Design and Evaluation of a Hybrid Georgian Voice Assistant for Public Services

Tamar Tutisani

Georgian American University, Tbilisi, Georgia.

Email: tamar.tutisani@gau.edu.ge

Nino Topuria

Candidate of Technical Sciences, Professor, Georgian Technical University, Tbilisi, Georgia.

Email: nino.topuria@gtu.ge

Abstract

Voice assistants have become an essential component of modern human-computer interaction, allowing users to access information through natural speech. However, most commercial voice assistants provide limited support for low-resource languages. One such language is Georgian, for which limited support for voice technologies makes it difficult for citizens to access local public services through voice interaction. This limitation is especially important for elderly users and people with visual impairments, who often use voice interfaces instead of traditional graphical user interfaces.

This paper presents the design, implementation, and evaluation of a hybrid Georgian language voice assistant that provides voice access to information about public transport and cultural events across Tbilisi. The proposed system combines cloud-based speech recognition, natural language understanding, and speech synthesis technologies within a web application architecture. Speech recognition is performed using Google Cloud Speech-to-Text Chirp 3 [3], while speech synthesis uses Microsoft Edge Neural Text-to-Speech. Intent recognition combines rule-based processing with Google Gemini 2.5 Flash, creating a hybrid architecture that improves robustness while maintaining low latency. Real-time information is obtained from TKT.ge and Google Maps services.

The developed prototype was evaluated through technical performance measurements and user-centered usability testing. Experimental results show that the proposed hybrid architecture provides reliable Georgian speech processing with acceptable latency and high usability.

Keywords: Voice Assistant, Speech-to-Text, Text-to-Speech, Intent Recognition, Natural Language Processing, Progressive Web Application, Accessibility.

Introduction

Voice-based human–computer interaction has rapidly evolved due to advances in artificial intelligence, cloud computing, and natural language processing. Modern voice assistants enable users to access digital services through natural speech, providing an alternative interaction mechanism to conventional graphical interfaces [1], [2].

Despite significant progress in speech technologies, language support remains uneven across linguistic communities. Low-resource languages often face limitations related to insufficient training data, limited availability of specialized models, and reduced integration into commercial voice platforms. Georgian represents such an environment, where the complexity of its morphological structure and limited language resources create additional challenges for speech recognition and natural language understanding systems.

The lack of effective voice-based access is particularly relevant in public information services, where users frequently need immediate access to transportation information, navigation assistance, and cultural event data. Traditional web and mobile platforms often require multiple manual interactions, which may reduce accessibility for users who prefer hands-free interaction or experience difficulties with conventional interfaces [1], [2].

Recent advances in cloud-based speech technologies and large language models have improved the capabilities of conversational systems [3]–[5]. However, relying exclusively on large language models for public information services may increase latency and computational costs.

Therefore, efficient voice assistants require hybrid approaches that combine deterministic processing for common tasks with advanced language understanding for complex interactions. This research presents the design and evaluation of a hybrid Georgian-language voice assistant developed for public transportation and cultural information services in Tbilisi. The proposed system investigates how existing cloud AI technologies can be adapted to support natural speech interaction in a low-resource language environment.

The main contributions of this work are summarized as follows:

1. A hybrid cloud-based architecture for Georgian-language voice interaction integrating speech recognition, intent processing, information retrieval, and speech synthesis.
2. A hybrid intent recognition approach combining rule-based processing with Google Gemini 2.5 Flash to improve efficiency and contextual understanding.
3. Georgian-specific text normalization techniques for improving speech synthesis quality.
4. Integration of real-time public information sources, including transportation and cultural event data.
5. Experimental evaluation based on technical performance measurements and usability assessment using the System Usability Scale (SUS) [6].

System Architecture and Implementation

Within this research, a Georgian-language voice assistant prototype, named the "Tbilisi Assistant", was developed to provide users with information about public transportation and cultural events through natural speech interaction. The system aims to improve access to public information by reducing reliance on conventional text-based interfaces. It adopts a three-layer architecture comprising the user interface layer, application logic layer, and data management layer, which enhances modularity, supports independent component development, and facilitates future system expansion.

The user interface is implemented as a Progressive Web Application (PWA), providing cross-platform accessibility without requiring separate installation. The interface follows a mobile-first design approach [7] and incorporates accessibility-oriented features, including high-contrast visual elements, screen-reader compatibility, and voice-centered interaction scenarios [8],[9].

The backend layer is implemented using the Python FastAPI framework, which provides asynchronous request processing and efficient communication with external services. The backend manages audio processing, user request analysis, information retrieval, and response generation. SQLite is used for temporary storage of cached transportation and cultural information.

The voice processing pipeline begins with speech recognition, where user audio input is captured through browser-based media interfaces [10] and converted into text. During system development, local Whisper-based processing was evaluated; however, the obtained Georgian recognition quality and CPU execution time were insufficient for real-time interaction requirements. Therefore, the final implementation uses Google Cloud Speech-to-Text Chirp 3 [3], which provides more stable recognition performance for Georgian speech, including local names and geographic terms.

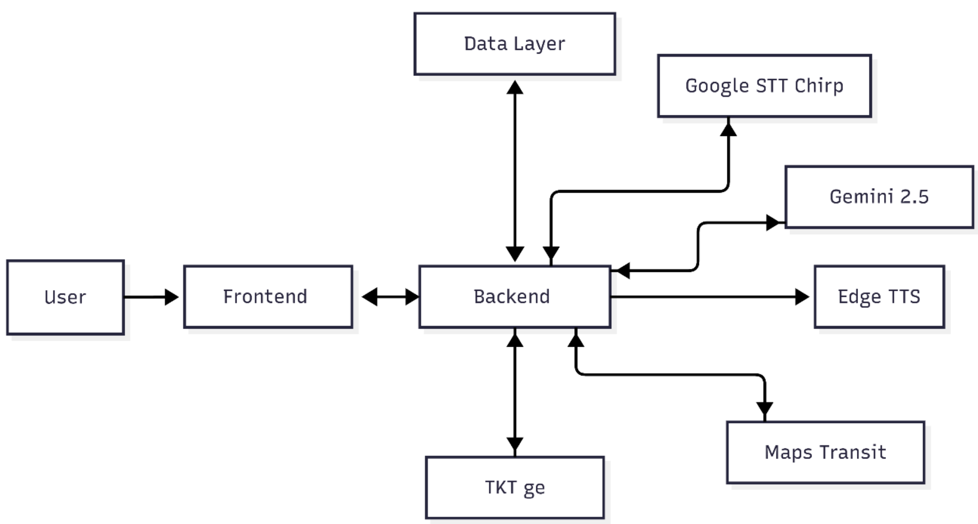
User intent recognition is implemented through a hybrid NLP approach combining rule-based processing and Google Gemini 2.5 Flash [4]. Frequently occurring structured queries, such as finding nearby stops or requesting route information, are processed using predefined rules to reduce latency and minimize unnecessary large language model requests. More complex natural language queries are forwarded to Gemini for contextual interpretation.

To address Georgian linguistic characteristics, preprocessing and text normalization techniques were applied. These techniques handle morphological variations, date and time expressions, abbreviations, and grammatical structures, improving both intent interpretation and speech output quality. Speech responses are generated using Microsoft Edge Neural Text-to-

Speech [5]. Before synthesis, generated responses undergo normalization to ensure more natural Georgian pronunciation.

The system integrates external data sources for real-time information retrieval. Transportation-related queries use Google Maps Platform services, including Directions API, Places API, and Geocoding API [11]. Cultural event information is obtained from the TKT.ge platform. To optimize response speed and data freshness, a stale-while-revalidate caching strategy is applied, allowing immediate access to cached information while updating data in the background. The complete processing pipeline is illustrated in Figure 1. The workflow includes audio acquisition, speech recognition, intent classification, information retrieval, response generation, and speech synthesis.

Figure 1. System architecture of the "Tbilisi Assistant"



Research Methodology and Evaluation Approach

The research was conducted using a system design and prototype-based development methodology. The objective was to design and evaluate a Georgian-language voice interface capable of providing public information services through natural speech interaction.

Initially, functional and non-functional requirements were defined. Functional requirements included voice input processing, speech recognition, intent identification, information retrieval, and voice response generation. The selected application scenarios focused on public transportation and cultural event information. Non-functional requirements addressed practical deployment considerations, including response latency, usability, accessibility, platform independence, and operational efficiency.

The development process consisted of several stages. First, different technological components were evaluated, including local and cloud-based speech recognition approaches. Second, the prototype architecture was implemented and integrated with external services. Finally, technical performance evaluation and user-centered usability testing were conducted.The technical evaluation focused on response latency, speech recognition stability, and overall system performance. User experience was assessed using the System Usability Scale (SUS), providing a quantitative measure of interface usability [6].

Architectural Design Rationale

The architectural design was driven by the need to provide an efficient and accessible voice interaction framework while maintaining flexibility for future extensions. The PWA-based interface was chosen to ensure platform independence and simplify deployment across different devices without requiring separate native applications. This approach also supports the integration of accessibility features required for inclusive public service applications.

The hybrid intent recognition mechanism was introduced to address the trade-off between response speed and natural language understanding. Structured requests, such as transportation searches and route-related queries, are processed using predefined rules to achieve low latency, while more complex user inputs are analyzed using the large language model component to improve contextual interpretation.

Cloud-based AI services were selected to provide reliable speech recognition and synthesis capabilities without requiring significant local computational resources. This approach is particularly appropriate for Georgian, where developing and maintaining fully local speech models remains challenging due to limited language resources.

The modular architecture enables future integration of additional public services, improved Georgian language models, and new data sources without significant changes to the existing system structure.

Conclusion

This study presented the design, implementation, and evaluation of a hybrid Georgian-language voice assistant for public transportation and cultural information services. The research demonstrated the feasibility of adapting existing artificial intelligence technologies to support natural speech interaction in a low-resource language context.

The experimental evaluation showed that the combination of cloud-based speech services, hybrid intent recognition, and Georgian-specific language processing techniques provides an effective solution for practical voice-based public information systems. The proposed approach enabled efficient processing of common requests while preserving the ability to interpret more complex natural language queries. The system achieved a SUS score of 89.17, indicating strong perceived usability among evaluation participants.

The developed prototype demonstrates that combining modern AI services with accessibility-oriented design principles can improve access to digital public services for Georgian-speaking users. The modular architecture also provides a foundation for future extensions, including additional service domains, enhanced language models, and multilingual support.

The study has several limitations. The current implementation relies on external cloud services and stable Internet connectivity, and the user evaluation was conducted with a limited participant group. Future research will focus on expanding real-world testing, improving Georgian language understanding capabilities, and investigating more adaptive conversational models for public service applications.

REFERENCES

[1] M. T. Harris and W. A. Rogers, "Social participation challenges and response strategies identified by adults aging with a disability", *Journal of Elder Policy*, vol. 3, pp. 95–109, 2024.

[2] P. Esquivel, K. Gill, M. Goldberg, S. A. Sundaram, L. Morris, and D. Ding, "Voice Assistant Utilization among the Disability Community for Independent Living: A Rapid Review of Recent Evidence, "*Human Behavior and Emerging Technologies*, Article no. 6494944, 2024.

[3] Google Cloud, "Speech-to-Text documentation", *Google Cloud Documentation*. https://cloud.google.com/speech-to-text/docs

[4] Google, "Gemini API documentation", *Google AI for Developers*. https://ai.google.dev/gemini-api/docs

[5] Microsoft, "Speech service documentation", *Microsoft Azure Documentation*. https://learn.microsoft.com/en-us/azure/ai-services/speech-service/

[6] J. Sauro and J. R. Lewis, Quantifying the User Experience: Practical Statistics for User Research, 3rd ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

[7] A. Firtman, Progressive Web Apps: The Future of Web Development, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2023.

[8] W3C, "Web Content Accessibility Guidelines (WCAG) 2.1", World Wide Web Consortium. https://www.w3.org/TR/WCAG21/

[9] W3C, "Accessible Rich Internet Applications (WAI-ARIA) 1.2," World Wide Web Consortium. https://www.w3.org/TR/wai-aria-1.2/

[10] Mozilla Developer Network, "MediaDevices.getUserMedia", *MDN Web Docs*. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/API/MediaDevices/getUserMedia

[11] Microsoft, "What is a Progressive Web App?", *Microsoft Edge Documentation*. https://learn.microsoft.com/en-us/microsoft-edge/progressive-web-apps-chromium/

[12] Google, "Google Maps Platform Documentation," *Google for Developers*. https://developers.google.com/maps/documentation