

The Role of Fo in Mongolian Stress

Yumei Sang¹, Philippe Martin¹

¹CLILLAC-ARP, EA 3967, UFR Linguistique
Université Paris Diderot Sorbonne Paris Cité

yumei.sang@linguist.jussieu.fr, philippe.martin@linguist.jussieu.fr

Abstract

This paper deals with vowel duration and Fo correlates of word stress in Halh Mongolian, the main dialect spoken in Mongolia. The position of word stress is predictable. The fundamental frequency (Fo) of stressed vowels shows a regular pattern compared to the Fo of unstressed vowels. Maximal and minimal values of Fo in the word are both used to indicate syllable prominence, as does longer vowel duration.

Index Terms: Fundamental frequency, Mongolian, Stress

1. Introduction

This paper deals with the word stress system and acoustic correlates in Halh Mongolian, the main dialect spoken in Mongolia. There are relatively few studies on stress in Mongolian and its position in the word remains still unsolved questions due to lack of consensus among linguists. In this paper we will analyze the stress for two word categories: words containing only short vowels and words containing at least one long vowel.

2. Mongolian Stress

2.1. Previous Studies

Stress and its place of realization in Halh Mongolian have been debated for a longtime and is still an unsolved issue until today. The only consensus about Mongolian stress is that the stress is not phonemic in the sense that it does not differentiate between different meanings. Previous studies proposed widely different interpretations about Mongolian stress and its place of realization. Different acoustics parameters such as Fo, intensity, duration, vowel quantity and quality, syllable weight etc. have been described as the main correlates in signaling stress in Mongolian. Two units, syllable and mora have been proposed for bearing the stress.

According to a point of view more traditional, and to most native speakers scholars [1], Mongolian has the stress on the first syllable.

Some western scholars, among which Poppe [2] assume that words containing no long vowels or diphthongs, have the stress on the initial syllable, whereas words containing long vowels or diphthongs, have the stress on the penultimate long vowel or diphthong.

Walker [3], claims that word stress falls on the rightmost heavy non final syllable; if there are no heavy syllables, stress falls on the first syllable. The syllable weight is considered as language specific: for some languages, both CVV and CVC are considered as heavy syllable. In Halh Mongolian, only CVV is considered as heavy syllable. VV here means a long vowel or a diphthong.

Harnud [4] in his study performed on di- and trisyllabic

words on Chakhar, one of the Mongolian dialects spoken in Inner Mongolia in China, shows that in the words with first short vowel, stress falls on the second syllable. While for the words with first long vowel, the author obtained contrasting results: the acoustic intensity seem to indicate that the first syllable is stressed whereas Fo features seem to indicate that the second syllable is stressed.

Svantesson [5, 6] and Karlsson [7] advanced a less common mora counting, based on bitonal LH (low-high) gesture analysis. The H tone is aligned with the first mora and the L tone is aligned with the second mora. A short vowel is counted as one mora and a long vowel is counted as two morae. According to this theory, long vowel occurs only in word initial syllable.

2.2. Data Analysis

2.2.1. Method

In our experiments, the data were collected in Ulaanbaatar, Mongolia in early 2011. Four 27-30 year-old Mongolian native speakers from Ulaanbaatar participated in this data collection. The recordings took place in a quiet room isolated from external noise. Recording used a high-quality digital recorder *Zoom H2*. In total, 328 target words of two and three syllables were recorded. We did the recordings in two times, so why one of our four speakers read a slightly different corpus. The other three speakers read the same corpus than the first recordings, but with some improvements. The target words were borne by frame-sentences such as 'Bi xxx gej helsen' ('I said xxx'). The data were analyzed using the WinPitch speech processing software.

In our acoustic analysis, target words were divided into two categories. The first category only contains words with short vowels whereas the second contains words with at least one long vowel. Short vowels are here referred to by the character V and long vowels referred to by VV (by writing twice the short vowel). The target words of the first category have structures such as V-V and V-V-V and the words of the second category have structures such as V-VV, VV-VV, VV-V, V-VV-V, V-V-VV, VV-VV-V, V-VV-VV and VV-VV-VV. The analysis of the target words is done in three steps: we start by investigating the Fo value, we then measure the vowel duration and observe the eventual absence of Fo movement within the vowel. The absence of Fo movement within the vowel is here referred by the term 'Fo absence'. That means in some cases where the unstressed vowels are reduced and barely audible, their Fo is not realized. In other words, some unstressed reduced vowels are devoiced.

2.2.2. F0 Value

Figure 1 shows the Fo value of short-vowel words for four speakers. In total 48 V-V structure target words and 31 V-V-

V structure target words were analyzed. Sp1, Sp2, Sp3 and Sp4 mean Speaker1, Speaker2, Speaker3 and Speaker4 respectively. Each line represents the Fo value for one speaker. We can see that for the four speakers, the Fo value of the first vowel is consistently lower than the Fo value of the second or the third vowel. There is no regularity for the Fo value of the second and the third vowel. The Fo of the second vowel is sometimes higher (Sp2, Sp3, Sp4) than the Fo value of the third vowel or sometimes lower (Sp1).

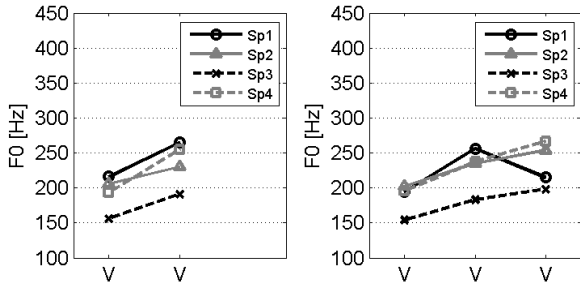


Figure 1: Average Fo value of the target words containing only short vowels for four speakers (Sp1, Sp2, Sp3 and Sp4).

Figure 2 shows the Fo value of the words containing at least one long vowel. In total 122 V-VV structure target words, 16 VV-VV structure target words, 1 VV-V structure target word of one speaker, 59 V-VV-V structure target words, 23 V-V-VV structure target words, 7 VV-VV-V structure target words, 12 V-VV-V structure target words and 3 VV-VV-VV structure target words are analyzed. We can see the leftmost syllable containing a long vowel always features the highest Fo value. In a very few cases only (words with structure VV-V for one speaker only and words with structure VV-VV for one speaker only) the Fo value of right syllable is 5Hz higher than the Fo value of the leftmost syllable containing a long vowel. However these very few cases are not significant since the difference between the Fo value of the two syllables never exceeds 5 Hz.

2.2.3. F0 Absence

Several cases of Fo absence were observed in this analysis; even for these target words carefully pronounced. Table 1 shows the percentage of the Fo non-realization in function of word structure for the words containing only short vowels. The total number of analyzed target words for each word structure is specified. Syll-1, Syll-2 and Syll-3 mean first syllable, second syllable and third syllable respectively.

W/structure	Nb/word	Syll-1	Syll-2	syll-3
V-V	23	4%	17%	
V-V-V	31	0%	13%	13%

Table 1: Percentage of Fo absence in words with only short vowels.

In the V-V structure words, the Fo of the vowel of the first syllable is almost always realized, except 1 case in 23 cases, or 4% of Fo absence. The Fo of the second vowel is not realized in the 17% of cases. In the V-V-V structure words, The Fo of the vowel of the first syllable is realized in the 100% of cases. The Fo of the vowel of the second and third syllable both is not realized in the 13% of cases.

Table 2 shows the percentage of Fo non-realization of the words containing at least one long vowel. The same the total number of analyzed target words for each word structure is specified. Syll-1, Syll-2 and Syll-3 mean first syllable, second syllable and third syllable respectively.

W/structure	Nb/word	Syll-1	Syll-2	syll-3
V-VV	90	9%	0%	
V-VV-V	59	5%	0%	29%
V-V-VV	23	13%	17%	
V-VV-VV	2	100%	0%	0%
VV-VV	16	0%	0%	
VV-VV-V	7	0%	0%	29%
VV-VV-VV	7	0%	0%	0%

Table 2: Percentage of Fo absence in the words containing at least one long vowel.

In table 2, we can see the Fo of the long vowel is always realized. The Fo of short vowels shows some interesting features: when there is a long vowel in a word, the Fo of the short vowel in this words is not realized in some cases; even this short vowel is in the initial syllable. In the V-VV, VV-VV, V-V-VV and V-VV-VV structure words, the percentages of Fo absence of the first short vowels are 9%, 5%, 13% and 100% respectively.

2.2.4. Vowel Duration

Figure 3 shows the vowel duration of words with only short vowels. In total, 48 V-V structure target words and 31 V-V-V structure target words are measured. Sp1, Sp2, Sp3 and Sp4 mean Speaker1, Speaker2, Speaker3 and Speaker4 respectively. Each line represents the average vowel duration for one speaker and the average vowel length is showed per syllable. The duration of the first vowel in figure 3 is always longer than the duration of other short vowels. This is observed in both types of V-V and V-V-V structure words. The vowel in initial syllable is in average 1.5 times longer than the vowels in non-initial syllables.

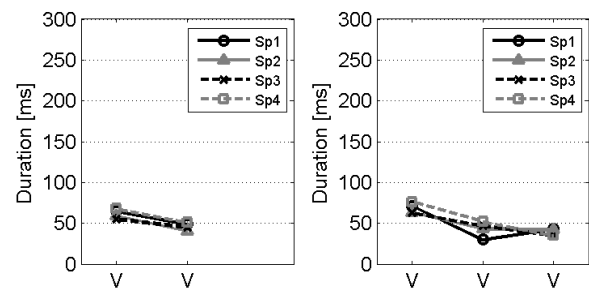


Figure 3: Average vowel duration per syllable for the words containing only shorts vowels for four speakers (Sp1, Sp2, Sp3 and Sp4).

Figure 4 shows the words containing at least one long vowel. In total 122 V-VV structure target words, 16 VV-VV structure target words, 1 VV-V structure target word of one speaker, 59 V-VV-V structure target words, 23 V-V-VV structure target words, 7 VV-VV-V structure target words, 12 V-VV-V structure target words and 3 VV-VV-VV structure target words are analyzed. We see that a long vowel is much longer than a short vowel. In average, the duration of a long vowel is

2.4 times longer than the duration of a short vowel. When there is more than one long vowel in a word, the result of our four speakers shows that the leftmost long vowel is in average 1.3 times longer than the other long vowels.

3. Discussion

In section 2.2, we showed that for the words with only short vowels, the Fo value of the first vowel is regular in the sense it always shows the minimal Fo value. However, the second and the third vowel do not show any regularities of Fo value. We also saw that in the words with short vowels only, the Fo of the first syllable is always realized whereas it is not always the case for the second and the third syllable. The vowel duration of the first syllable is also always the longest (1.5 times longer than the duration of vowels in non-initial syllables). The second short vowel is sometimes longer than the third short vowel or sometime shorter. To summarize, in the words with short vowels only, the first vowel has always the most regular Fo value, its Fo is always realized and is always the longest vowel of the word. The Fo value therefore seems to indicate the prominence of a particular vowel within a word and this vowel reaches its maximal length under the prominence. It shows that Fo is a relevant parameter for signaling stress in the words with short vowel only in Mongolian. This means that stress falls on the first vowel of the words with short vowels only. This result is in agreement with some previous studies [2, 3] claiming that Mongolian have the stress on the first syllable. However our result only applies to words with short vowels only.

For the words with at least one long vowel, according to the result in the section 2.2, when there is only one long vowel in a word, the Fo value of this long vowel is always higher than the Fo value of the short vowel. The duration of this long vowel is always longer (2.4 times) than the short vowel. The Fo of this long vowel is always realized. In the words with more than one long vowel, the Fo of the leftmost long vowel is regular in the sense it always shows the maximal Fo value. However the Fo value of the other long vowels do not show any regularities. The duration of the leftmost long vowel is also the longest, in average 1.3 times longer than the duration of the second long vowel and the third long vowel. In other words, in the words with at least one long vowel, the leftmost long vowel has always the highest Fo value, the longest vowel duration and its Fo is always realized. The Fo value therefore seems to signal the prominence on the leftmost long vowel in a word and this left most long vowel reaches its maximal duration under the prominence. It means that Fo is as well a relevant parameter for signaling the stress in the words with at least one long vowel in Mongolian. It shows that stress falls on the leftmost long vowel in the words with at least one long vowel.

Our result on vowel duration is in agreement with the study performed by Harnud [4]: a long vowel, what ever its position in a word, is always longer than a short vowel. Our result on the vowel duration does not support the theories saying that in Mongolian, long vowels appear only in the first syllable of the word [5, 7].

4. Conclusion

Regular and consistent behaviors of Fo in the stress signaling are observed in Mongolian. Fo maximal and minimal values are both used for signaling prominence at the word level. Vowels reach their maximal duration under the stress. The Fo value and vowel duration together signal stress. The stress can be

signaled by either a rising gesture or a falling gesture of Fo. The stressed vowel always features the most stable Fo value, has always F0 realized and is always the longest vowel. Based on these parameters of Fo value and vowel duration, we conclude that:

1. *Stress falls on the first syllable for the words with only short vowels.*
2. *Stress falls on the leftmost syllable containing a long vowel.*
3. *The acoustic parameter of stress results from a combination of longer vowel duration and rise or fall of Fo within the stressed syllable.*

In this study, we only studied words with two and three syllables. It would be useful to extend this study to words with more than three syllables or to words with diphthongs. This would allow a better understanding of stress in Mongolian.

5. Acknowledgements

We would like to thank our four speakers, Delgercuren Sanjaa, Enhmaa Chuluunbat, Monhceceg Batsuh and Onorbayr Bayraa for contributing to this study by participating in our recordings. We thank them for their time and patience during the recording process.

6. References

- [1] S. Luvsanvandan, *Orchin tsagiin Mongol helnii butets (1)*. M&D, 1968.
- [2] N. Poppe, *Mongolian Language Handbook*. Center for Applied Linguistics, Washington, D.C, 1970.
- [3] R. Walker, "Mongolian stress, licensing, and factorial typology," *Ms., University of California, Santa Cruz. ROA-171, Rutgers Optimality Archive*, 1997.
- [4] H. Harnud, *A Basic Study of Mongolian Prosody*. Publications of the Departement of Phonetics, University of Helsinki, 45. Helsinki, 2003.
- [5] J.-O. Svantesson, A. Tsendina, A. Karlsson, and V. Franzén, *The Phonology of Mongolian*. Oxford University Press, 2005.
- [6] J.-O. Svantesson, "Phonetic correlates of stress in mongolian," *Proceedings of the International Conference on Spoken Language Processing 1*, 617–620. Kobe, Japan, 1990.
- [7] A. Karlsson, *Rhythm and Intonation in Halh Mongolian*. Lund University, 2005.

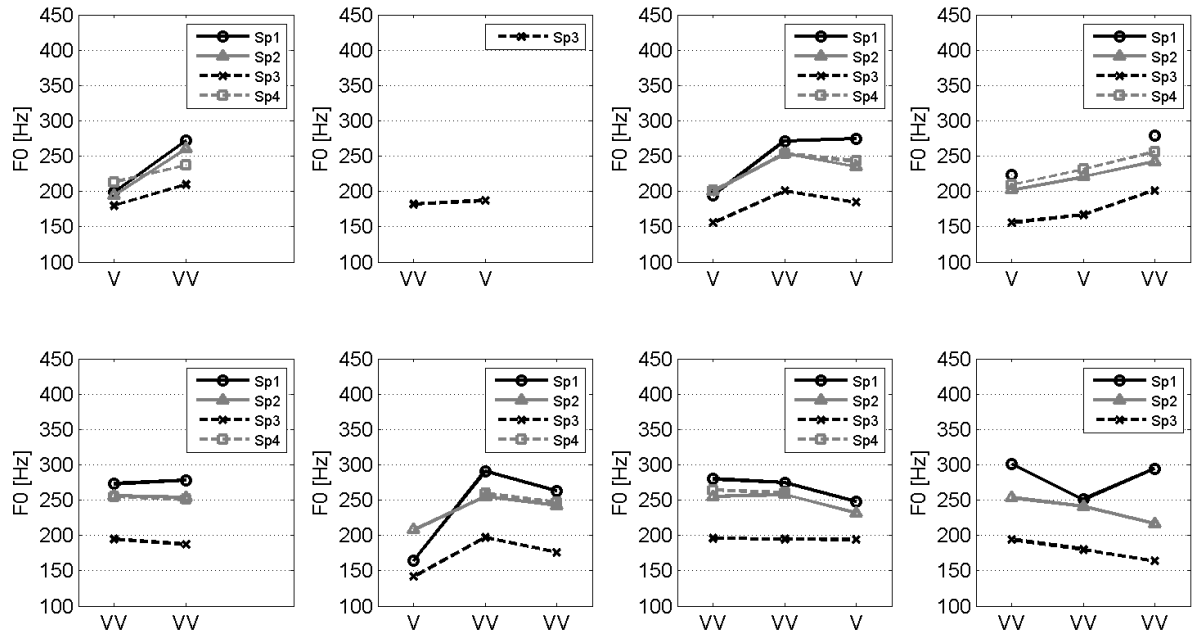


Figure 2: Average *Fo* value of the target words containing at least one long vowel for four speakers (Sp1, Sp2, Sp3 and Sp4).

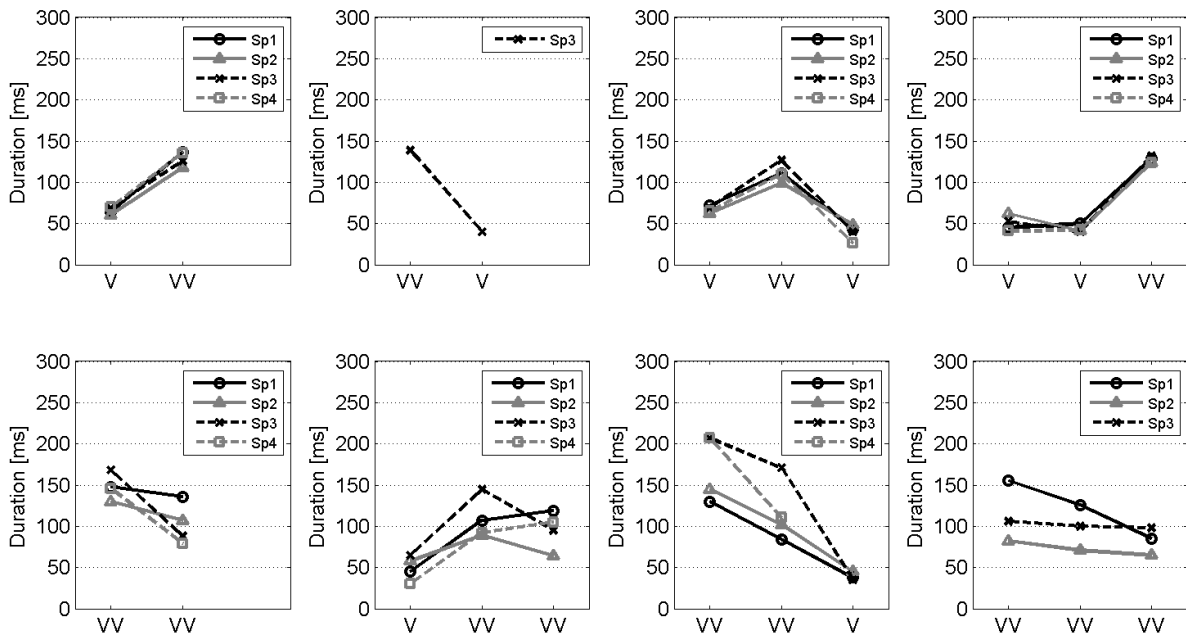


Figure 4: Average vowel duration per syllable for the words containing at least one long vowel for four speakers (Sp1, Sp2, Sp3 and Sp4).