Cue Primacy and Spontaneous Imitation: Is Imitation Phonetic or Phonological?

by

Harim Kwon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in the University of Michigan
2015

Doctoral Committee:

Professor Patrice S. Beddor, Co-Chair
Associate Professor Andries W. Coetzee, Co-Chair
Professor Julie Boland
Associate Professor Kuniko Y. Nielsen, Oakland University

To my parents,

with love and gratitude

ACKNOWLEDGEMENTS

It is a great pleasure to have this opportunity to thank so many special people without whom I could never have completed this dissertation.

First of all, thank you, my amazing advisors, Pam Beddor and Andries Coetzee. I have been extremely fortunate to have not only one but two most caring and supportive advisors. Pam and Andries have not only taught me how to write and how to do research, but have also been my role models of being an inspiring teacher, a dedicated researcher, and a happy and nice person, without sacrificing one for another. I could have never come this far without them. Andries, thank you for your calming and comforting words when I was stressed out or overwhelmed, and also for having faith in me even in the times I doubted myself. Pam, my graduate school mom, you have been incredibly encouraging, caring, and patient all along, and readily available whenever I needed you no matter how busy you were. Anything that is good about this dissertation, I owe almost entirely to Pam and Andries. Thank you as always!

I am also grateful for my terrific committee members, Julie Boland and Kuniko Nielsen. I sincerely appreciate all the inputs they have provided from the very beginning—Kuniko, I was first interested in spontaneous imitation thanks to you, and Julie, this project first became concrete as a term paper that I wrote for your course—till now. Also, thank you to the Phondi (Phonetics-Phonology discussion group) members, who have discussed this project with me many many times. I also wish to thank numerous anonymous participants without whom this project would have been literally impossible.

I have been lucky enough to be surrounded by supportive faculty here at the University of Michigan. A very special thanks goes to Jelena Krivokapić and Marlyse Baptista, who have supported me like one of their own students. I would also like to express deep appreciation to my other teachers, Sam Epstein, San Duanmu, Debbie Keller-Cohen, Acrisio Pires, Ezra Keshet, Robin Queen, Carmel O'Shannessy, and Anne-Michelle Tessier, for the lessons they have taught and the support they have given.

I also wish to thank my MA advisor, Chang-Yong Sohn, who first introduced me to academia. Without his guidance and inspiration, my first conference presentation at Bergamot, Italy—the spark that initiated my stride as a scholar—would never have happened. I also feel grateful to all my English teachers that I have met at some point of my life.

Many thanks are due to my friends in the department. My dear lab mates, Susan Lin, Kevin McGowan, Jon Yip, Cameron Rule, Kate Sherwood, Sagan Blue, Jiseung Kim, and Ian Calloway, and many other friends, David Medeiros, Joseph Tyler, Stephen Tyndall, Jae-Young Shim, Sujeewa Hettiarachchi, Mike Opper, Batia Snir, Hayley Heaton, Dave Ogden, Kenneth Lim, Theo Stern, Ariana Bancu, Alan Ke, Marjorie Herbert, and Jen Nguyen, I have never felt alone on the fourth floor mezzanine of Lorch, because of you. A special thanks to my cohorts, Tridha Chatterjee, Candice Scott, and Yan Dong, for sharing the ups and downs of my graduate years. And to Miyeon Ahn, thank you for being my big sister in Ann Arbor. My life in Ann Arbor would have been far less enjoyable had it not been for you.

A very special thanks goes to my dearest friends who were far away in distance but closest to my heart during the hardest times. Koeun, Hanbyul, and Yun Jung, thank you very much for being always right there for me despite the entire continent and the Pacific Ocean between us. I love you and thank you for assuring me that I am a decent human being.

Finally, I owe my deepest gratitude to my family in Korea. I wholeheartedly thank my parents for their unwavering faith in me even in the times I was too engrossed in my work to be there for them when they needed their daughter. I also feel grateful to Sumin, my little brother, for being there for them during my absence. I love you all.

엄마, 아빠, 나이만 먹었지 철은 안 든 딸이 들라는 철은 안 들고 논문을 썼어요. 불안하고 궁금하셨을텐데 묵묵히 믿고 지켜봐주셔서 고맙습니다. 짧지 않은 시간동안 타지에서 공부한다는 핑계로 딸 노릇도 제대로 못했어요. 죄송해요. 사랑합니다.

TABLE OF CONTENTS

LIST OF TABLES

<u>**Table**</u>

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

Previous research on spontaneous imitation examines how speaker-listeners' own production changes after hearing a few minutes of model speech, and suggests that speech perception and production are closely related. This dissertation investigates how cue primacy influences imitation by separately manipulating two co-varying cues differing in their primacy for one phonological category. By examining how similarly or differently primary and non-primary cues operate in spontaneous imitation, this dissertation studies the nature of the cognitive representations that are responsible for imitation.

In order to examine whether the cognitive representations that are involved in speech imitation are abstract phonological categories or individual phonetic properties, this study tests spontaneous imitation of aspirated stops by Seoul Korean speakers. In Seoul Korean, at least two distinct acoustic properties, stop voice onset time (VOT) and post-stop fundamental frequency (*f0*), differentiate aspirated stops from stops of different phonation types, with post-stop *f0* being the primary cue for aspirated stops. Seoul Korean participants heard and shadowed (i.e., immediately repeated what they heard without being told to imitate) target model speech in which initial aspirated /t^h/ was enhanced with either extended VOT or raised post-stop *f0*. Speakers' realization of these properties in their own /t^h/, /t/, and /t*/ productions were compared before, during, and after exposure.

The results show that enhancements of both primary and non-primary cues trigger imitative changes, and that exposure to an enhanced non-primary cue (long VOT) influences the production not only of that cue but also of the primary cue for aspirated stops (post-stop *f0*). However, an enhanced primary cue (high *f0*) does not have similar effects on the non-primary cue. Moreover, the imitative changes are generalized to maximize the relevant phonological contrast, as evidenced by lowering of *f0* after lax /t/ and sonorants. These findings suggest that imitation is not strictly tied to individual phonetic properties but it is rather phonological in that abstract categories are involved in the process of imitation. This dissertation provides a new insight on the role of phonology in spontaneous imitation.

CHAPTER I

Introduction

Language users are both producers and perceivers of speech. It is therefore perhaps unsurprising that a tight relation between production and perception has been proposed by different phonetic theories (e.g., the Motor Theory, Liberman & Mattingly, 1985; Direct Realism, Fowler, 1986, 1996; and exemplar models, Johnson, 1997). One experimental paradigm that has widely been used for a few decades to investigate this relation is spontaneous speech imitation. Researchers now have ample evidence that speakers shift their productions in the direction of the model speech that they have just heard (e.g., Goldinger, 1998); that is, what listeners perceive influences what listener-turned-speakers produce. An unresolved issue in the theoretical discussion of these substantial findings is the nature of the cognitive bridge between what listeners perceive and what listener-turned-speakers produce. This dissertation aims to contribute to the discussion by investigating the roles of two co-varying cues for one phonological category in spontaneous imitation when the two cues differ in their primacy. The guiding question of this study is, what is the cognitive representation that is involved in the process of spontaneous imitation. To address the question, I examine imitative changes in Seoul Korean speakers' stop production after they heard model speech including aspirated stop /t^h/ that is artificially enhanced using two co-varying cues for aspirated stops in Korean.

1.1 Spontaneous speech imitation

Spontaneous speech imitation, since Goldinger's (1998) seminal study, has been widely used as an experimental method to test the relation between speech perception and production. Speakers' productions become more similar to what they have recently perceived (i.e., model speech), both in conversational interactions and in non-conversational settings, and regardless of whether the model speech involves natural utterances or artificially manipulated ones. This phenomenon of perceptually induced, unintentional changes in speakers' subsequent productions is referred to as imitation (e.g., Babel, 2012; Goldinger, 1998; Nielsen, 2011), convergence (e.g., Pardo, 2006; Pardo, Gibbons, Suppes & Krauss, 2012), accommodation (e.g, Kim, 2012; Kim, Horton & Bradlow, 2011), or mimesis (Delvaux & Soquet, 2007). These different terms refer to different experimental conditions. Imitation commonly refers to findings from studies conducted in one-way settings in which participants hear model speech without interaction with an interlocutor (Babel, 2012; Goldinger, 1998; Nielsen, 2011, among others). Convergence, on the other hand, best captures what happens in natural conversational interactions in which the interlocutors often *converge* towards each other (Pardo and colleagues' studies). Accommodation is the most comprehensive among these terms, as pointed out by Kim (2012), as it includes not only convergence but also divergence and maintenance, which are often socially motivated. And finally, mimesis is a special kind of imitation, accompanying changes in relevant motor representations (Delvaux & Soquet, 2007). The experimental setting of the current study can be best described as one-way imitation, comparable to the designs used by Goldinger (1998), Nielsen (2011), and Babel (2012). Therefore, I will use *spontaneous (speech) imitation* to refer to unintentional changes in speakers' productions as a result of perceiving model speech.

1.1.1 Theoretical accounts of spontaneous speech imitation

Spontaneous speech imitation provides evidence that speech perception and production are closely related and, therefore, findings from spontaneous imitation studies are of great importance for phonetic theories that propose a strong link between the two processes. Here, I will focus on two theoretical accounts of spontaneous imitation, an episodic account (e.g., Goldinger, 1998) and a gestural account (e.g., Fowler, Brown, Sabadini & Weihing, 2003), and review their main findings and interpretations regarding speech imitation. There also exists another large body of research on social motivations for speech imitation, such as Communication Accommodation Theory (Giles, Coupland & Coupland, 1991; Shepard, Giles & Le Poire, 2001), but it will not be discussed here since social aspects of imitation (or more appropriately, accommodation) are outside the scope of the present study.

In episodic or exemplar-based models of speech perception (e.g., Johnson, 1997, 2006; Pierrehumbert, 2001), exemplars (i.e., detailed memories of specific linguistic experiences including phonetic details as well as non-linguistic information such as speaker- or situation-related details) are stored in memory and form perceptual categories that are defined as sets of all the exemplars belonging to the category. In this approach, when a listener perceives a word, a new exemplar is created and existing exemplars that are associated with the incoming exemplar are activated. Because activated exemplars contribute to subsequent productions in most approaches, detection of a new exemplar has the potential to lead to spontaneous imitation (e.g., Goldinger, 1998; Tilsen, 2009).

Goldinger (1998) examines spontaneous speech imitation using a single-word shadowing task in which participants heard and shadowed (i.e., immediately repeated) the model speech. The comparison between participants' (pre-shadowing) baseline production and shadowed production revealed that the participants indeed imitated the model speech. Also, the degree of imitation is greater in low-frequency than it is in high-frequency words, and in words that the participants heard more times. Goldinger demonstrates, using Hintzman's (1986) MINERVA2, that these effects of lexical frequency and repetition can be readily explained in an exemplar-based system without additional complexities. Low-frequency words show strong imitation effects because the weight of each new exemplar is relatively greater when only a small number of existing exemplars are associated with a given word. Also, more repetitions mean more exemplars associated with the model speech are newly added, increasing the degree of imitation.

Gestural theories of speech perception, particularly Direct Realism (Fowler, 1986, 1996), provide a rather different explanation for the mechanism underlying spontaneous imitation. The central claim of gestural theories is that listeners perceive vocal tract gestures. This intrinsic link between perception and production leads to the expectation that perception may have an immediate impact on succeeding production (e.g., Fowler et al., 2003; Honorof, Weihing & Fowler, 2011; Shockley, Sabadini & Fowler, 2004). For example, Fowler et al. (2003) find that response latency in a shadowing task in which participants were asked to repeat the syllable they heard is not particularly longer than that in a simple choice task in which participants were asked to produce a pre-assigned syllable. According to Fowler et al., the small latency difference between the shadowing task and the simple choice task indicates that the shadowing task does not involve choice making, but listeners directly perceive gestures and reproduce them. Because gestures are the common currency for speech perception and production, listeners can rapidly access the articulatory information upon listening, resulting in short response latency in the shadowing task.

Perceptually-guided changes in production become especially evident when using manipulated model speech because those manipulations are perceived as the acoustic consequences of different articulatory gestural configurations. For instance, upon hearing artificially lengthened voice onset time (VOT) of voiceless stops (Fowler et al., 2003; Shockley et al., 2004), listeners extract information about relative timing between the oral constriction gesture and the glottal opening-and-closing gesture, and that information guides what listener-turned-speakers produce. As another example, Honorof et al. (2011) examine imitation of American English velarized /l/ by independently manipulating two constriction gestures involved in the lateral independently, and find that participants reproduce aspects of articulatory configuration manifested in the model speech even when the specific articulatory pattern does not match the participants' own phonology. These findings are interpreted as support for Direct Realism's claim that the target of speech perception—and thus imitation—is articulatory rather than acoustic/auditory.

The two theoretical accounts are not in direct opposition to each other, nor do they make wholly incompatible predictions. Rather, they focus on different aspects of spontaneous imitation. While the exemplar account claims that the nature of the memory system gives rise to observed patterns of imitation with regard to the effects of lexical frequency, recency, and amount of exposure, the direct realist account limits itself to rapid shadowing and claims that the intrinsic link between speech perception and production at the articulatory level explain the rapid and direct imitation. In fact, Honorof et al. comment that they are not against the core claim of exemplar accounts that phonetic and non-linguistic details are stored in memory, if the phonetic details are not auditory but articulatory. And although some versions of exemplar models make it clear that the phonetic properties stored in exemplars are auditory that is output from the peripheral auditory system (e.g., Johnson, 1997, 2006), other models are agnostic, allowing phonetic detail to be either be holistic gestural or acoustic templates that are associated with word meanings, as long as the theory can correctly describe the interaction of word-specific phonetic detail with more general principles of phonological structure (e.g., Pierrehumbert, 2001). The results of the current study will be discussed related to the two theoretical accounts (gestural vs. exemplar-based), without attempting to evaluate the two theories.

1.1.2 What is imitated?

Spontaneous imitation has been examined both in conversational interactions and in non-interactional settings. Previous studies that tested spontaneous imitation in conversational interactions mostly focus on the impact of social factors on the extent of imitation (e.g. Kim et al., 2011; Pardo, 2006; Pardo et al., 2012; Pardo, Jay & Krauss, 2010). Of particular interest here are the findings of non-interactional studies, that is, of investigations that more closely match the design of the experiments conducted for this project. For example, Babel (2012) uses an imitation task in which participants heard and shadowed model speech with a photo of the purported speaker visually presented and finds imitation of the properties of the model speaker's vowels. Babel further reports that participants' attitude (social liking) towards the model speaker affects the degree of imitation, suggesting that imitation is socially mediated even in a supposedly non-social setting. Kim (2012) also shows spontaneous imitation of English words and sentences using multiple acoustic measures after passive auditory exposure to model speech. Results of both of these studies provide robust evidence for spontaneous imitation in non-interactional settings using naturalistic, unmanipulated model speech. Delvaux and Soquet (2007) and Mitterer and Müsseler (2013) are similar in this regard; their findings demonstrate that participants' productions shift towards those of model speakers showing characteristics of different regional dialects in a non-interactional setting.

As speech variation is ubiquitous, it can be desirable to make use of already existing variations to test speech imitation. However, testing imitation using natural model speech can become difficult if participants' productions are already similar to the model speech. Furthermore, it is not easy to assess the degree of imitation with natural model speech; precisely because of the substantial variation, researchers cannot know which property(s) of the model speech the listeners would attend to and adjust. As pointed by Pardo (2013), the imitative changes in production often occur on multiple phonetic properties, and it is also possible that different listeners attend to different properties of the same model speech.

Two different strategies have been used in the literature to address this issue. First, many studies provide perceptual judgments by a separate set of listeners as a holistic measure of degree of imitation (Goldinger, 1998; Goldinger & Azuma, 2004; Kim, 2012; Kim et al., 2011; Pardo, 2006; Pardo et al., 2010, 2012). Most commonly used is an AXB task, in which a separate set of listeners judge whether the participant's production A or B is a better match to model speech X. One of A or B is a production either before exposure to the model speaker (in passive listening tasks) or from early in conversational interactions with the interlocutor; the other is a post-exposure or late-interaction production. Although perceptual judgments obtained from AXB tasks provide a global perceptual measure of imitation fidelity, if the question of interest is about *what is imitated*, or which acoustic property is susceptible to imitation, global perceptual judgments are not especially informative.

A different approach, which focuses on participants' imitation of targeted properties, is to use model speech that is acoustically manipulated. Because listeners are especially likely to imitate salient properties of what they hear (Mitterer & Müsseler, 2013; Zellou, Scarborough & Nielsen, 2013, see §1.1.3 for more detailed discussion of this issue), using artificially manipulated stimuli can arguably narrow down the potential imitative adjustments of listener/speakers, which facilitates (and justifies) the choice of specific production measures. For instance, properties known to be susceptible to imitation include artificially lengthened Voice Onset Time (VOT) of English voiceless stops (Fowler et al., 2003; Nielsen, 2011; Shockley et al., 2004), increased and decreased coarticulatory vowel nasality in English (Zellou et al., 2013), and manipulated sub-phonemic details in vowel formants (Tilsen, 2009). In addition, Honorof et al. (2011) use articulatory measures and show that allophonic variations of /l/ in American English are imitated.

Overall, the evidence from previous studies clearly suggests that speakers shift their productions in the direction of the model speech that they have just heard and that various phonetic properties are susceptible to imitation. However, the question of *what is imitated* is not easy to answer with an inductive approach. That is, knowing that some phonetic properties are imitated while others are not does not appear to help answer the question of *what is imitated,* because lack of imitative changes might not mean that the properties are not susceptible to imitation, due to many factors that can affect imitation, including individual proclivity for imitation, social factors, and phonology. The role of phonology in imitation is discussed in detail in §1.1.3. This dissertation addresses the question *what is imitated* from a different perspective, without learning from what is not imitated. Instead, this study examines how two co-varying cues for one phonological category are differently imitated when the two cues differ in their primacy.

1.1.3 Phonology in spontaneous imitation

One intriguing often-asked question in the spontaneous speech imitation literature concerns the role of phonology in imitation. Previous studies have suggested that imitation is mediated by phonology in some ways, although interpretation of the findings is not always straightforward.

That phonology mediates imitation is suggested by negative evidence that shows that not all manipulated properties are imitated. For instance, Nielsen (2011) reports that although artificially lengthened VOT for English voiceless /p/ is imitated by English speakers, artificially shortened VOT is not, consistent with imitation being attenuated by the presence of a phonological boundary. This outcome suggests that the imitation is phonologically selective: imitation is avoided if it might threaten a phonemic contrast (here, /b/-/p/).

In addition, Mitterer and Ernestus (2008) propose a yet stronger role of phonology in spontaneous imitation. Mitterer and Ernestus examine whether different variants of Dutch /r/ (alveolar and uvular trills) that are different articulatorily but equivalent phonologically are imitated, and find that participants hardly ever deviate from their habitual articulation to imitate the variant they heard. Moreover, the response latency does not increase due to a gestural mismatch between the modeled speech and participant's response. Based on these findings, Mitterer and Ernestus claim that imitation occurs on an abstract phonological level. This claim is not uncontroversial, however. Honorof et al. (2011) suggest that the stimuli Mitterer and Ernestus used for the shadowing task may have inhibited the imitation and could have prevented the increase in response latencies. Mitterer and Ernestus's shadowing stimuli are disyllables whose two syllables are separated by 500ms, and a longer interval between perception of a syllable and its production can induce one to forget perceived details (Honorof et al., 2011).

Mitterer and Ernestus (2008) also report that duration of pre-voicing of Dutch voiced stops is not imitated in a shadowing task, attributing the lack of imitation to the phonological irrelevance of the properties. According to Mitterer and Ernestus, in Dutch, presence vs. absence of prevoicing, but not its temporal extent, is phonologically relevant and, consequently, listeners do not imitate longer vs. shorter pre-voicing. This claim is also controversial for several reasons. As mentioned in §1.1.2, many phonetic properties have been reported to be susceptible to imitation, and the list of imitable properties is not always limited to "phonologically relevant" properties. For instance, fundamental frequency (Babel & Bulatov, 2012) and duration (Kim, 2012; Pardo, Jordan, Mallari, Scanlon & Lewandowski, 2013) of English vowels in single-word productions is arguably not phonologically relevant. Moreover, as pointed by Honorof et al. (2011), if longer vs. shorter pre-voicing of Dutch voiced stops is not phonologically relevant, extended VOTs of English voiceless stops, one of the most robustly imitated properties, are arguably not relevant as well, because the lengthened VOTs (e.g., around 130 ms in Fowler et al., 2003 and 110 ms in Nielsen, 2011) are clearly within the aspirated allophonic category of English voiceless stops.

Mitterer and Müsseler (2013) provide an alternative interpretation of the non-imitation of duration of pre-voicing in Dutch: the difference between longer vs. shorter pre-voiced stimuli used in Mitterer and Ernestus (2008) might have not been perceptually salient, precluding imitation. And, in fact, some recent findings, including Mitterer and Müsseler's own, provide (unsurprising) evidence that perceptually more salient variation induces more robust imitation. Mitterer and Müsseler examine the imitability of two different regional variations in German, and show that not only more marked dialectal variations, but also more variation in stimulus presentation, lead to more robust imitation during shadowing. Specifically, fricative-stop cluster variants [st~ʃt] induce more imitation than do *-ig* variants [ɪk~ɪç], because the former is more marked, such that it clearly indexes a non-standard regional dialect, than the latter. Also, the imitation effect was greater when participants heard both variants than when they heard only one variant in an experiment, arguably because hearing both variants in the same experiment makes the variation more salient (Mitterer & Müsseler, 2013).

Perceptual salience is not independent from phonology. Phonologically less typical or natural variants often lead to more robust imitation, presumably because they are perceptually more salient (Honorof et al., 2011; Zellou et al., 2013). For instance, Honorof et al. (2011), in a series of shadowing experiments using manipulated variants of American English /l/, find that more velarized variants of /l/ in a syllable onset induced greater imitation in shadowing. Velarized /l/ in syllable onset is less typical in some varieties of American English and, therefore, their findings suggest that a less natural variant with regard to allophonic rules of a language facilitates greater imitation.

Zellou et al. (2013) provide further evidence of the likely role of perceptual salience due to phonological unnaturalness in spontaneous imitation. Zellou et al. investigate the imitation of coarticulatory vowel nasalization in English words from dense neighborhoods and suggest that the phonological unnaturalness—and consequent perceptual salience—of the model speech promoted long-term imitation. Zellou et al. find that both an increase and a decrease in coarticulatory nasality of English vowels were imitated during single-word shadowing tasks. However, only the imitation of decreased nasality persisted into a post-shadowing word-reading task. Because English words from dense neighborhoods naturally have a greater degree of coarticulatory nasality than those from sparse neighborhoods (Scarborough, 2004), their stimuli with decreased nasality were less natural and hence more perceptually salient, which could have contributed to the relatively long-term imitation of decreased nasality. A more salient difference in the stimuli arguably leads to longer-lasting imitation.

Zellou et al. also report that it was not the actual degree of nasality but the increase or decrease in degree of nasality that was imitated. That is, participants produced weaker vowel nasalization after hearing the stimuli that had been manipulated to be less nasal than the model speaker's natural productions, even when the stimuli still had more nasality than the participants' own baseline. (This was possible because the model speaker was a heavy nasalizer.) According to Zellou et al., the abstract degree of nasality was computed in comparison with oral fillers and then the normalized degree, instead of the actual degree, of nasality was imitated. This suggests that spontaneous imitation operates arguably at an abstract phonological level rather than at a physical level.

Another way to investigate potential impacts of phonology on the process of spontaneous imitation is through examining how imitation effects are generalized to unheard segments or words. In the study by Nielsen (2011), after hearing English target words beginning with /p/ with extended VOT, participants produced extended VOT on unheard /p/-initial words and /k/-initial words as well as on the exposed target words. These productions indicate both phoneme-level generalization (to new stimuli including the same segment of exposure) and feature-level generalization (to a new segment that shares a feature), and point toward the influences of these phonological units on the effects of imitation.

Kim (2012) also examines the generalizability of the imitation effects at the word level using various acoustic measurements, and at the sentence level using dynamic time warping analyses and perceptual judgments (XAB tests). Her acoustic measures (VOT of voiceless stops; duration, fundamental frequency, and formant frequencies of vowels) all reveal the generalization of imitation effects to unheard words. Also, both dynamic time warping analyses and perceptual judgments suggest imitation effects are generalized even at the sentence level. This, adding to Nielsen's findings, suggests that imitation is robustly generalized and that the effects of imitation might not be tied to certain phonetic properties or even to phonological categories.

Despite the flourishing literature on spontaneous speech imitation, research on the role of phonology in imitation remains in the relatively early stages. The growing body of literature alludes to the possibility that spontaneous imitation may not be tied to a certain acoustic property or an articulatory gesture. The current study attempts to gain further insights into the role of phonology spontaneous speech imitation, with an overarching goal of better understanding the cognitive representations that are involved in imitation. This complicated issue is broken down into three smaller questions: (1) which acoustic properties trigger imitation, (2) which aspects of articulation change as imitative adjustments, and (3) how widely the imitative changes are generalized. These questions are investigated by examining the respective role of two co-varying cues for aspirated stops of Seoul Korean in spontaneous imitation.

1.2 Seoul Korean voiceless stops

1.2.1 Post-stop f0 and stop VOT: primary and non-primary cues for aspirated stops

Korean has a three-way laryngeal contrast for voiceless stops: tense or fortis /p*, t*, k*/, lax or lenis /p, t, k/, and aspirated /p^h, t^h, k^h/. All three categories are phonetically voiceless in word-initial or phrase-initial positions, having a positive voice onset time (VOT) although the lax plosives are often voiced word- or phrase-medially (Jun, 1993). The three laryngeal categories differ in multiple acoustic properties, including stop VOT, *f0* (fundamental frequency) of the post-stop vowel, and *H1-H2* following stop release. The discussion here focuses only on VOT and *f0* differences, which the current study uses.

Earlier analyses of Korean stops describe tense stops as having short VOT and high *f0*, lax stops as having longer VOT and lower *f0*, and aspirated stops as having the longest VOT and higher *f0* (e.g., Cho, Jun & Ladefoged, 2002; Kagaya, 1974; C.-W. Kim, 1965; M.-R. C. Kim, 1994). More recent studies, however, report that VOT values are merged for the lax and the aspirated categories, and the lax-aspirated contrast is best differentiated by the *f0* difference on the following vowel (e.g., Choi, 2002; K.-H. Kang & Guion, 2008; Y. Kang, 2014; M. Kim, 2004; M.-R. Kim, 2000, 2008; Kong, Beckman & Edwards, 2011; Lee & Jongman, 2012; Oh, 2011; Silva, 2006). Relative to earlier measures, VOT values of aspirated stops have shortened while those of lax stops have lengthened, resulting in substantial VOT overlap with the post-stop *f0* being low for the lax stops and high for the aspirated stops. Table 1.1 provides comparison of VOT and post-stop *f0* from three previous studies, Kagaya (1974), M.-R. C. Kim (1994), and Oh (2011). Note that the values might not be directly comparable to one another, because of different experimental settings. Data of Kagaya (1974) are from isolated words starting with stops in various vowel contexts, while those of both Kim (1994) and Oh (2011) are from stops before /a/ produced in a carrier sentence. Also, Oh (2011) measured *f0* at the temporal midpoint of the post-stop vowel, unlike Kim (1994) and Kagaya (1974) who measured *f0* at the vowel onset.

			Aspirated		Lax		Tense	
			VOT	*f0*	VOT	*f0*	VOT	*f0*
Kagaya (1974)	male	speaker 1 speaker 2	160	162 150	60	148 162	15	160 192
M.-R. C. Kim (1994)	female	bilabial alveolar velar	71 80 86	294 305 310	78 72 95	224 224 225	8 9 18	286 284 284
	male	bilabial alveolar velar	77 87 88	163 159 163	46 40 67	119 117 114	9 9 21	143 144 145
Oh (2011)	female	bilabial velar	39 55	288	33 50	224	8 16	270
	male	bilabial velar	57 79	162	38 56	127	9 21	151

Table 1.1. Comparison of stop VOT (ms) and post-stop *f0* (Hz) of Seoul Korean stops. When there is only one number in a cell, it represents the mean value for conditions or speakers.

Along with these changes in production, Seoul Korean listeners are found to use the *f0* of the following vowel as a more reliable cue than VOT in perception to signal the contrast between the lax stops and the aspirated stops (M. Kim, 2004; M.-R. Kim, Beddor & Horrocks, 2002; M.-R. C. Kim, 1994; Kong et al., 2011; Lee, Politzer-Ahles & Jongman, 2013). M.-R. C. Kim (1994) reports that in stop identification tasks using synthetic speech, Seoul Korean listeners poorly identified different laryngeal stops based on mere VOT manipulation. M.-R. Kim et al. (2002) test Seoul Korean listeners' perception of the stop with cross-spliced and vowel-only stimuli and report that the vowel portion alone is a necessary and largely sufficient cue for perceiving lax stops whereas perception of the tense and aspirated stops relies on the combination of both consonant and vowel properties.

Taken together, these findings demonstrate that post-stop *f0* has become the primary cue for *stop aspiration*[1] in Seoul Korean both in production and perception. The contrast between aspirated and lax stops is maintained by the *f0* difference instead of the now-neutralized VOT. The loss of the VOT distinction and development of a tonal distinction between aspirated and lax stops hold more true for younger than older speakers (K.-H. Kang & Guion, 2008; Y. Kang, 2014; Silva, 2006), and more for female than male speakers (Kang, 2014; Oh, 2011). In addition, Kong et al. (2011) report that listeners were more sensitive to *f0* for the aspirated-lax distinction when they heard stops produced by female voices than those by male voices.

This development of a tonal contrast and loss of the VOT contrast in Seoul Korean accords with the typical pattern of tonogenesis (e.g., Hombert, Ohala, & Ewan, 1979), in which consonantal contrasts of voicing or aspiration are replaced with tonal contrasts. Due to physiological factors, *f0* at vowel onset is intrinsically correlated with the voicing of the preceding consonants. To explain this physiological relation between stop VOT and post-stop *f0*, two different hypotheses are often entertained, the aerodynamic hypothesis and the vocal fold tension hypothesis. As for effects of stop VOT on the following *f0,* both hypotheses predict that long VOT will lead to high *f0* and voicing to low *f0*, although they differ in how they reach the predictions (see Hombert et al., 1979). Regarding the effects of post-stop *f0* on stop VOT, however, the two hypotheses make different predictions. The aerodynamic hypothesis predicts that high *f0* reduces the amount of airflow across glottis and thus prevents vocal folds from vibrating, resulting in long VOT. The vocal fold tension hypothesis predicts that the high *f0*

[1] Because the commonly used phonological label for Korean /p^h, t^h, k^h/ is *aspirated stops*, I will reserve the term *aspiration* or *aspirated* to refer to the abstract phonological category and use long VOT to refer to the acoustic property, although I am fully aware that "aspirated" usually means "having a long positive VOT".

stiffens the vocal folds, allowing them to return quickly to the adducted position and start vibrating, resulting in short VOT (Narayan & Bowden, 2013; McCrea & Morris, 2005).

Narayan and Bowden (2013) investigate this physiological relation between post-stop *f0* and VOT, examining the effects of increase in *f0* on stop VOT in Seoul Korean and English. Comparing utterances produced in different *f0* ranges, Narayan and Bowden find that, with an increase in *f0,* VOT of Seoul Korean aspirated and lax stops and English voiceless stops decreases. However, VOT of Seoul Korean tense stop and English voiced stop is not influenced by *f0* changes. This finding suggests that VOT and post-stop *f0* in Seoul Korean aspirated stops are indeed physiologically related, but the relation can be better explained by the vocal fold tension hypothesis than the aerodynamic hypothesis.

In tonogenesis, the high or low *f0,* which is redundant for the consonantal contrast, first develops into a robust pitch distinction coexisting with the original consonantal contrast, and then the tonal distinction becomes primarily contrastive (Hombert et al. 1979). Therefore, as the sound change progresses, the difference in vowel *f0* not only increases in its magnitude but also extends beyond the onset of the vowel. Kim (2000) shows that this is indeed the case in Seoul Korean, where *f0* differences following lax onsets vs. aspirated onsets are greater than those following English voiced vs. voiceless stops. On average, *f0* following Korean aspirated stops is 50 Hz higher at vowel onset and is still 30 Hz higher at vowel offset than that after lax stops. On the other hand, *f0* contours following Korean tense vs. aspirated stops are similar to those following English voiced vs. voiceless aspirated stops; in both languages, mean *f0* at vowel onset is 20 Hz higher following aspirated/voiceless stops and the difference falls to 5 Hz or less by vowel midpoint. Kim's (2000) findings suggest that the *f0* difference observed between lax and aspirated stops in Seoul Korean is not an automatic consequence of laryngeal articulation.

Y. Kang (2014) provides further evidence that the *f0* distinction between lax and aspirated stops is being phonologized into a full-fledged tonal contrast. Her data drawn from a read speech corpus show that, for younger female speakers of Seoul Korean, the *f0* difference between aspirated and lax stops has been generalized as a tonal contrast between H and L tones, where the H context includes aspirated stops, tense stops, and /h/-initial conditions and L context includes lax stops and sonorant-initial conditions. Along similar lines, K.-H. Kang and Guion (2008) report that, in clear speech, younger speakers of Seoul Korean who primarily use the *f0* cue to differentiate aspirated and lax stops rely more on *f0* enhancement and produce only small VOT enhancement, whereas older Seoul Korean speakers solely use VOT to enhance the aspirated-lax contrast. However, contrary to Y. Kang's (2014) findings, K.-H. Kang and Guion do not find that young speakers enhance the *f0* contrast between the /h/ and /n/ conditions in clear speech as they do for aspirated and lax stops. Y. Kang suggests that the tonal contrast in sound change and on-line enhancement in clear speech may have different targets such that the sound change is mediated by phonological categories such as natural classes whereas the clear speech enhancement targets segmental contrasts.

In sum, previous findings on Seoul Korean stop productions suggest that post-stop high *f0* has become the primary cue for aspirated stops in Seoul Korean. The current study makes use of this situation, and tests whether primary and non-primary cues operate similarly or differently in spontaneous imitation. Seoul Korean participants are exposed to aspirated stops manipulated to have either higher post-stop *f0* or longer stop VOT, and the possible influences of these enhancements on participants' subsequent productions are assessed. The goal is to investigate not only whether primary and non-primary cues behave differently in imitation but also whether imitation is strictly tied to the manipulated phonetic property (e.g., high *f0* or long VOT) or it is rather phonological in that other typically co-occurring phonetic properties are also enhanced in imitated productions. The design also allows for testing whether *f0* and VOT are positively correlated (as predicted by the aerodynamic hypothesis) or negatively correlated (as predicted by the vocal folds tension hypothesis) in Seoul Korean speakers' imitative enhancements of aspirated stops.

1.2.2 Phonological accounts

Phonological representations of Korean stops have mainly been described using feature systems. Here I summarize two different feature specifications, one from Halle and Stevens (1971) and the other from Cho et al. (2002).

Halle and Stevens (1971) propose a feature-system of four binary features for laryngeal articulation, namely [±spread glottis], [±constricted glottis], [±stiff vocal cords] and [±slack vocal cords], to specify different stop types. Their account of Korean stops is summarized in Table 1.2.

	Aspirated	Lax	Tense
[spread glottis]	+	+	–
[constricted glottis]	–	–	–
[stiff vocal cords]	+	–	+
[slack vocal cords]	–	–	–

Table 1.2. Feature specification of Korean stops in Halle and Stevens (1971)

Cho et al. (2002), on the other hand, propose a privative feature system with underspecification (Table 1.3). This system can better account for their finding that lax stops have intermediate VOT, presumably resulted from intermediate glottal opening. To explain high *f0* after tense and aspirated stops, Cho et al. propose two redundancy rules, (1) [constricted glottis] → [stiff vocal cords], and (2) [spread glottis] → [stiff vocal cords].

	Aspirated	Lax	Tense
Laryngeal features	[spread glottis]	unspecified	[constricted glottis]
Redundancy rules	[stiff vocal cords]	not applicable	[stiff vocal cords]

Table 1.3. Feature specification of Korean stops in Cho et al. (2002)

Relevant to the current study, the two feature systems provide different sets of natural classes. Under the binary system of Halle and Stevens (1971), aspirated and lax stops form a natural class sharing [+spread glottis] whose acoustic correlate is long positive VOT. Under the privative system (Cho et al., 2002), aspirated and lax stops do not form a natural class.

This study investigates the generalizability of spontaneous imitation, by asking how lax and tense stops change as a consequence of hearing model speech with manipulated aspirated stops. If imitation is generalized at the feature level, as suggested in Nielsen (2011), different natural classes predict different patterns of generalization (see §2.5.3 for specific predictions).

1.3 Current Study

This study examines spontaneous speech imitation by separately manipulating two co-varying cues for one phonological contrast differing in their primacy. Seoul Korean speakers hear and shadow model speech in which the phonetic information for stop aspiration is manipulated. The aspirated stops in the model speech have either raised post-stop *f0* or extended

VOT, and the imitative enhancements in participants' own productions are assessed by measuring stop VOT and post-stop *f0*.

By examining how these cues operate in spontaneous imitation, this study aims to reveal whether the cognitive representations that are responsible for speech imitation are detailed phonetic properties, such as long VOT or high post-stop *f0*, or phonological categories, such as stop aspiration. The complicated question of "what is the cognitive unit that is responsible for speech imitation" is addressed by breaking it down into several specific questions. First, which phonetic properties trigger imitative enhancement? Is it only the primarily contrastive property for a phonological category or does the secondary property also trigger imitation? Second, which aspects of articulation are adjusted in imitative enhancements? Is only the property that is enhanced in the model speech imitated? Or are other properties associated with the enhanced phonological category also adjusted? If other properties change together, do they change in the direction predicted by the physiological relation between the properties, if any? Finally, what is the nature of the generalizability of spontaneous imitation? How wide is the scope of imitative generalization? For detailed predictions regarding these questions, see §2.5.

The remainder of this dissertation is organized as follows: In Chapter II, I present the methodology of the imitation experiment and state the predicted outcomes. Chapter III presents the results of the imitation experiment. Chapter IV discusses the current findings, taking into account the principal literature summarized in this chapter, and offers future research directions.

CHAPTER II

Methodology and Predictions

This study investigates the influence of cue primacy on spontaneous speech imitation by speakers of Seoul Korean. As mentioned in §1.2, in Seoul Korean, at least two distinct acoustic properties, stop VOT and post-stop *f0*, differentiate aspirated stops from stops of different phonation types, with post-stop *f0* being the primary cue for the contrast. Previous studies have shown that English speakers imitate extended VOTs of voiceless stops both in immediate shadowing (Fowler et al., 2003; Shockley et al., 2004) and delayed imitation (Nielsen, 2011). In this study, Seoul Korean speakers heard Korean aspirated /t^h/ with either extended VOT or raised post-stop *f0*. The realization of these properties in the speakers' own /t^h/, /t/, and /t*/ in (pre-shadowing) baseline, shadowing and (post-shadowing) test productions are compared. The rest of this chapter describes the details of the methodology (§2.1-§2.4) and predictions (§2.5).

2.1 Participants

Nineteen native speakers of Seoul Korean (12 female and 7 male) participated in the study. Figure 2.1 shows participants' demographic information. The participants are undergraduate or graduate students at the University of Michigan, currently living in Ann Arbor, Michigan. All participants are self-identified as native speakers of Seoul Korean. In a language proficiency and background questionnaire, most of them reported that they were born and raised in Seoul/Gyeonggi in Korea. Three participants were born in the United States but returned to Korea before the age of 5, and lived in Seoul for 15-23 years. Ages of the 19 participants range from 20 to 31 with the mean age of 25.2 years (s.d. = 3.3). All participants are proficient speakers of both Korean and English, but dominant in Korean. In the questionnaire, participants also reported their own proficiency of speaking and reading in two languages on a 1-7 scale. The self-reported proficiency values are as follows: speaking Korean (m = 6.7, s.d. = 0.6), reading Korean (m = 6.6, s.d. = 0.7), speaking English (m = 5.1, s.d. = 1.1), and reading English (m = 5.4, s.d. = 1.0). The mean length of residence in an English-speaking environment is 5.0 years (s.d. = 2.9, range = 2.0 ~ 12.7). No participants reported any history of speech or hearing impairments or extensive contact with any additional languages other than Korean and English. Each participant was paid $25 for completing the two experimental sessions.

2.2 Stimuli

Korean words with initial /t^h/, /t/, or /t^*/ were selected as test words from the NIKL corpus of modern Korean (morphologically parsed, corpus size = 15.3 million eojeols[1]) by the National Institute of Korean Language (2005). In addition, words with initial sonorants were selected as fillers. All test words were disyllabic, highly familiar (word familiarity scores being higher than 6.0 on a 7-point scale), and low in lexical frequency (below 50 in the NIKL corpus). The word familiarity score was obtained from ten native speakers of Korean who are different individuals from the participants of the main study. They were asked to rate the familiarity of the target words presented with fillers on a 7-point scale, and only words that obtained a familiarity score of higher than 6.0 on average were selected.

[1] An eojeol is an orthographic unit for Korean morphological analysis, which is separated by spaces. An eojeol can have one or more morphemes, or even words.

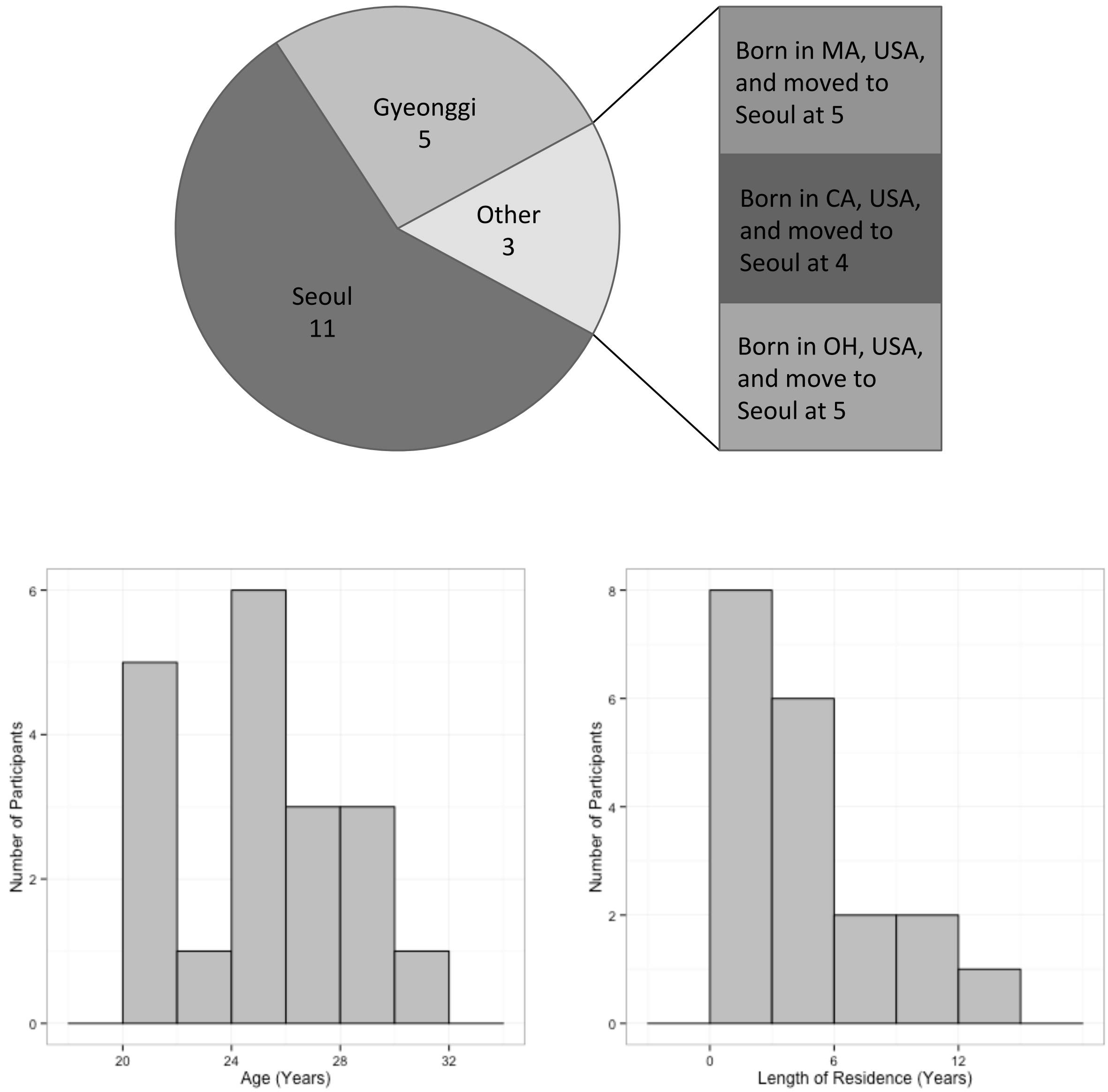


Figure 2.1. Demographic information of participants: birthplace (top), age (bottom left histogram), and length of residence in English-speaking countries (bottom right histogram)

Using the selected words, two wordlists (reading and shadowing lists) were constructed. The reading list contained 150 words: 50 /t^h/-initial words, 25 /t/-initial words, 25 /t^*/-initial words, and 50 sonorant-initial fillers. The shadowing list was a subset of the reading list (50 words), comprising half of the /t^h/-initial words and half of the fillers from the reading list. The remaining words in the reading list occurred only in pre- and post-test and were never presented auditorily. For a complete list of stimuli, see Appendix A.

A male native speaker of Seoul Korean (age = 25) recorded the words from the shadowing list. He produced the words in isolation three times in different randomized orders. He was instructed to speak naturally, at a normal speaking rate. His speech was digitally recorded onto a Macbook Pro laptop computer, using an AKG C 4000 B microphone and an external Edirol UA-25 preamlifier, with a sampling rate of 44.1 kHz via the Praat program (Boersma & Weenink, 2014). From the three repetitions, the best token of each item (free of unintended noises, or mispronunciations) was selected for inclusion. All selected tokens were equalized to have an average intensity of 65 dB using the Scale intensity function in Praat. The mean VOT for the initial /t^h/s was 58.38 ms (s.d. = 8.64) and the mean *f0* at the midpoint of post-stop vowels was 153.6 Hz (s.d. = 4.48).

The model speech for the targeted /t^h/-initial words was manipulated in two ways. The high *f0* stimuli were created by raising the first pitch period of the post-/t^h/ vowel by 20% (calculated in Hz value); *f0* of the rest of the first vowel was also raised proportionately. After manipulation, mean f0 at the midpoint of post-stop vowels was 176.16 Hz (s.d. = 7.05). The long VOT stimuli were created by extending the VOT of word-initial /t^h/ by 60ms. The manipulation to lengthen the VOTs followed the splicing method of Shockley et al. (2004). For each word, the medial portions of the aspiration with low steady amplitude were selected, copied and pasted back into the aspiration section of the waveform. The splicing was done in a way that did not induce any audible discontinuities. After manipulation, the mean VOT for /t^h/-initial targets was 119.82 ms (s.d. = 8.11). Both manipulations were performed in Praat, and *f0* manipulation was done using the PSOLA method (Boersma & Weenink, 2014).

2.3 Procedure

Each participant was tested in two experimental sessions that were conducted at least two weeks apart from each other. Each experimental session involved target stimuli with one of the two manipulations, either raised *f0* or extended VOT. The order of the two experimental sessions was counterbalanced across participants to prevent any potential confounding effect of the testing order. Each session lasted approximately 30 minutes, consisting of an imitation experiment followed by an oddity discrimination test. Participants were also tested in two additional sessions involving English stimuli on different days. The English data were collected for a separate study, and therefore are not analyzed in this dissertation. On the last day of participation, after all other procedures, participants completed a questionnaire on their language background. The entire experiment was conducted in a sound-attenuated booth in the Phonetics Laboratory at the University of Michigan.

2.3.1 Imitation experiment

The imitation experiment, using a slightly modified version of the word-naming imitation paradigm (Babel, 2012; Goldinger, 2000; Nielsen, 2011), consisted of warm-up, baseline production, shadowing, and test production blocks.

Participants were seated in front of a MacBook Pro laptop in a sound-attenuated booth. In the warm-up block, the words from the reading list were visually presented on the laptop screen, and the participants were asked to read them silently without pronouncing them. Each word was presented in the middle of the screen in Korean alphabet *Hangeul* one at a time, every 2 seconds, in a randomized order. In the baseline production block, the words were presented in the same way, but in a different randomized order. This time, the participants were instructed to read the words they saw on the screen aloud as clearly and promptly as possible. In the shadowing block, the words in the shadowing list with either the *f0* or VOT manipulation were played via AKG K271 MK II headphones with nothing presented visually on the screen. The shadowing list was repeated three times, each time in different random orders without any break between repetitions. The participants were instructed to say aloud what they heard as clearly and promptly as possible. They were not instructed to imitate the stimuli. The inter-stimulus interval was 1.5 seconds. Finally, after the shadowing block, the test production block was conducted in the same fashion as the baseline production. Between blocks, participants were allowed to take a short break but no one rested more than a minute. The structure of the imitation experiment is summarized in Table 2.1.

<table>
<tr><td colspan="2">Block</td><td>Warm-up</td><td>Baseline</td><td>Shadowing (x3)</td><td>Test</td></tr>
<tr><td colspan="2">Task</td><td>Read silently</td><td>Read aloud clearly and promptly</td><td>Say aloud the words heard</td><td>Read aloud clearly and promptly</td></tr>
<tr><td colspan="2">Wordlist</td><td>Reading list</td><td rowspan="5">Reading list:

Identical to Warm-up

In a differently randomized order</td><td>Shadowing list</td><td rowspan="5">Reading list:

Identical to Warm-up and Baseline

In a differently randomized order</td></tr>
<tr><td rowspan="2">Target words</td><td>Shadowed</td><td>25 /t^h/-initial words</td><td>25 /t^h/-initial words</td></tr>
<tr><td>Unheard</td><td>25 /t^h/-initial words
25 /t/-initial words
25 /t*/-initial words</td><td></td></tr>
<tr><td rowspan="2">Fillers</td><td>Shadowed</td><td>25 sonorant-initial words</td><td>25 sonorant-initial words</td></tr>
<tr><td>Unheard</td><td>25 sonorant-initial words</td><td></td></tr>
</table>

Table 2.1. Structure of the imitation experiment.

Stimulus presentation was implemented using SuperLab stimulus presentation software (version 4.0.8, Cedrus Corporation). All instructions during the experiment were given in Korean. For the written instructions displayed on the screen, see Appendix B. The participants' baseline, shadowing, and test productions were digitally recorded onto a separate MacBook Pro laptop, using an AKG C 4000 B microphone and an external Edirol UA-25 preamlifier, with a sampling rate of 44.1 kHz via the Praat program (Boersma & Weenink, 2014).

2.3.2 Discrimination test

After completing each imitation experiment, the participants performed an oddity discrimination task that tested perception of the cue manipulation (stop VOT or post-stop *f0*) used in the imitation experiment of that visit. The purpose was to determine whether the difference between the manipulated stimuli and the original recording was reliably perceived.

For each cue, 100 triplets were created from the same manipulated tokens that were used for the shadowing block in the imitation experiment and their original unmanipulated counterparts. Each triplet consisted of two identical tokens and an odd one. Half of the odd tokens were the manipulated ones, and the other half were the original ones. The task for the participants was to identify the odd one. The triplets were concatenated in Praat so every participant heard the same set of triplets. The place of the odd one in each triplet was decided pseudo-randomly, with the odd one appearing a roughly equal number of times in each of the three possible positions of each triplet (VOT condition: first position 33, second position 34, third position 33; *f0* condition: first position 34, second position 33, and third position 33). The interval between tokens within each triplet was 0.5 second.

Participants were seated in front of a Macbook Pro laptop with a response pad (model RB-740, Cedrus Corporation) attached. The stimulus presentation was implemented using SuperLab stimulus presentation software (version 4.0.8, Cedrus Corporation). One hundred triplets were presented over AKG K271 MK II headphones in two experimental blocks with a self-paced break between blocks. Within each block, stimuli were differently randomized for each participant. Participants were asked to choose the odd one from the triplet using the button box. Each new triplet was played one second after the participant hit the button for the previous item. No feedback was provided during the test. The button box responses as well as the response times were collected and analyzed.

2.3.3 Questionnaire

On the last day of participation, after all other procedures, participants completed a background questionnaire. The participants reported their language background, language usage and proficiency. In addition, they were asked to report if there were any words that they did not know during the experiment, and no such case was reported.

2.4 Measurements

For each token recorded during the imitation experiment, the following measures were taken by the author and a phonetically-trained research assistant.

- Voice onset time (VOT): VOT of word-initial stops was measured from the beginning of the release burst (marked as A in Figure 2.2) to the beginning of glottal pulsing in the waveform and/or the appearance of a voicing bar in the spectrogram (marked as B in Figure 2.2).

- F2 onset time (F2OT): F2OT of word-initial stops was measured from the beginning of the release burst (as for VOT) to the beginning of true modal voicing of the following vowel as identified spectrographically by onset of F2 and higher formants (marked as C in Figure 2.2). Thus, unlike VOT, F2OT includes any breathy voiced portion of the following vowel.
- V1 duration: Duration of the first vowel of each word was measured as illustrated from B to D in Figure 2.2. V1 duration was not included in the statistical analyses, but was used to identify vowel midpoint for the *f0* measurement.
- Word duration: Duration of each word was measured as illustrated from A to E in Figure 2.2.
- Post-stop *f0*: Post-stop *f0* was measured at the temporal midpoint of the first vowel of each word using the pitch tracking function in Praat. *f0* measurements for all tokens were checked for tracking errors, and those with *f0* doubling or halving errors were hand-corrected.

Several studies on Seoul Korean stops have used F2 onset time (F2OT) instead of, or in addition to, the traditional measure of VOT (Cho et al., 2002; Cho & Keating, 2001; Choi, 2002; Silva, 1992). F2OT is measured “from the point of the stop release to the voice onset of the second formant in the following vowel” (Cho et al., 2002; Cho & Keating, 2001; Choi, 2002). Silva (1992) uses a derivative measure called “vowel lag” which is defined as the greater of VOT or F2OT. Other studies use the usual measure of VOT, which is measured from the release burst to the first periodic cycle of the following vowel (e.g., Kang, 2014; Kong et al., 2011; Lee & Jongman, 2012; Oh, 2011). In this study, both VOT and F2OT of stops were measured, because, unlike traditional VOT, F2OT includes breathy voiced portion of the vowel with only low-frequency harmonics, which arguably makes it a better measure of the onset of true modal voicing.

Both waveforms and spectrograms were examined in taking the measurements. Prior to making the measurements, the file name of each recording was coded so that neither the author nor the assistant would know which production block (baseline, shadowing, or test) and manipulation type (VOT-extension or *f0*-raising) the recording belongs to.

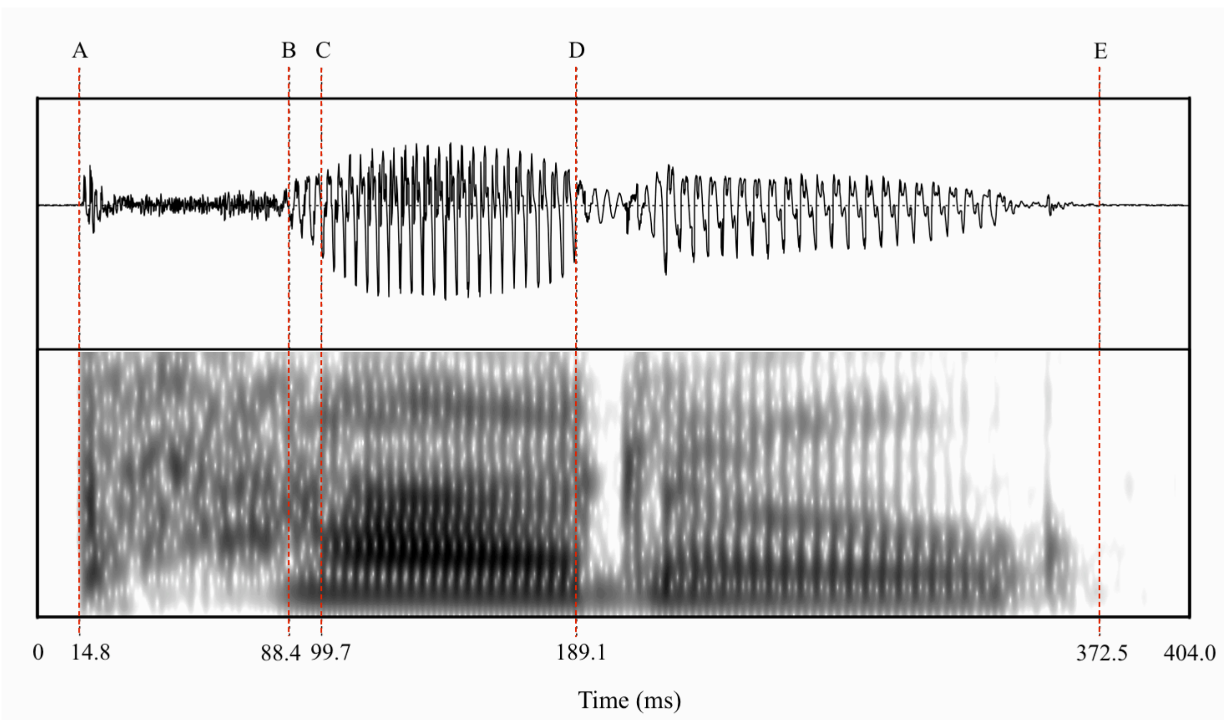


Figure 2.2. Waveform and spectrogram of a sample token 탈옥 [t^harok]. A is the stop burst for the initial stop; B is the onset of voicing; C is the onset of F2 of the first vowel; D is the end of the first vowel; and E is the end of the word.

2.5 Predictions

This study aims to investigate the nature of the cognitive representations that are involved in spontaneous imitation, by examining the effects of two co-varying cues for one phonological contrast that differ in their primacy on imitation. In order to reveal whether the cognitive representations that are responsible for speech imitation are detailed phonetic properties (e.g., long VOT or high post-stop *f0*) or phonological categories (e.g., stop aspiration), the following specific questions are asked in this study: (1) which phonetic properties trigger imitation, (2) which aspects of articulation are adjusted by listener-turned-speakers during imitation, and (3) the breadth of the scope of imitative adjustments. I will examine predictions regarding these questions in the following sections.

2.5.1 What triggers imitation?

Concerning the question of which phonetic properties trigger imitation, two distinct hypotheses can be offered. One possibility is that enhanced phonetic properties trigger speech imitation regardless of their cue primacy. Under this hypothesis, both manipulations used in this study, raised post-stop *f0* and extended stop VOT, will induce some type of imitation. Another possibility is that only the primary cue for a phonological contrast triggers speech imitation. In the latter case, for speakers of Seoul Korean, only the enhanced primary cue, post-stop *f0*, will trigger imitation. That is, aspirated stops with extended VOTs will not facilitate any imitation effects because VOT is a non-primary cue for aspirated stops in Seoul Korean. The existing literature is more consistent with the first hypothesis. As mentioned in §1.1, phonetic properties that do not arguably play a primarily contrastive role, such as English vowel duration (Kim, 2012; Pardo et al., 2012, 2013) and vowel *f0* (Babel & Bulatov, 2012, Pardo et al., 2013) are reported to be spontaneously imitated. Based on these findings, I predict that enhanced phonetic properties, regardless of their cue primacy, will trigger imitation.

In addition, phonologically unnatural manipulations often lead to more robust or longer-lasting imitation (e.g., Honorof et al., 2009; Zellou et al., 2013). Since younger speakers of Seoul Korean depend mainly on *f0* enhancement to enhance aspirated stops (Kang & Guion, 2008) and the model speaker of the current study is a young Seoul Korean speaker (age = 25), participants may perceive extended VOT stimuli to be less natural than raised *f0* stimuli. In that case, extended VOT is predicted to induce more robust imitation than raised *f0*, especially in the post-shadowing test productions.

One important caveat is that, of course, only those properties perceived by the listeners can trigger imitation. Clearly, if listeners do not detect (even sub-consciously) anything special or different about the stimuli, they would not adjust their subsequent productions based on what they have heard. In this study, regardless of the cue primacy, if the manipulations are not large enough to be perceived by the participants as being “different”, no imitation effects are expected. This is the raison d’être of the discrimination test. Although good performance in the oddity discrimination test may not be a sufficient condition for the participants to detect the target manipulation in the imitation experiments, it might arguably be a necessary condition. That is, if the difference between manipulated and original stimuli is not reliably discriminated when they are juxtaposed with each other (as in the discrimination test in this study), the same manipulation is unlikely to be detected as being “different from the typical” when presented alone without any basis for comparison (as in the imitation experiment). To recapitulate, the predictions for the question of which phonetic property triggers imitation are:

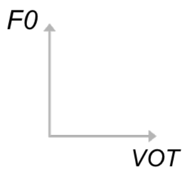
1. Both raised post-stop *f0* and extended VOT are expected to trigger imitation.
2. Extended VOT is expected to induce larger and longer-lasting imitation effects than raised post-stop *f0*.
3. If participants do not perform better than chance in the discrimination test for the specific cue manipulation, no imitation is expected.

2.5.2 Imitative enhancements: what is adjusted?

Assuming that the listener has detected the enhanced phonetic property, the next question is which phonetic property the listener-turned-speaker will adjust, if any, in her subsequent productions. For this question again, two contrasting hypotheses can be offered. The first possibility is that participants adjust the specific property that has been enhanced in the stimuli. That is, when the target model speech has aspirated stops with long VOTs, participants will lengthen their VOTs for aspirated stops. Likewise, when the stimuli are aspirated stops with high post-stop *f0*, participants will raise their *f0* after aspirated stops. Crucially, under this hypothesis, the unmanipulated cue is not expected to change as a consequence of hearing the other cue enhancement. For instance, hearing high post-stop *f0* would not have an enhancing effect on stop VOT in the participants' productions, and neither would long VOT on post-stop *f0*. Imitation, in this case, is strictly tied to a certain phonetic property, and thus I will refer to it as <u>phonetic</u> imitation.

The alternative hypothesis is that imitation is instead <u>phonological</u>, in which case, irrespective of the enhanced cue in the target stimuli, the listeners will enhance the property or properties they would normally use to enhance the relevant phonological category. For example, upon hearing aspirated /t^h/ with high post-stop *f0* or long VOT, participants perceive "exaggerated aspirated stop" and accordingly shift their production in that direction. This shift might involve both properties or, if a single property, might involve a property other than that manipulated in the heard stimuli. This mechanism assumes that imitation is mediated by language-specific associations between phonetic properties and phonological categories.

If imitation is indeed mediated by phonology and if it is triggered by both primary and non-primary cues, which phonetic property is enhanced in the stimuli is not crucial as long as the participant detects the manipulation as enhancing, in this case, aspirated stops. In other words, the imitative patterns will be the same under two experimental conditions (high post-stop *f0* and long VOT) of this study. The common imitative patterns are expected to involve enhancement of the phonetic property(s) the speakers would normally employ to enhance /t^h/. According to Kang and Guion (2008), younger speakers of Seoul Korean, who primarily use post-stop *f0* to distinguish aspirated and lax stops, rely mainly on *f0* enhancement to exaggerate aspirated stops in clear speech. If imitative enhancement were to parallel clear speech, participants in this study, young Seoul Koreans, are predicted to adjust mostly post-stop *f0* in both experimental conditions (see (a) in Figure 2.3). On the other hand, as mentioned in §1.2, tonal development is an ongoing sound change in which younger female speakers exhibit the most advanced stage (e.g., Kang, 2014). If this is the case, although all participants in this study are young, it is possible that at least some speakers, especially male speakers, may still rely on stop VOT at least to some extent. For these speakers, if they exist, the post-stop *f0* is not the exclusive cue for stop aspiration, hence they would be expected to adjust both post-stop *f0* and stop VOT with regard to imitation (as in (b) in Figure 2.3).



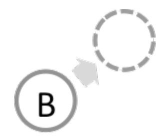
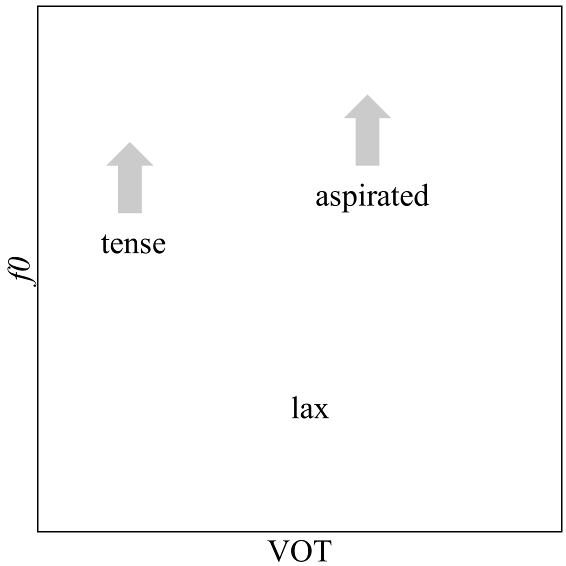
Condition (Manipulations)	Phonetic Imitation	Phonological Imitation
Enhanced primary cue (High post-stop *f0*)		(a) If post-stop *f0* is exclusive 
Enhanced non-primary cue (Long VOT)		(b) If stop VOT also plays a role 

Figure 2.3. Schematic diagrams for phonetic and phonological imitation hypotheses. (B) denotes baseline /t^h/ production on the hypothetical plane of VOT * *f0* shown on the left side of the table. Arrows and dotted circles show the direction of the imitative changes. Note that this table assumes that both primary and non-primary cues trigger imitation. In case a non-primary cue does not trigger imitation, no change is expected in the long VOT condition.

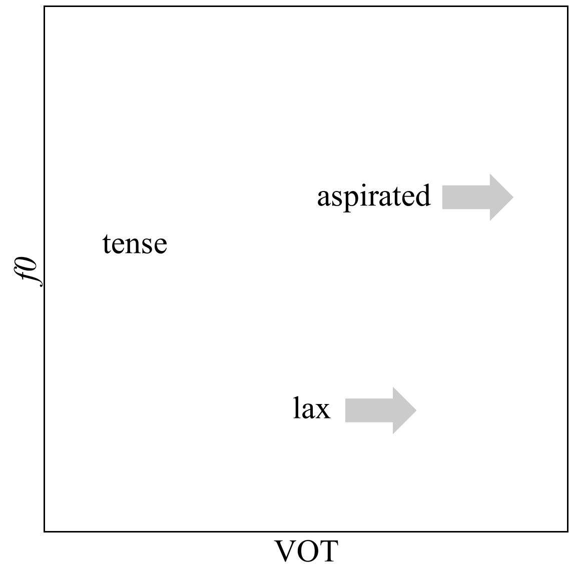
Taken together, the phonological imitation hypothesis predicts that the two manipulations used in this study (high post-stop *f0* and long VOT) will induce identical imitative patterns for a given speaker, and that the imitative patterns will vary across speakers according to their baseline productions. That is, if participants vary in their baseline productions of Korean stops, specifically in VOT differences between lax and aspirated stops, their imitative patterns are expected to vary together. There are two important assumptions that lie behind this hypothesis. First, for a given speaker, perception grammar and production grammar can be different though related. Although a speaker may employ *f0* primarily to exaggerate aspirated stops in her own speech, she can still use extended VOT as a cue for exaggerated aspirated stops when hearing other speaker's speech. Second, speakers in the same speech community can still have different production grammars from each other, in terms of which phonetic property they primarily use for a given phonological contrast, especially with an ongoing (or nearly completed) sound change, as in the case of Seoul Korean. Specific predictions made by the phonetic and phonological hypotheses are illustrated schematically in Figure 2.3.

2.5.3 Generalizability of imitation

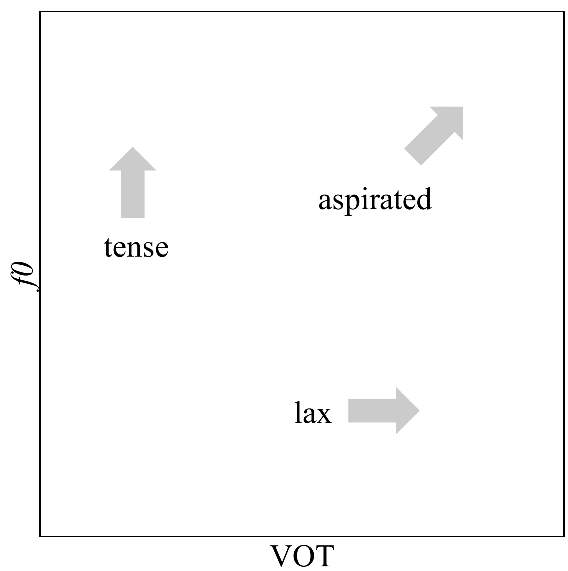
Another important component of the overarching question of what is the cognitive unit that is responsible for speech imitation is the scope of imitative adjustments. Nielsen (2011) found English speakers not only imitated /p/ with long VOT but also generalized the imitative behavior to new unheard /p/-initial words as well as to new /k/-initial words. This suggests that the imitation was generalized to unheard words sharing the initial phoneme (phoneme-level generalization) and those beginning with different phonemes that share the feature [+spread glottis] (feature-level generalization). This study investigates the generalizability of speech imitation more rigorously, by testing how lax and tense stops change as a consequence of hearing the target speech with manipulated aspirated stops.



(a) When *f0* after aspirated stops increases, so does post-tense-stop *f0*.



(b) When VOT of aspirated stops increases, so does lax stop VOT.



(c) When both *f0* and VOT of aspirated stops increase, post-tense-stop *f0* and lax stop VOT also increase.

Figure 2.4. Schematic representation of feature-level generalization.
It is assumed that tense and aspirated stops share [+stiff vocal folds] and lax and aspirated stops share [+spread glottis] (Halle & Stevens, 1971).

To test the extent to which imitative behavior (whether it is to raise post-stop *f0*, lengthen VOT or both) is generalized, the reading list in this study includes three types of words that are not included in the shadowing list: /t^h/-initial, /t/-initial and /t^*/-initial words. First, I predict that unheard /t^h/-initial words will also show the imitative effect, replicating Nielsen's (2011) phoneme-level generalization. Second, as for feature-level generalization, specific predictions depend on the feature system that is adopted. Figure 2.4 presents a set of predictions based on the feature system of Halle and Stevens (1971) (see §1.2.2). Assuming that lax and aspirated stops share [+spread glottis] and that tense and aspirated stops share [+stiff vocal cords], the VOT of aspirated and lax stops should shift together whereas the post-stop *f0* of aspirated and tense stops should shift in tandem. More generally, the important point is that stops that share a feature relevant to the acoustic properties (VOT or *f0*) will change together in the same direction.

Still another level of generalization is feasible. To maximize the relevant contrast, stop aspiration in this study, it is possible that participants will, in post-shadowing test productions, shorten the tense stop VOT or lower the post-lax-stop *f0*. I will refer to this type of *f0* lowering or VOT shortening as <u>phonological readjustment</u>. This phonological readjustment, unlike the phoneme- or feature-level generalization of imitation, is predicted strictly under phonological imitation. If the imitative adjustments are strictly tied to the single acoustic property, without phonology meditating imitation, phonological readjustment is improbable as its motivation is to maximize phonological contrast.

CHAPTER III

Results

This chapter presents the results of the two experiments (imitation and oddity discrimination). The participants performed the imitation task before the discrimination task out of the concern that that the latter task, in which manipulated and original tokens are heard side by side, may have unwanted influence on the results of the imitation experiment. Despite the order of testing, here I report the results of the discrimination test first (§3.1), and then proceed to the imitation experiment (§3.2), because the results of the discrimination test will work as the premise for the imitation experiment (see §2.5.1). After reporting the results pooled across participants, I examine productions of individual participants in §3.3.

All statistical analyses described in this chapter were conducted using R (R Development Core Team, 2014) with packages lme4 (Bates, Maechler, Bolker & Walker, 2014), lmerTest (Kuznetsova, Brockhoff & Christensen, 2014), and irr (Gamer, Lemon, Fellows & Singh, 2012).

3.1 Discrimination test

After completing each imitation experiment for each visit, the participants were tested in an oddity discrimination test that used the same cue manipulation (stop VOT or post-stop *f0*) as the imitation experiment for that session. In order to determine whether the participants could reliably discriminate the differences between the manipulated stimuli and the original ones and whether discriminability was comparable for the two cue manipulations, both accuracy and response time data were examined.

3.1.1 Accuracy analyses

Accuracy score is the number of correct responses out of 100 test triplets for each participant. Table 3.1 summarizes the pooled accuracy scores for the 19 participants.

Condition (Manipulation)	Accuracy Score			
	N	Mean	Std. Dev.	Range
Post-stop *f0*	19	64.68	14.39	48~91
Stop VOT	19	66.68	15.55	41~95

Table 3.1. Summary of descriptive statistics of accuracy score

In order to evaluate whether the accuracy score in each condition was significantly above chance (33.3%), two separate one-sample *t* tests were conducted on participants' accuracy scores in each manipulation condition. The accuracy scores in both conditions were found to be significantly better than chance [*f0* condition: t (18) = 9.60, $p < 0.001$; VOT condition: t (18) = 9.44, $p < 0.001$]. These results suggest that, at least when juxtaposed with unmanipulated stimuli, both the *f0* and VOT manipulations used in the imitation experiments were large enough to be detected by the participants.

To determine whether the participants' performance differed in the two manipulation conditions, a generalized linear mixed effects model with a binary accuracy response ('incorrect', 'correct') and a logit link function was performed using the lme4 package (Bates et al., 2014) for R. Manipulation condition (*f0* vs. VOT) and order of presentation (1~100) were entered into the model as fixed effects. Including interaction between fixed effects did not improve the model, so I will present the model without interaction terms. Random intercepts for speakers and words as well as by-speaker random slopes for manipulation condition were included in the model.

The effects of manipulation condition were not significant [$z = 0.741$, $p = 0.459$], which suggests the participants' performance was not better in one condition than the other. The effects of order of presentation were not significant either [$z = -0.272$, $p = 0.786$], suggesting that participants' performance did not improve (or deteriorate) during the test.

The size of the two manipulations used in this study is not directly comparable since they are on different dimensions, spectral and temporal. Nevertheless, the current result confirmed that one manipulation did not have a more salient impact than the other on the discriminability of the stimuli for Seoul Korean listeners. Furthermore, participants' accuracy scores in the two manipulation conditions were found to be correlated with each other. The scatter plot in Figure 3.1 represents the accuracy scores of the 19 participants. A Pearson's product-moment correlation coefficient was computed to assess the relationship between the accuracy scores in the *f0* and VOT manipulation conditions. A moderate, positive correlation was found between participants' performance in the two conditions [$r = 0.51$, $p = 0.027$]. These results suggest that participants who are more sensitive to one manipulation are, in general, more sensitive to the other manipulation.

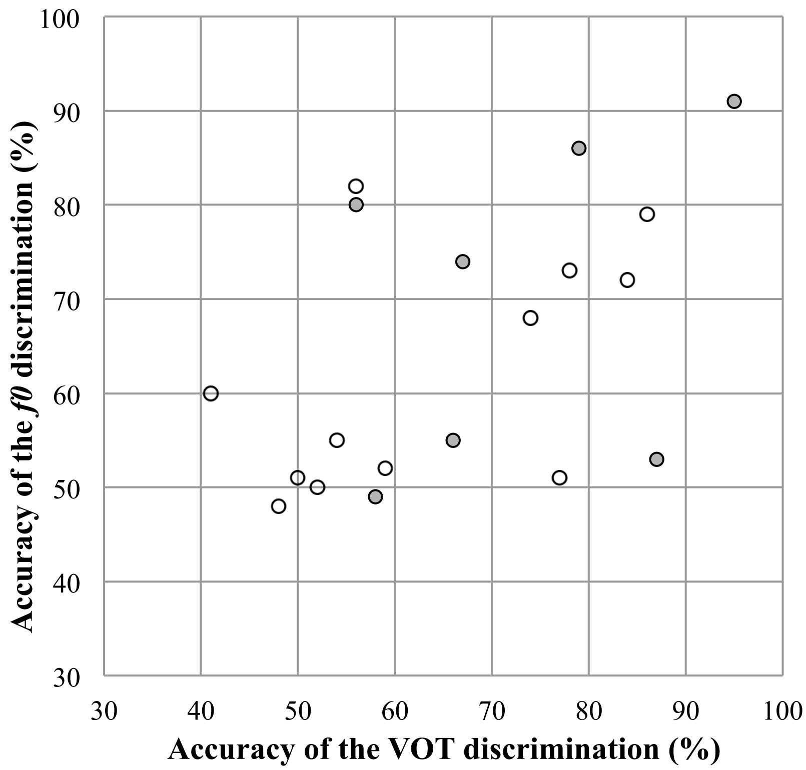


Figure 3.1. Performance of individual participants on two discrimination tests. Each data point represents one participant, with the shaded dots indicating male speakers. Accuracy scores for the VOT manipulation condition are plotted on the x-axis, and those for the *f0* manipulation condition are plotted on the y-axis.

3.1.2 Response time analyses

Table 3.2 summarizes the descriptive statistics for the response time data from the discrimination tests. Response time is measured as the duration between the end of the last token of the triplet and the button press. Excluded from analyses were response times from error trials as well as those greater than two standard deviations away from the mean for a particular participant and condition.

Condition (Manipulation)	Response Time (ms)				
	N	Mean	Median	Std. Dev.	Std. Error
Post-stop *f0*	1159	550.03	444	420.40	12.35
Stop VOT	1203	484.90	394	378.81	10.92

Table 3.2. Summary of descriptive statistics for response times in discrimination tests

In order to reconfirm that one manipulation was not easier or harder to discriminate than the other, a linear mixed effects model was fitted to the response time data using the lme4 package (Bates et al., 2014) for R. Manipulation condition (*f0* vs. VOT) and order of presentation (1~100) were entered into the model as fixed effects (without interaction terms). Including interaction between fixed effects did not improve the model, so the model without interaction terms is presented here. Random intercepts for speakers and words as well as by-speaker random slopes for manipulation condition were included in the model. Parameter specific *p*-values were obtained by using the Satterthwaite approximation, which was implemented in the lmerTest package (Kuznetsova et al., 2014) for R.

The effects of manipulation condition were not significant [$t = -1.268$, $p = 0.221$], confirming that it did not take the participants significantly longer to respond to one set of manipulated stimuli than to the other. The effects of order of presentation were found to be significant [$\beta = -1.462$, $t = -6.13$, $p < 0.001$], with response times decreasing as the test proceeded (albeit with no change in accuracy, as reported in §3.1.1).

3.1.3 Summary of discrimination results

Taken together, the results of the two discrimination tests provide evidence that the acoustic manipulations of *f0* and VOT were equally discriminable (relative to original productions) for Seoul Korean listeners. The magnitudes of the two manipulations are not directly comparable because they are on different acoustic dimensions, namely, spectral and temporal. Nonetheless, the accuracy and response latency data suggest that the two sets of manipulated stimuli were similarly perceptible for these participants.

More importantly, both manipulations resulted in perceptible differences in the stimuli, as evidenced by the results of the one-sample *t* tests on the participants' accuracy scores: participants were performing significantly better than chance in both conditions. Certainly, good performance in these oddity discrimination tests in which the manipulated stimuli are juxtaposed with the original ones does not guarantee that the participants would notice high post-stop *f0* or long stop VOT in the imitation experiments. However, if the difference had not been detected in discrimination testing, it would have been unlikely to be noticed in the imitation experiments as well. Therefore, the results of these discrimination tests serve as a prerequisite for the imitation experiments.

3.2 Imitation experiments

3.2.1 Statistical procedures

Prior to statistical analyses, tokens with a disfluency were excluded. Disfluency was defined as when participants did not utter a word that they read or heard, said a different word, repeated a part of a word (including self-correction), or had some extra-verbal interruption such as coughing or clearing the throat. 0.8% of the total productions were excluded from further analyses due to disfluency. For *f0* analyses, an additional 17.1% of the total productions were excluded because the first vowels of the tokens were creaky voiced or completely voiceless and the pitch-tracking function in Praat provided no *f0* value. Creaky vowels occurred most commonly after tense /t*/. Complete devoicing occurred frequently for /ɨ/ following /tʰ/ (e.g., 특가 /tʰɨkka/, 특진 /tʰɨktɕin/).

A subset (10.8%) of the remaining data was randomly chosen and analyzed to determine inter-rater consistency. The consistency score was computed using the Intraclass Correlation Coefficient (ICC) in the irr package (Gamer et al., 2012) for R. All duration measurements were highly consistent between two raters (VOT ICC = 0.997; F2OT ICC = 0.972; V1 duration ICC = 0.989; and word duration ICC = 0.992; all p values < 0.001).

The results were analyzed in linear mixed effects regression models using the lme4 package (Bates et al., 2014) for R. Changes in aspirated stops in the different production blocks (baseline, shadowing 1, 2, 3, and test), and changes in all stop types (aspirated /tʰ/, lax /t/, and tense /t*/) in baseline and test productions were separately analyzed (first column in Table 3.3). Also separately analyzed were different manipulation types (High post-stop *f0* vs. Long stop VOT, second column in Table 3.3). For each analysis, separate linear mixed effects models were built for each of the three dependent variables (VOT, F2OT, and Post-stop *f0*, third column in Table 3.3). A total of 12 separate linear mixed effects models were used to analyze different aspects of the data collected. Parameter specific *p*-values were obtained by using the Satterthwaite approximation, which was implemented in the lmerTest package (Kuznetsova et al., 2014) for R.

Target data under analyses	Manipulation Type	Dependent Variable	Relevant Section
Aspirated /t^h/s in Baseline - Shadowing 1, 2, 3 - Test	High post-stop *f0*	Stop VOT	§3.2.2.1
		Stop F2OT	
		Post-stop *f0*	
	Long stop VOT	Stop VOT	§3.2.3.1
		Stop F2OT	
		Post-stop *f0*	
Aspirated /t^h/s, lax /t/s, and tense /t*/s in Baseline - Test	High post-stop *f0*	Stop VOT	§3.2.2.2
		Stop F2OT	
		Post-stop *f0*	
	Long stop VOT	Stop VOT	§3.2.3.2
		Stop F2OT	
		Post-stop *f0*	

Table 3.3. Summary of statistical modeling and relevant sections

3.2.2 High f0 condition: primary cue in imitation

This section presents the effects of the enhanced primary cue for stop aspiration—high post-stop *f0*—on phonetic imitation of aspirated stops. Tables 3.4-3.6 summarize the descriptive statistics in VOT, F2OT, and *f0* measurements, respectively, in the high *f0* condition.

Consistent with previous findings on Korean stop productions (e.g., Kang, 2014; Oh, 2011), the VOT/F2OT difference between aspirated stops and lax stops was larger for male participants than it was for female participants. Nonetheless, both male and female participants produced higher post-stop *f0* for aspirated stops than for lax stops. Post-lax-stop *f0* was comparable to post-sonorant *f0*, as seen in the filler words in Table 3.6.

Word Type	Production Block	Stop VOT (ms)			
		Males		Females	
		Mean	Std. Dev.	Mean	Std. Dev.
Shadowed /t^h/	Baseline	67.33	17.70	64.55	19.32
	Shadowing 1	60.40	17.45	59.39	17.49
	Shadowing 2	61.34	16.96	59.34	19.63
	Shadowing 3	64.22	16.91	58.52	19.67
	Test	68.84	20.89	63.79	18.92
Unheard /t^h/	Baseline	66.31	18.22	66.53	19.66
	Test	69.02	18.96	64.13	21.00
Unheard /t/	Baseline	52.94	18.68	58.59	19.91
	Test	55.46	18.80	59.54	19.37
Unheard /t*/	Baseline	15.28	5.61	12.05	3.85
	Test	14.75	6.00	12.04	4.05

Table 3.4. Summary of descriptive statistics for VOT (msec) in the high *f0* condition

Word Type	Production Block	Stop F2OT (ms)			
		Males		Females	
		Mean	Std. Dev.	Mean	Std. Dev.
Shadowed /t^h/	Baseline	76.33	17.73	72.32	19.79
	Shadowing 1	68.56	17.68	67.44	17.95
	Shadowing 2	70.92	17.10	67.83	19.65
	Shadowing 3	72.95	16.55	66.95	19.81
	Test	78.27	21.81	72.53	18.92
Unheard /t^h/	Baseline	75.81	18.43	73.93	19.70
	Test	78.74	19.64	73.20	20.80
Unheard /t/	Baseline	61.70	18.78	67.79	19.92
	Test	64.66	19.32	69.66	20.75
Unheard /t*/	Baseline	20.12	7.07	15.34	5.28
	Test	19.92	7.82	15.60	5.48

Table 3.5. Summary of descriptive statistics for F2OT (msec) in the high *f0* condition

Word Type	Production Block	Post-stop *f0* (Hz)			
		Males		Females	
		Mean	Std. Dev.	Mean	Std. Dev.
Shadowed /tʰ/	Baseline	129.43	12.65	271.74	25.50
	Shadowing 1	132.59	17.07	271.41	21.69
	Shadowing 2	133.85	19.26	273.09	26.09
	Shadowing 3	134.17	17.79	274.49	24.74
	Test	132.82	18.05	277.87	25.61
Unheard /tʰ/	Baseline	129.75	12.65	271.04	23.52
	Test	131.97	17.49	277.69	26.82
Unheard /t/	Baseline	107.99	11.33	203.12	14.51
	Test	108.36	10.78	205.86	16.32
Unheard /t*/	Baseline	122.39	11.48	254.37	21.50
	Test	125.53	16.71	258.18	27.27
Shadowed fillers	Baseline	104.82	8.61	199.47	13.27
	Shadowing 1	106.01	9.22	201.39	16.33
	Shadowing 2	106.12	9.52	202.46	16.89
	Shadowing 3	106.81	9.48	202.67	17.41
	Test	105.59	9.46	200.41	15.28
Unheard fillers	Baseline	105.22	8.99	199.58	13.90
	Test	105.97	10.15	200.98	16.40

Table 3.6. Summary of descriptive statistics for post-stop *f0* (Hz) in the high *f0* condition

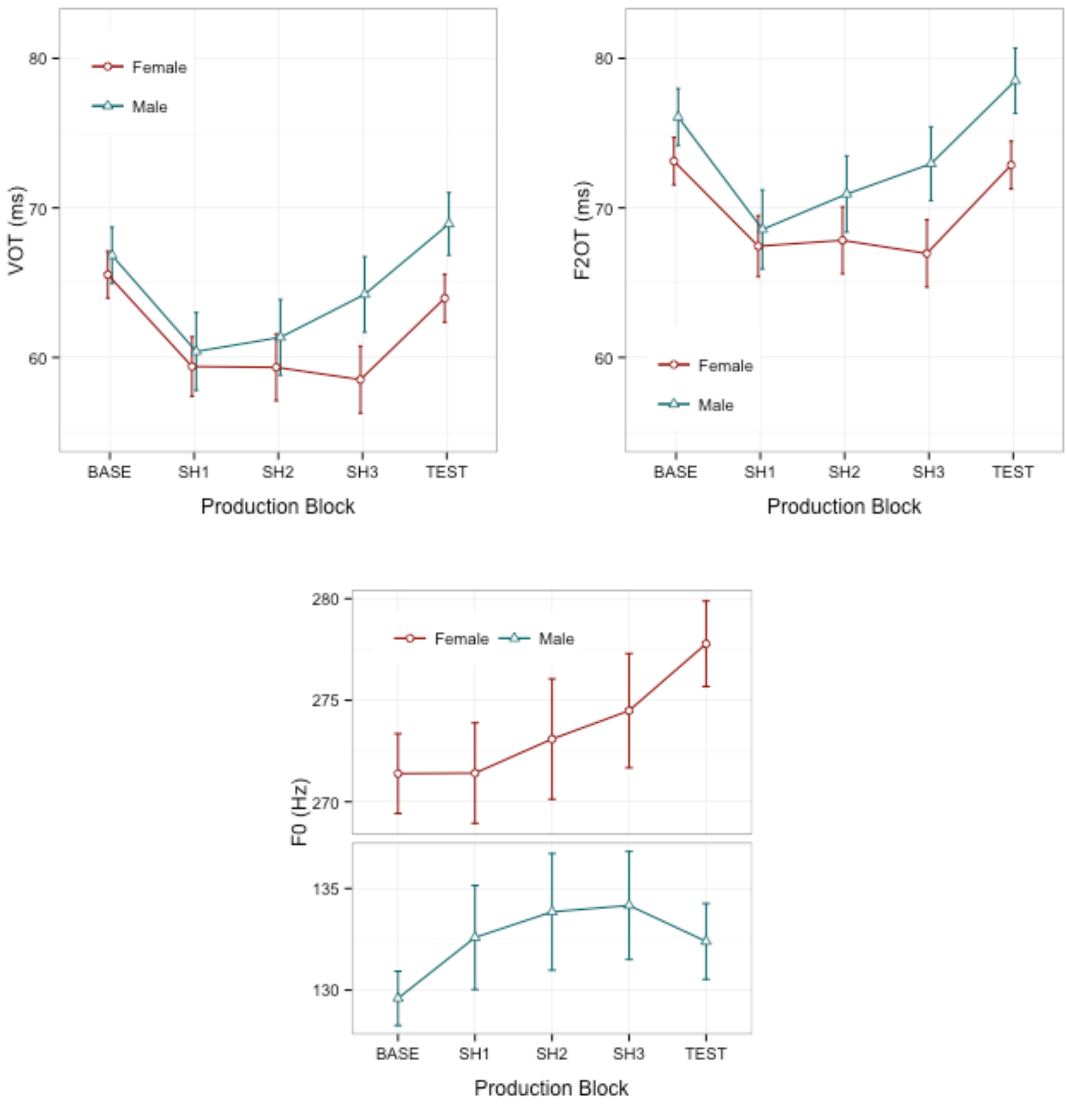


Figure 3.2. Changes in aspirated $/t^h/$ in the high *f0* condition: stop VOT (top left), stop F2OT (top right), and post-stop *f0* (bottom). Error bars represent 95% confidence intervals.

3.2.2.1 Effects of high *f0* /t^h/ stimuli on productions of /t^h/ in shadowing and test blocks

Figure 3.2 presents mean VOT, F2OT and post-stop *f0* of aspirated /t^h/s across speakers in the different production blocks (baseline, shadowing 1, 2, 3, and test) in the high *f0* condition.

To statistically analyze the changes in aspirated /t^h/ productions in the five production blocks, a linear mixed effects model was fitted to each of the dependent variables, stop VOT, F2OT, and post-stop *f0*. For all three models, production block (baseline, shadowing 1, 2, 3, and test), presence of exposure (shadowed vs. unheard), and speaker gender were entered into the models as fixed effects (without interaction terms). For the VOT/F2OT models, the remaining word duration (Remainder = total word duration – VOT/F2OT) was included in the models to verify that the changes in VOT/F2OT were not due to global changes in speech rate. Similarly, to make sure that the observed changes in post-stop *f0* were not due to global pitch shift, filler words were also included in the mixed effects model for *f0* by including the interaction between word type (aspirated /t^h/-words vs. fillers) and production block.

As random effects, intercepts for speakers and words were included in all three models. In the VOT and F2OT models, by-speaker random slopes for the effect of production block were also included. In the *f0* model, by-speaker random slopes for both word type and production block with interaction were included. Including by-word random slopes for the effect of production block did not improve the models, so it was not included in the models presented here. In the *f0* model, word random effects were nested in word type. Parameter specific *p*-values were obtained by using the Satterthwaite approximation (Kuznetsova et al., 2014).

Dependent variable (Outcome)	Fixed effect (Predictor)	Estimate (β)	t	p
Stop VOT	**(Intercept)**	**73.657**	**17.997**	**< 0.001**
	Production block:			
	Base-Shadowing 1	**-6.446**	**-5.276**	**< 0.001**
	Base-Shadowing 2	**-6.079**	**-4.658**	**< 0.001**
	Base-Shadowing 3	**-5.608**	**-3.619**	**0.002**
	Base-Test	-0.360	-0.347	0.739
	Exposure (unheard-shadowed)	-1.639	-0.522	0.610
	Gender (female-male)	2.911	0.769	0.457
	Word duration (Rest)	**-0.018**	**-3.630**	**< 0.001**
Stop F2OT	**(Intercept)**	**80.399**	**19.482**	**< 0.001**
	Production block:			
	Base-Shadowing 1	**-6.424**	**-4.949**	**< 0.001**
	Base-Shadowing 2	**-5.273**	**-4.142**	**< 0.001**
	Base-Shadowing 3	**-5.136**	**-3.249**	**0.004**
	Base-Test	0.598	0.593	0.561
	Exposure (unheard-shadowed)	-1.626	-0.487	0.628
	Gender (female-male)	3.602	0.982	0.339
	Word duration (Rest)	**-0.015**	**-3.110**	**0.002**
Post-stop *f0*	**(Intercept)**	**255.013**	**39.706**	**< 0.001**
	*Word type * Production block:*			
	Aspirated-Filler, Base	**-54.430**	**-8.394**	**< 0.001**
	Aspirated, Base-Shadowing 1	0.946	0.384	0.705
	Aspirated, Base-Shadowing 2	2.473	0.981	0.339
	Aspirated, Base-Shadowing 3	3.460	1.515	0.147
	Aspirated, Base-Test	**5.104**	**3.322**	**< 0.001**
	Filler, Base-Shadowing 1	0.760	0.326	0.748
	Filler, Base-Shadowing 2	-0.051	-0.024	0.981
	Filler, Base-Shadowing 3	-0.653	-0.320	0.753
	Filler, Base-Test	**-4.084**	**-2.662**	**0.008**
	Exposure (unheard-shadowed)	0.014	0.013	0.989
	Gender (female-male)	**-97.395**	**-40.484**	**< 0.001**

Table 3.7. Summary of linear mixed models for aspirated stops in different production blocks in the high *f0* condition. **Bolded** data are significant at $p < 0.05$.

Table 3.7 reports the results of the three mixed effects models, including the parameter estimate *β*, *t* value, and *p* value for the fixed effects included in the models. The pairwise comparisons for production block and word type are presented when the baseline production is compared against other production blocks or when baseline productions of different word types are compared against each other.

The results of the two duration models—VOT and F2OT—were almost identical to each other. The effects of production block were robust in both VOT and F2OT models, revealing significant shortening of VOT/F2OT during each block of shadowing relative to baseline. The VOT/F2OT difference between baseline and test productions was not significant.

In both duration models, word duration was a significant predictor of VOT/F2OT. Although this effect seems to be highly significant statistically [$p < 0.005$], *β* values were extremely small [$\beta = -0.018$ (VOT) and -0.015 (F2OT)], which means this effect is likely to be negligible in reality. Furthermore, the direction of the effect indicates that the changes in VOT/F2OT and those in the rest of word duration are negatively correlated, if at all. This means that lengthening (or shortening) in VOT/F2OT is not due to overall slowing down (or speeding up) of speech rate, and that the entire word duration was constant despite changes in VOT/F2OT. The relatively small changes in overall word duration can be explained as follows: Because stop VOT or F2OT is voiceless or breathy portion of the following vowel, VOT/F2OT is expected to be in inverse relation to the duration of the (modal) voiced portion of the vowel.

The results of the *f0* model showed significant effects of production block but only the baseline-test pairwise comparisons were found to be significant. None of the three blocks of shadowing productions was different from baseline [$| t | < 2$, $p > 0.1$]. Individual speakers' productions were investigated by examining the coefficients of the model by speaker, which revealed that there are three outlier speakers. These outlier speakers were all females, and they apparently imitated the pitch of the male model speaker by lowering their pitch during the shadowing blocks. (See Appendix C for the individual speakers' production patterns. The outlier speakers are F04, F12, and F20.) Excluding these three speakers, the same linear mixed effects model was fitted again, whose results are presented in Table 3.8.

Fixed effect (Predictor)	Estimate (β)	t	p
(Intercept)	**253.614**	**36.483**	**< 0.001**
*Word type * Production block:*			
Aspirated stop-Filler, Base	**-49.172**	**-7.036**	**< 0.001**
Aspirated stop, Base-Shadowing 1	3.883	1.853	0.083
Aspirated stop, Base-Shadowing 2	**6.025**	**3.381**	**0.004**
Aspirated stop, Base-Shadowing 3	**6.455**	**3.855**	**0.001**
Aspirated stop, Base-Test	**6.162**	**3.647**	**0.002**
Filler, Base-Shadowing 1	-0.853	-0.385	0.706
Filler, Base-Shadowing 2	-2.380	-1.423	0.174
Filler, Base-Shadowing 3	-2.545	-1.476	0.159
Filler, Base-Test	**-4.543**	**-3.047**	**0.008**
Exposure (unheard-shadowed)	0.028	0.028	0.977
Gender (female-male)	**-102.329**	**-50.046**	**< 0.001**

Table 3.8. Summary of the new linear mixed model for *f0* analysis (excluding outliers). **Bolded** data are significant at $p < 0.05$.

Without the three outlier participants, both shadowing 2 and 3 were significantly different from the baseline [$| t | > 3$, $p < 0.005$] while the post-stop *f0* difference between shadowing 1 and baseline did not reach significance [$t = 1.853$, $p = 0.083$]. A separate linear mixed effects model was fitted to the outlier speakers only, and it revealed that their *f0* in shadowing production was indeed lower than their baseline especially in the earlier blocks of shadowing [Shadowing 1: $\beta = -14.677$, $t = -3.404$, $p = 0.039$; Shadowing 2: $\beta = -16.425$, $t = -4.053$, $p = 0.023$; Shadowing 3: β

= -12.478, t = -2.548, p = 0.080]. These outlier speakers' post-stop *f0* in the test production was not different from their baseline [β = -0.469, t = -0.182, p = 0.867].

Other significant fixed effects in the *f0* model were the effect of speaker gender and word type (aspirated stop words vs. sonorant-initial fillers): female speakers had higher *f0* than male speakers and post-aspirated-stop *f0* was higher than post-sonorant *f0*, as expected.

The effects of presence of exposure (shadowed vs. unheard) were not significant in any of the three models [$| t | < 1$, $p > 0.1$]. Figure 3.3 presents the comparison between shadowed /t^h/-initial words and unheard ones in baseline and test productions. Words beginning with initial aspirated stops showed the same imitation effects regardless of whether the specific word was present or not during the shadowing blocks. This is consistent with previous findings (e.g., Nielsen, 2011) that phonetic imitation is not limited to specific words, but generalizes to novel words sharing the same phoneme.

Taken together, the results from the *f0* model and the durational models show a decrease in the duration of stop VOT/F2OT with increasing post-stop *f0* during shadowing blocks. This reverse relation between stop VOT and post-stop *f0* supports the vocal cord tension hypothesis (McCrea & Morris, 2005; Narayan & Bowden, 2013). Aspirated voiceless stops produced with a particularly high *f0* have shorter VOTs than productions made in a lower pitch range, presumably because high pitch can involve relatively stiff vocal cords, reducing the time for the vocal cords to adduct and start voicing. (This point is discussed further in Section §4.1.1.)

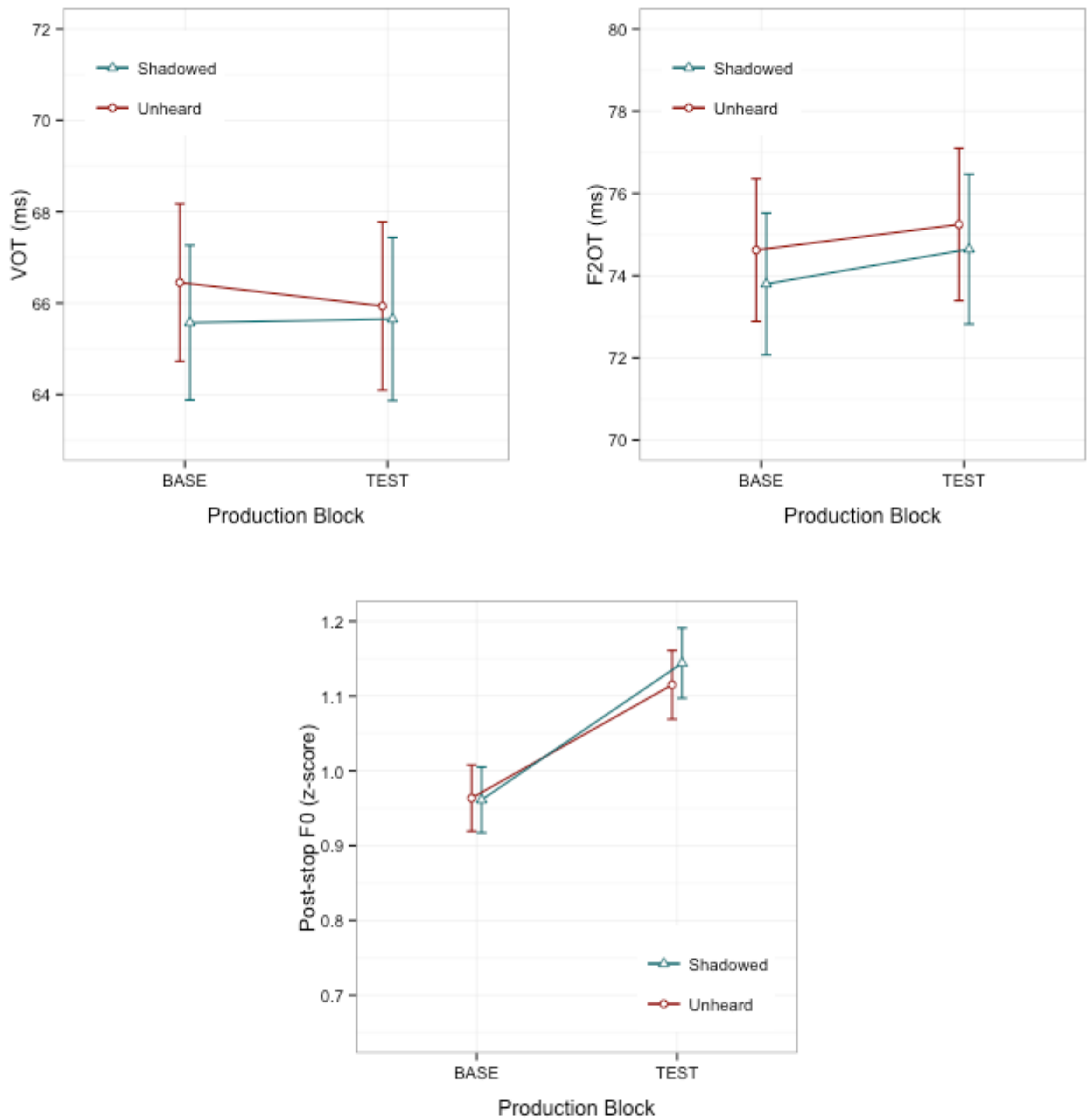


Figure 3.3. Shadowed and unheard $/t^h/$-initial words in the high *f0* condition: stop VOT (top left), stop F2OT (top right), and post-stop *f0* (bottom). To plot speakers of both genders together, post-stop *f0* is given in z-score. Error bars represent 95% confidence intervals.

3.2.2.2 Effects of high *f0* /t^h/ stimuli on productions of /t/ and /t*/

Another question that this study aims to answer is whether phonation types other than aspirated stops change as a result of hearing and shadowing the target stimuli with raised post-stop *f0.* Figures 3.4-3.6 presents mean VOT, F2OT and post-stop *f0*, respectively, of the three stop types in baseline and test productions of the high *f0* condition.

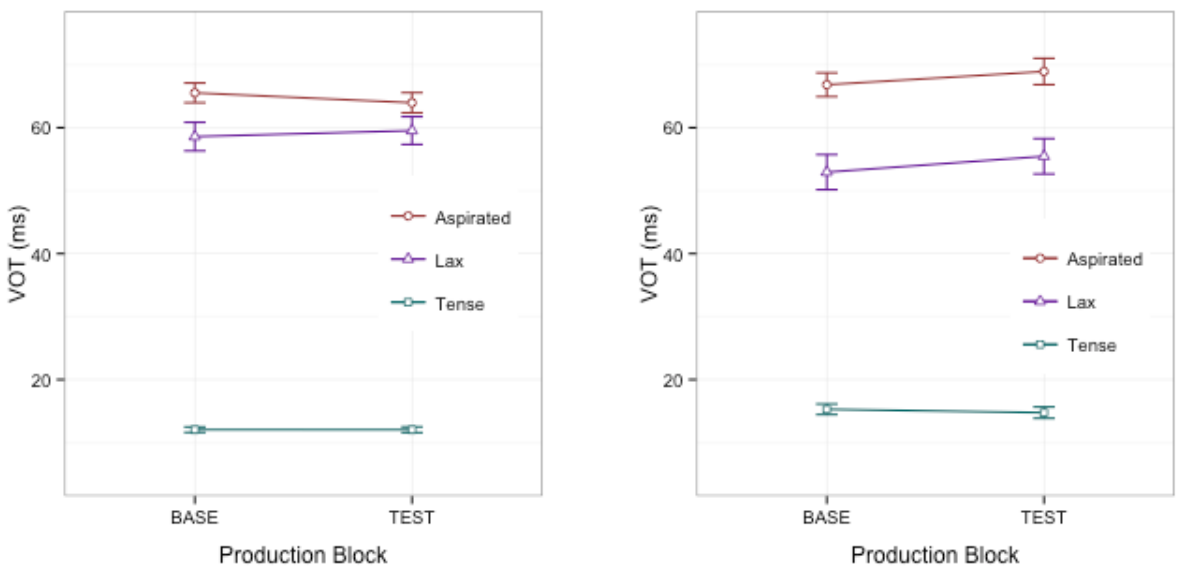


Figure 3.4. VOT changes in the high *f0* condition for different stops (aspirated /t^h/, lax /t/, and tense /t^*/) produced by female (left) and male (right) speakers. Error bars represent 95% confidence intervals.

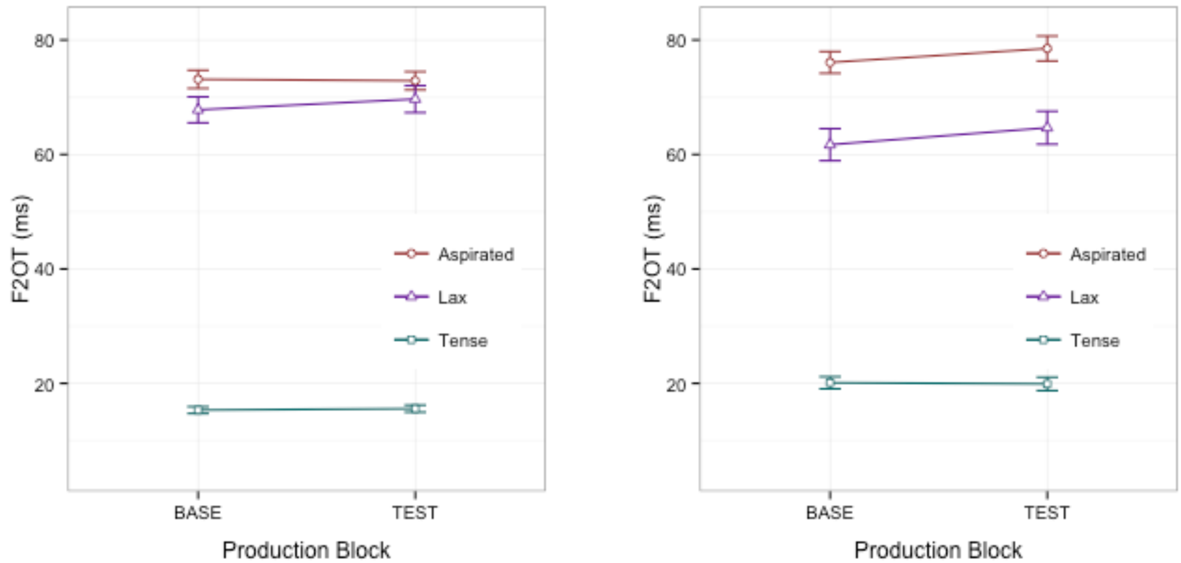


Figure 3.5. F2OT changes in the high *f0* condition for different stops (aspirated /t^h/, lax /t/, and tense /t*/) produced by female (left) and male (right) speakers. Error bars represent 95% confidence intervals.

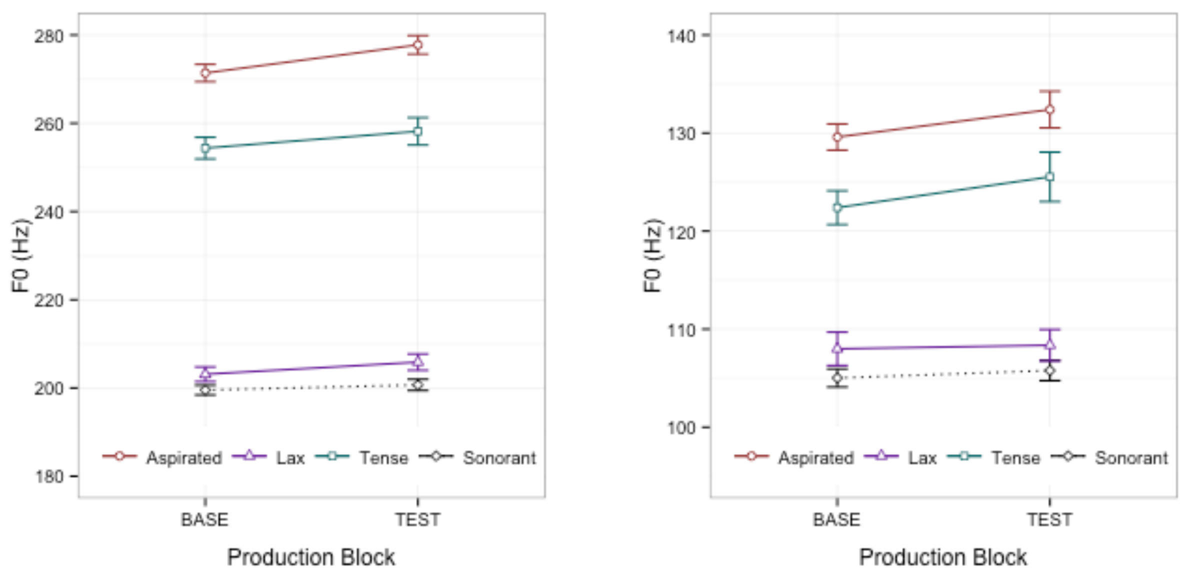


Figure 3.6. Post-onset *f0* changes in the high *f0* condition for different consonants (aspirated /t^h/, lax /t/, tense /t*/, and sonorant) produced by female (left) and male (right) speakers. Error bars represent 95% confidence intervals.

To statistically analyze the changes in production of all stop types (aspirated, lax, and tense stops) between baseline and test productions, three additional linear mixed effects models were fitted to three dependent variables (stop VOT, F2OT, and post-stop *f0*). For the two duration (VOT/F2OT) models, stop type (aspirated, lax, and tense), production block (baseline, test), and their interaction were entered into the models as fixed effects. For the *f0* model, word type (aspirated, lax, tense, and fillers) was entered instead of stop type in order to include filler words in the analysis. This was done to make sure that the observed changes in post-stop *f0* were not due to global pitch shift. For similar reasons, the remaining word duration (Remainder = total word duration – VOT/F2OT) was entered in the duration models as a fixed effect. In addition, all three models included presence of exposure (shadowed vs. unheard) and speaker gender as fixed effects (without interaction terms).

As random effects, intercepts for speakers and words as well as by-speaker random slopes for both stop/word type and production block with interaction were included in all three models. Including by-word random slopes did not provide better fits, so they were not included in the final models. In all three models, word random effects were nested in stop/word type. Parameter specific *p*-values were obtained by using the Satterthwaite approximation (Kuznetsova et al., 2014).

Table 3.9 summarizes the parameter estimate β, t value, and p value for the fixed effects included in the three fitted models. The pairwise comparisons for production block and word type are presented when baseline production of aspirated stop is compared against other word types or when the baseline production is compared against the test production.

Dependent variable (Outcome)	Fixed effect (Predictor)	Estimate (β)	t	p
Stop VOT	**(Intercept)**	**73.943**	**21.665**	**< 0.001**
	*Word type * Production block:*			
	Aspirated-Lax, Base	**-10.035**	**-3.113**	**0.002**
	Aspirated-Tense, Base	**-52.629**	**-15.010**	**< 0.001**
	Aspirated, Base-Test	-0.348	-0.324	0.750
	Lax, Base-Test	1.736	1.693	0.100
	Tense, Base-Test	-0.056	-0.044	0.965
	Exposure (unheard-shadowed)	-1.566	-0.557	0.579
	Gender (female-male)	0.496	0.471	0.642
	Word duration (Rest)	**-0.016**	**-4.490**	**< 0.001**
Stop F2OT	**(Intercept)**	**81.938**	**23.623**	**< 0.001**
	*Word type * Production block:*			
	Aspirated-Lax, Base	**-9.277**	**-2.720**	**0.008**
	Aspirated-Tense, Base	**-56.957**	**-16.218**	**< 0.001**
	Aspirated, Base-Test	0.586	0.575	0.573
	Lax, Base-Test	1.540	1.515	0.137
	Tense, Base-Test	-0.708	-0.553	0.586
	Exposure (unheard-shadowed)	-1.693	-0.572	0.568
	Gender (female-male)	0.886	0.617	0.545
	Word duration (Rest)	**-0.016**	**-4.268**	**< 0.001**
Post-stop *f0*	**(Intercept)**	**258.488**	**44.350**	**< 0.001**
	*Word type * Production block:*			
	Aspirated-Lax, Base	**-51.065**	**-7.617**	**< 0.001**
	Aspirated-Tense, Base	**-13.335**	**-5.953**	**< 0.001**
	Aspirated-Filler, Base	**-54.428**	**-8.152**	**< 0.001**
	Aspirated, Base-Test	**5.107**	**3.250**	**0.004**
	Lax, Base-Test	**-3.235**	**-2.535**	**0.020**
	Tense, Base-Test	-1.610	-1.180	0.253
	Filler, Base-Test	**-4.087**	**-2.606**	**0.018**
	Exposure (unheard-shadowed)	0.013	0.011	0.991
	Gender (female-male)	**-106.830**	**-44.041**	**< 0.001**

Table 3.9. Summary of linear mixed models for three stop types in baseline and test productions in the high *f0* condition. **Bolded** data are significant at $p < 0.05$.

First, the comparisons among different stop/word types within baseline production confirmed that the primary difference between aspirated stops and lax stops is maintained not by stop VOT/F2OT but by post-stop *f0*: (1) *f0* was significantly higher after aspirated stops than after lax stops with a mean difference of around 50 Hz [$| t | > 7$, $p < 0.001$], and (2) VOT/F2OTs of aspirated stops were significantly longer that those of lax stops [$| t | > 2.5$, $p < 0.01$], but the difference of roughly 10 ms seems quite small considering the large variation in VOT/F2OT within each stop category (see the descriptive statistics in Table 3.4-3.6). Although all three Korean stop categories under investigation are voiceless word-initially, *f0* following a lax stop seems comparable to post-sonorant *f0*. This pattern of results supports the interpretation that the post-lax-stop low *f0* in Seoul Korean is a phonological tonal contrast (e.g., Kang, 2014).

Turning to the comparisons between baseline and test productions, the effects of production block (baseline vs. test) were not significant in either the VOT or F2OT models. That is, participants did not adjust their stop VOT/F2OTs after hearing and shadowing aspirated stops with extra high post-stop *f0*. On the other hand, production block (baseline vs. test) had significant effects on post-stop *f0* for aspirated stops, lax stops, and fillers. Post-aspirated-stop *f0* increased in test production [$t = 3.250$, $p = 0.004$], while post-lax-stop and post-sonorant *f0* decreased [$t = -2.535$, $p = 0.020$; $t = -2.606$, $p = 0.018$, respectively]. These results suggest that speech imitation may not be limited to a specific phonological category or feature. It seems the participants "imitated" the enhanced aspiration by maximizing the difference in post-stop *f0* between aspirated stops and other stops, as post-stop *f0* is the primary cue for stop aspiration. (This point is discussed further in Section §4.1.2.)

Effects of the remaining word duration on the two duration measures and effects of speaker gender on pitch in general were found in the current models as well, which were not different from those reported in §3.2.2.1. Also, exposure (shadowed vs. unheard) was again not significant in any of the current models [$| t | < 1$, $p > 0.1$], indicating that shadowed words were not different from unheard words in this imitation task. This result supports the claim that the target of speech imitation is not individual words.

3.2.2.3 Interim summary

To recapitulate important findings of this section, the effects of /t^h/ with high post-stop *f0* are as follows.

During shadowing blocks, relative to baseline:

1. VOT/F2OT of /t^h/ decreased.
2. Excluding three female outlier speakers, post-stop *f0* of /t^h/ increased while post-sonorant *f0* did not increase.
3. For three outlier speakers who seem to have imitated the male model speaker, post-stop *f0* decreased; these speakers' patterns demonstrate cross-individual variation in spontaneous imitation.

During test block, relative to baseline:

1. VOT/F2OT of /t^h/, /t/, and /t*/ did not change significantly.
2. Post-stop *f0* of /t^h/ increased while *f0* following lax /t/ and sonorant onsets decreased.
3. Shadowed and unheard /t^h/-initial words were not different in their imitative changes.

Additionally, for all production blocks, VOT/F2OT were negatively correlated with the rest of word duration. Also, with the exception of the outlier speakers, male and female participants did not differ in their imitative patterns.

3.2.3 Long VOT condition: non-primary cue in imitation

Previous findings for spontaneous speech imitation show that English speakers imitate artificially extended VOTs of voiceless stops both in shadowing (Fowler et al., 2003; Shockley et al., 2004) and in delayed imitation (Nielsen, 2011). Since VOT is a primary cue for English voiceless stops, it remains unclear whether extended VOT facilitates imitation when it is not the primary cue for a phonation contrast, as is the case for Seoul Korean. Also, if imitation does occur, it is unknown whether the imitation patterns for Seoul Korean speakers will be comparable to those of English speakers. To answer these questions, the effects of enhancing a non-primary cue for aspirated stops (long stop VOT) on spontaneous imitation are examined. Tables 3.10-3.12 give summaries of the descriptive statistics for stop VOT, F2OT, and post-stop *f0* measurements, respectively, in the extended VOT condition.

Word Type	Production Block	Stop VOT (ms)			
		Males		Females	
		Mean	Std. Dev.	Mean	Std. Dev.
Shadowed /t^h/	Baseline	65.33	17.54	65.34	18.87
	Shadowing 1	65.18	16.65	64.19	20.61
	Shadowing 2	66.11	16.07	65.70	21.35
	Shadowing 3	66.29	16.41	66.07	20.29
	Test	67.25	19.03	67.03	20.88
Unheard /t^h/	Baseline	64.81	16.81	66.42	18.78
	Test	70.12	18.27	69.48	21.62
Unheard /t/	Baseline	50.68	17.47	60.19	18.44
	Test	53.42	17.59	61.03	18.80
Unheard /t*/	Baseline	14.28	4.44	13.10	4.86
	Test	14.29	4.82	12.99	6.75

Table 3.10. Summary of descriptive statistics for VOT (msec) in the long VOT condition

Word Type	Production Block	Stop F2OT (ms)			
		Males		Females	
		Mean	Std. Dev.	Mean	Std. Dev.
Shadowed /t^h/	Baseline	72.77	17.56	72.69	19.63
	Shadowing 1	72.68	16.19	72.45	21.76
	Shadowing 2	74.77	16.77	73.71	21.87
	Shadowing 3	74.29	16.16	74.13	20.92
	Test	75.58	18.81	74.70	22.22
Unheard /t^h/	Baseline	72.09	16.84	74.10	20.01
	Test	79.44	18.25	77.27	22.32
Unheard /t/	Baseline	57.82	16.43	68.40	18.07
	Test	61.48	17.04	70.12	18.89
Unheard /t*/	Baseline	18.32	5.93	16.02	6.36
	Test	18.18	6.47	16.45	7.85

Table 3.11. Summary of descriptive statistics for F2OT (msec) in the long VOT condition

Word Type	Production Block	Post-stop *F0* (Hz)			
		Males		Females	
		Mean	Std. Dev.	Mean	Std. Dev.
Shadowed /tʰ/	Baseline	129.40	12.53	259.47	28.15
	Shadowing 1	133.71	14.77	264.86	31.03
	Shadowing 2	133.66	14.45	261.96	32.45
	Shadowing 3	135.00	13.55	263.99	31.57
	Test	136.78	14.69	269.74	36.18
Unheard /tʰ/	Baseline	128.44	12.27	260.40	29.06
	Test	136.29	15.02	270.87	37.53
Unheard /t/	Baseline	105.75	9.34	198.53	19.16
	Test	107.81	9.19	202.84	20.89
Unheard /t*/	Baseline	121.82	11.08	244.16	25.95
	Test	127.32	12.10	251.39	33.11
Shadowed fillers	Baseline	103.64	8.82	195.04	17.21
	Shadowing 1	105.14	8.41	197.91	21.76
	Shadowing 2	105.98	8.92	197.88	22.48
	Shadowing 3	107.18	7.85	199.54	21.93
	Test	105.73	7.94	197.01	20.05
Unheard fillers	Baseline	104.39	9.01	194.85	17.26
	Test	105.81	8.94	197.22	20.27

Table 3.12. Summary of descriptive statistics for post-stop *f0* (Hz) in the long VOT condition

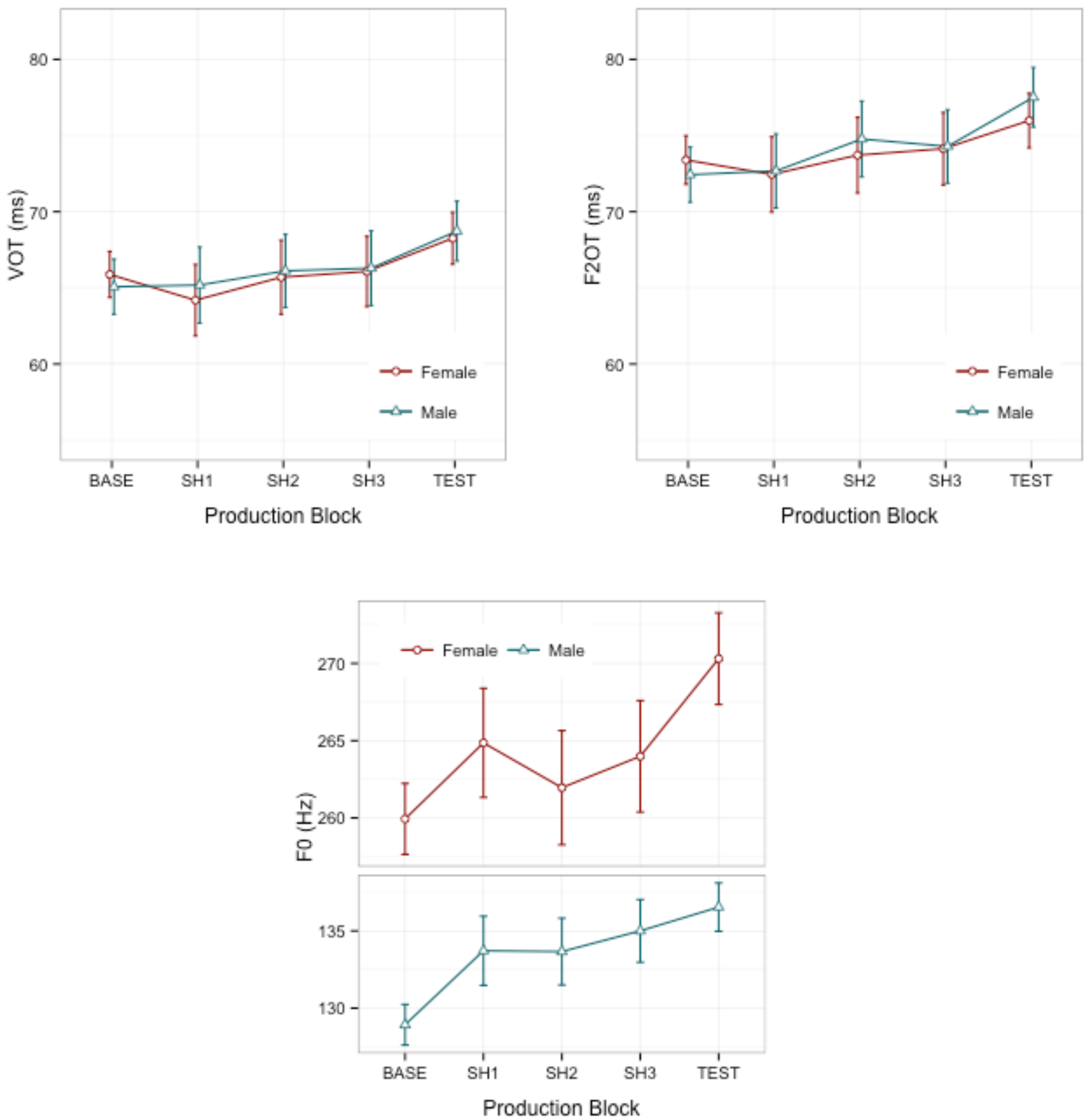


Figure 3.7. Changes in aspirated /t^h/ in the long VOT condition: stop VOT (top left), stop F2OT (top right), and post-stop *f0* (bottom). Error bars represent 95% confidence intervals.

3.2.3.1 Effects of long VOT /t^h/ stimuli on productions of /t^h/ in shadowing and test blocks

Figure 3.7 presents mean VOT, F2OT and post-stop *f0* of aspirated /t^h/s across speakers in the different production blocks (baseline, shadowing 1, 2, 3, and test) in the long VOT condition.

Three separate linear mixed effects models with different dependent variables (stop VOT, F2OT, and post-stop *f0*) were fitted in order to analyze changes in aspirated /t^h/ productions in five production blocks (baseline, shadowing 1, 2, 3, and test) as a result of exposure to the enhanced non-primary cue for aspirated stops. The details of the statistical modeling, such as fixed effects, random effects, and interaction terms included, were identical to the models used in the *f0* condition described in 3.2.2.1.

Table 3.13 shows the parameter estimate *β*, *t* value, and *p* value for fixed effects included in the three fitted models. The pairwise comparisons for production block and word type are presented when the baseline production is compared against other production blocks or when baseline productions of different word types are compared against each other.

In the two duration models, neither VOT nor F2OT was different between baseline and shadowing productions. That is, aspirated stops with extended VOT did not induce significant imitation effects in VOT [$|t| < 0.5, p > 0.1$], or F2OT [$|t| < 1, p > 0.1$] in shadowing productions. For the baseline-test comparison, the effects of production block were significant only in the F2OT model [$t = 2.102, p = 0.035$]. Changes in F2OT seem more robust than those in VOT; aspirated stop VOT was not significantly longer in test productions than in the baseline counterpart [$t = 1.922, p = 0.071$]. Word duration effects were significant in both VOT/F2OT models, in the same fashion as in the high *f0* condition.

Dependent variable (Outcome)	Fixed effect (Predictor)	Estimate (β)	t	p
Stop VOT	**(Intercept)**	**83.010**	**21.172**	**< 0.001**
	Production block:			
	Base-Shadowing 1	-0.965	-0.449	0.659
	Base-Shadowing 2	0.306	0.143	0.888
	Base-Shadowing 3	0.619	0.348	0.732
	Base-Test	2.401	1.922	0.071
	Exposure (unheard-shadowed)	-3.352	-1.116	0.270
	Gender (female-male)	-3.201	-0.914	0.373
	Word duration (Rest)	**-0.033**	**-6.501**	**< 0.001**
Stop F2OT	**(Intercept)**	**92.299**	**23.978**	**< 0.001**
	Production block:			
	Base-Shadowing 1	-0.414	-0.201	0.841
	Base-Shadowing 2	1.126	0.538	0.591
	Base-Shadowing 3	1.231	0.711	0.477
	Base-Test	**2.990**	**2.102**	**0.035**
	Exposure (unheard-shadowed)	-3.819	-1.221	0.222
	Gender (female-male)	-3.994	-1.190	0.234
	Word duration (Rest)	**-0.037**	**-7.103**	**< 0.001**
Post-stop *f0*	**(Intercept)**	**244.682**	**34.676**	**< 0.001**
	*Word type * Production block:*			
	Aspirated-Filler, Base	**-50.211**	**-8.325**	**< 0.001**
	Aspirated, Base-Shadowing 1	**4.985**	**2.010**	**0.044**
	Aspirated, Base-Shadowing 2	3.228	1.315	0.188
	Aspirated, Base-Shadowing 3	**4.930**	**2.003**	**0.045**
	Aspirated stop, Base-Test	**9.326**	**4.554**	**< 0.001**
	Filler, Base-Shadowing 1	-2.613	-1.826	0.068
	Filler, Base-Shadowing 2	-0.570	-0.397	0.691
	Filler, Base-Shadowing 3	-0.800	-0.473	0.636
	Filler, Base-Test	**-7.311**	**-3.745**	**< 0.001**
	Exposure (unheard-shadowed)	-0.239	-0.219	0.827
	Gender (female-male)	**-89.319**	**-24.165**	**< 0.001**

Table 3.13. Summary of linear mixed models for aspirated stops in different production blocks in the long VOT condition. **Bolded** data are significant at $p < 0.05$.

Contrary to the results in both duration models, the *f0* model showed significant effects of VOT extended stimuli on post-aspirated-stop *f0* in different production blocks. Post-stop *f0* in the first and last blocks of shadowing was significantly higher than that in baseline [$t = 2.010$, $p = 0.044$; $t = 2.003$, $p = 0.045$, respectively]. In test production, *f0* after aspirated stops was significantly higher than the baseline counterpart [$t = 4.554$, $p < 0.001$], while post-sonorant *f0* in filler words significantly decreased [$t = -3.745$, $p < 0.001$]. Post-stop *f0* in the second shadowing block was not significantly different from the baseline counterpart [$t = 1.315$, $p = 0.188$].

Taken together, the effects of extended VOT stimuli on /t^h/s in different production blocks indicate that participants imitated exaggerated stop aspiration cued by longer VOTs primarily by raising post-stop *f0* with a small (trending) increase in F2OT (VOT). Recall that the stimuli that the participants heard during shadowing blocks in this task never included *f0* manipulation. Even so, strong effects on post-stop *f0* emerged as a result of exposure to aspirated stops with extended VOTs. This result presumably occurred because post-stop *f0* is the primary cue for aspirated stops for most of the participants in this study. The current finding is not due to overall increase in *f0* range. To rule out the possibility of global shift in *f0* range, sonorant-initial fillers were included in the statistical modeling and the post-sonorant *f0* did not change in shadowing blocks and decreased in test block compared to the baseline (Table 3.13).

Other significant fixed effects in the *f0* model were the effect of speaker gender and word type (aspirated stop words vs. sonorant-initial fillers): female speakers had higher *f0* than male speakers and post-aspiration *f0* was higher than post-sonorant *f0*, as expected.

The effects of presence of exposure (shadowed vs. unheard) were not significant in any of the three models [$|t| < 1.3$, $p > 0.1$]. Words beginning with initial aspirated stops showed the same imitation effects regardless of whether the specific word was present or not during the shadowing blocks, as illustrated in Figure 3.8.

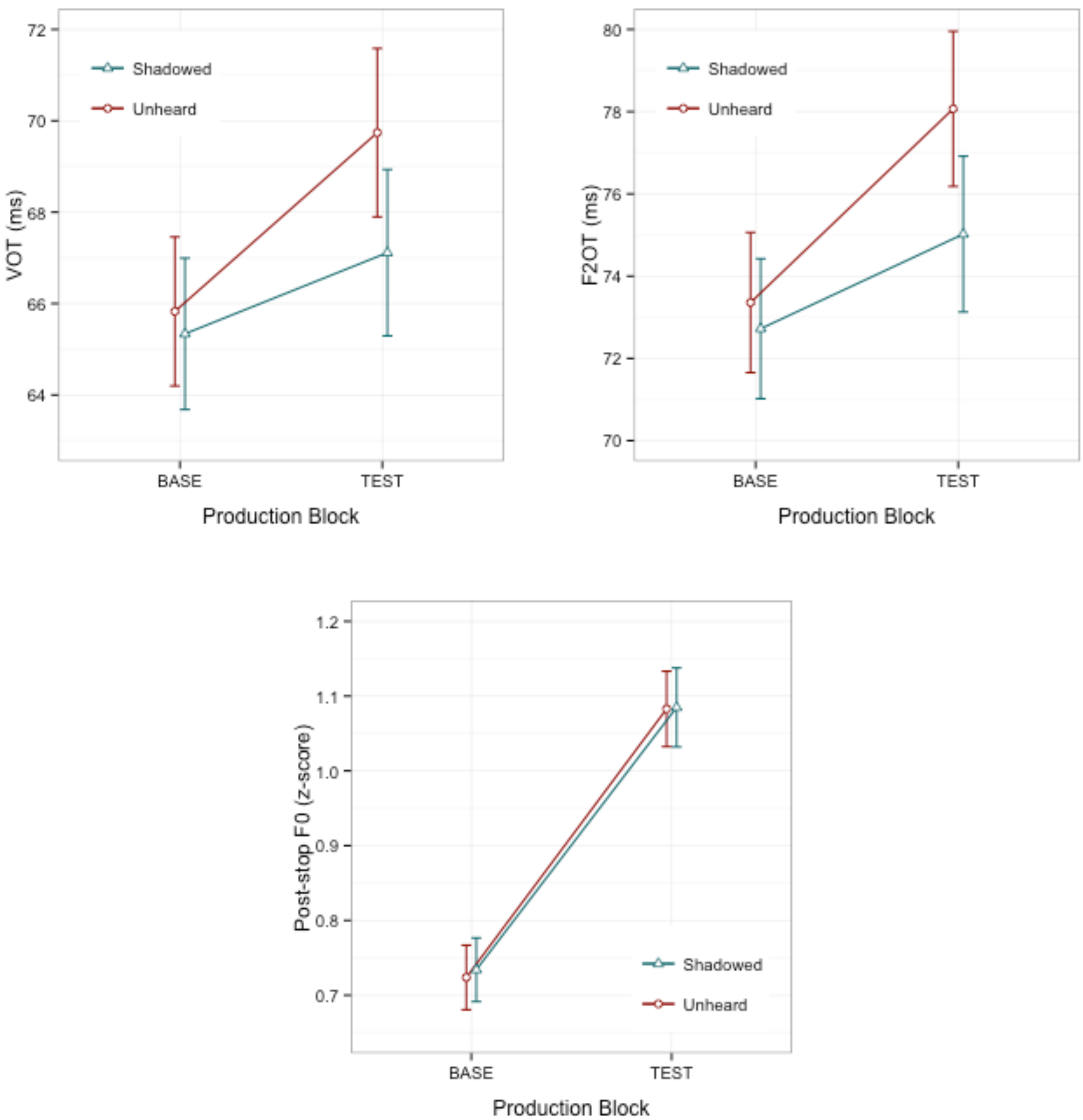


Figure 3.8. Shadowed and unheard /t^h/-initial words in the long VOT condition: stop VOT (top left), stop F2OT (top right), and post-stop *f0* (bottom). To plot speakers of both genders together, post-stop *f0* is given in z-score. Error bars represent 95% confidence intervals.

3.2.3.2 Effects of long VOT /t^h/ stimuli on productions of /t/ and /t*/

Figures 3.9-3.11 present mean VOT, F2OT and post-stop *f0*, respectively, of different stop types in baseline and test productions of the long VOT condition. To analyze productions of /t^h/, /t/, and /t*/ in baseline and test productions in the long VOT condition, three additional linear mixed effects models—one for each of the three dependent variables (stop VOT, F2OT, and post-stop *f0*)—were conducted. The details of statistical modeling, such as fixed effects, random effects, and interaction terms included, were identical to the models used in the VOT condition described in 3.2.2.2.

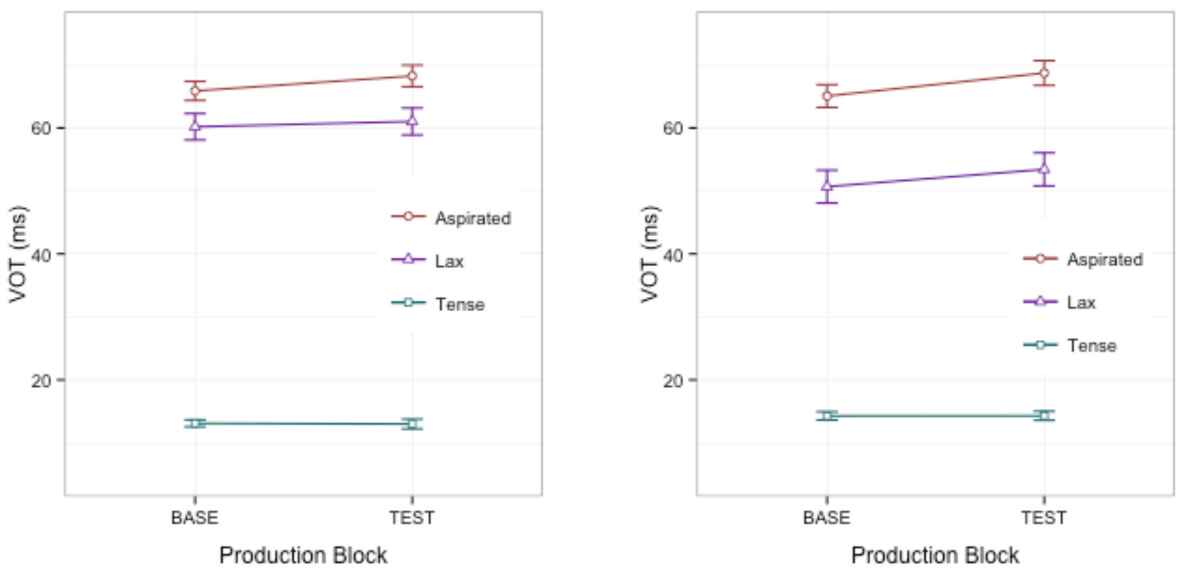


Figure 3.9. VOT changes in the long VOT condition for different stops (aspirated /t^h/, lax /t/, and tense /t^*/) produced by female (left) and male (right) speakers. Error bars represent 95% confidence intervals.

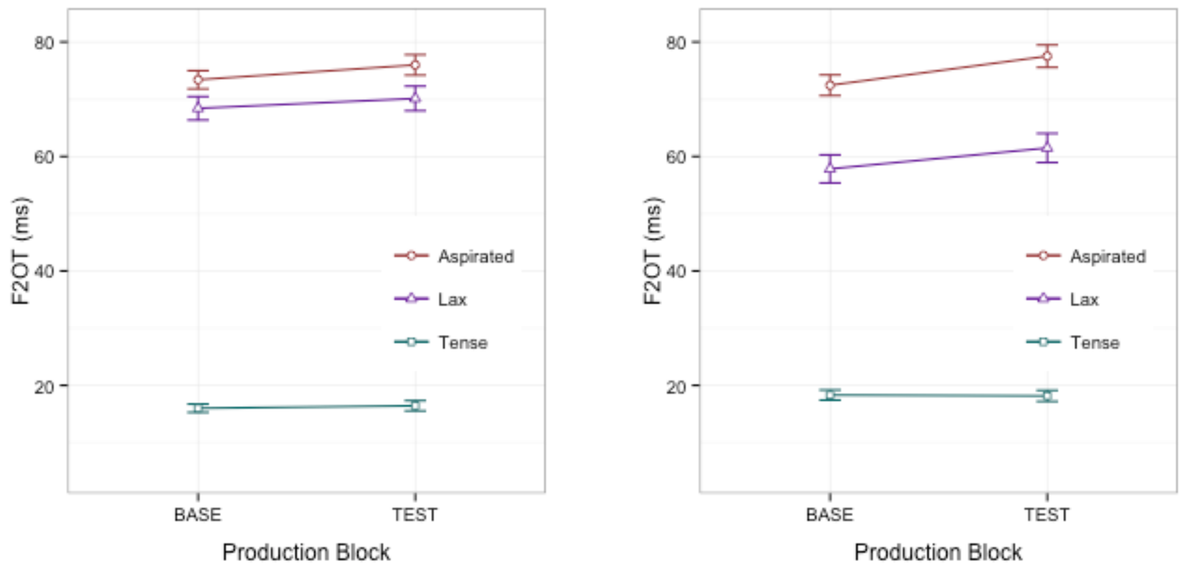


Figure 3.10. F2OT changes in the long VOT condition for different stops (aspirated $/t^h/$, lax /t/, and tense $/t^*/$) produced by female (left) and male (right) speakers. Error bars represent 95% confidence intervals.

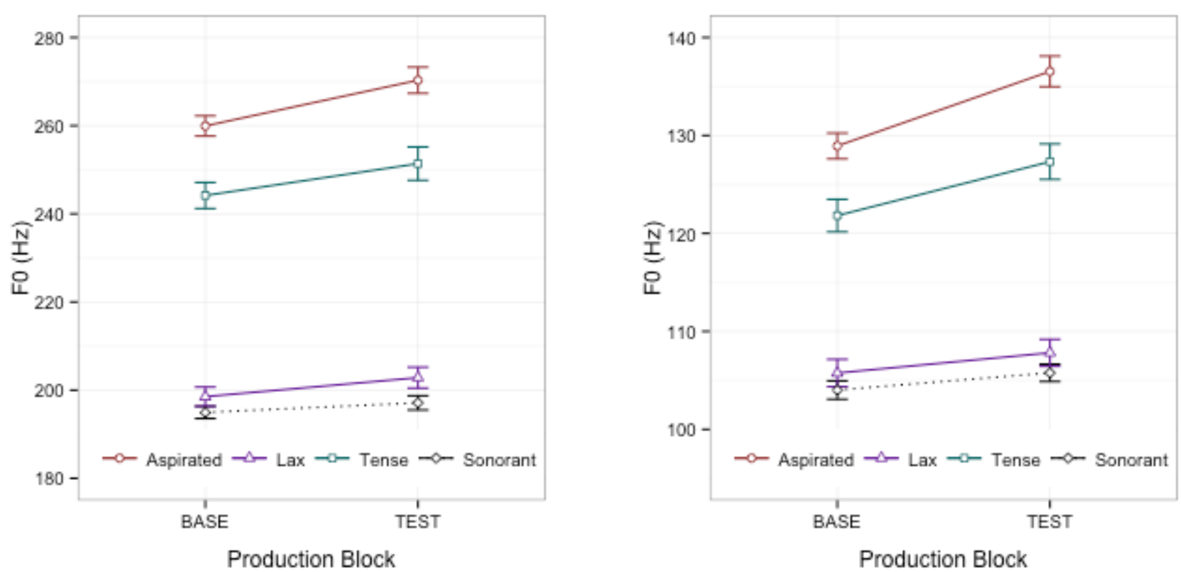


Figure 3.11. Post-onset *f0* changes in the long VOT condition for different consonants (aspirated $/t^h/$, lax /t/, tense $/t^*/$, and sonorant) produced by female (left) and male (right) speakers. Error bars represent 95% confidence intervals.

Dependent variable (Outcome)	Fixed effect (Predictor)	Estimate (β)	t	p
Stop VOT	**(Intercept)**	**81.105**	**24.832**	**< 0.001**
	*Stop type * Production block:*			
	Aspirated-Lax, Base	**-10.135**	**-3.284**	**0.001**
	Aspirated-Tense, Base	**-52.150**	**-16.963**	**< 0.001**
	Aspirated stop, Base-Test	2.451	1.961	0.065
	Lax stop, Base-Test	-1.161	-0.899	0.379
	Tense stop, Base-Test	-2.766	-1.800	0.087
	Exposure (unheard-shadowed)	-3.081	-1.128	0.262
	Gender (female-male)	-2.821	-1.678	0.111
	Word duration (Rest)	**-0.029**	**-7.464**	**< 0.001**
Stop F2OT	**(Intercept)**	**90.710**	**27.473**	**< 0.001**
	*Stop type * Production block:*			
	Aspirated-Lax, Base	**-9.939**	**-3.132**	**0.002**
	Aspirated-Tense, Base	**-56.137**	**-17.715**	**< 0.001**
	Aspirated stop, Base-Test	**3.040**	**2.144**	**0.032**
	Lax stop, Base-Test	-0.917	-0.655	0.513
	Tense stop, Base-Test	-3.124	-1.835	0.066
	Exposure (unheard-shadowed)	-3.466	-1.230	0.219
	Gender (female-male)	**-4.299**	**-2.180**	**0.029**
	Word duration (Rest)	**-0.033**	**-8.210**	**< 0.001**
Post-stop *f0*	**(Intercept)**	**248.739**	**39.489**	**< 0.001**
	Word type * Production block:			
	Aspirated-Lax, Base	**-47.434**	**-7.901**	**< 0.001**
	Aspirated-Tense, Base	**-12.682**	**-5.787**	**< 0.001**
	Aspirated-Filler, Base	**-50.213**	**-8.340**	**< 0.001**
	Aspirated stop, Base-Test	**9.321**	**4.570**	**< 0.001**
	Lax stop, Base-Test	**-5.864**	**-3.012**	**0.007**
	Tense stop, Base-Test	**-2.824**	**-3.045**	**0.005**
	Filler, Base-Test	**-7.306**	**-3.735**	**0.002**
	Exposure (unheard-shadowed)	-0.243	-0.213	0.831
	Gender (female-male)	**-100.319**	**-23.761**	**< 0.001**

Table 3.14. Summary of linear mixed models for three stop types in baseline and test productions of the long VOT condition. **Bolded** data are significant at $p < 0.05$

Table 3.14 shows the parameter estimate *β*, *t* value, and *p* value for the fixed effects included in the three models. The pairwise comparisons for production block and word type are presented when the baseline aspirated stop production is compared with other types of stops or production blocks. In keeping with the results for the high *f0* condition, the comparisons among different stop/word types within baseline production confirmed that the aspirated stops and lax stops are primarily distinguished by post-stop *f0* rather than stop VOT/F2OT. One unexpected finding was the significant gender effect in F2OT. To identify the source of the gender effect, another linear mixed model that included an interaction term among stop type, production block and speaker gender was fitted to the data. The results of this model revealed that male speakers have significantly shorter F2OT than female speakers only for lax stops [$\beta = -9.255$, $t = -3.792$, $p = 0.001$] (see Figure 3.10 for visual demonstration of this gender difference). This suggests the possibility that male participants in this study still maintain the F2OT contrast between lax and aspirated stops at least to some extent. Male and female speakers were not different in terms of their baseline-test comparisons.

Turning to the comparisons between baseline and test productions, the effects of production block (baseline vs. test) were very robust in the *f0* models, but not as robust in the two durational models. The only baseline-test comparison that reached the significance level ($p < 0.05$) in the two duration models was the F2OT difference in aspirated stops: aspirated stops in test production showed longer F2OT than those in baseline [$t = 2.144$, $p = 0.032$]. The same pairwise comparison in the VOT model was marginally significant [$t = 1.961$, $p = 0.065$]. On the other hand, the *f0* model showed highly significant effects for all word types: after hearing and shadowing extended VOT stimuli, *f0* after aspirated stops increased [$t = 4.570$, $p < 0.001$] while other word types (lax stops, tense stops, and fillers) all showed a significant decrease in post-onset *f0* [$| t | > 3, p < 0.01$]. These changes in post-consonantal *f0* for all word types indicate that the effects of speech imitation are not limited to a specific segment or word. Participants in this study seem to have imitated the enhanced phonological contrast signaled by longer VOTs in the aspirated stop stimuli, perhaps using their own way of exaggerating the contrast. This robust change in post-stop *f0* induced by the long VOT stimuli forms a striking contrast with the (lack of) imitative effects of the high *f0* stimuli on VOT, as reported in §3.2.2. This presumably suggests that primary and secondary cues play different roles in speech imitation. As the primary cue for phonological aspiration for most of the participants in this study is post-stop *f0*, they "imitated" the enhanced aspiration by maximizing the *f0* difference between aspirated stops and other stops/sonorants. (This point is discussed further in Section §4.1.2.)

Similar to the high post-stop *f0* condition, and as would be expected, the remaining (non-VOT/F2OT) word duration is negatively correlated with VOT/F2OT. Although the correlation is highly significant statistically, the mean differences are negligible in reality [β = -0.029 (VOT) and -0.033 (F2OT)]. Effects of speaker gender on *f0* in general were found, again as expected. Finally, the effects of shadowing (vs. unheard) were not significant in any of the models.

3.2.3.3. Interim summary

In sum, the effects of /t^h/ with long VOT are as follows.

During shadowing blocks, relative to baseline:

1. VOT/F2OT of /t^h/ did not change.
2. Post-stop *f0* of /t^h/ increased in the first and third repetitions of the shadowing blocks. The effect was not significant in the second shadowing block.

3. Post-sonorant *f0* did not change.

During test block, relative to baseline:

1. F2OT of /t^h/ increased significantly. The increase in VOT of /t^h/ was marginally significant ($p = 0.065$). VOT/F2OT of tense /t*/ during the test block showed a marginal decrease [$p = 0.087$ (VOT) and 0.066 (F2OT)].
2. Post-stop *f0* of /t^h/ increased while *f0* following lax /t/, tense /t*/, and sonorant onsets decreased.
3. Shadowed and unheard /t^h/-initial words were not different in their imitative changes.

In addition, for all production blocks, VOT/F2OT were negatively correlated with the rest of word duration. Male and female participants were not different in their imitative patterns; however, in their stop productions, male speakers had overall shorter F2OT for lax stops than did female speakers.

3.2.4 Summary of imitation results

The current findings demonstrate a clear asymmetry between primary and non-primary cues in spontaneous imitation of Seoul Korean aspirated stops. Although both primary and non-primary cues facilitated imitative effects, the details of the imitation patterns were distinct. An enhanced non-primary cue for stop aspiration (long VOT) induced increases in both primary and non-primary cues. That is, the Seoul Korean speakers "imitated" exaggerated stop aspiration cued by long VOT not only by lengthening their own VOT for aspirated stops but also by raising their *f0* after those stops. However, an enhanced primary cue (high post-stop *f0*) did not have similar effects on the non-primary cue. After hearing aspirated stops with raised post-stop *f0*, the same participants only imitated the manipulated property; they did not lengthen VOT. These results indicate that, in speech imitation, exposure to an enhanced phonetic property can influence production not only of that property but also of other phonetic properties if they are important for the targeted phonological category. The participants in this study appear to have adjusted the cue that they would primarily use to enhance stop aspiration (i.e., post-stop *f0*), in addition to imitating the cue that the stimuli they heard employed to enhance aspirated stops.

3.3 Analyses of productions of individual participants

As is typical for imitation experiments, not all participants behaved the same. To assess the variation across participants, the data for each participant were separately analyzed. Twelve separate linear mixed effects models were fitted to the data from each participant (the six summarized in Table 3.15 * two manipulation conditions (high *f0* and long VOT)). Fixed effects were identical to the corresponding models used in the pooled analyses except that speaker gender was not included for individual analyses (as summarized in the third column of Table 3.15). Only a single random effect of word (nested in stop or word type, when appropriate) was included. Parameter-specific *p*-values were obtained by using the Satterthwaite approximation (Kuznetsova et al., 2014).

For most speakers, the statistical results for the two manipulation conditions conformed to the general patterns found in the pooled analyses, although individual speakers' results were far more variable. The remainder of this chapter discusses these individual results, focusing on the following aspects: (1) whether or not each speaker's behavior differs in the two experimental conditions, (2) the relation between each speaker's baseline productions and their imitative patterns, and (3) the relation between each speaker's imitative patterns in the two manipulation conditions and their discrimination accuracy. Because individual shadowing results are too variable and complex for patterns to emerge, the discussion in this section is based only on the baseline-test comparison (see §4.2 for a possible reason for the high variability in shadowing blocks). The full statistical results including the shadowing analyses for each speaker are reported in Appendix C.

Target data under analyses	Dependent Variable	Fixed Effects
Aspirated /t^h/s in Baseline - Shadowing 1, 2, 3 - Test	Stop VOT Stop F2OT	Production block (base, sh1, sh2, sh3, test) + Presence of exposure (shadowed, unheard) + Rest of word duration
	Post-stop *f0*	Word type (/t^h/ vs. fillers) * Production block (base, sh1, sh2, sh3, test) + Presence of exposure (shadowed, unheard)
Aspirated /t^h/s, lax /t/s, and tense /t*/s in Baseline - Test	Stop VOT Stop F2OT	Stop type (/t^h/, /t/, /t*/) * Production block (base, test) + Presence of exposure (shadowed, unheard) + Rest of word duration
	Post-stop *f0*	Word type (/t^h/, /t/, /t*/, fillers) * Production block (base, test) + Presence of exposure (shadowed, unheard)

Table 3.15. Summary of statistical modeling for individual analyses

Figures 3.12-3.17 provide mean VOT, F2OT, and post-stop *f0* of /t^h/ in baseline and test productions in the two manipulation conditions for individual speakers. For each manipulation condition, speakers are ordered according to their discrimination accuracy scores for the corresponding property from left to right. For instance, in Figure 3.12, the rightmost speaker

M16 obtained the highest accuracy score in the discrimination task using high *f0* stimuli, followed by M14, F07, and so on. Statistical significance ($p < 0.05$) is indicated by two different symbols, ☆ for an increase from baseline to test, and * for a decrease. As participants heard /t^h/ with increased VOT and *f0*, only the ☆ indicates significant imitation. Recall that ascending changes in the non-manipulated cue (e.g., increase in post-stop *f0* after hearing /t^h/ with long VOT) are also referred to as imitation in this study, as it is viewed as evidence of phonological imitation (see §2.5 for phonetic and phonological imitation hypotheses).

In both manipulation conditions, only a small number of participants lengthened VOT and F2OT of /t^h/ (Figures 3.12-3.15) whereas almost all participants increased their post-stop *f0* (Figures 3.16-3.17). Most of the participants produced /t^h/s with higher post-stop *f0* after hearing /t^h/ with high *f0,* and even more participants did so after hearing /t^h/ with long VOT. All participants except for speakers F11, F20, and F21 produced significantly higher post-stop *f0* in test production than in baseline production in the long VOT condition, as presented in Figure 3.17. Note that the speakers who significantly decreased their *f0* in the high *f0* condition (F04, M16, and M17 in Figure 3.16) are different individuals from the three female outlier speakers who appear to have imitated the male model speaker by lowering their own *f0*s. As mentioned in §3.2.2.1, their lowering of *f0* is limited to shadowing production and their *f0* in the test production was not different from their baseline.

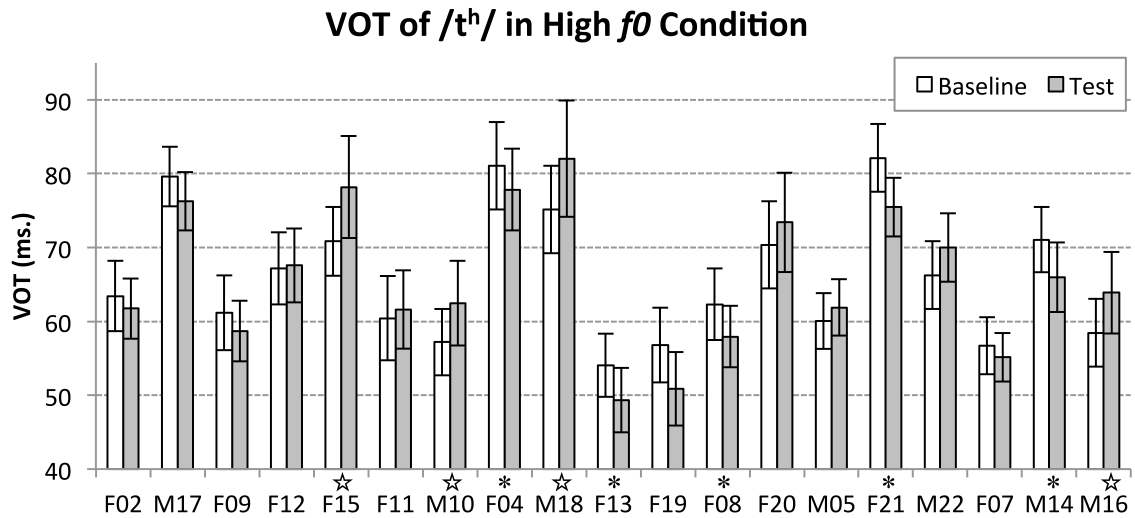


Figure 3.12. VOT of /t^h/ in baseline and test productions of the high *f0* condition, for each speaker. ☆ indicates a significant increase and * indicates a significant decrease ($p < 0.05$). Speakers are presented from left to right in order of increasing discrimination accuracy scores for high *f0* stimuli. Error bars represent 95% confidence intervals.

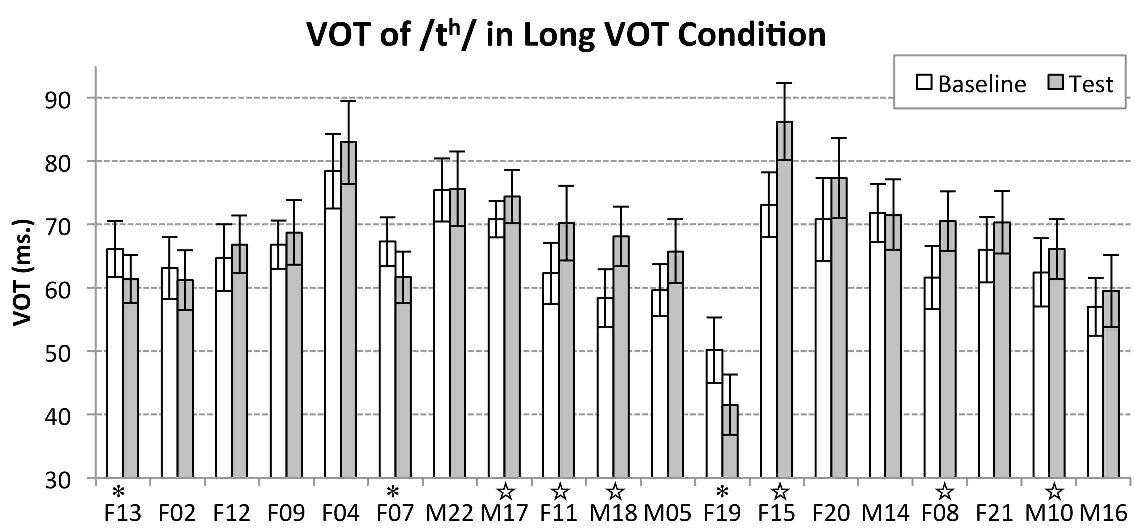


Figure 3.13. VOT of /t^h/ in baseline and test productions of the long VOT condition, for each speaker. ☆ indicates a significant increase and * indicates a significant decrease ($p < 0.05$). Speakers are presented from left to right in order of increasing discrimination accuracy scores for long VOT stimuli. Error bars represent 95% confidence intervals.

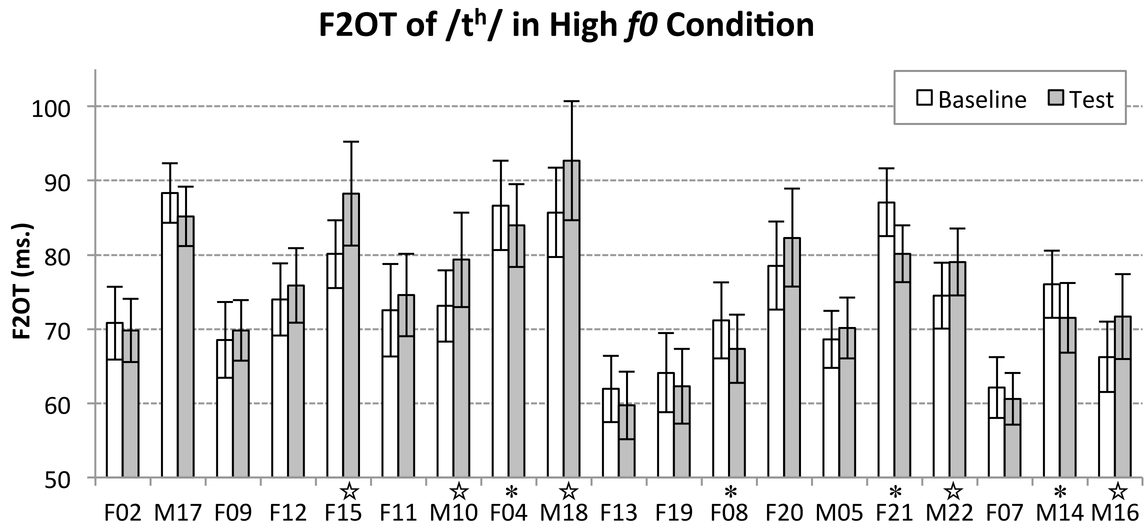


Figure 3.14. F2OT of /t^h/ in baseline and test productions of the high *f0* condition, for each speaker. ☆ indicates a significant increase and * indicates a significant decrease ($p < 0.05$). Speakers are presented from left to right in order of increasing discrimination accuracy scores for high *f0* stimuli. Error bars represent 95% confidence intervals.

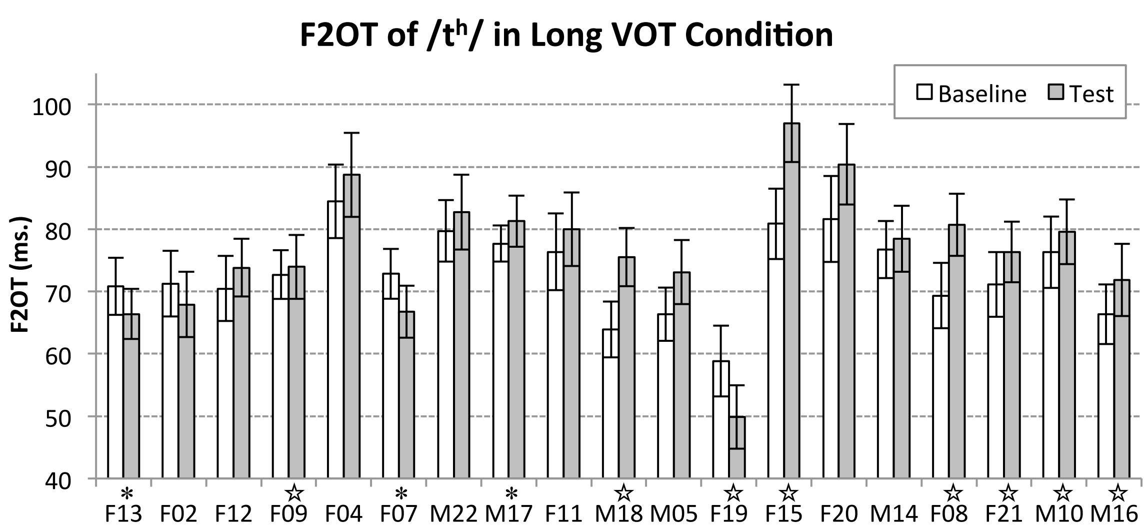


Figure 3.15. F2OT of /t^h/ in baseline and test productions of the long VOT condition, for each speaker. ☆ indicates a significant increase and * indicates a significant decrease ($p < 0.05$). Speakers are presented from left to right in order of increasing discrimination accuracy scores for long VOT stimuli. Error bars represent 95% confidence intervals.

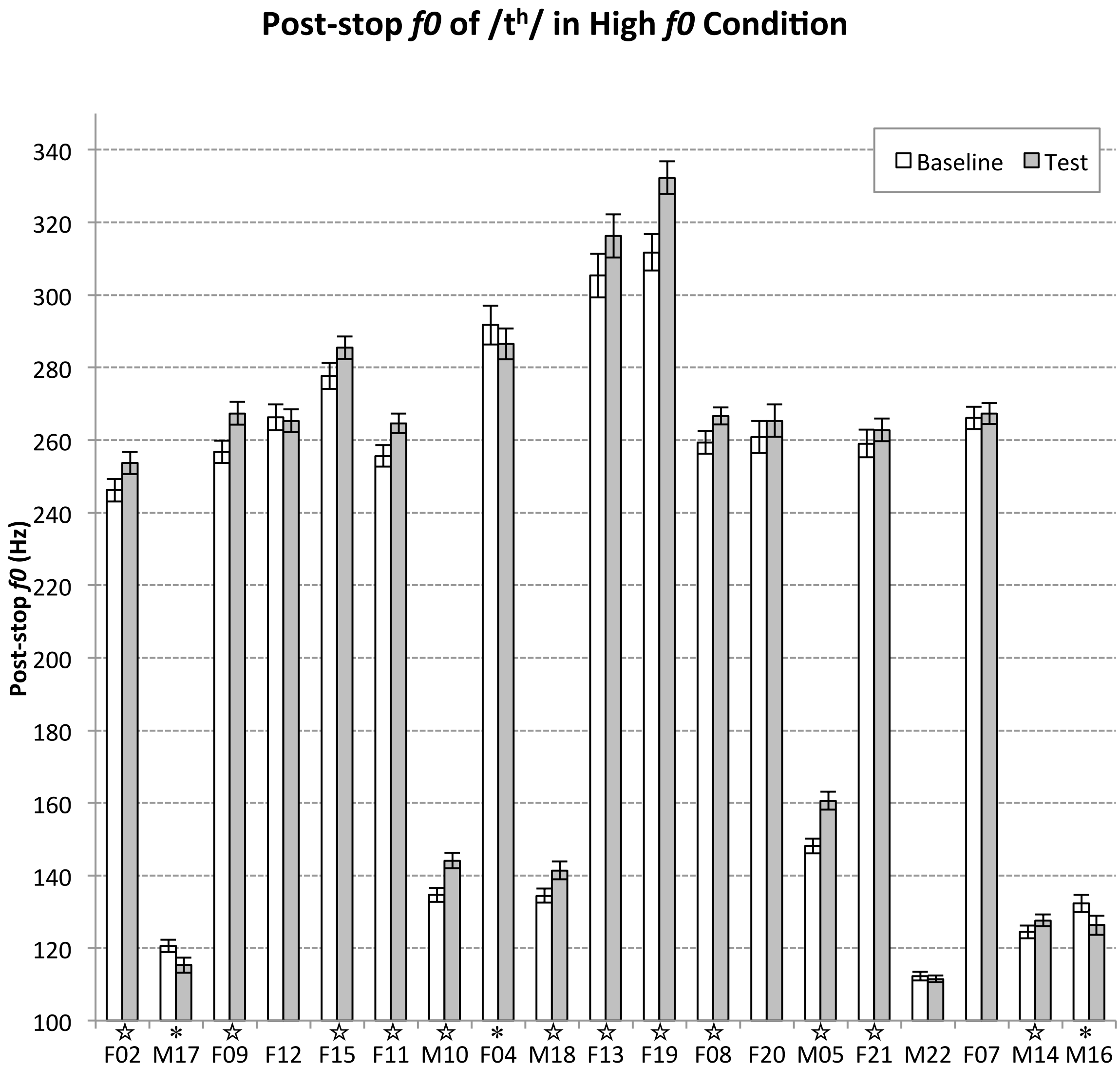


Figure 3.16. Post-stop *f0* of /tʰ/ in baseline and test productions of the high *f0* condition, for each speaker. ☆ indicates a significant increase and * indicates a significant decrease ($p < 0.05$). Speakers are presented from left to right in order of increasing discrimination accuracy scores for high *f0* stimuli. Error bars represent 95% confidence intervals.

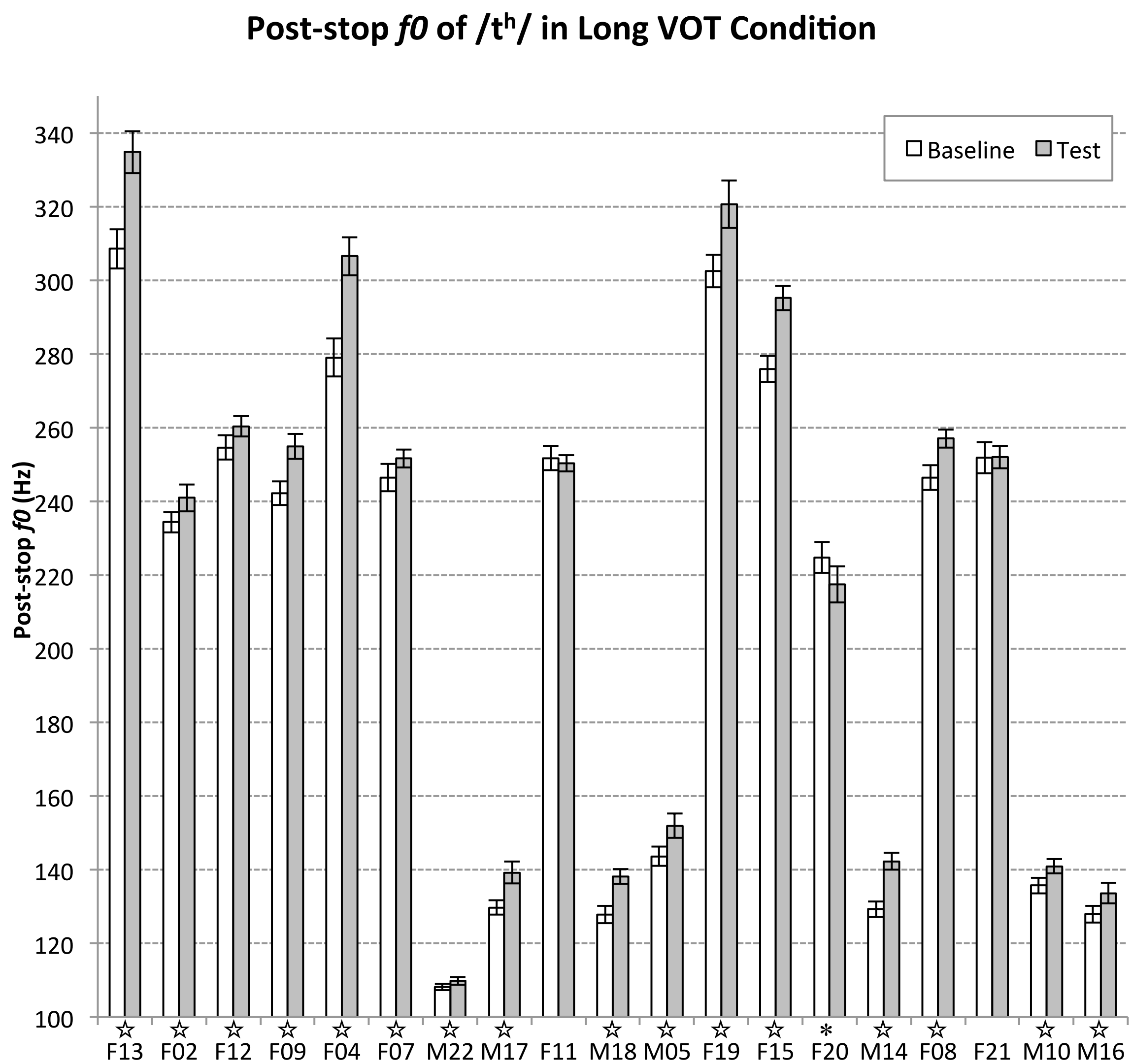


Figure 3.17. Post-stop *f0* of /t^h/ in baseline and test productions of the long VOT condition, for each speaker. ☆ indicates a significant increase and * indicates a significant decrease ($p < 0.05$). Speakers are presented from left to right in order of increasing discrimination accuracy scores for long VOT stimuli. Error bars represent 95% confidence intervals.

Table 3.16 presents comparisons of each participant's productions in the high *f0* and long VOT conditions. Two speakers (F11 and F21) seem to be "phonetic imitators" who raised *f0* after hearing high *f0* stimuli and lengthened VOT/F2OT after hearing long VOT stimuli. On the other hand, for half of participants, their imitative patterns were similar in the two manipulation conditions. Speakers F02, M05, F09, F13, F14, and F19 produced /t^h/ with higher post-stop *f0* in both conditions, and speakers M10, F15, and M18 produced /t^h/ with longer VOT and higher *f0* in both conditions. No participant imitated by lengthening VOT without raising *f0* in both conditions. This outcome is consistent with the phonological imitation hypothesis. The phonological imitation hypothesis predicted that the two manipulations (high *f0* and long VOT) would induce identical imitative patterns for a given speaker, and that the shared imitative patterns for the two manipulation conditions would involve enhancement of the phonetic property(s) the speakers would normally employ to enhance /t^h/ (§2.5.2). Participants in the current study are all young speakers of Seoul Korean, and many of them enhanced /t^h/ by increasing either "only post-stop *f0*" or "both stop VOT and post-stop *f0*".

To further verify whether participants' imitative patterns are consistent with their baseline productions of Korean lax and aspirated stops, VOT, F2OT and post-stop *f0* of /t^h/, /t/, and /t*/ in each speaker's baseline productions were analyzed by performing three separate one-way ANOVAs (dependent variable: VOT, F2OT, post-stop *f0*, respectively for each model; independent variable: stop type) for each speaker. The results of these ANOVAs confirmed that no speaker used only VOT for the aspirated-lax contrast, as should be expected (Table 3.17). (It has already been noted that no speaker showed only VOT/F2OT enhancement in the two test conditions.) The full results of these ANOVAs are summarized in Appendix C.

	High *f0* condition			Long VOT condition		
	f0	VOT	F2OT	*f0*	VOT	F2OT
F02	☆			☆		
F04	*	*	*	☆		
M05	☆			☆		
F07				☆	*	*
F08	☆	*	*	☆	☆	☆
F09	☆			☆		
M10	☆	☆	☆	☆	☆	☆
F11	☆				☆	
F12				☆		
F13	☆	*		☆	*	*
M14	☆	*	*	☆		
F15	☆	☆	☆	☆	☆	☆
M16	*	☆	☆	☆		☆
M17	*			☆	☆	☆
M18	☆	☆	☆	☆	☆	☆
F19	☆			☆	*	*
F20				*		
F21	☆	*	*			☆
M22			☆	☆		

Table 3.16. Summary of imitative changes in two manipulation conditions, by speakers. ☆ indicates a significant increase and * indicates a significant decrease ($p < 0.05$). Speakers who increased or decreased the same set of properties in the two conditions are **bolded**. Those who imitated only by raising *f0* (ignoring significant decreases in VOT/F2OT) are shaded.

	Aspirated – Lax		
	VOT (ms.)	F2OT (ms.)	Post-stop *f0* (Hz)
F02	- 0.51 (n/s)	0.42 (n/s)	**56.69 (*p* < 0.0001)**
F04	3.93 (n/s)	4.30 (n/s)	**89.49 (*p* < 0.0001)**
M05	**11.25 (*p* < 0.0001)**	**10.77 (*p* < 0.0001)**	**21.40 (*p* < 0.0001)**
F07	1.65 (n/s)	0.70 (n/s)	**52.48 (*p* < 0.0001)**
F08	**12.14 (*p* < 0.0001)**	**11.37 (*p* < 0.0001)**	**45.96 (*p* < 0.0001)**
F09	**14.43 (*p* < 0.0001)**	**10.86 (*p* < 0.0001)**	**45.37 (*p* < 0.0001)**
M10	**18.37 (*p* < 0.0001)**	**20.16 (*p* < 0.0001)**	**26.19 (*p* < 0.0001)**
F11	**8.61 (*p* = 0.0098)**	**8.24 (*p* = 0.0388)**	**43.06 (*p* < 0.0001)**
F12	3.48 (n/s)	2.40 (n/s)	**58.63 (*p* < 0.0001)**
F13	**12.71 (*p* < 0.0001)**	**12.88 (*p* < 0.0001)**	**90.43 (*p* < 0.0001)**
M14	**9.80 (*p* = 0.0003)**	**10.35 (*p* = 0.0001)**	**23.48 (*p* < 0.0001)**
F15	**6.65 (*p* = 0.0226)**	**6.40 (*p* = 0.0362)**	**59.50 (*p* < 0.0001)**
M16	**17.47 (*p* < 0.0001)**	**17.56 (*p* < 0.0001)**	**29.81 (*p* < 0.0001)**
M17	**15.55 (*p* < 0.0001)**	**14.98 (*p* < 0.0001)**	**19.24 (*p* < 0.0001)**
M18	**11.92 (*p* = 0.0011)**	**12.02 (*p* = 0.0016)**	**22.04 (*p* < 0.0001)**
F19	-4.78 (n/s)	**-7.10 (*p* = 0.0462)**	**107.58 (*p* < 0.0001)**
F20	6.21 (n/s)	4.34 (n/s)	**72.66 (*p* < 0.0001)**
F21	**10.97 (*p* = 0.0003)**	**6.86 (*p* = 0.0386)**	**56.16 (*p* < 0.0001)**
M22	**14.61 (*p* < 0.0001)**	**15.58 (*p* < 0.0001)**	**14.19 (*p* < 0.0001)**

Table 3.17. Baseline aspirated-lax contrast, by speakers. Numbers represent mean difference between aspirated and lax stops. *P*-values are from post-hoc Tukey tests. **Bolded** data are significant at $p < 0.05$, and shaded speakers show merged (or reversed) VOT/F2OT patterns.

The speakers whose cells are shaded in Tables 3.16 and 3.17 are those who seem to rely exclusively on post-stop *f0* in imitation (Table 3.16) and baseline (Table 3.17). Except for one speaker (F20), who does not seem to imitate in the test production at all, speakers with shaded cells in Table 3.17 are a subset of those shaded in Table 3.16. In other words, speakers with VOT/F2OT aspirated-lax merger in baseline apparently do not enhance VOT or F2OT in imitation. The converse, though, is not true, as is evident from speakers M05, F09, F13, and F14, who maintain the VOT/F2OT contrast for aspirated-lax stops in their baseline but produce only *f0* enhancement in imitation. A tentative conclusion can be reached that, in spontaneous imitation, speakers adjust the property(s) that they use in their own speech, but not all properties are necessarily adjusted together. And the property that has a primary contrastive role (here, high *f0*) appears to have precedence over a less important property.

Another asymmetry that emerges from the imitative patterns in Table 3.16 is that no participant imitated only the high *f0* stimuli, although several imitated only the long VOT stimuli. For instance, speakers F07 and F12 raised their post-stop *f0* only in the long VOT condition, without imitative changes in the high *f0* condition. Moreover, speakers F04 and M17 seem to have diverged from the model speech in the high *f0* condition but converged in the long VOT condition. I suggest that this asymmetry may be due to the difference in perceptual "atypicality" or saliency of specific variants (Mitterer & Müsseler, 2013; Zellou et al., 2013). As the model speaker was also a young speaker, lengthening VOT of aspirated stops is a less typical enhancement than raising post-stop *f0*. Possibly, then, participants may have perceived /t^h/ with long VOT produced by a young voice as more "atypical" than /t^h/ with high *f0*, resulting in more imitation for the long VOT stimuli.

While speakers' baseline productions seem to be related to their imitative patterns at the individual level, the relation between each speaker's discrimination accuracy and their imitative patterns is less evident. Table 3.18 presents the discrimination accuracy scores of each participant in the two conditions.

	F02	F04	M05	F07	F08	F09	M10	F11	F12	F13	M14	F15	M16	M17	M18	F19	F20	F21	M22
f0	48	55	74	82	72	50	53	52	51	60	86	51	91	49	55	68	73	79	80
VOT	48	54	67	56	84	52	87	59	50	41	79	77	95	58	66	74	78	86	56

Table 3.18. Discrimination accuracy of individual participants for the two manipulation conditions.

In Figures 3.12-3.17, speakers are ordered according to their discrimination accuracy scores for the appropriate property (low to high from left to right), and the ☆ below each speaker indicates a significant imitation. If better performance in the discrimination tasks were to lead to greater probability of imitation, the ☆s should be concentrated to the right side of the plots. For all three measures of the high *f0* condition (Figures 3.12, 3.14, and 3.16), there is no sign of a higher proportion of ☆s on the right side, suggesting that more accurate discrimination of the high *f0* stimuli does not lead to imitation of that property.

For the two duration measures (VOT/F2OT) in the long VOT condition (Figures 3.13 and 3.15), however, there is a slight rightward tendency in the distribution of the ☆s. It is also worth noting that speakers M16 and F21's VOT changes in the long VOT condition (Figure 3.13) were marginally significant (M16: $p = 0.083$; F21: $p = 0.095$). Thus, there is some indication that speakers who are more accurate in discriminating /t^h/s with lengthened VOT from regular /t^h/s are more likely to enhance VOT when they hear /t^h/s with VOT enhancement.

Note that discrimination accuracy is not related to the likelihood of imitating VOT enhancement itself but to the likelihood of enhancing VOT in response to the VOT enhanced stimuli. As evident in Figure 3.17, regardless of their accuracy in discriminating the long VOT stimuli, almost all participants raised their *f0* after hearing the long VOT stimuli. As mentioned in §3.1, all participants in the current study performed at better than chance level in discrimination tasks with both types of manipulated /t^h/ variants, which is arguably the necessary condition to trigger imitation. The production comparisons show that speakers who are more likely to enhance VOT (along with raising post-stop *f0*) as they imitate /t^h/s with VOT enhancement are those who are more accurate in discriminating /t^h/s with lengthened vs. unmanipulated VOT. Together with the findings that stop VOT is a non-primary cue for aspirated stops in Seoul Korean, this arguably suggests that enhancing a non-primary cue is at least weakly related to keen discrimination of the cue.

In sum, analyses of individual data provide evidence that the phonetic property(s) (VOT, *f0*) a Seoul Korean speaker enhances when imitating aspirated stops enhanced by either of the two properties is related to whether the imitator uses the property(s) in her ordinary speech (as seen in baseline productions in this study). Participants with merged VOT/F2OT between aspirated and lax stops did not lengthen their VOT when they imitated enhanced /t^h/. A rather weak relation is also suggested between discriminating a non-primary property of a phonological contrast and enhancing the non-primary property under imitation. The use of the non-primary property in baseline production and relatively accurate discrimination of that property are arguably related to each other. That is, a speaker who maintains a VOT difference between lax and aspirated stops is arguably perceptually more sensitive to VOT cues, and enhances VOT (along with the primary property) more readily in imitating enhanced aspirated stops.

CHAPTER IV

Discussion and Conclusion

The current study investigated Seoul Korean speakers' spontaneous imitation of aspirated stops. Seoul Korean aspirated stops are, as discussed in §1.2, differentiated from stops of other phonation types by at least two distinct acoustic properties, stop VOT and *f0* of the post-stop vowel, with the post-stop high *f0* being the primary cue. This study examined how these primary and non-primary cues for stop aspiration exhibit different imitation patterns, and an investigation of the effects of cue primacy on spontaneous imitation yielded a richer picture than has been presented in the literature as to the role of phonology in the process of spontaneous imitation.

In the imitation experiment, Seoul Korean speakers heard and shadowed model speech that contained aspirated stops manipulated by either raising post-stop *f0* or lengthening VOT. Their realization of these properties in /t^h/, /t/, and /t*/ productions were compared before, during, and after exposure. Although both high *f0* and long VOT induced imitative changes, as summarized in Table 4.1, the results pooled across participants revealed a clear asymmetry between primary and non-primary cues in imitation of Seoul Korean aspirated stops. In addition, separate discrimination tests were conducted to test if participants of the imitation study could discriminate the acoustic manipulations of *f0* and VOT from original unmanipulated productions. The relation among discriminating of a phonetic property, producing that property in ordinary speech production, and using it in imitative enhancement was investigated by analyzing data from individual speakers.

Production block	Phonological category	Manipulation Condition	
		High *f0*	Long VOT
Shadow	Aspirated /t^h/	Increase in *f0* Decrease in VOT	Increase in *f0* No change in VOT
	Sonorant	No change in *f0*	No change in *f0*
Test	Aspirated /t^h/	Increased in *f0* No change in VOT	Increased in *f0* (Increase in VOT)
	Lax /t/	Decrease in *f0* No change in VOT	Decrease in *f0* No change in VOT
	Tense /t*/	No change in *f0* or VOT	Decrease in *f0* (Decrease in VOT)
	Sonorant	Decrease in *f0*	Decrease in *f0*

Table 4.1. Overall imitation patterns, pooled across participants, relative to baseline. Parentheses indicate the effect is marginally significant.

In the following sections, I return to the guiding question of this dissertation—what is the nature of the cognitive representations involved in the process of spontaneous imitation?—and discuss the current findings in relation to the questions and predictions presented in §2.5.

4.1 Review of predictions

4.1.1 What triggers imitation?

With regard to the question of which phonetic properties trigger imitation, it was predicted that both raised post-stop *f0* and extended VOT would trigger imitation if participants reliably discriminate the manipulated stimuli from the original recordings. The results are consistent with this prediction. All participants in this study performed better than chance in the discrimination test for both manipulation types, and both cues induced imitative changes for the majority of participants. (Note that increases in post-stop *f0* after having heard /t^h/ with long VOT are also referred to as imitative changes in this study.) However, as shown in §3.3, further analyses of individual participants revealed that the better performance in the discrimination tests did not lead directly to a greater extent of imitation fidelity.

Because the long VOT stimuli can be perceived as a less typical pattern for the young model speaker, it was also predicted that extended VOT would be especially salient and might therefore induce more robust imitation than raised *f0*, especially in the post-shadowing test productions. Table 4.2 provides comparison of imitative changes in /t^h/ productions in different production blocks of the two manipulation conditions.

Manipulation Condition		High *f0*		Long VOT	
Measurements		VOT (ms)	Post-stop *f0* (Hz) [1]	VOT (ms)	Post-stop *f0* (Hz)
Production blocks	Shadow 1	-6.45 ($p < 0.001$)	3.88 ($p = 0.083$)	N/S	4.99 ($p = 0.044$)
	Shadow 2	-6.08 ($p < 0.001$)	6.03 ($p = 0.004$)	N/S	N/S
	Shadow 3	-5.61 ($p = 0.002$)	6.46 ($p = 0.001$)	N/S	4.93 ($p = 0.045$)
	Test	N/S	6.16 ($p = 0.002$)	2.40 ($p = 0.071$)	9.33 ($p < 0.001$)

Table 4.2. Comparison of changes in /t^h/ relative to baseline in two manipulation conditions, pooled across participants. Values presented are effect sizes (β) from linear mixed models for the 5 production analyses (presented in §3.2.2.1 for the high *f0* condition, and §3.2.3.1 for the long VOT condition).

[1] The values reported here are results of the analysis excluding three outlier speakers. See §3.2.2.1 for more details.

The results for the test productions appear to be consistent with this prediction. Specifically, although post-stop *f0* increased significantly in both manipulation conditions, the effect size was larger in the long VOT condition [β = 9.33] than the high *f0* condition [β = 6.16]. Although post-stop *f0* in the high *f0* stimuli is much higher than that in the long VOT stimuli, participants raised their post-/t^h/ *f0* more after having heard and shadowed the long VOT stimuli than after having heard and shadowed the high *f0* stimuli. In addition, VOT of /t^h/ is not significantly different from the baseline counterpart in either manipulation condition, but in the long VOT condition the increase in VOT trends towards significance with $p = 0.071$.

Turning to the results in the shadowing blocks, comparison of the changes in the two manipulation conditions is less straightforward. In the high *f0* condition, VOT of shadowed /t^h/ decreased with a significant increase in its post-stop *f0*, whereas in the long VOT condition, VOT did not change significantly and post-stop *f0* increased. The increase in post-stop *f0* was greater in the first shadowing block in the long VOT condition than that in the high *f0* condition, but not the second and third blocks of shadowing, which makes a direct comparison of the effect sizes inconclusive.

The decrease in VOT in the high *f0* condition makes comparison between the two manipulation conditions even more convoluted. Prior to making the comparison, we need to understand why speakers decreased their VOT so robustly when shadowing the high *f0* model speech, which had naturally produced VOT. It probably is not a result of participants' diverging from the model speech because it accompanies an increase in post-stop *f0*, which is evidence of a clear convergence. What, then, caused the decrease in stop VOT? One possibility is that participants imitated the model speaker's relatively short VOT. The unmanipulated VOT of the high *f0* stimuli was actually shorter than the most of the participants' baseline productions (model speaker mean = 58.4 ms, participants' baseline mean = 66.2 ms). However, given that the same participants hardly imitated extended VOT (mean = 119.8 ms) in their shadowing productions of the long VOT condition by lengthening VOT, it seems unlikely that they imitated a small difference in one condition but not a much larger difference in another condition, especially when perceptual saliency plays a role in the process of imitation.

The decrease in VOT in the high *f0* condition is more likely a physiological epiphenomenon of the increase in *f0*, rather than an imitative change. This is in line with Narayan and Bowden's (2013) finding that the VOT of Seoul Korean aspirated stops produced in a high *f0* range is shorter than that produced in a low *f0* range, for a given speaker. As pointed out in §1.2.1, this pattern may be due to high *f0* causing the vocal folds to be stiff, reducing the time for the vocal folds to adduct and start voicing, which in turn results in shorter VOT (Narayan & Bowden, 2013; McCrea & Morris, 2005).

Returning to the comparison of the imitation in shadowing blocks of the two enhancement conditions, there is no evidence that the long VOT stimuli facilitated more robust imitation than the high *f0* stimuli did. Unlike the post-shadowing test productions, where it was quite evident that the long VOT stimuli induced greater degree of imitation than the high *f0* stimuli did, no clear patterns emerge for the imitation effects induced by the two sets of stimuli during the shadowing productions.

To sum up, the answer this study provides to the question of *which phonetic properties trigger imitation* is as follows. First, the two types of manipulated /t^h/ variants used in the current study, one with the primary cue enhancement and the other with the secondary cue enhancement, triggered imitative effects in both shadowing and test productions. Second, the imitation effects were larger in the long VOT condition than they were in the high *f0* condition, at least in the post-shadowing test productions. These outcomes suggest that spontaneous imitation is triggered by a phonetic property regardless of its primacy for a phonological contrast, as long as it is sufficiently perceptually salient. In addition, the more "atypical" variant, which is presumably more perceptually salient, seems to induce longer-lasting imitation effects.

These results corroborate previous findings that phonologically less natural or expected variants often lead to more robust or longer-lasting imitation (e.g., Honorof et al., 2011; Mitterer & Müsseler, 2013; Zellou et al., 2013). For instance, as discussed in §1.1.3, Zellou et al. (2013) demonstrate that a less natural coarticulatory pattern induces longer-lasting imitation effects than a more natural pattern: only the imitation of a less natural pattern—a decrease (vs. an increase) in coarticulatory vowel nasality in English words from dense neighborhoods—persisted into a post-shadowing test production. On a related, but slightly different note, less natural hence more salient variants are reported to facilitate greater imitation in immediate shadowing (Honorof et al., 2011; Mitterer & Müsseler, 2013). Allophonically less natural variants (darker /l/ in syllable onset in American English) facilitate greater imitation in shadowing (Honorof et al., 2011); so does a more marked feature clearly indexing a non-standard dialect (Mitterer & Müsseler, 2013). In addition, Mitterer and Müsseler argue that more variation in stimulus presentation also results in more robust imitation in shadowing (see §1.1.3 for a more detailed discussion of these findings).

Adding to these previous findings, the findings of this dissertation suggest that a mismatch with the expectation about the voice may make a variant more salient, and therefore more susceptible to imitation. Younger speakers of Seoul Korean depend mainly on *f0* enhancement to enhance aspirated stops (Kang & Guion, 2008) and the participants in the current study presumably expect (unconsciously) that the young Seoul Korean model speaker would rely on *f0* to enhance his /t^h/. When participants heard /t^h/ with extended VOT, that is, when the actual signal mismatches with their expectation, the atypicality makes the long VOT variant more salient, and thus facilitates longer-lasting, strong imitation effects. Unlike Honorof et al. (2011) and Mitterer and Müsseler (2013), this study does not find a clear difference between the imitation patterns for the two manipulations during shadowing, probably due to considerable variation across participants. The physiological relation between stop VOT and post-stop *f0* also obscures the comparison between the two manipulation conditions. However, the current findings are very much in line with Zellou et al.'s (2013) findings, showing a clear difference between /t^h/ variants with long VOT and those with high *f0* in post-shadowing test production.

4.1.2 Imitative enhancements: what is adjusted?

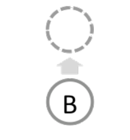
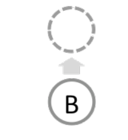
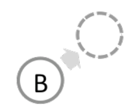
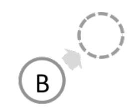
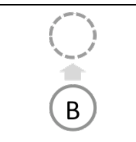
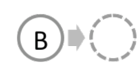
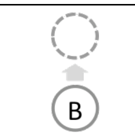
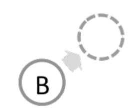
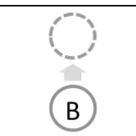
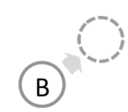
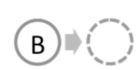
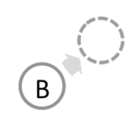
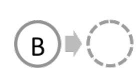
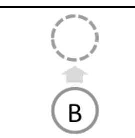
Another question that this dissertation posed is: when a listener detects the enhanced phonetic property, which phonetic property, if any, will the listener-turned-speaker adjust in their subsequent productions? With regard to this question, two distinct hypotheses were offered in §2.5.2. The phonetic imitation hypothesis predicted that the phonetic property manipulated in the model speech will match the property listener/speakers enhance in subsequent productions. The phonological imitation hypothesis predicted that, regardless of the cue manipulated in the model speech, listener/speakers will enhance the phonetic property(s) that they would normally use to enhance the relevant phonological category (in this study, aspirated stops).

Overall, imitation patterns pooled across participants (summarized in Table 4.1) are more consistent with the phonological hypothesis. Table 4.3 recapitulates the relation between cue primacy and the observed imitation patterns. In both the long VOT and high *f0* conditions, post-stop *f0*, which is the primary cue for the relevant target phonological category (i.e., aspirated stops), increased significantly. The phonetic imitation hypothesis cannot explain why participants raised post-stop *f0* (the primary property) after having heard /t^h/ with extended VOT (the non-primary property) without as robust an increase in the property that was manipulated in the stimuli.

	Production Block	Property enhanced in the model speech	
		Primary cue	Non-primary cue
Is the property manipulated in the model speech enhanced in imitative productions?	Shadow	Yes	NO
	Test	Yes	(Yes)
Is the primary cue for the enhanced phonological category in the model speech enhanced in imitative productions?	Shadow	Yes	Yes
	Test	Yes	Yes

Table 4.3. Cue primacy and imitation. Parentheses indicate the effect is marginally significant.

However, the results raise one complication for the phonological imitation hypothesis. Under the phonological imitation hypothesis, irrespective of the enhanced cue in the model speech, participants are predicted to enhance the property or properties they would normally use to enhance the relevant phonological category, resulting in identical imitation patterns in the two conditions. Contrary to this prediction, the results from pooled analyses seem to be different in the two conditions. For example, test productions showed a significant F2OT increase (and a trending VOT increase) in the long VOT condition (reported in §3.2.3.1) but not in the high *f0* condition (reported in §3.2.2.1).

	High *f0* condition	Long VOT condition	Participants
(1)			F02, M05, F09, F13, M14, F19
(2)			M10, F15, M18
(3)			F11, F21
(4)			F04, F07, F12
(5)			M17
(6)			F08
(7)			F16
(8)			M22
(9)			F20

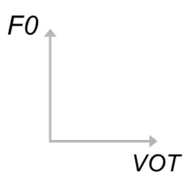


Figure 4.1. Schematization of imitative enhancement patterns in the test production, by individual speakers.
(B) denotes baseline /tʰ/ production on the hypothetical plane of VOT * *f0* shown on the left side of the table. Arrows and dotted circles show the direction of the imitative changes.

The results from individual analyses (reported in §3.3) provide some insights into this asymmetry. Figure 4.1 schematizes the patterns of imitative enhancements shown by individual participants in the test productions of the two manipulation conditions. Note that the schematization in Figure 4.1 focuses only on imitative *enhancements,* ignoring any decrease in either VOT/F2OT or *f0*.

For the nine speakers in (1) and (2) in Figure 4.1, the patterns of imitative enhancements in the two manipulations are identical, suggesting that these speakers are *phonological imitators*. Irrespective of the enhanced cue in the target stimuli, these speakers enhanced the property (only *f0* for the six speakers in (1)) or properties (both *f0* and VOT for the speakers in (2)) they would normally use to enhance the aspirated /t^h/. On the other hand, the two speakers in (3) in Figure 4.1 are *phonetic imitators*, in that they enhance only the cue that is manipulated in the stimuli. The remaining eight speakers in (4)-(8) show more complicated patterns in relation to the two hypotheses. However, except for the one speaker in (9) who is a *non-imitator* in test productions, these speakers' productions do show strong effects on the primary cue even when that cue is not being manipulated. This indicates that the target of speech imitation is not just the detailed acoustic parameters but rather abstract units (such as the phoneme /t^h/ or the natural class of aspirated stops, in this study).

At least for the six speakers in (1), their production and perception grammars would appear to be related, but not identical. Although these speakers increased *f0* to enhance /t^h/ in their own speech, perceptually they must also be sensitive to VOT as information for /t^h/ because otherwise they could not have interpreted extended VOT as a cue for enhanced /t^h/. This production-perception difference is not surprising. As speech is highly variable (perhaps even more so with a recent sound change as in Seoul Korean), a speaker-listener's perception grammar is necessarily more comprehensive than one's own production grammar for effective communication. Speakers need to perceive distinctions among speech patterns that they do not normally produce themselves but others around them produce.

In conclusion, the findings of this dissertation show that speech imitation is neither exclusively *phonetic* nor exclusively *phonological*. In this study, only a few speakers (2 out of 19) showed imitation patterns consistent with pure *phonetic imitation*, and about half of the speakers (9 out of 19) showed those consistent with pure *phonological imitation* in their post-shadowing test productions. We do not yet know what may predispose a speaker to imitate phonetically or phonologically. Nevertheless, this study provides a new insight that cross-individual differences in speech imitation are limited neither to the proclivity to imitate nor to the degree of imitation, but extend to which properties are imitated. Furthermore, different imitative patterns are not necessarily a consequence of speakers' diverging from one another. Instead, different production patterns can emerge as speakers converge using different strategies. In this study, different individuals imitated the same stimuli by enhancing different phonetic properties, and they did so in post-shadowing test productions, suggesting that the effect is perhaps not restricted to short-term memory.

Phonological imitators become more similar to speakers whose speech is being imitated by becoming more different from them in a targeted dimension. This means, corroborating Pardo's (2013) suggestion, that measuring a single phonetic property is not sufficient to assess spontaneous imitation. When perceptual judgments of listeners who are asked to assess imitated speech do not match the acoustic measurements (Pardo, 2013; Pardo et al., 2013), it could be that speakers are imitating phonologically. Phonological categories are signaled by multiple, often co-varying, phonetic properties, and it is possible that a given speaker will prefer to use one cue over another in spontaneous imitation.

Crucially, the process of phonological imitation is governed by language-specific associations between phonological categories and phonetic properties. That is, the same acoustic property could have different phonological significance for speakers of different languages (as is well-known in cross-language speech perception literature) and, therefore, have different impact on the subsequent productions of speakers of different languages. Specifically, the phonological imitation observed in this study, a robust increase in post-stop *f0* after hearing /t^h/ with longer VOT, is not expected for speakers of a language in which *f0* is not a primary cue for the stop phonation types. This calls for further study of this issue (see §4.2 for future study suggestions).

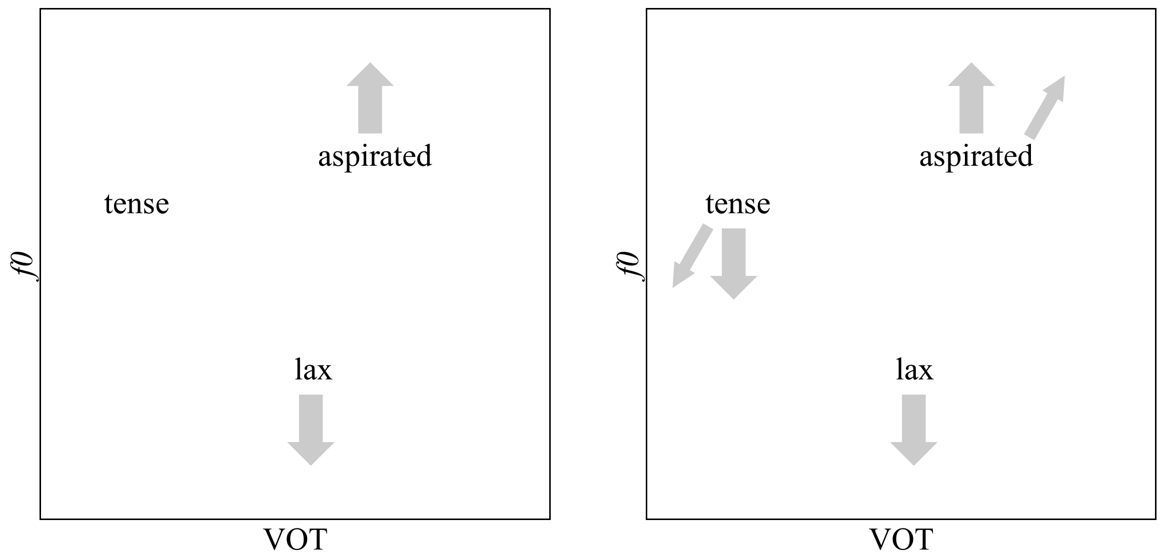
4.1.3 Generalizability of imitation

Another important component of the overarching question of what is the cognitive unit that is responsible for speech imitation is the scope of imitative adjustments. To test the extent to which imitative behavior (whether it is to raise post-stop *f0*, lengthen VOT or both) is generalized, the imitative enhancements in /t^h/-initial, /t/-initial and /t^*/-initial words that were not included in the shadowing list but were in the reading list were examined. Three levels of generalization were predicted: phoneme-level generalization, feature-level generalization, and phonological readjustment (§2.5.3).

First, the current results support phoneme-level generalization, corroborating Nielsen's (2011) finding that English speakers not only imitated /p/ with long VOT but also generalized the imitative behavior to new unheard /p/-initial words. In the current study, the imitative changes in shadowed /t^h/-initial words (raising post-stop *f0* and/or lengthening VOT) are generalized to unheard /t^h/-initial words in test productions, revealing no statistically significant difference between the shadowed and unheard words (see Figures 3.3 and 3.8). In other words, hearing and shadowing specific words three times did not result in greater imitation of those words of exposure, contrary to the effect of token repetition predicted by exemplar models (Goldinger, 1998).

Second, the predictions for feature-level generalization based on the feature system of Halle and Stevens (1971) are not confirmed by the current data. It was predicted that, if lax and aspirated stops share [+spread glottis] and tense and aspirated stops share [+stiff vocal cords], the VOT of aspirated and lax stops should shift together whereas the post-stop *f0* of aspirated and tense stops should shift in tandem. The current results show no enhancement effects (increase in VOT/F2OT or *f0*) for stops other than aspirated stops. Figure 4.2 schematically presents the changes in three stop categories in the test productions of the two manipulation conditions (see also Table 4.1 for a summary of imitative changes).

However, *f0* following lax /t/ as well as sonorant onsets (filler words) decreased in test productions of both manipulation conditions, with a decrease in *f0* following tense /t*/ only in the long VOT condition. This is consistent with the pattern predicted by the most abstract level of generalization, phonological readjustment. Participants readjusted their productions in the direction that maximizes the relevant contrast (i.e., stop aspiration) by lowering *f0* of other phonological categories, which include sonorant-initial filler words. The decrease in *f0* following sonorant onsets was not significant during the shadowing condition (see Table 4.1), suggesting that phonological readjustment is less likely to occur with immediate targets to imitate. Why post-tense-stop *f0* decreased only in the long VOT condition remains unclear. I speculate, however, that because overall imitative effects were more robust in the long VOT condition (as discussed in §4.1.1), the generalization of these imitative effects was more robust as well. This account can also explain the marginally significant decreases in tense stop VOT in the long VOT condition.



(a) High *f0* condition (b) Long VOT condition

Figure 4.2. Schematization of generalization of imitative changes in test productions. Thicker arrows show significant changes relative to baseline, and thinner arrows show marginally significant effects. (a) In the high *f0* condition, aspirated stop *f0* increased and lax stop *f0* decreased. (b) In the long VOT condition, aspirated stop *f0* increased and both lax and tense stop *f0* decreased. VOT of aspirated stops increased marginally, and tense stop VOT decreased marginally.

The discrepancy between the current results and Nielsen's (2011) finding of feature-level generalization may be due to the fact that the relation between Seoul Korean /t^h/ and /t/ is not equivalent to that between English /p/ and /k/, despite the small VOT difference between Seoul Korean /t/ and /t^h/. The current finding of non-generalization at the feature level may indicate that the feature specification that Seoul Korean lax and aspirated stops share [+spread glottis] and tense and aspirated stops share [+stiff vocal cords] (e.g., Halle & Stevens, 1971) needs to be reconsidered. As suggested by Cho et al. (2002), the VOT of /t^h/ and /t/, as well as the *f0* of /t^h/ and /t^*/, may not be featurally related to each other.

Can the phonological readjustments found in this study be interpreted as imitation of a careful speech mode instead of phonologically mediated generalization of imitation? Given the previous finding that imitation effects are generalized even at the sentence level (Kim, 2012), this interpretation cannot be ruled out, although it seems unlikely. The stimuli used in this study do not have characteristics of careful speech other than the enhancements of the two phonetic properties of /t^h/, making it unlikely that the participants heard the stimuli as careful speech in general. The model speech was actually rather fast and casual for laboratory speech, and it was shorter in duration than most participants' baseline productions. As a result, participants imitated shorter word duration as well as enhanced /t^h/, which suggests that participants imitated the signal as a whole but, at the same time, the imitation is, in part, governed by their phonological grammar.

4.2. Implications and suggestions for future study

The results of this dissertation provide evidence that the cognitive representations involved in the process of imitation, which bridges speech perception and production, draw on complex phonological categories such as stop aspiration rather than isolated acoustic properties, such as long VOT or high *f0*. The observed patterns of imitation are mediated by language-specific phonological grammar, and the imitative changes are not limited to a single acoustic cue, phoneme, or feature, but are generalized to maximize the relevant phonological contrasts. These findings have interesting implications, and at the same time pose challenges, for the two theoretical accounts of imitation introduced in §1.1.

First, the finding that words beginning with initial aspirated stops showed the same imitation effects regardless of whether the specific word was shadowed or not poses a challenge to Goldinger's (1998) (among others) exemplar account. According to Goldinger (1998), imitation should increase with more token repetitions, as more exposure to a specific token would increase its activation and thereby contribute to the subsequent production. The current results, however, provide evidence that hearing and shadowing multiple instances of specific words does not result in greater imitation of those words, calling into question the word-specificity of imitation suggested by Goldinger (1998).

The asymmetry between the primary and the non-primary cues in spontaneous imitation found in this study is consistent with an exemplar view that allows abstract linguistic levels (e.g., Pierrehumbert 2001, among others). In exemplar models, phonological categories are clouds of detailed perceptual memory traces. Production patterns governed by phonological associations between different phonetic properties can emerge automatically as a result of hearing ambient speech with the relevant properties. Specifically, if a speaker of Seoul Korean hears instances of /t^h/ with consistently high *f0* and variably long VOT, her category of /t^h/ will have many high *f0* instances and fewer long VOT instances. When she hears an enhanced /t^h/, all the exemplars associated with the enhanced /t^h/ will be activated and contribute to the subsequent production. In this system, the probability that a speaker would use high *f0* /t^h/ or long VOT /t^h/ variants is determined by the proportion of the two variants stored in memory and the associations among those traces. Because linguistic experiences themselves are the phonological categories, governing both the perception and production of speech, the role of the language-specific association between phonetic properties and phonological categories in imitation can be handled easily.

How would phonological imitation be explained in a gestural—in particular, a direct realist—account of imitation? The central claim is that listeners directly perceive vocal tract gestures, and therefore rapidly access articulatory information that can have an immediate impact on subsequent production. The current finding that participants enhance the primary cue for a contrast after having heard the contrast enhanced by a non-primary cue does not follow from a direct realist account of imitation, unless the two phonetic cues are linked at the articulatory level. If the link is purely physiological, then it would follow that we would find it in all languages. Stop VOT and post-stop *f0* are physiologically related, and the link is at the level of laryngeal articulation. However, as discussed in §1.2.1, high post-stop *f0* in Seoul Korean does not appear to be a physiological epiphenomenon of long VOT. This suggests that Seoul Korean speakers' raising *f0* after having heard long VOT /t^h/ is due to language-specific phonological association between the two properties rather than a physiological relation. In other words, perceiving a specific timing between the oral constriction gesture and the glottal opening-and-closing gesture that gives rise to extended VOT leads to a different laryngeal configuration with stiff vocal folds and narrower glottal opening only because the Seoul Korean phonology specifies the association between the two phonetic properties.

This pattern of phonological imitation seems to come across as a challenge to the direct realist account of imitation. According to Honorof et al. (2011), if listeners perceive gestures directly, they are not expected to imitate one gesture using a different gesture. This prediction of direct gestural perception and imitation remains the same regardless of how similar the two gestures' acoustic consequences are or how closely the two gestures are related in terms of phonology. The current finding that listeners substitute (or supplement) long VOT with high *f0* in spontaneous imitation would suggest that abstract linguistic categories mediate what listeners perceive and what listener-turned-speakers produce.

However, phonological imitation is not a threat to the direct realist account of imitation if it is accompanied by longer response latency. The direct perception and imitation of gestures (without intervening abstract category) are predicted during rapid shadowing. Therefore, it is still possible that speech gestures are directly perceived, but in the succeeding production with longer latency, the constellation of co-occurring gestures specified by the phonological grammar changes together. Hypothetically, *phonological imitation* involving gestural substitution or supplement might take longer than direct (*phonetic) imitation* of perceived gestures. If so, the difference between *phonetic imitators* and *phonological imitators* found in this study can be due to the reaction time differences between these speakers. That is, *rapid shadowers* with shorter response latency could be *phonetic imitators* whereas *slow shadowers* could be *phonological imitators,* since the short response latency is evidence for the rapid and direct access to the articulatory information.

The current study was not designed to answer this question. Although the instruction for the shadowing block emphasized "quick" responses (see Appendix B.4), participants were not especially pressed for time during shadowing; they could take as much time as they needed to shadow the word they heard within a 1.5 second inter-stimulus interval (a relatively long time for a shadowing task; c.f., the 180-250 ms latency lag for shadowing reported by Fowler et al. (2003)). This could have contributed to speakers' imitative patterns being considerably more variable in shadowing productions than in test productions in the current study. And again, hypothetically, the difference in shadowing productions could in some way give rise to speakers' imitative patterns in their test productions, resulting in some speakers imitating phonetically and others phonologically. Future testing will be needed to verify whether the difference in reaction time actually results in different patterns (phonetic vs. phonological) of imitation.

Another possible difference between *phonetic imitators* and *phonological imitators* is that individuals are different in their processing styles, which might contribute to different imitation patterns. For example, according to Yu (2010, 2013), individual listeners with different socio-cognitive processing styles exhibit different patterns of perceptual compensation for coarticulation. Future study can ask if phonetic imitators and phonological imitators also differ in their cognitive processing styles.

The results of this dissertation also inspire further investigation of the role of phonology in spontaneous imitation. As mentioned in §4.1.2, phonological imitation predicts that speakers of different languages would imitate the same acoustic manipulation in different ways. That is, the robust increase in post-stop *f0* after hearing /t^h/ with longer VOT observed in this study is not expected for speakers of English, for example, because the primary cue for English voiceless stops is long VOT (e.g., Abramson & Lisker, 1985; but see Whalen, Abramson, Lisker & Mody, 1993, as well). An interesting question to ask is, what happens in imitation in a bilingual setting, if the two languages of a bilingual speaker differ from each other in terms of which phonetic properties are associated with corresponding phonological categories. Evidence from imitation will provide insights into the nature of the relation between two phonetic systems of a bilingual speaker. To that end, the second phase of this project, which builds upon the results of the current dissertation, asks if the two cues for aspiration (long VOT and high *f0*) exhibit different imitation patterns in Seoul Korean-English bilingual speakers' production of English stops. To answer this question, bilingual speakers of Seoul Korean and English were tested with English voiceless stops with long VOT as well as with high *f0*. Preliminary results show that VOT-extended English stimuli do not induce *f0* increase in imitation, which suggests that the bilingual speakers do not associate high *f0* with phonological aspiration (or voicelessness) in English in the same way as they do in Seoul Korean.

Practically, the results of this dissertation suggest that future studies on speech imitation should consider taking more than one phonetic measurement to assess imitation. This is especially so when there exist multiple salient co-varying cues for a phonological category, as in this study. However, arguably all phonological categories have multiple phonetic cues associated with the category, and researchers cannot always be sure which phonetic cues listeners will attend to and listener-turned-speakers will adjust in the process of speech imitation. Therefore, future studies on speech imitation should take into consideration that what might appear to be divergence from the model speaker might rather be phonological convergence along distinct acoustic dimensions.

4.3 Conclusion

This dissertation investigated questions concerning the nature of cognitive representations involved in the process of spontaneous imitation by examining the patterns of imitation facilitated by enhancement of the primary and the secondary cues for aspirated stops in Seoul Korean. Both the primary and the secondary cues are shown to trigger spontaneous imitation, and exposure to an enhanced non-primary cue influences production not only of that property but also of the primary cue for the targeted phonological category. Further, the observed imitative changes are not limited to a single acoustic cue, phoneme, or feature, but are generalized to maximize the relevant phonological contrast. The results suggest that the cognitive bridge between speech perception and production involved in the process of spontaneous imitation is not simply an individual acoustic property, but rather involves abstract phonological categories. This dissertation provides a new insight on spontaneous imitation by demonstrating that production patterns different from the model speech can emerge through spontaneous imitation. Speakers can differ from one another in their imitation strategies (phonetic vs. phonological) and phonological imitation can result in diverging individual phonetic properties.

APPENDIX A

Stimulus List

This section contains a complete list of 150 words (50 /t^h/-initial words, 25 /t/-initial words, 25 /t^*/-initial words, and 50 sonorant-initial fillers) used in the experiment, along with their Korean spellings (Hangeul), IPA transcriptions, English glosses, lexical frequencies (LF), and familiarity scores (FS). Reported frequencies were taken from the NIKL (National Institute of Korean Language) corpus of modern Korean. The familiarity scores were obtained from ten native speakers of Korean who are different individuals from the participants of the main study. They were presented with all disyllabic words beginning with /t^h/, /t/, /t^*/, /m/, /n/, /l/, /w/ and /j/ from the NIKL corpus, and asked to rate the familiarity of the words on a 7-point scale.

Table A.1 presents 50 words (25 /t^h/-initial words and 25 sonorant-initial fillers) contained in both the shadowing list and the reading list. Table A2 presents additional 100 words (25 /t^h/-initial words, 25 /t/-initial words, 25 /t^*/-initial words, and 25 sonorant-initial fillers) that are included only in the reading list.

Type	Hangeul	Transcriptions		English Glosses	LF	FS
		Phonemic	Phonetic			
Shadowed /tʰ/-initial words	타박	/tʰapak/	[tʰabak]	'faultfinding'	49	6.4
	타지	/tʰatɕi/	[tʰadʑi]	'foreign land'	26	6.4
	탁상	/tʰaksaŋ/	[tʰaks*aŋ]	'table'	20	6.4
	탄로	/tʰanlo/	[tʰalo]	'disclosure'	30	6.6
	탄산	/tʰansan/	[tʰansan]	'carbonic (acid), carbonated'	22	6.2
	탄핵	/tʰanhɛk/	[tʰanhɛk ~ tʰanɛk]	'impeachment'	44	6.4
	탈색	/tʰalsɛk/	[tʰals*ɛk]	'bleaching'	28	6.5
	탈옥	/tʰalok/	[tʰaɾok]	'prison break'	31	6.2
	태교	/tʰɛkjo/	[tʰɛgjo]	'prenatal education'	38	6.9
	태백	/tʰɛpɛk/	[tʰɛbɛk]	name of a mountain	37	6.0
	택배	/tʰɛkpɛ/	[tʰɛkp*ɛ]	'parcel delivery service'	25	6.9
	탯줄	/tʰɛstɕul/	[tʰɛtɕ*ul]	'umbilical cord'	46	6.3
	털이	/tʰʌli/	[tʰʌɾi]	'tool for dusting off'	10	6.0
	털옷	/tʰʌlos/	[tʰʌɾot]	'fur coat'	9	6.3
	텃세	/tʰʌsse/	[tʰʌs*ɛ ~ tʰʌts*ɛ]	'territorial imperative'	24	6.1
	톱밥	/tʰoppap/	[tʰop*ap ~ tʰopp*ap]	'sawdust'	19	6.3
	통달	/tʰoŋtal/	[tʰoŋdal]	'mastery'	22	6.1
	통닭	/tʰoŋtalk/	[tʰoŋdak]	'whole chicken'	13	7.0
	통학	/tʰoŋhak/	[tʰoŋhak ~ tʰoŋak]	'commute to school'	49	6.1
	투숙	/tʰusuk/	[tʰusuk]	'lodge/stay in'	43	6.3
	투시	/tʰusi/	[tʰuʃi]	'see through'	27	6.3
	투합	/tʰuhap/	[tʰuhap ~ tʰuap]	'mutual understanding'	5	6.0
	특가	/tʰɨkka/	[tʰɨk*a ~ tʰɨkk*a]	'special bargain price'	8	6.6
	튼실	/tʰɨnsil/	[tʰɨnʃil]	'sturdy, solid'	15	6.4
	틀니	/tʰɨlni/	[tʰɨli ~ tʰɨlni]	'denture'	24	6.5

Table A.1. Korean orthography (Hangeul), IPA transcriptions, English glosses, lexical frequencies (LF), and familiarity scores (FS) of the 25 /tʰ/-initial words included in both the shadowing list and the reading list.

Type	Hangeul	Transcriptions		English Glosses	LF	FS
		Phonemic	Phonetic			
Shadowed sonorant-initial fillers	나래	/nalɛ/	[naɾɛ]	‘wing’	20	6.3
	나루	/nalu/	[naɾu]	‘dock, port’	32	6.4
	날림	/nallim/	[nalim]	‘shoddy’	19	6.6
	남남	/namnam/	[namnam]	‘total strangers’	25	6.7
	내란	/nɛlan/	[nɛɾan]	‘rebellion’	30	6.1
	노망	/nomaŋ/	[nomaŋ]	‘senility’	43	6.3
	녹말	/nokmal/	[noŋmal]	‘starch’	42	6.3
	놀음	/nolɨm/	[noɾɨm]	‘play, gambling’	47	6.4
	누명	/numjʌŋ/	[numjʌŋ]	‘false accusation’	46	6.1
	망언	/maŋʌn/	[maŋʌn]	‘reckless remark’	18	6.4
	맹물	/mɛŋmul/	[mɛŋmul]	‘plain water’	42	6.6
	멀미	/mʌlmi/	[mʌlmi]	‘motion sickness’	30	6.6
	멍울	/mʌŋul/	[mʌŋul]	‘lump’	9	6.2
	목련	/mokljʌn/	[moŋnjʌn ~ moŋljʌn]	‘magnolia’	40	6.5
	몰매	/molmɛ/	[molmɛ]	‘group beating’	28	6.0
	물레	/mulle/	[mule]	‘spinning wheel’	24	6.3
	물엿	/muljʌs/	[muljʌt]	‘starch syrup’	22	6.7
	미완	/miwan/	[miwan]	‘incomplete’	41	6.6
	열무	/jʌlmu/	[jʌlmu]	‘young radish’	29	6.7
	예능	/jenɨŋ/	[jɛnɨŋ]	‘entertainment’	50	6.9
	예매	/jemɛ/	[jɛmɛ]	‘purchase in advance’	30	6.7
	요람	/jolam/	[jolam]	‘cradle’	46	6.2
	용암	/joŋam/	[joŋam]	‘lava’	42	6.0
	원목	/wʌnmok/	[wʌnmok ~ wʌmmok]	‘hardwood’	31	6.3
	원양	/wʌnjaŋ/	[wʌnjaŋ]	‘ocean, pelagic’	16	6.2

Table A.2. Korean orthography (Hangeul), IPA transcriptions, English glosses, lexical frequencies (LF), and familiarity scores (FS) of the 25 sonorant-initial filler words included in both the shadowing list and the reading list.

Type	Hangeul	Transcriptions		English Glosses	LF	FS
		Phonemic	Phonetic			
Unheard /tʰ/-initial words	타국	/tʰakuk/	[tʰaguk]	‘foreign country’	38	6.3
	탄원	/tʰanwʌn/	[tʰanwʌn]	‘petition’	13	6.1
	탄환	/tʰanhwan/	[tʰanhwan ~ tʰanwan]	‘bullet’	25	6.0
	탈모	/tʰalmo/	[tʰalmo]	‘hair loss’	9	6.6
	탈선	/tʰalsʌn/	[tʰals*ʌn]	‘derailment’	47	6.5
	탈수	/tʰalsu/	[tʰals*u]	‘dehydration’	30	6.6
	탈진	/tʰaltɕin/	[tʰaltɕ*in]	‘exhaustion’	43	6.5
	탐라	/tʰamla/	[tʰamla ~ tʰamna]	old name for Jeju Island	7	6.0
	탕진	/tʰaŋtɕin/	[tʰaŋdʑin]	‘squander’	48	6.1
	태만	/tʰɛman/	[tʰɛman]	‘negligence’	25	6.4
	태몽	/tʰɛmoŋ/	[tʰɛmoŋ]	‘conception dream’	10	6.4
	태안	/tʰɛan/	[tʰɛan]	name of a place	20	6.0
	턱뼈	/tʰʌkp*jʌ/	[tʰʌkp*jʌ]	‘jawbone’	7	6.6
	털실	/tʰʌlsil/	[tʰʌlʃil]	‘woolen yarn’	12	6.3
	톱날	/tʰopnal/	[tʰomnal]	‘saw blade’	15	6.2
	톱니	/tʰopni/	[tʰomni]	‘saw tooth’	38	6.4
	통근	/tʰoŋkɨn/	[tʰoŋgɨn]	‘commutation’	14	6.4
	통금	/tʰoŋkɨm/	[tʰoŋgɨm]	‘curfew’	27	6.1
	통뼈	/tʰoŋp*jʌ/	[tʰoŋp*jʌ]	‘big bone’	7	6.2
	투병	/tʰupjʌŋ/	[tʰubjʌŋ]	‘struggle against disease’	30	6.9
	투옥	/tʰuok/	[tʰuok]	‘imprisonment’	41	6.0
	투하	/tʰuha/	[tʰuha ~ tʰua]	‘jettison’	45	6.0
	투혼	/tʰuhon/	[tʰuhon ~ tʰuon]	‘fighting spirit’	33	6.8
	특례	/tʰɨklje/	[tʰɨŋnje ~ tʰɨŋlje]	‘exceptional case’	42	6.6
	특진	/tʰɨktɕin/	[tʰɨktɕ*in]	‘special promotion’	19	6.4

Table A.3. Korean orthography (Hangeul), IPA transcriptions, English glosses, lexical frequencies (LF), and familiarity scores (FS) of the 25 /tʰ/-initial words included only in the reading list.

Type	Hangeul	Transcriptions		English Glosses	LF	FS
		Phonemic	Phonetic			
/t/-initial words	다과	/takwa/	[tagwa]	'refreshments'	10	6.7
	단오	/tano/	[tano]	name of a traditional holiday	42	6.3
	단짝	/tantɕ*ak/	[tantɕ*ak]	'best friend'	32	6.6
	달인	/talin/	[taɾin]	'expert'	13	6.7
	닭살	/talksal/	[taks*al]	'goose bumps'	10	6.6
	담소	/tamso/	[tamso]	'chat'	23	6.4
	대궐	/tɛkwʌl/	[tɛgwʌl]	'(royal) palace'	34	6.3
	대박	/tɛpak/	[tɛbak]	'jackpot'	32	6.8
	대범	/tɛpʌm/	[tɛbʌm]	'generous'	47	6.4
	대졸	/tɛtɕol/	[tɛdʑol]	'college graduate'	47	6.5
	더덕	/tʌtʌk/	[tʌdʌk]	name of a mountain herb	19	6.5
	덧니	/tʌsni/	[tʌnni]	'snaggletooth'	6	6.5
	덧셈	/tʌssem/	[tʌs*ɛm ~ tʌts*ɛm]	'addition'	15	6.8
	덮밥	/tʌpʰpap/	[tʌpp*ap ~ tʌp*ap]	'rice with topping'	17	6.8
	도보	/topo/	[tobo]	'walking'	32	6.7
	도예	/toje/	[tojɛ]	'pottery'	10	6.4
	독사	/toksa/	[toks*a]	'poisonous snake'	34	6.3
	독약	/tokjak/	[togjak]	'poison'	37	6.2
	돌솥	/tolsotʰ/	[tolsot ~ tols*ot]	'stone pot'	13	6.4
	동갑	/toŋkap/	[toŋgap]	'the same age'	31	6.6
	두건	/tukʌn/	[tugʌn]	'hood, headscarf'	12	6.2
	두유	/tuju/	[tuju]	'soymilk'	42	6.2
	들꽃	/tɨlk*otɕʰ/	[tɨlk*ot]	'wild flower'	46	6.5
	듬직	/tɨmtɕik/	[tɨmdʑik]	'reliable'	27	6.5
	등심	/tɨŋsim/	[tɨŋʃim]	'sirloin'	25	6.6

Table A.4. Korean orthography (Hangeul), IPA transcriptions, English glosses, lexical frequencies (LF), and familiarity scores (FS) of the 25 /t/-initial words included only in the reading list.

Type	Hangeul	Transcriptions		English Glosses	LF	FS
		Phonemic	Phonetic			
/t*/-initial words	따귀	/t*akwi/	[t*agwi]	‘cheek’	47	6.1
	따끈	/t*ak*ɨn/	[t*ak*ɨn]	‘steaming’	12	6.9
	따끔	/t*ak*ɨm/	[t*ak*ɨm]	‘stinging’	34	6.5
	따님	/t*anim/	[t*anim]	‘daughter (honorific)’	28	6.2
	땀띠	/t*amt*i/	[t*amt*i]	‘prickly heat’	7	6.5
	땀샘	/t*amsɛm/	[t*amsɛm]	‘sweat gland’	12	6.5
	땅굴	/t*aŋkul/	[t*aŋk*ul]	‘underground tunnel’	21	6.3
	땔감	/t*ɛlkam/	[t*ɛlk*am]	‘firewood’	37	6.6
	땡땡	/t*ɛŋt*ɛŋ/	[t*ɛŋt*ɛŋ]	‘ding-dong’	10	6.0
	땡볕	/t*ɛŋpjʌtʰ/	[t*ɛŋp*jʌt]	‘blazing sun’	44	6.6
	땡전	/t*ɛŋtɕʌn/	[t*ɛŋdʑʌn]	‘single coin, penny’	44	6.4
	떡국	/t*ʌkkuk/	[t*ʌkk*uk ~ t*ʌk*uk]	‘rice-cake soup’	40	6.8
	떡밥	/t*ʌkpap/	[t*ʌkp*ap]	‘paste bait’	5	6.2
	떡잎	/t*ʌkipʰ/	[t*ʌŋnip ~ t*ʌŋip]	‘seed leaf’	9	6.5
	떨기	/t*ʌlki/	[t*ʌlgi]	‘bunch *(of flowers)*’	13	6.1
	뗏목	/t*esmok/	[t*ɛnmok ~ t*ɛmmok]	‘raft’	33	6.3
	똘똘	/t*olt*ol/	[t*olt*ol]	‘brainy’	40	6.5
	똥배	/t*oŋpɛ/	[t*oŋp*ɛ]	‘potbelly’	13	6.5
	뚜벅	/t*upʌk/	[t*ubʌk]	onomatopoeia for walking noise	7	6.4
	뚝딱	/t*ukt*ak/	[t*ukt*ak]	‘clattering noise’	31	6.8
	뚝방	/t*ukpaŋ/	[t*ukp*aŋ]	‘(river) bank’	30	6.1
	뚝심	/t*uksim/	[t*ukʃ*im]	‘perseverance’	26	6.5
	뜨끈	/t*ɨk*ɨn/	[t*ɨk*ɨn]	‘steamy, hot’	12	6.9
	뜨악	/t*ɨak/	[t*ɨak]	onomatopoeia to express shock	25	6.2
	띵동	/t*iŋtoŋ/	[t*iŋtoŋ]	‘ding-dong’	7	6.1

Table A.5. Korean orthography (Hangeul), IPA transcriptions, English glosses, lexical frequencies (LF), and familiarity scores (FS) of the 25 /t*/-initial words included only in the reading list.

Type	Hangeul	Transcriptions		English Glosses	LF	FS
		Phonemic	Phonetic			
Unheard sonorant-initial fillers	나방	/napaŋ/	[nabaŋ]	‘moth’	26	6.2
	남극	/namkɨk/	[namgɨk]	‘the Antarctic’	40	6.5
	냉방	/nɛŋpaŋ/	[nɛŋbaŋ]	‘air conditioning’	20	6.7
	노끈	/nok*ɨn/	[nok*ɨn]	‘string, twine’	29	6.2
	눈금	/nunkɨm/	[nunk*ɨm ~ nuŋk*ɨm]	‘markings *(on a ruler)*’	20	6.5
	눈꽃	/nunk*otɕʰ/	[nunk*ot ~ nuŋk*ot]	‘snowflake’	20	6.6
	눈병	/nunpjʌŋ/	[nunp*jʌŋ ~ nump*jʌŋ]	‘eye disease’	39	6.8
	늦잠	/nɨtɕtɕam/	[nɨtɕ*am ~ nɨttɕ*am]	‘oversleeping’	48	6.9
	말벌	/malpʌl/	[malbʌl]	‘wasp’	6	6.1
	맑음	/malkɨm/	[malgɨm]	‘lucidity’	5	6.6
	맘씨	/mams*i/	[mamʃ*i]	‘nature, intention’	7	6.1
	맨땅	/mɛnt*aŋ/	[mɛnt*aŋ]	‘bare ground’	12	6.0
	모듬	/motɨm/	[modɨm]	‘assorted’	10	6.3
	목발	/mokpal/	[mokp*al]	‘crutch’	34	6.1
	몸값	/momkaps/	[momk*ap ~ moŋk*ap]	‘ransom’	49	6.0
	무적	/mutɕʌk/	[mudʑʌk]	‘invincibility’	29	6.0
	물병	/mulpjʌŋ/	[mulp*jʌŋ]	‘water bottle’	9	6.7
	밑줄	/mitʰtɕul/	[mitɕ*ul ~ mittɕ*ul]	‘underline’	41	6.6
	야경	/jakjʌŋ/	[jagjʌŋ]	‘night scene’	29	6.0
	야식	/jasik/	[jaʃik]	‘late-night snack’	9	6.8
	여울	/jʌul/	[jʌul]	‘rapids *(in a river)*’	33	6.0
	예습	/jesɨp/	[jɛsɨp]	‘preparation for class’	18	6.3
	왕관	/waŋkwan/	[waŋgwan]	‘crown’	31	6.1
	외박	/wɛpak/	[wɛbak]	‘stay out overnight’	43	6.6
	육교	/jukkjo/	[juk*jo ~ jukk*jo]	‘footbridge’	19	6.6

Table A.6. Korean orthography (Hangeul), IPA transcriptions, English glosses, lexical frequencies (LF), and familiarity scores (FS) of the sonorant-initial filler words included only in the reading list.

APPENDIX B

Imitation Experiment – Screens and Instructions

This appendix presents screenshots from the imitation experiment including the experimental instruction screens and a stimuli presentation screen. Figures B.1-B.5 provide Korean instructions that were actually presented to participants during the imitation experiment, along with their English translations. Figure B.6 demonstrates an example screen of visual presentation of a word for the warm-up, baseline, and test blocks.

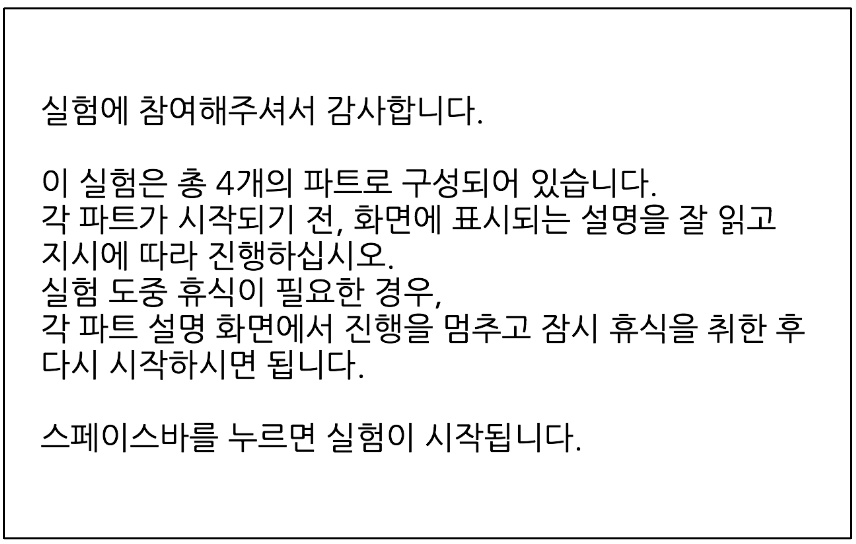


Figure B.1. Introduction

Translation: Thank you for your participation.

This experiment consists of 4 parts.
Before each part, you will see instructions on screen.
Please read and follow the instructions carefully.
If you need a short break during the experiment, you may rest at the instruction screens between parts. Please follow the instructions to resume when you are ready.

Hit the spacebar to continue.

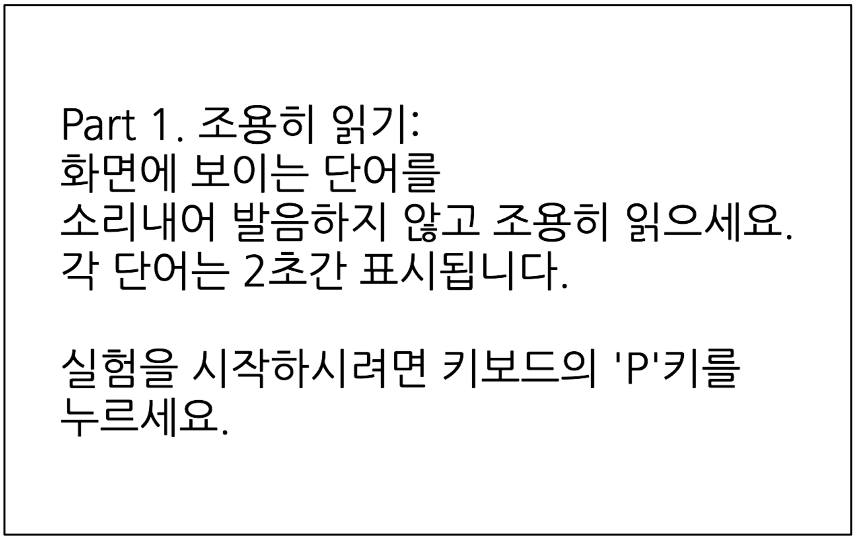


Figure B.2. Instruction for the warm-up block.

Translation: Part 1. Read silently:
Read the words on the screen silently without pronouncing them.
Each word will be shown for 2 seconds.

Hit ‘P’ on the keyboard to continue.

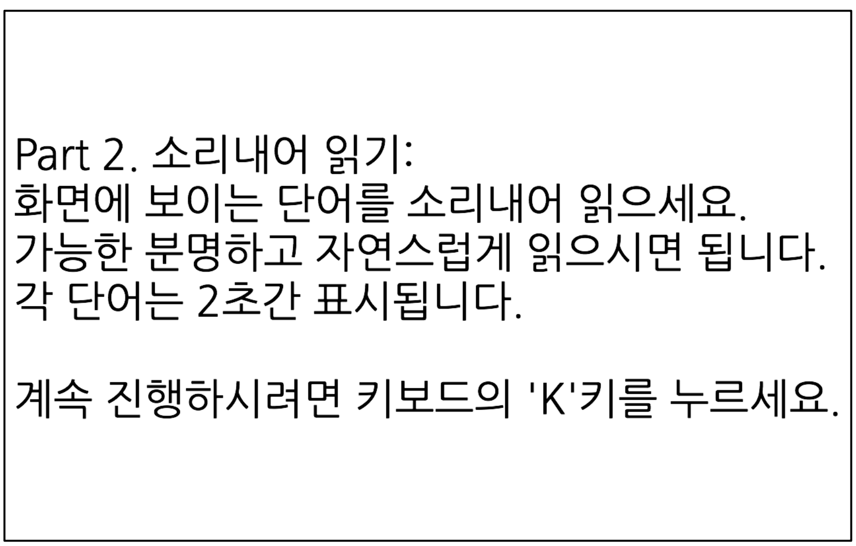


Figure B.3. Instruction for the baseline block.

Translation: Part 2. Read aloud:
Read aloud the words on the screen, as clearly and naturally as possible.
Each word will be shown for 2 seconds.

Hit ‘K’ on the keyboard to continue.

Part 3. 듣고 말하기:
이제 헤드셋을 통해 단어를 듣게 됩니다.
단어를 듣는 즉시 신속하고 정확하게
들리는 단어를 소리내어 말하십시오.

이제 헤드셋을 착용하시고, 준비가 되면
키보드의 'L'키를 누르세요.

Figure B.4. Instruction for the shadowing block.

Translation:

Part 3. Listen and say:
In this part, you will hear spoken words through the headset.
Upon hearing each word, say it aloud quickly and clearly.

Now please put on the headset.
When you are ready, hit ‘L’ on the keyboard to continue.

Part 4. 소리내어 읽기:
이제 헤드셋을 벗으셔도 됩니다.

화면에 보이는 단어를 소리내어 읽으세요.
가능한 분명하고 자연스럽게 읽으시면 됩니다.
각 단어는 2초간 표시됩니다.

계속 진행하시려면 키보드의 'P'키를 누르세요.

Figure B.5. Instruction for the test block.

Translation:

Part 4. Read aloud:
You can take off the headset now.

Read aloud the words on the screen as clearly and naturally as possible.
Each word will be shown for 2 seconds.

Hit ‘P’ on the keyboard to continue.

Figure B.6. An example screen of visual presentation of stimuli.

APPENDIX C

Statistical Results for Individual Speakers

This appendix provides the results for individual speakers. Data for each speaker include (1) VOT, F2OT and post stop *f0* of /t^h/, /t/, and /t*/ in their baseline productions, (2) imitative changes in VOT, F2OT and post stop *f0* of /t^h/, /t/, and /t*/ in baseline and test productions of the two experimental conditions (high *f0* and long VOT), and (3) imitative changes in VOT, F2OT and post stop *f0* of /t^h/ in five production blocks (baseline, shadowing 1, 2, 3, and test) of the two experimental conditions (high *f0* and long VOT).

VOT, F2OT and post-stop *f0* of /t^h/, /t/, and /t*/ in baseline productions were analyzed by performing three separate one-way ANOVAs (dependent variable: VOT, F2OT, post-stop *f0*, respectively for each model; independent variable: stop type) for each speaker. All ANOVAs showed significant effects of stop types on VOT, F2OT, and post-stop *f0*, so post-hoc Tukey tests were performed for pair-wise comparisons. Imitative changes for each speaker were analyzed by linear mixed effects models. For details of statistical modeling of imitative changes, see §3.3.

F02

Measure	ANOVA	$/t^h/$	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 203.2, *p* < 0.0001	63.29	63.79			0.9794
		63.29		**14.05**		**< 0.0001**
			63.79	**14.05**		**< 0.0001**
F2OT (ms.)	F(2,197) = 214.3, *p* < 0.0001	71.02	70.60			0.9867
		71.02		**18.19**		**< 0.0001**
			70.60	**18.19**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 435.8, *p* < 0.0001	**240.27**	**183.58**			**< 0.0001**
		240.27		**229.12**		**< 0.0001**
		240.27			**184.24**	**< 0.0001**
			183.58	**229.12**		**< 0.0001**
			183.58		184.24	0.9906
				229.12	**184.24**	**< 0.0001**

Table C.1. Measurements and statistics for baseline production of speaker F02. **Bolded** data are significant at *p* < 0.05.

		Base	Test	*t*	*p*
VOT (ms.)	$/t^h/$	89.58	86.77	-1.42	0.1590
	/t/	89.92	90.58	0.20	0.8460
	/t*/	39.32	38.01	-0.38	0.7030
F2OT (ms.)	$/t^h/$	99.03	96.76	-1.08	0.2830
	/t/	99.06	101.78	0.76	0.4520
	/t*/	45.63	43.00	-0.73	0.4690
Post-onset *f0* (Hz)	**$/t^h/$**	**245.60**	**253.12**	**4.79**	**< 0.0001**
	/t/	193.52	192.80	-0.27	0.7916
	/t*/	**235.56**	**228.88**	**-2.46**	**0.0152**
	filler	193.34	189.86	-1.57	0.1192

		Base	Test	*t*	*p*
VOT (ms.)	$/t^h/$	88.98	88.61	-0.17	0.8637
	/t/	83.26	81.95	-0.36	0.7224
	/t*/	39.32	38.59	-0.20	0.8433
F2OT (ms.)	$/t^h/$	101.43	100.07	-0.57	0.5714
	/t/	94.02	97.66	0.89	0.3734
	/t*/	48.63	48.60	-0.01	0.9947
Post-onset *f0* (Hz)	**$/t^h/$**	**235.14**	**241.72**	**2.97**	**0.0035**
	/t/	173.64	175.30	0.43	0.6657
	/t*/	222.68	222.54	-0.04	0.9709
	filler	**175.34**	**169.10**	**-1.99**	**0.0481**

Table C.2. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F02. **Bolded** data are significant at *p* < 0.05.

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	93.57		
	Sh1	**85.52**	**-2.97**	**0.0035**
	Sh2	**84.83**	**-3.13**	**0.0021**
	Sh3	**81.24**	**-4.22**	**< 0.0001**
	Test	90.58	-1.54	0.1252
/tʰ/ F2OT (ms.)	Base	99.55		
	Sh1	94.99	-1.69	0.0925
	Sh2	**92.23**	**-2.65**	**0.0091**
	Sh3	**87.82**	**-4.05**	**0.0001**
	Test	97.25	-1.21	0.2301
Post-stop *f0* (Hz)	Base	245.60		
	Sh1	**255.45**	**4.95**	**< 0.0001**
	Sh2	**254.73**	**4.59**	**< 0.0001**
	Sh3	**255.61**	**5.03**	**< 0.0001**
	Test	**253.12**	**4.89**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	193.34		
	Sh1	193.59	0.09	0.9283
	Sh2	**200.75**	**2.70**	**0.0075**
	Sh3	**199.23**	**2.14**	**0.0330**
	Test	189.86	-1.60	0.1107

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	101.66		
	Sh1	102.54	0.27	0.7910
	Sh2	103.27	0.49	0.6280
	Sh3	101.91	0.08	0.9390
	Test	102.14	0.19	0.8520
/tʰ/ F2OT (ms.)	Base	114.57		
	Sh1	117.27	0.72	0.4740
	Sh2	115.85	0.34	0.7330
	Sh3	113.27	-0.35	0.7270
	Test	114.14	-0.15	0.8820
Post-stop *f0* (Hz)	Base	235.14		
	Sh1	**251.62**	**7.48**	**< 0.0001**
	Sh2	**243.46**	**3.78**	**0.0002**
	Sh3	**246.50**	**5.16**	**< 0.0001**
	Test	**241.72**	**3.89**	**0.0001**
Post-sonorant *f0* (Hz)	Base	175.34		
	Sh1	170.31	-1.64	0.1017
	Sh2	**185.35**	**3.27**	**0.0012**
	Sh3	**184.15**	**2.88**	**0.0043**
	Test	**169.10**	**-2.61**	**0.0096**

Table C.3. Aspirated /t^h/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F02. **Bolded** data are significant at $p < 0.05$.

F04

Measure	ANOVA	$/t^h/$	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	$F(2,196) = 270.4$ $p < 0.0001$	79.75	75.82			0.4129
		79.75		**11.25**		**< 0.0001**
			75.82	**11.25**		**< 0.0001**
F2OT (ms.)	$F(2,196) = 297.9$ $p < 0.0001$	85.55	81.24			0.3508
		85.55		**13.19**		**< 0.0001**
			81.24	**13.19**		**< 0.0001**
Post-onset *f0* (Hz)	$F(3,294) = 660.2$ $p < 0.0001$	**285.37**	**195.88**			**< 0.0001**
		285.37		**261.29**		**< 0.0001**
		285.37			**191.44**	**< 0.0001**
			195.88	**261.29**		**< 0.0001**
			195.88		191.44	0.4245
				261.29	**191.44**	**< 0.0001**

Table C.4. Measurements and statistics for baseline production of speaker F04. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**$/t^h/$**	**105.65**	**100.01**	**-2.57**	**0.0116**
	/t/	95.57	99.78	1.18	0.2408
	/t*/	34.46	39.22	1.35	0.1800
F2OT (ms.)	**$/t^h/$**	**110.77**	**105.59**	**-2.21**	**0.0295**
	/t/	101.72	106.86	1.34	0.1822
	/t*/	37.25	41.62	1.16	0.2507
Post-onset *f0* (Hz)	**$/t^h/$**	**290.42**	**285.34**	**-2.27**	**0.0247**
	/t/	199.62	203.48	1.00	0.3214
	/t*/	**266.28**	**276.12**	**2.54**	**0.0123**
	filler	**191.70**	**203.20**	**3.66**	**0.0004**

		Base	Test	*t*	*p*
VOT (ms.)	$/t^h/$	107.04	109.45	1.17	0.2455
	/t/	100.81	100.45	-0.11	0.9138
	/t*/	38.49	33.18	-1.58	0.1171
F2OT (ms.)	$/t^h/$	115.60	117.49	0.96	0.3394
	/t/	107.56	108.56	0.31	0.7561
	/t*/	42.43	39.08	-1.04	0.2996
Post-onset *f0* (Hz)	**$/t^h/$**	**279.96**	**307.46**	**13.09**	**< 0.0001**
	/t/	**192.12**	**166.54**	**-7.03**	**< 0.0001**
	/t*/	256.72	251.62	-1.40	0.1630
	filler	**190.82**	**161.26**	**-9.95**	**< 0.0001**

Table C.5. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F04. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
/t^h/ VOT (ms.)	Base	111.68		
	Sh1	**97.69**	**-3.85**	**0.0002**
	Sh2	**97.84**	**-3.84**	**0.0002**
	Sh3	**91.82**	**-5.05**	**< 0.0001**
	Test	**105.37**	**-2.50**	**0.0136**
/t^h/ F2OT (ms.)	Base	114.66		
	Sh1	**102.26**	**-3.33**	**0.0011**
	Sh2	**102.29**	**-3.34**	**0.0011**
	Sh3	**97.33**	**-4.27**	**< 0.0001**
	Test	**109.04**	**-2.18**	**0.0313**
Post-stop *f0* (Hz)	Base	290.44		
	Sh1	**265.22**	**-8.24**	**< 0.0001**
	Sh2	**265.87**	**-8.15**	**< 0.0001**
	Sh3	**269.99**	**-6.78**	**< 0.0001**
	Test	**285.34**	**-2.17**	**0.0308**
Post-sonorant *f0* (Hz)	Base	191.72		
	Sh1	**217.54**	**6.12**	**< 0.0001**
	Sh2	**215.69**	**5.72**	**< 0.0001**
	Sh3	**215.37**	**5.65**	**< 0.0001**
	Test	**203.23**	**3.49**	**0.0006**

	Block	Estimate	*t*	*p*
/t^h/ VOT (ms.)	Base	114.47		
	Sh1	**89.56**	**-6.29**	**< 0.0001**
	Sh2	**96.20**	**-4.80**	**< 0.0001**
	Sh3	109.68	-1.35	0.1796
	Test	116.28	0.74	0.4610
/t^h/ F2OT (ms.)	Base	120.30		
	Sh1	**97.63**	**-6.08**	**< 0.0001**
	Sh2	**104.46**	**-4.42**	**< 0.0001**
	Sh3	114.40	-1.79	0.0760
	Test	121.82	0.67	0.5051
Post-stop *f0* (Hz)	Base	279.96		
	Sh1	274.92	-1.80	0.0734
	Sh2	**274.24**	**-2.04**	**0.0424**
	Sh3	274.80	-1.84	0.0669
	Test	**307.46**	**12.81**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	190.82		
	Sh1	196.04	1.33	0.1834
	Sh2	196.36	1.42	0.1581
	Sh3	196.88	1.55	0.1227
	Test	**161.26**	**-9.74**	**< 0.0001**

Table C.6. Aspirated /t^h/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F04. **Bolded** data are significant at $p < 0.05$.

M05

Measure	ANOVA	/t^h/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 222.8 *p* < 0.0001	**59.83**	**48.58**			**< 0.0001**
		59.83		**14.32**		**< 0.0001**
			48.58	**14.32**		**< 0.0001**
F2OT (ms.)	F(2,197) = 220.6 *p* < 0.0001	**67.49**	**56.72**			**< 0.0001**
		67.49		**19.26**		**< 0.0001**
			56.72	**19.26**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,294) = 267.4 *p* < 0.0001	**145.89**	**124.49**			**< 0.0001**
		145.89		**137.04**		**< 0.0001**
		145.89			**118.41**	**< 0.0001**
			124.49	**137.04**		**< 0.0001**
			124.49		**118.41**	**< 0.0001**
				137.04	**118.41**	**< 0.0001**

Table C.7. Measurements and statistics for baseline production of speaker M05. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	/t^h/	78.12	80.99	1.57	0.1188
	/t/	68.41	70.40	0.64	0.5271
	/t*/	35.40	31.12	-1.37	0.1741
F2OT (ms.)	/t^h/	88.75	91.44	1.36	0.1781
	/t/	81.23	80.66	-0.17	0.8690
	/t*/	44.08	39.88	-1.23	0.2223
Post-onset *f0* (Hz)	**/t^h/**	**147.86**	**160.36**	**11.15**	**< 0.0001**
	/t/	**127.88**	**114.66**	**-6.74**	**< 0.0001**
	/t*/	137.91	138.10	0.10	0.9250
	filler	**118.20**	**107.30**	**-6.88**	**< 0.0001**

		Base	Test	*t*	*p*
VOT (ms.)	/t^h/	98.37	99.53	0.54	0.5910
	/t/	83.42	81.61	-0.57	0.5720
	/t*/	52.36	47.84	-1.41	0.1620
F2OT (ms.)	/t^h/	109.78	110.81	0.47	0.6430
	/t/	92.85	91.57	-0.39	0.6994
	/t*/	60.50	56.24	-1.28	0.2046
Post-onset *f0* (Hz)	**/t^h/**	**143.94**	**152.24**	**6.17**	**< 0.0001**
	/t/	**121.28**	**114.06**	**-3.10**	**0.0023**
	/t*/	**136.24**	**131.46**	**-2.05**	**0.0421**
	filler	**118.64**	**108.64**	**-5.26**	**< 0.0001**

Table C.8. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker M05. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	79.40		
	Sh1	**68.05**	**-4.31**	**< 0.0001**
	Sh2	**73.78**	**-2.14**	**0.0343**
	Sh3	78.19	-0.46	0.6450
	Test	82.33	1.46	0.1470
/tʰ/ F2OT (ms.)	Base	89.19		
	Sh1	**77.17**	**-4.38**	**< 0.0001**
	Sh2	**83.21**	**-2.18**	**0.0311**
	Sh3	87.79	-0.51	0.6116
	Test	91.92	1.30	0.1958
Post-stop *f0* (Hz)	Base	147.86		
	Sh1	**157.70**	**7.80**	**< 0.0001**
	Sh2	**164.34**	**13.06**	**< 0.0001**
	Sh3	**160.82**	**10.27**	**< 0.0001**
	Test	**160.36**	**12.92**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	118.20		
	Sh1	**109.08**	**-5.18**	**< 0.0001**
	Sh2	**102.16**	**-9.12**	**< 0.0001**
	Sh3	**106.00**	**-6.93**	**< 0.0001**
	Test	**107.30**	**-7.96**	**< 0.0001**

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	89.94		
	Sh1	84.33	-1.55	0.1236
	Sh2	83.09	-1.90	0.0593
	Sh3	89.69	-0.07	0.9426
	Test	92.20	0.92	0.3588
/tʰ/ F2OT (ms.)	Base	96.79		
	Sh1	91.01	-1.54	0.1258
	Sh2	91.09	-1.52	0.1307
	Sh3	97.93	0.32	0.7464
	Test	99.57	1.09	0.2784
Post-stop *f0* (Hz)	Base	143.94		
	Sh1	**150.65**	**4.20**	**< 0.0001**
	Sh2	**150.38**	**3.97**	**0.0001**
	Sh3	**149.53**	**3.50**	**0.0006**
	Test	**152.24**	**6.75**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	118.64		
	Sh1	**108.24**	**-4.68**	**< 0.0001**
	Sh2	**110.83**	**-3.49**	**0.0006**
	Sh3	**110.80**	**-3.53**	**0.0005**
	Test	**108.64**	**-5.75**	**< 0.0001**

Table C.9. Aspirated /tʰ/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker M05. **Bolded** data are significant at $p < 0.05$.

F07

Measure	ANOVA	/tʰ/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 261.2, $p < 0.0001$	61.99	60.34			0.7186
		61.99		**15.74**		**< 0.0001**
			60.34	**15.74**		**< 0.0001**
F2OT (ms.)	F(2,197) = 283.6, $p < 0.0001$	67.47	66.77			0.9452
		67.47		**18.30**		**< 0.0001**
			66.77	**18.30**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 363.5, $p < 0.0001$	**256.28**	**203.80**			**< 0.0001**
		256.28		**242.02**		**< 0.0001**
		256.28			**200.12**	**< 0.0001**
			203.80	**242.02**		**< 0.0001**
			203.80		200.12	0.3891
				242.02	**200.12**	**< 0.0001**

Table C.10. Measurements and statistics for baseline production of speaker F07. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	/tʰ/	70.28	68.33	-1.17	0.2433
	/t/	68.21	68.22	0.01	0.9963
	/t*/	28.54	30.07	0.54	0.5927
F2OT (ms.)	/tʰ/	78.30	76.29	-1.21	0.2290
	/t/	77.02	76.77	-0.09	0.9320
	/t*/	32.98	34.64	0.58	0.5630
Post-onset *f0* (Hz)	/tʰ/	266.58	267.80	0.96	0.3412
	/t/	206.40	205.26	-0.52	0.6072
	/t*/	249.04	251.94	1.31	0.1921
	filler	**201.82**	**197.98**	**-2.13**	**0.0353**

		Base	Test	*t*	*p*
VOT (ms.)	**/tʰ/**	**89.50**	**83.49**	**-4.03**	**0.0001**
	/t/	88.44	91.44	1.17	0.2465
	/t*/	**41.03**	**46.96**	**2.31**	**0.0232**
F2OT (ms.)	**/tʰ/**	**94.70**	**88.27**	**-4.19**	**0.0001**
	/t/	94.50	97.89	1.28	0.2051
	/t*/	**44.08**	**50.27**	**2.33**	**0.0217**
Post-onset *f0* (Hz)	**/tʰ/**	**247.18**	**252.38**	**2.17**	**0.0317**
	/t/	201.20	196.68	-1.09	0.2782
	/t*/	**235.00**	**226.28**	**-2.10**	**0.0375**
	filler	**199.62**	**193.28**	**-1.87**	**0.0635**

Table C.11. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F07. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	68.32		
	Sh1	66.46	-0.84	0.4015
	Sh2	64.30	-1.83	0.0699
	Sh3	**62.13**	**-2.81**	**0.0058**
	Test	66.43	-1.13	0.2624
/tʰ/ F2OT (ms.)	Base	73.46		
	Sh1	70.05	-1.59	0.1153
	Sh2	**68.91**	**-2.12**	**0.0359**
	Sh3	**68.63**	**-2.25**	**0.0267**
	Test	71.60	-1.13	0.2607
Post-stop *f0* (Hz)	Base	266.58		
	Sh1	**277.26**	**7.03**	**< 0.0001**
	Sh2	**272.74**	**4.05**	**0.0001**
	Sh3	**270.26**	**2.42**	**0.0162**
	Test	267.80	1.05	0.2937
Post-sonorant *f0* (Hz)	Base	201.82		
	Sh1	**197.63**	**-1.97**	**0.0498**
	Sh2	198.35	-1.63	0.1039
	Sh3	198.11	-1.75	0.0822
	Test	**197.98**	**-2.34**	**0.0200**

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	87.16		
	Sh1	83.48	-1.61	0.1108
	Sh2	**80.02**	**-3.10**	**0.0024**
	Sh3	**79.49**	**-3.35**	**0.0011**
	Test	**81.19**	**-3.45**	**0.0008**
/tʰ/ F2OT (ms.)	Base	95.32		
	Sh1	**90.47**	**-2.05**	**0.0425**
	Sh2	**87.16**	**-3.43**	**0.0008**
	Sh3	**86.36**	**-3.79**	**0.0002**
	Test	**88.88**	**-3.62**	**0.0004**
Post-stop *f0* (Hz)	Base	247.18		
	Sh1	**257.09**	**5.78**	**< 0.0001**
	Sh2	**255.09**	**4.61**	**< 0.0001**
	Sh3	**257.57**	**6.06**	**< 0.0001**
	Test	**252.38**	**3.94**	**0.0001**
Post-sonorant *f0* (Hz)	Base	199.62		
	Sh1	**191.66**	**-3.34**	**0.0010**
	Sh2	**193.30**	**-2.65**	**0.0085**
	Sh3	**191.46**	**-3.43**	**0.0007**
	Test	**193.28**	**-3.40**	**0.0008**

Table C.12. Aspirated /tʰ/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F07. **Bolded** data are significant at $p < 0.05$.

F08

Measure	ANOVA	/tʰ/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 187.2 $p < 0.0001$	**61.98**	**49.84**			**< 0.0001**
		61.98		**13.75**		**< 0.0001**
			49.84	**13.75**		**< 0.0001**
F2OT (ms.)	F(2,197) = 194.7 $p < 0.0001$	**70.26**	**58.90**			**0.0001**
		70.26		**18.40**		**< 0.0001**
			58.90	**18.40**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 459.4 $p < 0.0001$	**252.92**	**206.96**			**< 0.0001**
		252.92		**235.56**		**< 0.0001**
		252.92			**199.40**	**< 0.0001**
			206.96	**235.56**		**< 0.0001**
			206.96		**199.40**	**0.0005**
				235.56	**199.40**	**< 0.0001**

Table C.13. Measurements and statistics for baseline production of speaker F08. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**/tʰ/**	**91.71**	**86.57**	**-2.66**	**0.0091**
	/t/	78.09	82.66	1.37	0.1738
	/t*/	42.82	46.87	1.21	0.2283
F2OT (ms.)	**/tʰ/**	**102.32**	**97.59**	**-2.31**	**0.0232**
	/t/	88.28	92.45	1.18	0.2415
	/t*/	48.96	54.48	1.56	0.1222
Post-onset *f0* (Hz)	**/tʰ/**	**258.81**	**266.11**	**5.73**	**< 0.0001**
	/t/	**210.76**	**204.82**	**-2.69**	**0.0079**
	/t*/	240.60	238.18	-1.10	0.2745
	filler	**199.93**	**192.55**	**-4.10**	**0.0001**

		Base	Test	*t*	*p*
VOT (ms.)	**/tʰ/**	**94.18**	**101.75**	**3.44**	**0.0008**
	/t/	83.71	82.12	-0.42	0.6750
	/t*/	**49.92**	**39.00**	**-2.89**	**0.0048**
F2OT (ms.)	**/tʰ/**	**105.08**	**114.89**	**4.74**	**< 0.0001**
	/t/	95.83	92.50	-0.94	0.3490
	/t*/	**58.12**	**43.63**	**-4.08**	**0.0001**
Post-onset *f0* (Hz)	**/tʰ/**	**246.81**	**257.35**	**6.53**	**< 0.0001**
	/t/	**203.16**	**196.22**	**-2.48**	**0.0142**
	/t*/	**230.52**	**224.58**	**-2.12**	**0.0354**
	filler	**198.65**	**188.69**	**-4.36**	**< 0.0001**

Table C.14. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F08. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	106.81		
	Sh1	100.84	-1.95	0.0536
	Sh2	**99.87**	**-2.26**	**0.0258**
	Sh3	**94.97**	**-3.83**	**0.0002**
	Test	**101.26**	**-2.39**	**0.0183**
$/t^h/$ F2OT (ms.)	Base	118.90		
	Sh1	**110.03**	**-2.81**	**0.0058**
	Sh2	**109.41**	**-3.00**	**0.0033**
	Sh3	**105.78**	**-4.12**	**0.0001**
	Test	113.70	-2.17	0.0319
Post-stop *f0* (Hz)	Base	258.81		
	Sh1	**268.37**	**4.35**	**< 0.0001**
	Sh2	**263.53**	**2.15**	**0.0325**
	Sh3	258.97	0.08	0.9402
	Test	**266.11**	**4.33**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	199.93		
	Sh1	**192.93**	**-2.29**	**0.0229**
	Sh2	194.77	-1.69	0.0929
	Sh3	201.93	0.65	0.5139
	Test	**192.55**	**-3.09**	**0.0022**

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	88.00		
	Sh1	**75.69**	**-3.72**	**0.0003**
	Sh2	84.10	-1.19	0.2365
	Sh3	85.13	-0.87	0.3877
	Test	**95.80**	**3.18**	**0.0019**
$/t^h/$ F2OT (ms.)	Base	98.76		
	Sh1	**88.03**	**-3.25**	**0.0015**
	Sh2	97.95	-0.25	0.8042
	Sh3	96.88	-0.57	0.5678
	Test	**108.83**	**4.14**	**0.0001**
Post-stop *f0* (Hz)	Base	246.81		
	Sh1	**267.59**	**10.41**	**< 0.0001**
	Sh2	**266.63**	**9.93**	**< 0.0001**
	Sh3	**266.91**	**10.07**	**< 0.0001**
	Test	**257.35**	**6.87**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	198.65		
	Sh1	**191.95**	**-2.42**	**0.0163**
	Sh2	**192.83**	**-2.10**	**0.0366**
	Sh3	**188.63**	**-3.62**	**0.0004**
	Test	**188.69**	**-4.59**	**< 0.0001**

Table C.15. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F08. **Bolded** data are significant at $p < 0.05$.

F09

Measure	ANOVA	/tʰ/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 215.2 $p < 0.0001$	**63.98**	**49.54**			**< 0.0001**
		63.98		**13.86**		**< 0.0001**
			49.54	**13.86**		**< 0.0001**
F2OT (ms.)	F(2,197) = 257.7 $p < 0.0001$	**70.62**	**59.77**			**< 0.0001**
		70.62		**16.14**		**< 0.0001**
			59.77	**16.14**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 519.8 $p < 0.0001$	**249.45**	**204.08**			**< 0.0001**
		249.45		**239.58**		**< 0.0001**
		249.45			**198.27**	**< 0.0001**
			204.08	**239.58**		**< 0.0001**
			204.08		**198.27**	**0.0063**
				239.58	**198.27**	**< 0.0001**

Table C.16. Measurements and statistics for baseline production of speaker F09. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	/tʰ/	72.05	69.13	-1.92	0.0578
	/t/	51.88	57.02	1.98	0.0510
	/t*/	22.31	24.82	0.97	0.3349
F2OT (ms.)	/tʰ/	78.12	78.94	0.54	0.5920
	/t/	62.09	62.39	0.11	0.9110
	/t*/	23.75	22.30	-0.56	0.5800
Post-onset *f0* (Hz)	**/tʰ/**	**257.42**	**268.08**	**8.18**	**< 0.0001**
	/t/	204.00	200.98	-1.34	0.1831
	/t*/	**243.80**	**236.18**	**-3.38**	**0.0009**
	filler	**200.66**	**194.72**	**-3.22**	**0.0016**

		Base	Test	*t*	*p*
VOT (ms.)	/tʰ/	87.33	88.83	0.77	0.4408
	/t/	72.93	70.43	-0.75	0.4573
	/t*/	33.44	30.79	-0.79	0.4299
F2OT (ms.)	/tʰ/	95.77	96.52	0.38	0.7045
	/t/	84.11	79.85	-1.24	0.2186
	/t*/	37.95	33.85	-1.19	0.2363
Post-onset *f0* (Hz)	**/tʰ/**	**242.81**	**255.53**	**7.92**	**< 0.0001**
	/t/	**204.16**	**190.04**	**-5.08**	**< 0.0001**
	/t*/	235.36	234.33	-0.37	0.7151
	filler	**197.21**	**185.77**	**-5.04**	**< 0.0001**

Table C.17. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F09. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	86.03		
	Sh1	**79.63**	**-2.57**	**0.0112**
	Sh2	**73.30**	**-5.10**	**< 0.0001**
	Sh3	**70.70**	**-6.12**	**< 0.0001**
	Test	**82.48**	**-1.99**	**0.0490**
$/t^h/$ F2OT (ms.)	Base	91.86		
	Sh1	**83.70**	**-3.15**	**0.0020**
	Sh2	**80.95**	**-4.17**	**0.0001**
	Sh3	**78.01**	**-5.27**	**< 0.0001**
	Test	91.96	0.05	0.9575
Post-stop *f0* (Hz)	Base	257.42		
	Sh1	**266.45**	**4.13**	**< 0.0001**
	Sh2	**267.77**	**4.73**	**< 0.0001**
	Sh3	**276.05**	**8.52**	**< 0.0001**
	Test	**268.08**	**6.35**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	200.66		
	Sh1	196.82	-1.26	0.2090
	Sh2	196.78	-1.27	0.2040
	Sh3	**187.54**	**-4.31**	**< 0.0001**
	Test	**194.72**	**-2.50**	**0.0130**

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	96.57		
	Sh1	**91.14**	**-2.00**	**0.0479**
	Sh2	**89.63**	**-2.56**	**0.0116**
	Sh3	**86.51**	**-3.67**	**0.0004**
	Test	97.84	0.62	0.5341
$/t^h/$ F2OT (ms.)	Base	105.84		
	Sh1	100.73	-1.81	0.0731
	Sh2	**98.65**	**-2.56**	**0.0118**
	Sh3	**94.81**	**-3.88**	**0.0002**
	Test	106.35	0.25	0.8072
Post-stop *f0* (Hz)	Base	242.81		
	Sh1	**266.42**	**9.96**	**< 0.0001**
	Sh2	**262.38**	**8.25**	**< 0.0001**
	Sh3	**267.30**	**10.33**	**< 0.0001**
	Test	**255.53**	**6.96**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	197.21		
	Sh1	**183.77**	**-4.09**	**0.0001**
	Sh2	**184.85**	**-3.76**	**0.0002**
	Sh3	**182.73**	**-4.40**	**< 0.0001**
	Test	**185.77**	**-4.43**	**< 0.0001**

Table C.18. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F09. **Bolded** data are significant at $p < 0.05$.

M10

Measure	ANOVA	/t^h/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 155 *p* < 0.0001	**59.82**	**41.45**			**< 0.0001**
		59.82		**17.03**		**< 0.0001**
			41.45	**17.03**		**< 0.0001**
F2OT (ms.)	F(2,197) = 173.1 *p* < 0.0001	**74.71**	**54.55**			**< 0.0001**
		74.71		**25.65**		**< 0.0001**
			54.55	**25.65**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,293) = 475.4 *p* < 0.0001	**135.17**	**108.98**			**< 0.0001**
		135.17		**129.63**		**< 0.0001**
		135.17			**105.63**	**< 0.0001**
			108.98	**129.63**		**< 0.0001**
			108.98		**105.63**	**0.0098**
				129.63	**105.63**	**< 0.0001**

Table C.19. Measurements and statistics for baseline production of speaker M10. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**/t^h/**	**82.58**	**89.12**	**3.14**	**0.0023**
	/t/	62.69	60.70	-0.56	0.5797
	/t*/	44.22	37.83	-1.79	0.0763
F2OT (ms.)	**/t^h/**	**99.98**	**107.56**	**3.21**	**0.0018**
	/t/	77.68	74.61	-0.76	0.4522
	/t*/	54.33	47.39	-1.71	0.0897
Post-onset *f0* (Hz)	**/t^h/**	**134.71**	**144.15**	**9.42**	**< 0.0001**
	/t/	**111.00**	**102.20**	**-5.07**	**< 0.0001**
	/t*/	129.80	128.46	-0.77	0.4447
	filler	**107.67**	**99.73**	**-5.60**	**< 0.0001**

		Base	Test	*t*	*p*
VOT (ms.)	**/t^h/**	**81.40**	**86.12**	**2.40**	**0.0185**
	/t/	63.59	57.65	-1.76	0.0815
	/t*/	36.49	32.52	-1.18	0.2424
F2OT (ms.)	**/t^h/**	**99.16**	**103.77**	**2.07**	**0.0410**
	/t/	80.83	74.11	-1.76	0.0817
	/t*/	50.75	46.71	-1.06	0.2921
Post-onset *f0* (Hz)	**/t^h/**	**135.23**	**140.43**	**5.15**	**< 0.0001**
	/t/	106.96	104.83	-1.23	0.2211
	/t*/	**129.48**	**124.27**	**-2.98**	**0.0034**
	filler	**103.15**	**99.98**	**-2.23**	**0.0270**

Table C.20. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker M10. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	90.30		
	Sh1	86.21	-1.42	0.1596
	Sh2	87.54	-0.96	0.3396
	Sh3	88.08	-0.77	0.4410
	Test	**97.24**	**3.14**	**0.0021**
$/t^h/$ F2OT (ms.)	Base	110.49		
	Sh1	104.24	-1.94	0.0551
	Sh2	104.65	-1.82	0.0718
	Sh3	108.03	-0.77	0.4458
	Test	**118.61**	**3.30**	**0.0013**
Post-stop *f0* (Hz)	Base	134.71		
	Sh1	**138.26**	**2.78**	**0.0058**
	Sh2	**143.01**	**6.58**	**< 0.0001**
	Sh3	**144.93**	**8.11**	**< 0.0001**
	Test	**144.15**	**9.73**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	107.67		
	Sh1	104.77	-1.64	0.1022
	Sh2	**100.63**	**-4.02**	**0.0001**
	Sh3	**99.87**	**-4.45**	**< 0.0001**
	Test	**99.73**	**-5.78**	**< 0.0001**

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	97.74		
	Sh1	100.36	0.77	0.4440
	Sh2	101.89	1.20	0.2310
	Sh3	101.97	1.23	0.2200
	Test	**103.32**	**2.17**	**0.0320**
$/t^h/$ F2OT (ms.)	Base	114.79		
	Sh1	116.65	0.51	0.6082
	Sh2	119.26	1.23	0.2215
	Sh3	118.97	1.15	0.2514
	Test	**120.27**	**2.01**	**0.0461**
Post-stop *f0* (Hz)	Base	135.22		
	Sh1	**140.87**	**4.51**	**< 0.0001**
	Sh2	**139.08**	**3.12**	**0.0020**
	Sh3	**142.40**	**5.81**	**< 0.0001**
	Test	**140.43**	**5.44**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	103.15		
	Sh1	100.26	-1.68	0.0942
	Sh2	100.41	-1.61	0.1094
	Sh3	**99.33**	**-2.24**	**0.0259**
	Test	**99.98**	**-2.36**	**0.0190**

Table C.21. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker M10. **Bolded** data are significant at $p < 0.05$.

F11

Measure	ANOVA	/t^h/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,196) = 125.5 $p < 0.0001$	**61.36**	**52.75**			**0.0098**
		61.36		**15.41**		**< 0.0001**
			52.75	**15.41**		**< 0.0001**
F2OT (ms.)	F(2,196) = 129.3 $p < 0.0001$	**74.45**	**66.21**			**0.0388**
		74.45		**21.15**		**< 0.0001**
			66.21	**21.15**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,294) = 387.3 $p < 0.0001$	**253.70**	**210.64**			**< 0.0001**
		253.70		249.86		0.2339
		253.70			**204.82**	**< 0.0001**
			210.64	**249.86**		**< 0.0001**
			210.64		**204.82**	**0.0214**
				249.86	**204.82**	**< 0.0001**

Table C.22. Measurements and statistics for baseline production of speaker F11. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	/t^h/	91.88	90.06	-0.83	0.4103
	/t/	78.11	76.43	-0.46	0.6450
	/t*/	46.79	46.22	-0.16	0.8752
F2OT (ms.)	/t^h/	110.56	108.89	-0.71	0.4783
	/t/	97.43	98.51	0.28	0.7795
	/t*/	58.69	58.40	-0.07	0.9415
Post-onset *f0* (Hz)	**/t^h/**	**255.35**	**264.37**	**4.88**	**< 0.0001**
	/t/	211.16	205.18	-1.87	0.0640
	/t*/	**250.60**	**242.70**	**-2.44**	**0.0158**
	filler	**205.55**	**193.81**	**-4.49**	**< 0.0001**

		Base	Test	*t*	*p*
VOT (ms.)	**/t^h/**	**84.44**	**90.41**	**2.48**	**0.0146**
	/t/	78.23	74.33	-0.97	0.3357
	/t*/	40.41	36.41	-0.98	0.3288
F2OT (ms.)	/t^h/	105.57	106.73	0.42	0.6750
	/t/	98.07	97.15	-0.20	0.8450
	/t*/	54.73	54.60	-0.03	0.9780
Post-onset *f0* (Hz)	/t^h/	252.20	251.06	-0.59	0.5540
	/t/	210.12	211.54	0.43	0.6650
	/t*/	248.42	244.71	-1.10	0.2730
	filler	204.41	204.99	0.22	0.8290

Table C.23. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F11. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	106.90		
	Sh1	102.29	-1.53	0.1284
	Sh2	101.94	-1.66	0.0986
	Sh3	106.91	0.00	0.9983
	Test	103.69	-1.43	0.1539
$/t^h/$ F2OT (ms.)	Base	119.56		
	Sh1	**113.02**	**-2.16**	**0.0329**
	Sh2	114.49	-1.68	0.0949
	Sh3	119.48	-0.03	0.9775
	Test	117.03	-1.12	0.2656
Post-stop *f0* (Hz)	Base	255.35		
	Sh1	**262.73**	**4.00**	**0.0001**
	Sh2	**260.67**	**2.92**	**0.0038**
	Sh3	**260.39**	**2.77**	**0.0061**
	Test	**264.37**	**6.45**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	205.55		
	Sh1	**197.77**	**-3.05**	**0.0026**
	Sh2	**199.56**	**-2.37**	**0.0187**
	Sh3	200.88	-1.84	0.0663
	Test	**193.81**	**-5.93**	**< 0.0001**

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	102.61		
	Sh1	109.31	1.87	0.0639
	Sh2	106.91	1.11	0.2688
	Sh3	108.26	1.49	0.1383
	Test	106.96	1.59	0.1136
$/t^h/$ F2OT (ms.)	Base	123.29		
	Sh1	129.29	1.51	0.1330
	Sh2	125.66	0.56	0.5790
	Sh3	128.18	1.17	0.2440
	Test	122.97	-0.10	0.9180
Post-stop *f0* (Hz)	Base	252.32		
	Sh1	**259.67**	**3.88**	**0.0001**
	Sh2	251.84	-0.25	0.7998
	Sh3	253.03	0.37	0.7150
	Test	251.01	-0.89	0.3732
Post-sonorant *f0* (Hz)	Base	204.36		
	Sh1	199.30	-1.95	0.0527
	Sh2	201.69	-1.03	0.3057
	Sh3	203.46	-0.34	0.7330
	Test	205.11	0.36	0.7162

Table C.24. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F11. **Bolded** data are significant at $p < 0.05$.

F12

Measure	ANOVA	$/t^h/$	/t/	/t*/	Filler	Tukey p
VOT (ms.)	$F(2,197) = 191$ $p < 0.0001$	65.97	62.49			0.4389
		65.97		**12.65**		**< 0.0001**
			62.49	**12.65**		**< 0.0001**
F2OT (ms.)	$F(2,197) = 234.4$ $p < 0.0001$	72.22	69.82			0.6673
		72.22		**14.34**		**< 0.0001**
			69.82	**14.34**		**< 0.0001**
Post-onset *f0* (Hz)	$F(3,296) = 594.4$ $p < 0.0001$	**260.45**	**201.82**			**< 0.0001**
		260.45		**240.82**		**< 0.0001**
		260.45			**199.17**	**< 0.0001**
			201.82	**240.82**		**< 0.0001**
			201.82		199.17	0.5348
				240.82	**199.17**	**< 0.0001**

Table C.25. Measurements and statistics for baseline production of speaker F12.
Bolded data are significant at $p < 0.05$.

		Base	Test	t	p
VOT (ms.)	$/t^h/$	97.81	97.18	-0.35	0.7300
	/t/	97.82	96.59	-0.40	0.6940
	/t*/	45.01	44.74	-0.09	0.9320
F2OT (ms.)	$/t^h/$	101.41	102.26	0.43	0.6650
	/t/	103.42	100.47	-0.87	0.3860
	/t*/	45.80	44.27	-0.45	0.6510
Post-onset *f0* (Hz)	$/t^h/$	266.63	265.67	-0.50	0.6193
	/t/	203.40	198.92	-1.34	0.1819
	/t*/	248.52	247.00	-0.46	0.6497
	filler	202.03	197.29	-1.74	0.0843

		Base	Test	t	p
VOT (ms.)	$/t^h/$	97.37	98.82	0.60	0.5480
	/t/	90.35	91.32	0.23	0.8160
	/t*/	47.06	48.13	0.26	0.7980
F2OT (ms.)	$/t^h/$	102.11	104.73	1.06	0.2930
	/t/	95.94	97.75	0.42	0.6740
	/t*/	46.82	50.09	0.76	0.4480
Post-onset *f0* (Hz)	**$/t^h/$**	**253.66**	**259.42**	**3.12**	**0.0022**
	/t/	200.24	194.40	-1.83	0.0697
	/t*/	233.12	236.52	1.06	0.2892
	filler	**195.70**	**188.34**	**-2.82**	**0.0055**

Table C.26. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F12. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	107.16		
	Sh1	109.43	0.73	0.4680
	Sh2	107.64	0.15	0.8780
	Sh3	108.76	0.52	0.6010
	Test	106.21	-0.41	0.6820
/tʰ/ F2OT (ms.)	Base	112.08		
	Sh1	114.60	0.80	0.4230
	Sh2	113.21	0.36	0.7190
	Sh3	111.95	-0.05	0.9650
	Test	112.53	0.19	0.8490
Post-stop *f0* (Hz)	Base	266.63		
	Sh1	**259.51**	**-2.81**	**0.0052**
	Sh2	**259.03**	**-3.00**	**0.0029**
	Sh3	265.63	-0.39	0.6944
	Test	265.67	-0.49	0.6258
Post-sonorant *f0* (Hz)	Base	202.03		
	Sh1	203.42	0.40	0.6871
	Sh2	**212.26**	**2.95**	**0.0034**
	Sh3	207.66	1.63	0.1051
	Test	197.29	-1.71	0.0895

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	109.26		
	Sh1	106.49	-0.88	0.3800
	Sh2	113.16	1.23	0.2200
	Sh3	110.85	0.50	0.6200
	Test	110.46	0.50	0.6150
/tʰ/ F2OT (ms.)	Base	112.80		
	Sh1	110.86	-0.62	0.5360
	Sh2	116.71	1.24	0.2170
	Sh3	115.27	0.78	0.4370
	Test	115.16	1.00	0.3210
Post-stop *f0* (Hz)	Base	253.66		
	Sh1	**247.31**	**-2.71**	**0.0071**
	Sh2	**241.15**	**-5.35**	**< 0.0001**
	Sh3	251.23	-1.04	0.2995
	Test	**259.42**	**3.16**	**0.0018**
Post-sonorant *f0* (Hz)	Base	195.70		
	Sh1	195.69	0.00	0.9965
	Sh2	**202.09**	**1.99**	**0.0474**
	Sh3	194.29	-0.44	0.6596
	Test	**188.34**	**-2.86**	**0.0047**

Table C.27. Aspirated /t^h/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F12. **Bolded** data are significant at $p < 0.05$.

F13

Measure	ANOVA	/t^h/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 219.7 *p* < 0.0001	**60.10**	**47.39**			**< 0.0001**
		60.10		**11.32**		**< 0.0001**
			47.39	**11.32**		**< 0.0001**
F2OT (ms.)	F(2,197) = 247.6 *p* < 0.0001	**66.38**	**53.50**			**< 0.0001**
		66.38		**13.23**		**< 0.0001**
			53.50	**13.23**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 826 *p* < 0.0001	**306.95**	**216.52**			**< 0.0001**
		306.95		**285.64**		**< 0.0001**
		306.95			**214.72**	**< 0.0001**
			216.52	**285.64**		**< 0.0001**
			216.52		214.72	0.8982
				285.64	**214.72**	**< 0.0001**

Table C.28. Measurements and statistics for baseline production of speaker F13.
Bolded data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**/t^h/**	**71.38**	**66.23**	**-2.62**	**0.0101**
	/t/	61.60	66.01	1.30	0.1970
	/t*/	31.82	35.67	1.13	0.2598
F2OT (ms.)	/t^h/	83.87	80.95	-1.44	0.1521
	/t/	72.20	75.92	1.07	0.2882
	/t*/	39.53	41.07	0.44	0.6592
Post-onset *f0* (Hz)	**/t^h/**	**306.48**	**317.40**	**3.74**	**0.0003**
	/t/	216.64	213.16	-0.69	0.4927
	/t*/	286.08	288.40	0.46	0.6472
	filler	216.52	213.92	-0.63	0.5301

		Base	Test	*t*	*p*
VOT (ms.)	**/t^h/**	**90.50**	**83.55**	**-3.39**	**0.0010**
	/t/	72.78	76.78	1.16	0.2493
	/t*/	34.26	39.73	1.59	0.1156
F2OT (ms.)	**/t^h/**	**98.25**	**91.21**	**-3.41**	**0.0009**
	/t/	81.30	87.74	1.86	0.0657
	/t*/	38.47	44.63	1.78	0.0777
Post-onset *f0* (Hz)	**/t^h/**	**309.33**	**335.57**	**11.01**	**< 0.0001**
	/t/	**216.40**	**204.28**	**-2.94**	**0.0039**
	/t*/	285.20	280.96	-1.03	0.3061
	filler	**214.83**	**201.65**	**-3.91**	**0.0001**

Table C.29. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F13. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	85.10		
	Sh1	80.61	-1.50	0.1358
	Sh2	**78.94**	**-2.06**	**0.0416**
	Sh3	83.19	-0.64	0.5243
	Test	**79.62**	**-2.42**	**0.0171**
/tʰ/ F2OT (ms.)	Base	97.21		
	Sh1	93.84	-1.14	0.2560
	Sh2	94.20	-1.02	0.3110
	Sh3	98.83	0.55	0.5820
	Test	93.90	-1.48	0.1420
Post-stop *f0* (Hz)	Base	306.48		
	Sh1	**296.68**	**-2.58**	**0.0103**
	Sh2	**317.84**	**3.00**	**0.0030**
	Sh3	313.52	1.86	0.0645
	Test	**317.40**	**3.71**	**0.0003**
Post-sonorant *f0* (Hz)	Base	216.52		
	Sh1	**233.91**	**3.33**	**0.0010**
	Sh2	216.31	-0.04	0.9686
	Sh3	223.79	1.39	0.1647
	Test	213.92	-0.63	0.5326

	Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	99.15		
	Sh1	**91.24**	**-2.88**	**0.0047**
	Sh2	93.91	-1.89	0.0615
	Sh3	94.40	-1.72	0.0888
	Test	**91.34**	**-3.64**	**0.0004**
/tʰ/ F2OT (ms.)	Base	107.38		
	Sh1	**101.05**	**-2.29**	**0.0238**
	Sh2	101.85	-1.98	0.0500
	Sh3	103.70	-1.32	0.1898
	Test	**99.41**	**-3.69**	**0.0003**
Post-stop *f0* (Hz)	Base	309.33		
	Sh1	**319.23**	**3.12**	**0.0020**
	Sh2	315.51	1.95	0.0523
	Sh3	**317.19**	**2.48**	**0.0138**
	Test	**335.57**	**10.77**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	214.83		
	Sh1	214.69	-0.03	0.9760
	Sh2	219.05	0.96	0.3383
	Sh3	222.25	1.69	0.0931
	Test	**201.65**	**-3.82**	**0.0002**

Table C.30. Aspirated /tʰ/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F13. **Bolded** data are significant at $p < 0.05$.

M14

Measure	ANOVA	/tʰ/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 286 *p* < 0.0001	**71.44**	**61.64**			**0.0003**
		71.44		**12.87**		**< 0.0001**
			61.64	**12.87**		**< 0.0001**
F2OT (ms.)	F(2,197) = 330.7 *p* < 0.0001	**76.38**	**66.03**			**0.0001**
		76.38		**13.71**		**< 0.0001**
			66.03	**13.71**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,287) = 323.1 *p* < 0.0001	**126.87**	**103.44**			**< 0.0001**
		126.87		**121.51**		**< 0.0001**
		126.87			**102.23**	**< 0.0001**
			103.44	**121.51**		**< 0.0001**
			103.44		102.23	0.6771
				121.51	**102.23**	**< 0.0001**

Table C.31. Measurements and statistics for baseline production of speaker M14. **Bolded** data are significant at *p* < 0.05.

		Base	Test	*t*	*p*
VOT (ms.)	**/tʰ/**	**96.11**	**90.18**	**-3.77**	**0.0003**
	/t/	90.39	88.88	-0.56	0.5761
	/t*/	**40.39**	**45.88**	**2.03**	**0.0451**
F2OT (ms.)	**/tʰ/**	**103.13**	**97.68**	**-3.52**	**0.0007**
	/t/	95.94	95.83	-0.04	0.9671
	/t*/	43.38	47.96	1.72	0.0890
Post-onset *f0* (Hz)	**/tʰ/**	**123.67**	**126.84**	**2.97**	**0.0035**
	/t/	103.36	100.12	-1.80	0.0748
	/t*/	121.74	118.12	-1.97	0.0510
	filler	**101.91**	**96.03**	**-3.97**	**0.0001**

		Base	Test	*t*	*p*
VOT (ms.)	/tʰ/	103.22	102.85	-0.23	0.8162
	/t/	90.43	95.08	1.72	0.0882
	/t*/	45.90	45.01	-0.33	0.7423
F2OT (ms.)	/tʰ/	107.44	108.94	0.96	0.3400
	/t/	94.03	99.24	1.93	0.0572
	/t*/	45.52	42.94	-0.96	0.3411
Post-onset *f0* (Hz)	**/tʰ/**	**128.86**	**141.89**	**13.12**	**< 0.0001**
	/t/	**103.52**	**94.97**	**-5.07**	**< 0.0001**
	/t*/	121.24	119.57	-0.99	0.3230
	filler	**101.45**	**93.07**	**-6.03**	**< 0.0001**

Table C. 32. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker M14. **Bolded** data are significant at *p* < 0.05.

	Block	Estimate	*t*	*p*		Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	101.94			/tʰ/ VOT (ms.)	Base	110.95		
	Sh1	**93.16**	**-4.03**	**0.0001**		Sh1	112.92	0.86	0.3940
	Sh2	**91.26**	**-4.91**	**< 0.0001**		Sh2	114.90	1.77	0.0791
	Sh3	**94.83**	**-3.29**	**0.0013**		Sh3	111.82	0.38	0.7036
	Test	**95.84**	**-3.71**	**0.0003**		Test	110.56	-0.23	0.8172
/tʰ/ F2OT (ms.)	Base	107.70			/tʰ/ F2OT (ms.)	Base	113.08		
	Sh1	**99.45**	**-3.61**	**0.0004**		Sh1	113.40	0.14	0.8857
	Sh2	**97.64**	**-4.41**	**< 0.0001**		Sh2	116.21	1.47	0.1452
	Sh3	**100.87**	**-3.01**	**0.0032**		Sh3	112.35	-0.34	0.7346
	Test	**102.10**	**-3.24**	**0.0015**		Test	114.53	0.90	0.3702
Post-stop *f0* (Hz)	Base	123.84			Post-stop *f0* (Hz)	Base	128.86		
	Sh1	122.57	-1.04	0.2982		**Sh1**	**123.74**	**-4.04**	**0.0001**
	Sh2	**126.15**	**1.92**	**0.0557**		Sh2	130.50	1.34	0.1809
	Sh3	**128.27**	**3.64**	**0.0003**		Sh3	130.18	1.06	0.2888
	Test	**126.88**	**3.27**	**0.0012**		**Test**	**141.89**	**13.51**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	101.81			Post-sonorant *f0* (Hz)	Base	101.45		
	Sh1	100.47	-0.81	0.4194		**Sh1**	**106.24**	**2.75**	**0.0064**
	Sh2	99.32	-1.51	0.1327		Sh2	102.08	0.37	0.7154
	Sh3	**97.33**	**-2.70**	**0.0073**		Sh3	104.08	1.52	0.1290
	Test	**96.05**	**-4.44**	**< 0.0001**		**Test**	**93.07**	**-6.21**	**< 0.0001**

Table C.33. Aspirated /tʰ/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker M14. **Bolded** data are significant at $p < 0.05$.

F15

Measure	ANOVA	/t^h/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	$F(2,196) = 280.9$ $p < 0.0001$	**71.96**	**65.31**			**0.0226**
		71.96		**13.88**		**< 0.0001**
			65.31	**13.88**		**< 0.0001**
F2OT (ms.)	$F(2,196) = 331.7$ $p < 0.0001$	**80.48**	**74.09**			**0.0362**
		80.48		**15.76**		**< 0.0001**
			74.09	**15.76**		**< 0.0001**
Post-onset *f0* (Hz)	$F(3,296) = 709.5$ $p < 0.0001$	**276.78**	**217.28**			**< 0.0001**
		276.78		**251.52**		**< 0.0001**
		276.78			**210.86**	**< 0.0001**
			217.28	**251.52**		**< 0.0001**
			217.28		**210.86**	**0.0042**
				251.52	**210.86**	**< 0.0001**

Table C.34. Measurements and statistics for baseline production of speaker F15. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**/t^h/**	**96.61**	**103.97**	**2.68**	**0.0086**
	/t/	86.97	88.98	0.42	0.6722
	/t*/	39.85	32.56	-1.52	0.1321
F2OT (ms.)	**/t^h/**	**103.69**	**111.82**	**2.93**	**0.0043**
	/t/	95.54	95.49	-0.01	0.9918
	/t*/	39.36	30.94	-1.73	0.0864
Post-onset *f0* (Hz)	**/t^h/**	**278.05**	**285.83**	**5.12**	**< 0.0001**
	/t/	213.04	210.02	-1.15	0.2530
	/t*/	251.20	249.34	-0.71	0.4810
	filler	207.53	206.37	-0.54	0.5900

		Base	Test	*t*	*p*
VOT (ms.)	**/t^h/**	**96.98**	**110.51**	**6.38**	**< 0.0001**
	/t/	**91.92**	**79.97**	**-3.25**	**0.0016**
	/t*/	**42.51**	**25.23**	**-4.70**	**< 0.0001**
F2OT (ms.)	**/t^h/**	**106.17**	**122.52**	**7.23**	**< 0.0001**
	/t/	**99.11**	**85.89**	**-3.37**	**0.0011**
	/t*/	**45.71**	**25.40**	**-5.19**	**< 0.0001**
Post-onset *f0* (Hz)	**/t^h/**	**275.48**	**294.93**	**11.09**	**< 0.0001**
	/t/	**221.52**	**206.62**	**-4.92**	**< 0.0001**
	/t*/	251.84	246.74	-1.69	0.0941
	filler	**214.16**	**197.74**	**-6.64**	**< 0.0001**

Table C.35. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F15. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	117.97		
	Sh1	**106.33**	**-3.16**	**0.0020**
	Sh2	114.31	-1.00	0.3206
	Sh3	112.73	-1.42	0.1571
	Test	**125.35**	**2.66**	**0.0089**
$/t^h/$ F2OT (ms.)	Base	120.39		
	Sh1	**109.42**	**-2.97**	**0.0036**
	Sh2	117.62	-0.75	0.4539
	Sh3	117.38	-0.82	0.4163
	Test	**128.50**	**2.92**	**0.0042**
Post-stop *f0* (Hz)	Base	278.05		
	Sh1	**288.64**	**5.46**	**< 0.0001**
	Sh2	**289.16**	**5.73**	**< 0.0001**
	Sh3	**289.80**	**6.06**	**< 0.0001**
	Test	**285.83**	**5.24**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	207.53		
	Sh1	202.73	-1.77	0.0773
	Sh2	204.81	-1.01	0.3160
	Sh3	**201.37**	**-2.28**	**0.0237**
	Test	206.37	-0.55	0.5809

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	112.77		
	Sh1	**123.77**	**3.49**	**0.0007**
	Sh2	**121.71**	**2.83**	**0.0054**
	Sh3	**119.67**	**2.19**	**0.0306**
	Test	**126.56**	**5.78**	**< 0.0001**
$/t^h/$ F2OT (ms.)	Base	122.08		
	Sh1	**134.44**	**3.76**	**0.0003**
	Sh2	**131.03**	**2.72**	**0.0074**
	Sh3	**129.94**	**2.40**	**0.0182**
	Test	**138.57**	**6.65**	**< 0.0001**
Post-stop *f0* (Hz)	Base	275.46		
	Sh1	**279.59**	**1.91**	**0.0577**
	Sh2	278.63	1.46	0.1447
	Sh3	**284.23**	**4.05**	**0.0001**
	Test	**294.97**	**11.68**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	214.14		
	Sh1	**206.86**	**-2.42**	**0.0162**
	Sh2	212.06	-0.69	0.4898
	Sh3	208.74	-1.80	0.0738
	Test	**197.68**	**-6.99**	**< 0.0001**

Table C.36. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F15. **Bolded** data are significant at $p < 0.05$.

M16

Measure	ANOVA	$/t^h/$	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 186.5 *p* < 0.0001	**57.71**	**40.24**			**< 0.0001**
		57.71		**11.19**		**< 0.0001**
			40.24	**11.19**		**< 0.0001**
F2OT (ms.)	F(2,197) = 210.5 *p* < 0.0001	**66.28**	**48.72**			**< 0.0001**
		66.28		**15.11**		**< 0.0001**
			48.72	**15.11**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,295) = 401.4 *p* < 0.0001	**130.10**	**100.29**			**< 0.0001**
		130.10		**118.16**		**< 0.0001**
		130.10			**98.62**	**< 0.0001**
			100.29	**118.16**		**< 0.0001**
			100.29		98.62	0.5236
				118.16	**98.62**	**< 0.0001**

Table C.37. Measurements and statistics for baseline production of speaker M16. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**$/t^h/$**	**72.41**	**78.72**	**3.30**	**0.0013**
	/t/	54.63	55.78	0.35	0.7255
	/t*/	26.00	22.36	-1.12	0.2648
F2OT (ms.)	**$/t^h/$**	**83.82**	**90.30**	**3.17**	**0.0020**
	/t/	64.76	64.44	-0.09	0.9272
	/t*/	33.40	27.33	-1.75	0.0835
Post-onset *f0* (Hz)	**$/t^h/$**	**132.58**	**126.58**	**-5.65**	**< 0.0001**
	/t/	**102.75**	**106.48**	**2.01**	**0.0466**
	/t*/	120.08	118.92	-0.63	0.5293
	filler	**100.26**	**106.36**	**4.06**	**0.0001**

		Base	Test	*t*	*p*
VOT (ms.)	$/t^h/$	82.74	85.53	1.75	0.0830
	/t/	62.26	57.38	-1.75	0.0829
	/t*/	35.12	32.56	-0.93	0.3563
F2OT (ms.)	**$/t^h/$**	**95.85**	**101.53**	**3.50**	**0.0007**
	/t/	**74.68**	**67.80**	**-2.42**	**0.0173**
	/t*/	**42.06**	**34.85**	**-2.56**	**0.0120**
Post-onset *f0* (Hz)	**$/t^h/$**	**127.47**	**133.19**	**5.36**	**< 0.0001**
	/t/	**97.68**	**93.00**	**-2.53**	**0.0125**
	/t*/	116.24	116.44	0.11	0.9141
	filler	**96.83**	**92.65**	**-2.77**	**0.0064**

Table C.38. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker M16. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*		Block	Estimate	*t*	*p*
/tʰ/ VOT (ms.)	Base	71.63			/tʰ/ VOT (ms.)	Base	86.21		
	Sh1	**65.52**	**-2.26**	**0.0256**		Sh1	91.14	1.87	0.0639
	Sh2	**65.69**	**-2.20**	**0.0302**		Sh2	90.73	1.72	0.0886
	Sh3	70.75	-0.32	0.7466		Sh3	88.51	0.88	0.3822
	Test	**77.89**	**2.99**	**0.0033**		Test	89.04	1.43	0.1566
/tʰ/ F2OT (ms.)	Base	83.65			/tʰ/ F2OT (ms.)	Base	97.35		
	Sh1	**77.31**	**-2.21**	**0.0290**		**Sh1**	**103.77**	**2.50**	**0.0140**
	Sh2	**77.91**	**-2.00**	**0.0482**		**Sh2**	**103.66**	**2.46**	**0.0155**
	Sh3	83.60	-0.02	0.9868		Sh3	101.39	1.58	0.1179
	Test	**90.12**	**2.92**	**0.0042**		**Test**	**103.04**	**2.94**	**0.0040**
Post-stop *f0* (Hz)	Base	132.58			Post-stop *f0* (Hz)	Base	127.47		
	Sh1	**127.09**	**-3.72**	**0.0002**		Sh1	127.09	-0.25	0.8019
	Sh2	**121.49**	**-7.51**	**< 0.0001**		Sh2	127.65	0.13	0.8982
	Sh3	**124.46**	**-5.51**	**< 0.0001**		Sh3	127.65	0.13	0.8982
	Test	**126.58**	**-5.28**	**< 0.0001**		**Test**	**133.19**	**5.05**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	100.26			Post-sonorant *f0* (Hz)	Base	96.83		
	Sh1	**104.41**	**2.03**	**0.0433**		Sh1	96.66	-0.08	0.9350
	Sh2	**109.53**	**4.54**	**< 0.0001**		Sh2	96.78	-0.02	0.9814
	Sh3	**108.21**	**3.89**	**0.0001**		Sh3	96.62	-0.10	0.9196
	Test	**106.36**	**3.80**	**0.0002**		**Test**	**92.65**	**-2.61**	**0.0096**

Table C.39. Aspirated /tʰ/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker M16. **Bolded** data are significant at $p < 0.05$.

M17

Measure	ANOVA	$/t^h/$	/t/	/t*/	Filler	Tukey p
VOT (ms.)	F(2,197) = 436.7 $p < 0.0001$	**75.23**	**59.68**			**< 0.0001**
		75.23		**16.29**		**< 0.0001**
			59.68	**16.29**		**< 0.0001**
F2OT (ms.)	F(2,197) = 439.7 $p < 0.0001$	**83.01**	**68.03**			**< 0.0001**
		83.01		**20.54**		**< 0.0001**
			68.03	**20.54**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 180.1 $p < 0.0001$	**125.18**	**105.94**			**< 0.0001**
		125.18		**115.86**		**< 0.0001**
		125.18			**103.67**	**< 0.0001**
			105.94	**115.86**		**< 0.0001**
			105.94		103.67	0.2415
				115.86	**103.67**	**< 0.0001**

Table C.40. Measurements and statistics for baseline production of speaker M17. **Bolded** data are significant at $p < 0.05$.

		Base	Test	t	p
VOT (ms.)	$/t^h/$	80.13	76.69	-1.92	0.0581
	/t/	64.48	63.91	-0.19	0.8521
	/t*/	16.70	18.54	0.60	0.5507
F2OT (ms.)	$/t^h/$	89.55	86.42	-1.62	0.1090
	/t/	74.53	74.05	-0.15	0.8850
	/t*/	21.31	25.32	1.21	0.2290
Post-onset *f0* (Hz)	**$/t^h/$**	**120.81**	**115.45**	**-5.18**	**< 0.0001**
	/t/	**101.44**	**110.04**	**4.80**	**< 0.0001**
	/t*/	**111.44**	**116.88**	**2.98**	**0.0034**
	filler	**100.09**	**107.07**	**4.77**	**< 0.0001**

		Base	Test	t	p
VOT (ms.)	**$/t^h/$**	**73.86**	**77.43**	**2.04**	**0.0441**
	/t/	55.66	56.92	0.41	0.6812
	/t*/	16.79	12.89	-1.28	0.2029
F2OT (ms.)	**$/t^h/$**	**83.26**	**86.90**	**2.03**	**0.0457**
	/t/	65.64	66.75	0.35	0.7238
	/t*/	24.43	19.46	-1.59	0.1149
Post-onset *f0* (Hz)	**$/t^h/$**	**129.93**	**139.41**	**8.36**	**< 0.0001**
	/t/	**110.44**	**97.88**	**-6.39**	**< 0.0001**
	/t*/	**120.28**	**117.04**	**-1.65**	**0.1010**
	filler	**107.63**	**96.85**	**-6.72**	**< 0.0001**

Table C.41. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker M17. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	76.18		
	Sh1	74.96	-0.47	0.6400
	Sh2	75.71	-0.18	0.8600
	Sh3	75.21	-0.37	0.7140
	Test	72.40	-1.91	0.0580
$/t^h/$ F2OT (ms.)	Base	84.32		
	Sh1	83.20	-0.42	0.6751
	Sh2	83.54	-0.29	0.7756
	Sh3	83.09	-0.45	0.6503
	Test	80.73	-1.78	0.0773
Post-stop *f0* (Hz)	Base	120.81		
	Sh1	119.00	-1.43	0.1546
	Sh2	**116.12**	**-3.71**	**0.0003**
	Sh3	119.96	-0.67	0.5045
	Test	**115.45**	**-5.47**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	100.09		
	Sh1	**103.55**	**1.99**	**0.0481**
	Sh2	**104.11**	**2.31**	**0.0218**
	Sh3	102.83	1.57	0.1171
	Test	**107.07**	**5.04**	**< 0.0001**

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	73.39		
	Sh1	74.37	0.39	0.6995
	Sh2	76.42	1.16	0.2468
	Sh3	72.70	-0.27	0.7869
	Test	76.96	1.88	0.0632
$/t^h/$ F2OT (ms.)	Base	82.25		
	Sh1	84.01	0.69	0.4910
	Sh2	86.62	1.67	0.0967
	Sh3	82.35	0.04	0.9695
	Test	85.89	1.90	0.0597
Post-stop *f0* (Hz)	Base	129.93		
	Sh1	**137.59**	**5.00**	**< 0.0001**
	Sh2	**137.39**	**4.87**	**< 0.0001**
	Sh3	**137.55**	**4.97**	**< 0.0001**
	Test	**139.41**	**8.02**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	107.63		
	Sh1	**102.01**	**-2.64**	**0.0087**
	Sh2	**101.69**	**-2.80**	**0.0056**
	Sh3	**101.29**	**-2.98**	**0.0031**
	Test	**96.85**	**-6.45**	**< 0.0001**

Table C.42. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker M17. **Bolded** data are significant at $p < 0.05$.

M18

Measure	ANOVA	/tʰ/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 136.1	**66.75**	**54.83**			**0.0011**
		66.75		**12.70**		**< 0.0001**
	$p < 0.0001$		**54.83**	**12.70**		**< 0.0001**
F2OT (ms.)	F(2,197) = 135.7	**74.81**	**62.79**			**0.0016**
	$p < 0.0001$	**74.81**		**18.95**		**< 0.0001**
			62.79	**18.95**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 204.4	**131.14**	**109.10**			**< 0.0001**
	$p < 0.0001$	**131.14**		**125.10**		**< 0.0001**
		131.14			**109.14**	**< 0.0001**
			109.10	**125.10**		**< 0.0001**
			109.10		109.14	1.0000
				125.10	**109.14**	**< 0.0001**

Table C.43. Measurements and statistics for baseline production of speaker M18. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**/tʰ/**	**112.33**	**119.16**	**2.13**	**0.0360**
	/t/	92.84	103.14	1.85	0.0672
	/t*/	53.97	47.32	-1.19	0.2357
F2OT (ms.)	**/tʰ/**	**126.69**	**133.56**	**2.04**	**0.0438**
	/t/	**106.61**	**118.61**	**2.06**	**0.0419**
	/t*/	66.27	60.04	-1.07	0.2881
Post-onset *f0* (Hz)	**/tʰ/**	**134.96**	**141.90**	**6.01**	**< 0.0001**
	/t/	**112.88**	**107.06**	**-2.91**	**0.0042**
	/t*/	128.04	128.66	0.31	0.7571
	filler	112.32	109.32	-1.84	0.0684

		Base	Test	*t*	*p*
VOT (ms.)	**/tʰ/**	**87.23**	**98.23**	**5.12**	**< 0.0001**
	/t/	79.89	74.75	-1.40	0.1656
	/t*/	**40.92**	**31.44**	**-2.58**	**0.0115**
F2OT (ms.)	**/tʰ/**	**90.81**	**103.52**	**5.84**	**< 0.0001**
	/t/	83.87	77.67	-1.66	0.1007
	/t*/	**43.49**	**32.66**	**-2.89**	**0.0047**
Post-onset *f0* (Hz)	**/tʰ/**	**128.48**	**138.74**	**9.18**	**< 0.0001**
	/t/	**105.32**	**99.86**	**-2.82**	**0.0055**
	/t*/	122.16	122.10	-0.03	0.9753
	filler	**107.12**	**101.82**	**-3.35**	**0.0010**

Table C.44. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker M18. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	116.44		
	Sh1	**101.31**	**-3.30**	**0.0012**
	Sh2	**107.13**	**-2.10**	**0.0379**
	Sh3	**105.45**	**-2.41**	**0.0172**
	Test	**123.26**	**2.08**	**0.0398**
$/t^h/$ F2OT (ms.)	Base	128.58		
	Sh1	**106.56**	**-4.78**	**< 0.0001**
	Sh2	**117.90**	**-2.37**	**0.0193**
	Sh3	**111.78**	**-3.66**	**0.0004**
	Test	**135.45**	**2.06**	**0.0419**
Post-stop *f0* (Hz)	Base	134.96		
	Sh1	**150.65**	**11.09**	**< 0.0001**
	Sh2	**151.29**	**11.54**	**< 0.0001**
	Sh3	**148.17**	**9.34**	**< 0.0001**
	Test	**141.90**	**6.35**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	112.32		
	Sh1	**102.66**	**-4.94**	**< 0.0001**
	Sh2	**102.66**	**-4.94**	**< 0.0001**
	Sh3	**106.02**	**-3.22**	**0.0014**
	Test	109.32	-1.94	0.0533

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	92.88		
	Sh1	98.32	1.79	0.0757
	Sh2	**106.47**	**4.27**	**< 0.0001**
	Sh3	**108.86**	**5.15**	**< 0.0001**
	Test	**104.14**	**4.92**	**< 0.0001**
$/t^h/$ F2OT (ms.)	Base	95.28		
	Sh1	100.62	1.76	0.0810
	Sh2	**112.84**	**5.59**	**< 0.0001**
	Sh3	**109.68**	**4.63**	**< 0.0001**
	Test	**108.18**	**5.66**	**< 0.0001**
Post-stop *f0* (Hz)	Base	128.48		
	Sh1	**144.13**	**11.90**	**< 0.0001**
	Sh2	**142.73**	**10.83**	**< 0.0001**
	Sh3	**142.53**	**10.68**	**< 0.0001**
	Test	**138.74**	**10.15**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	107.12		
	Sh1	**99.08**	**-4.39**	**< 0.0001**
	Sh2	**102.44**	**-2.55**	**0.0112**
	Sh3	**103.08**	**-2.20**	**0.0283**
	Test	**101.82**	**-3.71**	**0.0003**

Table C.45. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker M18. **Bolded** data are significant at $p < 0.05$.

F19

Measure	ANOVA	/tʰ/	/t/	/t*/	Filler	Tukey p
VOT (ms.)	F(2,197) = 149 $p < 0.0001$	53.48	58.26			0.2156
		53.48		**8.95**		**< 0.0001**
			58.26	**8.95**		**< 0.0001**
F2OT (ms.)	F(2,197) = 183 $p < 0.0001$	**61.49**	**68.59**			**0.0462**
		61.49		**10.77**		**< 0.0001**
			68.59	**10.77**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 1190 $p < 0.0001$	**307.12**	**199.54**			**< 0.0001**
		307.12		**285.90**		**< 0.0001**
		307.12			**195.50**	**< 0.0001**
			199.54	**285.90**		**< 0.0001**
			199.54		195.50	0.4123
				285.90	**195.50**	**< 0.0001**

Table C.46. Measurements and statistics for baseline production of speaker F19. **Bolded** data are significant at $p < 0.05$.

		Base	Test	t	p
VOT (ms.)	/tʰ/	98.62	99.48	0.41	0.6860
	/t/	91.08	94.76	1.03	0.3080
	/t*/	39.87	36.31	-1.00	0.3220
F2OT (ms.)	/tʰ/	83.73	81.98	-0.83	0.4083
	/t/	**88.64**	**97.67**	**2.47**	**0.0152**
	/t*/	32.11	33.21	0.30	0.7654
Post-onset *f0* (Hz)	**/tʰ/**	**310.37**	**330.91**	**7.98**	**< 0.0001**
	/t/	**199.08**	**188.42**	**-2.39**	**0.0181**
	/t*/	289.52	285.46	-0.91	0.3642
	filler	**194.27**	**178.49**	**-4.33**	**< 0.0001**

		Base	Test	t	p
VOT (ms.)	**/tʰ/**	**66.66**	**58.37**	**-4.47**	**< 0.0001**
	/t/	**71.29**	**81.32**	**3.13**	**0.0023**
	/t*/	**24.38**	**33.38**	**2.80**	**0.0062**
F2OT (ms.)	**/tʰ/**	**75.40**	**66.73**	**-4.49**	**< 0.0001**
	/t/	**81.41**	**91.18**	**2.93**	**0.0043**
	/t*/	**26.40**	**36.04**	**2.88**	**0.0049**
Post-onset *f0* (Hz)	**/tʰ/**	**302.51**	**320.65**	**6.71**	**< 0.0001**
	/t/	200.00	191.42	-1.83	0.0691
	/t*/	**282.28**	**272.86**	**-2.01**	**0.0462**
	filler	**195.37**	**182.27**	**-3.42**	**0.0008**

Table C.47. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F19. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	83.80		
	Sh1	**67.97**	**-5.22**	**< 0.0001**
	Sh2	**63.39**	**-6.73**	**< 0.0001**
	Sh3	**63.85**	**-6.58**	**< 0.0001**
	Test	**78.25**	**-2.42**	**0.0170**
$/t^h/$ F2OT (ms.)	Base	90.51		
	Sh1	**77.34**	**-4.35**	**< 0.0001**
	Sh2	**73.59**	**-5.57**	**< 0.0001**
	Sh3	**72.34**	**-6.00**	**< 0.0001**
	Test	88.79	-0.75	0.4544
Post-stop *f0* (Hz)	Base	310.37		
	Sh1	**296.79**	**-4.09**	**0.0001**
	Sh2	316.39	1.81	0.0708
	Sh3	**319.19**	**2.66**	**0.0084**
	Test	**330.91**	**7.96**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base			
	Sh1	**211.68**	**3.82**	**0.0002**
	Sh2	191.68	-0.57	0.5720
	Sh3	190.76	-0.77	0.4435
	Test	**178.49**	**-4.32**	**< 0.0001**

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	69.84		
	Sh1	**54.15**	**-6.13**	**< 0.0001**
	Sh2	**53.37**	**-6.43**	**< 0.0001**
	Sh3	**57.47**	**-4.82**	**< 0.0001**
	Test	**61.62**	**-4.24**	**< 0.0001**
$/t^h/$ F2OT (ms.)	Base	85.11		
	Sh1	**70.99**	**-4.82**	**< 0.0001**
	Sh2	**71.60**	**-4.61**	**< 0.0001**
	Sh3	**77.80**	**-2.49**	**0.0141**
	Test	**76.66**	**-3.80**	**0.0002**
Post-stop *f0* (Hz)	Base	302.51		
	Sh1	304.25	0.51	0.6127
	Sh2	**310.85**	**2.42**	**0.0161**
	Sh3	305.05	0.74	0.4606
	Test	**320.65**	**6.81**	**< 0.0001**
Post-sonorant *f0* (Hz)	Base	195.37		
	Sh1	200.70	1.12	0.2637
	Sh2	193.30	-0.43	0.6652
	Sh3	201.54	1.30	0.1959
	Test	**182.27**	**-3.48**	**0.0006**

Table C.48. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F19. **Bolded** data are significant at $p < 0.05$.

F20

Measure	ANOVA	/t^h/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 182 *p* < 0.0001	70.57	64.36			0.1460
		70.57		**9.50**		**< 0.0001**
			64.36	**9.50**		**< 0.0001**
F2OT (ms.)	F(2,197) = 209 *p* < 0.0001	80.08	75.74			0.4040
		80.08		**13.77**		**< 0.0001**
			75.74	**13.77**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,295) = 312.8 *p* < 0.0001	**243.02**	**170.36**			**< 0.0001**
		243.02		**225.38**		**< 0.0001**
		243.02			**167.68**	**< 0.0001**
			170.36	**225.38**		**< 0.0001**
			170.36		167.68	0.8598
				225.38	**167.68**	**< 0.0001**

Table C.49. Measurements and statistics for baseline production of speaker F20. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	/t^h/	98.62	99.48	0.41	0.6860
	/t/	91.08	94.76	1.03	0.3080
	/t*/	39.87	36.31	-1.00	0.3220
F2OT (ms.)	/t^h/	105.15	106.74	0.74	0.4640
	/t/	100.73	105.97	1.42	0.1580
	/t*/	44.10	39.49	-1.25	0.2130
Post-onset *f0* (Hz)	/t^h/	261.89	266.35	1.91	0.0585
	/t/	177.76	170.98	-1.67	0.0963
	/t*/	**240.64**	**223.70**	**-4.18**	**< 0.0001**
	filler	**177.79**	**166.09**	**-3.54**	**0.0005**

		Base	Test	*t*	*p*
VOT (ms.)	/t^h/	112.82	114.66	0.82	0.4164
	/t/	**98.61**	**88.60**	**-2.69**	**0.0084**
	/t*/	50.79	47.15	-0.98	0.3282
F2OT (ms.)	/t^h/	125.38	128.76	1.41	0.1602
	/t/	**112.51**	**102.93**	**-2.43**	**0.0170**
	/t*/	57.41	55.27	-0.54	0.5878
Post-onset *f0* (Hz)	**/t^h/**	**224.47**	**216.87**	**-3.76**	**0.0002**
	/t/	**162.96**	**169.95**	**2.02**	**0.0454**
	/t*/	210.12	206.95	-0.91	0.3625
	filler	**157.98**	**165.16**	**2.53**	**0.0125**

Table C.50. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F20. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
/t^h/ VOT (ms.)	Base	105.30		
	Sh1	**92.53**	**-3.85**	**0.0002**
	Sh2	103.18	-0.64	0.5237
	Sh3	**97.20**	**-2.45**	**0.0157**
	Test	105.62	0.13	0.8962
/t^h/ F2OT (ms.)	Base	108.53		
	Sh1	**99.82**	**-2.69**	**0.0080**
	Sh2	106.91	-0.50	0.6158
	Sh3	**100.54**	**-2.50**	**0.0139**
	Test	109.85	0.55	0.5849
Post-stop *f0* (Hz)	Base	261.89		
	Sh1	**250.33**	**-3.98**	**0.0001**
	Sh2	**244.81**	**-5.88**	**< 0.0001**
	Sh3	**245.93**	**-5.49**	**< 0.0001**
	Test	**266.35**	**2.01**	**0.0457**
Post-sonorant *f0* (Hz)	Base	177.79		
	Sh1	178.17	0.09	0.9253
	Sh2	180.29	0.62	0.5386
	Sh3	176.45	-0.33	0.7419
	Test	**166.09**	**-3.73**	**0.0002**

	Block	Estimate	*t*	*p*
/t^h/ VOT (ms.)	Base	123.61		
	Sh1	**140.30**	**4.44**	**< 0.0001**
	Sh2	**143.23**	**5.22**	**< 0.0001**
	Sh3	**138.65**	**4.03**	**0.0001**
	Test	124.13	0.18	0.8615
/t^h/ F2OT (ms.)	Base	132.81		
	Sh1	**147.91**	**4.00**	**0.0001**
	Sh2	**151.17**	**4.86**	**< 0.0001**
	Sh3	**148.59**	**4.20**	**0.0001**
	Test	135.23	0.80	0.4235
Post-stop *f0* (Hz)	Base	224.46		
	Sh1	**198.75**	**-10.24**	**< 0.0001**
	Sh2	**195.55**	**-11.51**	**< 0.0001**
	Sh3	**195.47**	**-11.54**	**< 0.0001**
	Test	**216.87**	**-3.89**	**0.0001**
Post-sonorant *f0* (Hz)	Base	157.98		
	Sh1	**166.59**	**2.46**	**0.0144**
	Sh2	**165.99**	**2.29**	**0.0227**
	Sh3	**169.55**	**3.31**	**0.0011**
	Test	**165.14**	**2.62**	**0.0094**

Table C.51. Aspirated /t^h/ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F20. **Bolded** data are significant at $p < 0.05$.

F21

Measure	ANOVA	$/t^h/$	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 268.5 $p < 0.0001$	**74.09**	**63.12**			**0.0003**
		74.09		**10.59**		**< 0.0001**
			63.12	**10.59**		**< 0.0001**
F2OT (ms.)	F(2,197) = 279.8 $p < 0.0001$	**79.08**	**72.22**			**0.0386**
		79.08		**15.03**		**< 0.0001**
			72.22	**15.03**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,296) = 487.1 $p < 0.0001$	**255.42**	**199.26**			**< 0.0001**
		255.42		**244.74**		**< 0.0001**
		255.42			**200.62**	**< 0.0001**
			199.26	**244.74**		**< 0.0001**
			199.26		200.62	0.9105
				244.74	**200.62**	**< 0.0001**

Table C.52. Measurements and statistics for baseline production of speaker F21. **Bolded** data are significant at $p < 0.05$.

		Base	Test	*t*	*p*
VOT (ms.)	**$/t^h/$**	**103.00**	**96.15**	**-3.87**	**0.0002**
	/t/	87.40	88.91	0.49	0.6233
	/t*/	**32.92**	**39.45**	**2.11**	**0.0370**
F2OT (ms.)	**$/t^h/$**	**108.48**	**101.39**	**-3.98**	**0.0001**
	/t/	97.82	97.53	-0.09	0.9264
	/t*/	**38.32**	**45.26**	**2.23**	**0.0279**
Post-onset *f0* (Hz)	**$/t^h/$**	**258.72**	**262.48**	**2.24**	**0.0267**
	/t/	201.72	199.12	-0.89	0.3729
	/t*/	251.32	252.76	0.50	0.6213
	filler	**202.82**	**196.60**	**-2.62**	**0.0098**

		Base	Test	*t*	*p*
VOT (ms.)	$/t^h/$	87.02	90.23	1.68	0.0954
	/t/	79.19	80.15	0.30	0.7674
	/t*/	33.06	29.09	-1.23	0.2237
F2OT (ms.)	**$/t^h/$**	**92.99**	**97.07**	**2.09**	**0.0387**
	/t/	87.69	88.38	0.21	0.8369
	/t*/	37.83	33.96	-1.17	0.2460
Post-onset *f0* (Hz)	$/t^h/$	251.56	251.82	0.17	0.8694
	/t/	**196.80**	**207.38**	**3.87**	**0.0002**
	/t*/	**238.16**	**244.22**	**2.22**	**0.0282**
	filler	**197.86**	**206.20**	**3.74**	**0.0003**

Table C.53. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker F21. **Bolded** data are significant at $p < 0.05$.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	107.56		
	Sh1	**91.50**	**-6.07**	**< 0.0001**
	Sh2	**91.96**	**-5.94**	**< 0.0001**
	Sh3	**95.33**	**-4.63**	**< 0.0001**
	Test	**100.66**	**-3.55**	**0.0005**
$/t^h/$ F2OT (ms.)	Base	112.84		
	Sh1	**97.24**	**-5.70**	**< 0.0001**
	Sh2	**98.04**	**-5.44**	**< 0.0001**
	Sh3	**100.64**	**-4.47**	**< 0.0001**
	Test	**105.71**	**-3.54**	**0.0006**
Post-stop *f0* (Hz)	Base	258.72		
	Sh1	**267.31**	**4.27**	**< 0.0001**
	Sh2	262.63	1.95	0.0529
	Sh3	**266.59**	**3.91**	**0.0001**
	Test	**262.48**	**2.44**	**0.0154**
Post-sonorant *f0* (Hz)	Base	202.82		
	Sh1	**192.44**	**-3.70**	**0.0003**
	Sh2	199.80	-1.07	0.2838
	Sh3	**193.72**	**-3.24**	**0.0014**
	Test	**196.60**	**-2.85**	**0.0047**

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	94.24		
	Sh1	97.50	1.17	0.2442
	Sh2	92.18	-0.73	0.4683
	Sh3	92.68	-0.55	0.5820
	Test	97.07	1.33	0.1876
$/t^h/$ F2OT (ms.)	Base	101.21		
	Sh1	104.38	1.12	0.2639
	Sh2	100.19	-0.36	0.7213
	Sh3	100.69	-0.19	0.8537
	Test	104.84	1.67	0.0965
Post-stop *f0* (Hz)	Base	251.56		
	Sh1	**258.02**	**3.21**	**0.0015**
	Sh2	254.14	1.28	0.2008
	Sh3	253.66	1.04	0.2973
	Test	251.82	0.17	0.8662
Post-sonorant *f0* (Hz)	Base	197.86		
	Sh1	197.94	0.03	0.9766
	Sh2	203.22	1.91	0.0577
	Sh3	202.86	1.78	0.0765
	Test	**206.20**	**3.83**	**0.0002**

Table C.54. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker F21. **Bolded** data are significant at $p < 0.05$.

M22

Measure	ANOVA	/t^h/	/t/	/t*/	Filler	Tukey *p*
VOT (ms.)	F(2,197) = 222.1 *p* < 0.0001	**70.86**	**56.25**			**< 0.0001**
		70.86		**19.06**		**< 0.0001**
			56.25	**19.06**		**< 0.0001**
F2OT (ms.)	F(2,197) = 276.3 *p* < 0.0001	**77.09**	**61.52**			**< 0.0001**
		77.09		**21.33**		**< 0.0001**
			61.52	**21.33**		**< 0.0001**
Post-onset *f0* (Hz)	F(3,295) = 332.7 *p* < 0.0001	**110.23**	**96.04**			**< 0.0001**
		110.23		**107.87**		**0.0056**
		110.23			**93.89**	**< 0.0001**
			96.04	**107.87**		**< 0.0001**
			96.04		**93.89**	**0.0142**
				107.87	**93.89**	**< 0.0001**

Table C.55. Measurements and statistics for baseline production of speaker M22. **Bolded** data are significant at *p* < 0.05.

		Base	Test	*t*	*p*
VOT (ms.)	/t^h/	90.43	93.78	1.78	0.0781
	/t/	78.50	74.21	-1.31	0.1919
	/t*/	42.62	39.86	-0.85	0.3972
F2OT (ms.)	**/t^h/**	**97.73**	**101.84**	**2.39**	**0.0186**
	/t/	83.77	81.94	-0.61	0.5409
	/t*/	44.33	43.03	-0.44	0.6622
Post-onset *f0* (Hz)	/t^h/	112.49	111.71	-1.43	0.1550
	/t/	97.04	98.06	1.08	0.2830
	/t*/	108.32	109.14	0.87	0.3870
	filler	94.99	95.05	0.08	0.9380

		Base	Test	*t*	*p*
VOT (ms.)	/t^h/	97.91	97.42	-0.24	0.8145
	/t/	82.01	86.71	1.29	0.2006
	/t*/	44.77	46.61	0.51	0.6144
F2OT (ms.)	/t^h/	104.87	106.99	0.95	0.3444
	/t/	88.73	93.09	1.14	0.2568
	/t*/	49.48	49.83	0.09	0.9268
Post-onset *f0* (Hz)	**/t^h/**	**107.98**	**109.59**	**2.76**	**0.0066**
	/t/	95.04	95.97	0.91	0.3625
	/t*/	107.42	105.91	-1.47	0.1431
	filler	92.83	93.30	0.56	0.5789

Table C.56. VOT, F2OT, and *f0* estimates of different stop types in the high *f0* condition (top) and the long VOT condition (bottom), by speaker M22. **Bolded** data are significant at *p* < 0.05.

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	98.77		
	Sh1	94.20	-1.52	0.1320
	Sh2	**90.07**	**-2.92**	**0.0042**
	Sh3	98.62	-0.05	0.9622
	Test	101.97	1.42	0.1597
$/t^h/$ F2OT (ms.)	Base	103.50		
	Sh1	101.87	-0.59	0.5590
	Sh2	102.34	-0.42	0.6750
	Sh3	106.69	1.16	0.2485
	Test	107.50	1.92	0.0572
Post-stop *f0* (Hz)	Base	112.49		
	Sh1	112.39	-0.13	0.8994
	Sh2	**114.27**	**2.50**	**0.0131**
	Sh3	112.07	-0.57	0.5669
	Test	111.71	-1.42	0.1577
Post-sonorant *f0* (Hz)	Base	94.99		
	Sh1	**97.68**	**2.70**	**0.0073**
	Sh2	95.96	0.98	0.3302
	Sh3	96.92	1.94	0.0535
	Test	95.05	0.08	0.9386

	Block	Estimate	*t*	*p*
$/t^h/$ VOT (ms.)	Base	100.13		
	Sh1	94.99	-1.58	0.1173
	Sh2	**90.46**	**-3.01**	**0.0032**
	Sh3	**89.64**	**-3.26**	**0.0014**
	Test	99.57	-0.23	0.8195
$/t^h/$ F2OT (ms.)	Base	104.47		
	Sh1	102.08	-0.73	0.4691
	Sh2	**97.19**	**-2.24**	**0.0269**
	Sh3	99.35	-1.58	0.1174
	Test	106.60	0.86	0.3931
Post-stop *f0* (Hz)	Base	107.97		
	Sh1	107.70	-0.38	0.7073
	Sh2	106.94	-1.45	0.1484
	Sh3	**113.14**	**7.24**	**< 0.0001**
	Test	**109.60**	**2.95**	**0.0035**
Post-sonorant *f0* (Hz)	Base	92.84		
	Sh1	93.72	0.88	0.3780
	Sh2	**94.92**	**2.11**	**0.0355**
	Sh3	93.73	0.90	0.3674
	Test	93.29	0.59	0.5550

Table C.57. Aspirated $/t^h/$ in baseline, shadowing, and test productions in the high *f0* condition (left) and the long VOT condition (right), by speaker M22. **Bolded** data are significant at $p < 0.05$.

BIBLIOGRAPHY

Abramson, A. S., & Lisker, L. 1985. Relative power of cues: *F0* shift versus voice timing. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged,* 25-33.

Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* **29**, 155-190.

Babel, M., & Bulatov, D. 2012. The role of fundamental frequency in phonetic accommodation. *Language & Speech* **55**, 231-248.

Bates, D., Maechler, M., Bolker, B., & Walker, S. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, <http://CRAN.R-project.org/package=lme4>.

Boersma, P. & Weenink, D. 2014. Praat: Doing phonetics by computer (version 5.3.84). [Computer program], <http://www.praat.org/>.

Cho, T., Jun, S.-A., & Ladefoged, P. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* **30**, 198-228.

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

Choi, H. 2002. Acoustic cues for the Korean stop contrast - Dialectal variation. *ZAS Papers in Linguistics* **28**, 1-12.

Delvaux, V., & Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* **64**, 145-173.

Fowler, C. A. 1986. An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics* **14**, 3-28.

Fowler, C. A. 1996. Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* **99**, 1730-1741.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language* **49**, 396-413.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. 2012. *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84, <http://CRAN.R-project.org/package=irr>.

Giles, H., Coupland, N., & Coupland, J. 1991. Accommodation theory: Communication, context, and consequences. In H. Giles, N. Coupland, & J. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics,* 1-68. Cambridge: Cambridge University Press.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* **105**, 251-279.

Goldinger, S. D. & Azuma, T. 2004. Episodic memory reflected in printed word naming, *Psychonomic Bulletin & Review* **11**, 716-22.

Halle, M., & Stevens, K. N. 1971. A note on laryngeal features. *Quarterly progress report* **101**, 198-212. Research Laboratory of Electronics, MIT.

Hintzman, D. L. 1986. "Schema abstraction" in multiple-trace memory model. *Psychological Review* **93**, 411-428.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. 1979. Phonetic explanations for the development of tones. *Language* **55**, 37-58.

Honorof, D. N., Weihing, J., & Fowler, C. A. 2011. Articulatory events are imitated under rapid shadowing. *Journal of Phonetics* **39**, 18-38.

Jun, S.-A. 1993. *The phonetics and phonology of Korean prosody*. PhD dissertation, Ohio State University.

Johnson, K. 1997. Speech perception without speaker normalization: an exemplar model. In K. Johnson and J.W. Mullennix (Eds.) *Talker Variability in Speech Processing*, 145-166. San Diego: Academic Press.

Johnson, K. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* **24**, 485-499.

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

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.

Kang, Y. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* **45**, 76-90.

Kim, C.-W. 1965. On the autonomy of the tensity feature in stop classification. *Word* **21**, 339-359.

Kim, M. 2004. Correlation between VOT and *F0* in the perception of Korean stops and affricates. *Proceedings of the 8th International Conference on Spoken Language Processing (INTERSPEECH-2004)*, 49-52. Jeju Island, Korea: International Speech Communication Association.

Kim, M. 2012. *Phonetic accommodation after auditory exposure to native and nonnative speech*. PhD dissertation, Northwestern University.

Kim, M., Horton, W. S., & Bradlow, A. R. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* **2**, 125-156.

Kim, M.-R. 2000. *Segmental and tonal interactions in English and Korean: a phonetic and phonological study*. PhD dissertation, University of Michigan, Ann Arbor.

Kim, M.-R. 2008. "Lax" stops in Korean revisited: VOT neutralization. *Studies in Phonetics, Phonology and Morphology* **14**, 185-203.

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, M.-R. C. 1994. *Acoustic characteristics of Korean stops and perception of English stop consonants*. PhD dissertation, University of Wisconsin-Madison.

Kong, E. J., Beckman, M. E., & Edwards, J. 2011. Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics* **39**, 196-211.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. 2014. *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-20, <http://CRAN.R-project.org/package=lmerTest>.

Lee, H. & Jongman, A. 2012. Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association* **42**, 145-169.

Lee, H., Politzer-Ahles, S., & Jongman, A. 2013. Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics* **41**, 117-132.

Liberman, A., & Mattingly, I. 1985. The motor theory revised. *Cognition* **21**, 1-36.

McCrea, C. R., & Morris, R. J. 2005. The effects of fundamental frequency levels on voice onset time in normal adult male speakers. *Journal of Speech, Language, and Hearing Research* **48**, 1013-1024.

Mitterer, H., & Ernestus, M. 2008. The link between perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* **109**, 168-173.

Mitterer, H., & Müsseler, J. 2013. Regional accent variation in the shadowing task: Evidence for a loose perception–action coupling in speech. *Attention, Perception, & Psychophysics* **75**, 557-575.

Narayan, C., & Bowden. M. 2013. Pitch affects voice onset time (VOT): A cross-linguistic study. *Proceedings of Meetings on Acoustics* **19**, 060095.

Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* **39**, 132-142.

Oh, E. 2011. Effects of speaker gender on voice onset time in Korean stops. *Journal of Phoentics* **39**, 59-67.

Pardo, J. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* **119**, 2382-2393.

Pardo, J. 2013. Reconciling diverse findings in studies of phonetic convergence. *Proceedings of Meetings on Acoustics* **19**, 060140.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* **40**, 190-97.

Pardo, J. S., Jay, I. C., & Krauss, R. M. 2010. Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* **72**, 2254-2264.

Pardo, J. S., Jordan, K., Mallari., R., Scanlon, C., & Lewandowski, E. 2013. Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures, *Journal of Memory and Language* **69**, 183-195.

Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure*, 137-157. Amsterdam: John Benjamins.

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

Scarborough, R. 2004. *Coarticulation and the Lexicon*. PhD Dissertation, UCLA.

Shepard, C. A., Giles, H., & Le Poire, B. A. 2001. Communication accommodation theory. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology*, 33-56. New York: Wiley.

Shockley, K., Sabadini, L., & Fowler, C. A. 2004. Imitation in shadowing words. *Perception and Psychophysics* **66**, 422-429.

Silva, D. J. 1992. *The phonetics and phonology of stop lenition in Korean.* PhD Dissertation, Cornell University.

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

The National Institute of the Korean Language. 2005. *A speech corpus of reading-style standard Korean* [DVDs]. Seoul, Korea: The National Institute of the Korean Language.

Tilsen, S. 2009. Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics* **37**, 276-296.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. 1993. F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America* **93**, 2152-2159.

Yu, A. C. L. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE* **5**, e11950.

Yu, A. C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. L. Yu (Ed.) *Origins of Sound Change: Approaches to Phonologization*, 201-227. Oxford: Oxford University Press.

Zellou, G., Scarborough, R., & Nielsen, K. 2013. Imitability of contextual vowel nasalization and interactions with lexical neighborhood density. *Proceedings of Meetings on Acoustics* **19**, 060083.