A Clustering Analysis of Chinese Consonants Based on Functional Load

Bin WU¹, Jinsong ZHANG¹,² * and Yanlu XIE¹,²

¹ College of Information Science, Beijing Language and Culture University
² Center for Studies of Chinese as a Second Language, Beijing Language and Culture University
E-mail: wubin7416850@outlook.com; jinsong.zhang@blcu.edu.cn; xieyanlu@blcu.edu.cn

Abstract — This paper attempts to provide some insights about the relationship between the differentiability and the classification importance of consonants in Chinese speech communication. The two characteristics can be modelled by the perceptual distance and the functional load respectively. We have a clustering analysis of Chinese consonants based on functional load (FL) relied on mutual information (MI) between the text and its phoneme transcription. Then we compare our clustering result with that based on the perceptual distance by articulation tests. By experimenting on the Chinese newspaper corpus with millions of sentences, we find most phonemes at the same place of articulation with different manners tend to have large FLs pairwise. It is consistent with the result that those phonemes tend to have long perceptual distance pairwise.

Index Terms: functional load, clustering analysis, perception

I. INTRODUCTION

Phonemes are described as the smallest contrastive linguistic units which may bring about a change of meaning [1], and their combinations form meaningful units such as words or morphemes. From the view of speech communication, the important roles of phonemes are two-sided: one is that a phoneme set provides a classification of speech acoustics. The other is that a sequence of phonemes forms the mental key for human to access specific words or meanings. Therefore, there are several interesting questions arising from phoneme classifications: Firstly, how differentiable is a phoneme set of a language? Especially, are the phonemes equally differentiable? Secondly, how important is a classification of each pair of phonemes in a language? And thirdly, what is the relationship between the differentiability and the classification importance of a pair of phonemes in a language? Answering these questions is not only important for understanding the mechanism of human speech processing, but also helpful for improving automatic speech recognition systems.

As the answer for the first question, phonemes are classified in the previous studies according to a number ways: different articulatory configurations, distinct features, distributional differences of phonetic acoustics, perceptual dissimilarities, and etc. [1-3] Among them, perceptual discrimination/confusion of different phonemes offered a direct image about how human ear can distinguish one phoneme or a class of phonemes from another based on acoustic cues only [2, 3]. Findings suggested that human ear is more easily to distinguish consonants with different articulation manners than different constriction places. Such a phenomenon exists in English [2], also in Chinese [3].

For the second question, the study of functional loads (FL) of phonemes or phonetic contrasts can be an answer [5, 6]. The FL measurements based on frequencies or entropies showed that different pairs of phonemes or classes may have significantly different roles in one language [5, 6]. For example, alveolar consonants are much more important than fricatives in English, Dutch and German. The differentiation of /s/ and /sh/ is more important than that of /f/ and /h/ in Chinese [6].

Compared to the relatively rich studies on the first two questions, few studies could be found with respect to the third one. Moreover, to find an acceptable answer to the third question, one might need a reasonably appropriate model for human speech cognitive process, which is assumed to be a mixing process consisting of a bottom-up acoustic-to-category perception and a top-down knowledge-based interpretation [4]. If perceptual discrimination can be regarded as a bottom-up classification of acoustics into phoneme categories, FLs based on frequencies or simple entropies as calculated in [5] cannot sufficiently well model the top-down process, thus inappropriate for the task.

However, the FL measurement based on the mutual information (MI) between text and its pronunciation, we proposed in [6], offers a new way to compute phoneme FLs, with a rational modeling of context effects including words and word coherences. It not only provides a more reasonable answer to the second question, but also opens the door to answer the third question.

Taking the example of 21 Chinese consonants, the previous study [2] concluded with a perceptual confusion tree, which can be regarded as the bottom-up classification structure by normal Chinese. Here, we will compute FL of each pair of the 21 consonants using a large text corpus, then cluster them into a hierarchical tree driven by the criterion of maximum FL. Finally, a comparison of the perceptual confusion tree and the FL structure will shed some light on the question what is the relationship between the consonant differentiability and their importance in Chinese speech communication.

The paper is arranged as follows: Section 2 introduces the theorem of FL based on MI. Section 3 briefly describes the clustering method and defines the distance based on FL.

* Corresponding author
Section 4 introduces the experiments based on corpus, shows the clustering result based on FL and compares it with the clustering result based on perceptual distance.

II. THEOREM OF FL BASED ON MI

A. Mutual Information

Suppose that we have a text \( W \) including sentences sampled from text language; each sentence is composed of word tokens. The word sequence \( W \) is transcribed as the phonemic transcription \( F \). Let

\[
MI(W, F) = H(F) - H(F|W)
\]

be the mutual information of the random text and its phonemic transcription. Suppose that the phoneme sequence \( F \) is both stationary and ergodic, and that we ignore the polyphones, we finally get (See the appendix for details of the proof.) [7]

\[
-\frac{1}{n} \log \sum_{i=1}^{m} P(W'_i) \rightarrow MI(W, F) \text{ with probability 1}
\]

(2)

where \( W'_1, W'_2, ..., W'_m \) are all text sequences have the same transcription \( F \) and \( n \) is the size of the sequence \( F \).

The probability of the text sequence \( P(W'_i) \) above can be computed by the bigram or the trigram language model, then the summation \( \sum_{i=1}^{m} P(W'_i) \) can be efficiently computed by the forward algorithm.

The \( MI \) can be represented as a Word Hypothesis Graph (WHG). For example, a part of WHG of the word sequence of "哺乳动物" with its pinyin transcription of "bu ru dong wu" can be:

![Word Hypothesis Graph](image)

Fig. 1 A part of WHG of the word sequence of "哺乳动物" (mammal) with its pinyin transcription of "bu ru dong wu".

Intuitively, more text sequence paths \( \{W'_i\}_{i=1}^{m} \) within a WHG tend to cause more confusability, and smaller \( MI \) between the Chinese text \( W \) and its phonemic transcription \( F \).

B. The functional load based on the MI

Suppose we want to measure the functional load of the phoneme \( x \) and the phoneme \( y \). We merge all \( x \) and \( y \) in the language, replace them with same phonemes \( x \), and keep other phonemes unchanged. Then the variation of the phonemic system decreases, causing more words sharing same pronunciations. WHG might grow around positions of \( x \) and \( y \) of original phonemic system as we do phoneme-text transmission. Consequently, more text sequences share the merged transcription, which causes confusability and indicates the loss of information. The relative change of mutual information can be defined as functional load based on \( MI \) [6].

\[
FL(x, y) = \frac{MI(W, F) - MI(W, F')}{{MI(W, F)}}
\]

where \( F \) is the transcription of \( W \) of original phonemic system; \( F' \) is the transcription with phonemes \( x \) and \( y \) merged. \( MI \) is calculated by formula (2).

III. CLUSTERING METHOD

We use agglomerative complete-link clustering algorithm, which ‘yields a dendrogram representing the nested grouping patterns and similarity levels at which groups change’ [8].

1) \( S1 \) Initially, treat each Initial as a group and we will have 21 groups.

2) \( S2 \) Compute the distance of all group pairs, choose the two groups with the smallest distance.

3) \( S3 \) Create a new group with the two groups merged. Delete the two groups and insert the new group.

4) \( S4 \) Repeat \( S2, S3 \) until all the initials are in one group.

The distance of two groups of the compete-link algorithm is defined as the maximum distance of their phonemes pairwise; Let

\[
\text{Distance (phoneme1, phoneme2)} = 1 - FL(\text{phoneme1, phoneme2})
\]

be the distance of a phoneme pair because we are interested in the phoneme groups that do the most work in keeping the language apart; the phonemes in same group tend to have high functional loads pairwise.

IV. EXPERIMENTS

A. Chinese phonology

<table>
<thead>
<tr>
<th>TABLE I. INITIALS WITH UP- UN-ASPIRATED, AP-ASPIRATED, VL- VOICELESS AND VC- VOICED.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Bilabial</td>
</tr>
<tr>
<td>Velar</td>
</tr>
<tr>
<td>Alveolar</td>
</tr>
<tr>
<td>Retroflex</td>
</tr>
<tr>
<td>Alveo-palatal</td>
</tr>
<tr>
<td>Labiodental</td>
</tr>
</tbody>
</table>

A word in Chinese is made up of one or several Chinese characters; each character can be phonemically transcribed as a pinyin – a monosyllable with a pitch tone. The base syllable of a pinyin usually begins with a consonant called an ‘Initial’ and ends with a vowel called a ‘Final’, discarding the pitch tone. There are 21 Initials in Table 1 categorized by the manner and the place of articulation.
B. Experimental setup

We sampled 1,000,000 sentences, with 7 to 9 characters each sentence, from the China Daily Newspaper. After word segmentation, those sentences are converted to pinyin sequences. The language model is trained from 3,000,000 sentences the China Daily Newspaper after segmentation using the SRILM TOOLKIT. We have computed functional loads of all pairs of different Initials for clustering. As the functional load is based on the mutual information of phoneme sequence and Chinese text, we experiment (A) on pinyin sequence with tone and on base syllables of pinyin sequence without tone; (B) on text language modeled by the bigram and pinyin sequence.

We divide the corpus into 5 groups, each with 200,000 sentences, and finally the 5 groups are merged.

C. FLs of Initial contrasts

FLs of contrasts of the five groups and the merged one are very similar, their correlations pairwise are more than 0.994. It means the FLs tend to converge. 21 Initials with 210 pairs are listed. Since pairs symmetric such as ‘m n’ and ‘n m’ have the same FL, only one is listed. Phonemes with themselves as pairs are not listed for their functional loads are zeros. To show the rough distribution of the FL, some labels of 210 pairs are missing at certain intervals.

In Fig. 2, three cases tend to have similar shapes while their values vary because of different levels of the context constrain.

D. FL and perception

According to the distance defined in the section of the CLUSTERING METHOD, larger FL means smaller distance. Those mergers causing great confusability in the language, as the WHGs grow, would appear in the same group with low ‘similarity’, implying smaller distance and larger FL, as showed in the y axis.

Fig. 2 Line chart of (A) pinyin sequence with or without tone and (B) text modeled by the bigram or the trigram.

Fig. 3 The dendrogram based on functional load with the mutual information between text modeled by the bigram and pinyin sequence with tone.

Fig. 4 Dendrogram based on psychological similarity of speech sounds adapted form [3].
In Fig. 3, phonemes grouped at the lowest levels like ‘i’, ‘d’, ‘h’, and ‘sh’ might refer to text words: ‘I’, ‘的’, ‘和’, and ‘是’. Those are among the most frequently used words in Chinese language [9]. That the group of ‘z’, ‘c’ and ‘s’ is always with higher ‘similarity’ compared to that of ‘zh’, ‘ch’ and ‘sh’ might because the former appears more often [9]. Phonemes used more frequently tend to be more important in the sense of FL.

In Fig. 3, we can find Initials at the same place with different manners tend to be in the same group. The dendrograms with pinyin sequence without tone or with the text computed by the trigram have similar results, which are not shown here. It suggests the robustness of our clustering results.

In Fig. 4, we can deduce most phonemes at the same place of articulation with different manners have long perceptual distance for they tend to be in different groups at low cutting level of the dendrogram.

Fig. 3 and Fig. 4 are the dendrograms based on perceptual distance and FL, which suggests the bottom-up and top-down processing of the speech processing in perception. Both of them show manners work more for Chinese language, particularly Initials at the same place with different manners.

Fig. 4, we can find Initials at the same place with different manners tend to be in the same group. The dendrograms with pinyin sequence without tone or with the text computed by the trigram have similar results, which are not shown here. It suggests the robustness of our clustering results.

In Fig. 4, we can deduce most phonemes at the same place of articulation with different manners have long perceptual distance for they tend to be in different groups at low cutting level of the dendrogram.

V. CONCLUSIONS

In this paper, we conduct a clustering analysis of FLs of Initials based on MI. We compare it with the clustering result of Initials based on perceptual distance. Results show the phonemes at the same place of articulation with different manners tend to have large FLs and long perceptual distance pairwise, which shows some relationship between the classification importance and differentiability of phoneme pairs in speech communication. In the future we will continue with the topic and try to find some applications.

ACKNOWLEDGMENT

The research underlying this paper was supported by National Nature Science Foundation of China (61175019).

REFERENCES