International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 11 November-2026 | Impact Factor: 3.5

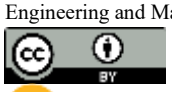


AI-Based Voice to Form Filling System: A Review of Speech Recognition, Information Extraction and Intelligent Form Automation

Prof. Vijaya Jadhav[1], Shubham Sakhare[2], Manisha Suryawanshi[3], Unnati Tambe[4]
[1] Project Guide, [2] MCA Student, [3] MCA Student, [4] MCA Student Department of Master of Computer Applications
ASM's Institute of Business Management & Research, Pune, Maharashtra, India
Roll No.: Shubham Sakhare – 253319 | Manisha Suryawanshi – 253163 | Unnati Tambe – 25339 Assessment / Guide: Prof. Vijaya Jadhav

How to Cite this Article:
Sakhare, S., Suryawanshi, M. & Tambe, U. (2026). AI-Based Voice to Form Filling System: A Review of Speech Recognition, Information Extraction and Intelligent Form Automation. International Journal of Creative and Open Research in Engineering and Management, 02(11), 1-9. https://doi.org/10.55041/ijcope.v2i11.036

License:
This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.
© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.





https://doi.org/10.55041/ijcope.v2i11.036

Abstract— Digital forms support registrations, applications, service requests, surveys, and institutional processes, but completing them can remain difficult for people who face typing barriers, limited digital literacy, language differences, or accessibility challenges. An AI-based voice-to-form filling system aims to make data entry more natural by allowing a user to speak information and converting it into correctly structured form values. This review examines the technologies relevant to the project: automatic speech recognition (ASR), speech-to-text conversion, natural-language processing (NLP), information extraction, semantic field mapping, validation, multilingual interaction, and user confirmation. It considers how voice-driven form automation can coordinate these components and how optical character recognition (OCR) may support document-assisted entry. The discussion emphasizes that transcription accuracy alone does not guarantee successful form completion. Correct extraction, normalization, mapping, correction, privacy, and usability must also be addressed. The review identifies research and implementation priorities for the project, including handling accents and noisy environments, reducing errors that propagate between stages, supporting multiple languages, and providing a secure human-in-the-loop workflow.

Keywords— Artificial Intelligence, Voice-Based Form Filling, Speech Recognition, Speech-to-Text, NLP, Information Extraction, Accessibility, Form Automation, OCR.

I. INTRODUCTION

Digital forms are widely used to collect structured information in education, healthcare, public services, employment, banking, and online applications. The web-based form has made submission and record management easier, but users still need to understand each field, enter information in the required format, and correct mistakes. Repeated typing can be slow, and form labels may be unfamiliar or written in a language that the user does not comfortably read. These barriers affect both convenience and accessibility.

Voice interaction provides an alternative to keyboard-based data entry. In a voice-to-form system, a user can state details naturally—for example, a name, address, date, or purpose—and the application processes the utterance before assigning the resulting value to a suitable field. The workflow is more complex than simply displaying a speech transcript. A form requires data in a defined structure, while natural speech may contain pauses, informal phrasing, corrections, and multiple values in one sentence.

The proposed project, AI-Based Voice to Form Filling System, connects speech input with structured form completion. The system's intended capabilities include user authentication, voice input, speech-to-text conversion, AI/NLP information extraction, intelligent auto-filling, multilingual support, validation and voice-based correction, dynamic form management, document upload with OCR, drafts and submission history, an administrative dashboard, and security/accessibility features. This review focuses only on these project-related topics and the research concepts that support them.

II. REVIEW METHODOLOGY

A focused literature-review method is used to organize the technologies and related systems relevant to voice-enabled form completion. The review considers work on speech recognition, natural-language processing, conversational interfaces, structured information extraction, and document recognition. Literature is examined by its problem, input type, processing approach, output, contribution to form entry, and limitations. Since this paper reviews the basis and design of the proposed project, proposed system features are not presented as experimentally proven outcomes.

The review process follows these stages: identification of related work; screening for relevance to voice and form automation; extraction of methods and system capabilities; thematic comparison; identification of limitations; and formulation of project implications. The central question is how a system can convert a user's spoken information into valid, correctly assigned form fields while allowing the user to verify the result.

III. TECHNICAL BACKGROUND

A. Automatic Speech Recognition and Speech-to-Text

Automatic speech recognition converts an audio signal into a sequence of words. It is a foundational component because the downstream NLP process depends on the recognized text. Speech recognition systems may use acoustic and language models or neural architectures trained on large speech collections. Recognition quality is affected by background noise, microphone characteristics, speaking speed, pronunciation, accent, vocabulary, and the language selected. Names, addresses, dates, and numerical strings are especially sensitive to recognition errors.

Speech-to-text conversion should therefore be treated as a distinct processing stage with a user-visible result. The application can show the recognized sentence, allow the user to repeat an utterance, and provide a correction option. If a recognition result is uncertain, asking the user to clarify is safer than filling a critical field with an unverified



value. For number fields, the application should preserve the recognized meaning while normalizing the display to the form's required format.

B. Natural-Language Processing and Semantic Extraction

Natural-language processing allows the application to interpret the transcript beyond its literal word sequence. A user might say, "I live in Pune and my date of birth is the fifth of May," while the form expects separate city and date fields. Information extraction identifies the entities and values, and semantic interpretation determines which value belongs to which field. A field schema can provide field names, types, required status, and validation rules that guide this process.

Extraction must also handle incomplete answers, user corrections, and references such as "change that to Monday." Normalization converts values into a standard representation—for example, a date format or a phone-number pattern. Rules and constraints can complement AI-based extraction: the model proposes a value, while deterministic validation checks whether it satisfies the field's expected type and format.

C. Conversational Form-Filling Workflow

A conversational interface can guide the user through a form by asking for one value at a time or by accepting a natural multi-field utterance. Field-oriented prompts can help users who do not know the form's terminology. The application should maintain a record of completed, missing, and uncertain fields. When the user revises an answer, the system should update the relevant value rather than create a duplicate or leave an obsolete value in another field.

A voice-first workflow should not remove user control. A review screen should display extracted values in the form itself, identify missing or uncertain information, and allow correction before final submission. This makes the system a guided assistant for form entry rather than an automatic submitter that acts without confirmation.

IV. RELATED WORK AND SUPPORTING TECHNOLOGIES

A. Voice-Based Form Automation

Voice-based form automation combines speech recognition with the structure of a target form. A related bank-form automation study describes a Kannada voice interface in which form prompts are spoken, responses are captured through speech recognition, and recognized information is processed for placement into fields. This type of work is relevant because it demonstrates the importance of language-specific interaction, spoken prompts, and the connection between free speech and structured data. A system designed for a particular language or form, however, may not directly generalize to other languages, form layouts, or user groups.

Conversational form-filling research also highlights the role of interaction design. A system may ask for missing fields, confirm ambiguous values, and let the user revisit earlier answers. These functions can reduce the burden of navigating long forms, but the conversation must remain aligned with the form's data model. A reliable application needs to track which field is currently being discussed and ensure that corrections update the intended field.

B. Multilingual Voice Support

Multilingual support involves more than selecting a speech-recognition language. The application must also interpret the recognized words, normalize local date and number expressions, understand field labels, and provide prompts and correction feedback in the selected language. Users may mix languages within a single utterance, especially for names of institutions, addresses, or technical terms. Language coverage should be considered across the complete pipeline.

For the proposed project, multilingual support is a functional module that can improve accessibility for users who prefer to speak in a familiar language. Its implementation should clearly indicate the active language and provide a way to change it. When a value cannot be reliably interpreted, the system should retain the transcript and ask the user to confirm rather than silently translating or altering personal details.

C. OCR and Document-Assisted Form Entry

Optical character recognition converts text contained in an image or scanned document into machine-readable text. In a voice-to-form application, OCR may be used when users upload a document that already contains some of the information requested by a form. Extracted values can be offered as candidate entries, reducing repeated typing or speaking. However, document quality, rotation, blur, handwriting, layout variation, and similar labels can reduce extraction reliability.

OCR output should be validated and shown to the user before it is accepted into the form. The application should avoid assuming that a document's extracted value is correct or that it belongs to a particular field solely because the text appears nearby. A schema-aware matching step and user confirmation are important when documents contain personal or official information.

D. Accessibility, Privacy, and Trust

Voice interaction can help users who find typing difficult, but the interface should also provide alternatives. Clear instructions, visible status feedback, repeat controls, editable text, keyboard access, and readable labels help users with different needs. Speech interfaces can be challenging in public or noisy environments, so voice should be an option rather than the only possible input method.

Audio recordings, transcripts, uploaded documents, and completed forms may contain personal information. The system should collect only what is needed, define how long data is retained, restrict access, and protect data during transmission and storage. Users should understand when speech is being captured and whether an external service processes it. Security and privacy are essential design requirements, not optional additions.

V. PROPOSED SYSTEM AND PROJECT MODULES

The proposed system follows a pipeline that connects the user interface, speech processing, language understanding, form management, validation, and data storage. The user first registers or signs in and selects a form. The system captures spoken input, generates a transcript, extracts values, maps those values to fields, validates them, and displays a reviewable form. The user can correct values by voice or by editing the displayed fields. Drafts can be saved, and final submission occurs only after confirmation.

A. Module 1: User Registration and Authentication

This module manages account creation, login, and user identity. Authentication helps associate saved drafts and submission history with the appropriate user. The design should protect credentials and restrict access to personal form data. Authentication should be straightforward, with clear error messages and a secure way to recover access.



B. Module 2: Voice Input and Speech Recognition

This module captures audio through the device microphone and supplies it to the speech-recognition component. It should indicate when recording begins and ends, handle microphone permission, and provide a retry option. Recognition should account for the selected language and common environmental limitations. The interface should not continue as though a value were confirmed when audio capture or recognition has failed.

C. Module 3: Speech-to-Text Conversion

The speech-to-text module converts captured speech into readable text. The transcript is displayed so the user can inspect what the system heard. Punctuation, spacing, and common number formats may be normalized while preserving the original meaning. The transcript is an intermediate representation and should not automatically be treated as a final form value.

D. Module 4: AI/NLP Information Extraction

This module identifies the information expressed in the transcript. It can extract entities such as name, address, date, contact information, and application-specific details. The system uses the form's field definitions and context to associate each extracted value with its intended field. When a sentence contains multiple values or ambiguous wording, the system should ask a targeted follow-up question.

E. Module 5: Intelligent Form Auto-Filling

The auto-filling module places extracted and normalized values into the corresponding form controls. It should respect field types, required status, and form-specific rules. A field mapping layer separates the language-processing output from the visual form, allowing different forms to use a common extraction workflow. The user should be able to inspect and edit every populated value.

F. Module 6: Multilingual Voice Support

This module supports voice interaction in the languages selected for the project. It includes recognition-language selection, prompts, interpretation, and normalization of language-specific expressions. Multilingual operation should be tested as an end-to-end capability; accurate transcription in a language does not by itself guarantee correct extraction or field assignment.

V. PROPOSED SYSTEM AND PROJECT MODULES (CONTINUED)

G. Module 7: Validation and Voice-Based Correction

Validation checks whether a value is present where required and whether it follows the field's expected format or constraints. Voice-based correction allows the user to replace an incorrect value by speaking a revised answer. The system must identify which field is being corrected, confirm the replacement when needed, and update the form state consistently. Validation messages should be understandable and should explain how to correct the issue.

H. Module 8: Dynamic Form Management

Dynamic form management supports different forms and changing field structures. Form definitions can specify labels, data types, required fields, options, and validation rules. By using metadata rather than hard-coding every form, the system can reuse the voice-to-field workflow across supported forms. Changes to form definitions should be reflected in prompts, extraction mapping, and validation.

I. Module 9: Document Upload and OCR

This module allows a user to upload a supported document and extract text using OCR. Extracted values may be proposed for matching form fields. The application should show the source and candidate value where practical, allow correction, and avoid accepting low-confidence text without review. File type, size, processing errors, and privacy should be considered.

J. Module 10: Drafts, Submission and History

Draft functionality allows users to pause and resume a form without losing entered information. Submission history can show the status and date of completed forms, subject to the project's data-retention rules. Before submission, the system should check required fields and display a final review. The distinction between saving a draft and submitting a form should be clear to prevent accidental submission.

K. Module 11: Admin Dashboard and Analytics

The administrative dashboard can support management of form templates, field definitions, supported languages, and operational information. Analytics should be limited to appropriate system-level measures, such as completion counts or common validation issues, and should avoid exposing unnecessary personal information. Administrative access should be protected and separated from ordinary user privileges.

L. Module 12: Security, Accessibility and Notifications

This module provides cross-cutting support for privacy, secure handling, accessible interaction, and relevant notifications. It can include access controls, secure communication, appropriate retention, readable interface elements, keyboard alternatives, and status notifications for saved or submitted forms. Notifications should be useful and should not reveal sensitive form details on an insecure channel.

VI. COMPARATIVE ANALYSIS

The technologies reviewed serve complementary roles. Speech recognition converts audio to text; NLP interprets and extracts values; form automation manages prompts and field state; OCR supplies document-derived candidates; and validation plus confirmation checks the result. None of these components alone guarantees a complete and correct form.

A meaningful evaluation should measure end-to-end form completion rather than transcript accuracy alone. Useful measures include field-level extraction accuracy, correct field mapping, completion rate, number of corrections, time to complete, user satisfaction, and errors remaining at submission. These measures should be tested across form types, languages, and realistic audio conditions.



VII. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

A. Robustness Across Accents and Environments

Speech input can vary by accent, speaking speed, microphone, and ambient noise. Future testing should include representative speakers and realistic environments. The system can offer repeat and clarification prompts, but should not claim reliable recognition for languages or conditions that have not been tested.

B. Error Propagation and Field-Level Reliability

A small recognition error may change an extracted value, and a correct value may still be mapped to the wrong field. The system should keep the transcript, extracted candidates, and accepted field values conceptually separate. Confidence-aware clarification, field-schema constraints, and user review can reduce the risk of an incorrect final submission.

C. Multilingual Understanding and Normalization

Future development should consider language mixing, regional expressions, date and number conventions, and correct handling of names. Multilingual support needs evaluation at each stage, including prompt quality, transcript accuracy, extraction, and correction. The system should preserve user intent instead of making unverified translations or assumptions.

D. User Experience and Accessibility

Voice input may reduce typing but can also introduce uncertainty or social discomfort. User testing should assess ease of use, perceived control, task completion time, and the clarity of correction messages. A well-designed system will provide voice and text-based alternatives and will let users inspect the final form.

E. Privacy, Security, and Responsible Data Handling

The project should define how audio, transcripts, documents, drafts, and submitted records are protected. Consent, data minimization, secure transfer, access restrictions, retention periods, and deletion procedures should be considered. The application should communicate what is captured and how it is used.

F. Evaluation and Generalization

A prototype should be evaluated using varied forms and test cases that were not used during development. Testing should include incomplete speech, corrections, multiple values in one utterance, invalid formats, and OCR uncertainty. Results should distinguish speech-recognition performance from extraction and overall form-completion performance.

VIII. CONCLUSION

This review examined the technical foundation and design considerations of the AI-Based Voice to Form Filling System. Speech recognition and speech-to-text conversion produce a transcript; NLP extracts and normalizes information; field mapping assigns values to form controls; and validation, correction, and confirmation help ensure that the final data is appropriate. Multilingual support and OCR can broaden the ways users provide information, while authentication, drafts, history, administration, accessibility, and security support the complete application.

The twelve project modules establish a coherent scope for development. The main challenge is not merely converting speech into text, but ensuring that the right information is placed in the right field and remains reviewable by the user. Future implementation should prioritize end-to-end accuracy, clear correction, privacy, inclusive usability, and testing across languages and form structures.

IX. REFERENCES

[1] "Bank Form Automation Using AI-Powered Speech Recognition and Transliteration in Kannada," 2025 International Conference on Smart & Sustainable Technology (INCSST), IEEE. doi: 10.1109/INCSST64791.2025.11210295.

[2] A. Dalmia, S. Midha, and J. Doshi, "FormBharo: Designing and Evaluating a Voice Agent for Conversational Form Filling in Rural India," arXiv:2608.06027, 2026.

[3] A. Graves, "Sequence Transduction with Recurrent Neural Networks," arXiv:1211.3711, 2012.

[4] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[5] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. draft, 2024.

[6] R. Smith, "An Overview of the Tesseract OCR Engine," in Proc. International Conference on Document Analysis and Recognition (ICDAR), 2007, pp. 629–633.