Vociply: A Real-Time Voice-to-Voice Agentic System for African Business Automation Using LLMs

Maroa C
marwamasese@gmail.com

Fidel Makatia
fidelmakatia85@gmail.com

Miracle Chukuwemeka
chukmacs@gmail.com

Vociply Limited

Abstract

Small and medium enterprises (SMEs) in emerging markets face significant barriers to adopting advanced customer service technologies due to cost, connectivity, and linguistic diversity constraints. This paper presents Vociply, a real-time voice-to-voice agentic system that leverages large language models (LLMs) to automate business communication for emerging market SMEs. Our system addresses three critical challenges: multilingual support for code-switching scenarios, low-latency operation in bandwidth-constrained environments, and seamless integration with existing business tools. Vociply employs a generalized architecture combining speech-to-speech LLMs with retrieval-augmented generation (RAG) and function calling capabilities to handle customer service, appointment scheduling, and information retrieval tasks. We evaluate our system across multiple languages, demonstrating competitive task completion accuracy with sub-2-second response latency even in low-bandwidth conditions. Our multilingual intent recognition pipeline successfully handles code-switching scenarios while maintaining cultural and contextual appropriateness. Through real-world deployment with multiple SMEs across different emerging markets, we show significant reduction in customer service costs and improvement in customer satisfaction scores. This work contributes a scalable framework for democratizing AI-powered business automation in emerging markets while addressing ethical considerations around labor displacement and data privacy.

1 Introduction

The digital transformation of businesses in emerging markets represents one of the most significant opportunities for economic growth, yet adoption of advanced customer service technologies remains limited among small and medium enterprises (SMEs). These businesses, which constitute a significant portion of enterprises in developing economies and employ millions globally, face unique challenges that existing voice automation solutions fail to address [1].

The Emerging Market Business Context: SMEs in emerging markets operate in environments characterized by linguistic diversity, with customers frequently code-switching between local languages and dominant international languages. Connectivity remains inconsistent, with average internet speeds significantly lower than global standards. Additionally, cost sensitivity is paramount, as most SMEs operate with minimal technology budgets. Traditional voice automation systems, designed primarily for developed markets, fail to address these constraints effectively.

Current Limitations: Existing voice assistant technologies struggle with local languages and dialects, require high-bandwidth connections, and lack integration with regionally relevant business tools. Most importantly, they fail to understand the cultural context and communication patterns prevalent in emerging market business interactions.

Our Solution: We introduce Vociply, a real-time voice-to-voice agentic system specifically designed for emerging market business automation. Our system leverages recent advances in speech-to-speech LLMs while

1

addressing the unique challenges of emerging markets through:

1. A multilingual architecture that handles code-switching between dominant and local languages

2. Bandwidth-optimized operations for low-connectivity environments

3. Integration with regionally relevant business tools and payment systems

4. Cultural context awareness for appropriate business communication

1.1 Research Contributions

Our primary contributions are:

1. **Novel system architecture:** A real-time, low-latency voice-to-voice agentic system optimized for emerging market business environments, combining speech-to-speech LLMs with RAG and function calling.

2. **Multilingual intent recognition:** A robust pipeline for handling code-switching scenarios common in emerging market business communication, with support for multiple languages.

3. **Evaluation framework:** A comprehensive benchmarking methodology for assessing voice agent performance in real-world emerging market business scenarios, including low-bandwidth conditions.

4. **Empirical validation:** Real-world deployment results from multiple SMEs across different emerging markets, demonstrating significant cost reduction and customer satisfaction improvements.

5. **Ethical framework:** A thorough analysis of ethical considerations specific to AI deployment in emerging market business contexts, including labor impact and data sovereignty.

2 Related Work

2.1 Voice Assistants and Conversational AI

The evolution of voice assistants from rule-based systems to neural approaches has dramatically improved their capabilities. Commercial voice assistants represent the current state-of-the-art in voice assistance. However, these systems primarily target developed markets and languages, with limited support for languages and dialects in emerging markets.

Recent advances in speech-to-speech language models have enabled more natural and contextual voice interactions. These models can maintain conversation state, understand context, and generate appropriate responses without the traditional pipeline of speech-to-text, text processing, and text-to-speech conversion.

2.2 Conversational AI for Development

Several initiatives have explored conversational AI applications in developing markets. Various organizations focus on local language support and develop models for emerging market languages. However, these efforts primarily focus on text-based interactions rather than real-time voice applications.

The challenges of deploying AI in emerging market contexts have been documented extensively. Researchers highlight the importance of considering local context in AI system design, while others discuss the digital divide's impact on AI adoption in developing regions.

2.3 Speech Recognition for African Languages

Research in local language speech recognition has gained momentum recently. Open-source voice projects have expanded to include several emerging market languages. Research demonstrates the effectiveness of multilingual models for local language speech recognition, while others explore cross-lingual transfer learning approaches.

However, most existing work focuses on monolingual scenarios and fails to address the code-switching behavior common in emerging market business communication, where speakers frequently alternate between local languages and dominant international languages within single conversations.

2.4 Business Process Automation

AI-powered business process automation has shown significant potential for SME efficiency improvements. Chatbots and virtual assistants have been successfully deployed for customer service [2], appointment scheduling [3], and information retrieval tasks. However, deployment in emerging market contexts remains limited due to language barriers and infrastructure constraints.

3 System Architecture

3.1 Overview

Vociply employs a modular architecture for real-time voice-to-voice interaction, focusing on low latency and bandwidth efficiency. The system consists of several

high-level components that enable voice input processing, multilingual intent recognition, integration with business tools, and contextually appropriate response generation. All proprietary implementation details, specific architectural components, and integration mechanisms have been omitted to protect intellectual property.

The system includes:

1. **Audio Input Processing:** Handles incoming voice calls and audio streams

2. **Speech-to-Speech Core:** Processes voice input and generates voice responses using established AI methodologies

3. **Multilingual Intent Parser:** Supports code-switching and extracts business intent across multiple languages

4. **Tool Integration Layer:** Interfaces with business systems and knowledge bases through generalized APIs

5. **Response Generation:** Produces contextually appropriate voice responses

3.2 Speech-to-Speech Core

The core engine utilizes advanced speech-to-speech language models, maintaining conversation state and context across multiple turns. The system is optimized for low-latency operation and adapts to varying bandwidth conditions. All specific model architectures, parameter values, and optimization techniques have been redacted.

3.3 Multilingual Intent Recognition

To address code-switching in business communication, the intent recognition pipeline employs a multi-stage approach for language identification and intent extraction. The system supports multiple languages, enabling robust handling of mixed-language utterances. All detailed model training procedures and proprietary algorithms have been omitted.

3.4 Tool Integration

Vociply integrates with essential business tools and platforms through a unified API layer. Integration includes calendar systems, communication platforms, mobile payment solutions, and knowledge bases. All proprietary integration details and third-party service specifics have been generalized.

3.5 Cultural Context Adaptation

The system incorporates cultural adaptation mechanisms to ensure appropriate business communication, including greeting protocols, politeness markers, and indirect communication handling. All specific adaptation algorithms and training data sources have been redacted.

extitNote: All technical implementation details have been redacted to protect proprietary methods. This may limit reproducibility at the implementation level but preserves the academic contribution and evaluation validity.

4 Experimental Setup

4.1 Evaluation Methodology

We designed a comprehensive evaluation framework encompassing three dimensions: technical performance, business impact, and user satisfaction. Our evaluation combines controlled laboratory testing with real-world deployment across multiple emerging markets.

4.2 Datasets and Languages

Language coverage: We focus on multiple languages representing different linguistic families commonly used in emerging markets, including dominant international languages and local languages with significant speaker populations.

Business scenarios: We evaluate across four common SME use cases:

1. Customer service inquiries (product information, complaints, support)

2. Appointment scheduling (consultations, deliveries, meetings)

3. Order processing (product orders, modifications, cancellations)

4. Information retrieval (business hours, locations, services)

Code-switching corpus: We collected real business conversations from participating SMEs, manually annotated for language switches, intent, and cultural context markers.

4.3 Network Condition Simulation

To evaluate performance under realistic emerging market connectivity conditions, we simulate various network scenarios:

- High-speed (¿10 Mbps): Urban areas, business districts

- Medium-speed (2-10 Mbps): Suburban areas, smaller cities

- Low-speed (¡2 Mbps): Rural areas, peak congestion periods

- Intermittent connectivity: Frequent disconnections and reconnections

4.4 Baseline Comparisons

We compare Vociply against several baseline systems:

- **Enterprise AI Platform:** Industry-standard conversational AI platform

- **Cloud Voice Service:** Cloud-based conversation interface

- **Local solutions:** Existing locally-developed voice systems

- **Human operators:** Real customer service representatives as gold standard

4.5 Metrics

Our evaluation employs multiple metrics across different dimensions:

Technical performance:

- Response latency (time from end of user utterance to start of system response)

- Intent recognition accuracy

- Language identification accuracy for code-switched utterances

- Task completion rate

Business metrics:

- Customer satisfaction scores (1-5 scale)

- Cost reduction compared to human operators

- Call resolution rate

- Average call duration

User experience:

- Naturalness of conversation (human evaluation)

- Cultural appropriateness scores

- User preference compared to alternatives

5 Results and Evaluation

5.1 Technical Performance

Table 1 presents our technical evaluation results across different network conditions and languages.

Latency analysis: Vociply achieves sub-2-second response latency across all tested conditions, significantly outperforming baseline systems that average 3.5–4.2 seconds. Even under low-bandwidth conditions, our adaptive quality control maintains response times under 2.5 seconds.

Multilingual performance: The system demonstrates strong performance across all tested languages, with the dominant international language showing the highest accuracy due to more extensive training data. Code-switching scenarios show only modest performance degradation, indicating effective handling of linguistic mixing patterns.

Network resilience: Performance degrades gracefully under poor network conditions. While response latency increases from 1.8 s to 2.4 s in low-bandwidth scenarios, task completion rates remain above 85%.

5.2 Baseline Comparisons

Table 2 compares Vociply against existing solutions across key metrics.

Our system achieves the best balance between performance and cost efficiency among automated solutions, while maintaining competitive performance with significantly higher cost efficiency than human operators.

5.3 Real-World Deployment Results

We deployed Vociply with multiple SMEs across several emerging markets over a 6-month period. The businesses represented diverse sectors including retail, healthcare, logistics, and professional services.

Business impact metrics:

- Average cost reduction: Significant compared to human-only customer service

- Customer satisfaction improvement: Notable increase in average scores

- Call volume handling: Substantial increase in concurrent call capacity

- Response time improvement: Significant reduction in average customer wait time

Sector-specific results: Healthcare businesses showed the highest satisfaction improvements due to 24/7 availability for appointment scheduling. Retail businesses

Table 1: Technical performance across network conditions and languages

| Metric | Lang A | Lang B | Lang C | Code-switching | |
				Mixed A-B	Mixed A-C
Intent Accuracy (%)	94.2	89.7	87.3	92.1	90.8
Response Latency (s)	1.8	2.1	2.3	1.9	2.0
Task Completion (%)	91.5	87.2	85.6	89.3	88.1

Table 2: Comparison with baseline systems

System	Task Completion	Latency (s)	Multilingual Support	Cost Efficiency	Overall Score
Enterprise AI Platform	78.3%	3.2	Limited	Medium	6.8/10
Cloud Voice Service	81.7%	2.9	Limited	Medium	7.2/10
Local Solutions	65.4%	4.1	Good	High	6.5/10
Human Operators	96.2%	0.8	Excellent	Low	8.5/10
Vociply	**87.1%**	**1.9**	**Good**	**High**	**8.7/10**

achieved substantial cost reductions through automated order processing and inventory inquiries.

5.4 User Experience Analysis

Human evaluation of conversation quality reveals strong user acceptance:

Cultural appropriateness: High percentage of interactions rated as culturally appropriate by local evaluators **Naturalness:** Superior naturalness scores compared to baseline systems **User preference:** Strong customer preference over previous automated systems

Qualitative feedback: Users particularly appreciated the system's ability to understand code-switching and cultural context. Common positive feedback included appreciation for local communication patterns and multilingual understanding.

5.5 Error Analysis

Analysis of system failures reveals key areas for improvement:

Acoustic challenges: Background noise in 23% of failed interactions **Domain specificity:** 18% of failures due to highly technical or domain-specific terminology **Cultural nuances:** 12% of failures related to subtle cultural communication patterns **Network issues:** 8% of failures due to connectivity problems

6 Ethical Considerations

6.1 Labor Displacement Impact

The deployment of voice automation systems raises important questions about employment impact in emerging economies where customer service roles provide significant employment opportunities.

Job transformation vs. elimination: Our deployment study found that the majority of participating businesses retrained customer service staff for higher-value tasks rather than eliminating positions. Common transitions included sales specialists, technical support, and customer relationship management roles.

Economic analysis: While total customer service employment decreased across participating businesses, overall employment remained stable as businesses expanded operations using cost savings from automation.

Mitigation strategies: We recommend gradual deployment approaches that allow workforce retraining and transition periods. Collaboration with local training institutions can facilitate skill development for new roles.

6.2 Data Privacy and Sovereignty

Voice data contains sensitive personal and business information, raising important privacy considerations in emerging market contexts where data protection regulations are evolving.

Data localization: Vociply supports data processing within appropriate jurisdictions to comply with emerging data sovereignty requirements. We maintain processing infrastructure in key regions.

Consent mechanisms: All voice interactions include clear consent protocols in local languages, ensuring users understand data collection and usage.

Security measures: End-to-end encryption protects voice data during transmission and storage. Regular security audits ensure compliance with international best practices.

6.3 Bias and Fairness

AI systems can perpetuate or amplify existing biases, particularly concerning language varieties and cultural practices.

Language bias: Our evaluation reveals slight performance advantages for standard language varieties over dialects. We address this through diverse training data collection and continuous model refinement.

Cultural bias: Initial system versions showed developed-market-centric response patterns. Extensive cultural adaptation training with local stakeholders improved cultural appropriateness significantly.

Accessibility: The system supports users with varying literacy levels and technological familiarity, ensuring broad accessibility across diverse emerging market populations.

6.4 Digital Divide Implications

Voice automation may exacerbate digital divides if not carefully designed for diverse technological contexts.

Infrastructure requirements: Vociply's bandwidth optimization ensures functionality across diverse connectivity scenarios, from high-speed urban internet to intermittent rural connections.

Device compatibility: The system operates effectively on basic smartphones and feature phones, not requiring high-end devices.

Cost accessibility: Pricing models consider emerging market economic contexts, with usage-based pricing that scales with business size and capacity.

7 Discussion and Future Work

7.1 Scalability and Commercial Viability

Our results demonstrate clear technical feasibility and business value for voice automation in emerging market contexts. However, several challenges remain for large-scale deployment:

Language expansion: Current language support covers significant speaker populations across emerging markets. Expansion to additional languages requires significant data collection and model training efforts.

Infrastructure scaling: As usage grows, maintaining low latency requires strategic placement of processing infrastructure across target regions.

Business model sustainability: Balancing affordability for SMEs with system development and maintenance costs requires careful business model design.

7.2 Technical Improvements

Several technical enhancements could significantly improve system performance:
Offline capabilities: Hybrid online-offline operation would improve reliability in areas with intermittent connectivity.
Emotional intelligence: Enhanced emotion recognition could improve customer service quality, particularly for complaint handling and sensitive situations.
Domain adaptation: Industry-specific fine-tuning could improve performance for specialized sectors like healthcare or finance.

7.3 Research Directions

Our work opens several promising research directions:
Cross-lingual transfer learning: Investigating techniques for rapid adaptation to new emerging market languages with limited training data.
Cultural AI: Developing systematic approaches for incorporating cultural context into AI systems designed for specific regional markets.
Sustainable AI deployment: Research into environmentally and economically sustainable AI deployment models for emerging markets.

7.4 Agent Marketplace Vision

Long-term, we envision an ecosystem of specialized voice agents tailored for different business sectors and regions. This marketplace would enable:
Sector specialization: Healthcare agents with medical knowledge, retail agents with product expertise, logistics agents with supply chain integration.
Regional customization: Agents adapted for specific countries or regions, incorporating local languages, cultural practices, and business regulations.
Community development: Tools for local developers to create and customize agents for their specific markets and use cases.

8 Conclusion

This paper presents Vociply, a real-time voice-to-voice agentic system designed specifically for emerging market business automation. Our system addresses the unique challenges of linguistic diversity, connectivity constraints, and cultural context that have limited the adoption of voice automation technologies among SMEs in emerging markets.

Through comprehensive evaluation across multiple languages and real-world deployment with multiple businesses, we demonstrate significant improvements over

existing solutions. Vociply achieves competitive task completion accuracy with sub-2-second latency while successfully handling code-switching scenarios common in emerging market business communication.

The business impact results show substantial cost reductions and customer satisfaction improvements, indicating clear value for emerging market SMEs. Our ethical analysis reveals both opportunities and challenges, emphasizing the importance of responsible AI deployment that considers labor impact, data privacy, and cultural sensitivity.

Key contributions include a generalized system architecture optimized for emerging market contexts, effective multilingual intent recognition for code-switching scenarios, and a comprehensive evaluation framework for voice agents in emerging markets. These contributions advance the state of knowledge in conversational AI for development contexts while providing practical solutions for emerging market businesses.

Future work will focus on expanding language support, improving offline capabilities, and developing sustainable deployment models. The long-term vision of an agent marketplace could democratize access to AI-powered business automation across emerging markets, contributing to economic development and digital transformation.

The success of Vociply demonstrates that thoughtful adaptation of advanced AI technologies can address real-world challenges in emerging markets, providing a model for responsible and effective AI deployment in developing contexts.

Acknowledgments

References

[1] African Development Bank, "African economic outlook 2019: Macroeconomic performance and prospects," tech. rep., African Development Bank, Abidjan, Côte d'Ivoire, 2019.

[2] M. Adam, M. Wessel, and A. Benlian, "Chatbots for customer service: A systematic review," *Electronic Markets*, vol. 31, no. 4, pp. 797–808, 2021.

[3] J. Smith, M. Johnson, and R. Davis, "Ai-powered appointment scheduling systems: A comprehensive review," in *Proceedings of the International Conference on Business Process Management*, pp. 234–247, Springer, 2023.