

# Examining the Lexical Effect on Categorical Perception of Stops in Taiwan Southern Min

Yunglin Tai

Institute of Linguistics

National Chung Cheng University

[yunglin2@gmail.com](mailto:yunglin2@gmail.com)

## Abstract

The goal of this study is to examine if there is a word superiority effect on perception of three-way contrast of stops in Taiwan Southern Min (TSM). Based on Ganong's (1980) findings that English participants showed a significant lexical effect in phonetic perception, I hypothesize that there exists a difference in perception between real words and nonwords in TSM. The prediction is that the categorical boundary shifts as lexical status plays a role in perception of stops. Experiment 1 was conducted as a neutral set of perception of TSM bilabial stops *b-p-p<sup>h</sup>*. In experiment 2, cases of "nonword-word-nonword" set along the *b-p-p<sup>h</sup>* continuum were conducted. Results showed that real words corresponded to a wider range of VOT in the continuum compared to the neutral pattern. The categorical boundaries (both between /b/ and /p/ and between /p/ and /p<sup>h</sup>/) were found to shift away from the real word sides towards the nonword sides. The lexical effect may be explained by parallel processing in which a higher level of processing (lexical level) interacts with a lower level (phonemic level) in speech perception.

Keywords: perception, categorical boundary, VOT, lexical effect, processing

## 1. Introduction

Categorical perception, which represents that sounds within a phonemic category are perceived as indistinguishable regardless of the correct identification of each sound, supports the modular view that speech is perceptually special. In literature, consonants were found to be perceived categorically, but not gradual; that is, although listeners were able to distinguish the consonant stimuli between different phonetic categories with ease, mostly they failed to distinguish the stimuli within categories.

Sometimes the perceptual difference between stops attributes only to the initial voicing feature. For instance, the difference between [b], [p] and [p<sup>h</sup>] is due to voice onset time (VOT): the interval between the stop release and the beginning of vocal cord vibrations. A

positive VOT value means such a lag exists (e.g. the aspirated [p<sup>h</sup>]); a VOT value of zero represents no delay in voicing; a negative VOT value refers to the phenomenon where the vocal fold vibrations begin before the articulatory release of the stop (the prevoiced [b]). As VOT varies gradually, the identification changes abruptly from one stop to another. This categorical perception drives how the modular view regards speech—as a modular system.

However, this raises an interesting question—is it modular in such a way that consonants are always perceived without the affection by other processing information?

There are some parallel models proposed for linguistic processing. One of those is the Trace model of speech perception (McClelland & Elman 1986). The model assumes that different levels of processing—features, phonemes, and words—are activated simultaneously during speech perception, which contradicts with the modularity view that phonemic processing is unaffected by higher levels of processing.

Ganong (1980) investigated whether auditory word perception affected phonetic categorization. He constructed acoustic continua varying in voice onset time, each with an end being a real word and the other end a nonword (e.g., *dash—tash* vs. *dask—task*). The results showed a categorical boundary shift. Thus he argued that the lexical effect must arise at a processing stage sensitive to both lexical and auditory information.

Based on previous studies on categorical perception and Ganong's (1980) findings on the lexical effect that English participants showed a significant lexical effect in phonetic perception, it is hypothesized that there exists a lexical effect—a difference in perception between real words and nonwords. The present study examines a language of a three-way contrast in initial stops, Taiwan Southern Min, to see if there is a word superiority effect on speech perception. The prediction is that real words will correspond to a wider range of VOT in the continuum either comparing to the neutral pattern or to the nonword situation. The logic behind it is that listeners tend to perceive an ambiguous sound as a real word rather than a nonword. If the prediction is true, it should be found that the phoneme boundaries (both between /b/ and /p/ and between /p/ and /p<sup>h</sup>/) shift away from the real word sides towards the nonword sides.

The current research questions are:

- (1) Where is the categorical boundary between *b* and *p* and that between *p* and *p<sup>h</sup>* in Taiwan Southern Min (TSM)?
- (2) Does lexical status (a real word or nonword) of a sound sequence affect the perception of TSM stops? If it does, how does the categorical boundary shift?
- (3) Categorical perception may be better explained by the modular theory or parallel processing models?

The pinyin system (including consonants, vowels and tones) adopted in this paper follows “台灣閩南語音標系統” released by Ministry of Education in 1998. Example lexemes were obtained through *臺灣閩南語辭典* by 董 et al. (2001).

Table 1. Tones used in the current study

Tone category	Tone value	Example
陰上	53	飽 /pa53/ ‘full’
陰去	11	富 /pu11/ ‘rich’
陽去	33	密 /ba33/ ‘closely’

## 2. Experiment 1—identification of a neutral set

Experiment 1 is an identification task designed to test the categorical boundaries of bilabial stops *b*, *p*, and *p<sup>h</sup>* in TSM. Serving as a neutral set, stimuli were not informed to participants beforehand about their lexical or non-lexical status in TSM. Rather, they were instructed to pay attention to any change of the consonant category only.

### 2.1 Methods

#### 2.1.1 Participants

Ten native speakers of TSM participated. All of them were graduate students from National Chung Cheng University, with self-reported fluency in TSM.

#### 2.1.2 Materials

21 stimuli from the bilabial neutral continuum pair *ba—pa—p<sup>h</sup>a*, with VOTs of -100, -90, -80, -70, -60, -50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 msec manipulated in Praat (Boersma & Weenink 2007). Voice onset time (VOT) in msec of each syllable-initial stop was measured from the release of the stop closure to the beginning of the oscillating line which demonstrates voicing in the following vowel.

These stimuli are adjusted from a TSM male speaker’s natural speech. Among the 21 stimuli, all attributes of sounds are identical except the VOT in syllable initials.

The manipulation of prevoicing or aspiration is through deletion or addition from a range of repetitive circular noise lines. One thing to be careful is that the beginning and ending point in selection must match each other in terms of their amplitude position. Otherwise, low pitch noise or unnatural burst would be created during manipulation.

Here is what some examples of stimuli look like:

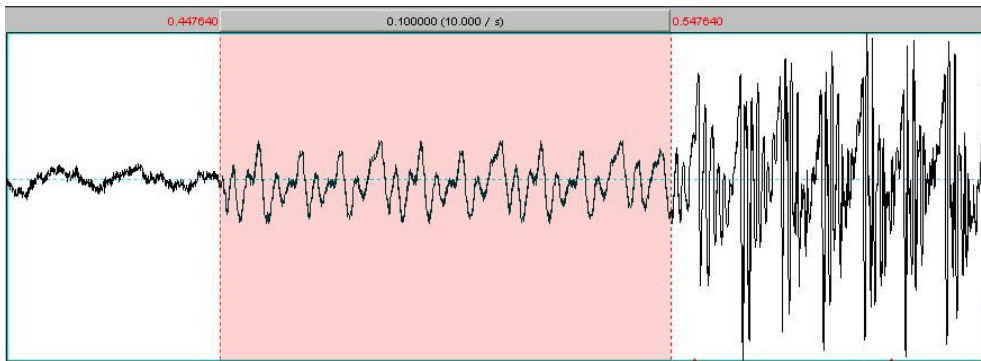


Figure 1. A Stimulus with -100ms in VOT (prevoicing)

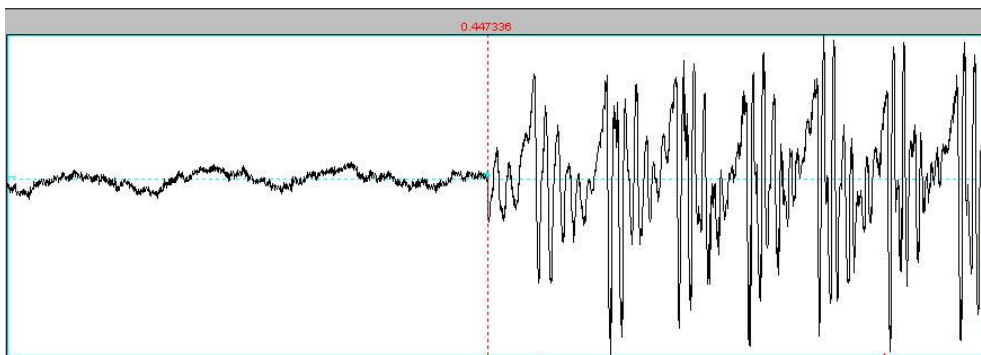


Figure 2. A Stimulus with 0ms in VOT

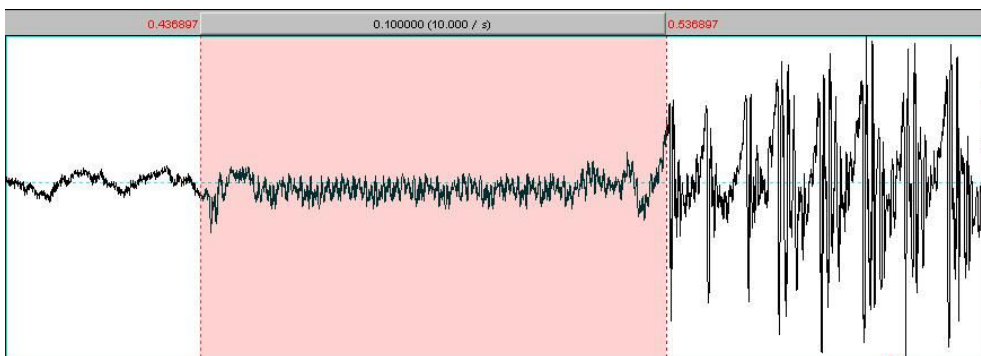


Figure 3. A Stimulus with 100ms in VOT (aspiration)

### 2.1.3 Procedure

Stimuli were saved as a WAV file and played in a notebook computer. Participants were given a piece of paper where there are three columns (*ba*, *pa*, and *pha*) along the 21 stimuli. They are instructed to judge whether the CV syllable they are listening begins with *b*, *p* or *p<sup>h</sup>* and then fill in the corresponding column with a check.

## 2.2 Results and discussion

Stimuli and the corresponding VOT values are shown below:

Table 2. Stimuli & Corresponding VOT values

Stimulus number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
VOT (ms)	-100	-90	-80	-70	-60	-50	-40	-30	-20	-10	0	10	20	30	40	50	60	70	80	90	100

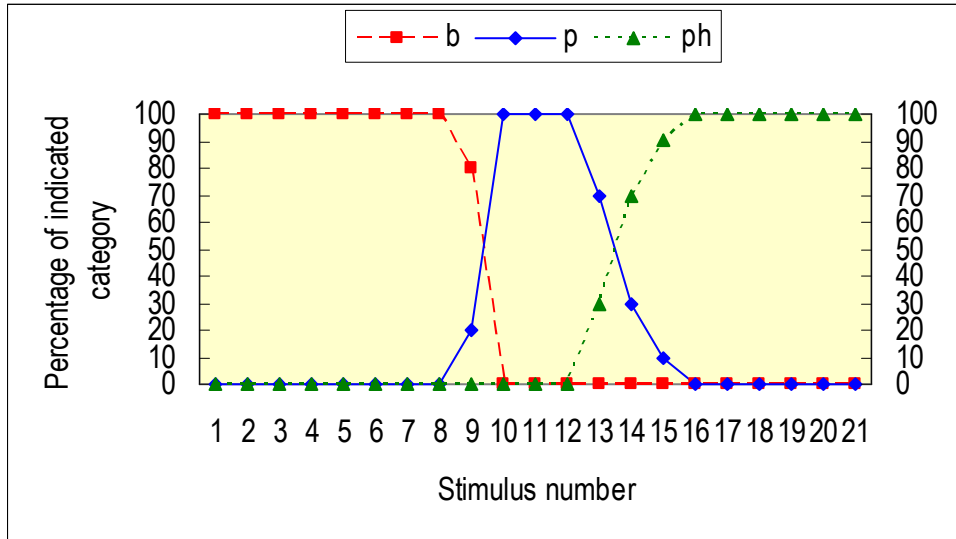


Figure 4. Perception of the Neutral Set

Inter-participant variation (standard deviation) is 0.422 for *b*, 1.16 for *p* and 0.994 for *ph*.

As for the categorical boundaries, their locations are to refer to the intersection points of each line. Thus, according to the graph above, the categorical boundary is -15ms between *b* and *p* and 25ms between *p* and *ph*.

### 3. Experiment 2—identification of nonword-word-nonword sets

Experiment 2 contains two identification tasks conducted to examine if there is a difference comparing with the result in Experiment 1, by adding a linguistic variable: the lexical status of a sound sequence along a continuum. In this experiment, combination of “nonword-word-nonword”<sup>1</sup> is chosen to examine this effect.

#### 3.1 Methods

##### 3.1.1 Participants

Ten native speakers of TSM (different from the participants in Experiment 1) participated. All of them were graduate students from National Chung Cheng University, with self-report fluency in TSM.

### 3.1.2 Materials

There are two groups<sup>2</sup> of stimuli used. The first group is 21 stimuli from the bilabial pair “*ba53 - pa53* 飽'full' - *pha53*”; the second group is 21 stimuli from the bilabial pair “*bu11 - pu11* 富 'rich' - *phu11*”. VOTs are -100, -90, -80, -70, -60, -50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 msec along the continuum respectively.

### 3.1.3 Procedure

Materials were in WAV file format and played in a notebook computer. All the participants do both groups of stimuli. This experiment was conducted with a counterbalance design in order. Half of the participants did the first group of stimuli first, and the other half did the second group first. Participants were given a piece of paper where there were three columns of *ba53*, *pa53* 飽 and *pha53* along the 21 stimuli and another piece of paper where there are three columns *bu11*, *pu11* 富 and *phu11* along the 21 stimuli (Or the reversed order). They were instructed to judge whether the CV syllable they were listening belonged to *ba53*, *pa53* 飽 or *pha53* and to *bu11*, *pu11* 富 or *phu11*, in order to fill in the corresponding column with a check.

## 3.2 Results and discussion

### 3.2.1 The “*ba53 - pa53* 飽'full' - *pha53*” set

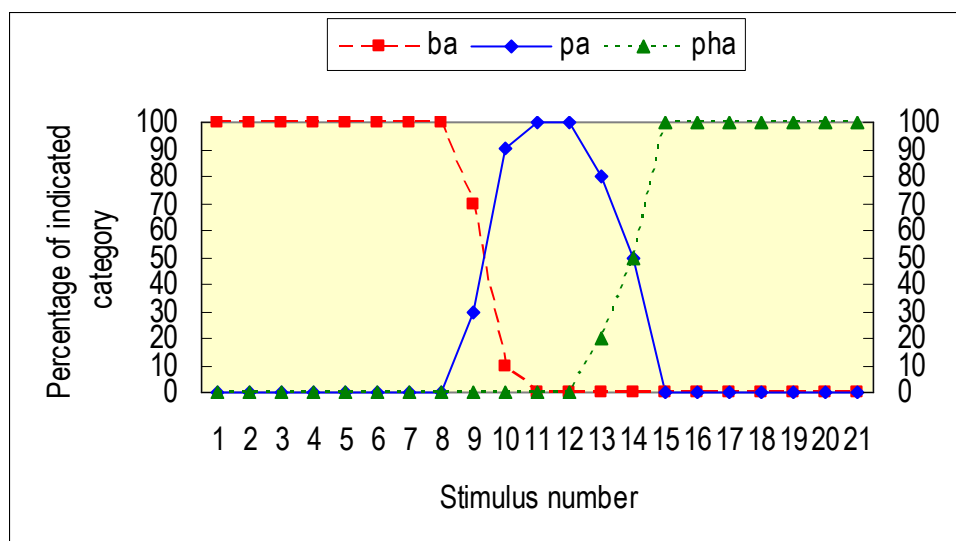


Figure 5. Perception of the “*ba53 - pa53* 飽 'full' - *pha53*” set

Inter-participant variation (standard deviation) is 0.632 for *ba*, 1.08 for *pa* and 0.823 for *pha*.

As for the categorical boundaries, their locations are to refer to the intersection points of each line. According to each of the solution of two formulas in mathematics, the number of intersection points in x axis is 28/3 and 14 respectively. Thus, after conversion to the corresponding VOTs, the categorical boundary is -16.667ms between *ba* and *pa* and 30ms between *pa* and *pha*.

We can see that comparing to those of the neutral set (-15ms and 25ms), the two categorical boundaries of this nonword-word-nonword set shift towards the nonword sides. Hence, the real word *pa53* 飽'full' corresponds to a wider range of VOT in the continuum comparing to the neutral pattern.

### 3.2.2 The “*bull* - *pull* 富 'rich' - *phull*” set

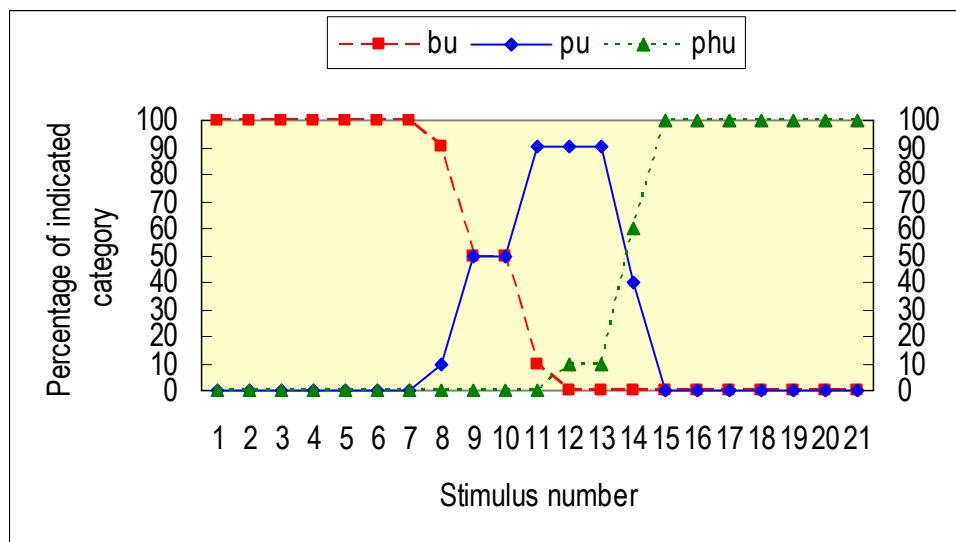


Figure 5. Perception of the “*bull* - *pull* 富 'rich' - *phull*” set

Inter-participant variation (standard deviation) is 1.333 for *bu*, 1.033 for *pu* and 0.919 for *phu*. Note that the variation of *bu* is higher than both *b* in the neutral set and *ba*. One of the possible factors may be due to the more marked CV co-occurrence patterns<sup>3</sup> (Davis & MacNeilage 1995) of *bu* than *ba*, which further affects the salience in speech perception.

As for the categorical boundaries, the number of the intersection points of each line in x axis is 9 to 10 (the common part is actually a line segment) and 13.8 respectively. Thus, after conversion to the corresponding VOTs, the categorical boundary is -15ms between *bu* and *pu* and 28ms between *pu* and *phu*.

Comparing to those of the neutral set (-15ms and 25ms), the categorical boundary between *pu* and *phu* in “*bull* - *pull* 富 'rich' - *phull*” set shift towards the nonword sides. But it does not show the leftward spreading clearly in this case. However, the real word *pull*

富 'rich' also corresponds to a wider range of VOT in the continuum compared to the neutral pattern.

#### 4. General discussion & Conclusion

To compare the three sets done in Experiment 1 and Experiment 2, see the following graphs:

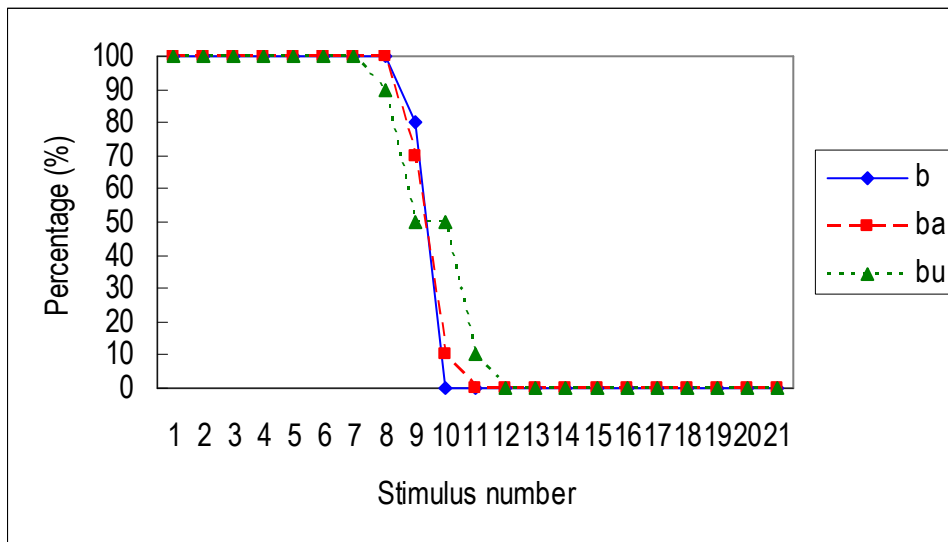


Figure 6. Comparison of the neutral /b/ with *b* in /ba/ and /bu/

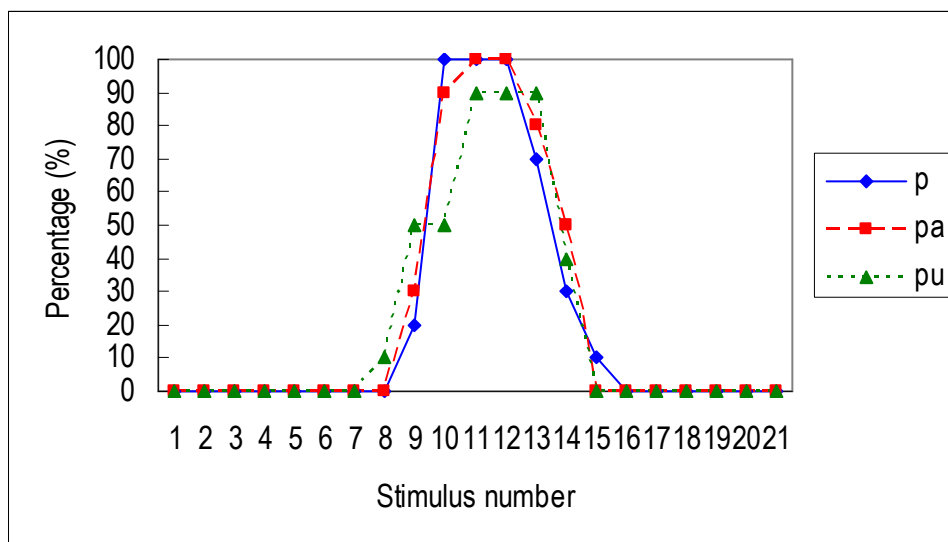


Figure 7. Comparison of the neutral /p/ with *p* in /pa/ and /pu/



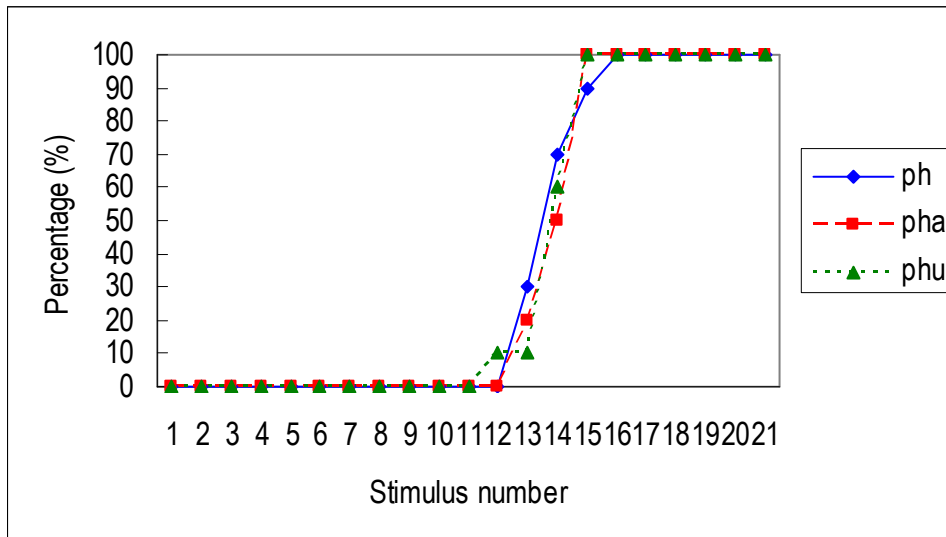


Figure 8. Comparison of the neutral /p<sup>h</sup>/ with p<sup>h</sup> in /p<sup>h</sup>a/ and /p<sup>h</sup>u/

If you see the broken line representing *ba* in the first graph, the location is a little bit left to the solid line representing the neutral pattern. For the *pa* line in the second graph, the location is outside the neutral solid line. In terms of the *pha* line in the third graph, it locates obviously at the right side of the neutral line. All of these consistent patterns suggest that there exists a lexical effect from the word *pa53* 飽'full'—its lexical status affects the perception of consonants, making categorical boundaries shift towards the nonword sides.

As for the dotted line representing *bu* in the first graph and *pu* in the second graph, the categorical boundary does not shift clearly; that is, the lexical effect is not found in that the categorical boundary does not spread leftward. However, *phu* in the third graph shows a similar pattern with *pha*, indicating a shift of the categorical boundary between *p* and *ph*. These suggest there is still a lexical effect found in the real word *pull* 富 'rich'.

However, since the results of the two sets in Experiment 2 show some difference themselves, there may be other factors other than only the lexical status to affect perception of consonants. Newman et. al. (1997) examined the lexical neighborhood effect and found that it affected word recognition in much the same way as lexical status. Difference in CV combinations, tone or neighborhood density of a possible word may be taken into account.

All in all, results in the present study show the pattern as what is predicted—real words correspond to a wider range of VOT in the continuum comparing to the neutral pattern. The categorical boundaries (both between /b/ and /p/ and between /p/ and /p<sup>h</sup>/) are found to shift away from the real word sides towards the nonword sides. Besides, the phenomenon that the lexical status affects the phonetic categorization suggests that there exists an interactive processing between different levels, consistent with the Trace model. Hence the lexical effect

may be explained by parallel processing in which a higher level of processing (lexical level) interacts with a lower level (phonemic level) in speech perception.

Other combination of lexical status (nonword-word-word and word-word-nonword) must be examined in future study to determine the lexical effect of mono-syllables in TSM. Besides, disyllabic word pairs in TSM are possible materials to examine the lexical effect in the future since disyllabic words in spoken TSM are more frequent compared to monosyllabic words.

## References

- [1] P. Boersma and D. Weenink, *Praat: doing phonetics by computer (Version 5.0.02)* [Computer program]. Retrieved December 27, 2007, from <http://www.praat.org/>.
- [2] B. L. Davis and P. F. MacNeilage, “The articulatory basis of babbling” in *Journal of Speech and Hearing Research*, Vol. 38, 1995, pp. 1199-1211.
- [3] W. F. Ganong, “Phonetic categorization in auditory word perception” in *Journal of Experimental Psychology: Human Perception Performance*, Vol. 6, 1980, pp. 110–125.
- [4] J. L. McClelland and J. L. Elman, “The TRACE model of speech perception” in *Cognitive Psychology*, Vol. 18, 1986, pp. 1-86.
- [5] R. S. Newman, J. R. Sawusch, and P. A. Luce, “Lexical neighborhood effects in phonetic processing” in *Journal of Experimental Psychology: Human Perception Performance*, Vol. 23, 1997, pp. 873–889.
- [6] 董忠司、城淑賢、張屏生, *臺灣閩南語辭典*. 五南, 2001.

---

<sup>1</sup> The combination of “word-nonword-word” is not chosen because there is only one case in TSM: “*bu53-pu53-phu53*”.

<sup>2</sup> Another group “*bu55-pu55-phu55*” is not used as materials for *pu55* serves as an onomatopoeic word only.

<sup>3</sup> Davis & MacNeilage (1995) proposed a universal unmarked CV co-occurrence patterns to be [bilabial+central], [alveolar+front] and [velar+back].