BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2025
http://excel.fit.vutbr.cz

Basic school math drill platform with entering answers by voice

Dominik Horut*

Abstract

Learning mathematics can be compared to building a house — it is crucial to establish a solid and high-quality foundation upon which to build further knowledge. Achieving this requires consistent practice. Mathematics practice can be carried out in various ways, and one of the most efficient methods is through drills — solving as many examples as possible in a short period of time. The aim of this work is to develop a web application that facilitates mathematics practice using this method. To make the process even faster and more convenient, the application allows users to input answers via voice. After testing the web application with elementary school students, a dataset was created to compare the most popular automatic speech recognition solutions and to determine which one is best suited for this use case.

*xhorut01@vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Mathematics, often referred to as the queen of the sciences, is undoubtedly one of the most useful subjects that students learn in elementary school. However, if a student struggles with mathematics, regular practice is essential; without it, they may encounter difficulties with more complex topics.

Today, almost every student has access to a phone or computer and can practice on a variety of platforms designed to support mathematics learning. However, very few platforms – and none currently available in Czech – support voice input of answers, which could significantly speed up and simplify the entry of more complex responses, such as fractions.

As mentioned before, there are not many platforms for practicing mathematics that offer voice input, and most that do are available only as mobile applications. The only known web application with a similar focus is VoiceMath[1], which supports voice input exclusively in English and provides only randomly generated examples for basic arithmetic.

Therefore, the goal of this work is to develop a web application for practicing mathematics through drills. Its content is based on the Czech mathematics curriculum, allowing elementary school students to easily navigate to the topics they wish to practice. The application also enables users to enter answers by voice in Czech.

[1]VoiceMath – https://voicemath.vercel.app/

2. Implemented Solution

The application is built on a client-server architecture, using the Vue.js framework on the client side and the Django framework on the server side.

The main and most important part of the application is shown in Figure 1. The user interface was designed to be as simple and user-friendly as possible, allowing even the youngest users to navigate and use it with ease.

This part of the application is responsible for sending user's answers to the server, which then evaluates them. If the user inputs the answer into the provided fields, the evaluation is straightforward. However, the process of evaluating voice input is more complex and will be described in the following section.

3. Evaluating answers by voice

The evaluation of voice answers sent by the client via the WebSocket protocol consists of two stages. First, a transcript of the user's speech is generated using

the Azure Speech-to-Text (STT) service. The transcription is completed once the user stops speaking.

The next step depends on the type of answer provided. If it is a simple numerical answer (for example, 12, 62, etc.), which Azure STT can transcribe directly in numeric format, a Python script compares the transcribed answer with the correct one.

For more complex answer formats, such as fractions or variable values, Azure STT may not consistently produce the desired numeric transcription. In these cases, the evaluation is handled by the Gemini large language model, which enables more accurate and reliable evaluation.

The entire evaluation pipeline is illustrated in Figure 2.

4. Testing

After the application was implemented and deployed, it was tested by the target users – elementary school students. They were asked to practice topics they were currently studying in their mathematics classes and to try entering their answers by voice. During the testing, users were asked questions about the user experience (UX) to assess whether the application met their needs, and their voice responses were also recorded.

In total, approximately 800 recordings were collected and manually annotated. The recordings were then categorized into four groups:

- clean recordings with no faults
- recordings with background noise
- recordings containing speaker errors (e.g., poor pronunciation, mumbling)
- recordings in English

The most popular automatic speech recognition (ASR) systems were subsequently selected based on [1] and evaluated on this dataset using the Word Error Rate (WER) metric. The results are shown in Figure 3. For clean recordings, Azure achieved the best performance. Across all four categories, Deepgram and its Nova-2 model performed the best overall.

5. Results

A web-based math drill application was developed that allows users to enter answers by voice. The application was tested by the target users, resulting in a collection of audio recordings of their inputs. These recordings formed the basis of a dataset that was subsequently used to evaluate popular automatic speech recognition (ASR) systems.

6. Acknowledgements

I would like to thank my supervisor, Ing. Igor Szőke, Ph.D., for his valuable guidance and support throughout the development of this work. I would also like to thank everyone who participated in the testing for their time and for providing valuable data.

References

[1] Voice Writer. The best speech recognition api in 2025: A head-to-head comparison. blogpost, Feb 2025. https://voicewriter.io/blog/best-speech-recognition-api-2025.