Date: Aug 2021

Lexical Effect on Mandarin Chinese Onset Perception in Noise

Jhe-Yu Jheng

Peking University, Department of Chinese Language and Literature, Phonetics Lab

Abstract— This article investigates how lexical knowledge influences the perception of Mandarin initial consonants in white noise. In particular, when a specific phoneme is activated by lexical knowledge, how do the activated phoneme influence the perception of other phonemes that are actually presented? Cocurrently, how do distinctive features play a related role in? The TRACE model (McClelland and Elman, 1986) assumed that inhibition occurred within the pre-lexical levels, for example, one phoneme could inhibit another once it was activated by the lexical knowledge. The present article thus concerns the inhibitory lexical effect between Mandarin phonemes. In the experiment, identification tasks were conducted by forced-choice technique, and collected data were analyzed by Multi-dimensional Scaling (MDS), Signal Detection Theory (SDT), Hierarchical Clustering (HC), and Max Entropy Grammar (MaxEnt). In terms of phonemes and natural classes respectively (the latter of which, representing the level of distinctive feature, are constituted with phonemes that exclusively share some common distictive features), this article analyzes the perceptual results with/without lexicality in white noise: In the condition that a phoneme is activated by lexical knowledge, the other speech are more likely to be perceived as a presentation of the lexically activated phoneme. Simultaneously, the article examines whether the results of phonemes/natural classes conformed to the predictions which are made in light of the inhibitory lexical effect (that is, the lexically incompatible phoneme targets will have higher probabilities to be perceived as (1) the lexical activated phoneme or (2) the other phonemes which share a natural class with the lexical activated one; the identification rates of the lexically incompatible phoneme targets may decrease.) The article finds that inhibitory lexical effect is better accounted for at the distinctive feature level.

Keywords— inhibitory lexical effect, speech perception, consonant perception, Madarin Chinese

I. Introduction

To begin with, an instance of related case to this article could be laid out in a conversation like: A: "Hey, it's time for lun–" B: "lunch? Come on, you've just finished your breakfast." In many cases like this, although a word is not completely represented, people can recognize it when hearing only a part of it. However, what if the someone says "hey, it's time for [lʌndʒ]"? And if that happens in a noisy environment, will the expected /tʃ/ affect the perception of [dʒ]?

As the questions above, how information acts in speech perception is among the crucial issues for linguists and cognitive scientists. Known as the botttom-up analysis, a consensus has been reached that speech recognition includes various processing levels from acoustic signals to concepts, e.g. acoustical analysis, lexical identification, syntatic analysis and semantics (Baars and Gage, 2010). Nevertheless, top-down process like lexical feedback remains controversial. Autonomous models (e.g., Cutler and Norris, 1979) argue that bottom-up analysis was the only fashion in the entire process, in other words, perception at the lower representational level can not be influenced by the information from the higher one. Proponents of interactive models (e.g., McClelland and Elman, 1986), on the contrary, allow lexical influences on the pre-lexical level. Although both autonomous models and interactive models can account for most effects on speech perception, several findings in neuroscience form the arguments in favor of the interactive view (see McClelland et al., 2006). It is proposed that the areas in brain are connected bi-directionally (Crick and Asanuma, 1986), and the top-down processing seems to appear as a general property regardless of processing vison or language (Baars and Gage, 2010).

Researches on lexical effects are the basis of the provided models. Lexical effect is the phenomenon that lexical knowledge influences the perception of phonemes, and it typically refers to the Ganong effect (Ganong, 1980). Ganong (1980) found that lexicality could shift the category boundary between phonemes. Taken as example, an ambiguous speech between /d/ or /t/ would be more likely to be identified as /d/ in the context "*-ash*", where /daʃ/ formed the word 'dash' and /taʃ/ did not. In subsequent researches, the relation between reaction time and the degree of the boundary shift are considered (e.g., Connine and Clifton, 1987; Kingston et al., 2016). Additionally, word frequency is another factor of the effect (e.g., Connine et al., 1993b; Shen and Politzer-Ahles, 2018).

Phoneme restoration is also an example of lexical effects. It was observed that the blank would be identified in the word

Contact data: Jhe-Yu Jheng, jheyujheng@gmail.com

with a deletion of one segment; however, auditory illusion would be likely to appear when the segment was replaced with a irrelevant sound, such as a cough or a pure tone (Warren, 1970). And the missing speech was even not able to be detected (Warren and Obusek, 1971).

Futhermore, lexical knowledge affects phoneme detection in the temporal aspect; in specific, targets are detected faster in a word than a non-word. In Frauenfelder et al. (1990), facilitatory lexical effect was attested that there were large differences in detection latencies to targets in words and matched non-words when the targets were at the positions after the uniqueness point[1] of the target-bearing word. The effects above illustrate that lexical knowledge facilitates speech perception in diverse perspectives, and in the interactive view, the facilitation is due to the lexical feedback through the bi-directional flow of information (McClelland et al., 2006).

Besides, the nature of interaction within the levels is of great interest to researchers as well. The TRACE model (McClelland and Elman, 1986), one of the interactive models, assumed that inhibition would occur within levels. The TRACE model (McClelland and Elman, 1986) contains numerous units which refers to the perceptual objects related to the utterance, and the units are organized into three levels, which are the feature level, the phoneme level and the word level. Between levels, information flow bi-directionally and lead to activations of units; on the other hand, the activated units inhibit the others within levels. As the assumption of the model, lexical knowledge will mediately inhibit phoneme identification, that is: The lexical knowledge will excite the corresponding phonemes after a word is activated, and the lexically activated phonemes will further inhibit other phonemes.

Phoneme monitoring experiments were consequently conducted by Frauenfelder et al. (1990) to examinate the extended prediction: Longer duration would be required to monitor a target phoneme when the target phoneme was incompatible with the lexical knowledge; for example, in the condition that the target phoneme (e.g., /t/) replaced one other phoneme (e.g., /l/) within a word (e.g., *vocabulary*) and thus turned it into a non-word (e.g. *vocabutary*). In Frauenfelder et al.'s Experiment 2 and the after ones, Inhibiting Non-words (INWs) were designed as the previous example (i.e., *vocabutary*), and it indicated that the word would be activated (e.g., *vocabulary*) and the lexically activated phoneme (e.g., /l/) would inhibit the actually presented one (e.g., /t/); Control Non-words (CNWs), as the control group, were formed by replacing the inital phonemes of the matched items in INWs with another (e.g., *socabutary*). However, the reaction times for monitoring the target phonemes were not significantly different between the two groups.

Wurm and Samuel (1997) pointed out the possibile problems of Frauenfelder et al.'s experiments. Wurm and Samuel (1997) obtained the similar result by replicating the Experiment 2 in Frauenfelder et al. (1990) with English rather than French stimuli; they argured, INWs might not be lexically processed as expected because Frauenfelder et al. (1990) placed more than half of the targets near middle of the stimuli and subjects might over-concentrate on the specific position. Also, CNWs were not appropriate to serve as the control group because the mapping between acoustic signal and lexical representation was not depend on matching in all-or-none fashion but the goodness-of-fit (Connine et al., 1993a); the lexical activation could not be simply blocked by replacing the inital phonemes, and the CNWs might be performed similarly as INWs.

The later experiments were thus improved by adding a new condition (TNWs, True Non-words) where stimuli were less word-like than CNWs, increasing fillers with various target positions, and taking attentional resource into consideration. It was found that targets in INWs required more processing resources than the ones in TNWs, which was consistent with the prediction, i.e., the lexically activated phoneme would inhibit the others which were actually presented.

Priming paradigm is another approach to explore inhibition within representational levels. In priming paradigms, subjects receive a prime and a following target in each token, and how activations of primes influence on the targets is mainly concerned. For example, phonological priming is adopted to study the competition between lexical representations. Phonological priming is performed in a case of the phonemic overlapped primes and targets. Dufour and Peereman (2003) indicates that inhibitory priming effect occurrs when the prime exclusively differs from the target in the last phoneme, as /bagaʀ/-/bagaʒ/, but not in the condition that the prime and the target are different in the last two phonemes (/bagɛt/-/bagaʒ/). That is, the required time to identify the target phoneme will delay if the preceding prime is phonemic overlapped in great extent.

Resembling the phonological priming, phonetic priming is conducted with the high degree of distinctive feature overlap between primes and targets, and it reveals how the priming activated phonemes affect the target ones. It has been found that both word (e.g., /bak/) and non-word primes (e.g., /bap/), which are phonetic highly similar to the targets (e.g., /dɛt/), delay the reaction time of identifying targets (Dufour and Frauenfelder, 2016).

Previous studies show that the identification of presented phonemes, which are incompatible with the activated or expected one, is delayed in the temporal aspect. In the studies, data that subjects failed to identify the targets (reflected as excessive biases of reaction time) were taken off for research purposes. Nonetheless, the failure data then bring up a puzzle: What do listeners "hear" when the presented speech conflicts with the expectation and the targets are not correctly identified? Moreover, the spatial relation between the representations or phonemes is not specified during the condition either. The current article thus investigates whether lexical knowledge influences the perception of Mandarin initial consonants in white noise. In particular, when a specific phoneme is activated by lexical knowledge, it is concerned how activated phoneme influences the percetpion of other phonemes that are actually presented, and cocurrently, whether distinctive features play a related role.

In light of the inhibition within the representation levels that is predicted by the TRACE model (McClelland and Elman, 1986), the main research questions in this article is illustrated as follows:

[1]Uniqueness point, or recognition point, is considered as the time when a word is recognized.

(1) In white noise, how do a specific phoneme, which is lexically activated, affect the perception of the incompatible and actually presented speech? And how do the particular distinctive features take part in?

(2) Do the perceptual results pattern at the phonemic level? That is, do the phonemes except for the lexically activated one show the same trend in perception?

It should be noticed that several reasons are provided for taking white noise into consideration. Replacing a sound with an unrelative one is likely to cause auditory illusion of the missing sound, mentioned as phoneme restoration, and white noise is therefore added to increase the effect. In order to reveal the internal structures between phonemes (see Miller and Nicely, 1955), white noise is also taken in for increasing the possibility of misperception because speech recorded in laboratory can be easily identified (Johnson, 2004). And, the noise is added to conform to the reality that having a converstion in a "clear" scene without noises is not often in real life.

Another point should be mentioned is, nine Mandarin Chinese onsets are concerned in the experiment for the purpose of investigating the lexical influence on phonemic perception in three dimensions, i.e., place of articulation, aspiration and manner of articulation. The phonemes are /p, p^h, f, t, t^h, s, k, k^h, x/, and are fit in a 3 by 3 matrix in terms of the contrast of the three.

II. Method

a. Subjects

Twenty students (14 males; 6 females) 20-28 years old (M=24; SD=1.97) at universities in Beijing city were paid for their participation in the 30-min experiment. All were native speakers of Mandarin Chinese with no known articulatory and hearing disorders.

b. Materials

The test stimuli, shown as Table 1, consisted of 9 target phonemes, which were /p, p^h, f, t, t^h, s, k, k^h, x/, presenting in a word context and a matched non-word context respectively, like INWs and TNWs in Wurm and Samuel (1997). The carrier word /a.li.pa.pa/ 'Alibaba (corp. name)'was a four-syllable monomorphemic word, which were familiar to people living in China. The matched non-word context mantained the same syllable numbers, syllable structure and tones[2] as the word context. In each stimulus, target phoneme was located at the onset of the fourth syllable, where was the third phoneme after the uniqueness point in the word context group. To avoid coarticulation, all target phonemes existed after vowels.

Also, another 18 target-bearing stimuli were added as fillers. The filler stimili which were constructed as the similar fashion of the test stimuli with different carriers and target position.

All stimuli were recorded in a sound-proof room by a male native speaker of Mandarin Chinese, who had been lived in Beijing for seven years. Recording equiptments include a laptop and an undiectional microphone (Audio-Technica, AT2020USB+); all stimuli were recorded at the sample rate 22050 Hz at 16 bit, and at the speaking rate of one stimulus every 0.9 s. Then, white noise was added to the recorded stimuli (SNR= -6 dB), begining from the preceding vowel of the target phoneme and ending at the following one of the target phoneme.

c. Procedure

Identification tasks with 9 forced-choice technique were conducted by Eprime 3.0 in a quiet room. Speech stimuli were played via a studio headphone (ATH-M20x) at a constant volume. Subjects were told that one of the "OOOX" or "OXOO" sequences would display on the screen in parallel with the speech stimuli. They were required to identify the onset of the syllable coresponding to the visually "X", and response by clicking the button labelled with pin-yin.

Before the experiment, subjects were confirmed to understand the procedure and finish a practice including 25 items. The 30 min task included 6 repetitions of 36 stimuli, and it was divided into two blocks and a 3 min compulsory break was provided.

d. Data analysis

The collected data are analyzed by Multi-dimensional Scaling, Hierarchical Clustering, Signal Detection Theory and Max Entropy Grammar.

Multi-dimensional Scaling is a statistic method according to simiarities or dissimilarties between objects. It is able to converse similarities and distances between multiple objects into the relative direction in lower dimension, and represents the differences of objects by perceptual map. In linguistics, the perceptul similarities between phonemes are caculated and used as the distances for Multi-dimensional Scaling. Based on the confusion matrices in Miller and Nicely (1955), Shepard (1972) proposed the fomula of perceptual similarity, which is:

$$S_{ij} = \frac{P_{ij} + P_{ji}}{P_{ii} + P_{jj}} \tag{1}$$

In (1), S_{ij} refers to the similarity between phoneme *i* and phoneme *j*, and P_{ij} represents the probability of perceiving phoneme *i* as phoneme *j*. Furthermore, the perceptual distances (D_{ij}) are derived by taking negative natural log of perceptual similarities, as (2).

$$D_{ij} = -ln(S_{ij}) \tag{2}$$

Another method is Hierarchical Clustering which also takes similarities or distances as input. Different from Multidimensional Scaling, Hierarchical Clustering reveals the internal structure between objects by classifying them into groups, so called clusters, layer by layer. Hierarchical Clustering works by considering each object as a separate cluster firstly, measuring the distances between clusters, and merging the two "closest" clusters into one. Then, it continues the previous steps (measuring and merging) until all objects are grouped as one cluster. In current article, perceptual distances, which represent spatial relation between phonemes at

[2] Yang and Jin (1988) pointed out that the perception of the stop consonants in Mandarin Chinese could be affected by different tones.

Table 1: Stimuli list

Target phoneme	Word context	Non-word context	Filler (word)	Filler (non-word)
p	/a.li.pa.**p**a/	/i.so.va.**p**a/	/puo.**p**uo.li.kɤ/	/xai.**p**uo.ma.kɤ/
p^h	/a.li.pa.**p^h**a/	/i.so.va.**p^h**a/	/puo.**p^h**uo.li.kɤ/	/xai.**p^h**uo.ma.kɤ/
f	/a.li.pa.**f**a/	/i.so.va.**f**a/	/puo.**f**uo.li.kɤ/	/xai.**f**uo.ma.kɤ/
t	/a.li.pa.**t**a/	/i.so.va.**t**a/	/puo.**t**uo.li.kɤ/	/xai.**t**uo.ma.kɤ/
t^h	/a.li.pa.**t^h**a/	/i.so.va.**t^h**a/	/puo.**t^h**uo.li.kɤ/	/xai.**t^h**uo.ma.kɤ/
s	/a.li.pa.**s**a/	/i.so.va.**s**a/	/puo.**s**uo.li.kɤ/	/xai.**s**uo.ma.kɤ/
k	/a.li.pa.**k**a/	/i.so.va.**k**a/	/puo.**k**uo.li.kɤ/	/xai.**k**uo.ma.kɤ/
k^h	/a.li.pa.**k^h**a/	/i.so.va.**k^h**a/	/puo.**k^h**uo.li.kɤ/	/xai.**k^h**uo.ma.kɤ/
x	/a.li.pa.**x**a/	/i.so.va.**x**a/	/puo.**x**uo.li.kɤ/	/xai.**x**uo.ma.kɤ/

perceptual level, are taken as input for the analysis. Additionally, s distances, which are founded on articulation, are analyzed as comparison, and feature distances are derived like (2) from feature similarities according to the natural classes similarity model employed by Frisch (1996), provided as (3).

$$Similarity = \frac{shared\,natural\,classes}{shared\,natural\,classes + non-shared\,natural\,classes} \quad (3)$$

Signal Detection Theory analysis is widely adopted in Psychophysics and other research fields including speech perception, e.g., Iverson and Kuhl (1995); Żygis and Padgett (2010). Signal Detection Theory analysis focuses on the accuracy of the signal in the information output system, that is to say, it deals with how subjects distinguish the signal from the noise. Following Macmillan and Creelman (2004), measures like detectability (d′) and criterion location (c) are derived from the relationships between the distribution of signal presentation and whether subjects respond correctly (which are *hit*, *miss*, *false alarm*, and *correct rejection* in Table 2). The d′ is used commonly in measuring accuracy without response bias. On the other hand, c is a basic bias measure which is indenpent of d′ in statisticial perspective. Both detectability (d′) and criterion location (c) are considered in this article, and fomulas are given in (4) where H is the hit rate and FA is the false alarm rate respectively.

$$d' = Z_H - Z_{FA}; \quad c = -0.5 * (Z_H + Z_{FA}) \quad (4)$$

Besides, an analysis of Max Entropy Grammar (MatEnt, Goldwater et al., 2003; Hayes, 2020; Hayes and Wilson, 2008) is presented in this article. The MaxEnt is a constraint-based linguistic theory, which adopts logistic regression method in mathematics. As a version of Optimality Theory (OT, Prince et al., 1993), the MaxEnt also comprises the Generator (GEN), the Constraint component (CON) and the Evaluator (EVAL), In the classical OT, the GEN generates a list of candidates from an input, and then the EVAL decides the optimal candidate as the only output by evaluting candidates in the light of the CON, which includes violable constraints and the ranking of them for the competition between candidates. However, various outputs are allowed in considerable conditions like allophones caused by social factors, and they can not be explained by the classical OT. The MaxEnt thus replaces the ranking of constraints with the weights, and the possibilities of outputs are derived by considering the weights and the violated frequencies of constraints; see (5).

Table 2: Stimulus and response for SDT analysis

		Response	
		"Signal present"	"Signal absent"
Stimulus	Signal present	*hit*	*miss*
	Signal absent	*false alarm*	*correct rejection*

hit rate (H) = hits / (hits + misses)
false alarm rate (FA) = false alarms / (false alarms + correct rejections)

$$Pr(x) = \frac{exp(-\sum_i w_i f_i(x))}{Z}, \, where \quad Z = \sum_j exp(-\sum_i w_i f_i(x)) \quad (5)$$

That is to say, the MaxEnt takes all constraints into consideration and it is not limited to a single output. To be brief, the MaxEnt is a mechanism for decision (Hayes, 2020), and the violation of constraints is evidential to the winning of candidates: the more degree of a candidate violating constraints, the less possible it is to be an output. In this article, the actually presented speech in the experiment are taken as inputs of the MaxEnt analysis, the outputs are the responses of subjects and the weights of designed constraints are mainly concerned.

e. Predictions

It is assumpted that the speech which are actually presented in the word context is easier to be perceived as the lexical activated phoneme /p/ or a other phoneme sharing a natural class with /p/, rather than in the non-word context. Noted that the "natural class" analysis is exclusively due to the nine phonemes this article concerned, and the distinctive feature system is according to Hayes (2011). On the basis of the methods mentioned above, the assumption is specified at the phonemic level and distinctive feature level respectively. Comparing with stimuli being in the non-word context, in the word context: if the lexically activated /p/ takes place at the phonemic level, the possibilities of speech being perceived as /p/ will increase (H1-1), the identification rates of phonemes expect for /p/ will decrease (H1-2), the perceptual similarities between /p/ and other phonemes will increase (H1-3), the detectabilities (d′) of the phonemes except for /p/ will decrease (H1-4), and taking /p/ as signal, the criterion locations (c) between /p/ and each of the other phonemes will decrease (H1-5).

On the other hand, if the lexically activated /p/ takes place at the distinctive feature level, the possibilities of speech being perceived as the phonemes that /p/ shares natural classes with will increase (H2-1). the identification rates of the natural classes that /p/ do not share will decrease (H2-2), the perceptual similarities between natural classes that /p/ shares and the other natural classes that /p/ do not share will increase (H2-3), the detectabilities (d') of natural classes that /p/ do not share would descrease (H2-4), and taking natural classes that /p/ shares as signal, the criterion locations (c) between signal and each natural classes that /p/ do not share will decrease (H2-5).

To the preceding predictions, the nine phonemes are grouped into diverse natural classes in accordance with the place of articulation, manner of articulation, aspiration or the combinations of them; the natural classes that /p/ shares and the members of them are listed as follows: [+labial], /p, p^h, f/; [−continuant], /p, p^h, t, t^h, k, k^h/; [−spread gl], /p, f, t, s, k, x/; [+labial, −continuant], /p, p^h/; [+labial, −spread gl], /p, f/; [−continuant, −spread gl], /p, t, k/.

III. Results

The data (with 2155 responses in total) were collected and represented as two confusion matrices in Table 3. The rows were the stimuli and the columns referred to the responses.

a. Identification rate and the /p/ perceived probability

Fig. 1 shows the possibilities of each presented speech being perceived as the phoneme /p/ and the identification rates of the phonemes in the two contexts. First of all, stimuli were more likely to be perceived as the lexically activated phoneme, if there were one, in the word context group. In the word context group, all stimuli were partially perceived as /p/, particularly /t/ (37.82%); diversely, stimuli were less being perceived as /p/ in the non-word context group and most of them (/p^h, t, t^h, s, k^h, x/) were not responded as /p/ at all. Nevertheless, the paired t test showed no significant differences (t=2.049, p=0.075, d= 0.92) of the possibilities being perceived as /p/ between the word context group (M= 0.1059, SD= 0.0490) and the non-word context group (M= 0.0093, SD= 0.0051). Although it was not significant statistically, there was still no reason to accept the null hypothesis that the two were the same.

Secondly, phonemes which shared the same place of articulation with the lexically activated one in the word context were easier to be identified. As the line chart presented, /p, p^h, f, t^h/ remained higher identification rates in the word context group than in the non-word context group, and the identification rates of other phonemes were lower. If /t^h/ was regarded as an exception, it could be concluded that the phonemes with higher identification rates in the word context group shared the place feature [+labial] with the lexically activated /p/.

b. Multi-dimensional Scaling (MDS)

Perceptual maps derived in different dimensions were provided in this section. Fig. 2 and 3 illustrate the spatial relationships of perception between phonemes in the word context group. Shown as the results of the three dimensional analysis in Fig. 2, the phonemes were divided into /p^h, t^h, k^h/ ([+spread gl]) and the others ([−spread gl]) in the x dimension at the margin of about x=-0.75, and thus the x dimension represented the contrast of aspiration. In the y dimension, labial phonemes were separated from non-labial phonemes at the margin of y=0.5, and dorsal stops were distinguished /k, k^h/ ([+dorsal, −continuant]) from the others. Cocurrently, fricatives /x, f, s/ could be grouped in the x-y dimensional space ($x > -1$, $z > 0$). The two dimensional result (see Fig. 3) was similar to the x-y dimensional space in Fig. 2; it referred to the [±spread gl] contrast in the x dimension and the [±labial] contrast in the y dimension. As the one dimensional analysis, also in Fig. 3, [−dorsal, +spread gl] and [+dorsal, −spread gl] were differentiated, yet the one dimension did not represent contrasts of any single distinctive features.

Results of the non-word context group were given as Fig. 4 and 5. As Fig. 4, in the y dimension, labial phonemes were and non-labial phonemes were distinguished at about y=−0.2, and /t^h, k^h/ ([−labial, +spread gl]) were separated from others. To the two dimensional perceptual map in Fig. 5, phonemes were presented as the distribution of [−labial,

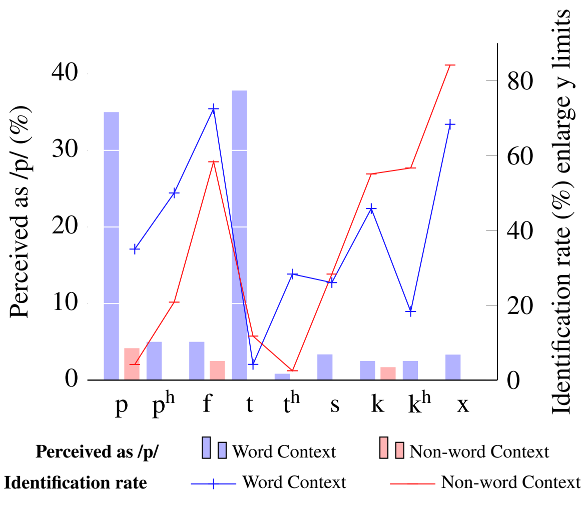


Fig. 1: Perceived possibilities and identification rates

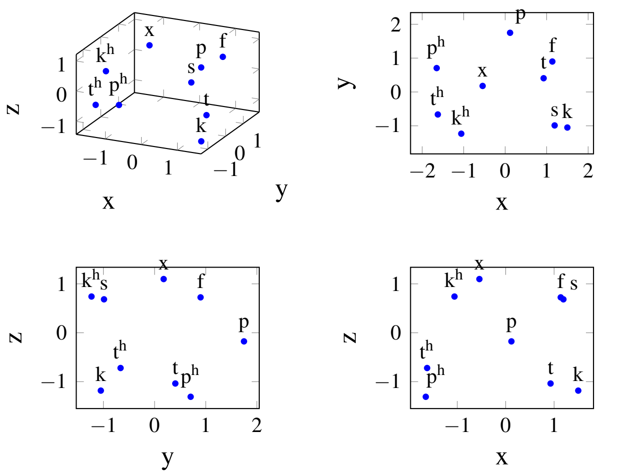


Fig. 2: The three dimensional perceptual map with two dimensional spaces of any two dimensions of the word context group (stress= 0.05, RSQ= 0.96)

Table 3: Confusion matrices of the collected data

	Word context										Non-word context								
	p	pʰ	f	t	tʰ	s	k	kʰ	x		p	pʰ	f	t	tʰ	s	k	kʰ	x
p	42	7	11	5	1	1	1	0	52	p	5	8	50	0	1	5	9	9	33
pʰ	6	60	8	0	5	0	0	2	39	pʰ	0	25	13	0	3	1	2	11	65
f	6	5	87	0	1	2	0	0	19	f	3	0	70	3	3	15	3	6	17
t	45	6	26	5	4	9	16	0	8	t	0	0	13	14	3	18	71	0	0
tʰ	1	60	3	0	34	0	0	10	12	tʰ	0	14	0	0	3	9	0	28	66
s	4	1	43	1	3	31	3	4	29	s	0	0	2	13	6	29	65	0	1
k	3	0	10	11	3	29	55	4	5	k	2	0	2	13	6	29	65	0	1
kʰ	3	17	3	0	9	4	0	22	62	kʰ	0	1	0	0	5	13	3	68	30
x	4	20	13	0	1	0	0	0	82	x	0	8	2	0	0	4	0	5	101

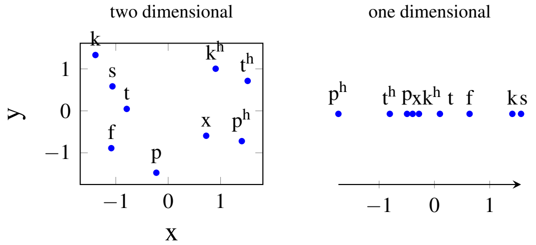


Fig. 3: The two dimensional perceptual map (stress= 0.13, RSQ= 0.86) and the one dimensional perceptual map (stress= 0.38, RSQ= 0.53) of the word context group

+spread gl] (/tʰ, kʰ/), [+labial, −spread gl] (/p, f/), /x, pʰ/ and /k, s, t/. To the result of one dimensional analysis, coronal and dorsal non-aspirated stops ([−continuant, −spread gl]) were distinguished from the others.

To summarize, in the results of one dimensional analysis, there were no contrasts of any single distinctive features in both word context group and non-word context group. In the results of two and three dimensional analysis, both groups contained one dimension representing the contrast of [±labial]. Furthermore, against to the non-word context group, the word context group remained one dimension indicating to the contrast of [±spread gl]. In other words, the distinction between labial and non-labial was perceptually prominent regardless of the word/non-word context, but aspiration was more identifible in the word context group with the lexically activated /p/.

c. Hierarchical Clustering (HC)

Tree diagrams in this section revealed the internal structures of the nine phonemes. Taken as reference, the feature tree (Fig. 6a) showed that the nine phonemes were first divided into two classes, which were /f/ and /pʰ, tʰ, kʰ, p, t, k, s, x/, by the feature [labiodental]. Then the following was dichotomized as /pʰ, tʰ, kʰ/ and /p, t, s, k, x/, relating to [+spread gl] and [−spread gl] respectively. The class with [+spread gl] was further classified as /pʰ, tʰ/ and /kʰ/ in accordance with the place feature [dorsal]; the class with [−spread gl] was sorted into groups, /t, s/ and /k, x, p/, by the place feature [coronal]. At last, /k, x, p/ was separated

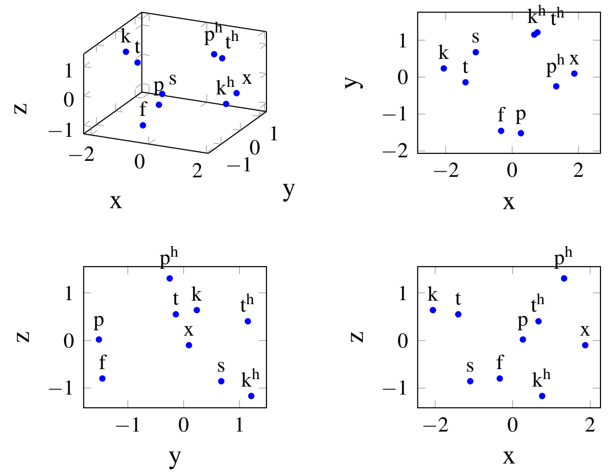


Fig. 4: The three dimensional perceptual map with two dimensional spaces of any two dimensions of the non-word context group (stress= 0.04, RSQ= 0.98)

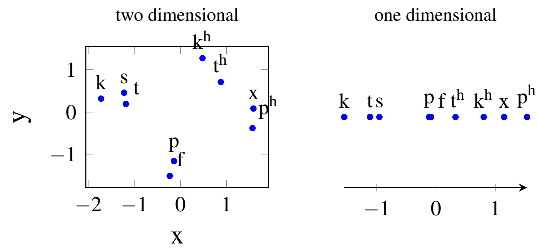
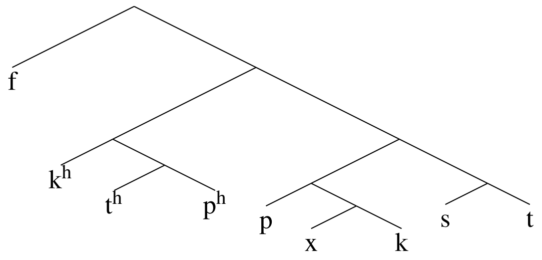
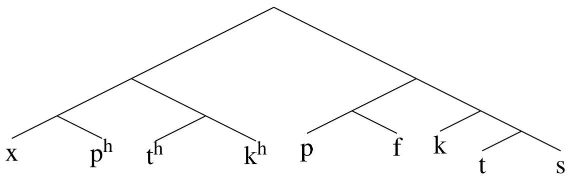


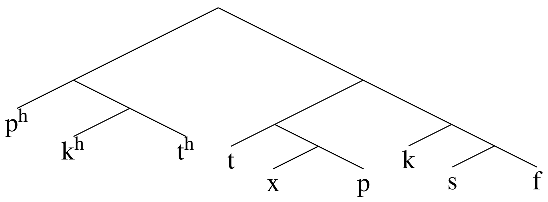
Fig. 5: The two dimensional perceptual map (stress= 0.09, RSQ= 0.94) and the one dimensional perceptual map (stress= 0.27, RSQ= 0.74) of the non-word context group



(a) feature tree



(b) Perceptual tree of the non-word context group



(c) Perceptual tree of the word context group

Fig. 6: Tree diagrams of hierarchical clustering (average linkage)

with the difference of the features, [dorsal] and [labial]. It could be stated that in the feature tree, [labiodental] were used for the clustering in the first layer of the top-down analysis; [spread gl] was used in the second layer; [dorsal] and [coronal] were used independently in the third layer; [dorsal] and [labial] were jointly used in the fourth layer.

The perceptual tree of the non-word context group was given as Fig. 6b. In this tree, /x, p^h, t^h, k^h/ and /p, f, k, t, s/ were separated in the first place. Secondly, the previous class was divided into /x, p^h/ and /t^h, k^h/; the latter one was divided into /p, f/ and /k, t, s/. Afterwards, the /k, t, s/ class was classified as /k/ and /t, s/ according to features, [coronal] and [dorsal]. If /x/ was taken out of consideration, it could be argued that in the non-word context group, [spread gl] was used to classify the phonemes in the first layer; [labial] was used in the second layer; [coronal] and [dorsal] were cooperatively used in the third layer.

In the perceptual tree of the word context goup (Fig. 6c), firstly, nine phonemes were classified into /p^h, t^h, k^h/ and /f, s, k, p, x, t/ on the basis of the feature [spread gl]. Subseqently, the preceding class was dichotomized as /t, x, p/ and /k, s, f/, and the latter one was separated as /p^h/ and /t^h, k^h/ with the difference of [+labial] and [−labial]. Lastly, /f, s, k/ was divided into /k/ and /s, f/ by [dorsal] and [continuant]; /t, x, p/ was classified by [coronal]. Generally, in the word context group, [spread gl] was used for classifing nine phonemes into clusters in the first layer; [labial] was used in the second layer; [coronal] and cooperation of [dorsal] and [continuant] were used respectively in the third layer.

Generally, hierachical clustering results suggested that the laryngeal feature, which represented aspiration in this case, was primarily considered in all three condidtions. The laryngeal feature [spread gl] was primarily used in distinguishing phonemes in all trees, especially in the perceptual ones (yet /x/ in the non-word context group was excluded), and then was the place features; the manner feature [continuant] was less used, which worked with the place feature [dorsal] exclusively in the third layer of the word context group.

In additional, /x/ and /t/, the two closest phonemes to /p/ in Fig. 6c, were the only two phonemes whose similarities to /p/ were greater in the word context group; see Table 4.

d. Signal Detection Theory (SDT)

The SDT analysis was presented in terms of phonemes and natural classes, and two measures, detectability (d′) and criterion (c) were provided. The criterion demonstrated the preference to the confusion of the signal and the noise. If the c values were positive, it indicated that the confusing targets were more responded as noises rather than signals; otherwise, if the values were negative, it indicated that confusing targets were tend to be reponded as signals. Adjustment used for avoiding infinite values in the analysis was to replace hit rates as $1-1/2N$ when they were 1, and to replace false alarm rates as $1/2N$ when they were 0 (Macmillan and Creelman, 2004).

Perspective of phonemes was focused in the first place. It was found that comparing to the non-word context group, in the word context group, most of the nine phonemes were more detectable, and response biases between the lexically activated /p/ and the others were tend to approach the /p/ side. As Table 5 shows, d′ values of most phonemes (except /t, k^h, x/) were greater in the word context group. Also, considering /p/ as the signal and each of the others as noises independently, all c values were smaller in the word context group than in the non-word context group.

In the perspective of natural classes, in each classification, the natural class that /p/ shared was regarded as the signal, and the others were noises independently. For instance, if the phonemes were classified as [−continuant](/p, p^h, t, t^h, k, k^h/) and [+continuant] (/f, s, x/) by the manner feature, the preceding one would be taken as the signal. If the phonemes were classified as [+labial, −coronal, −dorsal] (/p, p^h, f/), [−labial, +coronal, −dorsal] (/t, t^h, s/) and [−labial, −coronal, +dorsal] (/k, k^h, x/) by the place features, the natural class [+labial, −coronal, −dorsal] would be taken as the signal, and the followings would be taken as noises respectively. All data were shown in Table 6. In terms of d′, it should be pointed out that firstly, only few

Table 4: Similarities between /p/ and the other phonemes

Phonemes	Similarities to /p/		
	Word context		Non-word context
p^h	0.1275	<	0.2667
f	0.1318	<	0.7067
t	1.0709	>	0
t^h	0.0263	<	0.125
s	0.0687	<	0.1282
k	0.0412	<	0.1552
k^h	0.0469	<	0.1233
x	0.4516	>	0.3313

Table 5: SDT analysis results (phonemes)

phonemes	d′		phonemes (N)	c	
	word context	non-word context		word context	non-word context
p	1.0531	0.8289		*(SN= /p/)*	
p^h	1.1696	1.0345	p^h	0.1338	1.4658
f	1.7622	1.4667	f	0.3515	1.5466
t	0.3754	0.8491	t	-1.2638	0.6515
t^h	1.3352	-0.0189	t^h	-0.0443	0.8354
s	1.0336	0.7188	s	-0.3933	1.0890
k	1.9313	1.2428	k	-0.1812	1.1245
k^h	1.1332	1.6514	k^h	-0.7316	1.5022
x	1.1966	1.7269	x	0.9067	1.8786

natural classes that /p/ shared were more detectable in the word context group, which were [−continuant] and [−spread gl, −continuant], and most of the natural classes tha /p/ did not share contained smaller d′ values in the word context group. Secondly, confusing targets were preferred to be responded as natural classes that /p/ shared in the word context group. Observing the c values of taking natural classes that /p/ did not share as noises, all of them were smaller in the word context group with three exceptions ([+continuant], [+labial, −coronal, −dorsal, +spread gl], and [−spread gl, +continuant]).

e. Max Entropy Grammar (MaxEnt)

Collected data were fit into the MaxEnt analysis with a key operation, which was the settings of faithfulness constraints and markedness constraints. Faithfulness constraints and markedness constraints were designed with accordance to two extreme cases in the experiment, which were the perfect accuracy of identification and the "zero" accuracy of identification. In the perfect accuracy case, the responses (outputs) should completely match with the stimuli (inputs), and it could be demonstrated by the values of distinctive features. As the results, faithfulness constraints as followings:

Id-[labial]: The output value of the place feature [labial] should be consistent with the input.

Id-[coronal]: The output value of the place feature [coronal] should be consistent with the input.

Id-[dorsal]: The output value of the place feature [dorsal] should be consistent with the input.

Id-[spread gl]: The output value of the laryngeal feature [spread gl] should be consistent with the input.

Id-[continuant]: The output value of the manner feature [labial] should be consistent with the input.

In the other case, the white noise in the stimuli prevented subjects from identifying all phonemes. Therefore, none of the phonemes could be responded, and the markedness constraints were designed as:

*$\mathbf{p}$: Assign a violation mark for each /p/.
*$\mathbf{p^h}$: Assign a violation mark for each /p^h/.
*$\mathbf{f}$: Assign a violation mark for each /f/.
*$\mathbf{t}$: Assign a violation mark for each /t/.
*$\mathbf{t^h}$: Assign a violation mark for each /t^h/.
*$\mathbf{s}$: Assign a violation mark for each /s/.
*$\mathbf{k}$: Assign a violation mark for each /k/.
*$\mathbf{k^h}$: Assign a violation mark for each /k^h/.
*$\mathbf{x}$: Assign a violation mark for each /x/.

Based on the constraints above, two models were constructed corresponding to the word context group and the non-word context group. Within the models, collected data were served as the observations of the outputs; the stimuli were the inputs; the weights of each constraints were assigned by the Slover. The constraints and their weights were given in Table 7.

Two aspects of comparison between the two groups' results were provided. The first was the magnitudes of the weights, and it indicated that among the markedness constraints, *p, *p^h, *f and *t^h were assigned smaller weights in the word context group. Namely, in the word context group, phonemes /p, p^h, f, t^h/ were more possibile to be responded. And among the faithfulness constraints, Id-[labial] was the exclusive one that smaller weight was assigned in the word context group. The orders of the weights, as the second aspect, suggested the importance of the constraints. In the word context group, the constraint with the greatest weight was Id-[spread gl], and then Id-[continuant], Id-[dorsal], Id-[labial], Id-[coronal] in order. The one with the greatest weight in the non-word group was Id-[labial], and then Id-[spread gl], Id-[continuant], Id-[dorsal], Id-[coronal] in order.

It seemed that Id-[continuant] weighed as the second in the word context group and as the third in the non-word context group, and it thus could be drawn that the manner feature was more important than place features in the word context group. However, it should be mentioned that zero or two place features were always violated at once in this case. For example, a labial phoneme stimulus ([+labial, −coronal, −dorsal]) could be responded as a labial, coronal, or dorsal phoneme. If it was responded as a labial phoneme, no values of place features were changed. And if it was responded as a coronal ([−labial, +coronal, −dorsal]) or dorsal phoneme ([−labial, −coronal, +dorsal]), two of the place feature constraints were violated. Consequently, weights of any two place features should be added when measured the weights' order of place features, the laryngeal feature and the manner feature. Afterwards, the weights' orders would be arranged like: In the word context group, LARYNGEAL > $\text{PLACE}_{labial,\ dorsal}$ > $\text{PLACE}_{coronal,\ dorsal}$ > MANNER > $\text{PLACE}_{labial,\ coronal}$, and in the non-word context group, $\text{PLACE}_{labial,\ dorsal}$ > $\text{PLACE}_{labial,\ coronal}$ > LARYNGEAL > MANNER > $\text{PLACE}_{coronal,\ dorsal}$.

By regarding the orders, it could be seen that the contrast of the aspiration was more considerable in the word con-

Table 6: SDT analysis results (natural classes)

natural classes	d′ word context	d′ non-word context	natural classes (N)	c word context	c non-word context
Classification according to the place features					
[+labial, −coronal, −dorsal]	0.7157	1.3210	*(SN= [+labial, −coronal, −dorsal])*		
[−labial, +coronal, −dorsal]	0.5849	0.3962	[−labial, +coronal, −dorsal]	-1.0149	-0.317
[−labial, −coronal, +dorsal]	0.9635	0.6501	[−labial, −coronal, +dorsal]	0.1283	0.7762
Classification according to the laryngeal feature					
[−spread gl]	1.1065	1.5345	*(SN= [−spread gl])*		
[+spread gl]	1.6565	1.2269	[+spread gl]	-0.8282	-0.6135
Classification according to the manner feature					
[−continuant]	1.3202	0.7556	*(SN= [−continuant])*		
[+continuant]	1.3202	0.7556	[+continuant]	0.3865	0.3411
Classification according to the place features and the laryngeal feature					
[+labial, −coronal, −dorsal, −spread gl]	1.0980	1.6208	*(SN= [+labial, −coronal, −dorsal, −spread gl])*		
[−labial, +coronal, −dorsal, −spread gl]	0.6731	0.6827	[−labial, +coronal, −dorsal, −spread gl]	-1.1038	-0.2800
[−labial, −coronal, +dorsal, −spread gl]	0.7923	0.7384	[−labial, −coronal, +dorsal, −spread gl]	0.2488	0.6830
[+labial, −coronal, −dorsal, +spread gl]	1.1695	1.0345	[+labial, −coronal, −dorsal, +spread gl]	-0.2760	-0.5790
[−labial, +coronal, −dorsal, +spread gl]	1.3352	-0.01887	[−labial, +coronal, −dorsal, +spread gl]	-0.4795	0.3809
[−labial, −coronal, +dorsal, +spread gl]	1.1332	1.6514	[−labial, −coronal, +dorsal, +spread gl]	-0.9233	0.6921
Classification according to the place features and the manner feature					
[+labial, −coronal, −dorsal, −continuant]	0.7715	0.8310	*(SN= [+labial, −coronal, −dorsal, −continuant])*		
[−labial, +coronal, −dorsal, −continuant]	0.7498	0.2630	[−labial, +coronal, −dorsal, −continuant]	-0.9740	-0.5431
[−labial, −coronal, +dorsal, −continuant]	1.2958	0.9306	[−labial, −coronal, +dorsal, −continuant]	-0.5922	0.9473
[+labial, −coronal, −dorsal, +continuant]	1.7596	1.4668	[+labial, −coronal, −dorsal, +continuant]	0.0712	1.02674
[−labial, +coronal, −dorsal, +continuant]	0.7498	0.2630	[−labial, +coronal, −dorsal, +continuant]	-0.2594	0.7707
[−labial, −coronal, +dorsal, +continuant]	1.1934	1.7269	[−labial, −coronal, +dorsal, +continuant]	0.3021	1.0178
Classification according to the laryngeal feature and the manner feature					
[−spread gl, −continuant]	1.7875	1.4142	*(SN= [−spread gl, −continuant])*		
[+spread gl, −continuant]	1.6565	1.2269	[+spread gl, −continuant]	0.2678	0.3748
[−spread gl, +continuant]	1.3202	0.7556	[−spread gl, +continuant]	0.7364	0.4460

text group, and the manner feature was less distinguished in both groups. The previous one was possibly due to the lexically activated /p/ or the “lexical context”, and a potential reason for the latter was the influence of white noise in the two groups.

f. Comparing to predictions

In the current section, all predictions proposed in section II:e were examined at the phonemic level and distinctive feature level. Predictions were restated as (with the lexically activated /p/ in the word context group, and in the word context group:) At the phonemic level, the possibilities of speech being perceived as /p/ would be greater (H1-1); the identification rates of phonemes expect for /p/ would be smaller (H1-2); the perceptual similarities between /p/ and other phonemes would be greater (H1-3); the d′ value of the phonemes excluding /p/ would be smaller (H1-4); the c taking /p/ as signal and each of the phonemes as noises would smaller (H1-5). And at the distinctive feature level, the possibilities of speech being perceived as the phonemes that /p/ shared natural classes would be greater (H2-1); the identification rates of the natural classes that /p/ did not share would be smaller (H2-2); the perceptual similarities between natural classes that /p/ shared and /p/ did not share would be greater (H2-3); the d′ value of natural classes that /p/ did not share would be smaller (H2-4); the c taking natural classes /p/ shared as signals and each of the natural classes /p/ did not share as noises would be smaller (H2-5).

Results were shown in Table 8 and 9. In the Tables, the ones conformed to the predictions were marked as “✓”, and the not consistent ones were marked as “×”. At the phonemic level, predictions that best matched with the results were H1-1 (speech would be more perceived as /p/ in the word context group) and H1-5 (smaller c values in the word context group) and then H1-2 (smaller identification rate except for /p/ in the word context group), H1-4 (smaller d′ values of phonemes except for /p/ in the word context group), H1-3 (greater perceptual similarities to /p/ in the word context group) in order. Otherwise, from the point of each phonemes, the degree of conformity with the predictions was in order as /t/ > /k^h, x/ > /s, k/ > /p^h, f, t^h/.

Similarly, H2-1 (natural classes that /p/ did not share were more perceived as the /p/ shared ones in the word context group) and H2-5 (smaller c values in the word context group) were the best predictions conformed to the results at the distictive feature level. And the second were H2-2 (identification rates of natural classes that /p/ did not share were smaller in the word context group) and H2-3 (perceptual similarities to the /p/ shared natural classes were greater in the word context group). The last was H2-4 (smaller d′ values of natural classes that /p/ did not share were smaller in the word context group). Alternately, in the aspect of each classifications according to diverse distinctive features, the degrees of confomity with the predictions were in order as (classification according to:) PLACE at 80% ($= 8/10$) > PLACE,LARYNGEAL at 72% ($= 18/25$) > PLACE,MANNER at 64%($= 16/25$) > LARYNGEAL,MANNER = LARYNGEAL at 20% > MANNER at 0.

Table 7: The MaxEnt analysis results

Non-word context group (log-likelihood=-15.0787)				word context group (log-likelihood=-14.9986)			
faithfulness		markedness		faithfulness		markedness	
constraints	weights	constraints	weights	constraints	weights	constraints	weights
Id-[labial]	1.0634	*p	3.6404	Id-[labial]	0.2891	*p	1.3642
Id-[coronal]	0.1046	*p^h	1.9264	Id-[coronal]	0.2581	*p^h	0.9745
Id-[dorsal]	0.1843	*f	0.6859	Id-[dorsal]	0.6585	*f	0.5322
Id-[spread gl]	0.9458	*t	2.4100	Id-[spread gl]	1.2067	*t	3.0569
Id-[continuant]	0.3964	*t^h	2.6202	Id-[continuant]	0.8019	*t^h	2.0868
		*s	0.9700			*s	1.5656
		*k	0.6465			*k	1.6943
		*k^h	1.0250			*k^h	2.3753
		*x	0.0101			*x	0.0101

Table 8: Predictions examination (phonemic level)

Phonemes	H1-1	H1-2	H1-3	H1-4	H1-5
p^h	✓	×	×	×	✓
f	✓	×	×	×	✓
t	✓	✓	✓	✓	✓
t^h	✓	×	×	×	✓
s	✓	✓	×	×	✓
k	✓	✓	×	×	✓
k^h	✓	✓	×	✓	✓
x	✓	✓	×	✓	✓

Table 9: Predictions examination (distinctive feature level)

Classification according to the place features					
natural classes	H2-1	H2-2	H2-3	H2-4	H2-5
[−labial, +coronal, −dorsal]	✓	✓	✓	×	✓
[−labial, −coronal, +dorsal]	✓	✓	✓	×	✓
Classification according to the laryngeal feature					
natural classes	H2-1	H2-2	H2-3	H2-4	H2-5
[+spread gl]	×	×	×	×	✓
Classification according to the manner feature					
natural classes	H2-1	H2-2	H2-3	H2-4	H2-5
[+continuant]	×	×	×	×	×
Classification according to the place features and the laryngeal feature					
natural classes	H2-1	H2-2	H2-3	H2-4	H2-5
[−labial, +coronal, −dorsa, −spread gll]	✓	✓	✓	✓	✓
[−labial, −coronal, +dorsal, −spread gl]	✓	✓	✓	×	✓
[+labial, −coronal, −dorsal, +spread gl]	✓	×	×	×	×
[−labial, +coronal, −dorsa, +spread gll]	✓	×	✓	×	✓
[−labial, −coronal, +dorsal, +spread gl]	✓	✓	✓	✓	✓
Classification according to the place features and the manner feature					
natural classes	H2-1	H2-2	H2-3	H2-4	H2-5
[−labial, +coronal, −dorsa, −continuant]	✓	×	✓	×	✓
[−labial, −coronal, +dorsal, −continuant]	✓	✓	×	×	✓
[+labial, −coronal, −dorsal, +continuant]	✓	×	×	×	✓
[−labial, +coronal, −dorsa, +continuant]	✓	✓	×	×	✓
[−labial, −coronal, +dorsal, +continuant]	✓	✓	✓	✓	✓
Classification according to the laryngeal feature and the manner feature					
natural classes	H2-1	H2-2	H2-3	H2-4	H2-5
[+spread gl, −continuant]	✓	×	×	×	✓
[−spread gl, +continuant]	×	×	×	×	×

IV. Discussion and Conclusion

Back to the research questions of the current article:

(1) In white noise, how do a specific phoneme, which is lexically activated, affect the perception of the incompatible and actually presented speech? And how do the particular distinctive features take part in?

(2) Do the perceptual results pattern at the phonemic level? That is, do the phonemes except for the lexically acti-

vated one show the same trend in perception?

For the first question, it is found that comparing to the perceptual results in the matched non-word context, when the phoneme /p/ is activated by the word knowledge (i.e., /a.li.pa.pa/ 'Alibaba'):

- The other phonemes except for the lexically activated one are more likely to be misperceived as the lexically activated one.
- In the confusion with the lexically activated phoneme, the perceptual results of the other phonemes are tend to approach the side of lexically activated one.
- The identification rates of most of the other phonemes decrease in contradiction to /t/ and /p^h, f/, which share the place of articulation with the lexically activated one.

However, the results of perceptual similarities are highly inconsistent with the prediction that perceptual similarities between lexically activated phoneme and each of the others will increase. One possibile account is, high perceptual similarities in non-word context are caused by the white noise (SNR= -6 dB) that increases the misperception of /p/ in large degree; on the other hand, lexical knowledge takes part in and highly decreases the misperception of the lexically activated /p/ but slightly increases the possibilities that other phonemes are perceived as /p/, and finally leads to lower perceptual similarities in the word context group.

In addition, over half of the d' results are opposed to the prediction that phonemes excluding the lexically activated one will be less detectable in the word context. In d' results, /t, k^h, x/ are the merely ones which conform to the prediction, and nevertheless, it is lack of reasonable explanation in this article.

And at the distintive feature level, it is indicated that the laryngeal feature and the place features are relatively considerable when the phoneme /p/ is lexically activated:

- The inhibition between the natural classes that lexically activated phoneme shares and the others, represented as the degree of the conformity to the predictions, is most prominent in the classification according to the place features, and then the laryngeal feature, the manner feature in order.
- In the MDS results, the contrast of [$\pm$labial] and the contrast [$\pm$spread gl] are referred to the two separated dimensions respectively with the latter contrast not presented in the non-word context.
- According to HC analysis, the laryngeal feature is adopted for the distictions in the higher layer than the place features in the perceptual tree.
- MaxEnt analysis shows that the value of laryngeal feature is the least violable one in accordance with the highest weight of Id-[spread gl].

As for the second quesion, it is hard to argue that inhibitory lexical effect occurs at the phonemic level but not distinctive feature level. Indeed, in terms of the possibilities of other phonemes except for /p/ being perceived as /p/ (H1-1) and the c values (H1-5), perceptual results of all phonemes excluding /p/ act in same patterns, yet the results of identification rates, perceptual similarities and d' values are not. However, it seems to be the same at the disdinctive feature level.

Although the current article can not prove where inhibition takes place, the phonemic or the distinctive feature level, it can be proposed where the inhibitory lexical effect is better observed. It should be noticed that the phonemic classification, in this case, is able to be seen as a classification according to the place features, the laryngeal feature and the manner feature. Comparing the phonemic classification to the other provided classifications at the distinctive feature level, the degree of the conformity to the predictions in the phonemic classification is 62.5%, and it is in the order of the classifications as: PLACE > PLACE, LARYNGEAL > PLACE, MANNER > phonemic classification > LARYNGEAL, MANNER = LARYNGEAL > MANNER. And the order indicates that the inhibitory lexical effect is better observed at the distinctive feature level.

In conclusion, the current article illustrates, when the actually presented speech is incompatible with the lexically driven expectation, lexical knowledge facilitates listeners to misperceive other phonemes as the lexically activated one; the aspiration and the place of articulation are relatively influential in the case. Furthermore, the inhibition within prelexical levels, known as the inhibitory lexical effect, is preferably observed at the distinctive feature level.

REFERENCES

Baars, B. J. and Gage, N. M. (2010). Chapter 11 - language. In Baars, B. J. and Gage, N. M., editors, *Cognition, Brain, and Consciousness (Second Edition)*, pages 370–396. Academic Press, London, second edition edition.

Connine, C. M., Blasko, D. G., and Titone, D. (1993a). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32(2):193–210.

Connine, C. M. and Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human perception and performance*, 13(2):291.

Connine, C. M., Titone, D., and Wang, J. (1993b). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(1):81.

Crick, F. and Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. *Parallel distributed processing*, 2:333–371.

Cutler, A. and Norris, D. (1979). Monitoring sentence comprehension.

Dufour, S. and Frauenfelder, U. H. (2016). Inhibitory phonetic priming: Where does the effect come from? *Quarterly Journal of Experimental Psychology*, 69(1):180–196.

Dufour, S. and Peereman, R. (2003). Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition*, 31(8):1271–1283.

Frauenfelder, U. H., Segui, J., and Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, 16(1):77.

Frisch, S. (1996). *Similarity and frequency in phonology*. Northwestern University.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental psychology: Human perception and performance*, 6(1):110.

Goldwater, S., Johnson, M., Spenader, J., Eriksson, A., and Dahl, Ö. (2003). Learning ot constraint rankings using a maximum entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory. pp*, volume 111, page 120.

Hayes, B. (2011). *Introductory phonology*, volume 32. John Wiley & Sons.

Hayes, B. (2020). Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Ms. UCLA*.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, 39(3):379–440.

Iverson, P. and Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, 97(1):553–562.

Johnson, K. (2004). Acoustic and auditory phonetics. *Phonetica*, 61(1):56–58.

Kingston, J., Levy, J., Rysling, A., and Staub, A. (2016). Eye movement evidence for an immediate ganong effect. *Journal of experimental psychology: Human perception and performance*, 42(12):1969.

Macmillan, N. A. and Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

McClelland, J. L. and Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, 18(1):1–86.

McClelland, J. L., Mirman, D., and Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in cognitive sciences*, 10(8):363–369.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *The Journal of the Acoustical Society of America*, 27(2):338–352.

Prince, A., Smolensky, P., and Prince, C. A. (1993). Optimality theory 3.

Shen, L. and Politzer-Ahles, S. (2018). Analysis of the influence of word frequency in auditory perception.

Shepard, R. N. (1972). Psychological representation of speech sounds. In David, E. E. and Denes, P. B., editors, *Human Conmmunication: A unified view*, pages 67–113. McGraw-Hill, New York.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917):392–393.

Warren, R. M. and Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, 9(3):358–362.

Wurm, L. H. and Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language*, 36(2):165–187.

Yang, Y.-F. and Jin, L.-J. (1988). Stop consonats and tone perception. *Acta Psychologica Sinica*, 20(3):14–20.

Żygis, M. and Padgett, J. (2010). A perceptual study of polish fricatives, and its implications for historical sound change. *Journal of Phonetics*, 38(2):207–226.