

Public Speaking Web Trainer

Daniil Pliushchenko, Higher School of Economics
Mark Zaslavskiy, Saint Petersburg Electrotechnical University “LETI”

Introduction

A large number of students and researchers give public talks with their presentations.

A thesis defense, a university project presentation, or a speech at a conference.

Usually, they train before the actual talk (repeating their speech to themselves, in front of a mirror, or to another person).

In order to evaluate the quality of a speech and a presentation, there must be at least one other person having a high level of expertise in a particular domain.

However, it is not always possible to do so because of lack of experts' time.

Introduction

However, some properties related to the quality of a talk can be checked automatically.

For instance:

- Speech duration and pace
- Filler words / phrases usage
- Some criteria related to speech, slides, and speaker's behaviour are described in the literature ^{1, 2}

¹ Kopf, S., Schön, D., Guthier, B., Rietsche, R. & Effelsberg, W. (2015). A Real-time Feedback System for Presentation Skills.

² Hanani, A., Al-Amleh, M., Bazbus, W. & Salameh, S. (2017). Automatic Estimation of Presentation Skills Using Speech, Slides and Gestures.

Purpose and tasks

The purpose of this work is to design and develop an interoperable, scalable in terms of potential load and possible steps of evaluation, open-source web-based application that performs the preliminary evaluation of a given speech with a presentation and gives recommendations for improvement.

Tasks:

- Review analogues
- Review speech recognition libraries
- Develop flexible architecture that allows you to add and change components related to public speaking analysis and data storage
- Develop a web application according to the proposed architecture

Analoguees

Applications that evaluate presentation, speaking and / or pronunciation skills and return quantitative and measurable feedback were considered.

Overall conclusions:

- Usually, the only supported language is English
- Impossible to attach a presentation
- Impossible to adjust criteria
- Mobile applications are platform-dependent
- Mostly, no integration with external tools such as LTI / API

Existing applications

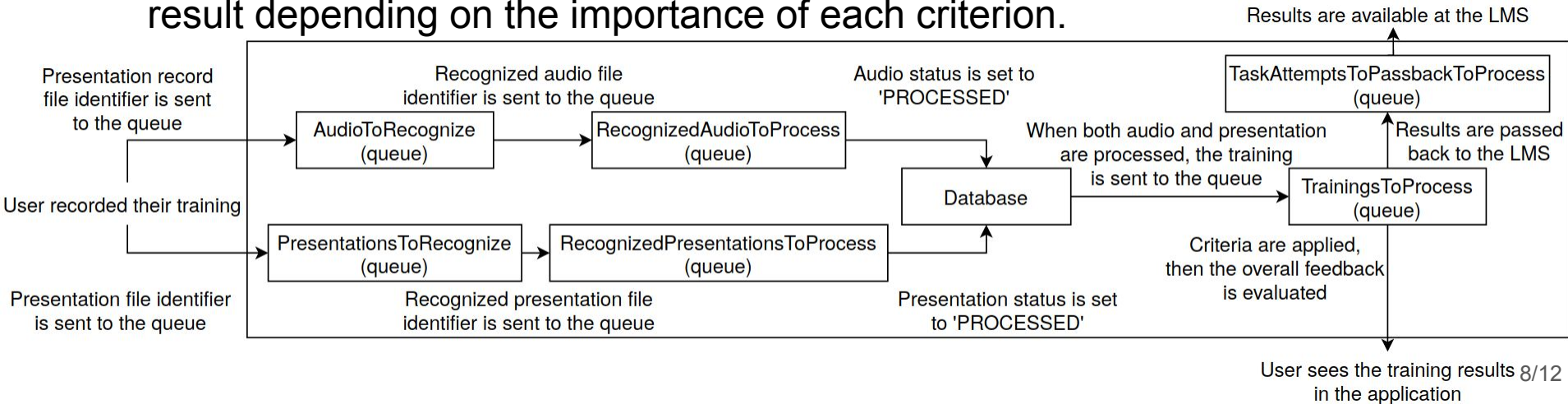
Name	Platforms	API?	Presentation attachment	Speech duration limits	Speech recording	Languages	Custom criteria?	Paid?
Speechace	Web	Yes	No	15 seconds for free	Yes	English	No	Free demo
Voice Notebook	Web	No	No	No	Yes	8	No	No
Speakit	Android	No	No	One phrase	Fixed words and phrases	American English	No	No
Aksent	iOS	No	No	One phrase	Yes	20+	No	No
Speeko	iOS	No	Text notes for the current record	No	Yes	English	No	Free trial
LikeSo	iOS	No	No	30 minutes	Yes	English	Subset of fillers	Yes
Orai	iOS, Android	No	Text notes	No	Yes	English	Subset of fillers	Yes
ELSASpeak	iOS, Android	No	No	One phrase	Fixed words and phrases	American English	No	Free demo
Говорилло (Govorillo)	Android	No	No	No	Yes	Russian	No	Yes

Speech recognition libraries

Name	Offline?	Russian language?	Timestamps?	Price
Vosk	Yes	Yes	Yes	Free
Speech-to-Text (Russian)	Yes	Yes	No	Free
Picovoice	Yes	No	No	Free
At16k	Yes	No	Yes	Free for personal use
Google Web Speech API	No	Yes	Yes	Free one hour per month
Google Cloud Speech API	No	Yes	Yes	Free one hour per month
Microsoft Speech Service	No	Yes	Yes	Free 5000 requests per month
IBM Speech to Text	No	No	Yes	Free 50 hours per month
Wit.ai	No	Yes	Yes	Free but query length limit is 20 seconds, no more than 60 requests per minute

Training processing

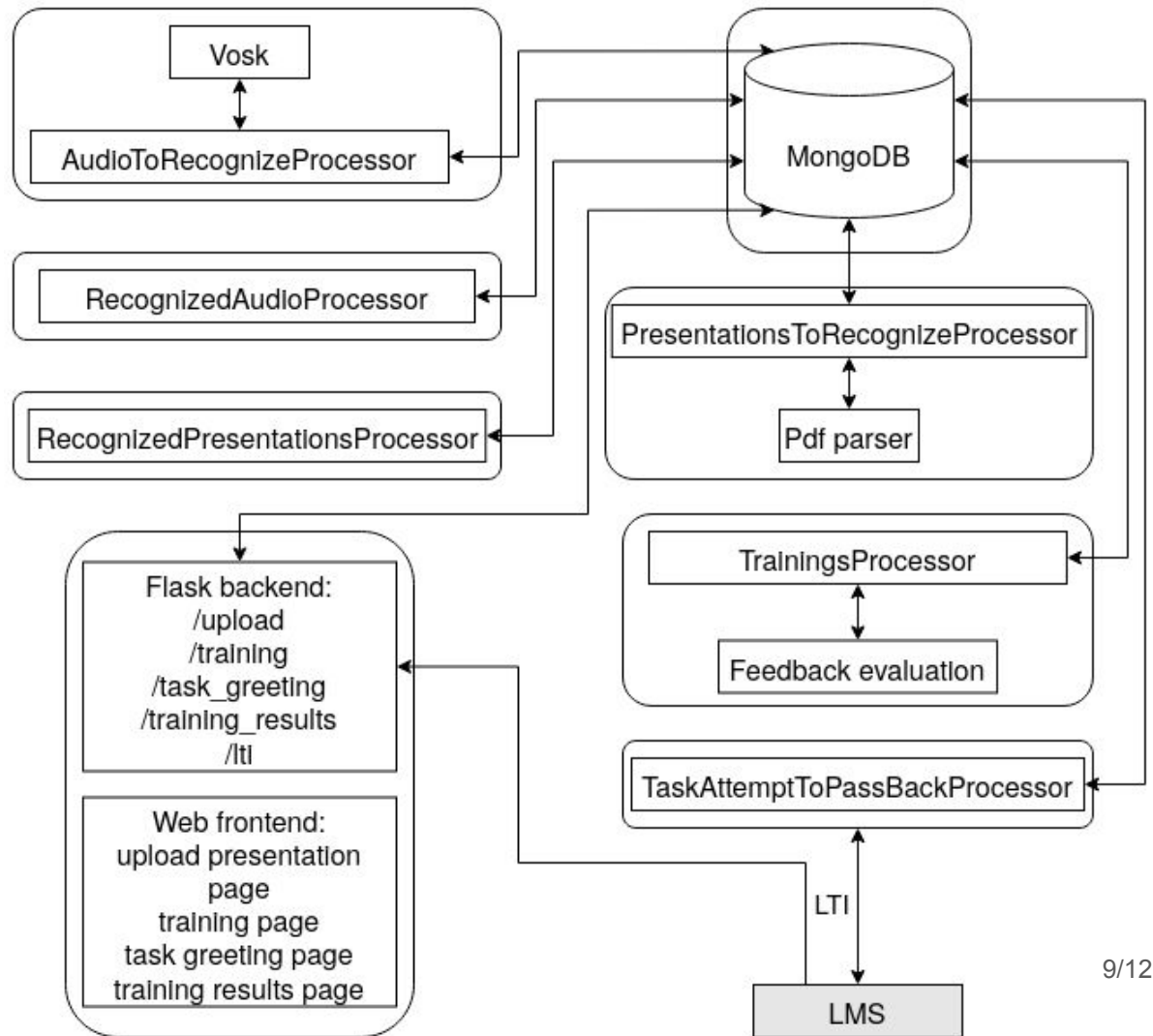
- An example of a criterion: the duration of the speech does not exceed X.
- An example of a parameterized criterion: the duration of the report does not exceed 7 minutes.
- We combine parameterized criteria into sets with information about the dependence of the criteria from each other, calculate in the desired order (topological sorting or just a given order).
- We use an arbitrary evaluation function. The function determines the final result depending on the importance of each criterion.



Architecture

Services: web, database, queue-based services.

API and LTI support are implemented within the backend part of the 'web' service.



Web Interface

Record Next Done Page: 4 / 12
Training time: 0:01:44

Purpose and tasks

The purpose of this work is to design and develop an interoperable, scalable in terms of potential load and possible steps of evaluation, open-source web-based application that performs the preliminary evaluation of a given speech with a presentation and gives recommendations for improvement.

Tasks:

- Review analogues
- Review speech recognition libraries
- Develop flexible architecture that allows you to add and change components related to public speaking analysis and data storage
- Develop a web application according to the proposed architecture

4/13

ten_minutes_training
fifteen_minutes_training
twenty_minutes_training



Score = 0.60 = 0.600 * 0.33 + 0.200 * 1.00 + 0.200 * 1.00

SpeechDurationCriterion = 0.33
SpeechPaceCriterion = 1.00
FillersNumberCriterion = 1.00

Training id: 608bcc835f7c5a592090dcce. Status: processed

Presentation id: 608bcc835f7c5a592090dcc6. Status: processed

Presentation name: FRUCT Public Speaking Web Trainer.pdf

Presentation record id: 608bcce95f7c5a592090dcdd. Status: processed

Training number: 2 / 3

Total score for previous trainings: 0.40

Next training



Evaluation

So far the longest step is audio recognition, presentation parsing takes a couple of seconds, and criteria evaluation takes less than a second, hence the overall training processing time is almost equal to the audio processing time.

12 copies of the same training that lasts 7:39

Number of speech recognition instances	Total processing time, mm:ss	Time processed for 1 second, seconds
1	25:31	3.60
2	17:47	5.16
3	14:31	6.32

Results

- Speech recognition libraries and analogues were reviewed
- Application architecture was proposed
- Web application prototype was developed ¹

Further work:

- Add more criteria, especially related to content
- Conduct usability studies and collect user feedback and analyze it in order to study the efficiency of the application.

¹Repository: https://github.com/OSLL/web_speech_trainer