PRODUCTION AND PERCEPTION OF LONG AND SHORT VOWEL CONTRAST IN MONGOLIAN

Ao Min[1], Dahu Baiyila[2], Aijun Li[3]

[12]Inner Mongolia University, China
[3]University of Chinese Academy of Social Sciences (UCASS) and IL, CASS, China
aomin@imu.edu.cn dabhvrbayar@163.com liaj@cass.org.cn

ABSTRACT

Long and short vowels are phonologically contrasted in many languages, but their related acoustic features are not always identical. In this study, we aim to explore the encoding and decoding mechanism of long-short vowel contrast in Mongolian. To this end, production and perception experiments were carried out for long-short vowel contrast which were embedded in meaningful and meaningless words.

This study found that, firstly, vowel duration is the most important acoustic production parameter, which has a relatively stable duration ratio pattern between long and short vowels. Secondly, the perception on a physical continuum of duration for meaningful and meaningless words indicates that there is a clear categorical boundary, while long vowels have a broader perceptual space than that of short vowels. Finally, there is a strong complementary relationship between word meaning and vowel duration in the process of long-short vowel encoding and decoding.

Keywords: Mongolian, long-short vowel contrast, acoustic feature, categorical perception

1. INTRODUCTION

1.1. Long-short vowel contrast

Long-short vowel contrast is ubiquitous in the vowel systems of world ' s languages, such as Uyghur, English, Korean, Czech, Finnish, Swedish, Arabic, Danish, Estonian and Ewe. Studies show that the difference between long and short vowels lies in duration, F0, formant, accented or not, adjacent consonants, and prosodic boundary types, among which duration is the most stable distinguishing feature [1]. Although duration is positively correlated with change of F0 for both long and short vowels in Swedish [2], there is a relatively stable duration ratio of 2.4: 1 between them in Japanese [3].

On the other hand, duration of long and short vowels change under individual conditions. The adjacent consonant and syllable structure will influence the acoustic duration pattern between long and short vowel [4]. However, few studies have been done on the variation patterns of vowel duration between long-short vowel contrast in continued speech.

1.2. Long-short vowel contrast in Mongolian

There are stable long-short vowel contrast in Mongolian, one of the Mongolian languages in Altaic language family [5,6]. However, the pairs of long and short vowels and specific vowels are differ in Mongolian dialects. There are 8 pairs of long-short vowel contrast in standard Mongolian [7], while there are 11 pairs in Khalkha dialect [8]. For example, the duration contrast of vowel /ɐ/ distinguishes the meanings between the words "ab" (/ɐw/, means " Take ") and "abu" (/ɐ:w/, means " Father "). Figure 1 is the spectrogram, where the vowels of both words have basically the same formant pattern but remarkably different durations.

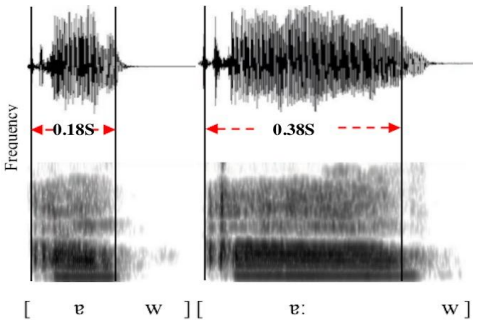


Figure 1: The spectrogram of the words /ɐw/ (left side, means " Take") and /ɐ:w/ (right side, means"Father") .

Long-short vowel contrast in Mongolian is widely studied in both segmental and suprasegmental aspects. For example, it is found that long and short vowels in Mongolian are different not only in duration, but also in tongue position [9]. According to the length of duration, vowels can be divided into ultra-long vowels, long vowels, short vowels and ultra-short vowels [10]. The prolongation of non-syllabic vowel has prosodic function in Mongolian [11]. However, there are many problems to be further discussed about the duration of long-short vowel contrast in Mongolian.

To this end, this study investigates the duration patterns and the encoding and decoding mechanism

of long-short vowel contrast in Mongolian under the regulation of word meaning.

2. METHOD

2.1. Corpus design

The speech data of this study includes seven short vowels in Mongolian, namely /ɐ/, /ə/, /i/, /ɔ/, /ʊ/, /o/ and /u/, and their corresponding long counterparts /ɐː/, /əː/, /iː/, /ɔː/, /ʊː/, /oː/ and /uː/. Two kinds of materials were designed based on monosyllabic words in fixed context, including 50 pairs of meaningful words (common words) and 116 pairs of meaningless words, with long-short contrast vowels in the same context. The meaningful words include 15 pairs of VC and 35 pairs of CVC, such as el_e(/əl/, each) and egeli(/əːl/, affectionate), and the meaningless words include 84 pairs of CVC and 32 pairs of VC obtained by the combination of /m/, /s/, /ʧʰ/ and /p/ with seven pairs of contrasted vowels respectively, such as /sɐs/ ~ /sɐːs/.

2.2. Production and perception experiments

Six speakers (gender balanced) participated in the production experiment. Another twenty listeners participated in the perception experiment. All the speakers, aged 20 to 30, were students with the Class I Type A certificate of standard Mongolian test. They didn't live in other dialect areas before going to senior high school. Speakers read each word which randomly presented on the screen three times at a normal speed. Finally 5,976 speech sounds were collected.

The perception experiment was conducted on duration continuum manipulated from 30 meaningful and 28 meaningless words with long vowel in CVC context by linear test method [12]. Altogether 10 stimulus sounds were synthesized for each vowel contrast. finally 920 stimuli were made in Praat and 11,600 perceptual results were obtained.

2.3. Tools for data collection and processing

Three kinds of experimental tools were used in the data processing: xSpeech tools [13] for recording, annotation and perception experiments; ProsodyPro [14] for extracting vowel duration. Formulas (1) for calculating the prolongation rate of vowel duration and formulas (2) for calculating the multiple of duration between long and short vowels are shown below [15].

$$(1) \quad VDxv = \frac{VD_{LV} - VD_{SV}}{VD_{LV-SV}} \times 100\%$$

$$(2) \quad K_V = \frac{VD_{LV}}{VDsv}$$

In Formula (1), VD_{XV} is the prolongation rate of the average vowel duration for a certain long-short vowel contrast (xv), VD_{LV} and VD_{SV} are the average duration of long-short vowel contrast (xv) respectively, and $VD_{LV\text{-}SV}$ represents the difference between the average duration of all long and short vowels. In Formula (2), K_V is the multiple of duration between long and short contrasted target vowels.

3. RESULTS

3.1. Acoustic features of vowel production

The result of acoustic analysis are shown in Table 1, and the values in brackets are meaningful monosyllabic words, while the values before the brackets are meaningless monosyllabic words. Each value is an average from three repeated readings by six speakers.The IF, PR and PD represent for influence factors, prolongation rate and prolongation direction,respectively.

No	IF	Type	Mean duration	PR	PD
1	Lip shape	L	Unround: 0.227 (0.286) Round: 0.231 (0.261)	1.5% (10%)	+(+)
		S	Unround: 0.123 (0.116) Round: 0.123 (0.104)	0.2% (11%)	+(+)
2	Tongue high and low dimensions	L	High: 0.232 (0.657) Medium: 0.227 (0.283) Low: 0.226 (0.305)	1.5% (6.5%)	+(+)
		S	High: 0.123 (0.657) Medium: 0.123 (0.122) Low: 0.123 (0.137)	0.2% (11%)	+(+)
3	Tongue front and low dimensions	L	Front: 0.228 Middle : 0.227 Back: 0.230	0.80%	+(+)
		S	Front: 0.123 Middle: 0.2123 Back: 0.123	0.20%	+(+)
4	Adjacent consonants	L	Unvoiced: 0.268 (0.238) Voiced: 0.227 (0.265)	18%	+(-)
		S	Unvoiced: 0.151 (0.078) Voiced: 0.097 (0.0965)	16%	-(+)
5	Syllable features	L	VC: 0.301 (0.301) CVC: 0.252 (0.251)	15% (6.5%)	-(-)
		S	VC: 0.146 (0.147) CVC: 0.087 (0.087)	16% (11%)	-(-)

Table 1: The influence degree and statistical significance of each factor on vowel durations.

The results indicate that adjacent consonants, syllable length and word meaning have a statistically significant influence on vowel duration in the processing of acoustic production although the duration is affected by multiple factors in varying degrees. First, the influence of vowel inherent features on vowel duration is not statistically significant (P>0.05, df=426, F=0.004). Although there are differences in the actual vowel duration in the classification of vowel inherent features, the dispersion is small, and the prolongation rate is between 0.2% and 1.5%, there is no significant difference. Second, syllable features are significant to vowel duration. The syllable length is inversely proportional to the vowel duration, and the duration shortening of short vowels (40%) is larger than that of long vowels (16%). Similar studies showed that [16], the vowel duration gradually decreases with the gradual increase of phonemes in syllables. Third, the influence of adjacent consonants on vowel duration is complicated. Although, vowel duration is shortened by unvoiced consonants it is prolonged by voiced consonants, which proved to be a cross-language phonetic feature [17].

Duration is the distinctive feature in the acoustic production of long-short vowel contrast in Mongolian. Figure 2 shows the duration distribution of 7 long-short vowel contrast according to their categories, where black represents short vowels and gray represents long vowels.

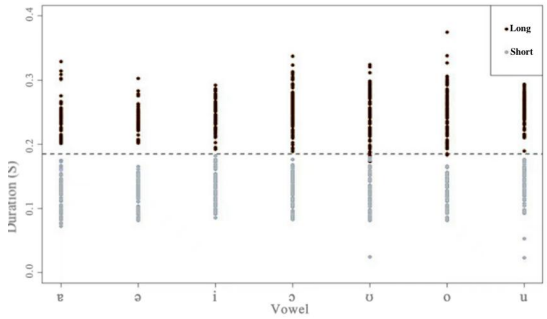
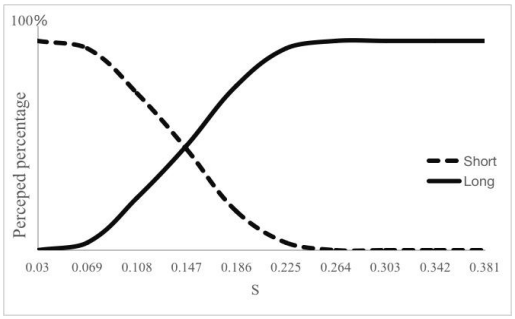


Figure 2: Duration contributions of 7 long and short vowel contrasts.

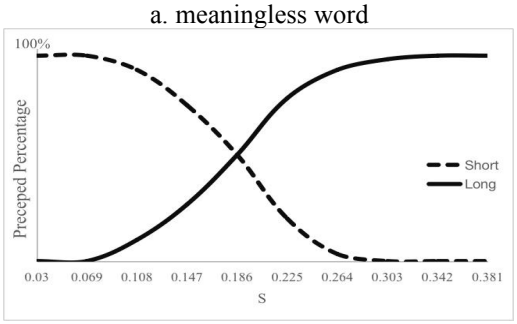
The physical duration of long and short vowels as a whole shows a remarkable boundary at 0.18s as indicated by black dotted line in the Figure. 95% of vowel duration is distributed around this boundary. Short vowels are shorter than 0.18s and long vowels are longer than 0.18s. As far as all the data are concerned, the standard deviation of duration among seven short vowels is 0.038s, and that among seven long vowels is 0.033s, and the difference between short and long vowels is 0.069s, which is much larger than the internal standard deviation of short vowels or long vowels. That is, on the duration dimension, the distribution between short and long vowels is scattered, while the internal distribution is concentrated. The duration difference between long and short vowels is statistically significant (P<0.01, df=2987, F=0.02).

3.2. Perceptual features

The identification results for continuum perception experiment are plotted in Figure 3. The total number of stimulus sounds in the identification task for meaningless words (Figure 3a) is 5,600, while that for meaningful words (Figure 3b) is 6,000.



a. meaningless word



b. meaningful words

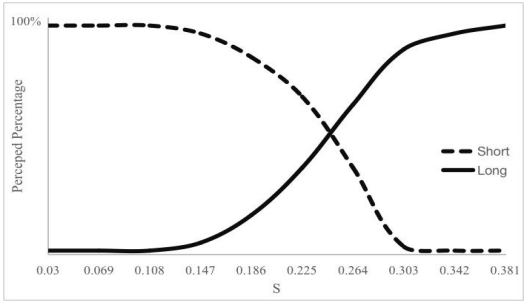
Figure 3: Perception results of long and short vowels.

Based on the binomial distribution of the identification scores and the sigmoid shape of the response function, the results show that:

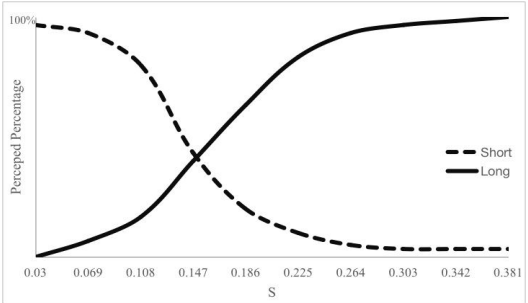
First of all, the perception boundary of long-short vowel contrast in Mongolian is 5/10 for meaningful words (see Figure 3b). When the duration of long vowels is shortened by half, it has a very obvious influence on vowel category. Most listeners can not identify them as a long vowel when the duration is shorter than half. Thus, duration is the primary phonetic feature for the classification of long-short vowel contrast in Mongolian at phonological level. Listeners basically identify long or short vowels by vowel duration.

Second, the perception space of long vowels is broader than that of short vowels. The experimental results show that the auditory space of long vowels covers all the duration segments after stimulus 2/10, while the space of short vowels ranges from stimulus 1/10 to 7/10. For meaningless words, the space of short vowels ranges from 1/10 to 6/10, and that of long vowels spans from 2/10 to 10/10.

The production and perception experiments of long-short vowel contrast prove that duration is the most important acoustic feature between long vowels and short vowels, but it is not the only one. The experimental results presented in Table 1 and Figure 2 fully illustrate that word meaning is a critical factor. Based on this finding, all experimental words, including 30 meaningful words and 28 meaningless words, are divided into long vowel words with meaning, whose short counterparts had no meaning, and short vowel words with meaning, whose long counterparts had no meaning, so as to re-examine the actual influence of word meaning on vowel type and duration pattern. For example, the /mus/ (ice) is meaningful, but the /mu:s/ is meaningless; the /mɐm/ is meaningless, but the /mɐ:m/ (Lama) is meaningful. The identification results in Figure 4 illustrates there is a complementary relationship between word meaning and vowel duration in decoding or perceiving vowel category. When experimental words are meaningless, listeners perceive long vowels or short vowels depending on vowel duration. However, when experimental words are meaningful, listeners first identify the vowel category based on the phonological and phonetic knowledge of the mother tongue. This is proved by the fact that meaningful short vowel words move the perceptual boundary in Figure 4a (1/10-6.5/10) more rightward than that of the meaningless words in Fig. 4b (1/10- 4/10); meaningful long vowels move the boundary more leftward in Figure 4b (4/10-10/10) than that of the meaningless words in Figure 4a (6.5/10-10/10) .



a. short (meaningful)-long (meaningless)



b. short (meaningless) - long (meaningful)

Figure 4: *P*erceptual results of vowel types in meaningful meaningless pairs.

The influence of word meaning on vowel duration is polarized. On the one hand, word meaning will shorten the acoustic duration of vowels. When the speech data are classified by syllable features, the influence of vowel duration in meaningless words is greater than that in meaningful words. As you can see in Table 1, the influence rate of long vowels is 15% (6.5% for meaningful words) and that of short vowels is 16% (11% for meaningful words). On the other hand, word meaning expands the perceptional space of short vowels and moves the perceptual boundary rightward as mentioned above which was shown in Figure 3. The identification curve of long and short vowels of meaningful words is less sharpen than that of meaningless words. Therefore, word meaning is an important factor besides duration in the identification of vowel category. For meaningful words, listeners identify the vowel comprehensively by duration and word meaning; or, they identify only by duration for meaningless words.

According to the duration data of this study, duration pattern of long-short vowel contrast in Mongolian is 2:1. that is, long vowels are twice as long as short vowels. Acoustically, although vowels differ in the proportion of duration, a comprehensive survey shows that long vowels are twice as long as short vowels. In term of perception boundary, the dividing point of duration between long-short vowel contrast is at 1/2, but it moves slightly when word meaning is involved. When vowel duration is less than or greater than 1/2 of the long vowel, the vowel type changes in phonology.

4. CONCLUSION

This study analyzed the acoustic production and perception category of Mongolian long-short vowel contrast. It was preliminarily found that duration is the most significant distinguishing feature between long vowels and short vowels in acoustic production. In addition, it also found the complementary mechanism between word meaning and vowel duration in the process of perception and identification of long and short vowels. When experimental words are meaningless, listeners perceive long vowels or short vowels depending on vowel duration. However, when experimental words are meaningful, listeners first identify the vowel category based on the phonological and phonetic knowledge of the mother tongue.

5. ACKNOWLEDGMENTS

This work is supported by national social science fund "The Construction of Mongolian Speech Database and the In-depth Study of Intonation

Related Issues" (22XYY003), Key projects of National Social Science Fund "Quantitative Syntactic Research Based on the Mongolian Annotated Corpus' (19AYY018) and, Project of Cultural Experts and "Four Batches" of Talents awarded to Aijun LI .

6. REFERENCES

[1] An acoustic study of the Japanese short and long vowel distinction. Master's thesis, Lawrence: University of Kansas, 2008.

[2] Lyberg. B, "Some observations on segment duration and its interaction with the fundamental frequency contour in Swedish," J. Acoust. Soc. Am, 77, S54, 1985.

[3] Kozasa.T, An acoustic and perceptual investigation of long vowels in Japanese and Phonpeian. PhD dissertation, Honolulu: University of Hawai'I at Manoa, 2005.

[4] Myers.S, "Vowel duration and neutralization of vowel length contrasts in Kinyarwanda," Journal of Phonetics, vol. 33: 2005, pp. 427-446.

[5] Svantesson. J. O. , Tsendina. A. , & Karlsson. A, Phonology of Mongolian, Oxford: Oxford university press, 2013, pp. 183-185.

[6] Huhe, Qoijungjab, Acoustic Analysis on the Mongolian Phonemes, Hohhot: Inner Mongolia University Press, 1999, pp. 131–153.

[7] Chinggeltei. Mongolian Grammer, Hohhot: Inner Mongolia People's Publishing House. 1991, pp. 11-21.

[8] Batdvrj. Acoustic Analysis of Khalha Dialect in Mongolian. Ph.D dissertation, Hohhot: Inner Mongolia university, 2014.

[9] Huhe, "Long and Short Vowels in Mongolian: A Multivariate Analysis of Variance Approach," Minority Languages of China.vol. 5, 2021, pp. 101-107.

[10]Yurong, Analysis of Vowel Duration based on the Mongolian Spoken Corpus, Journal of Inner Mongolia University, vol, 6, 2012, pp. 1-9.

[11]Aomin, "The Prosodic value of Mongolian non-Syllabic vowel," Applied Linguistics, vol. 1, 2018, pp. 108-115.

[12]F.Shi, Auditory Patterns: A Preliminary Study on the Characteristics of Chinese Speech Perception, Beijing: Commercial Press, 2019.

[13]Ziyu Xiong, "xSegmenter: Segment automatic segmentation and annotation tools," Chinese journal of phonetics, vol. 1, 2009, pp. 27-34.

[14] Xu Y, ProsodyPro - A tool for large-scale systematic prosody analysis, Tools and Resources for the Analysis of Speech Prosody, 2013.

[15]LI Bin. REN Rui-chao, "Mathematic Analysis of English Monophthongs' Duration and its Modeling," Mathematics in Practice and Theory, vol. 46, 2016, pp. 158-166.

[16]Xueguang, Aomin, "Statistical analysis of Mongolian standard phonetic vowel duration," Mongolian Language, vol. 7, 2019, pp. 47-51.

[17]Mcallister.R. , Flege.J.E. , & Piske.T, "The Acquisition of Swedish Long vs. Short Vowel Contrasts by Native Speakers of English, Spanish and Estonian," Applied Psycholinguistics, vol. 22, 2001, pp. 283-299. https://scripts.sil.org/cms/scripts/page.php?item_id=IPAhome.