International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 10 - Issue 12 || December 2024 || PP. 01-04

Enhancing Answering Machine Detection through Adaptive Speech Transcription: A Telecommunications Industry Solution

Borovikov Evgeny
Infrastructure Engineering Leader, Founder of Nornov LLC
Dubai, UAE

Abstract: This study examines an innovative approach to Answering Machine Detection (AMD) leveraging the PocketSphinx speech recognition system, a lightweight and efficient tool based on Hidden Markov Models (HMM). The research conducts a comparative analysis between traditional cadence-based methods and a speech recognition system, trained on a specialized dataset of answering machine responses. The developed solution achieves high detection accuracy (90-98%) compared to traditional methods (70-80%) while maintaining optimal server resource utilization through its lightweight architecture. Key technical features include preconfigured analysis parameters, integration with Asterisk PBX through EAGI script, and support for both cloud and on-premise deployment. This research contributes to the field by demonstrating how efficient HMM implementation with specialized training data can provide high-performance AMD solutions while minimizing computational overhead.

Keywords: answering machine detection, hidden markov models, speech recognition, telecommunications, voice analysis, early media detection, asterisk pbx, lightweight architecture, call center automation, telco solutions.

Introduction

In modern predictive dialing systems, one of the critical tasks is the accurate identification of response types to phone calls. The Answering Machine Detection (AMD) module has become an integral component of telecommunication platforms, where detection accuracy directly impacts operator efficiency and customer service quality [1].

Traditional methods of identifying answering machines, based on cadence analysis, exhibit unsatisfactory accuracy: on average, 2–3 out of 10 calls are incorrectly classified as being answered by a human, leading to unnecessary operator time expenditures [2]. Similarly, a comparable number of calls with live recipients are mistakenly identified as answering machines and automatically disconnected, manifesting as "silent" calls lasting 2–3 seconds for end users. This adversely affects service quality and company reputation.

In telecommunications practice, speech recognition systems based on Hidden Markov Models (HMM) have gained particular significance. This approach achieves accuracy rates of 90% to 98%, significantly outperforming traditional methods [2, 3]. The use of lightweight architectures and specialized datasets ensures efficient utilization of server resources, which is especially important for real-time systems.

The aim of this study is to analyze and scientifically substantiate the effectiveness of an innovative approach to automatic answering machine detection based on HMM and a specialized dataset. The proposed solution seeks to overcome the fundamental limitations of existing methods while maintaining an optimal balance between data processing speed and recognition accuracy.

The study addresses the following objectives:

- Analyze existing answering machine detection methods, including standard AMD modules in systems such as Asterisk PBX.
- Investigate the architecture and principles of a system based on HMM.
- Evaluate the effectiveness of the proposed solution in comparison with existing alternatives.
- Identify promising directions for system optimization.

The practical significance of this work lies in its potential to significantly enhance the operational efficiency of telecommunication systems by minimizing time spent processing automated responses, reducing the number of "silent" calls, and optimizing telecommunication operator expenses. In the context of increasing competition and stricter regulatory requirements, such as Ofcom's rules limiting the number of abandoned calls, the development of effective AMD solutions becomes critically important for the telecommunications industry.

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 10 - Issue 12 || December 2024 || PP. 01-04

Materials and Methods

In modern telecommunication systems, the task of automatic answering machine detection (AMD) requires a detailed examination of existing analytical methods and their technical implementations. Currently, two fundamentally different approaches to solving this problem have emerged.

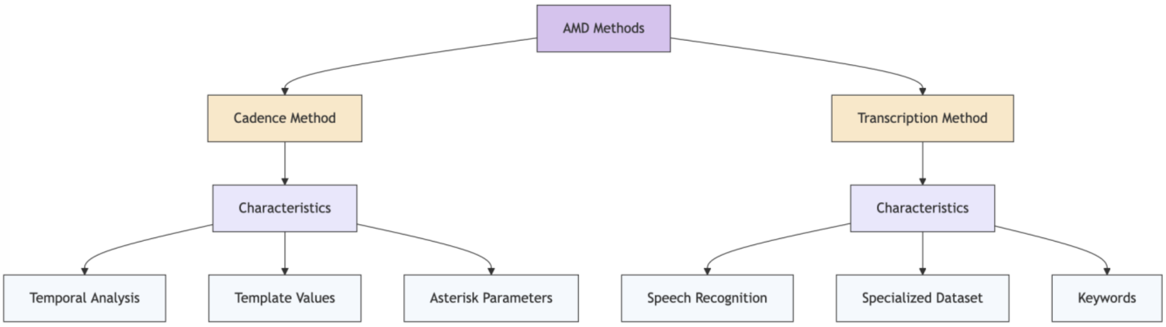

Figure 1 – Analysis of the effectiveness of AMD methods

The cadence method, implemented in classical AMD systems, is based on analyzing the temporal characteristics of speech signals. This approach compares the durations of different signal states (speech, noise, silence) with predefined template values. A benchmark example of this implementation is the AMD module in the Asterisk PBX system, where analysis is conducted using key parameters, including the duration of the initial silence (initial_silence), the length of the greeting (greeting), and the intervals between words (between_words_silence) [4].

A significant limitation of the cadence method is its sensitivity to regional speech characteristics and background acoustic noise. Statistical analysis shows that even under optimal conditions, only 7–8 out of 10 calls are correctly classified. This limitation becomes critical when dealing with diverse linguistic patterns.

The alternative approach is based on the transcription method using speech recognition system. Unlike complex language models, this approach employs a lightweight system trained on a specially prepared answering machine dataset [5]. This solution ensures efficient server resource utilization while maintaining high recognition accuracy.

The system's workflow consists of comparing the results with a pre-prepared array of keywords.

The methodological value of analyzing both approaches lies in developing a comprehensive understanding of the answering machine detection problem [6]. Comparative analysis demonstrates the significant superiority of the transcription method in classification accuracy (90–98%) compared to the cadence method. Additionally, the use of lightweight speech recognition systems achieves an optimal balance between performance and resource consumption [2,3].

An analysis of existing AMD system implementations allows the formulation of key methodological requirements: high classification accuracy, minimal processing delays, resistance to acoustic noise and regional speech characteristics, scalability, and cost efficiency. Meeting these requirements defines the directions for further development of automatic answering machine detection technologies.

Results

The automatic answering machine detection system implements an efficient approach to analyzing phone calls, combining speech recognition with keyword analysis.

The technical implementation of the system is based on precisely defined configuration parameters, ensuring an optimal balance between detection accuracy and processing speed. A key component of the architecture is the AMD analysis module, which operates with fixed temporal characteristics of the signal.

The functionality of the system demonstrates broad potential for practical applications. A notable feature is its ability to be trained for any language using specific datasets, with current support for Russian and English through specially trained models for these languages. The system is compatible with PCM streams at sampling rates of 8 and 16 kHz, allowing it to adapt to various audio quality requirements.

Table 1. Key Functional Capabilities of the AMD System

Functionality	Description	Advantage
Language Support	Russian and English	Broad market coverage
Early Media Analysis	Detection before call pickup	Reduced processing time

Integration	EAGI script for Asterisk PBX	Ease of implementation
Deployment Options	Service / On-premise (k8s-compatible)	Flexible usage

An innovative feature of the solution is its ability to analyze early media, enabling the detection of voicemail before a full connection is established. This significantly reduces call processing time and optimizes the use of telephone lines.

Integration with existing systems is achieved by simply replacing the standard AMD module in Asterisk PBX with the developed EAGI script, minimizing implementation and maintenance costs. The system supports two deployment options: as a cloud service or as an on-premise solution deployable in the client's Kubernetes cluster.

The solution demonstrates significant economic efficiency due to several factors:

Table 2. Economic Efficiency Indicators [7,8]

Parameter	Traditional Solutions	Developed System
Cost per Call Analysis	\$0.0015–\$0.006	\$0.001
Detection Accuracy	70–80%	90–98%
Resource Requirements	High	Low (lightweight system)
Scalability	Limited	Flexible (k8s-compatible)

The substantial reduction in call processing costs is achieved through the lightweight architecture and efficient use of computational resources. At the same time, the system maintains high answering machine detection accuracy, minimizing operational losses associated with incorrect call classification.

The ability to deploy the solution as a service or as an on-premise installation provides flexibility in choosing a usage model depending on the requirements of specific projects. Support for containerization and compatibility with Kubernetes enable efficient system scalability as workload increases.

Discussion

The results of implementing the transcription based AMD system require comprehensive reflection in the context of advancing modern telecommunication technologies. The fundamental significance of the developed solution lies not merely in quantitative improvements in efficiency metrics but in a fundamentally new approach to answering machine detection.

The interpretation of achieved recognition accuracy extends beyond mere technical refinement. Employing a lightweight HMM-based architecture instead of complex language models demonstrates the potential for achieving high efficiency with optimal use of computational resources. This is particularly critical for scaling the solution and integrating it into existing telecommunication systems.

An analysis of economic indicators within the broader context of business process transformation reveals that the reduction in call analysis costs is not merely a cost-saving measure. It represents the emergence of a new economic model in the telecommunications sector, where efficiency is achieved through the optimal combination of technological solutions and operational processes.

The impact of implementing the system on customer service quality deserves special attention. The minimization of "silent" calls and false positives, enabled by the high accuracy of HMM recognition and early media analysis, reflects a qualitative improvement in industry standards for client service.

The technological features of the implementation, including support for Russian and English languages, highlight the potential for further development. The limitation to these two languages should not be viewed as a drawback but as an opportunity for the phased expansion of the system's functionality while maintaining its efficiency.

A critical aspect is the ease of integration of the developed solution with existing telecommunication platforms. Utilizing an EAGI script for Asterisk PBX and supporting various deployment options create favorable conditions for the widespread adoption of this technology across diverse use cases.

An analysis of the current implementation allows for forecasting future directions in technology development. Optimizing system configuration parameters and expanding language support can be achieved without significantly complicating the architecture, preserving the advantages of the lightweight solution.

Discussing the results would be incomplete without considering their influence on related technological fields. The developed approaches to speech signal analysis based on HMM provide a methodological foundation for advancing new directions in natural language processing and speech technologies.

Conclusion

The conducted study of answering machine detection technology demonstrates the effectiveness of lightweight solutions based on HMM in the telecommunications sector. The fundamental feature of the developed approach lies in achieving high recognition accuracy with optimal utilization of computational resources.

The analysis of results leads to three key conclusions. First, the application of speech recognition systems in combination with a specially prepared dataset provides an effective solution to the task of answering machine classification. Second, optimizing the system architecture and utilizing preconfigured parameters establishes a foundation for economically efficient solution scaling. Third, the ability to analyze early media and the high recognition accuracy significantly enhance customer service quality.

The obtained results open up prospects for further research in optimizing system parameters and expanding supported languages while maintaining the advantages of the lightweight architecture. Advancing this direction may result in the emergence of new quality standards in the field of telecommunication technologies.

References:

[1]. Jain S., Chourasia T., Jain S., Joshi A. End to end analytics for call center // 2023 Second International Conference on Informatics (ICI). – IEEE, 2023. – P. 362–366.

[2]. Nornov. Answering Machine Analysis Module (AMD) [Electronic resource]. Dzen, 2020. Available at: https://dzen.ru/a/Xmf1c1nHswNk6z78?sid=161045703595407231 (accessed: 10.11.2024).

[3]. Walker B. Factors That Affect Transcription Accuracy Rates [Electronic resource]. Ditto, 2023. Available at: https://www.dittotranscripts.com/blog/factors-that-affect-transcription-accuracy-rates/ (accessed: 10.11.2024).

[4]. Chou W. Method and apparatus for answering machine detection in automatic dialing: patent 6850602 USA. – 2005. Available at: https://patents.google.com/patent/US6850602B1/en (accessed: 10.11.2024).

[5]. Sperber M., Niehues J., Ha T.-L., Waibel A. Consistent transcription and translation of speech // Transactions of the Association for Computational Linguistics. – 2020. – Vol. 8. – P. 695–709.

[6]. Mikhailova S. A., Anikeev K. G. Application of natural language processing technologies for voice control based on an open dictionary // Intelligent Technologies in Transport. – 2023. – No. 4 (36). – P. 19–27.

[7]. Google Cloud. Speech-to-Text Pricing [Electronic resource]. Available at: https://cloud.google.com/speech-to-text/pricing (accessed: 15.11.2024).

[8]. Yandex Cloud. SpeechKit Pricing Policy [Electronic resource]. Available at: https://yandex.cloud/en/docs/speechkit/pricing (accessed: 15.11.2024).