Coda Discrimination is Governed by Acoustic and Phonological Constraints: A Case Study of the Generalized Linear Model*

Chu, Man-ni**
Department of Applied English
Der Lin Institute of Technology

Lin, Hui-wen
Biostatistics Research and Consulting Center
Taipei Medical University

ABSTRACT

A series of studies has shown that consonants can be categorized by their acoustic character. A gating experiment was designed to recognize the factors used to identify terminal plosives correctly. Nine native residents of Shantou, in China's Guangdong Province, were recruited as subjects. The dialect spoken in Shantou does not allow */ap/ and */at/. Participants were asked to identify /taC/ and /kaC/ under conditions where C is one of the four choices /p, t, k, ʔ/. The overall rate of accuracy, from highest to lowest, was /p/ > /k/, /ʔ/ > /t/. In general, the similarities in the formant transitions of /t, k, ʔ/, also observed by Tartter et al. (1983), confused subjects, which resulted in a higher recognition rate for the labial stop. Results suggested that the subjects' language background helped discriminate the remaining sounds; for example, /k, ʔ/ were correctly identified more often than /t/.

Key Words: Shantou, terminal plosives, accuracy rate, second formant transitions

---

* Heartfelt thanks to two anonymous reviewers for insightful comments and criticisms on earlier drafts of this paper. It goes without saying that any errors or infelicities that remain are ours.

** The corresponding author's email address: mannichu@gmail.com

1. Introduction

The acoustic nature of consonants has been described by previous researchers (Stevens 1989; Stevens and Blumstein 1978, 1981). In general, the relationship between consonants and their adjacent vowels has been described as follows. Ladefoged (1975) reported that the formants of the vowel (e.g., F2 and F3) are crucial in discriminating the place of adjacent stops. In short, F2 and F3 decrease when the vowel is adjunct to /p/; they are marked as 'velar pinch' in /k/; however, F2 and F3 are fairly steady when they occur before or after /t/. Although no specific description of the interaction between /ʔ/ and its adjacent vowel has been given, /ʔ/'s frequently occur as an allophone of /t/, which presumably makes us guess that its formant transition is similar to /t/. Because one phoneme may be in-various in terms of its acoustic nature, variability was shown in the behavior of phone or multi-phoneme (Klatt 1980; Ohala 1992). In addition, Ohala et al. (1994) claimed that the identification of a sound is influenced by its neighboring sounds. In that study, subjects were asked to restore the vowel when its adjacent consonants /b/ or /d/ were masked by VCə utterances. Results indicated that the neighboring context did affect subject vowel identification. Thus, in this study a perceptual experiment was conducted to examine the ability of individuals to discriminate between the four final stops.

Lin and Chen (1996) reported that Shantou, a local dialect spoken by some 370,000 people in Guangdong Province, China, has three codas, /p/, /k/ and /ʔ/. Therefore, the /t/ identification was presumed to be the coda that would be least recognized correctly, as participants could be expected to miss sounds that do not exist in their own system. van Wieringen (1995) reported that native speakers of Japanese, whose language does not allow coda consonants, face greater difficulty in identifying terminal plosives as compared to native speakers of Dutch, whose language allows the CVC structure. Tsukada et al. (2007) found that subjects of different language backgrounds react differently in identifying word-final stops. In Tsukada's study, stimuli were drawn from American English, varying between released and non-released sounds, and Thai, with only unreleased terminal plosives. To Cantonese, Korean, and Vietnamese subjects, who always allow unreleased terminal stops, English pairs are more easily recognized than Thai pairs. The conclusion is that familiarity with a particular sound in a foreign language affects non-speaker abilities to discriminate and identify sounds correctly. In pair discriminations, /t/ and /k/ are the least easy to discriminate from one another. This means that /t/ and /k/ are perceptually close.

The gating experiment gives us information regarding the process by which individuals perceive sounds because subjects are presented with information gradually as they perceive a sound at each gate. Ciocca et al. (1994:1134) mentioned that F2 and F3 can differ /p, t, k/ in the middle of a vowel, inferring an assumption of the experiment that subjects perceive terminal stops in the middle of vowel formants.

In the present experiment, we attempted to discover the factors which dominate in subject identification by identifying which fragment causes subjects to detect coda characters. Specifically, we examined accuracy rates and tried from such observations to explain coda identification from both the acoustic and phonological aspects.

2. Methodology

2.1 Materials and Procedure

In order to examine the four sets of coda,[1] stimuli were recorded by a male native speaker of Tainan Southern-Min, which has /p/, /t/, /k/, and /ʔ/, using CoolEdit Pro 2.0 with a 44.1 kHz sampling rate and a 16 bit quantization level. The experiment presented six H-tone tokens of CVC wave forms, which varied in the extent to which they were cut short. This was done in 20ms steps from the end, such that the longest was intact and the shortest ended 100ms earlier. The start point was set at the beginning of the onset stop, where the initial closure was included. Onset C varied between /t/ and /k/, so that we had 6(steps)× 2(C-)×4(-C)=48 stimuli (see Appendix A). Shantou is a non-alphabetic language that uses Chinese characters to represent its lexicon. In order to prevent subjects from misrecognizing terminal stops, subjects were first trained to gain an appropriate level of familiarity with IPA. Subjects were asked to indicate which terminal stop was located in a series of sounds. A 24-stimuli pretest containing the four choices [p, t, k, ʔ] was run to make sure the subjects knew what they would be asked to do during the formal experiment. Subjects who passed the pretest may participate in the next experiment. Then, a 48-stimuli identification test presented subjects with a total of four choices [p, t, k, ʔ]. In order to focus their

1. It has been reported that different languages may have different phonetic features in producing the same consonant; however, since Tainan Southern-Min and Shantou belong to the same Southern-Min system and it is difficult for Shantou natives to produce the final /-t/ which is non-existent in their language, we believe that the languages difference between the listener and the speaker can be ignored.

attention, all the sounds were immediately preceded by a warning signal. All the sounds, presented in a random order determined by the computer software, appeared on the computer screen for subjects to select the one choice that best matched the most recent consonant heard. Each consonant sound was presented three times, giving a total of 48×3(times)$=144$ stimuli. In each section, the 48 stimuli were tested randomly and subjects needed to attend to the same experiment three times. Subjects were required to guess what they heard if they were not sure. The whole experiment took about 30 minutes. There were two options for the data-calculation. In the first option, all the answers in the raw data were counted. In the second, an answer was counted only when a subject consistently gave two same answers out of three experiments. We adopted the second method for stronger reliability, on the one hand, and for simplification, on the other hand. That is, if all the responses were collected, another fixed random effect model should be proposed since the three answers were highly correlated. In order to simplify the model and provide evidence to show that the logistic regression model is adequate to analyze the data, the same two answers were counted as the answer to one particular stimulus. If none of the answers were the same, the answer would be classified as 'undecided'.

2.2 Participants

A total of 18 native speakers of the Shantou dialect of Chinese who self-reported having no hearing ϕ problems participated in the experiment. Only the data of nine subjects were used for the experiment because they passed the threshold of 10/24 in the pretest. Evidence indicates that the probability that the subjects had made a guess by passing the threshold is 0.03, $p<0.05$. That means that those 9 (4 males, 5 females, aged 9-52, Mean 33, SD 15) participants who passed the threshold by guessing the answer correctly 10 out of 24 is less than .05.

2.3 Statistical Method

Among classical linear models (e.g., ANOVA, t-test), we usually assume independent outcomes that are normally distributed with equal variance. These assumptions are manifestly uncertain when the outcome variable is, for example, the correctness or incorrectness of a test, the success or failure of an operation, etc. The characteristics of such cases mean that: (a) normality is not a reasonable assumption, (b) a linear model is usually not appropriate, and (c) variance is generally dependent upon the mean. It is natural to model an outcome variable as binary or dichotomous with probability (Hosmer and Lemeshow 1989), so we propose a logistic regression model (Hosmer and Lemeshow 1989;

Agresti 1995) to model our data.

The designed logistic regression model is:

$$\log\frac{\Pr(Y=1|coda,onset,gate)}{\Pr(Y=0|coda,onset,gate)}=\beta_0+\beta_{1j_1}coda+\beta_{2j_2}onset+\beta_{3j_3}gate$$

Where the fixed variables coda, onset, and gate represent categorical data and Y is the outcome variable. Y=1 represents a correct result; Y=0 represents an incorrect result. Parameter β is a vector, which can be estimated using Maximum Likelihood Estimates (MLEs). The interpretation of the β_j parameter estimates is as the additive effect on the log **odds ratio** (log **odds ratio** is a log ratio of the percentage correct and percentage of incorrect.) for a unit change in the jth explanatory variable. In the case of a dichotomous explanatory variable, for instance, coda, e^{β} is the estimate of the odds ratio of having the outcome for, say, /p/, /k/, and /ʔ/ compared with /t/. The model has an equivalent formulation

$$P=\frac{e^{\beta_0+\beta_{1j_1}coda+\beta_{2j_2}onset+\beta_{3j_3}gate}}{1+e^{\beta_0+\beta_{1j_1}coda+\beta_{2j_2}onset+\beta_{3j_3}gate}}$$

P is percentage correct given in coda, onset and gate condition.

2.3.1 The Stepwise Method of Selecting Variables (Cohen 1991)

As we were also interested in determining which variables have the most significant effect on the results, we used a forward stepwise method, i.e., the Wald statistic for variable selection. The Wald statistic was calculated for variables in the model to determine whether a variable should be removed.

Where β is the vector of maximum likelihood estimates associated with the J-1 dummy variables and **V** the asymptotic covariance matrix for β:

the Wald statistic is $Wald=\hat{\beta}V^{-1}\hat{\beta}$

The asymptotic distribution of the Wald statistic is chi-square, with degrees of freedom equal to the number of parameters estimated.

2.3.2 Forward Stepwise (FSTEP)

FSTEP Algorithms obtain a final model by starting with all the variables suspected to affect the coda identification ability of subjects. To determine the subset of suspect variables, all the variables are estimated as a function of parameters and likelihood functions. In examining variable validity, we obtain the necessary information by checking the MLE parameters, the prediction probabilities, and the likelihood function for the current model. A variable is