Speech technologies applied to second language learning. A use case on Bulgarian.

Yaneva, Alexandrina
Curs 2020-2021

Director: MÓNICA DOMÍNGUEZ
GRAU EN ENGINYERIA INFORMÀTICA

Treball de Fi de Grau

Speech technologies applied to second language learning.
A use case on Bulgarian.

TREBALL FI DE GRAU DE

Alexandrina Yaneva

Director: Mónica Domínguez

Grau en Enginyeria en Informàtica

Curs 2020-2021

Universitat **Pompeu Fabra** *Barcelona*   Escola d'Enginyeria

To my mother and grandfather.

Acknowledgments

I would like to thank my family for the endless support, love, and understanding during the whole process of getting this university degree. For always believing in me and for all the patience during the times of a pandemic.

To all my friends, who are always next to me, who listen and support me when times are tough.

I am extremely grateful to Dr. Mónica Domínguez as my thesis supervisor. Her passion for languages and her positive approach in guiding me through this project have motivated me and helped me stay determined to strive to do my best.

To my new coworkers for the flexibility and all the knowledge that has helped me in this project.

Abstract

The Bulgarian language has specific phonetics as every other language. Some specifications make learning the pronunciation of Bulgarian more challenging for second language learners. Together with all technological advances, speech technologies have progressed notably in recent years. Systems such as Text-to-Speech and Automatic-Speech-Recognition belong to the subfields of speech technology - speech synthesis and speech recognition. They have various applications in language learning, and many studies have proven the positive benefits of their implementation. The Bulgarian language is not as rich in technological resources as other languages. With the technology available, I conducted a couple of experiments with native and non-native speakers of the language, which aimed to test how it can be used as a tool for the improvement of pronunciation of second language learners of Bulgarian. Then I designed a simple demo, which demonstrates an example of how they could be implemented.

Резюме

Българският език има специфична фонетика като всеки друг език. Някои спецификации правят изучаването на произношението на български език по-голямо предизвикателство за изучаващите го като чужд език. Заедно с всички технологични постижения, речевите технологии са напреднали значително през последните години. Системи като преобразуване на текст в реч и автоматично разпознаване на реч принадлежат към подполета на речевата технология - синтез на реч и разпознаване на речта. Те имат различни приложения в изучаването на езици и много изследвания са доказали положителните ползи от тяхното прилагане. Българският език не е толкова богат на технологични ресурси, колкото други езици. С наличните технологии проведох няколко експеримента с родни и чужди носители на езика, които имаха за цел да тестват как той може да се използва като инструмент за подобряване на произношението на изучаващите български като чужд език. След това програмирах семпла демонстрация, която да служи за пример за това как те могат да бъдат приложени.

Resumen

El idioma búlgaro tiene una fonética específica como cualquier otro idioma. Algunas especificaciones hacen que el aprendizaje de la pronunciación del búlgaro sea más desafiante para los estudiantes de un segundo idioma. Junto con todos los avances tecnológicos, las tecnologías del habla han progresado notablemente en los últimos años. Los sistemas como Text-to-Speech y Automatic-Speech-Recognition pertenecen a los subcampos de la tecnología de voz: síntesis de voz y reconocimiento de voz. Tienen varias aplicaciones en el aprendizaje de idiomas y muchos estudios han demostrado los beneficios positivos de su implementación. El idioma búlgaro no es tan rico en recursos tecnológicos como otros idiomas. Con la tecnología disponible, realicé un par de experimentos con hablantes nativos y no nativos del idioma, con el objetivo de probar cómo se puede utilizar como herramienta para mejorar la pronunciación de los estudiantes de segundo idioma de búlgaro. Luego diseñé una demostración simple que demuestra un ejemplo de cómo podrían implementarse.

Index

List of figures

List of tables

1. Introduction

1.1 Motivations

The Bulgarian language is neither among the most spoken languages nor are there numerous non-native speakers who are learning it as a second language. Consequently, the technological resources for Bulgarian are not as rich as those of other languages since the demand is not as high. Therefore, technologies and applications specifically aiming to aid second language learners to improve their pronunciation in Bulgarian are limited.

From my personal perspective, as a native speaker of Bulgarian, from a young age, we are taught that we should learn as many foreign languages as possible because knowing only Bulgarian 'won't take us anywhere in life'. Undeniably, mastering languages is a fortune, but with the decrease of the Bulgarian population in the past decades (fig.1), the number of native speakers, who are the majority of Bulgarian speakers, is also diminishing.
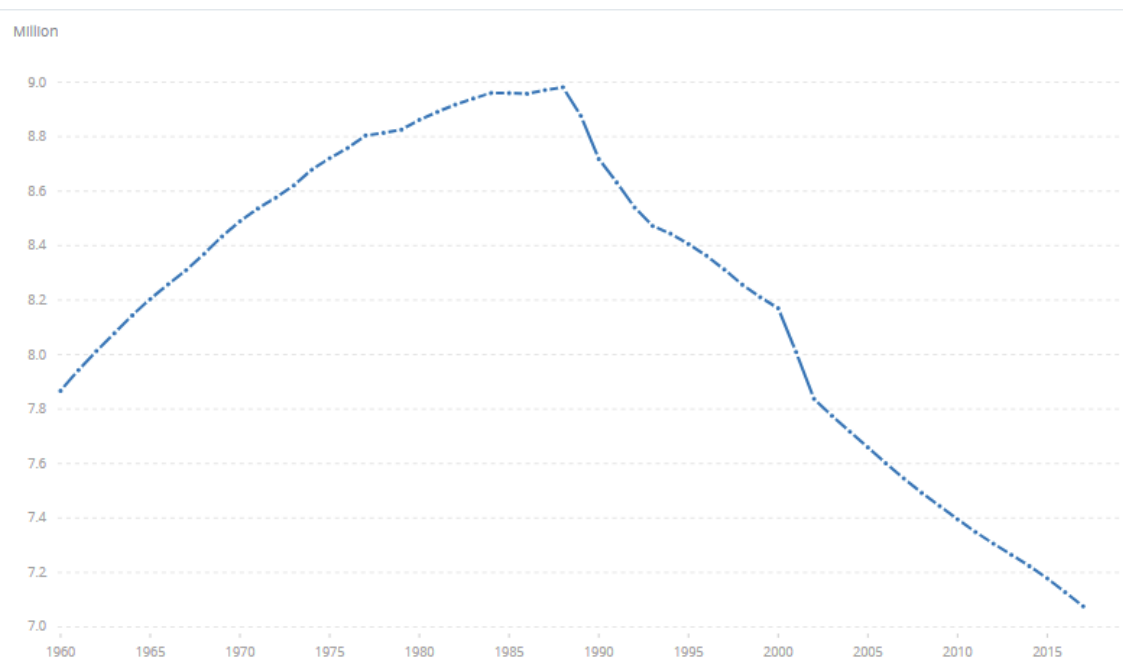
Figure 1: Bulgarian population ("Население на България", 2021)

Before, due to what we are taught, I believed that no foreigner would ever want to learn Bulgarian, what would it serve them for? But after a visit to Bulgaria, a friend of mine, who is a native Spanish speaker, got very motivated to learn Bulgarian. That is when we realized the insufficiency of the resources for second language learning of Bulgarian, and more specifically, tools targeting pronunciation.

As a native speaker, I appreciate the beauty of the language, and I believe that there are ways to make it more accessible to non-native speakers so that it does not become a minority language in the near future.

1.2 Objective

The objective of this thesis is to target a specific problem in Bulgarian language learning and to search for a solution with the help of speech technologies. More precisely - how speech technologies can be used, so that they can aid second language learners of Bulgarian to master their pronunciation.

1.3 Main hypothesis

The main hypothesis of this project is that Automatic-Speech-Recognition and Text-to-Speech systems can be implemented into a pipeline, which can be used as a tool for pronunciation improvement of second language learners of Bulgarian.

1.4 Approach

The approach followed in this thesis is based on research and experiments with available speech technologies in Bulgarian. The research targets specific linguistic problems in the second language learning of Bulgarian, as well as speech technologies and their implementation into language learning. The experiments are conducted with both native and non-native speakers of the language, so the evaluation of the technology is more

accurate. The information gathered from the experiments shows which technology would be helpful for pronunciation improvement. Then the systems, which perform well are implemented in a simple demo, which aims to demonstrate the utility of speech technologies for the improvement of pronunciation of second language learners of Bulgarian.

2. Fundamentals

2.1 The Bulgarian Language

2.1.1 Overview and history

The Bulgarian language is an Indo-European language, which belongs to the Slavic language group. It is the only official language of the Bulgarian Republic. Since the admission of Bulgaria to the European Union in 2007, it became one of the twenty four official languages of the European Union. From then on, Cyrillic has also become the third official script of the EU, following the Latin and Greek scripts.

Bulgarian is currently spoken by an estimated total of 6.8 million people worldwide, 5.7 million of which live in Bulgaria, counting for around 85% of the population.

Moreover, Bulgarian is the first written Slavic language. During the second half of the ninth century, the state of the First Bulgarian Empire brothers St. Cyril and St. Methodius (fig.2) created the Glagolitic script (fig.3), also known as the first Slavic alphabet. Its purpose was to translate liturgical and christian literature from Greek to Bulgarian language.[1]

---

[1] National Geographic България. (2019, May 24). *Глаголица и кирилица*. National Geographic България. https://www.nationalgeographic.bg/a/glagolica-i-kirilica.

Figure 2: St. Cyril and St. Methodius ("Кирил и Методий", 2021)

.

Figure 3: Glagolitic script, the first Bulgarian alphabet
(National Geographic България, 2019)

At the end of the ninth century or the beginning of the tenth century, St. Clement of Ohrid, one of the most prominent students of St. Cyril and St. Methodius participated in the creation of the Cyrillic script (fig.4). It was developed at the Preslav Literary School, and

it was aimed to replace the Glagolitic script. The script was named in honor of St. Cyril, and in 893 it became an official part of the Bulgarian writing system.[2]

Figure 4: Cyrillic script (Мишев, 2019)

The two scripts were used in parallel until the end of the tenth and the beginning of the eleventh century, when the Cyrillic script took over the Glagolitic one due to its ease of writing.

Nowadays, the Cyrillic script is used in more than 50 languages - Slavic (Belarusian, Bulgarian, Macedonian, Russian, Ukrainian, etc.) and Non-Slavic (Abkhaz, Bakshir, Kazakh, Komi, Mongolian, Tajik, Tatar, etc.)(fig.5).[3,4]

---

[2] Britannica, T. Editors of Encyclopaedia (n.d.). *Cyrillic alphabet*. Encyclopedia Britannica. https://www.britannica.com/topic/Cyrillic-alphabet.

[3] Cyrillic script. (2021, June 10). *Wikipedia*. https://en.wikipedia.org/wiki/Cyrillic_script.

[4] Кирилица. (2021, May 27). *Wikipedia*. https://bg.wikipedia.org/wiki/%D0%9A%D0%B8%D1%80%D0%B8%D0%BB%D0%B8%D1%86%D0%B0#%D0%A0%D0%B0%D0%B7%D0%BF%D1%80%D0%BE%D1%81%D1%82%D1%80%D0%B0%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B8_%D1%80%D0%B0%D0%B7%D0%BD%D0%BE%D0%B2%D0%B8%D0%B4%D0%BD%D0%BE%D1%81%D1%82%D0%B8.

Figure 5: Worldwide distribution of Cyrillic alphabet ("Cyrillic alphabets", 2021)

It is a common misconception that the Cyrillic script has Russian origins, which is not the case. After its invention in Bulgaria, it spread to other Slavic countries such as Serbia, Croatia, and Russia during the 10th century.[5,6]

In fact, the Russian version of Cyrillic has three more letters than the Bulgarian one. As well, they have some differences in terms of pronunciation. Such are the letters **и**, **е**, **ъ**, **ь**, **й**, and **щ** (detailed explanation on Bulgarian phonetics and pronunciation in section 2.2.2).[7]

---

[5] Iliev, I. G. (2013, February). *SHORT HISTORY OF THE CYRILLIC ALPHABET*. IJORS International Journal of Russian Studies. http://www.ijors.net/issue2_2_2013/articles/iliev.html.

[6] *Cyrillic language alphabets and how they diverge from one another*. Yale University Library. (n.d.). https://web.library.yale.edu/cataloging/music/cyrillic.

[7] Jakobson, R. (2018). *In Remarks on the phonological evolution of Russian in comparison with the other Slavic languages* (p. 175). essay, The MIT Press.

2.1.2 Phonetics and Second language learning

Second language learning or SLL is 'the process and study of how people acquire a second language', where second language refers to any language studied in addition to the native language.[8]

For second language learners of Bulgarian, more specifically for those with no Slavic phonetic background and mostly people with mother tongues, which use the Latin script, the most challenging part of learning the language is the Cyrillic script.
What makes adaptation to the Cyrillic script easier is the transliteration to the Latin script. Transliteration is a type of conversion of a text from one script to another that involves swapping letters in predictable ways.[9]

The Bulgarian version of Cyrillic has 30 letters (fig. 6), corresponding to 45 sounds or phonemes, of which 6 vowels and 39 consonants.

Аа Бб Вв Гг Дд Ее Жж Зз
Ии Йй Кк Лл Мм Нн Оо Пп
Рр Сс Тт Уу Фф Хх Цц Чч
Шш Щщ Ъъ Ьь Юю Яя

Figure 6: Bulgarian alphabet (Tanya, 2020)

Phoneme or speech sound is the smallest unit of speech distinguishing one word from another, it may have more than one variant, called allophone, which functions as a single sound.[10] Phonemes are represented visually through phonetic transcription.[11] It is usually

---

[8] Rieder-Bünemann A. (2012) *Second Language Learning. In: Seel N.M.* (eds) Encyclopedia of the Sciences of Learning. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1428-6_826

[9] Transliteration. (2021, May 22). *Wikipedia.* https://en.wikipedia.org/wiki/Transliteration.

[10] Encyclopædia Britannica, inc. (n.d.). *Phoneme*. Encyclopædia Britannica. https://www.britannica.com/topic/phoneme.

[11] *Transcription, Pronunciation and Translation of English Words*. Myefe. (2021, March 1). https://myefe.com/transcription-pronunciation.

written in the International Phonetic Alphabet (IPA), which provides a unique symbol for each distinctive phoneme in a language.[12]

The following table represents the letters of the Bulgarian alphabet, their pronunciation in Bulgarian, as well as how they are pronounced in English according to the official Bulgarian-English transliteration and how they are transcribed in accordance with the International Phonetic Alphabet (IPA). I have also included columns with example words in Bulgarian and English, which contain the specific phoneme.

	Bulgarian letter	Pronunciation in Bulgarian	Pronunciation in English (Official transliteration)	Transcription (International Phonetic Alphabet (IPA))	Example words in Bulgarian	Example words in English
1	А а	a	a	[a/ɐ]	мравка (mravk**uh**) - ant	c**a**r/g**u**n
2	Б б	бъ	b	[b/p]	**б**о**б** (**b**o**p**) - beans	**b**oy/po**p** (in the end of a word)
3	В в	въ	v	[v/f]	**в**агон (**v**agon) - wagon краси**в** (krasi**f**) - beautiful	**v**oice/hal**f** (in the end of a word)
4	Г г	гъ	g	[g/k]	**г**ло**г** (**g**lo**k**) - hawthorn	**g**ood/coo**k** (in the end of a word)`
5	Д д	дъ	d	[d/t]	**д**ебел (**d**ebel) - fat гра**д** (gra**t**) - city	**d**og/par**t** (in the end of a word)
6	Е е	e	e	[ɛ]	**е**л**е**н (**e**l**e**n) - deer	p**e**n
7	Ж ж	жъ	zh	[ʒ/ʃ]	**ж**ълт (**zh**ult) -	plea**s**ure/pu**sh** (in

					yellow колаж (kola**sh**) - collage	the end of a word)
8	З з	зъ	z	[z/s]	**з**елен (**z**elen) - green пра**з** (pra**s**) - leek	**z**oo/plu**s** (in the end of a word)
9	И и	и	i	[i]	**и**гла (**i**gla)	b**i**t
10	Й й	и-кратко	y/j	[j]	**й**од (**y**od) - iodine	**y**outh
11	К к	къ	k	[k/g]	**к**он (**k**on) - horse	**k**ite
12	Л л	лъ	l	[l/ł]	**л**илав (**li**lav) - purple	**l**ove
13	М м	мъ	m	[m]	**м**айка (**m**aika) - mother	**m**ine
14	Н н	нъ	n	[n]	**н**ос (**n**os) - nose	**n**ote
15	О о	о	o	[o/ɔ]	**о**гън (**o**gan) - fire	m**o**re
16	П п	пъ	p	[p]	**п**арк (**p**ark) - park	**p**ork
17	Р р	пъ	r	[r]	**р**ана (**r**ana) - wound	**r**ed/pe**r**o in Spanish (more rolled)
18	С с	съ	s	[s/z]	**с**утрин (**s**utrin) - morning	**s**it
19	Т т	тъ	t	[t/d]	**т**ози (**t**ozi) - this	**t**ime
20	У у	у	u	[u/o/w]	**у**тре (**u**tre) - tomorrow	r**u**le
21	Ф ф	фъ	f	[f]	**ф**илм (**f**ilm) -	**f**ish

					movie	
22	Х х	хъ	h	[x]	**х**рана (**h**rana) - food	**h**ot
23	Ц ц	цъ	tz/ts	[t͡s]	**ц**аревица (**ts**arevi**ts**a) - corn	**ts**unami
24	Ч ч	чъ	ch	[t͡ʃ]	**ч**овек (**ch**ovek) - human	**ch**eap
25	Ш ш	шъ	sh	[ʃ]	**ш**ише (**sh**ish**e**) - bottle	**sh**ot
26	Щ щ	штъ	sht	[ʃt]	**щ**астие (**sht**astie) - happiness	sma**shed**
27	Ъ ъ	ер-голям	u/a	[ɤ/ɐ]	**ъ**гъл (**u**gal) - corner	**a**bout/**a**
28	Ь ь	ер-малък	y	[j/not pronounced]	син**ь**о (sin**y**o) - blue	not pronounced (softens the previous consonant)
29	Ю ю	йу	yu	[ju/u/jo/o]	кл**ю**н (kl**yu**n) - beak	**you**
30	Я я	йя	ya	[ja/a/jɐ/ɐ]	гр**я**х (gr**ya**h) - sin	**ya**rn

Table 1: Bulgarian alphabet - pronunciation, transcription, examples.[13,14,15,16]

---

[13] Българска азбука. (2021, June 7). *Wikipedia.*
https://bg.wikipedia.org/wiki/%D0%91%D1%8A%D0%BB%D0%B3%D0%B0%D1%80%D1%81%D0%BA%D0%B0_%D0%B0%D0%B7%D0%B1%D1%83%D0%BA%D0%B0.

[14] *Bulgarian (Български)*. Omniglot.com. (2021, April 23). https://omniglot.com/writing/bulgarian.htm.

[15] *Learn the Bulgarian pronunciation.* coLanguage. (n.d.). https://www.colanguage.com/learn-bulgarian-pronunciation.

[16] Букви и звукове в българския език. (2020, October 19). *Wikipedia*.
https://bg.wikipedia.org/wiki/%D0%91%D1%83%D0%BA%D0%B2%D0%B8_%D0%B8_%D0%B7%D0%B2%D1%83%D0%BA%D0%BE%D0%B2%D0%B5_%D0%B2_%D0%B1%D1%8A%D0%BB%D0%B3%D0%B0%D1%80%D1%81%D0%BA%D0%B8%D1%8F_%D0%B5%D0%B7%D0%B8%D0%BA.

In pronunciation, difficulties for second language learners arise in phonemes, not present in the native languages of the learners, such as **ъ, [ɤ/ɐ]**.[17] More challenges in pronunciation come from phonemes, which are combinations of more than one phoneme, such as tz, sht, yu, ya, etc, as well as from accumulations of consonants.[18]

Furthermore, letters in Bulgarian might be pronounced differently depending on their position in a word. For example, when voiced consonants are at the end of a word they are pronounced as voiceless. An example is **боб** [**b**oп] - beans.[19] This can also be considered as a specification of the language, which is more difficult for second language learners to master.

From the 30 letters of the alphabet, the 6 vowels are:

Cyrillic	Transliteration	Transcription (IPA)
а	a	[a/ɐ]
ъ	u/a	[ɤ/ɐ]
о	o	[o/ɔ]
у	u	[u/o/w]
е	e	[ɛ]
и	i	[i]

Table 2: Vowels

[17] *МЕТОДИКА НА ОБУЧЕНИЕТО ПО БЪЛГАРСКИ ЕЗИК ЗА МИГРАНТИ.* (n.d.). https://download.ei-ie.org/Docs/WebDepot/SEB%20Handbook.pdf.

[18] *Трудности при овладяване на българската фонетична система.* elearn.uni-sofia. (n.d.). https://elearn.uni-sofia.bg/mod/resource/view.php?id=13605.

[19] *Как звучи българският език на чужденците?* Omega LS. (2019, September 30). https://omegals.bg/kak-zvuchi-bulgarskiqt-ezik-na-chuzdencite/.

The consonants are divided into two groups - hard and soft.[20,21] The hard ones are represented in the following table, where the phonemes **(дж, dzh, [dʒ])** and **(дз, dz, [dz])** are written as a combination of two letters.

Cyrillic	Transliteration	Transcription (IPA)
б	b	[b]
в	v	[v]
г	g	[g]
д	d	[d]
ж	zh	[ʒ]
дж	dzh	[dʒ]
з	z	[z]
дз	dz	[d͡z]
к	k	[k]
л	l	[ł]
м	m	[m]
н	n	[n]
п	p	[p]
р	r	[r]
с	s	[s]
т	t	[t]

[20] Davies, R. (2015, September 28). *Basic Bulgarian, Pronunciation - Consonants, Round 1*. Duolingo. https://forum.duolingo.com/comment/10741347/Basic-Bulgarian-Pronunciation-Consonants-Round-1.

[21] Bulgarian phonology. (2021, June 10). *Wikipedia.* https://en.wikipedia.org/wiki/Bulgarian_phonology.

ф	f	[f]
х	h	[x]
ц	tz/ts	[t͡s]
ч	ch	[t͡ʃ]
ш	sh	[ʃ]

Table 3: Hard consonants

The soft consonants are:

Cyrillic	**Transliteration**	**Transcription (IPA)**
б'	b	[b']
в'	v	[v']
г'	g	[k']
д'	d	[d']
з'	z	[z']
дз'	dz	[d͡z']
к'	k	[k']
л'	l	[l]
м'	m	[m']
н'	n	[n']
п'	p	[p']
р'	r	[r]
с'	s	[s']

т'	t	[t']
ф'	f	[f']
х'	h	[x']
ц'	tz/ts	[t͡s']
й	u/j	[j]

Table 4: Soft consonants

In the Bulgarian language, stress is free and non-fixed, which makes it more difficult for second language learners to learn how to pronounce correctly. An example is м'ъж (m'uzh) [mɤʃ] - man and мъж'ът (muzh'ut) [mɤʒɤt] the man. Here the word is articulated, and the stress moves from the first to the second syllable. It is also an example of how when voiced consonants are at the end of a word they are pronounced as voiceless. In the first word **ж** is pronounced as **sh** [ʃ], which is voiceless, while in the second one, where it is not at the end, it is voiced **zh** [ʒ].[22]

2.2 Speech Technologies

Speech technology is a type of computing technology that can recognize, analyze, duplicate, understand and respond to spoken human language. It has many uses and applications. Subfields of speech technology include speech synthesis, speech recognition, speaker recognition and verification, and multimodal interaction.[23]
Speech technology allows communication with computers without the usage of a keyboard. Nowadays, it is implemented in every smart device, often as virtual assistants. Some commercial personal assistants are Amazon Alexa, Apple's Siri, Google Assistant, Microsoft's Cortana.

---

[22] Innovative Language Learning. (2014). Top 5 Tips for Avoiding Common Mistakes in Bulgarian. In *Learn Bulgarian - Level 1 Introduction to Bulgarian, Volume 1: Volume 1: Lessons 1-25.* essay.
[23] Contributor, T. T. (2019, February 14). *What is speech technology?* SearchUnifiedCommunications. https://searchunifiedcommunications.techtarget.com/definition/speech-technology.

As well, speech technology, in the form of Text-to-Speech, Automatic Speech Recognition, or both, has been introduced in the majority of language learning applications, which are applications that aim to help users to learn and practice a specific language (further described in section 3.2).

2.2.1 Text-to-Speech

2.2.1.1 Overview

Text-to-Speech (TTS) technology is a form of speech synthesis, which is the conversion of normal language text into synthesized speech or the artificial production of human speech. It also includes converting phonetic transcriptions into speech.[24]

Most TTS systems work in the following manner. The first step is tokenization - the written input is analyzed, and abbreviations, numbers, dates, etc are converted into their written form. Next, grapheme-to-phoneme conversion takes place, where the text is split into prosodic units - phrases, clauses, sentences, and every word is assigned a phonetic transcription. Afterwards, the synthesizer converts into sound the symbolic linguistic representation. The general process can be observed in the figure below (fig.7).
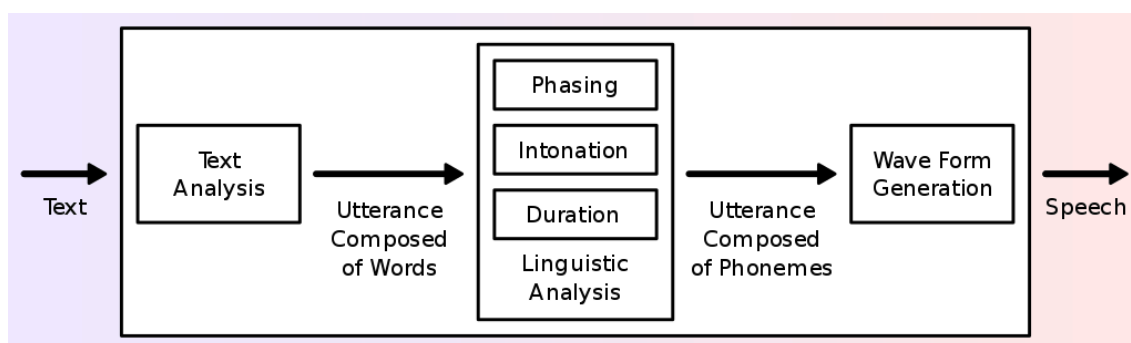

Figure 7: Typical TTS system ("Speech synthesis", 2021)

TTS was initially developed to aid visually impaired, people with reading difficulties and learning disabilities, thus helping them overcome literacy challenges. Nowadays, TTS

---

[24] Speech synthesis. (2021, June 9). *Wikipedia.* https://en.wikipedia.org/wiki/Speech_synthesis.

technologies have a broader range of applications in various industries. Some of them include Finance, Tourism, Telecommunications, and E-learning.

2.2.1.2 TTS in language learning

E-learning is a learning system based on formalized teaching but with the help of electronic resources.[25] A subcategory of E-learning is Computer Assisted Language Learning (CALL) (further explained in section 3.1). Many CALL systems implement TTS as a tool for language learning.

Some examples include a study (Huang & Liao, 2015) conducted in Taiwan with second language learners of English, where the implementation of TTS into the learning process during one semester was reported to strengthen students' spelling ability and to increase their self-learning motivation. Another study (Bione, Grimshaw & Cardoso, 2016) with Brazilian English language learners reported their positive view on using TTS as a pedagogical tool.

Furthemore, an experiment (Meihami & Husseini, 2014) conducted with English language learners at the Azad University of Ghorveh, which used IVONA UK Brain 1.4.21 TTS showed that in general TTS had positive effects on students' Total Fluency, which is a combination of features, such as word stress, word intonation, pitch contour, and fluency.

Currently, the majority of language learning applications, such as Duolingo (further described in section 3.2) implement Text-to-Speech technology as a tool for improvement of student pronunciation, understanding, and listening skills.

---

[25] *What is E-learning? Definition of E-learning, E-learning Meaning.* The Economic Times. (n.d.). https://economictimes.indiatimes.com/definition/e-learning.

2.2.2 Automatic Speech Recognition

2.2.2.1 Overview

Automatic Speech Recognition (ASR) or Speech-to-Text technology is the conversion of human speech into text, or the process of deriving the transcription of an utterance, given the speech waveform.[26]

Most ASR systems work in the following principle. First, the speaker talks to the system. Then, their audio is broken down into phonemes, normally using acoustic language modeling. Acoustic modeling is the relationship between phonemes/linguistic units of speech and audio signals, and language modeling uses statistical and probabilistic analysis on how linguistic units are connected in a sequence. After 'analyzing' the audio input, the system returns a text, which is supposed to correspond to the spoken audio.[27]

Just like TTS, ASR also has a majority of applications in a large number of areas. Some of them are Finance, Marketing, Healthcare, the Internet of Things, and again E-learning.

2.2.2.2 ASR in language learning

ASR is also used as a tool for language learning so that students can practice their pronunciation and speaking skills. A large number of studies have researched the benefits of its application to second language learning.

One example is a study (Junining, Alif & Setiarini, 2020), conducted with English language learners in Indonesia, which shows that ASR can be used as a tool for students to practice speaking individually so that their level of anxiety can be reduced when speaking in front of other people.

---

[26] Vijaya, Samudra K. (2017, November) *Automatic Speech Recognition.*
http://www.iitg.ac.in/samudravijaya/tutorials/asrTutorial.pdf
[27] Acoustic model. (2020, January 4). *Wikipedia.* https://en.wikipedia.org/wiki/Acoustic_model.

At present, some of the language learning applications, such as Babbel (further described in section 3.2) implement Automatic Speech Recognition technology as a tool for improvement of student pronunciation, understanding, and speaking skills.

Regarding ASR in Bulgarian, BulPhonC has been developed (Hateva, Mitankin & Mihov, 2016). It is a Bulgarian speech corpus, which aims to be used for the development of ASR Technology. But no ASR technology is available in the paper and does not seem to have been developed yet.

In section 4.2.1 I explain further about ASR technologies and their availability in Bulgarian.

3. State of the art

3.1. Computer Assisted Language Learning

Computer Assisted Language Learning (CALL), also known as Computer-Aided Instruction (CAI) or Computer-Aided Language Instruction (CALI) is "the search for and study of applications of the computer in language teaching and learning".[28,29]

I did not find any papers on CALL used for the teaching of Bulgarian. However, English is the language with the largest number of resources on the topic, such as CALL applied to teaching English as a foreign language (EFL) in Saudi Arabia (Hashmi, 2016), in Iran (Pirasteh, 2014), etc.

In the past Computer Assisted Language Learning attempted the implementation of ASR systems as a tool for English language teaching, but due to low-level accuracy, they gained a poor reputation (Carrier, 2017). However, with the rapid development of technology in recent years, CALL systems have risen in popularity. Some benefits include interactivity, accessibility at any time, and a stress-free environment to learn.

Many applications are currently focused on Computer aided pronunciation training or CAPT and are being used as a tool by non-native speakers to improve their pronunciation. According to a paper (Agarwal & Chakraborty, 2019), CAPT systems can be divided into four categories: Visual simulation based systems, Game based systems, Comparative phonetics based systems, and Artificial neural network based systems (fig.8).

---

[28] Levy M. (1997) *CALL: context and conceptualisatio*n, Oxford: Oxford University Press.
[29] Computer-assisted language learning. (2021, May 7). *Wikipedia.*
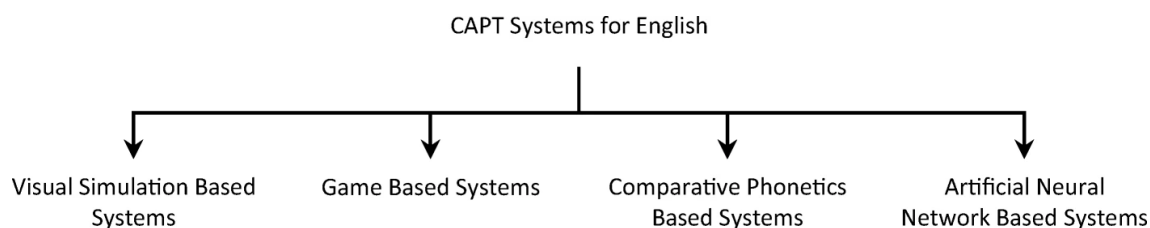https://en.wikipedia.org/wiki/Computer-assisted_language_learning.

Figure 8: CAPT systems (Agarwal & Chakraborty, 2019)

Visual simulation based systems are normally used for younger speakers. They record their speech and after analyzing it provide feedback through images, videos, and animated characters.

Game based systems can simulate real-world situations and include both formal and informal situations.

Comparative based systems are used by adult learners, fluent in their native languages. The learners record themselves and then the system provides feedback based on their mother tongues by comparing phonemes between both languages.

Artificial neural network based systems have to be trained using a corpora of hundreds of sentences, then they can be used for the detection of mispronunciations in learners' speech. An example is a deep neural network, designed by Li et al. (2017) to detect and correct mispronunciations caused by the differences in phonetics between the speakers' native language and English, incorrect conversion from letter to sound, and misreading text prompts.

3.2. Language learning applications

Most language learning applications are game based systems. Currently, some of the most used and among the top-rated applications for second language learning include Duolingo, Babbel, Mondly, Pimsleur, Rosetta Stone, Mango, Drops, Busuu, AudioNote, Rocket Languages, etc.[30] From them, Mondly, Babbel, Busuu, AudioNote, Rocket Languages, and Rosetta Stone implement speech recognition in their systems to facilitate

---

[30] Schumer, L. (2021, March 24). *9 Best Language Apps for Learning on the Go.* Good Housekeeping. https://www.goodhousekeeping.com/life/g32175725/best-language-learning-apps/.

students' speech practices.[31] ELSA Speak is another application, which aims to enhance the English pronunciation skills of language learners through the implementation of ASR and has shown positive results in a study (Kholis, 2020).

The speech recognition software is normally used, so that students can have a simulated conversation with the application, to practice their pronunciation of specific words and sentences, and also to help them perfect their accent. When the learner records their speech, from the ASR output they can know if their pronunciation is correct, clear, and how they can improve.

From all of the applications enumerated above, Mondly is the only one, which offers a course for learning the Bulgarian language.[32] In Mondly, speech practices consist of dialogues between phrases recorded by Bulgarian actors, which also appear written on the screen, and a list of responses, from which the learner can choose to say to the device. The speech recognition functionality is available for premium users only, but on the webpage, there is a demo[33] demonstrating how it works. Moreover, in 2017, Mondly launched Mondly VR, where students can learn in a virtual reality environment by communicating with chatbots and speech recognition systems. And in 2019, they introduced multiplayer rooms, in which users can connect and practice with each other.[34]

Some other applications, which offer courses in Bulgarian include BulgarianPod101, which is one of the best for listening comprehension but does not implement speech recognition.[35] Another one is FunEasyLearn, which implements speech recognition as a tool for pronunciation improvement but does not seem to be highly developed for the Bulgarian language.[36]

---

[31] Meredithkreisa. (2021, January 18). *6 Language Apps That Use Speech Recognition for Well-rounded Learning.* FluentU Language Learning. https://www.fluentu.com/blog/speech-recognition-language-learning/.

[32] *Learn Bulgarian Online in Just 10 Minutes a Day.* Mondly Blog. (2020, September 11). https://www.mondly.com/blog/2020/09/11/learn-bulgarian-online/.

[33] *Play your way to a new language with Mondly.* Mondly. (n.d.). https://www.mondly.com/ph.

[34] *Mondly VR Is Now Available on Steam.* Mondly Blog. (2020, June 9). https://www.mondly.com/blog/2019/09/25/mondly-learn-languages-in-vr-is-now-available-on-steam/.

[35] *Bulgarian Language with a Free App.* BulgarianPod101. (n.d.). https://www.bulgarianpod101.com/app/.

[36] *Learn Bulgarian - Free, Fast & Effective.* FunEasyLearn. (n.d.). https://www.funeasylearn.com/learn-bulgarian.

4. Experiments and evaluation

In order to understand whether and how speech technologies can be applied to the second language learning of Bulgarian, some experiments had to be performed.

For the experiments, I started by researching and gathering information on which words and phonemes are considered more challenging for second language learners of Bulgarian to pronounce. The resources were scarce, but I managed to create an initial list of 20 words (table 5). 10 of them I took from BulgarianPod101's (mentioned in the previous section - 3.2) webpage as 'Top 10 Hardest Words to Pronounce'.[37] The rest I chose based on the accumulation of consonants or the presence of specific phonemes. The idea was to analyze the most common sources of errors to further develop my project, which ideally had to be done in collaboration with a Bulgarian language teacher because even though I am a native speaker, I am not a professional linguist.

As a native speaker, I cannot clearly distinguish whether these words have the same level of difficulty, but the general idea was that they did, which I wanted to confirm through the experiments.

	Cyrillic	**English transliteration**	**English translation**
w1	Благодаря	Blagodarya	thank you
w2	Довиждане	Dovizhdane	goodbye
w3	Здравей	Zdravey	hello
w4	Птицечовка	Ptitsechovka	platypus
w5	Патладжан	Patladzhan	eggplant
w6	Цветарница	Tsvetarnitsa	flower shop
w7	Круша	Krusha	pear

	Cyrillic	English transliteration	English translation
w8	Щъркел	Shtarkel	stork
w9	Странник	Strannik	stranger
w10	Джудже	Dzhudzhe	dwarf
w11	Дрънкулка	Drankulka	trinket
w12	Триъгълник	Triagalnik	triangle
w13	Блясък	Blyasak	shine/glow
w14	Учреждение	Uchrezhdenie	establishment
w15	Площад	Ploshtad	square
w16	Спречкване	Sprechkvane	argument
w17	Сключвам	Sklyuchvam	conclude
w18	Лекарство	Lekarstvo	medicine
w19	Взрив	Vzriv	explosion
w20	Държава	Darzhava	country

Table 5: Initial target words

Then, based on the list, I created six sentences, which combine the different words (in bold) into complicated sentences, which I aim to have the same level of difficulty. The following table contains the sentences, their transliteration, and translation to English:

	Cyrillic	**English transliteration**	**English translation**
e1_s1	Имаше **спречкване** между **джуджето** и **странника** на **площада** пред **учреждението**.	Imashe **sprechkvane** mezhdu **dzhudzheto** i **strannika** na **ploshtada** pred **uchrezhdenieto**.	There was an **argument** between the **dwarf** and the **stranger** at the **square** in front of the **establishment**.
e1_s2	**Птицечовката** видя ярък	**Ptitsechovkata** vidya yarak	The **platypus** saw a bright

	блясък над гората след **взрива**.	**blyasak** nad gorata sled **vzriva**.	**glow** over the forest after the **explosion**.
e1_s3	**Държавата** **сключва** ли договор с **щъркелите** да пази техните **дрънкулки**?	**Darzhavata** **sklyuchva** li dogovor s **shtarkelite** da pazi tehnite **drankulki**?	Does the **country conclude** a contract with the **storks** to protect their **trinkets**?
e1_s4	В **цветарницата** не се продават **лекарства**, а само **патладжани** и **круши**.	V **tsvetarnitsata** ne se prodavat **lekarstva**, a samo **patladzhani** i **krushi**.	At the **flower shop** they don't sell **medicine**, only **eggplants** and **pears**.
e1_s5	**Здравей**, **страннико**, **благодаря** ти за **помощта**!	**Zdravey**, **stranniko**, **blagodarya** ti za **pomoshtta**!	**Hello stranger**, **thank you** for your **help**!
e1_s6	**Джуджето**, което имаше глава с формата на **триъгълник**, каза **довиждане** на **щъркела**.	**Dzhudzheto**, koeto imashe glava s formata na **triagalnik**, kaza **dovizhdane** na **shtarkela**.	The **dwarf** , whose head was in the shape of a **triangle**, said **goodbye** to the **stork**.

Table 6: Initial target sentences

Then, after conducting the first ASR experiment (point 4.2.2.1) with a second language learner of Bulgarian, whose mother tongue is Spanish, I analyzed the results based on the participant's performance and the sentence difficulty they reported. I came to the conclusion that it is not an efficient solution to include more than one target word in a single sentence. That is why I created a new list of simpler sentences, each one targeting one word (in bold) from the word list. Again, the idea was that the sentences have a similar level of difficulty. The sentences, together with their translation and translation to English are included in the following table:

	Cyrillic	**English transliteration**	**English translation**
e2_s1	**Благодаря** ти!	**Blagodarya** ti!	**Thank you**!
e2_s2	**Довиждане**, до скоро!	**Dovizhdane**, do skoro!	Goodbye, see you soon!
e2_s3	**Здравей**, как си?	**Zdravey**, kak si?	**Hello**, how are you?

e2_s4	**Птицечовката** се храни с насекоми.	**Ptitsechovkata** se hrani s nasekomi.	The **platypus** eats insects.
e2_s5	Тя много обича да яде **патладжани** на грил.	Tya mnogo obicha da yade **patladzhani** na gril.	She really likes to eat grilled **eggplants**.
e2_s6	**Цветарницата** се намира отсреща.	**Tsvetarnitsata** se namira otsreshta.	The **flower shop** is located opposite.
e2_s7	В пекарната предлагат пай с **круши**.	V pekarnata predlagat pay s **krushi**.	The bakery offers **pear** pie.
e2_s8	Когато настъпи есента **щъркелите** отлитат на юг.	Kogato nastapi esenta **shtarkelite** otlitat na yug.	When autumn comes, **storks** fly south.
e2_s9	Имаше висок **странник** пред магазина.	Imashe visok **strannik** pred magazina.	There was a tall **stranger** in front of the store.
e2_s10	Приказката за Снежанка и седемте **джуджета** ми е любима.	Prikazkata za Snezhanka i sedemte **dzhudzheta** mi e lyubima.	The story of Snow White and the Seven **Dwarfs** is my favorite!
e2_s11	Тя има много ръчно изработени **дрънкулки**.	Tya ima mnogo rachno izraboteni **drankulki**.	She has many handmade **trinkets**.
e2_s12	**Триъгълникът** е една от основните форми в математиката.	**Triagalnikat** e edna ot osnovnite formi v matematikata.	The **triangle** is one of the main shapes in mathematics.
e2_s13	Как да придадем повече **блясък** на косата.	Kak da pridadem poveche **blyasak** na kosata.	How to add more **shine** to hair?
e2_s14	Подадох документите в най-близкото **учреждение**.	Podadoh dokumentite v nay-blizkoto **uchrezhdenie**.	I submitted the documents to the nearest **institution/establishment**.
e2_s15	По празниците хората се събират на **площада**.	Po praznitsite horata se sabirat na **ploshtada**.	During the holidays, people gather at the **square**.
e2_s16	По време на изборите имаше много	Po vreme na izborite imashe mnogo **sprechkvaniya**.	During the elections, there were many

	спречквания.		**arguments/disputes**.
e2_s17	**Сключвам** договор за наем на новия апартамент.	**Sklyuchvam** dogovor za naem na noviya apartament.	I am **signing** a rental agreement for the new apartment.
e2_s18	Когато се разболея пия **лекарства**.	Kogato se razboleya piya **lekarstva**.	When I get ill, I take **medication**.
e2_s19	Чу се силен **взрив**.	Chu se silen **vzriv**.	There was a loud **explosion**.
e2_s20	В **държавата** се провеждат избори.	V **darzhavata** se provezhdat izbori.	Elections are being held in the **country**.

Table 7: Simplified target sentences

In the following section (4.1) is described the analysis from the experiments with TTS systems. In points 4.2.2.1 and 4.2.2.2 are explained the details and results from experiments 1 and 2, conducted with ASR systems.

4.1. Analysis of TTS systems

The analysis of the Text-to-Speech systems was aimed to see whether this technology could be applied as a tool for improving pronunciation of second language learners of Bulgarian. The way it could be implemented is as a guide on how to correctly pronounce words and sentences.

I analyzed two TTS systems - Nuance TTS and OpenTTS. Nuance TTS offers natural-sounding speech synthesis in 53 languages, one of which is Bulgarian. There are also 119 voice options, but for Bulgarian, there is only one - Daria, and it is not open source.[38] Open TTS is an open-source text-to-speech system, running on docker, available in a large number of languages, including Bulgarian.[39]

---

[38] *Text-to-Speech (TTS) Engine in 119 Voices: Nuance: Nuance.* Nuance Communications. (n.d.). https://www.nuance.com/omni-channel-customer-engagement/voice-and-ivr/text-to-speech.html#.

[39] Synesthesiam. (n.d.). *synesthesiam/opentts*. GitHub. https://github.com/synesthesiam/opentts.

The evaluation of TTS systems was based on three criteria - intelligibility, expressiveness, and naturalness, ranging from 1 to 5. Intelligibility is the quality of being understandable.[40] Expressiveness is the quality of effectively conveying a thought or feeling.[41] And naturalness is the quality or state of being natural, and not sounding like a robot.[42] The following table shows what each score means for each criteria:

Score	Intelligibility	Expressiveness	Naturalness
1	not understandable at all	not expressive at all	very robotic
2	a bit understandable	a bit expressive	robotic
3	understandable	expressive	a bit natural/a bit robotic
4	very well understandable	very expressive	very natural
5	perfectly understandable	perfectly expressive	perfectly natural

Table 8: TTS evaluation criteria

First of all, I rated the systems myself in order to get a general idea of their performance. In general, Daria performs very well. The system is very well understandable, and also performs well in terms of expressiveness - the interrogative questions are read with the correct intonation for a question and makes pauses where there are commas. Regarding exclamatory sentences, they are not expressed very well. For some words, the stress is mispronounced, which makes them sound a bit unnatural, but in general, the voice is very natural. For the majority of the words from the list (Table 5 in section 4), the system performs on a satisfactory level.

On the other hand, OpenTTS is very robotic. For me, as a native speaker, regarding intelligibility, the majority of words and phrases are very well understandable. It reads interrogative sentences correctly and pauses where there are commas. As in Daria,

---

[40] *intelligibility*. Cambridge Dictionary. (n.d.).
https://dictionary.cambridge.org/dictionary/english/intelligibility.
[41] Lexico Dictionaries. (n.d.). Definition of *EXPRESSIVENESS*. Lexico Dictionaries | English.
https://www.lexico.com/definition/expressiveness.
[42] Lexico Dictionaries. (n.d.). Definition of *NATURALNESS*. Lexico Dictionaries | English.
https://www.lexico.com/definition/naturalness.

exclamatory sentences are not really expressive. Stress is again mispronounced for some words, which combined with the robotic voice makes it more unintelligible in some sentences.

After rating the systems myself, I ran a perception test with an equal number of native and non-native speakers of Bulgarian.
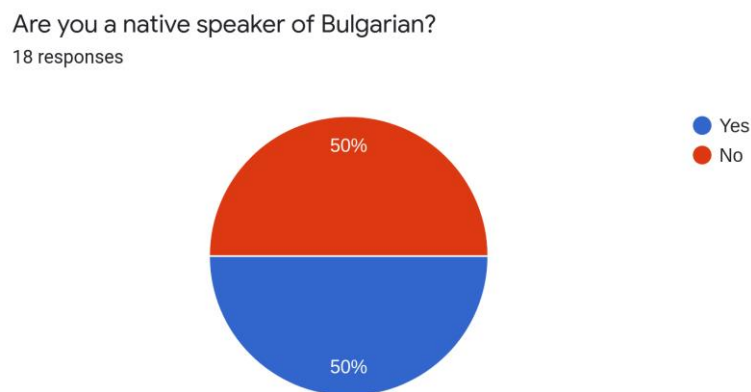
Figure 9: Perception test participants ratio

It consisted of a Google forms questionnaire (fig.10), where I asked them to evaluate the recordings of Daria for the first six sentences (Table 6 in section 4) and of OpenTTS for all the sentences (Tables 6 and 7 in section 4).

Figure 10: TTS perception test questionnaire

The following table includes links to the recordings of the six sentences for experiment 1 (Table 6 in section 7). In the first column are the recordings of Daria (NuanceTTS), in the second one - of OpenTTS, and in the third one - the recordings of a native speaker. The recordings of the native speaker serve as an example of how the sentences should be pronounced correctly.

Daria	OpenTTS	Native speaker
e1_s1	e1_s1	e1_s1
e1_s2	e1_s2	e1_s2
e1_s3	e1_s3	e1_s3
e1_s4	e1_s4	e1_s4
e1_s5	e1_s5	e1_s5
e1_s6	e1_s6	e1_s6

Table 9: TTS experiment 1

The following two bar charts show the average of all the participants' scores of the intelligibility for each sentence for both Daria and OpenTTS. We can observe that the participants perceive Daria as understandable to very well understandable, while OpenTTS as a bit understandable.
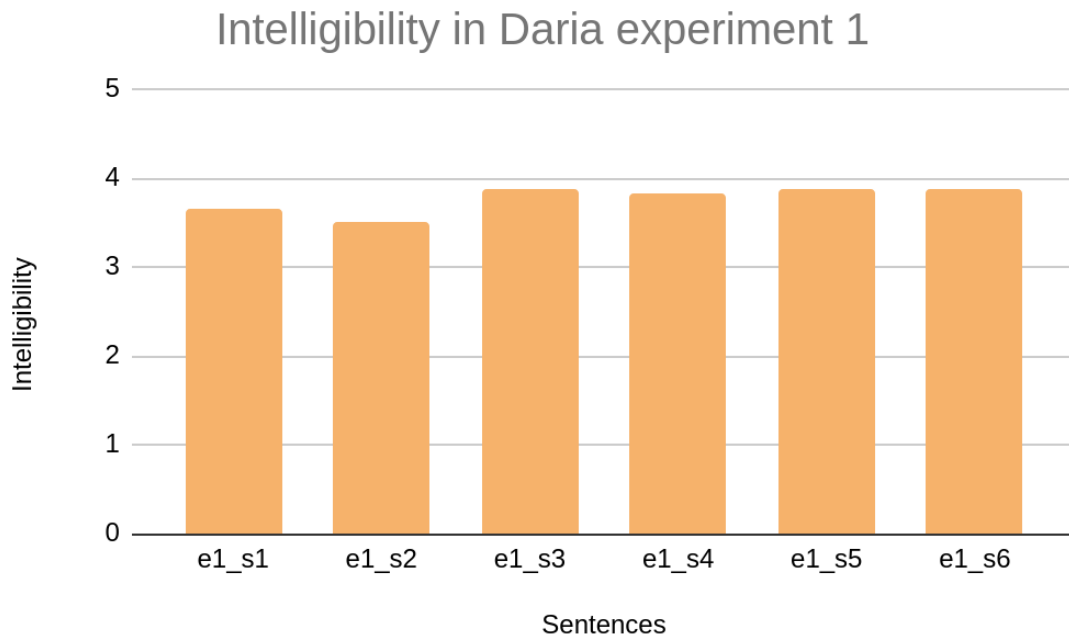
Figure 11: Intelligibility Daria experiment 1
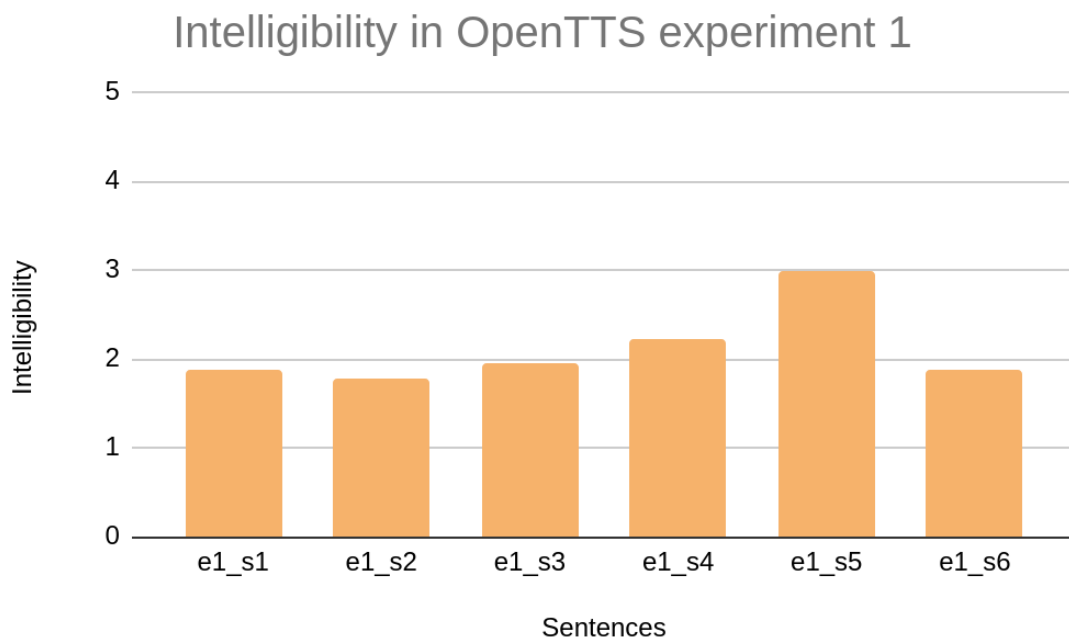
Figure 12: Intelligibility OpenTTS experiment 1

The next two bar charts depict the average of all the participants' scores of the expressiveness for each sentence for both Daria and OpenTTS. According to the participants, Daria is expressive, while OpenTTS ranges from not expressive at all to a bit expressive.
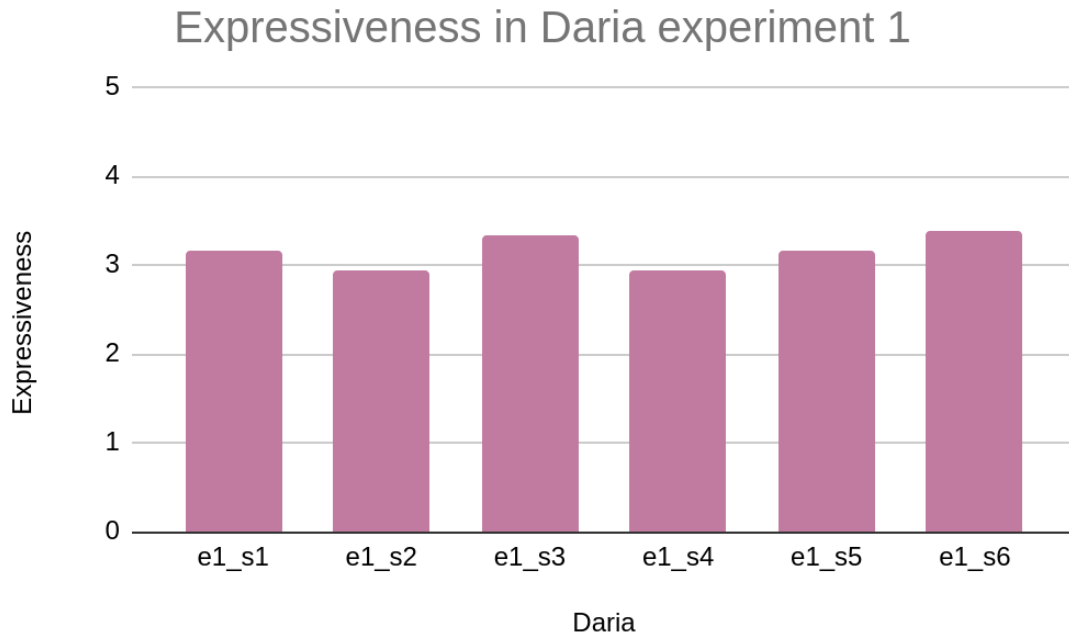
Figure 13: Expressiveness Daria experiment 1
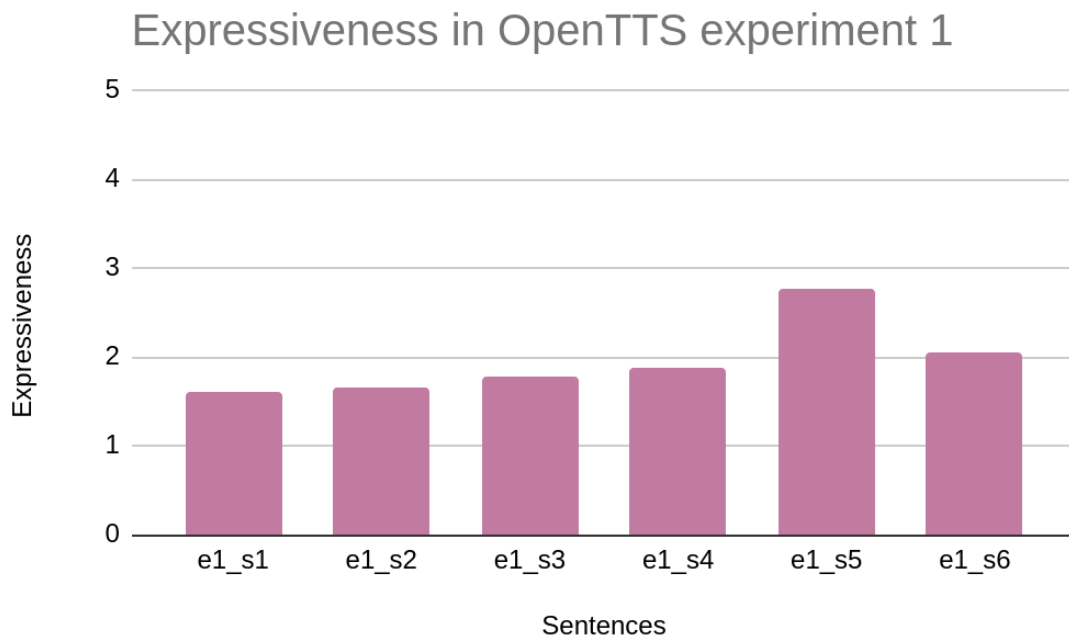
Figure 14: Expressiveness OpenTTS experiment 1

In the following two bar charts, we can see the average of all the participants' scores of naturalness for each sentence for both Daria and OpenTTS. The speakers rated Daria from a bit natural/a bit robotic to very natural, while OpenTTS as very robotic to robotic.
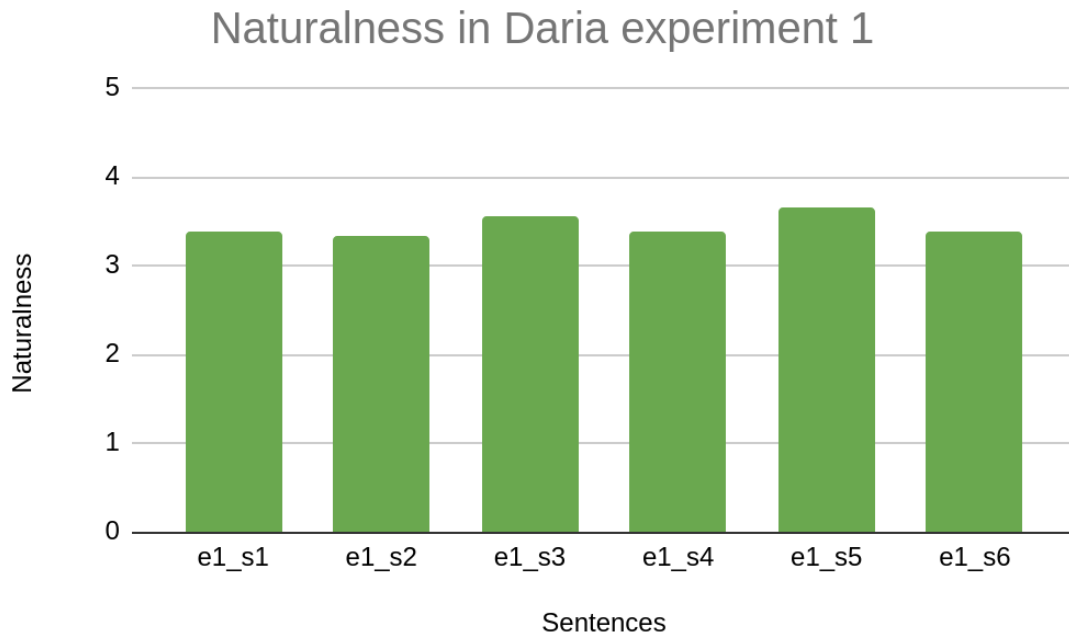
Figure 15: Naturalness Daria experiment 1
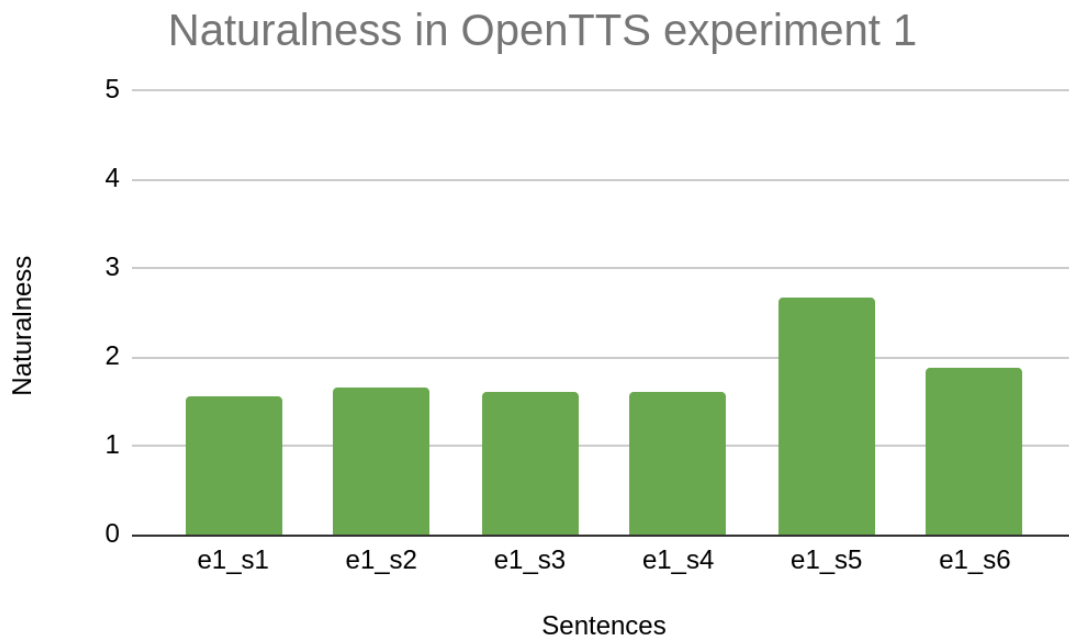
Figure 16: Naturalness OpenTTS experiment 1

From the first experiment, we can conclude that Daria performs better in terms of all the criteria compared to OpenTTS. Even when asked whether they think Daria could be used as a tool to help second language learners with their pronunciation, 100% of the participants replied positively.

Do you think Daria could be used as a tool for improving pronunciation of second language learners of Bulgarian? (that is if a student listens to the audio, it would be useful for them to learn)
18 responses
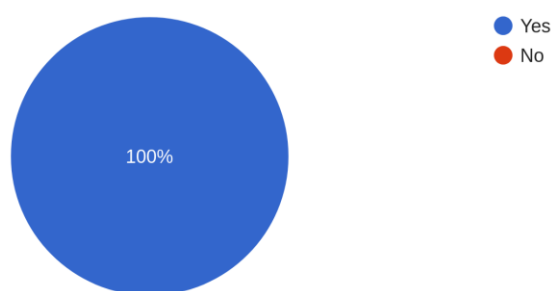
Figure 17: Daria survey

But after conducting the experiment, I realized that NuanceTTS, consequently Daria, is not open-source software, and thus, unless paid for, it cannot be implemented into a demo, where it could serve as a tool for aiding second language learners' pronunciation.

However, OpenTTS is open-source software. That is why, the participants were asked to rate the recordings of the sentences from experiment 2 (Table 7 from section 4), based on the same criteria - intelligibility, expressiveness, and naturalness (Table 8)

The following table includes links to the recordings of the sentences for experiment 2 (Table 7 in section 4). In the first column are the recordings of OpenTTS, and in the second one - the recordings of a native speaker. The recordings of the native speaker serve as an example of how the sentences should be pronounced correctly.

OpenTTS recordings	Native speaker
e2_s1	e2_s1
e2_s2	e2_s2
e2_s3	e2_s3
e2_s4	e2_s4
e2_s5	e2_s5
e2_s6	e2_s6
e2_s7	e2_s7
e2_s8	e2_s8
e2_s9	e2_s9
e2_s10	e2_s10
e2_s11	e2_s11
e2_s12	e2_s12
e2_s13	e2_s13
e2_s14	e2_s14
e2_s15	e2_s15
e2_s16	e2_s16
e2_s17	e2_s17
e2_s18	e2_s18
e2_s19	e2_s19
e2_s20	e2_s20

Table 10: TTS experiment 2

The following bar charts show the average of all the participants' scores for the three criteria for each sentence for OpenTTS.
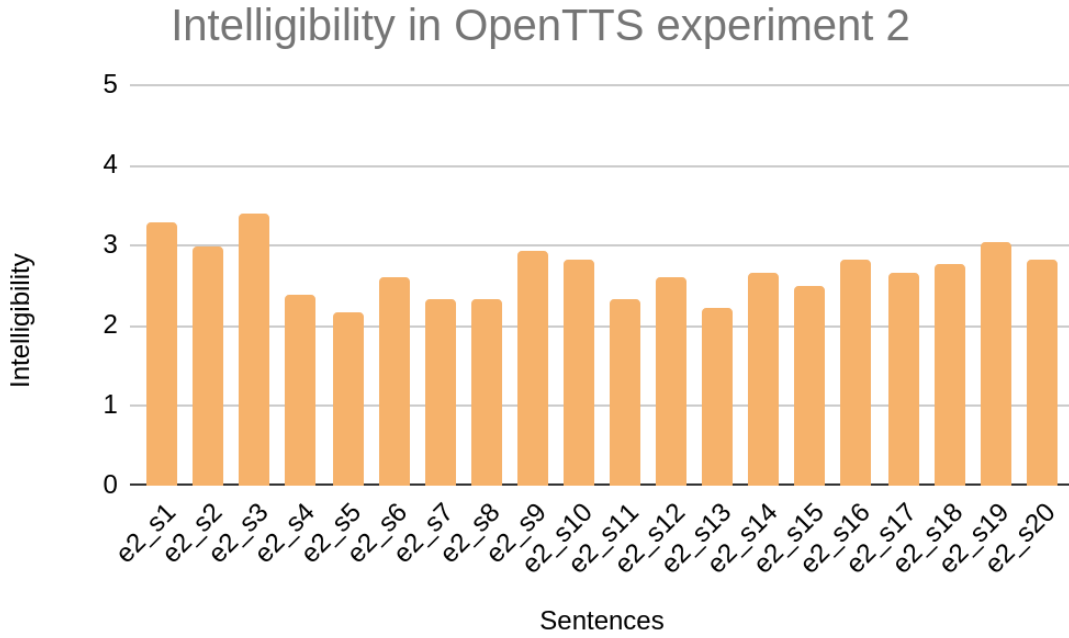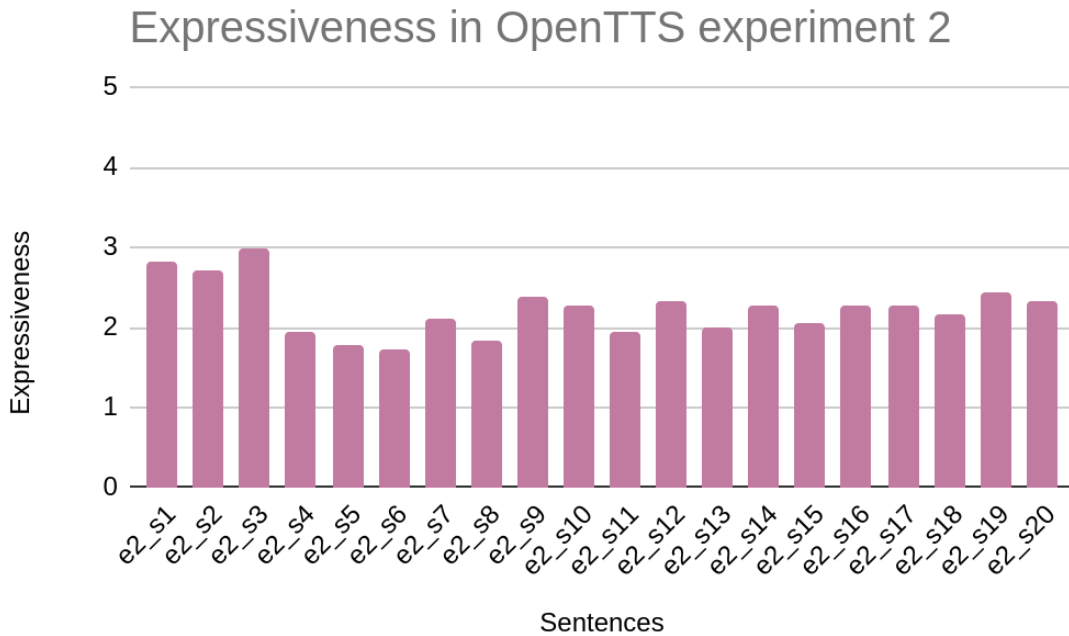

Figure 18: Intelligibility OpenTTS experiment 2

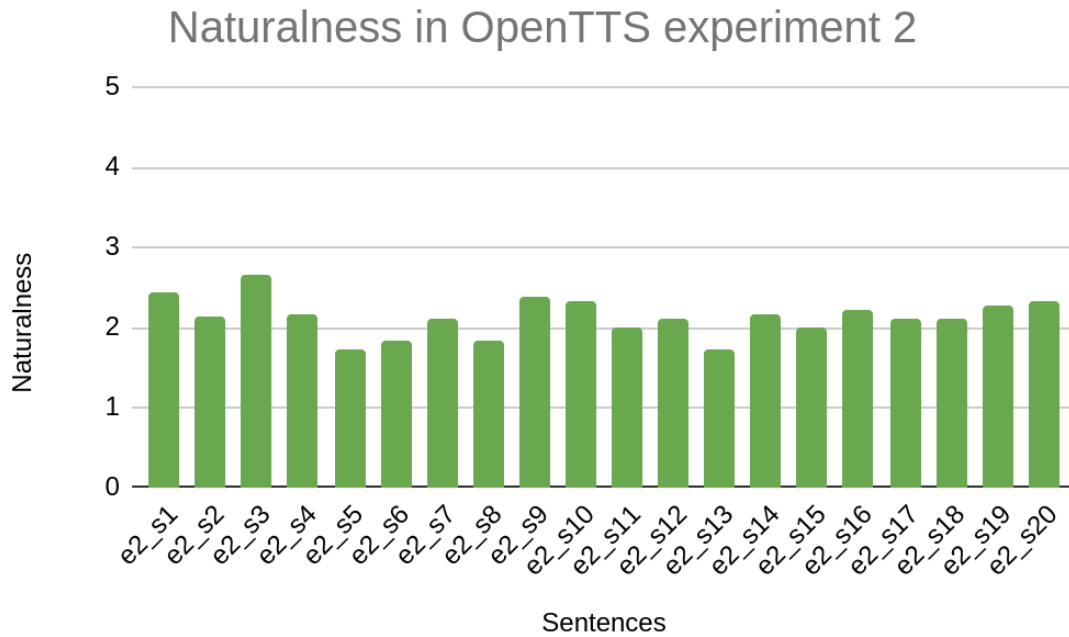*Figure 19: Expressiveness OpenTTS experiment 2*

Figure 20: Naturalness OpenTTS experiment 2

In general, we can observe a very steady pattern for all the sentences and all the criteria. According to intelligibility, the TTS ranges from a bit understandable to understandable. It is perceived as a bit expressive and the voice is considered robotic.

When asked whether they believe that OpenTTS could be used as a tool for the improvement of pronunciation for second language learners, more than half of the participants replied negatively.
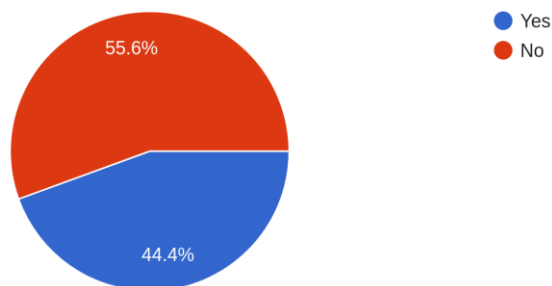

Figure 21: OpenTTS survey

After conducting the experiments and analyzing the results, we can see that OpenTTS might not be useful enough in helping second language learners of Bulgarian in improving their pronunciation. That is why I will not be implementing it in my demo (described later in point 5).

4.2. Analysis of ASR systems

4.2.1 Difficulties

A problem I encountered while searching for open-source ASR software supporting the Bulgarian language was that most high-quality and free systems do not support Bulgarian.[43]

I also checked all of the following resources - Dragon NaturallySpeaking, VoxSigma, Kaldi, CMUSphinx, Julius, Mozilla DeepSpeech (only possibility to donate), none of which support Bulgarian.[44] I only encountered a tool, which allows typing keys and mouse clicks by speaking into the microphone, but it is only for keyboard keys such as colon, period, shift, etc, as well as mouse clicks.[45]

Furthermore, I discovered iFLYTEK Open Platform, which is a quick start Chinese Artificial Intelligence open platform, which has the following features - TTS, ASR, and NLP SDKs, and seems to support the Bulgarian language.[46] Unfortunately, I read some negative reviews related to data privacy and chose not to work with it.

---

[43] *The Best 7 Free and Open Source Speech Recognition Software Solutions.* GoodFirms. (2020, January 28). https://www.goodfirms.co/blog/best-free-open-source-speech-recognition-software.

[44] Speech recognition software for Linux. (2021, March 13). *Wikipedia.* https://en.wikipedia.org/wiki/Speech_recognition_software_for_Linux#cite_note-9.

[45] *Configuration - Sphinx documentation. Sphinx.* (n.d.). https://www.sphinx-doc.org/en/master/usage/configuration.html.

[46] *Quick Start | iFLYTEK Open Platform Documents.* iFLYTEK. (n.d.). https://global.xfyun.cn/doc/platform/quickguide.html.

4.2.2 Methodology

The methodology of testing ASR systems involved two roles - a second language learner with no notions of Bulgarian, whose mother tongue is Spanish, as a student, and a native speaker as a teacher of Bulgarian. We performed two experiments testing ASR systems in order to see whether they can be used as a tool for second language learning (described in detail in sections 4.2.2.1 and 4.2.2.2).

After all the problems finding a free open-source software, I decided to perform the experiments using two ASR systems - SpeechTexter[47] and TalkTyper[48], which don't seem to be open-source but are available online for free.

For both experiments, we followed the same methodology. First of all, the second language learner recorded themselves saying the sentences in Bulgarian, and then the native speaker did that, as well. Afterwards, I ran the recordings on the ASR systems and wrote down the outputs.

Taking the original sentence and the ASR output, I ran a python script, which computed the Word Error Rate (WER) for both the outputs from her recordings and mine. Word Error Rate (WER) is a metric used to perform quantitative analysis of ASR systems, the formula is the following:

$$WER = \frac{substitutions\ +\ insertions\ +\ deletions}{total\ number\ of\ words}$$

Where substitutions are anytime a word gets replaced, insertions - anytime a word gets added that wasn't said, and deletions - anytime a word is omitted from the transcript.

---

[47] *SpeechTexter: Type with your voice online.* Speech Texter. (n.d.). https://www.speechtexter.com/.
[48] *TalkTyper - Speech Recognition in a Browser.* TalkTyper.com. (n.d.). https://talktyper.com/.

4.2.2.1 Experiment 1

The following chart depicts a plot of the Word Error Rates from experiment 1 - 20 words and 6 sentences (Tables 5 and 6 in section 4) for both ASRs for the second language learner and the native speaker's recordings. Positively, we can observe that most errors are from the second language learner's recordings, which was the expected outcome, as the other one is a native speaker. That shows that the ASR systems are accurate enough to be applied to second language learning.
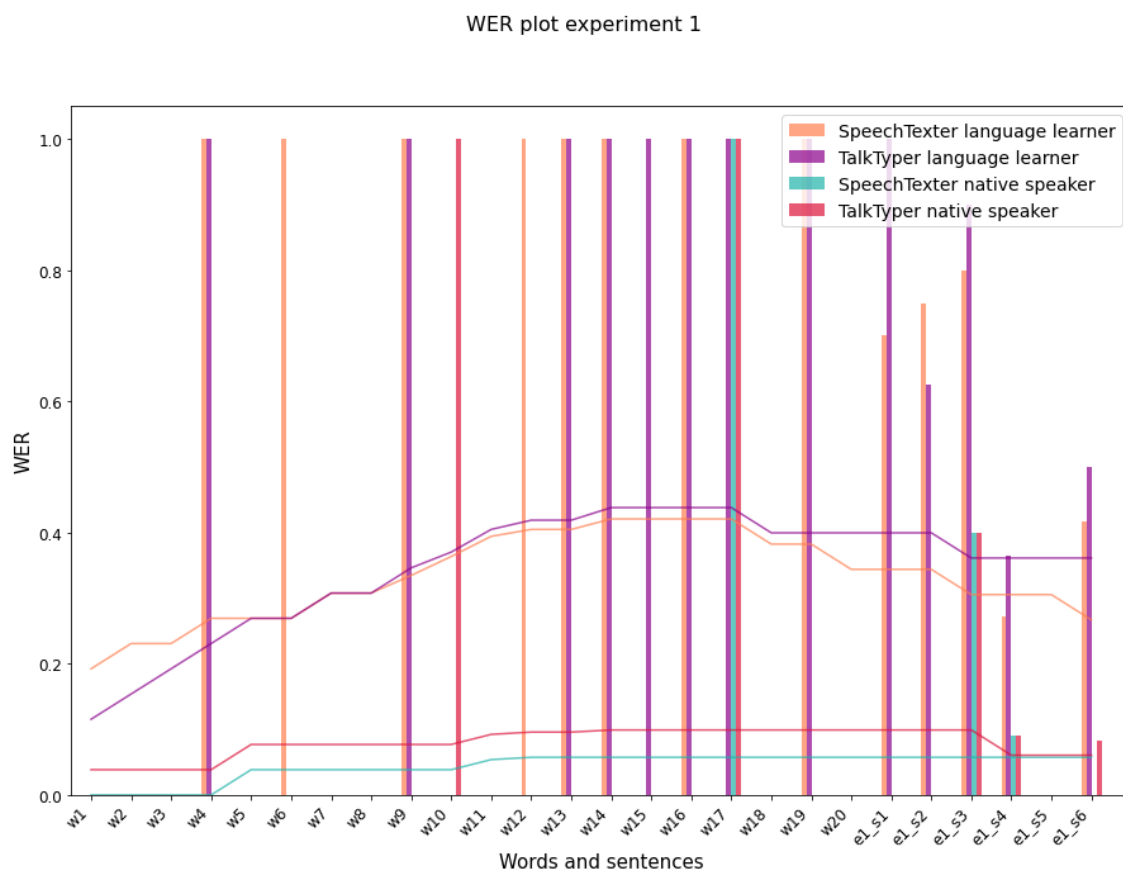

Figure 22: WER plot experiment 1

After computing the WER, I also evaluated the performance of the second language learner, to make sure that the automated evaluation was accurate.

When doing the recordings, the second language learner also rated the difficulty of the words and sentences on a 1 to 5 likert scale, and then I assessed them based on their

42

performance again on a 1 to 5 likert scale. The following table shows what each score means for each criteria:

Score	Difficulty	Assessment
1	very easy	very poor - many severe errors, or nothing is correct
2	easy	poor - some severe errors
3	moderate	average - there are 2 or more small errors
4	difficult	very good - there is a small error
5	very difficult	excellent

Table 11: Difficulty, assessment criteria

The following pie charts represent the distribution of difficulty and performance for experiment 1.
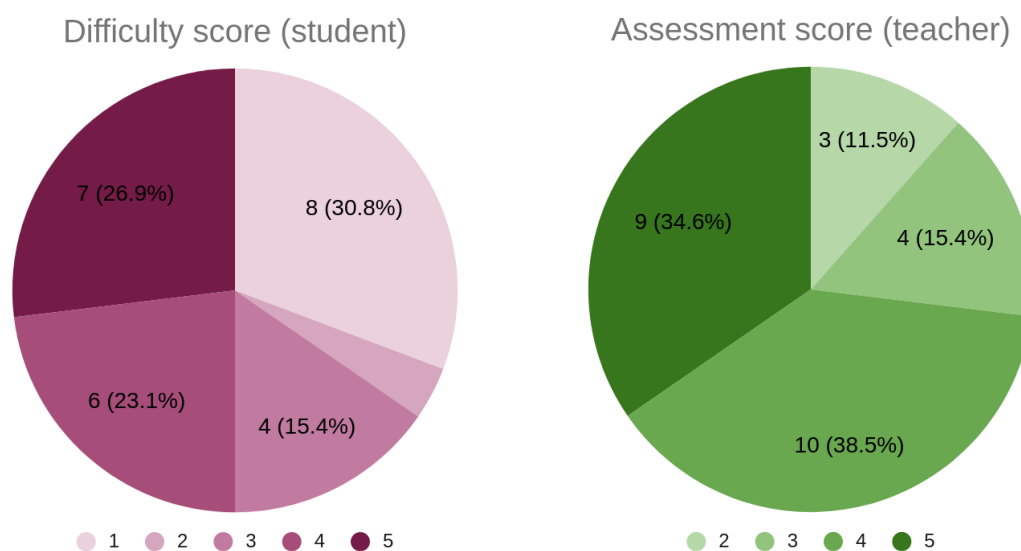

Figure 23: Pie charts - difficulty, assessment experiment 1

We can observe that half of the words and sentences the second language learner rated with a difficulty 4 or higher, which shows that the chosen target words and sentences seem to be complicated for a second language learner, which was also the wanted result.

Based on these two criteria, as well as the WER, I could evaluate which words and sentences required more practice, and which did not.

The following bar charts show the difficulty and the assessment for each word and sentence from experiment 1 (fig.24 and fig.25). In green are marked the ones, which do not require more practice, and in yellow those that do.
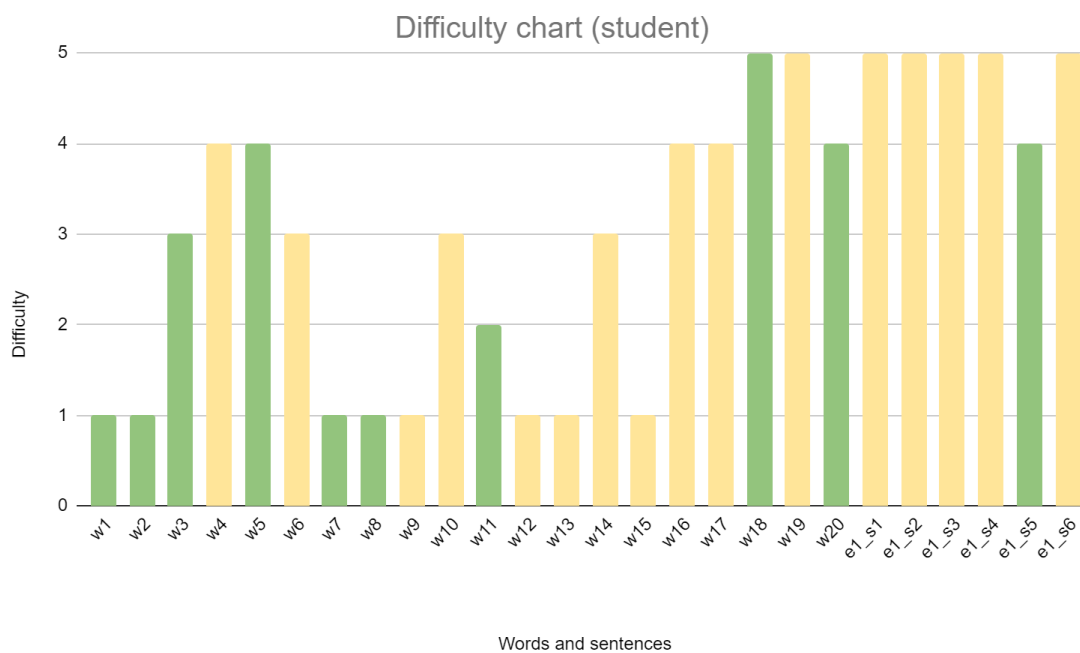

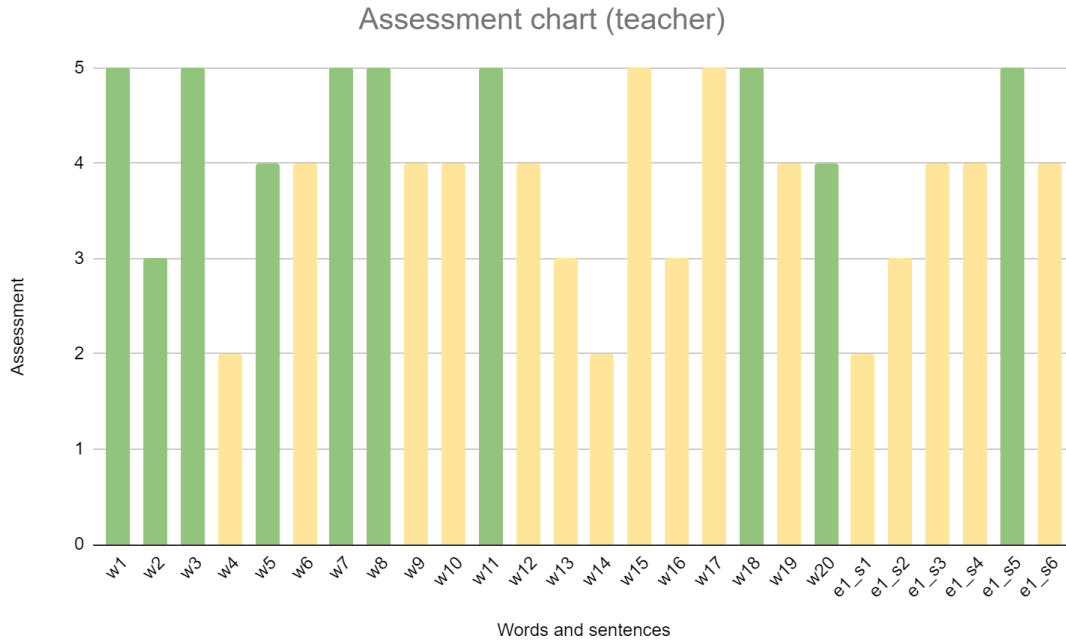*Figure 24: Difficulty chart experiment 1*

Figure 25: Assessment chart experiment 1

4.2.2.2 Experiment 2

Analogously to experiment 1 (section 4.2.2.1), the following chart depicts a plot of the Word Error Rates from experiment 2 - 20 sentences (Table 7 in section 4) for both ASRs for the second language learner and the native speaker's recordings. Again, we can observe that the majority of errors come from the second language learner's recordings (in orange and purple), which shows that the ASR systems can be applied to second language learning.
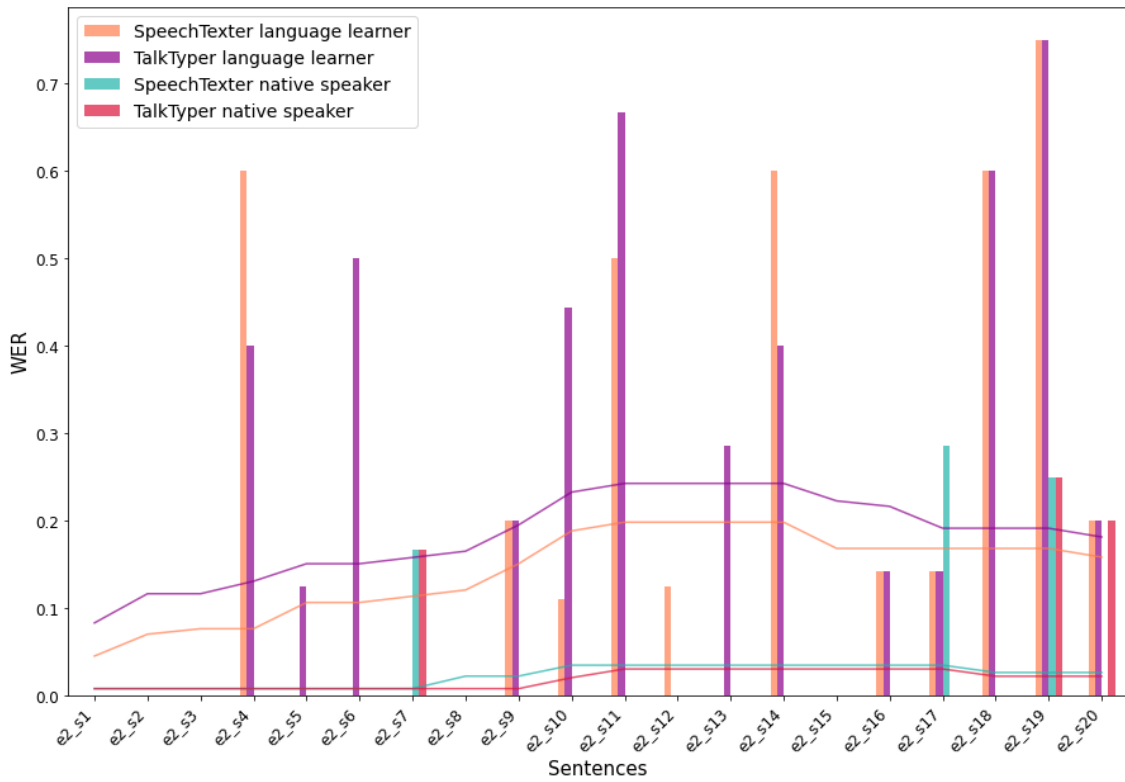
Figure 26: WER plot experiment 2

After computing the WER, I evaluated the second language learner's performance based on the same assessment scale (Table 11), to make sure that the automated evaluation was accurate.

The following pie charts represent the distribution of difficulty and performance for experiment 2.
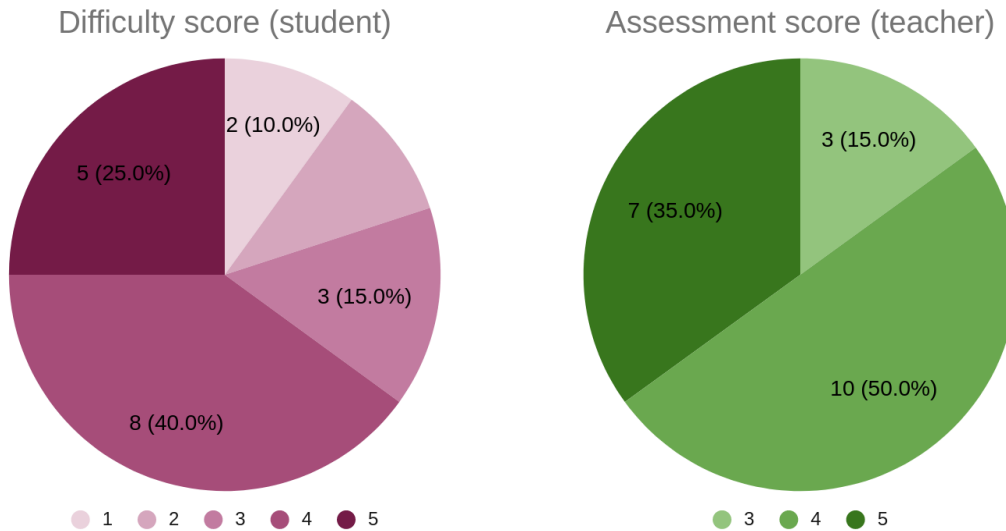
Figure 27: Pie charts - difficulty, assessment experiment 2

It can be seen that 65% of the sentences have a difficulty score of 4 or 5, which was again the wanted result. Even if the assessment score is high, the student would require more practice in order to feel more confident with the pronunciation of the target words and sentences. Again, based on these two criteria, and also the WER, I evaluated which sentences require more practice, and which do not.

The following bar charts show the difficulty and the assessment for each sentence from experiment 2 (fig.28 and fig.29). In green are marked the ones, which do not require more practice, and in yellow those that do.

We can observe that this time, when we have only target sentences and not a mix of words and sentences, only 25% of the sentences do not require to be practiced more and that it is related to their difficulty and WER being lower.
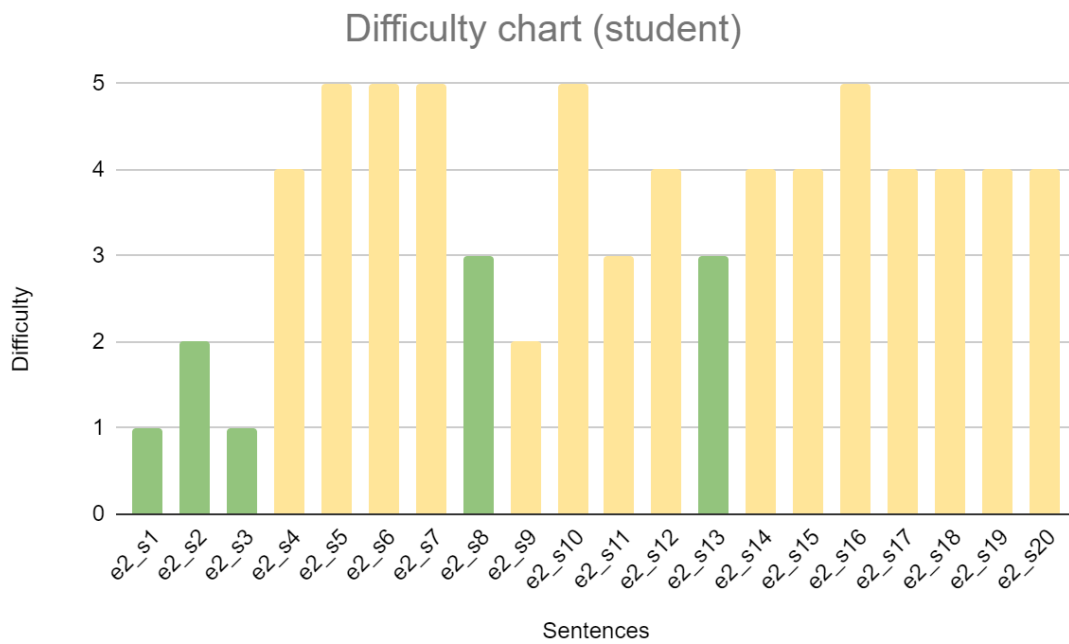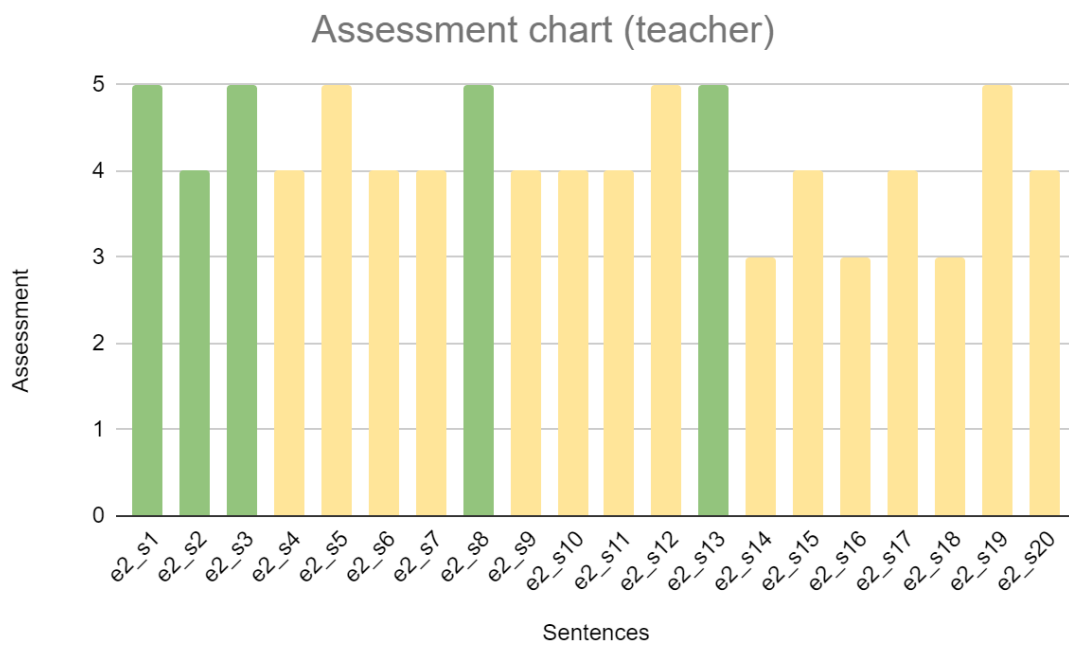
Figure 28: Difficulty chart experiment 2

Figure 29: Assessment chart experiment 2

4.2.3 Conclusions

After conducting the experiments, I analyzed the most common sources of errors, and they seem to come from consonants, which are not present in Latin languages, and also from accumulations of consonants. Stress is a common mistake, as well. In the table below, I have written down the problematic phoneme(s) - in Cyrillic, transliterated, and transcribed with IPA. In the second column are the target words from experiments 1 and 2 - in Bulgarian and their translations in English. And in the third column, I gathered new words, which target the specific problem, to be used in the logical algorithm (described in point 5) - again in Bulgarian and their English translations.

Problem (Cyrillic, transliteration, transcription(IPA))	Word from experiments Bulgarian (English translation)	New words Bulgarian (English translation)
ж, zh, [ʒ]	довиждане (goodbye), учреждение (establishment), държава (country)	Жарава (embers), жираф (giraffe), жълт (yellow), животно (animal), кръжа (hover/circle), пържола (steak), пържен (fried), дъжд (rain), чужд (foreign), ръжда (rust)
дж, dzh, [dʒ]	патладжан (eggplant), джудже (dwarf)	Джапанка (flip flop), джунгла (jungle), тенджера (pot), бояджия (dyer)
ч, ch, [t͡ʃ]	птицечовка (platypus), учреждение (establishment), ръчно (manual), сключвам (conclude)	Качвам (climb), тичам (run), капачка (cap), бръчка (wrinkle), проучвам (research)
ъ, u/a, [ɤ/ɐ]	триъгълник (triangle), блясък (shine/glow), ръчно (manual), държава (country),	Спътник (satellite), кътник (molar), ущърб (harm/detriment), дъжд (rain)

	дрънкулка (trinket)	
нн, nn, [nn]	стра**нн**ик (stranger)	съвреме**нно** (contemporary), есе**нно** (autumnal), неви**нн**ост (innocence)
чкв, chkv, [t͡ʃkv]	спре**чкв**ане (argument)	сма**чкв**ам (crumple), сбръ**чкв**ам (wrinkle), налу**чкв**ам (guess)
взр, vzr, [vzr]	**взр**ив (explosion)	**взр**ях (stared), **взр**ивен (explosive), не**взр**ачен (plain)
х (отпред), h (in front), [x]	**х**ора (people)	**х**ляб (bread), **х**авлия (bathrobe), **х**обот (trunk), **х**оро (horo/round dance)
бл, bl, [bl]	**бл**агодаря (thank you)	**бл**изък (close), **бл**ок (), **бл**ажен, **бл**ясък
здр, zdr, [zdr]	**здр**авей (hello)	**здр**авец (geranium), на**здр**аве (cheers), по**здр**ав (greeting)
ц, tz/ts, [t͡s]	**ц**ветарни**ц**а (flower shop)	**ц**ари**ц**а, пар**ц**ал, кори**ц**а, кошни**ц**а
ш, sh, [ʃ]	кру**ш**а (pear)	**ш**унка (ham), **ш**и**ш**е (bottle), па**ш**кул (cocoon), ко**ш**ница (basket)
щ, sht, [ʃt]	**щ**ъркел (stork), пло**щ**ад (square)	къ**щ**а (house), **щ**астие (happiness), по**щ**альон (postman), кръ**щ**ене (baptism)
ств, stv, [stv]	лекар**ств**о (medicine)	**ств**ол (trunk), цар**ств**о(kingdom), приятел**ств**о (friendship)
ударение/accent	дов'иждане (dov'zhdane), цвет'арница (tsvet'arnitsa), джудж'е (dzhudzh'e), бл'ясък (bl'yasak), яд'е (yad'e), пек'арната (pek'arnata), едн'а (edn'a), х'ората (h'orata),	

50

	лек'арства (lek'arstva), държ'авата (darzh'avata)	

Table 12: Sources of errors

From the new words gathered, I created new simple sentences, which target each word specifically. They are implemented in the logical algorithm described in point 5.

5. Proof of concept - Logical algorithm

Based on the experiments conducted, I designed an algorithm, which could help second language learners of Bulgarian with their pronunciation of Bulgarian. The general idea is that it is a web page, running on Flask[49], where the student can practice their pronunciation. For the moment it is only possible to run it locally. The following image shows the home page.
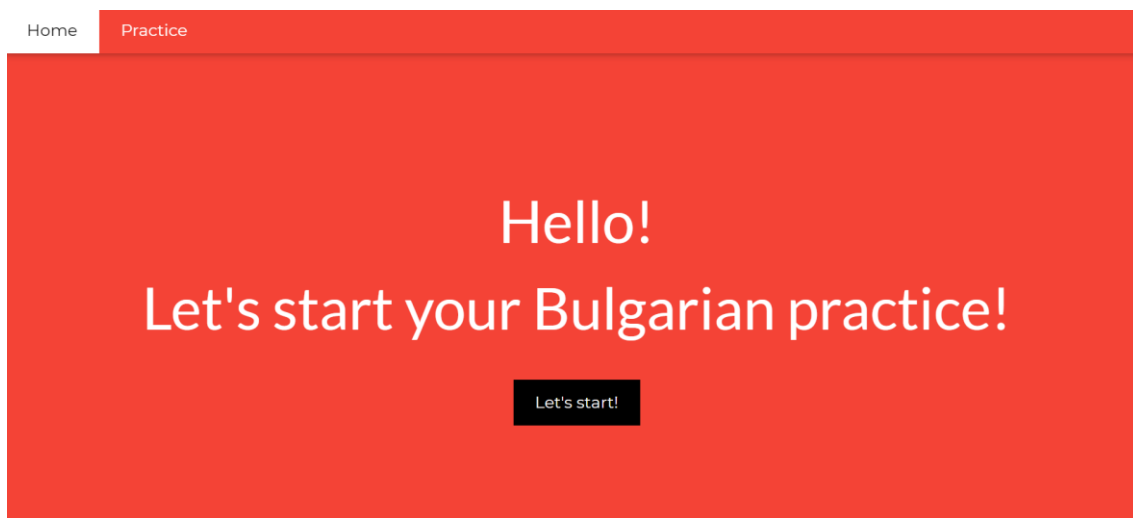

Figure 30: Home page demo

After clicking the 'Let's start!' button, the users sees the following page:
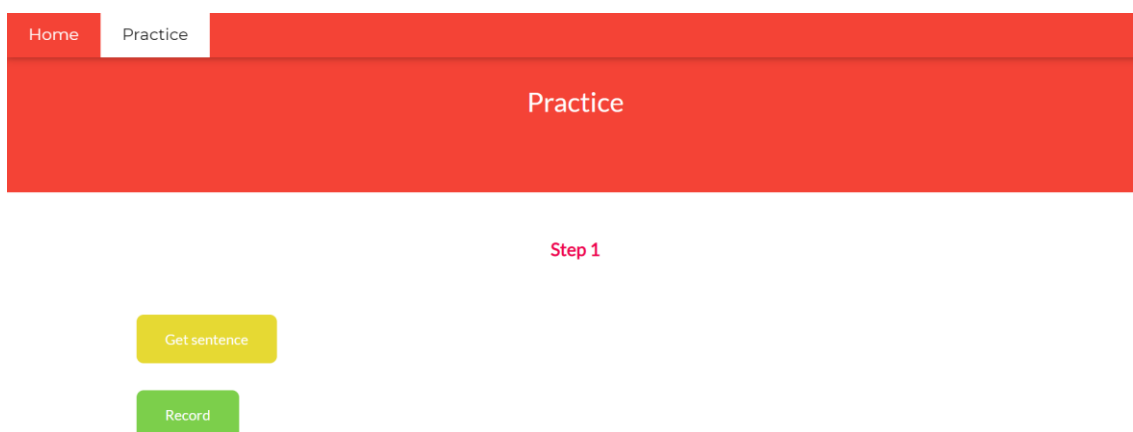

Figure 31: Step 1 demo

---

[49] *Welcome to Flask* . Flask. (n.d.). https://flask.palletsprojects.com/en/2.0.x/.

On this page, the student can click a button 'Get sentence', which randomly selects a target word and also a target sentence, containing the specific word. The target word and sentence are displayed together with their transliteration and translation in English. To transliterate, I used a python library called 'transliterate'.[50] I also tried using a library for the Bulgarian-English translation, which is based on Google translate called 'googletrans'[51], but it did not translate some of the sentences very accurately, that is why I translated them myself.

Figure 32: Step 1 demo sentence displayed

Then the student records themselves saying the sentence by clicking the button 'Record', which records for 10 seconds. The .wav file from the recording is saved in the folder from where the code is running.

After that they are redirected to another page, where they have to choose the file containing the recording. The student also has to select a number from 1 to 5 on how difficult they found the sentence, based on the scale described in section 4.

---

[50] *transliterate*. PyPI. (n.d.). https://pypi.org/project/transliterate/.

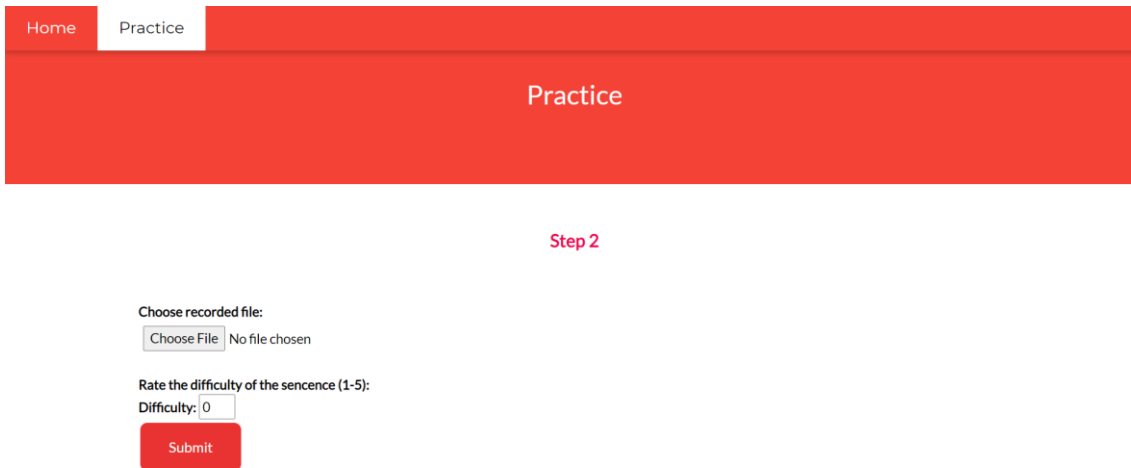[51] *googletrans*. PyPI. (n.d.). https://pypi.org/project/googletrans/.

Figure 33: Step 2 demo

Then Google Cloud's Speech-to-Text system gets the output from the recording.[52] The reason I used different tools for the experiments and the demo is that Google STT can be used for free for a limited number of minutes, which by my calculations was not enough to conduct both my experiments and design my demo. On the other hand, I could not implement SpeechTexter and TalkTyper (the tools used for the experiments described in sections 4.2.2.1 and 4.2.2.2) into the web demo as they are not open-source.

Afterwards, the WER is computed by comparing the target sentence to the ASR output. To compute the WER, I used a python library called 'jiwer'.[53]

Proceeding from experiments 1 and 2 (sections 4.2.2.1 and 4.2.2.2), I figured a simple logic, based on the WER and the difficulty score to decide whether the student requires more practice on the phoneme that the target word and sentence target. In case that the difficulty is 4 or higher, they would better practice more until they do not find the target phoneme as challenging, regardless of the Word Error Rate. And in the other case, where the difficulty is equal to 3 or below, the algorithm would also check whether the WER is higher or lower than 0.2. If it is higher, then the student would have to practice more:

---

[52] Google. (n.d.). *Speech-to-Text: Automatic Speech Recognition | Google Cloud*. Google. https://cloud.google.com/speech-to-text.

[53] *jiwer*. PyPI. (n.d.). https://pypi.org/project/jiwer/.

Figure 34: Step 3 demo - more practice necessary

By clicking 'Continue', they go back to Step 1, where by pressing 'Get sentence' they get a target word and sentence from the same target group as the previous one:

Figure 35: Step 1 demo - word and sentence from the same target group

If the WER is lower than 0.2, and the difficulty selected is smaller than 4, then more practice on this target group in not necessary:

Practice

Step 3

Well done!

Original sentence: Триъгълникът е една от основните форми в математиката.
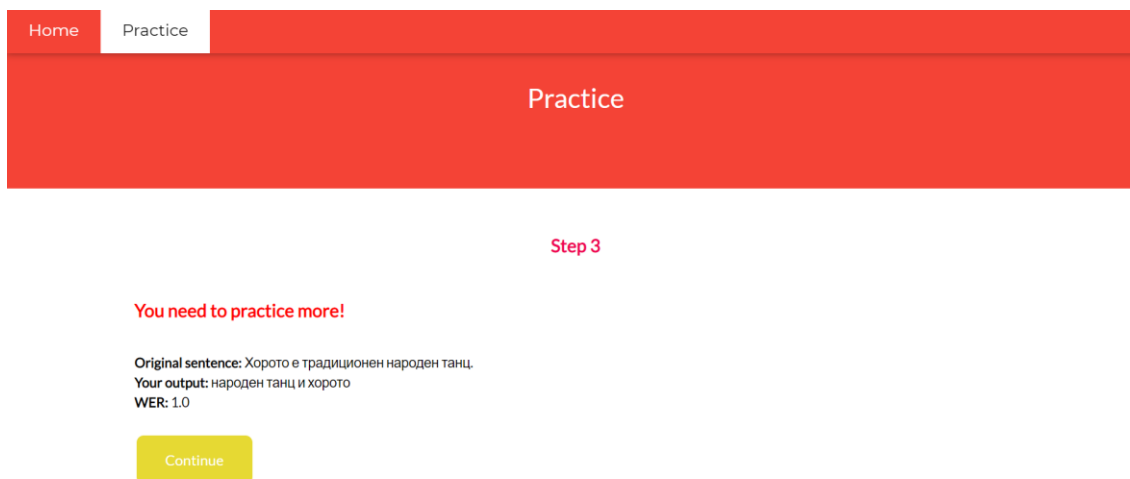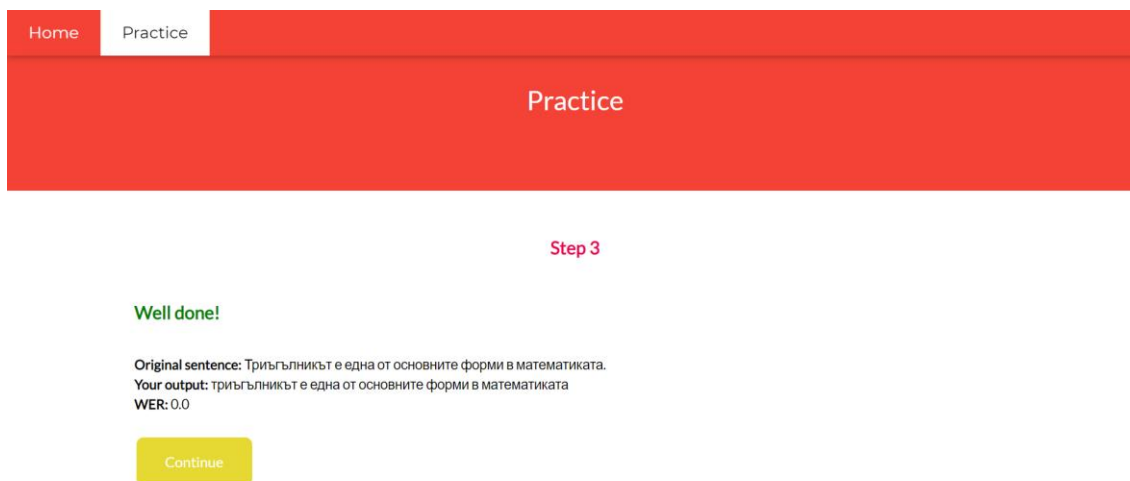**Your output:** триъгълникът е една от основните форми в математиката
WER: 0.0

Continue

Figure 36: Step 3 demo - more practice not necessary

By pressing 'Continue', they go back to Step 1, where by pressing 'Get sentence' the algorithm randomly selects another target word and sentence, which is focused on practicing another challenging phoneme:

Step 1

Get sentence

Target word: триъгълник
Transliteration: triagalnik
Translation: triangle

Target sentence: Триъгълникът е една от основните форми в математиката.
Transliteration: Triagalnikat e edna ot osnovnite formi v matematikata.
Translation: The triangle is one of the main shapes in mathematics.
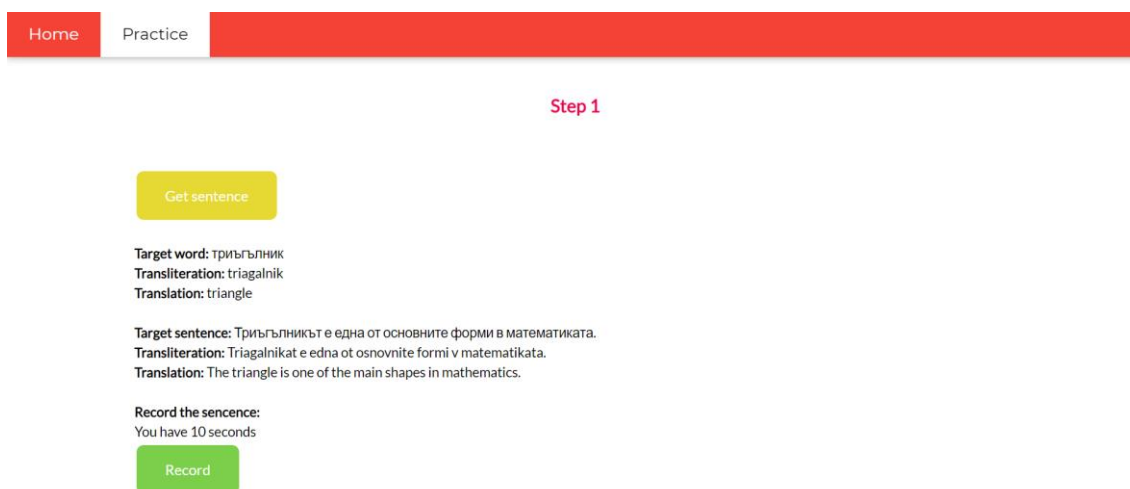
Record the sencence:
You have 10 seconds

Record

Figure 37: Step 1 demo - new random word and sentence

In the following links can be seen demonstrations on two scenarios:
1) When the student requires more practice on the target phoneme - video.
2) When more practice on the target phoneme is not necessary - video.

6. Conclusions

The aim of this thesis was to research and evaluate the available speech technologies in Bulgarian, and more precisely Text-to-Speech and Automatic-Speech-Recognition systems, and how they could be applied to second language learning of Bulgarian. It also targets a specific problem for second language learners of Bulgarian, which is the pronunciation of certain challenging phonemes.

The perception test of TTS systems showed that they could be used as a tool to aid second language learners of Bulgarian to master their pronunciation. But for the moment open-source systems, such as OpenTTS do not seem to be promising for the purpose.

The experiments and evaluation of ASR systems also demonstrated that this technology can be implemented into a system for pronunciation improvement of second language learners of Bulgarian. It has been confirmed through the demo, as well.

To conclude, I am extremely pleased that speech technologies could be applied in a way that my mother tongue could be more accessible to non-native speakers of Bulgarian, who have the desire to learn but are lacking the means.

7. Future work

In the future, the demo could be further developed into a web page, which is available online, and also into an application. In that case, to securely handle voice data from users, the Legal and Contractual obligations under the Terms and Conditions of third parties should be taken into consideration.[54]

The application could also implement a high-quality Text-to-Speech system, which could serve as a tool to demonstrate to the learners how the sentence should be pronounced correctly. Moreover, TTS and ASR could be used to develop a bot with which the second language learner could conversate.

Furthermore, an algorithm could be implemented to select words and generate sentences automatically. The identification of errors could also be done through an algorithm, which would detect the specific problems of the learner. As well, depending on certain needs of the user, the application could be more personalized by targeting particular problems that they have in terms of pronunciation. This part would be done in collaboration with a Bulgarian language teacher.

Later, the application could also be developed for other languages, for which such functionalities do not exist yet. As well, it could offer the possibility to learn through other languages other than English.

---

[54] B., R. (2020, February 19). *Voice Assistants and Privacy Issues*. TermsFeed.
https://www.termsfeed.com/blog/voice-assistants-privacy-issues/.

References

Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies*, 1-13. https://doi.org/10.1007/s10639-019-09955-7.

Bione, T., Grimshaw, J., & Cardoso, W. (2016). An evaluation of text-to-speech synthesizers in the foreign language classroom: learners' perceptions. https://files.eric.ed.gov/fulltext/ED572021.pdf.

Carrier, M. (2017). Automated Speech Recognition in language learning: Potential models, benefits and impact. https://rudn.tlcjournal.org/archive/1(1)/1(1)-03.pdf.

Cyrillic alphabets. (2021, May 27). *Wikipedia*. https://en.wikipedia.org/wiki/Cyrillic_alphabets.

Hashmi, N. (2016). Computer-Assisted Language Learning (CALL) in the EFL Classroom and its Impact on Effective Teaching-learning Process in Saudi Arabia. *International Journal of Applied Linguistics and English Literature, 5*, 202-206. https://www.journals.aiac.org.au/index.php/IJALEL/article/view/2152.

Hateva, N., Mitankin, P., & Mihov, S. (2016). BulPhonC: Bulgarian Speech Corpus for the Development of ASR Technology. *LREC*. https://www.aclweb.org/anthology/L16-1123.pdf.

Huang, Y., & Liao, L. (2015). A STUDY OF TEXT-TO-SPEECH (TTS) IN CHILDREN'S ENGLISH LEARNING. *Teaching english with technology, 15*, 14-30. https://files.eric.ed.gov/fulltext/EJ1140575.pdf.

Junining, E., Alif, S., & Setiarini, N. (2020). Automatic speech recognition in computer-assisted language learning for individual learning in speaking. https://journal.umsida.ac.id/index.php/jees/article/view/867/1083.

Kholis, A. (2021). Elsa Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills. *Pedagogy : Journal Of English Language Teaching, 9*(1), 01-13. doi:10.32332/joelt.v9i1.2723

Li, K., Qian, X., & Meng, H. (2017). Mispronunciation Detection and Diagnosis in L2 English Speech Using Multidistribution Deep Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25*, 193-207. https://ieeexplore.ieee.org/document/7752846.

Meihami, H., & Husseini, F. (2014). BRINGING TTS SOFTWARE INTO THE CLASSROOM: THE EFFECT OF USING TEXT TO SPEECH SOFTWARE IN TEACHING READING FEATURES. *Teaching english with technology, 14*, 23-34. https://files.eric.ed.gov/fulltext/EJ1143397.pdf.

National Geographic България. (2019, May 24). Глаголица и кирилица. National Geographic България. https://www.nationalgeographic.bg/a/glagolica-i-kirilica.

Pirasteh, P. (2014). The Effectiveness of Computer-assisted Language Learning (CALL) on Learning Grammar by Iranian EFL Learners. *Procedia - Social and Behavioral Sciences, 98*, 1422-1427. https://core.ac.uk/download/pdf/82473139.pdf.

Speech synthesis. (2021, June 9). *Wikipedia.* https://en.wikipedia.org/wiki/Speech_synthesis.

Tanya. (2020, June 22). *How to Learn the Bulgarian Language and Cyrillic Alphabet.* All Language Resources. https://www.alllanguageresources.com/learn-bulgarian-language/.

Кирил и Методий. (2021, May 28). *Wikipedia*. https://bg.wikipedia.org/wiki/%D0%9A%D0%B8%D1%80%D0%B8%D0%BB_%D0%B8_%D0%9C%D0%B5%D1%82%D0%BE%D0%B4%D0%B8%D0%B9

Мишев, М. (2021, April 29). *10 любопитни факта за кирилицата*. Българска история. https://bulgarianhistory.org/kirilitza/.

Население на България. (2021, June 2). *Wikipedia.* https://bg.wikipedia.org/wiki/%D0%9D%D0%B0%D1%81%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%BD%D0%B0_%D0%91%D1%8A%D0%BB%D0%B3%D0%B0%D1%80%D0%B8%D1%8F

Footnotes

1. National Geographic България. (2019, May 24). Глаголица и кирилица. National Geographic България. https://www.nationalgeographic.bg/a/glagolica-i-kirilica.

2. Britannica, T. Editors of Encyclopaedia (n.d.). Cyrillic alphabet. Encyclopedia Britannica. https://www.britannica.com/topic/Cyrillic-alphabet.

3. Cyrillic script. (2021, June 10). *Wikipedia.* https://en.wikipedia.org/wiki/Cyrillic_script.

4. Кирилица. (2021, May 27). *Wikipedia.* https://bg.wikipedia.org/wiki/%D0%9A%D0%B8%D1%80%D0%B8%D0%BB%D0%B8%D1%86%D0%B0#%D0%A0%D0%B0%D0%B7%D0%BF%D1%80%D0%BE%D1%81%D1%82%D1%80%D0%B0%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B8_%D1%80%D0%B0%D0%B7%D0%BD%D0%BE%D0%B2%D0%B8%D0%B4%D0%BD%D0%BE%D1%81%D1%82%D0%B8.

5. Iliev, I. G. (2013, February). SHORT HISTORY OF THE CYRILLIC ALPHABET. IJORS International Journal of Russian Studies. http://www.ijors.net/issue2_2_2013/articles/iliev.html.

6. Cyrillic language alphabets and how they diverge from one another. Yale University Library. (n.d.). https://web.library.yale.edu/cataloging/music/cyrillic.

7. Jakobson, R. (2018). In Remarks on the phonological evolution of Russian in comparison with the other Slavic languages (p. 175). essay, The MIT Press.

8. Rieder-Bünemann A. (2012) Second Language Learning. In: Seel N.M. (eds) Encyclopedia of the Sciences of Learning. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1428-6_826

9. Transliteration. (2021, May 22). *Wikipedia*. https://en.wikipedia.org/wiki/Transliteration.

10. Encyclopædia Britannica, inc. (n.d.). Phoneme. Encyclopædia Britannica. https://www.britannica.com/topic/phoneme.

11. Transcription, Pronunciation and Translation of English Words. Myefe. (2021, March 1). https://myefe.com/transcription-pronunciation.

12. Encyclopædia Britannica, inc. (n.d.). International Phonetic Alphabet. Encyclopædia Britannica. https://www.britannica.com/topic/International-Phonetic-Alphabet.

13. Българска азбука. (2021, June 7). *Wikipedia*. https://bg.wikipedia.org/wiki/%D0%91%D1%8A%D0%BB%D0%B3%D0%B0%D1%80%D1%81%D0%BA%D0%B0_%D0%B0%D0%B7%D0%B1%D1%83%D0%BA%D0%B0.

14. Bulgarian (Български). Omniglot.com. (2021, April 23). https://omniglot.com/writing/bulgarian.htm.

15. Learn the Bulgarian pronunciation. coLanguage. (n.d.). https://www.colanguage.com/learn-bulgarian-pronunciation.

16. Букви и звукове в българския език. (2020, October 19). Wikipedia. https://bg.wikipedia.org/wiki/%D0%91%D1%83%D0%BA%D0%B2%D0%B8_%D0%B8_%D0%B7%D0%B2%D1%83%D0%BA%D0%BE%D0%B2%D0%B5_%D0%B2_%D0%B1%D1%8A%D0%BB%D0%B3%D0%B0%D1%80%D1%81%D0%BA%D0%B8%D1%8F_%D0%B5%D0%B7%D0%B8%D0%BA.

17. *МЕТОДИКА НА ОБУЧЕНИЕТО ПО БЪЛГАРСКИ ЕЗИК ЗА МИГРАНТИ*. (n.d.). https://download.ei-ie.org/Docs/WebDepot/SEB%20Handbook.pdf.

18. *Трудности при овладяване на българската фонетична система*. elearn.uni-sofia. (n.d.). https://elearn.uni-sofia.bg/mod/resource/view.php?id=13605.

19. Как звучи българският език на чужденците? Omega LS. (2019, September 30). https://omegals.bg/kak-zvuchi-bulgarskiqt-ezik-na-chuzdencite/.

20. Davies, R. (2015, September 28). Basic Bulgarian, Pronunciation - Consonants, Round 1. Duolingo. https://forum.duolingo.com/comment/10741347/Basic-Bulgarian-Pronunciation-Consonants-Round-1.

21. Bulgarian phonology. (2021, June 10). *Wikipedia.* https://en.wikipedia.org/wiki/Bulgarian_phonology.

22. Innovative Language Learning. (2014). Top 5 Tips for Avoiding Common Mistakes in Bulgarian. In *Learn Bulgarian - Level 1 Introduction to Bulgarian, Volume 1: Volume 1: Lessons 1-25*. essay.

23. Contributor, T. T. (2019, February 14). *What is speech technology?* SearchUnifiedCommunications. https://searchunifiedcommunications.techtarget.com/definition/speech-technology.

24. Speech synthesis. (2021, June 9). *Wikipedia.* https://en.wikipedia.org/wiki/Speech_synthesis.

25. *What is E-learning? Definition of E-learning, E-learning Meaning*. The Economic Times. (n.d.). https://economictimes.indiatimes.com/definition/e-learning.

26. Vijaya, Samudra K. (2017, November) *Automatic Speech Recognition.* http://www.iitg.ac.in/samudravijaya/tutorials/asrTutorial.pdf

27. Acoustic model. (2020, January 4). *Wikipedia.* https://en.wikipedia.org/wiki/Acoustic_model.

28. Levy M. (1997) *CALL: context and conceptualisation*, Oxford: Oxford University Press.

29. Computer-assisted language learning. (2021, May 7). *Wikipedia.* https://en.wikipedia.org/wiki/Computer-assisted_language_learning.

30. Schumer, L. (2021, March 24). *9 Best Language Apps for Learning on the Go*. Good Housekeeping. https://www.goodhousekeeping.com/life/g32175725/best-language-learning-apps/.

31. Meredithkreisa. (2021, January 18). *6 Language Apps That Use Speech Recognition for Well-rounded Learning*. FluentU Language Learning. https://www.fluentu.com/blog/speech-recognition-language-learning/.

32. *Learn Bulgarian Online in Just 10 Minutes a Day*. Mondly Blog. (2020, September 11). https://www.mondly.com/blog/2020/09/11/learn-bulgarian-online/.

33. *Mondly VR Is Now Available on Steam*. Mondly Blog. (2020, June 9). https://www.mondly.com/blog/2019/09/25/mondly-learn-languages-in-vr-is-now-available-on-steam/.

34. *Play your way to a new language with Mondly*. Mondly. (n.d.). https://www.mondly.com/ph.

35. *Bulgarian Language with a Free App*. BulgarianPod101. (n.d.). https://www.bulgarianpod101.com/app/.

36. *Learn Bulgarian - Free, Fast & Effective*. FunEasyLearn. (n.d.). https://www.funeasylearn.com/learn-bulgarian.

37. *Top 10 Hardest Bulgarian Words to Pronounce*. BulgarianPod101. (n.d.). https://www.bulgarianpod101.com/bulgarian-vocabulary-lists/top-10-hardest-words-to-pronounce.

38. *Text-to-Speech (TTS) Engine in 119 Voices: Nuance: Nuance*. Nuance Communications. (n.d.). https://www.nuance.com/omni-channel-customer-engagement/voice-and-ivr/text-to-speech.html#.

39. Synesthesiam. (n.d.). *synesthesiam/opentts*. GitHub. https://github.com/synesthesiam/opentts.

40. *intelligibility*. Cambridge Dictionary. (n.d.). https://dictionary.cambridge.org/dictionary/english/intelligibility.

41. Lexico Dictionaries. (n.d.). *Definition of EXPRESSIVENESS*. Lexico Dictionaries | English. https://www.lexico.com/definition/expressiveness.

42. Lexico Dictionaries. (n.d.). *Definition of NATURALNESS*. Lexico Dictionaries | English. https://www.lexico.com/definition/naturalness.

43. *The Best 7 Free and Open Source Speech Recognition Software Solutions*. GoodFirms. (2020, January 28). https://www.goodfirms.co/blog/best-free-open-source-speech-recognition-software.

44. Speech recognition software for Linux. (2021, March 13). *Wikipedia.* https://en.wikipedia.org/wiki/Speech_recognition_software_for_Linux#cite_note-9.

45. *Configuration - Sphinx documentation*. Sphinx. (n.d.). https://www.sphinx-doc.org/en/master/usage/configuration.html.

46. *Quick Start | iFLYTEK Open Platform Documents*. iFLYTEK. (n.d.). https://global.xfyun.cn/doc/platform/quickguide.html.

47. *SpeechTexter: Type with your voice online*. Speech Texter. (n.d.). https://www.speechtexter.com/.

48. *TalkTyper - Speech Recognition in a Browser*. TalkTyper.com. (n.d.). https://talktyper.com/.

49. *Welcome to Flask* . Flask. (n.d.). https://flask.palletsprojects.com/en/2.0.x/.

50. *transliterate*. PyPI. (n.d.). https://pypi.org/project/transliterate/.

51. *googletrans*. PyPI. (n.d.). https://pypi.org/project/googletrans/.

52. Google. (n.d.). *Speech-to-Text: Automatic Speech Recognition | Google Cloud. Google.* https://cloud.google.com/speech-to-text.

53. *jiwer*. PyPI. (n.d.). https://pypi.org/project/jiwer/.

54. B., R. (2020, February 19). *Voice Assistants and Privacy Issues*. TermsFeed. https://www.termsfeed.com/blog/voice-assistants-privacy-issues/.