IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15527

A Review of Voice-Driven Accessible Vocational Training Platforms for Persons with Visual Impairment

Bharath Reddy[1], Chinta Vishnu Vardhan[2], H M Sharanabasava[3], K Ravi Kumar[4], Dr. Muhibur Rahman[5]

6th Sem B.E.(CS&E), Ballari Institute of Technology and Management (BITM), Ballari, Karnataka-583104, India1-4

Associate Professor, Department of Computer Science and Engineering.

Ballari Institute of Technology and Management (BITM), Ballari, Karnataka 583104, India5

Abstract: The increasing demand for inclusive education and skill development has brought significant attention to assistive technologies designed for persons with visual impairments. Voice-driven vocational training platforms represent a transformative approach to delivering accessible learning experiences without reliance on visual interfaces. This review examines existing literature, system architectures, and technological frameworks used in voice-based e-learning and vocational training tools tailored for visually impaired users. The paper presents a four-tier taxonomy of platform sophistication, analyses key methods including Text-to-Speech (TTS), Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and screen reader integration, and identifies critical research gaps. Comparative analysis of reviewed systems highlights strengths and limitations. The review concludes by recommending a unified, adaptive voice-driven platform with offline capabilities and multi-language support to bridge the vocational accessibility gap.

Keywords: Visual Impairment, Voice-Driven Learning, Vocational Training, Assistive Technology, Text-to-Speech, Speech Recognition, Accessible Education, NLP

I. INTRODUCTION

Vocational training is a cornerstone of economic independence and social inclusion for individuals with disabilities. For persons with visual impairments, conventional training programs built around visual content, printed materials, or screen-based interfaces present substantial barriers. These individuals often depend on sighted assistance or fail to access skill-building opportunities altogether, widening the employment gap between sighted and visually impaired populations.
With the rapid proliferation of voice-enabled devices, smart speakers, and advanced speech synthesis technologies, there now exists a viable pathway to deliver vocational education entirely through audio and voice interaction. Voice-driven platforms eliminate the dependency on visual output, enabling learners to navigate course content, receive instructions, complete assessments, and obtain certifications through speech alone.

This paper reviews the current state of voice-driven accessible vocational training platforms, examines methodologies employed across existing solutions, proposes a classification taxonomy, and identifies open research challenges. The goal is to provide a structured foundation for researchers and developers working towards inclusive vocational education systems.

II. THEORETICAL BACKGROUND

Voice-driven accessible vocational training platforms integrate multiple technologies to create barrier-free learning environments. This section describes the foundational concepts and models that underpin such systems.

A. System Architecture

A voice-driven vocational training system can be modelled as:

$$O = f(I, P)$$

Where:

- Input (I): Learner voice commands, profile data, course metadata, connectivity status
- Processing (P): ASR, NLP intent parsing, TTS synthesis, adaptive content delivery
- Output (O): Audio instructions, assessments, feedback, and progress reports

DOI: 10.17148/IJARCCE.2026.15527

B. Text-to-Speech (TTS) and Speech Synthesis

TTS converts written training content into spoken audio. Modern neural TTS systems (e.g., WaveNet, Tacotron) produce natural-sounding speech with prosody and intonation that improve comprehension. TTS quality is critical in vocational contexts where terminology accuracy is essential.

C. Automatic Speech Recognition (ASR)

ASR systems transcribe spoken learner inputs into text for processing. Challenges include:

- Accent and dialect variability
- Background noise interference
- Domain-specific vocabulary in vocational courses

State-of-the-art ASR models such as Whisper and wav2vec 2.0 have substantially reduced error rates across diverse acoustic conditions.

D. Natural Language Processing (NLP)

NLP enables the platform to interpret learner intent, parse questions, extract responses, and manage conversational dialogue. Key NLP tasks in vocational training platforms include:

- Intent classification: Determining what the learner wishes to do
- Named entity recognition: Identifying skill names, tools, and procedures
- Dialogue management: Maintaining context across multi-turn interactions

E. Screen Reader Integration

Screen readers such as NVDA, JAWS, and VoiceOver allow visually impaired users to interact with software interfaces through audio. Vocational platforms designed for WCAG 2.1 compliance and ARIA landmark integration ensure full screen reader compatibility.

F. Adaptive Learning Algorithms

Adaptive systems personalize content difficulty and pacing based on learner performance. Techniques include:

- Knowledge tracing models (e.g., Bayesian Knowledge Tracing, Deep Knowledge Tracing)
- Reinforcement learning for dynamic content sequencing
- Collaborative filtering for vocational module recommendations

G. Offline and Low-Connectivity Support

Many visually impaired learners in developing regions lack consistent internet access. Offline-capable platforms pre-download audio course modules and synchronize progress when connectivity is restored, ensuring uninterrupted learning.

H. Evaluation Metrics

Platform effectiveness is assessed through:

- Task Completion Rate: Percentage of learners completing modules
- Word Error Rate (WER): ASR accuracy in course interactions
- Learning Gain: Pre- and post-assessment score improvement
- User Satisfaction: Measured via accessible survey interfaces

III. FOUR-TIER TAXONOMY

To systematically classify voice-driven vocational training platforms for visually impaired users, we propose a four-level taxonomy based on technological sophistication, accessibility depth, and adaptability:

Level 1: Static Audio Content Systems

The most basic approach to accessible vocational education.

- Pre-recorded audio lessons delivered via CD, radio, or basic apps
- No interactivity or speech recognition
- Learners passively consume content with no navigation flexibility

Limitations:

- No assessment capability
- Unable to address individual learner queries
- Content quickly becomes outdated

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15527

Level 2: Screen Reader-Compatible E-Learning Platforms
Web-based or app-based courses with screen reader support.

- WCAG-compliant interfaces readable by JAWS, NVDA, or VoiceOver
- TTS output for course materials and assessments
- Keyboard-only navigation without voice command support

Limitations:

- Requires learners to understand assistive technology operation
- Still relies partially on visual layout awareness
- Limited interactivity in vocational skill demonstrations

Level 3: Interactive Voice Response (IVR) Training Systems
Systems incorporating basic voice command inputs alongside audio output.

- ASR enables spoken navigation and response submission
- Structured dialogue trees guide learners through modules
- Eliminates need for keyboard or touch input

Limitations:

- Rigid dialogue structures limit free-form interaction
- Poor handling of complex vocational queries
- Minimal personalization or adaptive learning support

Level 4: Adaptive Voice-Driven Vocational AI Platform (Proposed System)
The most advanced classification, representing the proposed direction of development.

- Full natural language interaction for course navigation, Q&A, and assessments
- Adaptive content sequencing based on learner performance
- Neural TTS with domain-specific vocational vocabulary
- Offline mode with synchronization on reconnection
- Multi-language and multi-dialect support
- Real-time feedback and certification issuance

Taxonomy Summary:
The proposed system integrates ASR, NLP, neural TTS, adaptive learning, and offline functionality into a unified platform, addressing all limitations identified in Levels 1-3. It represents a significant advancement towards truly inclusive vocational training for visually impaired learners.

IV. LITERATURE REVIEW

Research on accessible education and assistive technology for visually impaired persons has expanded considerably over the past decade. Early work focused on screen reader compatibility and WCAG compliance, while more recent studies have moved towards intelligent voice interfaces. Speech synthesis quality, ASR robustness, and NLP-driven interaction management have emerged as central research concerns. Studies across multiple regions highlight the pressing need for offline-capable, multi-language platforms serving low-resource environments. Comparative analyses consistently indicate that adaptive and voice-first platforms outperform static or screen-reader-dependent alternatives in both learner engagement and knowledge retention.

TABLE I: LITERATURE REVIEW SUMMARY

Sl.	Author(s)	Year & Title	Method / Technique	Key Findings	Venue & Index
1	Kumar R. et al.	2021 – Voice-Based Learning for Visually Impaired Students	TTS, Screen Reader, WCAG	Improved content accessibility; high learner satisfaction	IEEE Access (SCI)
2	Patel S. & Joshi M.	2020 – ASR Integration in Accessible E-Learning	Automatic Speech Recognition, NLP	Reduced navigation barriers; 85% task completion rate	ACM CHI Conference

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15527

3	Adewale O. et al.	2022 – Adaptive Voice Platforms for Low-Resource Settings	Offline TTS, Adaptive Learning	Effective in low-connectivity regions; multi-language support identified as critical gap	IJAIED (Scopus)
4	Ramesh K. & Devi P.	2019 – NLP-Driven Vocational Chatbots for the Blind	NLP, Dialogue Management	Natural language Q&A improved learner engagement significantly	Springer LNCS
5	Wang L. et al.	2023 – Deep Learning TTS for Accessibility	WaveNet, Tacotron 2	Neural TTS outperformed concatenative TTS in intelligibility scores	Interspeech (SCI)
6	Nkosi B. & Fourie I.	2020 – Vocational Skills Training via IVR in Africa	IVR, Pre-recorded Audio	IVR effective for basic skills; scalability noted as a limitation	IFIP WCCE
7	Sharma A. et al.	2022 – JAWS and NVDA Compatibility in LMS Platforms	Screen Readers, ARIA	Significant usability gaps found in major LMS platforms	Journal of Disability Studies
8	Chen Y. et al.	2024 – Whisper ASR for Visually Impaired Learners	Whisper, Transfer Learning	Low WER across diverse accents in vocational training tasks	EMNLP Workshop

V. COMPARATIVE ANALYSIS

Existing accessible vocational training platforms vary substantially in technological maturity, accessibility depth, and suitability for visually impaired users. Static audio tools provide basic access but lack interactivity. Screen reader-dependent platforms require high technical literacy. IVR systems improve interaction but rely on rigid dialogue trees that fail in complex vocational learning scenarios.

The proposed adaptive voice-driven platform addresses these limitations through natural language interaction, adaptive content delivery, and offline-first architecture. The following comparative table summarises key distinctions across reviewed platforms:

TABLE II: COMPARATIVE ANALYSIS OF REVIEWED SYSTEMS

Sl. No	System / Approach	Technique Used	Performance	Advantages	Limitations
1	Static Audio Systems	Pre-recorded TTS	Low	Simple and inexpensive	No interactivity or adaptivity
2	Screen Reader LMS	JAWS/NVDA + WCAG	Moderate	Web-accessible content delivery	Requires high technical literacy
3	IVR-Based Training	ASR + Dialogue Trees	Moderate	Voice navigation without keyboard	Rigid; poor complex query handling

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15527

4	Chatbot-Enhanced Platforms	NLP + Rule-Based Chatbot	High	Natural Q&A interactions	Limited vocational domain coverage
5	Adaptive E-Learning (Sighted)	ML + Recommendation	High	Personalised content sequencing	Not designed for voice-only access
6	Offline Audio Apps	Cached TTS	Moderate	Works without internet	No adaptive or interactive capability
7	Proposed Adaptive Voice Platform	ASR + NLP + Neural TTS + Adaptive ML	Very High	Fully voice-driven, adaptive, offline-capable, multi-language	Dependent on high-quality ASR in noisy environments

VI. RESEARCH GAP

Despite notable progress in assistive technology for visual impairment, critical gaps remain in the development of voice-driven vocational training platforms:

Gap 1 - Absence of End-to-End Voice-First Vocational Platforms
No existing system integrates ASR, NLP, adaptive learning, and vocational certification into a seamless, voice-only experience. Learners are frequently required to supplement platform interactions with sighted assistance.

Gap 2 - Insufficient Multi-Language and Dialect Support
Most reviewed platforms are English-centric. Visually impaired learners in South Asia, Sub-Saharan Africa, and Latin America are underserved due to limited TTS and ASR support for regional languages and dialects.

Gap 3 - Limited Offline Capability
The majority of voice-driven platforms require consistent internet connectivity. Rural and low-income learners with visual impairments frequently lack reliable broadband access, making such platforms inaccessible in practice.

Gap 4 - Weak Adaptive Personalisation
Existing accessible platforms rarely adapt content difficulty, pacing, or module sequencing to individual learner needs. This is particularly problematic in vocational training where skill acquisition follows non-linear progressions.

Gap 5 - Poor Vocational Domain Coverage in NLP Models
General-purpose NLP models lack domain-specific vocabulary for trades such as carpentry, healthcare assistance, textile production, or agriculture. This results in high error rates when learners ask vocationally specific questions.

Gap 6 - Absence of Voice-Accessible Assessment Frameworks
Formal assessment and certification mechanisms for voice-only learners are largely absent. Existing online assessment systems depend on visual interfaces incompatible with the needs of totally blind learners.

Gap 7 - Scalability Limitations in Current Assistive Platforms
Many assistive technology solutions are developed as research prototypes and have not been validated at scale. Deployment to large populations of visually impaired learners across multiple vocational domains remains untested.

Gap 8 - Inadequate Evaluation Frameworks for Accessible Platforms
No standardised benchmarking methodology exists specifically for voice-driven vocational training accessibility. Inconsistent evaluation metrics across studies make cross-study comparison difficult.

VII. CONCLUSION

This review has systematically examined voice-driven accessible vocational training platforms for persons with visual impairment, identifying key technologies, classifying systems into a four-tier taxonomy, and highlighting critical research gaps. The analysis demonstrates that while significant progress has been made in assistive technology and accessible e-learning, no existing solution fully addresses the requirements of visually impaired vocational learners across diverse linguistic, geographic, and socioeconomic contexts.

The proposed direction towards an adaptive, voice-first vocational training platform combines ASR, NLP, neural TTS, adaptive learning algorithms, and offline-first architecture to create a truly inclusive solution. Future research should prioritise multi-language model development, scalable deployment validation, standardised accessibility evaluation frameworks, and industry-standard vocational certification integration.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15527

Planned directions for future work include:

- Development of vocational domain-specific ASR and NLP models
- Integration with national disability employment schemes and certification bodies
- Multilingual expansion covering major regional languages of developing nations
- Longitudinal studies evaluating employment outcomes for platform graduates

REFERENCES

[1]. Kumar, R., Singh, A., and Mehta, V., 2021. Voice-based accessible learning for visually impaired students using TTS and screen reader integration. IEEE Access, 9, pp.45231-45244.
[2]. Patel, S. and Joshi, M., 2020. Integrating automatic speech recognition in accessible e-learning platforms. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems, pp.1-12.
[3]. Adewale, O., Oyedeji, S., and Adeyemo, A., 2022. Adaptive voice platforms for visually impaired learners in low-resource settings. International Journal of Artificial Intelligence in Education, 32(3), pp.512-535.
[4]. Ramesh, K. and Devi, P., 2019. NLP-driven vocational training chatbots for the blind: a conversational AI approach. In Lecture Notes in Computer Science, Springer, pp.221-233.
[5]. Wang, L., Li, J., and Zhang, Y., 2023. Neural text-to-speech for accessibility: benchmarking WaveNet and Tacotron 2 in educational contexts. In Proceedings of Interspeech 2023, pp.4112-4116.
[6]. Nkosi, B. and Fourie, I., 2020. Vocational skills training for blind learners via interactive voice response in Sub-Saharan Africa. In Proceedings of IFIP WCCE 2020, pp.88-97.
[7]. Sharma, A., Gupta, R., and Bansal, P., 2022. Usability audit of JAWS and NVDA compatibility in major LMS platforms. Journal of Disability Studies, 14(2), pp.78-93.
[8]. Chen, Y., Liu, H., and Park, S., 2024. Whisper ASR for visually impaired learners: low word error rates across accents in vocational tasks. In Proceedings of the EMNLP Workshop on Accessibility in NLP, pp.55-63.
[9]. Bhattacharya, S. and Das, M., 2021. Screen reader-based e-learning for Indian languages: challenges and design recommendations. Assistive Technology: The Official Journal of RESNA, 33(4), pp.195-207.
[10]. Guerreiro, T., Nicolau, H., Jorge, J., and Gonçalves, D., 2019. Multimodal real-time feedback for blind learners in vocational training applications. ACM Transactions on Accessible Computing, 11(2), pp.1-28.
[11]. Seelman, K.D. and Hartman, L., 2020. Assistive technology: what it is and how it works in vocational rehabilitation. In Assistive Technology for Persons with Disabilities, 3rd edition, AOTA Press.
[12]. Aziz, N.A., Yusoff, M., and Abdullah, M.T., 2022. Voice user interface accessibility standards and their application in mobile learning for visually impaired adults. Universal Access in the Information Society, 21(1), pp.103-118.
[13]. Alper, M. and Goggin, G., 2021. Digital technology and rights of persons with disabilities in vocational education. Disability & Society, 36(5), pp.724-741.
[14]. Montague, K. and Hanson, V.L., 2023. Towards conversational AI tutors for the visually impaired: design principles and user evaluation. In Proceedings of the ACM ASSETS Conference, pp.1-14.

