USING FUNCTIONAL LOAD TO SIMULATE INFORMATION-MISSING AMBIGUITY PROCESSING OF HUMAN BRAIN

Rian Bao

School of Information Science
Beijing Language and Culture University
15 Xueyuan Road, Beijing, 100083, China

Diya A

School of Chinese Ethnic Minority Languages and Literatures
Minzu University of China
27 South Zhongguancun Avenue, Beijing, 100081, China

Jinsong Zhang

School of Information Science
Beijing Language and Culture University
15 Xueyuan Road, Beijing, 100083, China

ABSTRACT

Human being has the natural ability to recover speech with incomplete information. If the missing information of speech causes ambiguity, human brain would recover it as the most likely answer. This paper tries to simulate the recovering process using functional load (FL) theory. Correlation between the processing of ambiguity caused by loss of Mandarin tone information and the functional load of the tone is investigated. 14 ambiguous pinyin sequences without tone were selected, and every syllable was recorded as level tone and time-normalized. 30 native Mandarin speakers were asked to dictate the level-tone speech. The result of the dictation task showed that the tendency of recognizing the level-tone speech is significantly correlated to the functional load of the tone, and thus FL can be used to predict human's decision to ambiguous speech.

Index Terms— Tone information, functional load

1. INTRODUCTION

Natural language is always ambiguous. Listeners need contexts to get the most likely meaning the speakers intend to express. Unambiguous language is inefficient because of its redundant context [1], and language users do not use complicated words to avoid ambiguity, so ambiguity exists in every language. Therefore, human brain has the natural ability to process ambiguous speech, even when some information is missing, and this process is related to word frequency, context, syntactic structure, etc. In this paper, we intend to use information theory to simulate human brain's processing of ambiguous speech, to be specific, to predict the ranking of the possible candidates.

Functional load (FL) is a information-theoretic measure that computes phonological contrast's contribution to successful word identification. Phonetic errors with high FL have relatively larger effect on comprehensibility than those with low FL [2]. A tentative assumption can be put forward that when the speech's information is incomplete, listeners tend to recognize it adding the least amount of information.

Mandarin is a tonal language. There are five lexical tones in Mandarin Chinese, namely, high level (T1), mid rising (T2), low dipping (T3), high falling (T4) and neutral (T5), and previous studies have found lexical tone contrast has a very important role in the language system. The importance of tone in speech recognition has been measured by reduction of word uncertainty [3], and the result showed that conditioning on tone information can reduce word uncertainty in conversational speech by 11%-20%. An entropy-based functional load (FL) study has shown that lexical tone contrast has a comparable FL to that of vowels for Mandarin [4]. Another FL study based on mutual information of Chinese text and phonemes has found FLs of some tone contrasts are much larger than that of phoneme pairs [5]. In a dictation study, tone information is proved to be important in speech with poor segmental quality [6].

If Mandarin lost its tonal information, words that share same phonemes but different tones would definitely cause confusion for human brain. However, listeners would still come up with an answer on which their brain have already add the tones according to their language experience and linguistic knowledge. We intend to simulate this process using FL theory. In other words, to make computer deal with ambiguity like human being.

According to mutual-information-based FL theory, the information of tone in different sentences are different [5]. Previous studies consider tone information as an invariant value [3]. In present study, we are going to choose sentences with exceedingly high tone information, which means a great ambiguity would be caused without tone. When Mandarin sentences with high tone information lose their tones, how would the listeners deal with the ambiguity? To illustrate this point with a simple example, consider when people hear the word "gu shi" with flat tone and ask them to guess what the word is, I believe most of them would recognize the word as "故事 gu4 shi5 (story)" rather than "股市 gu3 shi4 (stock market)" or "古诗 gu3 shi1 (ancient poem)". And we found the FL of the tone of "故事 (story)" is much less than the other two, which means tone is less important for this word. Therefore, listeners need the least effort to recognize it as "故事 (story)".

The purpose of the current paper is to assess whether FL is able to predict the most likely answer of the ambiguous speech.

2. METHOD

2.1. Calculation of FLs

FL is used to evaluate the importance of phonetic contrasts [5]. In this study, FL is calculated using FL model based on mutual information between text corpus and phoneme transcription [5]. FL of a phonemic contrast can be calculated as the change in mutual information of spoken texts (W) and phoneme sequence (F) induced by the merger of the corresponding phonemic contrast. In other words, the merger of tones causes an increase of the number of word sequences which shares the same pinyin sequences, and the mutual information between the W and F will decrease. The reduction of mutual information reflects the loss of information caused by the merger of the phonemic contrast, and thus it is a optimal method to quantify information contribution of phonemic contrasts.

The FL of tone contrast is calculated as:

$$FL(T1, T2, T3, T4, T5) = \frac{MI(W, F) - MI(W, F_{T1-5})}{MI(W, F)} \qquad (1)$$

In (1), FL(T1,T2,T3,T4,T5) represents the functional load of 5 tones in the sentence. F is phoneme sequence, and $MI(W,F_{T1\text{-}5})$ represents the mutual information of the pinyin transcription and word sequence when 5 tones are merged in this language.

The mutual information is computed as:

$$\mathrm{MI}(\mathrm{W},\mathrm{F}) = \lim_{n\to\infty} -\frac{1}{n}\log\sum_{i=1}^{m} \mathrm{P}(\mathrm{W}'_i) \qquad (2)$$

In (2), W'_1, W'_2, ..., W'_m are word sequences sharing the same pinyin transcription F. The probability of word sequence P(W'_i) is computed using bi-gram and tri-gram language model. The language model is trained using pinyin transcription of Chinese TV programs of 2007.

2.2. Materials

14 target sentences were selected (see in Table 1), all of which cause ambiguity in the absence of tone information. In other words, after removing all the tone marks of the pinyin sequence of a selected sentence, it can represent more than one grammatically smooth sentences. The average length of the sentences is 5.6 syllables (SD = 0.9). 15 filler sentences were designed as syntactically and semantically similar to the target sentences. Besides, 3 random sentences were selected as the practice sentences in the experiment.

Each sentence was spoken by a female native Mandarin speaker who was graded first class B (an excellent speaker) in Putonghua Proficiency Test. In order to fully remove tone information. We decided to use natural speech instead of resynthesized speech, because the latter cannot fully remove tone information by flattening f0 contours only [7][8]. Tone information not only exists in f0, but also amplitude envelope and duration [9].

In order to avoid any tone information in the stimuli, the speaker was given a list of the pinyin sequences of the sentences, all of which had been marked as tone 1 (level tone). The speaker was asked to read the list without knowing the lexical meaning of the sentences. After the recording, the duration of each syllable was normalized to 386ms (average score of all the syllables' durations) using Praat script. A sample of the normalized speech can be seen in Fig. 1.

The Pinyin sequence "zhe ge hua ti hen hao" in Fig. 1 can either be understood as "zhe4 ge5 hua4 ti2 hen3 hao3" meaning "this topic is very good" or "zhe4 ge5 hua2 ti1 hen3 hao3" meaning "this slide is very good".

Table 1. Pinyin Sequences Selcted in the experiment

Sequence 1	wo xi huan zhe ge gu shi
Sequence 2	lao ma xi huan shui jiao
Sequence 3	ta xi huan shi zi
Sequence 4	wo yao bao shi
Sequence 5	wo xi huan ta de mei mao
Sequence 6	ta han yu shuo de hao
Sequence 7	dui bei jing zuo jie shao
Sequence 8	ta kan le kan yan se
Sequence 9	zhe shi wo de tu di
Sequence 10	wo xi huan hua xue
Sequence 11	zhe ge hua ti hen hao
Sequence 12	ta shi zhu li
Sequence 13	wo tao yan wei qi
Sequence 14	wo rang ta ban zou

2.3. Participants

Participants were 30 native Mandarin speakers (16 females, mean age = 23.2 yr, SD = 1.95) and were students in University. All of them reported normal hearing and no history of language disorder.

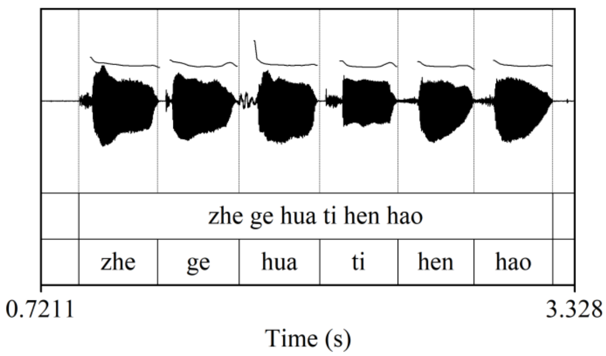

Fig. 1. A sample pinyin sequence without tone information

2.4. Procedures

The experiment was designed and carried out on the Finding Five platform, and the dictation task was played through headphones. First, participants were asked to listen to 3 practice sentences to familiarize the experiments. Then the formal trials were played randomly. Participants were asked to write down the sentences on a paper, and one answer only for each trial. After writing down a sentence, participants click “next” button to listen to the next sentence. When the experiment was over, the answer sheets were collected.

2.5. Data Analysis

The orthographic outputs of the participants are analyzed statistically. Answers with segmental errors and grammatical errors are excluded from the analysis. The frequency of each possible answer’s occurrence and FL of tone contrast (T1-T5) in each answer sentence based on Bi-gram and Tri-gram are calculated. The partial correlation between FL and was analyzed SPSS.

3. RESULT

The overall result of the dictation task can be seen in Fig.2. The FL value of each possible answer was computed and ranked. In most cases, listeners tend to recognize ambiguous sequence as the sentence with lower tone FL. However, there are still some exception, the detailed sentences can be seen in Table 2.

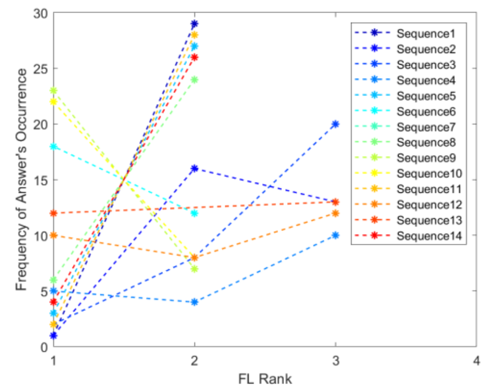

Fig. 2. Relation between FL rank and frequency of answer’s occurance

Table 2. The Overall Result of the Dictation Task (FAO: frequency of answer’s occurrence, R: rank of FL)

Seq.	Sentence	FAO	FL of tone (bi-gram)	FL of tone (tri-gram)	R
1	我喜欢这个古诗	1	0.212423	0.212675	1
	我喜欢这个故事	29	0.0354665	0.0354724	2
2	老妈喜欢水饺	1	0.319602	0.274547	1
	老妈喜欢睡觉	16	0.19628	0.143051	2
	老马喜欢睡觉	13	0.166315	0.129739	3
3	他喜欢柿子	2	0.101686	0.101999	1
	他喜欢识字	8	0.0803496	0.0808796	2
	他喜欢狮子	20	0.0333407	0.0341947	3
4	我要报失/报诗	3	0.208108	0.209306	1
	我要保湿	8	0.175284	0.169542	2
	我要保释/保试	5	0.100174	0.0956029	3
	我要暴食/报时	4	0.0881178	0.0896066	4
	我要宝石	10	0.0699603	0.0658836	5
5	我喜欢她的美貌	3	0.0655989	0.0687145	1
	我喜欢她的眉毛	27	0.0229971	0.0230068	2
6	他韩语说得好	12	0.0081987	0.00447535	2
	他汉语说得好	18	0.0080793	0.00744726	1
7	对背景做介绍	2	0.119911	0.120393	1
	对北京做介绍	28	0.0013591	0.00141433	2
8	他看了看眼色	6	0.0067322	0.00558842	1
	他看了看颜色	24	0.0030807	0.00105537	2
9	这是我的徒弟	23	0.114892	0.127433	1
	这是我的土地	7	0.0052898	0.00351844	2
10	我喜欢滑雪	22	0.0612612	0.0613863	1
	我喜欢化学	8	0.041318	0.0414578	2
11	这个滑梯很好	2	0.0339113	0.0539165	1
	这个话题很好	28	0.0003053	0.00031338	2
12	他是主力	10	0.0442044	0.0438012	1
	她是朱莉	8	0.0387232	0.0431461	2
	他是助理	12	0.0385793	0.0382034	3
13	我讨厌尾气	12	0.092122	0.0919378	1
	我讨厌违期	3	0.0855643	0.0853056	2
14	我让她伴奏	4	0.0438743	0.0452531	1
	我让她搬走	26	0.0156656	0.016112	2

The most ambiguous sequence was recognized as 5 different sentences (sequence 4). It can be seen from Table 2 that the listeners tend to recognize pinyin sequences as the sentence with lowest tone information (FL), eg. “wo xi huan zhe ge gu shi” has been recognized as “I like this story” and “I like this ancient poem”, and the former answer occurred 29 times, the latter only occurred once. The FL of the tone in the former sentence is much lower than the latter one, which means tone is less important in the former sentence. However, there are still some answer didn’t show a relation with FL, such as sequence 9. The answer with “徒弟 (apprentice)” occurred more than “土地 (land)”, which is probably because of the syntactic structure of this sentence suits “徒弟 (apprentice)” more than “土地 (land)”, and the FL is low probably because the word “徒弟(apprentice)” has few occurrence in the training corpus, i.e. TV program speech.

Table 3. Correlation between FAO and FL

	FL (Bigram)	FL (Trigram)	FAO
FL (Bigram)	1.000	.982 ***	-.461 **
FL (Trigram)	.982***	1.000	-.505 **
FAO	-.461**	-.505**	1.000

Pearson correlation coefficient (PCC) was used to quantify the degree of the relation between FL and answers’ frequency. A significant negative correlation between FL (based on bi-gram and tri-gram) and frequency of answer’s occurrence was observed ($p < 0.01$, see in Table 3), which means when listening to an ambiguous sentence without tone information, listeners tend to recognize it as the one with tones of lower FL. Moreover, FL based on tri-gram has a higher correlation ($PCC = -0.505$, $p < 0.01$) with frequency of occurrence compared to bi-gram.

4. CONCLUSION

The main purpose of this study is to prove FL can simulate ambiguity processing of human brain. Based on the present study, following conclusions can be drawn:

First, FL can predict the human’s decision to ambiguous speech to a certain extent.

Second, tone is not a redundant information for Mandarin. Loss of tone information cause a certain degree of ambiguity.

The tendency of the recognition of the speech without tone information is related to the FL of the tone in its context. To be specific, when a sentence loses its tone information, listeners tend to recognize the sequence as the sentence with the lowest tone FL value, where tone is not important. When a sentence with high FL tone lose its tone information, it would be very difficult for the listeners to recognize it correctly.

The corpus used in present study is not diverse enough to predict human’s decision, and a better language model can be used to compute the probability of word sequence in the further study.

5. ACKNOWLEDGMENT

This work is supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Beijing Language and Culture University (21YCX177).

6. REFERENCES

[1] S. T. Piantadosi, H. Tily, and E. Gibson, "The communicative function of ambiguity in language," *Cognition*, 122.3, pp. 280-291, 2012.

[2] M. J. Munro, T. M. Derwing, “The functional load principle in ESL pronunciation instruction: An exploratory study,” *System*, 34.4, pp. 520-531, 2006.

[3] D. Surendran, and G. A. Levow. "The functional load of tone in Mandarin is as high as that of vowels," *Speech Prosody*, International Conference. 2004.

[4] J. Zhang, W. Li., Y. Hou, W. Cao and Z. Xiong "A study on functional loads of phonetic contrasts under context based on mutual information of Chinese text and phonemes," *7th International Symposium on Chinese Spoken Language Processing*, IEEE, 2010.

[5] Q. J. Fu, and F. G. Zeng. "Identification of temporal envelope cues in Chinese tone recognition," *Asia Pacific Journal of Speech, Language and Hearing*, 5.1, pp. 45-57, 2000.

[6] H. Wang, et al. "Relative importance of tone and segments for the intelligibility of Mandarin and Cantonese," *Proceedings of the 17th International Congress of Phonetic Sciences*, City University of Hong Kong, 2011.

[7] J. Chen, et al. "The effect of F0 contour on the intelligibility of speech in the presence of interfering sounds for Mandarin Chinese," *The Journal of the Acoustical Society of America*, 143.2, pp. 864-877, 2018.

[8] Chen, Fei. "Mandarin tone identification with F0-flattening processed single-vowels,” *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2019.

[9] T. Ng, M. Siu, and M. Ostendorf. "A quantitative assessment of the importance of tone in Mandarin speech recognition," *IEEE Signal Processing Letters*, 12.12, pp. 867-870, 2005.