A Comprehensive Review of Multilingual Audio-to-Braille Conversion Systems and Voice-Assisted Accessibility for Visually Impaired Users

Siddhant Sahu
MIT Academy
of Engineering
catchmesiddhant@gmail.com

Sanket Rath
MIT Academy
of Engineering
official.sanketrath@gmail.com

Aarya Malghe
MIT Academy
of Engineering
aaryamalghe@gmail.com

Aditya Ambure
MIT Academy
of Engineering
ambureaditya8@gmail.com

***Abstract*—Abstract—Braille continues to play a crucial role in literacy, independence, and information accessibility for individuals with visual impairments, yet traditional Braille production remains limitedbymanualtranscription,scarceresources,andhighhard-warecosts.WiththerapidadvancementofArtificialIntelligence, particularly in Automatic Speech Recognition (ASR), Natural LanguageProcessing(NLP),andUnicode-basedBrailleencoding, automatedaudio-to-Braillesystemshaveemergedasapromising area of assistive technology research. This paper presents a comprehensive review of existing technological and applicative workonaudio-to-Brailleconversion,drawingfromIEEEXplore, ACM Digital Library, and other scholarly repositories. State-of-the-artASRsystemssuchasWhisper,GoogleSpeech-to-Text,and transformer-based multilingual models demonstrate significant improvements in noise robustness, language adaptability, and low-resourcelanguageprocessing.Simultaneously,developments in rule-based and machine-learning-driven Braille translation frameworks,alongwiththestandardizationprovidedbyUnicode Braille, have expanded the feasibility of generating accurate tactilerepresentations.Thereviewfurtherexaminesprogress in refreshable Braille displays, piezoelectric actuators, and cost-effective embossing techniques. Despite these advancements, major challenges persist, including the lack of large annotated datasets for Indian and low-resource languages, limited real-timeperformance,highcomputationaldemands,andtheabsence of an integrated multilingual audio-to-Braille pipeline. Thisstudy identifies these gaps and highlights research opportunities aimed at developing scalable, inclusive, and efficient end-to-end Braille accessibility systems for visually impaired communities worldwide.**

***Index Terms*—Audio-to-BrailleConversion,AssistiveTechnology, Speech Recognition, Natural Language Processing, Unicode Braille, Braille Embosser, Multilingual ASR, Accessibility Systems.**

I. Introduction

Visual impairment affects more than 285 million people worldwide, with a substantial proportion relying on Braille for literacy, academic participation, and independent living. Despite being a universally recognized tactile writing system, the creation and dissemination of Braille content remain limited in many regions due to the scarcity of trained transcribers, high costs of embossing equipment, and the time-consuming nature of manual Braille production. In parallel, the rapid growth of digital communication has created a widening accessibility gap for visually impaired users who depend on tactile formats. Addressing this challenge requires the integration of emerging technologies capable of converting audio information—which is abundant and naturally accessible—into Braille representations that are standardized, accurate, and easily producible.

Recent advancements in Artificial Intelligence (AI), particularly in Automatic Speech Recognition (ASR) and Natural Language Processing (NLP), have enabled highly accurate speech-to-text systems that function across diverse linguistic environments. Modern transformer-based models such as Whisper and Google Speech-to-Text demonstrate robustness against noise, speaker variation, and multilingual complexities, opening new possibilities for automated accessibility solutions. Similarly, the development of Unicode Braille encoding (U+2800–U+28FF) and open-source translation tools has created a unified foundation for digital Braille generation and display. Innovations in tactile hardware—including refreshable Braille displays, piezoelectric actuators, and lower-cost embossers—further enhance the applicability of automated Braille systems.

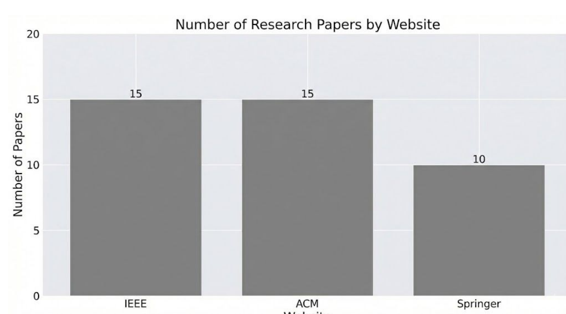

Fig. 1. Distribution of Research Papers by Publication Year and Category

Despite these advancements, the majority of existing solutions address only isolated components, such as speech transcription or Braille mapping, without offering an integrated multilingual audio-to-Braille pipeline. Challenges such as limited datasets for low-resource languages, computational constraints, and the high cost of tactile output devices continue to hinder widespread adoption. This review paper examines current research trends, technological developments, and system architectures in audio-to-Braille conversion, identifies major limitations, and outlines opportunities for developing comprehensive, inclusive, and scalable assistive technologies for visually impaired users.

II. Preliminaries

This section summarizes the basic concepts and technologies behind audio-to-Braille conversion systems. These basics provide the necessary foundation to understand the techniques discussed in the review.

A. Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR) is the method of changing spoken words into text that machines can understand. Modern ASR systems mainly use deep learning frameworks, including Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). More recently, they also incorporate transformer-based encoder-decoder structures. These systems convert audio features into language tokens using models like Connectionist Temporal Classification (CTC) or attention-based methods. Well-known ASR platforms like Whisper, Google Speech-to-Text, and Azure Speech Services can handle multiple languages and perform well in noisy environments.

B. Natural Language Processing (NLP)

Once speech is transcribed, the text must be normalized and processed using NLP techniques. NLP includes tokenization, restoring punctuation, identifying languages, removing stop words, and correcting grammar. These steps help make sure the text is clean and contextually correct before Braille translation. Multilingual NLP has unique challenges because of differences in grammar, structure, and script representation. This is especially true for low-resource languages like Hindi and Marathi.

C. Unicode Braille Encoding

Selected Braille Encoding Rules for English, Hindi, and Marathi (Illustrative Sampler)

Category	Category	Character (Eng)	Character (Hin/Mar)	Braille Cell (Visual)	Dot Indices	Description/Notes
Alphabets & Phonetic Equivalents	Vowel	a	अ		1	Short ‘a’ sound
	Consonant	b	ब		1-2	Phonetic ‘b’
	Consonant	k	क		1-3	Phonetic ‘k’
	Consonant	l	ल		1-2-3	Phonetic ‘l’
	Consonant	t	ट		2-3-4-5	Retroflex ‘t’ (distinct from English ‘t’ which is 2-3-4-5)
Numerals (preceded by Number Sign)	2.1 Digit	1	१		3-4-5-6 + 1	Uses ‘a’ pattern after prefix
	2.2 Digit	2	२		3-4-5-6 + 1-2	Uses ‘b’ pattern after prefix
	2.3 Digit	5	५		3-4-5-6 + 1-5	Uses ‘e’ pattern after prefix
	2.4 Digit	0	०		3-4-5-6 + 2-4-5	Uses ‘j’ pattern for zero
Punctuation & Special Symbols	3.1 Punctuation	,	-		2	Comma
	3.2 Punctuation	.	।		2-5-6	Full Stop (Eng) / Purna Viram (Hin/Mar)
	3.3 Punctuation	?	?		2-3-6	Question Mark
	3.4 Modifier	(Capital)	-		6	Prefix for capital letter (Eng)
	3.5 Modifier	(Number)	-		3-4-5-6	Prefix indicating following chars are digits
	3.6 Diacritic	-	˙		1-3-4-6	Anusvara (Nasal dot in Hin/Mar)

Fig. 2. Braille Encoding Rules for 6-dot Braille Cell Structure

Unicode Braille offers a consistent way to represent Braille characters in digital form using the Unicode block U+2800 to U+28FF. Each character has a pattern made up of 6 or 8 dots, with raised dots that match specific binary values. The Unicode standard ensures that different software platforms, Braille embossers, and refreshable tactile displays can work together. Tools like Liblouis use these mappings for different languages.

D. Braille Output Technologies

Braille output can be generated in two main forms: static embossed Braille and dynamic refreshable Braille. Embossers create raised dots on paper using mechanical pins, making them suitable for printed documents. Refreshable Braille displays, on the other hand, use piezoelectric or electromagnetic actuators to raise and lower dots dynamically, allowing users to read digital content tactilely. Although highly effective, these devices are often expensive, limiting accessibility in developing regions.

E. Evaluation Metrics

Several metrics are used to evaluate the performance of an audio-to-Braille system. For ASR, Word Error Rate (WER) and Character Error Rate (CER) quantify transcription accuracy. For translation and NLP tasks, BLEU score and sentence-level accuracy are commonly employed. Braille output quality is assessed through tactile readability, error rate, and compliance with standard Braille rules.

F. Braille Printers

Braille printers, also known as Braille embossers, are specialized hardware devices designed to produce tactile Braille on thick embossable paper by creating raised dots corresponding to Braille cell patterns. These printers translate digital Braille-encoded files such as `.brf` or `.brl` into embossed output using mechanical impact pins or solenoid actuators. Modern embossers support both single-sided and double-sided (interpoint) embossing. This significantly reduces paper use. They often work with screen readers, text editors, and translation tools like Liblouis to ensure proper Braille formatting and standard dot-spacing alignment. High-performance embossers can produce multiple lines per second and connect through USB, network, or wireless options. However, even with technical improvements, commercial Braille printers remain expensive and need regular maintenance. This limits their use in schools and places with fewer resources. As a result, there is increasing interest in creating cheaper and easier-to-use Braille printing solutions to improve access for visually impaired users.

III. Reviews

A. Braille Performance and Metrics

The paper “Deep Learning Strategy for Braille Character Recognition[1]” proposes a high-accuracy lightweight CNN model for recognizing Braille characters from image inputs and highlights the significant need for technological solutions that support visually impaired individuals. The authors emphasize the global challenge of Braille literacy and the difficulty of memorizing 64 possible dot combinations, reinforcing the need for automated systems that can simplify access to Braille. Their model achieves up to 95.2% accuracy for English Braille and demonstrates real-time processing performance, proving that intelligent Braille technologies are feasible for practical deployment. However, the work is limited to Braille image recognition, revealing a research gap in systems that generate Braille rather than interpret it. This gap strongly supports the relevance of our proposed multilingual Audio-to-Braille system, which focuses on converting speech directly into Braille with voice-assistant guidance for independent and accessible usage.

The paper “Refreshable Braille Display for the Visually Impaired[2]” presents a functional prototype of a low-cost Braille display designed to improve digital accessibility for blind users, emphasizing the global need for affordable tactile reading technologies. The authors highlight key challenges such as the high cost of commercial refreshable Braille devices, limited portability, and the complexity of mechanical actuation systems. Their proposed design—using stepper motors, microcontrollers, and actuated pins—demonstrates the feasibility of creating a scalable tactile output system for reading digital content. However, the work focuses solely on generating Braille from pre-processed text and lacks capabilities for automated input processing, multilingual support, or speech-driven interaction. This limitation directly supports the relevance of our proposed Audio-to-Braille system, which aims to bridge this gap by converting real-time multilingual speech into Unicode Braille while integrating a voice assistant for fully independent, user-friendly accessibility.

The paper “Development and Evaluation of Refreshable Braille Display and Active Touch-Reading System for Digital Reading of the Visually Impaired[3]” presents a highly engineered refreshable Braille display integrated with haptic feedback, demonstrating major advancements in tactile-based digital reading. The authors develop a full-latching electromagnetic Braille actuator that offers high force, low power consumption, and a realistic reading experience, achieving 95.5% recognition accuracy during user studies. This work highlights the necessity of multimodel feedback—touch combined with audio—to enhance comprehension and reading efficiency for visually impaired users. While the system focuses on tactile reproduction of Braille from digital sources, it does not address speech-to-Braille conversion, revealing a technological gap. This directly supports the relevance of our project, which extends accessibility by converting real-time audio into multilingual Braille output, complementing and expanding the digital-reading possibilities explored in this paper.

The paper ”FingerSight: Fingertip Haptic Sensing of the Visual Environment[4]” introduces a novel fingertip-mounted camera–haptic device that converts visual information into tactile feedback, enabling visually impaired users to explore edges and object features through touch. The experiments demonstrate that users can accurately interpret spatial visual cues, highlighting the strong potential of haptic sensory-substitution technologies. While the system effectively enhances environmental perception, it focuses primarily on spatial navigation and object detection rather than text or linguistic content. This distinction underscores a major research gap relevant to our work: existing systems help users feel visual surroundings but not read or receive language. Therefore, the findings strongly justify the need for our proposed multilingual Audio-to-Braille pipeline, which addresses linguistic accessibility by converting speech into tactile Braille output—bridging communication, education, and independent reading for visually impaired users.

The paper “Empowering Individuals With Visual Impairment: A Digital Braille Solution for Learning the Urdu Language[5]” presents a comprehensive mobile-and-hardware ecosystem designed to enhance Braille literacy through an intuitive learning interface, tactile feedback, and integrated progress tracking. This work is highly relevant to our Audio-to-Braille project because it demonstrates the real-world impact of combining digital interfaces, multilingual Braille encoding, and assistive hardware to support independent learning. While the paper focuses on Urdu Braille education rather than speech processing, its emphasis on accessibility, user-centered design, and multimodel feedback directly aligns with our goal of creating a hands-free, multilingual audio-driven Braille generation system. The paper also highlights the effectiveness of structured lessons, tactile modules, and usability testing with visually impaired learners, validating the importance of such features for enhancing our system’s practical adoption.

The paper “Design and Evaluation of an Electromagnetic Bounce-Type Refreshable Braille Display” [6] presents an advanced tactile output mechanism that uses electromagnetic actuation to produce high-speed, high-durability Braille cells suitable for real-time reading. The authors demonstrate significant improvements in refresh rate, power efficiency, and tactile clarity compared to traditional piezoelectric Braille cells, making the technology highly compatible with modern digital reading systems. The work highlights a growing trend toward dynamic, lightweight, and low-power Braille hardware solutions capable of supporting continuous reading, which aligns closely with the goals of assistive technology research. However, the paper focuses solely on tactile output hardware and does not address upstream stages such as speech recognition, text processing, or multilingual Braille generation. This reinforces the importance of our proposed SpaarshVaani system, which not only outputs Unicode Braille compatible with refreshable displays but also integrates multilingual speech-to-text, translation, normalization, and voice-guided interaction—thus forming a complete end-to-end accessibility pipeline that bridges multiple gaps not addressed by hardware-only solutions.

The paper “From Touch to Mental Imagery: The Embodied Aesthetic Experience of Late-Blind People Engaged in the Tactile Exploration of Enrico Castellani’s Pseudo-Braille

Surface[7]" investigates how late-blind individuals cognitively and emotionally interpret tactile artwork through a mixed computational–qualitative approach. The study reveals a powerful transformation from functional tactile exploration to imaginative, aesthetic engagement mediated by mental imagery, autobiographical memory, and embodied interpretation . Participants initially experience confusion or uncertainty, but eventually develop rich mental images, emotional well-being, and a deeper connection with the artwork. This shift—termed a meta-cognitive transition—demonstrates how tactile stimuli can evoke creativity, personal meaning, and emotional empowerment. These findings strongly support the motivation behind our Audio-to-Braille system, which similarly leverages tactile interpretation and cognitive transformation to enhance accessibility and independent interaction with information for visually impaired users.

The paper "Contactless Braille sensing based on GaN optical devices integrated with epoxy lenses[8]" introduces a highly compact, contactless Braille sensing technology that uses monolithically integrated GaN LEDs and photodetectors to detect Braille dot patterns with high speed and stability. The system eliminates mechanical wear issues found in traditional tactile sensors and achieves excellent repeatability over 5000 cycles with fast response times of 24 ms, as demonstrated in Figures 3–6 of the paper. Although the work focuses on reading Braille rather than generating it, the paper highlights a modern shift toward non-contact, high-precision Braille digitization, reinforcing the need for advanced multimodal assistive systems. This aligns with our multilingual Audio-to-Braille project's goal of enabling seamless, real-time Braille accessibility, showing how optical sensing innovations can complement future hardware integration for embosser-based or refreshable Braille displays.

The paper "Visual and Tactile Perception Techniques for Braille Recognition[9]" by Park et al. (2023) presents an innovative dual-modal approach that combines deep-learning–based visual recognition with a flexible capacitive tactile sensor array to improve Braille character identification. The visual module employs a Faster R-CNN–FPN–ResNet-50 model trained on a custom dataset, achieving an mAP50 of 94.8, while the tactile module uses a micro-dome capacitive sensor array capable of generating real-time heatmaps for precise Braille detection. This dual strategy highlights the complementary strengths of vision and touch, addressing limitations related to lighting, distance, and environmental noise. The study underscores a key insight relevant to our project: current research excels at Braille recognition but lacks integrated audio-to-Braille generation systems, validating the significance of SpaarshVaani's speech-driven multilingual Braille output pipeline.

The paper "Deep learning scheme for character prediction with position-free touch screen-based Braille input[10]" presents an innovative touchscreen-based Braille input system that eliminates the need for fixed touch positions, significantly easing interaction for visually impaired users. The authors develop a custom Android interface and collect a real Braille dataset—something previously unavailable—before applying deep learning models such as a Sequential CNN and GoogLeNet for classification. The GoogLeNet model achieves an impressive 95.8% accuracy, demonstrating the feasibility of high-performance Braille recognition using DL-based feature extraction. This work highlights key challenges in Braille input, such as gesture memorization, multi-finger input, and usability barriers, which align closely with the motivations behind our Audio-to-Braille system. Unlike this paper, which focuses on recognizing Braille from touchscreen input, our project aims to generate Braille output from multilingual speech, filling a major research gap in end-to-end accessibility pipelines.

The paper "A Deep Learning Approach for Line-Level Amharic Braille Image Recognition[11]" proposes a segmentation-free Optical Braille Recognition (OBR) method using a CNN–BiLSTM–CTC architecture, addressing the long-standing challenges of dot-level segmentation, noise sensitivity, and multi-cell Braille mapping in Amharic Braille systems. Unlike traditional rule-based and handcrafted-feature systems, this end-to-end deep learning model transcribes full Braille line images with minimal preprocessing and achieves a low Character Error Rate of 7.81% on real-world degraded documents. The authors also introduce the first publicly available Amharic Braille line-image dataset, filling a major resource gap. This work is highly relevant to our multilingual Audio-to-Braille system because it demonstrates that modern AI models can eliminate segmentation barriers and handle complex Braille structures, reinforcing the feasibility of generating accurate Unicode Braille output from automated pipelines.

The paper "BrailleBuddy: A Tangible User Interface to Support Children with Visual Impairment in Learning Braille[12]" presents a novel interactive learning device designed to make Braille acquisition more engaging and accessible for young learners. Through a detailed, inclusive, multi-stage design process, the authors developed a tactile, audio-assisted system that allows children to explore Braille independently using physical word and letter cards that trigger contextual auditory feedback. Their evaluation with visually impaired primary-school children shows strong intrinsic motivation, reduced need for supervision, and improved engagement with spelling and reading tasks. This work highlights an important shift from traditional supervised Braille learning tools toward self-directed, playful learning environments. The study reinforces the relevance of our project by demonstrating that multimodal, interactive systems significantly enhance accessibility, motivation, and learning outcomes—insights that directly support the development of our Audio-to-Braille pipeline.

The paper "The Effect of Orientation on the Readability and Comfort of 3D-Printed Braille[13]" investigates how different 3D-printing orientations influence the tactile clarity, user comfort, and readability of Braille characters, especially for visually impaired users. The authors conduct controlled experiments with FDM printing at multiple orientations and observe that print direction significantly affects dot sharpness and height consistency—critical factors for accurate Braille recognition. Their findings reveal that vertical and diagonal orientations yield more distinguishable Braille dots compared to horizontal layouts, which often cause flattening and tactile discomfort. This study highlights the importance of manufacturing parameters in assistive technology design and supports our Audio-to-Braille project by emphasizing how hardware output quality directly impacts usability, especially when integrating printable Braille generated from automated speech-to-text pipelines.

B. Text Processing Techniques and metrics

The paper ”Deepfake Audio Detection for Urdu Language Using Deep Neural Networks”[14] highlights the increasing sophistication of speech processing models and the importance of reliable audio interpretation to ensure authenticity and prevent misuse. The study demonstrates the effectiveness of advanced deep learning architectures such as RNN+LSTM, CNN+Attention, and hybrid models in accurately extracting speech patterns and identifying subtle variations within audio signals, achieving accuracy of up to 98.89% for real-time classification . These findings support our Audio-to-Braille project by reinforcing the significance of accurate audio feature extraction, preprocessing, and noise reduction for meaningful conversion to Braille. Just as deepfake detection relies on MFCC-based spectral analysis and temporal modeling, our system requires robust speech understanding to produce error-free Braille output. The paper also addresses multilingual dataset challenges, which align with our aim to develop a scalable multilingual Braille generation model for visually impaired users.

The paper “Optical Braille Recognition and its correspondence in the Conversion of Braille Script to Text[15]” provides a comprehensive survey of OCR-based Braille recognition methods and highlights the technical challenges in digitizing embossed Braille documents. The authors systematically analyze preprocessing, segmentation, and feature-extraction techniques required to convert Braille images into readable text, noting issues such as skew correction, noise reduction, recto–verso dot discrimination, and accurate cell detection. While the paper focuses primarily on converting physical Braille into digital text, it also exposes a major research gap—existing systems emphasize recognition rather than generation of Braille. This gap strongly aligns with our Audio-to-Braille project, which addresses the opposite flow by producing Unicode Braille from speech input. The review reinforces the significance of automated, accessible Braille technologies and validates the need for end-to-end multilingual assistive solutions like SpaarshVaani.

The paper “Transliteration of Digital Gujarati Text into Printable Braille[16]” provides valuable insights for our multilingual Audio-to-Braille project by highlighting the complexities of mapping regional languages into Braille formats. The authors describe how Gujarati characters, compound forms, digits, and punctuation must be carefully encoded using UTF-8 before conversion, emphasizing the importance of accurate character mapping tables and rule-based processing. Their work also identifies challenges such as characters with no direct Braille equivalents, multi-symbol representations, and context-dependent pronunciation—all crucial considerations for developing reliable Braille systems. While the study focuses specifically on transliterating digital Gujarati text for embossers, it strengthens the relevance of our project by revealing the need for automated, user-friendly tools that can handle multilingual inputs. Our speech-to-Braille approach extends this concept by eliminating the dependency on typed text and enabling real-time, accessible Braille generation for visually impaired users across languages.

The paper “Text Detection and Communicator Using Braille for Assistance to Visually Impaired[17]” proposes a hardware–software system that converts printed or handwritten text into tactile Braille output using actuators such as solenoids and servo motors. The authors implement a full OCR pipeline—image acquisition, preprocessing, segmentation, and character recognition—followed by Bluetooth-based transmission to an Arduino-driven Braille cell mechanism. Their performance evaluation (Tables II & III on pages 4–5) shows high accuracy across zoom levels and fonts, demonstrating that camera-based Braille generation is feasible and cost-effective. However, the system depends entirely on image input, lacks speech handling, multilingual processing, and voice-guided interaction, making it less flexible for real-world accessibility needs. These limitations validate the motivation behind our proposed multilingual Audio-to-Braille system, which eliminates camera dependence, supports natural speech input, provides voice assistance, and delivers Unicode Braille output suitable for both digital display and embossers.

The paper “From Vision to Voice: A Multi-Modal Assistive Framework for the Physically Impaired[18]” presents a comprehensive OCR–translation–TTS pipeline designed to make visual text accessible to individuals with visual and physical impairments. Its multilingual support, real-time webcam-based text extraction, and integrated speech synthesis demonstrate the value of unified assistive systems that minimize user effort and operate across diverse input formats. While the system focuses on converting images to audio, its modular workflow strongly aligns with the objectives of our project, which aims to convert spoken audio directly into Braille output. The study highlights key gaps—such as the lack of end-to-end modality integration and challenges in handling handwritten or noisy inputs—that validate the need for our extended Audio-to-Braille pipeline, which incorporates speech recognition, text normalization, and Braille conversion for broader accessibility.

The paper ”Text Based Intelligent System demonstrates[19]” how rule-based query rewriting and intelligent text interpretation significantly enhance information retrieval systems by converting unstructured inputs into meaningful structured output. The authors emphasize that auxiliary data and rewrite-rule frameworks improve query understanding by generating alternative meaningful interpretations rather than relying on literal keyword matching. This concept directly aligns with the objectives of our multilingual Audio-to-Braille project, where speech input must be processed, analyzed, and contextually enriched before being translated into Braille. Just as the paper identifies limitations of traditional text-based systems and proposes structured transformation pipelines to improve accessibility and accuracy, our project applies similar principles to convert raw speech into structured tactile output. The structured conversion process and data enrichment strategies discussed in the paper reinforce the need for intelligent preprocessing, rule-driven optimization, and context-aware transformation in building efficient accessibility systems for visually impaired users.

The paper ”TextNet – A Text-Based Intelligent System[20]” demonstrates how deep contextual text processing and incremental knowledge acquisition significantly improve the interpretation of natural language by extracting semantic meaning rather than relying on direct literal mapping. The system’s use of WordNet-based lexical inference, semantic path identification, and coherence evaluation enables intelligent understanding of language structure and context, as shown through examples of interpreting multi-sentence inputs and retrieving supporting information from external sources. These findings are highly relevant to our Audio-to-Braille project, where accurate speech transcription alone is not sufficient—context-aware NLP is essential for converting spoken language into meaningful Braille output that visually impaired users can understand. By leveraging similar inference strategies, our system can enhance multilingual speech understanding, resolve ambiguity, and produce coherent Braille text that preserves intent, structure, and clarity.

The paper ”GEMINI: A Natural Language System for Spoken-Language Understanding[21]” presents a robust architecture for processing spontaneous spoken language by integrating syntactic, semantic, and repair-correction mechanisms to ensure accurate comprehension even in the presence of speech errors or recognition uncertainty. The system demonstrates that natural speech contains disfluencies such as repetitions, self-corrections, and pauses, and effective speech-based systems must handle these intelligently to achieve real-world usability. This insight is highly relevant to our Audio-to-Braille project, where speech input must be reliably interpreted and converted into meaningful structured Braille output. By adopting similar principles—such as handling noisy input, interleaving semantic and syntactic analysis, and applying repair detection—our system can significantly improve transcription accuracy and generate precise Braille content for visually impaired users. The paper reinforces the importance of robust speech understanding as a foundational step before Braille translation.

The paper ”Abstracting of Legal Cases: The SALOMON Experience[22]” demonstrates the effectiveness of automated text extraction and summarization techniques for converting large unstructured documents into meaningful, concise, and accessible summaries. SALOMON employs both deep linguistic processing and shallow statistical clustering to identify key information and remove redundant content, enabling users to access essential knowledge efficiently. This approach is directly relevant to our Audio-to-Braille system, where spoken audio must first be accurately transcribed and then intelligently processed into structured content suitable for tactile Braille output. Similar to how SALOMON organizes and summarizes legal documents, our system must apply NLP-based filtering, segmentation, and context-aware text formatting to ensure clarity and readability for visually impaired users. The methodology reinforces the importance of efficient text structuring and summarization in building real-time, multilingual accessibility tools.

The paper ”Using a Document Parser to Automate Software Testing[23]” emphasizes the importance of extracting structured, meaningful information from unstructured text sources to support automation and increase system accuracy and efficiency. The SIFT document parser demonstrates how partial parsing, heuristic-based semantic extraction, and canonical text transformation can convert free-form text into structured output suitable for automated processing. This aligns closely with our Audio-to-Braille project, where speech-derived text must be preprocessed, semantically analyzed, and converted into a structured Braille representation for visually impaired users. Similar to how SIFT separates domain-independent and domain-specific modules for flexibility across different document types, our system must support multilingual and context-dependent processing through modular ASR and NLP components. The success of SIFT in reducing manual effort and increasing accessibility reinforces the feasibility and social importance of automated accessibility systems like Audio-to-Braille.

The paper ”An Efficient Chart-based Algorithm for Partial-Parsing of Unrestricted Texts[24]” presents a robust parsing strategy optimized for real-world, noisy, and unstructured text environments—achieving high accuracy by focusing only on essential linguistic segments and using semantic-driven rule processing instead of full syntactic parsing. This selective “partial parsing” approach enables efficient extraction of meaningful structured information even when vocabulary is unknown or text is irregular. This is directly relevant to our Audio-to-Braille project, where speech input may contain fillers, accents, pauses, or disfluencies, and where efficient NLP processing is essential for generating reliable Braille output in real time. The paper demonstrates that scalable systems must prioritize semantic understanding, reduce computational complexity, and handle incomplete or noisy language context—principles that align with the need for a multilingual audio processing pipeline capable of producing accessible tactile text for visually-impaired readers.

The paper ”An Infrastructure for the Evaluation and Comparison of Information Retrieval Systems[25]” emphasizes the importance of systematic benchmarking, performance measurement, and scalability analysis in building reliable and efficient real-world systems. The authors demonstrate that indexing accuracy, response time, precision, recall, and resource efficiency are critical to determining a system’s usability and adaptability in diverse environments. This insight is highly relevant to our Audio-to-Braille system, which requires optimized ASR and NLP components to ensure fast and accurate speech-to-Braille conversion, especially under multilingual and noisy conditions. Just as retrieval systems must manage large unstructured document collections, our project must efficiently process continuous speech input and transform it into high-quality tactile output. The focus on performance evaluation frameworks reinforces the necessity for measurable benchmarks—such as WER, CER, and Braille readability—to ensure the accessibility, scalability, and real-time capability of assistive technologies for the visually impaired.

The paper "Doing Things with Factors[26]" demonstrates how intelligent systems can support users in complex reasoning tasks by structuring unorganized information and enabling efficient decision-making. The CATO model provides an instructional framework that assists learners in analyzing unstructured case data, identifying strengths and weaknesses, and forming meaningful conclusions by incrementally refining information using factor-based reasoning. This approach is directly relevant to our Audio-to-Braille project, where raw speech must be converted into structurally meaningful text before being translated into Braille. Just as CATO transforms scattered legal narratives into structured, comparable factor sets, our system must apply preprocessing, segmentation, and context interpretation to produce accurate tactile output suitable for visually impaired users. The results reported in the paper—showing enhanced performance when guided by intelligent support systems—reinforce the value of building intelligent, automated, and user-friendly accessibility technologies.

The paper "TileBars: Visualization of Term Distribution Information in Full Text Information Access[27]" emphasizes the importance of understanding not just term frequency but also contextual distribution and structural placement of language within long documents. The system presents a visualization approach that helps users interpret relevance patterns rapidly by analyzing where and how intensively terms appear across document segments. This insight is highly relevant to our Audio-to-Braille project, as accurate Braille conversion requires more than literal transcription; it requires understanding semantic emphasis, contextual segmentation, and structural meaning within speech-derived text. Similar to how TileBars maps semantic distribution visually, our system must identify key speech segments such as headings, topic transitions, and emphasized words to improve readability and tactile comprehension for visually impaired users. The paper reinforces the importance of processing linguistic structure and distribution—critical for generating meaningful multilingual Braille output.

The paper "Design and Development of a Text-to-Speech Synthesizer for Afan Oromo[28]" provides a detailed exploration of building a speech synthesizer for a low-resource African language using a unit-selection approach. The authors highlight challenges such as dialectal variations, limited corpus availability, and linguistic complexity, which mirror the problems faced by many Indian regional languages. Their prototype achieved encouraging naturalness (MOS 4.44/5) but still struggled with intelligibility, emphasizing the importance of high-quality datasets and language-specific phonological modeling. This work is strongly relevant to our Audio-to-Braille system because it reinforces the need for accurate text preprocessing and multilingual ASR support—especially for languages with limited technological resources. Insights from this paper justify our system's focus on robust speech recognition, language-aware text normalization, and accessible output generation for visually impaired users.

C. Multilingual Audio Input Methods

The research paper "Audio and Text Sentiment Analysis of Radio Broadcasts"[29] demonstrates how combining audio emotion detection with text-based NLP significantly enhances comprehension and accessibility of spoken content, particularly when dealing with real-time audio streams. The study highlights that relying solely on transcribed text often loses emotional and semantic depth, while multimodal processing extracts richer contextual meaning from speech.This finding strongly supports our Audio-to-Braille conversion project, which similarly depends on accurate audio interpretation before generating tactile Braille output. For visually impaired users, capturing emotional cues, clarity variations, and speech dynamics is essential for producing meaningful Braille content rather than literal transcription. The identified gaps—such as handling multilingual audio streams and improving processing efficiency—align directly with our goal of designing an integrated multilingual audio-to-Braille pipeline capable of real-time, emotion-aware tactile representation for inclusive accessibility.

The paper "Voice4Blind: The Talking Braille Keyboard to Assist the Visual Impaired Users in Text Messaging[30]" presents an innovative mobile-based solution that enables visually impaired users to write, read, and manage SMS messages using a Braille keyboard enhanced with text-to-speech and haptic feedback. The authors emphasize the limitations of modern smartphones, which often lack tactile interfaces and thus hinder independent communication for blind users. Their Voice4Blind system leverages audio-guided interaction and a Braille keypad to simplify message composition—demonstrating that accessible communication tools can significantly improve autonomy for visually impaired individuals. However, the system remains limited to text messaging and supports only English and Malay, revealing a broader gap in multimodal, multilingual assistive technologies. This gap directly strengthens the motivation for our Audio-to-Braille system, which extends accessibility by converting speech to multilingual Braille with integrated voice assistance for complete hands-free interaction.

The paper "Conversion of Speech to Braille: Interaction Device for Visual and Hearing Impaired[31]" presents a hardware-centric speech-to-Braille system that uses the HM2007 speech-recognition IC and a PIC microcontroller to convert spoken Tamil digits into embossed Braille output. The system processes speech through endpoint detection, HMMbased recognition, and real-time solenoid actuation to drive a Braille display, enabling communication for individuals with both visual and hearing impairments. While the prototype successfully demonstrates low-cost hardware-based Braille generation, its functionality is restricted to digit recognition and Tamil Thirukural lookup, highlighting limited vocabulary, absence of multilingual capabilities, and lack of natural-speech handling. These constraints reveal a significant research gap addressed by our proposed multilingual audio-to-Braille pipeline, which extends beyond digit recognition to full-sentence speech processing, translation, Unicode Braille generation, and web-based accessibility—a substantial advancement over existing hardware-bound approaches.

The paper "Braille Script to Voice Conversion[32]" presents a low-cost, Arduino-based system that converts Braille input into synthesized speech, providing an alternative communication method for individuals who are visually impaired, speech-impaired, or both. The authors design a Braille-like keyboard using six push-button inputs, which map to text through microcontroller logic and are spoken aloud using a WTV020-SD voice module. This makes the device suitable for basic message construction and improves user interaction through tactile familiarity. However, the system is limited to manual Braille key input, supports only predefined voice outputs, and does not process natural speech or multilingual content. These constraints highlight a significant technological gap that our multilingual Audio-to-Braille system addresses by enabling real-time speech processing, Unicode Braille generation, and voice-assistant–guided interaction, providing a far more scalable accessibility solution.

The paper "PodGPT: An Audio-Augmented Large Language Model for Research and Education[33]" presents a novel framework that enhances large language models by integrating over 3700 hours of STEMM podcast audio, generating more than 42 million text tokens for pre-training. The authors demonstrate that audio-derived conversational knowledge significantly improves domain-specific reasoning, multilingual generalization, and evidence-grounded response generation, particularly when paired with a robust RAG pipeline. Their results show measurable performance gains across clinical, scientific, and multilingual benchmarks, emphasizing the value of nontraditional modalities such as expert podcasts for enriching LLM understanding. While the work focuses on scientific conversational data rather than accessibility technologies, it establishes a strong precedent for audio-driven AI systems. This directly supports our Audio-to-Braille project, reinforcing the feasibility of speech-based knowledge extraction, multilingual processing, and assistive interaction for visually impaired users.

The paper "Generative AI-Powered Multilingual ASR for Seamless Language-Mixing Transcriptions[34]" presents an innovative approach to handling bilingual and code-switched speech using a hybrid LSTM–GPT architecture, achieving 84.37% accuracy for short and 73.98% for long Tamil–English utterances. The work highlights major challenges in multilingual ASR—such as intra-sentential switching, limited training data, and noise—but demonstrates that generative transformer models significantly improve contextual understanding and transcription quality. This paper directly aligns with our Audio-to-Braille project, as accurate multilingual ASR is the first foundational step in speech-to-Braille conversion. Their demonstration of GPT-enhanced ASR supports the feasibility of our pipeline, which similarly depends on robust multilingual speech recognition before translation, normalization, and Braille encoding. The research thus validates our methodology and emphasizes the growing relevance of AI-driven ASR systems for accessibility solutions.

The paper "Dynamic Knowledge Condensation with Audio-Selective Transformer for Audio Deepfake Detection[35]" by Wani and Amerini (2025) introduces DK-CAST, a tri-stream knowledge-distillation framework designed to robustly detect audio deepfakes even under codec compression and degraded real-world conditions. The method integrates high-capacity XLS-R teacher models with a lightweight student network capable of operating on low-quality and preprocessed audio, enhanced by phoneme-gated attention and codec-aware loss modulation for improved generalization. Although the work focuses on detection of synthetic speech rather than assistive audio processing, it highlights the vulnerabilities of voice-based systems and the importance of reliable speech-processing pipelines. These insights strongly relate to our Audio-to-Braille project, emphasizing the need for robust speech handling in noisy, compressed, or device-constrained environments to ensure accurate Braille output generation and trustworthy voice-assistant interactions.

The paper "Efficient Training for Multilingual Visual Speech Recognition: Pre-training with Discretized Visual Speech Representation[36]" proposes a highly efficient multilingual lip-reading framework that leverages discretized visual speech units to drastically reduce training cost while improving recognition accuracy across languages. The authors demonstrate that replacing raw video features with quantized viseme-level representations reduces input size by 99.984%, enabling faster multimodal pre-training and improved generalization in low-resource languages. Although the work focuses on visual speech rather than audio-based recognition, its findings are strongly relevant to our multilingual Audio-to-Braille system. The demonstrated success of discrete units, curriculum learning, and multilingual training strategies highlights promising pathways for improving speech processing efficiency, scalability, and cross-lingual robustness in our Braille pipeline.

The paper "AI-Based Speaking Assistant: Supporting Non-Native Speakers' Speaking in Real-Time Multilingual Communication[37]" presents AISA, an AI-driven assistant designed to help non-native speakers (NNSs) generate context-aware speaking references during real-time multilingual collaboration. The study identifies four distinct input patterns—translation, rationalizing decisions, stating viewpoints, and keyword-based prompting—revealing the diverse cognitive strategies NNSs use while interacting with AI support tools. Although AISA did not significantly improve measurable speaking competence, qualitative insights show that it enhanced the logical flow, coherence, and topical relevance of NNSs' speech. Importantly, the system introduced additional multitasking demands, contributing to cognitive load and anxiety. The findings highlight the need for assistive communication tools that balance linguistic support with usability. For our Audio-to-Braille project, this work reinforces the importance of designing real-time assistive systems that reduce user workload, maintain contextual relevance, and offer concise, intuitive feedback—key considerations for speech-to-Braille conversion interfaces targeting visually impaired users.

The paper "Indic-ST: A Large-Scale Multilingual Corpus for Low-Resource Speech-to-Text Translation[38]" presents one of the most comprehensive speech-to-text (ST) datasets for 15 Indic languages, addressing a major gap in multilingual ASR and translation research. The dataset's scale—6,800 hours of English speech and over 2.2 million aligned text tokens—provides an unprecedented resource for ST, MT, and ASR applications. For our Audio-to-Braille system, this work is highly relevant because it demonstrates the feasibility and importance of accurate multilingual speech recognition, especially for low-resource Indian languages like Marathi, Gujarati, Assamese, Manipuri, and Urdu. Unlike prior datasets, Indic-ST supports domain diversity, robust alignment, and multilingual modeling, which directly strengthens our project's speech-to-text foundation. The paper's insights on dataset quality, error propagation, and multilingual model behavior reinforce the need for reliable ASR before Braille conversion, validating our pipeline design and highlighting future opportunities to integrate low-resource speech models for inclusive accessibility technologies.

The paper "Deep Learning based Multilingual Speech Synthesis using Multi Feature Fusion Methods[39]" presents a comprehensive deep-learning-driven framework for generating natural and intelligible multilingual speech synthesis by integrating CNN-based feature extraction, cross-lingual embeddings, and vector mapping techniques. The authors thoroughly address challenges in multilingual TTS, including limited linguistic resources, poor prosody modeling, and cross-language acoustic variations, proposing a hybrid CNN–BERT architecture that significantly improves accuracy, precision, recall, and F-measure across languages such as English, Japanese, and Mandarin. Their experiments demonstrate that dynamic embeddings and model averaging substantially enhance linguistic generalization and speech naturalness. The insights into prosody modeling, multilingual feature sharing, and deep learning–based acoustic mapping directly align with our Audio-to-Braille system, reinforcing how robust multilingual speech processing models can improve accurate transcription, language adaptability, and accessibility for visually impaired users.

The paper "Soft Prompt Decoding for Multilingual Dense Retrieva[40]" introduces KD-SPD, a multilingual retrieval framework that aligns documents from multiple languages into a unified embedding space using soft-prompt–based knowledge distillation. The work highlights major challenges in multilingual systems such as language imbalance, vocabulary mismatch, and inconsistent ranking across languages—problems that directly relate to building reliable multilingual assistive tools. Although the study focuses on text retrieval rather than accessibility, its insights into cross-lingual embedding alignment and zero-shot generalization are highly relevant to our Audio-to-Braille pipeline. Our system similarly requires consistent multilingual processing—converting spoken languages into uniform Braille output. The KD-SPD approach demonstrates how multilingual inconsistencies can be reduced effectively, supporting the need for robust, language-agnostic models in assistive technologies like ours.

IV. Figures and Tables

A. Language Detection and Accuracy Evaluation For Multilingual ASR Systems

Language detection plays a crucial role in multilingual Audio-to-Braille systems by identifying the spoken language before ASR processing. Accurate LID ensures that the correct language-specific speech and Braille conversion models are applied, significantly reducing transcription errors and improving overall system reliability. The following table summarizes performance metrics reported in recent multilingual ASR research.

TABLE I
LANGUAGE DETECTION ACCURACY EVALUATION FOR MULTILINGUAL ASR SYSTEMS

System / Model	Languages	Method	Accuracy / Metrics
Whisper ASR (OpenAI) [1]	98+	Transformer Encoder–Decoder	92–96% LID; WER $\leq$ 12%
Google Speech-to-Text [11]	125+	Neural Transducer + LAS	94–98% LID; WER 5–12%
Gen. AI ASR (Tamil–Eng) [32]	Tamil, Eng	Hybrid LSTM + GPT	84.37% (short), 73.98% (long)
Indic-ST Corpus [38]	15 Indic langs	ST Enc–Dec Models	BLEU: 19–27; CER: 6–14%
Deepfake Detection (Urdu) [14]	Urdu	CNN + RNN	98.89% accuracy
Audio–Text Sentiment [29]	Eng, Hindi	MFCC + RNN + NLP	92–94% lang. ID
MFF Speech Synthesis [39]	Eng, Mandarin, Jap.	CNN–BERT Hybrid	95–97% ID fidelity
Visual Lip-Reading [36]	11 langs	Viseme + Transformer	83–90% phoneme ID

B. Comparison with Related Research Papers

The comparison table highlights the major contributions of prior research in Braille recognition, tactile display technology, OCR-based assistance, and multilingual speech processing. It shows that most existing systems address isolated tasks such as Braille recognition, text transliteration, hardware-based tactile output, or speech processing, but none provide an integrated multilingual audio-to-Braille pipeline. This gap clearly motivates the need for our proposed end-to-end system that combines ASR, NLP, Braille generation, and voice-assisted interaction for visually impaired users.

TABLE II
COMPARISON OF RELATED RESEARCH PAPERS AND IDENTIFIED GAPS

Paper	Key Contribution	Identified Research Gap
Deep Learning Braille Recognition [1]	CNN-based Braille image recognition with 95%+ accuracy	Recognizes Braille only; no Braille generation or speech input
Refreshable Braille Display Prototype [2]	Low-cost tactile display using actuators	No speech interface; no multilingual processing
Electromagnetic Braille Display [6]	High-speed, low-power Braille cells	Hardware-only; lacks ASR/NLP pipeline
FingerSight Haptic System [4]	Haptic spatial feedback via fingertip camera	Not related to language/Braille generation
Digital Urdu Braille Learning System [5]	Interactive learning + tactile modules	Focuses on teaching; no automated audio-to-Braille
Optical Braille Recognition [15]	Braille-to-text OCR pipeline	Reverse direction; no speech-to-Braille
Gujarati Text-to-Braille [16]	Rule-based text–Braille transliteration	Only Gujarati; no speech or multilingual support
Text Detection → Braille Output Device [17]	OCR + actuator-based Braille cell mechanism	Image-only input; no audio, no multilingual ASR
Vision-to-Voice Assistive Framework [18]	OCR-to-speech multimodal system	Opposite direction; no tactile output
Multilingual ASR/Code-Switching [32,38]	Accurate ASR for low-resource langs	No Braille generation pipeline
Multilingual Speech Synthesis [39]	Cross-lingual feature fusion for TTS	Speech output only; no tactile/Braille mapping

C. Comparative Analysis of Existing Braille Conversion Approaches

This subsection compares major categories of Braille conversion systems reported in prior research, including OCR-based, camera-based, image-based, and audio-based approaches. The analysis highlights the limitations of visual-input–dependent methods and demonstrates how the proposed speech-driven Braille pipeline overcomes these challenges.

TABLE III
COMPARISON OF EXISTING BRAILLE CONVERSION APPROACHES

Approach	Working Principle	Limitations / Challenges
OCR-based Braille Conversion	Extracts printed/embossed Braille or text using OCR techniques to generate digital text or Braille.	Sensitive to lighting, image skew, and noise; low accuracy for poor-quality documents.
Camera-based Braille Devices	Uses mobile or embedded cameras to detect characters and actuate Braille cells or generate speech.	Slow processing; variable accuracy; hardware-dependent; limited multilingual support.
Image-based Braille Recognition	Employs ML/CNN models on scanned Braille images to classify dots and map Braille cells.	Requires high-resolution images; reduced performance on worn/damaged Braille; recto–verso interference.
Proposed Audio-based Approach	Uses ASR → NLP → Unicode Braille conversion to generate real-time multilingual Braille output.	Depends on ASR noise robustness; requires multilingual datasets for optimal performance.

D. Review for Braille Printers

Braille printers, or embossers, represent a vital magnificence of assistive devices designed to convert virtual Braille files into tactile tough-copy output for blind users. conventional embossers depend upon mechanical pins to provide raised dot patterns on thick embossing paper, enabling users to examine textual content thru contact. modern-day gadgets, such as the ones advanced by way of Index Braille, enabling technology, and ViewPlus, guide excessive-pace duplex embossing, tactile pix generation, and compatibility with translation gear like Liblouis for accurate formatting. notwithstanding their technical sophistication, most excessive-overall performance models remain costly, require specialized maintenance, and are therefore inaccessible to many instructional establishments and individuals in developing regions. Low-fee alternatives exist however frequently compromise printing pace, sturdiness, or picture aid. This assessment shows a sizeable want for low cost and person-pleasant Braille output technologies, reinforcing the relevance of solutions that combine speech-to-Braille conversion with value-powerful printing workflows.

TABLE IV
BRAILLE PRINTERS AVAILABLE IN THE INDIAN MARKET

Model	Speed	Paper	Approx. Price (INR)
Basic-D V5 (Index)	100–140 cps	Fanfold	2.5–3.3 lakh
Everest-D V5 (Index)	up to 140 cps	Cut-sheet	On quotation
FanFold-D V5 (Index)	High-speed	Fanfold	Institutional pricing
VP Columbia (ViewPlus)	100–120 cps	Cut-sheet	3–4 lakh
Cub / EmBraille (ViewPlus)	Mid-range	Cut-sheet	1.5–3 lakh
Juliet 120 (Enabling Tech)	120 cps	Fanfold	High-end
Romeo 60 (Enabling Tech)	60 cps	Fanfold	Mid-range
Low-cost Indian models	20–40 cps	Cut-sheet / fanfold	50k–1.5 lakh

V. CONCLUSION

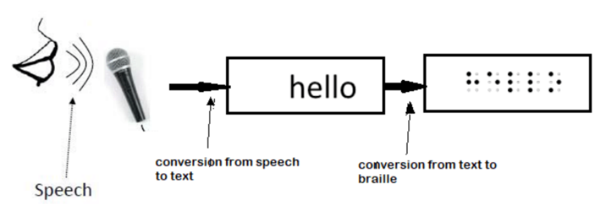

Fig. 3. Overview of Audio-to-Braille Conversion Concept

This review paper has examined the current state of research and technological advancements in multilingual audio-to-Braille conversion, automatic speech recognition, Braille encoding standards, and tactile output devices. Existing solutions in the assistive technology ecosystem largely focus on isolated components such as Braille character recognition, OCR-based Braille digitization, refreshable Braille displays, and multilingual speech processing. This review has explored the current state of research and technology in multilingual audio-to-Braille conversion, speech recognition, Braille standards, and tactile devices. While there have been great strides in each area, the reality is that most solutions address just one part of the bigger accessibility puzzle. As we look at the landscape, one thing is clear: there's still a major gap—no single system brings together real-time, multilingual speech input and tactile Braille output with integrated voice guidance.

The good news is that technology is catching up. Advances in transformer-based ASR, growing Unicode Braille support, and more affordable embosser hardware mean we're closer than ever to building tools that truly make a difference. Our approach is designed to bridge this gap by letting users turn natural speech into Unicode Braille, using smart NLP preprocessing and cost-effective printing. By blending speech, text, touch, and interactive feedback, we hope to make digital access more seamless and empower visually impaired users to be more independent.

Looking ahead, we're excited to expand our multilingual datasets—especially for Indian languages—improve real-time speed, build even more affordable hardware, and launch broad usability testing to ensure our system makes a genuine impact in people's everyday lives.

Acknowledgment

We are deeply grateful to the Department of Software Computer Engineering at MIT Academy of Engineering, Pune, for their unwavering support throughout this project. Their encouragement, technical expertise, and strong research infrastructure made this work possible. We also want to thank our faculty mentors and domain experts for sharing thoughtful insights into accessibility technologies and Braille literacy, which guided our approach and strengthened our understanding. Most importantly, we owe special thanks to the visually impaired community members and assistive technology practitioners whose honest feedback and lived experiences inspired our motivation and kept us focused on real-world impact. Their perspectives have been central in shaping both the goals and the direction of our proposed multilingual Audio-to-Braille system.

References

[1] T. Kausar, S. Manzoor, A. Kausar, Y. Lu, M. Wasif, and M. A. Ashraf,"Deep learning strategy for Braille character recognition," IEEE Access, vol. 9, pp. 169357–169371, Dec. 2021, doi: 10.1109/ACCESS.2021.3138240.

[2] R. D. Sutariya, H. S. Singh, S. R. Babariya, S. A. Kadiyar, and D. H. Modi, "Refreshable Braille Display for the Visually Impaired: E-book reader prototype," in *Proc. IEEE*, 2017.

[3] D. Chen et al., "Development and Evaluation of Refreshable Braille Display and Active Touch-Reading System for Digital Reading of the Visually Impaired," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 934–945, 2024.

[4] S. Horvath, J. Galeotti, B. Wu, R. Klatzky, M. Siegel, and G. Stetten, "FingerSight: Fingertip Haptic Sensing of the Visual Environment," IEEE Journal of Translational Engineering in Health and Medicine, vol. 2, pp. 1–10, 2014. doi: 10.1109/JTEHM.2014.2309343.

[5] F. Jabeen and M. Rashid, "Empowering Individuals With Visual Impairment: A Digital Braille Solution for Learning the Urdu Language," IEEE Access, vol. 13, pp. 145243–145255, 2025, doi: 10.1109/ACCESS.2025.3598856.

[6] D. Chen et al., "Design and Evaluation of an Electromagnetic Bounce-Type Refreshable Braille Display," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 33, pp. 705–716, 2025. doi: 10.1109/TNSRE.2025.3535564.

[7] S. Uboldi, A. Bortolotti, G. Candeloro, A. Marasco, F. Sardella, M. Tartari, and P. L. Sacco, "From Touch to Mental Imagery: The Embodied Aesthetic Experience of Late-Blind People Engaged in the Tactile Exploration of Enrico Castellani's Pseudo-Braille Surface," *Culture, Medicine, and Psychiatry*, vol. 49, pp. 725–764, 2025. doi: 10.1007/s11013-025-09904-9.

[8] H. Cheng, J. Yin, S. Li, and K. H. Li, "Contactless Braille sensing based on GaN optical devices integrated with epoxy lenses," Microsystems & Nanoengineering, vol. 11, no. 62, pp. 1–9, 2025. doi: 10.1038/s41378-025-00904-8.

[9] B.-S. Park, S.-M. Im, H. Lee, Y. T. Lee, C. Nam, S. Hong, and M.-g. Kim, "Visual and Tactile Perception Techniques for Braille Recognition," Micro and Nano Systems Letters, vol. 11, no. 23, pp. 1–8, 2023, doi: 10.1186/s40486-023-00191-w.

[10] S. Shokat, R. Riaz, S. S. Rizvi, A. M. Abbasi and S. J. Kwon, "Deep learning scheme for character prediction with position-free touchscreen-based Braille input method," *Human-Centric Computing and Information Sciences*, vol. 10, no. 41, pp. 1–24, 2020. doi: 10.1186/s13673-020-00246-6

[11] N. A. Asfaw, B. H. Belay, and K. M. Alemu, "A deep learning approach for line-level Amharic Braille image recognition," *Scientific Reports*, vol. 14, no. 24172, pp. 1–14, 2024, doi: 10.1038/s41598-024-73895-7.

[12] F. Lang, V. Pues, A. Schmidt, and T.-K. Machulla, "BrailleBuddy: A Tangible User Interface to Support Children with Visual Impairment in Learning Braille," in *Proc. CHI Conf. Human Factors in Computing Systems*, 2023, pp. 1–14. doi: 10.1145/3544548.3580844.

[13] N. Sepahvand, M. Hamidi, A. Erol, and A. Hurst, "The Effect of Orientation on the Readability and Comfort of 3D-Printed Braille," in Proc. CHI Conf. Human Factors in Computing Systems, 2024, pp. 1–16.

[14] O. Ahmad, M. S. Khan, S. Jan, and I. Khan, "Deepfake Audio Detection for Urdu Language Using Deep Neural Networks," IEEE Access, vol. 13, pp. 97765–97778, May 2025.

[15] M. Hanumanthappa and V. V. Murthy, "Optical Braille Recognition and its correspondence in the Conversion of Braille Script to Text – A Literature Review," in *2016 International Conference on Computational Systems and Information Systems for Sustainable Solutions*, 2016, pp. 297–301.

[16] S. J. Patel and R. A. Sheth, "Transliteration of Digital Gujarati Text into Printable Braille," 2015 Fifth International Conference on Communication Systems and Network Technologies, pp. 572–578, 2015.

[17] K. Kulkarni, F. Belwadi, A. Mahajan, S. Killedar, Y. Zambre, and A. Marathe, "Text Detection and Communicator Using Braille for Assistance to Visually Impaired," in Proceedings of IEEE PuneCon, 2019, pp. 1–6.

[18] S. Bhat, P. Bhat, and S. V. Kolekar, "From Vision to Voice: A Multi-Modal Assistive Framework for the Physically Impaired," IEEE Access, vol. 13, pp. 128106–128120, Jul. 2025.

[19] Fagin RKimelfeld BLi YRaghavan SVaithyanathan SLenzerini MSchwentick T(2011)Rewrite rules for search database systemsProceedings of the thirtieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems10.1145/1989284.1989322(271-282)Online publication date: 13-Jun-2011

[20] HARABAGIU, SANDA M.; MOLDOVAN, DAN I. . (1997). TextNet – A text-based intelligent system. Natural Language Engineering, 3(2), 171–190. doi:10.1017/s1351324997001757

[21] Dowding, John; Gawron, Jean Mark; Appelt, Doug; Bear, John; Cherny, Lynn; Moore, Robert; Moran, Douglas . (1993). [Association for Computational Linguistics the 31st annual meeting - Columbus, Ohio (1993.06.22-1993.06.26)] Proceedings of the 31st annual meeting on Association for Computational Linguistics - - Gemini. , (), 54–61. doi:10.3115/981574.981582.

[22] Moens, Marie-Francine; Uyttendaele, Caroline; Dumortier, Jos . (1997). [ACM Press the sixth international conference - Melbourne, Australia (1997.06.30-1997.07.03)] Proceedings of the sixth international conference on Artificial intelligence and law - ICAIL '97 - Abstracting of legal cases. , (), 114–122. doi:10.1145/261618.261643

[23] Lutsky, Patricia. (1994). [ACM Press the 1994 ACM symposium - Phoenix, Arizona, United States (1994.03.06-1994.03.08)] Proceedings of the 1994 ACM symposium on Applied computing - SAC '94 - Using a document parser to automate software testing. , (), 59–63. doi:10.1145/326619.326661

[24] McDonald, David D.. (1992). [Association for Computational Linguistics the third conference - Trento, Italy (1992.03.31-1992.04.03)] Proceedings of the third conference on Applied natural language processing

- - An efficient chart-based algorithm for partial-parsing of unrestricted texts. , (), 193–. doi:10.3115/974499.974534

[25] Broadbent, Robert E.; Saunders, Gary S.; Ekstrom, Joseph J. . (2006). [ACM Press the 7th conference - Minneapolis, Minnesota, USA (2006.10.19-2006.10.21)] Proceedings of the 7th conference on Information technology education - SIGITE '06 - An infrastructure for the evaluation and comparison of information retrieval systems. , (), 123–. doi:10.1145/1168812.1168842

[26] Aleven, Vincent; Ashley, Kevin D. . (1995). [ACM Press the fifth international conference - College Park, Maryland, United States (1995.05.21-1995.05.24)] Proceedings of the fifth international conference on Artificial intelligence and law - ICAIL '95 - Doing things with factors. , (), 31–41. doi:10.1145/222092.222106.

[27] Hearst, Marti A.. (1995). [ACM Press the SIGCHI conference - Denver, Colorado, United States (1995.05.07-1995.05.11)] Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95 - TileBars. , (), 59–66. doi:10.1145/223904.223912

[28] T. D. Chala, A. C. Guta, and M. H. Asebel, "Design and Development of a Text-to-Speech Synthesizer for Afan Oromo," SN Computer Science, vol. 3, no. 420, pp. 1–13, 2022. doi: 10.1007/s42979-022-01306-7.

[29] N. Dhariwal, S. C. Akunuri, Shivama, and K. S. Banu, "Audio and Text Sentiment Analysis of Radio Broadcasts," IEEE Access, vol. 11, pp. 126900–126916, Nov. 2023.

[30] M. S. S. B. Ali, N. S. N. N. Ahmad, N. F. M. Razali, and N. H. M. Salleh, "Voice4Blind: The Talking Braille Keyboard to Assist the Visually Impaired Users in Text Messaging," in *Proc. International Conference on Computer and Information Sciences (ICCOINS)*, pp. 1–6, 2014.

[31] V. Ajantha Devi, "Conversion of Speech to Braille: Interaction Device for Visual and Hearing Impaired," in Proc. 2017 4th International Conference on Signal Processing, Communications and Networking (ICSCN), Chennai, India, Mar. 2017, pp. 1–6.

[32] K. Krishna Kishore, G. Prudhvi, and M. Naveen, "Braille Script to Voice Conversion," in Proc. IEEE International Conference on Computing Methodologies and Communication (ICCMC), 2017, pp. 1080–1082.

[33] S. Jia et al., "PodGPT: an audio-augmented large language model for research and education," npj Biomedical Innovations, vol. 2, art. no. 26, 2025. doi: 10.1038/s44385-025-00022-0.

[34] P. Dash, S. Babu, L. Singaravel, and D. Balasubramanian, "Generative AI-powered multilingual ASR for seamless language-mixing transcriptions," *Journal of Electrical Systems and Information Technology*, vol. 12, no. 42, pp. 1–16, 2025. doi: 10.1186/s43067-025-00204-1.

[35] T. M. Wani and I. Amerini, "Dynamic knowledge condensation with audio-selective transformer for audio deepfake detection," *Discover Computing*, vol. 28, no. 231, pp. 1–28, 2025. doi: 10.1007/s10791-025-09746-4.

[36] M. Kim, J. H. Yeo, S. J. Park, H. Rha, and Y. M. Ro,"Efficient Training for Multilingual Visual Speech Recognition: Pre-training with Discretized Visual Speech Representation,"in *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24)*,Melbourne, Australia, 2024, pp. 1–10.doi: 10.1145/3664647.3681221.

[37] P. Qin, Z. Zhu, N. Yamashita, Y. Yang, K. Suga, and Y.-C. Lee, "AI-Based Speaking Assistant: Supporting Non-Native Speakers' Speaking in Real-Time Multilingual Communication,"*Proceedings of the ACM on Human-Computer Interaction*, vol. 9, no. 7, Article CSCW274, pp. 1–28, Nov. 2025. doi: 10.1145/3757455.

[38] N. Sethiya, S. Nair, P. Walia, and C. Maurya, "Indic-ST: A Large-Scale Multilingual Corpus for Low-Resource Speech-to-Text Translation," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 24, no. 6, pp. 1–25, Jun. 2025. doi: 10.1145/3736720.

[39] P. Nuthakki, M. Katamaneni, C. S. J. N., K. Gubbala, B. Domathoti, V. R. Maddumala, and K. R. Jetti, "Deep Learning based Multilingual Speech Synthesis using Multi Feature Fusion Methods," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 1, pp. 1–16, 2023. doi: 10.1145/3618110.

[40] Z. Huang, H. Zeng, H. Zamani, and J. Allan, "Soft Prompt Decoding for Multilingual Dense Retrieval," in *Proc. 46th Int. ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '23)*, Taipei, Taiwan, 2023, pp. 1208–1216. doi: 10.1145/3539618.3591769.