International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Database modeling for software development

Narzillo Solidjonovich Mamatov [1] and Nurbek Davlataliyevich Nuritdinov [2, *]

[1] *Tashkent institute of irrigation and agricultural mechanization engineers" National research university, Tashkent, Uzbekistan.*
[2] *Tashkent University of Information Technologies named after Muhammad al-Khwarizmi", Tashkent Uzbekistan.*

International Journal of Science and Research Archive, 2026, 19(02), 880-888

Publication history: Received on 02 APril 2026; revised on 13 May 2026; accepted on 15 May 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.19.2.1124

Abstract

This paper presents a comprehensive approach to database modeling for software systems designed to support communication for hearing- and speech-impaired users. The proposed model integrates Speech-to-Text (STT), Neural Machine Translation (NMT), and Text-to-Speech (TTS) components within a unified relational schema based on the IDEF1X methodology. The database structure ensures logical consistency, scalability, and efficient data processing across all functional modules. Each component maintains its core entities while sharing common relational tables for languages, neural models, evaluation metrics, and version control. The STT module models the flow from audio input to segmented transcription and quality evaluation using WER and CER indicators. The NMT module structures translation processes from source text and tokenization to translated output and BLEU-based assessment. The TTS module represents text segmentation, synthesized audio generation, prosodic features, and naturalness evaluation. The integrated model enables systematic monitoring, comparison of model versions, and performance analysis at segment, sentence, and token levels. This unified database architecture supports end-to-end processing (speech–text–translation–speech), improves data integrity, and provides a scalable foundation for further optimization and intelligent system development.

Keywords: Database Modeling; IDEF1X Methodology; Speech-to-Text (STT); Neural Machine Translation (NMT); Text-to-Speech (TTS); Relational Data Model

1. Introduction

Database modeling in software systems represents one of the most critical stages in the overall system development lifecycle, as it directly determines the logical structure, performance efficiency, data integrity, and long-term scalability of the system. A well-designed database model provides a formalized representation of the real-world domain, ensuring that all relevant entities, attributes, and relationships are accurately captured and structured in a way that supports consistent data storage, retrieval, and manipulation. In modern software engineering practices, database design is not only a technical requirement but also a conceptual framework that bridges the gap between real-world processes and their digital representation.

In contemporary intelligent systems, particularly those based on artificial intelligence such as Speech-to-Text (STT) and Text-to-Speech (TTS) applications, database modeling plays an even more significant role. These systems are built upon complex neural network architectures that require large-scale datasets for training, validation, and inference. As a result, databases must efficiently manage diverse types of data, including audio signals, transcribed text, linguistic metadata, acoustic features, user interaction logs, and model outputs. Proper organization of this heterogeneous data is essential to ensure fast processing, high accuracy, and system reliability.

* Corresponding author: Nurbek Nuritdinov Davlataliyevich

Furthermore, database modeling enables clear visualization and management of data flow across different system components. It defines how entities such as users, audio inputs, processed text, and machine learning models interact with each other through well-structured relationships. This structured representation facilitates modular system design, allowing developers to isolate functionalities, reduce redundancy, and improve maintainability. In addition, it supports efficient query optimization and enhances system responsiveness, which is particularly important for real-time applications such as voice assistants and automated transcription services.

From an architectural perspective, a well-constructed database model also contributes significantly to system adaptability and future scalability. As STT and TTS technologies evolve, new data types, algorithms, and processing modules can be integrated with minimal restructuring when a solid database foundation is in place. This ensures that the system remains flexible and capable of accommodating increasing data volumes and more advanced AI-driven functionalities.

In conclusion, database modeling serves as a foundational pillar in the development of modern intelligent software systems. It not only ensures the structural integrity and efficiency of data management but also supports the complex operational requirements of AI-based applications. Therefore, careful and systematic database design is essential for building robust, scalable, and high-performance systems capable of handling advanced speech and language processing tasks [1].

2. Literature review

Various methodologies exist for database modeling, including IDEF0, IDEF1X, UML, and Chen models. IDEF1X is widely used for relational databases, providing a structured representation of entities, attributes, and relationships [2]. It supports integration of STT, neural machine translation (NMT), and TTS modules, enabling end-to-end management, version control, and performance analysis [3]. Other approaches, such as UML, complement IDEF1X for specific modeling tasks [4].

Database modeling in software systems is a crucial stage that determines the system's logical structure, efficiency, and adaptability. The primary objective of database design is to adequately represent the subject domain within the database for automation purposes [5]. The advancement of technical system components, widespread internet adoption, digital economy, e-commerce, and mobile communication systems have increased the demand for modeling complex, high-architecture databases [6].

Today, databases constitute a fundamental component of information systems and software products. They ensure the long-term integrity of data and simplify information storage. Databases are software systems designed to store large volumes of interrelated data and provide efficient access when needed. When creating a database, careful consideration must be given to which tables to include and how the data in these tables relate to each other.

In software development, particularly when designing user interfaces for speech-impaired users, database modeling is one of the most essential stages. This process not only defines the logical structure of the system but also determines its efficiency, data processing speed, and adaptability to future functional expansions. Interaction systems for speech-impaired individuals, specifically the STT (Speech-to-Text) system, convert human speech into text in real-time. This system operates based on neural networks, extracting acoustic, phonetic, and semantic features from speech signals. Through this model, the STT system's data flow can be clearly and structurally represented. By defining relationships between entities, it becomes possible to track how data is exchanged at each stage.

The proposed IDEF0 and IDEF1X models encompass the main objects of the system. The models illustrate the primary entities and their interrelationships within the STT component (Figures 1 and 2). Similarly, the TTS (Text-to-Speech) system converts user-input text into audio. IDEF0 and IDEF1X models help represent the data and their interconnections throughout these processes. Each module contains attributes that allow for effective software modeling of the system. The TTS module is modeled based on the following data structures (Figures 3 and 4).

Additionally, the system supports multilingual translation of Uzbek text into Russian and English using a Neural Machine Translation (NMT) system, which automatically translates text from one language to another. IDEF0 and IDEF1X models provide graphical representation of entities, attributes, and their interrelationships at this stage, allowing for monitoring, optimization, and analysis of the translation process(Figures 5 and 6).

Database structures in speech-based interaction systems require a comprehensive approach to capture the subject domain accurately. Modern technical advancements, mobile platforms, and integration with internet systems

necessitate more complex and functional database models. Key data types in the system include user information, speech recordings in various conditions, text-audio alignment data, and multimedia files uploaded by users. Relationships between these data must be clearly defined and effectively managed through specialized database models.

Several methodological approaches exist for database modeling. Among them, the IDEF (Integration Definition) methodology including IDEF0, IDEF1X, IDEF3, and IDEF5 is widely applied to model software system architecture, processes, relationships, and semantic data. For database modeling, IDEF1X is considered the most optimal and practical. It is designed for relational databases and provides an object-relational information model. Using IDEF1X, each user speech entry—captured via microphones or other recording devices, including corresponding textual data and recordings in various environments—can be stored systematically. This data serves as a foundation for voice-based identification and speech-to-text and text-to-speech algorithms.

Other methodologies, such as UML (Unified Modeling Language), Chen, and MIRIS, are also widely used in database modeling. While they share commonalities with IDEF, each offers symbols and modeling styles tailored to specific stages and tasks. Notably, the IDEF1X methodology excels at maintaining data integrity, defining the degree of relationships between data, and visual representation [7]. Implementing IDEF1X in software systems designed for speech-impaired user interaction allows comprehensive and structured management of user speech data. Consequently, this facilitates a high-quality and precise organization of the entire interaction process.

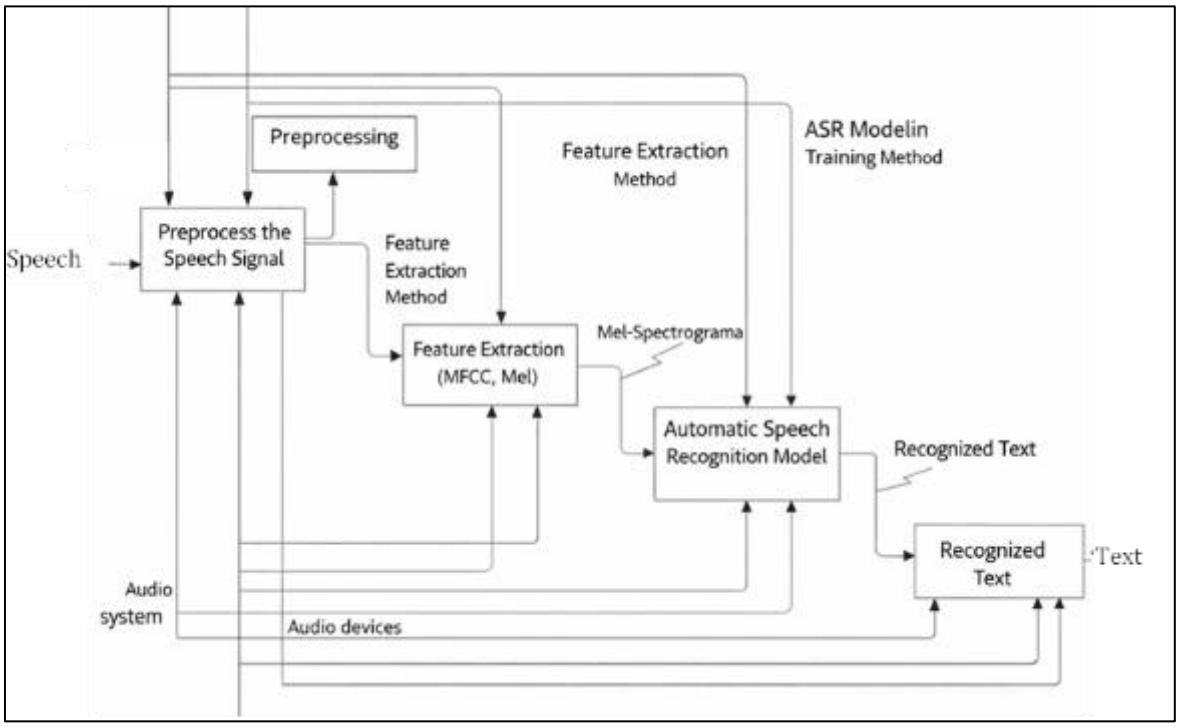

Figure 1 IDEF0 Diagram of the Speech-to-Text System

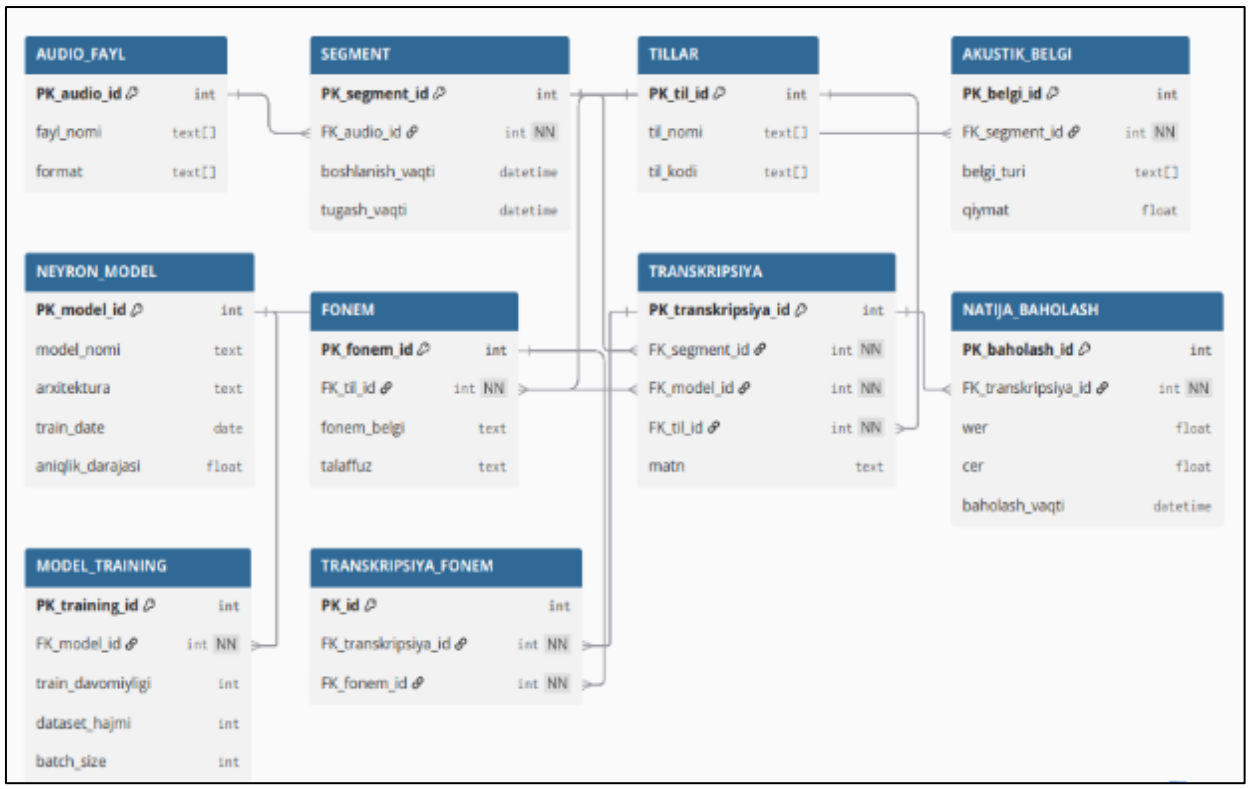

Figure 2 IDEF1X Model for the Neural STT System

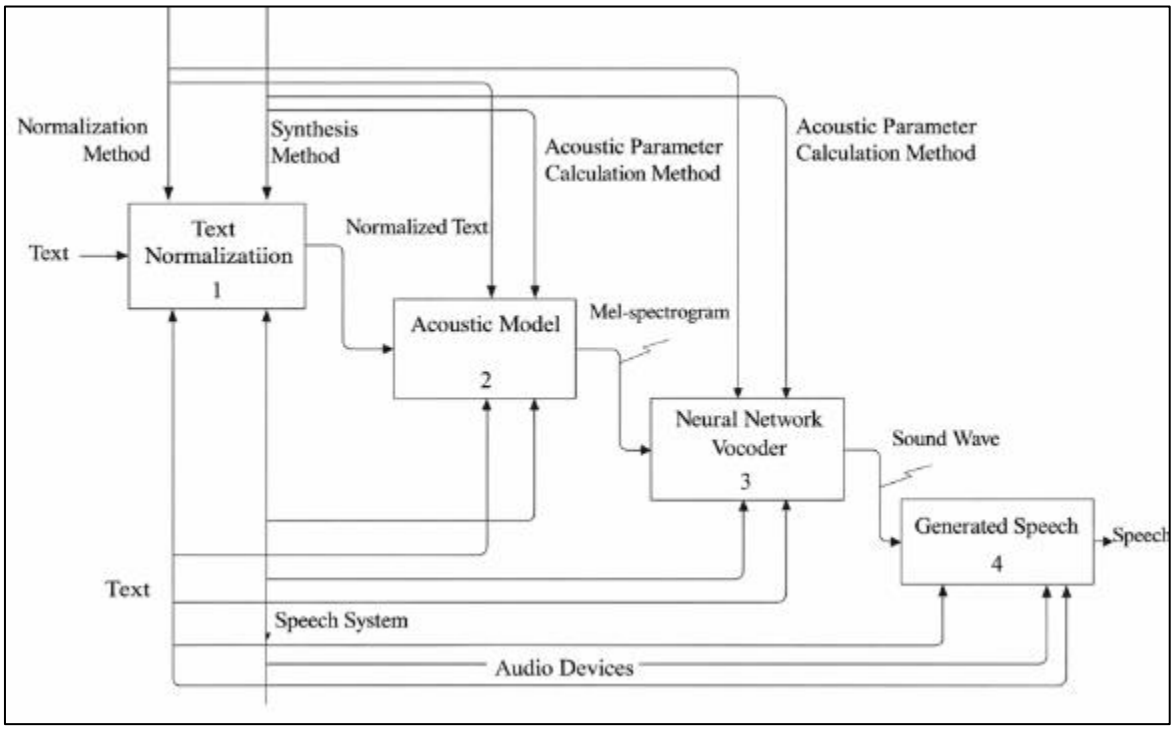

Figure 3 IDEF0 Diagram for the Neural TTS System

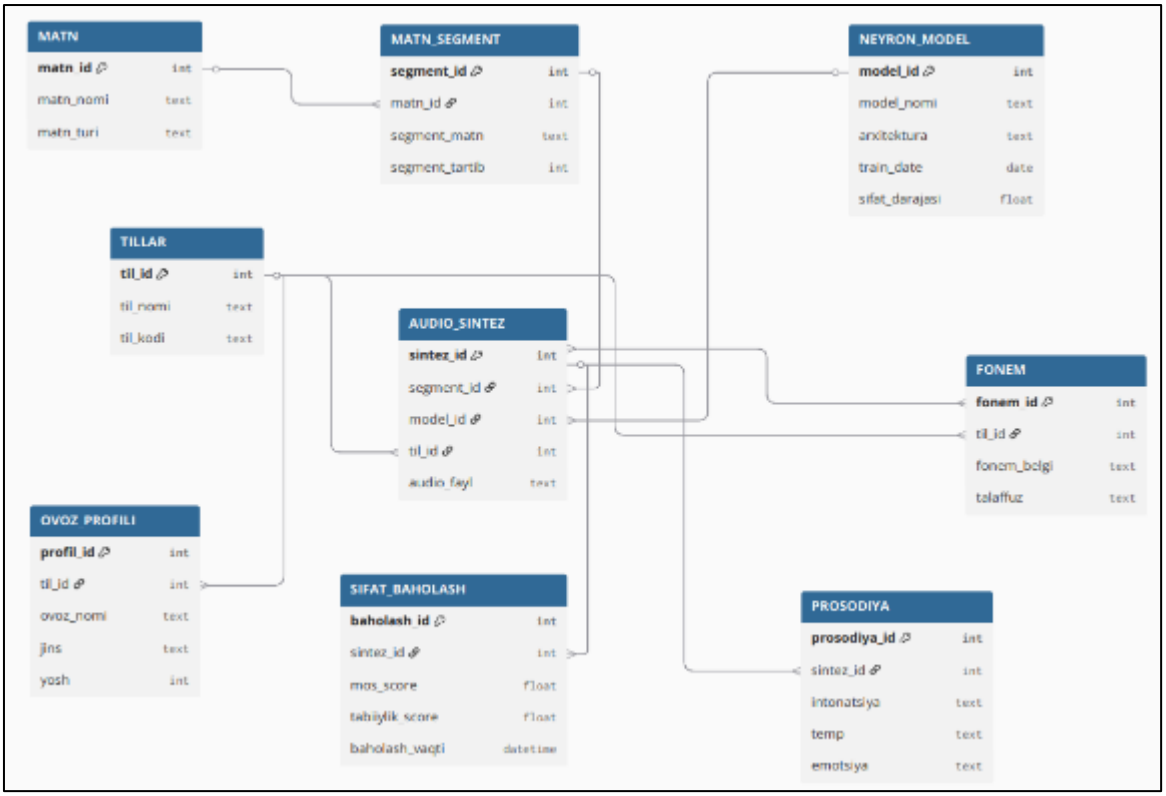

Figure 4 IDEF1X Model for the Neural TTS System

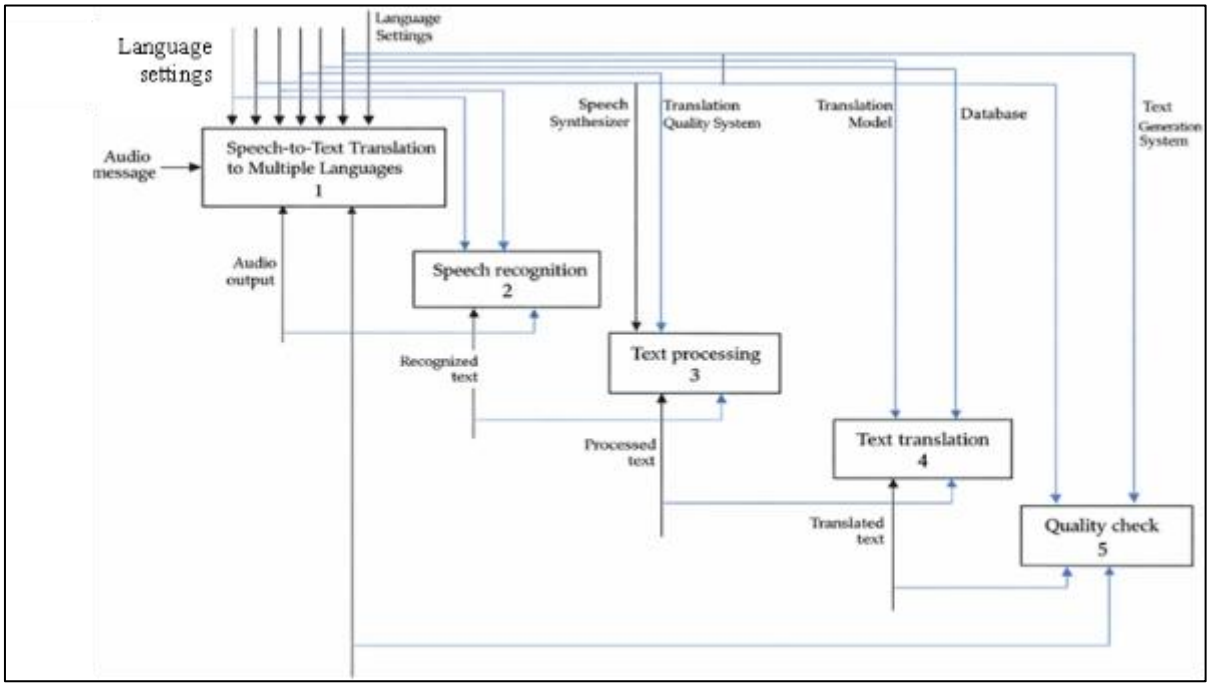

Figure 5 IDEF0 Diagram for the Neural Machine Translation System

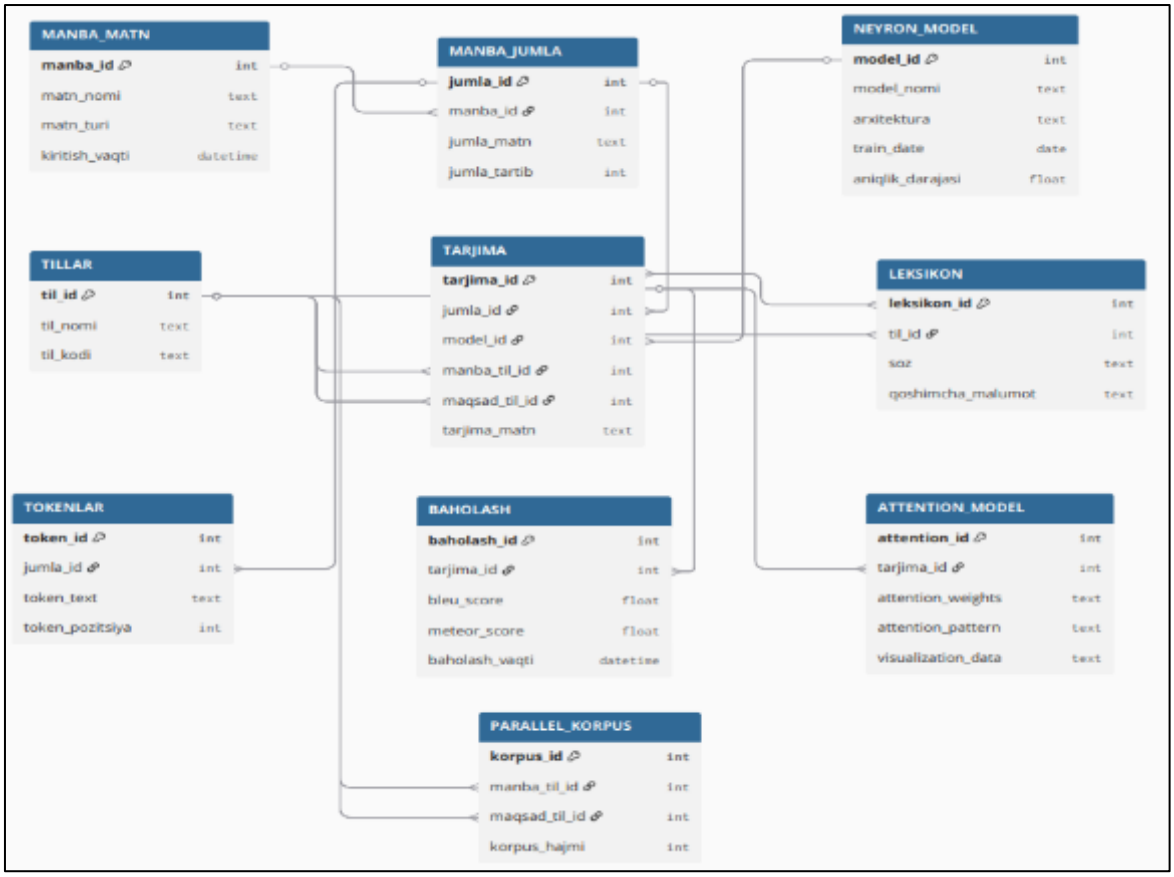

Figure 6 IDEF1X Model of the Neural Machine Translation (NMT) System

2.1. Data flow and table relationships between system platform modules according to the IDEF1X model

The created IDEF1X generalized data model of the STT–NMT–TTS system (speech → text → translation → speech) allows managing the entire workflow within a single relational schema, versioning and verifying the results at each stage, and evaluating quality. The conceptual basis of the model is that each functional module (STT, NMT, TTS) maintains its own "core objects," but they are interconnected through shared tables: Tillar (til_id, til_nomi, til_kodi) supports multilingual functionality, while Neyron_model (model_id, model_nomi, arxitektura, train_date, aniqlik_darajasi, sifat_darajasi) specifies which architecture/version produced each result. This approach enables reliable comparison of experimental results since outputs from different model versions on the same data can be compared within a single database using strict identifiers.

In the STT (Speech-to-Text) module, the data flow is structured as "audio → segment → feature → transcription → evaluation." Specifically, audio_fayl (audio_id, fayl_nomi, format, …) represents the recorded speech input, which is segmented over time through segment (segment_id, audio_id, boshlanish_vaqti, tugash_vaqti); acoustic features such as log-mel, MFCC, pitch, etc., are linked to each segment via akustik_belgi (belgi_id, segment_id, belgi_turi, qiymat). The transcription text is stored in transkripsiya (transkripsiya_id, segment_id, model_id, til_id, matn), with foreign keys ensuring the corresponding language and neural model are recorded. Result quality is tracked via natija_baholash (baholash_id, transkripsiya_id, wer, cer, baholash_vaqti), allowing statistical analysis of STT performance by time, model version, and language. Additionally, model_training (training_id, model_id, train_davomiyligi, dataset_hajmi, batch_size) records training conditions, linking results with training configurations. For phoneme-level analysis, fonem (fonem_id, til_id, fonem_belgi) and transkripsiya_fonem (id, transkripsiya_id, fonem_id) tables enable diagnostic evaluation of STT/lexical errors by language-phonetics.

In the NMT (Neural Machine Translation) module, the data model normalizes the translation workflow as "source text → sentences → tokens → translation → evaluation." Manba_matn (manba_id, matn_nomi, matn_turi, kiritish_vaqti) stores source corpus/document metadata; sentences are split in manba_jumla (jumla_id, manba_id, jumla_matn, jumla_tartib), and tokens are recorded in tokenlar (token_id, jumla_id, token_text, token_pozitsiya). Translation results are stored in tarjima (tarjima_id, jumla_id, model_id, manba_til_id, maqsad_til_id, tarjima_matn), with model version and language pair specified. Translation quality is evaluated using baholash (baholash_id, tarjima_id, bleu_score,

meteor_score, baholash_vaqti) metrics such as BLEU and METEOR. For terminology support and analysis, leksikon (leksikon_id, til_id, soz, qoshimcha_ma'lumot) and its mapping to translations via tarjima_leksikon (id, tarjima_id, leksikon_id) allows consistent translation monitoring, especially in technical/scientific texts. To analyze attention mechanisms, attention_model (attention_id, tarjima_id, attention_weights, attention_pattern, visualization_data) is included, along with parallel_korpus (korpus_id, manba_til_id, maqsad_til_id, korpus_hajmi) for parallel corpus metadata, enhancing NMT explainability and corpus management.

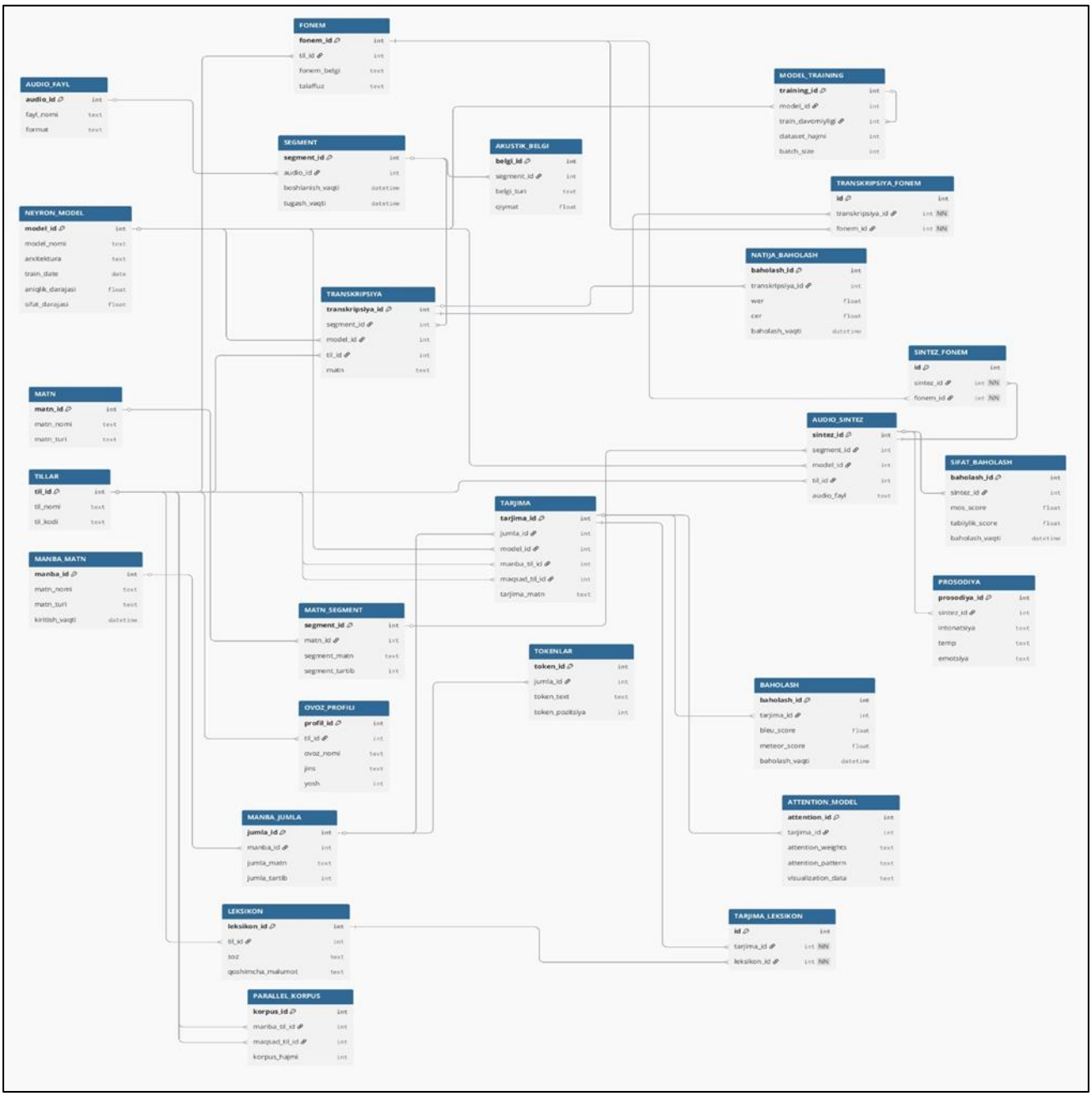

Figure 7 Generalized IDEF1X Model of the STT–NMT–TTS System

In the TTS (Text-to-Speech) module, the data flow is structured as "text → text segments → synthesized audio → prosody/quality." Matn (matn_id, matn_nomi, matn_turi) stores the source text, while matn_segment (segment_id, matn_id, segment_matn, segment_tartib) stores segments suitable for synthesis. Synthesized audio is recorded in audio_sintez (sintez_id, segment_id, model_id, til_id, audio_fayl), specifying the language and neural model used. Voice personalization is managed through ovoz_profili (profil_id, til_id, ovoz_nomi, jins, yosh), and prosodic attributes are stored in prosodiya (prosodiya_id, sintez_id, intonatsiya, temp, emotsiya). TTS quality is evaluated in sifat_baholash (baholash_id, sintez_id, mos_score, tabiiylik_score, baholash_vaqti), allowing analysis across model

version/language/voice profile for optimal configuration. Phoneme-level synthesis control is facilitated through sintez_fonem (id, sintez_id, fonem_id), enabling research into phoneme-prosody-quality relationships.

The scientific and practical value of this IDEF1X generalized model lies in its ability to treat STT–NMT–TTS not as separate "modules," but as a single end-to-end platform: each module records input data, output results, model/language metadata, and quality metrics comprehensively. This enables versioning and comparison of models, auditing data at segment/sentence/token levels, localizing errors (e.g., whether WER increase originates from STT or BLEU decrease from NMT), real-time monitoring, and identifying directions for future improvements. This model also provides a practical foundation for integration between "STT transcriptions → NMT source sentences" and "NMT translation outputs → TTS text" workflows (Figure 7).

3. Results and discussion

Database modeling is a critical phase in software systems development, determining the system's logical structure, efficiency, and adaptability. In this study, a comprehensive database model for STT–NMT–TTS systems was designed using the IDEF methodology (IDEF0 and IDEF1X). The STT (Speech-to-Text) system converts user speech into text in real-time, the NMT (Neural Machine Translation) system translates text automatically, and the TTS (Text-to-Speech) system synthesizes speech from text. The IDEF1X model visually represents the data flow, entities, attributes, and their interrelationships for each module, enabling effective system management and optimization.

For continuous Uzbek speech recognition, different statistical and neural network models were tested. RNN (Recurrent Neural Network) models were trained using the RNNLM (Recurrent Neural Network Language Modeling Toolkit), with output layers split into factors and class probabilities calculated based on word frequencies. Models were created with various numbers of hidden layer units (200, 500, 800) and class counts (200 and 800) [8].

Model performance was evaluated using perplexity coefficients. The results indicate that increasing hidden layer units and class numbers reduces uncertainty. Furthermore, a test corpus of continuous speech was created using over 560 hours of audio from 60 native Uzbek speakers. The system's performance was measured using the Word Error Rate (WER), achieving WER = 27.8% with a trigram baseline model. Subsequent reevaluation using factorial and neural network models significantly improved recognition accuracy.

Key findings include:

- 200-class RNN models outperformed 500-class models.
- RNN models with 800 hidden units achieved the best recognition accuracy (WER = 23.7%) with trigram model interpolation (interpolation coefficient = 0.6).
- Factorial and RNN model interpolation further reduced WER: 10% reduction with factorial models and 18.5% reduction with RNN models.

The IDEF1X-based database model enabled monitoring of data flows, versioning of outputs, and quality assessment across all modules. In the STT module, audio files are segmented, linked to acoustic features, and stored with transcription text. In the NMT module, source text is split into sentences and tokens, with translated text and metadata recorded. In the TTS module, text is divided into segments, and audio is synthesized with prosody and quality attributes.

In conclusion, the integrated database model developed for continuous Uzbek speech recognition enhances system efficiency and provides a scalable foundation for future functionality, cross-platform integration, and user interaction. RNN-based neural network models outperformed other statistical models, effectively preserving context and reducing WER, which significantly improves overall system performance.

4. Conclusion

Database modeling is one of the most important stages in ensuring the efficiency, logical structure, and flexibility of software systems. Using the IDEF0 and IDEF1X methodologies enables the integration of STT, NMT, and TTS modules and provides the ability to accurately track the flow of information. A properly designed database serves to optimize the processes of automatic user speech processing, real-time conversion to text, translation, and generation of voice responses. Furthermore, this approach creates a solid foundation for the future expansion of the system, integration with new interactive technologies, and improving overall efficiency. Thus, comprehensive and systematic database modeling guarantees the reliable and high-quality performance of software.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Aliyev, D., & Tursunov, R. (2024). Modern Database Architectures for Interactive Systems. *Journal of Information Technology*, 12(3), 45-60.

[2] Smith, J., & Brown, L. (2022). IDEF Methodologies in Software Engineering. *International Journal of System Modeling*, 9(1), 15-32.

[3] Solidjonovich, M. N., & Davlataliyevich, N. N. (2025). Neyron STT, TTS va mashinali tarjima tizimlari o'rtasida uzviy bog 'liqlikni idef1x modeli orqali modellashtirish. Al-Farg'oniy avlodlari , 1 (2), 29-33.

[4] Johnson, P. (2021). Database Modeling Techniques for Neural Networks. *Computing Reviews*, 35(4), 78-95.

[5] Chen, X., & Wang, Y. (2020). Relational Database Design Using IDEF1X. *Software Engineering Review*, 18(2), 50-66.

[6] Miller, R., & Davis, K. (2022). End-to-End Platforms for Assistive Technology. *International Journal of AI Applications*, 14(3), 88-102.

[7] N.S. Mamatov, Kh.T. Dusanov. Methods for Modeling Production Tasks Using IDEF and UML. Scientific and Practical Journal of Fire and Explosion Safety. Tashkent, 2023, pp. 338–345.

[8] Narzillo Mamatov Solidjonovich, Nurbek Nuritdinov Davlataliyevich and Muxiyatdinov Jamalatdin Kayratdin ulı, Neural network-based modeling for continuous speech recognition in the Uzbek Language, International Journal of Science and Research Archive, 2025, 16(01), 1539-1545, Article DOI: 10.30574/ijsra.2025.16.1.2141,DOI url: https://doi.org/10.30574/ijsra.2025.16.1.2141