GRAFFITI OR GRAVITY? MISPRONUNCIATION AT ITS BEST

Improving English Pronunciation through the language learning app ELSA (English Language Speech Assistant)

Word count: 17,154

Femke De Vrieze
Student number: 01903505

Supervisor: Prof. Dr. Orphée De Clercq

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Arts in Multilingual Communication

Academic year: 2022 – 2023

Abstract

Technology has become indispensable in today's digital society and hence also in language learning contexts. Many teachers have already implemented certain technological tools to help students improve their reading, writing, listening and speaking skills. This study aims to understand whether the Automatic Speech Recognition (ASR) system incorporated in the English Language Speech Assistant (ELSA) app is capable of improving students' speaking skills. Instead of only trying to improve pronunciation during class hours, the app would allow them to practice pronunciation whenever and wherever they want. To test this, a group of sixteen learners was evaluated in terms of overall improvement and in terms of improvement in the sounds that are particularly difficult for speakers of Dutch. The present study followed an experimental design and randomly divided the participants into an experimental group and a control group. The experimental group partook in a pre-test, a training period of one to two weeks in which they were asked to use the app and a post-test. The control group only partook in the pre- and post-test. Afterwards, the experimental group was also asked to fill in a short questionnaire about their attitudes and perceptions towards the ELSA app. The findings that were obtained by performing a paired samples t-test, a Wilcoxon signed-rank test and a Chi-square test were accomplished through SPSS Statistics 28 and can be linked to multiple causal factors. It would be interesting to take these causal factors into consideration in the future in order to improve the ELSA app and help students improve their pronunciation individually.

- 256 words -

Preface

After writing my bachelor's thesis on how ASR-systems are capable of providing both implicit and explicit feedback on a student's pronunciation and grammar, I quickly realised I wanted to pursue this subject in greater detail for my master's thesis. However, to avoid making my master's thesis too elaborate, I decided to focus solely on pronunciation. Therefore, I was rather fascinated when I came across an app called ELSA, which claimed it could improve people's English pronunciation. I wanted to investigate whether the creators of the app were truthful and for that purpose, I decided to set up an experiment. Personally, I found it rather challenging to find participants who were interested in my experiment and were willing to take the time to take part. Additionally, I also found it quite difficult and time-consuming to interpret the results correctly, but in the end, everything fell into place.

To successfully complete this thesis, I received an incredible amount of support from many people who wanted to see me succeed. First of all, I am very grateful to my supervisor, Prof. Dr. Orphée De Clercq, who guided me in the right direction whenever I felt a little lost. Her kind words and her willingness to answer all my questions and give me adequate feedback, helped me get back on my feet. I would also like to thank her for guiding me towards guidance counsellor Dr. Klaar Vanopstal and other services provided by the University of Ghent. I appreciate the help these services were able to provide whenever I felt like giving up.

In terms of the actual experiment, I would like to thank my brother and his friend for letting me record their voices as the example pronunciation for the sentences in my pre- and post-test design. I would also like to thank Mr. David Chan for his willingness to evaluate my participant's pronunciation for the pre- and post-test design. I would also like to thank Ms. Kloet, one of my teachers in secondary school, for allowing me to find participants in her classroom and for helping me motivate those participants to partake in my experiment. Lastly, I am, of course, very grateful to my participants, who wanted to see this through to the end, donating their valuable time to participate in this experiment.

Of course, I cannot forget my friends and family, without whom I would not have been able to finish this project. Thanks to their kind words of encouragement and heart-warming support, I found myself being able to do what I did not think possible.

Femke De Vrieze

May 2023

Table of contents

- 17,154 words -

List of abbreviations and symbols

AI   Artificial Intelligence

ASR   Automatic Speech Recognition

CALL   Computer Assisted Language Learning

CAPT   Computer Assisted Pronunciation Training

EFL   English as a Foreign Language

ELSA   English Language Speech Assistant

ESL   English as a Second Language

IT   Information Technology

TTS   Text-To-Speech

List of tables and figures

Chapter 1: Introduction

In today's digital society, technology has become an indispensable part of our lives. We tend to use it almost every day, either for work or as a means of relaxation and distraction. Also, schools are implementing certain types of technology, such as iPads and projectors, into their daily routine. In Flanders, for instance, the government is starting to implement Information Technology (IT) equipment for both students and teachers, who will also be taught how to use this particular equipment (Flemish Ministry of Education and Training, 2023). The question remains whether schools should take it one step further by implementing tools to aid language learning in particular. In recent years, Computer-Assisted Language Learning (CALL) has become a widely studied research field, mainly because of the rapid developments in technology. This technological evolution has also led to the use of new instruments in language learning classrooms, such as Grammarly[1], Quizlet[2], Kahoot![3] and many others. These instruments can be used for all four skills of the language learning process: reading, writing, listening and even speaking. The focus of this study will be on the latter of those skills. To become a proficient speaker in a certain language, a learner should properly master the following four skills: vocabulary, pronunciation, fluency and grammar (Binus University, 2018; Derakhshan et al., 2016; Twinkl, 2022). First of all, vocabulary is essential to understand certain concepts and to express yourself in the desired language. Furthermore, mistakes in pronunciation or disfluency can cause misunderstanding between two speakers. Especially in cases with minimal pairs in which two words only differ from each other in one single sound, poor pronunciation tends to lead to misunderstanding (Pourhosein Gilakjani, 2011; Pourhosein Gilakjani & Sabouri, 2016). To be fluent, on the other hand, is mostly a matter of being confident in the way you say something (BBC, n.d.). If you are not confident enough, you will give the impression of disfluency. Lastly, grammar may not seem very important during face-to-face conversations, but if someone does not understand how to use certain tenses or how to structure a sentence, it may become difficult for the listener to comprehend what the speaker is saying.

---

[1] Grammarly is a tool that can be used to correct grammar, spelling and punctuation. It is also able to suggest alternatives in order to improve a person's writing skills. (https://app.grammarly.com/)

[2] Quizlet is a tool that helps somebody revise for their exams by means of flashcards or games. (https://quizlet.com/nl)

[3] Kahoot! is a tool that can be used by teachers to keep their students engaged. It quizzes them on the subject at hand and adds a scoring system to safeguard competitiveness. (https://kahoot.it/)

For this study, we will zoom in on how certain technology can be employed for oral pronunciation training during individual speaking sessions at home. The technology that will be examined in closer detail is Automatic Speech Recognition or ASR. This technology has developed greatly in recent years and is being used for a variety of applications, such as Apple's Siri or Microsoft's Cortana which are both virtual assistants allowing – to some extent – human-computer interaction. Our focus will be on the ASR-application called ELSA (English Language Speech Assistant), which is a "personal AI-powered English Speaking coach" that provides short and fun dialogues which have been specifically designed to practice English pronunciation. The app also offers its users immediate feedback to improve their pronunciation in the form of implicit feedback. (Elsaspeak, 2023).

Since ASR is a relatively new technique which is mainly used for a variety of voice applications, such as the aforementioned Apple's Siri or Microsoft's Cortana, little research has been devoted to using ASR for language learning. Certain applications such as Novolearning, IVI and Nuance Dragon Dictation (section 2.3.1) have been tested before but are still relatively unknown to the general public. Even the ELSA app, which is considered to be one of the top five AI apps, along with Apple's Siri, Microsoft's Cortana, Google's Assistant and Amazon's Alexa (Davies, 2022), remains relatively unknown to the general public. The present study, therefore, aims to determine whether ELSA is capable of improving non-native students' English oral pronunciation performance.

To investigate this, the present study focused on sixteen native Dutch speakers who were randomly divided into an experimental group and a control group. The experimental group partook in a pre-test, a training period of one to two weeks in which they were asked to use the app and a post-test. The control group only partook in the pre- and post-test. The tests were specifically designed to focus on certain sounds that are particularly difficult for native Dutch speakers. Afterwards, the experimental group was also asked to fill in a short questionnaire about their attitudes and perceptions towards the ELSA app. The four main research questions resulting from this experimental design were:

1. Are there any improvements in oral pronunciation performance in general after having used the ELSA app for two weeks?

2. Is there a significant difference between the experimental group who worked with the app for one or two weeks and the control group who only took part in the pre- and post-test?

3. Are there any specific difficulties in English pronunciation for speakers of Dutch that have significantly improved after having used the ELSA app for two weeks?

4. What are the attitudes and perceptions of the experimental group towards the ELSA app?

Our results reveal that the experimental group did not perform as expected. Of course, there was to be noted some improvement within the experimental group, but the question remains whether this was entirely attributable to the ELSA app. It should be emphasised that multiple other factors could have influenced the outcome of this experiment.

The remainder of this thesis is subdivided as follows. Chapter 2 provides a literature study, which explains the importance of pronunciation skills in today's society, sheds light on the specific difficulties in English pronunciation for speakers of Dutch, clarifies the meaning and function of ASR and how it can be used in CALL environments and defines the significance of the ELSA Speak app. Chapter 3 explains the methodology behind the experiment: what were the research questions and hypotheses? Which participants were recruited? How was the design developed? How does the ELSA Speak app function? What type of questions were asked in the questionnaire? And which statistical tests were used? Chapter 4 then provides the findings of that particular experiment. The last chapter presents a conclusion of the findings and offers some prospects for future work.

Chapter 2: Literature review

2.1 English pronunciation

2.1.1 Importance of pronunciation

To become a proficient speaker in a specific language a learner should properly master the following four skills: grammar, pronunciation, fluency and vocabulary (Binus University, 2018; Derakhshan et al., 2016; Twinkl, 2022). The focus of this study will be on pronunciation, according to some the most significant skill within the speaking component of a language (Pourhosein Gilakjani, 2011). Even if the grammar and vocabulary of a sentence are undoubtedly correct, errors in pronunciation can still cause the conversation to become strained.

Firstly, if the pronunciation of a specific word or sentence is incorrect, it may become more difficult for the listener to understand the speaker. In some cases, especially in those with minimal pairs in which two words only differ from each other in one single sound, poor pronunciation tends to lead to misunderstanding (Pourhosein Gilakjani, 2011; Pourhosein Gilakjani & Sabouri, 2016). For instance, if 'cheer' were to be mispronounced as 'jeer' in a sentence such as 'the crowd cheered at the game last night', the meaning of that sentence would become the complete opposite, and as a result, may lead to misunderstanding. This example only leads to a minor misunderstanding, but should such an error occur in an international business meeting, a call centre or a medical situation, mispronunciation can eventually cause several major and more serious misunderstandings (Pennington & Rogerson-Revell, 2019).

Secondly, mispronunciation not solely leads to general misunderstanding; it can also give rise to particular grammatical mistakes, especially when dealing with final /t/ versus /d/ sounds (Gleason, 2012). For instance, the word 'stopped' is supposed to be pronounced as /stɑpt/, but some second language learners may drop the /t/-sound altogether as a consequence of their native language. Instead of /stɑpt/, they are now pronouncing it as /stɑp/, which is an infinitive form instead of a past tense (Gleason, 2012). These types of pronunciation errors will then result in a grammatical mistake, which makes it even more difficult for the listener to understand the utterance of the speaker.

A third and final reason why mispronunciation can become rather problematic is that it can undermine a person's credibility. When a non-native speaker pronounces a specific statement with a clear accent, the listener could perceive that statement, not as more difficult, but as less truthful (Lev-Ari & Boaz, 2010). This is especially the case with heavy-accented speakers. For mild-accented speakers, this issue may have less of an impact on their daily conversations.

Unfortunately, errors in pronunciation are still quite common in English as a Second Language (ESL) because teachers tend to pay more attention to grammar and vocabulary rather than to pronunciation (Pourhosein Gilakjani, 2011; Pourhosein Gilakjani & Sabouri, 2016). According to Pourhosein Gilakjani (2011) and Pourhosein Gilakjani and Sabouri (2016), there are several reasons for teachers to omit pronunciation during class, ranging from personal reasons to reasons involving their students.

First and foremost, learners need to be motivated to study a specific language. Their interest in pronouncing words accurately plays a vital role in achieving native-like pronunciation (Pourhosein Gilakjani, 2011). Therefore, teachers should not purely provide pronunciation exercises in class but should also encourage and motivate their students to acquire the correct pronunciation outside of school hours. This type of motivation immediately aligns with attitude. Learners who are more motivated and more concerned about their pronunciation of the target language will tend to have a better native-like pronunciation (Pourhosein Gilakjani, 2011).

Another way of improving a learner's pronunciation is by exposing them to the target language (Pourhosein Gilakjani, 2011). The more a learner is exposed to a particular target language, the easier it will be for that person to acquire the correct pronunciation. Pourhosein Gilakjani (2011) explains that although children are exposed to the target language for hours on end at school, adults often do not have that same advantage and thus live in what Pourhosein Gilakjani (2011) calls "linguistic ghettos". The lack of exposure to a target language can also be linked to fossilization. Fossilization refers to the phenomenon of critical periods in a learner's language acquisition, particularly the periods before the age of twelve (Pourhosein Gilakjani, 2011). After the age of twelve, it becomes increasingly more difficult to attain the correct pronunciation, since by then the wrong pronunciation is already deeply engrained in a learner's brain. In addition, a person's native language can also have a serious impact on the correct native-like pronunciation of a second language (Pourhosein Gilakjani, 2011). The mother tongue's

sound system is often transferred by many language learners onto the pronunciation of the target language. This can cause the learner to develop an accent, which may lead to misunderstanding and reduced credibility as mentioned in Lev-Ari and Boaz (2010).

Apart from language learners' personal reasons to neglect the target language's pronunciation, teachers also have multiple reasons for not incorporating pronunciation into their curriculum. They often perceive pronunciation as most futile in second language development. They find it less important than grammar and vocabulary and also think of it as the most difficult component of language to teach second language learners, which is why they prefer to focus less on pronunciation and more on grammar and vocabulary (Pourhosein Gilakjani & Sabouri, 2016). In general, most teachers do not have the time to teach every component as thoroughly either. They are bound to a certain time slot and thus only have a limited amount of time to teach whatever is on the curriculum; it is often pronunciation that is subsequently omitted (Pourhosein Gilakjani & Sabouri, 2016). Apart from time restrictions, teachers frequently do not possess the right tools to be able to evaluate pronunciation accordingly (Pourhosein Gilakjani & Sabouri, 2016). They often lack the educational resources as well as the required knowledge to evaluate a learner's pronunciation. Sometimes, they even "lack confidence regarding their own English pronunciation" (Van Hattum, 2014).

However, when teachers are able to evaluate a learner's pronunciation accordingly, they often focus on both the segmental and suprasegmental levels of language acquisition (Pourhosein Gilakjani, 2011; Pourhosein Gilakjani & Sabouri, 2016). The segmental level of language acquisition only focuses on "individual units of speech, such as phonemes or phones" (Cambridge Dictionary, 2022). These individual units of speech are more straightforward than suprasegmental features, which makes them easier to teach and evaluate. Suprasegmental features, on the other hand, can be more complicated and are rather versatile; they can include, inter alia, pitch, intonation, stress, tone or word juncture (Encyclopedia Britannica, 2022; Pourhosein Gilakjani, 2011). Nonetheless, even though the suprasegmental features are more difficult to teach and evaluate, they are not any less important, which is why certain teachers already include practice on those features in their classrooms (Pourhosein Gilakjani, 2011).

In what follows, this study will mainly focus on segmental features or individual sounds. Table 1 and Figure 1 below display all British consonants according to their manner of articulation, place of articulation and voicedness, as well as all British vowels and diphthongs according to their length, lip roundedness, frontness or backness and tongue height.

	Bi-labial	Labio-dental	Dental	Alveolar	Post-alveolar	Palato-alveolar	Pala-tal	Velar	Glottal
Plosive	**b**, p			t, **d**				k, **g**	
Fricative		f, **v**	θ, **ð**	s, **z**		ʃ, **ʒ**			h
Affricate						tʃ, **dʒ**			
Nasal	**m**			**n**				**ŋ**	
Lateral approximant				**l**					
Central approximant	**w**				**r**		**j**		

Table 1: British consonants according to their manner of articulation, place of articulation and voicedness (voiced = in bold, voiceless = not in bold).
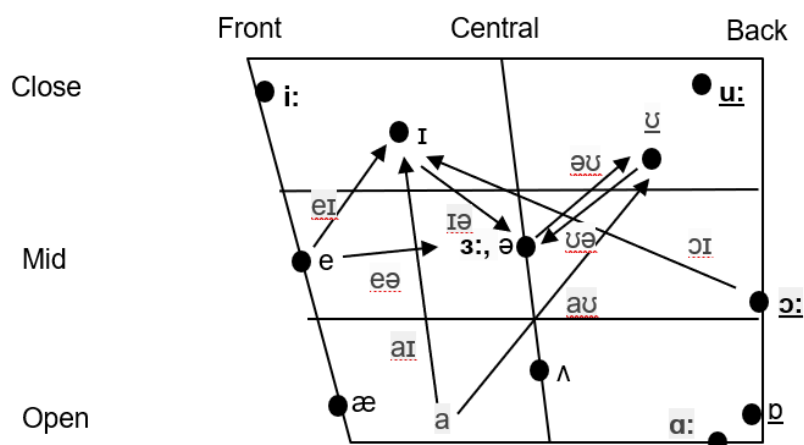
Figure 1: British vowels and diphthongs according to their length (long = in bold, short = not in bold), lip roundedness (round = underlined, unround = not underlined), frontness or backness and tongue height.

2.1.2   Difficulties in English pronunciation for speakers of Dutch

Since the experiment that will be conducted later on in this research focuses on Flemish secondary school students and first-year university students specifically, it can be practical to outline the difficulties speakers of Dutch particularly experience with regard to English pronunciation. It has to be noted, however, that these difficulties are mainly based on speakers of Dutch originating from the Netherlands, whereas the participants

---

[4] This information is based on the 'Engels Taalpraktijk A' syllabus taught in the Applied Linguistics course at Ghent University and written by Prof. Dr. Sabine De Vreese.

of this research all originated from Flanders, Belgium. Nonetheless, Collins and Mees (2003) and Collins and Vandenbergen (1998) incorporated both variants into their study and saw many similarities between the two varieties. For example, the /r/-sound in Ghent and Bruges (Flemish regions) and The Hague and Rotterdam (Dutch regions) are both pronounced in a uvular manner (Collins & Mees, 2003; Collins & Vandenbergen, 1998). For this study, we can thus assume that the differences are small enough to assess Flemish and Dutch students in a similar manner. The focus will first be on consonant sounds, then on vowel sounds and lastly on some suprasegmental features, such as assimilation and stress errors.

The Dutch particularly struggle with the aspiration of /t/, /p/ and /k/ at the beginning of British words (Cucchiarini, 2011; Hermans & Sloep, 2015; Kruitbosch, 2020; Van Hattum, 2014), the reason being that they do not aspirate these sounds in their native language. At the end of words, they also tend to struggle with voiced plosives such as /b/, /d/ and /g/, pronouncing them as their voiceless alternatives /p/, /t/ and /k/, respectively (Cucchiarini, 2011; Hermans & Sloep, 2015; Kruitbosch, 2020; Van Hattum, 2014). This has everything to do with the lenis (weak) and fortis (strong) contrast in end position. Several Dutch people also switch up the sounds /f/ and /v/ in English, especially in end position (Hermans & Sloep, 2015; Kruitbosch, 2020), which may also be due to the fortis and lenis contrast.

The sounds /θ/ and /ð/ generally do not exist in Dutch, which implies that the Dutch will not be able to pronounce them. The closest they can get to these sounds is by replacing them with /s/, /t/ or /f/ and /d/, respectively (Cucchiarini, 2011; Hermans & Sloep, 2015; Kruitbosch, 2020; Van Hattum, 2014). Similarly, the sounds /ʃ/ and /ʒ/ generally do not exist in Dutch, but these do not particularly pose a problem for the Dutch, since they have similar sounds in their native language (Kruitbosch, 2020). However, /tʃ/ and /dʒ/, sounds that also do not occur in Dutch, are still often mispronounced as /ʃ/ or /tʃ/ and /ʃ/, respectively (Cucchiarini, 2011; Kruitbosch, 2020).

Even though the /l/-sound does occur in their native language, the Dutch still tend to mispronounce it, making it appear too pharyngealized (Kruitbosch, 2020). In addition, they tend to insert an /ə/-sound when /l/ is followed by /p/, /f/, /m/ or /k/ (Kruitbosch, 2020), similar to what Arabs do, but with a different insertion sound. According to Van Hattum (2014), this is also known as "epenthesis". Lastly, similar to how an Arabic or Spanish trilled /r/ is incorrect in British English, the Dutch uvular /r/ can also be considered incorrect in British English (Kruitbosch, 2020).

In terms of vowels, the Dutch often find it hard to distinguish between three pairs of vowels in particular. The first vowel pair that is often confused is the difference between /æ/ and /e/ (Cucchiarini, 2011; Hermans & Sloep, 2015; Kruitbosch, 2020; Van Hattum, 2014). The Dutch often replace both /æ/ and /e/ with the Dutch vowel /ɛ/, which is close to both English vowels, but is still not completely correct and can lead to confusion between minimal pairs and thus misunderstanding (Kruitbosch, 2020; Pourhosein Gilakjani, 2011; Pourhosein Gilakjani & Sabouri, 2016). A second vowel pair the Dutch often confuse is the difference between /u:/ and /ʊ/ (Cucchiarini, 2011; Hermans & Sloep, 2015; Kruitbosch, 2020). They tend to relate both sounds to the Dutch /oe/-sound, which is not completely correct and can once again lead to confusion between minimal pairs and thus misunderstanding (Kruitbosch, 2020; Pourhosein Gilakjani, 2011; Pourhosein Gilakjani & Sabouri, 2016). A last vowel pair the Dutch often have difficulty with distinguishing between is the difference between /ʌ/ and /ə/ (Cucchiarini, 2011; Hermans & Sloep, 2015; Kruitbosch, 2020). They are quite close together and do not belong to the Dutch phonetic system, which is why the confusion is understandable.

As for the English diphthongs, none belong to the Dutch phonetic system, implying that they could all pose an issue for speakers of Dutch. However, certain English diphthongs require more practice than others for speakers of Dutch. The diphthongs that present the most problems are /aɪ/, /ʊə/ and /ɪə/ (Cucchiarini, 2011; Hermans & Sloep, 2015; Kruitbosch, 2020).

Apart from the mistakes made by speakers of Dutch in terms of segmental features, they often also make mistakes regarding some important suprasegmental features. Once again, there are three main ones that are quite striking. First of all, the Dutch assimilation pattern is different from the British one (Kruitbosch, 2020). Therefore, it can be quite difficult for the Dutch to adapt to a new assimilation pattern when they already have one of their own. Additionally, they often struggle with recognizing gradation words in English. These gradation words are pronounced strongly (fortis) when in isolation, but weakly (lenis) when used in a specific sentence (Hermans & Sloep, 2015; Kruitbosch, 2020). The Dutch then often make the mistake of pronouncing these gradation words strongly in every context (Kruitbosch, 2020). Lastly, when a certain word can both be a verb as well as a noun, Dutch people often use the same intonation pattern, causing them to confuse the verb with the noun (Van Hattum, 2014).

Kruitbosch (2020) combined the "error hierarchies for English pronunciation" created by both Collins et al. (2003) and van den Doel (2006). A summary of the most significant, the significant and the less significant errors in British Received Pronunciation made by the Dutch is provided below. The most significant errors will be the basis of the experiment later on in this research. However, it should be noted that these errors are based on Dutch speakers originating from the Netherlands, whereas the participants in this research are Dutch speakers originating from Flanders, Belgium.

Most significant	Loss of the fortis-lenis contrast, in particular the confusion between /f/ and /v/ in initial and medial position and the confusion between /t/ and /d/.
	The confusion between /æ/ and /e/.
	The loss of contrast, or confusion, between /ʊ/ and /u:/.
	The substitution of /θ/ with /t/.
	The epenthesis of /ə/ between l and a following non-alveolar consonant.
	The use of uvular /r/.
	The incorrect realisation of /ð/ in general (Collins et al., 2003), or the substitution of /ð/ with /d/ (Van den Doel, 2006).
	Assimilation errors.
	The incorrect realisation of /æ/ and /ʊ/.
	The replacement of word-final /dʒ/ and /tʃ/ by /ts/.
	The confusion of initial /g/ with /k/.
	The replacement of /θ/ with /s/ or /f/.
	The confusion of the contrast between /æ/, /e/, and /eə/.
	The confusion between /ʃ/ and /s/.
	The production of /aɪ/ that is too long before fortis.
Significant	Lack of aspiration in initial fortis stops.
	The use of pharyngealized l.
	The production of reduced and too rounded /ɜ:/, /ʌ/, and /ɒ/, resulting in /ɤ:/, /ʏ/, and /ɔ/ respectively.
	Substitutions of θ with t and ð.
	Overlong /aɪ/.
Less significant	The production of an /ɪ/ that is too close and an /ɑ:/ that is too fronted.
	Pharyngealized l.

Table 2: Summary of the errors in British Received Pronunciation made by the Dutch according to Kruitbosch (2020) based on Collins et al. (2003) and van den Doel (2006).

As mentioned in 2.1.1, teachers often do not have the time to pay attention to pronunciation in class. To eliminate the errors made by the Dutch in British Received Pronunciation and thus to avoid misunderstandings, it may be beneficial to ask students to use Computer Assisted Language Learning (CALL) and Computer Assisted Pronunciation Training (CAPT) systems at home (Cucchiarini, 2011; Kruitbosch, 2020). In this way, the Automatic Speech Recognition (ASR) systems incorporated in these CAPT systems could provide these students with feedback on their pronunciation and consequently help them improve their pronunciation (Cucchiarini, 2011; Kruitbosch, 2020).

The following section will therefore first explain what ASR entails and when we tend to use it in our daily lives. Section 2.2.2 will zoom in on some of the limitations ASR still comprises to this day, followed by a short summary of ASR's manner of functioning. The last section will then observe the state-of-the-art ASR-systems that are most popular in this day and age.

2.2   Automatic Speech Recognition or ASR

2.2.1   ASR in everyday life

Automatic Speech Recognition or ASR can be defined as "the technology that allows human beings to use their voices to speak with a computer interface in a way that, in its most sophisticated variations, resembles normal human conversation" (Zajechowski, 2014). The term may not immediately sound familiar, but we tend to use ASR-systems daily on a great number of devices. Jurafsky & Martin (2021) describe some of these uses in their seminal work on Speech and Language Processing:

- Nowadays, ASR is a built-in feature in smart home appliances, cell phones and personal assistants, such as Siri for the iOS, Amazon's Alexa, Google's Assistant, Windows' Cortana and many others (Hoy, 2018).
- ASR-systems are also very useful for transcribing movies, videos and live discussions. Take the field of law as an example: instead of a person having to write down everything another person dictates, ASR-systems could easily take their place in the future when they start producing output with a zero per cent error rate or at least close to zero per cent.
- Lastly, ASR could also play an important role in the interaction between computers and people with a disability, making them unable to type anything (e.g. blind or injured people).

User interfaces are by far the most used ASR-systems to date. Thanks to those ASR-systems, personal assistants such as Siri or Alexa can now execute simple tasks (Hoy, 2018, p. 83), such as:

- Sending and reading text messages or making phone calls.
- Answering basic questions asking about the weather or the time.
- Setting timers, alarms and calendar entries.
- Setting reminders, making lists and doing basic maths calculations.

- Controlling certain apps such as Spotify, Netflix, Amazon, Google Play and many more.
- Controlling items around the house connected to the internet, such as thermostats, lights, alarms and locks.
- Telling jokes and stories.

2.2.2   Limitations of ASR

ASR-systems have come a long way since their very beginning in the 1970s. The improvements made over the last decades have been enormous. Nevertheless, there are still some limitations as to what ASR-systems are capable of nowadays (Jurafsky & Martin, 2021; McTear et al., 2016):

- In terms of vocabulary size, ASR-systems started with only being able to recognize a few words, including the words *yes* and *no.* During the past decade, ASR-systems have evolved into being able to recognize over millions of words. However, they can still not recognize every word in every language.
- ASR-systems used to be speaker-dependent, which means that they were trained to only recognize the voice speaking to it. It took hours for them to be trained well enough and even then, not everyone could use it. Nowadays, ASR-systems are speaker-independent, but can still not recognize every voice, especially if there is a strong accent or speech disability involved.
- In the past, users of an ASR-system had to make sure they articulated well and they had to take small pauses in between words so that the system could understand every single word. Nowadays, ASR-systems are able to recognize continuous speech as well. However, it is still easier for them to understand read speech (speech that was already planned) than conversational speech (spontaneous speech between multiple people).
- When it comes to technical issues, there is a possibility that there is too much background noise (cars passing by, music playing …) for the ASR-system to understand what has been said. The robustness of the system still needs to be improved in order for background noises to be cancelled out.
- Another technical issue that can arise refers to the microphone quality. The microphone in question may not always be of the highest quality or be at the same distance from the speaker's mouth. This also influences whether the ASR-system is able to understand the speaker or not.

2.2.3   How does ASR work?

Most of the time, like with user interfaces for example, ASR is accompanied by text-to-speech (TTS): when asking something to a user interface, it is expected to give an appropriate reply. Although TTS will not be discussed because most studies in this literature overview only focus on the ASR-system used, it can be said that the process of TTS is almost the same as ASR, but in the opposite direction (Jurafsky & Martin, 2021).

ASR remains an incredibly complex process, which is why this literature study will not provide a complete and elaborate technical explanation of how it all works. Instead, a high-level overview is presented. Essentially, there are three components to the entire process (McTear et al., 2016):

- An acoustic model
- A language model
- A decoder

Chazen (2019) briefly explains how an ASR-system essentially works. Initially, somebody has to talk to the ASR-system. It will then detect the person's speech and create a wave file of the words spoken to it. The system then tries to cancel out all background noise and tries to normalize the volume. Afterwards, the wave file is broken down into sequences, which will be linked to the words or sentences that appear to be most likely by means of statistical probability. This is what McTear et al. (2016) call the acoustic model. The language model will then determine which words are most likely to appear next to each other by means of n-grams (a sequence of n words). These n-grams will predict the words that follow by analysing the preceding words. Lastly, the decoder will combine the acoustic model and the language model to determine the most likely word sequence (McTear et al., 2016).

2.2.4   State-of-the-art ASR-systems

As previously mentioned, such state-of-the-art ASR-systems have already been incorporated into many user interfaces, such as Apple's Siri, Amazon's Alexa, Google's Assistant and Microsoft's Cortana (Hoy, 2018). The question remains which one of these voice assistants is most competent at recognising human speech. After thoroughly testing the interfaces with some English commands, it seemed that Google was able to

outperform all other voice assistants in both isolated word recognition and speech recognition in its entirety (Kinsella, 2018). Recently, the company has also conducted some research into an even more improved voice recognition system of the English language that can detect people's speech "in real time when multiple people are talking by comparing it to an existing voice sample" (Schwartz, 2020). However, commercial systems, such as Google, Watson and Microsoft "offer little access to detailed model outputs", which make them less flexible (Sciforce, 2021). In that case, it could be better to consult a more open-source ASR-system. Nonetheless, the type of ASR-system you choose to use heavily depends on whether you are searching for high quality or flexibility (Sciforce, 2021). If it is high quality you are looking for, Google remains at the top of the list, but if you want your ASR-system to be more flexible, it may be best to use a more open source ASR-system such as Kaldi[5] (Sciforce, 2021).

The primary goal of an ASR-system in user interfaces is to understand the person talking to it so that it can form an appropriate reply (Jurafsky & Martin, 2021). To that purpose, the speaker needs to have the correct pronunciation and a good command of grammar. Of course, vocabulary and fluency matter as well, but an ASR-system is more likely to misunderstand a sentence due to incorrect pronunciation or grammar than it is due to disfluency or a vocabulary being too wide (Knill et al., 2018).

Now that it has been made clear how ASR-systems exactly operate, it may also be interesting to examine how they are used in several research fields, such as CALL, language learning and pronunciation. To this purpose, other studies were sought with participants whose native language was Dutch. However, only one Dutch study and one French study could be found. All other studies had an Asian origin (Chinese and Korean). It seems that those countries are more focused on these specific research areas.

---

[5] Kaldi was initially made for researchers, but it has made a name for itself fast. Kaldi is a John Hopkins University toolkit for speech recognition and […] can actually compete with Google" (Sciforce, 2021).

2.3 ASR-systems in learning pronunciation

2.3.1 Improvements in pronunciation by means of an ASR-system

A first method that will be discussed and that uses an ASR-system to try and improve pronunciation is relatively basic. That is to say, it was only based on one practising session in which 39 Dutch primary school pupils were asked to practice their Dutch pronunciation by means of the ASR-system NovoLearning[6] which provided direct feedback (Bai et al., 2020). This type of feedback produces spoken output, meaning that the system recognises the error in pronunciation and utters the word or sentence the way it was supposed to be pronounced, thus making it possible for the learner to repeat after the system. The researchers focused mainly on whether the ASR-system detected any advancements in the students' pronunciation accuracy during the practising session. This appeared to be the case since their pronunciation accuracy improved significantly from the first to the second and third attempts of reading a word, as measured by regression analysis, which could indicate that ASR is a very interesting tool to use (Bai et al., 2020). Nevertheless, this rather straightforward research method was not detailed enough to draw any real conclusions. Some limitations should definitely be taken into account such as the small number of participants and the relatively basic method that was applied.

A more intricate method was adopted by Inceoglu et al. (2020) and Liu et al. (2019): instead of participating in only one practising session, the students were now required to partake in a pre- and post-test to measure whether they had improved after having completed the practising sessions with the ASR-system. In the first experiment, 29 Chinese university students majoring in English were asked to take a preliminary test which determined their pronunciation level before practising with the iFlytek Voice Input (IVI) application[7] (Liu et al., 2019). After practising with the app for twenty minutes four times a week for a month, a post-test was carried out to determine whether the participants' pronunciation and intonation, among other things, had improved. However, unlike the previous experiment (Bai et al., 2020), there was no direct feedback. The system only highlighted the words that were pronounced incorrectly, which made it more

---

[6] NovoLearning is an ASR-system that can specifically be used for pronunciation training and reading practice. (https://www.novo-learning.com/language)

[7] The iFlytek Voice Input (IVI) application can be used on mobile devices only and is known for its intricate automatic speech recognition system. (https://iflytek-voice-input.en.uptodown.com/android)

difficult for the participants to rectify their mistakes. In a similar experiment nineteen intermediate Korean university students that studied English as a foreign language (EFL) were asked to participate in a pre-test, six practising sessions, each lasting three to six minutes, and a post-test (Inceoglu et al., 2020). During the tests and the practising sessions, they were allowed to use the ASR-systems provided by either Google (Google's Assistant) or Apple (Siri). As in the previous experiment (Liu et al., 2019), these provided only written output in order for the learners to correct their mistakes. The tests consisted of three tasks: reading aloud an excerpt from a text, describing a series of pictures and reading aloud a list of fourteen minimal sound pairs.

The more intricate methods introduced by Inceoglu et al. (2020) and Liu et al. (2019) offered more reliable results regarding the use of ASR-systems in language learning classrooms. Also, in these experiments, the participants' pronunciation improved significantly from pre- to post-test (Inceoglu et al., 2020; Liu et al., 2019) further corroborating the added value of ASR for university students of English as a second language. However, in these studies, the number of participants also remained quite small. Moreover, some other observations could be made, such as the accuracy of the ASR-system not being able to achieve 100% just yet (Liu et al., 2019) and the similarity between the participants' L1 and the target L2 sounds playing an important role in improving one's pronunciation (Inceoglu et al., 2020).

In another study, aside from adding pre- and post-tests, participants were also divided into two groups (Yuan & Liu, 2020). These groups were a low-proficiency group that had not been studying the language in question for that long and a high-proficiency group that already had a rather advanced language level. In compliance with this requirement, fifty Chinese senior high school students were divided into a low-proficient group (25) and a high-proficient group (25). They all participated in a pre-test, a training session and a post-test in which they had to read aloud and imitate a few sentences in English. During the training sessions, they were provided with direct feedback, similar to Bai et al. (2020), by the ASR-application Oral English Drill & Test[8]. In the next study by Wang & Young (2015), a similar distinction can be found, albeit implicitly. The participants were divided into a group of eighteen college students with seven years of English learning experience (high-proficiency) and a group of sixteen junior high school students with only three years of English learning experience (low-proficiency). Each of them was required to partake

---

[8] The Oral English Drill & Test app is an application that helps people improve their oral English and includes ASR technology developed by Associate Professor Liu Xiao-bin. (https://link.springer.com/content/pdf/10.1007/978-981-33-4594-2.pdf)

in a pre-test, eight weeks of practice and a post-test in which they had to listen to and repeat eight English sentences. During the practice sessions, they were allowed to use an ASR-system "developed by the Learning Technology (LT) Lab and the Multimedia Information Retrieval (MIR) Lab of the National Tsing Hua University in Taiwan" (Wang & Young, 2015, p. 495). The system provided both implicit and explicit feedback. The implicit feedback only offered a pronunciation score, whereas the explicit feedback enabled the participants to listen to how the word or sentence was supposed to be pronounced.

In respect to the distinction between high-proficiency and low-proficiency learners, the results remain indecisive. In the study by Yuan & Liu (2020), both groups showed an improvement from pre- to post-test, but only the low-proficiency learners saw a significant difference in their overall pronunciation accuracy. The similar experiment by Wang & Young (2015) produced a completely different outcome. The ASR-system did indeed help both groups in improving their pronunciation accuracy, but this time it was the high-proficiency group that witnessed a noticeable improvement. This group also seemed to prefer implicit over explicit feedback, whereas the low-proficiency learners probably need both explicit feedback and feedback from a teacher to further develop their pronunciation accuracy.

Up until now, the studies discussed have only taken into account the improvements of pronunciation from pre- to post-test, but the question remains whether it is indeed the ASR-system that ensures those improvements or whether the advancements in pronunciation depend entirely on the amount of time practised. This question can be answered by involving not only an experimental group that has access to an ASR-system, but also a control group that receives little to no feedback from the system. A first experiment compared a rather elaborate ASR-system to a simpler one. 38 adults were divided into an experimental group (18) that was allowed to use the three-level feedback intelligent computer-assisted speaking learning (iCASL)[9] system and a control group (20) that could solely use the one-level feedback system (Wang & Young, 2014). As in the previous experiment (Wang & Young, 2015), the researchers worked with both implicit and explicit feedback. The three-level feedback system provided both, whereas the one-level feedback system only provided implicit feedback. After the pre-test and the

---

[9] "The iCASL system could be regarded as a web-based speaking learning management system that assists learners in arranging their English study progress and trace their speaking performance" (Wang & Young, 2014, p. 222)

practising sessions, the participants took a post-test in which they had to listen to and repeat a few sentences.

However, instead of having just one experimental and one control group, a third group that combines an ASR-system with other types of feedback can be added as well. In this way, 180 Chinese students enrolled in a College English course were divided into three groups of 60 students each (Dai & Wu, 2021). The first group was centred around the best students giving feedback to the worst students (co-non-ASR group). The second group received both feedback from their peers as well as from the ASR-application called WeChat[10] (co-ASR group) and the final group only received feedback from the ASR-system (ASR group). The pre-test, practising session and post-test in which the participants had to read a few sentences were all taken at the same time and were followed by a supplementary unexpected post-test one week later. In a similar experiment in which 34 students of French as a second language participated, the three groups were divided differently, in the sense that one group received feedback from an ASR-system, another group from a teacher and a final group received no feedback whatsoever (Liakin et al., 2014). During the practising sessions, the students in the ASR group were required to use the ASR-application called Nuance Dragon Dictation[11] for twenty minutes five times a week. The other groups performed the same exercises but without the written feedback provided by the ASR-system, which was similar to the feedback provided by Liu et al. (2019). The pre- and post-test consisted of a production task in which the participants had to read a few sentences and a perception test in which they had to recognise the sound [y] in 45 monosyllabic French words.

The division between an experimental and a control group should be able to reveal whether it is truly the ASR-system that ensures improvements in pronunciation. In the first experiment (Wang & Young, 2014), both groups were allowed to use an ASR-system, but the first one received more elaborate feedback from the system than the second one. The researchers proved that the elaborate ASR-system that provided both explicit and implicit feedback is more capable of improving a student's overall pronunciation than a simpler one that solely provided implicit feedback. These findings seemed to be in line with the observations made about low-proficiency learners discussed in one of the previous paragraphs (Wang & Young, 2015). At the same time,

---

[10] WeChat is an app that can be used for many activities, such as chatting, calling, ordering food or taxis etc. It is not usually used for language learning, but it does involve an excellent speech-to-text recognition tool. (https://www.wechat.com/)

[11] Nuance Dragon Dictation is available on Apple's iPhone, iPod touch and iPad. It has a built-in ASR-system. (https://www.nuance.com/index.html)

these findings also align with the findings of the study by Dai & Wu (2021) which revealed that students need more input than solely implicit feedback to improve their pronunciation accuracy. However, this time the explicit feedback was mainly provided by the co-non-ASR and the co-ASR group, whereas the ASR group mainly received implicit feedback from the system, which means that the ASR group was outperformed by the other two groups. In addition, there was no real difference between the co-non-ASR and the co-ASR group, which indicates that ASR is not necessary for in-class contexts. However, it should be noted that the ASR group did improve from pre- to post-test which could indicate that ASR is still a good addition to individual practising sessions. The study by Liakin et al. (2014), however, contradicts these findings once again: the researchers discovered that when it comes to pronunciation, the ASR group outperformed the other two groups that only received feedback from a teacher or none at all. In general, these authors believe that incorporating ASR into in-class speaking sessions can be promising for developing better pronunciation.

2.3.2   Attitudes towards ASR-systems

Almost every study mentioned in the previous section was conducted as a mixed-methods experiment, which means that it not only measured the improvements in pronunciation by means of an ASR-system, but also investigated the attitudes that students adopted towards the ASR-system in question. The most common ways of interrogating the participants about their perceptions were multiple-choice questions, open-ended questions and interviews.

In the first experiment, the participants were asked to fill out an exit survey that consisted of seven statements on a 10-point Likert scale and two open-ended questions in which they had to answer truthfully whether there were any issues regarding the ASR-system and whether they would use it again to practice their English pronunciation (Inceoglu et al., 2020). The next experiment adopted a similar approach in the sense that there were thirteen questions relating to the perceptions of the participants that could all be answered by means of a 5-point Likert scale (Yuan & Liu, 2020). The following two researchers decided to ask only ten questions with a 5-point Likert scale and one open-ended question (Wang & Young, 2015). Their previous research (Wang & Young, 2014) was more extensive in this respect in that they used thirty questions with a 5-point Likert scale and two open-ended questions. The study by Dai & Wu (2021) also included six questions on a 6-point Likert scale. However, in this study also a longer interview was

conducted, comprising around 30 to 60 minutes, in which students were asked about their experiences, the advantages and disadvantages of the ASR-system and their suggestions to improve the system. Table 3 below provides some examples of the type of Likert-scale questions asked in the aforementioned studies (Dai & Wu, 2021, p. 12; Inceoglu et al., 2020, p. 833-835; Wang & Young, 2014, p. 228, 2015, p. 500).

How useful is ASR for pronunciation practice?
How well does ASR recognize your pronunciation?
I feel embarrassed while practising English with the system.
I enjoy practising English with the system.
I would like to use the system for further learning.
Using the system enhances the opportunities of English speaking.
Using the system promotes my motivation to speak English.
I think that the activity has improved my English pronunciation.
I think that the activity was interesting.
I think that I will repeat this activity to practice my English pronunciation.

Table 3: Questions relating to the perceptions and attitudes towards an ASR-system.

These studies all seemed to conclude that most students were of the opinion that ASR could be very helpful when practising their pronunciation (Inceoglu et al., 2020). The few students that were rather pessimistic complained mainly about the system not being able to recognize their voices. This could be due to ASR-systems not having reached their full potential just yet (Liu et al., 2019). Nevertheless, most learners, regardless of their proficiency, still evaluated the ASR-system rather positively (Wang & Young, 2014, 2015; Yuan & Liu, 2020). The participants stated that they were often nervous and embarrassed when talking in front of the class, whereas they were more relaxed and comfortable when speaking to the ASR-system (Wang & Young, 2014, 2015). They also elaborated more on all the different advantages and disadvantages that the new technology entails (Dai & Wu, 2021). For example, they mentioned that the system has a certain convenience and that it allows for more interaction and autonomy on the one hand, but can also display some technical issues on the other hand.

The above-mentioned studies were all rather limited when it comes to studying perception. In this respect, the study by Van Doremalen et al. (2016) offers a more complete picture in that their research was completely dedicated to this topic. This study is divided into three sections: a usability review, an expert review and user testing. In the expert review, nine experts of the Dutch language were introduced to the ASR-system

DISCO[12]. Afterwards, they had to fill in a questionnaire about their perceptions of the system and participate in an interview based on that questionnaire. During the user testing, nine teachers and five students of Dutch as a second language were also required to fill in a questionnaire regarding their attitudes towards the system. All of the previous findings (Dai & Wu, 2021; Inceoglu et al., 2020; Wang & Young, 2014, 2015) were confirmed by the nine teachers and five students of Dutch as a second language who used the DISCO-system as an example of an ASR-system (van Doremalen et al., 2016). However, it has to be noted that this system has not yet reached its full potential, considering that there were still some technical and content issues.

2.4 The ELSA Speak App

After having discussed various applications meant to improve a person's pronunciation, such as the Novolearning app, the IVI application and the Oral English Drill & Test app, the focus of this study will be on the rather innovative application called the English Language Speech Assistant (ELSA)[13]. ELSA was first launched in 2015 by CEO Vu Van and co-founder Dr. Xavier Anguera. Since she is of Vietnamese origin and struggles greatly with her English pronunciation, CEO Vu Van came up with the idea of creating an app that could help people improve their English pronunciation to sound more trustworthy (Anguera & Van, 2016). Dr. Xavier Anguera, a speech technologist, helped her build the AI-system with its built-in automatic speech recognition technology that is currently being used by more than thirteen million users in over one hundred countries (Nushi & Sadeghi, 2021). The ASR-system behind the app is, therefore, an in-house system for which little background information can be retrieved. In 2022, the app was even considered one of the top five AI apps, along with Apple's Siri, Microsoft's Cortana, Google's Assistant and Amazon's Alexa (Davies, 2022).

However, as with the other applications that have already been mentioned in this research, the ELSA Speak app is, of course, not one hundred per cent failproof. Becker

---

[12] DISCO stands for Development and Integration of Speech Technology into COurseware for language learning. It is a system specifically designed by the researchers of the study to give their students feedback on their Dutch pronunciation and grammar.

[13] ELSA is a "personal AI-powered English Speaking coach" that provides short and fun dialogues for you to practice your pronunciation and that gives immediate feedback for you to improve your pronunciation (https://elsaspeak.com/en/).

and Edalatishams (2019) have identified several shortcomings related to the app. The largest shortcoming according to them is "the lack of balance between segmental and suprasegmental practice opportunities" (Becker & Edalatishams, 2019, p. 3-4). The ELSA app tends to focus solely on segmental features, whereas suprasegmental features are considered equally important. Moreover, when a speaker mispronounces a specific word or sentence, the ASR-system behind ELSA will occasionally still indicate that word or sentence as correct. According to the researchers, this shortcoming may be due to ELSA being more focused on quantity than quality. Lastly, Becker & Edalatishams (2019) are of the opinion that the app should take the user's background, such as demographic information and native language, into consideration. When you first open the app, you are asked to indicate your mother tongue, but it remains unclear whether this is taken into account in the exercises themselves. Another small detail that makes the app less trustworthy is the fact that the app contains multiple typos. That is why the researchers plead to accept applied linguists to strengthen the ELSA team rather than having solely software developers and engineers improve the app.

Ever since the launch of the app, CEO Vu Van and co-founder Dr. Xavier Anguera have been open to feedback from ELSA's users (Anguera & Van, 2016). They listen to the users' input and try to improve the app's interface and educational purpose. Multiple studies have already been conducted to determine the users' perceptions of the ELSA app. Most students who have already worked with the language learning app appear to have a positive attitude towards it (Anggraini, 2022; Kholis, 2021; Silaen & Rangkuti, 2022). They state that they are motivated to improve their pronunciation through ELSA and they particularly like that the app can provide immediate feedback (Anggraini, 2022; Kholis, 2021). Especially during the lockdown, when teaching methods had to be adapted to the online environment, the ELSA app became a much-appreciated means of learning. Many students were satisfied with the content, the pedagogical approach and the interface provided by the ELSA app. Consequently, these positive attitudes towards the ELSA app show that students are willing to use the app to improve their pronunciation.

However, determining the actual benefit of the ELSA app on English pronunciation is much more difficult. As the app has not been around for that long, not many studies can be found on this particular subject. Those that can be found originate primarily from Indonesia, but they all come to the same conclusion.

Anggraini (2022) and Kholis (2021) both conducted a quantitative research study in which they let their participants take part in a pre-test, a training session and a post-test, as in the previous studies. The first experiment counted 30 students who followed the Easy English Course at Universitas Bina Sarana Informatika in Indonesia (Anggraini, 2022). At the end of the experiment, it was determined that the ELSA Speak app could indeed improve students' pronunciation, since the students' scores went from an average of 60 out of 100 to 80 out of 100 before and after using the app. The second experiment counted 18 students from the Ulama University of Yogyakarta in Indonesia (Kholis, 2021). This research also claimed that the ELSA Speak app was effective because the experimental group had shown an improvement from pre- to post-test and had also improved more than the control group who was only taught through regular teaching. Anggraini (2022) and Kholis (2021) both included a short qualitative study as well, in which it was made clear that the students had a positive attitude towards the ELSA Speak app, mentioning that they had the impression their pronunciation had improved after using the app. Pinontoan et al. (2022) only wanted to determine the perceptions of 25 students towards the ELSA app, because they thought it would be interesting to be able to use the app during the Covid-19 pandemic while it can be difficult for teachers to give feedback through asynchronous instructions. Once again, most students showed a positive attitude towards the app, stating that it had an easy-to-use interface and had good-quality content. The creators of the app, Vu Van and Xavier Anguera, also did their own research and came to the same conclusion as the previous studies. They analysed the data of a few regular users of the app and saw a clear improvement in pronunciation from the first time they used the app (Anguera & Van, 2016).

However, the aforementioned studies are not all as reliable. The study by Anguera and Van (2016), for example, is quite subjective, since they are the owners of the app. Anggraini (2022) and Kholis (2021) both used quantitative methods, but did not process their results statistically. Moreover, the studies by Anggraini (2022), Kholis (2021) and Pinontoan et al. (2022) were all conducted from an Indonesian point of view. The current research, therefore, proposes to investigate the benefit of ELSA from a Dutch and objective perspective. The design of the experiment that will be conducted in the next section will therefore be the same as the experiments cited in this literature study (Section 2.3). First, a pre-test for the target audience will be set up, after which that same target audience will work with the app for about two weeks, trying to improve their pronunciation. Afterwards, a similar post-test will be conducted to observe whether a change in pronunciation has effectively taken place.

Chapter 3: Methodology

The following chapter will explain the methodology behind the experiment. The first section will elaborate on the four main research questions posed in the introduction and link them to a hypothesis based on the literature review. Afterwards, the background of the participants and actual data collection will be discussed. The next section will then explain the design of the pre- and post-tests, followed by a short summary of ELSA's functions which the participants have used during the experiment. The penultimate section will specify the design of the questionnaire, which was filled in by the experimental group only. Lastly, the statistical tests used in this experiment will be explained.

3.1   Research questions and hypotheses

As mentioned in the introduction, the four main research questions were designed to determine whether ASR-systems, and in this case ELSA's ASR-system, could effectively help to improve the English pronunciation of Dutch speakers. This research will first take into account the general improvement (or stagnation) in pronunciation performance of Dutch speakers by comparing their pronunciation scores from a pre- and post-test. These scores have been determined by an expert in the field, namely a native speaker of English with ample experience in teaching English to Dutch-speaking students.

As demonstrated in the literature study (section 2.3), multiple experiments have shown a significant improvement in students' English pronunciation from pre- to post-test after using an ASR-system (Inceoglu et al., 2020; Liu et al., 2019). This is why the first research question and hypothesis can be stipulated as follows:

R1: Are there any improvements in oral pronunciation performance in general after having used the ELSA app for two weeks?

H1: After having used the ELSA app for two weeks, the experimental group will have improved their general pronunciation performance.

To determine whether it was in fact the ELSA app that provided the potential improvements in pronunciation from pre- to post-test, a control group, who only participated in the pre- and post-test, was formed. Some experiments mentioned in the

literature study (section 2.3) also incorporated the distinction between an experimental and a control group to ensure that the improvement in pronunciation was entirely owed to the ASR-system (Liakin et al., 2014; Wang & Young, 2014). It can be concluded that the experimental group of most studies saw an improvement in pronunciation after working with the ASR-system, except for one (Dai & Wu, 2021). Even though there is some disagreement, most studies lead us to believe that ASR-systems can help to improve pronunciation and thus this leads to the second research question and hypothesis.

R2: Is there a significant difference between the experimental group who worked with the app for one or two weeks and the control group who only took part in the pre- and post-test?

H2: The experimental group will have improved more than the control group.

Apart from the general improvement (or stagnation) in oral pronunciation performance, this research also set out to find out whether certain English pronunciation difficulties for speakers of Dutch can be remedied by using the ELSA Speak app. As mentioned in section 2.1.2, the pre- and post-test will include some of the difficulties presented by Kruitbosch (2020) in her research on "Pronunciation errors made by Dutch secondary school students in English". It should be taken into account, however, that these errors are based on Dutch speakers of the Netherlands, whereas the participants of this research are from Flanders, Belgium (section 2.1.2). Some examples of the errors considered in this research are the incorrect realisation of the /r/-sound, the substitution of /θ/ with /t/ and assimilation errors. Although there are no particular studies that prove that ASR can help to improve certain English pronunciation difficulties for Dutch speakers, many studies have demonstrated that ASR can help with English pronunciation in general (Bai et al., 2020; Dai & Wu, 2021; Inceoglu et al., 2020; Liakin et al., 2014; Liu et al., 2019; Wang & Young, 2014, 2015; Yuan & Liu, 2020). The third research question and the third hypothesis can therefore be formulated as follows:

R3: Are there any specific difficulties in English pronunciation for speakers of Dutch that have significantly improved after having used the ELSA app for two weeks?

H3: After having used the ELSA app for two weeks, certain difficulties in English pronunciation for speakers of Dutch (but not all) will have been eliminated for some of the participants.

The final research question is not measured by means of a quantitative method, but by means of a qualitative method. This research question will take into account the attitudes and perceptions the experimental group had towards the ELSA app. The group will be asked to indicate their agreement on a five-point Likert scale for statements such as 'I understood the feedback that was given to me' or 'I would use the app again to improve my pronunciation'. Previous research has indicated that, overall, students react positively towards an ASR-system, stating that it can be helpful, it allows you to be more relaxed and comfortable and it creates a certain degree of autonomy (Inceoglu et al., 2020; van Doremalen et al., 2016; Wang & Young, 2014, 2015; Yuan & Liu, 2020). The final research question and hypothesis can therefore be formulated as follows:

R4: What are the attitudes and perceptions of the experimental group towards the ELSA app?

H4: It is hypothesised that most participants will have a positive attitude towards the ELSA app.

3.2 Participants, data collection and analysis

For this study, only students who spoke Dutch as their mother tongue and who had a certain notion of the English language but still had an underdeveloped English pronunciation were allowed to participate. The participants that were selected were, therefore, students in the sixth year of secondary school at College O.-L.-V. Ten Doorn in Eeklo, East Flanders and first-year university students in the study programme Applied Linguistics at Ghent University. It is assumed that these participants all have a B1 level of English according to the Common European Framework of Reference for Languages (Council of Europe, n.d.), as is expected at the end of general secondary education in Flanders (Onze Taal, n.d.). For privacy reasons, all data mentioning these participants were anonymised. In the end, a total of nineteen students participated in the experiment. They were randomly split into a group of eleven, acting as the experimental group and a group of eight, acting as the control group. However, three of the participants in the experimental group were not able to finish the task that was assigned to them, which caused them to be omitted from the research. In the end, a total of sixteen students participated in the experiment, eight of them acting as an experimental group and eight of them as a control group.

All participants first had to take part in a small pre-test of about two to three minutes, in which they read out loud ten English sentences with multiple risk factors for speakers of Dutch. The students were first allowed to hear how these sentences were pronounced by a Dutch speaker who is proficient in English before pronouncing them themselves. The reason being that the ELSA Speak app also allows the users to listen to the sentence before they pronounce it themselves. The same principle applied to the post-test. Section 3.3.1 further explains the design of the pre-test and Appendix A shows the pre-test that was presented to the participants. The experimental group then started working with the free version of the ELSA app for one or two weeks in total. Before they started their lessons, they were asked to complete the 'assessment test' provided by the ELSA app itself. Thanks to this assessment test, ELSA was able to create a personal trajectory for every participant. During the first week, the participants were asked to follow this personal trajectory with five lessons each day. During the second week, they were given the choice to stop or to continue working with the app. If they continued, they were urged to follow the lessons provided by the button 'study by topic'. Section 3.4 further explains what the ELSA app looks like and how it can be used.

To ensure that all participants did what was asked of them, they had to fill in a logbook in which they indicated which lessons they had followed every day (Appendix B). Unfortunately, three participants were not able to finish the assigned tasks and were thus omitted from the experiment. When the two weeks had passed, all participants were asked to partake in a post-test. The design of the post-test was entirely the same as the pre-test, but with different sentences. Section 3.3.2 further explains the design of the post-test and Appendix D shows the post-test that was presented to the participants.

When all pre- and post-test recordings were completed, they were evaluated and given a score out of twenty by an expert in the field, that is to say by a native speaker of English who teaches English courses in the study programme Applied Linguistics at Ghent University. The pre- and post-test scores were compared to each other, as were the scores of the experimental and control groups, to discover whether those scores had improved from pre- to post-test and whether it was in fact the app that had effectuated those improvements. Apart from the general scores, the individual sounds that were specifically difficult for speakers of Dutch were also observed to discover whether there was to be noticed an improvement from pre- to post-test.

A week after the final post-tests were taken, the experimental group received a link to a Google form with ten statements asking them to indicate their agreement on a five-point Likert scale. The statements ranged from how they had perceived their training to whether they would recommend the app to others. The further design of this questionnaire is elaborated on in section 3.5.

3.3 Design of the pre- and post-test

As mentioned in section 2.1.2, the sentences included in the pre- and post-tests were specifically chosen to contain various pronunciation difficulties, as presented in Kruitbosch (2020), which native Dutch speakers experience when speaking English. Once again, it has to be noted that these difficulties are based on speakers of Dutch originating from the Netherlands, whereas the participants of this research all originated from Flanders, Belgium (section 2.1.2). Appendices A, C, D and E represent the exact pre- and post-tests that were presented to the participants and a summary of the types of mistakes incorporated in the pre- and post-tests. In what follows, each sentence will be further dissected to find out where exactly the pronunciation difficulties lie.

3.3.1 Pre-test

(S1) Did you know that I had to return my new favourite black sweater to the store? [14]

The first sentence in the pre-test already contains multiple difficulties, including the phrase 'Did you', which is often assimilated wrong by Dutch students of English. Another difficulty lies in the contrast between the sounds /æ/ and /e/, which are especially difficult to tell apart for Dutch students. This contrast is clearly visible in the words 'that', 'had', 'black' and 'sweater'. Another challenging contrast for Dutch students is the fortis-lenis contrast of the sounds /f/ and /v/ and the sounds /t/ and /d/, which are visible in the words 'favourite' and 'sweater'. The fortis sounds /f/ and /t/ are often weakened into the lenis sounds /v/ and /d/ respectively, but the opposite is also true. The words 'you', 'to' and 'new' all contain the sound /u:/, which Dutch students often confuse with the /ʊ/- sound.

---

[14] The colours indicate the types of mistakes that could have been made in each sentence. The types of mistakes for each colour can be found in Appendices C and E.

The /ð/-sound is then often mispronounced as /d/, especially in the function word 'the'. Lastly, Dutch students tend to struggle with the /r/-sound, which is supposed to be an alveolar approximant, but is often replaced by a uvular /r/ by speakers of Dutch. This /r/-sound can be found in the words 'return' and 'favourite'.

(S2) I think one of the previous customers must have tried it on, tearing apart the seams without noticing.

A first pitfall in this sentence is the /θ/-sound in the word 'think', which Dutch speakers often turn into either a /t/-, an /s/- or an /f/-sound. The /ð/-sound in the words 'the' and 'without' are often also mispronounced as /d/. Similarly, the /t/-sounds in 'customers', 'it', 'tearing', 'apart', 'without' and 'noticing' are also often replaced by /d/. The /v/-sound in the word 'previous' sometimes turns into its lenis-counterpart /f/ as well. At the same time, the /r/ in 'previous' has a tendency to sound uvular instead of alveolar, as with 'tried' and 'tearing'. 'Tearing' then again contrasts with 'have' because of the difficulty to distinguish between the /æ/- and /eə/-sounds.

(S3) I asked the girl behind the counter for a refund, but she only spoke Dutch and French.

Apart from the aforementioned pitfalls, such as the /r/-sound in 'for a', 'refund' and 'French', the /ð/-sound becoming a /d/ in 'the', the /t/-sound becoming a /d/ in 'asked' and 'counter' and the /f/-sound becoming a /v/ in 'refund', a few new problems also pose a risk for speakers of Dutch. The initial /g/-sound of a word is often replaced by its fortis counterpart /k/, for example in the word 'girl'. Some Dutch speakers also experience problems with the /dʒ/-, /tʃ/- and /ʃ/-sounds, replacing them /ts/ or /s/, respectively, as in the words 'she', 'Dutch' and 'French'.

(S4) I felt like a complete fool having to explain myself through gestures.

The words that could potentially pose a problem in this sentence and that have not yet been mentioned are 'felt', 'like' and 'myself'. With 'felt' and 'myself', Dutch speakers tend to insert an /ə/-sound between the /l/ and the non-alveolar consonant that follows. This is also called an "epenthesis" (section 2.1.2). With 'like' and 'myself', Dutch speakers tend to pronounce the /aɪ/-sound a little too long before the fortis-sound. Other issues in this sentence are: the contrast between /f/ and /v/ ('felt', 'fool'), between /e/ and /æ/ ('felt', 'myself', 'gestures') and between /u:/ and /ʊ/ ('fool', 'through', 'to'). Moreover, the

mispronunciation of the /r/-sound ('trying', 'through') and the /θ/-sound ('through') can also be problematic.

(S5) Luckily, another woman walked by asking me if I needed any help with my purchase.

This sentence does not contain as many possible risk factors as previous or upcoming sentences, but the word 'woman' seems to be particularly difficult for Dutch speakers because of the /ʊ/- sound that is often confused with the /uː/-sound. Furthermore, Dutch speakers will also struggle to pronounce the /ð/-sound in 'another', the /e/-sound in 'any' and 'help' and the /t/-sound in 'walked'. Lastly, 'help' is often considered to be the poster word of epenthesis.

(S6) When I explained the problem and showed her the receipt, she insisted on finding a solution.

This sentence solely contains errors that have already been discussed, such as the contrast between /e/ and /æ/ ('when', 'and'), the mispronunciation of /ð/ ('the'), /ʃ/ ('showed', 'solution', 'she') and /uː/ ('solution') and the contrast between /t/ and /d/ ('receipt', 'insisted') and between /f/ and /v/ ('finding'). Furthermore, the /r/ may again pose problems in the words 'receipt' and 'problem'.

(S7) I was delighted when she told me I could exchange the sweater for anything of the same value.

This sentence combines many of the previous possible risk factors into one simple sentence. It contains the contrast between /t/ and /d/ ('delighted', 'sweater'), between /e/ and /æ/ ('when', 'sweater', 'anything', 'value'), between /ʊ/ and /uː/ ('could', 'value') and between /v/ and /f/ ('value'), but it also contains the possible mispronunciation of /ð/ ('the'), /ʃ/ ('she'), /θ/ ('anything') and word-final /dʒ/ ('exchange'). Moreover, it includes the possible overlong pronunciation of /aɪ/ ('delighted') and the possibility of pronouncing a uvular /r/ ('for anything').

(S8) That is when a pair of boots of the exact same value caught my attention.

Once again, this sentence does not particularly show anything new. The dangers not only lie in the contrast between /eə/, /æ/ and /e/ ('that', 'when', 'exact', 'value', 'attention',

43

'pair'), between /ʊ/ and /u:/ ('boots', 'value'), between /f/ and /v/ ('value') and between /t/ and /d/ ('attention'), but also in the mispronunciation of /ð/ ('that', 'the') and /ʃ/ ('attention'). Additionally, the /r/ could also pose a problem in the phrase 'pair of'.

(S9) Fortunately for me, they still had size eleven in stock at the store, so I could try them on.

The contrasts between /f/ and /v/ ('fortunately', 'eleven'), between /æ/ and /e/ ('had', 'eleven', 'at', 'them') and between /ʊ/ and /u:/ ('could') are clearly evident in this sentence. Not to mention, the words 'they', 'the', 'them' and 'try' are all prone to being mispronounced because of the /ð/- and /r/-sounds.

(S10) When I eventually felt satisfied with my new perfect pair of boots, I thanked the woman and left the store.

The last sentence then again combines multiple possible risk factors for speakers of Dutch, such as the confusion between /e/, /eə/ and /æ/ ('when', 'eventually', 'felt', 'satisfied', 'thanked', 'pair', 'and', 'left'), between /f/ and /v/ ('eventually', 'felt', 'satisfied', 'perfect'), between /t/ and /d/ ('satisfied', 'thanked') and between /ʊ/ and /u:/ ('new', 'boots', 'woman'). The word 'felt' may be subject to an epenthesis as in sentences four and five. 'Pair of' may constitute difficulty due to the /r/-sound and the /th/-sound in 'with', 'the' and 'thanked' is also often prone to mispronunciation.

3.3.2 Post-test

(S1) Did you take a look at the fridge door yet like I asked you to?

The first sentence of the post-test already contains multiple possible errors for speakers of Dutch, including the confusion between /ʊ/ and /u:/ ('look', 'you', 'to'), between /t/ and /d/ ('at', 'asked'), between /æ/ and /e/ ('at', 'yet') and between /f/ and /v/ ('fridge'). The word 'fridge' also contains the risks of mispronouncing the /r/-sound and the /dʒ/-sound at the end of the word. 'Did you' at the beginning of the sentence is often assimilated wrong and the /aɪ/-sound in 'like' is regularly pronounced too long. Lastly, there is, of course, still the danger of mispronouncing the /ð/-sound as a /d/ in 'the'.

(S2) Ever since last night, I have not been able to close it all the way and I think I know why.

This sentence contains the risk of mispronouncing the /ð/- and the /θ/-sound in 'the' and 'think', respectively, as well as the risk of confusing the pairs /e/ and /æ/ ('ever', 'have', 'and'), the pairs /f/ and /v/ ('ever'), the pairs /ʊ/ and /u:/ ('to') and the pairs /t/ and /d/ ('last', 'night', 'not', 'it'). The word 'night' may also be prone to mispronunciation because of the /aɪ/-sound that is pronounced too long.

(S3) I saw one of your guests at the party last night slamming the door shut a little too enthusiastically.

On the one hand, the difficulties in sentence three lie in the contrast between /e/ and /æ/ ('guests', 'at', 'slamming', 'enthusiastically'), between /ʊ/ and /u:/ ('too', 'enthusiastically') and between /t/ and /d/ ('at', 'party', 'last', 'night', 'shut', 'little'). On the other hand, there is a possibility of mispronouncing /g/ as /k/ in 'guests', /ð/ as /d/ in 'the', /θ/ as /t/, /s/, or /f/ in 'enthusiastically' and /ʃ/ as /s/ in 'shut'. Lastly, the word 'night' may again be prone to mispronunciation because of the /aɪ/-sound that is pronounced too long.

(S4) I actually already tried talking to her, but she was only able to speak Dutch and French.

The number one complication in this sentence is the /tʃ/-sound at the end of the words 'Dutch' and 'French'. Additionally, the /ʃ/ in 'she' and the /r/ in 'already', 'tried' and 'French' also have a chance of being mispronounced. As per usual, the contrast between /e/ and /æ/ ('actually', 'and'), between /ʊ/ and /u:/ ('to'), between /t/ and /d/ ('already', 'but') and between /f/ and /v/ ('French') may once again pose problems for some Dutch speakers.

(S5) I felt like a complete fool trying to make her understand what the problem was.

Even though 'felt' is a rather small word, it does contain a few complications Dutch speakers may struggle with, such as the contrast between /f/ and /v/ at the beginning (as with 'fool' as well), the contrast between /e/ and /æ/ and the epenthesis between /l/ and the following non-alveolar consonant /t/. As in the first sentence, the /aɪ/-sound in 'like' is prone to being pronounced too long and the /r/- and /ð/-sound in 'her understand', 'trying', 'problem' and 'the' may be mispronounced as well. Finally, there is still the

contrast between /ʊ/ and /uː/ ('fool', 'to') and between /t/ and /d/ ('what') that causes difficulties.

(S6) Eventually, I just had to give up, because she clearly could not understand a word I said.

Sentence six contains a few possible risk factors, but compared to other sentences, there are not as many. The contrasts between /f/ and /v/ ('eventually'), between /e/ and /æ/ ('eventually', 'had', 'understand', 'said'), between /ʊ/ and /uː/ ('could') and between /t/ and /d/ ('had', 'not', 'understand', 'said') are immediately noticeable. Moreover, the /g/ in 'give' and the /ʃ/ in 'she' could be mispronounced as /k/ and /s/, respectively.

(S7) Still, I would like for you to take another look at the fridge door as soon as possible.

The main contrast in this sentence is the one between /ʊ/ and /uː/, which is visible in the words 'you', 'to', 'look', 'soon' and 'would'. Other contrasts to be found in this sentence are the ones between /e/ and /æ/ ('at', 'as') and between /f/ and /v/ ('fridge'). As in the first sentence, the word 'fridge' also contains the risks of mispronouncing the /r/ and the /dʒ/-sound at the end of the word and the /aɪ/-sound in 'like' is regularly pronounced too long. The /ð/-sound in 'another' and 'the' could be mispronounced as /d/ as well.

(S8) I would be delighted if I could cook my favourite meal tomorrow without the ingredients going bad.

Once again, the main part that can cause confusion in this sentence is the contrasts between /ʊ/ and /uː/ ('would', 'could', 'cook'), between /t/ and /d/ ('delighted', 'favourite', 'without', 'ingredients'), between /e/ and /æ/ ('bad') and between /f/ and /v/ ('favourite'). /R/, /ð/ and /g/ could also be mispronounced in the words 'favourite', 'tomorrow', 'ingredients', 'without', 'the' and 'going'. Lastly, 'delighted' may include a prolongation of the /aɪ/-sound.

(S9) Should you not be able to repair it, I'll call the maintenance company for help tomorrow.

The penultimate sentence includes some of the errors that have already been mentioned in previous sentences, such as the mispronunciation of /r/ ('repair', 'tomorrow'), /ð/ ('the')

and /ʃ/ ('should'). It also contains the confusion between /e/ and /eə/ ('help', 'repair'), between /ʊ/ and /u:/ ('you', 'to', 'should') and between /t/ and /d/ ('not', 'it', 'maintenance'). However, it also encompasses rarer mistakes, such as the possible assimilation error in 'should you' and the possible epenthesis in 'help'.

(S10) They will probably know how to fix the issue for a small price.

The last sentence is reasonably short and thus does not contain that many possible risk factors. First of all, the /ð/-sound in 'they' and 'the', the /ʃ/-sound in 'issue' and the /r/-sound in 'probably', 'for a' and 'price' are prone to mispronunciation. 'Price' also contains the /aɪ/-sound, which is easily pronounced too long. Lastly, 'to' and 'fix' embody the contrasts between /ʊ/ and /u:/ and between /f/ and /v/, respectively.

3.4 Short summary of ELSA's functions

Section 2.4 already gave a brief overview of what ELSA entails and what the app is specifically used for. Before the participants started using the app, the researcher briefly explained which features of the app the participants were allowed to use for the experiment. First, they had to download the free version of the app from the Play Store on an Android phone or the App Store on an iPhone. When they first opened the app, they had to log in, either through their Facebook account or through their e-mail address and a password. Once logged in, ELSA asked them a few questions to discover their level of English proficiency. They were first asked to indicate their mother tongue, which had to be Dutch for this particular experiment. Afterwards, they were asked why they were practising their English, to which the researcher told them to answer 'education'. They also had to estimate their own English proficiency, going from 'beginner' to 'intermediate' to 'advanced'. Assuming that their level of English was at a B1 level according to the Common European Framework of Reference for Languages (Council of Europe, n.d.), the researcher urged them to indicate 'intermediate'. Lastly, the participants had to choose for themselves how long and when they wanted to practice with the app each day. The minimum was set at five lessons every day for a week; the maximum could be decided for themselves, but no one exceeded the two weeks' time frame. Thanks to the logbook (Appendix B), the number of lessons for each participant could be recorded.

Once all the administration was taken care of, the participants landed on the homepage of the app. The researcher first wanted them to go to their own profile and scroll down to take the 'assessment test' (Figures 2 and 3), in order for ELSA to determine which items of the English language were particularly difficult for them individually. The test made them listen to and repeat sixteen English sentences. Once the test had been completed, they were able to start their ELSA journey for one or two weeks, depending on how motivated they were.

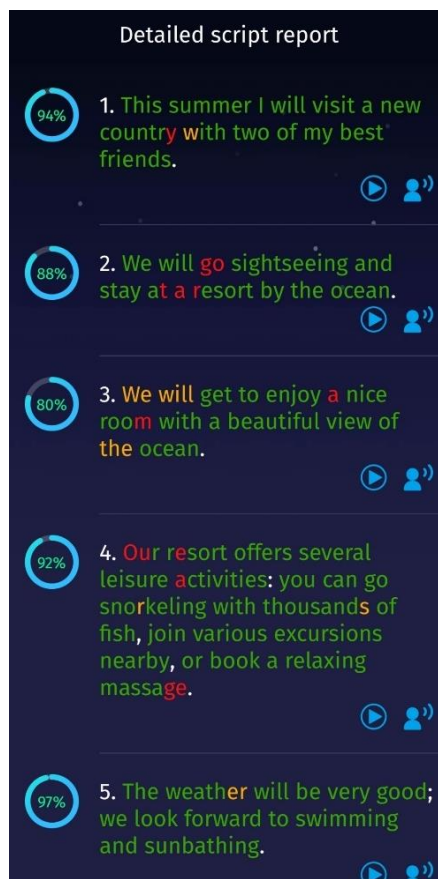*Figure 2: Elsa app assessment test score*   *Figure 3: Elsa app assessment test example*

During the first week, the participants were asked to follow their individual trajectories, which had planned five lessons a day for seven days in total. They simply had to click on 'improve pronunciation' when on the homepage (Figure 4) to view their individual lessons (Figure 5). These lessons contained multiple small assignments to be fulfilled at the end of the day. The participants either had to pronounce a few words or sentences (Figure 6), watch a small one-to-two-minute video (Figure 7), listen to and indicate the word they had just heard (Figure 8) or read a small explanation about how to pronounce a certain sound (Figure 9). Sometimes ELSA also included a small conversation in which the participant had to read and record their part of the conversation (Figure 10). During the second week, the participants were free to choose whether they wanted to continue

using the app. Those that did, were urged to click on 'study by topic' when on the homepage (Figure 4) and to choose whichever topic appealed to them. A few examples of topics to choose from are: 'basics', 'introductory English', 'lifestyle' and 'small talk' (Figure 11). Within these topics, there were also subtopics the participants could choose from (Figure 12). Because they were working with the free version of the app, however, certain lessons could not be accessed. Nevertheless, there were enough lessons to choose from so that they could continue on for the rest of the week. These lessons contained the same types of exercises as mentioned in the individual trajectory.
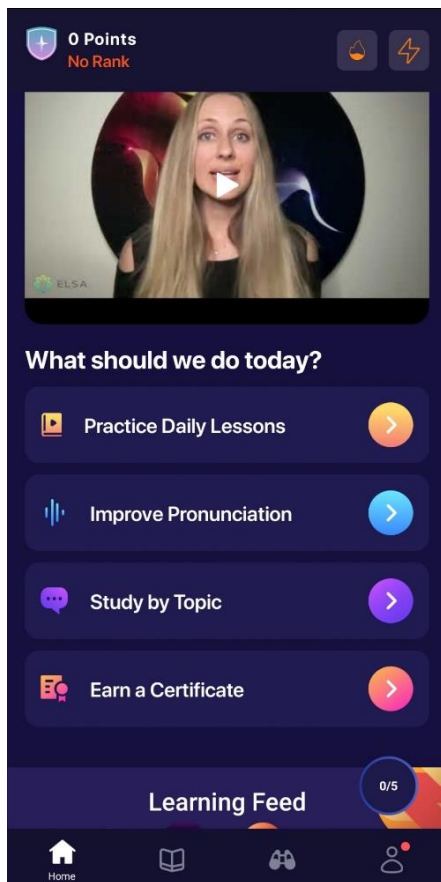

Figure 4: Elsa app homepage

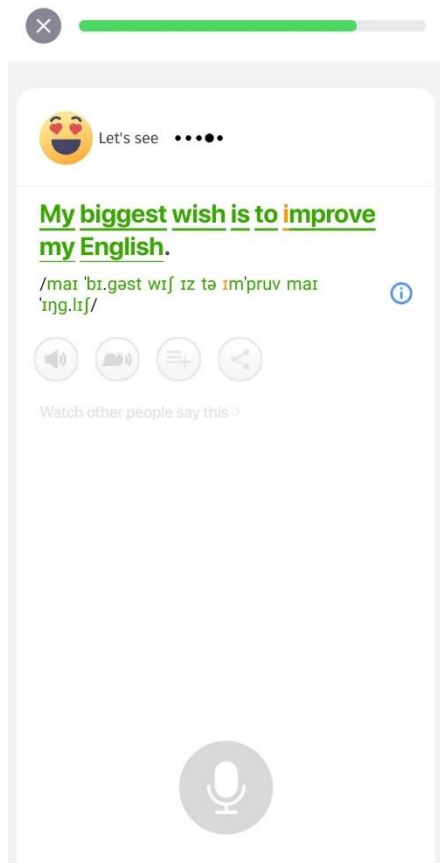*Figure 5: Elsa app individual lessons*

Figure 6: Elsa app pronounce a sentence
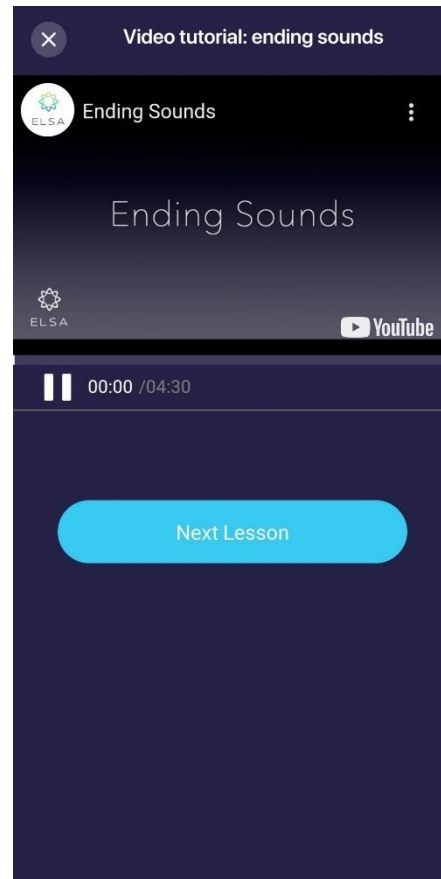

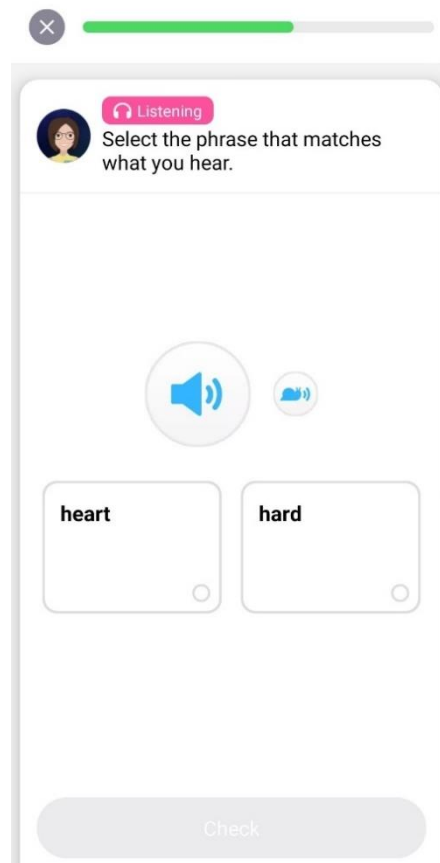*Figure 7: Elsa app watch a video*

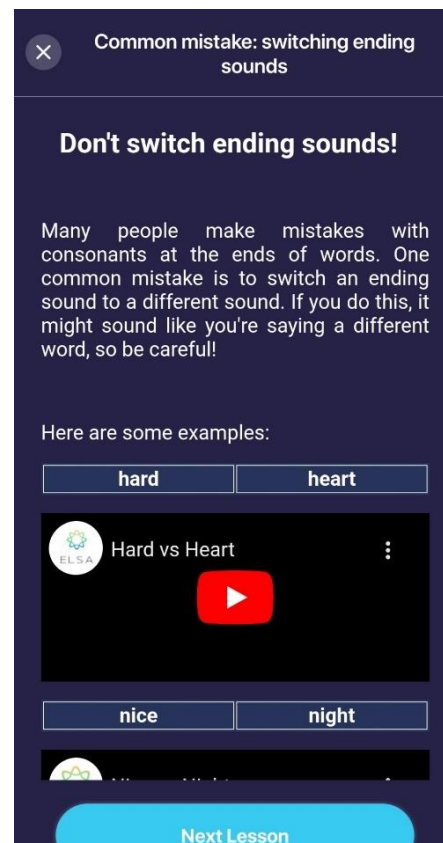*Figure 8: Elsa app listen and indicate*

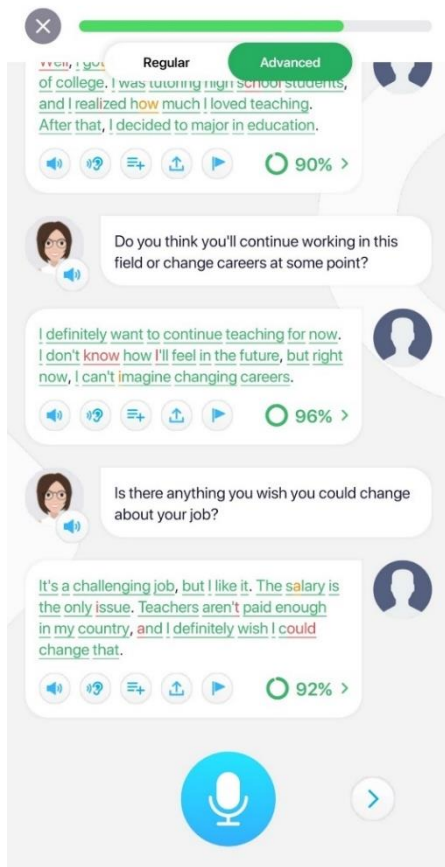*Figure 9: Elsa app read an explanation*

Figure 10: Elsa app conversation

Figure 11: Elsa app topics

Figure 12: Elsa app subtopics

3.5 Questionnaire

As mentioned in sections 3.1 and 3.2, this section will describe the questionnaire that was presented to the experimental group in terms of the attitudes and perceptions they had towards the ELSA app. The eight participants involved were asked to fill in a Google form in which they were presented with ten statements about how they had perceived their training with the ELSA app. They were prompted to indicate their agreement by means of a five-point Likert scale in which one equalled 'strongly disagree' and five equalled 'strongly agree'. The ten statements that were included in the form are presented below. The original form (in Dutch/English) can be found in Appendix F.

The app is user-friendly and has an easy interface.
I liked the fact that the lessons were personalised.
I would have preferred more pronunciation lessons instead of listening and viewing exercises.
I understood the feedback the system gave me.
I think the feedback the system gave me was not always correct.
I feel like my pronunciation improved while practising with the app.
I would use the app again to improve my pronunciation.
I would use the app if it were offered in other languages as well.
I would recommend the app to others.
I was disappointed by the fact that I had to pay for the app if I wanted to do more than five lessons a day.

Table 4: Statements relating to the perceptions and attitudes towards the ELSA app.

3.6 Statistical tests

To solve the first research question, a paired samples t-test was executed. Since there were two groups to be compared (pre- and post-test scores of the experimental group) and the dependent variable was a ratio variable, a t-test had to be conducted (van der Zee, 2016). To ensure that the t-test could indeed be carried out, in other words, to guarantee that the data were normally distributed, a Shapiro-Wilk's test had to be performed. The t-test also needed to be paired because the scores of the same group (the experimental group) had to be compared. The control group underwent the same test, as the results of all groups should be analysed similarly to allow accurate and reliable comparisons and as the assumption of normality needs to be validated for the second research question.

The second research question required a Wilcoxon signed-rank test. At first, it seemed that an independent samples t-test had to be conducted because there were two groups to be compared (post-test scores of the experimental and control group) and the dependent variable was a ratio variable (van der Zee, 2016). The t-test also needed to be independent because the scores of different groups (experimental and control groups) had to be compared. However, the Shapiro-Wilk's test showed that the data were not normally distributed, which meant a t-test could not be performed. Instead, a Wilcoxon signed-rank test needed to be carried out, as this is the non-parametrical equivalent of the students' t-test (Statistisch Handboek Studiedata, 2021).

To solve the third research question, a Chi-square test was selected. Since there were two groups to be compared (pre- and post-test of the experimental group), the dependent variable was nominal and there were multiple characteristics, there was no doubt that it needed to be a Chi-square test.

For the fourth research question, no statistical test was conducted. Instead, each statement was considered separately. The first five statements dealt with the characteristics of the app, which are similar to the statements mentioned in previous studies (Inceoglu et al., 2020; Liakin et al., 2014; van Doremalen et al., 2016; Wang & Young, 2014, 2015). First of all, any app needs to have an easy-to-use interface, which is why the first statement is of the utmost importance. The content of the exercises is, of course, equally as important, if not more important. That is why it is useful to know whether the students appreciated the personalised and diverse lessons. The literature study (section 2.3) also particularly focussed on the feedback that was provided by the ASR-systems, which is why it is important to know whether these students appreciated the implicit feedback that ELSA was able to provide. On a second level, it may also be interesting to understand how the students perceived their own progress. Did the students feel like they had improved after practising with the app? Would they use it again to improve their pronunciation? Would they use it if it were offered in other languages? Would they recommend it to anybody else? And were they disappointed to learn that there was a paying version with more content?

Chapter 4: Results

The following section is designed to try and formulate a response to the four research questions that were raised in the introduction and which were further explained in section 3.1. The first part (section 4.1) will formulate a reply to the first two research questions which set out to verify whether the experimental group had improved significantly from pre- to post-test and whether that group had improved more significantly than the control group. Section 4.2 will further elaborate on the third research question and will, therefore, focus on whether the experimental group was able to correct some of the typical mistakes Dutch speakers tend to make in English. The last section (section 4.3) will then concentrate on some of the attitudes and perceptions the experimental group had towards the ELSA Speak app.

4.1   General scores

As mentioned in the methodology section, all participants, both in the experimental group and in the control group, received a grade out of 20 from a native speaker of English on both the pre- and post-test. The scores for those tests can be found in the table and figures below.

Participant	ELSA	Score pre-test	Score post-test
1	X	13/20	12/20
2	X	15/20	16/20
3	X	14/20	15/20
4	X	10/20	11/20
5	X	14/20	14/20
6	X	12/20	13/20
7	X	12/20	13/20
8	X	12/20	14/20
9		11/20	12/20
10		13/20	14/20
11		13/20	13/20
12		13/20	13/20
13		15/20	16/20
14		13/20	13/20
15		13/20	14/20
16		14/20	14/20

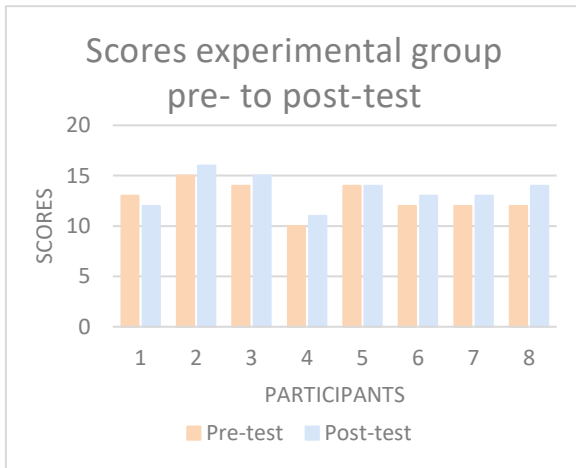*Table 5: General scores for the pre- and post-test.*

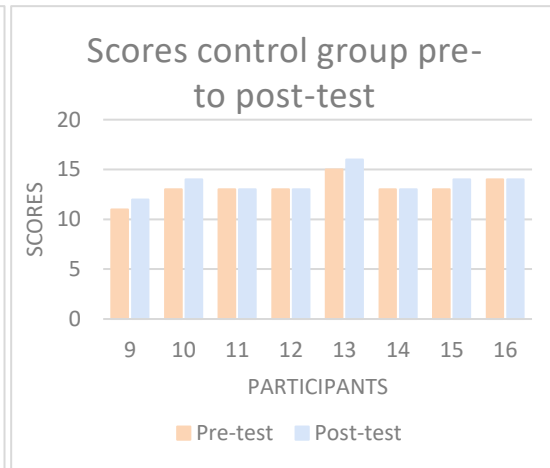*Figure 12: Scores for the experimental group from pre- to post-test.*

Figure 14: Scores for the control group from pre- to post-test.

Table 5 and Figures 13 and 14 show that the experimental and control groups scored quite similarly on both the pre- and post-test. It can be inferred from the graphs that the experimental group scored between 10 and 15 on the pre-test and between 11 and 16 on the post-test. The control group obtained similar results with a score between 11 and 15 for the pre-test and a score between 12 and 16 for the post-test.

To examine whether the experimental group had improved significantly from pre- to post-test, a Shapiro-Wilk's test was first required to determine whether the data were distributed normally. For this purpose, the scores of the pre-test were first subtracted from the scores of the post-test (the postminpre variable). Figure 15 then depicts the results of the Shapiro-Wilk's test for the postminpre variable and shows an outcome that is narrowly non-significant (p=0.054). A non-significant result for the Shapiro-Wilk's test signifies that a one-sided paired samples t-test may be performed to determine whether the experimental group's pre- and post-test show a significant difference. It appears that Figure 16 then does indeed show a significant outcome (p=0.024), which means that the first hypothesis can be confirmed:

C1: After having used the ELSA app for two weeks, the experimental group improved their general pronunciation performance.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
postminpre	,361	8	,003	,826	8	,054

a. Lilliefors Significance Correction

Figure 15: Normality test for the experimental group from pre- to post-test.

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	pretestapp - posttestapp	-,750	,886	,313	-1,491	-,009	-2,393	7	,024	,048

Figure 16: Paired samples t-test for the experimental group from pre- to post-test.

To ensure that it was indeed the ELSA Speak app that generated this improvement in pronunciation, the control group underwent the same test. Their pre-test results were first subtracted from their post-test results, which formed the postminprenonapp variable. The Shapiro-Wilk's test for this variable, presented in Figure 17, depicts a clearly significant result (p<0.001), which means that the data for the pre- and post-test are not normally distributed. This also means that the one-sided paired samples t-test cannot be performed. Instead, the data underwent a Wilcoxon signed-rank test. Figure 18 depicts the two-sided outcome for this test, but to find the one-sided result, this outcome must first be divided by two. In this case, the pronunciation of the control group also significantly improved from pre- to post-test (p=0.023). This could have multiple explanations and will therefore be further discussed in Chapter 5. Nevertheless, to be absolutely certain, the post-tests of both groups will be compared by conducting an unpaired t-test.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Postminprenonapp	,325	8	,013	,665	8	<,001

a. Lilliefors Significance Correction

Figure 17: Normality test for the control group from pre- to post-test.

Test Statistics[a]

	posttestnietapp - pretestnietapp
Z	-2,000[b]
Asymp. Sig. (2-tailed)	,046

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Figure 18: Wilcoxon signed-rank test for the scores of the control group from pre- to post-test.

Once again, the post-tests first needed to undergo a Shapiro-Wilk's test to ensure that the data were normally distributed and to ensure that the one-sided unpaired t-test could indeed be performed. Figure 19 shows the outcomes of the Shapiro-Wilk's tests for both the post-test of the experimental group (posttestapp variable) and for the post-test of the control group (posttestnonapp variable). It can be concluded that both tests were non-

significant (posttestapp: p=0.975; posttestnonapp: p=0.245), which once again means that the data are normally distributed. Apart from a Shapiro-Wilk's test, a Levene test also needed to be carried out to ensure that the one-sided unpaired t-test could indeed be performed. For this test, a new variable, the valueposttest variable, was designed, which included all post-test scores of both the experimental and the control group. The Levene test in Figure 20 then showed a non-significant result (p=0.862), which means that the one-sided unpaired t-test could indeed take place. Finally, Figure 21 reveals a non-significant result for the one-sided unpaired t-test (p=0.431), showing, therefore, that the second hypothesis needs to be rejected.

C2: The experimental group was not able to improve more significantly than the control group.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
posttestapp	,128	8	,200[*]	,983	8	,975
posttestnonapp	,251	8	,147	,892	8	,245

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure 19: Normality test for the post-tests of the experimental and control groups.

ANOVA

Valueposttest

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,063	1	,063	,031	,862
Within Groups	27,875	14	1,991		
Total	27,938	15			

Figure 20: Levene test for the post-tests of the experimental and control groups.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p			Lower	Upper
Valueposttest	Equal variances assumed	,851	,372	-,177	14	,431	,862	-,125	,706	-1,638	1,388
	Equal variances not assumed			-,177	12,904	,431	,862	-,125	,706	-1,650	1,400

Figure 21: Independent samples t-test for the post-tests of the experimental and control groups.

4.2   Specific difficulties in English pronunciation for speakers of Dutch

Apart from testing whether the participants in the experimental group underwent a general improvement from pre- to post-test, this research was also designed to examine whether the experimental group was able to rectify some of the pronunciation mistakes they tended to make in English. To try and prove this, the participants in the experimental group were evaluated on a few criteria, that is to say, the most significant errors stated in Kruitbosch (2020) (sections 2.1.2 and 3.3). Table 6 and Figure 22 below depict the frequency of each type of mistake a participant made during the pre- and post-test.

Type of mistakes	Number of participants that made the mistake in the pre-test	Number of participants that made the mistake in the post-test
Loss fortis-lenis contrast	8	5
/æ/ vs /e/	0	0
/ʊ/ vs /u:/	1	0
/θ/ vs /t/	2	5
epenthesis of /ə/	0	0
uvular /r/	2	1
/ð/ vs /d/	8	4
Assimilation	1	2
incorrect /æ/ and /ʊ/	1	0
/dʒ/ vs /tʃ/ and /ts/	0	0
/g/ vs /k/	0	0
/θ/ vs /s/ and /f/	0	0
/æ/ vs /e/ vs /eə/	2	1
/ʃ/ vs /s/	1	0
/aɪ/ that is too long before fortis	6	4

Table 6: Types of mistakes each participant made during the pre- and post-test
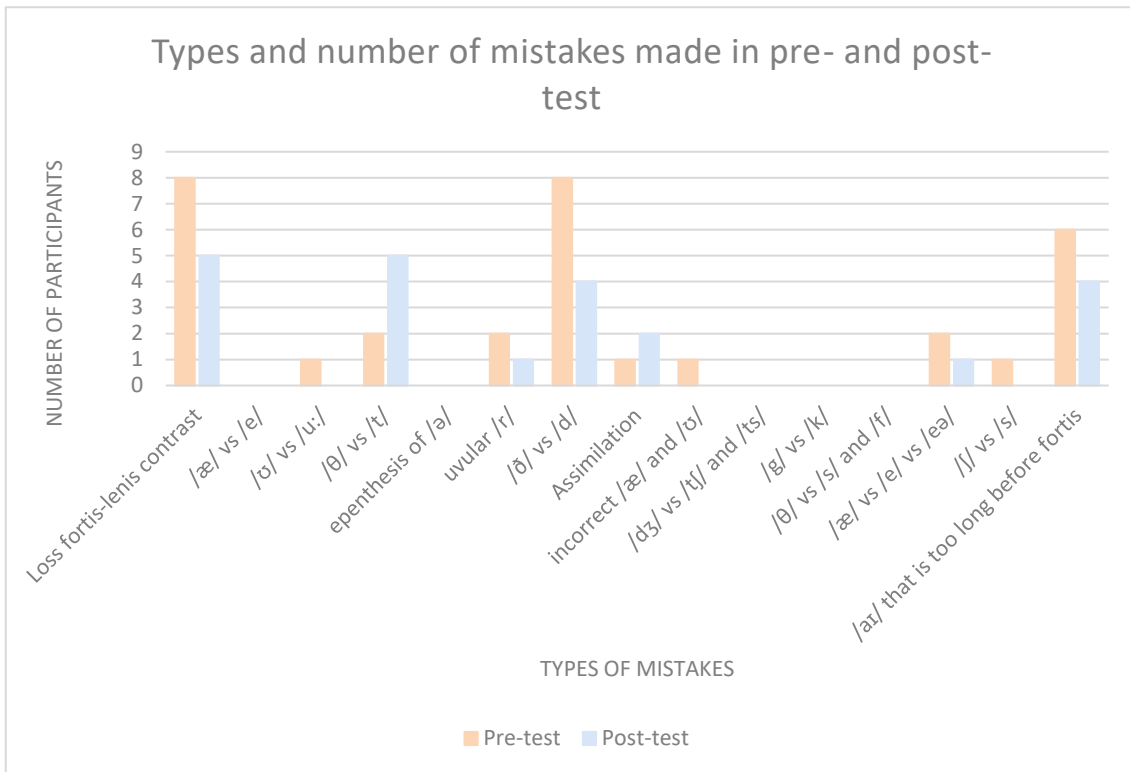
Figure 22: Frequency of the types of mistakes made by the experimental group from pre- to post-test.

It can be inferred from the graph that certain mistakes are more common than others within the experimental group. The three most common mistakes are the contrast between fortis and lenis sounds, especially between the sounds /t/ and /d/ and between the sounds /f/ and /v/, the mispronunciation of /ð/ as /d/ and the /aɪ/-sound that is pronounced overlong. Others, on the other hand, seemed to pose no problem at all. For instance, the experimental group did not make any mistakes, neither in the pre-test nor in the post-test, with regard to the contrast between /æ/ and /e/ or between /g/ and /k/ in initial position. Nor did they mispronounce /dʒ/ or /tʃ/ as /ts/ or /θ/ as /s/ or /f/. Lastly, no one added an /ə/ between an l-sound and a following non-alveolar consonant, as in the words 'help' or 'felt'. In addition, the graph also clearly shows that the three most common errors greatly declined from pre- to post-test. This could indicate that the app did indeed aid in rectifying some of the typical mistakes the Dutch make against the English language. Nevertheless, to ensure that this result was significant, a chi-square test had to be performed. Figure 23 depicts the results of this chi-square test, but with a p-value of 0.733; the outcome was clearly non-significant. Unlike the third hypothesis stipulated, the experimental group did not improve some of the typical mistakes they tend to make against the English language. Subsequently, the third hypothesis also had to be rejected.

C3: After practising with the ELSA app for two weeks, the participants were not able to rectify some of the mistakes the Dutch typically make against the English language.

pre_or_post_test * Types_of_mistakes * Frequency Crosstabulation

Frequency: Total
Count

		/æ/ vs /e/ vs /ee/	/aɪ/ that is too long before fortis	/ð/ vs /d/	/ʃ/ vs /s/	/ɜ/ vs /uː/	/θ/ vs /t/	Assimilation	incorrect /æ/ and /ɒ/	Loss fortis-lenis contrast	uvular /r/	Total
pre_or_post_test	post	1	4	4	0	0	5	2	0	5	1	22
	pre	2	6	8	1	1	2	1	1	8	2	32
Total		3	10	12	1	1	7	3	1	13	3	54

Chi-Square Tests

Frequency		Value	df	Asymptotic Significance (2-sided)
1	Pearson Chi-Square	6,000[b]	5	,306
	Likelihood Ratio	7,638	5	,177
	N of Valid Cases	6		
2	Pearson Chi-Square	8,000[c]	3	,046
	Likelihood Ratio	8,997	3	,029
	N of Valid Cases	8		
4	Pearson Chi-Square	.[d]		
	N of Valid Cases	8		
5	Pearson Chi-Square	.[d]		
	N of Valid Cases	10		
6	Pearson Chi-Square	.[e]		
	N of Valid Cases	6		
8	Pearson Chi-Square	.[d]		
	N of Valid Cases	16		
Total	Pearson Chi-Square	6,068[a]	9	,733
	Likelihood Ratio	7,104	9	,626
	N of Valid Cases	54		

Figure 23: Chi-square test for the types of mistakes made by the experimental group from pre- to post-test.

4.3 Questionnaire: attitudes and perceptions towards the ELSA app

One week subsequent to the experiment, the participants in the experimental group were asked to fill in a short questionnaire with ten statements on a 5-point Likert scale about how they had perceived their pronunciation training with the ELSA app. In what follows, each statement will be discussed in further detail.

The first statement (Figure 24) dealt with the app's layout, stating that the app is user-friendly and has an easy interface. With regard to this statement, all participants seemed to be in full agreement. Six of them stated that they agreed with this statement and two of them even strongly agreed. It can therefore be inferred that the app is very easy to use and that the app's functions all prove to be entirely clear. The users can easily navigate their way around the platform.

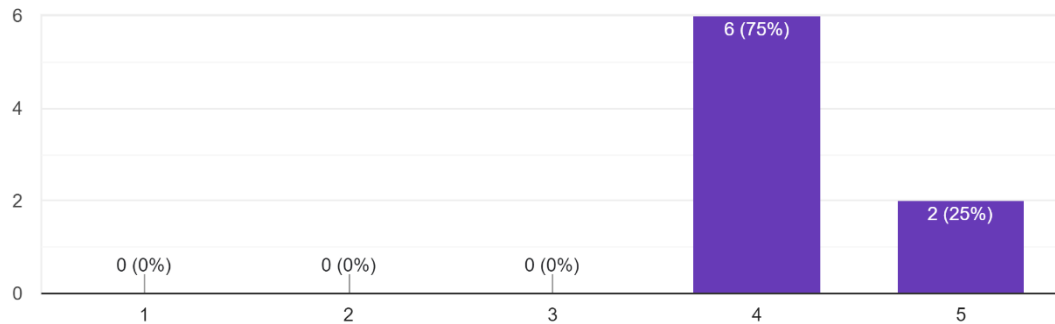De app is gebruiksvriendelijk en heeft een makkelijke interface.
8 antwoorden

Figure 24: Questionnaire statement 1

The second statement (Figure 25) also seemed to have been received rather well among most participants. Five of them either agreed or strongly agreed with the statement declaring that they liked the fact that the lessons were personalised; the remaining three stayed neutral to this statement. Perhaps the remaining three did not perceive the lessons as personalised or perhaps they preferred general pronunciation lessons. Nevertheless, it can be concluded that most participants preferred the personalised lessons ELSA is able to provide.

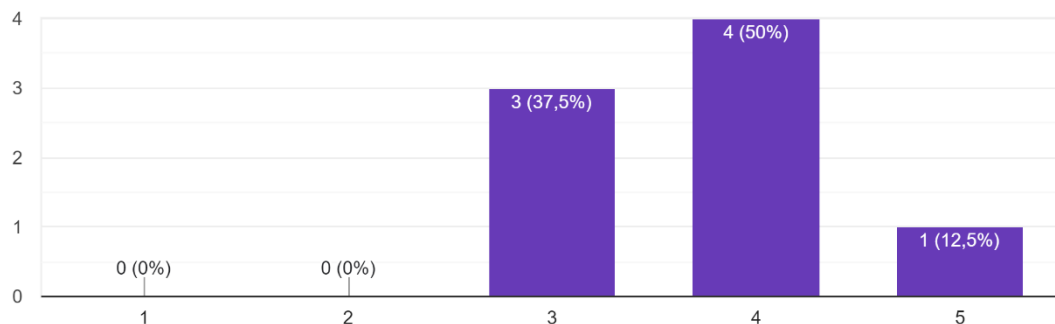Ik vond het goed dat er gepersonaliseerde lessen waren.
8 antwoorden

Figure 4: Questionnaire statement 2

Seven out of eight participants would prefer it if the app included some more actual pronunciation lessons instead of the listening and viewing exercises they often received (Figure 26). This is, of course, understandable since the users will only learn something, in this case pronunciation, through repeated practice. One person, however, did not seem to agree with this statement, from which can be deduced that he or she preferred the balance between pronunciation lessons and listening and viewing exercises that the app already provided.

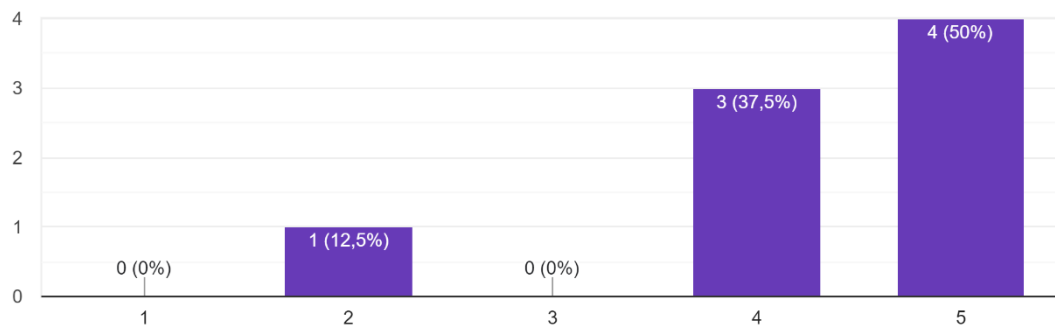Ik had liever wat meer uitspraaklessen gehad en wat minder luister- en kijkoefeningen.
8 antwoorden

Figure 5: Questionnaire statement 3

More than half of the participants declared that they understood the feedback that the system gave them (Figure 27). Two people remained neutral and one person stated they did not understand the system's feedback. Overall, however, it can be concluded that according to these participants, ELSA's ASR-system provides helpful feedback to improve one's pronunciation. On the other hand, the statement 'I think the feedback the system gave me was not always correct' proves otherwise.

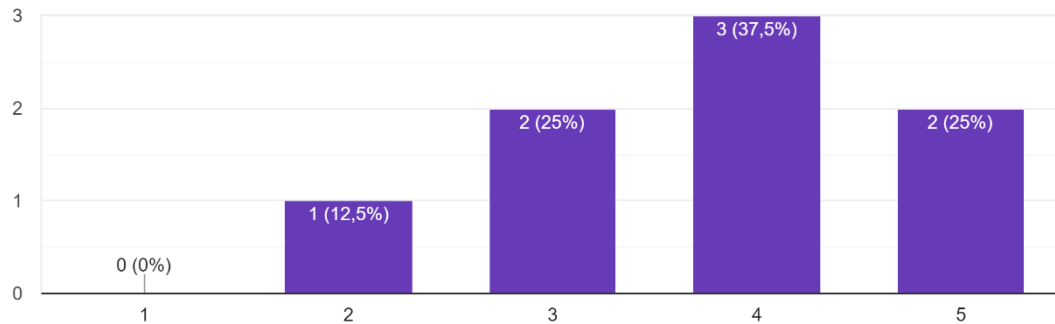Ik begreep de feedback die het systeem mij gaf.
8 antwoorden

Figure 6: Questionnaire statement 4

The opinions regarding that statement (Figure 28) were rather divided. Three people agreed, three others disagreed and two participants remained neutral. Those that agreed must have either said something to the system they thought was correct but was interpreted by the system as incorrect or the other way around. As Becker and Edalatishams (2019) already mentioned, this can be explained by the fact that ELSA's ASR-system is not yet 100% failproof. The app still contains several shortcomings in that regard.

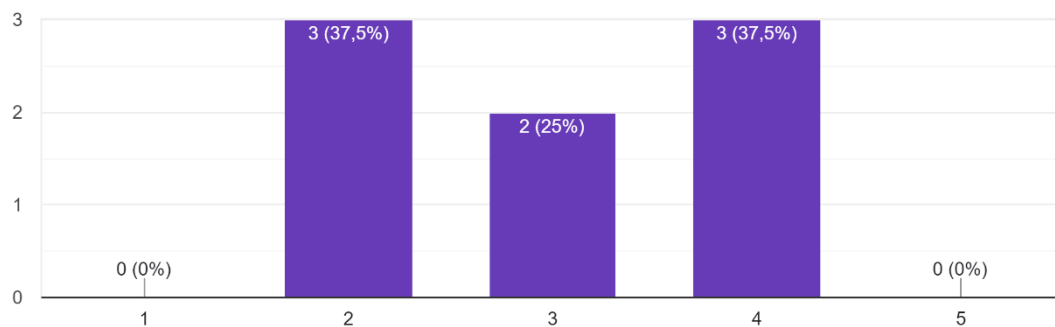Ik vond de feedback van het systeem niet altijd even correct.
8 antwoorden

Figure 7: Questionnaire statement 5

Still, half of the participants thought their pronunciation had improved after using the app for a period of one to two weeks (Figure 29). According to the paired samples t-test, there was indeed a significant improvement from pre- to post-test for the experimental group, but it was not significant in comparison to the control group. This could also explain why the other half of the participants did not think that their pronunciation had improved from pre- to post-test.

Ik heb het gevoel dat mijn uitspraak verbeterd is tijdens het oefenen met de app.
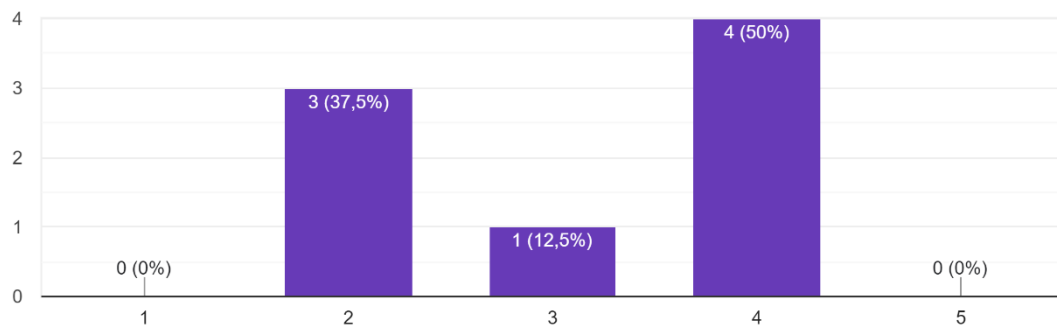8 antwoorden

Figure 8: Questionnaire statement 6

With regard to the number of participants who would want to use the app again to further improve their pronunciation, the results did not look promising (Figure 30). Half of the participants stated that they would not use the app again, another three remained neutral and only one person declared wanting to use the app again. This could have multiple reasons and is probably linked to the last statement (see below).

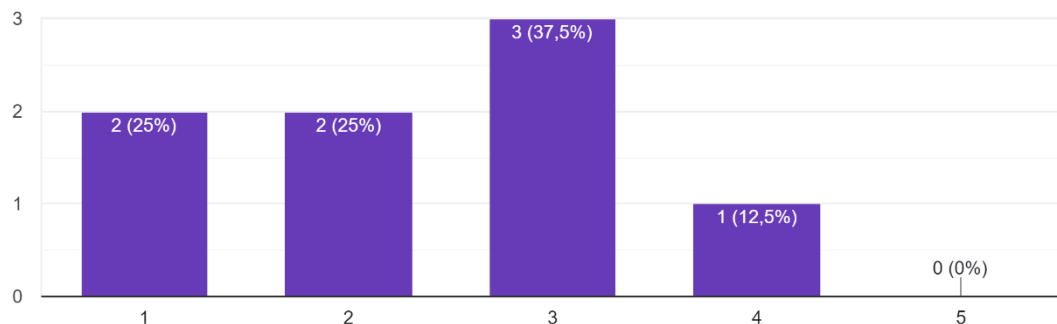Ik zou de app zelf nog gebruiken om mijn uitspraak te verbeteren.
8 antwoorden

Figure 9: Questionnaire statement 7

When asked whether they would use the ELSA Speak app if it were offered in other languages as well (Figure 31), three participants stated they wanted to try, another four remained neutral and only one person disagreed. Since most of the participants are in an Applied Linguistics course at Ghent University and thus study two foreign languages, it is understandable that some of them would want to try and improve their pronunciation for the second foreign language through a language learning app. As a consequence, these results can be seen as something relatively positive.

Ik zou de app ook gebruiken als die in andere talen wordt aangeboden.
8 antwoorden
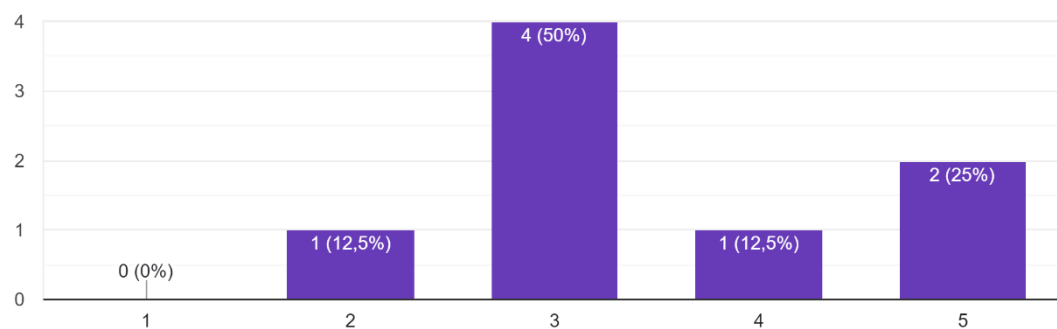
Figure 10: Questionnaire statement 8

As for the question of whether the participants would recommend the app to others (Figure 32), the opinions were, once again, rather divided. Two people stated they would absolutely not recommend the app to others, whereas two other people declared they would. The remaining four participants remained neutral. Once again, this could have multiple reasons and is probably linked to the last statement (see below).

Ik zou de app ook aanraden aan anderen.
8 antwoorden
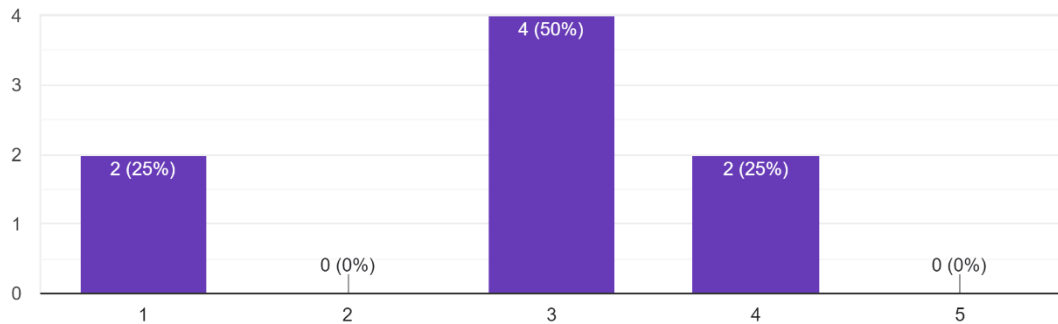
Figure 11: Questionnaire statement 9

Lastly, the participants were asked whether they were disappointed in the fact that they had to pay for the app if they wanted to do more than five lessons a day. Almost all participants, except for one, agreed that they did find it rather disappointing. To link back to statements 7 and 9, the reason why many participants would not use the app again or would not recommend the app to others is probably that the free version is very limited in comparison to the version for which you have to pay.

Ik vond het jammer dat de app betalend was als ik meer dan vijf lessen per dag wilde doen.
8 antwoorden
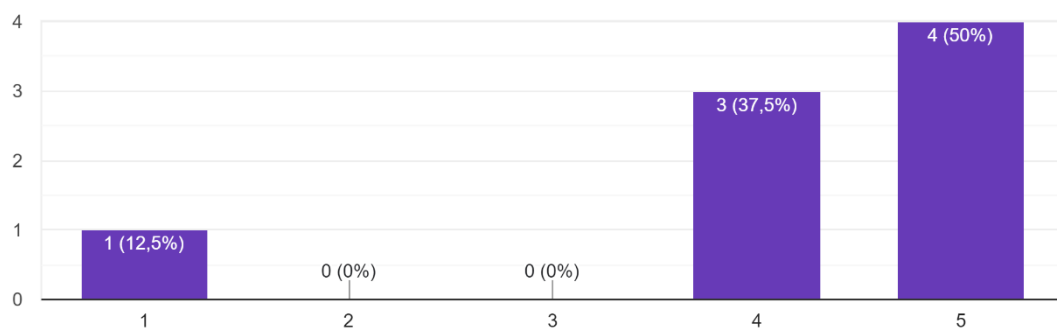
Figure 12: Questionnaire statement 10

To conclude, most participants remained rather positive in terms of the ELSA Speak app, stating that it has an easy-to-use interface, that they understood the feedback provided by the app and that they felt like their pronunciation had improved after working with the app for one or two weeks. Of course, there are a few exceptions to these positive reactions, but all in all most participants had a positive attitude towards the ELSA app. This thus signifies that the fourth hypothesis can be accepted.

C4: Most participants have a positive attitude towards the ELSA app.

Chapter 5: Discussion and conclusion

As is widely known and as was already mentioned earlier on in this thesis, Artificial Intelligence (AI) has become increasingly important in today's day and age. The new technology is being deployed in countless research fields, one of them being Computer-Assisted Language Learning (CALL) and more specifically Automatic Speech Recognition (ASR). The ELSA Speak app, which was examined in close detail in this study, also makes use of this new technology. As the developers mention on the ELSA Speak website, "ELSA's artificial intelligence technology was developed using voice data of people speaking English with various accents. This allows ELSA to recognize the speech patterns of non-native speakers, setting it apart from most other voice recognition technologies" (Elsaspeak, 2023).

ELSA's developers thus claim that the app is able to improve its users' English pronunciation by means of the feedback its AI-incorporated ASR-system can provide. Several other apps, however, have already claimed and proven that they are capable of doing the exact same thing. The experiment in which the Novolearning app was employed, for instance, showed a clear improvement in pronunciation accuracy from the first to the second and third attempts of reading a word (Bai et al., 2020). The IVI application also showed a clear improvement in its users' pronunciation from pre- to post-test (Liu et al., 2019). Even in the experiment in which some of the participants were allowed to use the Nuance Dragon Dictation app and others only received feedback from a teacher or none at all, the ASR group was clearly able to outperform the other two groups (Liakin et al., 2014). Since ELSA has not been around for an extended period of time and since the number of studies on this topic is quite limited, this research study wanted to test whether ELSA was capable of improving its users' pronunciation accuracy, as it so claims on its website.

For this experiment, sixteen Dutch students in the sixth year of secondary school and in the first year of university were divided into two equal groups. The experimental group (8) was asked to partake in a pre-test, a practising phase of one to two weeks in which they had to do five lessons a day with the app and a post-test. The control group (8) on the other hand only participated in the pre- and post-test. To discover whether the app was indeed capable of improving a person's English pronunciation, four research questions were examined.

The first two research questions wanted to determine whether the experimental group had improved significantly from pre- to post-test and whether they had improved significantly in comparison to the control group. For this purpose, both the pre- and post-tests were evaluated by a native speaker of English who gave each participant a score out of twenty. The one-sided paired samples t-test indicated a significant result for the experimental group (p=0.024), which seemed very promising at first, but the one-sided Wilcoxon signed-rank test for the control group also indicated a significant result (p=0.023). Additionally, in comparison to the control group, the one-sided unpaired t-test marked a non-significant result (p=0.431) for the experimental group.

The third research question went into a little more detail by examining whether the participants were able to improve some of the typical mistakes Dutch students tend to make against the English language, based on the research carried out by Kruitbosch (2020). The number of 'most significant' mistakes in both the pre- and post-test for the experimental group were counted and compared to each other. However, the Chi-square test that was performed in the process also indicated a non-significant result (p=0.733), which signifies that the experimental group did not rectify any of the typical mistakes Dutch students tend to make against the English language.

All of these negative results could have multiple explanations; some of them could be related to the design of the experiment itself, for example. First of all, the time frame in which the experimental group practised with the app may have been far too short. The experimental group was asked to use the app for ten minutes every day for a time period of seven to fourteen days. However, other experiments, in which there was a significant difference from pre- to post-test, included practising sessions of twenty minutes four times a week for a month (Liu et al., 2019). Another reason could be associated with the fact that there was only one assessor who graded the participants of this research. It would be better and less subjective if more than one expert were to be involved in the assessment of the participants' pronunciation. Speaking of the participants, this experiment was also rather small-scale with a mere sixteen students taking part. To obtain a more significant result, a larger group of participants with less variation (only focusing on secondary school students or only focusing on university students, for example) would be preferred. Lastly, it may also be interesting to take a look at some other features. This experiment mainly focused on the 'most significant' errors Dutch speakers make, based on Kruitbosch (2020), but perhaps the ELSA app pays attention to other potential errors as well. For example, the Dutch also tend to struggle with the aspiration of /t/, /p/ and /k/ at the beginning of British words (Cucchiarini, 2011; Hermans

& Sloep, 2015; Kruitbosch, 2020; Van Hattum, 2014) and they tend to mispronounce the /l/-sound, making it appear too pharyngealized (Kruitbosch, 2020). Perhaps these are mistakes that ELSA does take into account when offering feedback to the user through its AI-incorporated ASR-system.

Other reasons why the experiment did not turn out as expected may be related to the app itself as well. It has been proven, for example, that the app is not yet one hundred per cent failproof. Becker and Edalatishams (2019, p 3-4) have identified several shortcomings, such as "the lack of balance between segmental and suprasegmental practice opportunities", the focus on quantity over quality and the failure to take into account the user's background (native language, demographic information etc.). All of the previous reasons could, therefore, explain the results that were obtained in this experiment.

Nevertheless, in terms of the fourth research question about the attitudes and perceptions of the experimental group towards ELSA, it can be concluded that most participants had a positive attitude towards the ELSA app. Many of them stated that the app had an easy-to-use interface and provided understandable feedback. Half of the participants in the experimental group even declared feeling as if their pronunciation had improved after practising with the app. This could indicate that, even though the ELSA Speak app may not be suited to be used in class, it could still become an important tool in individual pronunciation practising sessions at home.

However, before teachers advise their students to start using the app individually at home, more research should be conducted to ensure that the app can indeed effectively improve a person's pronunciation. To achieve this, it is suggested for future research to organise longer practising sessions with a larger group of participants who show less variation. For instance, it would be beneficial to compare two classes of twenty secondary school students each, with one class working with the app for ten minutes every day for a month and one class that does not use the app. It is assumed that this type of experiment would yield a significant result.

Bibliography

Anggraini, A. (2022). Improving students' pronunciation skill using ELSA Speak application. In *Journey: Journal of English Language and Pedagogy*, *5*(1), 135-141. https://doi.org/10.33503/journey.v5i1.1840

Anguera, X., & Van, V. (2016). English Language Speech Assistant. In *Interspeech 2016: Show&Tell Contribution*, 1962-1963.

Bai, Y., Hubers, F., Cucchiarini, C., & Strik, H. (2020). ASR-Based Evaluation and Feedback for Individualized Reading Practice. In *Proc. Interspeech 2020*, 3870–3874. https://doi.org/10.21437/Interspeech.2020-2842

BBC. (n.d.). Improve your pronunciation and fluency. Retrieved 12 May 2023, from https://www.bbc.co.uk/worldservice/learningenglish/radio/specials/1535_questionanswer/page65.shtml

Becker, K., & Edalatishams, I. (2019). ELSA Speak—Accent Reduction (Review). In *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*.

Chazen, D. (2019, December 26). How Does ASR Work? The New Generation of Transcription. In *Verbit*. Retrieved 19 January 2022, from https://verbit.ai/asr-and-the-next-generation-of-transcription/

Collins, B. S., & Vandenbergen, A.-M. (1998). *Modern English pronunciation: A practical guide for speakers of Dutch*. Academia Press. Retrieved 27 December 2022, from http://hdl.handle.net/1854/LU-110349

Collins, B. & Mees, I. M. (2003). *The phonetics of English and Dutch, Fifth Revised Edition*. (E.J. Brill). Retrieved 13 February 2023, from https://www.scribd.com/document/522509724/Beverley-Collins-Inger-M-Mees-The-Phonetics-of-English-and-Dutch-Brill-2003#

Council of Europe. (n.d.). Common European Framework of Reference for Languages: Learning,Teaching, Assessment (CEFR). Retrieved 12 May 2023, from https://www.coe.int/en/web/common-european-framework-reference-languages

Cucchiarini, C. (2011). Error selection for ASR-based English pronunciation training in 'My Pronunciation Coach'. In *Proceedings of Interspeech 2011*. 1165-1168.

Dai, Y., & Wu, Z. (2021, July 26). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. In *Computer Assisted Language Learning*, *0*(0), 1–24. https://doi.org/10.1080/09588221.2021.1952272

Davies, A. (2022, December 7). 10 Best AI Apps Of 2023. In *DevTeam.Space*. Retrieved 7 February 2023, from https://www.devteam.space/blog/10-best-ai-apps/

Elsaspeak. (2023). Elsa. Retrieved 25 October 2022, from https://elsaspeak.com/en/

Flemish Ministry of Education and Training (2023). Digisprong. Retrieved 12 May 2023, from https://onderwijs.vlaanderen.be/nl/directies-en-administraties/organisatie-en-beheer/ict/digisprong

Gleason, J. (2012). Beaches and Peaches: Common Pronunciation Errors Among L1 Spanish Speakers of English. In *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*. 205-215.

Hal Schwartz, E. (2020, November 16). Google Demonstrates New Speech Recognition Model for Mobile Use. In *Voicebot.Ai*. Retrieved 10 May 2022, from https://voicebot.ai/2020/11/16/google-demonstrates-new-speech-recognition-model-for-mobile-use/

Hermans, F., & Sloep, P. (2015). Teaching the Dutch how to pronounce English. In *International Journal of Language Studies*, *9*(4), 55–80.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. In *Medical Reference Services Quarterly*, *37*(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Inceoglu, S., Lim, H., & Chen, W.-H. (2020). ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs. In *The Journal of AsiaTEFL*, *17*(3). 824–840. https://doi.org/10.18823/asiatefl.2020.17.3.5.824

Jurafsky, D., & Martin, J. H. (2021, December 29). Automatic Speech Recognition and Tekst-to-speech. In *Speech and Language Processing*. Retrieved 14 January 2022, from https://web.stanford.edu/~jurafsky/slp3/

Kholis, A. (2021). Elsa Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills. In *Pedagogy : Journal of English Language Teaching*, *9*(1). 1-14. https://doi.org/10.32332/joelt.v9i1.2723

Kinsella, B. (2018, May 14). Google Home Beats Amazon Echo in Two Audio Recognition Performance Tests, But Alexa Delivers Highest Composite Score. In *Voicebot.Ai*. Retrieved 10 May 2022, from https://voicebot.ai/2018/05/14/google-home-beats-amazon-echo-in-two-audio-recognition-performance-tests-but-alexa-delivers-highest-composite-score/

Knill, K., Gales, M., Kyriakopoulos, K., Malinin, A., Ragni, A., Wang, Y., & Caines, A. (2018). Impact of ASR performance on free speaking language assessment. In *Interspeech 2018*. 1641–1645.

Kruitbosch, V. (2020). Pronunciation errors made by Dutch secondary school students in English. *Master's thesis, Radboud University*.

Lev-Ari, S., & Boaz, K. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. In *Journal of Experimental Social Psychology*. 46. 1093-1096. https://doi.org/10.1016/j.jesp.2010.05.025

Liu, X., Xu, M., Li, M., Han, M., Zejia, C., Mo, Y., Chen, X., & Liu, M. (2019). Improving English pronunciation via automatic speech recognition technology. In *International Journal of Innovation and Learning*, *25*(2), 126–140. https://doi.org/10.1504/IJIL.2019.097674

McTear, M., Callejas, Z., & Griol, D. (2016). *The Conversational Interface*. (Springer International Publishing). Retrieved 11 January 2022, from through https://doi.org/10.1007/978-3-319-32967-3

Nushi, M., & Sadeghi, M. (2021). A Critical Review of ELSA: A Pronunciation App. In *Computer Assisted Language Learning Electronic Journal*. 22(3). 287-302

Onze Taal. (n.d.). Taalniveaus volgens het ERK. Retrieved 12 May 2023, from https://onzetaal.nl/taalloket/taalniveaus-vreemde-talen#:~:text=Taalniveaus%20volgens%20het%20ERK%20(Europees,vreemde%20taal%20aan%20te%20geven.

Pennington, M. C., & Rogerson-Revell, P. (2018). Beyond the Language Classroom: Wider Applications of Pronunciation Research and Practice. In *English Pronunciation Teaching and Research: Contemporary Perspectives*. 343–399. https://doi.org/10.1057/978-1-137-47677-7_7

Pinontoan, K. A., Pelenkahu, N., & Muntuuntu, M. C. (2022). Using English Language Speech Assistant (ELSA) Speak application to improve students' of SMP Katolik St. Johanis Laikit Pronunciation Ability. In *Journal of Teaching English, Linguistics, and Literature*. 1(11). 1227-1237.

Pourhosein Gilakjani, A. (2011). A Study on the Situation of Pronunciation Instruction in ESL/EFL Classrooms. In *Journal of Studies in Education*. 1(1). 1-15. https://doi.org/10.5296/jse.v1i1.924

Pourhosein Gilakjani, A., & Sabouri, N. (2016). Why Is English Pronunciation Ignored by EFL Teachers in Their Classes? In *International Journal of English Linguistics*. 6(6). 195-208. https://doi.org/10.5539/ijel.v6n6p195

Sciforce. (2021, July 7). Automatic Speech Recognition (ASR) Systems Compared. In *Sciforce*. Retrieved 7 April 2023, from https://medium.com/sciforce/automatic-speech-recognition-asr-systems-compared-6ad5e54fd65f

Segmental. (n.d.). In *Cambridge Dictionary*. Retrieved 15 December 2022, from https://dictionary.cambridge.org/dictionary/english/segmental

Silaen, T. N., & Rangkuti, R. (2022). ELSA Speak app usage in blended learning during the Covid-19 pandemic: students' perspectives. In *Journal of Basic Education Studies*. *5*(1). 26-34.

Statistisch Handboek Studiedata. (2021). Wilcoxon signed rank toets. Retrieved 13 May 2023, from https://sh-studiedata.nl/R/07-Wilcoxon-signed-rank-toets-I-R.html

Suprasegmental (n.d.) In *Encyclopedia Britannica*. Retrieved 15 December 2022, from https://www.britannica.com/topic/suprasegmental

van den Doel, W. Z. (2006). *H*ow Friendly are the Natives? An Evaluation of Native-speaker Judgements of Foreign-accented British and American English. *Doctoral thesis, Utrecht University.* https://www.lotpublications.nl/Documents/144_fulltext.pdf

van der Zee, F. (2016). Welke statistische toets moet je gebruiken? *Hulp bij onderzoek, Groningen,* 1-16.

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. In *Computer Assisted Language Learning*, *29*(4), 833–851. https://doi.org/10.1080/09588221.2016.1167090

Van Hattum, J. (2014). English pronunciation teaching at Dutch secondary schools: Taught or caught. In *Pronunciation Matters; Accents of English in the Netherlands and Elsewhere*. 69-81.

Wang, Y.-H., & Young, S. S.-C. (2014). A Study of the Design and Implementation of the ASR-based iCASL System with Corrective Feedback to Facilitate English Learning. In *Journal of Educational Technology & Society*, *17*(2), 219–233.

Wang, Y.-H., & Young, S. s.-C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. In *Journal of Computer Assisted Learning*, *31*(6), 493–504. https://doi.org/10.1111/jcal.12079

Yuan, Y., & Liu, X. (2020). An Empirical Study of the Effect of ASR-Supported English Reading Aloud Practices on Pronunciation Accuracy. In *Technology in Education. Innovations for Online Teaching and Learning* (1302). 75–87. https://doi.org/10.1007/978-981-33-4594-2_7

Zajechowski, M. (2014, December 29). Automatic Speech Recognition (ASR) Software—An Introduction. In *Usability Geek*. Retrieved 4 May 2022, from https://usabilitygeek.com/automatic-speech-recognition-asr-software-an-introduction/

Appendices

Appendix A: Pre-test

Beluister onderstaande zinnetjes die eerder al werden ingesproken en herhaal ze. Neem jezelf ook op terwijl je de zinnetjes luidop voorleest.

Indien je dit op je eigen gsm opneemt, stuur de opname dan door naar het volgende e-mailadres: femke.de.vrieze@hotmail.be.

Ter info: deze opnames zullen enkel gebruikt worden voor dit experiment. Nadien zullen ze vernietigd worden.

English version:

Listen to the following sentences that have been recorded in advance and repeat them. Record yourself while reading these sentences out loud.

Should you record this on your own phone, send the recording to the following email address: femke.de.vrieze@hotmail.be

For your information: These recordings will only be used for this experiment. Afterwards, they will be destroyed.

- Did you know that I had to return my new favourite black sweater to the store?

- I think one of the previous customers must have tried it on, tearing apart the seams without noticing.

- I asked the girl behind the counter for a refund, but she only spoke Dutch and French.

- I felt like a complete fool having to explain myself through gestures.

- Luckily, another woman walked by asking me if I needed any help with my purchase.

- When I explained the problem and showed her the receipt, she insisted on finding a solution.

- I was delighted when she told me I could exchange the sweater for anything of the same value.

- That is when a pair of boots of the exact same value caught my attention.

- Fortunately for me, they still had size eleven in stock at the store, so I could try them on.

- When I eventually felt satisfied with my new perfect pair of boots, I thanked the woman and left the store.

Appendix B: Logbook

Day	Number of lessons?	Which lessons? Which topic?
Day 1		
Day 2		
Day 3		
Day 4		
Day 5		
Day 6		
Day 7		
Day 8		
Day 9		
Day 10		
Day 11		
Day 12		
Day 13		
Day 14		

Appendix C: Number of difficulties for speakers of Dutch in the pre-test

Most significant errors according to Kruitbosch (2020)	Words containing that specific error
Loss of the fortis-lenis contrast, in particular the confusion between /f/ and /v/ in initial and medial position and the confusion between /t/ and /d/.	favourite, refund, French, felt, fool, finding, fortunately, perfect, felt, satisfied, favourite, previous, having, value, value, eleven, eventually, return, favourite, sweater, customers, it, tearing, apart, without, noticing, asked, counter, walked, receipt, insisted, insisted, delighted, delighted, sweater, attention, satisfied, satisfied, thanked
The confusion between /æ/ and /e/.	that, had, black, have, having, and, value, that, exact, value, had, at, satisfied, thanked, and, sweater, French, felt, myself, gestures, any, help, when, when, sweater, anything, when, attention, them, when, eventually, left, felt, eleven
The loss of contrast, or confusion, between /ʊ/ and /u:/.	you, to, new, to, fool, to, through, solution, value, boots, value, boots, woman, could, could, new, woman
The substitution of /θ/ with /t/.	Think, anything, thanked, through
The epenthesis of /ə/ between l and a following non-alveolar consonant.	Felt, help, felt, myself
The use of uvular /r/.	return, favourite, previous, tried, tearing, for a, refund, French, problem, receipt, pair of, try, pair of, for anything, through
The incorrect realisation of /ð/ in general (Collins et al., 2003), or the substitution of /ð/ with /d/ (Van den Doel, 2006).	that, the, the, the, without, the, the, another, with, the, the, the, the, that, the, they, the, them, the, with, the
Assimilation errors.	Did you
The incorrect realisation of /æ/ and /ʊ/.	that, had, black, have, and, having, and, value, that, exact, value, had, at, thanked, and, satisfied, woman, could, could, woman
The replacement of word-final /dʒ/ and /tʃ/ by /ts/.	French, Dutch, exchange

The confusion of initial /g/ with /k/.	Girl
The replacement of /θ/ with /s/, or /f/.	Think, anything, thanked, through
The confusion of the contrast between /æ/, /e/, and /eə/.	that, had, black, have, and, having, and, value, that, exact, value, had, at, that, satisfied, thanked, and, sweater, French, felt, myself, gestures, any, help, when, when, sweater, anything, when, attention, them, when, eventually, left, felt, eleven, tearing, pair, pair
The confusion between /ʃ/ and /s/.	She, showed, she, solution, she, attention
The production of /aɪ/ that is too long before fortis.	Like, myself, delighted

Appendix D: Post-test

Beluister onderstaande zinnetjes die eerder al werden ingesproken en herhaal ze. Neem jezelf ook op terwijl je de zinnetjes luidop voorleest.

Indien je dit op je eigen gsm opneemt, stuur de opname dan door naar het volgende e-mailadres: femke.de.vrieze@hotmail.be.

Ter info: deze opnames zullen enkel gebruikt worden voor dit experiment. Nadien zullen ze vernietigd worden.

English version:

Listen to the following sentences that have been recorded in advance and repeat them. Record yourself while reading these sentences out loud.

Should you record this on your own phone, send the recording to the following email address: femke.de.vrieze@hotmail.be

For your information: These recordings will only be used for this experiment. Afterwards, they will be destroyed.

- Did you take a look at the fridge door yet like I asked you to?

- Ever since last night, I have not been able to close it all the way and I think I know why.

- I saw one of your guests at the party last night slamming the door shut a little too enthusiastically.

- I actually already tried talking to her, but she was only able to speak Dutch and French.

- I felt like a complete fool trying to make her understand what the problem was.

- Eventually, I just had to give up, because she clearly could not understand a word I said.

- Still, I would like for you to take another look at the fridge door as soon as possible.

- I would be delighted if I could cook my favourite meal tomorrow without the ingredients going bad.

- Should you not be able to repair it, I'll call the maintenance company for help tomorrow.

- They will probably know how to fix the issue for a small price.

Appendix E: Number of difficulties for speakers of Dutch in the post-test

Most significant errors according to Kruitbosch (2020)	Words containing that specific error
Loss of the fortis-lenis contrast, in particular the confusion between /f/ and /v/ in initial and medial position and the confusion between /t/ and /d/.	fridge, French, felt, fool, fridge, favourite, fix, ever, eventually, favourite, asked, last, night, not, it, at, party, last, night, shut, little, but, what, at, delighted, favourite, without, it, maintenance, already, understand, had, understand, said, delighted, ingredients, bad
The confusion between /æ/ and /e/.	at, have, and, at, slamming, enthusiastically, actually, understand, had, understand, at, as, as, bad, yet, ever, guests, already, French, felt, eventually, said, help
The loss of contrast, or confusion, between /ʊ/ and /uː/.	You, look, could, would, look, would, could, cook, should, you, you, to, to, to, enthusiastically, to, to, fool, to, to, you, to, soon, you, to, to, issue
The substitution of /θ/ with /t/.	Think, enthusiastically
The epenthesis of /ə/ between l and a following non-alveolar consonant.	Felt, help
The use of uvular /r/.	fridge, already, tried, French, trying, her understand, problem, fridge, favourite, tomorrow, ingredients, repair, tomorrow, probably, for a, price
The incorrect realisation of /ð/ in general (Collins et al., 2003), or the substitution of /ð/ with /d/ (Van den Doel, 2006).	the, the, the, the, the, another, the, without, the, the, they, the
Assimilation errors.	Did you, should you
The incorrect realisation of /æ/ and /ʊ/.	at, have, and, at, slamming, enthusiastically, actually, understand, had, understand, at, as, as, bad, look, could, would, look, would, could, cook, should,
The replacement of word-final /dʒ/ and /tʃ/ by /ts/.	fridge, fridge, Dutch, French
The confusion of initial /g/ with /k/.	guests, give, going
The replacement of /θ/ with /s/, or /f/.	Think, enthusiastically

The confusion of the contrast between /æ/, /e/, and /eə/.	at, have, and, at, slamming, enthusiastically, actually, understand, had, understand, at, as, as, bad, yet, ever, guests, already, French, felt, eventually, said, repair, help
The confusion between /ʃ/ and /s/.	Shut, she, she, should, issue
The production of /aɪ/ that is too long before fortis.	like, night, night, like, like, delighted, price

Appendix F: Questionnaire (Dutch/English)

Dutch version

Follow-up ELSA app

Als ik je gevraagd hebt om de ELSA app te gebruiken tijdens mijn masterproefonderzoek, kan het zijn dat je nu dus op deze enquête bent beland.

In deze enquête vraag ik je kort even om je mening te geven over de ELSA app aan de hand van tien stellingen waarbij je kunt aangeven in hoeverre je ermee akkoord gaat (1 = helemaal niet akkoord, 5 = helemaal akkoord). De enquête zal dus ook niet meer dan 3 minuutjes van je tijd innemen.

Alvast heel erg bedankt om hem in te vullen!

De app is gebruiksvriendelijk en heeft een makkelijke interface. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik vond het goed dat er gepersonaliseerde lessen waren. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik begreep de feedback die het systeem mij gaf. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik vond de feedback van het systeem niet altijd even correct. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik heb het gevoel dat mijn uitspraak verbeterd is tijdens het oefenen met de app. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik zou de app zelf nog gebruiken om mijn uitspraak te verbeteren. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik zou de app ook gebruiken als die in andere talen wordt aangeboden. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik zou de app ook aanraden aan anderen. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

Ik vond het jammer dat de app betalend was als ik meer dan vijf lessen per dag wilde doen. *

	1	2	3	4	5	
Helemaal niet akkoord	○	○	○	○	○	Helemaal akkoord

English version

Follow-up ELSA app

If I asked you to use the ELSA app during the research of my master's thesis, you may now have been directed towards this survey.

In this survey, I will briefly ask you to give your opinion about the ELSA app by indicating to what extent you agree with each statement (1 = completely disagree, 5 = completely agree). This survey will probably only take 3 minutes of your time at most.

Thank you very much in advance for filling it out!

femke.de.vrieze.fdv@gmail.com (niet gedeeld) Ander account

*Vereist

The app is user-friendly and has an easy interface. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I liked the fact that the lessons were personalised. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I would have preferred more pronunciation lessons instead of listening and viewing exercises. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I understood the feedback the system gave me. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I think the feedback the system gave me was not always correct. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I feel like my pronunciation improved while practising with the app. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I would use the app again to improve my pronunciation. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I would use the app if it were offered in other languages. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I would recommend the app to others. *

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree

I was disappointed by the fact that I had to pay for the app if I wanted to do more *
than five lessons a day.

	1	2	3	4	5	
Completely disagree	○	○	○	○	○	Completely agree