Statistical Analysis of Syllable Duration of Uyghur Language

Askar Hamdulla, Guzalnur Dilmurat
Institute of Information Science and Engineering
Xinjiang University
Urumqi 830046, China
askar@xju.edu.cn, 1453091752@qq.com

Gulnur Arkin, Mijit Ablimit
Institute of Information Science and Engineering
Xinjiang University
Urumqi 830046, China
gulnur407@163.com, 125195364@qq.com

Abstract—**Phonetics is both an ancient and a young subject. Syllables are important units of speech. Based on the data requirements of speech synthesis and speech recognition, this paper studies from the perspective of experimental phonetics. Firstly, different syllable words are counted from the large-scale "Speech Acoustic Parameters Database of Uyghur Language", including monosyllable words, two-syllable words, three-syllable words and four-syllable words. Secondly, the prosodic parameters are extracted, and statistical analysis is made. Accordingly, the duration distribution of different length words for male and female speakers are studied, and the fixed CV type syllable duration of consonant , the duration of vowel, the whole syllable duration and the pitch of syllable are extracted and analyzed. The effect of different vowels on the duration of CV syllables is further studied, and provided the reliable parameter basis for Uyghur speech synthesis and speech recognition.**

Keywords-Syllable(CV); Duration; Uyghur language; Statistical analysis; Speech synthesis

I. INTRODUCTION

Uyghur language belongs to the Altai language and is attributed to adhesive language type in form. Scholars agree that there are eight unit phonetic bits a[ɑ], ä[ɛ], e[e], i[i], o[ɔ], ö[ø], u[u], ü[y], no compound vowels [1-3]. There are six kinds of syllable types in Uyghur language, such as V, CV, VC, CVC, VCC and CVCC. Only one vowel appears in each syllable [4]. The phonetic syllable of Uyghur is composed of vowels and consonants. Since the syllable can be formed by a unitary sound, the vocal sounds are often subjected to precise acoustic measurement in speech recognition technology. Although more Uyghur phonetics research has done a lot of work in vowels, consonants, pronunciation parts, pronunciation methods and so on, there are many basic issues still need to be studied [5].

Phonetics is one of the most important part of language, is a manifestation of language. The science of phonetics is called "phonetics". Phonetics is divided into traditional phonetics and experimental phonetics. Traditional phonetics mainly describe the phonetic features of a language by means of listening and recording. However, people's hearing ability has some limitations and subjectivity. Therefore, later more modern phonetic experiments, data, technology and other more objective research methods gradually occupy the voice of the dominant position [6]. Duration is one of the important prosody features, and is acoustic parameters that are sensitive to the speech perception of an ordinary audience. The duration of syllables varies with interaction of syllable and rhythm features. For speech synthesis, natural speech can be synthesized only by effectively mimicking changes in the duration of natural speech [7].

Syllable is the smallest phoneme unit in which the phonemes are combined together. Syllable is also one of the most frequently debated problems in language learning [8]. The inherent words of the Uyghur language are basically monosyllabic or disyllabic. Syllables cannot consist of only a few consonants and must contain vowels. The number of vowel phonemes in a word must be equal to the number of syllables, in a word, how many vowels in the sentence is divided into how many syllables. In order to improve the accuracy of Uyghur speech recognition, it is necessary to analyze the duration of syllables of recognized speech and the duration of syllables in different contexts [9]. Although syllables can be directly divided by hearing, however, how to define the syllable exactly, how to scientifically explain the nature of syllable and how to determine the boundary between syllable are always one of the most difficult problems in phonetics [10].

II. DATA SOURCES

A. Sources of Corpus

In order to further analyze and study the duration of Uyghur syllables, considering the importance of analyzing syllables, this paper collected syllable duration from the "Speech Acoustic Parameters Database of Uyghur Language" which is built by the National Ministry of Education and the National Language Commission's National language standards construction and Informatization project on the "Speech Acoustic Parameter Database of Tibet, Uyghur, Yi", and carried out the statistics and analysis. According to statistics, we know that there are 520 (male, female) monosyllabic words, 979 (male) and 986 (female) two-syllable words, 345 (male) and 352 (female) three-syllable words and 103 (male, female) four-syllable words in the database. The corresponding proportions are shown in Figure 1 below.
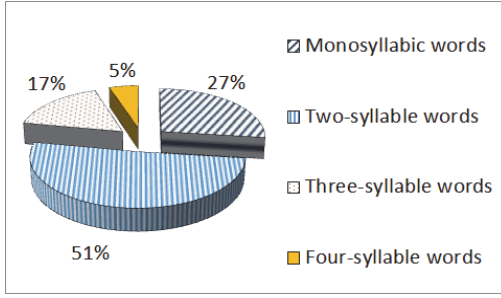
Figure1. Word Distribution in Uyghur Language Database

Also selected 1438 (male) and 1516 (female) CV type syllable. As shown in Table 1. (F - female, M - male).

TABLEI. THE DISTRIBUTION OF CV TYPE SYLLABLES IN DATABASE

	Monosyllable	prefix	Mid-fix	Suffix	Total
F	16	622	325	553	1516
M	16	599	325	543	1483

B. Recording and Data Acquisition

We carried out the recording task in the standard studio of the Ethnology and Anthropology Institute of the Chinese Academy of Social Sciences by using IBM R51 laptop and the external sound card. The pronunciation collaborators are male, female, at the age of 30 to 40, are professional announcers of China National Radio. Each word individually read twice. Acoustic parameters extracted by using the PRAAT Speech Analysis Software [11]. Such as: duration of syllable, duration of consonant, duration of vowel and overall syllable duration of CV type syllables.

III. EXPERIMENTAL DATA ANALYSIS

A. Syllable Duration Distribution Statistics of Uyghur Language

According to the distribution of the speaker's duration, draw male and female speakers monosyllabic words (520 for men and women), two-syllable words (male: 979, female: 986), three-syllable words (male: 345, female: 325) and the distribution of the duration of the four-syllable syllables (103 for men and women). Figure 2 to Figure 9 show the overall distribution of syllable duration for male and female speakers, M (male) for male speakers and F (female) for female speakers.

It can be seen from Fig. 2 and Fig. 3 that the monosyllabic effective duration of male speaker's is 661(ms), the shortest is 227(ms), the average duration of syllable is 438(ms), and the standard deviation is 75. As can be seen from the figure, there is a peak at 436(ms), and the overall distribution of duration is concentrated. of the longest duration of a female speaker's monosyllabic is 705(ms), the short is 105(ms), and the calculated average duration and standard deviation are 421(ms） and 83, respectively. As can be seen from the figure, there is a peak at 486(ms). It is as concentrated as the male speaker. Male

and female speaker duration distributions are consistent with positive distribution.
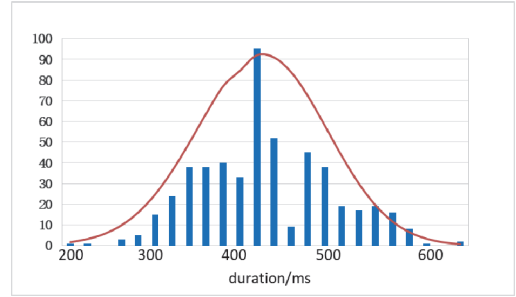

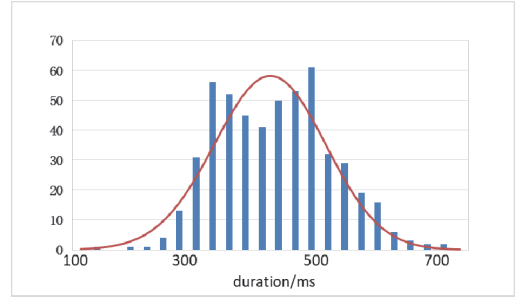Figure2. Duration Distribution of Monosyllabic

Words (M)

Figure3. Duration Distribution of Monosyllabic

Words (F)

As shown in the Fig. 4 and Fig. 5, that the male and female speaker's two-syllable duration are roughly the same.
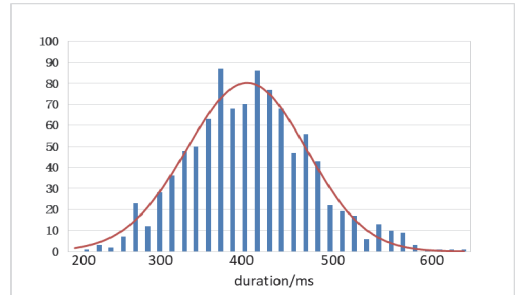

Figure4. Duration Distribution of Two-syllable

Words (M)

Further analysis of the duration distribution reveals that the male speaker's two-syllable duration is 827(ms), the shortest is 293(ms), and the average duration of syllable is 530 ms. Peaks appear at 497(ms)and 548(ms). The longest and shortest duration of female speakers are 943(ms) and 312(ms), respectively, with an average duration of 572(ms). The duration value is mainly concentrated in 557(ms). As a whole, duration of two-syllable words for male and female distribution pattern is uniform. In the general trend, female are longer than male.
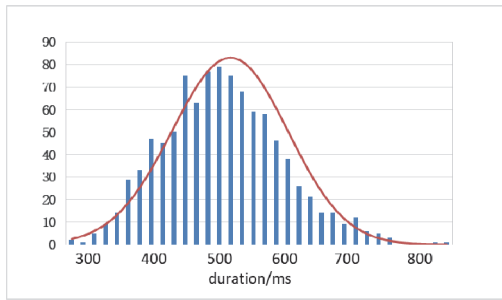
Figure5.   Duration Distribution of Two-syllable

Words (F)

As is shown in the above Fig. 6 and Fig. 7 that the distribution of the duration of the three-syllable words of the male and female speakers is scattered.
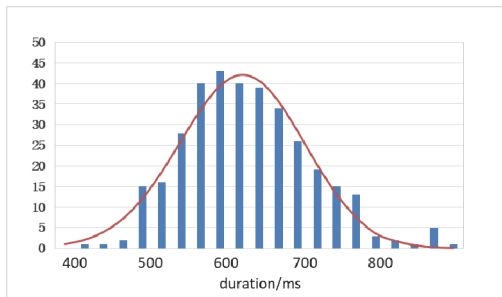

Figure6.   Duration Distribution of Three-syllable
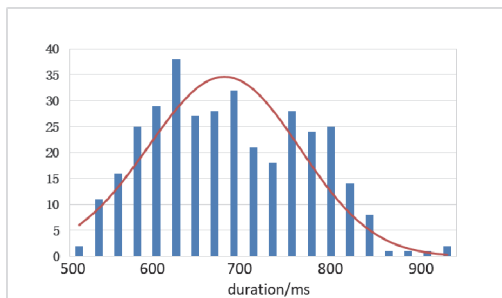
Words (M)

Figure7.   Duration Distribution of Three-syllable

Words (F)

Further analysis shows that the male speaker has a longest duration of 871(ms) and a minimum duration of 397(ms). The calculation shows that the average duration is 619 ms and there is a peak at 595(ms). The average length of the three-syllables of female speakers is 709(m), and the duration distribution is mainly concentrated at 654(m). The longest duration is 978(ms) and the shortest duration is 513(ms). The average duration of female speakers is greater than the average duration of male speakers.

From the Fig. 8 and Fig. 9 above that the average duration of the four-syllables of the male and female

speakers is 768(ms) and 850(ms), respectively, and the longest duration of the male speaker is 1058(ms), and the shortest is 488 ms. It can be seen that there is a peak at 849(ms) on the distribution map. The female speaker has a maximum duration of 1153(ms), a minimum of 564(ms), and a peak at 884(ms). Duration of four-syllable words for male and female overall decentralized distribution of female than male.
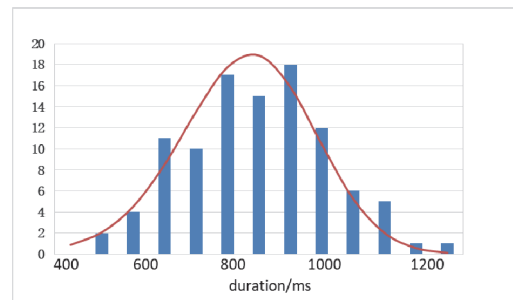

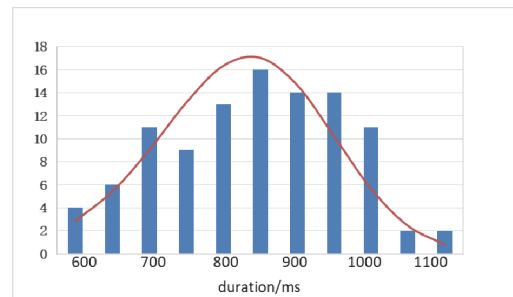Figure8.   Duration Distribution of Four-syllable Words (M)

Figure9.   Duration Distribution of Four-syllable Words (F)

Table 2 shows syllable duration statistics of male and Table 3 shows syllable duration statistics of female. (AVG - average, STDEV - standard deviation).

TABLEII.   DURATION STATISTICS OF AVERAGE SYLLABLE FOR

MALE

Male					
	Max(ms)	Min(ms)	AVG(ms)	STDEV	Sample Points
Monosyllabic words	661	227	438	75	520
Two-syllable words	827	293	530	85	979
Three-syllable words	871	397	619	82	345
Four-syllable words	1058	488	768	116	103

TABLEIII.   DURATION STATISTICS OF AVERAGE SYLLABLE FOR

FEMALE

Female

	Max(ms)	Min(ms)	AVG(ms)	STDEV	Sample Points
Monosyllabic words	705	152	421	83	520
Two-syllable words	943	321	572	99	986
Three-syllable words	978	513	709	95	352
Four-syllable words	1153	564	850	132	103

B. Duration Analysis of CV type Syllable

There are 1438(male) and 1516(female) CV type syllable in "Uyghur speech acoustic parameters database", C represents consonants and V represents vowels. As shown in Table 4, CV type syllable duration of the overall distribution. (CD - consonant duration, VD - vowel duration, SD - syllable duration).

TABLEIV.   DURATION DISTRIBUTION OF CV TYPE SYLLABLE

	Separate pronunciation	Polysyllabic words					
	Monosyllable		Prefix		Mid-fix		Suffix
	M	F	M	F	M	F	M	F
CD（ms）	120	111	98	86	68	69	78	82
VD（ms）	238	255	81	89	67	82	183	218
SD（ms）	358	366	179	180	134	149	261	298

C. Duration of Consonants，vowel and syllable

Generally speaking polysyllabic words pronunciation of consonant is inadequate, especially the consonant in the tail and pronunciation of a faster pronunciation speaker, there will be noise reduction phenomenon, a stop sounds and fricative when the tongue is not in place. With the increase of phonemes in syllables, the phenomenon of inadequate consonant pronunciation is more prominent. In the natural speech stream will be more obvious [12].
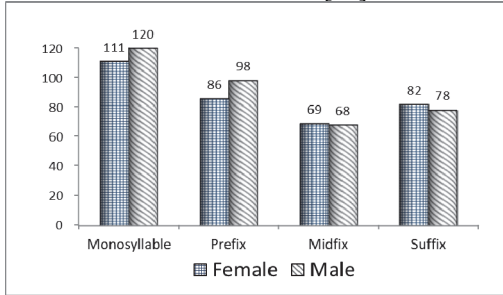

Figure10. Duration of Consonants

As shown in the Figure 10 consonant duration statistics in CV type syllables, overall, the duration of consonants when the CV type syllables are pronounced alone is longer than the prefix syllable, mid-fix syllable and suffix syllable. In a polysyllabic word the prefix syllable has the longest consonant duration, the consonants mid-fix have shortest consonant duration. Duration of the male consonant is longer than the female in the CV type syllables monosyllable and the prefix syllables of polysyllabic words.

Duration of female consonant longer than male in the mid-fix syllable and suffix syllable.

In the Uyghur language, the pitch of vowel is stronger than the consonant. Duration of vowel is lower than the aspiration consonant and fricative, but similar to semivowel, lateral and nasal duration [13]. The duration distribution of vowel pattern of CV type syllables are shown in the Figure11. The duration of vowels when the CV type syllables are pronounced alone is longer than the prefix syllable, mid-fix syllable and suffix syllable. The longest duration of vowels in polysyllabic words is the suffix syllable and the shortest is mid-fix syllable. Whether in monosyllabic words or polysyllabic words, the CV type syllable duration of vowel female higher than the male.
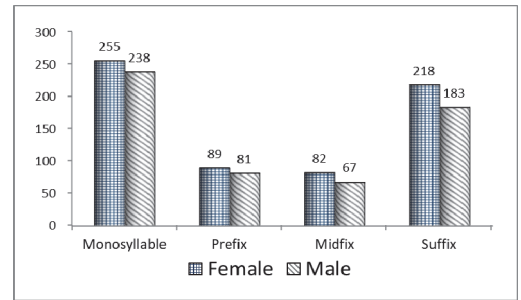

Figure11. Duration of Vowel

Figure 12 shows the duration distribution of CV type syllables contained in the word.
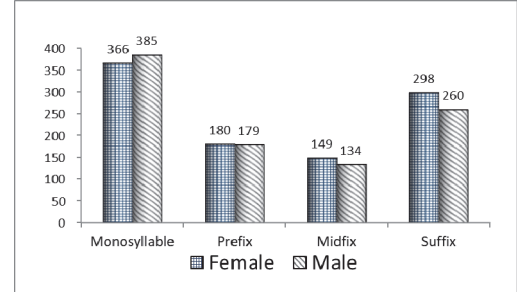

Figure12. Duration of CV type Syllable

Duration CV Type syllable pronounced alone are longer than duration of syllable for prefix syllable, mid-fix syllable and suffix syllable, duration of prefix syllable longer than the mid-fix syllable, the duration of mid-fix syllable is shortest, this means that syllables in the word position will affect duration of syllable. When the CV syllable is monosyllable, the duration of male longer than the female, in polysyllabic words female longer than the male.

D. Pitch distribution pattern of CV syllable

Pitch is determined by the frequency of vibration, if the frequency is high and the pitch is high. Figure 13 is a schematic of the pitch (average) distribution pattern of the CV syllables contained in the word. As can be seen from the figure, the starting frequency distribution pattern is: suffix > mid-fix > monosyllable words > prefix. Frequency

distribution pattern of turning point: monosyllable words > prefix > mid-fix > suffix. The mode of the end point frequency distribution is: prefix > mid-fix > suffix > monosyllable words. CV syllables appear at the end of the word with the highest starting frequency. The end point frequency is lowest. The frequency changes the most.
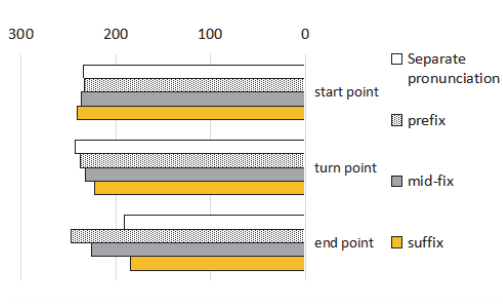

Figure13. pitch distribution pattern of the CV syllables

E. *Statistical Analysis of Different Vowel in CV type Syllables*

There are a total of 32 phonemes in the modern Uyghur standard language, eight of which are vowel phonemes. They are: a[ɑ]、ä[ɛ]、e[e]、i[i]、o[ɔ]、ö[ø]、u[u]、ü[y]. As shown in Table 5：

TABLEV.   VOWEL PHONEME PATTERN OF UYGHUR LANGUAGE

Vowel	Lip-opened		Lip-rounded	
	Front	Back	Front	Back
High and Half-high	i		y	u
	e		ø	ɔ
Low	ɛ	ɑ		

The patterns of the eight vowel phoneme of Uyghur in the speech system. The effect of the duration of the eight vowels on the duration of the inherent CV type syllable, the last syllable is easily prolonged, regardless of two-syllable words or three three-syllable words [14]. Variance analysis of syllable duration when the inherent CV type syllables contain different vowels gives $F_{(7, 1488)} = 42.39 > Fa$, $P < 0.05$. It is clear that eight vowels have a significant effect on the duration of a syllable. Figure 13 is the duration of CV syllables in different vowels. The CV syllables containing vowel "ɑ" and "ɛ" have longer duration and the syllables that contain "e" vowels has the shortest duration. To sum up, the syllables with a lip-rounded vowel have longer duration and longer than the syllables with the lip-opened vowels. In general, duration of syllable for female are longer than those of male.

Finally, we analyze the CV-type syllables in different vowels and get the result:
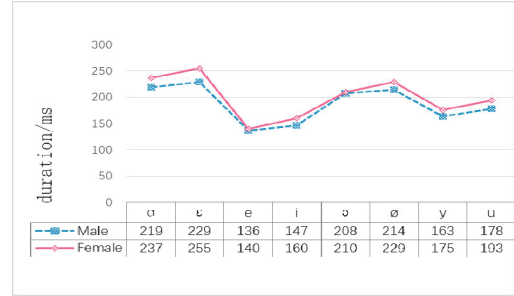
ɛ >ɑ > ø >ɔ > u > y > i > e.

Figure14. Duration of Different Vowel in CV type Syllables

IV. CONCLUSION

Through the statistics and analysis of the overall distribution patterns of syllable duration in Uyghur language, the distribution of monosyllabic words and two-syllable word in male and female is concentrated, and the syllable duration distributions of three-syllable words and four-syllabic words are relatively scattered. Generally speaking. The physiological difference in female vocal cords is shorter than that of male and the intonation of the female has a wide range of changes. male usually flat end, compared to the firm. Therefore, the duration of syllable for female is longer than the male. We can realize gender recognition from duration features.

In this article, we analyzed the duration of syllable when fixed CV syllables are in different positions of words, and extracted the duration of consonants, the duration of vowel and the duration of syllables. We found that the duration of consonant of CV syllable is the longest when it is pronounced alone. In a polysyllabic word, the prefix syllable is the longest, and the shortest syllables are the mid-fix syllable. In a word, duration of monosyllable > prefix > suffix > mid-fix. Duration of vowel is longer when pronounced alone, the longest as suffix syllable, and the shortest as mid-fix syllable in the polysyllabic words. In a word, monosyllable > suffix > prefix > mid-fix. Duration of CV type syllable in separate pronunciation is the longest. In the polysyllabic words the longest duration at the suffix syllable, the shortest duration at the mid-fix syllable. In a word, monosyllable > suffix > prefix > mid-fix. All this confirms that the duration of the syllables are affected by the syllables before and after. Pitch and duration are the main acoustic correlate of word stress [15]. The duration distribution pattern and pitch distribution pattern of the CV type syllables in polysyllabic words can be seen, The CV syllables are long at the suffix, and the pitch is dominant in the prefix and mid-fix. It can be proved that the traditional pronunciation of the "Uyghur word stress generally falls on the last syllable of the word". An analysis of the duration of syllables when CV type syllables contain different vowels reveals that the duration of syllables with low vowels is the longest, and the duration of syllables with high vowels is shorter. This explains that the tongue position can be a reason for the difference in duration of syllable.

REFERENCES

[1] Aynur Nurtay, Zulpiye Aman, Dilmurat Tursun, et al. "Research on vowel patterns of disyllable words in Uyghur language," Computer Engineering & Science, 35(1), 2013, pp. 185–190.

[2] Gulnur Arkin, Askar Hamdulla, Dilmurat Tursun. "Format Features of Vowel Harmonious Word in four-syllabic Uyghur Words." Communication Technology 11, 2013, pp. 47–50.

[3] Askar Hamdulla. "Experimental Phonetics Based Acoustical Analysis of Nasals in Uyghur." Journal of Chinese Information Processing, 2012, 26(1), pp. 110-118.

[4] Zilikam Kasim, Nasirjan Tursun and Wushour Silamu, "Acoustic Analysis of the Initial Syllabic Vowels in Uyghur Language", Journal of Chinese Information Processing, 2009,23(5), pp. 114-118.

[5] Wang Kunlun, Zhang guanhong and Turgunjan Abdukirim, "Analysis of Acoustic Frequency Feature of Uyghur Vowels and Their Identification", Journal of Chinese Information Processing, 24(2), 2010, pp. 122-125.

[6] Zhang Qian. "Phonetic Empirical Study on the Relationship between Syllable Duration and Corresponding Information Carried in Mandarin Chinese", Anhui: Anhui University, 2016.

[7] Feng yongqiang, Chu min, He lin, lv shinan. , "A statistical analysis to syllable duration of Chinese". In Proc. of 5th National Conference on Modern Phonetics, Beijing. 2001.

[8] Wu Zongji, Lin Maocan. Essential of Experimental Phonetics, Peking University Press, 2014, pp. 214.

[9] Zhao Xiangru, Zhu Zhining. The Uyghur Jian Zhi, national press, 1985, pp. 14.

[10] [Lin Tao, Wang Lijia. A Course in Phonetics, Peking University Press, pp. 91, 2013.

[11] Seyyare Imam, Aynur Nurtay, Akbar Pattar, Askar Hamdulla. "The Patterns of Vowels in Monosyllabic Words of Uyghur Language", International Journal of Database Theory and Application, 9(3), 2016, pp. 113-122.

[12] Wang Kunlun. "A Study of Uighur Syllable Speech Recognition and the Base Element of the Recognition", Computer science, 30(7), 2003, pp. 182-184.

[13] Wang Hui, Numemet Yolwas, Wushouer Silamu, "Acoustic Feature Analysis of the Uyghur Phonemes", Journal of Chinese Information Processing, 28(1), 2014, pp. 100-106.

[14] Gulnur Arkin, Askar Hamdulla. "Comparitive Analysis on Acoustic Parameters of Uyghur Word Vowel Harmony". Conference of the Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Technique (Cocosda), Seoul. , 2017, pp. 353-358.

[15] Ranagul Dagut, Askar Hamdulla, Dilmurat Tursun. "Acoustic Analysis on Prosodic Feature of CVC Type Syllable in Uyghur Language", Computer Engineering, 37(9), 2011, pp. 193-195.