Stress Gravity of Neutral Tone Words in Different Information Structures

Jingwen Huang[1], Aijun Li[1,2]

[1] Institute of Linguistics, Chinese Academy of Social Sciences
[2] University of Chinese Academy of Social Sciences
huangjw@cass.org.cn, liaj@cass.org.cn

Abstract

Although the issue whether Chinese has word stress is still controversial, there is a neutralization phenomenon of tone sandhi in Chinese for unstressed syllable. Some of neutral tone words even have counterpart, such as /tɔŋ1 ɕi0/ (things/a thing/the thing, 0 for neutral tone) and /tɔŋ1 ɕi1/(oriental and western / east and west). In this study, we studied the effect of sentence level stress on neutral tone by comparing the acoustic patterns of such kind of counterparts in two information structures, i.e. broad focus (BF) and narrow focus (NF). In addition to the popular prosodic features of f0, duration and intensity, a comprehensive feature called Stress Gravity was adopted, which is an efficient and robust parameter for measuring stress. The results showed that neutral tone syllable got the highest increment in Stress Gravity when the neutral tone word is focused, where the pitch-range expansion of neutral syllable contributes the most. Besides, the focused neutral tone word is realized by raising the pitch of its full syllable, and increasing the difference in duration and intensity between the full syllable and the following neutral tone syllable.

Index Terms: neutral tone, focus, stress, Stress Gravity

1. Introduction

In mandarin, a full syllable is stressed and carries one of four lexical tones in Mandarin, and the syllable can lose the original tonal categories and become a neutral tone under certain conditions [29].

Previous studies on neutral tone in Mandarin have conducted from a wide range of perspectives, in both phonology and phonetics, including traditional descriptions based on auditory sense and explorations by experiments [3, 20, 2, 25, 16, 17, 15]. These studies cover various aspects of syntactic, phonological, acoustic and perceptual properties about neutral tone.

From the acoustic aspect, the neutral tone exhibits the obvious sense of sounds in terms of the compression of pitch range and the shortening of duration, then the weakening of intensity. It is generally assumed that syllables in neutral tone do not have independent pitch values, and their phonetic pitch is determined by the preceding lexical tone [4, 5, 22, 24]. The pitch curve of neutral tone following Tone1(H), Tone2 (LH) and Tone4 (HL) performs as a drop in tone, as a flat or rising tone when after Tone 3 (LL(H)). Also, the two types of pitch levels show that the pitch is high after Tone3 and low after the other tones. The average duration of the neutral-tone syllable is 45% [20, 23, 14] or 60% [2, 6] of the corresponding duration of lexical tone syllable. However, as for intensity, it is believed not an essential feature of neutral tone [23, 2, 21].

The scope and material on neutral tone in recent research has expanded. Li examined the realization of disyllabic neutral tone words in more complex information structures [15], while Chen & Xu and Li & Li investigated the properties of neutral tone in multiple syllables [6, 18]. Huang & Shi and Huang compared various types of neutral-tone syllables such as *de* (的, auxiliary particle), *shang* (上, localizer), repetition of nouns (e.g. "爸爸(/pa4 pa0/father)") in different information structures [12, 13].

We have understood the characteristics of neutral tone at the lexical level, but far from enough at the sentence level. Taking information structure as the starting point, the focus is the means to convey new information [10]. Bolinger put forward the view of relative semantic weight, that the prominence of speech and emotion determines the stress of sentences [7]. Gussenhoven proposed the Focus-to-Accent theory, using pitch accents to mark the most phonetically prominent parts of a sentence (i.e., focus) [9].

In the research on focus and sentence stress in Mandarin, the opinions basically coincided, that the stress expands the pitch range and duration of the focused syllable, then increases the intensity and compresses the pitch range of syllable after focus [5, 29, 31, 37]. A well-known view of post-focus compression (PFC) was proposed by Xu et al. [34], which was validated in many languages or dialects in China [35, 33].

Yin investigated the relativity between various acoustic features and stress when exploring rhythm, finding that the combination of the product of pitch and duration is the highest correlation with stress among all the variables [38]. We adopted the method of combining the product of pitch-range and duration to analyze the stress relationship between neutral tone stressed and unstressed, the product named Stress Gravity.

Now there are many existing studies on neutral tone and focus, but relatively few have examined the combination of the two. Thus, in this study, we seeks to explore how does the stress on sentence level affect neutral-tone syllable in different information structures in terms of prosodic features as well as a novel feature called Stress Gravity.

2. Method

2.1 Experimental design and material

Referring to [26] and [27], the disyllabic words with neutral tone (SN) and their counterparts (S1S2) in normal stress were selected from [39].

Considering the four citation tone (T1~T4), we selected 16 tone combinations (4×4) in S1S2. Also, there were 16 words in SN as counterparts, 32 target words obtained in all. Then, based on [37], we put the target words in the middle of the declarative sentences and produced them in two focus contexts, i.e. BF and NF, as listed in Table 1. 64 sentences were obtained for the

979-8-3503-9796-3/22/$31.00 ©2022 IEEE

recording, with half carrying target words S1S2 and another half SN.

Table1: *Target words in BF and NF*

Contexts	Target words	Eliciting sentences	Sentences with target words
BF	(SN) 东・西	你说什么？ What are you saying?	大陆 东・西 充足。 On the mainland *things* are sufficient.
	(S1S2) 东西		文化 东西 融合。 Cultures *Oriental and Western* integrated
NF	(SN) 东・西	大陆什么充足？ (What is sufficient on the mainland?)	大陆 **东・西** 充足。 On the mainland ***things*** are sufficient.
	(S1S2) 东西	文化怎么融合？ (How do cultures integrate?)	文化 **东西** 融合。 Cultures ***Oriental and Western*** integrated

2.2 Participants

Eight graduate students (4 males and 4 females, with average age of 28) were recruited in the recording experiment for a small stipend. They were native speakers of northern Mandarin without a diagnosed reading or hearing disability, living in Beijing for a long time and their Mandarin proficiency level is at Level 2A or above.

2.3 Recording procedure

The recording was conducted at a sampling rate of 11025Hz with a 16bit quantitative resolution in a sound-proof booth in the phonetic lab affiliated with Beijing Language and Culture University. The recordings obtained were monitored and screened by the first author to make sure they were of good quality and without mispronunciations. Before the statements of NF pronounced, the experimenter produced a question in order to provide a context for the participant to pronounce the focal sentences naturally. In the end, there were 256 sentences read in BF statements and 256 in NF statements, 512 sentences in total, among which there were 256 SN and S1S2 respectively.

2.4 Data analysis

All sound files were first automatically segmented by an alignment program (developed by SAIT-Lab) to generate syllable-level and phone-level boundaries. Visual inspection and manual correction were taken to ensure the segmental boundaries accurately marked.

F0 contours of each recording were analyzed by Praat [1]. F0 values (Hz) were then extracted for each final in a syllable at nine points in equal intervals to normalize in duration dimension, with creaky voice excluded. Duration (ms) and intensity (amplitude-duration integral, [19]) of each syllable were also extracted, using Praat scripts.

F0 values (Hz) were converted into semitone values (St), where the reference frequency was 55Hz for the male, 64Hz for the female.

After that, the original data of the pitch range, duration, intensity of the target words were converted into the corresponding proportional relationship, and the original were also conducted for statistics in R [28].

3. Acoustic Features

We concentrated on the acoustic features of the target words instead of analyzing the intonation of whole sentence, because the aim of this study is to investigate the prosodic features of stressed vs. unstressed words under two information structures.

3.1 Pitch

Figure 1 shows the tonal curve and pitch range for each target word and syllable in BF and NF. Table 2 lists specific values of pitch range of target syllables and the ratio of them.

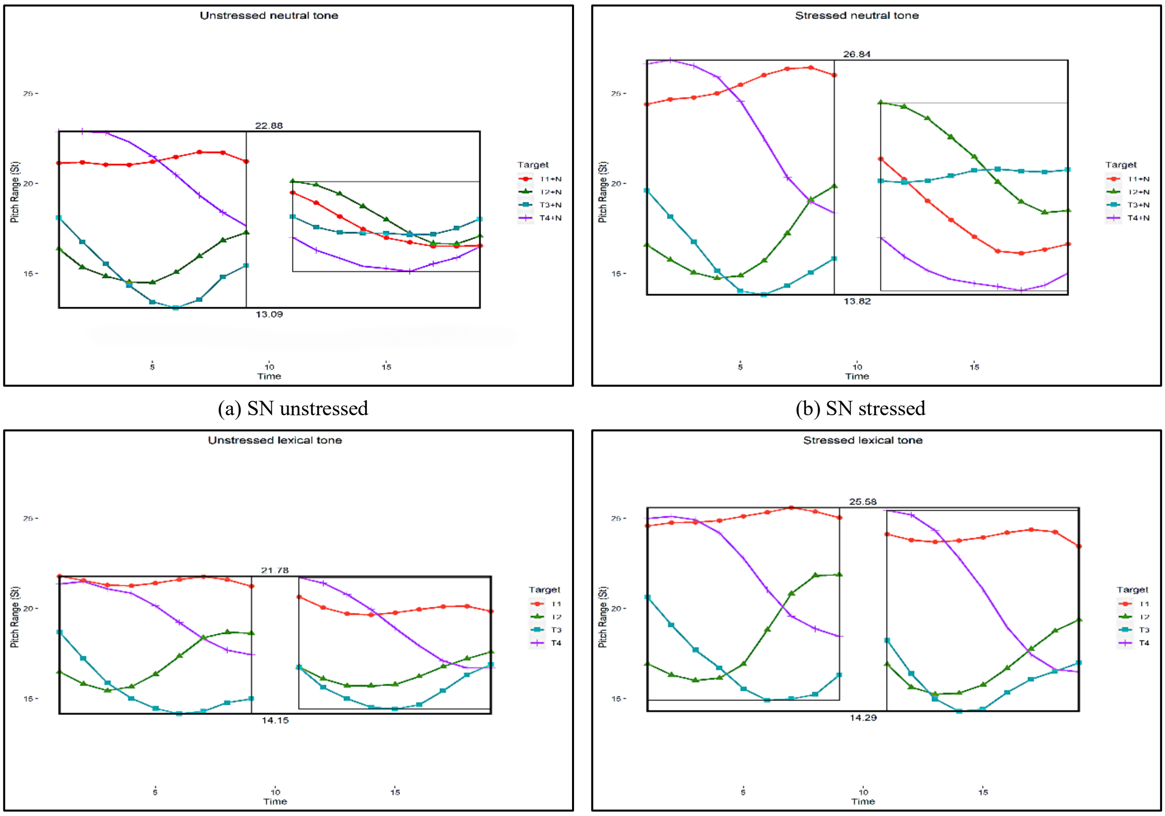
The pitch contours of N mostly performs as a falling tone after Tone1, Tone2 and Tone4, as a flat tone when following Tone3. And the pitch value is the highest after Tone2 and the lowest after Tone4. The top line of pitch of S in SN is higher than that in the target first syllable of S1S2 when the target word in BF and NF, and the pitch top line of N is lower than that in S2. Pitch top lines of the target disyllable of SN and S1S2 are both higher when located in NF than in BF. Compared to BF sentences, each top line of pitch in NF has been raised to a certain extent, by 3.96st and 3.37st in SN, 3.38st and 3.72st rising in S1S2 respectively.

The bottom lines of pitch of S in SN are lower than S1 in S1S2 in both BF and NF. In BF, the pitch bottom line of N is higher than that of the counterpart, while lower than that of S2 in NF. Bottom lines of S and S1 in the two types of target words in NF are both higher than the corresponding in BF, while bottom lines of N and S2 in NF lower than their counterparts in BF. The bottom lines of pitch of S and S1 increase by 0.73st and 0.14st respectively, those of the second syllables decrease by 1.06st and 0.13st separately, the falling degree in bottom line of N being the greatest.

Regardless of the two contexts, the pitch range of S in SN is wider than that of S1 in S1S2, and that of N is narrower than that of S2. Comparing the pitch ranges of target syllables in NF to BF, ranges of the target disyllable in SN and S1S2 are all extended, a greater degree of expansion occurring in the range of target second syllable.

Table2: *The pitch range of target disyllable*

Contexts	Pitch range (St)		Range Ratio (NF/BF)		Range Ratio
	1st	2nd	1st	2nd	(2nd / 1st)
SN-BF	9.79 (22.88-13.09)	5.00 (20.11-15.11)	1.33	2.09	0.51
SN-NF	13.02 (26.84-13.82)	10.43 (24.48-14.05)			0.8
S1S2-BF	7.63 (21.78-14.15)	7.29 (21.71-14.42)	1.40	1.53	0.96
S1S2-NF	10.69 (25.58-14.29)	11.14 (25.43-14.29)			1.04



(a) SN unstressed (b) SN stressed

(c) S1S2 unstressed (d) S1S2 stressed

Figure1: *Pitch patterns of SN and S1S2 in BF and NF,* T1~T4 denote four tones respectively.

The present study investigated the difference between NF and BF via comparing the acoustic parameter of the target syllable in NF to that in BF. Since the ratio is the result of NF to BF, it indicates the acoustic parameter of NF is greater than that of BF when the ratio is more than 1, and NF is smaller than BF when the ratio is less than 1. This means that the ratio goes larger when the acoustic parameter compared increases.

According to Range Ratio (NF/BF) in Tab.2, among the four syllables of two disyllabic target words, the degree of range expansion in S of SN and disyllable of S1S2 is similar, while the largest range expansion occurring in N.

Then the method of comparing second syllable to first syllable in the target word is used to investigate the difference between N and S. When the ratio goes smaller, it means the acoustic parameter of the second syllable is much less than that of the first syllable. Range Ratio (2^{nd} / 1^{st}) in Tab.2 shows that, the pitch-range difference between N and its preceding in BF is the biggest among the four contexts, the former approaching to 50% of the latter. When in NF, the pitch-range difference between the two becomes smaller, with still a 20% gap yet. In S1S2, the pitch range of S2 in BF is 4% narrower than that of S1, while in NF that becoming 4% wider than S1.

To sum up, the pitch ranges of SN and S1S2 have been expanded in NF, and the rising of top line has obvious effect on the expansion of pitch range, especially in S of SN. Among the syllables of the two types of disyllabic words, though pitch range of the second syllable is still less than that of its preceding, the pitch-range expansion of N is the greatest. The pitch range of S2 changes from narrower than that of S1 in BF to broader than that of S1 in NF.

3.2 Duration

Table3 exhibits the duration of each target syllable and the duration ratio between syllables. In both BF and NF, the duration of S in SN is the same or similar to its counterpart in S1S2. The duration of neutral-tone syllable is shorter than its preceding in both BF and NF. In S1S2, the S2's duration is similar to S1 in BF, but it is less than S1's duration in NF.

Table3: *The duration of target disyllable*

Contexts	Duration (ms)		Duration Ratio (NF/BF)		Duration Rati
	1^{st}	2^{nd}	1^{st}	2^{nd}	(2^{nd} / 1^{st})
SN-BF	249	195	1.30	1.18	0.78
SN-NF	324	229			0.7
S1S2-BF	249	248	1.29	1.19	1
S1S2-NF	320	295			0.92

Comparing the target syllable in NF and BF, the duration of syllables of SN and S1S2 in NF has been lengthened, and the extended degree of duration of the first syllables is larger than that of the second. Duration Ration (NF/BF) shows there is the greatest lengthened degree in S of SN, while the smallest lengthened degree in N in NF. The duration of the first syllable in S1S2 is less prolonged than that in SN, and duration-lengthening degree of S2 is larger than that of N. The obvious trend is that the prolonged degree in S and S1 is close, as is that in N and S2.

Duration Ratio (2^{nd} / 1^{st}) of SN and S1S2 decreases from BF to NF, both falling by 8%. This indicates that, whether in SN or S1S2, the duration difference between the first and second syllable becomes greater in NF than in BF, that is, the duration of second syllables of SN and S1S2 is much less than that of the first syllables.

3.3 Intensity

Similar to the feature of duration, in both BF and NF, the intensity of S in SN is close to its counterpart in S1S2. The intensity of the N and S2 is much weaker than that of the preceding in BF and NF, especially the SN.

Table4: *The intensity of target disyllable*

Contexts	Intensity		Intensity Ratio (NF/BF)		Intensity Ratio (2^{nd} / 1^{st})
	1^{st}	2^{nd}	1^{st}	2^{nd}	
SN-BF	333	177	1.78	1.57	0.53
SN-NF	591	276			0.47
S1S2-BF	348	284	1.73	1.66	0.82
S2S2-NF	601	471			0.78

Given the intensity ratio of NF to BF, the intensity of syllables of SN and S1S2 in NF has been strengthened, and there is a much larger enhanced degree in the intensity of first syllables than that of second syllables. Intensity Ratio (NF/BF) shows that the intensity of S in SN is enhanced the most among all syllables of the two types of words, the minimum enhancement in intensity of N. The strengthened degree of S1 in S1S2 is less than that of the counterpart in SN, while that of S2 more than N. It can be seen that the enhancement in S and S1 is close, that in S2 a little larger than in N.

Intensity Ratio ($2^{nd}/1^{st}$) of SN and S1S2 falls from BF to NF, SN by 6%, S1S2 by 4%, which indicates that regardless of the target words, the intensity difference between the first and second syllable becomes larger in NF than in BF. In particularly, the intensity of N varies from greater than 50% of its preceding in BF to less than 50% of its preceding in NF.

4. Stress Gravity

From the raw data, values of the acoustic parameters of target words in broad focus are significantly less than those in the focus contexts ($p<0.05$), in terms of pitch, duration and intensity via Growth Curve Analysis (GCA) and Mixed Linear Model. The pitch range of N and S2 is extended to a certain extent in focus sentences. The main manifestation is that pitch range of S2 changes from less than its first syllable in BF to more than the first in NF, while N's pitch range in both BF and NF is smaller than its preceding. Nevertheless, duration and intensity of the target second syllable in NF have both been compressed, the difference between the target disyllable becoming more distinct than in BF, whether the syllable is N or S. The result of expansion in pitch range of S2 stressed has been confirmed in previous studies [30, 31], while the shortening duration of S2 stressed is different from findings in most previous research [8, 32]. Huang explored the acoustic characteristics of target S1S2 stressed at different positions in statements [11] based on the materials in [29], finding that the duration of target second syllable of Tone1 at the middle of the sentence is shorter than its preceding in the word, but longer at the end of the sentence. This inspires us that the position of target stressed words in sentences may affect the acoustic features of targets and we would investigate the boundary effect on N further.

As Yin put forward that the product of average pitch and duration has a strong correlation with the strength of the syllable [38]. We propose a rather robust feature called Stress Gravity by calculating the product of pitch range and duration of a syllable to present the degree of phonetic strength efficiently.

According to the Stress Gravity in Tab.5, in the two information structures, N and S2 are both weaker than their preceding syllables respectively. In SN, the difference between the Stress Gravity of N and its preceding is great in both BF and NF, approximately 50% difference. In S1S2, S2's Stress Gravity is similar to S1.

Comparing NF to BF, the Stress Gravity of SN and S1S2 increases obviously in NF, and the enhancement of N and S2 is larger than that of the preceding, which is more evident in SN, the magnitude of increase in Stress Gravity of N being more than 70% of that in S of SN. The enhanced degree in Stress Gravity of S2 is close to S1 of S1S2, only 2% difference. Among the four target syllables, there is the greatest enhancement of Stress Gravity in N, while the increase in Stress Gravity of other three syllables similar. Based on the pitch range ratio (NF/BF) in Tab.2 and duration ratio (NF/BF) in Tab.3, there is the maximum expansion in pitch range of N, while the lengthening degree in N and S2 being close, which indicates that the pitch-range expansion makes the most contribution to Stress Gravity enhancement of N.

Table5: *The Stress Gravity (SG)*

Contexts	SG		SG Ratio (NF/BF)		SG Ratio (2^{nd} / 1st)
	1^{st}	2^{nd}	1^{st}	2^{nd}	
SN-BF	2435	973	1.73	2.45	0.40
SN-NF	4224	2386			0.56
S1S2-BF	1899	1808	1.80	1.82	0.95
S1S2-NF	3420	3284			0.96

Stress Gravity Ratio ($2^{nd}/1^{st}$) shows that in NF, the difference of Stress Gravity between the first and second syllable in SN and S1S2 is narrower than that in BF. Compared with BF, the Stress Gravity ratio in SN is greatly improved in NF, but the gap between the two syllables is still very obvious.

5. Conclusion

This study investigates the stress relationship between a neutral-tone syllable and a full syllable in target words on narrow focus and broad focus. Three acoustic parameters and one combination of Stress Gravity are adopted to measure the level in relationship. The results show that, among the target syllables of two types of words in NF, the maximum rising in the top line occurs to S of SN compared to the BF context. The pitch range of N is widened greatly, but it is still narrower than its preceding, and S2 pitch range is broader than its preceding in NF. Duration and intensity of the second syllables in NF are both enhanced, but the difference of duration and intensity between the second syllable and the first syllable is enlarged at the same time, particularly in SN. It can be concluded that compared to S1S2, SN is more likely to achieve focus by raising the top line of its first syllable and increasing the difference in duration and intensity between the two syllables. When in NF, N is also affected by focus to increase its pitch range, duration and intensity, but the enhancement in Stress Gravity of N mainly depends on the expansion of pitch range.

In addition, the result of pitch-range expansion in S2 stressed has been confirmed in previous studies. However, Stress Gravity of S2 is less than that of S1 when S1S2 on focus, which is another new finding, requiring further exploration. And the boundary effect on N would be also explored in the future.

6. Acknowledgements

This work is supported by the "Four Batches" Talent Project grant awarded to Aijun Li and the fellowship of China Postdoctoral Science Foundation awarded to Jingwen Huang.

Many thanks would like to be given to Prof. Feng Shi for his help and suggestions in this study.

7. References

[1] Boersma, P., Weenink, D., 2011. *Praat: Doing phonetics by computer.* Retrieved from: http://www.fon.hum.uva.nl/praat/.

[2] Cao, J. F., 1986. An analysis of properties of neutral tone syllables in Standard Chinese. *Journal of Applied Acoustics* 5.

[3] Chao, Y. R., 1932. A preliminary study of English intonation (with American variants) and its Chinese equivalents. In Wu Zongji, Chao Xinna (eds.), *Proceedings of Linguistics by Chao Yuenren*. Beijing: The Commercial Press, 2002.

[4] Chao, Y. R., 1948. *Mandarin primer: An intensive course in spoken Chinese*. Cambridge: Harvard University Press.

[5] Chao, Y. R., 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

[6] Chen, Y. Y., Xu, Y., 2006. Production of weak elements in speech – Evidence from F0 patterns of neutral tone in standard Chinese. *Phonetica* 63.

[7] Bolinger, D., 1972. Accent Is Predictable (If You're a Mind-Reader). *Language*, 48(3).

[8] Gao, M. M., 1993. *An experimental study on stressed prosodic features in Mandarin sentences*, Doctoral, Peking University.

[9] Gussenhoven, C., 1983. Focus, Mode and the Nucleus. *Journal of Linguistics* 19.

[10] Halliday, M. A. K., 1967. Notes on transitivity and theme in English. *Journal of Linguistic* 3.

[11] Huang, J. W., 2019. A study on the prosodic distribution of focus declarative sentences in Mandarin. *Chinese Language Learning* 6.

[12] Huang, J. W., Shi, F., 2019. Diversity of prosodic expressions on Chinese neutral tones. *Applied Linguistics* 1.

[13] Huang, J. W., 2021. *Suprasegmental feature of different types of neutral tone in Mandarin*, Doctoral, Beijing Language and Culture University.

[14] Lee, Wai-Sum, 2003. A Phonetic Study of the Neutral Tone in Beijing Mandarin. *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain.

[15] Li, A. J., 2017. Phonetic correlates of neutral tone in different information structures. *Contemporary Linguistics* 19.

[16] Li, A. J., Gao, J., Jia, Y., Wang, Y. R., 2014. Pitch and duration as cues in perception of neutral tone under different contexts in Standard Chinese. In Asia-Pacific Signal and Information Processing Association (ed.), *Proceedings of 2014 APSIPA Annual Summit and Conference*, Siem Reap, city of Angkor Wat, Cambodia, December 9–12.

[17] Li, A. J., Fan, S. S. 2015. Correlates of Chinese neutral tone perception in different contexts. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*.

[18] Li, A. J., Li, Z. Q., 2022. Prosodic realization of tonal target and F0 peak alignment in Mandarin neutral tone[J]. *Language and Linguistics* 1.

[19] Liang, L., Shi, F., 2010. The energy ratio of bi-syllable words in standard Chinese. *Nankai Linguistics* 16.

[20] Lin, M. C., Yan, J. Z., 1980. Acoustic characteristics of neutral tone in Beijing Mandarin. *Dialect* 3.

[21] Lin, M. C., Yan, J. Z., 1990. Neutral Tone & Stressed and Unstressed Syllables in Mandarin. *Language Teaching and Linguistic Studies* 3.

[22] Lin, T., 1962. The relation between stress and syntactic structures in Modern Chinese. *Studies of the Chinese Language* 7.

[23] Lin T., 1983. A preliminary study on the properties of neutral tone in Beijing Mandarin. *Essays on Linguistics 10*. Beijing: The Commercial Press. 1983.

[24] Lin, T., 1985. Preliminary experiments in the exploration of the nature of Mandarin neutral tone. In Lin, Tao & Wang, Lijia(eds.), *Working papers in experimental phonetics*, 1–26. Beijing: Peking University Press.

[25] Lu, J. L., Wang, J. L., 2005. On defining "qingsheng". *Contemporary Linguistics* 2.

[26] Lu, Y. Z., 1995. *Neutral tone and erhua in Mandarin*. Beijing: The Commercial Press.

[27] Jin, S., 2002. *The dynamic study of neutral tone in modern Chinese*. Beijing: The Ethnic Publishing House.

[28] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: https://www.Rproject.org/, 2019.

[29] Shen, J., 1982. *An analysis of pitch configuration on Mandarin intonation*, Master's, Peking University.

[30] Shen, J., 1992. A humble opinion on Chinese intonation model. *Linguistic Research* 4.

[31] Shen, J., 1994. Chinese intonation structure and type. *Dialects* 3.

[32] Shi, F., Liang, L., 2007. Acoustic performance of bi-syllable words with neutral-tone and normal stress in Mandarin. *Nankai Phonetics Annual Report* 3.

[33] Wang, B., Wang, L., Qadir, T., 2011. Prosodic encoding of focus in six languages/dialects in Chinse. *ICPhS*. Hong Kong.

[34] Xu, Y., et al., 2010. "Prosodic focus with post-focus compression: single or multiple origin?" *The Second Workshop on Evolutionary Linguistics.*

[35] Xu, Y., 2011. Post-focus Compression: Cross-linguistic Distribution and Historical Origin. *ICPhS*. Hong Kong.

[36] Xu, Y., Xu, C. X., 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33.

[37] Xun, E. D., Rao, G. Q,. Xiao, X. Y., Zang, J. J., 2016. Development of BCC corpus under the background of large data. *Corpus Linguistics* 3.

[38] Yin, Z. G., 2011. *A study on the rhythm of reading discourse in Mandarin*, Doctoral, Chinese Academy of Social Sciences.

[39] The Dictionary Editing Office of the Institute of Linguistics, Chinese Academy of Social Sciences. 2016. *Modern Chinese Dictionary (7th Edition)*, The Commercial Press, 2016.