Text to Braille Representation," Int J Comput Appl, vol. 4, no. 6, pp. 25–29, Jul. 2010.

[12] N. B. Jariwala and B. Patel, "Transliteration of digital Gujarati text into printable Braille," in Proc. CSNT 2015, pp. 572–577.

[13] S. Urolagin, K. V. Prema, and N. V. S. Reddy, "Kannada Alphabets Recognition with Application to Braille Translation," Int J Image Graph, vol. 11, no. 3, pp. 293–314, 2011.

[14] C. S. Baboo and V. Devi, "Tamil Braille System: A Conversion methodology of Tamil into Contracted Braille Script (Grade 2)."

[15] K. Parvathi, B. M. Samal, and J. K. Das, "Odia Braille: Text transcription via image processing," in Proc. ABLAZE 2015, pp. 138–143.

[16] R. Hariharan et al., "Dhvani Text to Speech System." [Online]. Available: https://dhvani.sourceforge.net/

[17] A. Mukhopadhyay et al., "Shruti: An embedded text-to-speech system for Indian languages," IEE Proc.: Software, vol. 153, no. 2, pp. 75–79, 2006.

[18] H. Jain, V. Kanade, K. Desikan, and V. Team, "Vani—An Indian Language Text to Speech Synthesizer," 2004.

[19] Y. Ren et al., "Prosospeech: Enhancing Prosody with Quantized Vector Pre-Training in TTS," in Proc. ICASSP 2022, pp. 7577–7581.

[20] Y. Yi et al., "Prosodyspeech: Towards Advanced Prosody Model for Neural TTS," in Proc. ICASSP 2022, pp. 7582–7586.

[21] F. Malfrère, T. Dutoit, and P. Mertens, "Fully Automatic Prosody Generator for Text-to-Speech."

[22] A. Trivedi et al., "Speech to text and text to speech recognition systems—A review," IOSR-JCE, vol. 20, no. 2, pp. 36–43, 2018.

[23] T. K. Bhatin, "The Evolution of Tones in Punjabi," Studies in the Linguistic Sciences, vol. 5, no. 2, pp. 12–24, 1975.

[24] B. Bowden, "Punjabi Tonemics and the Gurmukhi Script: A Preliminary Study."

[25] N. Kaur and P. Singh, "Segmentation of Punjabi Text into Prosodic Unit," ICCSE 2019.

[26] A. S. Al-Salman and H. S. Al-Khalifa, "Towards a computerized Arabic Braille environment," Softw Pract Exp, vol. 33, no. 6, pp. 497–508, 2003.

[27] G. S. Lehal and C. Singh, "A Gurmukhi Script Recognition System."

[28] N. Sadat-Tehrani, "The Intonation Patterns of Interrogatives in Persian," Linguistic Discovery, vol. 9, no. 1, 2012.

[29] S. Francis et al., "Architecture and Implementation of Bangla Text to Braille Translation," 2011.

[30] S. A. Hossain et al., "Analysis of Bangla-2-Braille Machine Translator," 2014.