International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 4, Jul.-Aug., 2024, pp: 109-121 www.ijhssm.org

Developing a Rule-Based Text-to-Speech System for Turkish

Özgün Koşaner[1], Murat Özgen[2], Çağdaş Can Birant[3]
[1,2]*Professor, Dokuz Eylül University, İzmir, Turkey*
[3]*Engineer, Gmbh, Dusseldorf, Germany.*
Corresponding Author: Özgün Koşaner

Date of Submission: 26-06-2024 Date of Acceptance: 04-07-2024

ABSTRACT: The sudden increase of the workload and transaction speed in daily life forced humans to develop new technologies to adapt these situations. One of these technologies is the complex software that transform text into speech. This study aims at developing a phoneme/phone-based text-to-speech software, which does not need large databases, and which can run on different platforms. To achieve this, the phonological structure of Turkish was investigated and then a phoneme/phone-based text-to-speech software is developed, differently from most of the previous studies. At the end of the study, a software that runs with a small database is developed. Since the rules, symbol and recordings used in the algorithm are kept separately from the code of the software, it allows to change these data without altering the code. Thus, it would not be necessary to write the code anew for future improvements. The powerful morphology of Turkish, and the complex morphology-phonology interaction prevented processing of prosodic features such as stress and intonation. To perform such a processing, a large and annotated database is needed to conduct morphological and syntactic parsing. Considering these limitations of our study we believe that the software we developed runs smoothly, however it requires improvement regarding prosody.

KEYWORDS: Turkish, Text to Speech, Phonology.

I. INTRODUCTION

The rapid increase in the workload and speed in daily life causes people to create new technologies to keep up with this workload and speed. One of these technologies is software infrastructures that convert text to sound or sound to text. Computer reading of written texts or speech perception and conversion of speech into texts are widely used in English. In our country, such applications have recently found commercial use (voice response systems of banks). The software that this study aims to develop, applications for text-to-speech conversion or voice synthesis, are applications that can be used in certain operating systems and require large databases.

The aim of this study is to develop a voice-based text-to-speech converter software that does not require large databases and can run on different operating systems and different hardware platforms. For this purpose;

i. firstly, a corpus of 10 million words will be created,
ii. then this whole will be divided into equal parts of 1 million words,
iii. these texts will be converted into phonetic alphabet symbols through the GrafoFon software developed in the project 2012.KB.SOS.003 "Development of a grapheme-phoneme converter software for Turkish"
iv. using these symbols, triphone frequencies in Turkish will be found,
v. the most frequently used triplet-sounds will be identified using statistical methods,
vi. a text containing the identified triplet-sound sequences will be created,
vii. this text will be read by different native speakers of Turkish and recorded in a studio environment (in a quiet and anechoic environment) in the Phonology Laboratory of the Department of Linguistics at Dokuz Eylul University,
viii. the audio recordings will be broken down into their sounds using PRAAT and AUDACITY software and each sound will be matched to its equivalent in the phonetic alphabet,
ix. A sound database will be obtained for Turkish,
x. The database obtained will be used in the software to be developed and the conversion of text to sound will be provided.

During the implementation of the software, GrafoFon software developed in the project 2012.KB.SOS.003 "Development of a Phoneme-to-Phoneme Converter Software for Turkish" and the infrastructure of Dokuz Eylül University

Department of Linguistics Phonology Laboratory will be used.

II. LITERATURE REVIEW

Grapheme to phoneme conversion software has been in use for a very long time. In their early days, this software produced a rather mechanical sound, but as hardware and software technologies have improved, they have come closer to natural speech. However, the software that is freely available to the end user for Turkish is either unnatural or requires internet connection or high hardware because they require large databases. Despite the fact that they contain large databases, the text to speech software of giant organisations such as Google still does not provide satisfactory results for Turkish. For example, when the vocalisation feature of the Google translation engine is used for the words “ı**l**ık” (lukewarm) or oku**l**u (“his school”), it is heard that there is a problem with the /l/ sounds in these words. The fact that Turkish is a relatively new language in this field leads to these problems. However, there are still a considerable number of studies in the literature. In this section, these studies will be both described and criticised with reference to the following sections of our study.

Some of the studies on text-to-speech conversion software are rule-based studies. In other words, the pronunciation rules are entered into the software by the researchers. Another group of studies uses statistical or learning machine approach. These studies train the software on a prepared data set and enable them to determine the rules themselves. Since it is not within the scope of our study to discuss the positive or negative aspects of these two methods, these issues are not addressed.

When the literature is analysed, it is seen that many studies have recently been conducted on this subject. In his master's thesis, Eker claimed that he developed a text-to-speech converter software based on the misconception that Turkish is a language that is pronounced as it is written [1]. It is an agreed upon fact in the linguistics literature that Turkish is not a language that is phonetically transparent. The fact that the letter <ğ> in the Turkish alphabet does not have any sound equivalent, in some cases it lengthens the vowel before it, and in some cases, it turns into the sound [j] (the pronunciation of the letter y) refutes this misconception. In his thesis, Eker uses the diphone ending method [1]. In this method, words are divided into sound, more precisely letter pairs. If these diphones are found in the database, he matches them with the sounds and vocalises all the diphthongs in the word by ending them. If a binary sound is not found in the database, the system gives an error [1].

Vural and Oflazer developed a morphology-based text-to-speech converter [2]. Unlike other studies, Vural and Oflazer took the sounds, not the letters of the Turkish alphabet, as a basis and created an algorithm that can process morphological phenomena such as accent and intonation by assuming that there are 8 vowel and 26 consonant sounds in Turkish. They also created a more comprehensive database by using 54 sounds instead of 29 letters. The authors claim that they produced more natural speech by synthesising the outputs of their algorithms in a system called Festival [2]

Sak, Güngör, and Safkan, like the previous study, took a sound-based approach and argued that there are more sounds than 29 letters in Turkish and identified 8 vowels and 24 consonants [3]. They created a lexicon of 3500 root words and their pronunciations and reached a sound synthesis by matching the pronunciations in this lexicon with the texts. They also made natural language speakers listen to the produced sounds and measured how much natural synthesis the system performed [3].

Orhan and Görmez developed a phoneme-based and contiguous text-to-speech conversion system. Orhan and Görmez started from the assumption that there are 21 consonants and 8 vowels in Turkish and suggested that there are seven different syllable patterns in Turkish [4]. As we will discuss in the following sections, apart from the letters in the Turkish alphabet, there are three other graphemes formed by using the correction sign (^). Therefore, the calculations made in this study are incomplete. In addition, the number of sounds in Turkish is not limited to the letters in the alphabet. While some letters have no sound equivalent (e.g. ğ), some letters may have more than one sound equivalent (e.g. kal - kel). The system developed by Orhan and Görmez uses the Zemberek project for phoneme segmentation and divides phonemes consisting of more than two units [4]. This situation causes these features of the sounds, which exhibit different interactions when they are in the same syllable and different interactions when they are in different syllables, to be overlooked.

Another recent study in this field is the paper presented by Sel, Hanbay and Karabatak. In this study, a phoneme-based algorithm was also used [5]. The authors also based on the assumption that there are seven different vowel patterns in Turkish. Unlike Orhan and Görmez [4], the authors used their own phoneme segmentation algorithm and obtained a phoneme database by segmenting the corpora into phonemes [5]. However, they did not provide detailed information on how the speech synthesis was performed.

III. DATABASE PREPARATION

Since the aim of this study is to develop an audio-based text-to-speech converter software that does not require large databases and can run on different operating systems and different hardware platforms, in the planning phase of the study, it was thought to first create a corpus of 10 million words, then divide this corpus into equal parts of 1 million words, convert these equal parts into phonetic alphabet symbols and find the triphone frequencies in Turkish using these symbols. During the execution of the study, a preliminary study was conducted to see how the method would work. In this preliminary study, 100 thousand words of text obtained from the Turkish National Corpus V.1 and TSCorpus corpora and the databases created by the researchers themselves were divided into ten thousand words. In the statistical study conducted on the first ten-thousand-word text segment, it was observed that the triplet frequencies were far from reflecting the features of spoken Turkish, and this method was abandoned and a rule-based database was created for a software that was already planned to be rule-based. While creating this database, care was taken to ensure that the rules were not operated according to the graphemes. The linguistic justification for this method change is presented below.

In the Turkish alphabet, there are a total of 29 letters, eight of which are vowels (a, e, ı, i, o, ö, u, ü) and twenty-one of which are consonants (b, c, ç, d, f, g, ğ, h, j, k, l, m, n, p, r, s, ş, t, v, y, z). However, the graphemes of Turkish are not limited to these. There are three other graphemes that we use to vocalise some borrowed words in Turkish. These graphemes are â, î and û, which are formed by adding the correction sign (^) to the three letters in the standard alphabet. Although the frequency of use of the last two is quite low today, â is important because it is a meaning modifier. If we take these three orthographies into consideration, we can say that Turkish has a writing system consisting of 32 graphemes. If we were to conduct our study on a grapheme-based database, we would have to analyse the frequency of 32X32X32 = 32,768 possible triplets. However, even such a high number of triplet patterns would have failed to capture the sound interactions in the rationale we present below. Therefore, instead of this method, it would be appropriate to adopt an approach that focuses entirely on the phonological rules of Turkish.

In this study, where we aim to develop a text-to-speech converter software that does not require large databases in Turkish, our starting point should be the sounds of Turkish. However, many studies in the literature try to capture the interaction between sounds through graphemes, in a sense letters. This leads to overlooking the features of Turkish phonology that are not reflected in writing. To overcome this problem and to capture the interactions between the vowel and consonant phonemes corresponding to the graphemes in the Turkish alphabet, the syllable patterns of Turkish were first extracted. In linguistic sources, it is generally suggested that there are six syllable patterns in Turkish [6]. These syllable patterns are determined by the interaction of consonant and vowel sounds and the rules of Turkish phonology. Accordingly, the phoneme patterns of Turkish in the literature are as follows:

Phoneme Patterns of Turkish

V	**a**.na
VC	**aç**.lık
CV	**ka**.ra
CVC	**kıt**.lık
VCC	**art**.çı
CVCC	**kırk**

However, although it is not very prominent in Turkish, before the vowels at the beginning of a word (in terms of phonology), the vocal cords in the larynx produce a popping sound when they open for the first time after a long interval. In phonology, this popping sound is called a glottal plosive and is denoted by /ʔ/ in the International Phonetic Alphabet. In languages such as English and Arabic, this sound exhibits a distinctive feature, but this is not the case in Turkish. However, considering that the qualities of the vowels at the beginning of the word and the vowels in the word are different, we have determined the vowel patterns presented above in our study by following Demircan as follows [7]:

Phoneme Patterns of Turkish (Revised)

CV	**ʔa**.na, **ka.**ra
CVC	**ʔaç**.lık, **kıt**.lık
CVCC	**ʔart**.çı, **kırk**

As can be seen from the syllable patterns above, all syllables in Turkish begin with a consonant. Therefore, the number of syllable patterns in Turkish is reduced to three. As can be seen here, the glottal plosive [ʔ], which is not shown in the script but has a phonological effect, has the

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 4, Jul.-Aug., 2024, pp: 109-121 www.ijhssm.org

quality to change all vowel patterns and thus the structure of the grapheme-to-phoneme software.

At this stage, after explaining why we are going to use the phonemes and not the graphemes of Turkish, we can present information about the phonemes of Turkish. We have already mentioned that there are 32 graphemes in the Turkish orthography. However, there are many more sounds in the phonetic system of Turkish. There are some differences between the studies in the literature in terms of the number of sounds and the use of signs indicating the sounds. We can present this situation better on a chart as follows.

Table 1. Phonemes and sounds in the phonetic system of Turkish

	Grapheme		Phoneme	Phone	Phoneme	Phone	Phoneme	Phone	Phoneme	Phone‡
	Capital	Lower	Özsoy (2000)		Ergenç (2002)		Göksel&Kerslake (2005)		Kılıç	
1	A	a	ɑ	ɑ a ɑ̃ ã	ɑ	ɑ a	a	a ạ	ɑ	ɑ a ɑː aː
2	B	b	b	b	b	b	b	b	b	b
3	C	c	ǰ	ǰ	ʤ	ʤ	ʤ	ʤ	ʤ	ʤ
4	Ç	ç	č	č	ʧ	ʧ	ʧ	ʧ	ʧ	ʧ
5	D	d	d	d	d	d	d	d	d	ʧ ʧʰ
6	E	e	e	e ɛ ẽ eˇ	e	e ɛ	e	e ɛ æ	ɛ	e ɛ
7	F	f	f	f ɸ	f	f	f	f ɸ	f	f ɸ
8	G	g	g / ɟ	g ɟ	g / ɟ	g ɟ	g / ɟ	g ɟ	g	g ɟ
9	Ğ	ğ	*	*	*	*	ɣ	*	ɰ	ɰ ɣ ː
10	H	h	h	h x ç	h	h	h	h x ç	h	h ɦ
11	I	ı	ɯ	ɯ ɯ̃ ɯˇ	ï	ï	ı	ı	ɯ	ɯ ɯ̆
12	İ	i	i	i ĩ iˇ	i	i ɪ	i	i I	i	i iː ĭ
13	J	j	ž	ž	ʒ	ʒ	ʒ	ʒ	ʒ	ʒ
14	K	k	k / c	k kʰ c cʰ	k / c	k c	k / c	k kʰ c cʰ	k	k kʰ c cʰ
15	L	l	ɫ / l	ɫ l	l	l ɫ	ɫ	ɫ l	l	l ɫ l̥ ɫ̥
16	M	m	m	m ɱ	m	m	m	m	m	m m̥
17	N	n	n	n ɲ ŋ	n	n ŋ	n	n ɲ ŋ	n	n n̥ ɲ ŋ
18	O	o	o	o ɔ̃ ɔ	o	o ɔ	o	o ọ	ɔ	ɔ o
19	Ö	ö	ø	ø œ̃ œ	ø	ø œ	ö	ö	œ	œ ø
20	P	p	p	p pʰ	p	p	p	p pʰ	p	p pʰ
21	R	r	r	r ɾ r̥	r	r ɾ ɣ	r	r ɾ̥	r	r ɾ ɹ ɹ̝ ɹ̥
22	S	s	s	s	s	s	s	s	s	s
23	Ş	ş	š	š	ʃ	ʃ	ʃ	ʃ	ʃ	ʃ
24	T	t	t	t tʰ	t	t	t	t tʰ	t	t̪ t t̪ʰ tʰ
25	U	u	u	u ũ uˇ	u	u ᴜ	u	u ụ ʊ	u	u uː ŭ
26	Ü	ü	y	y ỹ yˇ	y	y ʏ	ü	ü ʏ	y	y y̆
27	V	v	v	v β	v	v ʋ	v	ɷ v β	v	v ʋ β w
28	Y	y	j	j	j	j	j	j	j	j ʝ
29	Z	z	z	z	z	z ʐ	z	z	z	z
30	Â	â	†	†	†	†	†	†	†	†
31	Î	î	†	†	†	†	†	†	†	†
32	Û	û	†	†	†	†	†	†	†	†
33	’	’	†	†	†	†	†	†	ʔ	ʔ

As can be seen in Table 1, different researchers identified a minimum of 42 and a maximum of 69 sounds depending on how detailed their analyses were. In addition, the signs they use for the same sounds or phonemes are also different. Nevertheless, based on this chart, we can say that there are more than 32 phonemes. In our study, we created our own phoneme and phoneme clusters by reducing the detailed parts of these studies and adding the missing phoneme and phone symbols. These phonemes, the phones derived from the interactions between these phonemes, and the rules explaining these interactions are presented in the table 2:

Looking at the chart above, it is possible to say that regressive assimilation, that is, the process where the following sound resembles the previous sound completely or in certain aspects, is more dominant in sound interactions in Turkish. Progressive assimilation occurs mostly in morphophonological phenomena such as affixation (e.g. yok-**d**ur -> yok-**t**ur). In sound level, the sound that comes after usually changes the sound that comes before. For example, the sound [n], which is the last sound of the word you, resembles the sound [e] that comes before it, causing this sound to become nasalised. Similarly, the [f] sound in the word pistachio changes the sound [m] before it in terms of its origin and turns it into an [ɱ] sound.

Another result we obtained when we examined the table is that progressive similarity works together with regressive similarity. In other words, a sound can be affected by both the sounds that come before it and the sounds that follow it. For example, the phoneme [n] in the word "yenge" (aunt) changes by being affected by the [e] and [g] sounds that come before and after it and turns into the sound [ɲ]. A general rule that we can deduct from the table is that vowels in Turkish are lowered at the end of words. For example, the [i] sound in the word "bit" (flea) and the [iˇ] sound in the word "evi" (his house) are pronounced differently.

			Rule
a	A	α	The phoneme /α/ becomes [α̃] if it occurs before [m], [ɱ], [n], [ɲ] or [ŋ] in the same syllable. k**a**n, d**a**m
â	Â	a	If the phoneme /a/ precedes [m], [ɱ], [n], [ɲ] or [ŋ] in the same syllable, it becomes [ã]. müjg**â**n
b	B	b	**b**a**b**a
c	C	ʤ	**c**enk, e**c**e
ç	Ç	ʧ	**ç**ocuk, a**ç**
d	D	d	**d**e**d**e
e	E	e	1. If the phoneme /e/ precedes [l], [r], [r̥] in the same syllable, it turns into [ɛ]. d**e**lgi, y**e**r, 2. If the phoneme /e/ precedes [m], [ɱ] [n], [ɲ] or [ŋ] in the same syllable, it becomes [ɛ̃]. n**e**m, s**e**n 3. When the phoneme /e/ becomes the final sound of a word, it turns into [eˇ]. der**e**
f	F	f	If the phoneme /f/ is used with rounded vowels ([o], [u], [ø], [y]), it turns into [ɸ]. u**f**uk
g	G	ɡ	1. The phoneme /ɡ/ turns into [g] if it is used with [α], [α̃], [ɯ], [ɯˇ], [ɯˇ], [ɯ̃ˇ], [o], [ɔ], [ɔ̃], [u], [uˇ], [ũˇ] in the same syllable. dur**g**un 2. If the phoneme /ɡ/ is used in the same syllable with the sounds [a], [ã], [e], [ɛ], [ɛ̃], [eˇ], [i], [iˇ], [ø], [œ], [œ̃], [y], [yˇ], [ỹˇ], it turns into [ɟ]. A**g**âh, **g**üzel, **g**ez**g**in
ğ	Ğ	· or :	1. If the letter <ğ> is used at the end of a word, it lengthens the preceding vowel, shown as [:]. da**ğ** [dα:] 2. If the letter <ğ> is used after a vowel and before a consonant, it lengthens the preceding vowel. ya**ğ**mur [jα:muˇr̥] 3. If the letter <ğ> is between the same vowels, these two vowels are pronounced as a single long vowel. a**ğ**aç [α:ʧ] 4. The letter <ğ> causes a vowel shift between different vowels. ö**ğ**üt [ø·yt]
h	H	h	1. If the phoneme /h/ is used word-initially before the sounds [α], [α̃], [ɯ], [ɯˇ], [ɯ̃ˇ], [ɯ̃ˇ], [o], [ɔ], [ɔ̃], [u], [uˇ], [ũˇ], it becomes [h]. **h**arç, **h**urç, **h**ırlamak, **h**orlamak 2. If the phoneme /h/ is not word-initial and is used with the sounds [α], [α̃], [ɯ], [ɯˇ], [ɯ̃ˇ], [ɯ̃ˇ], [o], [ɔ], [ɔ̃], [u], [uˇ], [ũˇ] in the same syllable, it turns into [x]. a**h**val, a**h**ır 3. If the phoneme /h/ is not at the beginning of a word and is used in the same syllable with the sounds [a], [ã], [e], [ɛ], [ɛ̃], [eˇ], [i], [iˇ] [ĩˇ], [ø], [œ], [œ̃], [œ̃], [y], [yˇ], [ỹˇ], it becomes [ç].
ı	I	ɯ	1. The phoneme /ɯ/ becomes [ɯˇ] if it precedes [ɫ], [r], [r̥] in the same syllable or if it is the final sound of the word. **ı**lgın, sar**ı** 2. If the phoneme /ɯ/ precedes [m], [ɱ] [n], [ɲ] or [ŋ] in the same syllable, it becomes [ɯ̃ˇ]. ılg**ı**n, sar**ı**m
i	İ	i	1. The phoneme /i/ becomes [iˇ] if it precedes [l], [r], [r̥] in the same syllable or if it is the final sound of the word. **i**lg**i** 2. If the phoneme /i/ precedes [m], [ɱ] [n], [ɲ] or [ŋ] in the same syllable, it becomes [ĩˇ]. d**i**ng**i**n
j	J	ʒ	**J**ale, o**j**e
k	K	k	1. The phoneme /k/ turns into [k] if it is used with the sounds [α], [α̃], [ɯ], [ɯˇ], [ɯ̃ˇ], [ɯ̃ˇ], [o], [ɔ], [ɔ̃], [ɔ̃], [u], [uˇ], [ũˇ] in the same syllable. a**k**ıl 2. The phoneme /k/ turns into [c] if it is used with [a], [ã], [e], [ɛ], [ɛ̃], [eˇ], [i], [iˇ], [ø], [œ], [œ̃], [y], [yˇ], [ỹˇ] in the same syllable. i**k**i 3. The sounds [k] and [c] turn into [k^h] and [c^h] at the beginning of words and syllables. **k**uru, o**k**ul
l	L	ɫ	1. The phoneme /ɫ/ turns into [ɫ] if it is used with [α], [α̃], [ɯ], [ɯˇ], [ɯ̃ˇ], [ɯ̃ˇ], [o], [ɔ], [ɔ̃], [u], [uˇ], [ũˇ] in the same syllable. akı**l** 2. The phoneme /ɫ/ turns into [l] if it is used with [a], [ã], [e], [ɛ], [ɛ̃], [eˇ], [i], [iˇ], [ø], [œ], [œ̃], [y], [yˇ], [ỹˇ] in the same syllable. e**l,** a**l**â
m	M	m	If the phoneme /m/ is used before [f], it becomes [ɱ]. amfora, amfi
n	N	n	1. The phoneme /n/ becomes [ŋ] when used after [α], [α̃], [α̃], [ɯ], [ɯˇ], [ɯ̃ˇ], [ɯ̃ˇ], [o], [ɔ], [ɔ̃], [ɔ̃], [u], [uˇ], [ũˇ] and before [k] or [g]. ya**n**kı, ta**n**k, ta**n**go 2. / The phoneme /n/ becomes [ɲ] when used after [a], [ã], [e], [ɛ], [ɛ̃], [eˇ], [i], [iˇ], [ø], [œ], [œ̃], [y], [yˇ], [ỹˇ] and before [c] or [ɟ]. ye**n**ge, de**n**k
o	O	o	1. The phoneme /o/ turns into [ɔ] if it precedes [ɫ], [r], [r̥] in the same syllable or if it is the final sound of the word. **o**lgu 2. If the phoneme /o/ precedes [m], [ɱ], [n], [ɲ] or [ŋ] in the same syllable, it becomes [ɔ̃]. k**o**mşu
ö	Ö	ø	1. The phoneme /ø/ turns into [œ] if it precedes [l], [r], [r̥] in the same syllable or if it is the final sound of the word. g**ö**l 2. If the phoneme /ø/ precedes [m], [ɱ] [n], [ɲ] or [ŋ] in the same syllable, it becomes [œ̃]. g**ö**nder
p	P	p	The phoneme /p/ turns into [p^h] at word-initial and syllable-initial position. **p**is, a**p**açi
r	R	r	1. The phoneme /r/ becomes [r̥] when it is the final sound of a word. ki**r** 2. When the phoneme /r/ is between two vowels, it becomes [ɾ]. a**r**a
s	S	s	**s**arı, a**s**al, ka**s**
ş	Ş	ʃ	**ş**a**ş**ı, ka**ş**
t	T	t	The phoneme /t/ turns into [t^h] at word-initial and syllable-initial. **t**ik, a**t**ak
u	U	u	1. The phoneme /u/ becomes [uˇ] if it precedes [ɫ], [r], [r̥] in the same syllable or if it is the final sound of the word. d**u**rgun 2. The phoneme /u/ becomes [ũˇ] if it occurs before [m], [ɱ] [n], [ɲ] or [ŋ] in the same syllable. durg**u**n
ü	Ü	y	1. The phoneme /y/ turns into [yˇ] if it precedes [l], [r], [r̥] in the same syllable or if it is the final sound of the word. g**ü**r 2. The phoneme /y/ becomes [ỹˇ] if it precedes [m], [ɱ] [n], [ɲ] or [ŋ] in the same syllable. düzg**ü**n
v	V	v	When the phoneme /v/ is used with rounded vowels ([o], [u], [ø], [y]), it turns into [β]. ko**v**uk
y	Y	j	yazı
z	Z	z	Zaman
		ʔ	The glottal plosive that occurs at the beginning of a word.

In the second one, the mouth is more open and the tongue is in a lower position in the mouth. The lowering rule applies to all vowels except [a] and [α]. Another general rule that we can obtain from the table is the rounding rule. For example, since the /f/ phoneme in the word "üfürük" (puff) is located between two rounded vowels, it becomes rounded and turns into the [ɸ] sound. Considering that a software developed without taking these rules into consideration would not be very successful, the method of our study was changed and an attempt was made to develop a completely rule-based software. In order to separate the sounds that will form the database of the software we will develop in the study, it is necessary to first determine the sounds whose symbols and rules are given in the table above. For this purpose, based on the syllable patterns we mentioned before and the dominance of the regressive assimilation process, the following syllables and words containing these phonemes were determined. Since some syllables do not comply with the sound system of Turkish, examples of these sound sequences could not be found and they are stated as N/A in the list below. In order not to ignore the effect of the glottal plosive sound, firstly, the CV pronunciation pattern consisting of glottal plosive + vowel sound combinations was determined and words containing these were determined. In the determinations, attention was paid to the use of Turkish words, and borrowing and onomatopoeic words were used for sound sequences that are not found in Turkish words. Since there are eight vowel sounds in Turkish, 8 glottal plosive + vowel sequences have been determined. Then, consonant + vowel sequences that again exhibited the CV syllable pattern, but this time at the end of the word, were determined. Based on our determination that progressive assimilation does not operate at the phonetic level in Turkish, only Y + vowel sound sequences were determined for the word-final CV pattern.

For the CVC syllable pattern, sound sequences in the form of glottal plosive + vowel + consonant were determined, taking into account the effect of the glottal plosive sound, and words containing these were found. There was no need to determine the pronunciation patterns in the form of consonant + vowel + consonant. Because, based on our determination that regressive assimilation is more dominant in Turkish, it was thought that the data obtained from the glottal plosive + vowel + consonant sequence would be sufficient to keep the database of our study small. Since the effects of regressive assimilation can be achieved by separating the glottal plosive parts from the glottal plosive + vowel + consonant sequence during the parsing stage after the voice recordings are made, making the same recordings again will cause the database to grow. Instead, by focusing on the special situations such as nasalization, lowering and rounding that we mentioned in our explanations of Table 2, more specific sequences were determined and words containing them were found.

As a result of these decisions, a total of 208 phoneme patterns were determined. However, since 38 of them were in sequences that were not allowed by the Turkish sound system, it was decided to record a total of 170 words. The list of these syllable patterns and words is presented in the annex of this study.

The 170 words were recorded in the phonology laboratory of the Department of Linguistics, Faculty of Letters, Dokuz Eylül University in an isolated and echo-free environment. The recordings were taken from a university graduate whose family has been living in Izmir for two generations. Special care was taken to ensure that this person was not interested in linguistics. Before starting the voice recording, the person whose voice was to be recorded was informed about the study and a consent form was obtained stating that he/she voluntarily participated in the study and that there was no objection to the use of his/her voice. During the recording, each word was recorded in a single file with three-second gaps between each word and the original file was stored in wav format without conversion. After the recording file was listened to, the faulty sections were re-recorded and added to the first recording file in place of the faulty recording.

The recording file was then processed in the Levelator software produced by The Conversations Network to eliminate the level differences that may be caused by the change in the distance of the vocaliser from the microphone and the sound levels were equalised.

Then, this file was processed with Winpitch software in order to equalise some of the dominant vowels due to the stress pattern of Turkish. In this processing, the basic frequency of the recording file was determined and the pitch curve (F0 value) was reduced and equalised so that the sound did not turn into a mechanical sound in order to equalise the frequency changes caused by accent.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 4, Jul.-Aug., 2024, pp: 109-121 www.ijhssm.org

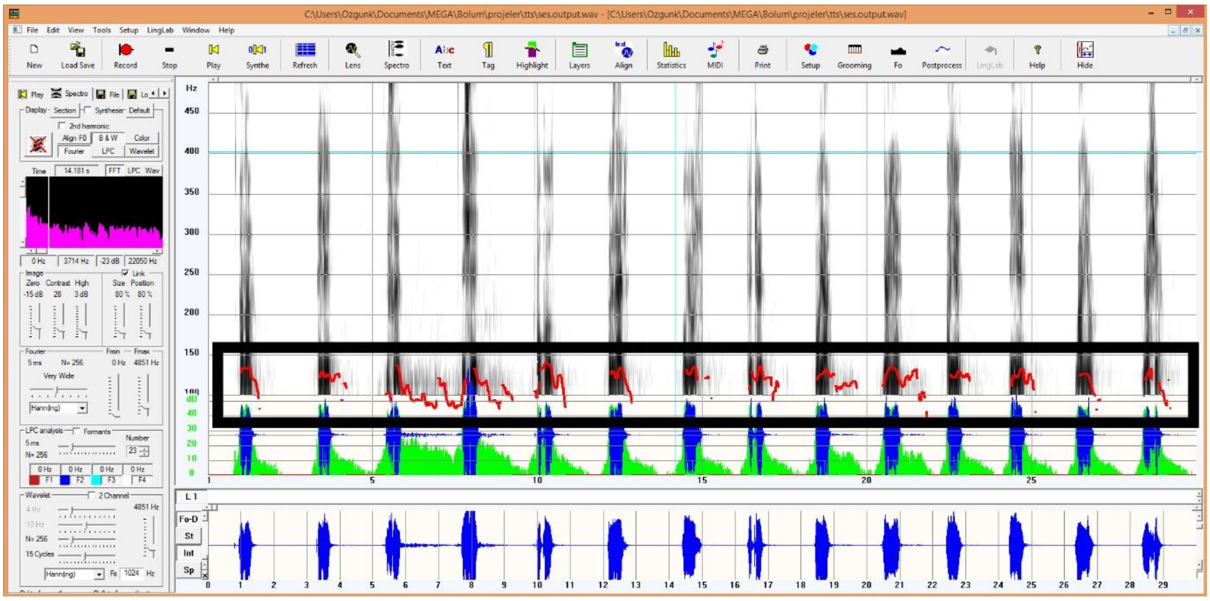

Figure 1. Pitch curves in the original format of the recording file

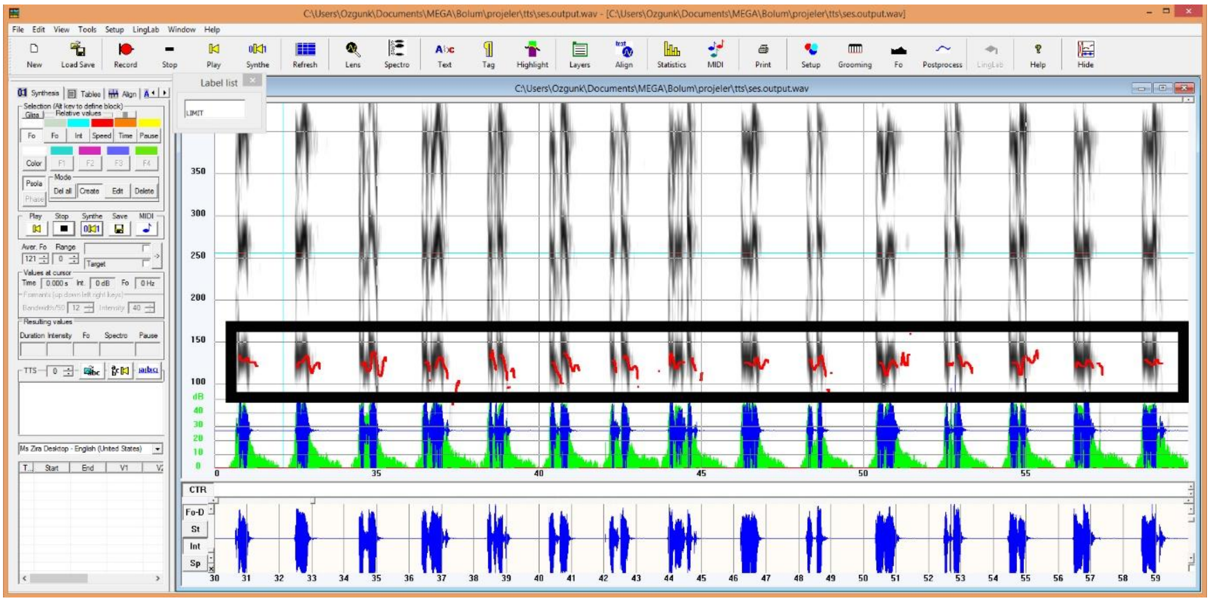

Figure 2. Pitch curves synchronized in the recording file

Then the recording file was opened with Praat software and the boundaries of each sound were determined and these boundaries were labelled. After the boundary determination and labelling process was completed, these labelled parts were extracted into separate files using Praat software. A total of 212 sounds obtained in the parsing process were numbered and the phonological markings of the sounds as well as their phonological environment features, i.e. which sound comes after and before which sound, were coded in a chart. Thus, the database of our study was created.

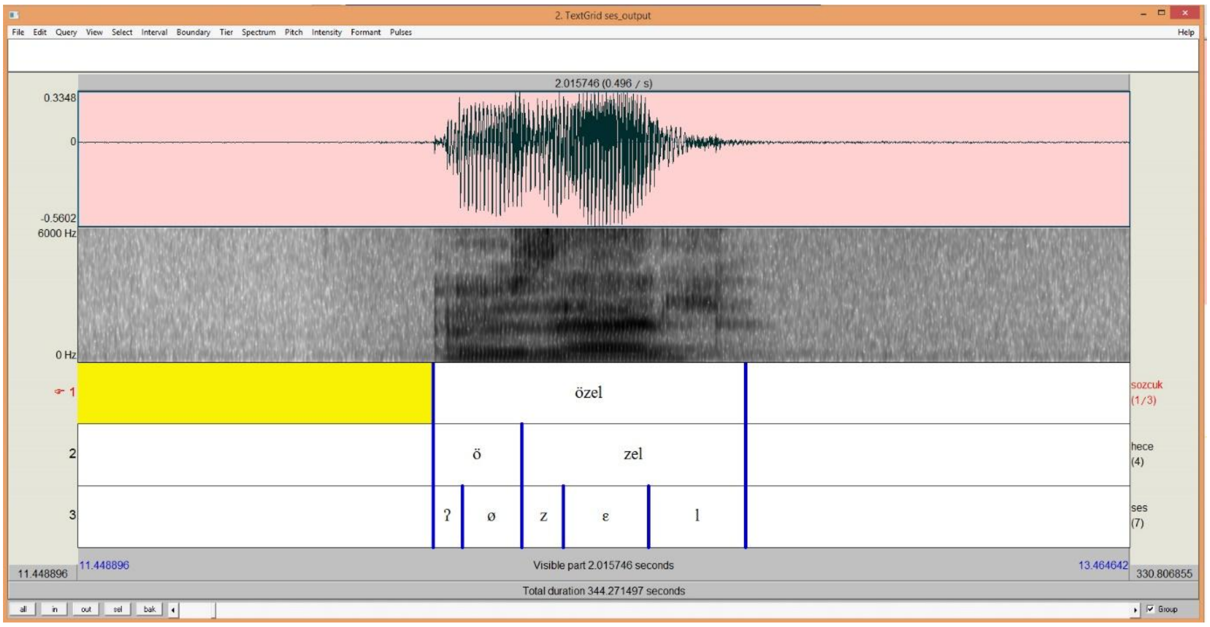

Figure 3. Words, sounds and sounds tagged in the Praat program

IV. SOFTWARE DEVELOPMENT

In the development phase of the software, the end-of-sentence detection algorithm (RBSDM) [8] and the script-phoneme conversion algorithm (Grafofon Project No: 2012.KB.SOS.3) [9] previously developed by our project team were used. The end-of-sentence detection algorithm is an algorithm that takes the text in a plain text (.txt) file, breaks it into its sentences and outputs it in XML format and works with an accuracy rate of 99.78%. Grafophon, on the other hand, is a software that converts the text input into phonetic cues and works with an accuracy of 96.12%. Since the first version of the software did not include the phoneme structure and the glottal burst features discussed in the previous section, the phonological rules and the phoneme/sound database were updated in this project. The flowcharts of these algorithms are as follows:

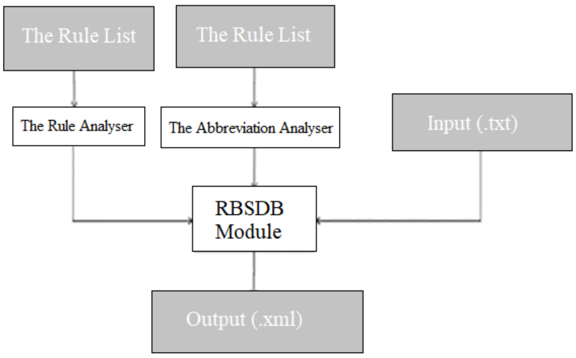

Figure 4. RBSDB - End of Sentence Detection Algorithm

When the output of the sentence separation algorithm is given to the Graphophone algorithm, this algorithm first separates the text into words. These words are then separated into their pronunciations by the speaker. The grapheme-phoneme converter module converts graphemes into phonemes using the grapheme list, phonological rules and the phoneme/sound list. The matcher gives output by matching the original form of the text with the form converted into phonemes.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 4, Jul.-Aug., 2024, pp: 109-121 www.ijhssm.org

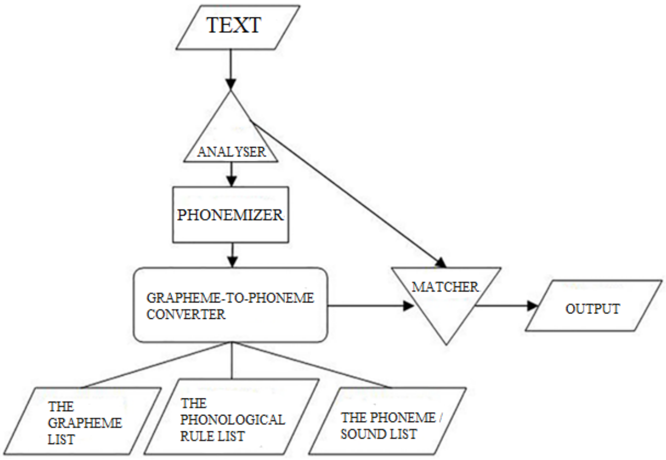

Figure 5. Grafofon – Grapheme-to-Phone Converter Algorithm

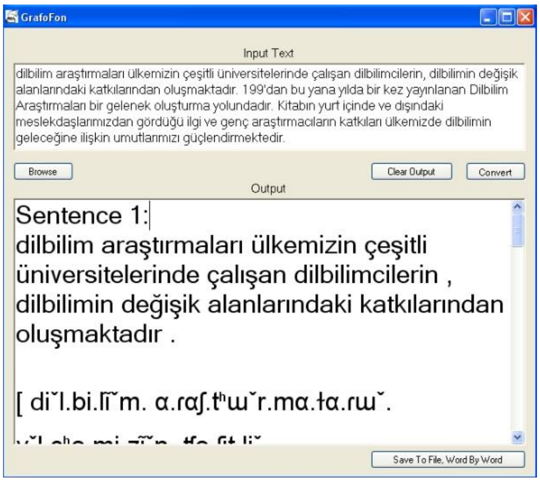

Figure 6. Grafofon – Grapheme-to-Phone Converter Output

As seen in Figure 6, the Graphophone algorithm provides both a detailed phonological description and shows phoneme boundaries (using the "." sign). The software we developed in this study contains a database of sound files that we recorded, similar to the phoneme/sound list of the Graphophone algorithm.

The algorithm we developed in this study, presented in Figure 7 takes a plain text file as input, first separates the text in this file into sentences and sends these separated sentences to the parser module. The parser module separates the sentences into their words. In addition, in this module, the numbers in the text are converted into their written form and the phonemes of these numbers are given as output. The data from the parser goes to the phonetician where the words are separated into their phonemes. Then, the units in this separated phoneme are converted into phonemes or phonemes using the orthographic list, the phonological rules list and the phoneme/phoneme symbol list. These phonemic symbols are then matched with the relevant phonemes using the phoneme-phoneme matching table we created earlier, and these phonemes are then concatenated. The output obtained from these matches is combined in the voice synthesiser to produce a sound file output.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 4, Jul.-Aug., 2024, pp: 109-121 www.ijhssm.org

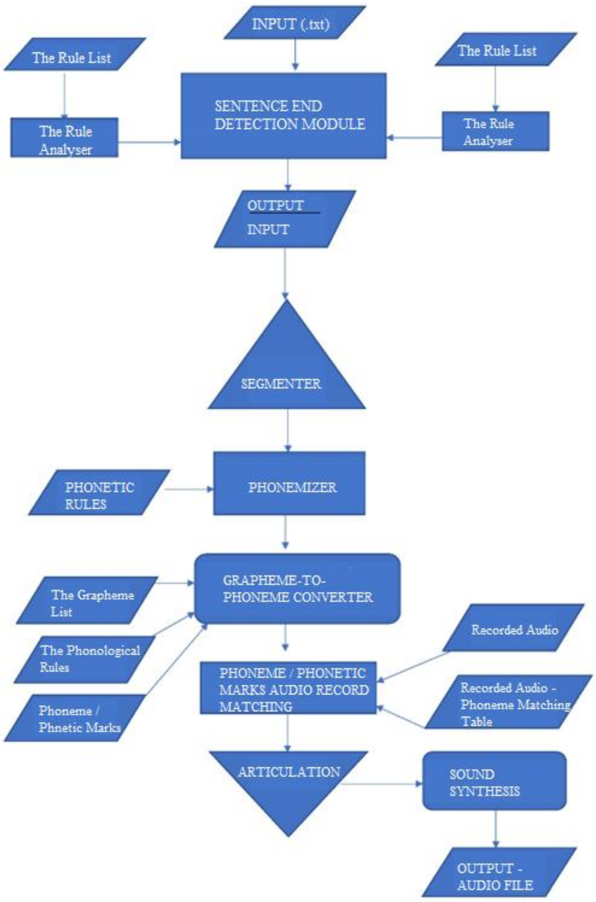

Figure 7. Grapheme-to-Speech Conversion Algorithm

V. SOFTWARE EVALUATION

In order to evaluate the performance of the software, 6 different short texts from different sources (newspapers, novels, academic texts) were prepared as input. These texts were then converted into audio files by running the programme. The audio files were presented to 10 individuals and they were asked to evaluate the correct extraction of phonemes, fluency of speech and traceability of the subject in these audio recordings. The 10 individuals chosen to listen to the audio recordings were selected from the students of the Department of Linguistics at Dokuz Eylül University. The selection of these participants was based on the fact that they had taken and passed the Phonetics and Phonology, and Turkish Phonology courses in the department. As a result of the feedback received from the participants, it was concluded that the text-to-speech converter software was able to produce phonemes and sounds correctly, but the traceability of the subject was low. The reason for this is that the software is not able to present suprasegmental features such as word stress and intonation. In the study, word stress was also tried to be included. However, although it is said that the stress of Turkish is located at the end of the word, the fact that there are more than thirty unstressed affixes in Turkish and that they change the stress pattern of words requires morphological, syntactic and even pragmatic analysis in addition to phonological analysis. For example, the stress patterns of the homophonous word “tokat” (slap) and the word “Tokat” (city in Turkey) are different. Similarly, the stress patterns of the word “gelin” (bride) and the word “gelin” (come) in the imperative mood (2nd plural) are different from each other. Therefore, processing stress and intonation patterns requires a more comprehensive and large-scale software.

VI. CONCLUSION

The aim of this study is to develop an audio-based text-to-speech converter software that does not require large databases and can run on different operating systems and different hardware

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 4, Jul.-Aug., 2024, pp: 109-121 www.ijhssm.org

platforms. In this project, a software that works with a database as small as 30 mb and is easy to implement on every platform has been developed. The fact that the list of rules, markings and audio recordings used in the algorithm are kept in separate files independent of the programme allows these rules to be changed without the need for coding. Thus, there will be no need to write code again for improvements to be made in the software. In addition, the programme can be easily adapted to other languages, especially languages related to Turkish, which have lists of rules, markers and phonemes. The fact that Turkish has a very rich morphological structure and that the interaction between morphology and phonology is very complex makes it difficult to process morphological features such as stress and intonation. Such processing requires morphosyntactic and syntactic analysis and large and labelled databases. Considering these limitations of our study, we think that our software works successfully but needs to be improved in terms of morphological features.

REFERENCES

[1]. Eker, B., (2002). *Turkish Text to Speech System*, Unpublished Master's Thesis, Bilkent University.
[2]. Vural, E. and Oflazer, K., (2004). A Prosodic Turkish text-to-Speech Synthesizer, in *Proceedings of IEEE 12th Signal Processing and Communications Applications Conference (SİU'04)*
[3]. Sak, H., Güngör, T., & Safkan, Y. (2005). Generation of synthetic speech from Turkish text. In *13th European Signal Processing Conference, EUSIPCO* 2005, 165-168.
[4]. Orhan, Z., & Görmez, Z. (2008). The Framework of the Turkish Syllable-Based Concatenative Text-to-Speech System with Exceptional Case Handling. *Trans. on Comp.*, *7*(10), 1525–1534.
[5]. Sel, İ., Hanbay, D., Karabatak, M., (2011). Beyin Bilgisayar Arayüzleri için Türkçe Metinden Konuşma Sentezleme Sistemi, *Fırat Üniversitesi Elektrik Elektronik Bilgisayar Sempozyumu, Bildiri Kitabı II, 273-276.*
[6]. Özsoy, A. S. (2004). Türkçe'nin Yapısı-I Sesbilimi [The Structure of Turkish-Phonology]. Boğaziçi Üniversitesi Yayınları.
[7]. Demircan, Ö. (1996). *Türkçenin ses dizimi: sesler, sesbirimler, ayırıcı özellikler, ses değişimleri, vurgu, vurgulama, ezgi, ezgileme*. Der Yayınları.
[8]. Aktas, Ö., & Çebi, Y. (2013). Rule-Based Sentence Detection Method (RBSDM) for Turkish. *International Journal of Language and Linguistics, 1, 1.*
[9]. Koşaner, Ö., Birant, Ç. C., Aktaş, Ö. (2013) Improving Turkish language training materials: Grapheme-to-phoneme conversion for adding phonemic transcription into dictionary entries and course books. In *Procedia Social and Behavioral Sciences, Proceedings of the 13th International Educational Technology Conference.*
[10].

ANNEX

List 1. Phonemes that form the database of the study

a. CV: Word-initial ʔ*V*

1. ʔa a.na
2. ʔe e.li
3. ʔı ı.sı
4. ʔi i.şi
5. ʔo o.kul
6. ʔö ö.zel
7. ʔu u.yuz
8. ʔü ü.tü

b. CV: Word-final CV

9. ya bo.ya
10. ye gön.ye
11. yı a.yı
12. yi i.yi
13. yo ban.yo
14. yö mös.yö
15. yu ku.yu
16. yü bü.yü

c. CVC: Word-initial ʔ*ÜZ*
a*

17. ʔab ab.dal
18. ʔac ac.ziyet
19. ʔaç aç.lık
20. ʔad ad.cıl
21. ʔaf af.gan
22. ʔag ag.nos.tik
23. ʔağ ağ.rı
24. ʔah ah.bap
25. ʔaj aj.var
26. ʔak ak.ça
27. ʔal al.gı
28. ʔam am.bar

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 4, Jul.-Aug., 2024, pp: 109-121 www.ijhssm.org

29. ʔan	an.lık
30. ʔap	ap.tal
31. ʔar	ar.sız
32. ʔas	as.kı
33. ʔaş	aş.kın
34. ʔat	at.kı
35. ʔav	av.cı
36. ʔay	ay.lak
37. ʔaz	az.gın
e*	
38. ʔeb	eb.leh
39. ʔec	ec.dat
40. ʔeç	eç.hel
41. ʔed	ed.na
42. ʔef	ef.kâr
43. ʔeg	eg.za.ma
44. ʔeğ	eğ.ri
45. ʔeh	eh.ven
46. ʔej	ej.der.ha
47. ʔek	ek.mek
48. ʔel	el.çi
49. ʔem	em.zik
50. ʔen	en.gin
51. ʔep	ep.ri.me
52. ʔer	er.baş
53. ʔes	es.ki
54. ʔeş	eş.lik
55. ʔet	et.çil
56. ʔev	ev.cil
57. ʔey	ey.lem
58. ʔez	ez.gi
ı*	
59. ʔıb	N/A
60. ʔıc	N/A
61. ʔıç	N/A
62. ʔıd	N/A
63. ʔıf	N/A
64. ʔıg	N/A
65. ʔığ	ığ.dır
66. ʔıh	ıh.la.mur
67. ʔıj	N/A
68. ʔık	ık.la.mak
69. ʔıl	ıl.gın
70. ʔım	N/A
71. ʔın	ın.ca.lız
72. ʔıp	N/A
73. ʔır	ır.gat
74. ʔıs	ıs.lak
75. ʔış	ış.kı
76. ʔıt	ıt.lak
77. ʔıv	N/A
78. ʔıy	ıy.diye
79. ʔız	ız.ga.ra
i*	
80. ʔib	ib.ra.him
81. ʔic	ic.raat
82. ʔiç	iç.ki
83. ʔid	id.man
84. ʔif	if.rit
85. ʔig	ig.lu
86. ʔiğ	iğ.renç
87. ʔih	ih.san
88. ʔij	N/A
89. ʔik	ik.ti.sat
90. ʔil	il.kin
91. ʔim	im.ge
92. ʔin	in.di.rim
93. ʔip	ip.lik
94. ʔir	ir.kil.me
95. ʔis	is.tek
96. ʔiş	iş.lik
97. ʔit	it.ki
98. ʔiv	iv.me
99. ʔiy	N/A
100. ʔiz	iz.ci
o*	
101. ʔob	ob.ruk
102. ʔoc	N/A
103. ʔoç	N/A
104. ʔod	od.yak.maz
105. ʔof	of.la.ma
106. ʔog	N/A
107. ʔoğ	oğ.lak
108. ʔoh	oh.la.ma
109. ʔoj	N/A
110. ʔok	ok.çu
111. ʔol	ol.gu
112. ʔom	om.buds.man
113. ʔon	on.luk
114. ʔop	op.ti.mist
115. ʔor	or.du
116. ʔos	os.man
117. ʔoş	N/A
118. ʔot	ot.lak
119. ʔov	ov.ma
120. ʔoy	oy.lum
121. ʔoz	oz.moz
ö*	
122. ʔöb	N/A
123. ʔöc	N/A
124. ʔöç	öç.len.me
125. ʔöd	öd.lek
126. ʔöf	öf.ke
127. ʔög	ög.le.na
128. ʔöğ	öğ.ren.ci
129. ʔöh	N/A
130. ʔöj	N/A
131. ʔök	ök.çe

132.ʔöl öl.gün
133.ʔöm öm.rün.ce
134.ʔön ön.lük
135.ʔöp öp.me
136.ʔör ör.gü
137.ʔös ös.ta.ki
138.ʔöş N/A
139.ʔöt öt.me
140.ʔöv öv.gü
141.ʔöy öy.kü
142.ʔöz öz.gü

u*

143.ʔub N/A
144.ʔuc N/A
145.ʔuç uç.kur
146.ʔud N/A
147.ʔuf uf.lama
148.ʔug N/A
149.ʔuğ uğ.run
150.ʔuh uh.re.vi
151.ʔuj N/A
152.ʔuk uk.te

153.ʔul ul.vi
154.ʔum um.re
155.ʔun un.luk
156.ʔup N/A
157.ʔur ur.gan
158.ʔus us.lu
159.ʔuş uş.kun
160.ʔut ut.ku
161.ʔuv N/A
162.ʔuy uy.gun
163.ʔuz uz.la.şı

ü*

164.ʔüb N/A
165.ʔüc üc.ra
166.ʔüç üç.lü
167.ʔüd N/A
168.ʔüf üf.le.me
169.ʔüg N/A
170.ʔüğ üğ.rüm
171.ʔüh N/A
172.ʔüj N/A
173.ʔük N/A
174.ʔül ül.ke
175.ʔüm üm.ran
176.ʔün ün.lü
177.ʔüp N/A
178.ʔür ür.per.ti
179.ʔüs üs.tün
180.ʔüş üş.me
181.ʔüt üt.me
182.ʔüv N/A
183.ʔüy N/A
184.ʔüz üz.gün

d. Exceptions

185.$V_{[round]}$f	ü**f**.lemek	bilabial f
186.f$V_{[round]}$	ü.**f**ür.mek	bilabial k f
187.h$V_{[back]}$	**h**a.yır	glottal h
188.Vh$V_{[back]}$	a.**h**ır	velar h
189.h$V_{[front]}$	**h**iç	palatal h
190.g$V_{[back]}$	dur.**g**un	velar g
191.g$V_{[back]}$	**g**er.**g**in	palatal g
192.k$V_{[back]}$	**k**a.tı.	aspirated velar k
193.k$V_{[front]}$	**k**e.sin.	aspirated palatal k
194.$V_{[back]}$k	a**k**.kor	velar k
195.$V_{[front]}$k	e**k**.şi	palatal k
196.V[back]l	a**l**.gı	velar l
197.l$V_{[back]}$.	**l**a.**l**a	velar l
198.$V_{[front]}$l	e**l**.çi	alveolar l
199.l$V_{[front]}$.	**l**â.**l**e	alveolar l
200.m$C_{[labiodental]}$	a**m**.fi	labiodental m
201.$V_{[back]}$n$C_{[velar]}$	ya**n**kı	velar n
202.$V_{[front]}$n$C_{[palatal]}$	ye**n**ge	palatal n
203.pV	**p**a.nik	aspirated p
204.r#	ki**r**	voiceless r
205.VrV	i**r**i	tap r
206.tV	**t**a.kı	aspirated t
207.$V_{[round]}$v	**v**ur.kaç	bilabial v
208.v$V_{[round]}$	a.**v**uç	bilabial v