CLEAR SPEECH PRODUCTION AND PERCEPTION OF KOREAN STOPS AND THE SOUND CHANGE IN KOREAN STOPS

by

KYOUNG-HO KANG

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2009

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Kyoung-Ho Kang

Title:

"Clear speech production and perception of Korean stops and the sound change in Korean stops"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Susan Guion, Chairperson, Linguistics
Eric Pederson, Member, Linguistics
Melissa Redford, Member, Linguistics
Kaori Idemaru, Outside Member, East Asian Languages & Literature

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

September 5, 2009

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Kyoung-Ho Kang for the degree of Doctor of Philosophy

in the Department of Linguistics to be taken September 2009

Title: CLEAR SPEECH PRODUCTION AND PERCEPTION OF KOREAN STOPS AND THE SOUND CHANGE IN KOREAN STOPS

Approved: ______________________________________

Dr. Susan G. Guion

The current dissertation investigated clear speech production of Korean stops to examine the proposal that the phonetic targets of phonological categories are more closely approximated in hyperarticulated speech. The investigation also considered a sound change currently underway in Korean stops: younger speakers of the Seoul dialect produce the aspirated and lenis stops differently from older speakers of the same dialect. Hyperarticulated, clear speech provided evidence for difference in the phonetic targets of the stops between the two age groups. Compared with conversational and citation-form speech, younger speakers primarily enhanced the F0 difference between the aspirated and lenis stops in clear speech, with only a small VOT enhancement, whereas older speakers solely enhanced VOT difference between the two stops. These different clear speech enhancement strategies were interpreted to indicate that younger speakers have developed different phonetic targets for stop production than older speakers.

The results from a perceptual experiment using re-synthesized stimuli indicated that the production differences between the younger and older speakers are linked to perceptual differences. The perceptual processing of the stops differed between the groups in a manner parallel to the production differences. When identifying aspirated and lenis stops, younger listeners evidenced greater cue weight for F0 than older listeners, whereas older listeners evidenced greater cue weight than younger listeners for VOT and H1-H2. In addition, the results from a perceptual experiment using noise-masked stimuli confirmed an intelligibility improvement effect of clear speech and also indicated that the three speaking styles were on a continuum from the most casual, conversational speech, to the most careful, clear speech, with citation-form speech in the middle. In the final chapter, the different findings of the current study were discussed in view of various theoretical models and hypotheses.

This dissertation includes previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Kyoung-Ho Kang

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
University of South Carolina, Columbia, South Carolina
Arizona State University, Tempe, Arizona
Dong-A University, Busan, South Korea

DEGREES AWARDED:

Doctor of Philosophy in Linguistics, 2009, University of Oregon
Master of TESL in Teaching English as a Second Language, 1999, Arizona State University
Master of Arts in English linguistics, 1994, Dong-A University
Bachelor of Arts in English Language and Literature, 1992, Dong-A University

AREAS OF SPECIAL INTEREST:

Speech perception
First/Second language acquisition
Bilingualism in Korean and English
Korean phonetics/phonology

PROFESSIONAL EXPERIENCE:

Instructor, Department of Linguistics, University of Oregon, 2009
Teaching Assistant, Department of Linguistics, University of Oregon, 2005-2009
Teaching Assistant, Department of East Asian Languages and Literatures, University of Oregon, 2001-200
Research Assistant, Linguistics Program, University of South Carolina, 2000-2001

GRANTS, AWARDS AND HONORS:

University of Oregon CAPS (The Center for Asian and Pacific studies) small professional grant, 2009

PUBLICATIONS:

Kang, K.-H. and Susan, G. Guion (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America, 124*, 3909-3917.

Kang, K.-H., and Susan G. Guion (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America, 119*, 1672-1683.

ACKNOWLEDGMENTS

Thanks to many people, this dissertation has come out to the world. First of all, I would like to express my deepest gratitude to my parents. My father and mother have supported me every way for a long period of time for my education. My years in America have never been possible without their support. I was able to endure lonely moments in a foreign country and keep my hopes, thinking of their expectation and sacrifice. At the same time, I also thank to my brother and sister for their sacrifice to serve our family for such a long time when the eldest son was away.

Even though her name is put in this paragraph, my sincerest appreciation for Professor Susan Guion, my advisor is second to none. I chose the University of Oregon to work with her and to learn from her. Her profound knowledge in many areas, as well as her thoroughness in work, has always amazed me. I have learned tremendously a lot from her many ways, to list only a few, how to study, how to teach, how to arrange things, and how to present in conference talks, not to mention the knowledge. Her positive and active attitude and mind also cheered me up all the time. She helped me to challenge and gain confidence when I was frustrated by adversities and hesitated in front of new challenges. I never doubt that what I learned from her will be one of the most precious assets for my life in the future.

My gratitude cannot cease to be given to other committee members for this dissertation. Professor Eric Pederson, head of the department and the man of kindness, Professor Melissa Redford, another phonetician in the committee and the best researcher in the field, and Professor Kaori Idemaru, excellent professor and my colleague were the essential part that made this dissertation completed. Their help and encouragement made this dissertation greatly improved, as well as their excellent comments on the manuscript.

Finally, I give my special thanks to my fellow graduate students, Grace Oh and Wookkyung Choe. For the long journey of the dissertation project, they provided me with heart-to-heart talks and friendship, especially when I was, mentally and physically, tired of the endless-looking process. Their support helped me to survive from the long trip.

August, 2009

In Eugene, Oregon

Kyoung-Ho Kang

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

The current dissertation investigated clear speech production and perception of Korean stops, while at the same time investigating the nature of sound change occurring in the Korean stop system. The fundamental goal of the study was to obtain insight into the nature of phonetic representation of Korean stops in light of on-line adjustments in speech production and diachronic changes in Korean stops.

The stop system of Korean is undergoing a sound change (Silva, Choi, & Kim, 2004; Silva, 2006; Wright, 2007). Younger speakers of the Seoul dialect of Korean produce the aspirated-lenis stop contrast differently from previous generations. For the younger speakers, the VOT contrast between the two stop types has diminished and as a result, F0 contrast on the following vowel has taken over the status of a primary acoustic correlate. In contrast, the older speakers maintain the traditional VOT values and F0 is considered secondary to VOT.

Based on the proposal that phonetic targets of phonological categories are more closely approximated in hyperarticulated speech (Lindblom, 1990; Johnson et al. 1993), this study predicted that the generational differences in the production of Korean stops should be more pronounced in hyperarticulated, clear speech. The production of Korean stops by younger and older speakers was investigated in terms of three acoustic correlates realized in three different speaking styles. Due to the potentially different phonetic targets conditioned by the sound change, the younger and older speakers were predicted to show different clear speech enhancement patterns: younger speakers were predicted to enhance F0 differences, whereas older speakers were predicted to enhance VOT differences in clear speech. Different clear speech enhancement patterns would indicate that younger and older speakers have developed different phonetic targets for the stop production.

As a next step, this dissertation examined the greater intelligibility characteristic of clear speech. Given that clear speech is a listener-oriented speech mode that aims at helping a listener to better access speech signal, this study investigated whether the hyperarticulated, clear stop productions were more intelligible than productions in more casual speaking styles, i.e., conversational and citation-form speech. At the same time, potential intelligibility improvements between citation-form speech and conversational speech were investigated. For these purposes, a perception experiment using noise-masked perception stimuli was conducted.

This dissertation also investigated production-perception interactions in relation to the sound change. Unlike the investigations of diachronic changes occurring in Korean stops by Silva (2004, 2006) and Wright (2007) which primarily investigated production, the current study also investigated the perception of the Korean stops. Whether differences in the phonetic targets for the production of Korean stops are linked to different perceptual behaviors across the two speaker groups was investigated. Another perception experiment using re-synthesized stimuli was performed to investigate whether the differentially enhanced acoustic correlates between the younger and older speakers lead to the correspondingly differing perceptual weighting of the acoustic correlates between the two groups.

The findings of the present study are reported in six chapters. Chapter I introduces the issues and questions of the study. In Chapter II, the literature is reviewed examining acoustic-phonetic properties of clear speech distinctive from conversational speech, as well as intelligibility improvement over conversational speech. Also, production and perceptual investigations on Korean stops, and studies on sound change in the Korean stop system are reviewed. In Chapter III, experiments examining productions of Korean stops in clear speech are reported compared with the productions in conversational and citation-form speech. By examining the acoustic-phonetic modifications observed in the production of the stops in three different speaking styles, differences in the clear speech enhancement patterns were investigated between younger and older Korean speakers. Specifically, three acoustic correlates to Korean stop contrasts, voice-onset time (VOT), voice quality (H1-H2), and fundamental frequency (F0), were examined to investigate the clear speech enhancement patterns for the two groups. Results of the two production experiments are reported in terms of changes in the acoustic correlates as a function of speaking style and speaker age. Chapter III includes published co-authored materials.

Chapters IV and V present findings from two perception experiments. The experiment reported in Chapter IV investigated whether the stops produced in clear speech were more intelligible than the stops produced in conversational or citation-form speech. For this purpose, speech tokens from the production tasks were submitted to the preparation of the test stimuli. The test stimuli were mixed with masking noise to test intelligibility improvement and presented to listeners in three speaking styles.

The findings from another perception experiment are reported in Chapter V. The goal of the experiment was to further investigate the findings of the two production experiments. In order to assess the extent to which each of VOT, H1-H2, and F0 contribute to the identification of the three types of Korean stops, a set of re-synthesized stimuli in which the three acoustic correlates were manipulated was created. The specific question of the experiment was whether, corresponding to the production differences, F0 was a more weighted perceptual correlate than VOT for younger listeners, whereas VOT was a more weighted correlate for older listeners. Chapter VI concludes the dissertation with the discussion of the findings in view of different theoretical and experimental models and hypotheses.

CHAPTER II

REVIEW OF LITERATURE

2.1 Introduction

The primary goal of speech is communication and this suggests that talkers speak adaptively to meet listeners' demands. When talkers estimate the listeners' speech perception difficulty due to limited access to speech signals (e.g., background noise, hearing problems, or limited experience of second language speakers with the target language), they will increase vocal effort accordingly to minimize perceptual confusions for the listener. At the same time, when talkers do not estimate a need to improve intelligibility, they will default to a more economical, talker-oriented production. Here, the former is referring to an intelligibility-enhancing mode of speech production and it is commonly termed "clear speech" (Smiljanic & Bradlow, 2005).

Another aspect of clear speech is acoustic-phonetic modifications involved in hyperarticulation. By the definition used in clear speech research (Smiljanic & Bradlow, 2009), clear speech aims to enhance intelligibility by providing listeners with enhanced acoustic cues in the speech signals and thus to help listeners to access and comprehend the message better. The enhancement of acoustic salience of speech signals is shared with other goal-oriented speaking styles which involve hyperarticulation modifications, such as infant-directed speech. One aspect of the acoustic-phonetic modifications involved in hyperarticulated speech is the enhancement of phonemic contrasts. The current study focuses on this phonemic contrast enhancement element of the hyperarticulation process involved in clear speech production.

Due to the difficulties in eliciting on-line modulations of speech production in a natural setting, clear speech research has typically used two speaking styles elicited in a laboratory setting. More casual, conversational speech is typically elicited from text-reading tasks in a citation-form, as opposed to more careful, clear speech (e.g., Bradlow, 2002; Bradlow & Bent, 2002). Clear speech is elicited by performing the same tasks with greater attention to the articulatory accuracy. This chapter presents findings of clear speech research, focusing on the intelligibility improvement benefit of clear speech over conversational speech and the acoustic-phonetic modifications occurring during the conversation-to-clear speech style change. In addition, as a background to the investigation of clear speech production and perception of Korean stops, previous investigations on acoustic characteristics of Korean stops, perception of the stops, and diachronic changes in Korean stops are reviewed.

2.2 Acoustic-phonetic characteristics of clear speech

2.2.1 General characteristics

In attempts to identify features contributing to the greater intelligibility of clear speech, clear speech research has reported a wide range of acoustic-phonetic modifications that characterize clear speech productions. In a seminal paper on clear speech, Picheny et al. (1986) reported various properties of clear speech. At a global level, clear speech was produced at a slower speech rate with increased segment duration and more frequent pauses. The overall F0 was also higher in clear speech. At a phonetic level, vowels were reduced less often, and coda stops were released more often. Voice onset time for word-initial voiceless stops increased. Relative consonant-vowel power ratio (in dB) was greater for clear speech. In addition, RMS intensity for stop bursts was much greater in clear speech than in conversational speech.

Ferguson and Kewley-Port (2002) observed acoustic differences in vowel productions between clear and conversational speaking styles. An experienced speaker produced ten English vowels /i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/ in the /bVd/ context embedded in carrier sentences. First formant (F1) and second formant (F2) frequencies at steady-state of the vowel, vowel duration, and dynamic formant movement indexed by vector length (length of a vector in F1 × F2 space connecting the 20% and 80% values of the formants) were measured. For all vowels except /ɪ/ and /ɛ/, F1 was higher in clear speech. As for F2, all of the front vowels (/i/, /ɪ/, /e/, /ɛ/, /æ/) had higher F2, and back vowels (/ɑ/, /ʌ/, /o/, /ʊ/, /u/) had overall lower F2 in clear speech with a significant difference for /ʌ/ and /ʊ/. As for the formant movement measure, differences between clear and conversational speech were only observed for more crowed regions (for /i/, /ɪ/, /e/, /ɛ/, and /o/, /ʊ/, /u/) of the speaker's vowel space. In the less crowded region in F1 × F2 space (/æ/, /ɑ/, and /ʌ/), significant formant movement was not found. Regarding the duration of vowel productions, vowel tokens in clear speech were on average twice as long as vowel in conversational speech. Perkell, Zandipour, Matthies, and Lane (2002) also reported increased vowel duration in clear speech. Seven American English speakers produced "bob", "dod", and "gog" in a carrier phrase, and four out of the seven speakers produced significantly longer vowels in the clear speech compared with conversational speech condition. The findings included greater sound pressure levels for the vowels produced in clear speech condition as well.

More recent investigation of clear speech production in English and Croatian has supported the previous findings. In Smiljanic and Bradlow (2005), five native speakers of each language produced twenty nonsense sentences in each language. Overall speaking rate (calculated by segment duration, number of syllables per second, number and duration of pauses), pitch range (indexed by the highest and lowest F0 points in sentences), and vowel space (indexed by $F1 \times F2$ space) were considered. The results suggested that, for both languages, clear speech production was characterized by a decreased speaking rate, an increased pitch range, and an increased vowel space.

In the previous section, the clear speech research was conducted on deliberately produced clear speech, that is, clear speech was elicited with deliberate instructions for experiment participants to speak clearly. However, some studies have examined the acoustic-phonetic characteristics of naturally more intelligible speakers identified from intelligibility assessments. With speaking conditions and listener-tasks controlled uniform, Bond and Moore (1994) reported that the less intelligible speakers, compared with the most intelligible speakers, produced test words and sentences with shorter segment duration, less differentiated vowel space, more variable amplitude on stressed vowels, less distinctive consonantal cues (VOT and stop release burst), and fewer systematic pauses.

Bradlow, Torretta, and Pisoni (1996) investigated speaker-related correlates of variability in speech intelligibility using a sentence database from 20 speakers. The results indicated that more intelligible speakers used a larger vowel space and a wider F0 range. Overall speaking rate measured by mean sentence duration did not correlate with intelligibility scores, but relative timing between segments were consistently correlated with listener errors. In addition, female speakers were overall more intelligible than male speakers.

Hazan and Markham (2004) reported findings inconsistent with previous findings. Forty five British English speakers produced 124 test words in a sentence frame and were measured for intelligibility. Various acoustic measures, long-term average spectrum, F0, word duration, consonant-vowel intensity ratio, and vowel formants were examined to investigate correlation with intelligibility. The results showed significant correlations only between the amount of energy in the 1000 to 3000 Hz region and intelligibility of the test words. Unlike previous studies, F0 range, speaking rate, and $F1 \times F2$ space did not significantly correlate with intelligibility. However, the findings included a significant correlation between talker voice characteristics and intelligibility. Six most and least intelligible talkers were identified from intelligibility measurements, and these talkers were judged by listeners for the subjective voice characteristics. Subjective dimensions, such as mumbly/precise, unpleasant/pleasant, muffled/clear, weak/powerful, and high/low for a (fe)male significantly separated "good" or "poor" talkers. The combined findings from the three studies on naturally more intelligible talkers suggest that overall, natural clear speech shares acoustic-phonetic characteristics with deliberately elicited clear speech.

The investigation of acoustic-phonetic characteristics of clear speech also revealed that clear speech production involves "closer" approximation of phonetic targets compared to conversational speech. Moon and Lindblom (1994) examined English front vowels (/i/, /ɪ/, /e/, /ɛ/) produced in a /wVl/ frame in conversational and clear speech. Productions of "wheel", "will", "wail", and "well" were examined for formant frequency movement in relation to the adjacent consonant contexts, and the patterns were compared to the values obtained from the /hVd/ context. The results showed less second formant displacement related to the adjacent consonant contexts in clear speech than in citation-form speech. The second formant movement was greater with shorter vowels in both citation-form and clear speech, but this context-duration relation was less obvious in clear speech. At the same time, the results also showed that the smaller target undershoot effect in clear speech was achieved by more rapid formant transitions in clear speech. These results were interpreted to indicate a reorganization of phonetic gestures to overcome coarticulation effects independent of speech rate.

Johnson, Flemming, and Wright (1993) found that vowels produced in hyperarticulated speech were better matched to speakers' best exemplar of the vowels than vowels produced in citation-form speech. American English speakers listened to various synthetic English vowels and were asked to choose best instances matching the vowel categories they had in their mind. Then, the same speakers produced the same vowels in citation and hyperarticulated productions. The results showed that the speakers' choice of vowels better matched the vowels produced in a hyperarticulated style than those produced in a citation style. These results were interpreted to mean that phonetic targets are mirrored better in hyperarticulated, clear speech than in casual, reduced speech.

To summarize the findings of different investigations on clear speech, the following acoustic properties are listed as general characteristics of clear speech allowing it to be distinctive from conversational speech: expanded vowel space (Picheny et al., 1986; Moon & Lindblom, 1994; Krause & Braida, 2004; Ferguson & Kewley-Port, 2002, 2007), decreased speaking rate (Picheny et al., 1986; Uchanski et al., 1996), wider pitch range (Bradlow et al., 1996; Smiljanic & Bradlow, 2005), and increased energy in 1000-3000 Hz range of long-term spectra (Ferguson & Kewley-Port, 2002, Krause & Braida, 2004).

2.2.2 Beyond slow speaking rate

As indicated by the review of general characteristics of clear speech in the previous section, clear speech production typically involves reduction in speaking rate (e.g., Picheny et al., 1986, Bradlow et al., 1996; Similjanic & Bradlow, 2005). Some studies, however, attempted to isolate the intelligibility benefit of clear speech apart from the reduced speaking rate character of clear speech. In a study by Picheny, Durlach, and Braida (1989), clear and conversational speech signals were cross-time manipulated, that is, clear speech was compressed and conversational speech was expanded in time. The processed speech signals were then submitted to intelligibility tests along with original speech signals. Intelligibility scores were higher for clear speech than conversational speech for both processed and unprocessed sets of speech signals. However, the intelligibility scores decreased for both processed conversational and clear speech signals compared to the unprocessed conversational and clear speech signals and thus, the scores for processed clear speech was lower than the scores for unprocessed conversational speech, indicating that the time manipulations rather deteriorated intelligibility.

In a subsequent study by Uchanski et al. (1996), similar experiments were conducted. In this case, however, the speech signals were non-uniformly time-scaled (adjusting segment duration to reflect the various durational differences between clear and conversation speech across different segment classes). Intelligibility tests were conducted on different sets of clear and conversational speech signals, such as cross-processed for speaking rate, doubly presented conversational speech, naturally produced fast clear speech by a professionally trained speaker, and unprocessed clear and conversational speech by three speakers. Both of the time-scaled speech styles were found to be less intelligible than unprocessed conversational speech. The natural clear speech produced in fast speaking rates by a trained speaker was only as intelligible as another speaker's conversational speech at the comparable speaking rate. Thus, the trained speaker was not successful at improving intelligibility of his clear speech without reducing speaking rate.

Krause and Braida (2002, 2004) continued the investigation of speaking rate in clear speech. In Krause (2002), five professionally-trained speakers' clear and conversational speech at slow, normal, and fast rates was examined for intelligibility.

Soft and loud modes of speech were also tested as other alternative forms of clear speech. The results suggested intelligibility enhancement of clear speech produced at a faster (i.e., normal, conversational) speech rate over clear speech produced at a slow speech rate. With properly trained speakers, intelligibility improvement was found for clear speech produced at speech rates faster than clear speech in other studies (e.g., Picheny et al., 1985; Uchanski et al., 1996). The intelligibility advantage of clear/fast speech over conversational/fast speech could not be found because none of the speakers succeeded at producing clear/fast speech with a comparable speaking rate to conversational/fast speech. Also, at a given speech rate, loud speech did not provide a comparable intelligibility advantage to clear speech.

In a related study, Krause and Braida (2004) investigated the acoustic-phonetic characteristics of clear speech produced at a comparable speaking rate to conversational speech. The clear and conversational speech signals showed differences even though speakers were not uniform in their strategies for producing clear speech at normal, conversational speaking rate. At a global-level, greater F0 average and range, greater energy in 1000-3000 Hz range of long-term spectrum, and greater modulation depth below 4 Hz of the intensity envelope were observed for normal rate-clear speech compared with normal rate-conversational speech. At phonological and phonetic levels, more frequent final stop release, longer VOT for word-initial voiceless stops, and greater energy near second and third formants were observed.

In sum, the findings from investigations of acoustic properties of clear speech and the series of investigations on speaking rate in clear speech suggested that the intelligibility benefit of clear speech is likely to be attributed to the combinations of various acoustic properties, and reduced speaking rates do not necessarily correlated with the higher intelligibility of clear speech

2.2.3 Phonemic contrast enhancement effect of clear speech

The investigation of acoustic-phonetic properties contributing to higher intelligibility of clear speech has also revealed that the acoustic-phonetic modifications occurring during the conversational-to-clear speech transformation include phonemic contrast enhancement as well as the enhancement of various acoustic properties summarized in section 2.2.1. In other words, distinctiveness between phonological categories is exaggerated in clear speech compared to conversational speech, and this contributes to overall intelligibility enhancement of clear speech with the effect of the phonological categories being perceptually less confusable in clear speech.

The contrast enhancement effect of clear speech has been investigated for various acoustic-phonetic features serving to differentiate phonological categories. First, acoustic correlates to vowel quality were found to be enhanced in clear speech. Studies on various languages have reported vowel space expansion (indexed by $F1 \times F2$ distance among vowel categories) in clear speech: Spanish (Bradlow, 2002), English (Bradlow, 2002; Ferguson & Kewley-Port, 2002; Picheny et al., 1986; Smiljanic & Bradlow, 2005), and Croatian (Smiljanic & Bradlow, 2005). In addition, for individual vowel categories, the formant targets were found to be more fully realized in clear speech. Ferguson and Kewley-Port (2002) and Hay et al. (2006) reported that formant movement over the vowel nucleus was significantly larger in clear speech than in conversational speech. In Moon and Lindblom (1994), English front vowels were less coarticulated with adjacent consonants in clear speech compared to citation-form speech.

Secondly, vowel duration differences have been found to be enhanced in clear speech in a way related to language-specific phonological structure. In Hay et al. (2006), durational difference was enhanced in a [+focus] context for German short and long vowels, but not for English tense and lax vowels. The duration ratio of German long to short vowels was significantly increased in a [+focus] condition. However, the duration ratio difference between English tense and lax vowels did not reach significantly increase in the [+focus] condition.

Recently, Smiljanic and Bradlow (2008) examined the effect of clear speech production on the duration of English tense and lax vowels, English vowels before voiced and voiceless coda stops, and Croatian short and long vowels. Durational analyses indicated a greater durational difference in clear speech for the Croatian short and long vowels and the English vowels before voiced and voiceless coda stops, but not for the English tense and lax vowels, suggesting an effect of language-specific phonological structure on durational enhancement. Additionally, when duration ratio (clear minus conversational divided by conversational) was considered, the proportional duration contrast for both the Croatian vowels and the English vowels before coda stops remained stable.

Finally, stop contrast enhancement has been investigated in different languages. However, the results provided so far are inconclusive. Languages are found to vary in

VOT modifications in clear speech. In a comprehensive acoustic study of conversationally and clearly produced English sentences, Picheny et al. (1986) found that the voicing contrast for English stops was enhanced in clear speech. Significant VOT lengthening for word-initial voiceless stops was found in clear speech, but not for voiced stops. However, Ohala (1994) examined VOT and found no enhancement in VOT difference between English voiceless and voiced stops in clear speech. Words produced in response to perceptual clarification requests did not show significant VOT enhancement. Smiljanic and Bradlow (2008) examined the voicing contrast between voiced and voiceless stops in English and Croatian. For English, VOT for voiceless stops lengthened in clear speech while the VOT for voiced stops remained unchanged. This voicing contrast enhancement occurred only in word-initial position, not in word-medial position. However, for Croatian prevoiced stops, VOT showed a non-significant trend toward greater negative values in clear speech, while the VOT of voiceless stops remained unchanged. Also, when a proportional measurement of VOT as a percentage of stop duration (i.e., closure + aspiration) was considered, the proportional VOT remained stable across the speaking styles.

Some researchers have also investigated the stop contrast enhancement of clear speech with infant directed speech. Infant directed speech may be considered a variant of clear speech in the sense that talkers adjust their speech on-line to meet the demands of their target audiences (Smiljanic & Bradlow, 2009). Englund (2005) investigated VOT for infant directed speech to the infants in their first six months. VOT was significantly longer in infant directed speech for both Norwegian voiced and voiceless stops. As such, the voicing contrast, as measured by VOT, was not exaggerated in infant directed speech. Sundberg and Lacerda (1999) reported similar results for Swedish stops. VOT was shorter in infant direct speech (with 3 month olds) than in adult directed speech for both Swedish voiced and voiceless stops. Also, the separation between the stop categories was less distinctive in infant directed speech. However, Malsheen (1980) found a greater VOT difference between English voiced and voiceless stops in speech directed to children aged 15-16 months as compared to adult directed speech.

2.3 Intelligibility improvement effect of clear speech

The perceptual clear speech effect, that is, the intelligibility advantage of clear speech over conversational speech has been well established in numerous studies with various listener populations: normal-hearing listeners (Moon, 1991; Ferguson, 2004; Krause & Braida, 2004), hearing impaired listeners (Picheny, Durlach, & Braida, 1985; Payton, Uchanski, & Braida, 1994; Ferguson & Kewley-Port, 2002), and listeners with limited experience in the test language (Bradlow & Bent, 2002; Bradlow & Alexander, 2007). Assuming that clear speech production aims to enhance access to acoustic cues in the speech signals, clear speech research has mostly examined the intelligibility advantage of clear speech with stimuli presented in a condition simulating degraded access to the speech signals, typically presented in noise (Payton, Uchanski, & Braida, 1994; Ferguson & Kewley-Port, 2002; Ferguson, 2004). The conversational and clear speech tokens are typically elicited in a laboratory setting with specific instructions, such as "read naturally as possible as you can as if when you are talking to friends" or "read as carefully and clearly as you can." The majority of studies were conducted on English with a few exceptions (e.g., Smiljanic & Bradlow, 2005 in Croatian).

Picheny et al. (1985) examined intelligibility difference between clear and conversational speech for five hearing-impaired listeners. Fifty nonsense sentences produced clearly and conversationally by three male talkers were submitted to intelligibility tests. Intelligibility score was calculated by counting correctly retrieved key words. The intelligibility benefit in clear speech over conversational speech was, on average, 17 percentage points independent of presentation level (in dB), frequency-gain characteristic, and listener. However, the intelligibility scores showed talker variability, suggesting that talkers vary in their ability to speak clearly. The intelligibility increase was also observed for all phoneme classes with the largest increase for plosives.

Payton et al. (1994) extended Picheny's work by examining the intelligibility benefit of clear speech for noisy, reverberant, and combined listening environments. Ten normal-hearing listeners and five hearing-impaired listeners were tested on the identification of key words embedded in the nonsense sentences used in Picheny et al. (1985). All listeners showed higher scores for the clearly produced sentences for all of the degraded listening environments. The clear speech effect increased as the noise levels and/or reverberation increased. The overall intelligibility improvement was 20 and 26 percentage points for normal-hearing and hearing-impaired listeners, respectively.

More recently, in Smiljanic and Bradlow (2005), 20 Croatian and 30 English listeners participated in sentence-in-noise perception tests for semantically anomalous sentences produced in conversational and clear speech. Five speakers produced the sentences for each language. The average sentence perception scores (computed using correct keyword scores) revealed an intelligibility gain for both Croatian and English for all speakers except one speaker who showed the highest intelligibility score in conversational speech.

Another notable finding of clear speech intelligibility research is that the intelligibility benefit of clear speech over conversational speech is not uniform across speakers. Gagne et al. (1994) reported, in a study with ten female speakers, that the size of clear speech effect varied from -3 percentage points to +24 percentage points. In Picheny et al. (1985), the clear speech effect was not independent of speakers. Three male speakers were not uniform in the size of the intelligibility enhancement. Ferguson (2004) expanded this line of research with a large pool of 41 speakers. Seven normal hearing listeners in a noise condition participated in the vowel intelligibility tests for /bVd/ words. The results showed a wide range of differences among the 41 speakers for intelligibility scores. In conversational and clear speech modes, the intelligibility scores ranged from 25% to 83% and 29% to 94%, respectively. The intelligibility benefit between conversational and clear speech varied by speaker from -12 to 33 percentage points. Female speakers had significantly higher intelligibility scores (78% vs. 68%) for clear speech and greater intelligibility increase than male speakers (11 vs. 5.6 percentage points). Speaker age and prior experience of communicating with hearing-impaired listeners did not affect the intelligibility benefit of clear speech over conversational speech.

So far, we have seen that more careful, clear speech has a wide range of acoustic-phonetic properties distinguishing it from more casual, conversational speech. Clear speech has higher intelligibility for various listener populations compared with conversational speech, and varying size of improvement across speakers. Because the current dissertation examines the these findings of clear speech research in the context of the Korean stop system and related sound changes, the next section will review the findings of previous research on Korean stops.

2.4 Acoustic correlates to Korean stop manner contrasts

The consonant system of the Korean language exhibits three-way manner contrasts among stops. The three types of stops are all voiceless in word-/phrase-initial position, but in word-medial (or phrase-medial) position the lenis stops are voiced intervocalically (Jun, 1993, 1995). The first type, referred to as aspirated, is generally described as strongly aspirated stop, and the second type, referred to as lenis, is generally described as slightly aspirated lax stop. The third type, referred to as fortis, is generally described as an unaspirated tense stop. All three stop types occur at the three places of articulation: bilabial, alveolar, and velar. Table 2.1 presents the nine Korean stop consonants. The voiceless fortis stops are designated with the diacritic " * " and the voiceless lenis stops are left unmarked, following the conventions used by Cho, Jun, and Ladefoged (2002) and Kim, Beddor, and Horrocks (2002).

Table 2.1 Korean stop consonants

	Bilabial	Alveolar	Velar
Aspirated	p^h	t^h	k^h
Lenis	p	T	k
Fortis	p*	t*	k*

2.4.1 Acoustic correlates to three-way manner contrasts

The acoustic-phonetic characteristics of Korean stops have been well documented in a number of studies. The research on Korean stops has identified various acoustic properties distinguishing the three types of Korean stops. Some of the acoustic properties include, among others, voice onset time (VOT), fundamental frequency (F0), and H1-H2 (amplitude difference between the first and second harmonics) on the vowel following stop release. First, regarding VOT, it has been reported that mean VOT values in word-initial position are shortest for fortis, longer for lenis, and longest for aspirated stops (Lisker & Abramson, 1964; Han & Weitzman, 1970; Hardcastle, 1973; C.-W. Kim, 1965; Cho et al., 2002). However, recent work (Silva, Choi, & Kim, 2004; Silva, 2006; Wright, 2007; Kang & Guion, 2008) has suggested that younger Seoul Korean speakers may be shortening the VOT for aspirated stops, such that the VOT for aspirated stops is similar to lenis stops. As for the relative difference in amplitude across first harmonic (H1) and higher harmonics (H2 and above) at the onset of the following vowel, lenis and aspirated stops show greater values than fortis stops (Ahn, 1999; Cho et al., 2002). Generally, the values for both lenis and aspirated stops are positive, and negative for fortis stops. This difference indicates that vowels following lenis and aspirated stops are produced with a breathy or modal quality and vowels following fortis stops are produced with a tense or creaky quality.

As for F0, it has been found that F0 at the onset of the following vowel is lowest for lenis, higher for fortis, and highest for aspirated stops (Han & Weitzman, 1970; M.-R. Cho Kim, 1994; M. R. Kim et al., 2002). In general, F0 for lenis stops is significantly lower than for aspirated and fortis stops (Kagaya, 1974; Jun, 1993; Shimizu, 1996; Ahn, 1999; Cho et al., 2002). These segment-induced F0 characteristics of Korean stops are phonologized in Korean rather than being local phonetic effect at the onset of the vowel (Jun, 1993, 1995)

2.4.2 Perceptual studies on Korean stops

In addition to the identification of acoustic prosperities distinguishing the three types of Korean stops, some studies have investigated the role of the acoustic properties in the perception of the stops. In the one of the earliest investigations of Korean stops, Han and Weitzman (1970) investigated the importance of VOT and intensity characteristic at voice onset in the perception of the three stop types. The results of identification tests using re-synthesized CV stimuli revealed that aspirated stops and lenis stops were reliably identified from each by VOT modulation. In contrast, for the distinction of lenis stop from fortis stops, the intensity characteristic at voice onset was the most important cue. When intensity build-up at voice onset was gradual, greater lenis response was observed, and on the other hand, when intensity build-up was relatively rapid, fortis response was preferred by the listeners.

M.-R. Cho Kim (1994) examined productions of Korean stop consonants for various acoustic features and investigated their effects on the perception of the stops. Each of twelve male and female Korean speakers who were aged 25 to 35 years participated in the perception tasks. Three types of synthetic syllables, /ta/, /t^ha/, and /t*a/ were created, varying in VOT, F0, and aspiration intensity. A total of 72 synthetic stimuli were synthesized using Klatt's (1980) cascade/parallel formant synthesizer. The stimuli had 250 ms duration for each syllable and 12 VOT levels ranging from 5 to 82 ms in 7-ms steps. Also, the three F0 levels, 100, 125, and 150 Hz were employed, and the aspiration intensity varied in two levels, -33 and -21 dB. The results showed that the Korean listeners had greater /t^ha/ and /t*a/ responses when F0 increased. When VOT increased, they showed greater /t^ha/ and lesser /t*a/ responses, and for /ta/ response they showed a rise-fall contour at about 20 to 50 ms values of VOT. For the modulation of aspirated intensity, the listeners showed significantly greater /t^ha/ response for the stimuli of greater intensity, even though the difference was quite small. This result was contrary to the findings of production experiments because the speakers did not show significant differences in the intensity of aspiration across the three stop types in the production of the Korean stops.

More recently, M.-R. Kim et al. (2002) conducted perceptual experiments to determine the contribution of consonantal and vocalic cues to Korean listener's perception of Korean stops. Listeners were 12 speakers of the Seoul dialect of Korean and their age ranged from 26 to 32. The relative importance of VOT, H1-H2, and F0 was investigated using re-synthesized /Caŋ/ and /Ci/ syllables. For one experiment, the test syllables were cross-spliced at the vowel onset point and conjoined together. For instance, the consonant portion had information of aspirated stops and the vowel portion had properties of lenis stops. For another experiment, the consonant portion was removed and the vowel portion only was presented to the listeners. The results indicated that low F0 was a sufficient cue for lenis stops to be distinguished from aspirated or fortis stops, regardless of absence of consonantal information or conflicting information of the consonant portion. The fortis and aspirated vowel portions without the corresponding consonant portion tended to induce fortis response if the vowel portion had a relatively high F0. For aspirated and fortis distinction, VOT overrode the role of the vocalic information. With consonantal information absent, H1-H2 provided reliable information for the distinction of aspirated and fortis contrast. Overall, the results suggested that neither consonant information alone nor vocalic information alone provided sufficient cues for the three-way distinctions, but F0 was a dominant cue for the two-way distinction of lenis versus fortis or aspirated stops.

In another recent study investigating production and perception of Korean stops, M. Kim (2004) discussed a trade-off relationship between VOT and F0 for the distinction of aspirated and lenis stops. Fourteen listeners of the Seoul dialect at the age of 20s were presented CV syllables varying in F0 and VOT. The results of the perception tests indicated that the F0 boundary for aspirated and lenis stops decreased as VOT values increased. In addition, short VOT values were a sufficient cue for the perception of fortis stops, due to the very short values typically produced for fortis stops.

Finally, Y.-H. Kim (2007) investigated the effects of selective attention to VOT and F0 on the distinction of /t^ha/ and /ta/ syllables. Four sets of perception tests were conducted to the speakers of the Seoul dialect in their early 20s. The results suggested that the Korean listeners attended to both VOT and F0 to distinguish the lenis and aspirated stops. Overall, with longer VOT or higher F0, greater /t^ha/ response was induced, and greater /ta/ response was observed with shorter VOT and lower F0. However, the results of multidimensional scaling analyses and a speeded classification task revealed that the attention weight was greater for F0. Selective attention to F0 was more frequently observed than attention to VOT. Attention to VOT was influenced by F0, whereas attention to F0 was rarely influenced by VOT. When conflicting information was presented across /t^ha/ and /ta/ types for VOT and F0, the listeners showed greater attention to F0. Attention to F0 was retained even when the VOT difference between /t^ha/ and /ta/ stimuli was large. In short, the young Korean listeners attended more to F0 than to VOT for the distinction of /t^ha/ and /ta/ syllables.

To summarize the findings described in sections 2.4.1 and 2.4.2, Korean lenis stops are most reliably identified from aspirated and fortis stops by the F0 difference on the following vowel (lowest and higher/highest), and Korean fortis and aspirated stops are reliably identified from each other by the VOT difference (shortest and longest) and difference in voice quality (creaky/tense and breathy). In the next section, I discuss how these acoustic-phonetic properties associated with Korean stops are encoded differently in the speech of younger speakers compared with older speakers of the Seoul dialect of Korean.

2.4.3 Sound change in the Korean stop system

As shown in the previous sections, it is a well known fact that the three acoustic correlates, VOT, H1-H2, and F0, as a set, reliably distinguish the three types of Korean stops. However, when considering the two-way contrast of aspirated and lenis stops, a few studies have reported potential changes in the relative weight of the three acoustic correlates to the Korean stops.

Early investigations of Korean stops have reported an overlap between aspirated and lenis stops in VOT ranges for individual speakers (Lisker & Abramson, 1964; C.-W. Kim, 1965; Han & Weitzman, 1970). However, the overlap in VOT was more apparent with lenis and fortis stops than aspirated and lenis stops. Han and Weitzman (1970) found some overlap between aspirated and lenis types in VOT ranges, but the degree of overlap was much smaller compared with lenis and fortis types. C.-W. Kim (1965) also reported a similar pattern. Individual VOT values for lenis and fortis stops had an overlap for all of the three places of articulation, whereas the overlap for aspirated and lenis stops was found only with alveolar place. Moreover, when mean values across speakers and places of articulation were considered, VOT was considered as a reliable cue to distinguish the lenis and aspirated stops.

With recent investigations of Korean stops, the VOT overlap between lenis and fortis stops is rare. Rather, the overlap between aspirated and lenis stops is more frequently observed, and sometimes reversed VOT values have been observed. M. R. Cho Kim (1994), using 12 male and female Korean speakers of the Seoul and Pusan dialects, reported that mean VOTs for lenis and fortis stops never overlapped with each other and instead, VOT for lenis stops was rather longer than VOT for aspirated stops for Seoul female speakers. This pattern was not found with the speakers of the Pusan dialect. In addition, there was greater variability in VOT values for the speakers of the Seoul dialect compared with the Pusan speakers. In more recent studies, Kang and Guion (2006, 2008) have reported that mean VOTs for aspirated and lenis stops were roughly indistinct from each other for young Korean speakers of the Seoul dialect, who are around 20 years of age.

Wright (2007) attended to this diachronic trend and investigated production and perception of Korean bilabial aspirated and lenis, /p^h/ and /p/ stops. The 20 speakers of the Seoul dialect were born in 1955 for the oldest and in 1987 for the youngest. The VOT trends showed that as the age of the speakers decreased, VOT difference between the two stops decreased and that VOT for /p/ was even greater than for /p^h/ with some young speakers. In addition, younger speakers had less distinctive VOTs with greater overlap than older speakers. However, these correlations with speaker age were not found for F0 differences between the aspirated and lenis stops. Perception tasks using re-synthesized stimuli investigated whether younger speakers had greater reliance on F0 than older speakers for the perception of the distinction of /p^h/ and /p/ stops. The results showed an age-dependent difference between the youngest and oldest speakers, but there was not a linear relation between age and F0 weighting among the speakers as a whole. The author interpreted these uncertainties in the findings to be attributed to inappropriate experimental design and limited range in speaker age.

Silva (2004 et al, 2006) also noted the age-related diachronic change with production of lenis and aspirated stops and investigated the issue with speakers having wide range of age. Silva (2006) reported that VOT difference between aspirated and lenis stops has reduced and was even neutralized for some young Korean speakers of the Seoul dialect. Thirty six Korean speakers of the Seoul dialect born between 1943 and 1982 were examined for Korean stop productions. VOT values showed binary clusters among the speakers born before and after 1965. Speakers born before 1965 had significantly distinctive VOTs for aspirated and lenis stops. In contrast, speakers born after 1965 had markedly lower VOT for aspirated stops, resulting in non-distinctive VOTs for the lenis and aspirated stops, except for one speaker. In addition, five speakers born in 1970s and 80s produced lenis stops with longer VOT than that of aspirated stops. For F0, the correlation with speaker's age was not found. Thus, based on acoustic data and literature on Korean stops, Silva and colleagues (Silva et al., 2004; Silva 2006) have proposed that Korean stops are undergoing a diachronic change. VOT values for Korean aspirated stops have shortened over the past two generations. As a result, for young Seoul speakers, the distinction between lenis and aspirated stops, which has traditionally been rendered, at least partly, by VOT, has come to be coded primarily by a F0 difference on the vowels following the stops.

In the literature on Korean stops, the insufficiency of VOT alone to the distinction of the three-way contrast of Korean stops has been reported in the earlier studies and in turn, the voice characteristics of the following vowels, such as F0, intensity have been suggested as supplementary cues to the stop contrast (Lisker & Abramson, 1964; C.-W.

Kim, 1965; Han & Weitzman, 1970). However, to my knowledge, attention to the diachronic change in VOT values is only recent, and only a few investigations are available to date (Silva et al., 2004; Silva, 2006; Wright, 2007).

The current dissertation is designed to further the findings of these studies with more comprehensive experimental designs, especially with more fine-tuned perception tasks and various statistical analyses. In addition, the diachronic changes in Korean stops are investigated by examining hyperarticulated, clear speech, which is considered to be a speech mode revealing the closest approximation to the phonetic targets speakers have for the phonological categories. In hyperarticulated Korean stop productions, phonetic targets to the stop categories are predicted to be more closely approximated and the closely approximated phonetic targets are predicted to provide evidence to the diachronic changes in Korean stops. In Experiments 1 and 2 in Chapter III, the three types of Korean stops are produced in more casual, conversational speech, more careful, clear speech, and medium careful, citation-form speech to examine the varying realization of the phonetic targets adaptive to the different speaking styles in light of diachronic changes in Korean stops. In Experiment 3 in Chapter IV, the intelligibility improvement effect of clear speech is examined through a perception task. Finally, in Experiment 4 in Chapter V, the production data from Experiments 1 and 2 are examined for their relation to the perception of the stops through another perception task in which relevant acoustic correlates are manipulated.

CHAPTER III

PRODUCTION EXPERIMENTS: CHANGING PHONETIC TARGETS AND ENHANCEMENT STRATEGIES IN CLEAR SPEECH PRODUCTION OF KOREAN STOPS

This work has been published in Volume 124 of Journal of the Acoustical Society of America in 2008. Kyoung-Ho Kang was the first author of the published paper and all of the experimental work was performed by him. He wrote the first drafts of the paper, and Prof. Susan Guion, his advisor provided guidance for the experimental design and writing of the subsequent drafts.

3.1 Introduction

Chapter III presents findings from two production experiments investigating the contrast enhancement effect of clear speech for Korean stop production and its implication for sound change in the Korean stop system. The goal of the experiments was to investigate the proposal that the phonetic targets of phonological categories are more closely approximated in hyperarticulated speech (Johnson et al., 1993; Lindblom, 1990). For this purpose, the enhancement of acoustic correlates to Korean stop contrasts in hyperarticulated, clear speech was examined, compared with production in more casual speaking styles. In addition, it was investigated whether there was a difference in the clear speech enhancement strategies between two groups of Korean speakers, one of which had undergone a diachronic change in the Korean stop system (Silva et al., 2004; Silva, 2006; Wright, 2007). As a result of the diachronic changes, younger speakers are thought to have different phonetic targets than older speakers for Korean aspirated and lenis stops. These different phonetic targets are predicted to result in different enhancement patterns in clear speech production between the two groups.

Clear speech research has revealed an intelligibility improvement effect of clear speech for normal-hearing listeners (Ferguson, 2004; Krause & Braida, 2004), hearing impaired listeners, and listeners with limited experience in the test language (Bradlow & Bent, 2002; Ferguson & Kewley-Port, 2002; Bradlow & Alexander, 2007). In addition to this perceptual "clear speech effect," clear speech production has the effect of acoustic-phonetic modification of speech signals. Distinctiveness between phonological categories is enhanced in clear speech compared with conversational speech. This phonemic contrast enhancement in clear speech production has been found in single language studies (Picheny et al., 1986; Ferguson & Kewley-Port, 2002) as well as in cross-linguistic studies (Bradlow, 2002; Smiljanic & Bradlow, 2005).

The H & H theory (Lindblom, 1990) provides an explanation for the acoustic-phonetic modifications in clear speech production. The theory hypothesizes that talkers actively adjust articulatory effort according to the perceived difficulty of intelligibility for the listener. When intelligibility could be degraded, talkers will adopt a speech mode called hyperspeech. By hypothesis, talkers use greater articulatory energy in an adaptive effort to increase intelligibility for listeners, and this includes the enhancement of distinctiveness between phonological contrasts. When talkers do not estimate a need to improve intelligibility, they will default to a more economical system-oriented production, which is termed hypospeech. Thus, speech signals vary along the continuum between the two extremes of hypo- and hyperspeech. Clear speech can be viewed as production along the hyperarticulated end of the continuum.

Hyperspeech production may be investigated by the examination of specific phonetic targets associated with phonological categories. These phonetic targets are envisioned as stable, but their attainment variable, depending on production constraints, such as articulatory effort, corarticulatory force, and speech rate (Johnson et al., 1993; Moon & Lindblom, 1994). In this view, hyperarticulated clear speech production would approximate phonetic targets most closely. In contrast, hypoarticulation is a phonetic process in which the phonetic targets are underrealized. The implication of this theoretical view is that contrastive phonological categories would be maximally distinctive in hyperarticulated speech due to the greater approximation of phonetic targets.

As summarized in Chapter II, the Korean stop system manifests a three-way manner contrast termed aspirated, lenis (or lax), and fortis (or tense). In utterance-initial position, aspirated stops have the longest voice onset time (VOT) values, fortis stops have the shortest values, while lenis stops have intermediate values. Vowels following aspirated and lenis stops are breathier than the vowels following fortis stops (indicated by greater H1-H2 values). In addition, fundamental frequency (F0) on the vowels following aspirated and fortis stops is higher than that of vowels following lenis stops. However, the acoustic correlates related to the aspirated and lenis contrast have undergone a sound change in speakers born after roughly 1965 (Silva, 2006). In early studies on Korean stops (Lisker & Abramson, 1964; C.-W. Kim, 1965; Han & Weitzman, 1970), VOT values reported for aspirated stops were high (93–108 ms on average) and values for reported lenis stops were low (30–35 ms on average), so that VOT values for these two stop types did not overlap. As such, VOT was considered to be the primary acoustic correlate to the lenis-aspirated stop distinction and F0 was considered a redundant correlate (Silva, 2006, p. 298). In contrast, recent studies, examining speakers born after about 1965, have reported nondistinctive VOT values between lenis and aspirated stops (Silva et al., 2004; Kang & Guion, 2006; Silva, 2006; Wright, 2007). Silva et al. (2004) and Silva (2006) have proposed that Korean stops are undergoing a diachronic change and that VOT values for Korean aspirated stops have shortened over the past two generations. As a result, the distinction between lenis and aspirated stops, which was previously coded primarily by a VOT difference, has come to be coded primarily by a F0 difference on the vowels following the stops. Thus, older (born before 1965) and younger speakers may have different phonetic targets associated with the aspirated and lenis stop contrast.

If younger and older speakers have different VOT and F0 targets for the production of Korean aspirated and lenis stops, the H & H theory (Lindblom, 1990) predicts that this difference in phonetic targets will lead to different enhancement strategies. As phonetic targets are thought to be more greatly approximated in clear speech, different phonetic targets are predicted to result in difference in the usage of VOT and F0 between younger and older speakers in clear speech. Specifically, the question addressed here is whether younger and older speakers use different strategies to produce Korean aspirated and lenis stops in clear speech. Younger speakers are predicted to enhance F0 differences in clear speech, whereas older speakers are predicted to enhance VOT differences in clear speech.

3.2 Experiment 1: clear speech production of Korean stops

3.2.1. Participants

The data were collected from two groups of 11 native speakers of Korean. Each group had five male and six female speakers. The age in the "younger" group ranged from 20 to 29 years old and the participants were all born after 1977 (mean=25.8 years). Age ranged from 40 to 60 years old for the "older" group and the participants were all born before 1966 (mean=46.7 years). The speakers were affiliates of the University of Oregon, either foreign students, visiting scholars, or their family members. Self-reported daily use of English was low, especially with older speakers (15% for the older group and 34% for the younger group). None of the participants reported any hearing or speech disorders. All were speakers of the Seoul dialect of Korean.

3.2.2 Speech stimuli

The target words used for recordings were prepared to elicit the Korean stops in different speaking styles. For each of the nine Korean stops (aspirated, lenis, and fortis stops at three places of articulation), a pair of target words was selected for a total of 18 target words. Each of the 18 words had an initial CVC syllable, where the onset C was one of the nine Korean stops and V was the low vowel /a/ (see Appendix).

3.2.3 Procedure

To investigate potential contrast enhancement in clear speech production among the stop types, recordings were made in three different speaking styles: conversational, citation-form, and clear speech. Unlike previous studies on clear speech, which employed citation-form and clear speech only, the present study included a production task that aimed to elicit more natural "conversational" speech signals as well. Since eliciting true conversational speech in an experimental situation is difficult, this study assumes that the elicited conversational speech should be "more conversational" than citation-form speech even though it is not literally conversational. It is expected that the conversational speech will differ more when compared to clear speech than citation-form speech.

As the target words were repeated using different speaking styles, it was crucial to keep the target words in comparable prosodic and discourse environments. The target words were produced in utterance-initial position in all of the speaking styles. Additionally, all words were in a focused context. In conversational speech, they were focused as the topic of conversation; in the citation-form condition, they were focused as new information in a carrier phrase that was repeated across the trials; and in the clear speech condition, they were focused as new or contrastive information in an isolated context. Thus, prosodic and discourse environments were controlled, allowing for the unconfounded manipulation of the instruction to speak clearly.

Overall, a total of 2376 tokens were obtained, 792 tokens (18 target words × 2 repetitions × 22 speakers) for each of the three speech styles. Interaction between the experimenter and the speakers was conducted exclusively in Korean. The recordings were made separately for each speaker in a sound-attenuated booth using a high quality head-mounted microphone (Shure SM 10A) and a Marantz digital recorder (PMD 670). The utterances were digitally recorded at a sampling rate of 22 500 Hz and saved as wave files in a personal computer.

3.2.3.1 Conversational speech

Speech produced in a conversational style was elicited by asking speakers to explain the meaning of the target words or elaborate on the contexts where the target words are commonly used. The target words were real words whose meaning could not be easily explained in one or two words. The aim of choosing real word stimuli with complex definitions was to orient the speakers' attention towards exploring the word meaning. This task was performed first, and speakers were not told about other tasks before beginning. This was to prevent the speakers from noticing the links between the production tasks in different speaking styles.

Each of the 18 target words were presented individually on flash cards using Korean orthography. The target words were presented to each speaker in a random order. The speakers were allowed enough time to explore the meaning of the words. While speakers were giving explanations, the experimenter responded to the speaker by agreeing, disagreeing, or giving his own opinion to the speaker's explanations. This was to make the interaction more 'conversational'.

To make sure that the target words were included in the explanations speakers gave, the speakers were asked to indicate which word they were explaining. The explanations they gave were typically constructed as Korean sentences beginning /______ + nɨn .../, or /______, igʌ + nɨn .../, meaning '_____ +TOPIC MARKER ....', or '______, this +TOPIC MARKER ....'. The first two productions of a given target word that occurred in utterance-initial position were submitted to analysis.

3.2.3.2 Citation-form speech

After the task for conversational speech ended, citation-form speech was recorded. The 18 target words were presented in a random order on 18 flash cards two times each. The speakers produced each target word in a frame sentence, /_______ haseyo./ In English this translates to 'Say _______.' Speakers were instructed to produce the sentences at a comfortable speech rate and loudness level. When producing the words, the speakers were asked to read one card at a time and hand it back to the experimenter. This was to prevent the speakers from reading too fast as is commonly observed when words are read in a list. Thus, they had to pause before reading the next flash card.

3.2.3.3 Clear speech

In order to elicit clearly produced speech, a situation in which listener-oriented careful speech is required was set up. The speakers were asked to imagine that they were teaching Korean to second language learners of Korean and reading out Korean words to their students. In order to assist the speaker reading the different target words clearly, each of the 18 target words was presented with another target word that shared a place of articulation but differed in the type of the initial stop (e.g., 탄탄하다 /tʰan.tʰan.ha.ta/ and 단단하다 /tan.tan.ha.ta/; 단단하다 /tan.tan.ha.ta/ and 딴딴하다 /t*an.t*an.ha.ta/; 딴딴하다 /t*an.t*an.ha.ta/ and 탄탄하다 /tʰan.tʰan.ha.ta/). So, each flash card had two of the 18 target words, but only the production of the first word of each pair was submitted to the acoustic measurements. Speakers were instructed to read the target words as clearly as possible and no other instructions, such as to read loudly or slowly, were given to the speakers. In this manner, two productions of each target word were elicited.

3.2.4 Measurements

For each of the tokens obtained from the production tasks, three acoustic properties encoding Korean stop contrasts were measured using PRAAT (Boersma & Weenink, 2005). VOT was measured as the time duration between the point of stop burst release and the onset of the periodic portion of the waveform for the initial vowel of the target words. In other words, the VOT measure used here was a true measure of voice onset and did not include a period of nonmodal phonation (if present), following the method described by Lisker and Abramson (1964). Other approaches to measuring VOT have been used for Korean stops. See Wright, 2007, pp. 14–30 for a discussion. H1-H2 and F0 were measured on the following vowel. H1-H2, which is the amplitude (dB) difference between the first (H1) and the second (H2) harmonic, was measured just after the first full glottal pulse of the waveform. The onset of the vowel in the waveform was determined by the onset of the first full glottal pulse of the vowel in conjunction with the second formant. The amplitude values were calculated using a narrowband fast Fourier transform spectrum (window length of 30 ms). F0 was measured at the temporal midpoint of the vowel using the pitch tracking function in PRAAT. When the pitch line abruptly moved or was discontinued, F0 was recorded by measuring the duration of the relevant period in milliseconds and dividing it by 1000.

3.2.5 Results

Several statistical tests were used to assess the predictions that the acoustic correlates encoding Korean stop contrasts would be enhanced in clear speech production and that the enhancement would be made differently by younger and older speakers. First, a multivariate mixed analysis of variance (MANOVA) was performed with the following factors: group (younger and older), speaking style (conversational, citation-form, and clear), and stop type (aspirated, lenis, and fortis), with the last two factors treated as repeated measures. Given that Korean stops contrast with one another in VOT, H1-H2, and F0, these three acoustic measures were all entered as dependent variables. The individual data points submitted to analysis were an average of the two repetitions of each test item for each speaking condition for each measure. The MANOVA returned significant main effects for group [$F(3,128) = 5.34$, $p < 0.001$], speaking style [$F(6,125) = 54.357$, $p < 0.001$], and stop type [$F(6,125) = 893.413$, $p < 0.001$]. Significant two-way and three-way interactions were found for all combinations of these three factors [F-values ranged from 6.21 to 19.207, $p < 0.001$]. These results indicate that the Korean stops were produced differently by the different groups, in different speaking styles. In order to further explore the interactions, the data were split by age group, and then the effects of speaking style and stop type on each of the three acoustic properties were investigated in separate repeated measure ANOVAs. Both subject (F1) and item (F2) analyses were run. Both factors were treated as repeated measures in the subject analysis, but only speaking style was treated as repeated in the item analysis. Due to the fairly large degrees of freedom, the most conservative statistic, the lower bound, is reported here, and the focus of the analysis is on the interaction of speaking style and stop type. Main effects are only reported in the absence of significant interactions. Our hypotheses revolved around the interactions, as we sought to determine the nature of the speaking style effect as modulated by stop type for each dependent measure.

3.2.5.1 Younger group

For the younger group, univariate tests for each of the dependent measures VOT, H1-H2, and F0, revealed significant interactions for factors speaking style and stop type for both subjects and items analyses (VOT [$F1(1,65) = 10.220$, $p = 0.002$, $F2(2,15) = 12.223$, $p = 0.001$], H1-H2 [$F1(1,65) = 11.730$, $p = 0.001$, $F2(2,15) = 10.091$, $p = 0.002$], and F0 [$F1(1,65) = 12.281, p = 0.001, F2(2,15) = 26.496, p < 0.001$]). Thus, the patterns found in the subjects analyses are consistent across different test items. These interactions are seen in Figure 3.1 and interpreted in the paragraphs below.

As shown in Figure 3.1 (a), the interaction for VOT is due to differences in the patterning for the stop types across the speech styles. VOT for aspirated stops slightly increased in citation-form and clear speech while that for lenis stops slightly decreased. As a result of this, the VOT distance for the aspirated and lenis stop contrast expanded in clear speech.

As for the H1-H2 measure, the interaction is due to the enhancement pattern of fortis stops (see Figure 3.1(b)). Vowels following fortis stops had a more negative H1-H2 value in clear speech, indicating greater creaky voice quality in this speech style. On the other hand, the other stop types showed minimal effects of speech style. The interaction for F0 is seen in Figure 3.1(c). F0 increased in clear speech production for all of the three stop types. However, because the F0 increase in clear speech for vowels following aspirated and fortis stops was relatively larger than that of lenis stops, the F0 distance between lenis stops and the other two stop types was expanded in clear speech.

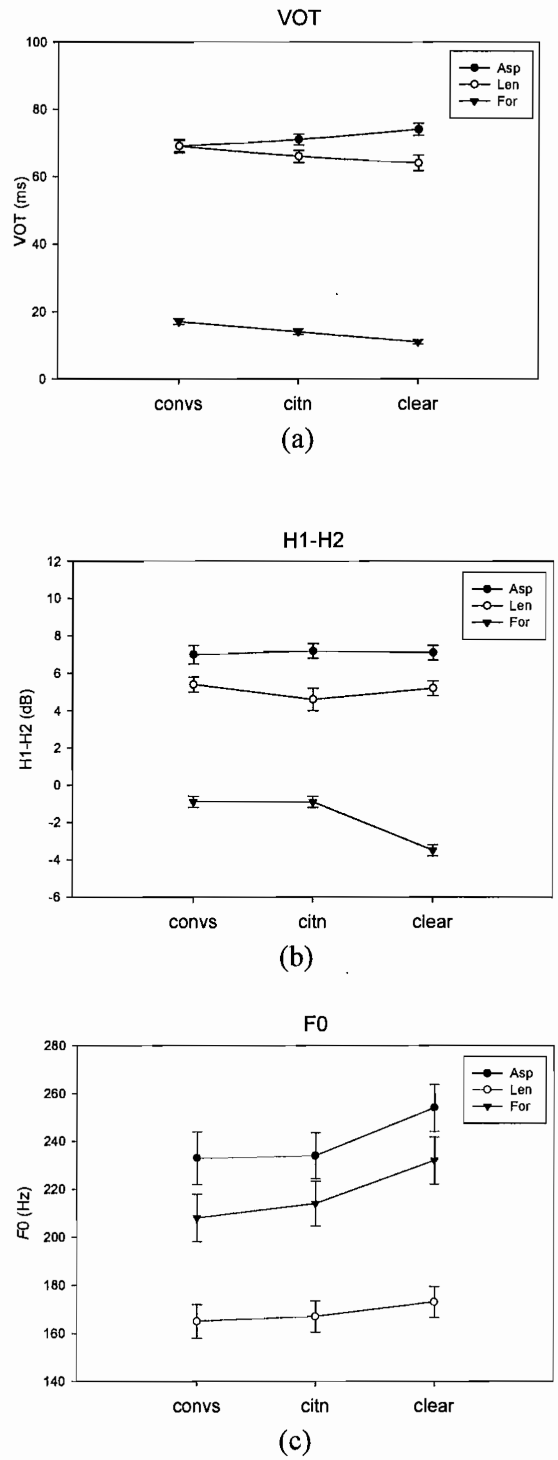


Figure 3.1 Mean values with standard errors for the production of Korean stops (aspirated, lenis, fortis) in conversation, citation-form, and clear speech styles by the younger group (n=11) for three acoustic correlates [(a) VOT; (b) H1-H2; (c) F0].

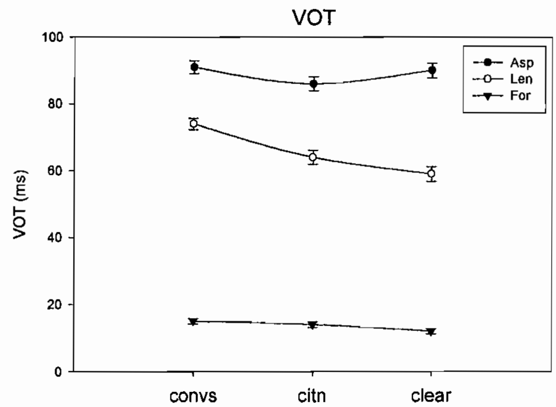
3.2.5.2 Older group

For the older group, univariate tests returned significant interactions of speaking style by stop type for each of VOT, H1-H2, and F0 (VOT [$F1(1,65) = 14.717, p < 0.001$, $F2(2,15) = 7.867, p = 0.005$], H1-H2 [$F1(1,65) = 5.826, p = 0.019$, $F2(2,15) = 5.850, p = 0.013$], and F0 [$F1(1,65) = 12.208, p = 0.001$, $F2(2,15) = 19.223, p < 0.001$]). Interactions between speaking style and stop type are displayed in Figure 3.2.

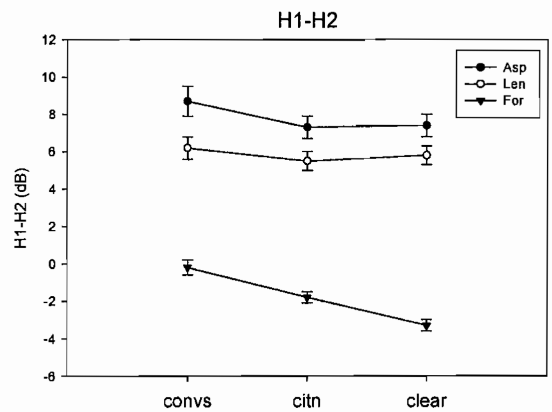
Figure 3.2(a) shows that the VOT distance between aspirated and lenis stops was expanded in clear speech. Mean VOT for lenis stops decreased substantially, whereas that for aspirated stops was roughly the same between conversational and clear speech.

As for the H1-H2 measure (see Figure 3.2(b)), the interaction was due to the substantial reductions found in citation-form and clear speech for fortis stops. These results indicate that fortis stops were produced with creakier voicing on the following vowels in these speaking styles, whereas the other two stop types showed little effect of speech style.

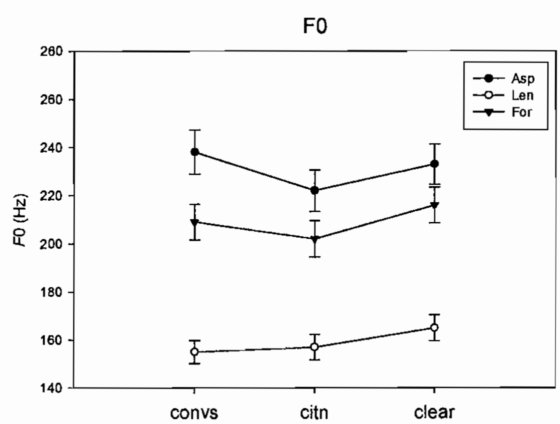
Finally, for the F0 measure, the interaction was primarily reflected in the patterning in citation-form speech (see Figure 3.2(c)). The mean F0 for aspirated and fortis stops decreased vis-a-vis the other two speaking styles. Also, in clear speech, mean F0 for aspirated stops slightly decreased compared to conversational speech, whereas mean F0's for lenis and fortis stops increased. Consequently, the overall F0 distance among the stop categories was slightly reduced in clear speech.



(a)



(b)



(c)

Figure 3.2 Mean values with standard errors for the production of Korean stops (aspirated, lenis, fortis) in conversation, citation-form, and clear speech styles by the older group (n=11) for three acoustic correlates [(a) VOT; (b) H1-H2; (c) F0].

3.2.6 Discussion

The results of the analyses suggest that the phonological contrasts between the Korean stops are enhanced in clear speech production for both the younger and older speakers, but in a different manner. The younger speakers produced Korean aspirated and lenis stops with merged VOT values in the conversational condition. However, in clear speech, the two stops were differentiated with a small VOT difference of 10 ms. In addition, the H1-H2 difference between the fortis and the other two stop types was more exaggerated in clear speech, as the following vowels became even creakier in the context of fortis stops, but not in other stop contexts (the difference between aspirated and fortis stops in conversational speech was 7.9 dB, and 10.7 dB in clear speech). The F0 difference between the lenis and the other two stop types was also enhanced in clear speech by a greater F0 increase in the vowels following the aspirated and fortis stops than the lenis stops (the difference between aspirated and lenis stops in conversational speech was 68 Hz, and 81 Hz in clear speech).

For the older speakers, a VOT difference between aspirated and lenis stops was found in conversational speech, but the difference became even greater in clear speech (18 ms difference in conversational speech and 31 ms difference in clear speech). As for the H1-H2 measure, fortis stops and the other two stop types, aspirated and lenis, became more distinctive from each other in clear speech due to enhanced creaky voicing on the vowels following fortis stops (the difference between aspirated and fortis stops in conversational speech was 9 dB, and 10.7 dB in clear speech). Finally, in contrast to the case of the younger speakers, the F0 difference between the stop categories reduced in clear speech production (the difference between aspirated and lenis stops in conversational speech was 84 Hz, and 68 Hz in clear speech). As a result, F0 did not enhance any phonological distinctions in clear speech production in the case of the older speakers.

To sum up, these results indicate that there is a difference in the use of VOT and F0 to enhance the aspirated and lenis stop contrast between the two age groups. The older group used VOT, whereas the younger group used both VOT and F0, with F0 being a stronger correlate. However, interpreting the difference in F0 enhancement in clear speech between the two groups requires closer examination. One question that arises is whether the greater F0 difference in clear speech for the younger speakers was induced by the enhancement of a phonetic target used to realize the phonological contrast of aspirated and lenis stops, or whether it was due to an overall greater F0 use by the younger speakers in the clear speech style. That is, an alternative interpretation of the results may be that the younger speakers generally use F0 more robustly than older speakers, and that the F0 enhancement in clear speech is simply attributable to a larger F0 range in clear speech. In order to test this alternative possibility, another experiment was conducted with a separate set of test items containing nonstop consonants.

3.3 Experiment 2: F0 modification in stop and non-stop consonants in clear speech production

Experiment 2 examines F0 variation of aspirated consonants as a function of speaking style in comparison with lenis consonants. The purpose was to investigate the question of whether the F0 enhancement of the younger speakers in experiment 1 was due to a phonological contrast enhancement between the aspirated and lenis stops or an expanded F0 range in clear speech. Specifically, the phrase initial tonal contrast between the high-tone associated Korean aspirated and fortis stops and the low-tone associated lenis stops (Jun, 1993; Cho et al., 2002) was investigated in comparison with other nonstop consonants showing a similar tonal specification phrase initially. Thus, a high-tone triggering voiceless fricative /h/ and a low-tone triggering voiced nasal /n/ were examined for their F0 realization in citation-form and clear speech. If the larger F0 difference in clear speech for the younger speakers were induced by a generally greater pitch enhancement regardless of segment type, the F0 difference for the vowels following /h/ and /n/ would be greater in clear speech than in citation-form speech, similar to the pattern found for the aspirated and lenis stops. On the contrary, if the F0 differentiation was part of the enhancement strategy for the phonological contrast of the aspirated and lenis stops, the F0 difference between the vowels following /h/ and /n/ would not be enhanced in clear speech for either the younger or the older speakers.

3.3.1 Participants

Twenty native Korean speakers who did not participate in Experiment 1 were recruited. The participants were grouped separately based on their age. The ten younger speakers were 19 – 29 years old (mean of 23 years) and the ten older ones were 40 – 58 years old (mean of 47 years). For both groups, three male speakers and seven female speakers participated in the experiment. All of the speakers reported that they were born in Seoul or Kyung-gi province (i.e., the greater Seoul area) and used the Seoul or Kyung-gi dialect of Korean.

3.3.2 Procedure

Each of the 20 speakers read 35 Korean words in two speaking styles, citation-form and clear speech. For the citation-form style, the speakers used their own comfortable speech rate and loudness level. For the clear style, they were instructed to read in a 'clear' way, as if speaking to a 'foreigner' audience who needs greater linguistic-phonetic resources to have full access to the linguistic information.

The carrier phrase, /____ la.ko mal.ha.se.yo/, meaning 'Say ____ + QUOTATIVE PARTICLE,' was used to ensure that the target words were produced in the same prosodic context in both speaking styles. Each word began with an aspirated or lenis consonant. The initial consonants examined in the analysis were /t^h, h/ for the aspirated series, and /t, n/ for the lenis series. The words were 타다 /t^ha.ta/ 'to ride', 하다 /ha.ta/ 'to do', 다도 /ta.to/ 'tea ceremony', and 나비 /na.pi/ 'butterfly.' Each word was produced three times in three separate blocks for each speaking style. The reading was recorded at a sampling rate of 22,050 Hz using a portable Marantz digital recorder (PMD 670) and a high quality headset microphone (Shure SM10A). The recording was made in a sound-attenuated booth or in a quiet room at the participants' home. The recorded utterances were saved to a computer in a wave sound format. F0 was measured on the first syllable of the utterance-initial target words. F0 was obtained at the temporal midpoint of the vowel using the pitch analysis function of Praat.

3.3.3 Results

The F0 measurements for the aspirated /t^h/ and /h/ and the lenis /t/ and /n/ consonants were submitted to a three-way repeated measures analysis of variance (ANOVA). Speaking style (citation-form, clear speech) and condition (stop /t^h, t/, continuant /h, n/) and consonant type (aspirated / t^h, h/, lenis /t, n/) were entered as factors. Mean F0 values across three repetitions of each test item were used as the dependant measure.

Analyses were conducted separately for each of the younger and older speaker groups. The test for the younger speakers returned significant main effects for speaking style [$F(1,9) = 9.284, p = 0.014$], condition [$F(1,9) = 29.269, p < 0.001$], and consonant type [$F(1,9) = 85.582, p < 0.001$]. A significant interaction was found for speaking style by consonant type [$F(1,9) = 5.471, p = 0.044$] and for speaking style by condition by consonant type [$F(1,9) = 25.081, p = 0.001$]. These results indicate that the effect of speaking style on the F0 difference between aspirated and lenis stop stops, /t^h/ and /t/, was different from that found for the fricative and the nasal, /h/ and /n/. As shown in the

Table 3.1, the F0 difference for the aspirated and lenis stop contrast increased in clear speech (63 Hz versus 76 Hz) due to a higher F0 for aspirated stops, whereas the F0 difference for the fricative and nasal contrast largely stayed the same (69 Hz versus 72 Hz) between citation-form and clear speech.

In contrast, in the case of the older speakers, the main effect of consonant type (aspirated / t^h, h/, lenis /t, n/) was the only significant finding [$F(1, 9) = 59.077, p < 0.001$]. None of the other main effects or interactions were significant ($p > 0.05$). The significant effect of consonant type is due to the large F0 difference between aspirated (/t^h, h/) and lenis (/t, n/) consonant types. The results indicate that the older speakers did not differentially enhance F0 for the aspirated and lenis stop contrast vis-à-vis that of the continuant contrast (/h/ and /n) between the speaking styles. As shown in the Table 3.1, the mean F0 differences for the stop contrast, as well as the continuant contrast, marginally increased in clear speech (stops: 65 Hz versus 67 Hz; continuant: 68 Hz versus 69 Hz).

Table 3.1. Mean F0 for stop and non-stop consonants for younger and older group with standard error in parenthesis

	Younger (n=10)		Older (n=10)	
	Citation-form	Clear	Citation-form	Clear
/t^h/	231 (20.2)	244 (20.7)	227 (20.8)	235 (22.7)
/t/	168 (14.8)	168 (13.7)	162 (13.9)	168 (14.8)
/h/	228 (20.2)	236 (20.6)	224 (20.0)	234 (22.3)
/n/	159 (14.0)	164 (13.7)	156 (12.5)	165 (13.9)

3.3.4 Discussion

The results indicate that the greater F0 difference in clear speech for the younger speakers is specific to the aspirated and lenis stops: F0 difference substantially increased in clear speech for the stop contrast (/th/ versus /t/), but increased only slightly for the continuant contrast (/h/ versus /n/). Older speakers showed a similar difference in the two speaking styles for both contrast types. Combined with the results from experiment 1, these results suggest that the greater F0 difference for the aspirated and lenis stops found for the younger speakers can be attributed to the greater approximation of the high F0 target associated with aspirated stops in clear speech. The results were not consistent with the alternative hypothesis of an overall expanded F0 range for the younger speakers.

3.4 General discussion

Based on the H & H theory (Lindblom, 1990), we predicted that phonological contrasts would be enhanced in clear speech by more fully realizing phonetic targets in an effort to aid listeners in speech perception. Given the sound change in the Korean stop system (Silva et al., 2004; Silva, 2006; Wright, 2007), younger and older Korean speakers were predicted to have different phonetic targets for stop production and, in turn, to display different enhancement patterns in clear speech production. Specifically, for the aspirated and lenis stop contrast, younger speakers were predicted to enhance F0 differences, whereas older speakers were predicted to enhance VOT differences.

The results of the two experiments upheld these predictions. The three-way manner contrast of Korean stops was found to be enhanced in clear speech compared with conversational or citation-form speech. Furthermore, there was a difference in the use of VOT and F0 between the younger and older speakers. The older group used VOT in order to enhance the aspirated and lenis contrast in clear speech. The younger group enhanced the difference between aspirated and lenis stops by increasing the relative F0 of aspirated stops in clear speech. VOT also slightly diverged for the two stop types in clear speech but to a much lesser degree than the older speakers.

This small VOT usage by younger speakers merits further discussion. One possibility is that the small VOT usage may reflect a small difference in VOT targets between aspirated and lenis stops that is only revealed in clear speech. Phonetic targets may have shifted such that F0 is used as a primary acoustic correlate and VOT as a secondary one by younger speakers. In other words, the small VOT usage for the younger speakers may reflect a decreased, yet residual, role of VOT in the distinction of aspirated and lenis stops. Another possibility is that the small VOT usage may not reflect differences in VOT targets between the aspirated and lenis stops. In other words, the stops have the same or quite similar VOT targets. The small VOT dispersion in clear speech may rather reflect an attempt to use acoustic correlates heard in the speech of others. That is, exposure to older speakers, who maintain VOT differences, has resulted in the partial ability to manipulate VOT, as a type of style shifting or mimicking of a more traditional speaking style that employs VOT.

Another finding from this study was that the older group enhanced the VOT difference by lowering the VOT for lenis stops and keeping the VOT for aspirated stops relatively stable across the speaking styles. One may wonder why the VOT for the aspirated stops did not also increase. We believe an explanation may lie in the prosodic environment in which the stops were produced. In studies investigating varying realizations of VOT of Korean stops as a function of different prosodic contexts, Cho and Jun (2000) and Cho and Keating (2001) reported that the VOT values for aspirated and lenis stops in higher domain-initial positions (such as intonational phrase or utterance initial) was greater than in lower domain initial positions (such as accentual phrase initial). As stimuli were elicited in the utterance-initial position in the current study, we would expect high VOT values for both aspirated and lenis stops due to the domain-initial strengthening effect[1]. In the case of aspirated stops, these high VOT values may be close to the phonetic target, but in the case of lenis stops, these high VOT values may overshoot the phonetic target. Thus, in clear speech production, the lower VOT target for lenis stops may be more closely approximated, resulting in lower VOT values for clear speech. On the other hand, if the utterance-initial position had already conditioned targetlike values for the aspirated stops, an increased VOT in clear speech would not be expected. The idea that lenis stops have a relatively low VOT target is supported by previous investigations. Early studies report VOT values for lenis stops in the range 21–41 ms, sometimes even overlapping with fortis stops (Lisker & Abramson, 1964; C.-W. Kim, 1965; Han & Weitzman, 1970). More recent studies with speakers born in the 1960s or earlier report VOT values from approximately 5 to 40 ms for lenis stops (Cho & Keating, 2001; Wright, 2007).

---

[1] The VOT strengthening effect in higher domain-initial positions has been found in several languages (Hsu and Jun, 1998, in Taiwanese; Jun, 1993, in Korean; and Pierrehumbert and Talkin, 1992, in English), suggesting that it may be a general articulatory process.

With respect to the nature of the Korean stop sound change, one possibility is that the sound change is incremental and, thus, related to speaker's age in a continuous fashion. Another possibility is that the sound change has occurred in the younger speakers but not in the older speakers and, thus, evidence for the change will only be found in the younger group. An examination of individual production revealed no appreciable F0 enhancement by any speakers in the older group, suggesting that they have not been affected by the sound change. In the case of the younger group, there was no apparent relationship between age (speakers ranged from 20 to 30 years old) and the amount of F0 or VOT enhancement. However, a sample larger than 21 younger speakers in the two experiments may be needed to detect such a gradient relationship. Additionally, data from speakers in the 30–40 year age range would prove probative to the question. Given these findings, we tentatively propose that speakers of the Seoul dialect born in the 1970s or later have, as a whole, undergone the sound change and that speakers before this have not been affected by it.

The H & H theory (Lindblom, 1990) hypothesizes that talkers make articulatory modifications in response to the estimated perceptual needs of the listeners, and such modifications are predicted to result in enhancement of phonological contrasts. This chapter presented an instance of stop contrast enhancement in clear speech in Korean, lending typological support to the language universal predictions of the H & H theory. Given that clear speech aims to provide listeners with more accessible speech signals, what remains for further study is to experimentally test whether such articulatory modifications result in the improvement of intelligibility for listeners. To this end, a perception study testing the intelligibility improvement effect of clear speech is required. Also, as the Korean stop system is experiencing changes in the acoustic correlates to the stops, differences are predicted between older and younger listeners in the extent to which each acoustic correlate contributes to the perception of the stops. Perceptual experiments investigating these questions are reported in Chapters IV and V.

CHAPER IV

PERCEPTION EXPERIMENT 1: INTELLIGIBILITY TEST

4.1 Introduction

Given that by definition, clear speech aims at greater intelligibility, the primary goal of Experiment 3 in this chapter was to investigate the "clear speech effect" of clear speech production on speech perception (e.g., Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Ferguson, 2004; Ferguson & Kewley-Port, 2002; Krause & Braida, 2004). Here the clear speech effect refers to the intelligibility improvement of clear speech compared with conversational speech. To investigate the intelligibility improvement effect of clear speech, the three different speaking styles in Experiment 1, conversational speech, citation-form speech, and clear speech productions were tested for intelligibility in this chapter. Based on the hypothesis that speakers produce more "exaggerated" or "hyperarticulated" speech to accommodate hearers experiencing speech perception difficulty (Lindblom, 1990), Experiments 1 and 2 in Chapter III predicted that articulatory modifications aiming at intelligibility improvement include enhancement of the distinctiveness between phonological categories. The results of the two production experiments indicated that the acoustic distances between Korean stop contrasts were expanded in clear speech compared with conversational and citation-form speech.

Experiment 3 aims to investigate whether this contrast enhancement in clear speech production would lead to the improvement in the intelligibility of the stops. For this purpose, the recordings from Experiment 1 were submitted for the preparation of the speech perception stimuli of Experiment 3. The specific questions addressed in Experiment 3 were whether Korean stops produced in a clear speech style were more intelligible than those produced in a conversational style, and at the same time whether there was a difference in intelligibility between the conversational and citation-form speaking styles and between the citation-form and clear speaking styles. Thus, I examined whether there was a progressive improvement across the speaking styles from conversational, to citation-form, to clear speech.

4.2 Experiment 3: noise-masked stimuli perception task

4.2.1 Participants

Eleven native speakers of Korean participated in the perception task. All of the eleven participants were female speakers and their length of staying in the U.S. ranged from 0.3 to 8 years (mean = 3.25 years). None of the listeners had participated in Experiment 1 or 2. All of the participants were speakers of the Seoul or Kyung-gi dialect[2] of Korean and all were in their twenties (mean = 23 years). The participants were all international students at the University of Oregon and reported that they used

[2] A Korean dialect spoken in a Kyung-gi province of South Korea. Kyung-gi province refers to the north-west part of South Korea which surrounds the city of Seoul. Kyung-gi dialect is commonly referred to as a Seoul-Kyung-gi dialect of Korean.

Korean in daily activities with their Korean friends or family members. None of the listeners reported any history of language disorder or hearing problems.

4.2.2 Perception stimuli

The syllables used for the syllable-in-noise perception task in Experiment 3 were taken from the recordings for Experiment 1. The initial syllable of the three target words, 탄탄하다 (/t^han.t^han.ha.ta/), 단단하다(/tan.tan.ha.ta/), and 딴딴하다 (/t*an.t*an.ha.ta/) was excised and used for the preparation of the perception stimuli. So, three types of test syllables, /t^han/, /tan/, and /t*an/ were used in the perception tasks. Because the 11 younger speakers (six females and five males) in Experiment 1 produced these target words twice for each of the three speaking styles (conversation, citation-form, and clear speech), 18 syllables (2 productions × 3 target words × 3 speaking styles) were submitted to the preparation of test syllables for each of the 11 speakers. Accordingly, 198 test syllables were prepared in total. The speech signals from older speakers in Experiment 1 were not used. This was because the primary interest of Experiment 3 was to examine the possible intelligibility benefit of clear speech over conversational and citation-form speech. The focus of the experiment was to investigate whether the manipulation of speaking style in Experiment 1 would lead to intelligibility differences across the speaking styles. To this purpose, the participant groups were aligned according to age, such that younger listeners listened to younger speakers' speech tokens. Generational difference in intelligibility improvement benefit of clear speech for matched speakers and listeners was not predicted. Therefore, only one generational group was studied in this experiment.

In order to equate amplitude over the test stimuli, speaking styles, and 11 speakers, the amplitude of all of the test syllables was normalized in terms of Root Mean Square (RMS) amplitude. Because most of the energy was distributed on the vowel /a/ and nasal /n/ portions and also because the boundary between the vowel and the nasal was hard to determine, the normalization process was performed to the combined amplitude of the vowel and nasal portion. In other words, the amplitude of the entire syllable, that is, stop + vowel + nasal was normalized to the combined amplitude level of the vowel + nasal portion. Each of the test syllables went through the following procedure. First, since the normalization level was set at 70 dB SPL, a sound pressure value in Pascal (p) equivalent to 70 dB sound pressure level (SPL) was obtained based on the following formula.

$$70 \text{ SPL (dB)} = 20 \log \frac{p}{p_0}$$

where p is a value in Pa, and p_0 is 0.00002 Pa

The obtained Pa value equivalent to 70 dB SPL was 0.063245. Then, the RMS amplitude of each of the test syllable was normalized to this Pascal value of 0.063245. To do this, the RMS amplitude level of the vowel and nasal portion was taken in Pascal, and the ratio of the amplitude of this vowel and nasal portion (in Pa) to 0.063245 Pa was obtained. Then, the obtained ratio was applied to the RMS amplitude of the entire test syllable.

The ratio was less than 1 when the amplitude of the vowel and nasal portion (in Pa) was less than 0.063245 Pa (70 dB), and the ratio was greater than 1 when the amplitude of the vowel and nasal portion was greater than the reference level of 0.063245 Pa. In this way, the test syllables were normalized in amplitude in a way that the average amplitude value over the entire syllable (that is, RMS amplitude of the entire syllable) is minimally affected by the varying amplitude structure over the three stop types[3].

As the next step in preparing the test stimuli, the normalized test syllables were then mixed with a multi-talker babble noise. First, the speech-like babble noise was created using the speech of six speakers. In order to produce multi-talker speech in a lab setting, six Korean speakers read a short Korean text simultaneously. Then, to make the reading unintelligible noise, each speaker started to read at different times so that the readings of the six speakers have different phases in time. As a result, the readings were completely unintelligible (as judged by the author) and sounded like the "buzzing" noise produced when a large group of people talk simultaneously. The recordings were made in a sound-attenuated booth using a Shure B.G 5.1 condenser microphone and a Marantz digital recorder (PMD 670) at the sampling rate of 22,500 Hz. The six speakers were standing around the microphone approximately 2 feet away while they were reading.

After saving and storing the recordings in a wave file, two 30 second stretches of the recorded noise signal were randomly cut out and were mixed with each other to make

---

[3] The fortis stop + vowel + nasal type, that is, /t*an/ type had the greatest ratio of vowel + nasal duration to the entire syllable duration compared with the aspirated or lenis stop vowel + nasal syllable (/t^han/ or /tan/) due to the shortest duration between stop release to the vowel onset of the /t*an/ type.

a 12-talker babble noise (Ferguson and Kewley-Port, 2002; Ferguson, 2004). From this 30 second stretch of the noise file, a one second portion without an apparent pitch contour was excised and used as the source noise to make a masking noise. This process was used to ensure that the noise signals were not intelligible and thus, they only functioned as a background noise masking the test syllables.

For the last step of creating the test stimuli, the multi-talker masking noise and the test syllables were mixed with each other. The mixing procedure was as follows. Each of the test syllables was mixed with the 1 second source noise. The length of the source noise was determined by the duration of the test syllables. The test syllable was centered within the noise with a head and tail noise of 100 milliseconds (ms). This required the noise to be 200 milliseconds longer than the test syllable. So, for example, if the test syllable was 450 ms, the noise signal was edited to 650 ms from the beginning out of the entire 1000 ms duration. After each of the noise files was edited according to the duration of the test syllables, amplitude of each of the noise signals was rescaled in dB. The signal to noise ratio (S/N ratio) was chosen at -6 dB S/N after a series of pilot tests at 0, -1, -3, and -6 dB S/N. At -6 dB, the rate of correct response reached up to 85 to 90 % for the clear speech stimuli. The other S/N ratios showed a ceiling effect and thus, they were not selected. Through the mixing processes, a total of 198 stimuli was created (2 productions × 3 types of test syllables (t^han, tan, t*an) × 3 speaking styles (conversational, citation-form, clear) × 11 speakers).

4.2.3 Procedure

The 198 test stimuli were presented in a random order across the 3 syllable types, 3 speaking styles, and 11 speakers. Pilot tests returned a ceiling effect when the test stimuli with speaker blocked, due to the salience of the role of F0 for the distinction of the lenis stops from the other two stop types. During an informal talk with the experimenter, the author, the participants reported that they easily noticed F0 difference and were inclined to overly attend to F0 in identifying the /ta/ type from the other two syllable types. Additional pilot tests with greater S/N ratio did not solve this problem. In order to reduce the reliance on given speaker's F0 and avoid a ceiling effect, the stimuli were presented with randomized across the speakers.

Each of the test stimuli was presented four times and thus, 792 responses were collected from each of the 10 listeners. The listeners participated in a forced-choice identification task. Three response categories of 탄 /t^han/, 단 /tan/, and 딴 /t*an/ were displayed on a computer screen in Korean orthography, and the participants were asked to choose one of these three choices when listening to the stimuli delivered binaurally through headphones (Sony MDR-7506). A small set of practice stimuli was provided before the main task to familiarize participants with the format of the task. The entire task took about 50 minutes including a practice session and it was conducted in the phonetics lab at the University of Oregon. The listeners who participated in the pilot tests were excluded from the participant pool of the current experiment.

4.2.4 Statistical design

For the dependent variable, the number of correct responses to each syllable type was entered. Because the question of interest was whether for each syllable type there was a difference in the level of intelligibility between the speaking styles, the number of correct responses was entered in a matrix of syllable type by speaking style. For each syllable type (/t^{h}an/, /tan/, and /t*an/), two test stimuli were created for each of the 11 speakers. In addition, each test stimulus was presented four times. So, for example, the /t^{h}an/ type had 88 instances in the entire stimulus set for each speaking style. The number of the correct responses for these 88 instances for each listener was entered as a data point for the dependent variable. Syllable type (/t^{h}an/, /tan/, and /t*an/) and speaking style (conversational, citation-form, and clear) were entered as independent variables and treated as repeated measures.

4.2.5 Results

A univariate repeated measures analysis conducted with factors, speaking style and syllable type returned significant main effects for speaking style [$F(2, 20) = 34.368$, $p < .001$] and syllable type [$F(2, 20) = 12.016$, $p < 0.001$]. The analysis also revealed an interaction of speaking style × syllable type [$F(4, 40) = 9.4$, $p < 0.001$]. These results indicate that the effect of speaking style varied depending on the syllable types.

Following the investigation of the overall effects of speaking style and syllable type on the intelligibility of the Korean stops, pairwise comparison tests were performed to compare two particular speaking styles (α-level was set at 0.016 for the three comparisons). The comparison between conversational and citation-form styles showed significant main effects for speaking style [$F(1, 10) = 9.691, p = .011$] and syllable type [$F(2, 20) = 14.719, p < 0.001$]. However, the interaction between these two factors was not significant ($p = 0.384$). For the comparison between citation-form and clear speech styles, significant main effects for speaking style [$F(1, 10) = 24.9, p = .001$] and syllable type [$F(2, 20) = 14.105, p < 0.001$] were found in addition to a significant interaction between speaking style and syllable type [$F(2, 20) = 14.459, p < 0.001$]. Lastly, the same results were found for the comparison between conversational and clear styles. Significant main effects for speaking style [$F(1, 10) = 58.075, p < .001$], syllable type [$F(2, 20) = 7.35, p = .004$], and a significant interaction between speaking style and syllable type [$F(2, 20) = 13.017, p < .001$] were returned. These results indicate that there was a difference in the level of intelligibility for all of the three compared sets of speaking styles. In addition, the results indicate that the intelligibility difference between conversational and clear speech styles and also between citation-form and clear speech styles varied depending on syllable type. These results are presented in Table 4.1. The left-most column in Table 4.1 shows the mean intelligibility level over the three syllable types for each speaking style, and thus represents the overall intelligibility differences between the speaking styles. The other three columns show the mean intelligibility level over the 11 listeners in a matrix of syllable type by speaking style.

Next, in order to investigate the effect of speaking style for each of the three syllable types and thus to further investigate the interactions found between the speaking style and syllable type, paired-samples t-tests were performed. For each of the /t^han/,

/tan/, and /t*an/ syllable types, conversational and clear speech styles were paired and compared to each other. In addition, conversational and citation-form styles on the one hand, and citation-form and clear speech styles on the other hand, were compared in the same manner. In total, nine comparisons were made. Table 4.2 presents the results of the nine paired samples t-tests (α-level was adjusted to 0.005 for the nine comparisons).

The overall intelligibility over the three syllable types shown in Table 4.1 indicates intelligibility improvement for both citation-form and clear speech compared with conversational speech. However, the size of the overall intelligibility improvement varied. Clear speech showed greater improvement than citation-form speech. Citation-form speech showed a 3% increase in the mean correct response percentage from conversational speech, whereas clear speech showed a 12 % increase from conversational speech. At the same time, the intelligibility improved progressively in the sequence of conversational, citation-form, and clear speech (see the left-most column of Table 4.1).

When each of the syllable types was examined individually, clear speech and citation-form speech revealed further differences. Clear speech showed significant intelligibility improvements over conversational speech for all of the three syllable types (see the middle rows for each syllable type in Table 4.2). However, for citation-form speech, the /t^han/ syllable type only showed significant improvement (p = 0.004) over conversational speech (see the top rows for each syllable type in Table 4.2). In addition, in clear speech, the intelligibility improvement over conversational speech varied greatly depending on syllable types. The increase was much greater for /t*an/ type, 18 % than those for /tan/ type, 6% and /t^han/ type, 11% (see the top-row conversational style and bottom-row clear speech style in Table 4.1). In contrast, the intelligibility improvement in citation-form speech over conversational speech had a less variation. The improvement was 3%, 5%, and 2% for /tan/, /t^{h}an/, and /t*an/ types, respectively (see the top-row conversational style and middle-row citation-form style in Table 4.1). These disproportionate improvements between the syllable types were reflected in the interaction of speaking style and syllable type reported earlier.

The clear speech style showed intelligibility enhancement over citation-form style as well. The size of overall improvement was 9 % (See Table 4.1). As was the case for citation-form style over conversational style, the improvement was not present for all of the three syllable types. Only the /t*an/ type was significantly more intelligible in clear speech compared with citation-form speech ($p < .001$) (see the bottom row for each syllable type in Table 4.2).

To summarize, clear speech showed an intelligibility improvement over conversational speech for all of the three syllable types. Clear speech also showed a trend toward intelligibility improvement over citation-form speech, but only had a significant improvement for the /t*an/ type. Similarly, citation-form speech showed a general improvement over conversational speech, but only had a significant improvement for the /t^{h}an/ type. However, clear speech had a greater intelligibility improvement than citation-form speech over conversational speech for all three syllable types.

Table 4.1. Mean percentage of correct response for each syllable type and speaking style over 11 listeners. Numbers in the parentheses represent the mean correct responses over 11 listeners.

Mean over /tan/, /t^{h}an/, and /t*an/ types	/tan/	/t^{h}an/	/t*an/
Conversational			
77 %	77 %	83 %	71 %
(67.7)	(67.6)	(72.8)	(62.6)
Citation-form			
80 %	80 %	88 %	73 %
(70.7)	(70.5)	(77.4)	(64.2)
Clear speech			
89 %	83 %	94 %	89 %
(77.9)	(73.1)	(82.5)	(78)

Table 4. 2. Results of paired-samples t-tests for each of the three syllable types between conversational - citation-form, conversational - clear speech, and citation-form – clear speech pairs. Asterisks (*) indicate significant differences in intelligibility between the two speaking styles investigated ($\alpha = 0.005$). Conversational = conversational style, Citation = citation-form style, and Clear = clear speech style.

	Pair	t	df	Sig. (2-tailed)
/tan/	Conversational – Citation	-1.443	10	.180
	Conversational – Clear	-4.353	10	.001*
	Citation – Clear	-1.219	10	.251
/t^{h}an/	Conversational – Citation	-3.714	10	.004*
	Conversational – Clear	-4.517	10	.001*
	Citation – Clear	-2.834	10	.018
/t*an/	Conversational – Citation	-1.111	10	.292
	Conversational – Clear	-8.863	10	.000*
	Citation – Clear	-7.864	10	.000*

4.2.6 Discussion

The goal of Experiment 3 was to investigate the intelligibility enhancement effect of clear speech. In Experiments 1 and 2 of Chapter III, Korean stops produced in the clear speech style showed an enhancement of acoustic correlates encoding the Korean stop contrasts and this gave rise to the enhancement of acoustic distance between the stop categories. Experiment 3 examined whether this acoustic contrast enhancement effect found in Experiments 1 and 2 lead to the enhancement in the intelligibility of the stop categories as well.

The findings from Experiment 3 support the clear speech intelligibility enhancement effect. Korean stops produced in the clear speech style were more intelligible than those produced in the conversational and citation-form speech styles. This finding is compatible with the findings of studies dealing with clear speech intelligibility over citation-form speech for syllables and vowels (Ferguson & Kewley-Port, 2002; Gagne et al., 2002), as well as with other studies that investigated the same question for sentences (Picheny et al., 1985; Uchanski et al., 1996; Krause & Braida, 2002; Smiljanic & Bradlow, 2005). In addition, Korean stops produced in citation-form style were more intelligible overall than those produced in conversational style.

The finding that Korean stops produced in the citation-form style were overall more intelligible than those in the conversational style relates to the work of Harnsberger, Wright, & Pisoni (2008). In Experiment 1, in an attempt to tackle the limit of other clear speech research dealing with only citation-form and clear speech, another speech mode named "conversational" speech was elicited. In order to minimize the degree of self monitoring commonly occurring in the text reading tasks, a meaning-explanation task was designed. The rationale for adding a more "hypoarticulated" speech style was to provide experimental data containing more distinctively hypoarticulated and hyperarticulated speech and thus to show the talkers' adaptive articulatory effort to listeners' perceptual need more clearly.

Having a similar motivation of eliciting reduced, naturalistic speech in a laboratory setting closer to natural, conversational speech, Harnsberger et al. (2008) attempted to elicit three different speaking styles, reduced, citation, and hyperarticulated. To elicit a reduced speech style, speakers were prompted to perform a distracter task recalling multiple digits of number from short-term memory while reading sentences. The results of Harnsberger et al.'s (2008) work revealed overall success of eliciting three distinctive speech styles in a laboratory setting. The reduced and citation speech produced by 6 out of 12 speakers were impressionistically judged distinctive from each other by native listeners. The citation and hyperarticulated speech styles were also distinctive from each other. Their acoustic analyses revealed a limited difference between reduced and citation speech primarily on the durational properties for key words and sentences among other measures.

Even though Harnsberger et al. (2008) did not directly measure the intelligibility difference between the speaking styles they elicited, the findings from Experiments 1 and 3 of the current study are parallel to Harnsberger et al.'s (2008) findings. In Experiment 1, conversational and citation-form speech were less distinctive from each other for the three acoustic measures investigated compared with the conversational and clear speech comparison (see Figures 3.1 and 3.2). The results of Experiment 3 indicated that citation-form speech was less distinctive from conversational speech than clear speech in its intelligibility. Similarly, in Harnsberger et al. (2008), the distinctiveness judgment score between the speaking styles revealed that the scores for reduced and citation-form speech were not as different as the scores for citation-form and hyperarticulated speech or reduced and hyperarticulated speech. Likewise, in Experiment 3, the percent correct response for the test syllables linearly improved over the three speaking styles (see Table 4.1). These findings suggest that the attempts in Experiment 1 to elicit speech signals residing on the hypo-speech and hyper-speech continuum (Lindblom, 1990) were successful.

In addition to the finding of the overall intelligibility improvements between the three speaking styles, the findings of Experiment 3 included a varying degree of intelligibility improvement across the three syllable types. The /t*an/ type had the greatest improvement of 16 % in clear speech over conversational speech, the /tan/ type had the least improvement of 6 %, and the /t^han/ type had the medium 11 % improvement (see Table 4.1). At the same time, when it comes to a comparison of citation-form speech with conversational speech, the /t^han/ type only had an improvement over citation-form speech. Similarly, only /t*an/ type had significant improvement for the comparison of clear speech with citation-form speech (see

Table 4.2). Because the tokens of the three syllable types were presented to the listeners across the three syllable types (/t^han/, /tan/, and /t*an/), a discussion on the varying intelligibility improvements must be also made in light of the perception of the three-way manner contrasts of Korean stops.

The significant improvement for the /t*an/ type in clear speech over citation-form speech may be related to the voice quality property for the /t*an/ type. The tense or creaky voice quality of the /t*an/ type enhanced in clear speech might contribute to the distinction from /tan/ and /t^han/ types, which are associated with breathy voicing quality. The results from investigation of acoustic-phonetic modifications in clear speech production in Chapter III support this explanation. As can be seen from Figures 3.1(b) and 3.2(b), for both younger and older speakers, the H1-H2 values for fortis stops greatly decreased in the clear speech style (indicating increased tense or creaky voicing), and this may explain the greatest intelligibility improvement of the /t*an/ type in clear speech over citation-form speech.

As for the /t^han/ type, the greater intelligibility compared with the /tan/ or /t*an/ type for all of the three speaking styles (ranging from 83 to 94% in Table 4.1) may be related to the acoustic characteristics associated with aspirated stops. With the highest F0 and the longest VOT values (see Figure 3.1(a) and (c), and Figure 3.2 (a) and (c)), perhaps the /t^han/ type was least confused with the lowest F0 /tan/ type and the shortest VOT /t*an/ type, and this might contribute to the highest intelligibility of the /t^han/ type overall among the three syllable types.

In contrast, regarding the least intelligibility improvement of the /tan/ type, the /tan/ type have a less distinctive VOT and F0 difference from the /t*an/ type compared with the /tan/-/t^{h}an/ contrast, and this might give rise to the greater level of confusion for the /tan/-/t*an/ distinction. In addition, the "tense" voice quality of the /t*an/ type was less salient for the tokens produced in the conversational and citation-form styles (see Figures 3.1(b) and 3.2(b)), and this might contribute, in conjunction with the less distinctive VOT and f0 difference, to the greater confusion with the /tan/ type. Lastly, the low F0 effect of the lenis /tan/ type on the distinction from the other two types might have been diminished by the between-speaker presentation of the test stimuli (6 female and 5 male speakers), and this may be related to the less extent of intelligibility improvement of the /tan/ type.

CHAPTER V

PERCEPTION EXPERIMENT 2: PERECEPTUAL CUE WEIGHT TEST

5.1 Introduction

The primary goal of Experiment 4 was to further examine the findings from the production experiments in Chapter III as they relate to speech perception. The results of the two production experiments suggested that F0 is a stronger correlate than VOT for the aspirated-lenis stop distinction for younger speakers, whereas F0 is a secondary correlate, with VOT being primary, for older speakers. In hyperarticulated, clear speech production of aspirated and lenis stops, the F0 difference was more enhanced than the VOT difference for younger speakers. In contrast, VOT was the only acoustic correlate enhanced for older speakers. These production differences were interpreted to mean that the two groups had different VOT and F0 phonetic targets for the production of aspirated and lenis stops and that the different phonetic targets induced different clear speech enhancement patterns.

Experiment 4 examined whether the production differences between the two speaker groups lead to perceptual differences. To this end, the speech tokens of Korean stop production were manipulated with respect to the three acoustic correlates of Korean stops and a speech perception experiment was conducted. Also, listeners were recruited separately into younger and older groups. Through the perception experiment, the work reported in Chapter V attempted to determine the extent to which VOT, H1-H2, and F0 contribute to the distinction of the three-way manner contrast in Korean stops. For this purpose, three sets of analyses were performed: multivariate analysis of variance, logistic regression, and correlation.

With respect to the generational differences found in the clear speech production of aspirated and lenis stops, more attention was given to VOT and F0 while investigating the contribution of the three acoustic correlates to the perceptual distinction of the three-way manner contrasts in Korean stops. Specifically, two possibilities were considered. One possibility was that F0 is a more heavily weighted perceptual correlate than VOT for younger listeners for the aspirated-lenis stop distinction, whereas VOT is more heavily weighted than F0 for older listeners. The other possibility was that F0 is simply relatively more weighted for younger listeners than older listeners for the distinction of aspirated and lenis stops without a weighting reversal between F0 and VOT.

5.2 Experiment 4: re-synthesized stimuli perception task

5.2.1 Participants

A total of 20 native speakers of Korean participated in a re-synthesized stimuli perception task. The 20 listeners belonged to two different groups depending on their age. As was the case of Experiment 1, 10 listeners, 20 to 29 years old were assigned to "younger group," and 10 listeners over 40 years old, were assigned to "older group." The mean age was 20.2 years for the younger group and was 48.5 years for the older group. All of the listeners were speakers of the Seoul or Kyung-gi province dialect. The recruitment was made at Hanyang University (HU) in Seoul, Korea. The younger listeners were all students at HU, and the older listeners were staff members at HU or their family members. At the time of the task, none of the listeners reported any history of language or hearing disorders. Speakers who participated in Experiment 1, 2, or 3 were excluded from the participant pool for Experiment 4. The listeners were paid for their participation.

5.2.2 Perception stimuli

A 36 year-old male Korean speaker produced the source recordings for the re-synthesized stimuli. He was a speaker of the Seoul dialect of Korean. The speaker produced each of 다 /ta/, 타 /t^{h}a/, and 따 /t*a/ syllables 11 times along with other distracter syllables, such as 가 /ka/, 카 /k^{h}a/, 까 /k*a/, 나 /na/, 마 /ma/, etc. The source speech signals were recorded using a high-quality headset microphone and a Marantz digital recorder (PMD 670) at the sampling rate of 22,500 Hz. The recordings were made in a sound-attenuated booth at the University of Oregon. Adobe Audition 2.0 was used for the editing work.

The syllables composed of the three types of alveolar stops plus the low vowel /a/ were submitted to acoustic analyses for VOT, H1-H2, and F0. After reviewing the measured values for the three acoustic properties, three tokens of /ta/, /t^{h}a/, and /t*a/ were chosen as bases for re-synthesis. The choice was based on the typicality of measured values to the acoustic characteristics of Korean stops summarized in section 2.4. The three tokens were primarily chosen for their H1-H2 values, which are a primary acoustic correlate of voice quality. The selected /ta/, /t^ha/, and /t*a/ tokens had 1.5, 6.5, and -3.6 dB of H1-H2 values at the vowel onset, respectively. These values demonstrate distinctive voice qualities of modal or breathy voicing for lenis /ta/ and aspirated /t^ha/ types (1.5 and 6.5 dB), and tense or creaky voicing for fortis /t*a/ type (-3.6 dB). So, in terms of an H1-H2 scale, there were three levels.

To equate the amplitude difference among the bases, the peak intensity of vowel portion of three base tokens were normalized to 50% using Audacity (ver.1.2.6). The entire stretch between stop release burst and the beginning of the first periodic cycle of the vowel was removed from the three base tokens. Then, the vowel portions were equated to 260 ms for the three base tokens by truncating the end of the vowel durations. 260 ms was chosen because the shortest duration of clear formant structure among the three tokens was 260 ms (for the /t*a/ base). The original vowel durations were 303 ms, 286 ms, and 267 ms for the /ta/, /t^ha/, and /t*a/ bases, respectively. The vowel for the /t*a/ base showed a good amplitude envelope in that it tapered off at the end of the vowel and sounded natural. The other two base tokens were manipulated so that they had quite similar amplitude envelopes and tapered off at the end of the vowel. With the amplitude envelope function of Audacity (ver.1.2.6), the last 60 ms of the vowel was ramped down from 100% to 50 % (/t^ha/ base) or 40% (/ta/ base) of the intensity of the original stimuli.

Following this, consonantal portions (release burst and aspiration to the beginning of the first periodic cycle) were conjoined to the three vowel bases prepared. The base for the consonantal portion was taken from another /t^ha/ token, which had an original VOT of 91 ms. Six different consonant portions were prepared by shortening the 91 ms VOT from the end (i.e. right-hand side) with a 15 ms interval. Thus, the stimuli had six different VOT levels: 10, 25, 40, 55, 70, and 85ms.

As the last step, the vowel portion was manipulated for F0 at four levels: 110, 120, 130, and 140 Hz (using Praat, ver. 4.6). The F0 levels were held constant throughout the entire duration of the vowel. The original F0 was 115, 144, and 128 Hz at the vowel onset for the /ta/, /t^ha/, and /t*a/ tokens, respectively. Through the entire re-synthesis process, 72 stimuli types (3 H1-H2 levels × 6 VOT levels × 4 F0 levels) were created in total.

5.2.3 Procedure

Each stimulus type was presented 10 times, 1 time each in 10 randomized blocks, eliciting 720 responses from each of the 20 listeners. The three response categories were provided in Korean orthography, such that 다 /ta/, 타 /t^ha/, and 따 /t*a/ were displayed on a computer screen for each of the 720 stimuli. Listeners were instructed to choose one of these three categories by clicking a mouse after listening to the stimuli delivered to both ears through a headset. Because most of the test stimuli were phonetically ambiguous, the listeners were told that the stimuli may not sound like any of the three response choices and asked to use their instant impressions to make a decision for the category indemnification of the stimuli. The Multiple Forced Choice (MFC) function of Praat for speech perception experiment was used to administer the entire listening task. The task progressed at the listeners' own pace and the participants were told that there was no time limit to complete the task. Listeners were offered short breaks between the blocks. A practice session of 30 stimuli was presented before the main task. On average each listener used about 35 to 40 minutes to complete the task.

5.2.4 Statistical design

In order to assess potential differences between younger and older listeners in the perceptual weight of VOT, H1-H2, and F0 for distinction of manner contrast in Korean stops, three kinds of statistical analyses were performed. First, a multivariate analysis of variance (MANOVA) was performed to examine the overall group (younger and older) differences in relation to the three acoustic cues (VOT, H1-H2, and F0) for the distinction of the aspirated and lenis stops. The number of /ta/ response and the number of /t^ha/ response for each of the 720 stimuli for each listener were entered as the dependent variables. As the primary interest of the analysis was in the listeners' distinction of the /t^ha/ - /ta/ contrast, the number of the /t^ha/ and /ta/ responses was entered as dependent variables. Response to all three categories /t^ha/, /ta/, and /t*a/ will be further investigated in the analyses discussed below. Each stimulus type was presented 10 times and thus, the number of /ta/ or /t^ha/ response ranged 0 to 10 for each of the 72 stimulus type for each of the 20 listeners. The three acoustic dimensions of the test stimuli, VOT, H1-H2, and F0, in addition to group, were entered as the independent variables. As a result, a four-way multivariate analysis with two dependent variables was performed.

After investigating overall potential group differences in the perceptual weight of the acoustic correlates, logistic regression analyses were performed to further examine group differences (Morrison, 2007). Three binary logistic regression analyses were performed for each younger and older listener group for the three response categories in a pairwise manner. The purpose of the analyses was to determine the relative perceptual weights of VOT, H1-H2, and F0 in the distinction of the three-way manner contrast of Korean stops. For the aspirated-lenis contrast, the /t^ha/ or /ta/ responses for the 720 stimuli were entered as a dependent variable for each listener. That is, the binary values, /ta/ or /t^ha/ response (/ta/ coded as 0 and /t^ha/ coded as 1) were entered as an outcome variable in the regression analysis. Regarding the other aspirated-fortis and lenis-fortis contrasts, two separate analyses examining /t^ha/ and /t*a/ responses (/t^ha/ coded as 0 and /t*a/ coded as 1) as one pair and /ta/ and /t*a/ responses (/ta/ coded as 0 and /t*a/ coded as 1) as another were performed in the same manner. The predictor variables were the three acoustic correlates, VOT, H1-H2, and F0 (the six levels of VOT were encoded in the sequence of 10, 25, 40, 55, 70, and 85 ms; the three levels of H1-H2 were encoded in the sequence of 6.5, 1.5, and -3.6 dB; the four levels of F0 were encoded in the sequence of 110, 120, 130, and 140 Hz). The primary interest of the analyses was in investigating whether there was a difference between younger and older listeners in the perceptual weight of the three acoustic correlates for the distinction of the three-way manner contrasts.

As the last set of analyses, the potential group differences in the perceptual weight of the acoustic correlates were also investigated through correlation analyses (Holt & Lotto, 2006). Relative perceptual weight among the three acoustic correlates was computed separately for each of the younger and older listeners. The computation process of the relative perceptual weight was as follows. First, the correlation coefficient, *r* was obtained between each of the three acoustic correlates (VOT, H1-H2, F0) and each of the three response category (/ta/, /t^{h}a/, /t*a/) for each listener (the six levels of VOT were encoded in the sequence of 10, 25, 40, 55, 70, and 85 ms; the three levels of H1-H2 were encoded in the sequence of 6.5, 1.5, and -3.6 dB; the four levels of F0 were encoded in the sequence of 110, 120, 130, and 140 Hz) . As a result, 9 correlation coefficient values (3 acoustic correlates × 3 response categories) were obtained for each of the 20 listeners. Next, for each response category for each listener, the obtained correlation coefficient values for the three acoustic correlates were summed together. Then, the coefficient value for each acoustic correlate was divided by the summed coefficient value of the three acoustic correlates. The equation below shows the computation process.

$$\text{Proportional coefficient } r = \frac{r \text{ of VOT (or H1-H2 or F0)}}{r \text{ of VOT} + r \text{ of H1-H2} + r \text{ of F0}}$$

Thus, the outcome values represented a proportion of the sum, and these values were summed to 1. The proportional values represented a relative proportion of each acoustic correlate over the three acoustic correlates, that is, the relative weight of each acoustic correlate among the three acoustic correlates for each response category. The relative weight of the acoustic cues was determined in this way for the two participant groups under investigation.

5.2.5 Results

5.2.5.1 Results of MANOVA

A MANOVA with factors, VOT (10, 25, 40, 55, 70, 85ms), H1-H2 (-3.6, 1.5, 6.5 dB), F0 (110, 120, 130, 140 Hz), and group (younger, older) returned significant main effects for all of the four independent variables (VOT [$F(10, 2590) = 128.391, p < 0.001$]; H1-H2 [$F(4, 2590) = 513.847, p < 0.001$]; F0 [$F(6, 2590) = 315.172, p < 0.001$], and group [$F(2, 1295) = 28.393, p < 0.001$]). The dependent variable was the number of /t^ha/ and /ta/ response, and the four independent variables were all treated as a between-subject factor. In addition, three-way interactions of group × VOT × F0 [$F(30, 2590) = 1.485, p = 0.044$] and group × H1-H2 × VOT [$F(20, 2590) = 1.826, p = 0.014$] were found, as well as significant interactions for all of the two-way combinations of the four factors. The four-way interaction of the four factors was not significant ($p = 0.986$). The three-way interactions indicate that the younger and older groups differed in the manner in which the level of VOT affected response choice depending on the level of F0. Likewise, the two groups differed in the manner in which the level of VOT affected response choice depending on the level of H1-H2.

The focus of the analysis was on the investigation of potential group difference in the perceptual weights of VOT and F0 for the distinction of the aspirated and lenis stops.

Therefore, the significant interaction in group × VOT × F0 was further examined through the following four figures. The four figures were plotted individually according to the four F0 levels. Each figure shows the number of /t^ha/ and /ta/ responses varying as function of VOT levels for both younger and older groups. Thus, the four figures, as a set, show an interaction between VOT and F0 for the choice of /t^ha/ and /ta/ response with the two speaker groups compared with each other. Specifically, Figure 5.1 shows the mean number of /t^ha/ and /ta/ responses for each of the 10 younger and older listeners for the stimuli with 110 Hz varying in VOT values. Figures 5.2, 5.3, and 5.4 represent the mean number of responses in the same manner for the stimuli with 120, 130, 140 Hz, respectively. For all of the four figures, the lines with filled circles or triangles represent the responses by younger listeners, and the lines with empty circles or triangles represent the responses by older listeners. In addition, the lines with a circle represent the /ta/ responses, whereas the lines with a triangle represent the /t^ha/ responses.

Let us begin with Figure 5.1, which represents responses for stimuli with 110 Hz. Given the F0 effect on the distinction of the aspirated and lenis stop contrast describe in section 2.4, with this low F0, the number of /ta/ responses was greater than /t^ha/ response at all of the six VOT levels for both younger and older listeners. For stimuli with long VOTs, 55, 70, 85ms, /ta/ response decreased, whereas /t^ha/ response increased. However, /ta/ responses were still greater than /t^ha/ responses for the stimuli with long VOTs. These results suggest that the effect of the low F0 held throughout the entire VOT range.

In addition to the role of F0 distinguishing aspirated and lenis stops, Figure 5.1 also shows a group difference in the degree of F0 effect on the distinction of the two stop types. Younger listeners (solid line with filled circles) had greater /ta/ responses than older listeners (dotted line with empty circles) for all VOT levels. This suggests that younger listeners were more affected by the low F0 than older listeners and identified more tokens as lenis stops overall at this low F0.

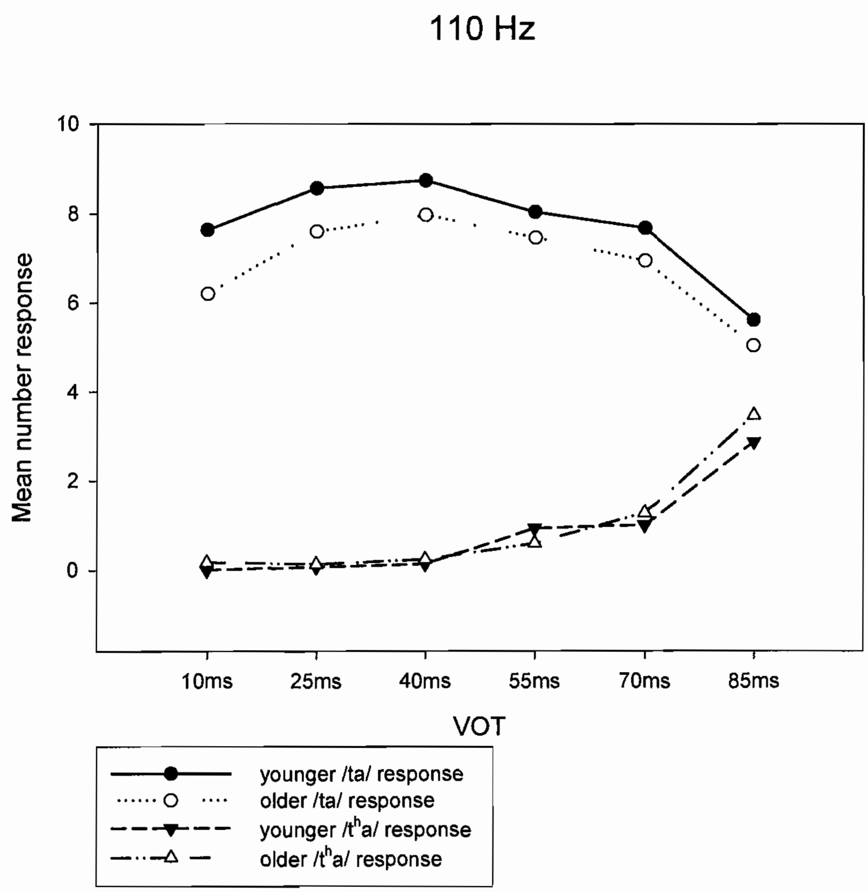


Figure 5.1. Mean number of response, collapsed across H1-H2, to /ta/ and /tʰa/ stimuli with 110 Hz F0 for younger and older listeners (n=10).

For the response to stimuli with 120 Hz (Figure 5.2), the number of /ta/ responses was lesser and the number of /t^ha/ responses was greater overall, compared with the responses to stimuli with 110 Hz. The responses were more strongly affected by VOT than was the case with the 110 Hz stimuli. In addition, the two listener groups differed in the extent to which they responded to the increased F0. For the younger listeners, as VOT increased, /t^ha/ responses increased more rapidly and at the same time, /ta/ response decreased more rapidly than was the case for older listeners. This suggests that the younger listeners were more greatly affected by the F0 increase. Also, the crossover points for /t^ha/ vs. /ta/ response were different between the two groups. The crossover point for younger listeners was at a lower VOT value (about 55 ms) than for older listeners (about 80 ms). These results indicate that for stimuli with shorter VOTs, the younger listeners gave more /t^ha/ responses. In other words, older listeners needed stimuli with longer VOTs to give /t^ha/ response.

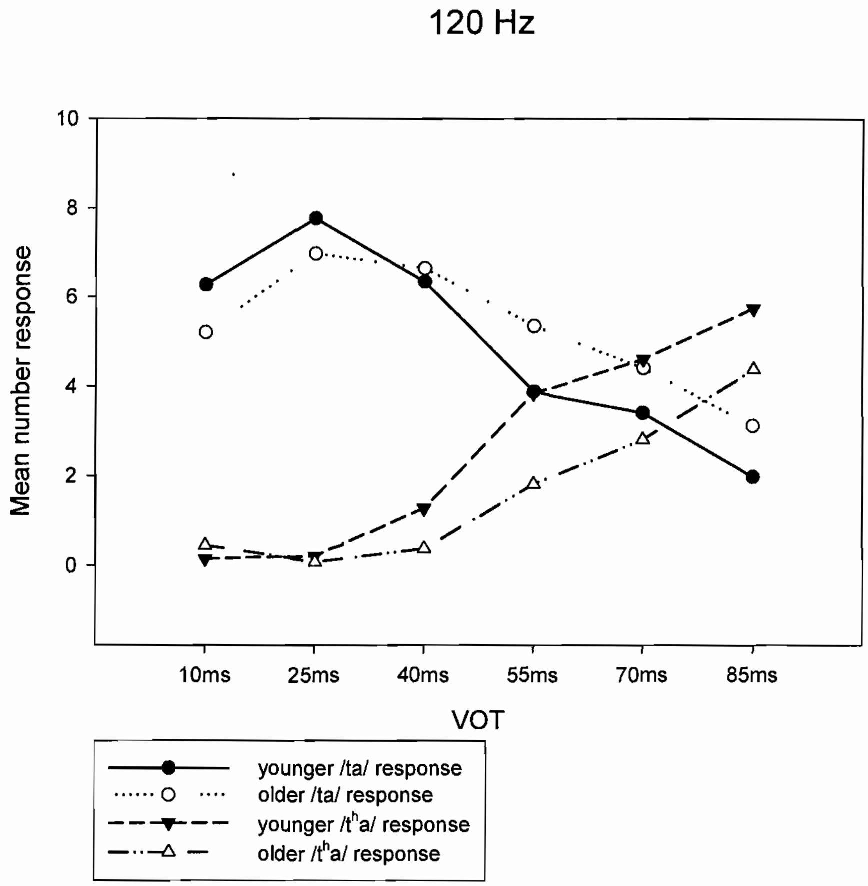


Figure 5.2. Mean number of response, collapsed across H1-H2, to /ta/ and /tʰa/ stimuli with 120 Hz F0 for younger and older listeners (n=10).

For the stimuli with 130 Hz (Figure 5.3), the patterns seen for stimuli with 120 Hz were largely repeated. Overall, with this increased F0, more /tʰa/ response and less /ta/ responses were given by both listener groups compared with the stimuli with the lower F0 values. Also, the increase of /tʰa/ response and the decrease of the /ta/ response over the VOT range were more rapid. As for the group difference, the crossover point was earlier for the younger listeners than for the older listeners, indicating that the younger listeners were more greatly affected by the increased F0. For both listener groups, the crossover points were made at lower VOT values compared with the responses to stimuli of 120 Hz. However, as was the case of stimuli with 120 Hz, the crossover point for younger listeners was at a lower VOT value (about 30 ms) than older listeners (50 ms).

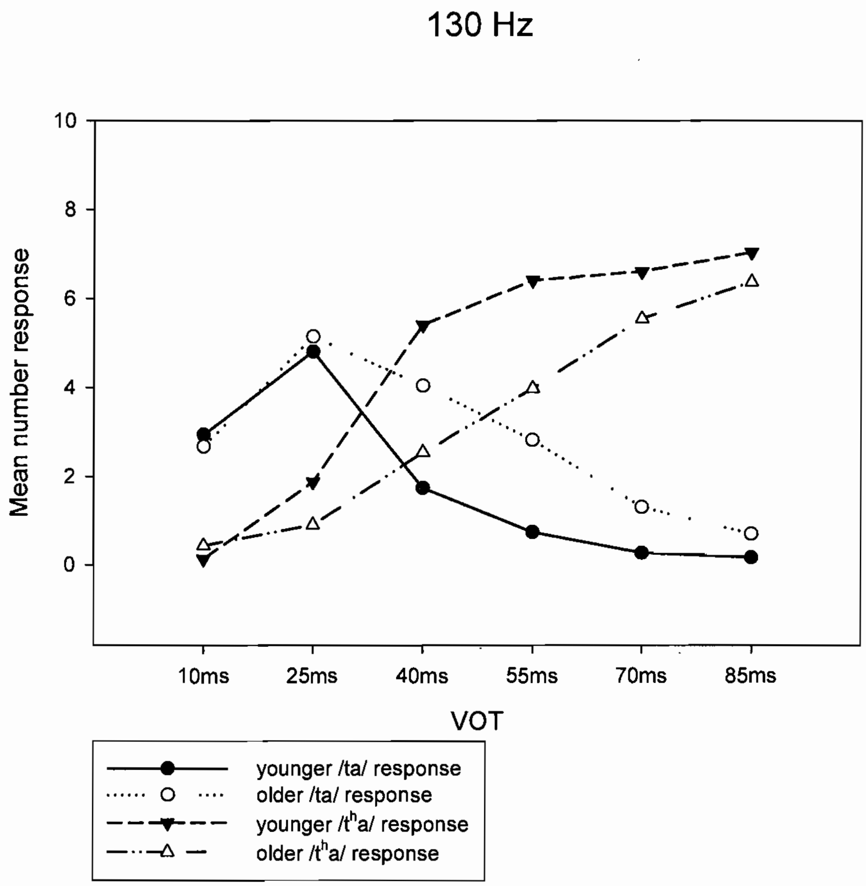


Figure 5.3. Mean number of response, collapsed across H1-H2, to /ta/ and /tʰa/ stimuli with 130 Hz F0 for younger and older listeners (n=10).

The response to stimuli with 140 Hz (Figure 5.4) was more or less the inverse of the response to stimuli with 110 Hz. The number of /tʰa/ responses was greater than /ta/ responses for nearly the entire VOT range, especially for younger listeners. This result replicates the F0 effect found with stimuli with 110 Hz in an inverse way. With stimuli of 110 Hz, /ta/ response was greater /t^ha/ response at all VOT levels. In addition, just as the other three Figures did, Figure 5.4 also shows a group difference. Note that, for both younger and older listeners, for the stimuli with 120 and 130 Hz, the 25 ms was a point where the greatest /ta/ response was induced and thus, /ta/ responses were distinctively greater than /t^ha/ responses at 25 ms (See Figures 5.2 and 5.3). However, for stimuli with 140 Hz, even at 25 ms, /t^ha/ responses were greater than /ta/ responses for younger listeners. In addition, the crossover point at a lower VOT level compared with older listeners was repeated with the stimuli with 140 Hz. These results suggest that younger listeners gave more /t^ha/ responses to the stimuli with short VOTs than older listeners for the stimuli of 140 Hz.

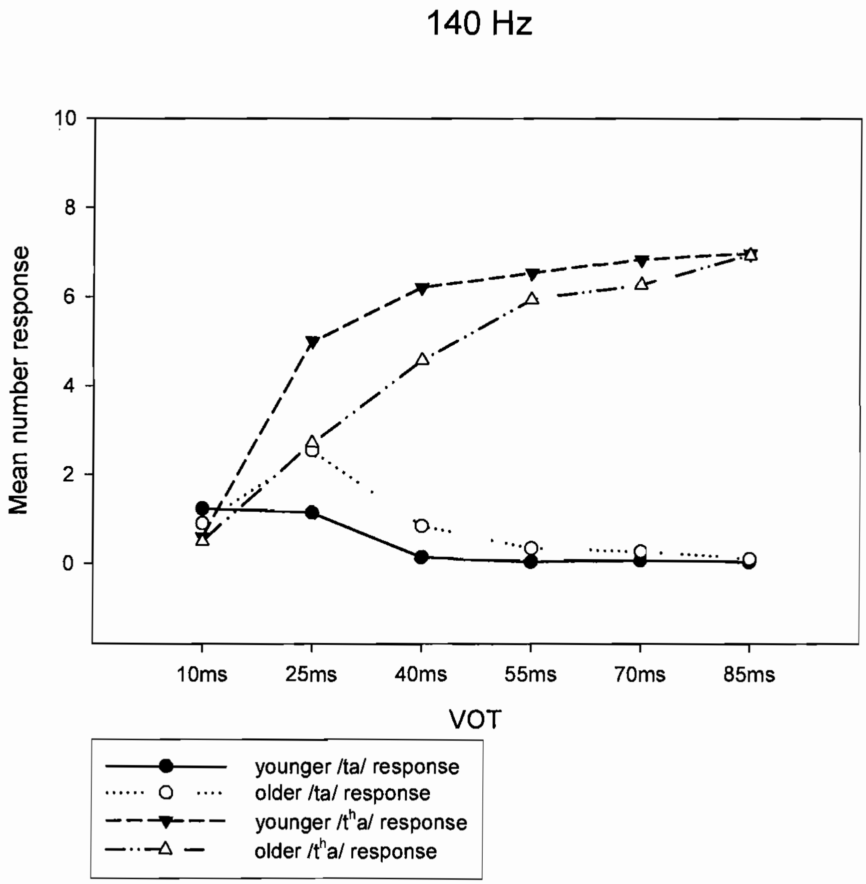


Figure 5.4. Mean number of response, collapsed across H1-H2, to /ta/ and /tʰa/ stimuli with 140 Hz F0 for younger and older listeners (n=10).

Examining all of the four figures together, one notes that, as F0 increased, the younger listeners gave more /tʰa/ responses than older listeners for the stimuli with short VOTs. As can be seen from Figures 5.2, 5.3, and 5.4, as F0 increased, the crossover point (where more /tʰa/ response began) shifted left, to the lower VOTs, for both groups. This indicates that, as the F0 of stimuli increased, the listeners gave more and more /tʰa/ response for the stimuli with shorter VOTs. However, at the same time, the crossover points for younger listeners were consistently to the left of those for older listeners, that is, at the lower VOT levels. These results suggest that younger listeners were more greatly affected by the F0 modulation than older listeners for the distinction of Korean aspirated and lenis stops.

5.2.5.2 Results of logistic regression analyses

The results from the multivariate analysis indicated that the younger and older speakers were affected by VOT and F0 modulation to a different degree with respect to the distinction of the aspirated and lenis stops. Younger speakers were more affected by F0 modulation than the older speakers for the /t^ha/ and /ta/ distinction. Logistic regression analyses exhibit this group difference in more detail by showing the relative contribution of the acoustic correlates for the distinction of the manner contrast in Korean stops.

Let us begin with the analyses for /t^ha/ and /ta/ distinction. For younger speakers, a test of the full model with all of the three predictors (VOT, H1-H2, and F0) against a constant-only model was significant, χ^2 (3) = 3870, $p < .05$. This indicates that the three predictors reliably distinguished /t^ha/ and /ta/ responses with a good model fit, Nagelkerke $R^2 = .71$. With all of the three predictor variables, correct classification rates were 88% for /ta/ response and 85% for /t^ha/ response. The overall correct classification rate was 86%.

For older speakers, as was the case of younger speakers, a test of the full model with all three predictors against a constant-only model was significant, χ^2 (3) = 2677, $p < .05$. This indicates that the three predictors reliably distinguished /t^ha/ and /ta/

responses with a good model fit, Nagelkerke $R^2 = .59$. With all of the three predictor variables, correct classification rates were 87% for /ta/ response and 75% for /t^{h}a/ response, with an overall correct classification rate of 82%.

Table 5.1 shows the relative contribution of the individual predictors to the model. For both younger and older listeners, VOT and F0 were significant predictors of the outcome. H1-H2 was only significant for younger listeners. The interesting difference was that the odds ratio for F0 was 9.6 for younger listeners and was 4.8 for older listeners. This result indicates that for the younger listeners, the likelihood of /t^{h}a/ response was 9.6 times greater after one-unit increase in F0 (10 Hz in the current stimuli) compared with before the change. By the same token, it means that the likelihood of /t^{h}a/ response for the older listeners was 4.8 times greater after one-unit increase in F0. Therefore, for a given increase in F0, the likelihood of /t^{h}a/ response by younger listeners was by far greater than the likelihood of /t^{h}a/ response by older listeners.

The results also showed that for both groups, the likelihood of /t^{h}a/ response after one-unit increase in VOT increased. The odds ratio was 3.2 for younger listeners and 2.5 for older listeners, indicating that the likelihood of /t^{h}a/ response was 3.2 times greater for younger listeners and 2.5 times greater for older listeners after one-unit VOT increase. These odds ratios for VOT are not as different from each other as the odds ratios for F0. This indicates that younger and older listeners were not much different from each other for the effect of VOT modulation on the perception of the /t^{h}a/ and /ta/ types, even though both groups were affected by VOT modulation of the stimuli.

In addition, older listeners were not affected by the H1-H2 modulation for the distinction of the /t^ha/ and /ta/ categories, whereas younger listeners were affected. Table 5.1 (the Wald statistic column) shows that H1-H2 was not a significant contributor to the distinction of the /t^ha/ and /ta/ categories for older listeners. However, the odds ratio for H1-H2 was around 1 for both listener groups (0.7 for younger and 0.9 for older listeners). This indicates that the younger and older listeners were not much different from each other for the likelihood of /t^ha/ response after one-unit H1-H2 decrease in the stimuli. Thus, the combined results from VOT, H1-H2, and F0 indicate that a prominent difference between younger and older listeners resides on the contribution power of F0 to the distinction of /ta/ and /t^ha/ types.

Table 5.1. Contribution of three predictors to the perception of /ta/ and /t^ha/

Group	Predictor variables	*B* (*SE*)	Wald (*df* = 1)	Odds Ratio (*Exp*(*B*))
Younger	VOT	1.17 (.04)	834.3*	3.2
	H1-H2	-.32 (.06)	26.9*	0.7
	F0	2.26 (.07)	1207.1*	9.6
	Constant	-9.25 (.29)	1001.5*	0.000
Older	VOT	0.91 (.03)	753.8*	2.5
	H1-H2	0.09 (.06)	1.8	0.9
	F0	1.57 (.05)	1040.2*	4.8
	Constant	-7.73 (.25)	972.7*	0.000

(*B*, the logistic coefficient represents the log of the odds of an event occurrence for one-unit change in the predictor variable. Wald statistic tells whether the predictor in question make a significant contribution to the outcome. The odds ratio (*Exp* (*B*)) represents the relative strength of the variables for the prediction of the outcome. An odds ratio of greater than 1 indicates that the odds of event occurrence increases when one-unit change is made in a predictor variable. An odds ratio of less than 1 indicates that the odds of event occurrence decreases when one-unit change is made in a predictor variable. Asterisks indicate significant effects, $p < .05$.)

As for the distinction of the /ta/ and /t*a/ types, the same types of analyses were performed. For both younger and older listeners, the three predictors, as a set, contributed to the model (χ^2 (3) = 3061, $p < .05$ for younger; χ^2 (3) = 3669, $p < .05$ for older) and the models reliably distinguished /ta/ and /t*a/ responses with a good model fit, Nagelkerke $R^2 = .63$ for younger group and Nagelkerke $R^2 = .66$ for older group. The models also correctly classified /ta/ and /t*a/ responses with overall rates of 83 % for younger (85 % for /ta/, 81 % for /t*a/) and 84 % for older listeners (84 % for /ta/, 85 % for /t*a/).

Table 5.2 shows the relative contribution of VOT, H1-H2, and F0 to the distinction of /ta/ and /t*a/ categories for both younger and older listeners. The two groups were different from each other in multiple ways. For H1-H2, the odds ratio was 15.4 for older listeners and 7.6 for younger listeners. This indicates that after a unit decrease in H1-H2, the likelihood of /t*a/ response was 15.4 times greater compared with before the change for older listeners, and the likelihood of /t*a/ response was 7.6 times greater for younger listeners. In other words, for the distinction of the /ta/ and /t*a/ types, the younger listeners were less affected by H1-H2 modulation than older listeners.

As for F0, the odds ratio was 5.1 for younger listeners and was 3.6 for older listeners. This indicates that for a unit increase in F0, the likelihood of /t*a/ response was greater for younger listeners. In other words, for the distinction of the /ta/ and /t*a/ types, younger listeners were more affected by F0 modulation of the stimuli than older listeners. Taking the results for H1-H2 and F0 together, H1-H2 was by far a greater contributor than F0 for older listeners to the distinction of the /ta/ and /t*a/ types (the odds ratios were 15.4 and 3.6 for H1-H2 and F0, respectively). For younger listeners, the difference in odds ratio between H1-H2 and F0 was not as large as the case of older listeners (7.6 and 5.1 for H1-H2 and F0, respectively), indicating that F0 had relatively a similar contribution power to H1-H2 compared.

The results for the lenis /ta/ and fortis /t*a/ responses also indicate that only older listeners were affected by VOT modulation of the stimuli for the distinction of the /ta/ and /t*a/ types. Table 5.2 (Wald statistic column) shows that VOT was a significant contributor only for older listeners. The odds ratio for VOT was 0.8 (that is, less than 1) for older listeners, and this indicates that for a unit increase in VOT, the likelihood of /t*a/ response decreased for older listeners. In other words, older listeners were slightly affected by VOT modulation for the distinction of the /ta/ - /t*a/ contrast, whereas younger listeners were not affected.

The finding that younger listeners were more affected by F0 change than older listeners for /ta/ - /t*a/ distinction relates to the previous findings for /t^ha/ - /ta/ distinction. The lenis /ta/ - fortis /t*a/ contrast has a large F0 difference with higher F0 for the /t*a/ type. The aspirated /t^ha/ - lenis /ta/ contrast is another pair that has a large F0 difference, with higher F0 for the /t^ha/ type. For the distinction of the /t^ha/ and /ta/ categories earlier, younger listeners were more sensitive to F0 modulation than older listeners. Thus, the combined findings from /t^ha/ - /ta/ and /ta/ - /t*a/ contrasts suggest that F0 is more heavily weighted for younger listeners than older listeners in the perception of Korean stop manner contrast.

Table 5.2. Contribution of three predictors to the perception of /ta/ and /t*a/

Group	Predictor variables	*B* (*SE*)	Wald (*df* = 1)	Odds Ratio (*Exp*(*B*))
Younger	VOT	0.13 (.03)	0.9	1.0
	H1-H2	2.03 (.07)	845.4*	7.6
	F0	1.6 (.05)	1101.8*	5.1
	Constant	-8.5 (.25)	1105.5*	0.001
Older	VOT	-.26 (.03)	92.7*	0.8
	H1-H2	2.7 (.08)	1287.2*	15.4
	F0	1.3 (.05)	753.5*	3.6
	Constant	-8.4 (.24)	1190.7*	0.000

(*B*, the logistic coefficient represents the log of the odds of an event occurrence for one-unit change in the predictor variable. Wald statistic tells whether the predictor in question make a significant contribution to the outcome. The odds ratio (*Exp* (*B*)) represents the relative strength of the variables for the prediction of the outcome. An odds ratio of greater than 1 indicates that the odds of event occurrence increases when one-unit change is made in a predictor variable. An odds ratio of less than 1 indicates that the odds of event occurrence decreases when one-unit change is made in a predictor variable. Asterisks indicate significant effects, $p < .05$.)

Lastly, let us examine the analyses for the /t^{h}a/ - /t*a/ contrast. For both younger and older listeners, the three predictors, as a set, contributed to the model (χ^2 (3) = 1549, $p < .05$ for younger, χ^2 (3) = 2199, $p < .05$ for older) and the models reliably distinguished /t^{h}a/ and /t*a/ responses with a good model fit, Nagelkerke R^2 = .39 for younger listeners; R^2 = .53 for older listeners. The models also correctly classified /t^{h}a/ and /t*a/ responses with overall rates of 75 % for younger (81 % for /t^{h}a/, 69 % for /t*a/) and 82 % for older listeners (76 % for /t^{h}a/, 87 % for /t*a/).

Table 5.3 shows the relative contribution of VOT, H1-H2, and F0 to the distinction of /t^{h}a/ and /t*a/ categories. For both younger and older listeners, the three predictors were significant contributors. The odds ratio for H1-H2 was greater than 1 and the largest among the three predictor variables for both groups. These results indicate that when a unit decrease in H1-H2 was made, the likelihood of /t*a/ response increased 6.5 times for younger listeners and 15.1 times for older listeners, and at the same time that H1-H2 was the greatest contributor to the distinction of the /t^ha/ - /t*a/ contrast. These results are parallel to the results for the /ta/ - /t*a/ contrast. The contribution of H1-H2 was greater for older listeners than younger listeners for the /ta/ - /t*a/ contrast (odd ratio was 15.4 for older and 7.6 for younger listener), and H1-H2 was the greatest contributor to the distinction of the /ta/ - /t*a/ contrast as well (see Table 5.2).

The results also indicate that odds ratios for VOT and F0 were less than 1 for both groups. These indicate that when one-unit increase in VOT and F0 was made, the likelihood of /t*a/ response decreased for both younger and older listeners. Unlike the /t^ha/ - /ta/ and /ta/ - /t*a/ contrasts, the different sensitivity to F0 modulation between younger and older listeners was not found with the /t^ha/ - /t*a/ contrast. This contrast does not have a F0 difference, as both are associated with high F0.

Table 5.3. Contribution of three predictors to the perception of /t^ha/ and /t*a/

Group	Predictor variables	*B* (*SE*)	Wald (*df* = 1)	Odds Ratio (*Exp*(*B*))
Younger	VOT	-0.64 (.03)	566.5*	0.5
	H1-H2	1.88 (.06)	1029.3*	6.5
	F0	-0.21 (.04)	23.2*	0.8
	Constant	-1.15 (.22)	28.4*	0.3
Older	VOT	-0.94 (.04)	669.1*	0.4
	H1-H2	2.72 (.08)	1058.6*	15.1
	F0	-.11 (.05)	4.8*	0.9
	Constant	-1.66 (.2)	47.6*	0.2

(*B*, the logistic coefficient represents the log of the odds of an event occurrence for one-unit change in the predictor variable. Wald statistic tells whether the predictor in question make a significant contribution to the outcome. The odds ratio (*Exp* (*B*)) represents the relative strength of the variables for the prediction of the outcome. An odds ratio of greater than 1 indicates that the odds of event occurrence increases when one-unit change is made in a predictor variable. An odds ratio of less than 1 indicates that the odds of event occurrence increases when one-unit change is made in a predictor variable. Asterisks indicate significant effects, $p < .05$.)

5.2.5.3 Results of correlation analyses

In this section, the relative perceptual weight of the three acoustic correlates for the perception of the three manner contrasts in Korean stops is investigated. The relative weight among the three acoustic correlates computed from correlation between acoustic correlates and response categories is presented in six different tables below, by response category and listener group. Each table lists the results for each listener and also the pooled results over the ten listeners of each group. The primary aim of the investigation was to examine whether there were group differences in the relative perceptual weight of the three acoustic correlates. Accordingly, the focus of the analyses was on the group comparison. Tables 5.4, 5.5, and 5.6 represent results for younger listeners and Tables 5.7, 5.8, and 5.9 represent results for older listeners.

First, for the perception of /ta/ category (see Tables 5.4 and 5.7), F0 was weighted most heavily for both listener groups. The pooled F0 weight was 0.58 for younger listeners and 0.49 for older listeners. These results are consistent with the findings from the MANOVA and logistic regression analyses, which showed the greater F0 effect for younger listeners on the /ta/ - /tʰa/ identification. The correlation coefficients, (*r*) for VOT, H1-H2, and F0 all had negative values, indicating that as VOT, H1-H2, and F0 increased in the stimuli, the number of /ta/ response decreased. The /ta/ category is associated with medium-long VOT, breathy voicing, and low F0, and the negative correlations correspond to these acoustic characteristics of the /ta/ category.

Secondly, for the perception of /tʰa/ category (see Tables 5.5 and 5.8), the pooled VOT weight shows that VOT was weighted most heavily for both younger and older listeners. Similarly to the case of /ta/ perception, the pooled F0 weight was slightly larger for younger listeners (0.35) than older listeners (0.31). However, unlike the case of /ta/ perception, the relative weight among the three acoustic correlates was more or less similar to one another, even though VOT weight was the largest for both groups. The correlation coefficients, (*r*) for H1-H2 were negative, corresponding to the breathy voicing characteristic of the /tʰa/ category. For VOT and F0, the correlation coefficients, (*r*) were positive, corresponding to the high VOT and high F0 characteristics of the /tʰa/ type.

Lastly, as for the /t*a/ category (see Tables 5.6 and 5.9), the pooled H1-H2 weight shows that H1-H2 weight was the largest for both listener groups. As was the case for the /t/ and /tʰa/ categories, the pooled F0 weight was a little larger for younger listeners (0.26) than older listeners (0.20). The correlation coefficients (*r*) for VOT were negative, corresponding to the low VOT characteristic of the /t*a/ type. For H1-H2 and F0, correlation coefficients, (*r*) were positive, corresponding to low H1-H2 and high F0 characteristics of the /t*a/ type.

Combining the results for /ta/, /tʰa/, and /t*a/ categories, the F0 weight for younger listeners was consistently larger than older listeners for the three categories. In addition, the H1-H2 weight for older listeners was consistently larger than younger listeners for the three categories. These results suggest that younger listeners were more sensitive to F0 and less sensitive to H1-H2 modulation of the stimuli than older listeners. This finding parallels the results from the MANOVA and logistic regression analyses as a whole.

Considering the results of the three types of analyses together, similar findings were obtained: younger listeners were more affected by F0 modulation than older listeners, whereas older listeners were more affected by H1-H2 modulations than younger listeners for the perception of the three manner types in Korean stops. In addition, older listeners tended to be more affected by VOT modulation than younger listeners.

Table 5.4. Relative weight of VOT, H1-H2, and F0 for the perception of /ta/ category for younger listeners

	Correlation coefficient (*r*)			Relative weight		
Younger listener	/t^h^a/ × VOT	/t^h^a/ × H1-H2	/t^h^a/ × F0	VOT weight	H1-H2 weight	F0 weight
101	-0.342	-0.397	-0.561	0.26	0.31	0.43
102	-0.247	-0.171	-0.813	0.20	0.14	0.66
103	-0.044	-0.317	-0.767	0.04	0.28	0.68
104	-0.559	-0.111	-0.524	0.47	0.09	0.44
105	-0.194	-0.106	-0.887	0.16	0.09	0.75
106	-0.241	-0.395	-0.658	0.19	0.31	0.51
107	-0.311	-0.403	-0.641	0.23	0.30	0.47
108	-0.489	-0.085	-0.717	0.38	0.07	0.56
109	-0.16	-0.359	-0.719	0.13	0.29	0.58
110	-0.193	-0.067	-0.861	0.17	0.06	0.77
Pooled for 10 listeners	-0.272	-0.231	-0.704	0.23	0.19	0.58

Table 5.5. Relative weight of VOT, H1-H2, and F0 for the perception of /t^{h}a/ category for younger listeners

	Correlation coefficient (*r*)			Relative weight		
Younger listener	/t^{h}a/ × VOT	/t^{h}a/ × H1-H2	/t^{h}a/ × F0	VOT weight	H1-H2 weight	F0 weight
101	0.462	-0.521	0.363	0.34	0.39	0.27
102	0.386	-0.531	0.435	0.29	0.39	0.32
103	0.401	-0.274	0.598	0.32	0.22	0.47
104	0.78	-0.076	0.415	0.61	0.06	0.33
105	0.215	-0.592	0.401	0.18	0.49	0.33
106	0.474	-0.514	0.43	0.33	0.36	0.30
107	0.501	-0.446	0.459	0.36	0.32	0.33
108	0.626	-0.385	0.465	0.42	0.26	0.32
109	0.504	-0.388	0.542	0.35	0.27	0.38
110	0.564	-0.107	0.639	0.43	0.08	0.49
Pooled for 10 listeners	0.474	-0.35	0.452	0.37	0.27	0.35

Table 5.6. Relative weight of VOT, H1-H2, and F0 for the perception of /t*a/ category for younger listeners

	Correlation coefficient (r)			Relative weight		
Younger listener	/tʰa/ × VOT	/tʰa/ × H1-H2	/tʰa/ × F0	VOT weight	H1-H2 weight	F0 weight
101	-0.074	0.851	0.228	0.06	0.74	0.20
102	-0.167	0.721	0.327	0.14	0.59	0.27
103	-0.33	0.641	0.349	0.25	0.49	0.26
104	-0.448	0.32	0.14	0.49	0.35	0.15
105	-0.024	0.608	0.599	0.02	0.49	0.49
106	-0.171	0.808	0.261	0.14	0.65	0.21
107	-0.075	0.78	0.305	0.06	0.67	0.26
108	-0.16	0.646	0.383	0.13	0.54	0.32
109	-0.347	0.749	0.174	0.27	0.59	0.14
110	-0.55	0.245	0.245	0.53	0.24	0.24
Pooled for 10 listeners	-0.2	0.611	0.286	0.18	0.56	0.26

Table 5.7. Relative weight of VOT, H1-H2, and F0 for the perception of /ta/ category for older listeners

	Correlation coefficient (r)			Relative weight		
Older listener	/tʰa/ × VOT	/tʰa/ × H1-H2	/tʰa/ × F0	VOT weight	H1-H2 weight	F0 weight
201	-0.225	-0.453	-0.649	0.17	0.34	0.49
202	-0.287	-0.32	-0.688	0.22	0.25	0.53
203	-0.021	-0.508	-0.613	0.02	0.44	0.54
204	-0.088	-0.279	-0.853	0.07	0.23	0.70
205	-0.05	-0.409	-0.609	0.05	0.38	0.57
206	-0.147	-0.402	-0.738	0.11	0.31	0.57
207	-0.464	-0.443	-0.439	0.34	0.33	0.33
208	-0.254	-0.493	-0.573	0.19	0.37	0.43
209	-0.23	-0.639	-0.413	0.18	0.50	0.32
210	-0.223	-0.401	-0.6	0.18	0.33	0.49
Pooled for 10 listeners	-0.194	-0.423	-0.598	0.16	0.35	0.49

Table 5.8. Relative weight of VOT, H1-H2, and F0 for the perception of /tʰa/ category for older listeners

	Correlation coefficient (*r*)			Relative weight		
Older listener	/tʰa/ × VOT	/tʰa/ × H1-H2	/tʰa/ × F0	VOT weight	H1-H2 weight	F0 weight
201	0.407	-0.382	0.571	0.30	0.28	0.42
202	0.595	-0.398	0.414	0.42	0.28	0.29
203	0.434	-0.273	0.512	0.36	0.22	0.42
204	0.304	-0.456	0.497	0.24	0.36	0.40
205	0.572	-0.473	0.396	0.40	0.33	0.27
206	0.469	-0.568	0.355	0.34	0.41	0.26
207	0.653	-0.395	0.201	0.52	0.32	0.16
208	0.651	-0.255	0.413	0.49	0.19	0.31
209	0.47	-0.39	0.425	0.37	0.30	0.33
210	0.579	-0.41	0.473	0.40	0.28	0.32
Pooled for 10 listeners	0.503	-0.384	0.406	0.39	0.30	0.31

Table 5.9. Relative weight of VOT, H1-H2, and F0 for the perception of /t*a/ category for older listeners

	Correlation coefficient (*r*)			Relative weight		
Older listener	/tʰa/ × VOT	/tʰa/ × H1-H2	/tʰa/ × F0	VOT weight	H1-H2 weight	F0 weight
201	-0.153	0.822	0.125	0.14	0.75	0.11
202	-0.274	0.723	0.324	0.21	0.55	0.25
203	-0.3	0.745	0.274	0.23	0.56	0.21
204	-0.121	0.625	0.566	0.09	0.48	0.43
205	-0.445	0.744	0.171	0.33	0.55	0.13
206	-0.213	0.766	0.387	0.16	0.56	0.28
207	-0.19	0.813	0.227	0.15	0.66	0.18
208	-0.344	0.743	0.204	0.27	0.58	0.16
209	-0.105	0.863	0.098	0.10	0.81	0.09
210	-0.37	0.757	0.112	0.30	0.61	0.09
Pooled for 10 listeners	-0.243	0.754	0.244	0.20	0.61	0.20

5.2.6 Discussion

Experiment 4 investigated whether the three acoustic correlates of manner contrast in Korean stops (VOT, H1-H2, and F0) were perceptually differently weighted between younger and older listeners. The findings from Experiment 4 suggested that F0 was more heavily weighted for younger listeners than older listeners. The MANOVA analyses (see Figures 5.2, 5.3, and 5.4) showed that the VOT threshold for /t^ha/ response was lower with younger listeners than older listeners consistently at all F0 levels, indicating a greater F0 effect on the /t^ha/ - /ta/ identification for younger listeners. The logistic regression analyses returned similar results. For the distinction of /t^ha/ - /ta/ and /ta/ - /t*a/ contrasts, the contribution of F0 was greater for younger listeners than older listeners. In addition, the correlation analyses showed that F0 weight was greater for younger listeners than older listeners for the perception of three response categories (/ta/, /t^ha/, and /t*a/)

The findings also indicated that H1-H2 was more heavily weighted for older listeners than younger listeners. The results of the logistic regression analyses showed that, for the /ta/ - /t*a/ and /t^ha/ - /t*a/ contrasts, the contribution of H1-H2 was greater for older listeners than younger listeners. The correlation analyses also showed that older listeners had greater H1-H2 weight than younger listeners for the perception of the three response categories (/ta/, /t^ha/, and /t*a/). In addition, the logistic regression analyses revealed that younger listeners were not affected by VOT for the distinction of the /ta/ - /t*a/ contrast, whereas older listeners were slightly affected.

The findings reported in Chapter V can be considered in light of the production differences between the younger and older speakers reported in Chapter II. In Chapter II, based on the sound change occurring among younger speakers of Korean (see section 2.4.3 for a summary of the sound change), different clear speech enhancement patterns were predicted for the production of Korean aspirated and lenis stops: younger speakers was predicted to enhance F0 difference in clear speech, whereas older speakers was predicted to enhance VOT difference. The results of the two production experiments revealed that younger speakers primarily enhanced F0 difference between the two stops with a marginal VOT enhancement, whereas older speakers enhanced the VOT difference with no enhancement in F0. The findings from Experiment 4 suggest that these production differences between the two groups are related to different perceptual weights of F0 between the two listener groups. The younger speakers primarily enhanced F0 difference of aspirated and lenis stops in clear speech production, and their perception of the aspirated and lenis stops were more greatly affected by the F0 modulation compared with the older listeners. The older speakers, who showed the enhancement of VOT difference in clear speech production, were far less affected by the F0 modulation than younger listeners for the perception of Korean aspirated and lenis stops.

Another finding from Experiment 4 is the primacy of F0 over VOT in its contribution to the distinction of /ta/ from the other /t^ha/ or /t*a/ types. The younger listeners who, in clear speech production, enhanced F0 difference with a marginal VOT enhancement had a greater contribution of F0 compared with VOT in the distinction of the /ta/ - /t^ha/ contrast (see Table 5.1 for odds ratio of F0 and VOT). The younger listeners also had greater F0 weight than older listeners for the perception of the /ta/, /t^{h}a/, and /t*a/ types (see Tables 5.4 and 5.7, 5.5 and 5.8, and 5.6 and 5.9 for group comparisons). These results indicate a matched relationship between production and perception. However, the older listeners who only enhanced VOT difference also had a greater contribution of F0 over VOT to the /ta/ and /t^{h}a/ distinction (odds ratios were 2.5 for VOT and 4.8 for F0 in Table 5.1), even though the contribution difference between F0 and VOT was not as large as the case of the younger speakers (odds ratios were 3.2 for VOT and 9.6 for F0. See Table 5.1 for odds ratio of F0 and VOT for both listener groups). These findings suggest that F0 is more heavily weighted than VOT for both younger and older speakers for the distinction of /ta/ and /t^{h}a/ contrast. For younger speakers, the weight of F0 might be heavy enough to induce the enhancement of this feature in clear speech production. For older listeners, the F0 weight might not be prominently heavier over VOT weight. Both younger and older listeners are exposed to the use of VOT and F0 by other speakers, young and old, all the time. They both might use F0 for the stop identification. However, F0 weight might be prominently heavy only for younger listeners in comparison to VOT weight.

The findings of Experiment 4 also mirror what has been found in previous investigations of Korean stop perception. In the perception experiments investigating the importance of consonantal and vocalic information in the perception of Korean initial stops, M.-R. Kim et al. (2002) found that low F0 of lenis stop + /a/ syllables provided dominant information over VOT and H1-H2 for the distinction of lenis stops from aspirated or fortis stops. This finding is repeated in the current experiment for both younger and older listeners. Figures 5.1 and 5.4 show the F0 effect of on the identification of lenis stops from aspirated stops. In Figure 5.1, the number of lenis /ta/ responses for the stimuli with the lowest 110 Hz was consistently greater than aspirated /t^ha/ response over the entire VOT range, and this was true for both younger and older listeners. Similarly, in Figure 5.4, the number of aspirated /t^ha/ responses for the stimuli with the highest 140 Hz was greater than the lenis /ta/ response at nearly all VOT levels, except for the lowest (10 ms).

The merger of /t^ha/ and /ta/ responses at 10 ms seen in Figure 5.4 is most likely to relate with /t*a/ responses. As seen in Figures 5.5 and 5.6 below, the mean number of /t*a/ responses was noticeably greater for the stimuli with 10 ms of VOT and 130 and 140 Hz of F0, compared with stimuli with lower F0 and longer VOT values. The mean number of /t*a/ responses for stimuli with 10 ms of VOT and 140 Hz of F0 reached up to 8 to 9. This great number of /t*a/ response is inversely related with /t^ha/ or /ta/ responses, and it is reflected in Figure 5.4.

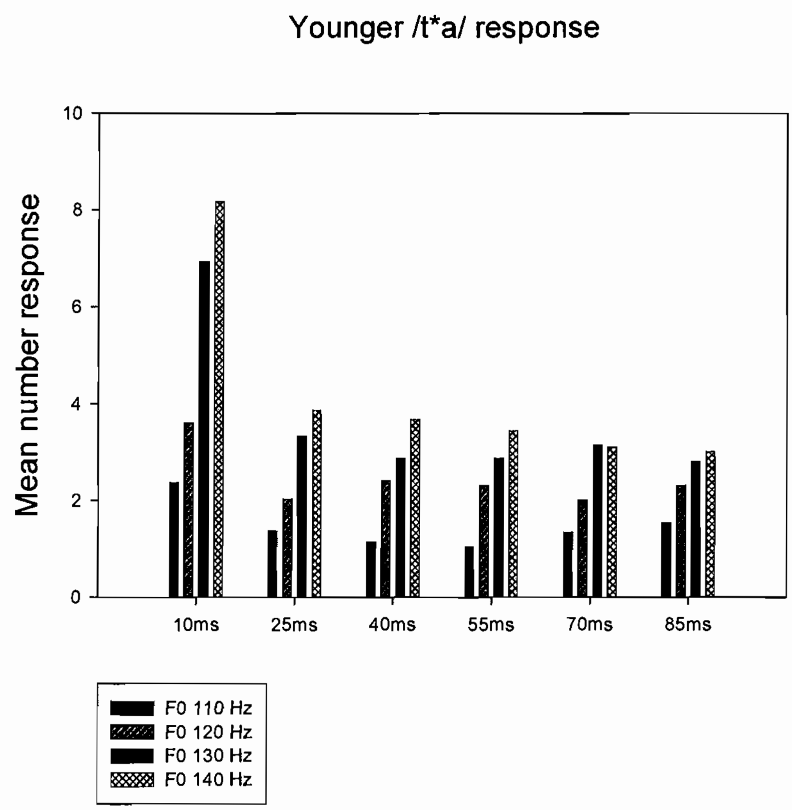


Figure 5.5. Mean number of /t*a/ response for younger group (n = 10).

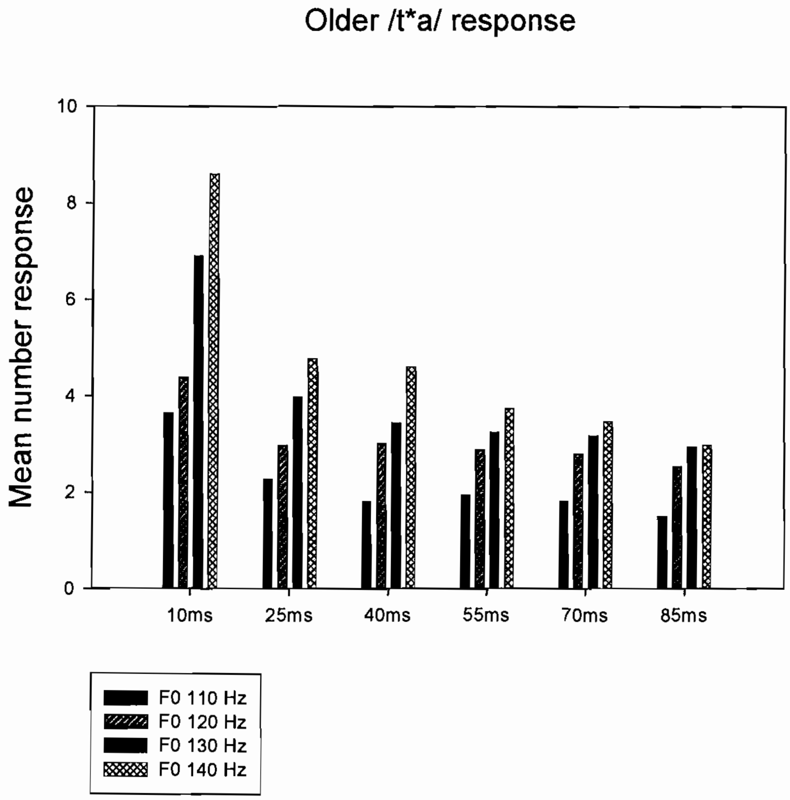


Figure 5.6. Mean number of /t*a/ response for older group (n = 10).

To summarize, the combined results of the production and perception experiments in Chapters III, IV, and VI suggest that younger and older speakers of the Seoul dialect of Korean have different phonetic targets for the production of Korean aspirated and lenis stops, such that the VOT distinction between the two stop types has diminished and F0 difference between the two stop types arose more salient for younger speakers. Accordingly, for the identification of aspirated and lenis stops, younger speakers are more greatly resort to the F0 contrast of the two stop types compared with older speakers.

CHAPTER VI

DISCUSSION AND CONCLUSION

The current dissertation examined clear, hyperarticulated productions of Korean stops to investigate the proposal that the phonetic targets of phonological categories are more closely approximated in hyperarticulated speech. This question was considered in light of sound change in Korean stops. The production and perceptual results indicated a perception-production link in the sound change, as well in clear speech production and perceptual processing. In this chapter, the experimental findings presented in Chapters III, IV, and V are further discussed in view of different theoretical models and empirical hypotheses.

6.1 Perception-production interaction

The generational differences in the production and perception of Korean stops found in the current study seem to be related and may be considered evidence for a perception-production link. Among various theories and studies that investigated perception-production interactions (Fox, 1982; Liberman & Mattingly, 1985; Guenther, 1995), Johnson et al. (2003) proposed the "hyperspace effect" for the link between speech production and perception. Adult native speakers of English produced / i, eɪ, ɛ, æ, ʌ, a, ɔ, oʊ, ʊ, u / in citation and hyperarticulated speech and the productions were compared to their selection of the best perceptual exemplars of the vowels. The results indicated that the perceptual vowel space had more extreme F1 and F2 than the vowels produced in citation speech and thus, the best exemplars were more closely aligned with the hyperarticulated vowels. Frieda (2000) replicated these findings with production and perception of English vowel /i/ using 35 participants. The findings suggested that the perceptual prototypes were more closely linked to hyperarticulated speech. In other words, perceptual prototypes for vowel categories were more fully approximated in hyperarticulated speech.

The results of the production experiments in Chapter III suggested different phonetic targets between the younger and older speakers regarding Korean aspirated and lenis stops. Consistent with the interpretations of Johnson et al. (1993) and Frieda (2000), if we assume different phonetic targets between the younger and older speakers, the perceptual prototypes would also be different between the two groups, and the different perceptual prototypes would be more explicitly realized in hyperarticulated speech. The results of the perceptual experiment in Chapter V suggested that the younger listeners had greater cue weight for F0 than older listeners, and the older listeners had greater VOT cue weight than younger listener for the perception of Korean aspirated and lenis stops. This perceptual behavior matched the production behavior observed through clear, hyperarticulated speech in Chapter III. The younger speakers, in clear speech, enhanced F0 difference between aspirated and lenis stops, along with a small enhancement of VOT difference, whereas the older listeners only enhanced VOT difference in the clear speech.

In short, the acoustic correlates were differentially enhanced in clear, hyperarticulated speech between the younger and older speakers, corresponding to the different perceptual cue weight between the two groups.

The interaction between production and perception may also be related to the reliability of the cue in speech processing. Holt and Lotto (2006) propose that listeners are sensitive to the distributional information of acoustic cues in auditory categorization, as well as absolute values of the acoustic information of the cues. VOT is a reliable cue for the American English voicing contrast distinction because the VOT range does not overlap between voiced and voiceless stops. VOT is informative for category identification, and this information is language specific. The informativeness of cues associated with distributional characteristics of acoustic dimensions may provide an explanation for the matched link between the production and perception of Korean stops. As described in Chapter III, younger speakers of the Seoul dialect are experiencing a diachronic change, such that VOT difference between aspirated and lenis stops is diminishing. As a result of the sound change, the VOT overlap between the two stop types are greater and more extensive among younger speakers compared with older speakers. The production data in Chapter III exhibited this trend. Hence, according to informativeness of cues, VOT should be a less informative cue for younger speakers for the aspirated-lenis stop distinction, compared with older speakers. The cue weighting patterns reflecting this line of perceptual behavior were observed in the perceptual data in Chapter V. VOT was relatively a less weighted cue for the younger listeners compared with the older listeners. By the same token, F0 functions as a more informative cue for the younger listeners corresponding to the enhanced F0 difference between the aspirated and lenis stops in clear, hyperarticulated production of the stops for the younger speakers.

Another finding related to perception-production interaction is that the older listeners had greater H1-H2 cue weight than younger listeners for the perception of fortis stops in distinction from lenis or aspirated stops (see Chapter V). This finding is surprising in that the younger and older speakers did not have significant differences in the enhancement of H1-H2 in clear speech. For the production of fortis stops, both younger and older speakers exaggerated the tense/creaky quality on the following vowel in clear speech fairly similarly, except that older speakers enhanced the tense/creaky quality in citation-form speech as well as clear speech. For aspirated or lenis stops, neither group enhanced the breathy quality on the following vowel in clear speech (see Figures 3.1(b) and 3.2(b)). Thus, the difference in H1-H2 weight between the younger and older speakers merits discussion. For the lenis-fortis stop distinction, the informativeness of cues associated with the distributional properties of acoustic dimensions predict that both F0 and H1-H2 are likely to be apparently informative cues because the two stop categories do not overlap in these two acoustic dimensions. Lenis stops are associated with low F0 and breathy voice quality, and fortis stops are associated with high F0 and tense/creaky voice quality on the following vowel (see section 2.4.1). The results in Chapter V showed that as was the case for the aspirated-lenis stop distinction, the younger listeners had greater sensitivity for F0 difference than older listeners for the lenis-fortis distinction. Having said this, for the younger listeners, the cue weights for F0 and H1-H2 are more evenly weighted with respect to each other, compared with the older listeners (see Table 5.2 for this pattern). In contrast, the older listeners had the relatively smaller cue weight of F0 and this might contribute to the greater dependency on H1-H2 for the perception of fortis stops in distinction from lenis or aspirated stops.

This sort of explanation assumes a trading relation in cue weighting among multiple cues (Repp, 1982). When there is a change in the setting of one acoustic cue and the phonetic percept of a category may be affected by that, a change in the setting of another cue may occur to compensate possible change in the percept. In this way, with the smaller cue weight of F0 for older listeners, H1-H2 might affect the older listeners' perception of the fortis stops more greatly. As described in section 5.2.3, the participants in the perceptual task were instructed to use their instant impressions when dealing with ambiguous test stimuli and were forced to give responses, even when they thought that none of the response choices matched what they listened. As mentioned in Repp (1983), a trading relation is observed among multiple acoustic cues when speech stimuli are phonetically ambiguous.

A different vein of explanation may be in the nature of sound change in the Korean stops. In the production data in Chapter III, the older speakers showed enhanced H1-H2 values for fortis stops in citation-form speech as well as clear speech, whereas the younger had the same enhancement only in clear speech (see Figures 3.1(b) and 3.2(b)). This observation suggests that H1-H2 may be more heavily weighted for older speakers than younger listeners. In fact, this pattern matches the greater H1-H2 cue weight for older listeners found from the perception experiment in Chapter V. What would account for the possibility of matched production-perception in H1-H2? The heavy cue weight for H1-H2 for older listeners may reflect a traditional cue weight. In the early literature on acoustic study of Korean stops (e.g., C.-W. Kim, 1965; Han & Weitzman, 1970), voice quality was considered a secondary cue to VOT or equivalently important as F0, but the importance of voice characteristic for the three-way manner distinction in Korean stops was clearly noticed. Keeping this in mind, what may be happening is that with the emergence of greater F0 weight, the cue weight for H1-H2 may be also withering along with VOT for younger listeners. As for the production side, this change is not yet vividly seen in hyperarticulated speech, unlike the cases of VOT or F0. Thus, in sound change, perception may precede production. In other words, perceptually, the different cue weight for H1-H2 is present there in the phonetic representations of the stops, and it does not appear in hyperarticulated speech and only realized in citation-form speech.

Another finding from the regression analyses in Chapter V also merits further discussion. For the distinction of the aspirated /t^ha/ and fortis /t*a/ contrast, VOT was not a great contributor to the distinction of the two stop types for both younger and older listeners, whereas H1-H2 was (see Table 5.3). This finding that VOT was not a great contributor was surprising because the aspirated and fortis stops have a large difference in VOT as well as in voice quality on the following vowel. VOT is longest for aspirated stops and shortest for fortis stops in the stop system. One possible explanation may be related to the nature of the test stimuli. In the current study, VOT and H1-H2 were not comparably manipulated. VOT values of the stimuli was manipulated by the equal increment of 15 ms in a range of 10 ms to 85 ms, whereas H1-H2 values had the characteristics of the natural speech tokens without manipulations, resulting in varying steps in H1-H2 values (1.5, 6.5, and -3.6 dB). According to Holt and Lotto (2006, p. 3060), even with equal physical steps in different acoustic dimensions, change in the phonetic percept may not be equivalent. Thus, the differential contribution of VOT and H1-H2 to the perception of /t^ha/ and /t*a/ types may be related to different perceptual effectiveness of the two cues derived from the variability in the robustness of the acoustic dimensions attributed to varying physical steps in the stimuli themselves. In other words, voice quality may be more robustly encoded by the auditory system than the durational property of timing mismatch between oral closure and onset of glottal vibration. This is also related to the reason why the focus of the analyses in Experiment 4 was on the group comparison.

The second stimuli-related explanation is that the robustness of the VOT dimension might be affected by the invariant stop burst in the test stimuli. As described in section 5.2.2 in Chapter V, VOT values of the stimuli were modified by cutting off the right end of the source VOT token, such that the burst energy remained constant across all of the stimuli. Cho et al. (2002) observed significantly greater burst energy for aspirated stops compared with lenis or fortis stops for the Seoul dialect of Korean. In a perceptual experiment on Korean stops, M.-R. Cho Kim (1994) found a small correlation between aspiration intensity and identification of the stop categories. With the greater intensity of aspiration, greater response of aspirated stop was elicited. However, in this study, the intensity of aspiration noise was not manipulated. Therefore, a direct relation between the burst energy and identification of the three stop types is not readily established. In order to overcome the potential limit derived from unnatural VOT characteristics of the stimuli, natural speech tokens containing full amplitude variation at stop burst across the stop types could be used in the future studies.

6.2 Conclusion

The current study provided experimental evidence to the proposal that phonetic targets are more closely approximated in hyperarticulated speech using the data from a language in which a sound change is in progress. Affected by a sound change in Korean stops, younger speakers of the Seoul dialect of Korean have different phonetic targets than older speakers of the same dialect for the production of Korean stops. The different phonetic targets between younger and older speakers were more vividly realized in hyperarticulated, clear speech. In addition, the differences in production were found to be linked to the differences in perceptual behaviors of the stops between the two groups. At the same time, the characteristics of greater intelligibility and enhanced phonemic contrast of clear speech were repeated with Korean stop production in this study.

The present study investigated the production and perception of Korean stops in three speaking styles: conversational speech, citation-from speech, and clear speech. The experimental results showed that the three speaking styles are different from one another in acoustic-phonetic and perceptual representations. The previous studies on clear speech only used two speaking styles, citation-form and clear speech, which may have some shared properties. The findings of this study may provide further insight into the reduction processes in speech production, as well as hyperarticulation processes. The findings are consistent with the hypo- and hyper-speech continuum hypothesis of the H & H theory (Lindblom, 1990).

As frequently mentioned in the literature (e.g., Cho et al. 2002; M.-R. Kim et al. 2002) the three-way stop contrast in VOT within the voiceless category is linguistically uncommon. Considering the typological atypicalness and the great inter- and intra-speaker variability of VOT, it is conceivable that the importance of VOT has reduced and speakers resort to cues other than durational property. What remains for future research is to investigate the relative importance of VOT and F0 in the production of Korean aspirated and lenis stops for the Kyungsang dialect of Korean. This dialect is particularly interesting to investigate because it has a trace of the tonal contrast in Middle Korean (10th ~ 16th C.). In the Kyungsang dialect, some lexical items still have a phonemic tonal contrast (Kenstowicz & Park, 2006). Would this more salient pitch contrast in the lexicon interfere with the pitch contrast of the stops and amplify the durational differences in the stops? Or, would the two cues work independently in the dialect and would they be equivalently great contributors to the distinction of the aspirated and lenis stops?

APPENDIX

KOREAN TEST WORDS

Korean words used in the production tasks for conversational, citation-form, and clear speech. The consonant-vowel sequences submitted to acoustic measurement are bolded.

1. Aspirated stop series

판판하다 /**p^han**.p^han.ha.ta/	'to be even/to be flat'
팔팔하다 /**p^hal**.p^hal.ha.ta/	'to be lively'
탄탄하다 /**t^han**.t^han.ha.ta/	'to be solid/to be firm'
탕감하다 / **t^haŋ**.kam.ha.ta/	'to write off a debt'
칼칼하다 /**k^hal**.k^hal.ha.ta/	'to be spicy (food)/to have a scratchy throat'
캄캄하다 /**k^ham**.k^ham.ha.ta/	'to be dark/to be gloomy'

2. Lenis stop series

반반하다 /**pan**.pan.ha.ta/	'to be comely/to be pretty'
발발하다 /**pal**.pal.ha.ta/	'to break out'
단단하다 /**tan**.tan.ha.ta/	'to be hard/to be strong'
당당하다 /**taŋ**.taŋ.ha.ta/	'to be dignified/to be imposing'
간간하다 /**kan**.kan.ha.ta/	'to be nicely salted'
감감하다 /**kam**.kam.ha.ta/	'to forget entirely/to have no news'

3. Fortis stop series

빤짝하다 /**p*an**.c*ak.ha.ta/	'to shine/to have ephemeral fame '
빵빵하다 / **p*aŋ**.p*aŋ.ha.ta/	'to be packed/to have good condition (person)'
딴딴하다 /**t*an**.t*an.ha.ta/	'to be strong/to be hard'
땅땅하다 /**t*aŋ**.t*aŋ.ha.ta/	'to be short and chubby'
깐깐하다 /**k*an**.k*an.ha.ta/	'to be fastidious/to be strict'
깜깜하다 /**k*am**.k*am.ha.ta/	'to be pitch dark/to be ignorant'

BIBLIOGRAPHY

Ahn, H. (1999). *Post-release phonatory processes in English and Korean: acoustic correlates and implications for Korean phonology*, Doctoral dissertation, University of Texas at Austin.

Boersma, P., & Weenink, P. (2005). Praat: doing phonetics by computer (Version 4.4) [computer program]. Retrieved December, 2005, from http://www. Praat.org/.

Bond, Z. S., & Moore, T. J. (1994). A Note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech communication, 14,* 325-337.

Bradlow, A. R. (2002). Confluent talker- and listener-related forces in clear speech production. In C. Gussenhoven & N.Warner (Eds.), *Laboratory phonology 7* (pp. 241-272). Berlin: Mouton de Gruyter.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancement for speech-in-noise recognition by native and non-native listeners, *Journal of the Acoustical Society of America, 121*, 2339-2349.

Bradlow, A. R, Torretta, G., & Pisoni, D. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech communication, 20*, 255-272.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112*, 272-284.

Cho, T., & Jun, S.-A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. In J. Boyle, J.-H. Lee, and A. Okrent (Eds.), *Chicago Linguistic Society 36* (pp. 31–44). Department of Linguistics, University of Chicago, Chicago, IL.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics 29*, 155–190.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives, *Journal of Phonetics 30,* 193-228.

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First language, 25*, 219-234.

Ferguson, S. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America, 116*, 2365-2373.

Ferguson, S., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*, 259-271.

Ferguson, S., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowel. *Journal of speech, language, and hearing research, 50*, 1241-1255.

Fox, R. A. (1982). Individual variation in the perception of vowels: implications for a perception-production link. *Phonetica, 39*, 1-22.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (2000). Adults' perception and production of the English vowel /i/. *Journal of speech, language, and hearing research, 43*, 129-143.

Gagne, J.-P. (1994). Across variability in auditory, visual, audiovisual speech intelligibility for conversational and clear speech. *Journal of the Academy of the Rehabilitative Audiology, 27*, 135-158.

Garner, W. R. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network of speech production. *Psychological Review, 102*, 594-621.

Han, M. S., & Weitzman, R. S. (1970). "Acoustic features of Korean /P,T, K/ /p, t, k/ and /p^h, t^h, k^h/," *Phonetica 22*, 112-128.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean, *Journal of Phonetics 1*, 263-272.

Harnsberger, J., Wright, R. and Pisoni, D. (2008). "A new method for eliciting three speaking styles in the laboratory," *Speech Communication* 50, 323-336.

Hay, J. F., Sato, M., Coren, A. E., Moran, C. L., & Diehl, R. L. (2006). Enhanced contrast for vowels in utterance focus: A cross-language study, *Journal of the Acoustical Society of America, 119*, 3022-3033.

Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America, 116*, 3108-3118.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implication for first and second language acquisition. *Journal of the Acoustical Society of America, 119*, 3059-3071.

Hsu, C.-S., & Jun, S.-A. (1998). "Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic?," *UCLA working papers in phonetics 96*, 69–89.

Johnson, K., Flemming, E., & Wright, R. (1993). "The hyperspace effect: Phonetic targets are hyperarticulated," *Language 69*, 505-528.

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*, Doctoral dissertation. Ohio State University, Columbus, Ohio.

Jun, S.-A. (1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean. In B. Connell & A. Arvaniti (Eds.), *Paper in laboratory phonology IV: Phonology and phonetic evidence* (pp. 235-253). Cambridge, UK: Cambridge University Press.

Kagaya, R. (1974). A fiberscopic and acoustic study of Korean stops, affricates, and fricatives, *Journal of Phonetics 2*, 161-180.

Kang, K.-H., & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America, 119*, 1672-1683.

Kang, K.-H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America, 124*, 3909-3917.

Kenstowicz, M., & Park, C. (2006). Laryngeal Features and Tone in Kyungsang Korean: a Phonetic Study. *Studies in Phonetics, Phonology and Morphology,12*, 247-264.

Kim, C.-W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops), *Word 21*, 339-359.

Kim, M. (2004). Correlation between VOT and F0 in the perception of Korean stops and affricates. *INTERSPEECH 2004*, 8th International Conference on Spoken Language Processing, Jeju Island, Korea, October 4 - 8, 2004.

Kim, M.-R. Cho. (1994). *Acoustic characteristics of Korean stops and perception of English stop consonants,* Doctoral dissertation, University of Wisconsin at Madison.

Kim, M.-R., Beddor, P.S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics, 30*, 77-100.

Kim, Y.-H. (2007). The effect of attention on discrimination between lenis /t/ and aspirated /t^h/ of Korean alveolar stops as shown by the Korean and the Japanese. Doctoral dissertation, Seoul National University. (in Korean).

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 67, 971-995.

Krause, J., & Braida, L. (2002). Investigating alternative forms of clear speech: The effects of speaking rates and speaking mode on intelligibility. *Journal of the Acoustical Society of America, 112*, 2165-2172.

Krause, J., & Braida, L. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America,115*, 362-378.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech revised. *Cognition, 21*, 1-36.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. J. Hard-castle and A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht, Netherlands: Kluwer Academic.

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

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mother to children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology (Vol. 2)* (pp. 173-184). San Diego: Academic Press.

Moon, S.-J. (1991). *An acoustic and perceptual study of undershoot in clear and citation-form speech.* Doctoral dissertation. University of Texas at Austin.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*, 40-55.

Morrison, G. S. (2007). Logistic regression modeling for first-and second-language perception data. In Eds.), *Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory 282*, 219-236.

Ohala, J. J. (1994). "Acoustic study of clear speech: A test of the contrastive hypothesis," *International symposium on prosody, 18*, 75-89.

Payton, K., Uchanski, R., & Braida, L. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America, 95*, 1581-1592.

Perkell, J. S., Zandipour, M., Matthies, M. L., & Harlan, L. (2002). Economy of effort in different speaking condition. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America, 96*, 1627-1641.

Picheny, M., & Durlach, N. (1979). Speaking clearing for the hard of hearing. *Journal of the Acoustical Society of America, 65, Supplement 1.*, S 135

Picheny, M., Durlach, N., & Braida, L. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech, *Journal of speech and hearing research, 28*, 96-103.

Picheny, M., Durlach, N., & Braida, L. (1986). Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of speech and hearing research, 29*, 434-446.

Picheny, M., Durlach, N., & Braida, L. (1989). Speaking clearly for the hard of hearing III: an attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of speech and hearing research, 32*, 600-603.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty and D.R. Ladd (Eds.), *Papers in laboratory phonology II: gesture, segment, prosody* (pp. 90-117). Cambridge, U.K; Cambridge Univ. Press.

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological bulletin, 92*, 81-110.

Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech communication, 2*, 341-361.

Shimizu, K. (1996). *A cross-language study of voicing contrasts of stop consonants in Asian languages*. Tokyo: Seibido Publishing.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology, 23*, 287-308.

Silva, D. J., Choi, Y.-J., & J. E. Kim. (2004). Diachronic shift in VOT values for Korean stop consonants. In S. Kuno, I.-H. Lee, J. Whitman, J. Maling, Y.-S. Kang, & Y.-J. Kim (Eds.), *Harvard studies in Korean linguistics X.S* (pp. 173-185). Cambridge, MA: Department of Linguistics, Harvard University.

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: talker and listeners factors in speaking style changes, *Linguistics and Language Compass, 3(1)*, 236-264.

Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English, *Journal of the Acoustical Society of America, 118*, 1677-1688.

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian, *Journal of Phonetics 36*, 91-113.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica, 56*, 186-199.

Uchanski, R., Choi, S., Braida, L., & Durlach, N. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of speech and hearing research, 39*, 494-509.

Wright, J. D. (2007). *Laryngeal contrast in Seoul Korean*, Doctoral dissertation, University of Pennsylvania.