An Exploratory Study for Quantifying the Contextual Information for Successful Chinese L2 Speech Comprehension

Rian Bao[1]*, Linkai Peng*[2]*, Yuchen Yan*[1]*, Jinsong Zhang*[1]

[1]School of Information Science, Beijing Language and Culture University, Beijing, China
[2]Youdao, NetEase, Beijing, China

boroooo@163.com, penglinkai96@gmail.com, yanyuchen1999@foxmail.com, jinsong.zhang@blcu.edu.cn

Abstract

Language comprehension requires context in order to function as communication between speakers and listeners. When inferring the speaker's intended meaning, listeners make use of contextual information. But the amount of contextual information has not been quantified in previous studies. The current study aims to introduce a model for quantifying the information loss of Chinese defective sentences uttered by second language (L2) learners and test the model through behavioral experiment using listener-based judgements of comprehensibility. Relative change of mutual information between text and phoneme sequence is used to represent the information loss. 52 speech samples elicited from 20 Urdu-speaking learners of Chinese are used for validating the predictability of relative information loss in inferring speaker's judgements of comprehensibility. The results showed that the information loss has significant negative correlation with comprehensibility judgements ($r = -0.649$, $p < 0.001$), and we observed that information loss around 1% has little effect on comprehension, and information loss above 5% has severe effect on comprehension for the sentences with 10-15 syllables.

Index Terms: language comprehension, comprehensibility, psycholinguistics, sentence processing

1. Introduction

The understanding of human language requires processing of input phoneme sequences and building of the message which the speaker's intended to transmit. As in Cohort model, human speech comprehension is achieved by processing incoming speech continuously as it is heard, and then computing the best interpretation of currently available input combining information in the speech signal with prior semantic and syntactic context [1]. In TRACE model, lexical representation can be activated through incomplete input (e.g., from "exting..." to "extinguish") [2], which can explain why native listeners are able to understand defective speech produced by second language (L2) learners.

Prior studies have proved that different phonemes have different importance for speech comprehension. According to their explanation, phonemic errors with high functional load (an information-theoretic measure that computes contribution of phonological contrast to successful word identification, hereafter referred to as FL) in L2 speech inhibit comprehension more than those with low FL [3, 4]. High and low FL classification in these studies was based on Brown's FL principle [5], in which FL is calculated through 12 related factors such as cumulative frequency and minimal pairs.

At the same time, an alternative line of research has focused on defective L1 speech. An attempt to explore word-level comprehension has employed word-onset gating, a paradigm in which a listener is presented with increasing amounts of a word's onset duration until the word can be correctly identified [6, 7]. Evidence showed that word-level comprehension is highly correlated with semantic contextual information that the semantic context of word candidates has a large facilitatory effect [8] in the gating task, while syntactic information has little effect. Studies also support that perceptual restoration of interrupted speech is easier when speech has high contextual information [9, 10]. To quantify the effect of contextual information, expectation-based theories adopted the concept *surprisal* to measure the processing difficulty for a word in a context [11, 12, 13], where the processing difficulty $D_{\text{surprisal}}$ is directly proportional to the surprisal of a word w_i in a context c, and the surprisal is equal to the negative log probability of the word in the context, see in (1).

$$D_{\text{surprisal}}\ (w_i \mid c) \propto -\log p\,(w_i \mid c) \tag{1}$$

Based on the expectation-based theories [14], it can be inferred that if a word is highly probable in a context, then most of the information has already been given from the context when the word is really encountered, so the word itself doesn't carry much information in this situation. If the encountered word has a low probability in this context, this word carries a large amount of information by itself. However, this model is hard to apply to phonemes, whose probability is relatively less meaningful to language comprehension compared to word probability. In this study, we intend to propose a phoneme-level model to quantify the relative contribution of a phoneme to comprehension in certain context, which can be used to evaluate L2 speech. We advance a model based on mutual-information-based functional load (FL) model [15], where the percentage of mutual information loss in a sentence caused by mispronunciation of phonemes can be calculated.

In order to test the model, the concept comprehensibility is used for behavioral experiment. Comprehensibility is an important concept for second language (L2) learning, which is defined as listeners' perception of how easily and smoothly they understand L2 speech [16]. From a methodological point of view, L2 comprehensibility is most often measured based on listeners' intuitive judgements on a 9-point scale (1 = difficult to understand, 9 = easy to understand) [16, 17, 18].

The goal of this paper is to introduce an information-theoretic model to quantify the information loss of L2 speech caused by mispronounced phonemes and to examine it through behavioral experiment. Then we will estimate the maximum level of information loss for successful language comprehension.

2. Method

2.1. Materials

The speech materials used in present study were selected from BLCU-SAIT corpus [19], which is an interlanguage speech corpus of L2 learners of Chinese. 52 read speech of simple sentences with segment and tone errors from 18 Urdu-speaking learners (9 male and 9 female) were selected in the present study, and each sentence contains 10 to 15 syllables. All the errors were annotated by a professional linguistic student. All the speakers were students from Beijing Language and Culture University.

2.2. The calculation of information loss

Based on mutual-information-based FL model [15], information loss is calculated through the relative change of mutual information (MI) between text corpus and phoneme transcription after phonemic merger which represents conflations of the phonemes. The difference between FL and information loss calculation is that the former is calculated from a large corpus and the latter is from a single sentence. When a phonemic pair merged, the Word Hypothesis Graph (WHG) will be extended and the MI between text transcription and corresponding pinyin sequence will decrease. Relative information loss (InfoLoss) is defined in Equation (3), where $\mathrm{MI(W,F)}$ and $\mathrm{MI(W,F}_\alpha)$ is the MI before and after all conflated phonemes α are merged, and MI is defined in Equation (2), where $\mathrm{W}_1^{'}, \mathrm{W}_2^{'}, \ldots, \mathrm{W}_m^{'}$ are all text sequences sharing the same phonemic transcription (in this case, pinyin transcription) F. $\mathrm{P}(\mathrm{W}_i^{'})$ is the probability of the text sequence, which is computed based on tri-gram language model trained with a 300,000-word transcribed text corpus of Chinese TV show.

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

$$\mathrm{InfoLoss}(\alpha) = \frac{\mathrm{MI(W,F)} - \mathrm{MI(W,F}_\alpha)}{\mathrm{MI(W,F)}} \tag{3}$$

MI can be represented as a WHG. For example, as shown in Figure 1, if a speaker is not able to distinguish *"n"* and *"l"* in Chinese and the pinyin sequence *"liu2 lao3 lao5 he1 niu2 nai3"* is pronounced as *"niu2 nao3 nao5 he1 niu2 nai3"*, the possible text sequence paths increase, and the WHG grows bigger, which represents more confusion and smaller MI between text W and phonemic transcription F. Therefore the relative decrease of MI can be viewed as the relative loss of information. In this study, five lexical tones (T1: high level, T2: mid rising, T3: low dipping, T4: high falling and T5: neutral) are also viewed as phonemes and calculated in the model.

2.3. Raters and comprehensibility judgements

We recruited six graduate students to participate in the comprehensibility rating experiment. All the raters are native speakers of Chinese, were all born in China and raised by monolingual parents, and none of them reported hearing disorder. All of them have no experience of teaching Chinese to speakers of other languages. They were all aged between 22 and 26 (M=23.5, 3 females and 3 males).

The comprehensibility rating tasks were conducted individually in a quiet room using the Praat's ExperimentMFC [20],

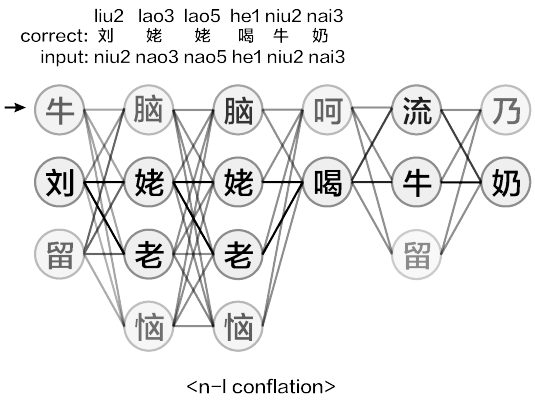

Figure 1: *A sample of Word Hypothesis Graph*

and all the rating results can be automatically recorded in the software. Each rater listened to the audio through a set of headphones on the researcher's laptop. Before the data collection, the investigator trained all the raters. First, the raters familiarized themselves with the listening materials (9 sentences with standard comprehensibility ratings). Then, each rater completed a practice rating session with three sample sentences. In the formal experiment, they were asked to pay attention on the effort it takes to understand the sentences. If they can understand the sentence very easily, then this sentence is highly comprehensible, and vice versa.

In the formal rating experiment, 52 sentences were divided into two groups following the principle of non-repetition of the sentence content. Each group was rated by three raters. Sentences were played for each rater in a randomized order. 12 sentences from other L2 speakers (outside the 18 speakers in 2.1) were included in both groups and used as computing inter-rater agreement. Each sentence can be played only one time. After hearing a sample, they made an intuitive judgement using a 9-point scale (1 = hard to understand, 9 = easy to understand). The detailed rules of comprehensibility rating can be seen in Table 1, where score 7-9 represent highly comprehensible, 4-6 represent moderately comprehensible, and 1-3 represent hardly comprehensible. The whole session took around 30 minutes.

3. Results

3.1. Inter-rater reliability

In terms of inter-rater reliability, Pearson's r was computed among three raters' scores from the same group. The strength of correlations is relatively high, which varied from r = .658 to r = .828, and the inter-group reliability is r = .805.

3.2. Correlation between comprehensibility judgements and information loss

Then Pearson correlations were computed to examine the strength of the relationship between mean comprehensibility and information loss. The result shows that log-transferred information loss have significant correlations with comprehensibility (r = -0.649, $p < 0.001$), indicating that the model is able to simulate human listeners' judgements.

Table 1: *Comprehensibility rating principles*

Comprehensibility Score	***Definition***
9	The speech is close to native speakers, and it is completely effortless to understand without active thinking. 100% of the content can be understood.
8	100% of the content can be understood, but listeners need to think for a while because of non-standard pronunciations.
7	100% of the content can be understood, but listeners need to think for a long time to completely get the correct words.
6	More than 80% of the content can be understood, and a word or two with mispronunciation might cause ambiguity or uncertainty.
5	70% of the content is understandable.
4	Main idea of the sentence, around 60% of the content, is understandable, but some details cannot be understood.
3	Listeners are unable to judge the main idea of the sentence, can only roughly guess the meaning of the sentence by understanding part of the words.
2	Listeners can only understand one or two words and are clueless about the main idea of the sentence.
1	Listeners are totally confused and clueless about the content.

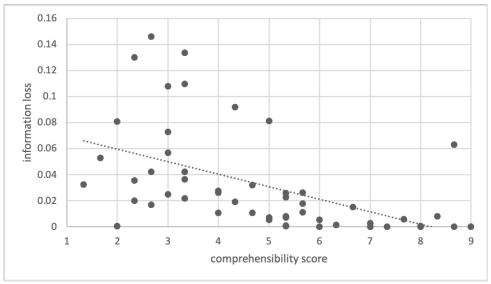

Figure 2: *The relationship between listener-based comprehensibility scores and relative information loss.*

3.3. Estimating the baseline of information loss for successful comprehension

The visualized overall information loss for three levels of comprehensibility can be seen in Figure 3. Mean information loss for comprehensibility score between 6-9 (highly comprehensible) is 0.011, for 3-6 (moderately comprehensible) is 0.032, and for 1-3 (hardly comprehensible) is 0.058. The analysis of independent t-test results showed significant difference in information loss between the three levels of comprehensibility ($p < 0.05$).

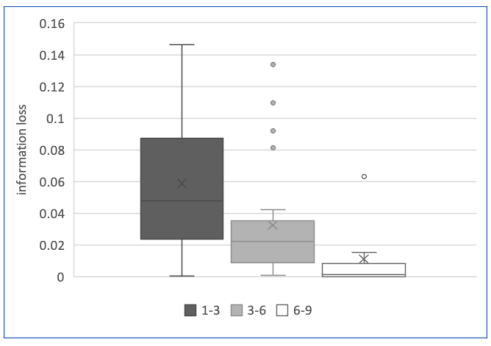

Figure 3: *The relative information loss of three levels of comprehensibility scores*

Figure 4-6 showed the distribution of information loss in three levels of comprehensibility. It can bee seen that most of the sentences that are highly comprehensible have less than 1% of information loss, and the most of the sentences which are hardly comprehensible have more than 5% of information loss.

We observed a sentence length effect that the tolerance of information loss for longer sentence is slightly greater than shorter sentence (See in Figure 7), but the difference between the information loss of short sentences and long sentences didn't show statistical significance ($p > 0.05$).

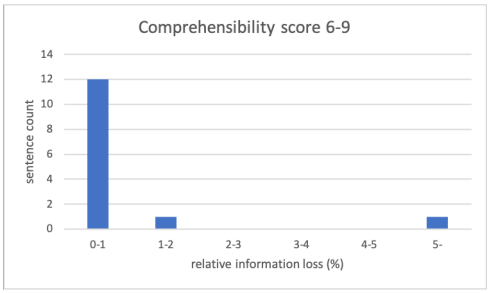

Figure 4: *Relative information loss of sentences with comprehensibility scores 6-9*

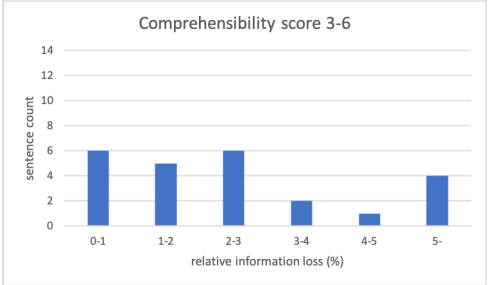

Figure 5: *Relative information loss of sentences with comprehensibility scores 3-6*

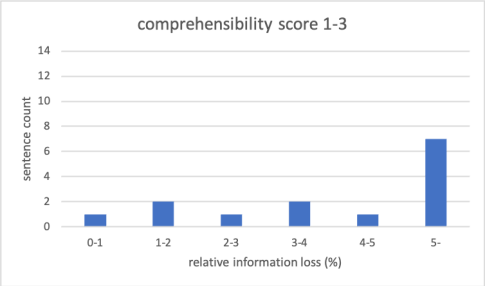

Figure 6: *Relative information loss of sentences with comprehensibility scores 1-3*

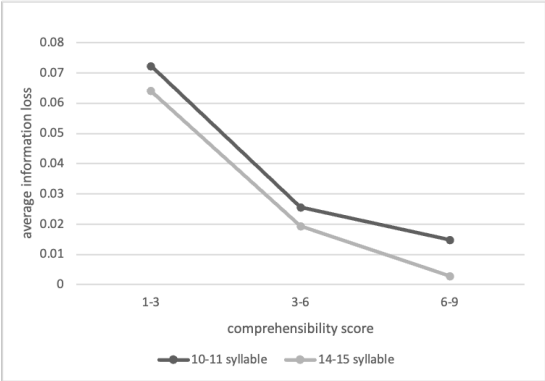

Figure 7: *Relative information loss of sentences with 10-11 syllables and 14-15 syllables*

4. Discussion

In the present study, our goal was to propose a model to quantify contextual information for Chinese sentences and to examine the model through listener-based comprehensibility scores. Relative information loss was calculated through the relative change of mutual information between text and phoneme. Mutual information can be interpreted as the probability of the target text deduced from the phoneme sequence. When two phonemes are merged, the probability will decrease, and the mutual information will decrease, which represents the relative information loss in the current study.

As noted in the Introduction, many theories of sentence comprehension used probabilistic expectation to quantify the probability of a word in certain context. Since the high probability words are pre-activated, high probability words don't carry much information itself. Pre-activation of low probability words is inefficient for language comprehension, so they carry much information when encountered. By assuming the mutual information between text sequence and phoneme sequence is the amount of contextual information, the information change caused by phoneme conflations can be calculated.

To validate the information loss model, we selected natural speech uttered from L2 speakers of Chinese. Speech data were obtained from a Chinese L2 speech corpus and rated by six naive listeners. The results of the behavioral experiment revealed that there is a significant negative correlation between information loss and listener-based comprehensibility judgements. Information loss below 1% had little to no effect on comprehension, and information loss above 5% had severe inhibitory effect on comprehension.

This work has relied on calculations done over only sentences with 10-15 words. There might be a sentence length effect that the tolerance of information loss for longer sentence might be greater than shorter sentence. In this study, we didn't provide a statistical significance for sentence length effect, which should be taken into consideration in the future study.

5. Conclusions

We have proposed a quantitative model for relative information loss of L2 speech, where information loss caused by phoneme conflation can be thought of in terms of relative change of mutual information between text and phoneme transcription. The model is tested by analyzing its correlation with comprehensibility judgement. In the behavioral experiment, we have provided a preliminary evidence for the significant negative correlation between information loss and comprehensibility. For sentence-level comprehension, when the relative information loss is around 1%, listeners are able to understand the sentence completely. When the relative information loss is above 5%, the listeners almost cannot understand the sentence. Our findings also observed that there might be a sentence length effect, but it didn't show statistical significance due to the limited materials.

This is an exploratory study for quantifying sentence-level speech information. Future studies should further explore the information model and examine the validity of this model with more speech data. Furthermore, sentence length effect can be further tested.

6. Acknowledgements

This study was supported by advanced Innovation Center for Language Resource and Intelligence (KYR17005), and Wutong Innovation Platform of Beijing Language and Culture University (19PT04), and the Fundamental Research Funds for the Central Universities, and the Research Funds of Beijing Language and Culture University (21YCX177). Jinsong Zhang is the corresponding author.

7. References

[1] W. D. Marslen-Wilson, "Functional parallelism in spoken word-recognition," *Cognition*, vol. 25, no. 1-2, pp. 71–102, 1987.

[2] A. G. Samuel, "Speech perception," *Annual review of psychology*, vol. 62, pp. 49–72, 2011.

[3] M. J. Munro and T. M. Derwing, "The functional load principle in ESL pronunciation instruction: An exploratory study," *System*, vol. 34, no. 4, pp. 520–531, 2006.

[4] Y. Suzukida and K. Saito, "Which segmental features matter for successful l2 comprehensibility? revisiting and generalizing the pedagogical value of the functional load principle," *Language Teaching Research*, vol. 25, no. 3, pp. 431–450, 2021.

[5] A. Brown, "Functional load and the teaching of pronunciation," *TESOL quarterly*, vol. 22, no. 4, pp. 593–606, 1988.

[6] S. Cotton and F. Grosjean, "The gating paradigm: A comparison of successive and individual presentation formats," *Perception & Psychophysics*, vol. 35, no. 1, pp. 41–48, 1984.

[7] F. Grosjean, "Spoken word recognition processes and the gating paradigm," *Perception & psychophysics*, vol. 28, no. 4, pp. 267–283, 1980.

[8] L. K. Tyler and J. Wessels, "Quantifying contextual contributions to word-recognition processes," *Perception & Psychophysics*, vol. 34, no. 5, pp. 409–420, 1983.

[9] S. Grossberg and S. Kazerounian, "Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise," *The Journal of the Acoustical Society of America*, vol. 130, no. 1, pp. 440–460, 2011.

[10] C. Patro and L. L. Mendel, "Role of contextual cues on the perception of spectrally reduced interrupted speech," *The Journal of the Acoustical Society of America*, vol. 140, no. 2, pp. 1336–1345, 2016.

[11] J. Hale, "A probabilistic earley parser as a psycholinguistic model," in *Second meeting of the north american chapter of the association for computational linguistics*, 2001.

[12] ——, "Information-theoretical complexity metrics," *Language and Linguistics Compass*, vol. 10, no. 9, pp. 397–412, 2016.

[13] R. Levy, "Expectation-based syntactic comprehension," *Cognition*, vol. 106, no. 3, pp. 1126–1177, 2008.

[14] D. Jurafsky, "Probabilistic modeling in psycholinguistics: Linguistic comprehension and production," *Probabilistic linguistics*, vol. 21, 2003.

[15] 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," in *2010 7th International Symposium on Chinese Spoken Language Processing*. IEEE, 2010, pp. 194–198.

[16] M. J. Munro and T. M. Derwing, "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," *Language learning*, vol. 45, no. 1, pp. 73–97, 1995.

[17] T. M. Derwing and M. J. Munro, "Accent, intelligibility, and comprehensibility: Evidence from four l1s," *Studies in second language acquisition*, vol. 19, no. 1, pp. 1–16, 1997.

[18] P. Trofimovich and T. Isaacs, "Disentangling accent from comprehensibility," *Bilingualism: Language and Cognition*, vol. 15, no. 4, pp. 905–916, 2012.

[19] B. Wu, Y. Xie, L. Lu, C. oCao, and J. Zhang, "The construction of a chinese interlanguage corpus," in *2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*. IEEE, 2016, pp. 183–187.

[20] P. Boersma and D. Weenink. (1992) Praat: doing phonetics by computer [computer program].