
IMPLEMENTATION OF TURKISH TEXT-TO-SPEECH SYNTHESIS ON A VOICE SYNTHESIZER CARD WITH PROSODIC FEATURES

Hakan TORA ^{1,*}, İbrahim Baran USLU ², Timur KARAMEHMET ³

¹ Avionics Department, School of Civil Aviation, Atılım University, Ankara, Turkey

² Electrical-Electronics Engineering Department, Engineering Faculty, Atılım University, Ankara, Turkey

³ Roketsan A.Ş., Ankara, Turkey

ABSTRACT

This study is on hardware implementation of the Turkish text-to-speech (TTS) synthesis with a voice synthesizer card. Here, a fully functional TTS system, capable of synthesizing every Turkish text, including abbreviations, numbers, etc. is designed and implemented. The system is additionally enriched by applying some prosodic attributes for more intelligible and natural speech production. A set of rules required for proper pronunciation and stress patterns are precisely defined in a lexicon utilized for synthesizing Turkish speech. Performance of the developed system is assessed by the Mean Opinion Score (MOS) test. An average score of 3.29 out of 5 is achieved. It indicates that the proposed synthesizer can be successfully integrated to many practical Turkish TTS applications.

Keywords: Text-to-Speech synthesis, RC8660 voice synthesizer, Phonetic transcription, Prosody, MOS test

1. INTRODUCTION

Text-to-Speech synthesis is an attractive area in speech technology. It is defined as an automatic conversion of text into speech. It finds many useful applications in daily life such as assistive products for visually handicapped people, reading and announcing systems. A typical TTS system is shown in Figure 1. There are three main blocks, which are text segmentation, phoneme conversion and waveform conversion. Two additional blocks are required to design a complete TTS system. They are phonetic transcription and prosody blocks. The former introduces pronunciation norms of a language while the latter describes stress and intonation rules. They together improve the intelligibility, an indicator of how well the synthesized words are understood, and the naturalness, specifying a degree of closeness to human voice, of the produced speech. The phoneme-to-waveform conversion block generates speech signal in three ways: articulatory, formant, and concatenative synthesis [1–3]. It is well known that, among the three methods, concatenative synthesis produces more natural and comprehensible speech since it uses prerecorded human voice samples [4]. This paper describes a Turkish synthesizer implemented with a voice synthesizer card (VSC) namely RC8660. This card is a text-to-speech processor that is originally designed for English. It comprises of the main components of a TTS system shown in Figure 1. We focus only on phoneme conversion, phonetic transcription rules, and prosodic attributes. The text segmentation and speech generation in our TTS system are handled by VSC. Our aim was to adopt the card's own capabilities so that it will speak Turkish. Consequently, the proposed system does not need any Turkish speech database. Instead, it exploits the phonemes already available in RC8660 recorded for English. Since Turkish is a phonetic language, that is each letter corresponds to a particular sound unit, it is always possible to find similar Turkish sounds in other languages. In other words, any multilingual synthesis system can easily be adapted to Turkish.

*Corresponding Author: hakan.tora@atilim.edu.tr

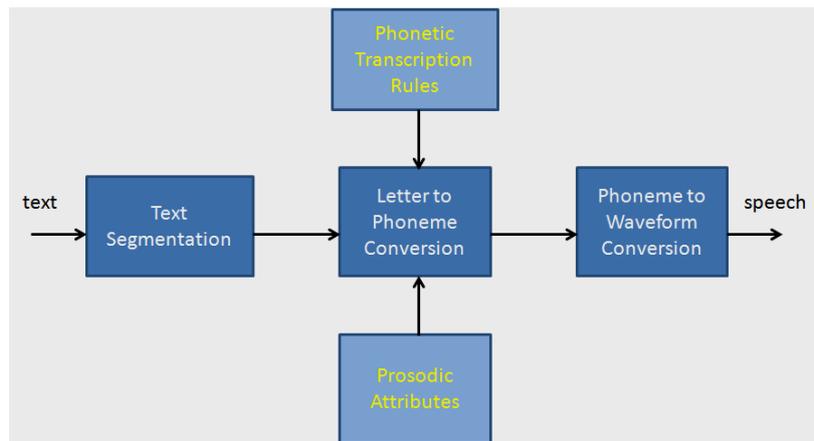


Figure 1. A typical TTS system

Besides being a phonetic language, Turkish is also agglutinative. Therefore, this is another advantage of Turkish for TTS systems. One can find many studies on Turkish TTS in literature. Some of the recent works are about realizing text-to-speech conversion in either software or hardware [5–11]. The others are specifically on prosodic improvements such as duration, pitch contour, stress, and tone analyses [12–16].

There exist also multilingual synthesizers developed by research teams for various languages, such as: MBROLA, FESTIVAL, TTSBOX, HTS, MaryTTS [17, 18]. They are all PC-based synthesizer systems.

The contribution of our work can be seen as building a Turkish TTS system with a card dedicated to voice synthesis in English. To our best knowledge, there are a less number of hardware realizations for Turkish TTS [5,7,10]. In [5], TMS5220 speech synthesizer is employed for purpose of generating synthetic speech. The system uses pre-recorded sounds represented by LPC parameters. Also, the quality of the produced speech is not evaluated in this work. In [7], a chip is developed for synthesizing Turkish Language. The authors propose a system with a language dependent database which employs prerecorded human speech examples. On the other hand, our system contains merely the phonemes, the smallest speech unit, instead of prerecorded speech samples. Furthermore, despite our TTS system is implemented for Turkish, it can be readily used for any other language by making the proper matching between the phonemes in VSC and the target language. Our study is not intended to develop a voice synthesizer IC card. Instead, an available synthesizer card in the market is purchased and utilized to implement a Turkish TTS system. In [10], the authors utilize a sound synthesizer integrated circuit called SpeakJet Magnevation to vocalize the Turkish text. They do not take prosodic attributes into consideration while producing speech. Moreover, the performance of the developed system is not assessed at all. The study in this paper fills that gap by offering an IC-based synthesizer system that has a feature of being embedded to many practical applications. The proposed synthesizer therefore differs from the PC-based systems. It is well known that TTS systems in the literature recruit not only phonemes but also larger speech units such as diphones, triphones, syllables and even words. It is needless to say that the larger the speech units, the less the number of concatenations. Thus this will lead to a more natural synthesis, which is a desired property for a TTS system, but also requires more storage. The latter makes a TTS system unpreferable for practical applications, because it causes a huge computational burden along with memory spaces. On the other hand, our system needs only a phoneme database, where each phoneme matches with a letter in the alphabet, which makes our speech inventory fixed-length. That makes the proposed system very useful for real world applications, because practical tasks are supposed to be realized in a fast way. In addition, the proposed phoneme-based TTS system is enriched by some Turkish prosodic attributes.

Consequently, our TTS system produces natural speech comparable to corpus-based TTS systems based on the MOS tests [4,11].

Organization of the paper is as follows. Section 2 introduces the TTS implementation platform. Turkish phoneme conversion and pronunciation rules, prosodic features, and implementation of Turkish TTS system are discussed in Section 3 in detail. Section 4 includes the findings and discussions of the tests. Conclusion is presented in Section 5.

2. TTS IMPLEMENTATION PLATFORM

We selected RC8660 voice synthesizer in this work as a platform to implement Turkish text-to-speech conversion practically. That chip is a multilingual voice synthesizer equipped with a text-to-speech (TTS) processor. In order for this processor to speak in other languages, set of phoneme mapping, pronunciation of special characters (abbreviations, numbers, and symbols), and target language rules are to be defined in a dictionary. Additionally, the processor provides us with control capabilities on prosodic attributes of phonemes such as amplitude, duration, and pitch.

RC8660 generates a phonetic transcription of the written words and also helps us hear the synthesized speech. For example, the word “akşam” (evening) is represented phonetically as “aa k sh aa m”. It allows voice attributes to be modified on phoneme boundaries within a word. Thus, we are able to change the stress applied to any part of a word. The VSC has six more prosodic features: expression, reverb, formant frequency, tone, articulation, and delay. Figure 2 illustrates a Graphical User Interface (GUI) designed for controlling all the TTS prosodic attributes.

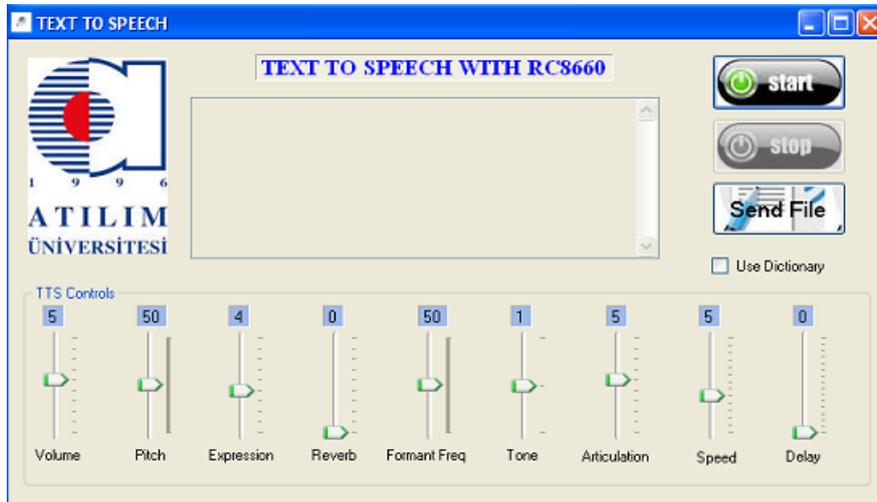


Figure 2. A GUI for TTS attributes

The software of VSC is RCStudio. Its screenshot can be seen in Figure 3. As seen, there are four functional parts from left to right namely, exception dictionary, control panels, text box and phonetic output. The text to be synthesized is entered through the text box. Phonetic output box represents the input text phonetically. TTS controls box allows prosodic features to be modified. Since RC8660 utilizes the lexicon and letter-to-sound conversion rules of English, one should prepare a dictionary including the rules for the target language. This dictionary can be readily constructed in the dictionary box appearing on left most side in Figure 3. One can activate the exception dictionary by selecting the “Use Dictionary” button in the control panel box. Note that the generated exception dictionary can be used simultaneously with the built-in English dictionary, if any rules belonging to the target language are not included.

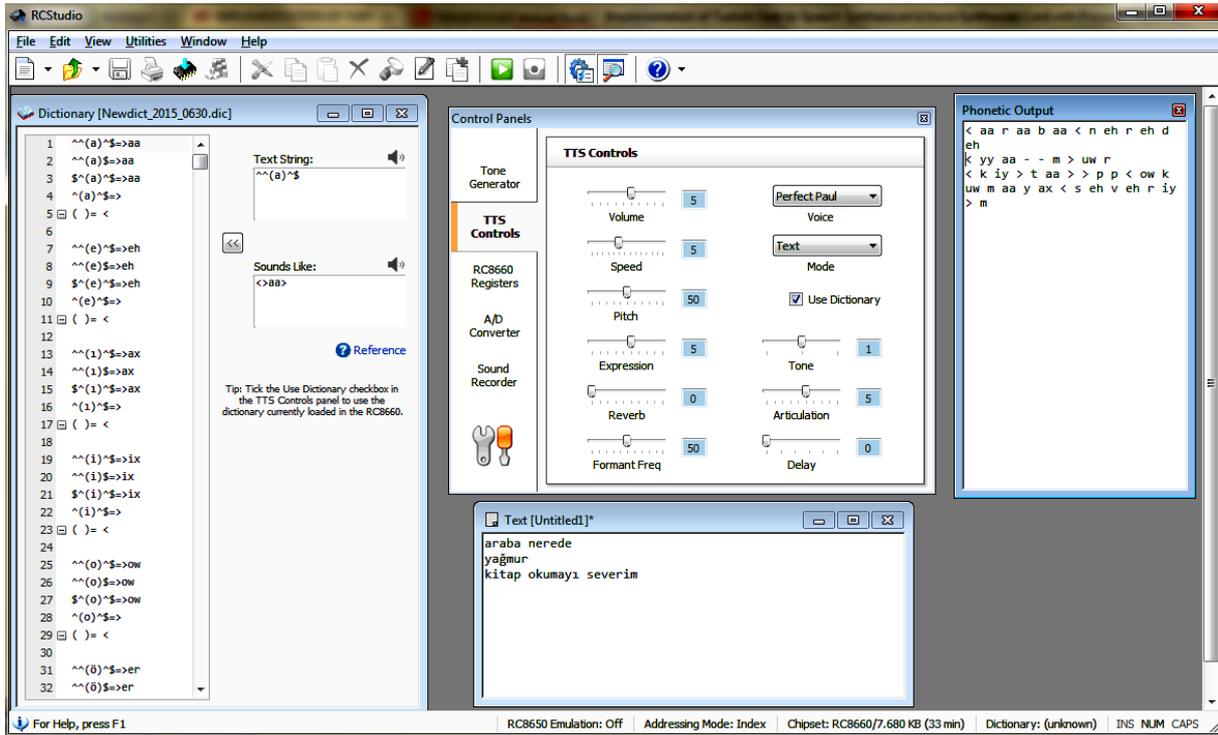


Figure 3. A Display of RCStudio

3. ADAPTATION OF THE VSC TO TURKISH TTS

Turkish TTS synthesis system is realized by the RC8660 speech processor in this paper. The processor uses the concatenation method, one of the TTS techniques, to produce spoken words. In concatenative synthesis, recorded speech units are joined to each other in a rule-based manner. It is also well known that this method produces more natural and comprehensible voices than that of articulatory and formant based synthesizers.

Concatenative synthesizers can be constructed by using either fixed length or variable length speech units. The developed system, on the other hand, uses built-in phonemes of RC8660 as speech units, which are inherently fixed-length. Phonemes play a very important role in TTS systems because they have a major influence on the quality of the produced speech.

Every concatenative TTS system needs a speech unit database. Creating such a database is time consuming and also requires severe effort. However, adapting VSC to Turkish negates such burdens, since VSC already comes with the recorded phonemes. Making use of this English phoneme database in VSC, we matched Turkish phonemes with their English counterparts. Hence, there was no need for recording a new speech database for Turkish. Such adaptation resulted in a considerable Turkish speech generation.

There are 8 vowels and 21 consonants in Turkish. Even though we tried to pair Turkish letters with proper English sounds, these pairings were limited with the recorded 55-phoneme database of VSC. As can be seen from Table 1, each Turkish letter is assigned to at least a single phoneme represented in the phonetic notations of both VSC and IPA (International Phonetic Alphabet). IPA is addressed frequently by the researchers from TTS groups. Therefore, in order to make mapping from Turkish letter to English phoneme in VSC, IPA was taken as reference and Table 1 was built by the following steps:

Step-1: Determine IPA notations corresponding to Turkish letters.

Step-2: Choose an English word having the same IPA representation as the Turkish letter.

Step-3: Vocalize the selected word in the VSC and determine the corresponding VSC's phonetic notation.

Step-4: Form the phoneme sets in Table 1.

On the other hand, there are some letters particular to Turkish, which are {ö, ü, ı, ç, ğ, ş}. Since these letters do not exist in English, the VSC does not have the corresponding phonemes. However, their pronunciations can be caught in the sounds of some English words. For instance, the letter “ö” is pronounced as in the English word “bird”. It is also worth noting that the letter “ü” is vocalized by employing the two VSC phonemes, namely “yy” (from the word: “yes”) and “uh” (from the word: “put”). We must also mention about “ğ” here. As seen in Table 1, it has a corresponding phoneme neither in IPA nor in VSC. This implies that the letter is not vocalized in English. Additionally, even though “ğ” takes place in transcription in Turkish, whether it is pronounced or not will not have a critical effect on the intelligibility of the word. Usually, it disappears in pronunciation and the vowel prior to it is lengthened. Hence, the duration of the vowel preceding ‘ğ’ is increased.

3.1. Creating a Lexicon for Turkish

As stated earlier, RC8660 is for multilingual usage. In order to enable VSC to speak in any language, one should construct a lexicon, called exception dictionary in RC8660 terminology. All possible pronunciation rules for the characters encountered must be defined in the lexicon. A useful and thorough study on Turkish pronunciation for TTS applications can be found in [19].

RC8660 accepts any text as a string of words. Each word is processed character by character. Each character is then pronounced according to the rules. These rules are searched sequentially for each character until a rule is found in the lexicon. In order to get a deep understanding, let us consider the VSC's syntax:

$$\mathbf{Left(Character)Right = Pronunciation}$$

where the *Character* indicates the text segment to be pronounced. The preceding and succeeding characters, presented in the *Left* and *Right* in that form, specify the way the *Character* is pronounced. The *Pronunciation* is the phonetic representation of the *Character*. As an example:

$$\mathbf{\wedge(a)\wedge = aa}$$

where ‘ \wedge ’ denotes any consonant and ‘a’ having a consonant before and after will be pronounced by the phoneme ‘aa’.

After the letters are converted into phonemes, next operation in TTS is to connect these phonemes according to Turkish syllabification rules. There exist seven types of syllables in Turkish shown in Table 2. As can be seen from Table 2, every vowel within a syllable attracts the preceding consonant. All Turkish syllable forms and their pronunciation rules are defined into the exception dictionary. The following is an example of the pronunciation rule of a syllable in the CV form:

$$\mathbf{a(r)a = r aa}$$

Here, ‘ara’ text fragment will be pronounced with ‘r aa’.

All the words including the text fragment of ‘ara’ will use the above defined rule. Hence, the word “araba” (car) is correctly syllabified and pronounced as “a-ra-ba” by the VSC. Unless these syllable forms were introduced, the VSC would use its built-in English pronunciation rules and syllabify that word as “a-rab-a”, which violates Turkish pronunciation rule.

Table 1. Turkish phoneme pairing set

Turkish Letter	VSC Phoneme	English Example Word	IPA
a	aa, ah, ay	<u>A</u> rm, C <u>u</u> t, F <u>i</u> ve	a:, ^, aɪ
ı	ax	Bot <u>o</u> mm	-
e	ey, ae, eh	<u>E</u> ight, C <u>a</u> t, B <u>e</u> d	eɪ, æ, e
i	ih, iy, ix	S <u>i</u> t, H <u>e</u> at, R <u>a</u> bb <u>i</u> t	ɪ, i:, i
o	ow, oy	H <u>o</u> me, B <u>o</u> y	oo, oɪ
ö	er	T <u>u</u> rn	ɜr
u	uh, uw	P <u>u</u> t, F <u>o</u> od	ʊ, u:
ü	yy uh	C <u>ü</u> re	ʊə
b	b	<u>B</u> ad	b
c	j	<u>J</u> oy	dʒ
ç	ch	<u>C</u> heck	tʃ
d	d	<u>D</u> ad	d
f	f	<u>F</u> ind	f
g	g	<u>G</u> et	g
ğ	No sound	Not Available	No sound
h	h	<u>H</u> at	h
j	zh	Meas <u>u</u> re	ʒ
k	k	<u>C</u> ut	k
l	l	<u>L</u> eft	l
m	m	<u>M</u> an	m
n	n	<u>N</u> ine	n
p	p	<u>P</u> en	p
r	r	<u>R</u> un	r
s	s	<u>C</u> ity	s
ş	sh	<u>S</u> he	ʃ
t	T	<u>T</u> ea	t
v	V	<u>V</u> oice	v
y	Yy	<u>Y</u> es	j
z	Z	<u>Z</u> oo	z

In every language, there are some types of text such as abbreviations, numbers, dates and special characters which need to be uttered according to the pronunciation rules of that language. Thus, a TTS system developed should consider the pronunciation of these non-standard characters. Vocalization of such texts is accomplished in our TTS. For example, each character in abbreviations in Turkish is separately read. For instance, “THY” (Turkish Airlines) is uttered as “t eh - h eh - y eh”. When it comes to numbers, they are not pronounced digit by digit as in English. For example “1923” is read as either “one thousand nine hundred twenty three” (as a date), or “nineteen twenty three” (as an extension) in Turkish according to the context.

Table 2. Turkish syllable forms

Type V: Vowel C: Consonant	Example Word	Syllabified Word
V	araba (car)	<u>a</u> -ra-ba
CV	şekil (shape)	<u>şe</u> -kil
VC	eldiven (glove)	<u>el</u> -di-ven
CVC	asansör (elevator)	a- <u>san</u> - <u>sör</u>
VCC	Altlık (base)	<u>alt</u> -lık
CVCC	korkmak (to scare)	<u>kork</u> -mak
CCVC	tren (train)	<u>tren</u>

3.2. Creating a Lexicon for Turkish

Prosodic attributes module in Figure 1 supplies amplitude, pitch, and duration adjustments for utterances. These features are effective on the intelligibility and naturalness of produced speech. We achieved some prosodic improvements by modifying these attributes. These manipulations on the attributes are general, not text-dependent. Thus our Turkish VSC is able to sound any given text with our prosodic rules.

In this context, *i*) word stress, *ii*) intonation in sentences with a question suffix, and *iii*) stress in plosive consonants (p, ç, t, k) were taken into consideration.

Stress in Turkish words applies generally to the last syllable. In this paper, the amplitude is chosen to control the word stress. To do so, the position of the last syllable in every word is first detected and then the volume of the last syllable is relatively increased. For example, the syllabified word “ki-tap” (book) is pronounced as “k iy > t aa p”, where the symbol “>” shows the position of the stress in the word, allowing to increase volume of the following syllable.

In this work, the intonation of interrogation form was also considered. There are such question suffixes as “mı, mi, mu, mü” in Turkish which are used for requesting a Yes/No reply. The intonation of all the sentences having this form is accomplished by highlighting the syllable before those suffixes. For example, the stress at the sentence “Geliyor mu?” (Is he coming?) is given to the syllable “-yor”. This sentence is phonetically represented as “< g eh l iy < / / y o r m uw”, where the set of symbols “< / /” decreases volume and increases pitch (2 steps), respectively. The increment/decrement step numbers of the prosodic attributes were determined empirically by listening to the output speech.

We noticed on our listening tests that words containing plosive consonants: (p, ç, t, k) were not occasionally perceptible enough. In order to overcome this, we emphasized these sounds by duplicating the phonemes and relatively increasing their amplitudes. For example, the word “topuk” (heel) is pronounced as “t o > p p uw > > k k”.

The Turkish particular character “ğ” is another issue. Although it occurs in texts, it is not vocalized; instead it has the effect of lengthening the duration of the preceding vowel. For example in our synthesizer, the word “yağmur(rain)” is sounded as: “yy – – aa ++ m uh r”, where the symbols “–,+” represent duration increment (speed decrease) and decrement (speed increase) operators, respectively. In that phonetic representation, it is seen that each letter except “ğ” has a corresponding phoneme. However, the phoneme “aa” is pronounced longer.

4. TESTS AND FINDINGS

The speech quality of the proposed TTS system is evaluated by subjective listening tests. Mean Opinion Score (MOS) test is utilized to evaluate the performance of the system [20]. A test set consisting of 30 statements (words, phrases, and sentences) is created, which is given in Table 3. The statements are grouped into three parts as *one-word*, *phrase with two words* and *sentences*. While we are preparing the test set, special attention is paid to include particularly the letters {ü, ö, ç, ı} within these statements. The reason that we focus on these letters is to test the adaptation performance of the VSC to Turkish. In addition, the number of the statements in each group is chosen to be the same.

The prepared test set is divided into two equal subsets that are graded at different times by the listeners based on both intelligibility and naturalness. Consequently, we obtain more reliable and objective MOS test grades. Subset-1 is formed by selecting the first 5 statements of each group and Subset-2 is obtained from the rest of the statements. 20 native Turkish listeners, who are not expert on TTS systems, are asked to score each subset according to MOS scale, where grades 1 through 5 correspond to quality measures: bad, poor, fair, good, and excellent.

Figure 4a shows the MOS scores for each statement of Subset-1. Each score is calculated by taking the average of the 20 grades. As clearly seen, overall average of 3.36 is obtained. The minimum score is 2.29 obtained at S11. The maximum score is 4.50 obtained at S5. The average score for each group of statements (*one-word* / *phrase with two words* / *sentences*) is 3.87, 3.09 and 3.11, respectively. Figure 4b shows the results for Subset-2. The overall, minimum, maximum, and the average of each group scores are 3.22, 2.21 (at S19), 3.86 (at S6), 3.47, 3.06, 3.13 respectively.

Table 3. Test set*

No	Group-1 One-word Statements	No	Group-2 Two-word Statements	No	Group-3 Sentences
S1	Araba (Car)	S11	Okulun kapısı (School's gate)	S21	Ali topu tutamadı. (Ali couldn't catch the ball.)
S2	Asansör (Elevator)	S12	Bahçe duvarı (Garden's fence)	S22	Tatile gidemediniz mi? (Couldn't you go on vacation?)
S3	Çarşamba (Wednesday)	S13	Meyve suyu (Fruit juice)	S23	Meyveli yoğurt yedim. (I ate fruit yogurt.)
S4	Dokuz (Nine)	S14	TBMM binası (TBMM building)	S24	Bugün 20 Temmuz 2015. (Today is 20 July 2015.)
S5	Nisan (April)	S15	Halkalı şeker (candy ring)	S25	Hesabı alabilir miyim? (May I have the bill?)
S6	Ramazan (Ramadan)	S16	Elma hoşafı (apple compote)	S26	Kitap okumayı severim. (I like reading a book.)
S7	Salı (Tuesday)	S17	200 TL (200 TL)	S27	Bu yemek çok tuzsuz. (This meal is very unsalted.)
S8	Sevecen (Loving)	S18	Anne sevgisi (Mother's love)	S28	Hava çok güzeldi. (The weather was very nice.)
S9	Telefon (Telephone)	S19	Genel kültür (General knowledge)	S29	Spor yapıyorlar mı? (Do they work out?)
S10	Kulaklık (Headphone)	S20	Horoz sesi (Rooster voice)	S30	Karadeniz yeşil bir bölgedir. (BlackSea is a green region.)

*: The synthesized sound files are available from the corresponding author if demanded.

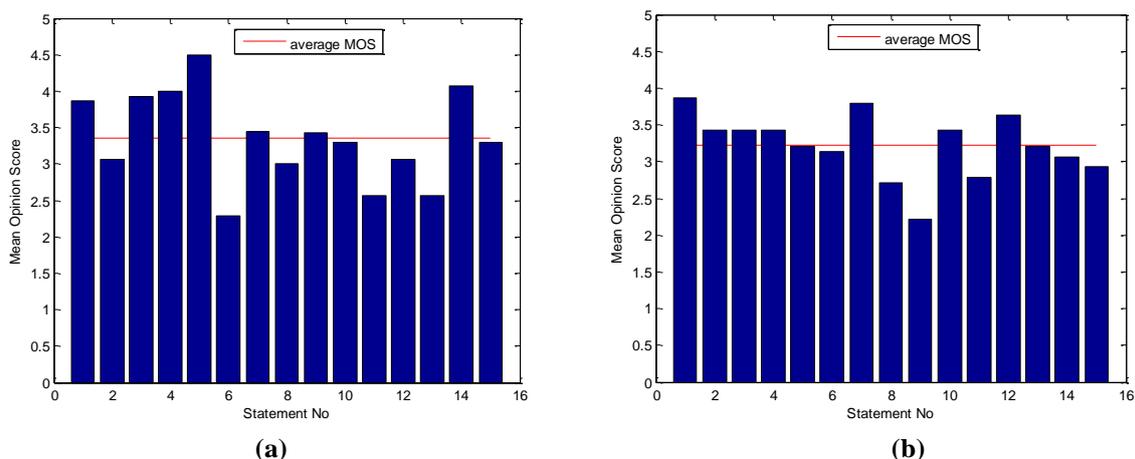


Figure 4. MOS results for Subset-1 and Subset-2

When we compare the scores for each group in Figure 4a and 4b, they show consistency with each other. Considering both subsets together, the total performance score is calculated as 3.29.

According to MOS grades, the obtained scores for each group are said to be good enough. As expected, it is easier to understand *one-word* statements than the ones in other groups; hence the maximum score is achieved. Also, the *sentences* are easier to comprehend than the *two-word* statements, since we, as a human being, are able to complete a sentence by only hearing a few words even if we could not hear each word of the sentence. That explains the reason of having moderate scores for this group.

According to our test results, only six statements received grades below 3.0, which indicate poor or bad quality. The low results can be blamed on the phonetic transcription rules. VSC originally is equipped with the English language. Therefore, all the phonetic transcription rules are already defined for English. As mentioned in Section 2, in order to use VSC for any language, an exception dictionary is to be defined for the used language. Even though we tried to cover all the Turkish phonetic transcription rules in the exception dictionary, one can still run into a possible case which is not considered in the dictionary. In such cases, VSC applies English phonetic transcription rules to the encountered statements. That results in vocalizing the Turkish words with sounds unfamiliar to a native Turkish speaker. In other words, the developed system utters Turkish statements partially with an English accent. In addition, it is observed that improper duration settings cause degradations on the comprehension of the words.

5. CONCLUSION

This study introduces a practical implementation of concatenative Turkish TTS adapting a voice synthesizer card named RC8660. This card is originally designed for vocalizing English statements. It possesses the letter-to-phoneme conversion rules for 55 English phonemes. That motivated us to use those phonemes for developing a practical Turkish TTS system without recording Turkish phonemes. First, English phonemes are coupled to Turkish phonemes as properly as possible. Second, both Turkish syllabic rules and prosodic attributes are adopted to the VSC. As well known, prosodic features are critical for a statement to be vocalized in a natural way. The naturalness of the synthetic speech significantly influences the performance of a synthesizer. It is observed that the performance of the proposed Turkish TTS system shows acceptable and promising results in terms of the intelligibility and naturalness of the synthesized speech. An average MOS score of 3.29 is achieved on the selected statements.

Since the speech unit database of the developed TTS system is sufficiently small compared to the corpus based systems, the system can be employed in real world applications requiring less computational complexity. Besides, that system can also be utilized as a test bed for speech technology developers. Having that system at hand, one can easily make prosodic tests and observe the influence of each prosodic parameter such as pitch, speed, volume, delay, and tone, on the output speech to evaluate the synthesized speech quality.

Consequently, with this work, a hardware Turkish TTS system equipped with a proper prosody is implemented.

REFERENCES

- [1] Dutoit T. An Introduction to Text-to-Speech Synthesis. Dordrecht: Kluwer Academic Publishers, 1997.
- [2] Taylor P. Text-to-Speech Synthesis. Cambridge University Press, 2009.
- [3] Aida-Zade KR, Ardil C, Sharifova AM. The main principles of text-to-speech synthesis system, International Journal of Computer, Electrical, Automation, Control and Information Engineering 2013; 7.
- [4] Sak H, Gungor T, Safkan Y. A corpus based concatenative speech synthesis system for Turkish, Turkish Journal of Electrical Engineering 2006; 14: 209-223.
- [5] Gercek A. A TMS 5220 based speech synthesis development system. Middle East Technical University, Graduate School of Sciences and Engineering, 1991.
- [6] Bozkurt B, Dutoit T, D'alessandro C, Pagel V, Prudon R. Improving quality of MBROLA synthesis for non-uniform units synthesis, Proceedings of the IEEE TTS Workshop, Santa Monica, 2002.
- [7] Aktan O, Baskaya IF, Dundar G. A single chip solution for text-to-speech synthesis, Proceedings of the European Conf. on Circuit Theory and Design, 2005.
- [8] Asliyan R, Gunel K. Türkçe metinler için hece tabanlı konuşma sentezleme sistemi, X. Akademik Bilişim Konferansı (in Turkish), 2008.
- [9] Orhan Z, Gormez Z. A concatenative Turkish text-to-speech system and evaluation process, 6th International Conference on Electrical and Electronics Engineering (ELECO), 2009.
- [10] Tora H, Cengizler Ç. Producing synthetic speech from Turkish text via a single sound synthesizer IC. National Conf. on Electrical, Electronics and Computer Engineering (ELECO) 2010.
- [11] Guner E. and Demiroğlu C. A small footprint hybrid statistical/unit selection text-to-speech synthesis system for agglutinative languages. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.
- [12] Oskay B, Salor O, Ozkan O, Demirekler M, Ciloğlu T. Intonation abstraction from text and its applications for Turkish sentences. 9th IEEE Signal Processing and Communications Applications Symposium, 2001.

- [13] Sayli O, Arslan LM, Ozsoy AS. Duration properties of the Turkish phonemes. 11th International Conference on Turkish Linguistics (ICTL), KKTC, 2002.
- [14] Kulekci MO, Oflazer K. An infrastructure for Turkish prosody generation in text-to-speech synthesis. 15th Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN), 2006.
- [15] Oflazer K, Inkelas S. The architecture and the implementation of a finite state pronunciation lexicon for Turkish. *Computer Speech & Language* 2006; 20.
- [16] Uslu IB, İlk HG, Yilmaz AE. A rule based prosody model for Turkish text-to-speech synthesis. *Technical Gazette* 2013; 20.
- [17] Yurtay N, Çelebi S, Gunduz AB, Bicil Y. A mobile product recognition system for visually impaired people with Iphone 4. *Global Journal on Technology* 2013; 3.
- [18] Schröder M, Trouvain J. The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology* 2003; 6: 365-377.
- [19] Akbulut A, Adiguzel T, Yilmaz AE. Statistical syllable analysis for pronunciation ambiguity detection and resolution in text-to-speech synthesis applications: A case study in Turkish. *Acta Polytechnica Hungarica* 2011; 8.
- [20] Streijl, RC, Winkler S, Hands DS. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems* 2016; 22(2).