말소리와 음성과학 제 1 권 제 3 호 (2009)
pp. 19~28

구간 신호 길이 자질과 한국인의 영어 파열음 지각
The Duration Feature of Acoustic Signals and Korean Speakers' Perception of English Stops

김문형[1] • 전종섭[2]
Kim, Munhyong • Jun, Jong Sup

ABSTRACT

This paper reports experimental findings about the duration feature of the acoustic components of English stops in Korean speakers' voicing perception. In our experiment, 35 participants discriminated between recorded stimuli and digitally transformed stimuli with different duration features from the original stimuli. 72 sets of paired stimuli are generated to test the effects of the duration feature in various phonetic contexts. The result of our experiment is a complicated cross-tabulation with 540 cells defined by five categorical independent variables plus one response variable. To find a meaningful generalization out of this complex frequency table, we ran logit log-linear regression analyses. Surprisingly, we have found that there is no single effect of the duration feature in all phonetic contexts on Korean speakers' perception of the voicing contrasts of English stops. Instead, the logit log-linear analyses reveal that there are interaction effects among phonetic contexts (=C), the places of articulation of stops (=P), and the voicing contrast (=V), and among duration (=T), phonetic contexts, and the places of articulation. To put it in mathematical terms, the distribution of the data can be explained by a simple log-linear equation, $\log F = \mu + \lambda CPV + \lambda TCP$.

Keywords: duration, acoustic signal, English stop, voicing perception, logit log-linear regression

1. Introduction

The voicing perception of English stops, i.e. [b, d, g] vs. [p, t, k], is an old topic in experimental phonetics. In classical studies, several acoustic cues like Voice Onset Time (VOT; Lisker and Abramson, 1964, 1967; Stevens and Klatt, 1974; Chen and Alwan 2001), F1 characteristics (Lisker, 1975), the closure interval (Lisker, 1957), and the length of the preceding vowel (Chen, 1970) are identified to explain how people discriminate between voiced and voiceless stops.

Earlier works, however, have focused on one or two acoustic cues in restricted environments. For instance, Lisker and Abramson (1964, 1967), focusing on the role of VOT, report that the VOTs of voiced stops are significantly shorter than those of voiceless stops. Likewise, Lisker (1975) points out that the F1 onset frequencies of voiced stops are comparatively lower than those of voiceless stops; also, the length and the steepness of F1 transition play an important role for the perception of voicing contrasts (Liberman et al., 1958; Stevens and Klatt, 1974). Thirdly, Lisker (1957) shows that intervocalic voiced stops have shorter closure intervals than their voiceless counterparts. Finally, Chen (1970), Raphael (1972), and House and Fairbanks (1953) report

---

[1] 서울대학교, likerainsun@snu.ac.kr
[2] 한국외국어대학교, jongsupjun@korea.com, 교신저자

(이 연구는 2009 학년도 한국외국어대학교 교내학술연구비의 지원에 의하여 이루어진 것임)

접수일자: 2009년 7월 30일
수정일자: 2009년 9월 10일
게재결정: 2009년 9월 11일

that people tend to recognize stops with longer preceding vowels as voiced. To get a comprehensive picture of listeners' perception of English stops, a thorough comparison of these acoustic components should be carried out.

Kang and Park (2005) is a pioneering work in this regard, since they tested effects of various acoustic signals in diverse phonetic contexts on the voice distinction with Korean learners of English and English native speakers. Their experimental design, however, is limited in scope, in that they did not include a test to see effects of the duration feature of the acoustic components of plosive sounds. Besides, they resort exclusively on descriptive statistics, which does not adequately account for the complicated data. Many previous studies in applied linguistics have faced the same problem of analyzing data with simple statistics that leads to shallow conclusions.

The goal of this paper is to clarify how the duration feature of various acoustic components of stops affects Korean speakers' perception of the voicing contrasts in diverse phonetic environments. To achieve this goal, we tested 35 Korean speakers with different commands of English in a phonetic discrimination experiment. Following Kang and Park (2005), voiced or voiceless stops, i.e. [b, d, g] and [p, t, k], occur in five different contexts in our experiment: voicing contrasts (1) at the onset of words, (2) between a stressed vowel and a non-stressed vowel (= intervocalic after stress), (3) between a non-stressed vowel and a stressed vowel (= intervocalic before stress), (4) at the coda of words with explosion, and (5) at the coda position without explosion. For a phonetic discrimination test, we created 72 sets of paired stimuli for the five contexts. In each paired stimuli, one is an original word and the other is a transformed word. To create a transformed stimulus, we divided acoustic signals into three components: the preceding vowel, the closure interval, and the duration between the onset of burst and the glottal signal. Each of these components of an original word is digitally lengthened or shortened; hence, we could see whether the subjects discriminate the original stimulus from the transformed counterpart with a different duration feature.

To overcome the limitation of basic statistical tests like χ^2, we ran an advanced modeling procedure of inferential statistics, namely logit log-linear regression that helps find a meaningful generalization from a cross-tabulation with more than three categorical variables plus dependent variables. Surprisingly, our study shows that there is no single effect of the duration feature on the voicing perception; rather, Korean speakers' perception of English stops is best explained by the interaction between duration and other independent parameters like phonetic contexts and the places of articulation of plosive sounds.

This paper is organized, as follows. Section 2 briefly reviews earlier studies that are closely related to our work. Section 3 presents the design of our experiment. Sections 4 and 5 are the result and the discussion of our experiment. Section 6 concludes the paper.

2. Earlier Studies

Classical studies have focused on the roles of the preceding vowel length, VOT, and the closure interval for the voicing perception of English stops. 2.1 sums up earlier studies with native English speaker subjects, and 2.2 presents representative studies with Korean EFL (= English as a foreign language) learners.

2.1 Voice distinction for native English speakers

Raphael (1972) investigated how the preceding vowel length affected speakers' recognition of the voice distinction of plosives, fricatives and clusters in synthetic speech. The vowel duration of CVC pairs is adjusted with reference to a range of values found from real-speech samples. The study concludes that subjects recognize the final consonants as voiceless when the duration of preceding vowels is comparatively short.

Krause (1982) studied the judgment of voicing for intervocalic stops controlling the subjects' age. Three age groups (i.e. three-year olds, six-year old, and adults) are tested for sets of stimuli ([bip]-[bib], [pot]-[pod], and [back]-[bag]). The result shows that the judgment of all age groups shifts from voiceless to voiced consonants as vowel lengths progressively get longer.

VOT has been considered as an important acoustic cue in classical studies. Lisker and Abramson (1964) studied eleven languages, and proposed three VOT categories: prevoicing [3], short-lag VOT, and long-lag VOT. They point out that VOT can

---

[3] The prevoicing VOT occurs when voice-leading starts before the onset of the burst of air.

be a universal acoustic cue for the voicing perception in languages with the voicing contrasts of stops.

The closure interval is another relevant parameter. Lisker (1957) reports that a short closure interval is judged to be voiced, and a long closure interval voiceless. Nonetheless, in the coda position, the closure interval does not have an effect on the voiced-voiceless discrimination (Raphael, 1981).

2.2 Voice distinction for Korean EFL learners

Kang and Park (2005) is an elaborate study of acoustic cues in various contexts, in which they tested native English speakers and three groups of Korean EFL learners classified by their TOEIC listening scores. They made stimuli in five different contexts for plosives: voice contrasts (1) at the onset of words, (2) between a stressed vowel and a non-stressed vowel (= intervocalic after stress), (3) between a non-stressed vowel and a stressed vowel (= intervocalic before stress), (4) at the coda of words with explosion, and (5) at the coda position without explosion. They recorded the test material, digitally transformed the recorded sound samples, and conducted an identification test. The result shows that both native English speakers and Korean EFL learners are prone to rely on VOT at the word onset, and between a non-stressed vowel and a stressed vowel. Unlike native English speakers who tend to rely on preceding vowels, Korean EFL learners show preference to the closure component in identifying voiced-voiceless stops.

As a subsequent study, Kang (2006) further investigated the recognition of the voicing contrasts only in the intervocalic context. He reported that the hierarchy of degree of reliance for both native English speakers and Koreans to recognize the voice distinction is “preceding vowel >> closure >> VOT >> following vowel in the intervocalic position after stress and VOT >> following vowel >> closure >> preceding vowel in the intervocalic position before stress.”

Besides, Ko (1998) studied whether or not the duration of preceding vowels for plosives affects the perception of the voice distinction for English native speakers and Korean native speakers. The study shows that the preceding vowel does not play a crucial role for the voice distinction for both groups. In later studies, Ko (1999), and Ko and Lee (1998) investigated if the duration of closure, and the release portion can be reliable cues for the voice distinction. They conclude that none of these features are decisive for the discrimination of the voicing contrasts.

So far, we have summed up earlier studies that present representative acoustic cues for the perception of voiced and voiceless stops for both native English speakers and Korean EFL learners. Kang and Park’s (2005) study is particularly important, since they tested effects of several acoustic cues in various contexts. Nevertheless, their experimental design does not test effects of the duration feature of stops in various phonetic contexts. Notice that the duration feature has been admitted as an important parameter in the literature, but that no study gives an overall description of the duration parameter in various contexts. For this reason, we aim at an experimental design that tests the effect of duration in various environments like the duration effect of the preceding vowel at the word onset, the duration effect of the closure interval between a stressed vowel and a non-stressed vowel, the duration effect of VOT between a non-stressed vowel and a stressed vowel, etc.

3. Experimental Procedure

3.1 Subjects

Thirty-five university students who are all native Korean speakers with different levels of English proficiency participated in our study. Before the main experiment, we gave the students the Minimal English Test (MET)[4] to categorize the subjects into three groups of different English proficiency. The obtained MET scores of the subjects range from 29 to 69 out of seventy. We also obtained directly from our subjects their TOEIC or other standard test scores if they had taken any standard test before. After comparing the obtained MET scores with other available test scores, we categorized the subjects into three groups of different English levels: 8 subjects as high level (MET score 63-69), 20 subjects as medium level (MET score 42-60), and the remaining 7 subjects as low level (MET score 29-40). This way, we wanted to

---

4) The Minimal English Test (MET) is developed by Hideki Maki and his colleagues as a five-minute test that correctly shows ESL learners’ level of English proficiency. Surprisingly, this five-minute quiz score significantly correlates with many standard test scores. See Maki et al. (2008) for details.

test the effect of Korean speakers' English proficiency on their perception of English stops in the discrimination test.

3.2 Stimuli

The stimuli were recorded by a 32-year-old native speaker of English, who had been born and raised in Chicago areas. As a recording device, SONY RM-D100K was used with an Audio Technica Microphone 72 (ATM75). The sampling rate was 44,000Hz, and the sound was digitalized in 16bit quantizing.

Since the effect of the duration feature may vary depending on the place of articulation of a stop sound, we varied the test material into bilabial, dental and velar stops. To make the stimuli natural, and to guarantee precise results, the test material should not be recorded in isolation, but should be read when it is uttered in a natural sentence (Lisker and Abramson, 1967). To this end, we adopted Kang and Park's (2005) test design, and used the test material in sentence contexts as shown in <Table 1>.

Table 1. Words in Sentence Contexts

Five Contexts of Plosives	Sentence Contexts	Words
Onset	"____again, please."	pad, bad, tad, dad, cad, gad
Intervocalic after stress	"Say ____ again."	dapper, dabber, dadder, datter, dacker, dagger
Intervocalic before stress	"Say the ____, again."	pad, bad, tad, dad, cad, gad
Coda	"Say ____."	dap, dab, dat, dad, dack, dag
Coda with no explosion[5]	"Say ____."	dap, dab, dat, dad, dack, dag

3.2.1 Generating transformed stimuli and the discrimination test

From the recorded sentences, we first extracted the target words, i.e. test stimuli, by using Praat 4.4.22. Then, we identified three components of acoustic signals out of each extracted stimulus: the interval before the closure, the closure interval and the interval of release before the burst with aspiration. We will call these three units the preceding vowel, the closure, and VOT respectively.[6]

To test the effect of the duration of each acoustic component of a stop, we transformed each recorded stimulus by digitally lengthening or shortening the durations of interval of the preceding vowel, the closure, and VOT to match the durations of interval for its voicing-contrasted stimulus. For instance, when a recorded stimulus *dapper* is given, the duration interval of the preceding vowel of *dapper* is digitally adjusted to match the duration interval of the preceding vowel of *dabber*; likewise, the duration intervals of the closure and VOT of *dapper* are lengthened or shortened to match the duration intervals of the closure and VOT of *dabber*.

The transformation is performed by Praat 4.4.22. Since we want to analyze the effects of three duration intervals (i.e. the preceding vowel, the closure, and VOT) in five phonetic contexts in <Table 1>, we generated 72 digitally transformed stimuli (i.e. onset 6 + intervocalic after stress 18 + intervocalic before stress 18 + coda 18 + coda with no explosion 12). For illustration, look at the signals and spectrograms of [pad] and [bad] in <Figure 1> and <Figure 2>. The VOT of [p] in [pad] is approximately 67.4 ms, and its counterpart [b] in [bad] has the VOT of about 20.4ms. Thus, we shortened the VOT of [pad] to 20.4ms to create a paired stimuli [pad] vs. [pad]$_{\text{transformed}}$. Likewise, the VOT of [bad] is digitally extended to 67.4ms to create a paired stimuli [bad] vs. [bad]$_{\text{transformed}}$.

Finally, the set of 72 paired stimuli, each of which consists of an original recorded stimulus and its transformed counterpart, is used in our discrimination experiment. The test is executed with the Experiment MFC of Praat 4.4.22 as shown in <Figure 3>.

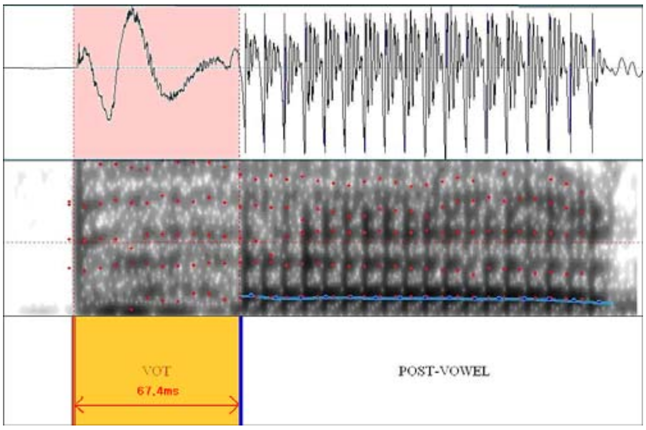


Figure 1. The Signal and Spectrogram of [pad]

---

5) The *Coda with no explosion* context is made by removing the explosion from the coda position of a target word.

6) The interval after VOT is not included in the present report, since we assume based on our pilot analysis that the duration of the interval after VOT will have much less effect on the voice discrimination than the acoustic signals before VOT.

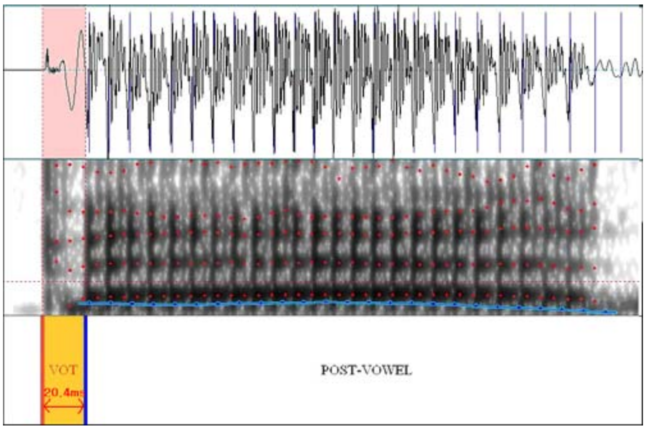
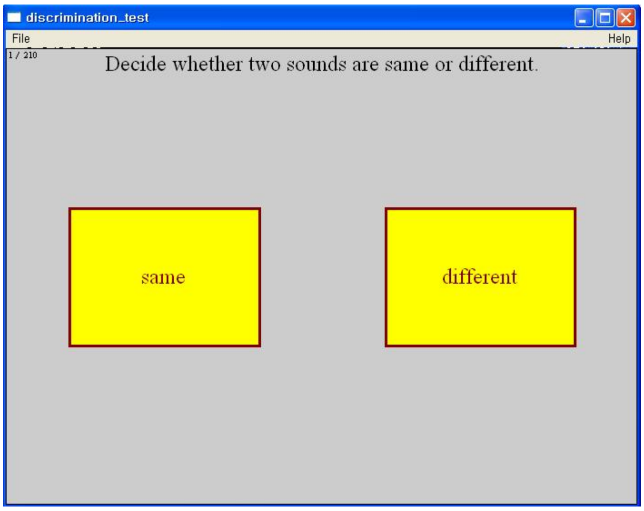


Figure 2. The Signal and Spectrogram of [bad]



Decide whether two sounds are same or different.

Figure 3. The Discrimination Test using Praat

4. Result

4.1 Data

To analyze the data, we set up five categorical independent variables and one dependent (response) variable: the English proficiency level, the phonetic context of a plosive, the voicing contrast, the place of articulation of a plosive, the transformed interval, and the response from subjects. <Table 2> summarizes all the variables and their values. Notice that the response variable is not an independent variable, but a dependent one.

Table 2. Variables and Values

Variables	Values
English Proficiency Level	High
	Middle
	Low
Place of Articulation	Bilabial
	Dental
	Velar
Voicing Contrast	Voiced
	Voiceless
Context of Plosive	Word Onset (WO)
	Intervocalic after Stress (IAS)
	Intervocalic before Stress (IBS)
	Word Coda with Explosion (WC)
	Word Coda with no Explosion (WC-NE)
Transformed Interval	Preceding Vowel
	Closure
	VOT
Response	Same
	Different

The six categorical variables produce a cross-tabulation with 540 cells (i.e. 3 values for English Proficiency Level * 3 values for Place of Articulation * 2 values for Voicing Contrast * 5 values for Context of Plosive * 3 values for Transformed Interval * 2 values for Response = 540 cells). <Table 3> shows part of the cross-tabulation for the better understanding of our data.

Table 3. Part of the Multi-way Contingency Table

Contexts of Plosives		Intervocalic after Stress					
Transformed Interval		Preceding Vowel					
Place of Articulation		Bilabial					
Voicing Contrast		Voiced			Voiceless		
English Proficiency Level		H	M	L	H	M	L
Response	Same	8	18	7	8	20	7
	Different	0	2	0	0	0	0
Contexts of Plosives		Intervocalic after Stress					
Transformed Interval		Preceding Vowel					
Place of Articulation		Dental					
Voicing Contrast		Voiced			Voiceless		
English Proficiency Level		H	M	L	H	M	L
Response	Same	8	19	7	8	18	7
	Different	0	1	0	0	2	0

The shaded area of <Table 3> shows only 24 out of the 540 cells. Clearly, it is impossible to analyze this kind of cross-tabulation with basic (non-parametric) statistics like χ^2. It is also impossible to come up with any meaningful generalization by seriously looking at a print-out of the 540-cell table. We need an advanced tool to analyze this table appropriately – logit log-linear regression.

4.2 Logit log-linear regression

To understand logit log-linear regression, we have to introduce log-linear models first. Log-linear regression is a specialized case of General Linear model (GLM) for Poisson distributed data.7) 8)

---

7) It is a distribution of random events occurring in a fixed period of time or area. For example, the events of finding bird droppings on the windshield within one month period of time

The purpose of log-linear analysis is to find the most parsimonious model from a set of possible mathematical models that fit the observed frequency data of categorical independent variables. The full or saturated model includes all possible interactions among variables. When we do log-linear analyses, we create numerous mathematical models, test the validity of each model, and look for the most restricted or parsimonious model that has the same explanatory power for the data as the saturated model.

The log-linear analysis helps us see the internal structure of a multi-way contingency table for frequency data. Unlike the χ^2 test for one or two categorical variables, log-linear regression enables us to analyze more than three categorical variables.

We can even run log-linear regression with several categorical independent variables plus dependent (response) variables. A specialized case of log-linear modeling that includes dependent variables is logit log-linear regression. In <Table 3>, the response variable is treated exactly like the other independent variables that define a frequency table. Nevertheless, it makes more sense to consider the response variable as a dependent variable. For this reason, we ran a series of logit log-linear regression analyses to test dozens of restricted models that account for the (seemingly random) distribution of the 540 cells of our table.

4.3 Results: Mathematical models for the data

The goal of (logit) log-linear modeling is to find the simplest model that fits the data. In the backward elimination procedure, we start with the saturated model, and deletes effects from higher-order effects (i.e. Main < Association < Interaction).[9] On the other hand, forward aggregation begins with main effects, and adds (immediate) upper-order effects until the model achieves an adequate goodness of fit to the saturated model.[10] We begin our analysis by forward aggregation. <Table 4> sums up the models we tested for the null hypothesis that a given restricted model is not different from the saturated model.[11]

Table 4. Logit Log-Linear Analyses of the Data

Model	Effects	*df*	L^2	*p*
(1)	{E}, {C}, {P}, {V}, {T}	258	653.878	.000
(2)	{E}, {C}, {P}, {V}, {T}, {E C}, {E P}, {E V}, {E T}, {C P}, {C V}, {C T}, {P V}, {P T}, {V T}	212	302.026	.000
(3)	{E}, {C}, {P}, {V}, {T}, {E C}, {E P}, {E V}, {E T}, {C P}, {C V}, {C T}, {P V}, {P T}, {V T}, {E C P}, {E C V}, {E C T}, {E P V}, {E P T}, {E V T}, {C P V}, {T C P}, {C V T}, {P V T}	120	85.616	.992

---

seem to be random. These events are independent with one another, and they are not easily analyzed by basic statistics. Poisson distribution is a mathematical function that enables us to find meaningful generalizations for these random events. If the expected number of occurrences in a given interval is λ, then the Poisson probability function for *x* occurrences is given as follows:

$$f(x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$$

8) See Li (2002) for a detailed introduction to log-linear modeling.

9) The main effect is the simple effect of an individual variable, and the association effect is the combined effect of any two variables. The interaction effect combines all variables, and presupposes the inclusion of main and association effects.

10) The goal of log-linear modeling is to find the most restricted model that accounts for the distribution as effectively as the saturated model. Hence, the goodness of fit statistic of an adequate restricted model should not be significant: the significant *p* value (i.e. < .05) means that the model is different from the saturated model.

11) Following the general convention of log-linear modeling, effects of the predictor variables are abbreviated to capital letters in curly brackets, as shown below:

E = English Proficiency Level
C = Context of Plosive
P = Place of Articulation
V = Voicing Contrast
T =Transformed Interval (*to test the duration effect*)

Two characters in curly brackets refer to an association effect; i.e. {E C} refers to association effect of English Proficiency Level and Context of Plosive. Likewise, three characters in curly brackets refer to an interaction effect of the three variables.

In log-linear modeling, a set of effects in curly brackets is a shorthand notation for a log-linear equation, as follows:

$\{X\} \Rightarrow \log F_{ij} = \mu + \lambda_i X$
$\{Y\} \Rightarrow \log F_{ij} = \mu + \lambda_j Y$
$\{X\}, \{Y\} \Rightarrow \log F_{ij} = \mu + \lambda_i X + \lambda_j Y$
$\{XY\} \Rightarrow \log F_{ij} = \mu + \lambda_i X + \lambda_j Y + \lambda_{ij} XY$

(4)	{E}, {C}, {P}, {V}, {T}, {E C}, {E P}, {E V}, {E T}, {C P}, {C V}, {C T}, {P V}, {P T}, {T C P}	196	226.130	.069
(5)	{E}, {C}, {P}, {V}, {T}, {E C}, {E P}, {E V}, {E T}, {C P}, {C V}, {C T}, {P V}, {P T}, {C P V}	204	227.172	.127
(6)	{E}, {C}, {P}, {V}, {T}, {E C}, {E P}, {E V}, {E T}, {C P}, {C V}, {C T}, {P V}, {P T}, {C P V}, {T C P}	188	160.147	.930
(7)	{E C}, {E P}, {E V}, {E T}, {C P}, {C V}, {C T}, {P V}, {P T}, {C P V}, {T C P}	188	160.147	.930
(8)	{C P V}, {T C P}	210	221.030	.467

<Table 4> shows a gradual process of finding the most parsimonious model by forward aggregation.[12] First, we began with only main effects, and added two-way association effects and three-way effects through models (1)-(3). Model (3), which incorporates all three-way effects and lower-order effects, fits the saturated model nearly perfectly (p = .992), but this model is too complicated to see what effects are important among all possible effects. Thus, we examine models that include main effects, two-way effects, and one three-way effect to see which three-way effect is important. Two of these models are adequate: models (4) and (5) have *p* values of .069 and .127 respectively. Next, model (6), which includes main, two-way, and two three-way effects from models (4) and (5), provides an adequate fit to the saturated model (p = .930).

As a next step, we started the forward elimination, not backward elimination[13]. We eliminate all main effects from model (6) as shown in model (7), and all two-way effects as shown in model (8). Surprisingly, model (8), which has only two three-way effects, is selected as the most parsimonious model (p = .467). Model (8) provides us with a meaningful generalization to account for the distribution of the 540 cells of the cross-tabulation shown in <Table 3>. Notice that "{C P V}, {T C P}" in model (8) is a shorthand notation for the following log-linear equation: $logF = \mu + \lambda CPV + \lambda TCP$. Model (8) indicates that there are clear interactions among phonetic contexts (=C), the places of articulation of stops (=P), and the voicing contrast (=V), and among the transformed interval (=T), phonetic contexts, and the places of articulation. Notice that the transformed interval (=T) is an operationally defined variable for the duration feature of acoustic components of stops.

5. Discussion

Log-linear modeling is a powerful tool which enables us to find a meaningful generalization out of a cross-tabulation of more than three categorical variables. Logit log-linear regression, as an extended version of general log-linear modeling, allows us to include independent variables in generating linear equations. <Table 3> shows just part of a large cross-tabulation with 540 cells defined by six categorical variables. If we consider the response variable as a dependent variable, we are still left with 270 cells with five categorical predictor variables. Without log-linear modeling, it is virtually impossible to find a meaningful generalization out of this complicated cross-tabulation data. With log-linear modeling, however, we can build up a surprisingly simple linear equation that accounts for the data.

Using the response variable as a dependent variable, we ran a series of logit log-linear regression analyses, and found out that the distribution of our data is best explained by a simple log-linear equation; i.e. $logF = \mu + \lambda CPV + \lambda TCP$. Then, how can we understand the two three-way interaction effects, namely {C P V} and {T C P}, in this linear equation?[14]

Parameter estimate analyses, as *post hoc* tests, help us understand model (8). <Table 5> shows how the two three-way association effects explain the observed data with detailed cases.

---

12) We tested 16 restricted models for this analysis, but present only eight models in <Table 4>. The other eight models are left out since they are not relevant for discussion.

13) Since only two of the three-way association effects are relevant, we need to eliminate lower-order effects to find the most parsimonious model.

14) As we pointed out in Footnote 9, the three-way interaction effects presuppose main and two-way association effects. This means that adding main or two-way terms to model (8) does not change the mathematical outputs.

Table 5. Parameter Estimates[15]

{C}	{P}	{V}	{T}	*p*	R (s/d)
WO	Bilabial		VOT	.007	39/31
WO	Dental		VOT	.000	9/61
WO	Velar		VOT	.000	33/37
IBS	Bilabial		VOT	.000	38/32
IBS	Dental		VOT	.000	20/50
IBS	Velar		Closure	.047	62/8
IBS	Velar		VOT	.000	32/38
WC	Bilabial		VOT	.009	49/21
WC	Dental		Closure	.012	38/32
WC	Velar		Preceding Vowel	.000	36/34
WC-NE	Bilabial		Preceding Vowel	.002	55/15
WC-NE	Dental		Preceding Vowel	.000	34/36
WO	Bilabial	Voiced		.002	13/22
WO	Velar	Voiced		.002	23/12
IBS	Dental	Voiced		.000	84/21
IBS	Velar	Voiced		.000	90/15
WC	Bilabial	Voiced		.001	98/7
WC	Dental	Voiced		.000	56/49
WC-NE	Dental	Voiced		.000	38/32
WC-NE	Velar	Voiced		.026	43/27

<Table 5> shows how particular combinations of values from three categorical variables are explained by the log-linear equation in model (8). For instance, the first row of <Table 5> shows an interaction effect of {TCP}. This particular case refers to a question in our discrimination test, where {C}ontext of Plosive is Word Onset (WO), {P}lace of Articulation is Bilabial, and {T}ransformed Interval is VOT. The *p* value for this case is .007, which shows significant difference between the subjects who responded to the question with 'same' and those who responded to the question with 'different.' This way, <Table 5> is self-explanatory, and shows us how model (8) explains the distribution of the data.

So far, we have focused on the restricted models with interaction effects that provide adequate fit to the saturated model. As a further analysis, we tested a model with only two-way effects, i.e. {EC}, {EP}, {EV}, {ET}, {CP}, {CV}, {CT}, {PV}, {PT}, {VT}, to figure out the importance of the main effects; to our surprise, the values of *df*, L^2, and *p* were the same as those in model (2). This means that single effects do not influence Korean speakers' perception of the voicing contrasts of English stops. This finding is particularly important, since it shows that Korean speakers' English proficiency alone does not have an effect on the perception of the voicing contrasts.

Another great advantage of log-linear analyses is that it allows us to assess relative importance of several variables. As *post hoc* tests, we examined the following models in <Table 6>.

Table 6. Logit Log-Linear Analyses: *Post Hoc* Tests

Model	Effect Name	*df*	L^2	*p*
(9)	{E}	267	1058.082	.000
(10)	{C}	265	760.340	.000
(11)	{P}	267	1023.343	.000
(12)	{V}	268	1057.261	.000
(13)	{T}	267	929.660	.000
(14)	{E}, {C}	263	760.213	.000
(15)	{E}, {P}	265	1023.229	.000
(16)	{E}, {V}	266	1057.149	.000
(17)	{E}, {T}	265	929.542	.000
(18)	{E}, {C}, {T}	261	695.794	.000
(19)	{E}, {P}, {V}	264	1022.284	.000
(20)	{E}, {C}, {P}	261	720.559	.000

In <Table 6>, each model has only main effects, and hence we can examine a single effect of a variable by comparing two models. In models (14)-(17), we tested two-way effects that include the English Proficiency parameter. By comparing one of these models with model (9), we can test whether other variables are more important than English Proficiency. Look at models (9) and (16), first. To see if model (16) fits better than model (9), we subtract the L^2 and the *df* of model (16) from the corresponding values of model (9), and evaluate the significance of the difference with reference to the chi-square distribution. The computation turns out to be not significant (χ^2 (1) = 0.933, p > .05); therefore, we conclude that the single effect of voicing is not important. On the other hand, models (14), (15), and (17) fit

---

[15] Refer to <Table 2> to see the values of {C}.

significantly better than (9) ($\chi2$ (4) = 297.869, $p < .005$; χ^2 (2) = 34.853, $p < .005$; $\chi^2(2) = 128.54$, $p < .005$, respectively).

A comparison of effects between phonetic contexts and the transformed interval can be done by comparing two relevant models in <Table 6>. Model (18) is better than model (14) and (17) ($\chi^2(2) = 64.419$, $p < .005$; $\chi^2(4) = 233.793$, $p < .005$, respectively). Likewise, we compare effects between the place of articulation and voicing: model (19) is not better than model (15) ($\chi^2(1) = 0.945$, $p > .05$), but is better than model (16) ($\chi^2(2) = 34.865$, $p < .005$).

Finally, we compare effects between the contexts of stops and the place of articulation. The subtraction of L^2 from model (14) to model (20) is significant ($\chi^2(2) = 39.654$, $p < .005$). Model (20) is also better than (15) ($\chi^2 = 302.67$, $p < .005$). To wrap up all the comparisons made so far, we conclude that the phonetic contexts, the place of articulation, and the transformed interval (i.e. the duration feature) are more important factors for the voiced-voiceless discrimination than English proficiency and the voicing contrast; i.e. ({C} / {P} / {T}) > ({E} / {V}).

6. Conclusion

In this paper, we have reported experimental findings about the role of the duration feature of acoustic signals in Korean speakers' perception of the voicing contrasts of English stops. The overall design of the experiment is adopted from Kang and Park (2005), but we have focused on the role of duration of the preceding vowel, the closure interval, and VOT in various phonetic environments.

The result of our experiment is a complicated cross-tabulation with five categorical independent variables plus one dependent response variable. To find a meaningful generalization out of the 540-cell table, we ran logit log-linear regression analyses. Surprisingly, we have found that there is no single effect of the duration feature in all phonetic contexts of stops. Also, the level of English proficiency does not influence Korean speakers' perception of the voicing contrasts of stops. Instead, it has turned out that there are interaction effects among phonetic contexts, the places of articulation of stops, and the voicing contrast, and among the transformed interval (i.e. the duration effect), phonetic contexts, and the places of articulation. To put this finding in a mathematical term, the distribution of the data can be explained by a simple log-linear equation, $\log F = \mu + \lambda CPV + \lambda TCP$.

When we resort to basic statistical analyses, we tend to limit the number of variables by dropping out *seemingly* unimportant factors or by collapsing several variables into one variable. Results from such analyses are not reliable, since even if we had found a significant effect, it might have come from other unknown factors. This is why advanced statistical procedures like log-linear regression are preferred to basic statistical tests. In this regard, the finding reported in this paper will not only contribute to our understanding of the voicing perception of stops, but also introduce a powerful method of data analyses to applied linguistics.

References

Chen, M. (1970). "Vowel length variation as a unreleased Danish stops consonants", *Phonetica*, Vol. 22, pp. 129-159.

Chen, M. & Alwan A. (2001). "On the perception of voicing for plosives in noise", *Journal of the Acoustical Society of America*, Vol. 107, No. 5, pp. 2917-2920.

Hause, A. S. & Fairbanks. G. (1953). "The influence of consonantal environment upon the secondary acoustical characteristics of vowels", *Journal of the Acoustical Society of America*, Vol. 25, No. 1, pp. 105-113.

Kang, S. (2006). "A perceptual study on the temporal cues of English intervocalic plosives for various groups depending on background language, English listening ability, and age", Speech Sciences, Vol. 13, No. 2, pp. 133-145.
(강석한, (2006). "언어별, 연령별, 수준별 집단에 의한 모음간 영어 파열음 유/무성 인지 연구", 음성과학, 13권, 2호, pp. 133-145.)

Kang, S. & Park, H. (2005). "A perception study of the temporal cues of English plosives for leveled groups of Korean English Learners", Malsori, Vol. 56, pp. 49-73.
(강석한, 박한상, (2005). "다양한 수준의 한국인 영어 학습자의 영어 파열음의 구간 신호 지각 연구", 말소리, 56권, pp. 49-73.)

Ko, H. (1998). "A study on consonantal voicing-dependent vowel duration in English", Proceedings of Acoustic Society of Korea Conference, Vol. 17, No.1, pp 399-420.
(고현주, (1998). "영어의 자음 유·무성 대립과 선행모음 길이와의 상관관계 연구", 한국음향학회 학술발표대회 논문집, 17권, 1호, pp. 399-402.)

Ko, H. (1999). "Clues to the voicing identification of word-final stops in English –focusing on their consonantal features-", *Malsori*, Vol. 37, pp. 13-21.
(고현주, (1999). "영어 어말 폐쇄음의 유·무성 인지 실마리에 관한 연구 -폐쇄음의 자음적 특징을 중심으로-", 말소리, 37권, pp. 13-21.)

Ko, H. & Lee, S. (1998). "On the perceptual cues to voicing of English word-final stops-focusing on the consonantal features", *Proceedings of Acoustic Society of Korea Conference,* Vol. 17, No. 2, pp. 89-92.

(고현주, 이숙향, (1998). "영어 어말 폐쇄음의 유·무성 인지 실마리에 관한 연구 -폐쇄음의 자음적 특징을 중심으로-," 한국음향학회 학술발표대회 논문집, 17권, 2호, pp. 89-92.)

Krause. S. E. (1982). "Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults", *Journal of the Acoustical Society of America*, Vol. 71, No. 4, pp. 990-995.

Li, P. (2002). "Loglinear models for the analysis of language acquisition data", *Journal of Cognitive Science*, Vol. 3, pp 27-41.

Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1958). "Some cues for the distinction between voiced and voiceless stops in initial position", *Language and Speech*, Vol. 1, pp. 153-167.

Lisker, L. (1957). "Closure duration and the intervocalic voiced-voiceless distinction in English", *Language*, Vol. 33, pp. 42-29.

Lisker, L. (1975). "Is it VOT or a first-formant transition detector?", *Journal of the Acoustical Society of America*, Vol. 57, pp. 1547-1551.

Lisker, L. & Abramson, A. (1964). "A cross-language study of voicing in initial stops: acoustical measurements", *Word*, Vol. 20, pp. 384-422.

Lisker, L. & Abramson, A. (1967). "Some effect of context on voice onset time in English stops", *Language and Speech*, Vol. 10, pp. 1-28.

Maki, H., Kasai, C., Goto, K.& Okada, A., Takahashi, K., Hasebe, M., Imamaki, H., Ishikawa, A., Ito, T., Oku, S., Yumoto, Y., Miyamoto, Y., Ochi, M., Hamasaki, M., Ueda, Y., Nagasue, K., Kasai, H., and Dunton, J. (2008). "The Minimal English Test: Its correlation with the university entrance examination", *Gifu University Regional Studies Research Report*, Vol. 23, pp. 79-86.

Raphael, L. J. (1972). "Preceding vowel duration as a cue to the perception of voicing characteristics of word-final consonants in American English", *Journal of the Acoustical Society of America*, Vol. 51, pp. 1296-1303.

Raphael, L. J. (1981). "Duration and contexts as cues to word-final cognate opposition in English", *Phonetica*, Vol. 51, pp. 126-147.

Stevens, K. N. & Klatt, D. H. (1974). "Role of formant transitions in the voiced-voiceless distinction for stops", *Journal of the Acoustical Society of America*, Vol. 55, pp. 653-659.

- **김문형 (Kim, Munhyong)**
  서울대학교 언어학과
  서울시 관악구 관악로 599 3동 325호
  Tel: 02-880-6164 Fax: 02-882-2451
  Email: likerainsun@snu.ac.kr
  관심분야: 음성학, 컴퓨터언어학
  현재 언어학과 대학원 석사과정 재학중
- **전종섭 (Jun, Jong Sup)** 교신저자
  한국외국어대학교 언어인지과학과
  경기도 용인시 모현면 왕산리 산89
  Tel: 031-330-4228 Fax: 031-330-4074
  Email: jongsupjun@korea.com
  관심분야: 심리언어학, 언어습득, 통계학
  2005~현재 언어인지과학과 조교수