

Modeling difficulty of listening comprehension items with machine learning

Thursday 24 July 2025 17:10 (20 minutes)

Abstract

Understanding the difficulty of items in educational assessment is crucial for test development, teaching, and learning. While prior research used text features and machine learning to model the difficulty of reading comprehension items (Štěpánek et al., 2023), listening comprehension (LC) presents additional complexity. Beyond textual factors, LC item difficulty can be influenced by several additional factors, including variations in speech rate, voice characteristics, and others.

In this study, we analyze data from English, German, and French listening sections of the Czech Matura exams. We extract text features commonly used in reading comprehension difficulty modeling, including readability indices and similarity measures. Additionally, we incorporate audio-specific features, such as speech rate, confidence scores from automatic speech recognition (ASR) models, and voice distinguishability derived from a speaker diarization model. This feature set allows us to capture the linguistic and auditory effects on LC item difficulty.

Using these features, we train a regression model to predict the final LC difficulty as described by student response data. The optimal model is selected based on accuracy measures of different models and model parameters, and compared to predictions of content experts to assess alignment between computational modeling and human evaluation.

The model is implemented into an interactive application to aid exam development. This tool aims to provide educators and test developers with insights into LC item difficulty, allowing for more informed test development.

Keywords

Item Difficulty Modeling, NLP, ML

Primary author: Mr MARTINEK, Filip (Institute of Computer Science of the Czech Academy of Sciences)

Co-authors: Mr NETÍK, Jan (Institute of Computer Science of the Czech Academy of Sciences; Faculty of Education, Charles University); MARTINKOVÁ, Patrícia (Institute of Computer Science of the Czech Academy of Sciences; Faculty of Education, Charles University)

Session Classification: Poster Session 4

Track Classification: Statistical analyses: Statistical analyses