# Phonetic and Cognitive Bases of Sound Change

by

Reiko Kataoka

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Linguistics

in the

Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor John J. Ohala, Co-Chair
Professor Keith Johnson, Co-Chair
Professor Andrew Garrett
Professor Yoko Hasegawa

Fall 2011

The dissertation of Reiko Kataoka, titled Phonetic and Cognitive Bases of Sound Change, is approved:

Co-Chair _____ Date 11 Aug 2011

Co-Chair _____ Date Aug 12, 2011

_____ Date August 11, 2011

_____ Date Aug. 12, 2011

University of California, Berkeley

Phonetic and Cognitive Bases of Sound Change



by Reiko Kataoka

# Abstract

Phonetic and Cognitive Bases of Sound Change

by

Reiko Kataoka

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor John J. Ohala, Professor Keith Johnson, Co-Chairs

In this dissertation I investigated, by using coarticulatory /u/-fronting in the alveolar context for a case study, how native speakers of American English produce coarticulatory variations and how they perceive and reproduce continuously varying speech sounds that are heard in coarticulatory and non-coarticulatory contexts.

The production study addressed the question of whether in American English coarticulatory fronting of /u/ in alveolar contexts is an inevitable consequence of production constraints or if it is produced by active speaker control. The study found that: (1) the relative acoustic difference between the fronted /u/ and the non-fronted /u/ remained across an elicited range of vowel duration; and (2) the degree of acoustic variability was less for the fronted /u/ than the non-fronted /u/. These results indicate that speakers of American English have a distinct and more narrowly specified articulatory target for the fronted /u/ in the alveolar context than for the non-fronted /u/.

The perception study addressed the issue of individual variation and compensation for coarticulation. The study found within-subject consistency in classification of /CVC/ stimuli both in compensatory and non-compensatory contexts. The study found no evidence for a within-subject perception-production link, but did find positive evidence for the relationship between linguistic experience and speech perception—the similarity between the distributional characteristics of the fronted and the non-fronted variants of /u/ in production data (a proxy for ambient language data) and the ranges of variation in perceptual responses toward /CVC/ stimuli in the fronting and the non-fronting contexts. Together, these results suggest that the source of individual variation in speech perception is the differences in the phonological grammar (perceptual category boundary) that guide speech perception, and that this perception grammar emerges in response to the ambient language data.

Finally, the vowel repetition study examined how perceptual compensation for coarticulation and individual differences in speech perception affect vowel repetition performance. This study found that: (1) ambiguous vowels were repeated with a significantly lower F2 when the vowels were heard in the fronting context than in the non-fronting context; (2) a given stimulus was repeated by some listeners un-ambiguously as the vowel belonging to the speaker's /i/ category for all trials, yet the same stimulus was repeated by other listeners un-ambiguously as vowels

belonging to that speaker's /u/ category for all trials; and (3) the perceptual category boundary was a significant predictor for the repeated vowel's F2 value. Based on these results, it was hypothesized that one source of pronunciation variation in a given community is individual variation in speech perception that contributes variable mental representations across listeners when they encounter ambiguous speech.

One general pattern that was found in all experiments was vowel-specific variability: responses to /i/ were less variable than responses to /u/ in a production task, and /i/-like stimuli were repeated less variably than /u/-like stimuli in a vowel repetition task. Similarly, between /u/ in fronting and non-fronting contexts, /u/ elicited less variability in the fronting context than in the non-fronting context consistently in the production, perception, and vowel repetition tasks. More broadly, I contend that speech forms a dynamic system, characterized by mutual dependency and multiple causal loops between and among speech perception, speech production, knowledge about pronunciation norm, and ambient language data. These properties in language use govern the output of communicative interactions among members in a speech community, and one such output is member's knowledge of multiple sub-phonemic pronunciation categories that exist in any speech community. Additionally, I argue that any speech community is in a constant state of readiness to respond to an innovative pronunciation as a new community norm, because members have a variable but rich pronunciation repertoire even when there is no observable community-level sound change.

# Contents

# List of Figures

# List of Tables

# Acknowledgments

I would never have been able to finish, or even start, my dissertation without the inspiration, guidance, and support I received from many people.

First and foremost I would like to express the deepest appreciation to my thesis committee: John Ohala, Keith Johnson, Andrew Garrett, and Yoko Hasegawa. John has been my academic mentor and champion for the past ten years. He has taught me the essential tools—the knowledge of the physics and physiology of speech, deductive reasoning, and the skills to examine speech data—that I needed in my quest to discover how humans produce and perceive speech. I also want to thank John for his friendship, for all the discussions, jokes, and weekly trips to Milano (across Bancroft) that helped me to keep a healthy mind, even in the tough times of the Ph.D. pursuit. I owe much to Keith for my knowledge and research skills in speech perception and appreciate his generosity in sharing ideas and resources. I am also thankful for his commitment to provide an excellent lab atmosphere for doing research while relaxing and having fun in the lab. I thank Andrew for convincing me, when I was an undergraduate student, that doing linguistics is cool and helping me to develop my background in language change. I also appreciate his truly constructive critiques of my work, which have improved my approach to the issue of sound change. I am deeply indebted to Hasegawa-sensei for her guidance and moral support through my long journey of completing the dissertation. I am particularly grateful for her constant encouragement during the beginning stages of writing, when the process was especially difficult. My dissertation research and writing also benefitted greatly from the guidance and discussions I had with Maria-Josep Solé, Pam Beddor, Meghan Sumner, Alan Yu, and Kiyoko Yoneyama. I also thank Maria-Josep and Kiyoko for their guidance on research and beyond and Meghan for providing me with research assistantship, during which I have learned so much about the mental representation of speech sounds.

I would also like to express my deepest gratitude to Ronald Sprouse, who has set up and maintained the Lab instruments that I needed, taught me basic scripting, and even offered me a weekly writing consultation for my English writing. Thank you, Ron. This dissertation would not have been possible without you. My gratitude also reaches out to the three enthusiastic undergraduate research assistants, Colin Alley, Jaime Lambert, and Matthew Quan, who collected data for the perception experiment and the vowel repetition experiment, and did segmentation of the audio data. It was my pleasure working with you: we were a good team.

Now looking back on my past ten years in the Linguistics department at Berkeley, both as an undergraduate student and graduate student, I realize what a positive influence this warm and friendly department has had on my decision to pursue my career as a linguist. For this, I really want to thank everyone in the department. Especially, I thank Leanne Hinton for taking me under your scholarly wings during my two years of research assistantship in Survey of California and Other Indian Languages; Larry Hyman for illuminating discussions at Phorum meetings, that helped me to gain a balanced view on the issues discussed each time; Sharon Inkelas for giving me big smiles every time spoke up in the classrooms and Phorum meetings, which boosted my confidence in expressing my thoughts and ideas; Ian Maddieson for showing great examples of direct, elegant, and precise argument each time he spoke, which I wish to be able to emulate one day; Robin Lakoff for her mentorship during my undergraduate training, during which she provided me with a good foundation for the social aspects of language use; and Gary Holland for

helpful advice and suggestions for literatures. I am also grateful to the extremely helpful department staff Paula Floro, Belén Flores, and Natalie Babler. Paula must have been my mother in our previous lives: she has been just so kind and supportive to me all the time. I cannot thank enough to Belén, who has saved me every time when my lack of organizational skills almost caused a disaster.

I am also grateful for having met and being surrounded by many wonderful fellow graduate students in the department, especially Nicole Marcus, Rebecca Cover, and Yuni Kim, who are also good friends of mine. I also want to thank everyone in the Phonology Lab (current and former), especially those whom I have had many opportunities to work, eat, and have stimulating discussions about each other's research with: Anne Pycha, Charles Chang, Christian DiCanio, Clara Cohen, Elsa Spinelli, Eurie Shin, Grant McGuire, John Sylak, Marc Ettlinger, Masako Fujimoto, Melinda Woodley, Molly Babel, Pawel Nowak, Rungpat Roengpithya, Russell Rhodes, Ryan Shosted, Sam Tilsen, Shinae Kang, Shira Katseff, Te-hsin, Liu, Will Chang, Yao Yao, and YiZhi Wang.

Last but not least, I am grateful to my friends and family for their support and encouragement, especially Erin Diehm for thoughtful advice and many passionate cheers, which helped me to keep on sprinting toward completion of this dissertation, and Kazuko Hönes, Hiroko Kakegawa, Susanne Nell, Kaori Okada, Kazumi Okano, and Mitsuko Yamashiro, to name just few, for the constant supply of warmth and light-hearted moments. Finally, I owe my deepest gratitude to Vincent and Sean. Thank you, Vincent, for giving me a true sense of security, peace, and comfort. Thank you, Sean, for loving me and trusting me even when I was too desperate in my work to do what is expected of moms. We have walked this long and strenuous journey together, and we did it very well. Vincent and Sean, this dissertation is for *you*.

# Chapter 1

# Introduction

Language is a complex dynamic system of human communication, and the parts and subparts that collectively define language interact and influence each other (Beckner, Blythe, Bybee, Christiansen, Croft, Ellis, Holland, Ke, Larsen-Freeman & Schoenemann, 2009; Oudeyer, 2005; Pierrehumbert, 2006). This fact was already captured in the classic speech chain[1] model (Denes & Pinson, 1973), in which a speaker is modeled as his or her own listener by a feedback loop that allows the speaker to monitor his or her speech and make corrections and adjustments as needed. In this conception of human speech, the speaker's ability to speak normally and intelligibly relies on the separate act of listening at the same time. The speech chain also represents dynamic speaker-listener interactions at a conversational level. In dyads, speakers constantly modulate their conversational contribution at the moment of interaction, as it is required, so as to ensure their speech to be understood by the listener (Grice, 1989; Lindblom, 1990). Again, speaking successfully in communicative interactions relies on effective feedback from the listener; therefore, the listener's reaction to the speaker's utterance inevitably influences the way subsequent utterances will be articulated by the speaker. In many respects, speech, and language in general, exhibits interdependence between different actions involved in the speech and its participants. The speech chain forms a complex system with multiple interactive loops, rather than a simple linear one.

This dissertation is about pronunciation variations within a single speech community that emerge through a system of mutual dependency between individual language users and their speech community, between speech perception and speech production, and between speech perception and knowledge of community pronunciation norms. The main part of the dissertation reports a three-part study that examined: (1) how a group of individuals in a speech community contributes to forming structured and phonetically constrained pronunciation variations in the pool of community speech data; (2) how listeners accommodate the expected types and ranges of pronunciation variation upon hearing other members' speech; and (3) how the listener's perceptual interpretation of another speaker's utterances influences the way the same listener reproduces the perceived utterances. In a nutshell, this study investigates how speech production

[1] For this and all other underlined terms, see Section 1.5 (Definition of Key Terms) for definition and additional clarifications.

and speech perception, and an individual language user and his or her speech community collectively form a series of complex interactions around speech pronunciation. Based on the results of the study, this dissertation offers a description of inherent stability and instability of pronunciation norms that arise from interactions between physical, physiological, and cognitive activities that co-occur in speech.

## 1.1 Background

The present study situates itself on an extension of a long lineage of inquiries into the phonetic bases of sound change, some of which date back to as early as the 17th century (Ohala, 2008, p. 369). The most influential early pioneers in this field of research were a group who came to be known as the Neogrammarians. They sought the cause of regularity in sound change within a phonetic bias toward more convenient articulation and human tendency for errors in articulation and in learning (Osthoff & Brugman, 1878/1967; Paul, 1888/1970). Paul, for example, conjectured that the origin of sound change is in altered articulatory representations of speech sounds.[2] For Neogrammarians, sound change was a natural consequence arising from constant interactions between phonetic and cognitive underpinnings of speech. Other early pioneers were functional linguists, who examined communicational function of language and sought the cause of sound change within conflicting pressures from the principles of minimizing articulatory effort and minimizing perceptual confusion (Grammont, 1939; Martinet, 1952, 1962). What is in common between these two groups of pioneers is their conviction that the key to understand sound change is to understand the sources of pronunciation variations and how language users react to these variations. Their approaches have not been free of criticism (e.g., Kiparsky, 1965; Weinreich, Labov & Herzog, 1968), but the ideas articulated in their works have inspired and influenced a large body of subsequent research and debates on the cause and mechanisms of sound change (e.g., Kiparsky, 1965, 1988; Labov, 1994, 2001, 2010; Lindblom, Guion, Hura, Moon & Wilerman, 1995).

While both Neogrammarians and functionalists considered speakers as primary contributors in initiating sound change, Ohala (1981, 1989, 1993) argued that the listener takes on the main role in sound change. In Ohala's view, as was also expressed by Baudouin de Courtenay (1910), sound change originates in listener "misperception" in a speech chain, whereby mental representation of speech sound(s) decoded and stored by the listener differs from what is encoded by the speaker. The same view that an innovative or a deviant pronunciation may arise due to misperception and mis-reproduction of sounds was presented by several predecessors (Jonasson, 1971; Paul, 1888/1970, p. 54; Sweet, 1888, p. 16).[3] Ohala hypothesized that assimilatory misperception would occur when a feature or multiple features induced by coarticulation is interpreted by a listener as a part of intended features for a target sound.

Assimilatory listener misperception defined as such is compatible with other models of sound change. First, misperception may be considered as equivalent to "phonologization" (Hyman,

[2] Paul (1888/1970) assumed that mental representation of speech sounds consists of articulatory representations and auditory representations. I share the same assumption in this study.
[3] According to Ohala (1981), the same view was also shared by Durand (1956) and Passy (1890).

1972, 1975, 1976, 2008) that occurs within the listener's mental representation. Phonologization is a process whereby a phonetic feature that speech sound has acquired or lost due to physical and physiological constraints in a given phonetic environment becomes exaggerated to the degree that the feature (or lack of feature) is no longer perceived by the language users as induced by the phonetic context but rather independently controlled as a distinctive specification of the sound (e.g., Hyman, 1976; see §2.4). Misperception obviates speaker exaggeration, though the listener who has perceived the coarticulatory feature to be controlled feature might subsequently exaggerate that feature in future production. Second, while the connotations in the terms are different, the outcome of assimilatory misperception would be identical to the outcome of the "how" mode of perception (Lindblom et al., 1995; see §2.1.3), wherein fine acoustic details, including distortions, are faithfully perceived and stored by a listener. Finally, Ohala's conceptualization of the origin of sound change converges with that of Paul in that the very first event of sound change, or *Initial Change* as it is called in this dissertation, is defined as a change in the mental representation of the speech sound. However, while Paul treated speaker-generated articulatory representation of speech as a target of Initial Change, Ohala proposed that it was an auditory representation of perceived speech that has ramification for sound change. The approach to explain the cause of sound change by using underlying phonetic conditioning in speech production and or speech perception has been adopted by many researchers today (e.g., Beddor, 2009; Blevins, 2004; Blevins & Garrett, 2004; Garrett & Johnson, in press; Guion, 1996; Hansson, 2008; Yu, 2004, 2010, in press).

Admittedly, phonetic conditioning of speech, which is based on universally present physical and physiological constraints on speech, does not solve an "actuation problem"—why a particular sound change occurs in one particular time or only in some languages (Weinreich et al., 1968). This, however, is not a problem in phonetic approach but only highlights the division of labor in research. Historical linguists have made the distinctions between the object of study: one is innovation, or a single person's usage (or grammar) that differs from the previous usage (or grammar) and the other is change, or the adoption of an innovation by all or at least much of the community members (Janda & Joseph, 2003, p. 13). Studies on Initial Change, including the present study, address the issue of innovation, not subsequent adoption by the rest of the community member. The actuation problem is concerned with the community-level change. "Triggering events" (in a sense of Labov, 2010, p. 90) of a specific community-level sound change, or more specifically, the motivations for the community members to adopt particular innovations or to suddenly exaggerate existing variations, need to be sought outside of the phonetic and other universal factors.

Some researchers look into social factors that trigger community-level sound changes (§2.3.1). For example, Labov's (1963) pioneering work on the vowel change in Martha's Vineyard clearly demonstrated how some regional variations can be suddenly exaggerated by heightened social attitude held by particular group of speakers in a particular speech community in a particular time. Eckert (1989) showed that the ways in which community members manifest their social memberships correlate with phonological variation. In these approaches, sound change needs to be studied within the specific speech communities in which it takes place. Other researchers have argued for a role of phonology, over phonetics, as a guiding force of sound change (§2.3.2). Bloomfield (1933) characterized regularity of sound change as "phonemes change" (p. 353). This statement is too simplistic, overlooking contextual constraints on many

regular changes (e.g. Grimm's law, by which Proto-Indo-European voiceless stops became voiceless fricatives, did not apply when the voiceless stops were preceded by */s/), but there are cases where structural analyses allow us to explain some patterns in sound change such as the difference between an irregular sound change that exhibits lexical diffusion from a regular sound change (Kiparsky, 1965, 1988, 2003).

Nonetheless, studies on universally available phonetic and cognitive factors have significant consequences in our understandings on the cause of community-level sound change as well. Consider the two paradoxical traits of sound change. First, sound change takes place only in a particular language in a particular time, but it is also universal in the sense that every language's sounds change over time. The very fact that a particular sound change occurs in some language but not in others implies that sound change can occur only when underlying phonetic and cognitive conditions meet certain other conditions—most likely social and structural. However, root causes, or preconditions, of sound change must be universally present at any given time regardless of the occurrence or non-occurrence of observable change. These preconditions include the language user's sensitivity to subtle sub-phonemic variations, as incipient changes are extremely subtle (Labov, 2001). Second, sound change is disruptive to speech intelligibility, yet within the community where a particular sound change progresses, communication has never been disrupted by sound change and language learners even increment the change (§2.2.2). Here, language users', and young learners' ability in particular, to discern multiple sub-phonemic variants as socially meaningful pronunciation categories (Labov, 1994) allow inception and progression of sound change. This ability, too, must be universally available to all language users. Sound changes are triggered by specific factors, but the human capabilities that allow incipient sound change to occur and progress without reducing communicative functioning of their languages are universal traits, and there is so much to learn from this aspect of human behavior. It follows that there are at least two different ways to study the cause of sound change. Doubtlessly, the most direct approach is to address the triggers of specific sound changes, examine each case separately as the change progresses within a particular speech community, and then draw general principles from accumulated body of findings. Another equally useful approach is to address the preconditions of sound change, especially the conditions that make the initiation of sound change as well as the adaptation to the change possible. These two approaches complement each other.

In addition, studies on pronunciation variation and language users' responses to this variation have important contributions to speech science in general, as these studies relate to issues in coarticulation, perceptual normalization of speech, mental representation of speech, and interaction of linguistic experience with these aspects of speech. These studies necessarily rely on theories, models, and methodologies developed in diverse fields of inquiry, including speech physiology, phonetics, phonology, psycholinguistics, and cognitive science. In return, these studies serve as a vehicle to promote mutually beneficial interactions among related research fields.

## 1.2 The Issues

A central topic for a general theory of sound change is phonetic variation and the consequences of this variation on the linguistic knowledge and language use both at an individual- and a community-level. Four issues are particularly important in this regard. The first is the nature of individual variation. The exemplar theory (e.g., Johnson, 1997; Pierrehumbert, 2001; Wedel, 2006) predicts that individual variation reflects systematic differences in stable behavioral patterns across individuals because this variation arises from different sets of previous experiences. While cross-linguistic and cross-dialect studies have provided evidence of language-specificity in the linguistic behaviors both in production (e.g., Beddor, Harnsberger & Lindemann, 2002; Öhman, 1966) and perception (e.g., Beddor & Krakow, 1999; Harrington, Kleber & Reubold, 2008) within a single speech community, expected predictability in individual variation has hardly been tested. More studies that directly test correlations between certain linguistic behaviors and linguistic experience of the language users are needed.

The second is language users' sensitivity to sub-phonemic and sub-allophonic pronunciation variation in production and perception. Theories of sound change assume the language user's capability to represent coarticulatory variation as a distinct pronunciation category (Labov, 2010; Lindblom, et al., 1995; Ohala, 1981, 1989, 1993). However, it has been well established that perceptual compensation for coarticulation normalizes contextual variation at least to some degree (e.g. Beddor & Krakow, 1999; Lindblom & Studdert-Kennedy, 1967; Mann & Repp, 1980; Ohala & Feder, 1994). Therefore, it is important to explicitly test language user's sensitivity to and knowledge about sub-phonemic variation both in production and perception, as well as in encoding.

The third is the relationship between speech production and speech perception within a single language user's performance. Many studies have suggested that speech perception is influenced by the internal production mechanism (e.g., Fowler, 1986; Liberman & Mattingly, 1985), and there is also evidence that speech production uses auditory feedback as a part of the control mechanism (e.g., Guenther, 2006; Katseff, 2010). However, the relationship between listener perception of other speaker's utterances and the listener's own production patterns has not been well understood. The link between these two components of speech needs to be investigated in a manner that identifies causal relationships between the two.

Finally, the last issue is the assumption behind the idea of the perception-production feedback loop over time (e.g., Oudeyer, 2005; Pierrehumbert, 2001). Models of sound change that use the perception-production feedback loop assume that listener perception has a systematic effect on the listener's speech production in later occasions. These models predict that if two listeners differ from each other in their perceptual interpretation of a given utterance, then these two listeners would later reproduce the utterance in two distinct pronunciations. This prediction needs to be tested in a controlled experiment.

## 1.3 Purpose of the Study

In order to address the above issues, this study examines how language users produce, perceive, and encode coarticulatory variation, and explores how these aspects of linguistic behavior are related to each other and to the speaker's phonological knowledge. A particular case examined in this study was coarticulatory variation of high back vowel /u/ in fronting contexts (i.e. adjacent to alveolar consonants) and non-fronting contexts (i.e. in isolation from or adjacent to bilabial consonants) in the American English. It has been widely documented that this vowel exhibits markedly different realizations in these two contexts (§3.1). It has been also reported that these variants have important implications for synchronic phonology and sound change (§3.1). Using a production, a perception, and a vowel repetition experiment, the present study explors (1) the nature of coarticulatory variations in speech production and speech perception in the two contexts, (2) the relationship between speech production and speech perception, and (3) the relationship between community members' knowledge about normative pronunciation and the actual distributional structures of these variants within the community. Specific research questions regarding coarticulation are as follows:

- Do speakers of American English have distinct production goals for contextual variation—one for fronted /u/ and another for canonical /u/?
- Do listeners systematically vary in terms of the amount of perceptual compensation they exhibit?
- What linguistic knowledge guides compensatory perception? Is it purely a perceptual phenomenon or is it linked to a listener's knowledge about one's own speech production patterns or speech patterns in a community?
- How compensation for coarticulation influences the encoding of spoken inputs? Will re-production of perceived speech be guided by compensatory perception?

Specific research questions regarding sensitivity for and encoding of sub-phonemic variation are as follows:

- Are listeners capable of encoding sub-phonemic variations and reproducing these variations as distinct pronunciation patterns? If so,
- How does a listener's sensitivity to sub-phonemic variation interact with compensation for coarticulation?

## 1.4 Theoretical and Methodological Assumptions

In the present study I have adopted basic assumptions shared in the Complex Adaptive System approach for language use (Beckner et al., 2009). In this view, the linguistic behavior of an individual is based on multiple factors such as (1) interactions between the individual and

other speech community members, (2) patterns of past linguistic experience, and (3) various physical and physiological constraints and social motivations. Structures of language are considered to emerge from mutually dependent relationships between social interactions and cognitive processes (Beckner et al, 2009). Although cognitive and biological endowments imply universal properties of language, variation among language users and their social interactions with each other explains the variation in the systems at the population level (Pierrehumbert, 2006). This approach thus implies that language systems and linguistic phenomena are best understood through the interaction between multiple components, calling for a holistic study of the system in its entirety, rather than focusing on isolated parts (Oudeyer, 2005).

In addition, I assume exemplar-based mental representations of speech (Goldinger, 1996, 1998; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002; Wedel, 2006). In exemplar-based models of lexicon words are stored in the mental lexicon as clusters of remembered instances, or exemplars, of words that the listener has experienced and consciously attended to, and each exemplar maintains detailed information of and about that particular instance (e.g., Goldinger, 1996; Johnson, 1997). The exemplar memory can store, with decay, auditory information such as speaker voice, pitch, and other acoustic-phonetic details, situational information such as when and where the remembered speech occurred, and indexical information about the speaker such as gender and dialect (Johnson, 1997, 2006). While some of the exemplar models assume that the abstract sub-lexical category such as phonemes are not stored as such but rather emerge from the pool of exemplars at the time of retrieval (Goldinger & Azuma, 2003; Johnson, 2006), I assume that phonemes are represented as stable and distinct category nodes, rather than just being an emergent properties. This assumption is guided by recent "hybrid" models (Beckman & Pierrehumbert, 2004; Chistovich, Fant & de Serpa-Leitano & Tjernlund 1966; Cutler, 2010; Hawkins, 2003; Hawkins & Smith, 2001; McLennan, Luce & Charles-Luce, 2003; Pierrehumbert, 2006; Sumner & Samuel, 2009). With phoneme nodes encompassing range of phonetic variants, an assumed exemplar-based knowledge is compatible with the notion of phonemes with internal structure (e.g., Miller, 2001; Volaitis & Miller, 1992). In addition, I assume that under phoneme nodes, there are intermediate nodes corresponding to contextual allophones and even some sub-allophonic pronunciation variants, which I call ***pronunciation categories***. This notion is similar to what Keating (1998) calls "categorical phonetic representations" (p. 324), but the pronunciation categories assumed in this study is not feature-based, but exemplar-based. Also, the assumed pronunciation categories are not structurally motivated, but are perceptually motivated categories. Thus, pronunciation variations such as regional and idiolectal accents can be represented as distinct pronunciation categories. Finally, in the present study I assume that exemplars correspond to representations that are discriminable by the listener, which is called ***phonetic representations*** in the present study. This assumption is in accord with a widely shared condition that exemplars are granular units and each exemplar may or may not be identical to the actual stimuli, as the two very similar exemplars that human auditory system cannot distinguish as different instances are stored as an identical exemplar (Johnson, 1997, pp. 152-153; Pierrehumbert, 2001, p. 141).

On the methodology, I assume that the best way to study speech variations in a manner that provides much insight into the preconditions of sound change is to focus on the speakers' and the listener's responses to *pivot* sounds. The term pivot is traditionally used to refer to a basis for analogical change (Hock, 1991, p. 215). For example, In Latin some nouns in a particular

paradigm type were leveled to follow another paradigm pattern when the nominative singular forms share the identical parts, serving as the pivot, as shown in (1):

(1) singular, nominative: *pater* 'father' *socer* 'son-in-law'
singular, genitive: *part-is* *socr-ī* > *socr-is*

In (1), shared *–er* ending between the consonant stem class words (e.g. *pater*) and the *o*-stem class words (e.g. *soccer*) in their singular nominative forms is considered to be the pivot, motivating the change in the singular genitive forms of some of the o-stem words (e.g. *socr-ī* > *socr-is*) to match the paradigm of the consonant stem class (Hock, 1991, p. 217). Recently the notion of pivot has been extended to refer to the critical linguistic form that allowed multiple interpretations about its structural make-up, and therefore gave rise to an innovative structural analysis that resulted in language change (Garrett, in press). For the significance of focusing one's analysis on the pivot construction, Garrett stated as follows: "We cannot understand how one thing has turned into another without locating the pivot context in which the change originated and understanding how the properties of that context invite the change."

I assume that coarticulatory variants become pivots for the potential sound change if the variants allow language users to respond differently, within a range of language-specific constraints. Some possibilities are as follows:

- In speech production, coarticulation can be analyzed (both by researchers and speakers) either as the result of physical and physiological constraints or the speakers' deliberate control.
- In speech perception, heavily coarticulated sounds can be perceived either as a member of the coarticulated variant of its plain counterpart (by normalization) or as a member of its own category.
- In encoding, coarticulated sounds can be encoded either categorically or gradiently.

## 1.5 Overview of Dissertation

Chapter 2 provides an extended theoretical background of the present study. This chapter reviews theories and models of preconditions and triggers of sound change. In this chapter I argue that studies from these two areas of inquiry mutually benefit each other and thus equally significant for holistic understanding of sound change. The next three chapters report a series of three experimental studies. Since each study uses a different method and experimental paradigm, each chapter first provides its own literature review for a conceptual framework of the study. Chapter 3 reports a production study, which was prompted by a question of whether or not in American English fronting of /u/ in alveolar contexts has been phonologized. The chapter begins with attestations of synchronic phonological patterns and historical sound change that are linked to coarticulatory fronting of /u/, and then reviews the phonetic bases of coarticulatory /u/-fronting. The chapter then describes and reports the results from the production study with the method used in Lindblom (1963), Solé (1992), and Solé and Ohala (2010). The results will be

discussed in the light of articulatory phonology (Browman & Goldstein, 1986, 1990, 1992) and exemplar- and usage-based phonology (e.g., Bybee, 2001; Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2001).

Chapter 4 reports a perception study, which investigated systematic individual variation in compensation for coarticulation as well as factors that affect the amount of perceptual compensation. The main question was "is compensation unanimous and complete?" The chapter begins with a review of known factors that influence speech perception. The chapter then describes and reports the results from the perception study using a phoneme classification paradigm (Harrington et al., 2008; Lindblom & Studdert-Kennedy, 1967; Mann & Repp, 1980; Ohala & Feder, 1994). The results will be compared with the findings from previous compensation studies and from Production Study (Chapter 3). A potential link between listener judgments on coarticulatory variation and exemplar- and usage-based knowledge (e.g., Bybee, 2001; Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2001) about the normative range of pronunciation variation in different contexts will be discussed.

Chapter 5 reports a vowel repetition study, which explored how listeners respond to continuous vowel sounds in fronting and non-fronting contexts, using the vowel imitation paradigm (Alibuotila, Hakokari, Savela, Happonen & Aaltonen, 2007; Chistovich et al., 1966; Kent, 1973, 1974, 1979; Repp & Williams, 1985, 1987; Schouten, 1977; Vallabha & Tuller, 2004). The results will be discussed in terms of categorical vs. continuous mode of speech perception (Chistovich et al., 1966; Liberman, Harris, Hoffman & Griffith, 1957; Pisoni, 1973a, b; Repp, 1984).

Finally, Chapter 6 presents a summary of the research questions and findings. In this chapter, I propose a model of speech perception and an extended model of the speech chain, both of which are characterized by (1) experience-based phonological knowledge, (2) stability and instability in pronunciation norm in the speech community, and (3) mutual dependency and mutual causality between and among speech perception, speech production, phonological knowledge, and ambient language data. The proposed models illustrate how these properties in human language and human language use govern the output of communicative interactions among members in a speech community. One such output, I argue, is knowledge of multiple sub-phonemic pronunciation categories that speech community members have. Further, I argue that any speech community is in a constant state of readiness to respond to a trigger and adopt an innovative pronunciation as a new community norm, because members have rich pronunciation repertoire even when there is no observable community-level sound change.

## 1.6 Definition of Key Terms

The following is a list of key terms that will be used in this dissertation. Each definition is accompanied by clarification of concepts that are particularly relevant for the present study.

Coarticulation The overlapping of articulation of neighboring speech sounds. As a result, articulation and acoustic quality of speech sounds are influenced by the phonetic context in which the sounds occur. It is a universal phenomenon due to physiological constraint and one major source of production variation.

Initial Change The present dissertation uses the term Initial Change to refer to an event wherein a *mental representation of speech sounds* (see below) is obtained by a single person upon hearing a single utterance, but the obtained mental representation differs from the representation encoded by the speaker in the utterance. I assume that this change occurs mainly in auditory representations, by hearing one's own (a listener = a speaker) or someone's utterances.

A similar notion is expressed by a more commonly used term "innovation," which refers to a single person's usage (or grammar) that differs from the previous usage (or grammar) (Janda & Joseph, 2003, p. 13). However, Initial Change does not necessarily involve a new pronunciation being used by the listener in his or her subsequent speech, or the new pronunciation of a particular phoneme becomes generalized for all words that contain the same phoneme in the mental lexicon of that listener. It only consists of an event that the new pronunciation for a particular word enters into a single person's lexicon as an alternative pronunciation of that particular word (cf. § 4.7.2).

Mental representation of speech sounds A decodable mental object based on the language user's subjective experience with the speech sounds. This is a theoretical construct rather than a directly observable object of study. Internal subjective mental representation is not completely understood by biology; however, it is mediated by neuronal representations and as such it must have a statistical relation to both the input and the output (deCharmes and Zador 2000). This statistical relation is the guiding principle in the present study. For example, in the repetition study reported in Chapter 5, it is assumed that if continuous inputs are repeated categorically, then such output patterns (with particular distributions of acoustic parameter values) would provide evidence for categorical representations for speech inputs.

The present study assumes two types of representations. One is a representation of the input stimuli, or what the brain holds during the stimulus-response process. This representation can be mapped onto an articulatory representation if articulatory response is called for or evoked (cf. Chapter 5). The other is a representation of previously encountered linguistic data. This is what the brain holds for a long term so that the processing system can use it as a reference when performing cognitive tasks such as phoneme identification, sound discrimination, etc. There are both auditory and articulatory representations of this type. Finally, this study assumes multiple levels of representation for speech sounds: minimally, these are phonetic and phonological representations. In addition, this study assumes that some of the sub-phonemic units can also be represented (see *pronunciation category* below).

Misperception Misperception occurs when a listener perceives a normally produced utterance (this excludes the case of speech errors) and arrives at a mental representation of the utterance that is different from what the speaker assumed at the time of the utterance (Baudouin de Courtenay, 1910/1972; Ohala, 1981, 1989, 1993). This process does not necessarily involve failure to detect or confuse acoustic signal of an utterance, as an object of misperception is speaker intent and not an acoustic signal. Whether a listener has misperceived a speaker's intent or not can be determined only by comparing the mental representation of an utterance that the listener arrives at with the mental representation that a speaker encoded at the time of utterance. This study conceptualizes misperception as an event within the *speech chain model* (see below).

Perceptual compensation for coarticulation An effect of segmental context on speech perception whereby a listener perceives coarticulated sequences of sounds as if there is no co-articulation. It is one source of perceptual constancy.

Pronunciation category An assumed unit of representation. As discussed in Section 1.4 (Theoretical and Methodological Assumptions), the present study is based on the assumption that knowledge of speech sounds can be conceptualized as three layers of pre-lexical representations in the long-term memory: these layers are for (1) phonetic representations, (2) pronunciation category representations, and (3) lexical phoneme representations. In this assumption, lexical phonemes encompass multiple pronunciation categories/variants that do not make lexical contrast but are recognized by language users as distinct articulatory and discernible auditory categories, and thus more abstract than phonetic representations. These pronunciation categories include contextual allophones (cf. Keating, 1998) as well as dialectal and individual pronunciation variations. In the context in which language user's performance is considered (over the phonetic properties of the sound), the term *pronunciation category* or *pronunciation variant* is used. When the articulatory and acoustic properties are concerned, the term *sub-phonemic variation* is used.

These three types of representations might be also called as phones, phonetic categories, and phonemes, but the use of term pronunciation category over phonetic category highlights (1) distinct cognitive status of this representation than what the term phonetic might suggest and (2) exemplars as a basis of mental representation. Also, the term phonetic representation over phones highlights the fact that any mental representations are the results of perceptual processing and therefore distinct from raw acoustic patterns.

Production target An assumed mental representation. This study assumes that each speaker has a unique set of articulatory and auditory representations in the long-term memory for speaker's own speech. This assumption entails that community pronunciation norm and one's own production target are not necessarily the same.

Sound change The change in the pronunciation of words, but only those changes for which it is possible to hypothesize phonetic constraint(s) as a root cause for the emergence of new pronunciation patterns. Changes in pronunciation due to analogy and other non-systematic processes such as clipping, folk etymology, etc. are thus not considered as instances of sound change. This dissertation concerns only those changes that are considered to have arisen form a pool of coarticulatory variations. Further, this study has adopted the distinctions between innovation—a change within a single person's usage (or grammar) and change—the adoption of an innovation by all or at least much of community members (Janda & Joseph, 2003, p. 13). Thus the term sound change is used in this dissertation only for a community-level change.

Speech chain A chain of events between a speaker-end and a listener-end in speech communication, as well as a model of a communication system whereby a speaker's message is recognized in the listener's brain. The term and the model were introduced by Denes and Pinson (1973). The model consists of (1) a linguistic level, where the intended message is put into a linguistic form in the speaker's brain, (2) a physiological level, where the brain's instructions in the form of the nerve impulses set the speech organs into movement, (3) an acoustic level, where movement of the speech organs create pressure variations in the atmospheric air (a.k.a. 'sound

wave'), (4) a physiological level at the listener's side, where pressure changes are transformed into nerve impulses that travels to the brain, and (5) a linguistic level, where nerve impulses are interpreted in the listener's brain (pp. 5-6). In this model the speaker is also one's own listener, as the acoustic signal of the speech is fed-back to the speaker's ears (auditory feedback). Viewing speech communication in terms of articulation, acoustics, and cognitive processes for planning, motor control, and speech perception (both by the speaker and the listener), the model helps in identifying various types of constraint at different level in the chain.

Speech Perception A process by which the input stream of acoustical patterns is interpreted by the brain as a sequence of one or more of linguistic codes (phoneme, syllable, word, etc.). This process is achieved by two (potentially interrelated) systems—the peripheral system that detects change in the spectro-temporal pattern in an acoustic waveform (sensation) and the central system that interprets input from the peripheral system (perception).

# Chapter 2

# Theoretical Background

This chapter has three goals: one is to provide a theoretical background in which three main themes of the present study—variation in speech production, variation in speech perception, and the interrelationship between the two—have emerged. Another is to review some major models developed in the studies of sound change. The last is to argue for the mutual benefit obtained from a study on the preconditions of sound change and a study on specific "triggering events" (Labov, 2010, p. 90) of sound changes to each other. This chapter is organized as follows: the chapter first reviews models of preconditions of sound change, starting from the models that account for a change that is triggered by producing or hearing a single utterance (§2.1), then proceeding to the models of transmission of change (§2.2). Next, the chapter briefly reviews theories on sound change that focus on community- and language-internal factors (§2.3). Finally, the chapter presents the link between preconditions and triggers of sound change, and argues that a study on preconditions and a study on triggers are mutually useful for each other's advancement (§2.4).

## 2.1 Models for Initial Change

In the late 19th century a group of German linguists (Neogrammarians: Junggrammatiker) proposed the Regularity Hypothesis, which states, "sound change (Lautwandel), in so far as it operates mechanically, proceeds according to exceptionless laws" (Osthoff & Brugmann, 1878/1967, p. xii]). Since then many scholarly works have been devoted to explain why sound change occurs and how it achieves characteristic regularity. One approach to this problem has been to focus on the fact that all language changes over time and to seek explanations in general properties in language and language use.

Some models that were developed in this general approach were designed to account for a change in mental representation of speech sounds that may occur when producing or hearing a single utterance. This type of change, which would be called *Initial Change*, is the theme of the first five models reviewed in this section.

## 2.1.1 Articulatory Drift: Paul (1886/1970)

In the third chapter, "On Sound Change" ("Der Lantwandel"), of *Principles of the History of Language* (*Prinzipien der Sprachgeschichte*), Paul (1886/1970) discussed the physiological and cognitive bases of learning how to speak a language and argued that these learning mechanisms serve as the driving force of sound change. Paul recognized three different types of change—(1) the change that is caused by the alteration of articulatory representation, (2) the change that occurs in the transmission of sounds (i.e., the creation of innovative pronunciation), and (3) the change that are caused by repeated mispronunciations (pp. 54-55). His main concern was on the first type, which would be called the *articulatory drift*.

Two components of learning, according to Paul, were: (1) the movement of speech organs that causes "motory" (i.e. articulatory) sensation, and (2) the speaker's hearing the speech of others and one's own, which causes auditory sensation. He suggested that these sensations, after they are gone, live in the speaker's memory as what he called a "memory-picture" (i.e. mental representations), and it is this articulatory representation that enables a speaker to reproduce similar articulatory movement (pp. 36-37). For Paul, then, the key element in learning to speak a language is to develop articulatory and auditory representations of the language's speech sounds.

Paul conjectured that the cause of sound change is the articulatory representation in one's memory, which is constantly modified each time one speaks, with recent sensations having a greater influence than earlier ones. He assumed that the shift in representation would occur with certain directionality rather than at random, because certain sound sequences are easier to articulate than others (e.g. Italian word *otto* is easier to pronounce than Latin *octo*) and this facilitation provides a bias toward a particular directionality in sound change (pp. 44-47).

Paul also conjectured that the auditory representation of speech sounds, which is based on all the sounds that a speaker has ever been exposed to, serves as a reference for the speaker in controlling one's speech (pp. 48-49). He also suggested that as speakers have a general desire to conform to the pronunciation of other members in the speech community, the auditory representation provides a safeguard for the community norm of pronunciation. As a consequence, changes that occur in one generation would be slight and the greater change would occur in the transmission of speech to new individuals (pp. 53-54).

Paul's approach to sound change has been criticized on several grounds. First, as articulatory bias is an inherent factor in speech production, it fails to explain why sound change occurs only in one particular time or only in some languages (Weinreich, Labov & Herzog, 1968, pp. 111-112). Second, it does not offer any mechanism that promotes a regularization of the change, where a word change becomes a phoneme change (Wheeler, 1901, p. 13). Third, it does not explain why a particular degree and variety of assimilation occur in a given language (Hock & Joseph, 1996, p. 146). Finally, it does not explain why, during successive transmissions, the deviation would accumulate in one direction (Hock & Joseph, p. 147; Weinreich et al., 1968, p. 112). These critiques highlight that sound changes are complex phenomena, and Paul's model explains only a fragment of the phenomena.

Recently two assumptions on which Paul based his conjecture gained empirical support. These assumptions are: (1) that auditory and articulatory (i.e. somatosensory) feedbacks are used in controlling speech production (Guenther, 2006; Houde & Jordan, 2002; Katseff, 2010; Tremblay, Shiller & Ostry, 2003); and (2) that a speaker matches his or her production to what

he or she has heard from the ambient language (Harrington, 2006; Sancier & Fowler, 1997). These findings support the following hypotheses that are directly drawn from Paul's model: (1) an individual would acquire a unique pronunciation habit by repeatedly using the similar articulation; (2) the individual pronunciation habit does not deviate from the community norm; and (3) during the time of on-going sound change, speakers would adapt to gradually shifting to community pronunciation norm. Further, his conception of incomplete learning adequately accounts for the individual variation in phonological grammar. Although Paul's focus was on a single event of articulatory drift, his work offered many useful concepts for the subsequent scholarly inquiry on sound change.

### 2.1.2 Misperception: Ohala (1981, 1989, 1993)

While Paul proposed a speaker-based theory, Ohala proposed a listener-based theory of change (Ohala 1981, 1989, 1993), which states that a sound change originates in listener "misperception"—an event wherein a listener misattributes phonetic features in the acoustic signal to a segment not intended by the speaker, and arrives at a representation that is different from what is encoded by the speaker.[1] This hypothesis is in agreement with that of Paul in that the first step toward a community-level sound change is a change in the mental representation within individuals. But Ohala focused on the change that occurs in a transmission of speech, emphasizing the importance of speech perception (the act of a listener) over speech production (the act of a speaker).

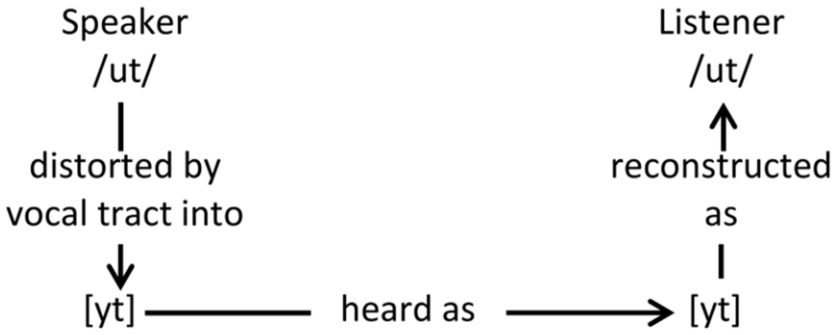
Ohala hypothesized that at the encounter of aerodynamically and physiologically conditioned speech variation listeners can do one of the following: (1) normalize predictable variations and arrive at the pronunciation intended by the speaker, or (2) fail to correct the distortion in the speech signal and take the phonetic form as its intended pronunciation. The second case leads to a type of misperception, which Ohala termed as "hypo-correction." These two cases are schematically represented in Figure 2.1 (A and B).

In the first case, a speaker intends to utter /ut/, but the segment /u/ is phonetically realized as a front vowel [y] due to coarticulatory fronting (cf. §3.1). A listener perceives the acoustic signal that represents [yt], but correctly attributes the [+front] feature on the vowel to the following context, interpreting the vowel as /u/ as intended by the speaker. In the second case, on the other hand, the listener fails to detect the coda [t] thus attributes the feature [+front] to the vowel itself, interpreting the entire utterance as /y/. When the listener utters the same word at a later occasion, the listener utters /y/, not /ut/.

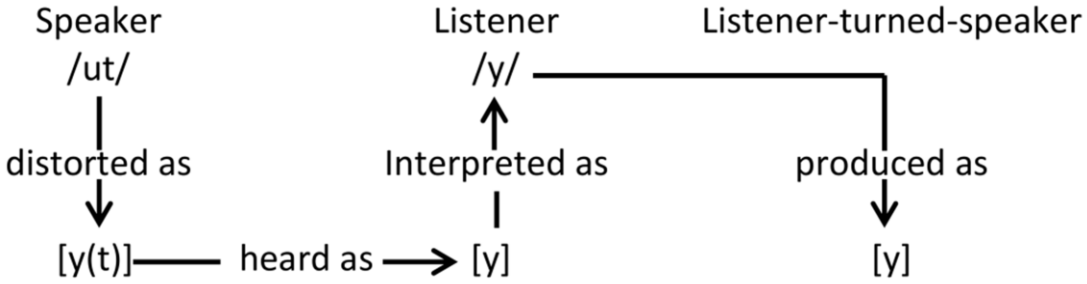
While assimilative sound change is explained in terms of hypo-correction, dissimilative sound change calls for another mechanism, which Ohala termed as "hyper-correction," which is illustrated in Figure 2.1 (C). A speaker intends to utter /yt/, which is realized as [yt]. A listener perceives the acoustic signal that represents [yt]. Since the listener knows that /ut/ is usually

[1] Baudouin de Courtenay (1910/1972, pp. 267-268]) has already named misperception (*lapsus auris*) as a factor of change, and the similar idea was expressed by other scholars (e.g. Jonasson, 1971; Paul, 1888/1970, p. 54; Sweet, 1888, p. 16) but these scholars' contributions were philosophical in nature, while Ohala proposed specific models with associated directions of change.

(A)




(B)




(C)

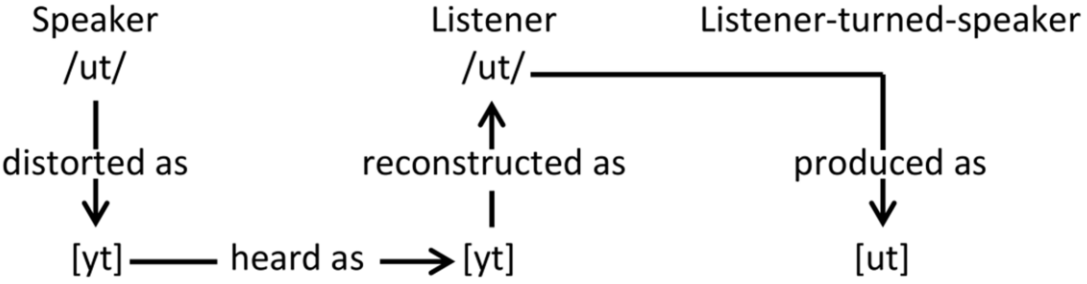



Figure 2.1 Schematic representations of the processes involved in (A) perceptual correction, (B) hypo-correction, and (C) hyper-correction. Adapted with permission from "The listener as a source of sound change," by J. Ohala, 1981, *Proceedings from CLS 21,* p. 182. Copyright 1981, Chicago Linguistic Society.

realized as [yt], the listener thinks that the speaker intended to utter /ut/. When the listener utters the same word at a later occasion, the listener aims to produce /ut/, not /yt/.

Thus, given the inherent ambiguities and contextual variations in natural speech, Ohala's models predict that

(1) A correction process prevents misperceptions; and
(2) Hypo- and hyper-corrections cause misperceptions, or "mini" sound changes.

These models are not meant to account for the triggering event of any specific community-level sound change. Rather, the models account for different analyses used in perceptual interpretations of speech sounds across individuals, which results in individual variation in encoding the phonetic pronunciation variation. Thus the models explain how a given word form may be uttered in different pronunciations across members in a speech community—an assumed precondition of sound change.

Bodies of evidence support the three models. Perceptual correction of the speech signal by listeners has been well documented in compensation for the coarticulation research (Beddor & Krakow, 1999; Elman & McClelland, 1988; Harrington et al., 2008; Lindblom & Studdert-Kennedy, 1967; Mann, 1980; Mann & Repp, 1980; Ohala & Feder, 1994; Repp & Mann, 1981), and other studies on listener sensitivity to covariations between phonetic properties in natural speech (e.g., the voicing of onset consonant and F0 in the following vowel (Fujimura, 1971), speech rate and rate in formant transition (Lindblom & Studdert-Kennedy, 1967; Miller & Liberman, 1979; Newman & Sawush, 1996). The loss of the conditioning context as predicted in the hypo-correction model is in accord with the development of the distinctive nasal vowel from the historical vowel-nasal sequence in French and the development of tonal distinctions out of former voiced vs. voiceless contrasts on prevocalic consonants in the East and Southeast Asian languages (Matisoff, 1973). Finally, laboratory studies have shown listener misperceptions that follow the patterns predicted in the hyper-correction model (Ohala & Busà, 1995: Ohala & Shriberg, 1990).

A remaining issue for Ohala's theory is a specific causal mechanism of hypo-correction when the context is retained. Guion (1996) pointed out that velar palatilization (/k/>/ʧ/) before high vowels and glides, which is a common sound change and commonly occurs without the loss of the conditioning environment, presents a challenge for the theory. Ohala considered the lack of experience of various contextual variations that enables a listener to do the perceptual correction as a cause of hypo-correction in such a case, and proposed that children acquiring language and adult second-language learners would hypo-correct for this reason. However, an assumption that this type of misperception occurs ONLY in immature listeners has been challenged by evidence showing that adult listeners exhibit incomplete correction (Beddor & Krakow, 1999). A mechanism that accounts for the /ut/ > /yt/ type of misperception in mature native listeners would be a welcome addition to the current hypo-correction model.

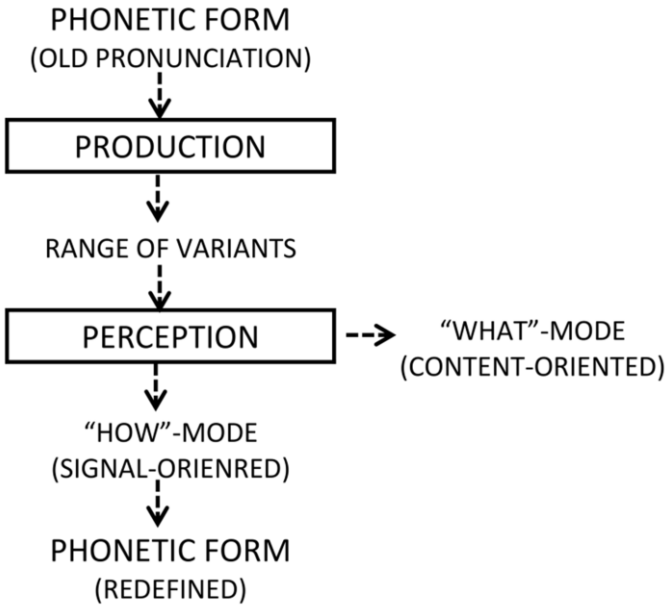
## 2.1.3 Variation-Selection Model: Lindblom et al. (1995)

While Ohala's model attributed the discrepancy in pronunciation targets between a speaker and a listener to the listener's misperception, the same outcome was attributed to different modes of perception in Lindblom, Guion, Hura, Moon, and Willerman's (1995) model, which would be called the *variation-selection model*. This model shares the functionalist tenet that sound changes are cures for a linguistic system—the system that strives towards an ideal balance between conflicting pressures from "the law of lesser effort," in speech production and "the law of greater effort," for intelligibility (Grammont, 1939, p. 176). As such, sound changes must not be allowed to conflict with communicative need or disrupt useful phonemic oppositions (Martinet, 1952, p. 5).

The variation-selection model was based on Ohala's theory of misperception and Lindblom's (1983, 1990) theory of Hype-Hypo Speech (H&H), which maintains that speakers can adapt to the particular communicative situations (e.g. formal, casual, etc.) and adopt different strategies to control the degree of coarticulation along a hyper/hypo-speech continuum. The variation-selection model has the following two-step mechanism: the first step (Fig. 2.2 (A)) involves a pronunciation variation along the hyper/hypo-speech that arises from the speaker's adaptation to communicative needs. The first step also involves a perception variation between a context oriented "what"-mode of perception, whereby a listener pays more attention to what is said over pronunciation, and a signal oriented "how"-mode of perception, whereby a listener pays more attention to pronunciation over informational contents. Given these sources of variation, Lindblom and his colleagues claimed that it is this incidental "how"-mode of, or decontextualized, perception that enables a listener to posit a mental representation close to the raw acoustic form of speech sound, with its coarticulatory distortion, reduction, etc. When an innovative pronunciation is selected and used by a listener at a later time, this form will be further filtered (i.e. selected or rejected) by the rest of the speech community (Fig. 2.2 (B)). Thus given a theoretically motivated range of pronunciation variation, the variation-selection model accounts for how these variants are either normalized (or ignored) or faithfully encoded in the language user's mental lexicon (i.e. the "mini" sound change in Ohala's term), and how the latter case still may or may not become a community-level sound change.

It is noteworthy that the model assumes that the native speaker's long-term memory holds multiple representations of a given linguistic item, including "canonical (should-be)" pronunciations and relatively unprocessed phonetic forms (p. 17). This assumption is supported by the fact that speakers are capable of altering their pronunciations depending on social factors such as the formality of the communicative setting (Fisher, 1958; Labov, 1966; Schilling-Estes, 2002; Trudgill, 1974) and addressee (Bell, 1984; Bradlow et al., 2003; Coupland, 1984; Fernand et al., 1989; Hay et al., 1999; Kuhl et al., 1997; Smith, 2007; Uther et al., 2007). The variation-selection model thus predicts that not only children acquiring language but also adult members can initiate sound change, because even if a listener already knows the canonical pronunciation of a particular word, this listener still has a chance to acquire an innovative form if the "how"-mode of perception allows them to capture a raw acoustic pattern. This prediction is in accord with the findings from studies on sound change (Labov, 1994, 2001, 2010) and dialect adaptation (Maye, Aslin & Tanenhaus, 2008; Sumner & Samuel, 2009).

(A)




(B)

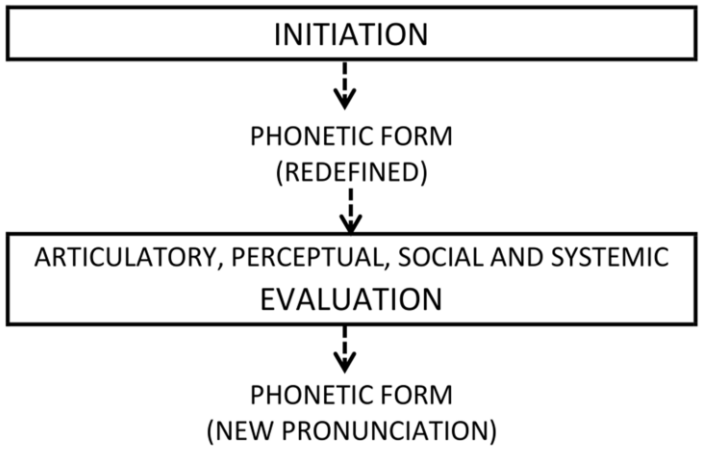



Figure 2.2 Schematic representations of the two-steps process in the variation-selection model—(A) variation in production and perception, and (B) selection of innovative forms. Adapted from "Is sound change adaptive," by B. Lindblom, S. Guion, S. Hura, S.-J. Moon, and R. Willerman, 1995, *Rivista di Linguistica,* 7, p. 16. 

A remaining issue for the variation-selection model is the precise mechanism of the "how"-mode of perception and its validity. In laboratory speech perception experiments, listeners tend to fail to capture the raw acoustic signals even though the listeners presumably engage in their tasks with the "how"-mode of speech perception. The listeners' perceptual responses are influenced by various factors such as the length of stimuli (Fujisaki & Kawashima, 1969, 1970, as cited in Pisoni, 1973a), the perceived speech rate (Lindblom & Studdert-Kennedy, 1967; Miller & Liberman, 1979; Newman & Sawush, 1996), the segmental context in which the target sound occurs (Beddor & Krakow, 1999; Elman & McClelland, 1988; Harrington et al., 2008; Lindblom & Studdert-Kennedy, 1967; Mann, 1980; Mann & Repp, 1980; Ohala & Feder, 1994; Repp & Mann, 1981), the acoustic characteristics of a precursor sentence, (Ladefoged & Broadbent, 1957; Ohala & Shriberg, 1990) and even by the perceived talker's identity (Hay Warren & Drager, 2006; Johnson 1990; Johnson, Strand & D'Imperio 1999). Regardless of this issue, the idea that variation in the mode of perception results in multiple mental representations of a given lexical item (and presumably any other linguistic units) is appealing. When different modes of perception are fully specified and accompanied with an empirical support, the model would become a powerful tool.

## 2.1.4 CCC Model: Blevins (2004)

Recently, Blevins (2004) proposed a theory called Evolutionary Phonology, the central tenet of which is that "recurrent synchronic sound patterns have their origins in recurrent phonetically motivated sound change" (p. 8). The theory provides the following three mechanisms of misperception in a situation of a single speaker-listener interaction (pp. 32-33):

CHANGE "The phonetic signal is misheard by the listener due to perceptual similarities of the actual utterance with the perceived utterance."

CHANCE "The phonetic signal is accurately perceived by the listener but … [the] listener associates a phonological form with the utterance which differs from the phonological form in the speaker's grammar."

CHOICE "Multiple phonetic signals representing variants of a single phonological form are accurately perceived by the listener, and due to this variation, the listener (a) acquires a prototype or best exemplar of a phonetic category which differs from that of the speaker; and/or (b) associates a phonological form with the set of variants which differs from the phonological form in the speaker's grammar."

This model may be better understood if one assumes that decoding spoken input into linguistic representations proceeds through a series of mappings and abstractions as follows: (1) the mapping of acoustic properties of speech input onto phonetic representations; (2) the mapping of the phonetic representations onto phonological representations; (3) the mapping of the linearly ordered phonological representations onto a word form. In this conception of speech perception, CHANGE occurs at the first stage of this process: acoustic properties are mapped

onto wrong phonetic representations due to perceptual confusion.[2] CHANCE occurs at the second stage: a given phonetic representation or a sequence of phonetic representations is mapped onto different phonological interpretation depending on the particular phonological grammar that is applied at this stage of processing, and the listener reconstructs a phonological representation that differs from what the speaker has encoded. The suggested mechanism for this discrepancy is "innocent misperception" in a sense of Sweet (1888, p. 16), Baudouin de Courtenay (1910/1972, pp. 267-268), and Ohala (e.g. 1981, 1983, 1990). Finally, CHOICE creates the same outcome as CHANGE does: its input is a faithfully represented phonetic form and its output is a phonological representation that differs from what the speaker has encoded at the time of utterance. A defining property of CHOICE is its representation-based mechanism, while CHANGE and CHANCE are process-based mechanisms (cf. Chapter 4 and 5 for reviews of process-based and representation-based models of speech perception). CHOICE assumes that phonological representations are prototypes or the best exemplars, specified by all stored phonetic forms. Therefore, the outcome of CHOICE depends on distributional properties of the stored phonetic forms in one's memory rather than the perceptual processing itself.

A unique feature of the models is that after acoustic signals are mapped onto phonetic representations, these forms can subsequently go through either CHANCE or CHOICE for further transformation. In this regard, the CCC model can be seen as an elaborate equivalent of the variation-selection model with CHANCE and CHOICE replacing the "how"-mode of speech perception. A remaining issue in this model is the precise mechanism of CHANCE: exactly what grammatical analysis leads to CHANCE?

## 2.1.5 Perceptual Grammar Model: Beddor (2009)

Beddor (2009) proposed a conceptual model of the listener's role in sound change, in which variation in the perception grammar plays a central role. This model is based on the observation that the temporal and spatial extents of coarticulatory effects are highly variable and that listeners vary in terms of their sensitivity to the multiple acoustic cues that co-occur in a given sequence of sounds. In this model, a given acoustic pattern has different representations across listeners, depending on the particular weights each listener assigns to each of the multiple acoustic cues. As a result, phonological representations for a given phonological unit, which the model assumes to encompass the full range of phonetic variation, vary across listeners in terms of the exact range of variation that the representations encompass. Individual variation in phonological grammar, according to Beddor, is a natural outcome of daily exposure to highly variable speech signals: "because multiple representations are consistent with the variants found in everyday communicative interactions, a given listener need not arrive at the same representation—and/or need not arrive at the same perceptual weighting of the acoustic properties that map to a representation—as that of other listeners, or as that of the speaker" (p. 788).

Beddor (2009) provided empirical support from a series of case studies on coarticulatory vowel nasalization in sequences of a vowel, coda nasal, and oral consonant (/VNC/) in American

[22] Whether this occurs at the peripheral or central processing system is not a crucial issue here.

English. Her production data revealed that the speakers employ relatively constant velum lowering gestures but temporally align the velum gestures variably relative to the oral gestures, resulting in an inverse relationship between the duration of the nasal consonant and the nasalized portion of the vowel. Listeners discriminated poorly between pairs of stimuli that had different nasalized vowel-to-nasal coda duration ratios ([Ṽ]-to-[N] ratio) but similar total nasalization durations, indicating that the listeners treated vocalic and consonantal nasality as perceptually equivalent. Additionally, two-alternative-forced-choice /CVNC/ vs. /CVC/ tasks that tested listeners' nasality judgments revealed that some listeners placed long-Ṽ short-N stimuli in the /CVNC/ category while others placed them in the /CVC/ category, showing that listeners vary in their perceptual weights on the relevant acoustic cues.

## 2.2 Models for Transmission of Change

While the above five models were designed to account for a change caused by a single utterance, other models were designed to account for the effect of exposure to multiple utterances through language acquisition and language use both at an individual level and at a community level. This type of long-term effect, which may cause both individual variation and adaptation to change, is the theme of the three models reviewed in this section.

### 2.2.1 Gradual Shift Model 1: Hale (2003)

Hale (2003) modeled a "change event" as "the set of differences between the grammar generating the primary linguistic data used by an acquirer and the grammar ultimately constructed by that acquirer" (2003, p. 345). In Figure 2.3, $O_1$ represents the primary language

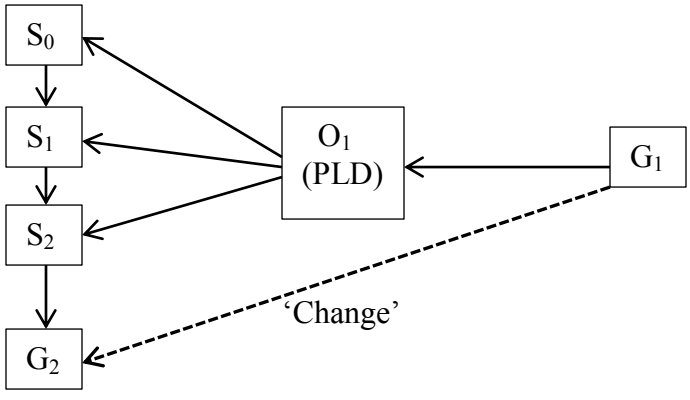



Fig. 2.3 The nature of "change event" Adapted with permission from "Neogrammarian sound change," by M. Hale, 2003, *The handbook of historical linguistics,* p. 345. Copyright 2003, Blackwell Publishing.

data produced by the source grammar ($G_1$), from which the language acquirer generates the grammar. This is a random subset of PLD, which differs from another random subset of PLD (not shown in the figure) from which $G_1$ had been generated. $S_0$ is the initial state of the acquirer's knowledge (Universal Grammar); $S_1$ and $S_2$ denote the intermediate states; and $G_2$ is the acquired grammar.

In this model the change event is a direct consequence of a random variation in sampling: $G_2$ differs from $G_1$ because the data (the subset of PLD) presented to the acquirer of $G_2$ is different from the data presented to the acquirer of $G_1$. This conception is similar to Initial Change in that the change event occurs within a single language user, but Hale's model captures more stable and robust differences among the grammars possessed by community members. The same conception of sampling-based variation in grammars was also expressed in Paul (1886/1970) with emphasis on the point that "the whole consideration is how often he hears them" (p. 52). Since no two individuals share exactly the same experience, this model predicts that everyone in a given speech community has slightly different grammar and different linguistic behavior.

## 2.2.2 Gradual Shift model 2: Labov (1994, 2007)

Labov (1994, 2007) modeled native language acquisition by children in a community with the on-going sound change, where continuous directional change in pronunciation occurs across generations by means of a "transmission of change". In this model, the transmission of change is carried out by native-language acquiring children when they faithfully learn the adult system including its variable elements and then further advance the variable elements in the direction indicated by the age vectors, thus incrementing the change in the direction that has been set by the previous generation (2007, p. 346). The key component in this model that accounts for directionality of sound change is a stratified ambient language data, from which the language learner can infer correlation between the speaker age and realization of certain linguistic variable.

The model successfully accounts for the data from apparent time studies, in which the systematic difference in the pronunciation in the speakers of different age groups is interpreted as evidence of generational sound change in progress. For example, in Atlas of North American English (Labov, Ash & Boberg, 2006) the data on the Northern City Shift (NCS) in Northern Illinois shows significant correlation between the speaker's age and the degree of NCS characteristics such as the raising of /æ/, the fronting of /o/, the backing of /e/, and the backing of /ʌ/: The younger the speaker is, the greater the degree of these features are realized. As the model predicts, children in a community that has on-going sound change are capable of learning not only static patterns of sounds (i.e. the community's norm in pronunciation) but also dynamic patterning of sounds (i.e. the trend in variable norm as a function of age group).

## 2.2.3 Gradual shift in Exemplar Model

Recently, attempts have been made to model the gradual shift for articulatory target by using the exemplar theory (e.g., Garrett & Johnson, in press; Johnson, 1997; Pierrehumbert, 2001;

Wedel, 2006). In some applications, gradual drift has been implemented as a systematic articulatory bias toward a certain direction in a production-perception loop (Pierrehumbert, 2001; Wedel, 2006). For example, Pierrehumbert (2001) simulated the lenition effect by applying a bias factor to a phonetic articulatory target for each round of production. Since this production is fed back to the same exemplar cloud by a feedback loop, exemplar distribution is affected by altered articulation, and the subsequent articulation of the same sound will be incrementally more lenited. Wedel (2006) modeled phonologization of the secondary contrast at the time of neutralization of the primary contrast (e.g. non-distinctive vowel length contrast becoming a distinctive feature between [kæt] and [kæd] when coda voicing is neutralized). This was implemented by amplifying the distinctness of the secondary feature variation for the tokens that have weakly implemented primary feature, gradually shifting the functional weights of the original primary and secondary distinctions along the cycles of perception-production iterations (Wedel, 2006). In a recent simulation study, Garrett and Johnson (in press) showed how a particular sound change can occur in one social community but not in another, depending on whether phonetically motivated/biased variant is taken for its face value or disregarded by the community members. Crucially, Garrett and Johnson attributed differential sensitivity to phonetic variation to the strength of language user's desire to signal one's social identity: "[s]peakers who seek to identify with the group may be more likely to notice phonetic variation among group members and thus include it in as s group indexical property." These simulations are compatible with not only diachronic interpretations for the progress of changes along the time course but also synchronic interpretations that more frequently used items exhibit greater degrees of variability and changes, than less frequently used items. These models' behaviors are in accord with the reported word-frequency effect on sound change (Bybee, 2000; Jurafsky et al., 2001).

Exemplar-based memory itself does not explain why the phenomena that the model represents occurs: for example, it does not explain why in the pre-sound change time the particular bias factor was not active, the balance between primary and the secondary contrast was maintained, or particular phonetic variant was not assigned any social indexical value. The memory itself only learns and implements the change when it takes place. However, exemplar-based models capture the plasticity of individual knowledge and knowledge-based performance. This property makes the exemplar-based model an extremely powerful tool that can be used to represent, among others, the current state of grammar and linguistic behavior of language users both at an individual-level and at a group-level, and predicts how the knowledge and the behavior would change in the future.

## 2.3 Triggers vs. Preconditions

The nine models reviewed in Section 2.1 and 2.2 are mutually compatible, collectively offering a coherent theoretical framework based on the shared principle of VARIATION in speech. Further, these models have adopted the same assumption that the speaker's knowledge about speech sounds, particularly the mental representation of speech sounds, forms a flexible system, allowing representations to be altered through language use. Each model accounts for

different ways in which language users exhibit variation (e.g. in production, in perception, in mental representations, in grammar), different sources of variations (e.g. perceptual analysis, attention, past linguistic experience), and the way language users react to variations (e.g. normalization, overlook, incorporation in the unified representation, addition of a new representation, etc.).

These models, however, can only explain how Initial Change (either by articulatory drift or by misperception) occurs or how community members adapt to on-going sound change. These models do not explain why an entire community adopts a specific innovation at a specific time. Triggering events of specific community-level sound changes (or "antecedent causes" as Labov regarded them as the *explanation* of the change (2001, p. xiv)) must be sought within social and structural domains.

## 2.3.1 Social Factors in Sound Change

Labov (1963, 1972) demonstrated how the use of the centralization of diphthongs (/ay/ and /aw/) in Martha's Vineyard correlated with Islander's social identity: middle aged Up-islanders—the community leaders who had maintained the maritime tradition, (and were also antagonistic to the temporary summer residents from the mainland)—used centralized variants much more frequently and in a more exaggerated manner than Down-islanders.

Sociolinguistic studies have reported that major triggers of sound change are changes in the demographic make-up in the speech community (caused mainly by immigration) and change in the relative social status of the community members (Labov, 2001, pp. 503-510). The latter cause has been also documented to interact with gender difference in the preferred tool for social "indexical work" (Eckert, 2008). Women display their social identity through symbolic resources such as language more than men do (Eckert, 1989). This difference between women and men in how they go about social negotiations explains why women are the leaders in the strategic use of "nonconformity" in socially "constructive" ways (Labov, 2001).

Another major triggering event that Labov (2010) proposed was the change in phonological structure caused by a preceding sound change (pp. 89-119). One of the cases discussed by Labov was the fronting of /uw/ in most parts of North American English dialect regions, which was preceded by the deletion of the historical /y/ glide after coronal consonants. Labov's analysis on this cause and effect is as follows: Step1: In the middle of the twentieth century, historical /yuw/ sequence was replaced with /iw/ in Northern speech, setting up the /iw/ vs. /uw/ contrast (being observed in minimal pairs such as *dew* and *do*, *lute* and *loot*, etc.). Step2: Because of this glide loss, the next change, the merger of /uw/ to /iu/, has occurred after the coronal[3]. Step 3: After the completion of the merger, those allophones of /uw/ that occur in non-coronal contexts have also moved to the front (2011, p. 107). Labov claimed that phonological structure, once disturbed, may motivate subsequent sound change.

[3] Another analysis is to consider the first two steps to have occurred in a single step of glide deletion (or, "later yod dropping") which had generalized "early yod dropping" before palatals, /r/ and a consonant-/l/cluster (e.g. chew, rude, and flew) before all coronals but retained it before labials and velars (Wells, 1982, as cited in Amos, 2007).

### 2.3.2 Structural Factors in Sound Change

Structures of language can explain certain aspects of sound change. Common chain shifts (e.g. Great Vowel Shifts in Middle English) evidence that phonetic laws in the sense of the Neogrammarians do not operate blindly (Martinet, 1952, p. 5). Kiparsky (1965) argued that the structures of the language can also determine the outcome of innovation. His example was Old High German, which exhibited two distinct outcomes of umlaut: one type (e.g., [u] > [ü] before /i/) produced only new allophones but the other type ([a] > [e] before /i/) was phonemic because /a/ and /e/ contrasted in other environments (Kiparsky, 1965, pp. 4-6). According to Kiparsky, sound change can be also conditioned by the syntactic and phonological structures (Kiparsky, 1965, pp. 27-30). In addition, Kiparsky (1988) argued that whether the change involves a lexical rule (that allows lexical exception) or a postlexical rule (that does not allow exception) determines if the change takes on diffusion or a regular sound change (but see Labov, 2007 for alternative view).

Kiparsky (2003, 2006, 2008) has recently argued that the most efficient explanation for sound change is to have a phonological level of Universal Grammar (UG) to constrain the possible patterns of sound change. Thus, Kiparsky attributed the unattested or hardly attested sound change (such as coda voicing neutralization to voiced obstruents) to a universal constraint (but see Yu, 2004 and Blevins, 2004 for alternative analyses). However, if the language, as an outcome of UG's designing, is in accordance with general constraints in speech, such as perceptibility of certain features in certain positions (see e.g., Hayes & Steriade, 2004), UG and cognitive constraints, between what is learnable as a system and what is doable as basic human performance, seem to fuse into a common human conditions.

## 2.4 A Link between Synchronic Variations to Triggers

While there is a wide range of variation in theoretical viewpoints and the models of sound change, one general agreement is that variation observed as a result of common and recurring sound changes (e.g. assimilation, umlaut, vowel harmony, palatalization, etc.) is also commonly observed in a pool of synchronic pronunciation variation (Kiparsky, 2003; Labov, 1994; Lindblom et al., 1995; Ohala, 1989). This section reviews two mechanisms that have been proposed to bridge synchronic pronunciation variations and phonologically meaningful structural change or socially meaningful phonetic change. These are "phonologization" (Hyman, 1972, 1975, 1976, 2008) and association of social value to phonetic variants (Labov, 2004, P. 462).

One widely shared hypothesis about sound change is that the common sound change occurs via process of phonologization of conditioned variation (e.g., Barnes, 2002; Blevins, 2004; Blevins & Garrett, 2004; Garrett & Johnson, in press; Yu, 2004, in press). Hyman (2008) conceptualized phonologization as one of the two distinct stages of sound change wherein the phonetic variation originally induced by universal phonetic constraint(s) becomes language-specific variation, while not being a part of phonological structure, as summarized in Table 2.1.

Table 2.1 Two stages in language change via phonologization (Hyman, 2008)

| before | | stage 1 | | stage 2 |
|---|---|---|---|---|
| universal phonetics ("automatic") | > | Language-specific phonetics ("speaker-controlled") | > | phonology ("structured") |

In the stage 1, coarticulation is implemented in language-specific way, exhibiting greater degree of effect than expected from mechanical interactions alone. However, the result is still phonetic if it is gradient rather than categorical. Coarticulation enters the domain of phonology when, as in the stage 2, the controlled property becomes structured and categorical (Hyman, 2008, p. 385). One assumption underlying the concepts of phonologization is that context-specific speech variation can become a pronunciation goal in its own right and thus be mentally represented as such.

Sociolinguistic studies have offered yet another hypothesis about the link between the low-level phonetic variation and sound change: It is an association of particular phonetic variants with a particular speech style or social group (Labov, 2001, P. 462), which would be termed as "indexing" in the sense of Eckert (2008). Thus, some of the categorical allophones such as English [p] and [$p^h$], no matter how distinct they are on the phonetic ground, have never developed sound change, as this contrast has never been indexed for gender or social class differences (Labov, 2006, p. 509).

An essential process that is involved in both phonologization and indexing is recognition of variants as variants that deviate from the normative pronunciation. I assume that this recognition is equivalent to the Initial Change. When variants are recognized as variants (even as idiosyncratic variants) these forms enter the cognitive domain, opening the door for social evaluation. Positively indexed variants would have a better chance to be repeatedly used by the speaker and by the rest of the community members than negatively indexed or non-indexed (e.g. recognized as speech error) variants. Therefore, a study on Initial Change will help understanding how trigger is made possible, and a study on triggers will help in understanding what types of variations are more likely to be recognized and assigned with distinct social meanings by human language users. Studies from these two different perspectives will complement each other to solve a grand puzzle of what causes sound change.

The next three chapters report a three-part study on the issue of Initial Change. In this study, I address the question "what causes pronunciation variations and how language users produce, perceive, and learn these speech variations?"

# Chapter 3

# Production Study

## 3.1 Introduction

The study reported in this chapter addresses the question of whether in American English fronted variants of the high back vowel /u/ in alveolar contexts are the result of physical and physiological constraints or a speaker's deliberate control. If fronting of /u/ is a result of purely biomechanical constraints, then the production of the fronted variant does not require any specification in the input to the motor control system. However, if the fronted variant is produced by the speaker's deliberate control over the articulatory sequence of the alveolar consonant and /u/, then such sequential control requires context-specific target specification for the vowel. What the question asks, then, is this: do speakers maintain separate articulation targets for fronted and non-fronted variants of /u/?

A larger question that motivates the present study is the issue of "phonologization," a process whereby a context dependent phonetic feature becomes a distinctive specification of the sound (Hyman, 1972, 1975, 1976, 2008; see §2.4). The above question is re-phrased as this: has contextual /u/-fronting in American English phonologized? However, this question cannot be answered straight-forwardly because exactly what types of coarticulatory variations should be considered as phonologized variations is still an open question at this moment. As phonologization could take place gradually it is difficult to know where phonetic details become phonological pattern, especially if we accept gradient, scalar, or probabilistic phonology (Cohn, 2006; Flemming, 2001; Silverman, 2006).

Instead of attempting to determine whether /u/-fronting should be considered as phonological or not, the present study focuses on obtaining sufficient evidence to determine whether /u/-fronting is purely automatic coarticulation due to production constraints or whether it has a controlled component. This study also documents some acoustic properties of fronted and non-fronted variants of /u/. In the situation where there is no generally accepted set of criteria for determining whether phonologization has or has not occurred, detailed descriptions of coarticulated sounds shed much needed light on the debate on phonologization.

The rest of the chapter is organized as follows. First, the chapter will survey attested cases that the allophonic split of the high back vowel in fronting and non-fronting contexts has become

phonemic (§3.2). Next, the chapter will review previous articulatory and acoustic studies that examined phonetic bases of contextual /u/-fronting (§3.3). Then the chapter will present a research hypothesis (§3.4), and justify the method of testing this hypothesis (§3.5). The chapter will then report the experimental study and its results (§3.6), and discuss the implications of the findings for the theory of control mechanism of coarticulation and theory on the role of coarticulation in phonology/phonologization (§3.7). The chapter ends with a prospectus for future research on the issue of mental representation of coarticulation, and introduces chapters 4 and 5 as a preliminary attempt in this direction.

## 3.2 Attestation

Strong associations between fronting of /u/ (and other back vowels) and adjacent coronal consonants have been observed in studies on historical sound changes as well as synchronic sound patterns of several languages. Two of the relevant diachronic cases have been found through comparison of Written Tibetan (WT) with spoken Lhasa Tibetan (Michailovsky, 1975) and Dzongkha (Mazaudon & Michailovsky, 1988). WT was established in about the eighth century, and it is considered as preserving pronunciation of the language at that time (Michailovsky, 1975). Lhasa Tibetan and Dzongkha (the national language of Bhutan) are two of the modern descendants from the same variety captured in WT (Mazaudon & Michailovsky,

*Table 3.1* Sound changes from WT (8th C.) to Lhasa Tibetan (Michailovsky, 1975) and to Dzongkha (Mazaudon & Michailovsky, 1988). The vowels affected by sound change were bold-faced.

| | WT | Lhasa Tibetan | gloss |
|---|---|---|---|
| (1a) | skad | /qɛ̄ːˋ/ | ‘language’ |
| | bal | /phɛ̲ː/ | ‘wool’ |
| | ston | /tœ̄ː/ | ‘autumn’ |
| | lus | /lʏ̲ːˋ/ | ‘body’ |
| (1b) | goŋ | /ghõ̲ːˋ/ | ‘price’ |
| | nub | /nu̲ːˋ/ | ‘west’ |
| | WT | Dzongkha | gloss |
| (2a) | skad | /$^{1}$ke:/ | ‘language’ |
| | khyod | /$^{1}$ʧoe:/ | ‘you (sg.)’ |
| | lud | /$^{3}$lue:/ | ‘manure’ |
| (2b) | slob | /$^{1}$lo:/ | ‘word, talk’ |
| | lug | /$^{3}$lu:/ | ‘sheep’ |

*Note.* IPA symbols are normalized in accord with the standard convention. Markers of tone and voice quality are same as the original. Dzongkha tone is expressed as a raised number (tone 1-4).

1988; Michailovsky, 1975). As shown in Table 3.1, (1a) and (1b), the modern Lhasa reflexes of WT show vowel fronting when the historical coda consonant was coronal (e.g. /us/ > /y:/), but the same historical vowel has been retained when the historical form had a non-coronal coda (e.g. /ub/ > /u:/). The similar relationship between the historical coda coronal and the preceding vowel is observed in the Dzongkha data (Table 3.1 (2a) and (2b)). In the modern Dzongkha reflexes, the historical vowel preceding a coda coronal is realized as either a front vowel or a diphthong with front off-glide, but the same historical vowel is retained when the historical coda was non-coronal. These data suggest that sometime in the history of the language the original back vowel had split into a fronted allophone before coronal and a back allophone otherwise.

Cognates in Bantu languages of the Ring (Nkom) subgroup spoken in Cameroon suggest an involvement of similar conditioned split in the history of languages as shown in Table 3.2. The cognates in (1) illustrate that Babanki shares the high back vowel /u/ with other Ring languages when a syllable onset is either a labial or velar consonant, but /u/ in the other languages corresponds to front vowel /y/ in Babanki when the syllable onset is coronal. These data imply that the historical vowel of the Ring languages has split into /u/ and /y/ in Babanki due to conditioned fronting by the coronal onset.

Table 3.2 Cognates of the Ring languages of Western Grassfields, Cameroon.

| | Babanki | Aghen | Isu | Kom | gloss |
|---|---|---|---|---|---|
| (1) | ə̀.kʊ́ | ƙɨ̀.kʊ̂ | ƙe.kʊ̄ | ā.kʊ́ | ‘forest’ |
| | mʊ́ʊ́ | mʊ́ʊ́ | mwɨ̄ | ə̄.mʊ́ | ‘water’ |
| | kʊ́ | ɪ́.fʊɔ́ | nɨ̌ | fʊ̄ | ‘to give’ |
| | kʊ̌ | ɪ́.kʊɔ̂ | kwɔ̌ | kʊ̀ | ‘to snore’ |
| (2) | ə̀.lɣ̀ | ʈɨ̀.zʊ́ | ʈə̀.zʊ́ | ə̄.lʊ̂ | ‘honey’ |
| | ʒɣ̌ | ɪ́.zʊ̂ | zʊ́ | ʒvɔ̀ | ‘to skin’ |
| | ʃɣ̌ | ɪ́.sʊ̂ | sʊ̌ʔ | sʊ̀ | ‘to wash’ |

*Note.* Data from Lexical database of Ring subgroup of Grassfields Bantu (1977), collected by L. Hyman, H. Jisa, and J.-M. Hombert.

Synchronically, the co-occurrence restriction of the back vowels and coronal consonants is observed in Cantonese (Cheng, 1991). Cantonese has both front and back non-low rounded vowels (/ü/, /u/, /ö/, /o/), but the high back rounded vowel /u/ does not occur with coronal onset (* /Tu/, where /T/ = coronal) and the mid back rounded vowel /o/ does not occur between coronal onset and coda (* /ToT/, where /T/ = coronal). These vowels can occur with coronal coda as long as the onset is non-coronal, and /o/ can occur with coronal onset if the coda is non-

coronal. In other words, the /T_(T)/ context is restricted to front members of the rounded vowels. Cheng proposed that these patterns have arisen due to assimilatory fronting of /u/ and /o/ in the coronal contexts (p. 121). Her analysis implies that there was an allophonic split for /u/ and /o/ into [u] and [ü], and [o] and [ö], respectively, and these front-back allophones had become contrastive in other environments except for the context of /T_(T)/, before the current sound patterns were established.

Similar affinity between the front vowel and the coronal consonants has been observed in Maltese (Hume, 1996). In Maltese, a vowel for an imperfective prefix /jV +/ is realized as the same vowel as the first stem vowel (e.g. /jV + kɔtor/ → *joktor* 'he/it increases'; /jV + hlɛs/ → *jehles* 'to set free'), except for when the stem initial consonant is a coronal obstruent (e.g. /jV + dalam/ → *jidlam* 'to grow dark'; /jV + skot/ → *jiskot* 'to be silent'). Hume argued that these examples evidence that [i] and coronal consonants are both specified with feature [coronal], because if front vowels are [-back], as in a traditional theory, the affinity between front vowels and coronal consonants becomes an arbitrary one.

The last example of vowel fronting caused by adjacent coronal consonants comes from Morocan Arabic, where a non-word final /o/ is realized as [ö] immediately after a coronal consonant (Hume, 1996). For example, a pair of related forms /ṣəmtɔ/ and /ma ṣəmtöʃ/ ('he killed him' and 'they killed him') shows vowel alternation, as the vowel /o/ occurs after a coronal consonant, non-word finally; but another pair /dərbɔ/ and /ma ḍərbɔʃ/ ('he hit him' and 'he didn't hit him') does not show such alternation, as /o/ is preceded by a non-coronal consonant (p. 175).

These attestations all illustrate, in one way or another, fronting effect of a coronal consonant on a back vowel, which is general enough to manifest itself both in diachronic change and synchronic patterning of sounds. Further, that these attestations come from unrelated languages indicates that the fronting of back vowels in coronal contexts is likely to have its cause in universal phonetic constraints, especially overlapping of articulation (i.e. coarticulation) in natural speech and its acoustic consequences. The next section examines these constraints from previous articulatory and acoustic studies on /u/-fronting.

## 3.3 Observations from Articulatory and Acoustic Studies

### 3.3.1 Articulation of /u/ in Fronting and Non-fronting Contexts

Coarticulatory effects of consonants on the movement of speech organs for vowels have been observed in various articulatory studies in the last 50 years (e.g. Farnetani & Recasens, 1993; Kent & Moll 1972; Kiritani, Itoh, Hirose & Sawashima, 1977; Kiritani 1986; MacNeilage & DeClerk, 1969; Öhman, 1966, 1967; Recasens, 1991; Recasens, Pallarés & Fontdevila, 1997). One characteristic of coarticulation is that the extent of coarticulatory influence a given segment receives from or exerts on an adjacent segment varies depending on the particular consonants, vowels, and even parts of the tongue that are involved in coarticulation (see, Recasens and Espiona (2009) for a review). For example, Kiritani et al.'s s x-ray microbeam study on a

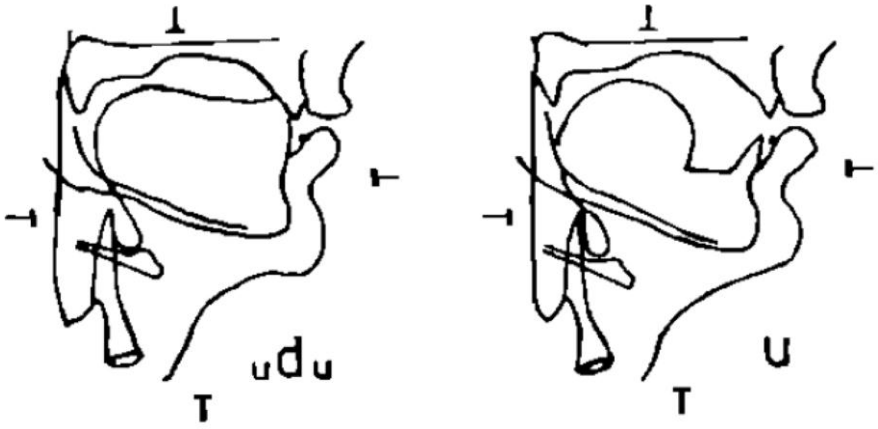



Figure 3.1 Contour tracings from x-ray motion pictures of /udu/ (left) and /u/ (right) uttered by a male Swedish speaker. The edges of the hyoid bone, the mandible, and the epiglottis are shown. Reprinted with permission from "Coarticulation in VCV utterances: Spectrographic measurements," by S. E. G. Öhman, 1966, *Journal of the Acoustical Society of America*, 39, p. 166. Copyright 1966, Acoustical Society of America.

Japanese speaker's articulations of C1VC2 sequences (V = /a, e, i, o, u/; C = /m, t, k, s/) shows that for the front vowels (/i/ and /e/) tongue tip positions are relatively stable across consonantal environments, but tongue tip positions vary considerably for the back vowels (/u/, /o/, /a/).[1] Interestingly, in the environment of /t/, the upper surface of the tongue is stretched out and becomes flat, and because of this "the difference in the tongue shapes for the different vowels tends to decrease" (p. 13).

MacNeilage and DeClerk (1969) reported time varying articulatory data collected from speakers of American English by using cinefluography. Their data of $C_1VC_2$ monosyllables show three main characteristics of coarticulated speech. One is that coarticulatory influence is stronger in $C_1V$ than in $VC_2$, in agreement with the data reported in Kiritani et al.. Another is that there are observable differences between the articulation of a vowel in /b_b/ context, which is a neutral context for tongue articulation for a vowel (p. 1218), and that same vowel in other symmetrical /C_C/ contexts. And finally, in either C1V or VC2, coarticulatory influence from consonant to vowel is for the most part on the front part of the tongue, especially on the tongue tip.

Öhman (1966) took contour tracings from lateral x-ray motion pictures of his own utterances. His tracings show the difference between the vocal tract shapes of the /d/ closure in the /udu/ (left) and the vowels /u/ (right) (Fig. 3.1). These data illustrate that during the alveolar closure in /udu/ the back of the tongue is slightly lowered and fronted than in plain /u/ and, more

[1] There is individual variation in the articulation so that for some speakers tongue tip positions are stable for back vowels, too, except for the environment of /ɾ/ (Y. Hasegawa, personal communication).

importantly, the tongue tip is markedly higher in /udu/, making its constriction at the alveolar ridge. This suggests that, in natural speech, the tongue tip tends to be in a relatively higher position for at least some part of the vowel in /ud/, /ut/, /du/, and /tu/ sequences, because of anticipatory/perseveratory influence from the tongue configuration for /d/ (or /t/).

Collectively, these findings suggest the following characteristics in the spatio-temporal interactions between /u/ and alveolar consonants:

1) Alveolar consonants exert greater constraints on vowel articulation than other consonants; and
2) Coarticulatory influence from consonant to vowel is mainly on the tongue tip.

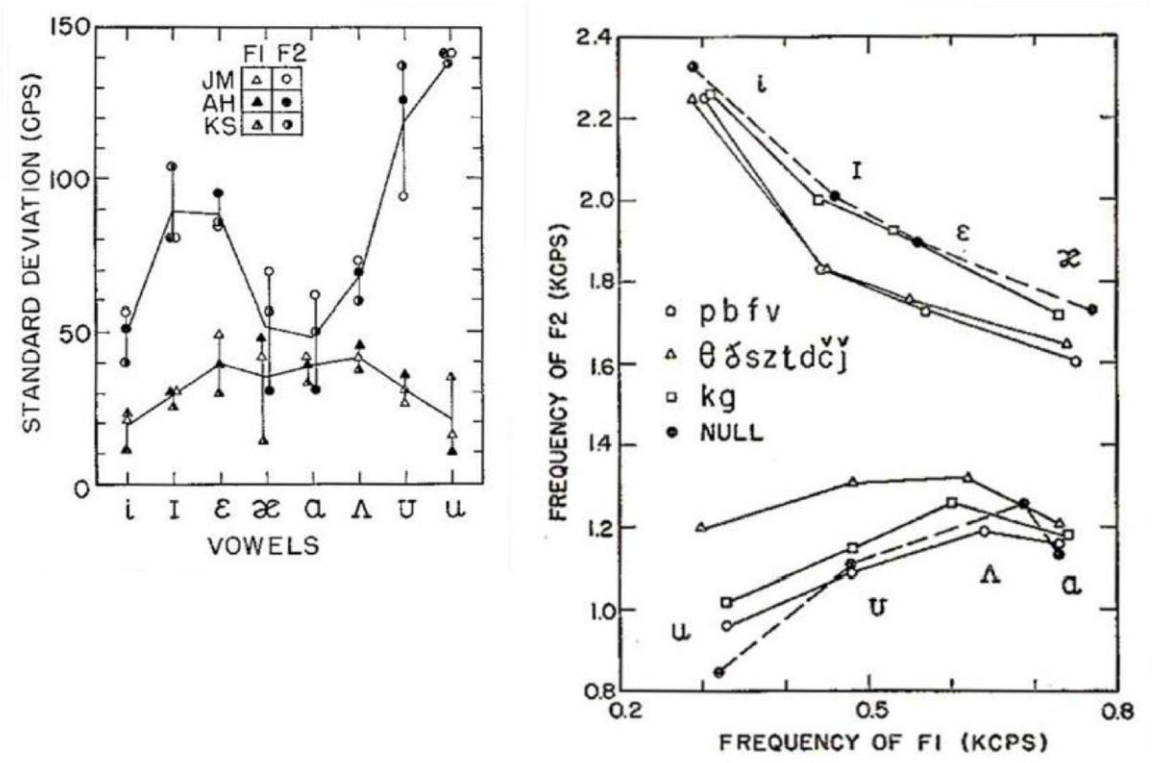



Figure 3.2 Stevens and House (1963) data showing the extent of variability of F1 and F2 across the 14 consonantal contexts (left) and the effect of the place of articulation of the consonants on vowels' F2 (right). Reprinted with permission from "Perturbation of vowel articulations by consonantal context: An acoustical study," by K. N. Stevens & A. S. House, 1963, *Journal of Speech and Hearing Research*, 6, p. 119. Copyright 1963, American Speech and Hearing Association.

### 3.3.2 Acoustic Properties of /u/ in Fronting and Non-fronting Contexts

Adjacent consonants exert coarticulatory influence on the vocal tract shape of a vowel, and alter acoustic properties of the vowel both in high-low and front-back dimensions, as often revealed by F1 and F2 measurements (e.g., Farnetani & Recasens, 1993; Lindblom, 1963; Öhman, 1966, 1967; Recasens, 1991; Stevens & House, 1963). On the effect of alveolar consonants on the back vowel /u/, previous studies unanimously reported a raising effect on F2 of the vowel. For example, Öhman (1966) reported that in /udu/ utterances, the vowel's F2 increased by 490 Hz at the VC juncture and by 690 Hz at the CV juncture compared with the point where F2 was steady. Dynamic changes in F2 are expected, as the tongue gradually moves from a vowel configuration to a consonant configuration (or from a consonant to a vowel), as observed by MacNeilage and DeClerk (1969). Stevens and House measured formant values at the middle of the English vowels (/i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/) produced by three male talkers (JM, AH, KS) in null environments (i.e. in isolation or in /hVd/ syllables) and in consonantal contexts (i.e. in symmetrical /C_C/ syllables, where C = /p,b,f,v,θ,ð,s,z,t,d,ʧ,ʤ,k,g/). Their results (Fig. 3.2) show that consonantal effects on F2 are much greater for the rounded vowels /u/ and /ʊ/ than for other vowels (left panel), and that it is the postdental (=alveolar) consonants that cause the greatest shift—as much as 350 Hz for /u/—in F2 (right panel). Taken together, these observations point to a particular vulnerability of F2 in the high back rounded vowels in the context of alveolar consonants. In the auditory vowel space, an upward shift of F2 corresponds to fronting of the vowel quality, thus the resulting vowel may sound like [ɯ], [ɨ], or [y] depending on the degree of the consonantal constriction made simultaneously with the [u] configuration (Ohala, 1981, p. 180).

## 3.4 Hypothesis

It is clear that when /u/ is produced before or after alveolar consonants the front part of the tongue is inevitably influenced by an apical configuration for the consonants, and as a consequence F2 of /u/ becomes higher than /u/ produced in the null environment. In auditory vowel space, higher F2 translates to a fronted vowel quality. The combination of articulatory, acoustic, and auditory factors is the phonetic basis of fronted variants of /u/ in alveolar and other coronal contexts. But this is not the end of the story. The question is: are such phonetically motivated allophonic variants mentally represented? In other words: do the speakers have a distinct articulatory goal for a fronted /u/ apart from the goal for a canonical /u/?

There are both rational and empirical reasons to hypothesize that this is the case. Rational support comes from an analysis of control mechanisms of coarticulation. On the issue of articulation of stop consonants in the context of vowels, Öhman remarks as follows:

> [F]or the purpose of speech description, the tongue may be regarded as three independently controllable mechanical systems,… . These systems may be called the apical articulator, the dorsal articulator, and the tongue body articulator. … We also observed that the production of vowel+stop consonant+vowel utterances of certain languages seemed to involve two simultaneous gestures, viz.,

> diphthongal gesture of the tongue-body articulator and a superimposed constrictor gesture of the apical or the dorsal articulators. Since motions of the three articulators individually have an effect on the whole vocal-tract (VT) shape, and since the effect of an individual articulator is different for different simultaneous motions of the other articulators, it is not possible to associate invariant-target VT shapes with the intervocalic stop consonants.
>
> (Öhman, 1967, p. 310)

By extending Öhman's account, one might expect that it is not possible to associate invariant-target VT shapes with the back vowel /u/. Even if we assume an invariant gesture of the tongue-body articulator for /u/, a constrictor gesture of the apical or the dorsal articulator is superimposed, in accord with the place of articulation of the adjacent consonant. In the case of alveolar consonant-vowel sequence, the dorsal articulator is not actively controlled during the closure, and therefore would freely coarticulate with the constriction gesture of the apical articulator. After the consonantal release, both apical and dorsum articulators, which are not controlled by the vowel, may remain in a relatively higher position due to inertia. As a result the whole VT shape during production of /u/ will be quite different adjacent to an alveolar consonant as opposed to null environments.

Empirical support comes from numerous cross-linguistic studies on coarticulation that have shown that phonetic implementations of speech signals consist of both mechanical components and controlled components, and that these controlled components shape phonetic output in language-specific ways. In his pioneering study, Öhman (1966) found greater degree of vowel-to-consonant coarticulation in Swedish and English than in Russian. The author hypothesized that in Swedish and English the precise shape of the vocal tract during the stop closure is phonemically irrelevant, leaving subsets of the tongue muscle to freely respond to the articulation for vowel, but this is not the case in Russian, where the stop series has distinctive palatalization/velarization in addition to place features. Similar types of language-specificity in the temporal extent and/or degree of coarticulation has been observed in cross-language comparison of, for example, vowel-to-vowel coarticulation between American English and Shona (Beddor, Harnsberger & Lindemann, 2002) and vowel nasalization between American English and French (Cohn 1993). These observations suggest that some portion of coarticulation can result from speakers' fine-tuned control over different speech organs in a context-specific manner rather than the result of interconnected articulatory movements of different musculature.

Nonetheless, one should not assume that every type of coarticulation is under speaker control. Solé (1992) presented evidence that there are both *automatic* types and *controlled* types of coarticulation (see §3.5 below). Thus the question of whether a particular coarticulation is an automatic type or a controlled type must be tested case by case. Now, the next question is how?

## 3.5 Methodology

Lindblom (1963) employed a vowel manipulation method to test whether vowel reduction involves centralization or coarticulatory assimilation. His data showed that the extent of coarticulatory influences of the flanking consonants (/b_b/, /d_d/, or /g_g/) on the eight Swedish

lax vowels (/ɪ, ɛ, ʏ, æ, a, ɵ, ɔ, ʊ/) reduced as the vowels' duration increased: the formant frequencies of each of the vowels approached asymptotic values as the duration increased. Further, the three regression models that were derived from the production data to predict each vowel's formant values in the /b_b/, /d_d/, and /g_g/ context were generally successful without including centralization as a predictive factor in the model. From these results Lindblom made the following claims: (1) vowel reductions are due to assimilation,[2] not centralization (pp. 1780-81); (2) vowel duration is the main determinant of the extent of vowel reduction (p. 1780); and (3) each vowel has a single articulatory target regardless of the consonantal context, and the articulator hits this target if there is sufficient time to do so (pp. 1778-9). This study illustrates how the dependency of the extent of contextual perturbations to the vowel's duration can be interpreted as evidence that the contextual variations arise from biomechanically-based articulatory constraints. The same method was used in more recent studies that investigated phonetic vowel reduction (Nowak, 2006) and phonological vowel reduction (Barnes, 2006).

Solé (1992) used the same method to test cross-linguistic variation in temporal extent of vowel nasalization in vowel-coda nasal sequences. Her results showed that in American English the duration of nasalization during the vowel was proportional to the overall vowel duration (thus the duration of the nasalized part of the vowel increased as the vowel duration increased), but in Continental Spanish the duration of nasalization remained constant regardless of overall vowel duration. With these data, Solé claimed that coarticulation may arise from purely phonetic constraints (as in Continental Spanish) or with additional control over its temporal degree (as in American English). This study illustrates how constant proportionality between the duration of coarticulated part of the segment and the entire segment duration serve as evidence that the observed degree of coarticulation results from speaker's deliberate control toward context-specific articulatory goals. The same method was used in a cross-language comparison in vowel duration variations that co-occur as a secondary feature with phonemic vowel height differences (Solé & Ohala, 2010).

The studies discussed in this section have demonstrated the usefulness of duration manipulation in determining the articulatory instructions executed by the speakers. Following these studies, the present study employs the duration manipulation method to investigate speakers' production goals for /u/ in alveolar contexts. The hypothesis to be tested is: in American English, the articulatory goal of /u/ in alveolar context is not the same as the goal for /u/ in nonalveolar context. If the degree of fronting of /u/ persists regardless of vowel duration, then this would be taken as evidence that a speaker has multiple production targets, one for plain /u/ and the other for a fronted /u/ in alveolar contexts.

[2] A basis of this claim becomes extremely clear when one consults Ohala's (1991) presentation of the model's predictions as F1-F2 plots of vowels as a function of consonantal contexts and vowel duration. These plots are clear visual summaries of the Lindblom's regression models.

# 3.6 Experiment

## 3.6.1 Participants

Thirty-two native speakers of American English (18 females and 14 males) between the ages of 19 and 45 participated in the experiment. The participants were all undergraduate students attending UC Berkeley at the time of experiment, and all of them reported that they had normal hearing and speaking. They received $10 for participation.

## 3.6.2 Materials

Table 3.3 Words elicited in the production experiment

| Test words (context = /D_D/) | | Control word (context = /b_d/) | | Reference words (context = /h_d/ or /h_t/) | |
|---|---|---|---|---|---|
| *dude* | [dud] | *booed* | [bud] | *heed* | [hid] |
| *toot* | [tut] | | | *hid* | [hɪd] |
| *zoos* | [zuz] | | | *head* | [hɛd] |
| *Seuss* | [sus] | | | *had* | [hæd] |
| *noon* | [nun] | | | *hot* | [hɑt] |
| *dune* | [dun] | | | *HUD* | [hʌd] |
| *tune* | [tun] | | | *hood* | [hʊd] |
| | | | | *who'd* | [hud] |

A list of English test words, a control word, and reference words was created (Table 3.3). The test words had the high back vowel /u/ in a symmetrical /C_C/ context, where the Cs were one of the alveolar consonants (/d, t, z, s, n/). These contexts were expected to elicit fronted variants of /u/. The control word had the same vowel phoneme but in the context /b_d/. This context was expected not to induce a significant quality difference on the /u/ because the place of articulation of the onset consonant and the place of greatest constriction of the vowel are the same. The purpose of eliciting the test vowels and the control vowel was to compare acoustic properties of /u/ in fronting and non-fronting contexts. Reference words had one of the eight English monophthongs (/i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,ʊ,u/) in the context /hVd/ or /hVt/, following the numerous previous studies that examined acoustic properties of English vowels (e.g., Hillenbrand, Getty, Clark & Wheeler, 1995; Peterson & Barney, 1952). The vowels in the /h_d/ context are

expected to be produced with a comparable articulatory configuration as the vowel in isolation, and therefore the context of /h_d/ has been often regarded as the null context (Stevens & House, 1963, p. 116).[3] The purpose of eliciting reference vowels was to construct a speaker-specific vowel space in which to calculate the degree of /u/-fronting for each speaker.

## 3.6.3 Procedure

Speakers were recorded individually in a sound attenuated room in the University of California, Berkeley, Phonology Lab. The microphone (Shure 10A) was connected to a preamp (M-Audio Audio Buddy) then to a computer. The microphone was positioned about three centimeters away from the speaker's lips, and the gain was adjusted for each speaker during a short test recording session prior to the data collection session.

The speakers were instructed to first repeat each of the test words in a carrier sentence ("*That's a ____ again."*) six times with a medium speech rate. They next repeated the same task with a fast rate, and then with a slow rate. They performed an identical set of repetitions at each rate for the control word ***booed*** and one of the reference words *who'd*. Finally, they were asked to perform the same set of repetitions with the rest of the reference words but only with the medium rate. The term ***medium rate*** was explained to the speakers as "the speech rate that you would use for most normal conversational situations." The term ***fast/slow rate*** was explained as "a faster/slower rate than what you used in the medium rate' tasks," and exactly how fast or slow was the speaker's own choice. The summary of elicitation conditions was as follows:

- Test words, ***booed***, and *who'd*: 9 words, 4-6 repetitions, 3 different speech rates
- Reference words: 7 words, 4-6 repetitions, 1 speech rate (medium)

## 3.6.4 Acoustic Measurements

The speakers' utterances were digitally recorded to the computer's hard drive at the sampling rate of 22050 Hz and quantized at 16 bits/sample. Two speakers' (subjects #18 and #30) data were removed from the analysis because of substantial clipping in the audio signals. For the reference vowels /i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,ʊ,u/ and the control vowel /u/ in booed, in which F1 and F2 generally exhibited steady-state formant contours except for the later part of the vowels, F1 and F2 values were measured at the temporal midpoint of a vowel (Fig. 3.3, upper panel). For the test vowel /u/, F1 and F2 values were measured from the temporal point where F2 reaches its minimum (Fig. 3.3, lower panel). This point was interpreted as the point where the adjacent consonants' coarticulatory influence on the vowel was smallest, or equivalently, the point where the articulator best approximates the target configuration for a given vowel (cf. Lindblom, 1963).

[3] As discussed in Section 3.3, previous study showed that in a C1VC2 syllable C2-to-V coarticulation is much smaller than C1-to-V coarticulation (Kiritani et al., 1977; MacNeilage & DeClerk, 1969). Also vowel formants in hVd syllable are not much different than in isolated vowels in English (Stevens & House, 1963).

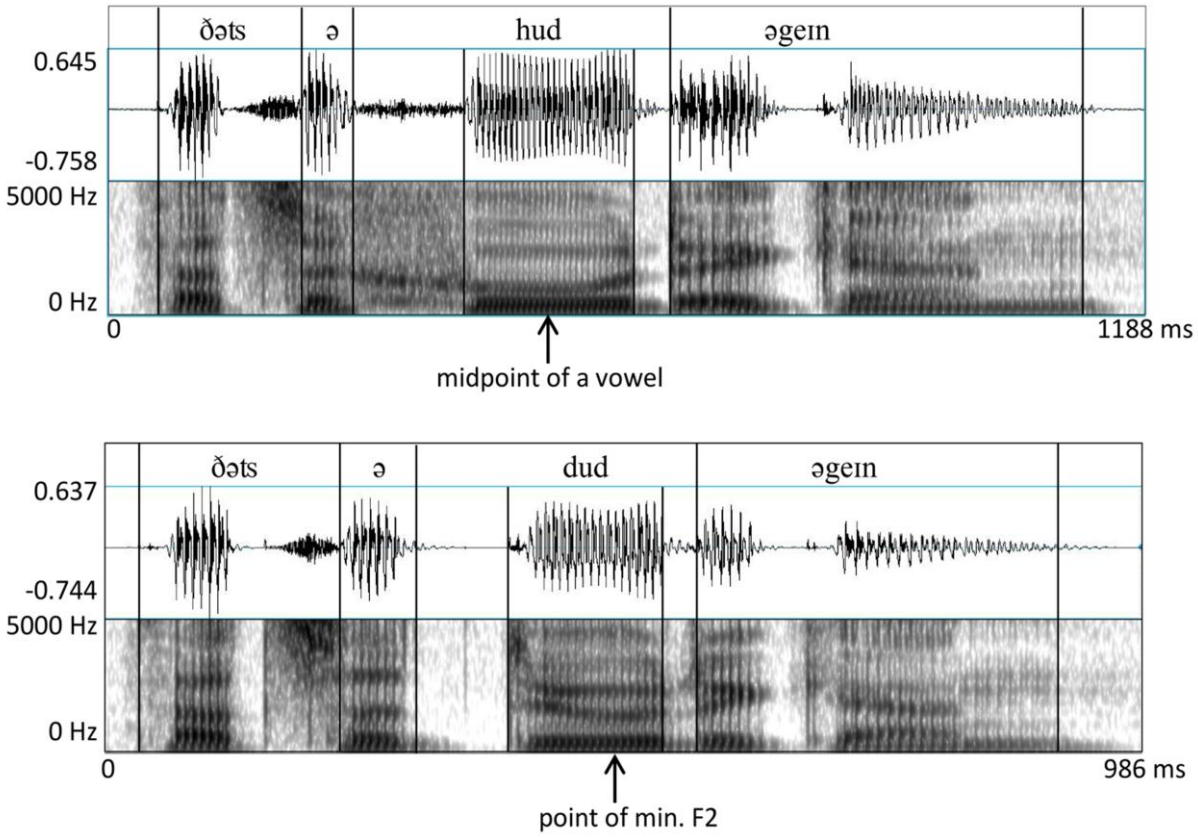



Fig. 3.3 Examples of a waveform and a spectrogram of a reference word (upper) and a test word (lower) embedded in a carrier sentence. Each example shows demarcation for formant measurements: If the onset was a fricative, the beginning of a vowel segment was set at the beginning of the vowel (upper); if the onset was a stop, the beginning was set to the onset release (lower). Arrows indicate the points from where F1 and F2 were measured.

The relative location of F2 minima varied across words and speakers, but the general tendency was that minimum F2 occurred during the last half of the vowel, often near the very end.

Formant measurement was done with Praat (Boersma & Weenink, 2007) by using a script that measures and records F1 and F2 values at the specified time point(s) from pre-specified segment intervals. The script was a modified version of the original script obtained from the following site:
http://www.helsinki.fi/~lennes/praat-scripts/public/collect_formant_data_from_files.praat.
The modification was minor, in that the measurement was taken at intervals for every 10% of the overall vowel duration (5% from edges excluded), rather than taking measurements only at the midpoint, as the original script does.

For the reference vowels and the control vowel, F1 and F2 were measured only at the 50% point of the vowel. From the measurements taken from the four-to-six repeated tokens, median F1 and median F2 were calculated for each speech rate for each speaker. The reason for using median values over mean values (or all measurements) was to reduce the influence of spurious measurements that arise occasionally from autocorrelation.

For the test vowels, F1 and F2 were measured at all ten (5%, 15%, …, 95%) points. From the measurements taken from the repeated tokens, the median F1 and F2 were calculated for each time point. These values yield time-normalized and stylized formant trajectories for the middle 90% of a given vowel. Then the lowest F2 value was found for each vowel and F1 from the same time point was also found.

## 3.6.5 Speaker Normalization

F1 and F2 values were transformed to talker normalized values so that data obtained from different speakers and from both sexes could be pooled in the analysis. For this purpose Nearey's (1978) individual log-mean method was employed. This method is based on the assumption that it is the relationships among the formants that enable listeners to overcome across-talker variations of the vowels (Joos, 1948, as cited in Nearey, 1978, p. 86) as well as the empirical data supporting this relational normalization with an additional point in a vowel space as a correction factor for individual variation. In this method, each speaker's vowels are located in a logarithmic F1-F2 space in relation to a single reference point, the mean of log-transformed formant values (pp. 90-95). The choice of this normalization method was motivated by Adank (2003) and Adank, Smits, and van Hout (2004), which showed that Nearey's method effectively reduces the effect of anatomical/physiological differences while preserving phonemic and sociolinguistic variation. One assumption made in this study was that if there is any sub-phonemic but deliberately controlled variation, then such variation should also be maintained after normalization.

Figure 3.4 illustrates the normalization process. First, F1 and F2 were transformed into their natural logarithms (LF1 and LF2). Then the mean of LF1 (MLF1) and the mean of LF2 (MLF2) were calculated for each speaker. These two log means define, in F1-F2 coordinates, an operationalized center of each talker's vowel space. Each vowel's normalized F1 (NF1) and F2 (NF2) were obtained as LF1 minus MLF1 and LF2 minus MLF2, respectively. The sign of NF1 indicates that the vowel is lower (+) or higher (-) than the center. Thus, in the vowel space in the figure, for example, the vowels /i/ and /u/ have negative NF1 and the vowel /æ/ has positive NF1. For NF2, positive/negative sign indicates that the vowel is more frontward/backward than the center. Again, in the same figure, the vowels /i/ and /æ/ have positive NF2 and the vowel /u/ has negative NF2.

The effect of speaker normalization can be appreciated by comparing the vowel plots based on un-normalized and normalized data. Figure 3.5 shows the distributions of the un-normalized formant frequencies of the reference vowels on the F1-F2 plane. Each data point represents median F1 and F2 calculated for each vowel for each speaker ($N$ = 240: 30 speakers x 8 vowels). The black colored symbols present female speakers' data (18 speakers) and the gray colored symbols represent male speakers' data (12 speakers). The boundary for each vowel category for

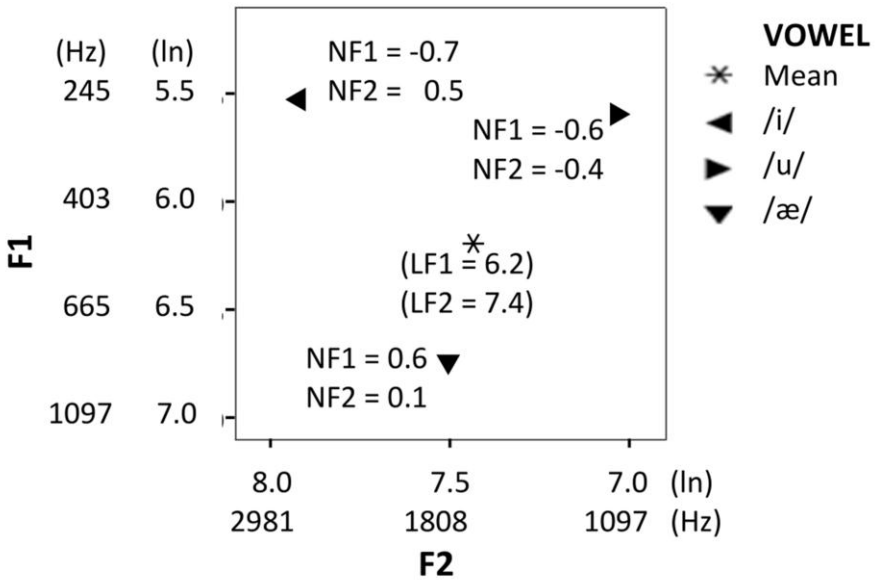



Fig. 3.4 A sample of vowel normalization for speaker #7 (female). Log mean was 6.2 for F1 and 7.4 for F2. Normalized F1 and F2 for /i/, /u/, and /æ/ are provided in the F1-F2 plots.

each sex group was defined as the 95% prediction confidence ellipse for F1 and F2. As expected from the un-normalized data, males and females had systematically different distributions in the F1-F2 plane, with males occupying generally lower frequency regions than females within each vowel category. The plots also reveal individual variations in F1 and F2, resulting in multiple overlaps of the vowel boundaries within the male and female data. This within-group variation is not surprising given that even within sex groups speakers vary considerably in physical size, and presumably also in vocal tract size. As a result of sex differences and individual differences in the formant values the plots exhibited considerable overlaps of the vowel categories. Figure 3.6 shows the distribution of the normalized formant frequencies (NF1 and NF2) of the same vowels. As the data were normalized for each speaker, sex differences of the formant values were reduced and more distinctive, tighter vowel categories have emerged.

Other than the effect of the speaker normalization, the plots also revealed an interesting pattern. There is a near-perfect convergence of the category centers of the female and male data for the lax vowels /ɪ/ and /ʊ/, but for the other vowels sex differences still remain. The trend is for male vowels to be more centralized on the F1 dimension than female vowels, and this trend was particularly noticeable for the high back vowel /u/.

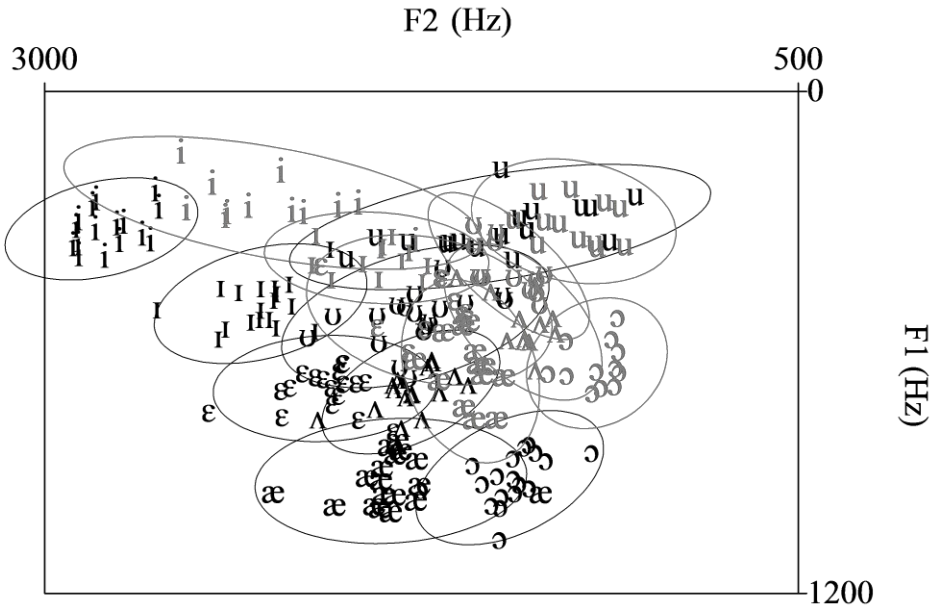



Figure 3.5 Mean F1 and F2 values (Hz) of each reference vowel for 18 females (black) and 12 males (gray) with 95 % confidence ellipses ($N$ = 240: 8 vowels x 30 speakers).

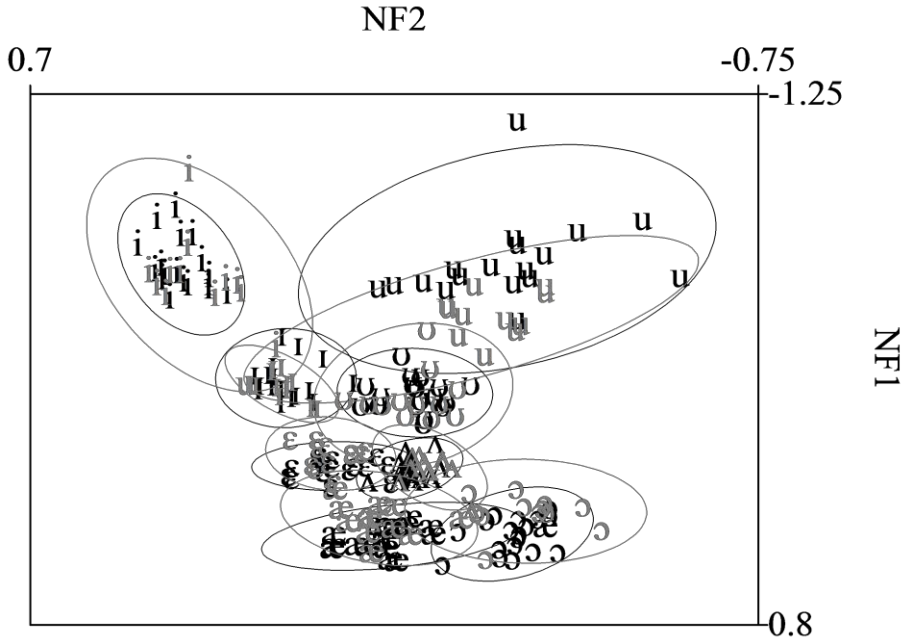



Figure 3.6 Mean NF1 and NF2 values of each reference vowel for 18 females (black) and 12 males (gray) with 95 % confidence ellipses ($N$ = 240: 8 vowels x 30 speakers).

## 3.6.6 Analyses and Results

### 3.6.6.1 Vowel Duration

Figure 3.7 shows mean vowel duration for the three speech rates for the thirty subjects. All the speakers except for #21 showed monotonically increasing vowel duration as they varied the speech rates from fast through medium to slow. Speaker #21 had slightly longer vowel duration for the medium rate than for the slow rate, but this reversal of vowel duration does not concern us here, because the difference was slight and both durations were longer than for the fast rate. A numerical summary of vowel duration is given in Table 3.4.

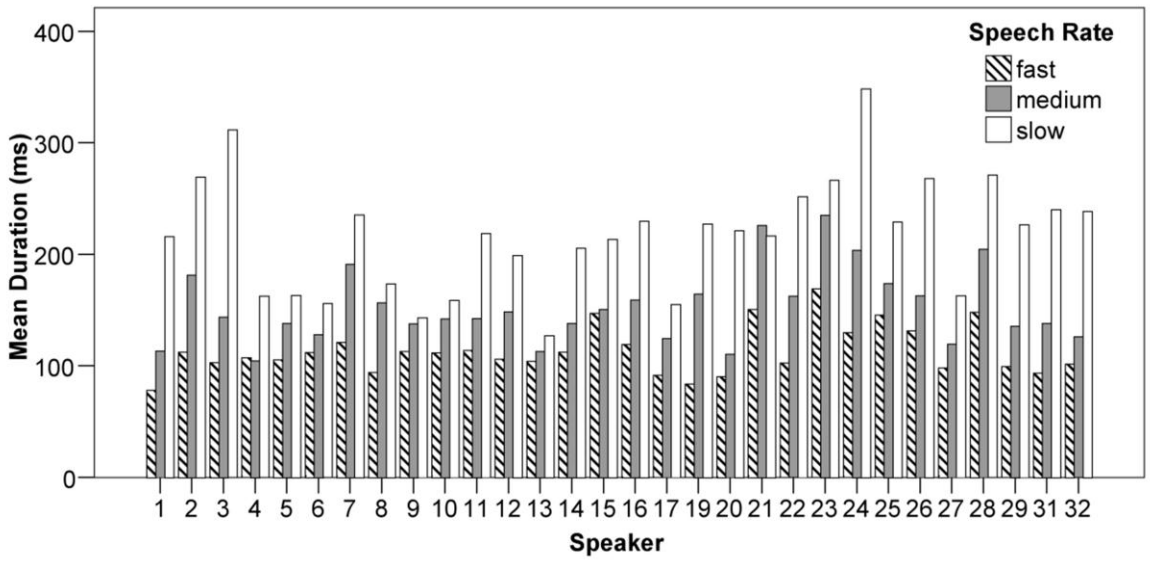



Figure 3.7 Mean vowel duration by speech rate for each of the 30 subjects (subjects #18 and #30 excluded).

Table 3.4 Summary for vowel duration (ms) by speech rate ($N = 270$).

| Rate | N | Mean | | SD | Min. | Max. |
|---|---|---|---|---|---|---|
| Fast | 90 | 113.18 | } Δ = 39.42 | 25.59 | 64.46 | 190.29 |
| Medium | 90 | 152.60 | } Δ = 64.26 | 37.41 | 77.74 | 245.40 |
| Slow | 90 | 216.86 | | 55.43 | 117.51 | 383.10 |
| Total | 270 | 160.88 | | 59.40 | 64.46 | 383.10 |

### 3.6.6.2 Variation of /u/

Before examining the effect of vowel duration on the degree of fronting, some preliminary observations on the acoustic properties of /u/ in fronting and non-fronting contexts were made. Figure 3.8 shows stylized NF2 trajectories based on the measurements taken from the ten equally-distanced points of the mid 90% of the vowel segment in each test word (*dude*, *dune*, and etc.), control word (*booed*) and reference word (*who'd*). For *dude*, *dune*, and *booed*, only the measurements from the second time point (15% point) and later were used. This is because the vowel segment for the stop-vowel-stop words was defined as an interval between the stop release and vowel offset. With the stop release included in the interval, formant measurements were inconsistent in the periods that contained aspiration noise. The formant measurements were reliable only from the second point. For *toot* and *tune*, only the measurements taken from the last four points were used because the vowel segment contained long aspiration noise, in which there were no well-defined formants and/or formant measurements were inconsistent.

Several patterns emerged from the plots. First, the vowel /u/ exhibited distinct NF2

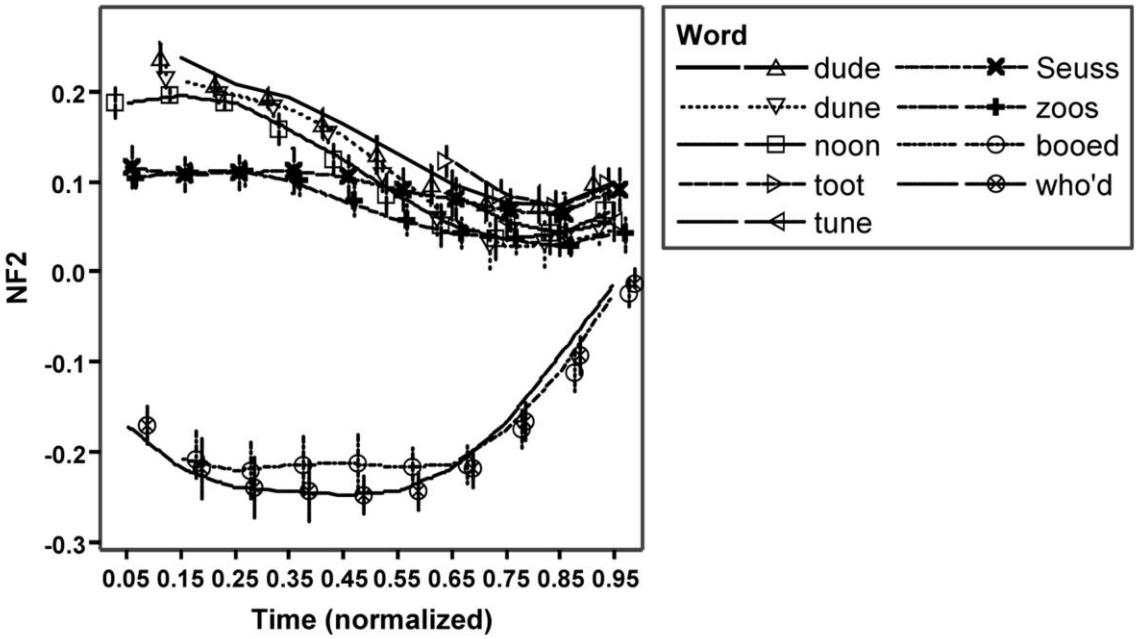



Figure 3.8 Averaged and time-normalized NF2 trajectories with 95% confidence intervals based on the measurements taken from the ten equally-distanced points of the mid 90% of /u/ in each test word (*dude*, *dune*, etc.), control word (*booed*) and reference word (*who'd*). The trajectories are aligned at onset stop release for the *dude*, *dune*, *noon*, *toot*, and *tune* and vowel onset for *Seuss*, *zoos*, and *who'd*. The trajectories begin at the nearest time point where vowels were visible (i.e., excluding aspiration noise).

trajectories in the test words than in *booed* and *who'd*. The difference persisted even at the point where NF2 of test words reached its minimum. Second, all test words had very similar NF2 trajectories in the later portion of the segment, and NF2 seemed to converge to a common value at the vowel offset. This is not surprising given that all words shared the same vowel-coda sequence. Third, at vowel onset, *Seuss* and *zoos* had lower NF2 than other test words. One

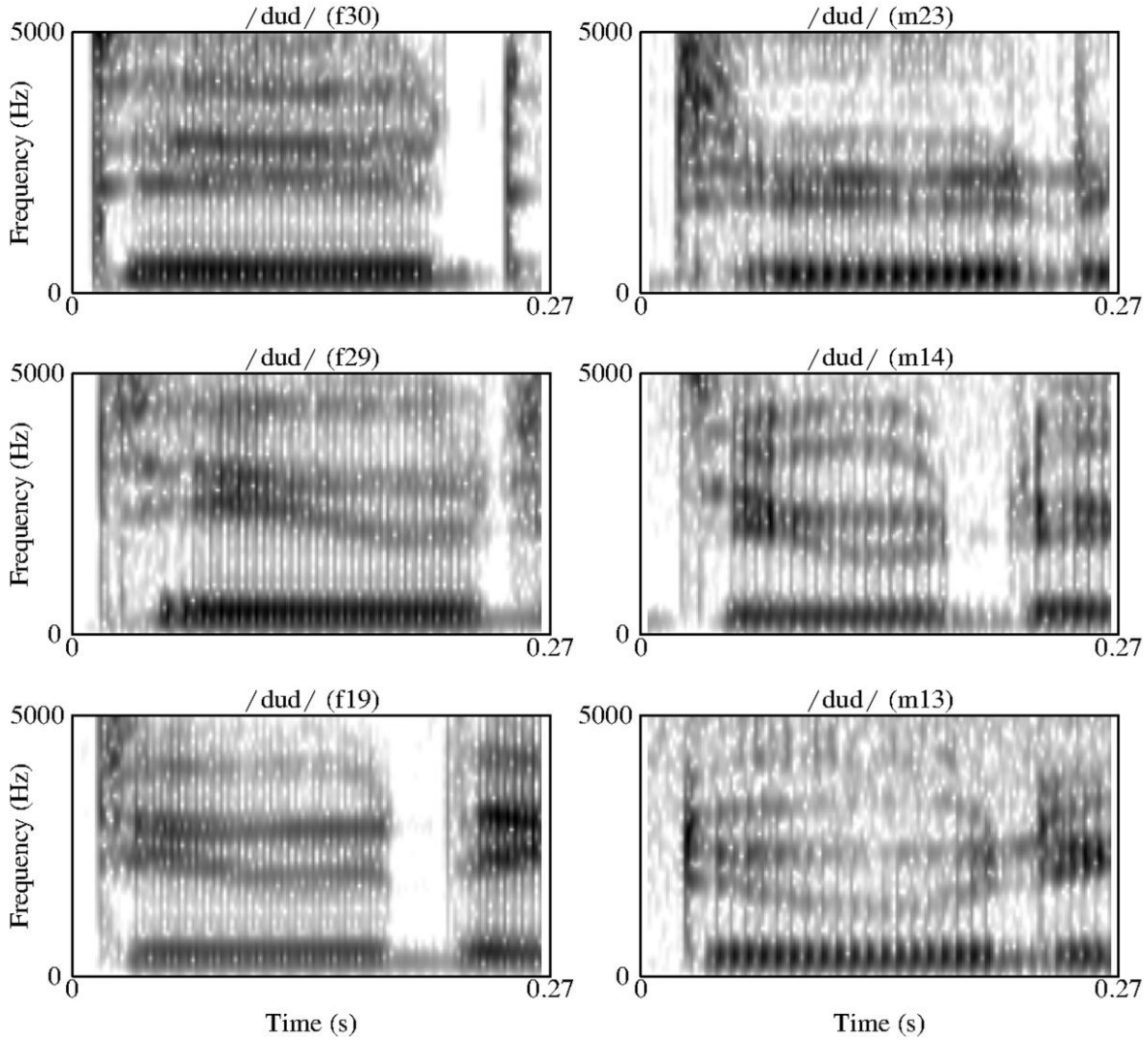



Figure 3.9 Spectrograms of the American English speakers' production of *dude* (/dud/) and a part of following *a* (/ə/) in a carrier sentence "That's a ___ again."
The spectrograms in the left and the right column represent utterances of female (#30, #29, #19) and male (#23, #14, #13) speakers, respectively, and the spectrograms in the top, the middle, and the bottom row represent *flat*, *up-and-down*, and *U-shape* trajectories, respectively.

possible interpretation of this pattern is that this was an artifact of the segmentation: for these words the vowel segments started at vowel onset, excluding release noise. This means that for any given time point, articulatory events during vowel segments were probably not identical between stop-vowel-stop words and fricative-vowel-stop words. Another interpretation is that this pattern reflects genuine difference between onset alveolar stops and onset alveolar fricatives in their F2 raising behavior. Since English has another set of fricatives in post-alveolar, speakers might try not to produce extremely high F2 in /su/ and /zu/ to avoid these sequences to sound like /ʃu/ and /ʒu/, respectively. Finally, despite the antagonistic relationship between alveolar onset and the vowel /u/, the vowel's NF2 did not fall immediately after vowel onset. Rather, F2 remained in its initial level or even rose for a short period of time after vowel onset. Indeed, many of the test word tokens exhibited rising-falling F2 contour, similar to the F2 contour in a sequence of palatal glide and a vowel (/ju/) as in words *beauty*, *youth*, and etc.

F2 trajectories for stop-vowel-stop words varied across speakers, but generally fell into one of the three types. For one type, though there were not many tokens that fell in this type, F2 started high at CV juncture, immediately started falling to reach its minimum toward the end of the vowel, and rose again toward the release of the coda stops. This would be called as a *U-shape* trajectory. For another type, F2 made a noticeable rise before it started falling, as described in the previous paragraph. This type would be called as an *up-and-down* trajectory. For the last type, F2 made relatively flat trajectory. This would be called as a *flat* trajectory. A sample spectrogram illustrating each type of F2 trajectory from female and male utterances of *dude* is presented in Figure 3.9. All three types of F2 trajectories were observed from both male and female speakers' production; however, there was a sex difference in that majority of the male speakers' vowels had *flat* trajectories, sounding monophthongal (e.g. [dyd]), while majority of the female speakers' vowels had *up-and-down* trajectories, sounding diphthongal (e.g. [djud]).[4]

### 3.6.6.3 Distribution of /u/ in NF1-NF2 Space

Figure 3.10 shows NF1-NF2[5] plots of test vowels (with a symbol "d") and the control vowel (with a symbol "b"), as spoken in the fast, medium, and slow rate conditions (*N* = 180: 2 vowels x 3 rates x 30 speakers), overlaid on the background of the 95% confidence ellipses for the reference vowels /i/ and /u/ (*/i/-landmark* and */u/-landmark*).[6] For the test vowels each data point represents the mean NF1 and the mean NF2 of all of the seven words (*dude*, *toot*, *zoos*, *Suess*, *noon*, *dune*, and *tune*). The plots illustrate the effect of fronting vs. non-fronting contexts on the phonetic realization of /u/. The acoustic distribution of /u/ in the non-fronting context was similar to the region of /u/-landmark. The vowel /u/ in the fronting context, on the other hand,

[4] These variants were similar to the variation observed in /u/-fronting in Houston (Koops, 2010), though the associated social dimension was different. Koops reported a generational difference, in which younger speaker's /u/ was monophthongal, exhibiting somewhat more flat F2 trajectory, and older speaker's /u/ was diphthongal, exhibiting more dynamic F2 trajectory.

[5] See Appendices A-D for F1 and F2 values for each speaker.

[6] Note that the ellipses of the reference vowels only serve as landmarks on the NF1-NF2 plane. These plots were not meant to approximate the full range of variability of the reference vowels due to speech rate variation.

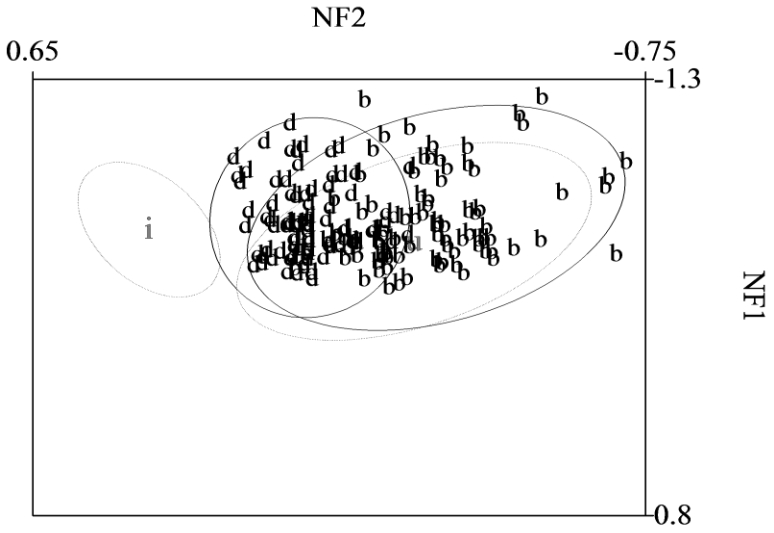



Figure 3.10 NF1-NF2 plots and 95% confident ellipses (black symbols with solid lines) of the vowels in the test vowels ("d") and in the control word ("b") spoken in three speech rates (*N* = 180: 2 word types x 3 rates x 30 speakers), overlaid on 95% confidence ellipses of reference vowels /i/ and /u/ (gray symbols with dotted lines, serving as the /i/-landmark and the /u/-landmark, respectively). For test words, each data point represents the mean of seven test words (*dude*, *toot*, *zoos*, *Suess*, *noon*, *dune*, and *tune*).

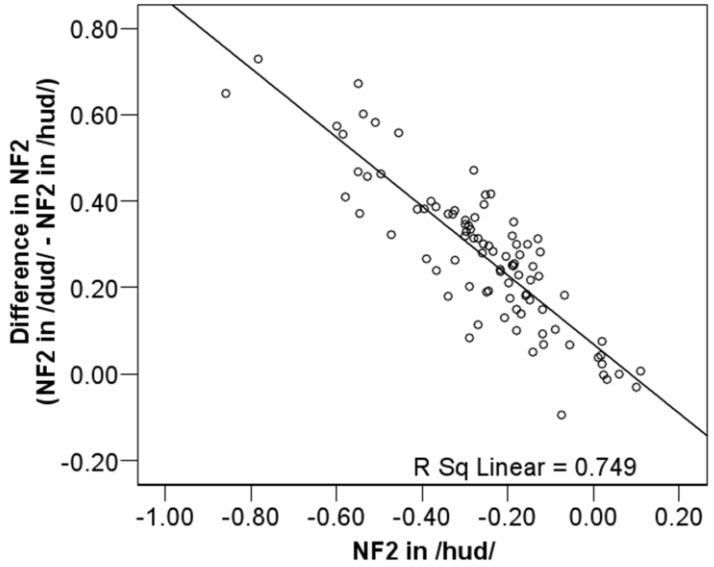



Figure 3.11 Scatter plots of the difference in NF2 of /u/ in *who'd* (/hud/) and *dude* (/dud/) spoken in three speech rates (a measure of degree of fronting) as a function of NF2 of /u/ in who'd (/hud/), together with a linear regression line (*N* = 90: 3 rates x 30 speakers).

occupied an entirely different space, right next to the /i/-landmark. It is clear that /u/ in alveolar contexts has different acoustic qualities compared with /u/ in non-fronting contexts.
In addition, the plots reveal that NF2 values of /u/ have a rather compact distribution in the context of /D_D/ compared with the other two contexts. That is, speakers who produce their canonical /u/ with relatively low NF2 made a greater shift in NF2 in the fronted /u/ than speakers who produce their canonical /u/ with relatively higher NF2. This trend was so robust that there was a strong correlation between NF2 values of /u/ in the null context (/h_d/) and the amount of shift in NF2 values of /u/ between the null context and the fronting contexts (Fig. 3.11). This trend suggests that shifts in NF2 in the fronting context are not a result of physiological constraints because each speaker exhibited different amount of fronting; instead, speakers seemed to aim at distinct acoustic patterns for the fronted /u/, which are more narrowly defined than their null-context counterparts.

#### 3.6.6.4 F2 as a Function of Vowel Duration

The effect of vowel duration manipulation varied across the contexts. Figure 3.12 shows the plots of NF2 of /u/ in /D_D/, /b_d/, and /h_d/ contexts as a function of vowel duration. Each data point represents the mean NF2 and the mean duration of /u/ in the fast, medium, and slow speech rate ($N$ = 270: 30 speakers x 3 contexts x 3 rates). Linear regression lines for each context were added to the plots.

Preliminary inspection of the plots and the regression lines revealed a few patterns. First, the regression lines for /D_D/ and /h_d/ contexts showed the trend that NF2 became lower as vowel duration increased; that is, in both fronting and null contexts F2 of /u/ at its minimum had a tendency to be lower as vowel duration became longer. Second, these two regression lines did not converge or approach each other as vowel duration became longer: the lines were near-parallel. That is, the extent of NF2 differences between these two contexts remained nearly the same across the observed range of the vowels. Finally, the regression line for the /b_d/ context had a near zero slope, indicating that there was no effect of vowel duration on the vowel's NF2.

Whether the degree of fronting of /u/ in the fronting context persisted regardless of duration manipulation or not can be determined by testing whether the slope and intercept of the regression lines for fronting and non-fronting contexts were the same. This test is typically done by analysis-of-covariance (ANCOVA), which tests a series of two null hypotheses: the first null hypothesis is that the slopes of the regression lines are all the same. If this hypothesis is not rejected, the second null hypothesis that the y-intercepts of the regression lines are all the same would be tested. For the present study rejection of the second null hypothesis would support the hypothesis that NF2 of the vowels in fronting and non-fronting contexts are not the same. However, ANCOVA is not appropriate for the present data because the assumptions of independent observation and constant variance were violated: the same subjects repeated vowel productions for different contexts and different speech rates, and NF2 for /hud/ and /bud/ was much more variable than NF2 of /DuD/ (Levene statistic = 10.488 (2, 267), $p < 0.001$). Therefore, the study hypothesis was examined by using a repeated-measure mixed-model with Subject as a random factor. Context (3 levels) and Rate (3 levels) were crossed to yield 9 conditions, which were used as repeated factors. Duration (of a vowel) was the covariate, the

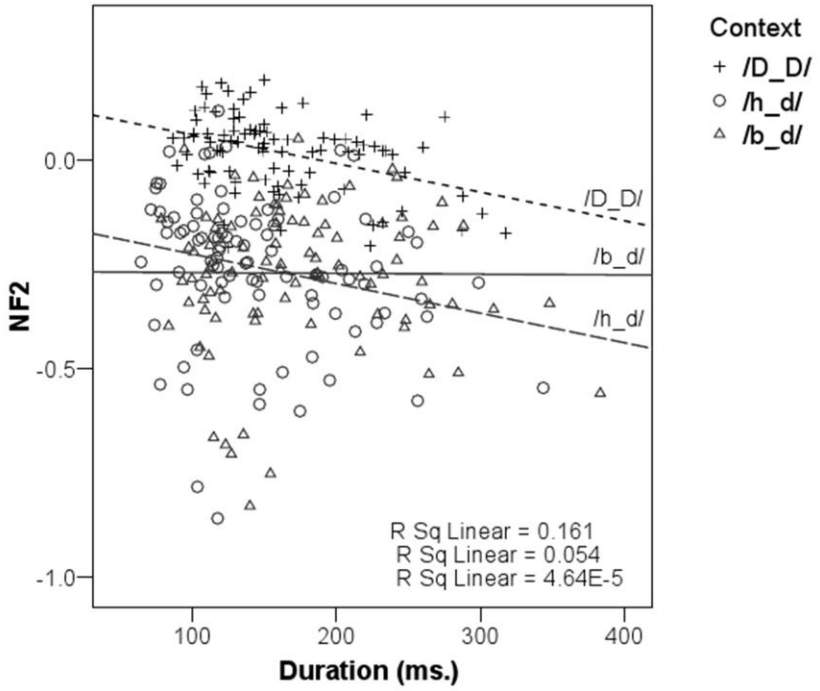



Figure 3.12 Scatter plots of NF2 values of /u/ as a function of segment duration. Each data point represents mean NF2 values of /u/ calculated from all test words (/DuD/), reference words (/hud/) and control words (/bud/) for each speech rate (fast, medium, and slow) for each speaker. ($N$ = 270: 3 contexts x 3 rates x 30 speakers). Linear regression line was added for each context.

fixed effect of which was to be controlled, and Context was the fixed factor, whose effect was to be evaluated. The model predicting NF2 of the repeated vowels is as follows:

$$\mathrm{NF2}_{ij} = (b0 + u0_j) + (b1 + u1_j)\,(\mathrm{Context})_{ij} + b2\mathrm{Duration}_{ij} + \varepsilon_{ij} \qquad (1)$$

In equation (1), the subscript $i$ represents a level of the random variable (i.e. each subject, in this case); the subscript $j$ represents a level of the fixed variable (i.e., /D_D/, /b_d/, /h_d/); $b0$ is fixed intercept; $u0_j$ reflects variability in intercepts; $b1$ is fixed slope for Context; $u1_j$ reflects variability in separate slopes; $b2$ is fixed slope for Duration; and finally $\varepsilon$ represents error. Note that the variable Rate was used as a repeated measure to accurately reflect clustering of the data, but this variable was not tested as a predictor (as the Duration replaced the same measure); therefore, in the results of the regression analyses the effect of Context reflects the effect across all three speech rates.

The results show that Context was significantly associated with NF2 after the effects of Duration was controlled (Table 3.5). Estimated mean NF2 values for /bud/, /hud/, and /DuD/ at the average value of vowel durations (161 ms) are shown in Table 3.6. Mean NF2 was highest for /DuD/, and the means for the other two contexts were very similar to each other. Thus, as shown in Table 3.7, the estimated parameter value for the /bud/ context (i.e. difference of NF2 between /b_d/ and /h_d/) was very small (0.015) and not significant [$t(21) = 0.885$, $p = 0.39$], while the parameter value for the /D_D/ context (i.e. difference of NF2 between /D_D/ and /h_d/) was large (0.283) and significant [$t(52) = 13.84$, $p < 0.01$]. From these results I conclude that /u/ produced in the fronted context is qualitatively distinct sound from /u/ in the non-fronting context.

Table 3.5 Type III tests of fixed effects on NF2: context (0=/h_d/, 1= /b_d/, 2 = /D_D/).

| Source | df 1 | df 2 | *F* | Sig. |
|---|---|---|---|---|
| Intercept | 1 | 27.224 | 3.305 | .080 |
| Context | 2 | 39.334 | 120.713 | .000 |
| Duration | 1 | 14.384 | 34.669 | .000 |

Table 3.6 Estimates of mean NF2 values at mean vowel duration (=160.88 ms).

| Context | Est. | SE | df | 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower Bound | Upper Bound |
| /b_d/ | -.246 | .025 | 129.681 | -.295 | -.196 |
| /D_D/ | .022 | .018 | 373.909 | -.014 | .058 |
| /h_d/ | -.261 | .024 | 29.904 | -.311 | -.211 |

Table 3.7 Estimates of fixed effects on NF2: Context (0=/h_d/, 1= /b_d/, 2 = /D_D/) across Duration (covariate) (N = 270), by repeated measures linear mixed model with Subject as a random factor and Context and Rate as repeated measures.

| Parameter | Est. | SE | df | *t* | Sig. |
|---|---|---|---|---|---|
| Intercept | -.145305 | .026772 | 13.392 | -5.428 | .000 |
| Conetxt /b_d/ | .015428 | .017431 | 20.881 | .885 | .386 |
| Context /D_D/ | .282967 | .020446 | 51.802 | 13.840 | .000 |
| Context /h_d/ | 0 | 0 | . | . | |
| Duration | -.000719 | .000122 | 14.384 | -5.888 | .000 |

*Note*. -2 log likelihood = -441.538

## 3.7 Summary and Discussion

The production study was conducted to answer the question of whether in American English coarticulatory fronting of /u/ in alveolar contexts is an inevitable consequence of production constraints or it is produced by deliberate speaker control, presumably as a context-specific articulatory target.

Two kinds of evidence were obtained to favor the conclusion that /u/-fronting in alveolar contexts is a controlled articulation. First, relative acoustic difference between fronted /u/ and canonical /u/ remained across differences in vowel duration (Min. = 64 ms; Max. = 383 ms). This result was further confirmed by statistically significant NF2 differences between fronted and canonical /u/. These results indicate that although vowel duration had an effect on NF2, longer vowel duration did not make these contextual variants more similar to each other. Rather, the effect of longer vowel duration was applied equally for /u/ in both fronting and non-fronting contexts. These results imply that the fronted and non-fronted variants of /u/ are distinct acoustic patterns, and speakers do not aim to make these vowels with the same articulatory patterns even in slow speech, when the articulator has more time to approximate the intended target articulation. Second, fronted /u/ did not exhibit the same degree of variability as canonical /u/. Thus, the lower the speaker's NF2 in canonical /u/, the greater the upward shift in NF2 the speaker's fronted /u/ exhibited. This result suggest that fronted /u/ has greater production constraint than canonical /u/; while speakers have relatively more freedom in articulating canonical /u/, they need to hit a more narrowly specified articulatory target for /u/ in alveolar contexts. From these interpretations, I conclude that speakers of American English have a distinctive production target for fronted /u/ in alveolar contexts separately from that for canonical /u/.

Assuming that above conclusion holds true, (1) how do we account for these production patterns in terms of control mechanisms of coarticulated sound sequences, and (2) what implication do these results have for the theory of phonologization?

There are several types of models that attempt to explain variable realizations of vowels in coarticulatory environments. One type of early models were inertia-based undershoot models (Lindblom, 1963; Moon & Lindblom, 1994; Stevens & House, 1963; Stevens, House & Paul, 1966). For example, the three regression models derived by Lindblom represent consonant-to-vowel coarticulations, wherein that vowel phonemes have invariant acoustic targets and that vowel-undershoot occurs when the articulator does not have sufficient time to hit the target. An undershoot model is clearly incompatible with the present results. Although NF2 values become lower as vowel duration increases, the NF2 differences between fronted /u/ and canonical /u/ remained the same. This is not to claim that the undershoot model is inadequate, but that observed /u/-fronting in American English is not an example of the undershoot type of coarticulation.

Another early model was Öhman's (1966, 1967) model, which is similar to the undershoot model in that it also assumes invariance. In Öhman's model, however, invariance is in the domain of neural commands rather than acoustic targets. The model represents a neural command to three independent regions of the articulatory systems—the apical, the dorsal, and the tongue body articulator, and it predicts that coarticulation would occur as long as articulatory

gestures are compatible with the gestures for adjacent segments. Consonant-vowel interactions are possible because each of the three regions responds independently to vowel commands and to consonant commands. According to this model, the tongue body responds to the vowel command, and the apical or the dorsal articulator responds to the consonant commands. In either case some parts of the articulator are left to freely coarticulate to the adjacent segment. By using this model, one might conceptualize /u/-fronting in terms of the behavior of the tongue tip and dorsum, which do not lower completely during the following vowel because these articulators do not receive a direct command for the vowel and thus are susceptible to carry-over coarticulatory effects from the previous gesture for a consonantal constriction. However, our results indicate that the effect of alveolar consonants on /u/ is much greater than what Öhman's model predicts. The smaller acoustic variability for fronted /u/ compared with canonical /u/ suggests that the configuration for the vowel is strongly constrained by the articulation of the preceding consonant. A model that accounts for variable strength of coarticulatory effects for a particular kind of consonant-vowel interaction may be more appropriate for the present results.

One recent model that explicitly accounts for variable degrees of coarticulatory effects is a gestural model within *articulatory phonology* (Browman & Goldstein, 1986, 1990, 1992). In articulatory phonology, the basic phonological unit is the articulatory gesture, which is defined as a member of a set of of functionally equivalent articulatory movements that are actively controlled to form a given phonetic goal (Saltzman & Munhall, 1989), and coarticulation is modeled as an overlap between gestures (Browman & Goldstein, 1992; Fowler & Saltzman, 1993). According to articulatory phonology, such gestural overlap may be organized into a gestural constellation such that the onset of a vowel gesture is phased with the onset of a preceding consonant, ensuring a strong coarticulatory effect. For example, by using the gestural activation wave (Fowler & Saltzman 1993), /u/-fronting in American English can be represented in terms of a tongue tip constriction gesture making an extended carryover field into the following vowel by combination of strengthened CV coupling by virtue of being in word-initial position (Goldstein, Byrd & Saltzman, 2006) and coupling between alveolar consonants and /u/ that is tighter than other types of CV coupling.

Viewing /u/-fronting as a case of gestural constellation has the merit of capturing greater acoustic effects of alveolar consonants on /u/ than other types of CV coarticulation, as observed in the previous studies (cf. §3.2) and lesser variability of fronted /u/ compared with canonical /u/, as observed in the present study. The articulatory phonology model fits the data nicely; however, an underlying assumption that gestural constellations emerge online as natural consequences of gestural coordination (Browman & Goldstein, 1995; Fowler & Saltzman, 1993) may or may not hold. The model would predict that a speaker of American English produce the correct vocal tract configuration for fronted /u/ without assuming a separate articulatory goal for a fronted /u/; however, there is a good reason to believe that such articulatory patterns are nonetheless mentally represented.

Mental representations are the brain's natural response to a repeatedly encountered experience, as stated in exemplar-based theories of phonological grammar (Bybee, 2001, 2006; Goldinger, 1996, 1998; Hale, 2003; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002, 2003, 2006; Wedel, 2006). The main idea of exemplar-based grammar is that all instances of speech that the speaker/hearer has experienced are stored in memory as phonetically detailed exemplars,

and grammar emerges as generalizations over these experiences (Johnson, 2006). Bybee (2006), for example, articulates the idea as follows:

> [T]he general cognitive capacities of the human brain, which allow it to categorize and sort for identity, similarity, and differences, go to work on the language events a person encounters, categorizing and entering in memory these experiences. The result is a cognitive representation that can be called a grammar. This grammar … is strongly tied to the experience that a speaker has had with language. (p. 711)

Supportive evidence for exemplar-based grammar includes word frequency effects on phonetic reduction (Bybee, 2001; Pierrehumbert, 2001), on sound change (Bybee, 2001; Jurafsky, Bell, Gregory & Raymond, 2001; Schuchardt, 1885/1972) and on word recognition (Broadbent 1967; Connine, Titone & Wang, 1993) as well as the effect of assumed talker identity on speech perception (Hay, Warren & Drager, 2006; Johnson 1997). One implementation of the exemplar-based memory (Pierrehumbert, 2003) assumes multiple layers of representation—one layer for phonetically detailed representations and higher layers for somewhat more abstracted perceptual category representations. If one accepts this type of multi-layered model, then it naturally follows that repeatedly experienced /u/-fronting would be mentally represented either as a phonetically distinct sound category, as a distinct articulatory category, or as both.[7] One would also assume that these distinct representational nodes are connected at the higher node representing lexical phonemes, because these perceptually distinct patterns do not make lexical contrast. Yet for the purpose of lexical access and producing speech, this type of model suggests that the representations used for making and understanding speech can be these intermediate levels of representations rather than lexical phonemes.

One implication of a multi-layered and exemplar-based approach to coarticulation for a theory of phonologization is that even a mechanical coarticulation can be phonologized if the output of coarticulation is acoustically and/or kinesthetically distinct and if this sound pattern is repeatedly experienced. Fronted variant of /u/ in American English certainly satisfy these conditions: it is a distinct acoustic pattern and speakers of American English repeatedly experience this sound. /u/-fronting is a likely candidate for phonologized coarticulation.

Ultimately, the question of whether coarticulated sound sequences are mentally represented or not has to be tested by a task other than speech production because it is possible for a speaker to produce, or at least for a researcher to model, contextual /u/-fronting either by (1) using a production pattern stored in memory or (2) by on-line planning for a strongly coupled CV sequence. The present study does not fully addresses the question of the mental representation of subphonemic variations, which remains a topic for future research. However, the vowel repetition study that will be reported in Chapter 5 partially addresses this question.

---

[7] Another interpretation of the results is that the mental representations that are used for speech production are diphone-based (as often used in text-to-speech synthesis). This and other larger-unit representations are compatible with the present proposal for distinct representations for distinct allophones. More studies are needed to determine exactly what units and what layers of representations are necessary and sufficient to account for empirical data.

# Chapter 4

# Perception Study

## 4.1 Introduction

Chapter 3 reported that the high back vowel /u/ that occurs between alveolar consonants is realized as its fronted variant, with much higher F2 at the vowel nucleus and optional palatal on-glide. However, listeners usually overlook coarticulatory distortions and categorize this fronted /u/ in the same way as its canonical counterpart. This phenomenon—*perceptual compensation for coarticulation*, a type of context effect whereby a listener's perception of speech segments is influenced by surrounding sounds so as to undo coarticulation—is the topic of the study reported in this chapter.

Over the past 30 years, compensation for coarticulation has been at the center of a theoretical debate in speech perception research.[1] Compensation is particularly interesting because it can be induced by multiple sources of contextual information including speech or non-speech sounds (Holt, Lotto & Kluender, 2000; Lotto & Kluender, 1998), visual input conveying information about vocal tract gestures (Fowler, 2006; Fowler, Brown & Mann, 2000), and lexical or explicit information that enables listeners to know, consciously or not, the categorical phonemic identity of the context (Elman & McClelland, 1988; Magnuson, McMurray, Tanenhaus & Aslin, 2003; Man & Repp, 1981; Ohala & Feder, 1994; Samuel & Pitt, 2003). In addition, compensation interacts with additional factor(s) such as speech rate (Lindblom & Studdert-Kennedy, 1967) and listener's linguistic background (Beddor & Krakow, 1999; Beddor, Harnsberger & Lindemann, 2002; Harrington, Kleber & Reubold, 2008) so that the amount of compensation varies depending on co-occurring conditions. Although many triggering and interacting factors have been found, exactly what mechanism in speech perception is responsible to cause and modulate compensation is not yet known. Competing explanations have been offered from different theoretical perspectives (§4.7).

[1] There is a general division of labor in the speech perception research. This division is based on the goal of a model—whether it is speech perception (how acoustic properties are interpreted in terms of basic linguistic units such as features and phonemes), word recognition (how strings of phonemes create percepts of words with associated meanings), or sentence processing (how full sentence comprehension is achieved): see, for example, Dahan and Magnuson (2006) and Samuel (2011) for discussion on this division of labor. The present paper uses the

Not only is compensation for coarticulation a central issue for a theory of speech perception, it is also an important area of inquiry for a theory of sound change. In sound change research, it is generally assumed that one precondition for common assimilatory sound change is variable listener interpretation of a coarticulated speech sound (Beddor, 2009; Blevins, 2004; Lindblom, Guion, Hura, Moon & Willerman, 1995; Ohala, 1981, 1989, 1993). However, compensatory perception normalizes contextual perturbations on a target sound in usual listening situations. In addition, listeners have multiple occasions to hear most of the words that they use. Therefore, any perception-based theory of sound change must explain how a given listener's interpretation of a coarticulated sound uttered by a speaker consistently differs from that of a speaker. Here, consistency means, for example, if a given speaker's utterance is perceived by a listener and that listener arrives at a mental representation of the utterance that differs from what the speaker assumes, then the same deviant perceptual interpretation must occur every time the same listener hears the same utterance.

Recent studies on compensation for coarticulation have reported that listeners do systematically vary in their perceptual interpretation of coarticulated sounds, and offered two factors that are responsible to this variation. Beddor (2009) proposed that individual differences in their sensitivity to the multiple acoustic cues that co-occur in a coarticulated sequence of sounds explain why a given contextual variation may be represented differently across listeners. Beddor's study on the vowel nasalization in /VNC/ sequences in American English has shown that listeners generally treat vocalic and consonantal nasality as perceptually equivalent but some listeners placed long-Ṽ short-N stimuli in the /CVNC/ category while others placed them in the /CVC/ category, indicating that listeners vary in their perceptual weights on the relevant acoustic cues. Yu (2010) proposed individual differences in cognitive processing style as yet another factor that correlates individual variation in perceptual judgment for context-induced speech variation. This proposal is based on the recent cognitive theories that have linked degree of "autistic" traits to enhanced cognitive performances such as retention of detailed information and systemizing that co-occur with difficulties in social development and communication (p. 2). Yu's study has shown that magnitude of perceptual compensation for coarticulatory /s/-retraction preceding a rounded vowel /u/ varied as a function of individual's "autistic" traits, as measured by the Autism Spectrum Quotient (AQ) and the Systemizing Quotient (SQ), and that the correlation varies between the female and the male listeners.

However, systematic variation in perception that has been found in these recent studies may or may not apply to other types of coarticulation, and there may be yet other factors that explain variation in speech perception. The study reported in this chapter addresses these gaps in our understanding of speech perception and reports on listener compensation for /u/-fronting in an alveolar context.

The purpose of this study is threefold. The first purpose is to replicate the previous findings of the compensation for contextual /u/-fronting in alveolar contexts (Harrington et al., 2008; Lindblom & Studdert-Kennedy, 1967; Ohala & Feder, 1994) and the effect of language-external factor of speech rate on perceptual compensation (Lindblom & Studdert-Kennedy, 1967). The second purpose is to investigate the range of individual variation in compensatory perception as well as to examine how systematic (or idiosyncratic) the variations is. Finally, the third purpose is to examine the relationship between speech perception and the distributional properties of speech sounds in the listener's native language.

The main arguments of this chapter are as follows: (1) compensation for coarticulation does not guarantee invariance in speech perception due to the wide range of individual variation in speech perception; (2) individual variation in compensatory perception is systematic; and (3) speech perception is guided by experience-based knowledge about the distributional properties of speech sounds in one's language.

The chapter is organized as follows. First, the chapter surveys various language-external and language-internal sources of variation in speech perception (§4.2). Next, the chapter will review the three previous studies on the compensation for coarticulatory /u/-fronting in alveolar contexts, which have provided a foundation for the present study (§4.3). Based on findings from previous studies, research questions and hypotheses will be formulated (§4.4). The chapter will then report experimental studies (§4.5 and §4.6). Finally, the chapter will discuss the implications of these findings for sound change and for theories of speech perception (§4.7). The chapter ends with a prospectus for future research on the issue of how listeners handle speech variation, with particular focus on the ways the listener mentally represents pronunciation variation.

## 4.2 Variation in Speech Perception

Speech signals are inherently variable, and one major source of speech variation is coarticulation. As reported in chapter 3, coarticulation consists of universal biomechanical and language-specific phonological components: biomechanical constraints determine the direction of coarticulatory perturbations (e.g. /u/ is *fronted* in the context of alveolar consonants), and phonological knowledge guides the degree of coarticulation. Universal and language-specific components are also found in speech perception. Listeners' can generally compensate for systematic covariations of the acoustic properties of natural speech (§4.2.1), but the degree of compensation varies systematically depending on listener's linguistic experience and listener expectation toward normative range of speech variation in certain linguistic contexts (§4.2.2).

### 4.2.1 Effects of Contexts on Speech Perception

There is a large body of experimental studies that examine the effects of context on the perception of target speech sounds. In a commonly used methodology, an experimenter would create one or more acoustic continua, where the acoustic property of each sound is systematically altered along a relevant dimension such as VOT or formant frequency, so that the perceived phonemic category of each sound in a given continuum transforms, either gradually or abruptly, from one category into another. Each of these target sounds is embedded in a particular context, and subjects are asked to determine the phonemic identity of each target sound. The null hypothesis is that subjects' identification of the target sound will remain constant regardless of the context in which the sound occurs, and the alternative hypothesis is that subjects'

identification of the target sound will vary in a way that demonstrates listener compensation for contextual effects.[2]

Mann and Repp's (1980) study represents one of the early studies that used sound continua to examine listener compensation for coarticulation. They examined listener identification of synthetic fricative noise from an [s]-[ʃ] continuum when followed by either [a] or [u]. In natural speech production, fricative noise in /s/ is realized as a little more [ʃ]-like when followed by [u] because anticipatory lip protrusion for an upcoming round vowel lowers the center frequency of fricative noise. Thus, if the listener compensates for coarticulation, then the listener would identify an ambiguous fricative noise stimulus more often as [s] in the [u] context than in the [a] context. Mann and Repp's results showed this pattern. Their listeners' [s]-[ʃ] category boundary shifted toward the [s]-end in the [u] context relative to the boundary in the [a] context. Since then, numerous studies have shown converging results—perceptual category boundary shift—for consonant contexts and vowel targets (Beddor, Krakow & Goldstein, 1986; Harrington et al., 2008; Holt et al., 2000; Lindblom & Studdert-Kennedy, 1967; Ohala & Feder, 1994), consonant contexts and consonant targets (Elman & McClelland, 1988; Repp, 1980; Repp & Mann, 1980), and vowel contexts and vowel targets (Beddor et al., 2002).

Compensatory perception occurs in the context of covarying features, as well. It has been repeatedly demonstrated that F0 tends to be lower for vowels immediately following voiced consonants than those following voiceless consonants (Hombert, 1974; Hombert, Ohala & Ewan, 1979; House & Fairbanks, 1953; Lehiste & Peterson, 1961; Ohde, 1984), so if listeners compensate for this covariation, then they would more likely hear an ambiguous onset to be [+voice] when followed by a low-F0 vowel than a high-F0 vowel. Fujimura (1971) tested this hypothesis by using a synthesized stimulus series varying perceptually from [k] to [g]. For the ambiguous tokens from the middle of the continuum, listeners more often reported hearing [g] when the F0 of the following vowel is low.

Further, compensatory effects can be triggered and the degree of the effects can be influenced by non-segmental contexts. For example, Ladefoged and Broadbent (1957) tested, among other things, listeners' identification of a synthesized /bVt/ word when played back after a precursor phrase, the F1 of which was shifted down from the standard precursor. The test word was identified as ***bit*** (/bɪt/) by 87% of the subjects when preceded by the standard precursor but the same word was identified as ***bet*** (/bɛt/) by 90% of the subjects when preceded by the precursor that had lower F1, presumably because listeners took the overall low- or high-F1 context into account when judging the height of the vowel in the test word. Later, Ohala and Shriberg (1990) showed that low-pass and high-pass filtering of the precursor phrase and the target vowel stimuli can alter listeners' perceptual judgments of the target vowels along the front-back dimension.

These findings offer two important insights. First, compensation and other contrastive context effects are closely related phenomena: compensation is achieved by a dynamic process involving both local-level adjustments of a target acoustic signal relative to the immediate

[2] Compensatory perception can be tested with rating task, as well. Kawasaki (1986) used a rating task to ask her English speaking subjects to evaluate perceived nasality of originally nasalized vowel (the vowel which was produced in a context of [m_m]) in both oral and nasal contexts. The subjects gave higher 'nasality' rating for the nasalized vowel that occurred in an oral context than in a nasal context. That is, Kawasaki's subjects interpreted nasalized vowels more as oral vowels when they occurred in a nasal environment.

context as well as larger-level adjustments of the perceptual scale. The second insight is that compensation is closely linked to listener knowledge about the various types of systematic and context-dependent surface variations found in day-to-day spoken communication. The next section reviews research on this second point—influence of linguistic knowledge on speech perception.

## 4.2.2 Effects of Linguistic Knowledge on Speech Perception

Speech perception and word recognition involve interpreting acoustic signals in terms of phonemes and then to words. In addition, there is a rich body of evidence that higher-level knowledge such as semantics and lexical knowledge influence perceptual judgments on the lower-level linguistic unit such as phonemes and features. For example, Marslen-Wilson and Welsh (1978) have shown that listeners are able to shadow (i.e., repeat what they have just heard) faster when the sentences they were asked to repeat were both semantically and syntactically well-formed. Subjects were least successful in shadowing random meaningless sequences of words. For well-formed sentences, their subjects shadowed them with very short latencies, about 250 ms, or roughly the length of a single syllable. This means that in polysyllabic words they were able to recognize and begin repeating a word even before it was presented completely. These results show that: 1) listeners start narrowing down lexical candidates the moment the speech signal starts; and 2) assuming lexical candidates expedite subsequent perceptual processing. In another study, Warren (1970) demonstrated that lexical knowledge causes the *phoneme restoration* effect. When a single segment within a word (i.e. /s/ in */legislature/*) was replaced by a cough-like sound, his subjects recognized the word without any problem, and could not even tell which segment was replaced by the sound of cough, presumably due to restoration of the missing phoneme, which is guided by lexical knowledge.

Later, Elman and McClelland (1988) showed that lexically restored phonemes can cause compensation for coarticulation. Prior to their study, Mann and Repp (1980) and Repp and Mann (1981) showed that American listeners shift perceptual phonemic category boundary location on a /t/-/k/ continuum toward the /k/-end (ambiguous sounds receive more /t/-responses) in a context of preceding /ʃ/ than in a context of preceding /s/, presumably because the listeners compensate for a coarticulatory retraction of /t/ when it is heard after /ʃ/. Elman and McClelland replicated this compensation effect by using a pair of words such as *progress* and *abolish*, for which the final phoneme is /s/ and /ʃ/, respectively, as contexts but with the final consonants replaced with a synthesized sound that is intermediate between [s] and [ʃ]. Their subjects tended to perceive the ambiguous final consonant as /s/ or /ʃ/ in a way to form a real word than a non-word context (e.g. *progress* is a real word but *progresh* is a non-word) and subsequently compensate for coarticulation on a target sound from a /t/-/k/ continuum.

Compensation for coarticulation is also induced by visual stimuli[3] (Fowler, 2006; Fowler, Brown, & Mann, 2000; Mitterer, 2006). For example, Fowler and her colleagues replicated

[3] However, other studies have shown that visual information for the context only influences the identification of the concurrently occurring context but not the perception of subsequently occurring target sound (Holt, Stephens & Lotto, 2005; Vroomen & Gelder, 2001), highlighting the need for more studies to understand at what stage of

Mann's (1980) finding for /da/ bias on /da/-/ga/ continuum when preceded by /ar/ but not /al/ (due to compensation for retraction and lowered F3 of /d/ after /r/) when the context syllable was perceptually ambiguous between /al/ and /ar/, but clearly disambiguated by a simultaneous video of a speaker hyperarticulating /alda/ or /arda/.

Another type of listener knowledge that influences perceptual judgments of phoneme identity is the knowledge about gender variation in speech sounds (Hay et al., 2006; Johnson, 1990, 1991; Johnson, Strand & D'Imperio, 1999; Strand, 1999). For example, in a vowel normalization study Johnson (1990) demonstrated that listeners actively adjust perceived vowel quality depending on perceived speaker identity. He used a *hood-hud* ([hʊd]-[hʌd]) continuum, and the target stimuli were embedded in a carrier sentence that had either a rising or falling F0 contour, ending at constant F0, which is same as target word's F0. These pitch contours were designed to mimic male speakers' interrogative (rising contour, starting with low F0) and female speakers' declarative (falling contour, starting with high F0) pitch contours. Listeners made more hood responses for the ambiguous tokens in the perceived female condition than in the perceived male condition. That the observed shift in perceptual judgment was not due to a formant shift in the precursor phrase as in the case of the Ladefoged and Broadbent (1957) highlights the role of listener expectation, in this case that male talkers tend to realize /ʊ/ as slightly lower variant, somewhat more similar to /ʌ/, than females.

Evidence for the link between speech perception and phonological knowledge also comes from cross-linguistic studies on speech perception variation, which correlates with language-specific sound patterns. For example, velum lowering in Thai and American English vowel-nasal coda (VN) sequences starts during the vowel, but Thai exhibits less overlap than English (Beddor & Krakow, 1999). Consistent with this shorter duration of the nasal portion of the vowel, Thai listeners exhibit less compensation for nasalization in nasal contexts than English listeners; that is, Thai listeners perceive greater nasality from the nasalized vowels in [NVN] context than English listeners do (Beddor & Krakow, 1999). In addition, speakers of languages that differ in the degree of nasal overlap prefer different amounts of nasalization and temporal patterns of overlap when judging stimulus naturalness (Stevens, Andrade & Viana, 1987). These studies show a language-specific relationship between patterns of vowel nasalization and the perceptual judgments on nasalized vowels. One of the significant implications of these studies is that the knowledge about the language-specific degree of coarticulation also influences perceived degree of coarticulatory perturbation on the segments.

Another aspect of speech perception where a cross-linguistic difference has been observed is weighting of acoustic cues. In a study on the acoustic cues for place of articulation of stops in Japanese and American English, Fujimura, Macchi, and Streeter (1978) showed, firstly, that CV release cues dominate over VC closure cues when these cues conflict. Thus, for example, a stimulus made up by splicing /ab/ (except for the release burst) onto /da/ (starting from the burst) was heard as /ada/, instead of /abda/. Secondly, and more importantly for the purpose of the present review, their study showed different response patterns that were influenced by the stress/accent patterns of the subjects' native languages. Only American subjects showed an attenuation of the dominance of the CV release cue when the [VCCV] stimuli had a high pitch V1 and low pitch V2 pattern compared with the opposite pitch pattern. American subjects

---

perceptual processing various information is integrated.

responded to the release cue more strongly when it was high-pitched than low-pitched, presumably because the American subjects interpreted high-pitched syllables as stresses syllables. This study suggested that in addition to any physical differences between VC and CV cues, listeners' linguistic experience dictates which cues they pay most attention to.

Collectively, findings from these studies suggest that memorized sound patterns and articulatory configurations for speech sounds and sequences of these sounds that make up words influences what listeners think they hear as well as how the perceptual system processes incoming acoustic signals.

## 4.3 Perceptual Compensation for /u/-fronting

Previous studies have shown that listeners compensate for coarticulatory fronting of back vowels in the context of palatal and alveolar consonants, and these studies also found that both universal and language-specific factors interact with compensatory perception (Harrington et al., 2008; Lindblom & Studdert-Kennedy, 1967; Ohala & Feder, 1994). For example, Lindblom and Studdert-Kennedy (1967) examined listener identification of vowels in series of [jVj] and [wVw] syllables varied along two perceptual continua from [jɪj] to [jʊj] and from [wɪw] to [wʊw]. Since the [j_j] context causes fronting of high back vowels, if the listener compensates for coarticulation, then the listener would identify ambiguous vowel stimuli more often as [ʊ] in the [j_j] context than in the [w_w] context. Their results showed this pattern: the [jVj] stimuli from the middle of continuum were more often judged as [jʊj] than the [wVw] stimuli with the identical vowel being judged as [wʊw]. In other words, listeners' [ɪ]-[ʊ] category boundary shifted toward the [ɪ]-end in the palatal context relative to the boundary in the labio-velar context. In addition to demonstrating compensation for coarticulation effects, this study also found greater compensation effects in 'fast speech' stimuli than in 'slow speech' stimuli, which suggests that compensation effect occurs in a gradient manner.

Ohala and Feder (1994) tested perceptual compensation for coarticulatory fronting of /u/ in alveolar contexts with speakers of American English and examined whether acoustic signals for the context is a necessary condition for perceptual compensation or whether the phonemic identity of the context alone can induce compensatory perception. Stimuli were synthesized along an /i/-/u/ continuum, and subjects heard each of the vowel stimuli in five different conditions: 1) in isolation; 2) followed by /bə/; 3) followed by /də/; 4) and 5) were the same as (2) and (3), respectively, but the consonant and the formant transition of the /ə/ was replaced by amplitude-adjusted white noise. The listeners' task was to determine whether the target vowel in the first syllable is /i/ or /u/. Subjects wrote their responses on an answer sheet on which the second syllable (either də or bə) was pre-printed so that even when hearing the noise stimuli the subjects were made to believe that the noise was masking an actual /də/ or /bə/ syllable. Subjects identified ambiguous vowels more often as /u/ in the /də/ context than in the /bə/ context or in isolation, showing compensatory perception. Crucially, the same pattern was found with the noise stimuli. That is, listeners compensated for expected coarticulation even when the conditioning context was not acoustically present. Thus, this experiment demonstrated that

compensation for coarticulation can be induced by listener's belief about the phonemic identity of the contexts.

Finally, more recently Harrington et al. (2008) studied the interaction of perceptual compensation for /u/-fronting and age-related production variation in speakers of Southern British. They compared younger and older listeners' identification of vowels on a yeast-used (/jist/-/just/) and a sweep-swoop (/swip/-/swup/) continuum (Fig. 4.1). Both groups' category boundaries were at comparable points on the palatal continuum and were closer to the /i/-end than on the labial continuum, showing a compensation effect. However, the younger group's boundary on the labial continuum was much more fronted, hence closer to the boundary on the palatal continuum, indicating less compensation than the older group. The authors attributed these results to a difference in the listeners' own speech production: in Southern British younger speakers' /u/ phonemes are generally more fronted than the same phoneme produced by older speakers (Fig. 4.2). This study thus shows a link between the grammar governing speech production and the grammar guiding speech perception.

## 4.4. Purposes and Assumptions

The main purpose of the present study is to replicate and extend the three findings from the previous works on perception of /u/-fronting. This purpose breaks down into the three specific aims. The first is to replicate Ohala and Feder's (1994) findings of perceptual compensation for /u/-fronting in an alveolar context that was induced both by the acoustic context and the contexts that are conveyed by visual stimuli in the absence of acoustic signals (= *assumed* contexts). The second is to replicate Lindblom and Studdert-Kennedy's (1967) findings of speech rate effects on compensation. The third is to extend Harrington et al.'s (2008) finding of the age-based difference in the phonemic category judgments by testing for systematic individual variation in phonemic category judgments in a much more homogeneous listener group.

Another purpose of the study is to address an issue regarding the mechanism underlying compensation for coarticulation. One particularly heated theoretical debate on this issue concerns whether compensation uses the mechanism specific to language processing, such as motor representations of speech (e.g., Fowler, 1986; Liberman & Mattingly, 1985) or the general auditory processing of spectral contrast (e.g., Lotto, Kluender & Holt, 1997). The spectral contrast view has strong support from the finding that both speech and non-speech contexts induce comparable compensation effects (Holt & Kluender, 2000), thereby avoiding the need to access motor representations used for speech production. However, as discussed in section 4.2.2, there have been numerous demonstrations that compensation can be mediated by non-acoustic linguistic cues such as visual information for articulatory configurations and lexically restored phonemic contexts. These findings suggest that compensation is mediated by linguistic knowledge and therefore spectral contrast alone cannot account for the full range of effects. The present study intends to contribute to this debate by investigating the effect of a sentential context, which is designed to manipulate modes of speech perception between an acoustic mode and a linguistic mode, on compensatory perception.

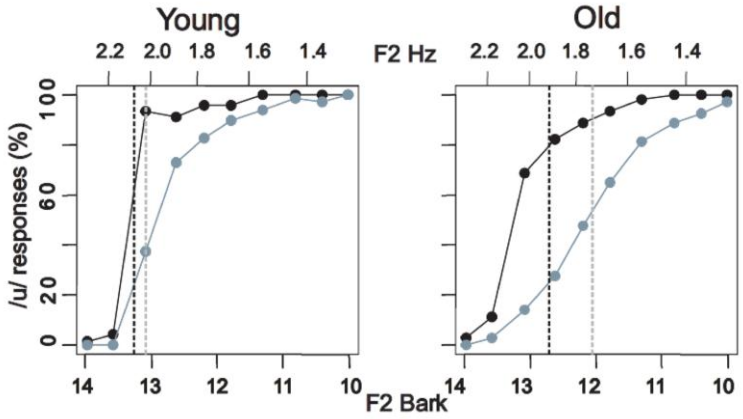



Figure 4.1 The /u/-response functions for the older (right) and younger (left) listeners of Southern British English, obtained from *used-yeast* (black) and *sweep-swoop* (gray) continua. Operationalized /i/-/u/ category boundaries are indicated by dotted lines. Reprinted with permission from "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," by J. Harrington, F. Kleber, & U. Reubold, 2008, *Journal of the Acoustical Society of America*, 123, p. 2832. Copyright 2008, Acoustical Society of America.

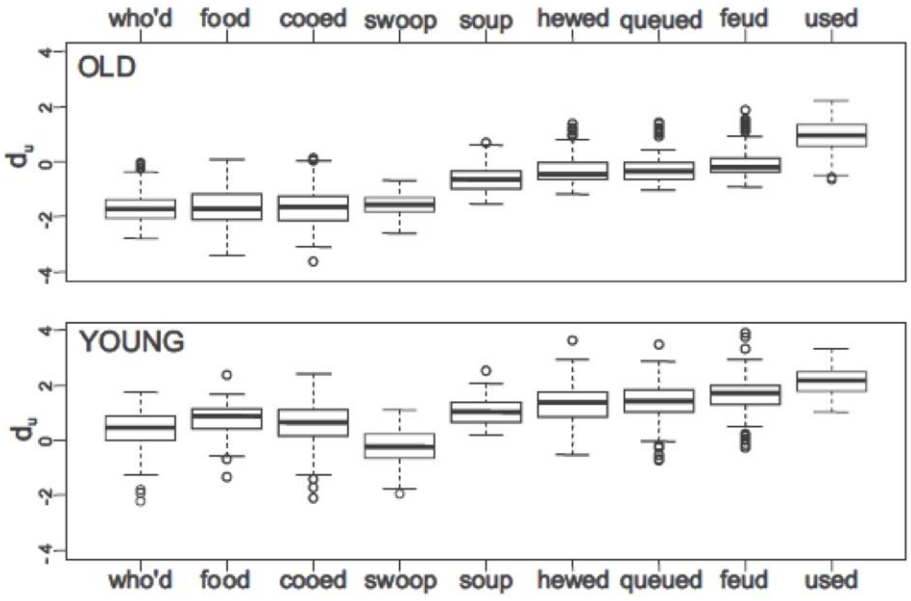



Figure 4.2 Boxplots showing relative degree of /u/-fronting (expressed by a unit-less parameter $d_u$) for the same older (above) and younger (below) listeners for various test words. The higher $d_u$ value indicates greater degree of fronting of /u/. Reprinted with permission from "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," by J. Harrington, F. Kleber, & U. Reubold, 2008, *Journal of the Acoustical Society of America*, 123, p. 2832. Copyright 2008, Acoustical Society of America.

The present study has two assumptions on phonological knowledge. The first assumption is that each language user holds in his or her long-term memory phonemic representations mapped to a range of lower-level phonetic representations, and exactly what range of phonetic representations map onto each phonemic representation is based on what the language user has previously experienced and classified as a phoneme category member. This assumption is based on the listener knowledge of internal structures of phoneme (Beddor, 2009; Blevins, 2004; Miller, 2001; Miller & Volaitis, 1989; Pierrehumbert, 2003; Volaitis & Miller, 1992; Wayland, Miller & Volaitis, 1994) and general experience- and/or exemplar-based phonological knowledge (Bybee, 2001, 2006; Hale, 2003; Goldinger, 1996, 1998; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002, 2003, 2006; Wedel, 2006). The second assumption is that the phoneme-to-phonetic mappings are available in context-specific ways (e.g. Miller 2001; Volaitis & Miller 1992). Thus language users know, for example, what the typical /u/ phoneme sounds like in /du/, where /u/ is typically fronted, and /bu/, where /u/ is a back vowel. Finally, the last assumption is that, following Johnson (1997), phonetic representations are indexed to salient socio-linguistic information, such as talker gender, dialect, etc.

## 4.5 Experimental Study 1[4]

Based on the assumptions stated above and the findings from the three previous studies on compensation for /u/-fronting, the first experiment tested the following hypotheses:

H1: The /i/-/u/ category boundary will be shifted towards the /i/-end (more stimuli that have ambiguous acoustic qualities will be heard as /u/) when the vowel is heard in the alveolar context as compared to the bilabial context (i.e. positive compensation effect resulting from the *acoustic* context).

H2: The /i/-/u/ category boundary will be shifted towards the /i/-end when the vowel is heard in the assumed alveolar context as compared to the assumed bilabial context (i.e. positive compensation effect from the *assumed* context).

H3: Greater boundary shifts (i.e. greater compensation) will be observed when the stimuli are spoken in fast speech as compared to slow speech.

H4: Listeners will vary systematically in terms of the category boundary; that is, a group of listeners whose category boundary is closer to the /i/-end than the other group in one condition will systematically exhibit the same difference in the other conditions.

In addition to testing these hypotheses, the present study addresses an issue of exactly how context alters perception of a target sound, by examining the effect of precursor phrase on the degree of compensation and on reaction time (RT). A specific research questions were:

[4] Experimental Study 1 reported here was previously published as Kataoka (2009). A Study on Perceptual Compensation for /u/-fronting in American English. In Proceedings of the 35th annual meeting of the Berkeley Linguistics Society (pp. 156-167). Berkeley, CA: Berkeley Linguistics Society.

Q1: Does an additional precursor induce greater degree of compensation, by possibly encouraging the listeners to engage in a speech mode of processing?

Q2: Does the precursor provide facilitative or impeding effects on phoneme classification that can be observed in RT data?

## 4.5.1 Method

### 4.5.1.1 Participants

Thirty-two native speakers of American English (18 female, 14 male), aged between 19 and 45 years, participated as listeners. These are the same subjects who participated in the production study reported in Chapter 3. The participants were paid $10 upon completion of the experiments. None of the participants indicated past or present speech or hearing disorders.

### 4.5.1.2 Stimuli

Six ten-step CVC continua were created by using Praat (Boersma & Weenink, 2007). Three of the continua (slow, medium, and fast speech rate) ranged perceptually between minimal pairs *beep* to *boop* (/bip/-/bup/), and the other three ranged from *deet* to *doot*[5] (/dit/-/dut/). The continua were created by concatenating three acoustic segments: (1) a natural onset stop burst, (2) a re-synthesized steady-state vowel without formant transitions, and (3) a natural coda stop burst. In addition, another two sets of CVC continua with white noise in place of the consonant intervals were created in the medium speech rate. The vowels were re-synthesized using a male speaker's natural voice source (extracted by inverse filtering) so that the stimuli would maintain the speaker's characteristic voice quality and thus sound natural when played after a precursor phrase spoken by the same speaker.

The process of vowel re-synthesis was as follows. First, a young male Californian's natural utterance of a sustained vowel /u/ was digitally recorded at 44.1 kHz and 16 bps. Then, a single period was selected from the middle of the vowel and iterated to obtain a vowel of 80 ms for the fast continua, 100 ms for the medium continua, and 120 ms for the slow continua. From each of these vowels, the source signal was extracted by: re-sampling the signal to 10 kHz; performing LPC analysis with 10 linear-prediction parameters, using an analysis window of 25 ms, time step of 5 ms, and pre-emphasis above 50 Hz; and applying inverse filtering of the LPC filter on the original sound. Next, the obtained source signal was applied to a filter, which was specified by five center frequency values and corresponding bandwidth values, to create a steady-state re-

[5] The listener's judgment might be biased toward a word-forming direction (i.e. towards *beep*/*deet* vs. *boop*/*doot*) as it has been repeatedly observed (e.g., Connine and Clifton 1987; Ganong, 1980; Pitt and Samuel 1993). However, this bias would not hide the hypothesized context effect, because the directionality of the bias is the same for both /dVt/ and /bVp/ continua.

synthesized vowel. Frequencies for each of the five formants for the end stimuli (i.e. prototypical /i/ and /u/) were determined by consulting published formant values (Hagiwara, 1997; Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952) as well as the formant frequencies of the speaker's natural utterances for /i/ and /u/. Formant frequencies and bandwidths for the /i/-end of the continuum (stimulus #1) are given in Table 4.1. To illustrate the re-synthesis process, vowel spectra of the original vowel, the same vowel after inverse filtering, and after subsequent application of a new filter, are given in Figure 4.3. The nine other vowels were created by applying nine different filters that had identical formant and bandwidth specifications except for the F2 and F3 values. These values are given in Table 4.2.

To each of the steady-state vowels, a smooth amplitude fade-in and fade-out was added by applying a half Hamming window to the first and the last 15 ms. Then, F0 contour was adjusted (by manipulating PitchTier on Praat) so that F0 varied, from 130 Hz (vowel onset) to 90 Hz (offset). Finally, from this /i/-/u/ continuum, /bip/-/bup/ and /dit/-/dut/ continua were created by adding a natural /b/ (or /d/) onset burst immediately before the vowel and a /p/ (or /t/) coda burst 70 ms after the vowel offset. The duration of each CVC syllable was 170 ms between the two stop bursts (20 ms VOT + 80 ms vowel + 70 ms coda closure) for the fast continua, 190 ms for the medium continua, and 210 ms for the slow continua. These stimuli will be referred to as the CVC stimuli.

Parallel continua were created by replacing the onset and the coda bursts with 20 ms of white noise. The white noise had a steady amplitude envelope and its amplitude matched the peak amplitude of the vowel. These stimuli will be referred to as the NVN (noise-vowel-noise) stimuli.

Three kinds of precursor phrases were created by altering the duration of a naturally uttered phrase "*I guess the word is*", spoken by the same speaker. This manipulation was done by using DurationTier on Praat. The durations of the *fast*, *medium*, and *slow* precursors were 800 ms, 1000 ms, and 1200 ms, respectively. These durations were chosen impressionistically by the experimenter for naturalness.

#### 4.5.1.3 Procedure

The experiment consisted of four blocks, each of which had two counter-balanced sub-blocks where only the bilabial stimuli (from the /bip/-/bup/ continuum) or the alveolar stimuli (from /dit/-/dut/ continuum) were presented. The first block tested the baseline compensation effect by using CVC and the NVN stimuli. Within the alveolar and the bilabial sub-blocks the medium rate CVC stimuli and the NVN stimuli were presented in isolation four times in random order. This block thus had 160 trials in total—2 contexts (alveolar vs. bilabial) x 2 conditions (acoustic vs. assumed contexts) x 10 vowels x 4 repetitions. Note that the NVN stimuli were identical in both sub-blocks. In the remaining three blocks, the fast, medium, and slow rate CVC stimuli were presented after a precursor phrase of the matching speech rate. Within the alveolar and bilabial sub-blocks, each stimulus from the /CiC/-/CuC/ continuum was presented four times in random order. Each of the three blocks thus had 80 trials in total—2 contexts (alveolar vs. bilabial) x 10 vowels x 4 repetitions.

Table 4.1 Formant frequencies and bandwidths of the stimulus #1 (/i/-end of the continuum).

| | F1 | F2 | F3 | F4 | F5 |
|---|---|---|---|---|---|
| Frequency (Hz) | 375 | 2372 | 2969 | 3500 | 4500 |
| Bandwidths (Hz) | 50 | 100 | 150 | 200 | 250 |

Table 4.2 F2 and F3 values in Hz and Bark scale for each of the ten-step vowel continuum ranges between /i/ (#1) and /u/ (#10). F2 and F3 values decrease by 0.5 and 0.18 Bark, respectively, for each subsequent step.

| | /i/ | ---------- | | | | | | | | /u/ |
|---|---|---|---|---|---|---|---|---|---|---|
| Stimulus # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| F3 (Hz) | 2969 | 2888 | 2808 | 2732 | 2658 | 2586 | 2516 | 2448 | 2382 | 2319 |
| F2 (Hz) | 2372 | 2201 | 2042 | 1895 | 1759 | 1632 | 1513 | 1402 | 1298 | 1200 |
| F3 (Bark) | 15.62 | 15.44 | 15.26 | 15.08 | 14.90 | 14.72 | 14.54 | 14.36 | 14.18 | 14.00 |
| F2 (Bark) | 14.15 | 13.65 | 13.15 | 12.65 | 12.15 | 11.65 | 11.15 | 10.65 | 10.15 | 9.65 |

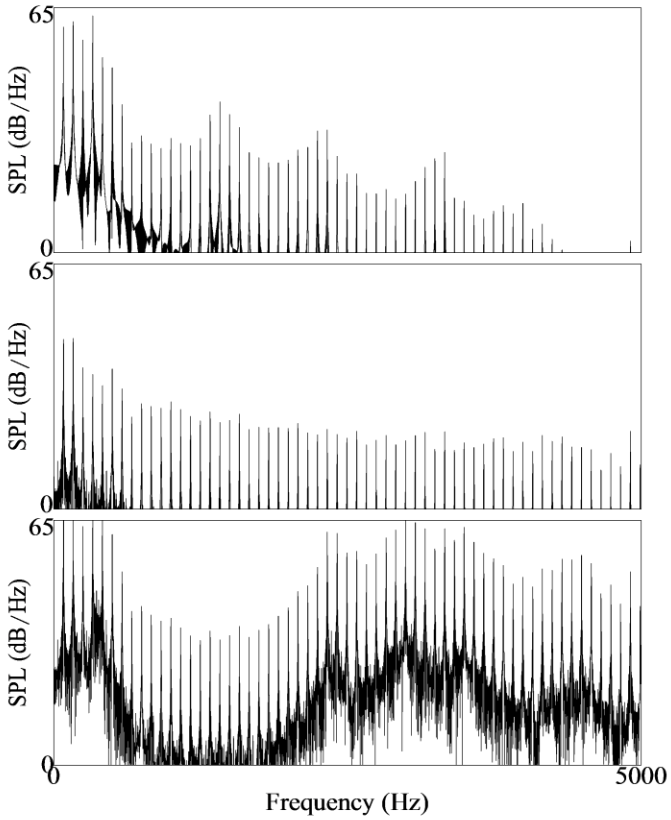



Figure 4.3 Spectra of the original vowel /u/ (A), after applying the inverse filtering to extract the voice source (B), and further applying a new filter to create a re-synthesized /i/ (C).

The listener's task was identical in all blocks: it was a classification task, where the listeners were asked, after listening to the test stimuli, to determine if the stimulus just heard was *deet* or *doot* (in the alveolar sub-block) or *beep* or *boop* (in the bilabial sub-block). Each block was preceded by a short practice block of four trials to familiarize the listeners with the task and stimuli.

The manner of stimulus presentation and response logging was also identical for all blocks. A computer monitor displayed instructions and answer options for each trial; for example, the display for the bilabial trials read "*Press [1] for 'beep'—Press [5] for 'boop'*". The listener was asked to listen carefully to each stimulus over headphones and to enter a response as quickly as possible by pressing the appropriate button on a five-button response box.

For the first block, the written instructions on the computer monitor also served the purpose of leading the listener to believe that each of the NVN stimuli was identical to its CVC counterpart (i.e., either a [dVt] or [bVp] depending on the sub-block in which the stimuli were presented), except that the onset and coda bursts were masked with white noise.

#### 4.5.1.4 Data Analyses

For each subject, a /CiC/-/CuC/ category boundary (henceforth *category boundary*) was calculated for the /bip/-/bup/ and /dit/-dut/ continuum separately, for all five conditions: (A) in isolation, acoustic contexts; (B) in isolation, assumed contexts; (C) with precursor, fast; (D) with precursor, medium; and (E) with precursor slow condition (cf. Fig. 4.4). The category boundary was defined as the location on the stimulus continuum (1-10) where the percentage for the /CuC/ response was estimated to be 50 %. Following Harrington et al. (2008), the estimated 50% boundary was calculated using probit analysis. Probit analysis is a special case of the generalized linear model, which is used to analyze binomial response variables (e.g. response with /CuC/ or not, in our case). As the original response functions were not linear, the proportions of /CuC/-responses were first transformed to probability units, or probits[6], so that, the response curve becomes linear, and then the linear regression lines were derived. Only the data between asymptotic regions (i.e. for the levels of stimulus for which estimated probability of an /u/-response was between 0.01 and 0.99) were used to fit the regression lines. Thus for all listeners, only the subset of the response data (mostly the responses to the stimuli between #3, and #7) were used. Once the regression equation was derived, then stimulus location for the 50% crossover point was obtained from the regression equation by solving for the stimulus position that corresponds to a probit value of 0.

Once the category boundaries were obtained for each subject, repeated-measures ANOVA analyses were performed to test main effects for Context (alveolar vs. bilabial), Precursor (with precursor vs. without precursor), and Rate (fast vs. medium vs. slow). The effect of Context was tested for each of the five conditions separately, by using the category boundary as the dependent

[6] The probit function (a.k.a. inverse standard normal function) is the inverse cumulative distribution function associated with the standard normal distribution. In other words, it transforms a proportion value into a Z-value at which a left-tail area under the standard normal curve corresponds to that proportion. For example, probit(0.025) = -1.96 = -probit(0.975); probit(0.5) = 0, and etc. This reflects the fact that the standard normal distribution (N(0,1)) places 95% of probability between -1.96 and 1.96, and is symmetric around zero.

variable. The null hypothesis was that the category boundaries were the same between the alveolar and the bilabial contexts. The effect of Precursor was tested by comparing the magnitude compensation effect (i.e. amount of boundary shift) between the conditions (A) and (D), between which the only difference was the absence or presence of a precursor (Fig. 4.4). The effect of Rate was tested by comparing the magnitude of compensation effect among the conditions (C), (D), and (E). For the effect of Precursor and Rate, the dependent variable was distance between the boundaries on the alveolar and bilabial continua.

In order to test for systematic individual variation in category boundaries, each listener was sorted into a *Fronter* or *Backer* group based on the results in the condition (A), and then the difference in category boundaries between the Fronter/Backer groups was tested for the conditions (C), (D), and (E). Those listeners who had mean category boundaries (midway between the boundaries on alveolar and bilabial continua) below position 4.5 were classified as Fronters, the others as Backers. Each group had 16 listeners.

Reaction time (*RT*) was measured from the stimulus onset. Out of 12800 total observations (32 listeners x 400 trials per listener), there were 71 (0.55%) missing responses.

## 4.5.2 Results

Figure 4.4 presents the mean percentage of /CuC/-responses for the /dVt/ and /bVp/ continua and a mean category boundary on each continuum in the acoustic context condition (A), assumed context condition (B), fast rate condition (C), medium rate condition (D), and slow rate condition (E). Significant Context effects were observed in the fast rate [$F(1, 31) = 18.27$; $p < 0.01$] and in the medium rate [$F(1, 31) = 4.98$; $p < 0.05$] conditions (panels C & D), partially supporting the hypothesis (H1) that listeners compensate for the fronting of a high back vowel in an alveolar context. No compensation effect was observed in the assumed context [$F(1, 31) = 0.66$; $p = 0.42$] condition, failing to support the hypothesized effect of the assumed context (H2). Rate had a significant effect on the degree of compensation [$F(2, 62) = 7.15$; $p < 0.01$] (panel C vs. D vs. E), supporting the hypothesized interaction of speech rate on the degree of perceptual compensation (H3). As for the question regarding the effect of precursor phrase on the degree of compensation (Q1), although there was a discernible increase in the degree of compensation in the with-precursor condition when compared with the no-precursor condition (panel C vs. A), this difference was not significant [$F(1, 31) = 1.69$; $p = 0.20$].

Figure 4.5 shows the distribution of the category boundary on the /dVt/ and /bVp/ continua in four conditions (acoustic, slow rate, medium rate, and fast rate) by the Fronters and by the Backers. The listeners were classified as Fronters or Backers based on the results in the acoustic-context, no-precursor condition (condition A), and then the difference between the two groups' boundaries was tested for the three different speech rate conditions. The Fronters' boundaries lie closer to the /i/-end in all three conditions. Two-tailed t-tests reveal a significant group difference in mean boundary on the /dVt/ continuum in the medium rate and the fast rate conditions, but the slow rate condition was not quite significant: slow [$t(30) = 1.93$; $p = 0.06$], medium [$t(30) = -2.37$; $p < 0.05$], fast [$t(30) = -2.12$; $p < 0.05$]. On the /bVp/ continuum, the group difference was significant in all three conditions: slow [$t(30) = -3.07$; $p < 0.01$], medium

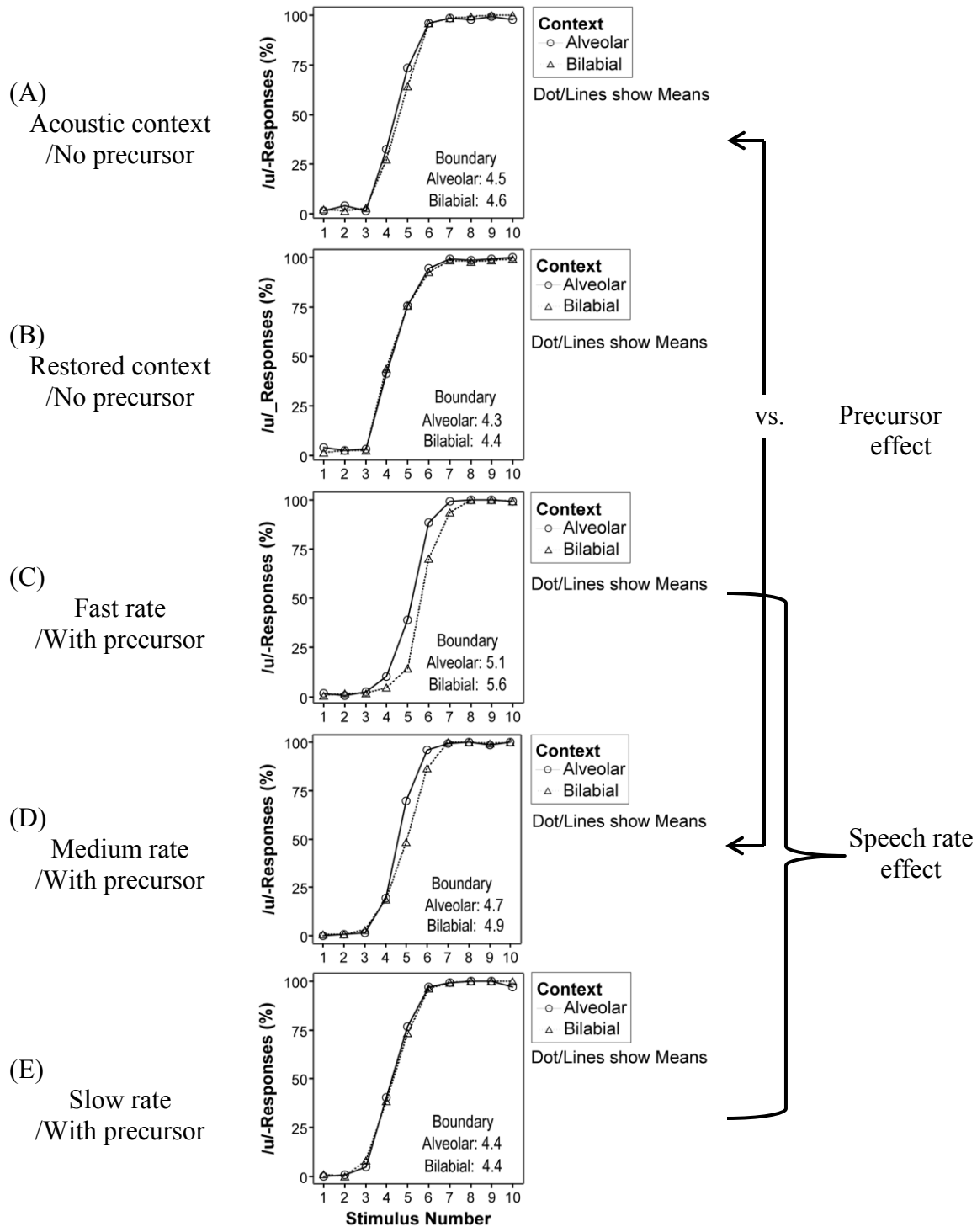



**Figure 4.4** Percentage of /CuC/-responses as a function of stimulus number on a /dVt/ continuum (solid) and a /bVp/ continuum (dotted) in the five conditions. Context effect (i.e. compensation) was tested in all five conditions. Precursor effect was tested by comparing the results from (A) and (D). Speech rate effect was tested by comparing the results from (C), (D), and (E).

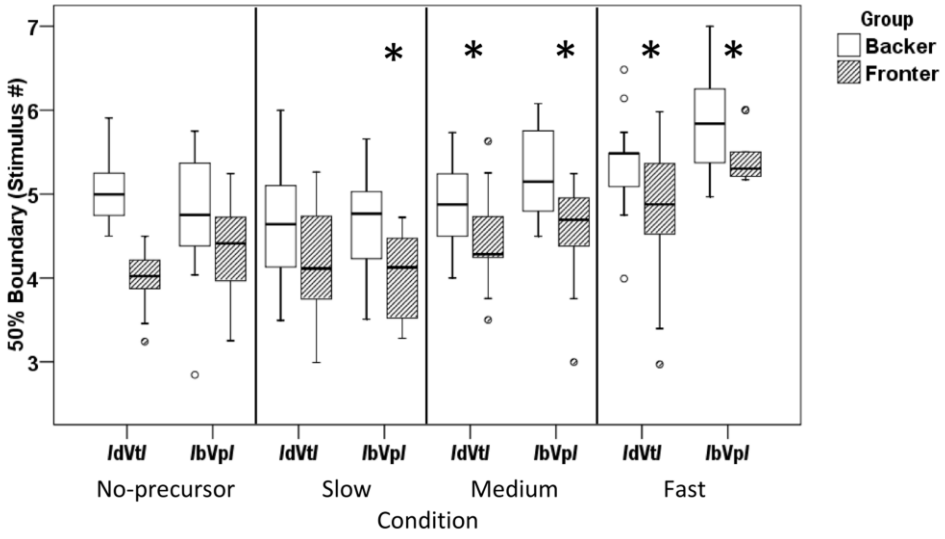



**Figure 4.5** Distribution of category boundary by Fronters (striped) and Backers (white) on a /dVt/ continuum (left two plots in a panel) and on a /bVp/ continuum (right two plots in a panel), in four conditions (no precursor, medium rate, slow rate, and fast rate). The box plots show median (thick horizontal bar), interquartile range (box), and outliers (circles). Asterisks mark continua for which there was a significant group difference in boundary.

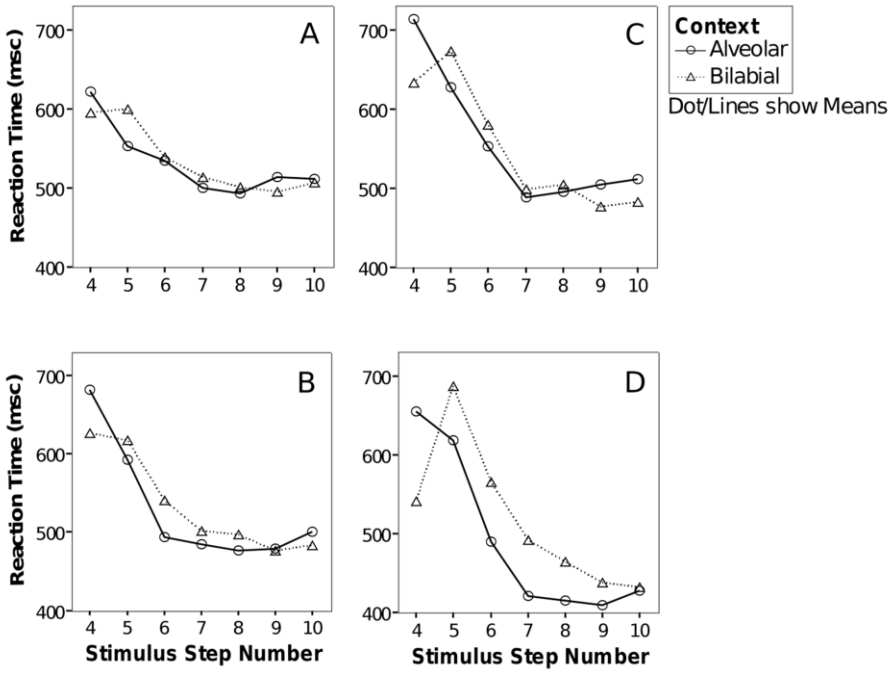



**Figure 4.6** Mean RT from /CuC/-responses as a function of stimulus number on a /dVt/ continuum (solid) and a /bVp/ continuum (dotted) in four conditions: A) no precursor; B) slow rate; C) medium rate; and D) fast rate.

[$t(30) = -3.61$; $p < 0.01$], fast [$t(30) = -2.79$; $p < 0.01$]. Thus the new hypothesis (H4) that there is a systematic individual variation in category boundary judgment was generally supported. The mean RT for the /CuC/-responses to the /dVt/ and /bVp/ stimuli (#4 and above) in all but the assumed context condition are presented in Figure 4.6. Some patterns emerge from the RT data. For stimuli #5, 6, 7, and 8, RT was shorter for /dVt/ stimuli than for /bVp/ stimuli. The RT data for stimuli #6 to #10, where within-condition RTs are relatively stable across stimuli, show that mean RTs are markedly shorter in the fast rate condition than in other conditions. The RT data show much smaller across-stimulus variation in the no-precursor (i.e., acoustic) condition as compared to the medium rate condition: this result is interesting since the target CVC stimuli were identical in duration in these conditions. This result might be interpreted as an indication that precursor phrase did influence the way subjects perceived the target stimuli (Q2). Finally, the RT trend lines uniformly exhibited down trends toward the /CuC/-end, showing that end stimuli were easier than middle stimuli to classify as a member of /CuC/ category. In the no-precursor and the medium rate conditions, the RT trend lines for the /bVp/ stimuli also had minima at a higher stimulus number than the corresponding RT minimum for the /dVt/ stimuli, showing that the hardest vowel stimulus for category identification (an indication of category boundary) varied depending on the context in which the target vowel occurred.

## 4.5.3 Discussion

The present study had mixed results in replicating the previous findings. The hypothesis that listeners compensate for the fronting of a high back vowel in an alveolar context and the hypothesis that the degree of compensation varies depending on speech rate were generally supported. These results confirm the previous findings that perceptual compensation of a target sound involves listener adjustments to both 1) the local phonetic context; and 2) global properties of the utterance, such as speech rate. The previous findings on compensation induced by assumed contexts, on the other hand, were not replicated in the present study.

The hypothesis that there is systematic individual variation in category boundary was supported in five out of the six comparisons: the Fronters, who had a category boundary closer to the /i/-end than the Backers in the no-precursor condition used to establish the groupings, had the same relative boundary location in all but the alveolar/slow condition as well. On the precursor effect on compensation and RT, there was non-significant trend that the with-precursor condition induced greater amount of compensation and greater variability in RT than the no-precursor condition even though the CVC stimuli were the same in these two conditions.

A couple of results did not confirm our expectations and resulted in the modification of the stimuli and experimental design in next study (§4.6). First, the compensation effect observed in this experiment was weaker than expected from the production data (Chapter 3). In the production experiment, mean vowel duration for the fast speech was 113 ms. The medium and slow rate stimuli used in this experiment (the vowel portion was 100 ms and 120 ms, respectively) thus correspond to, in natural situation, fast speech. The production experiment showed considerable coarticulation in CV sequences of similar vowel duration. While previous studies suggested that the degree of listener compensation reflects the degree of coarticulation in

the listeners' native language or dialect (e.g., Beddor & Krakow, 1999; Harrington et al., 2008), no compensation effect was observed in the slow speech condition.

Another surprising result was the null effect of the assumed context. Previous studies have successfully induced compensation for various types of coarticulation from lexically restored contexts as well as visual information about the contexts (e.g., Elman & McClelland, 1988; Fowler et al., 2000; Magnuson et al., 2003; Ohala & Feder, 1994; Samuel & Pitt, 2003).

The null effect in the slow speech and the assumed context conditions might be a result of the listeners attending only to the vowel portion of the stimuli and ignoring the conditioning C context. If so, how listeners can be guided to pay attention to the entire CVC sequence? One possibility is to emphasize the acoustic cues for the consonant's place of articulation in the stimuli. A second possibility is to change the experimental design. In the present experiment the conditioning contexts were constant within sub-blocks, which could allow listeners not to pay any attention to the consonant context at all. Mixing consonant contexts within a single block is a better approach. These issues will be addressed in the next section (§4.6).

Implications of the positive results (positive compensation effect, speech rate effect on compensation, systematic individual variation in perception, varying RT along stimulus continua) will be discussed in General Discussion section (§4.7).

## 4.6 Experimental Study 2

The first study confirmed listener compensation for /u/-fronting in an alveolar context and reported systematic individual variation in /CiC/-/CuC/ category boundaries. However, these findings need further reinforcement because the observed compensation effects were overall smaller than what was expected from the production data, and the first experiment did not use a wide range of experimental stimuli. Also, the previous finding of compensation from assumed contexts was not replicated. Against this background, the same experiment was conducted by using the different stimuli and design.

The second study had four purposes. The first was to re-attempt to replicate Ohala and Feder's (1994) finding that compensation for coarticulatory /u/-fronting can be elicited from assumed contexts. The second purpose was to re-test and improve the robustness of the results from the first study that show systematic individual variation in category boundaries. The stimuli used in the first study were identical across the experimental conditions, except for the stimulus duration. In this design we cannot tell whether the observed consistency in the listeners' responses across different conditions is due to the consistent acoustic properties of the stimulus or is due to consistent category judgments. The same degree of consistency may or may not be observed when different sets of stimuli are used. Therefore the same question is re-addressed in this study, this time with the stimuli based on multiple speakers' voices. The third purpose was to examine whether the degree of perceptual compensation for /u/-fronting correlates with the degree of coarticulatory /u/-fronting in the subject's own production. Harrington et al. (2008) showed this correlation in the younger and older subjects' speech productions and perceptions (see Fig. 4.1 and 4.2). A question that arises from their findings is whether this group-level correlation also holds for individuals. Finally, the fourth purpose was to

investigate a potential correlation between the distributional properties of the ambient language data and the range of acoustic variation that listeners of that speech community tolerate. As stated in section 4.4, the present study assumes that the range of acoustic variation that each listener maps onto an existing phonemic category is based on what the listener has previously experienced and classified as a phoneme category member, and this assumption requires empirical validation.

In order to achieve these purposes, Experiment 2 was designed to elicit a greater compensation effect to avoid null results due to floor effects. A probable reason that the first study did not find a strong compensation effect was that the experimental stimuli vowels lacked formant transitions, which are major acoustic cues for the place of articulation of adjacent consonants (Delattre, Liberman, & Cooper, 1955; Liberman, 1957; Stevens & Blumstein, 1987; Sussman, McCaffrey, & Matthews, 1991). Thus, the current study used improved stimuli that had more realistic acoustic specifications for the consonantal contexts. Another probable reason for the weak compensation effect was that listeners selectively diverted attention to the vowel, ignoring the consonantal context as much as possible. Since the alveolar and bilabial stimuli were presented in two separate sub-blocks, the task of judging a stimulus to be either /CiC/ or /CuC/ ultimately boiled down to determining the stimulus's vowel to be either /i/ or /u/, and listeners could ignore the consonantal context. Therefore, in the current study, both the /dVt/ and /bVp/ stimuli were presented in the same block to encourage listeners to pay attention to the entire CVC sequence.

## 4.6.1 Hypotheses and Research Questions

The three hypotheses tested in the second study were as follows:

H1: The /CiC/-/CuC/ category boundary will be shifted towards the /CiC/-end when the vowel is heard in the assumed alveolar context as compared to the assumed bilabial context.

H2: A greater assumed context effect will be observed in the mixed presentation condition than in the sub-blocked presentation condition.

H3: Listeners will vary systematically in terms of their category boundary locations; that is, /CiC/-/CuC/ category boundaries on the two different stimulus continua will positively correlate with each other, because relative rank order of the listeners along the Fronter-Backer dimension (in the sense of the study 1) should be stable across different stimuli.

In addition to testing these hypotheses, the present study also explored potential link between production and perception of coarticulation. A specific research question was:

Q1: Does the degree of perceptual compensation for /u/-fronting (as measured by amount of boundary shift) correlate with the degree of /u/-fronting in production (as measured by difference of F2 in /u/ in fronting and in non-fronting contexts)?

In addition to testing these hypotheses, the study addressed two issues: one is the range of individual variation in phonemic category judgment, and the other is the relationship between linguistic experience and speech perception. The relationship between linguistic experience and speech perception was examined by comparing the subjects' perception data with the production data obtained from the different group of subjects, who participated in the first study. In this comparison the production data are assumed to provide a model of ambient language data in the speech community from which the subjects were sampled.

## 4.6.2 Methods

### 4.6.2.1 Participants

Thirty native speakers of American English participated as subjects (15 female, 15 male; 19-29 years old). The participants were paid $10 upon completion of the experiment. None of the participants indicated past or present speech or hearing disorders.

### 4.6.2.2 Stimuli

Three sets of ten-step CVC continua ranging between minimal pairs *beep-boop* (/bip/-/bup/) and *deet-doot* (/dit/-/dut/) were created with Praat (Boersma & Weenink, 2007). Each CVC stimulus was a concatenation of a synthesized onset stop burst, a re-synthesized vowel, closure silence, and a synthesized coda stop burst. Duration for each part was: 15 ms for the onset burst, 100 ms for the vowel, 70 ms for the coda stop closure, and 15 ms for the coda stop burst (CVC total = 200 ms). The glottal source used for the re-synthesized vowels was obtained from an isolated utterance of a single vowel for each speaker so that the stimuli would sound natural when played after a precursor spoken by the same speaker. The differences between the current stimuli and the ones used in the first study were as follows:

1) The CVC continua (/bip/-/bup/ and /dit/-/dut/ continua) were created based on three different speakers' utterances so that one set modeled a female speaker's CVC syllables (*female* stimuli) and other two sets modeled two different male speakers' CVC syllables (*male1* stimuli and *male2* stimuli). Each set was given a unique F0 contour characteristic for each speaker. The female set was given formant frequencies distinct from the two male sets, which were the same.

2) The vowels had four formants instead of five formants as in the old stimuli. This was due to the frequency range of the re-synthesized vowel (0-5000 Hz). Typically, female vowels have only four formants in this range (vs. male vowels typically have five formants in the same range). The choice was made, arbitrarily, that both female and male stimuli have four formants.

3) The vowels had 40 ms of formant transitions in F2, followed by 60 ms of steady-state F2. Study 1 stimuli had no formant transitions. As in the Study 1 stimuli, the vowels did not

have transition to the following coda. This is because consonant-to-vowel coarticulation is much weaker in VC than in CV, as most of the transition is realized during the closure in the VC (see §3.3.2 and Fig. 3.8), and the vowel's terminal F2 is often comparable to F2 in the middle of the vowel.

4) Burst noise for the onset and coda was synthesized, not copied from the speaker's own utterance, in order to control the amount of acoustic information for the consonantal context across the three sets of stimuli.

The process of vowel re-synthesis was as follows. First, a natural utterance of a sustained vowel /u/ was digitally recorded at 44.1 kHz and 16 bps from one female speaker and two male speakers of American English. From each of these vowels, a glottal source of 100 ms was obtained by the same process as in Study 1 (see §4.5.2.2), then filtered. The filter was specified by four center frequency values and corresponding bandwidth values. Bandwidth for each formant had a constant value for the entire vowel duration, as did the F1, F3 and F4 frequencies. F2 frequencies were constant after the 40ms transition. During the first 40 ms of the vowel, F2 values were interpolated between the onset value and the target steady state value. F1 and F4 as well as all formant bandwidths were the same for each of the ten vowels on a given continuum. F2 and F3 were the only acoustic parameters differentiating vowel quality along the continuum. The frequency values and bandwidths for the /i/-end of the continuum for the female and the male stimuli are given in Table 4.3. The frequency values for the steady portion of F2 and F3 for each of the ten vowels in the female and the male voices are given in Table 4.4. These formant values were determined by consulting published formant values (Peterson & Barney, 1952; Hillenbrand et al., 1995; Hagiwara, 1997) as well as the formant frequencies of each speaker's natural utterances for /i/ and /u/. The vowels' onset F2 values are given in Table 4.5. These values were determined by consulting published values for the spectral peak for the onset stop release and F2 values at the vowel onset (Delattre et al., 1955; Stevens & Blumstein, 1987; Sussman et al., 1991) as well as the corresponding parameter values observed in each speaker's natural utterances for /bip/, /bup/, /dit/, and /dut/. First, the F2 locus at the stop release (constant for a given /dVt/ or /bVp/ syllable) and the vowel's target F2 (varying in ten-steps) were determined. Then the vowel's onset F2 for each vowel was set to a point midway between the F2 locus and each vowel's target F2 values. F2 transition patterns for the stimuli are shown schematically in Figure 4.7.

A smooth amplitude fade-in and fade-out was added by applying a half Hamming window to the first and the last 15 ms of each of these vowels. Then, a curved F0 contour was added to obtain natural-sounding vowels. All female stimuli had the identical F0 contours (onset = 200 Hz, offset = 170 Hz), and so did all male1 stimuli (onset = 110 Hz, offset = 80 Hz) and male2 stimuli (onset = 120 Hz, offset = 90 Hz). Finally, to each of these vowels, a synthesized onset stop burst (15 ms) was added immediately before the vowel and a synthesized coda stop burst (15 ms) was added 70 ms after the vowel offset to obtain a series of 200ms CVC syllables. The onset burst was a mixture of amplitude modulated noise and "voicing" (a single period from a 110 Hz sine wave for the male stimuli and two periods from a 200 Hz sine wave for the female stimuli), and the coda burst was amplitude-modulated white noise. These sounds were bandpass filtered to give appropriate peak frequencies. The lower edge, the higher edge, and the smoothing of the passband for each stop burst are given in Table 4.6.

Table 4.3 Formant Frequencies and bandwidths (in parentheses) in Hz for F1, steady portion of F2, F3, and F4 for the /i/-end of the continuum in the female-voice and the male-voice.

| | Frequency (and Bandwidth) in Hz | | | |
|---|---|---|---|---|
| Voice | F1 | F2 | F3 | F4 |
| Female | 355 (50) | 2000 (100) | 2700 (150) | 3500 (200) |
| Male | 300 (50) | 1600 (100) | 2300 (150) | 3500 (200) |

Table 4.4 F2 and F3 values on the ten-step vowel continuum in the female and in the male voice, in Hz and in Bark. Each continuum ranges between /i/ (#1) and /u/ (#10). In both female and male continua, F2 and F3 values decrease by 0.45 and 0.18 Bark, respectively, for each subsequent step.

| | | /i/ | | | | | | | | | /u/ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Stimulus # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Female | F3 (Hz) | 3500 | 3400 | 3303 | 3210 | 3120 | 3033 | 2949 | 2868 | 2789 | 2713 |
| | F2 (Hz) | 2758 | 2575 | 2405 | 2248 | 2102 | 1965 | 1838 | 1718 | 1606 | 1500 |
| | F3 (Bark) | 16.66 | 16.45 | 16.30 | 16.12 | 15.94 | 15.76 | 15.58 | 15.40 | 15.22 | 15.04 |
| | F2 (Bark) | 15.14 | 14.69 | 14.24 | 13.79 | 13.34 | 12.89 | 12.44 | 11.99 | 11.54 | 11.09 |
| Male | F3 (Hz) | 2969 | 2888 | 2808 | 2732 | 2658 | 2586 | 2516 | 2448 | 2382 | 2319 |
| | F2 (Hz) | 2394 | 2237 | 2092 | 1956 | 1829 | 1710 | 1598 | 1493 | 1394 | 1300 |
| | F3 (Bark) | 15.62 | 15.44 | 15.26 | 15.08 | 14.90 | 14.72 | 14.54 | 14.36 | 14.18 | 14.00 |
| | F2 (Bark) | 14.21 | 13.76 | 13.31 | 12.86 | 12.41 | 11.96 | 11.51 | 11.06 | 10.61 | 10.16 |

**Table 4.5** Onset F2 (Hz) for each of the ten vowels in the /dVt/ and the /bVp/ continuum in the female and in the male voice. Assumed F2 locus for each stop onset is indicated in parentheses next to the context label.

| | | /i/ | | | | | | | | | /u/ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Stimulus # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Female | /bVp/ (800) | 1779 | 1687 | 1603 | 1524 | 1451 | 1383 | 1319 | 1259 | 1203 | 1150 |
| | /dVt/ (2300) | 2529 | 2437 | 2353 | 2274 | 2201 | 2133 | 2069 | 2009 | 1953 | 1900 |
| Male | /bVp/ (300) | 1347 | 1269 | 1196 | 1128 | 1064 | 1005 | 949 | 896 | 847 | 800 |
| | /dVt/ (1900) | 2147 | 2069 | 1996 | 1928 | 1864 | 1805 | 1749 | 1696 | 1647 | 1600 |

Table 4.6 The lower, higher, and smoothing passband frequencies (Hz) for onset and coda bursts.

| | Onset/Coda | Lower | Higher | Smoothing |
|---|---|---|---|---|
| Female | /b/ and /p/ | 0 | 800 | 2500 |
| | /d/ and /t/ | 3200 | 3500 | 2500 |
| Male | /b/ and /p/ | 0 | 300 | 2500 |
| | /d/ and /t/ | 2900 | 3200 | 2500 |

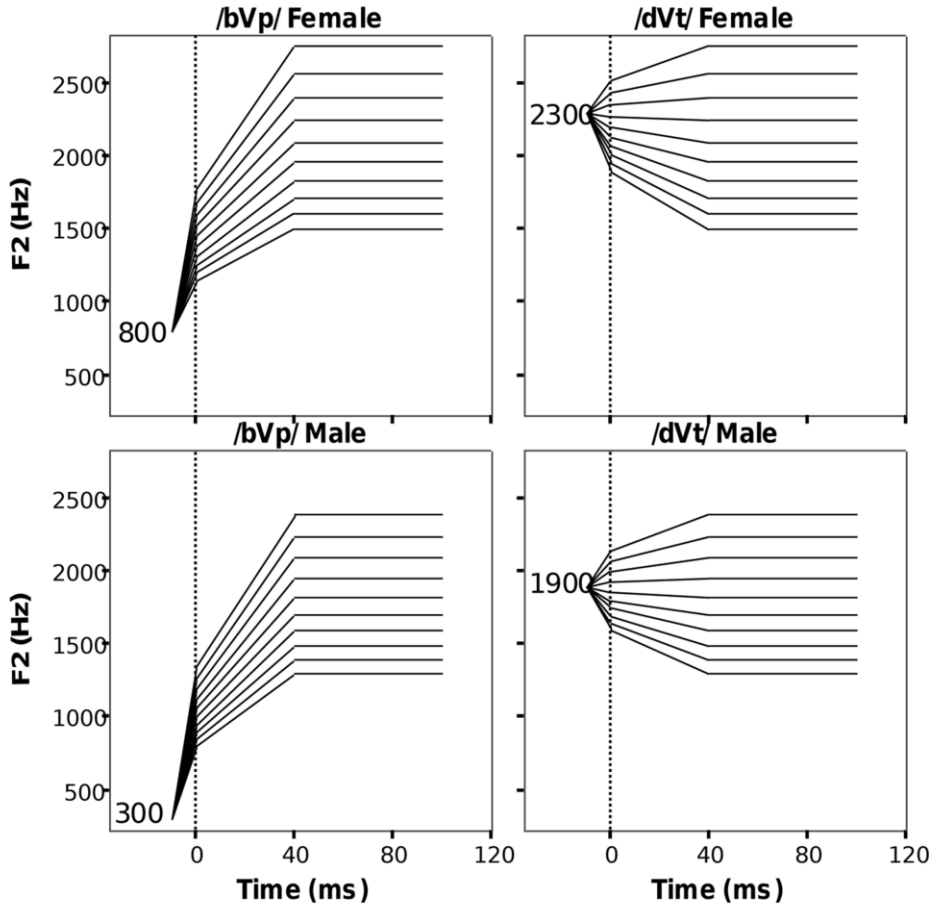



Figure 4.7 Stylized F2 trajectories for the ten vowels in the /bVp/ and /dVt/ continua (in column) in the female and the male voice (in row). The frequency values at the point where the trajectories converge indicate assumed F2 locus for each onset (female /b/, female /d/, male /b/, and male /d/), and the dotted lines indicate F2 at the vowel onset. F2 reaches to each vowel's target value at 40 ms after the vowel onset and remains constant for the rest of the vowel.

A set of ten step NVN stimuli (to be used in the assumed context condition) was created from the male2 voice. First, the ten-step vowel continuum was created in the same way as described above, except that the vowels had steady formants without transitions (no place cue for theadjacent consonant). Then amplitude-modulated white noise (50 ms) was added immediately before the vowel and 70 ms after the vowel offset to obtain sequences of noise-vowel-noise.

Finally, a precursor phrase (*"I guess the word is"*) was recorded from each speaker.

#### 4.6.2.3 Procedure

The experiment consisted of five blocks. The first block was a production task. A list of English words was created for the purpose of eliciting fronted /u/ in *dude*, canonical /u/ in *booed*, and eight English monophthongs as reference vowels (Table 4.7). The recording parameters were the same as in the previous production study (see §3.6.3 Procedure for details). The speakers were instructed to repeat each word in a carrier sentence (*"That's a ____ again"*) four times with a medium speech rate. As in the previous experiment, the term *medium rate* was explained to the speakers as "the speech rate that you would use for most normal conversational situations," and exactly how fast or slow was the speaker's own choice.

Table 4.7 Words elicited in the production part of the experiment 2

| Target sound | Word(s) elicited |
|---|---|
| Fronted /u/ | *dude* [dud] |
| Non-fronted /u/ | *booed* [bud] |
| Reference vowels | *heed* [hid], *hid* [hɪd], *head* [hɛd], *had* [hæd], *hot* [hɑt], *HUD* [hʌd], *hood* [hʊd], *who'd* [hud] |

The second block was a vowel repetition task with the male2 stimuli. This part of the experiment will be reported in the next chapter (Chapter 5: Vowel Repetition Study).

The next two blocks were for perception tasks with the female stimuli and male1 stimuli. These two sets of stimuli were presented in separate blocks in counterbalanced order. Within each block, all of the ten CVC stimuli from the /dit/-/dut/ and /bip-/bup/ continua were presented four times in random order for the classification (/CiC/ or /CuC/) tasks. There were 80 trials in each block (10 stimuli x 4 trials x 2 contexts).

The next block repeated the same task with the male2 stimuli for comparison. Note that the male2 stimuli presented in this block are identical to those presented in the second block for the vowel repetition task.

The last block tested the assumed context effect in a *sub-blocked* and *mixed* presentation conditions. Male2 NVN stimuli (and male2 acoustic stimuli as fillers) were used in this block. In a sub-blocked condition, NVN stimuli were presented only with /dVt/ fillers or /bVp/ fillers in a given counterbalanced sub-blocks.[7] Within each sub-block all of the ten stimuli from the NVN continua were presented four times, and all of the ten stimuli from the /dVt/ (or /bVp/) continua were presented once (50 trials per sub-block, random order). In a mixed condition, both fillers were presented with the NVN stimuli in a single block. All of the NVN stimuli were presented eight times, and all the fillers were presented once (100 trials, random order). Of the 30 subjects,

[7] As in the previous experiment, the written instruction for the response alternatives was expected to lead the listeners to believe that the NVN stimuli were the noise-added versions of the CVC stimuli. The ten fillers would further reinforce listeners' belief that the noise masked specific onset and coda consonants.

14 were assigned to a sub-blocked condition, and the other 16 were assigned to a mixed condition.

The procedure was the same for all perception tasks. For each trial, the target CVC (or NVN) stimulus was played after the precursor phrase with a matching voice. The manner of stimuli presentation and response logging was identical across blocks and the same as in the previous perception tasks (see §4.5.2.3 Procedure for description). Each block was preceded by a short practice block of two to four trials to familiarize the listeners with the task and stimuli.

#### 4.6.2.4 Data Analyses

*Production Data* Speakers' utterances were digitally recorded at the sampling rate of 22,050 Hz and quantized at 16 bits/sample. F1 and F2 values were measured at the vowel's temporal midpoint for all vowels (vowels in the reference words and /u/ in the test words and the control word). Next, median F1 and F2 were obtained for each word, and the median F2 values were transformed to talker normalized values (*NF2*) by using Nearey's (1978) individual log-mean method (see Chapter 3 section 3.6.5 for the detail of the normalization procedure). Finally, for each subject, the degree of /u/-fronting was calculated as the difference in NF2 (ΔNF2) in /dud/ vs. /bud/. F1 data were used only for inspection purpose to check the accuracy of the automated formant measurements.

*Perception Data* The /CiC/-/CuC/ category boundary was calculated for each listener for each experimental condition—acoustic alveolar and bilabial contexts in female, male1, and male2 voices and assumed alveolar and bilabial contexts (in male2 voice). As in Study 1, the category boundary was calculated by using probit analysis. An omnibus test was used to test for a Context main effect (H1), with Boundary as the dependent variable. Correlations among boundaries in three voices (H3) were tested by using Alveolar Boundary (i.e. boundaries on the /dVt/ continua), Bilabial Boundary (i.e. boundaries on the /bVp/ continua), and Mean Boundary (midway between the alveolar and the bilabial boundaries) as factors. In addition, the amount of boundary shift (Boundary Shift) was calculated for each listener for each condition separately. For the assumed context condition, the difference in mean boundary shift in the mixed and the sub-blocked conditions (H2) was tested by a two-sample t-test, with Boundary Shift as a dependent variable and Presentation as an independent variable. The boundary shift data from the acoustic conditions were used to test the correlation with the /u/-fronting data (H4).

### 4.6.3 Results

*Category boundary* Figure 4.8 presents the 30 subjects' mean percentage of /CuC/-responses and boundary locations on the /dit/-/dut/ and /bip/-/bup/ continua in the following conditions: female voice (A), male1 voice (B), male2 voice, acoustic context (C), male2 voice, assumed context, sub-blocked presentation (D), and male2 voice, assumed context, mixed presentation (E). Boundary data for each subject by condition are given in Table 4.8 and also visually presented in Figure 4.9.

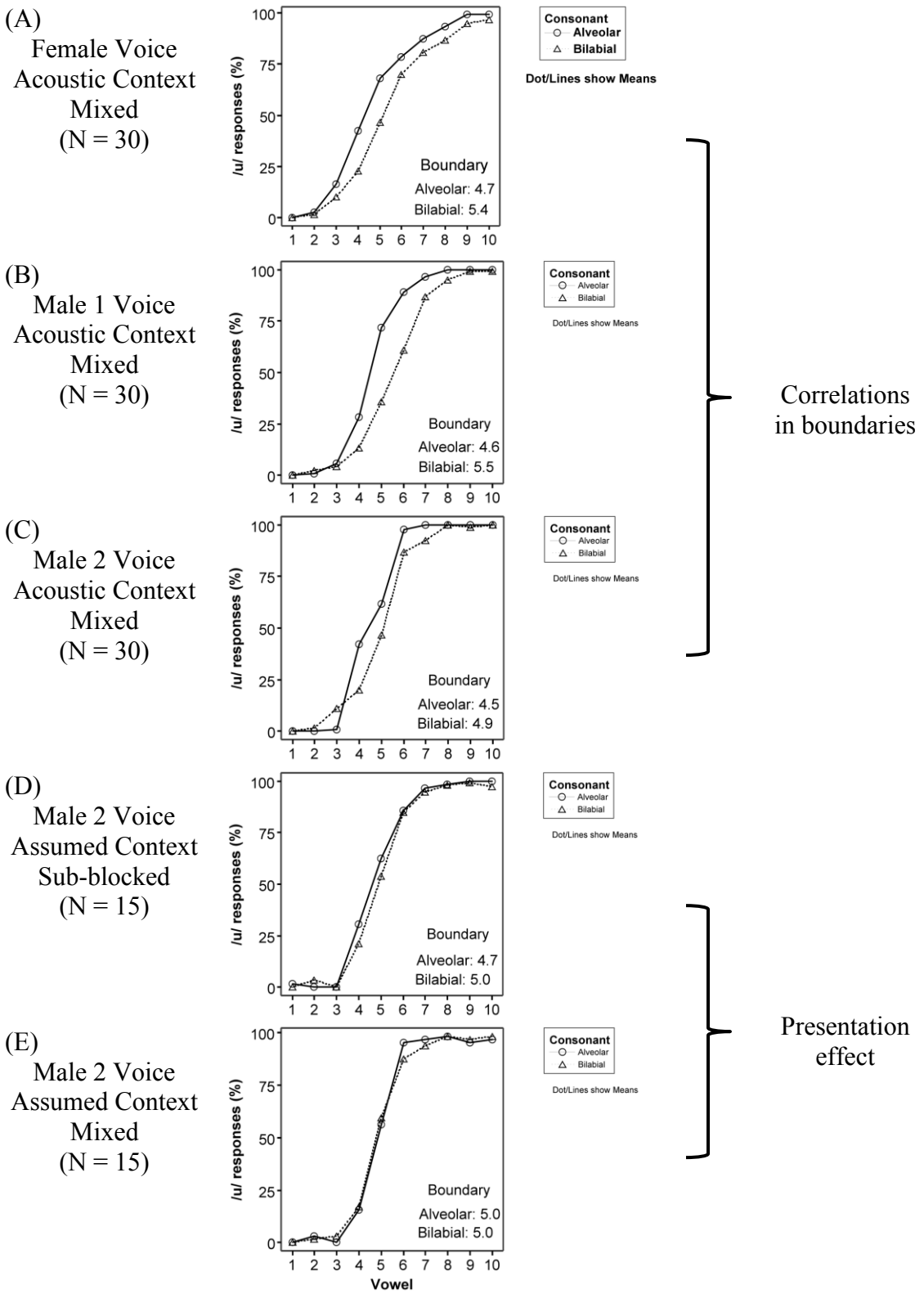



**Figure 4.8** Percentage of /u/-responses as a function of stimulus number on a /dVt/ continuum (solid) and a /bVp/ continuum (dotted) in the five conditions. Context effect (i.e. compensation) was tested in all five conditions.

**Table 4.8** /CiC/-/CuC/ category boundary as defined by 50% crossover point on a /CiC/-/CuC/ stimulus continuum for each subject for each condition.

| Voice: | Female | | Male1 | | Male2 | | Male2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Condition: | Acoustic Contexts | | | | | | Assumed Contexts | | | |
| Presentation: | Mixed | | | | | | Blocked | | Mixed | |
| Subj \ Context: | /dVt/ | /bVp/ | /dVt/ | /bVp/ | /dVt/ | /bVp/ | /dVt/ | /bVp/ | /dVt/ | /bVp/ |
| 1 | 4.3 | 4.0 | 5.0 | 5.3 | 5.5 | 5.0 | - | - | 5.7 | 4.8 |
| 2 | 4.5 | 5.0 | 4.5 | 6.5 | 3.5 | 5.8 | - | - | 4.8 | 3.6 |
| 3 | 3.0 | 5.3 | 3.0 | 5.7 | 3.5 | 4.5 | - | - | 4.0 | 4.5 |
| 4 | 4.5 | 4.5 | 5.3 | 5.7 | 4.8 | 4.3 | - | - | 4.8 | 4.5 |
| 5 | 7.9 | 9.5 | 4.0 | 5.2 | 5.8 | 5.5 | - | - | 4.5 | 5.0 |
| 6 | 4.5 | 5.0 | 5.0 | 6.5 | 5.5 | 5.5 | - | - | 4.8 | 5.4 |
| 7 | 4.1 | 4.6 | 5.0 | 5.2 | 5.5 | 5.0 | - | - | 5.2 | 7.7 |
| 8 | 5.5 | 5.9 | 7.0 | 5.9 | 5.8 | 6.3 | - | - | 6.8 | 5.3 |
| 9 | 5.0 | 6.0 | 3.7 | 5.5 | 4.5 | 6.8 | - | - | 4.5 | 4.5 |
| 10 | 4.3 | 6.7 | 4.7 | 7.7 | 3.5 | 4.8 | - | - | 5.0 | 4.3 |
| 11 | 3.5 | 4.2 | 5.0 | 6.0 | 3.8 | 4.5 | - | - | 4.8 | 5.5 |
| 12 | 4.1 | 3.9 | 4.0 | 5.4 | 3.5 | 4.8 | - | - | 5.3 | 5.2 |
| 13 | 3.5 | 4.5 | 4.4 | 5.5 | 4.3 | 4.8 | - | - | 5.0 | 4.5 |
| 14 | 4.3 | 4.5 | 5.0 | 5.8 | 5.3 | 5.5 | - | - | 4.7 | 5.3 |
| 15 | 3.5 | 4.2 | 5.0 | 5.5 | 4.5 | 5.5 | - | - | 5.2 | 4.8 |
| 16 | 3.6 | 2.8 | 4.0 | 4.5 | 4.5 | 4.5 | - | - | 4.3 | 4.0 |
| 17 | 2.8 | 3.3 | 3.8 | 4.3 | 5.8 | 3.8 | 5.0 | 5.0 | - | - |
| 18 | 3.9 | 4.0 | 5.5 | 5.5 | 4.0 | 5.5 | 5.3 | 6.4 | - | - |
| 19 | 7.0 | 9.1 | 4.0 | 4.3 | 3.5 | 3.0 | 4.0 | 4.0 | - | - |
| 20 | 4.0 | 5.7 | 3.7 | 4.0 | 3.5 | 3.0 | 3.5 | 4.5 | - | - |
| 21 | 7.6 | 7.2 | 2.8 | 3.5 | 3.8 | 3.0 | 3.5 | 3.2 | - | - |
| 22 | 6.3 | 5.2 | 4.5 | 4.5 | 3.5 | 2.8 | 4.3 | 4.6 | - | - |
| 23 | 6.8 | 8.4 | 6.0 | 8.5 | 5.5 | 7.5 | 7.3 | 7.0 | - | - |
| 24 | 3.5 | 4.7 | 4.3 | 5.0 | 3.5 | 5.3 | 4.2 | 5.0 | - | - |
| 25 | 6.0 | 7.0 | 4.5 | 6.0 | 5.2 | 5.5 | 4.7 | 4.5 | - | - |
| 26 | 3.0 | 5.7 | 4.0 | 5.3 | 3.5 | 5.0 | 4.0 | 4.2 | - | - |
| 27 | 6.8 | 8.0 | 5.0 | 5.1 | 5.5 | 7.5 | 6.0 | 5.9 | - | - |
| 28 | 4.5 | 5.0 | 5.0 | 6.8 | 4.8 | 5.5 | 5.5 | 5.7 | - | - |
| 29 | 4.5 | 4.5 | 4.3 | 6.5 | 5.5 | 5.5 | 5.3 | 5.5 | - | - |
| 30 | 2.8 | 4.0 | 5.3 | 4.8 | 3.2 | 2.0 | 4.0 | 4.4 | - | - |
| Average | **4.7** | **5.4** | **4.6** | **5.5** | **4.5** | **4.9** | **4.7** | **5.0** | **5.0** | **5.0** |

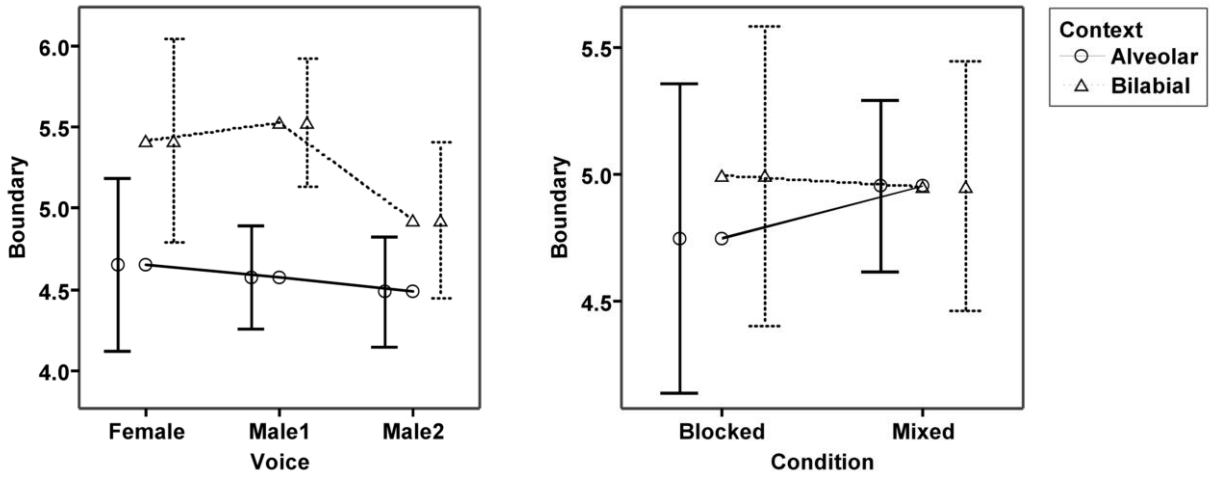



**Figure 4.9** Means (dots) and 95 % confidence intervals (error bars) of /CiC/-/CuC/ boundary location for each Context and Voice (panel A), and for each Context and Presentation.

A linear mixed-effect model was fit to the acoustic context data (Fig. 4.8, panel A) with Subjects as a random factor; Voice (female vs. male1 vs. male2) and Context (alveolar vs. bilabial) as fixed factors; and Boundary as the dependent variable. Voice and Contexts were specified as repeated measures. The results showed a significant Context main effect (i.e. compensation) [$F(1, 28) = 25.87$; $p < 0.001$] and a non-significant Voice main effect [$F(2, 28) = 2.85$; $p = 0.075$]. These results support the robustness of the perceptual compensation effect when the acoustic properties of the contexts are heard by listeners (H1). The results also showed significant Voice by Context interaction [$F(2, 29) = 3.38$; $p < 0.05$], reflecting that Context effect was significantly smaller in male2 stimuli than in the female or male1 stimuli.

A separate model was fit to the assumed context data (Fig. 4.8, panel B) with Subjects as a random factor; Presentation (sub-blocked vs. mixed) and Context (alveolar vs. bilabial) as fixed factors; and Boundary as the dependent variable. Context was specified as repeated measures. The results showed no significant Context effect [$F(1, 28) = 2.22$; $p = 0.15$], Presentation effect [$F(1, 28) = 0.07$; $p = 0.80$], or Context by Presentation interaction [$F(1, 28) = 2.27$; $p = 0.14$]. Thus, the compensation for coarticulation effect induced by assumed contexts (H2) was not replicated in this study.

Individual variation As an initial measure of within-subject consistency in the boundary locations across continua, discrepancies in the boundaries between pairs of continua were plotted (Figure 4.10). The discrepancies were noticeably smaller for the pairs of male stimuli (M1-M2 pairs) than the male-female pairs, indicating that listeners generally had similar boundaries on the two male continua and different boundary on the female continua. Also, the comparison of

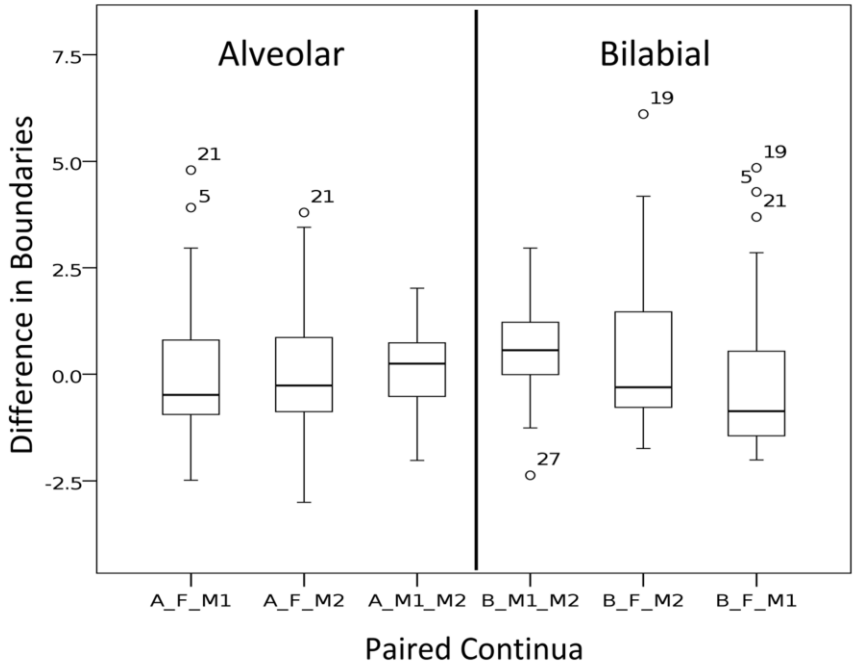



**Figure 4.10** Distribution of boundary difference on two different stimulus continua. The continua compared are: Alveolar female & male1 (F-M1), female & male2 (F-M2), and male1 & male2 (M1-M2); and Bilabial female & male1 (F-M1), female & male2 (F-M2), and male1 & male2 (M1-M2) continua.

**Table 4.9** Correlations between the paired continua in category boundary locations and their significant levels. Results are based on all data ($N$ = 30) and a subset of the data, where three outliers are excluded (N = 27).

| Paired Continua | All data ($N$ = 30) | | Outliers excluded ($N$ = 27) | |
|---|---|---|---|---|
| | Pearson Correlation | Sig. (2-tailed) | Pearson Correlation | Sig. (2-tailed) |
| A_Female vs. A_Male1 | -0.001 | 0.997 | 0.404 | 0.036 * |
| A_Female vs. A_Male2 | 0.276 | 0.140 | 0.411 | 0.033 * |
| A_Male1 vs. A_Male2 | 0.391 | 0.033 * | 0.407 | 0.035 * |
| B_Female vs. B_Male1 | 0.117 | 0.536 | 0.483 | 0.011 * |
| B_Female vs. B_Male2 | 0.218 | 0.247 | 0.528 | 0.005 * |
| B_Male1 vs. B_Male2 | 0.623 | < 0.001 * | 0.521 | 0.005 * |

* Correlation is significant at the 0.05 level (2-taied).

female-male pairs (F-M1 and F-M2) between Bilabial and Alveolar contexts revealed that discrepancies in the boundaries between female and male continua are generally greater in the bilabial context than in the alveolar context. Finally, the plots revealed that three subjects (#5, #19, and #21) had markedly larger discrepancies in their category boundaries between female and male stimuli than the other listeners. As shown in Table 4.8, these listeners had boundaries much closer to the /CuC/-end on the female continua than on the male continua. Correlations for category boundaries between the two continua were tested twice, once with all 30 subjects' data and a second time without these three outliers (Table 4.9). Correlations were significant in all paired stimuli when the three outliers are excluded. These results indicate that individual variation in perceptual phoneme category judgments is not random. Rather, listeners vary systematically in their perceptual category boundaries and resulting category judgments, supporting hypothesis H3.

Ranges of individual variation were examined in both the /u/-response functions and the boundary data. Figure 4.11 presents each subject's mean /u/-response function calculated from his/her responses on the female, male1, and male2 continua, in the alveolar (left panel) and the bilabial (right panel) contexts. These plots reveal a wide range of variation in perceptual judgments across the listeners. For example, the response functions for the [dVt] stimuli show that while one listener perceived [dVt] stimulus #4 as an instance of /dit/ 100% of the time across all three voices (12 trials in total), another listener perceived the same stimulus as a member of /dut/ 100% of the time. The same type of variation is observed in the response functions to the [bVp] stimuli. For example, one listener perceived [bVp] stimulus #5 as an instance of /bip/ 100% of the time across all three voices, and another listener perceived the same stimulus as a

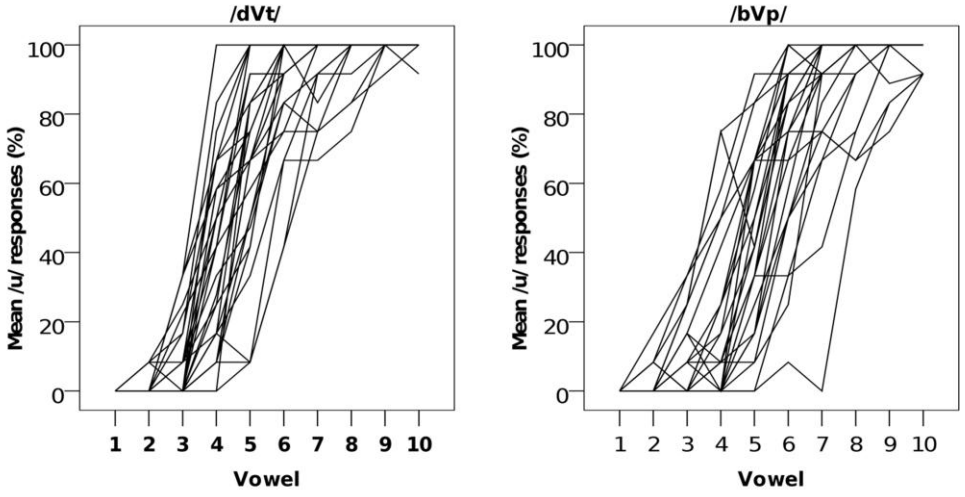



**Figure 4.11** 30 listeners' individual mean of the percentages of /u/-responses in the female, male1, and male2 stimuli, as a function of stimulus number on a /dVt/ continuum (left) and a /bVp/ continuum (right).

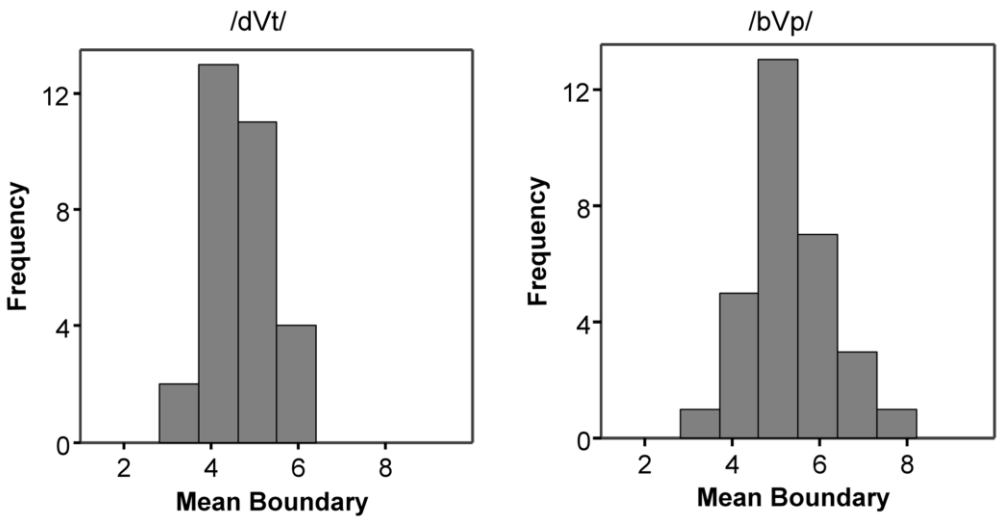



**Figure 4.12** Distribution of 30 listeners' individual mean of the boundaries in the female, male1, and male2 /dVt/ continua (left) and /bVp/ continua (right).

member of /bup/ nearly 100% of the time. Consistent responses in these cases suggest that [dVt] stimuli #4 and [bVp] stimuli #5 were ambiguous for the group of subjects as a whole but were not ambiguous for individual listeners

Context-specific range of variation Thus far the results show a systematic and considerable degree of individual variation in their perceptual judgments on the CVC stimuli. In addition, the range of across-listener variation in their perceptual judgments differed between the two contexts that were tested. Figure 4.12 presents histograms showing the distribution of individual subjects' mean boundary locations calculated from his/her boundaries on the female, male1, and male2 continua, in the alveolar (left panel) and bilabial conditions (right panel). These histograms reveal greater across-listener variability in the boundary location on the /bip/-/bup/ continua than on the /dit/-/dut/ continua. This finding is noteworthy because the directionality of this context-based difference in the range of individual variation is in the same direction as the context-specific difference in the range of production variation found in Chapter 3. As shown in figure 3.10, the back vowel /u/ had a wide range of phonetic realizations, depending on context and speaker, and across-speaker variation was much greater for /u/ in bilabial contexts than for the /u/ in alveolar contexts. That is, both in production and perception, subjects' responses were much more variable for back variants of /u/ than for fronted variants. Note that this similarity is not based on within-subject consistency between production and perception, as the production data were obtained from a different group of subjects. Rather, the results suggest American English speakers have a general tendency to produce /u/ more variably in its non-fronting contexts than in its fronting contexts, and perception of /u/ reflects that production variability.

**Table 4.10** Each subject's NF2 of /dud/ and /bud/ and their difference—a measure of degree of coarticulatory /u/-fronting. Left columns present female data, and right columns present male data. One subject's (m15) data were eliminated due to failure in recording.

| | NF2 | | ΔNF2 | | NF2 | | ΔNF2 |
|---|---|---|---|---|---|---|---|
| | /dud/ | /bud/ | | | /dud/ | /bud/ | |
| f1 | .22 | -.05 | .27 | m2 | .08 | -.27 | .35 |
| f3 | .11 | -.11 | .22 | m6 | .14 | -.46 | .60 |
| f4 | .01 | -.49 | .50 | m11 | -.06 | -.31 | .25 |
| f5 | .13 | -.29 | .42 | m14 | .07 | -.11 | .18 |
| f7 | .05 | -.21 | .26 | m15 | - | - | - |
| f8 | .06 | -.05 | .11 | m16 | .09 | -.04 | .13 |
| f9 | .10 | -.28 | .37 | m19 | .08 | -.15 | .22 |
| f10 | .13 | -.40 | .53 | m20 | .13 | -.11 | .23 |
| f12 | .09 | -.12 | .22 | m21 | -.22 | -.58 | .36 |
| f13 | .09 | -.04 | .13 | m22 | .11 | -.04 | .16 |
| f17 | .16 | -.02 | .18 | m23 | .12 | -.21 | .33 |
| f18 | .10 | -.19 | .30 | m25 | -.09 | -.33 | .24 |
| f24 | .20 | -.03 | .23 | m26 | .15 | -.24 | .39 |
| f27 | .10 | -.19 | .29 | m28 | -.06 | -.37 | .31 |
| f30 | .08 | -.48 | .56 | m29 | .10 | -.42 | .52 |

Table 4.11 Correlation between ΔNF2 and Boundary Shift.

| Paired Factors | Pearson Correlation | Sig. (2-tailed) |
|---|---|---|
| ΔNF2 vs. Shift in Female continua | 0.338 | 0.073 |
| ΔNF2 vs. Shift in Male1 continua | 0.290 | 0.127 |
| ΔNF2 vs. Shift in Male2 continua | 0.076 | 0.697 |

Production-perception link Finally, the within-subject correlation between the extent of coarticulatory fronting of /u/ and the extent of perceptual compensation for /u/-fronting was examined (Q1). Table 4.10 presents NF2 of /u/ in dude (/dud/) and booed (/bud/) and ΔNF2 (NF2 in /dud/ minus NF2 in /bud/) for each subject, and Table 4.11 presents correlations between ΔNF2 and the amount of boundary shift in the female, male1, and male2 continua. Significance levels of the correlations are also indicated. These results show no significant correlation

between the degree of fronting in production, measured as ΔNF2, and degree of perceptual compensation, measured as amount of boundary shift. Although lack of correlation cannot be proved, it is tentatively concluded that how much coarticulation one's speech exhibit is not directly linked to how much coarticulatory variation the listener tolerates.

## 4.6.4 Discussion

The results of this study supported the study's main hypotheses that listeners have stable "perception grammars" and that these grammars are listener-specific (Beddor, 2009, p. 815). Moderate but significant correlations in perceptual boundaries between two different continua with the same consonantal contexts indicate that the relative rank order of the listeners along the Fronter-Backer dimension was fairly stable within a consonantal context across different continua regardless of the differences in the F2 range of the continua and the F0 and the voice of each stimulus. This within-listener consistency in their categorical judgments suggests that individual variation in speech perception is not due to random behavior, but it is due to differences in the perceptual grammars across listeners. Listeners tacitly know the range of acoustic auditory patterns that maps onto phonemic representations and use this knowledge consistently when they hear and recognize speech sounds. Systematic individual variation was observed in perceptual judgments on ambiguous sounds in both fronting and in non-fronting contexts. Taken together, these results suggest that (1) listeners differ from each other in terms of the range of sub-phonemic variants that their grammars encompass under phonemic representations; and (2) the range of variation encompassed under a particular phoneme shifts depending on the contexts.

The lack of correlation between degree of fronting in production and degree of compensation in perception suggest dissociation between what speakers assume for their own personal production target and how much of contextual variation they tolerate for other speaker's production. This dissociation is in accord with two pieces of previous findings. One is that listeners can adapt to other speaker's idiosyncratic pronunciation pattern without altering their own production[8] (Kraljic et al., 2008), and the other is lack of negative effect of gestural mismatch between stimulus of a shadowing task (listener hears a stimulus and repeat it as quickly as possible) and a shadower's own gestural habit (Mitterer & Ernestus, 2008). These studies and the present study suggest that the perceptual knowledge and production target are only indirectly linked.

The results also revealed an interesting similarity between the range of context-specific variability between group-level production data and subjects' perceptual responses. Across-speaker variability in the phonetic realization of /u/ was much larger in non-fronting contexts than in fronting contexts, and across-listener variability in perceptual category boundaries was also wider in non-fronting contexts than in fronting contexts. The same parallel pattern between production and perception was found in Harrington et al.'s (2008) data (compare Figures 4.1 and 4.2). Their production data showed much larger between-group difference in the fronting index

[8] However, authors discussed numbers of reasons why the speakers did not change their productions and maintain that the effect of perceptual learning on the production is an open question.

($d_u$) in the non-fronting context (*booed*) than in the fronting context (*dude*). Their perception data also showed that the between-group difference in the boundary location was much larger for the *sweep-swoop* continua than for the *yeast-used* continua. Results from the present study and Harrington et al.'s study both show that: 1) in a given speech community, /u/ exhibits much wider across-speaker variation in non-fronting contexts than in fronting contexts; and 2) listeners' perceptual boundaries for the /u/ category are also much more variable when /u/ is heard in non-fronting contexts than in fronting contexts. That similar results were obtained from two different speech communities suggests that these patterns may also exist in other languages.

Assuming that the correlation between a range of production variation and a range of perception variation holds true, what mechanism(s) could underlie this correlation? One possible explanation is a direct link between the speaker's production target and the same individual's perceptual center, or idealized acoustic-auditory image. However, the results of this study challenge this explanation. If there is any such link between production targets and category centers in perceptual space, then we would expect to find significant correlations between F2 in /u/ and perceptual category boundaries, in both fronting and in non-fronting contexts. But the results contradict this expectation. Therefore, we will consider some alternative explanations in the General Discussion section (§4.7).

Finally, assumed context elicited no compensation effect. This may have been due to a floor effect. As shown in figure 4.13, compensation effect was relatively smaller for the male2 stimuli than the other stimuli even in the acoustic context condition. The compensation effects are generally smaller when induced by (1) categorical awareness of the context or (2) a lexically restored context than when induced by an acoustic context (Elman & McClelland, 1988; Magnuson et al., 2003; Ohala & Feder, 1994; Pitt & McQueen, 1998; Samuel & Pitt, 2003). Thus it is possible that the alveolar contexts that were evoked in the listeners' minds were not strong enough to cause an observable boundary shift. The fact that the stimuli were used in the vowel repetition task (Chapter 5) may also have weakened assumed context compensation effect. In the repetition task, subjects listened to the male2 CVC stimuli and repeated only the vowel. The task might have trained the listeners to dissociate the consonantal contexts from the vowel, presumably by encouraging them to pay closer attention to the vowel portion of the stimuli.

## 4.7 General Discussion

The two perception experiments described in this chapter found evidence for: 1) the robustness of perceptual compensation for coarticulatory fronting of /u/ in alveolar contexts; 2) a positive correlation between compensation effects and speech rate; 3) varying RT along stimulus continuum; 4) individual variation in perceptual judgments of ambiguous speech sounds as well as the amount of compensation; and 5) similarity between the range of context-specific within-group variation in the production and perception of /u/. In this last section, I will discuss the implications of these results for the theories of speech perception and sound change.

### 4.7.1 Implications for Theory of Speech Perception

As mentioned in Introduction, speech perception research has revealed many factors that can induce and modulate compensation for coarticulation. Currently, models of speech perception focus on only some of these aspects of the phenomenon. For example, the General Auditory theory (e.g., Holt & Kluender, 2000; Lotto et al., 1997) accounts for nonlinguistic factors in compensation. It has been well established that listeners (and even birds!) respond with compensatory perception to coarticulated speech sounds even when they have no previous experience with that particular source-target coarticulation (Mann, 1986). The same compensatory perception was observed in four-month-old infants as well (Fowler et al., 1990).

The General Auditory model is therefore well suited to account for speech rate effects, as faster formant transition rates cause greater degrees of contrastive perception effects. Speech rate effect might be explained in terms of the tendency to overshoot or extrapolate formant values for short stimuli with rapidly changing spectra (Fujisaki, 1980, p. 77; Lindblom & Studdert-Kennedy, 1967, p. 840), as this extrapolation effect has been demonstrated empirically (Divenyi, 2009; Pols & van Son, 1993). The vowels used in the Experiment 1 did not have formant transitions from which the listener could calculate transition rates; nevertheless, the spectral peak in the preceding stop burst and the beginning of vowel formants might have provided sufficient dynamism to cause perceptual extrapolation such that the vowels were perceived as having lower resonant frequencies than they actually had. This scenario explains the results of the Experiment 1 nicely: compensation would be stronger in shorter stimuli and reduced or even nullified in longer stimuli, where there is a sufficiently long steady-state region so that as the listener's spectral analysis proceeds, the extrapolated resonant frequency would match the actual frequency.

The weakness of the General Auditory model is that it does not explain language-specific effects such as listener language effects (Beddor & Krakow, 1999; Beddor et al., 2002; Harrington et al., 2008) or restored and assumed context effects (Elman & McClelland, 1988; Magnuson et al., 2003; Man & Repp, 1981; Ohala & Feder, 1994; Samuel & Pitt, 2003).

The Gestural models of speech perception (e.g., Fowler, 1986, 2006; Liberman & Mattingly, 1985; Liberman & Whalen 2000), on the other hand, focus on how speech perception is guided by articulatory knowledge, in particular knowledge of articulatory gestures. Abundant evidence shows that seeing a talker's face is beneficial for speech perception. In real life situations visual phonetic information helps people with decreasing hearing ability due to aging (Summerfield, 1987), and in the laboratory it enhances recognition accuracy against noise (Grant & Braida, 1991; Sumby & Pollack, 1954; Summerfield, 1979). The McGurk effect (McGurk & MacDonald, 1976) is an extreme example of the integration of visual articulatory information in phoneme perception. Thus from the gestural approach, compensation is conceptualized as listener sensitivity to gestural information, including visual information of the vocal gestures of the context sound.

According to the Gestural models, the speech rate effects observed in the Experiment 1 can be explained in terms of the listener's knowledge about speech production, which enables the listener to predict the degree of coarticulation from the perceived speech rate (Lindblom & Studdert-Kennedy, 1967, p. 839). This explanation is compatible with the analysis-by-synthesis explanation. Short RTs for /CuC/-responses in the alveolar context in the *fast* condition might be

taken as support for this analysis: in strong fronting contexts, low-frequency prominence might be mapped onto a back vowel more quickly than in other contexts.

However, since the model takes contextual information directly from the ongoing speech signal it does not seem to be able to account for restored and assumed context effects.

Recently, Sanderogger and Yu (2010) proposed a computational model that represents listeners' optimal categorization responses to coarticulated speech inputs in a more generalized manner. This model represents speech perception as a single computational system based on Bayesian inference, with a specific goal, in this case, of producing an optimal solution for the problem of phoneme categorization in the face of contextual variation. The model makes a conceptual departure from the traditional approaches that view speech perception as analogous to a mechanical system. The Sanderogger and Yu model's task is to take speech input $S$ in context $k$ and determine whether the input belongs to category $c1$ or $c2$ (within a 2AFC paradigm). Each category has a variable context-specific pronunciation target $T$, and $S$ is normally distributed around $T$. Perceptual compensation is represented as a shift in the 50% crossover point as a function of the context. Successful simulation of human listeners' perceptual responses suggests that the outcome of context-sensitive speech perception, irrespective of its actual mechanisms, might indeed follow a Bayesian model.

Regardless of whether a model focuses on mechanical aspects or computational aspects of the speech perception system, the challenge is to develop a more complete model that accounts for a wider range of empirical observations. Collection of more behavioral data, particularly those indicating complex interactions between phonetic context effects and other linguistically-relevant factors, which could serve as bases of these models, is needed.

Another important theoretical issue regarding speech perception is the nature of long-term mental representations of speech sounds and their relation to speech perception. The present results of varying RT along stimulus continua (Experiment1) suggest that phonemes are represented or mapped onto another layer of representation as structured distributions, with both good exemplars and poor exemplars (Iverson & Kuhl, 1995, 1996; Kuhl, 1991; Miller & Volaitis, 1989; Volaitis & Miller, 1992).

Further, the present study found evidence for systematic individual variation (i.e. within-listener consistency) in speech perception and the link between the structure of ambient language data and speech perception. These results constituted a micro-level counterpart of Harrington et al. (2008), which found systematic difference between younger British listeners and older British listeners in their context-specific phoneme category boundaries. These findings strongly suggest that the source of individual variation in speech perception is differences in phonological grammar that individual has, and that this phonological grammar that guides one's speech perception emerges in response to the ambient language data that the community members produce and the listener has exposed to through day-to-day language use. In this regard, the present study supports usage-based and exemplar-based model of phonological knowledge (Bybee, 2001, 2006; Hale, 2003; Goldinger, 1996, 1998; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002, 2003, 2006; Wedel, 2006), and extends Beddor's (2009) findings about individual variation in the phonological grammar to the case of /u/-fronting in alveolar context.

## 4.7.2 Implications for Theory of Sound Change

The findings made in the present study have significant implications for models of sound change. In "The Causes of Uniformity in Phonetic Change" (1901), Wheeler argued, against Paul's theory on articulatory drift (see Chapter 2), for the position that sound change advances from word to word until it achieves its characteristic generality. The central issue for Wheeler's position was to identify the "compelling force" (p. 15) that links a new pronunciation in one word (e.g., *home* Old English /hām/ > Middel English /hɔ̄m/) to another word (e.g. *stone* OE /stān/). After *home* had completed its vowel change, then a new form /hɔ̄m/ has no phonological link with /stān/ through which to exert the force for vowel change. Wheelers answer to this problem was as follows:

> [T]he new pronunciation of a word does not in the individual utterly displace the old. The two exist side by side… Certain conditions, a certain environment, the presence of certain hearers, suggest a preference for one above the other. … In my own native dialect I pronounced *new* as *nū*. I have found myself in later years inclined to say *nyu*, especially when speaking carefully and particularly in public; so also *tyuzdi* (Tuesday). There has developed itself in connection with these and other words *a dual sound-image* (italics added) *ū* : *yū* of such validity that whenever *ū* is to be formed after a dental (alveolar) explosive or nasal, the alternative *yu* is likely to present itself … (p. 14)

Here, a crucial argument is that the initial change does not immediately replace the old form with the new form, but rather adds the new form in a listener's mental lexicon together with the old form.[9] Wheeler's proposal explains how a speaker who has acquired an innovative pronunciation for any single word may generalize this new pronunciation to entire lexicon and become an innovator. This process thus provides one piece of the grand puzzle of how sound change takes place.

A model for the mechanism of Initial Change and Wheeler's proposal mutually complement with each other to offer a complete model of an innovator (whether the rest of the community members adopt this innovator's pronunciation or not is a separate issue, though). Ohala's (1993) model of misperception offers one such mechanism: when a listener fails to detect a coarticulatory source, and therefore also fails to associate a given coarticulatory perturbation in the speech signal with the conditioning environment, then the listener takes the coarticulated form as the intended pronunciation (Ohala, 1993, p. 246). Beddor's (2009) model offers another mechanism: individual variation in cue weighting is responsible for differences in phonological analyses of speech inputs across listeners. Yu (2010) proposed yet another mechanism: individual variation in cognitive processing style accounts for variable encoding of context-induced variation of speech signals. My study offers the fourth mechanism: individual variation

[9] Wheeler's proposal that a speaker can choose between co-existing sub-phonemic variants a more desirable form for a given communication situation has significant implications for a sociolinguistic theory of listener accommodation. This proposal is in accord with Kraljic et al.'s (2008) finding that learning more pronunciation variation does not immediately alter the listener's production, and the lack of correlation between within-speaker production and perception in the present study.

in perceptual category boundaries is responsible for differences in phonological analyses of speech inputs across listeners. Since listeners have multiple occasions to hear a given word, from different transmission channels and different speakers each time, all four mechanisms make the same prediction regarding word form learning: when word forms are learned from spoken inputs, learners may posit multiple sound-images for any word form in their mental lexicon. This prediction is compatible with what Lindblom's variation-selection model and Blevins' CHOICE model predict. I assume that these sub-phonemic variants are available for the future-reuse by the listener, and further assume this knowledge of categorical sub-phonemic and even sub-allophonic variations is necessary for style-shift in speech, or adaptation for other speaker's pronunciation habit as well as regional and social dialects. Based on these assumptions and the findings from the present study, I argue that language user's ability to have multiple sound images within one's pronunciation repertoire is the precondition of sound change.

One major weakness of the present study is that it does not offer direct evidence that listeners are capable of encoding sub-phonemic variant for a new pronunciation target; that is, this study did not show, for example, that when a listener encounters a word form /dut/ with a heavily coarticulated vowel, the listener encodes this word form as /dyt/. All it shows is that an ambiguous /dVt/ input was categorized variably as /dit/ or /dut/ across listeners. Given that English has only two high vowel phonemes, can a listener encode a novel pronunciation form other than /i/ and /u/? And if so, how? These issues are addressed in Chapter 5.

# Chapter 5

# Vowel Repetition Study

## 5.1 Introduction

The perception study reported in Chapter 4 found that: (1) subjects from a single speech community unanimously judged a certain range of variants as instances of a single phonemic category; and (2) the same subjects differed from each other in their phoneme category judgments on ambiguous speech sounds. In addition, Chapter 4 proposed, based on results from the reported experiments, that individual variation in perceptual category boundaries is responsible for differences in phonological analysis of speech inputs across listeners. The perception experiment, however, did not answer the question of whether or not variation in speech perception may result in the creation of an innovative pronunciation category. First, a classification task does not show exactly to what extent of variation a listener accepts for a given linguistic category because the task requires the listener to classify the stimuli only in terms of the categories specified by the experimenter. In the classification task with /CiC/-to-/CuC/ continua, listeners were not allowed the options "neither /CiC/ or /CuC/" or "could be both." Thus, a listener perceiving a /CVC/ stimulus from the middle of a continuum as an ambiguous auditory pattern must classify the stimulus in terms of the two given categories. As a consequence, one cannot determine what is responsible for a listener's fluctuating responses for stimuli from this mid-region of the continuum. At least three different explanations can be offered: (1) the two categories have an overlapping areas in the listener's perceptual space and the listener fluctuates in perception between /CiC/ and /CuC/; (2) the listener has another/other pronunciation category/categories between the /CiC/ and /CuC/ categories and fluctuates in response strategy (i.e. press /CiC/ or /CuC/) when the stimuli are perceived as members of this third category; and (3) the stimulus falls in a category vacuum, where there is no category to assign and the listener genuinely does not know how to interpret and how to name that stimulus, resulting in random choice between the two available options. The category boundary, as defined for the experimental study (50% crossover point), may correspond, in an assumed acoustic-auditory space of a listener, to either (1) the boundary of adjacent /CiC/ and /CuC/ categories, (2) a midpoint of overlapping /CiC/ and /CuC/ categories, or (3) a midpoint of non-overlapping, non-adjacent /CiC/ and /CuC/ categories. Data obtained from the classification task

alone does not answer the question of where context-specific phoneme boundaries[1] are in listeners' acoustic-auditory spaces.

Second, the validity of the proposal that individual variation in category boundary results in positing multiple sound-images for any word form in the listener's mental lexicon cannot be asserted unless we have evidence that listeners are capable of positing novel pronunciation categories[2] when they encounter extreme pronunciation variants. While the perception experiment in chapter 4 was successful in demonstrating that listener's judgments vary for ambiguous speech sounds, by design it could not show that any subjects categorized those stimuli as members of a unique pronunciation category.

The study reported in this chapter addresses these issues by using a different method that allows us to evaluate listener response to continuous vowel stimuli when they do not need to assign any explicit label to the perceived stimuli.

In addition, this study examines how spoken stimuli are mentally represented by a listener *for the purpose of future re-use*. In natural linguistic experiences, language users learn new words primarily through spoken communication, in which a listener hears a novel word uttered by a speaker and learns its form (pronunciation) and meaning so that the listener can produce the same word later. With this process of language learning and language use, individual language user and the rest of the community members (via speech samples produced by the community members) form a "perception-production loop" (Oudeyer, 2006; Pierrehumbert, 2001), through which community members jointly define phonological organizations of the speech sounds for their community (Bybee, 2001, 2002, 2006). To test the hypothesis of this perception-production loop, the present study examines how perceived spoken inputs are reproduced by the listeners.

Before going into methodology, the chapter briefly reviews current theories of speech perception that focus on the nature of the mental representation of speech that are stored in the long-term memory and used during perceptual processing of speech input.

## 5.2 Background: Representation-based Accounts of Speech Perception

The puzzling contrast between the variability of the acoustic signals of speech and the remarkable stability in listeners' perceptual interpretations (Blumstein & Stevens, 1981; Liberman & Mattingly, 1985; Stevens & Blumstein, 1978) has been the major issue in speech perception research, often dubbed the "lack of invariance" problem. Listeners convert variable and continuous acoustic signals into discrete linguistic units (e.g. distinctive features and phonemes) and match them to words in the lexical storage, where the meaning is accessed. To

[1] The term "context-specific phoneme" is synonymous to "allophone," which is assumed to encompass multiple phonetic variants. The present dissertation deliberately uses the term "context-specific phoneme" over "allophone" to emphasize the distinctness of certain allophones, such as /u/s in fronting and non-fronting contexts.

[2] As stated in §1.3, one main research question addressed in this dissertation is whether or not language users are capable of having distinct pronunciation categories within a single phoneme and even within a single contextual allophones.

explain the lack of invariance problem requires the understanding of how speech sounds and word forms are represented in the long-term memory (the representation problem) and how listeners map incoming speech signals onto discrete linguistic representations (the mapping problem), as pointed out in recent works (Lahiri & Reetz, 2002; Nguyen, to appear; Pisoni & Levi, 2005; Poeppel, Idsardi & van Wassenhove, 2008). These two components of speech processing play equally important roles when listeners interpret incoming speech sounds.

Many researchers have approached the lack of invariance problem by focusing on the process by which listeners achieve acoustic-to-linguistic unit mapping (process-based approach). The theories and models of speech perception and word recognition discussed in Chapter 4 exemplify this approach. Another approach, which is the focus of this chapter, is representation based. Theories from this approach vary considerably from each other in the degree of abstraction to which they assume on encoding word forms and speech sounds (see e.g. Hawkins, 2004; Nguyen, Wauquier & Tuller, 2009; Pitt, 2009 for reviews). Most traditional theories take the abstract approach (Lahiri & Reetz, 2002; Marslen-Wilson, 1987; Stevens, 2002, 2004) and assume that speech perception involves the "normalization" of fine phonetic details of speech inputs into context-independent abstract phonological units at the pre-lexical level of processing.

According to one abstract model, the Featurally Underspecified Lexicon (FUL) model (Lahiri & Reetz, 2002), words are stored in the mental lexicon as sequences of abstract, *underspecified* phonological features. For example, in the English lexicon the feature [CORONAL] is underspecified. As a consequence of this underspecification, when a listener hears an utterance, as in Lahiri and Reetz's example, "*Where could Mr. Bean be*?" with the word-final /n/ in *Bean* being pronounced as [m] due to anticipatory place assimilation, the detected feature [LABIAL] (of [m]) does not mismatch with a lexical representation for the word *Bean* because the /n/ in BEAN does not contain the feature [CORONAL]. A positive match between detected and lexical features increases the activation level for the lexical candidate, but the lack of match (e.g. [LABIAL] in the above example, which does not match any of the specified features for /n/) does not reject lexical candidates; the only "true mismatch" rejects a lexical candidate. In this mechanism, contextual variants of underspecified features do not hinder word recognition. Although some early researchers expressed doubt about the possibility of finding "acoustic invariants for the individual phonemes" (e.g. Cooper, Delattre, Liberman, Borst & Gerstman, 1952, pp. 604-605) or even about the perceptual relevance of the phoneme, most researchers have generally felt that it is important to seek acoustic correlates of phoneme-like units (Hawkins, 2004).

Recently, a growing number of researchers have taken an exemplar-based approach (Goldinger, 1996, 1998; Johnson, 1997, 2006; Pierrehumbert 2001, 2002, 2003, 2006; Wedel, 2006), which is conceptually opposed to the abstract view. The exemplar-based approach encodes linguistic categories in the mental lexicon as "exemplar clouds," which are large clusters of remembered instances, or "memory traces" of word forms that the listener has experienced and consciously attended to, and each exemplar maintains detailed information of and about the experienced words. For example, the exemplar memory stores: auditory information such as speaker voice, pitch, and other acoustic-phonetic details, situational information such as when and where the remembered speech occurred, and indexical information about the speaker such as gender and dialect (Johnson, 1997, 2006). The lack of invariance is not a problem for exemplar-based models because indexical and other information stored in the lexicon allow listeners to

deal with speech variation without altering the input signal to match a fixed internal representation. Only exemplars that match associated information are activated, while inappropriate exemplars are deactivated (Johnson, 2006, p. 383). Since listeners interpret the speech input by comparing it to stored exemplars, speech perception does not require invariant acoustic correlates to any linguistic units. Another merit of the exemplar-based model is its capacity to straightforwardly account for frequency effects of various phenomena, including sound change (Hooper, 1976; Phillips, 1984, 2001; Schuchardt, 1885/1972), allophonic alternations in production (Bybee, 2001, 2002), and spontaneous talker imitation (Goldinger, 1998; Shockley, Sabadini & Fowler, 2004).

However, simple exemplar-based models that do not allow sub-lexical and abstract representations are challenged by a body of recent empirical evidence suggesting that language users do have and use sub-lexical and abstract linguistic categories such as phonemes (see e.g. Beckman & Pierrehumbert, 2004; Pierrehumbert, 2006; Cutler, 2010 for reviews of evidence for abstract phonemic categories).

One particularly strong piece of evidence for the representation of phonemic categories comes from studies on phonemic category re-tuning, in which listeners flexibly re-tune phonemic categories and then re-apply, or generalize, this modified category to novel linguistic patterns. For example, in Norris, McQueen, and Cutler's (2003) study, Dutch listeners were exposed with /s/-final words (e.g. *naaldbos*, 'pine forest') and /f/-final words (e.g. *witlof*, 'chicory'), but the acoustic signal for the last segment was replaced with a sound that is ambiguous between /s/ and /f/ ([f-s]). A group of listeners who had heard the [f-s] replacing /f/ subsequently classified ambiguous sounds from the middle region of an [ɛs]-[ɛf] continuum more often as [ɛf] than those who had heard the [f-s] replacing the /s/-final words. The former group had expanded their /f/ categories and the latter group had expanded their /s/ categories, and both groups had applied these expanded categories in classifying the stimuli from a newly presented [ɛs]-[ɛf] continuum. Later studies have shown that listeners can re-tune phonemic categories for both consonants (Kraljic & Samuel, 2005, 2007) and vowels (Maye, Aslin & Tanenhaus, 2008).

In response to the evidence for both abstract phonemes and fine phonetic details in the language user's knowledge, recently "hybrid models" (Hawkins, 2003; Hawkins & Smith, 2001; McLennan, Luce & Charles-Luce, 2003; Pierrehumbert, 2002, 2006) have been considered as more realistic models of linguistic units in the mental lexicon.

## 5.3 Assumptions and Methodology

The present study assumes a hybrid representation, consisting of both abstract representations and phonetically detailed representations.

As an investigation tool, this study used a vowel imitation paradigm (Alibuotila, Hakokari, Savela, Happonen & Aaltonen, 2007; Chistovich, Fant, de Serpa-Leitao & Tjernlund, 1966; Kent, 1973, 1974, 1979; Repp & Williams, 1985, 1987; Schouten, 1977; Vallabha & Tuller, 2004). Implementations vary across studies, but the basic procedure is for the subjects to listen to the stimuli from vowel continua (e.g. /i/-/ɛ/-/a/ continuum, /i/-/u/ continuum, etc.) and to

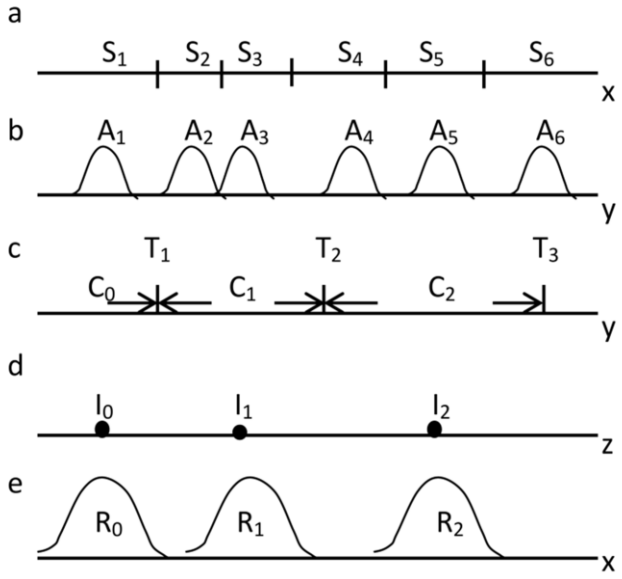



**Figure 5.1** Models of processes involved in vowel imitation. S1, S2, …, S6 represent stimulus vowels varying its quality along the x-scale. A1, A2, …, A6 are corresponding auditory patterns that reflect random error in this first stage of perception. These auditory patterns are transformed to discrete perception categories (C0, C1, and C2) with reference to category boundaries (T1, T2, and T3). The categorical output will be mapped onto a set of motor instructions (I0, I1, and I2). This set of instructions will elicit response vowels (R0, R1, and R2) that have range of variation due to random motor errors. Adapted from "Mimicking of synthetic vowels," by L. Chistovich, G. Fant, A. de Serpa-Leitao, & P. Tjernlund, 1966, *Speech Transmission Laboratory-Quarterly Progress and Status Report*, Fig. 1-A-1.

imitate[3] the vowel, one at a time, as closely as possible to the stimuli. In the present study, a modified version of the paradigm was used, in which the stimuli were in the CVC form, not the isolated vowels as in the previous studies. The listeners were asked to repeat only the vowel portion of the stimuli, not the entire CVC.[4] This modification allows testing compensation in the

[3] Chistovich et al. (1966) used the term "mimicking" in their study. Kent (1973) used the term "imitation" for the same task. Repp and Williams (1985) differentiate the term "shadowing" (listeners were asked to repeat the vowel as quickly as possible) from "mimicking" (oral responses were made after a specified time interval) but use the term "imitation" in describing their own study. In this paper, a general term "imitation" is used in describing these previous studies and the term "repetition" is used to describe this study. This choice of terminology reflects the purpose of the present study, which is to investigate how listeners would reuse words that they learn by hearing them.

[4] The idea behind this paradigm is to use re-produced vowels as proxy of mental representations. Therefore the consonantal contexts (/C_C/), the effect of which on perception of vowels was the object of the study, were kept only for the stimuli but were eliminated from the response vowels so that the response vowels would not be systematically distorted by coarticulation.

vowel repetition task while eliminating the confounding effect of coarticulation on speech production.

In order to interpret repetition data, the present study adopted a categorization model assumed by Chistovich et al., (1966) as presented in Figure 5.1. This model describes how continuous vowel stimuli are perceived, mentally represented, and reproduced. S1, S2, …, S6 represent stimulus vowels varying its quality along the x-scale. A1, A2, …, A6 are corresponding auditory patterns that reflect random errors in this first stage of perceptual processing. (One might assume, following the exemplar approach, that these patterns are all stored in the long-term memory.) For the purpose of reproduction, these auditory patterns are transformed into discrete perception categories (C0, C1, and C2) by comparing the auditory input to category boundaries (T1, T2, and T3), the output of which will be mapped onto a set of motor instructions (I0, I1, and I2) for programming and the control of the articulatory organs. This set of instructions will elicit response vowels (R0, R1, and R2) that have a range of variations due to random motor errors. By adapting this framework, the present study assumes that vowel repetition data adequately reflect mental representations of spoken inputs in the working memory as well as the number, size and internal structure of perception categories that are stored in the long-term memory.

Previous studies using the vowel imitation task found evidence that listeners have intermediate levels of representations that arise as a result of transforming continuous acoustic signals into discrete perceptual patterns, but before further processing the patterns in reference to the language's phoneme categories. For example, in Chistovich et al. (1966), where a single subject was asked to imitate a synthesized vowel (modeled after the subject's own speech) from a predetermined path of variation in the F1-F2-F3 space, the reproduced vowels did not follow the acoustic pattern of the stimuli. Instead, these vowels formed several clusters in F1, F2, and F3 dimensions, suggesting that continuous stimuli were perceived and reproduced as vowels belonging to several discrete categories. In Kent (1973), where subjects tested an /u/-/i/ continuum (beside an /i/-/æ/ continuum), the imitated vowels produced by two of the four subjects exhibited two peaks in standard deviations of the formant values. These variability peaks indicate that along the /u/-/i/ continuum the subjects had two uncertain regions for category membership, which was indicative of category boundaries, suggesting that these two subjects had a third category between /u/ and /i/. Further, relatively constant response latency observed in Chistovich et al. as well as Repp and Williams (1985) indicated that the subjects were able to imitate the vowel with the same speed regardless of uncertainties about the phonemic identities of the stimuli. These results suggest that vowel imitation is not mediated by phonemic classification (Repp & Williams, 1987). Finally, the manipulation of time interval between the stimuli presentation and the utterance of imitated vowels did not affect the imitation performance (Repp & Williams, 1985, 1987), suggesting the lack of involvement of the phonemic analysis for vowel imitation. Taken together, these findings suggest that each listener has his or her own unique set of sub-phonemic and pre-phonemic perception categories, and that vowel imitation utilizes this level of representation rather than the representations for phonemic categories.

## 5.4 Hypotheses and Research Questions

Using these conceptual and methodological tools, the present study tested the following hypotheses:

H1) When ambiguous vowels from the middle of the /CiC/-to-/CuC/ continua are repeated, response vowels will have significantly lower F2 when the stimuli are heard in the /d_t/ context than in the /b_p/ context, because perceptual compensation would make an ambiguous vowel in the alveolar context to be perceived and represented as belonging to a perceptually distinct back vowel category rather than the same vowel in the bilabial context.

H2) Individuals would vary systematically in their production category boundaries.

H3) The perceptual /CiC/-/CuC/ category boundary influences vowel imitation performance for ambiguous stimuli in that the closer to the /i/-end the subjects' perception boundaries are, the lower their response vowels' NF2 would be.

The motivation behind the first two hypotheses is to test whether the compensation for coarticulation and the systematic individual variation that were observed in the perception experiments (Chapter 4) hold in the vowel repetition experiment. The motivation behind the third hypothesis is to test the perception-production loop within subjects.[5]

In addition to testing these hypotheses, the present study also investigated how many perception categories listeners have, how listeners encode continuous vowel stimuli in their working memories, and how these vowels are reproduced by the listeners. Three research questions were as follows:

Q1) Will the /CVC/ stimuli that are better exemplars of the category be imitated in a different way from other stimuli that are not good exemplars of any category? If so, what are the acoustic characteristics of repeated vowels from the good and poor category exemplars?

Q2) Do subjects differ from each other in how categorically or continuously they repeat the continuous vowel stimuli?

Q3) How many distinct perception categories do subjects have in a /bip/-/bup/ and a /dit/-/dut/ continuum?

[5] This perception-production loop is different from the perception-production link that was tested (and rejected) in the perception experiment.

# 5.5 Experiment

## 5.5.1 Subjects, Stimuli, and Procedure

The same thirty subjects (15 female, 15 male) who participated in the perception experiment 2 (§4.6.2.1) also participated in this experiment. These two experiments were conducted in a single experimental session in two separate blocks, each of which had its own instructions, practice sessions, and test sessions. The two experiments were separated with a forced break, the duration of which was of the subjects' own choice. The vowel repetition experiment was completed before the perception experiment.

The stimuli used for the experiment were male2 stimuli that were used in the perception experiment 2 (§4.6.2.2). Briefly, these were two series of ten-step CVC syllables that form a /dit/-/dut/ continuum and a /bip/-/bup/ continuum. The duration of each CVC stimulus was 200 ms.

Preceding the vowel repetition experiment, each subject was asked to complete a short vowel production task, for which each of eight test words of /hVd/ or /hVt/ form (*heed, hid, head, had, hot, HUD, hood, and who'd*) was uttered four times in a comfortable speech rate in a carrier phrase "*That's a ___ again.*" The purpose of this production task was to obtain baseline vowel data for each subject.

The vowel repetition experiment consisted of two counter-balanced blocks, within which only the /bVp/ stimuli or the /dVt/ stimuli were presented. Within the blocks, all ten CVC stimuli were presented four times in random order, making forty trials per block. For each trial, a CVC stimulus was played back after a precursor phrase "I guess the word is __." The task of the subjects was to listen to each stimulus and to repeat ONLY the vowel as closely as possible when prompted. The prompt appeared 1,000 ms after the offset of the stimulus, and the subject was asked to utter a vowel after a precursor "*That's a __.*"

The precursor phrase was used to achieve two purposes. First, it was hoped that extra phrases added to the stimuli and the response vowels would encourage subjects to engage in the task in the speech mode rather than the acoustic mode of perception and repetition. Second, the precursor was placed before the response vowel in order to avoid potential coarticulation that might occur due to the unconscious or conscious mental rehearsal of the vowel in a consonantal context. Without a precursor, participants might utter a response vowel after rehearsing the entire syllable as in "(/dit/)-/i/" or "(/bup/)-/u/". If this occurs, the response vowels may exhibit coarticulatory distortions. In order to test the effect of *perceptual* compensation in the vowel imitation task, any potential source of coarticulation needed to be removed. Further, with the precursor phrase that ends with /ə/, subjects were made to utter a response vowel after configuring the articulator for the neutral vowel position. This way, it was hoped that the response vowels would faithfully reflect the way speech inputs are encoded in the subjects' working memory. Subjects were asked to be careful not to use /ən/ in the precursor phrase, because the alveolar nasal consonant would be coarticulated with the following response vowel. Although /ə/ and a vowel form an illegal sequence in English, subjects seemed to be generally confortable uttering /ə.V/ sequences.

## 5.5.2 Analyses and Results

The speakers' utterances were digitally recorded at the sampling rate of 22050 Hz and quantized at 16 bits/sample. Subject #15's data were eliminated due to failure in recording. The rest of the subjects' vowel imitation data were analyzed as follows. First, for all vowels, F1 and F2 values were measured at the temporal midpoint of a vowel by using a Praat script. Obtained data were plotted on a F1-F2 plane for each subject for the purpose of inspection for the accuracy of automated formant measurements. Outstanding formant data (i.e. potentially of wrong measurements) were compared with the spectrograms, and the measurements were hand-corrected if confirmed to be errors.[6] After this data verification process, F2 values were transformed to NF2 (talker normalized formant) values as in the production data reported in Chapter 3 and Chapter 4 (see Chapter 3 for the description of this procedure).[7] F1 data were not used for the analysis: these were used only for the initial inspection of the data. The NF2 data were analyzed both at a group level and at an individual level. The raw F2 data were used for the statistical analyses as described below.

To test the hypotheses H1 (the context effect) and H2 (the relevance of perceptual boundary to the repetition performance) statistically, the F2 data for stimuli #4, 5, 6, and 7 were submitted to a hierarchical multiple regression analysis.[8] The reason for using the F2 data instead of the NF2 data was that regression coefficients would become more easily interpretable when they are expressed in Hertz than unit-less numbers. The variable to be predicted was the F2 of the response vowels. Predictors were Context (2 levels), Vowel (4 levels), F2-i and F2-u (each subjects' F2 of [i] and [u], obtained from *heed* and *who'd*), and Boundary (each subject's mean /i/-/u/ category boundary on male2 continua (§4.6.3, Table 4.8)). In the first step of the analysis, only Constant, Context, and Vowel were included in the model; in the second step, F2-i and F2-u were added; and in the last step Boundary was added. The $R^2$ increment was tested at the second and the third step. The results follow.

Figure 5.2 presents group-level results—95% confidence intervals (CI) and mean NF2 values obtained from twenty-nine subjects' responses from each of the ten vowel stimuli presented in /dVt/ and /bVp/ syllables. Three characteristics emerge from these plots. First, the plots show that the stimuli #6 and #7 elicited a considerable amount of context effects on the response vowels. For stimulus #6, in particular, 95% CI of repeated vowels' NF2 exhibited complete separation, with much lower NF2 in the vowels elicited from the /dVt/ stimuli compared with the vowels elicited from the /bVp/ stimuli. The results of the regression analyses show that contexts were significantly associated with F2 (Table 5.1). Response vowels' F2 would be lower for the /dVt/ stimuli than for the /bVp/ stimuli for about 52 Hz. It is very unlikely that the observed context effect is induced by production constraints, since the subjects uttered their vowels in

---

[6] In the production study, where a single F1 and F2 value was needed per word per subject, taking median value from multiple tokens (4-6 tokens per word) ensures that occasional erroneous measurements are discarded (assuming that at least two tokens for each word provide accurate measurements). In this study, however, all tokens were used for analysis and thus automatically obtained formant measurements needed to be double-checked by eyes.

[7] F2 values of the repeated vowels are presented in Appendices E & F (Chapter 5: F2 of the repeated vowels).

[8] Responses for the end stimuli (#1-3 and #8-10) were not analyzed for these hypothesis tests because context effect was expected only for ambiguous stimuli.

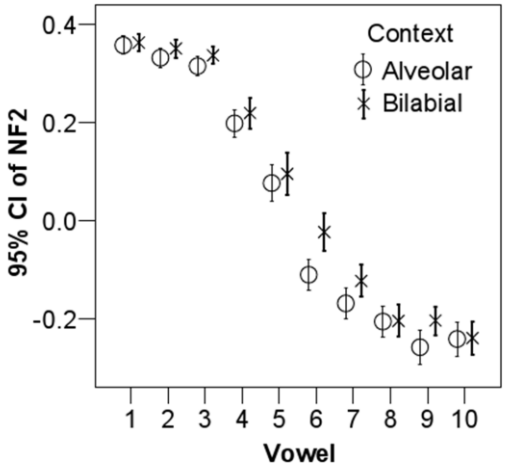



**Figure 5.2** 95% CIs (lines) and means (dots) of repeated vowels' NF2 as a function of stimulus vowels in /dVt/ syllables (Alveolar context) and /bVp/ syllables (Bilabial context). ($N$ = 2320: 10 vowels x 4 tokens x 2 contexts x 29 subjects).

**Table 5.1** Constant and coefficients of the three regression models of F2 (Hz) in the repeated vowels

| Model | | B | Std. Error | Beta | t | Sig. |
|---|---|---|---|---|---|---|
| 1 | (Constant) | 2725.06 | 33.01 | | 82.56 | .000 |
| | Vowel | -149.13 | 4.83 | -.79 | -30.89 | .000 |
| | Context | -52.87 | 27.73 | -.05 | -1.91 | .057 |
| 2 | (Constant) | 908.38 | 100.14 | | 9.07 | .000 |
| | Vowel | -149.13 | 3.74 | -.79 | -39.88 | .000 |
| | Context | -52.87 | 21.48 | -.05 | -2.46 | .014 |
| | F2.i | .61 | .04 | .32 | 14.39 | .000 |
| | F2.u | .19 | .04 | .12 | 5.37 | .000 |
| 3 | (Constant) | 753.64 | 106.87 | | 7.05 | .000 |
| | Vowel | -149.13 | 3.70 | -.79 | -40.35 | .000 |
| | Context | -52.87 | 21.23 | -.05 | -2.49 | .013 |
| | F2.i | .60 | .04 | .31 | 14.30 | .000 |
| | F2.u | .18 | .04 | .11 | 5.00 | .000 |
| | Boundary | 42.39 | 11.05 | .08 | 3.84 | .000 |

*Note*: $R^2$ = .62 for Model 1
$R^2$ = .76 for Model 2; $\Delta R^2$ = .151 for Model 2 ($p$ < .001)
$R^2$ = .78 for Model 3; $\Delta R^2$ = .005 for Model 3 ($p$ < .001)

isolation, after a neutral vowel /ə/. Thus these results are interpreted as a positive support for H1 that perceptual compensation for coarticulation would influence the way vowels are repeated. In the alveolar context, which is a fronting context for back vowels (cf. Chapter 3 & Chapter 4), ambiguous vowel stimuli were perceived to have lower F2 than they actually had, and this perceptual effect in turn influenced the way the stimuli were represented in the subjects' working memory for repetition.

Second, the significant $R^2$ increment at the final step of the regression analyses indicates that the perceptual /CiC/-/CuC/ boundary was a significant predictor for the imitation performance. The estimated effect was about 42 Hz. This means that if subject A's category boundary is one step higher (closer to /u/-end) than that of subject B (who has comparable characteristics except for category boundary), then A's response vowels would have higher F2 than B's response vowels for 42 Hz. This pattern is expected, as subjects who have their boundaries closer to the /CuC/-end (*Backers*, in the sense of chapters 3 & 4) would hear the middle stimuli as members of the /i/-like category than subjects who have their boundaries closer to the /CiC/-end (*Fronters*). This result thus supports H3.

Third, in response to Q1, the pooled data exhibited qualitative differences across vowel stimuli in how categorically and how consistently these stimuli were imitated. Stimuli #1, 2, and 3 were repeated in a more categorical manner and more consistently than any other stimuli: the mean NF2 were very similar for these three vowels and 95% CIs for these three vowels were much narrower than CIs for other vowels. These characteristics were observed in both alveolar and bilabial contexts. These results indicate that most subjects repeated the stimuli #1, 2, and 3 as instances of the same pronunciation category regardless of the context. Stimuli #8, 9, and 10, were also repeated somewhat in a categorical manner, and especially so in the bilabial context. However, their CIs were much wider than CIs for the three end stimuli on the /CiC/-side, and were just as wide as CIs for mid stimuli (#4, 5, 6, and 7). Finally, the mid stimuli were repeated variably across and within the stimuli, as indicated by the markedly different mean values and wide CIs. A question of whether the variable realization of response vowels was due to across subject variation, within subject variation, or a combination of both will be addressed in the following paragraphs with the examination of the individual results.

Figure 5.3 and 5.4 present individual results from the female and male subjects, respectively. In addition to the response vowel's NF2, the plots also show NF2 values of the stimuli, which were the stimulus F2 values transformed into a unique set of NF2 values for each subject. Thus stimulus NF2 were relatively lower for the female subjects (mostly 0.25 or below) than for the male subjects. Note that the range of the y-axis (NF2) varies across subjects. A variable range (rather than a fixed range for all subjects' data) was used so that the categorical trend of the response vowels can be better discerned: A qualitative characteristic of the data will not be appreciated if a uniform range is used for the y-axis because the uniform range is much wider than the individual range and thus data from each subject will be compressed along the y-axis.

These plots reveal considerable individual variation in three respects. First, subjects varied from each other in terms of response /i/-/u/ category boundaries. Response category boundaries were examined only for the data that exhibit categorical responses. Following Chistovich et al. (1966), response boundaries were identified as the stimulus location that elicited the largest variability in response vowels (i.e. point of maximally unreliable repetition). According to this criterion, the category boundary for female subject #27, for example, was at the stimulus #6,

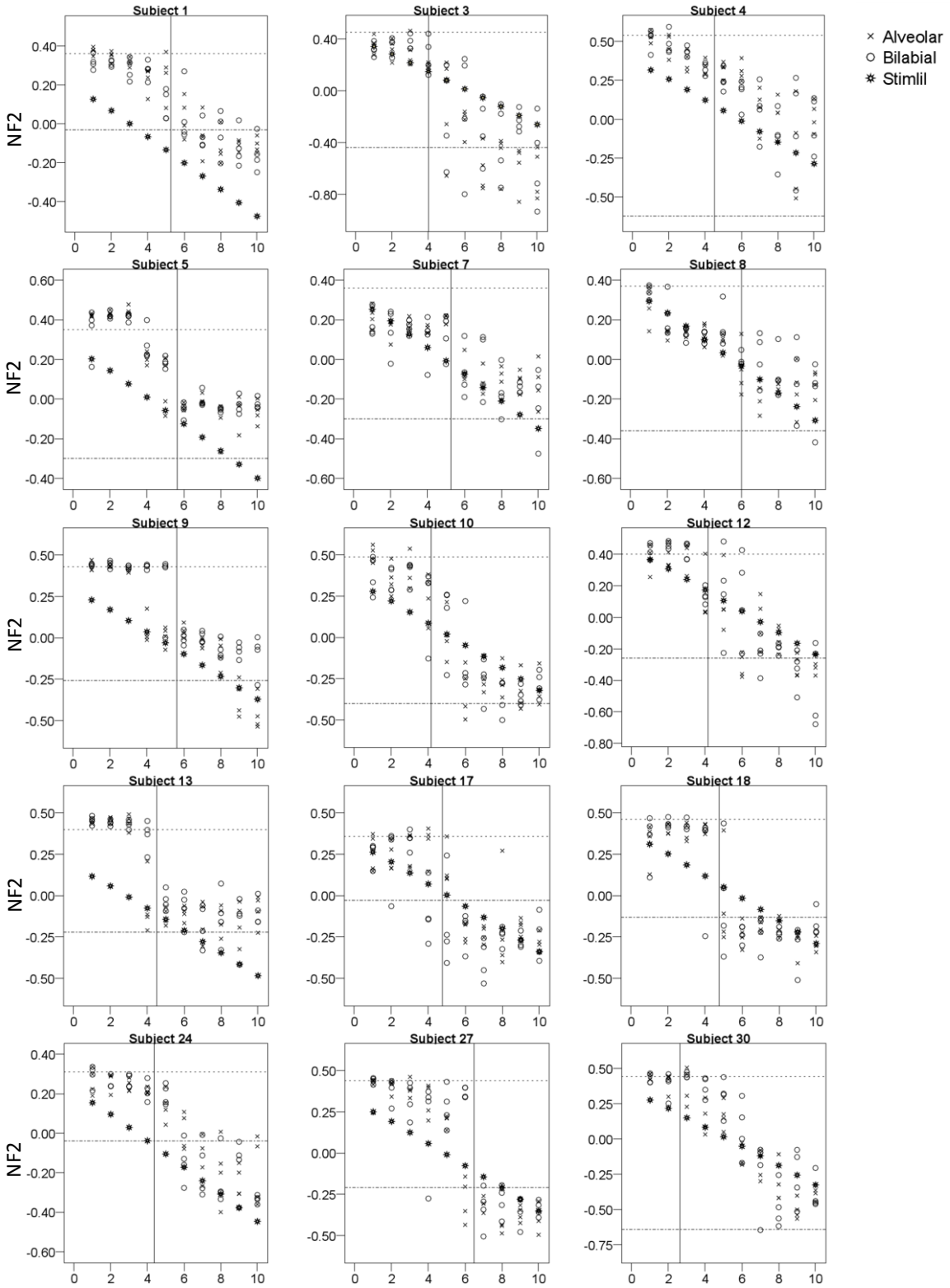



Figure 5.3 Results from fifteen female subjects: NF2 of response vowels for each stimulus vowel (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli. A vertical line indicates Boundary and horizontal lines indicate NF2 in *heed* (upper) and NF2 in *who'd* (lower) for each subject. (See Appendix G for larger figures.)

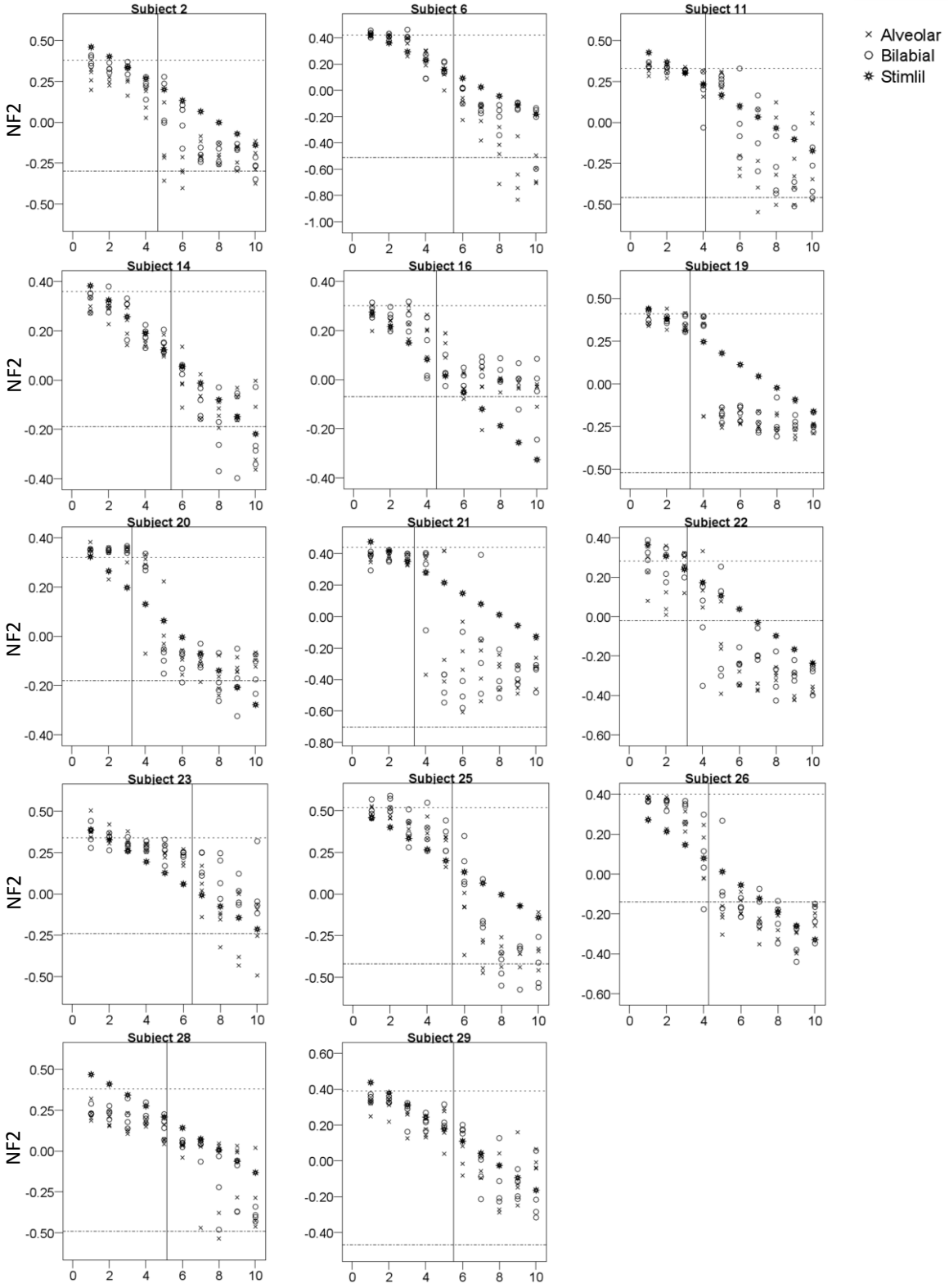



Figure 5.4 Results from fourteen male subjects: NF2 of response vowels for each stimulus vowel (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli. A vertical line indicates Boundary and horizontal lines indicate NF2 in *heed* (upper) and NF2 in *who'd* (lower) for each subject. (See Appendix H for larger figures.)

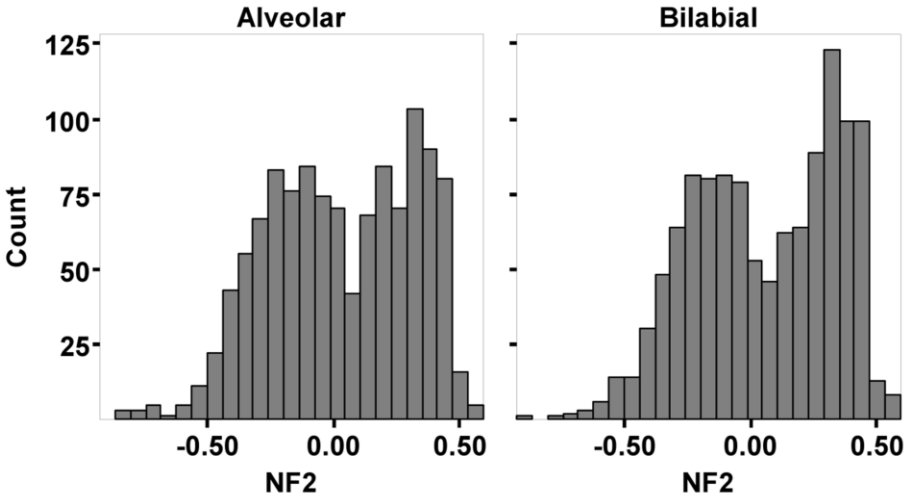



**Figure 5.5** Histograms of NF2 of response vowels by contexts (/dVt/ vs. /bVp/).

because she repeated stimuli #1 to #5 as /i/ all the time except for just once, stimuli #7 to #10 as /u/ all the time, and stimulus #6 variably as either /i/ or /u/. Crucially, her category boundary was not shared by all the subjects. For another female subject #13, the category boundary was at the stimulus #4. For male subjects #19 and #20, the boundary was between stimulus #4 and #5, and for female subject #30, it was between stimulus #6 and #7. These results echoed the results from the perception experiments (Chapter 4): members of a single speech community agree with each other in perceptual category judgments for prototypical sounds, but for ambiguous sounds their judgments vary. Importantly, this variation is not limited to be a matter of degree, where, for example, some subjects perceived a given stimulus as a slightly poor instance of /CuC/ while others perceived it as an extremely poor instance of /CuC/. Subjects varied in their absolute categorical judgments; that is, some perceived stimulus #5 or #6 unambiguously as an instance of /CiC/, while others perceived it unambiguously as an instance of /CuC/. These results indicated systematic perceptual category structure differences (cf. Fox, 1982, pp. 19-20) across subjects, supporting H2.

Second, in response to Q2, subjects also varied in terms of how categorically or continuously they responded to the stimuli. The majority of the subjects repeated the vowel stimuli categorically rather than continuously, reflected in the NF2 plots forming discernible clusters—one cluster near the top-left corner and the other near the bottom-right corner. Female subject #18 and male subject #19's data were the prototypes of this categorical repetition. That categorical repetition was a majority tendency is also clearly demonstrated in the histogram of response vowels' NF2 (Fig. 5.5). If all subjects repeated the stimulus vowels continuously, making equal or nearly equal acoustic distances between the pairs of two adjacent vowels, then the histogram would exhibit a near-flat envelope for most of distributional range. To the contrary, the obtained histograms show bimodal distribution with a dip near the point of NF2 = 0.00, indicating that majority of subjects had tendency to respond to stimulus vowels in

categorical-like manner than in continuous manner. This result is in accordance with the findings from all of the previous studies on imitation of equally distributed speech stimuli: naïve subjects have a categorical bias in their responses (Alibuotila et al., 2007; Chistovich et al., 1966; Kent, 1973, 1974, 1979; Repp & Williams, 1985, 1987; Vallabha & Tuller, 2004).

Few subjects were capable of repeating vowels in more continuous-like manner than others. This gradient response pattern was more frequently observed from male subjects than from female subjects: while only one female subject (#4) exhibited continuous response, several male subjects (#2, #6, #14, #23, #25, and #29) produced gradient response patterns for an entire range or a part of the stimulus continua. This gender asymmetry is probably caused by the fact that the stimuli were modeled after male speech. Gradient response entails that stimuli were repeated accurately, which requires a match between a subject's natural production range and the range of variation of the stimuli. Although a match itself does not guarantee that a listener can repeat speech sounds accurately, as listeners tend to have bias toward categorical response even when imitating self-produced speech (Repp & Williams, 1987; Vallabha & Tuller, 2004), the mismatch seems to cause additional difficulty not only in acoustically faithful repetition of stimuli but also in reproducing a gradually varying pattern of speech stimuli.

Third, in response to Q3, one subject (female, #5) made responses around three distinct pronunciation centers. Her vowel data formed one cluster around NF2 = 0.4, another cluster around NF2 = 0.2, and another large cluster covering the NF2 value of 0.1 and below. Presumably, the clusters that have higher and lower NF2 values represent this subject's /i/ and /u/ categories. Her responses to some of the stimuli #4 and #5 belong to neither of these two categories. That her responses to these stimuli formed an identifiable cluster suggests that she had the third distinct pronunciation category between the /i/ and /u/ regions. The source of this third category could be her familiarity to French (revealed from a pre-test language background questionnaire) that has three phonemic high vowels—/i/, /u/, and a front rounded vowel /y/. This result is thus in accord with Schouten's (1977) study, in which his Dutch-English bilingual subjects exhibited categorical response patterns using vowel categories from both languages.

In addition to finding effects of the compensation for coarticulation and a category boundary on repeated vowels as well as the systematic individual variation in vowel repetition patterns, the present study also found a difference across stimuli in how variably these stimuli were repeated. In order to explore this variability difference further, the range of NF2 in the repeated vowels was calculated as a difference between the maximum and the minimum of the NF2 values from the four repetitions elicited from the ten vowel stimuli, separately for the /d_t/ and /b_p/ contexts and for each subject. Figure 5.6 presents summary data—95% CIs and the means of the NF2 range for each vowel stimulus in the two contexts. These plots confirmed what other plots already revealed: stimuli #4-10 were responded to with much more variation than stimuli #1-3. This pattern both converges to and diverges from previous findings. On the one hand, previous studies found that stimuli that fall on high-mid to high back vowel region tend to be imitated more variably than stimuli that fall on high-front region (Kent, 1973; Repp & Williams, 1985, 1987). The present study's results converged to these findings. On the other hand, these previous studies found that stimuli that approximated prototypical /u/ elicited relatively more stable response patterns than the stimuli closer to the high-mid regions. The present study's results diverged from these findings, as the present data exhibited variable repetition for the /CuC/-end of stimuli. Since the variability in the response vowels reflects consistency in

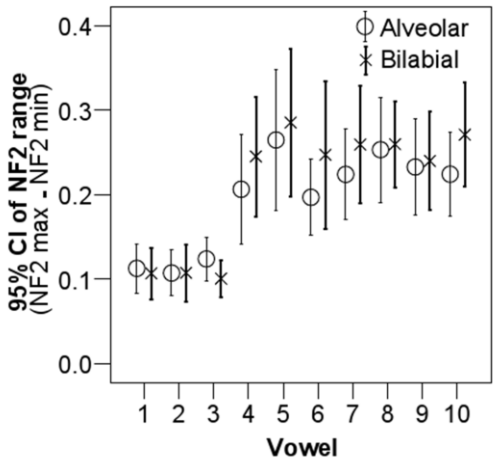



**Figure 5.6** 95% CIs (lines) and means (dots) of NF2 Range (= NF2 max - NF2 min) for each vowel stimulus in each context. (*N* = 580: 10 vowels x 2 contexts x 29 subjects).

repetition performance, it is understandable that there was a greater consistency in repeating familiar stimuli than unfamiliar ones (Kent, 1973, p. 7). Therefore, the present results are contrary to the results from previous studies and to intuition. A possible account for this result will be offered in the following section. The same plots also indicated that stimuli #6 and #7 elicited less variable repetition in alveolar context (and only in alveolar context). Interestingly, these are the stimuli that induced a significant amount of compensation for coarticulation: stimuli #6 and #7 in the /dVt/ syllable elicited response vowels with significantly lower NF2 than the same vowel stimuli in the /bVp/ syllable. In addition, these /dVt/ stimuli were responded to in a more categorical manner than their /bVp/ counterparts. Together, these results suggest that perceptual compensation results in more categorical and more stable representation of the speech inputs, which is reflected in categorical and more consistent repetition performances.

## 5.6 Summary and Discussion

The present study used a vowel repetition task as a way to examine how listeners encode in their working memory the continuous vowel stimuli and how these representations vary in different consonantal contexts and across listeners. The results are summarized as follows:

1) Phonetic contexts in which the test vowels occur influence the perceptual categorization and the subsequent repetition of ambiguous vowel stimuli; that is, the compensation for coarticulation influenced the repetition of ambiguous stimuli.

2) A portion of the variation in vowel repetition performance is attributable to the individual variation in perceptual /CiC/-/CuC/ category boundary.
3) Subjects also varied in terms of the number of perception categories they have in a given stimulus continuum.
4) Ambiguous /dVt/ stimuli that elicited significant compensation were responded to with more categorization and consistency than the comparable /bVp/ stimuli.
5) Stimuli #1, 2, 3 (/CiC/-side) and #8, 9, 10 (/CuC/-side) were repeated more categorically than the middle stimuli.
6) While the /i/-side of stimuli elicited consistent responses, the rest of the stimuli elicited variable responses.
7) Subjects varied in how categorically they responded to the stimuli.
8) Male subjects responded to stimuli more continuously than female subjects.

The most important results for the purpose of the present study were the first three. These results, in combination with the results from the Perception experiment (Chapter 4) suggest that within a given speech community there are stable and unstable acoustic-auditory regions for phonemes. Knowledge on the stable acoustic-auditory regions, reflected by 100% or near 100% agreement in subjects' perceptual judgments, unites the listeners as members of a single speech community. Presumably, these individuals are capable of both producing sounds from these stable regions in careful speech and understanding the speaker intent correctly when listening to another speaker producing sounds from the same regions. However, when these members encounter an extreme pronunciation variant such as the extremely fronted variant of /u/ in the alveolar context, each individual may encode the vowel in different forms, depending on their context-specific perceptual category boundary for /u/, and this would be the case even if all listeners employed perceptual compensation for coarticulatory fronting. The present study thus supports Beddor's (2009) model for the variation in perceptual grammar (§2.1.5). As claimed in the Introduction of Chapter 4, compensation does not guarantee invariance in speech perception, because individuals vary in their perceptual grammar.

In addition, the obtained results suggest that various factors influence the way incoming speech sounds are represented in the short-term memory. These factors include: the phonetic context, listener's own internal category boundaries, similarity to category prototypes, and the match between the acoustic properties of stimuli and the listener's own production range. The results also suggest that the nature of mental representations vary in terms of categoricality and consistency: some inputs are represented more categorically and/or consistently than others. In this section a possible mechanism through which various factors influence perceptual categorization and representation will be considered. Then the implications of the findings for the theories of speech perception and the theory of sound change will be discussed.

One of the major and converging findings from speech perception and vowel imitation studies alike, including the present study, is that speech stimuli can be perceived (and presumably represented) either categorically or continuously (see Repp, 1984 for review). Categorical (vs. continuous) perception is typically tested by using stimulus classification tasks

and discrimination tasks, and studies using these paradigms have found various factors that affect these two modes of perception.

First is the class of target sound: consonants are perceived in a categorical-like mode (Liberman, Harris, Hoffman & Griffith, 1957; Liberman, Harris, Kinney & Lane, 1961; Mattingly, Liberman, Syrdal & Halwes, 1971; Pisoni, 1973a), but vowels are perceived in a continuous-like mode (Fry, Abramson, Eimas & Liberman, 1962; Pisoni, 1973a; Repp, 1981). Second is stimuli naturalness: more naturally sounding stimuli are perceived more categorically than less naturally sounding stimuli (Schouten & Van Hessen, 1992). Third is the experimental task: the ABX task elicits more categorical responses than the 4IAX task (Pisoni 1973b) or the AX task (Crowder, 1982).[9] Finally, when the discrimination task is employed, longer ISI (inter-stimulus interval—time interval between or among the stimuli) tends to elicit more categorical responses than shorter ISI (Cowan & Morse, 1986; Pisoni, 1973a,b; Van Hessen & Shouten, 1992). These findings have invited various explanations of the difference between categorical and continuous modes of perception.

Fujisaki and Kawashima (1969, 1970, as cited in Pisoni, 1973a,b; see also Fujisaki, 1979, for an updated version of the model) made a pioneering proposal on the mechanism underlying the different modes of perception employed in discrimination tasks. Their model assumed that during discrimination tasks listeners use both acoustic, continuous auditory information and abstract, categorical phonetic information, and attributed the differences in perceptual response patterns to the differences in the relative use of each of the two types of information.[10] This hypothesis was supported by Pisoni (1973b), which showed that vowel discrimination can be made more or less continuously when auditory information is made more or less readily available (i.e. 4IAX task vs. ABX task). According to this model, categorical perception occurs due to a failure of retrieval or the loss of auditory information due to decay in memory tasks (Pisoni, 1973b, p. 115).

In a more general term, the hypothesis that listeners use both auditory information and categorical linguistic information in speech perception and speech repetition tasks has been supported by previous studies (Pisoni & Tash, 1974; Yoneyama, 2007). For example, Yoneyama showed that neighborhood density measures calculated based on segmental representation and on auditory representation were both significant predictors for Japanese speakers' word-naming latency (the time for reproducing the perceived spoken inputs) Yoneyama's model of word-naming allows acoustic input to directly map onto phonological representation or be mediated by auditory representations: either case, phonological representations serve as the basis of articulation, in accord with the previous models (Jusczyk, 1993; Plaut & Kello, 1999). However, relative contribution of the two types of information may vary, and it is possible that one type of information dominates the word-repetition performance (Mitterer & Ernestus, 2008).

---

[9] In ABX paradigm listeners hear two different references (A&B) and then determine whether the following third sound (X) matches to A or to B. In 4IAX paradigm listeners hear two sets of paired stimuli (thus 4 intervals) and determine which of the two paired stimuli contains difference. In AX paradigm listeners hear two stimuli and determine whether the two stimuli are same or different.

[10] In terms of the categorization model presented in Figure 5.1, this hypothesis can be implemented as two layers of representations at the "Category" level, one for phonemic category and the other for phonetic category, and the final articulatory instructions reflect the output of both types of resolution.

The hypothesis about the differential contribution of continuous auditory and categorical linguistic information as function of availability of acoustic information accounts for two factors that elicited categorical-like responses in the present study. One was the gender difference, where female subjects responded more categorically than male subjects. A possible explanation is that the female subjects had difficulties in retrieving accurate acoustic information from the stimuli modeled from male speech and resulted in categorical-like responses. Another was the effect of the compensation for coarticulation on the categorical-like responses. A possible explanation for this factor is that compensatory perception inherently involves loss of original acoustic information and thus induces categorical-like responses. In a broad sense, talker normalization and compensation for coarticulation are the similar process whereby speech variation is reduced and input signals are transferred so as to more closely approximate to familiar forms.

These observations lead to two related hypotheses about a mechanism for categorical-like responses. One is that given the speech mode of perception, the more unfamiliar the stimuli are for the perceivers, the more likely that the stimuli are represented in categorized forms than acoustically faithful forms. Another is that phonological categories provide common currencies for the members of the speech community, and also bridge speech perception and articulatory instructions that are used to reproduce spoken inputs. If all members share the same vocal tracts and produce an identical range of speech sounds, then listeners can encode others' speech faithfully, the long-term mental representations of speech sounds being very similar to the short-term representations of acoustic forms. In reality, however, speakers' utterances vary so much that some speaker's utterances are simply not producible by others. A listener transforms these acoustic patterns into more useful forms such as phonemes. This hypothesis is compatible with recent findings in a shadowing task that the subjects tended to more faithfully imitate phonologically relevant contrasts than phonologically irrelevant phonetic details (Mitterer & Ernestus, 2008). For native speakers of Dutch, for example, phonological two-way contrast in initial stops is realized by presence or absence of pre-voicing (van Alphen & Smits, 2004) but exact amount of pre-voicing is irrelevant for the contrast (van Alphen & McQueen, 2006), and also both alveolar and uvular variants of /r/ are familiar non-contrastive variants of /r/ (Mitterer & Ernestus, 2008). Shadowing performance of the Dutch subjects in Mitterer and Ernestus (2008) study was not affected by whether or not their own habitual production of /r/ (alveolar vs. uvular) matched or mis-matched with the production of stimulus. In addition, their responses had significantly different amount of reaction time and VOT depending on whether the stimuli had pre-aspiration or not, but their responses were comparable for the stimuli having different amount of pre-voicing.

Further, as a flip side of the amount-of-auditory-information account, one might hypothesize that clarity of the category identity itself may influence the retrieval of acoustic information. That is, there would be a negative correlation between the strength of the evoked category identity and the amount of acoustic information extracted from the speech stimuli. This hypothesis is based on an assumption that listeners divert just enough cognitive resources to complete the task at hand: the moment a categorical identity of a stimulus becomes obvious, listeners stop paying perceptual attention to resolve acoustic properties, because such attention and resolution is not called for in a normal communication situation and listeners have been habituated not to do so. In this scenario, then, categorical perception is still a result of the failure

of the retrieval of auditory information, but this failure does not arise due to the difficulty in extracting or holding information as in the case of the ABX task, but due to the lack of attention. Whether listeners reduce attention to the acoustic details when hearing prototypical sounds or not waits for future testing, but there is a body of evidence that the sounds near category prototypes are hard to discriminate from each other due to perceptual warping (Guenther & Gjaja, 1996; Iverson & Kuhl, 1995; Kuhl, 1991). In sum, I propose three hypotheses regarding how listeners fail to represent stimuli continuously: one is the difficulty to retrieve/hold information in their memory, another is the failure to retrieve information due to the lack of attention, and yet another is perceptual warping. All these mechanisms can lead to the same outcome—a bias toward categorical perception (and representation) than a continuous one. These hypotheses nicely account for the categorical imitation of the end stimuli (#1, 2, 3 and #8, 9, 10) in the present study. Needless to say, these hypotheses are rather speculative at this moment, and they need to be directly tested in future studies. In particular, it would be of interest to know if listeners can detect extreme pronunciation variants better from the same-sex speakers than from opposite-sex speakers. Such a result, if obtained, would have significant implications for the sociolinguistic theory of sound change: an innovative pronunciation would spread to the same-sex speakers first.

On the issue of variability in the response vowels, this study found more variable repetition for the /CuC/-side of stimuli than the /CiC/-side of stimuli, and this pattern was found equally frequently from both female and male subjects. It is important to note that this variability was within-subject variation. According to an assumed categorization model (Figure 5.1), this result could arise either from perception variation or production variation. In perception, an observed difference means either that the /CuC/-like stimuli are mapped onto more variable acoustic-auditory patterns (greater variance for As in Figure 5.1) and/or there are multiple perceptual categories for the sound that listeners broadly perceive as /CuC/ (good /CuC/, poor /CuC/, etc.) than /CiC/. In production, an observed difference means that there are greater errors in motor control (i.e. greater variance for Rs in Figure 5.1) for /u/ than /i/. The first scenario that the acoustic signal for /CuC/ is more variably perceived than the /CiC/ signal is supported by results from previous perception studies (Fox, 1982; Peterson & Barney, 1952) and the present study. In Peterson and Barney, for example, 70 adult listeners classified 1520 vowel tokens (produced by 76 different speakers, each producing 2 tokens of 10 English vowels) into 10 vowels categories. Their results showed that out of 152 [i] tokens 143 (94%) were unanimously classified by all listeners as [i], while the number of tokens unanimously classified by all listeners as [u] decreased to 109 (72%). This discrepancy in listener agreement rates cannot be attributed to the different production variation between the two vowels, as Peterson and Barney's speakers produced [i] and [u] with comparable degrees of speaker variation.

The support from the present study is imitation responses to the middle stimuli, which showed more variable responses for the /bVp/ stimuli than the /dVt/ stimuli. Take, for example, subjects' responses for the stimulus #6 in Figure 5.2. In the /dvt/ context these stimulus vowels were repeated as comparable vowels elicited from the stimulus #7 in the /bVp/ context. Yet these two vowels that are presumably executed by the comparable set of motor instructions have a markedly different variability (Fig. 5.6). The variable degree of motor errors might indeed be responsible for a portion of the variability differences in the response vowels, but a large portion of the differences should probably be explained by the perception factor.

The observed differences in the response variability between the /CiC/-like and the /CuC/-like stimuli as well as between the middle stimuli in the alveolar and bilabial context is noteworthy because the directionality of this difference is in the same direction as the difference in the range of production variation between /i/ and /u/ as well as /u/ in fronting and non-fronting contexts as reported in Chapter 3. As shown in figure 3.10, the back vowel /u/ had a wider range of phonetic realizations than /i/, and within the /u/ category the vowel had a wider range of realization in the non-fronting context than the fronting context. Further, we found in Chapter 4 that the /CiC/-/CuC/ category boundary was more variable across listeners in the non-fronting context than in the fronting context. Here, we observe the parallel among the production, perception, and repetition patterns. Together, these results suggest that the difference in the repetition variability is better attributed to the variability in acoustic-to-auditory mappings than in motor errors. As discussed in Chapter 4 there is a large body of evidence that speech perception is shaped by linguistic experience (see §4.2.2). The subjects in this study have been exposed to a wide range of acoustic patterns for the high back vowel /u/ due to an on-going fronting sound change in California and other varieties of American English. The vowel repetition data provided yet another piece of evidence that speech perception, and perceptual categorization in particular, develops in response to the structures of ambient language data.

# Chapter 6

# Findings, Conclusions, and Implications

Motivated by the importance of the question "what causes pronunciation variations and how do language users produce, perceive, and learn these variable forms?" for understanding the mechanism of sound change, this dissertation investigated, by using coarticulatory /u/-fronting in the alveolar context for a case study, how native speakers of American English produce coarticulatory variations and how, as listeners, these speakers perceive and reproduce continuously varying speech sounds that are heard in coarticulatory and non-coarticulatory contexts. The specific questions addressed in this investigation were: (1) Does a phonetically motivated coarticulatory variant have a distinct articulatory goal?; (2) How do listeners vary in their perceptual interpretation of vowels in coarticulatory and non-coarticulatory contexts?; (3) Is perceptual compensation linked to the structure of speech sounds in one's native language?; (4) How does speech perception guide the reproduction of spoken inputs?; and (5) To what extent can listeners discern and encode sub-phonemic variation? The answers to these questions offer a crucial key to solving the mystery in the trigger cause of sound change: What variations are possible for phonologization and social indexing?

## 6.1 Summary of the Study

The production study reported in Chapter 3 addressed the question of whether in American English coarticulatory fronting of /u/ in alveolar contexts is an inevitable consequence of production constraints or if it is produced by deliberate speaker control. This study found two kinds of evidence for a context-specific articulatory target for the fronted /u/. First, the relative acoustic difference between the fronted /u/ and the non-fronted /u/ remained across elicited ranges of vowel duration. Second, the degree of acoustic variability was less for the fronted /u/ than the non-fronted /u/. These results were compatible with a gestural model within articulatory phonology (Browman & Goldstein, 1986, 1990, 1992). Nonetheless, following experience-based and exemplar-based theories of phonological grammar (Bybee, 2001, 2006; Hale, 2003; Goldinger, 1996, 1998; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002, 2003, 2006; Wedel, 2006), the obtained results were interpreted as evidence that speakers of American English have

a distinct and more narrowly specified articulatory target for the fronted /u/ in the alveolar context apart from the target of the non-fronted /u/.

One implication of these results for a theory of coarticulation is that, although "coarticulation is a universal characteristic of human speech production" (Hardcastle & Hewlett, 1999, p. 3), implementation and characteristics of coarticulation vary case by case. The term coarticulation presupposes that at some level there be invariant unit underlying the variable speech output (Kühnert & Nolan, 1999, p. 7) and this notion is reflected in Lidblom's (1963) target undershoot model. But many cases, as exemplified in the present study, show that the extent of coarticulation is greater than what is expected from the undershoot model: a more active process needs to be considered to explain these cases (Kühnert & Nolan, 1999, p. 16). In some cases, articulatory instruction for a particular part of the tongue causes resistance to coarticulation (Öhman, 1966), or certain segments are inherent more resistant than others (Recasens, 1984). Or, any type of coarticulation may be adequately modeled in terms of gestural overlap (Browman & Goldstein, 1992; Fowler & Saltzman, 1993). My proposal is that not all coarticulations are equal, and each type of coarticulation needs to be explained and modeled separately. In addition, I propose that certain coarticulatory variants that have distinct perceptual-auditory properties are likely to be treated by speakers as distinct pronunciation categories.

The perception study in Chapter 4 consisted of the replication of previous findings and testing some new hypotheses. The study replicated previous findings of perceptual compensation for coarticulatory /u/-fronting in alveolar contexts (Harrington et al., 2008; Lindblom & Studdert-Kennedy, 1967; Ohala & Feder, 1994) as well as the effect of speech rate on the amount of compensation (Lindblom & Studdert-Kennedy, 1967). The study did not replicate the effect of the assumed phonemic identity of contexts (vs. the acoustic information of the contexts) in inducing compensatory perception as found previously (Ohala & Feder, 1994). It was speculated that the vowel repetition task that used the same stimuli and preceded the perception task had trained the listeners to dissociate the coarticulatory source from the target, causing the failure to elicit compensatory perception.

This study found positive evidence for systematic individual variation (i.e. within-listener consistency) in the classification of /CVC/ syllables both when the Cs formed fronting and non-fronting contexts for the vowel. The study also investigated the correlation between the degree of perceptual compensation for /u/-fronting and the degree of /u/-fronting in production, but found no evidence for this perception-production link. The study also addressed the issue of the relationship between linguistic experience and speech perception, and found one piece of positive evidence—the similarity between the distributional characteristics of the fronted and the non-fronted variants of /u/ in production data (, which was considered as a model of ambient language data) and the ranges of variation in perceptual responses toward /CVC/ stimuli in the fronting and the non-fronting contexts. Findings of the systematic individual variation in speech perception and of the positive relationship between the perceptual response patterns and structures of ambient language data constituted a micro-level counterpart of Harrington et al. (2008), which found a systematic difference between younger British listeners and older British listeners in their context-specific phoneme category boundaries. Together, these results were interpreted to suggest that the source of individual variation in speech perception is individual differences in phonological grammar (perceptual category boundary), which emerge in response to the ambient language data that the community members produce and that the listener has been

exposed to from day-to-day language use. In this regard, the present study supports the usage-based and exemplar-based model of phonological knowledge (Bybee, 2001, 2006; Hale, 2003; Goldinger, 1996, 1998; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002, 2003, 2006; Wedel, 2006), and extends Beddor's (2009) findings about individual variation in the phonological grammar to the case of /u/-fronting in the alveolar context.

One implication for theories of sound change is the importance of study on the mental representation of sub-phonemic variation. Studies on dialect, accent, and idiolect adaptation have shown that listeners are capable of adapting to various types of pronunciation variation (see Samuel & Kraljic, 2009 for a review). While "dialect borrowing" (Bloomfield, 1933, p. 444) is separated from regular sound change by Neogrammarians, even an Initial Change by articulatory drift needs to be adopted by new speakers to become a community-level sound change. The only theoretical distinction between "borrowing" and "change" seems to be whether a new listener maintains two distinct representations (borrowing) or one (change). Studies aiming to reveal how language users handle sub-phonemic variation provide valuable insight into the mechanism of sound change.

Finally, the vowel repetition study reported in Chapter 5 tested (1) the effect of compensation for coarticulation on the repeated vowels when the target vowels were heard in the fronting and the non-fronting contexts, (2) the systematic individual variation in categorically repeated vowel's phonemic identities, and (3) the effect of the perceptual category judgment of the stimuli on the acoustic quality of the repeated vowels. Ambiguous vowels were repeated with a significantly lower F2 when the vowels were heard in the fronting context than in the non-fronting context. A given stimulus was repeated by some listeners un-ambiguously as the vowel belonging to the speaker's /i/ category, for all trials, regardless of the contexts, yet the same stimulus was repeated by other listeners un-ambiguously as vowels belonging to that speaker's /u/ category for all trials. A regression analysis determined that the perceptual category boundary was a significant predictor for the repeated vowel's F2 value. Based on these results, it was hypothesized that one source of pronunciation variation in a given community is an individual variation in speech perception that contributes variable mental representations across listeners when they encounter ambiguous speech.

In addition to obtaining these results, the vowel repetition study also found some general tendencies in the vowel repetition performance of the subjects. First, the vowels were repeated rather categorically than continuously, as shown in the previous studies (Alibuotila et al., 2007; Chistovich et al., 1966; Kent, 1973, 1974, 1979; Repp & Williams, 1985, 1987; Vallabha & Tuller, 2004). Also, the vowels that approximated the typical /i/ and the typical /u/ were repeated more categorically than ambiguous vowels. The /i/-like vowels were repeated more consistently than other vowel stimuli. The vowel stimuli that elicited a greater amount of compensation were repeated more categorically than other vowels. Male subjects repeated vowel stimuli (modeled after male speech) more continuously than female subjects. These findings are in accord with the hypothesis that speech stimuli could be perceived (and presumably represented) either categorically or continuously depending on the availability (or use) of acoustic, continuous auditory information and abstract, categorical phonetic information (Fujisaki & Kawashima, 1969, 1970 (cited in Pisoni, 1973a, b); Pisoni, 1973b; Fujisaki, 1979). Finally, one subject's data exhibited a clearly discernible third cluster between the cluster of /i/-like vowels and another cluster of /u/-like vowels. This result and some other subjects' results

that showed continuous repetition patterns were interpreted as evidence that at least some listeners are capable of representing multiple sub-phonemic variants as their own pronunciation repertoire.

One general pattern that was found from all three experiments was that responses to /i/ were less variable than responses to /u/. Thus /i/ exhibited less variability than /u/ in a production task, and /i/-like stimuli were repeated less variably than /u/-like stimuli in a vowel repetition task. Similarly, between /u/ in fronting and non-fronting contexts, /u/ elicited less variability in the fronting context than in the non-fronting context consistently in the production, perception, and vowel repetition tasks. These findings suggest that native speakers of American English treat contextually defined front and back variants of /u/ as distinct as separate phonemes.

My contention is that language users are sensitive enough to notice, attend, and register multiple degrees of contextual variants as well as stylistic variants as distinct articulatory and/or auditory patterns in their memory. This sensitivity to variations and ability to register them as distinct patterns are the contents of linguistic knowledge that enables speakers to employ style shift depending on communicative contexts and interlocutors. I suggest that the same competence enable language users to index particular variants with particular social value. When only a small numbers of speaker adopts these variants, then their speech remain as socio-linguistic variation. When sufficiently large numbers of speakers adopt them, then it will become sound change. The production study reported in chapter 3 reported that the word 'dude' is realized by the young speakers variably as [dyd], [djud], etc. Any of these pronunciation variants might receive positive social evaluation and become a community-level sound change in the future.

## 6.2 Proposed Model of Speech Perception

Based on the previous models of speech perception—the hypo-correction model (Ohala, 1981, 1989, 1993), the variation-selection model (Lindblom, et al., 1995), the perceptual grammar model (Beddor, 2009), and the CCC model (Blevins, 2004)—and the results obtained from the three experiments, this section proposes a model of speech perception that accounts for Initial Change as arising from individual variation in perceptual category boundary (Fig. 6.1).

The proposed model is characterized by flexibly updatable multiple layers of phonological and phonetic representations in the long-term memory; these layers are for, minimally, phonetic representations, pronunciation category representations, and lexical phoneme representations[1]. The motivation to have an intermediate layer of representations under lexical phoneme representation is twofold. First, the results from the perception experiment showing qualitatively different response patterns in the fronting and non-fronting contexts (i.e. greater response variability in the non-fronting context than in the fronting context) as well as the results from the

[1] These three types of representations might be also conceptualized as *phonemes*, *phonetic categories*, and *phones*, but the use of term *pronunciation category* over *phonetic category* highlights distinct cognitive status of this representation than what the term *phonetic* might suggest. Also, the term *phonetic representation* over *phones* highlights the fact that any mental representations are the results of perceptual processing and therefore distinct from raw acoustic patterns.

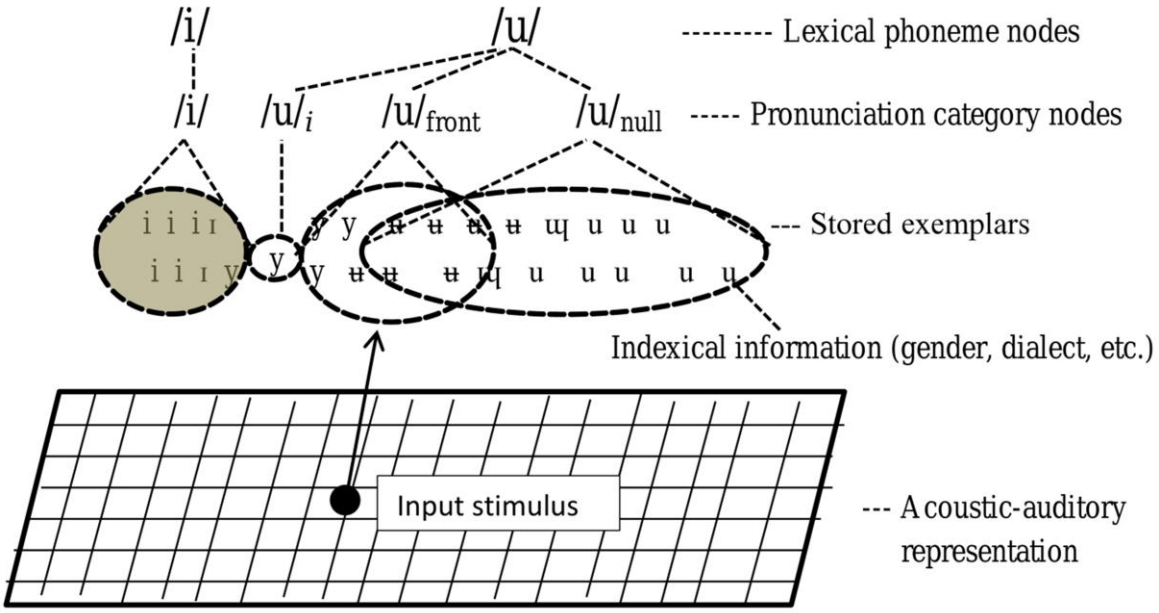



Figure 6.1 Schematic representations of acoustic-auditory-to-phoneme mappings in hypothesized layered representations for a model listener of American English. The upper three layers are represented in the long-term memory. The acoustic-auditory representation is available in the short-term memory. The top-most symbols /i/ and /u/ are the lexical phoneme representations. These representations encompass pronunciation category representations such as /i/ (in the second layer). The symbol /u/$_{front}$ is for /u/ in fronting contexts; and the symbol /u/$_{null}$ is for /u/ in non-fronting (null) contexts. The symbol /u/$_{i}$ is for a particular speaker's idiosyncratic and extremely fronted pronunciation that has been captured and represented by this model listener. A speech input is evaluated in terms of pronunciation category for usual word recognition purpose but can be also evaluated in terms of lexical phonemic identity if the task calls for. The number of pronunciation category and its boundary is listener-specific.

production experiment strongly suggest distinct articulatory and auditory representations for contextual allophones of /u/. Second, studies on perceptual learning of accented speech have shown that listeners are capable of learning new accents and apply this new knowledge to new speaker's speech and new words (see Samuel and Kraljic, 2009 for a review). These perceptual learning studies suggest that phonological knowledge includes knowledge of acceptable subphonemic variants at abstract level. In the proposed model, lexical phonemes are not specified in terms of a single acoustic pattern or gestural configuration, but are defined in terms of a collection of distinct pronunciation categories that do not form a lexical contrast but are learned by a listener as distinct, discernible auditory and articulatory patterns. This knowledge of subphonemic yet categorical pronunciation variations enables listeners, for example, to learn regional and social accents.

The pronunciation categories are defined in terms of a center (i.e. category prototype) and a spread (i.e. category boundary) of phonetic representations. A number of lexical phonemes and pronunciation categories are stable but alterable—more so in the pronunciation category level than in the lexical phoneme level.[2] In the phonetic level of representation, the center and category boundary are updated each time a new speech is experienced and categorized as a member of the same pronunciation category. The model holds individual phonetic representations of recently encountered speech sounds, but these individual representations will gradually fade (unless it is originally encoded with a particular salience) leaving their effects only in the center and category boundary.

A conjecture that mental representations of phonemes map onto ranges of acoustic-auditory patterns rather than existing as single abstract entities is also shared by Beddor's (2009) perception grammar model, and is compatible with Blevins' (1994) CHOICE model and the works of Miller and her colleagues (e.g., Volaitis & Miller 1992; Miller 2001). A conjecture that sub-phonemic units (the pronunciation category in the present model) are represented as stable and distinct categories, rather than just raw exemplars, has been proposed in previous works on speech perception and mental representations (Beckman & Pierrehumbert, 2003; Chistovich et al., 1966; Sumner & Samuel, 2009). Following Johnson (1997), this model also assumes that relevant ranges of phonetic representations and even a single representation can be indexed with salient socio-linguistic information of a talker (e.g., gender, dialect, etc.) as well as salient situational information.

A crucial property of the model is that the number of the pronunciation categories and the range of phonetic representations covered by each pronunciation category are determined by what the language user has previously experienced and classified as a particular pronunciation category member. This property has two important consequences for a theory of sound change. First, this variation in phonological grammar causes individual variation in perceptual phoneme judgments across listeners even in a single speech community. Second, these pronunciation categories are updatable, allowing listeners to adopt both a unique pronunciation of a particular speaker with whom the listener has frequent contact and a more stable pronunciation variation characteristic to a particular dialect or social group. In the case of /u/, a listener may learn, in addition to a pronunciation for a canonical /u/, contextual allophones as well as heavily coarticulated contextual allophones as distinct and separate pronunciation objects.

This model is a straightforward representation of the previous findings on the variation in speech perception that were discussed in Section 4.2. Speech perception is guided by previous linguistic experience, and no two individuals have exactly the same linguistic experience; therefore, it follows that no two individuals exhibit exactly the same perceptual responses to every kind of speech signal. In most communication situations within a single speech community, however, community members must arrive at the same interpretations of speech inputs because there is a substantial range of acoustic-auditory patterns on which the perceptual responses of community members converge. Only in the rare cases in which listeners encounter extreme variations falling near category boundaries does the model predict that listeners vary in their perceptual judgments of the phonemic identities of speech inputs.

---

[2] One might question the validity of this claim: the fact that structure changing sound changes (i.e. mergers and splits) occur indicates that even a number of phonemes needs to be alterable.

The proposed model is also a straightforward extension of the previously proposed models of listener misperception. It is guided by Ohala's (1981) hypo-correction model for its basic principles that (1) perceptual phonemic identity judgment takes coarticulatory variation into account, and (2) language users learn phonological representations of words primarily through perceptual analyses of the spoken word forms. The model is also guided by the concept of perceptual variation proposed by Lindblom et al. (1995), Blevins (2004) and Beddor (2009), which states that variable perception occurs without failure in the perceptual processing system. Finally, this proposal is based on Beddor's (2009) claim that individual variation in perceptual analysis of coarticulatory variation results in variation in phonological grammar. However, my model treats individual variation in perceptual grammar as both the source and the outcome of the variable perceptual interpretation, in contrast to Beddor's model in which individual variation in grammar is only an outcome.

## 6.3 Speech Chain as an Interactive System

The results obtained from the three experiments in the present study suggest, collectively, that speech is a dynamic interactive system, wherein individual language users and the rest of the community members mutually influence each other's pronunciation grammar, perception and production through multi-layered feedback loops. A proposed model of the speech chain is presented in Figure 6.2. The model is characterized by three levels of mutual dependencies. The first is between an individual language user and the speech community. Each speaker develops his or her own unique pronunciation grammars mainly in response to the speech data sampled from the speech community but when they speak their speech will be added to the pool of the community speech data for someone else to sample. The second is between speech perception and the pronunciation grammar. The pronunciation grammar develops as a generalization over previously perceived and classified speech data but each time a listener classifies an incoming speech token. This task is achieved in reference to the pronunciation grammar that the listener possesses at that time. The third is between speech perception and speech production. One's perceptual judgment determines how a pronunciation target for a given word is learned and stored by the listener but the basis of this learning of the normative pronunciation of a word relies on how the majority of the community members produce that word. The model also depicts that each language user has his or her unique and stable set of pronunciation targets in addition to a wider set of pronunciation grammar that is used when understanding another member's speech. This conjecture is based on the result from the Perception experiment showing no correlation between the amount of perceptual compensation for coarticulation and the degree of coarticulation in the same subject's production. This separation is desirable, allowing a speaker to learn various types of pronunciation variation (dialect, idiolect, foreign accent, etc.) while maintaining his or her own speech unaffected.

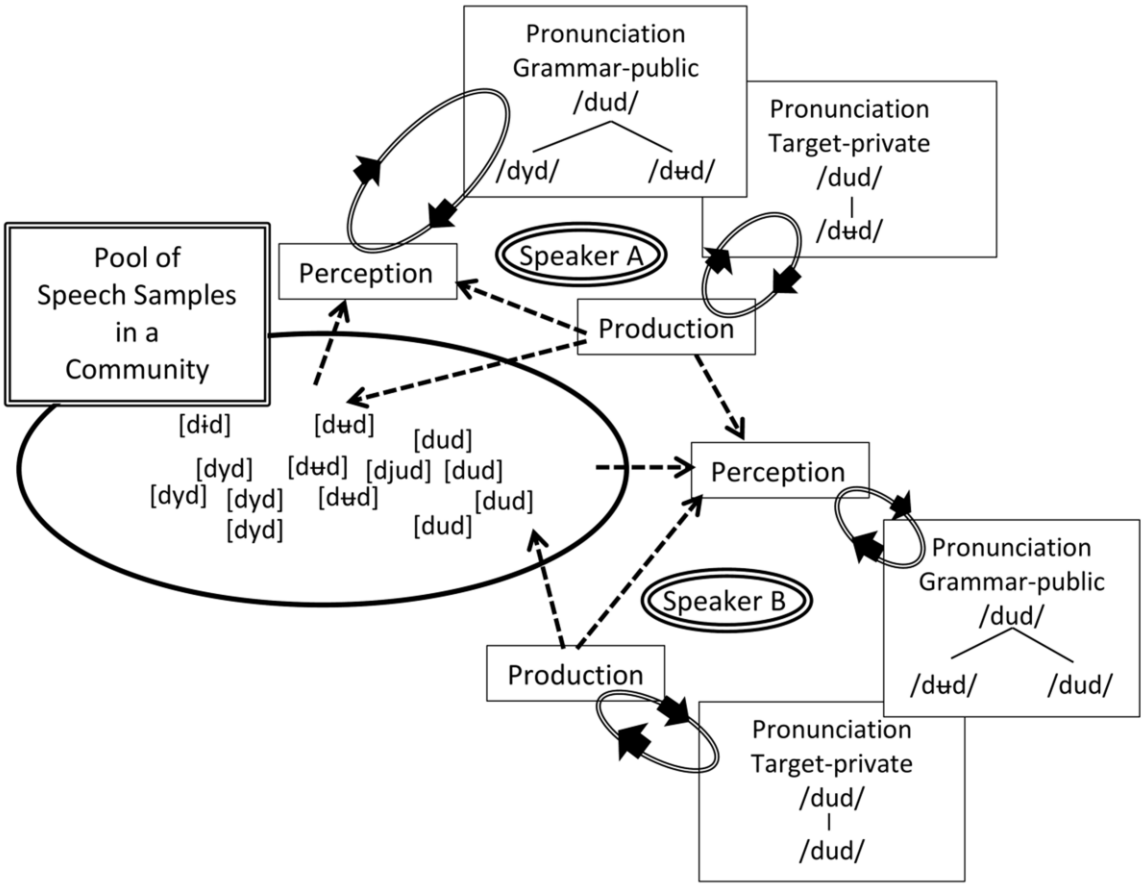



Figure 6.2 Schematic representation of interdependencies and looped causalities between: (1) individual language users and their speech community, (2) speech perception and pronunciation grammar, and (3) speech perception and speech production. Dotted arrows indicate local-level uni-directional interactions brought about by speaking and listening. The community provides a pool of speech samples that exhibit range of pronunciation variation. Language users develop: (1) pronunciation grammar that is used to understand other community member's speech, for which auditory representations play a major role, and (2) their own private pronunciation target, for which articulatory representations play a major role. How each word's pronunciation is learned by a speaker is determined by perceptual interpretation of speech samples of that word. The pronunciation grammar that serves as a frame of reference for speech perception is developed as a response to speech data produced by community members. Thus, speech perception is guided by pronunciation grammar but that grammar itself is shaped and influenced by other speakers' productions.

## 6.4 Implications for Synchronic and Diachronic Phonology

Conceptualizing speech as a dynamic interactive system has two implications for the linguistic study. One is that studying individual variation in language use provides much needed insight into the process of language change. The system depicted above is a stable system. Although each individual may have a distinct pronunciation habit, the normative pronunciation of a speech community, as might be defined as a set of mean values of particular acoustic parameters, will not be altered into any particular direction in this system. However, this system is ready to change when a particular variant is indexed with a positive social value. Many speakers already have multiple pronunciation categories in their pronunciation repertoire; therefore, given a trigger an entire community is able to respond to it with relative ease. Historical linguists have made the following distinctions among the object of study: (1) diachronic correspondence—comparing two sets of data obtained from two non-adjacent times, (2) innovation—a single person's usage (or grammar) that differs from the previous usage (or grammar), and (3) change—the adoption of an innovation by all or at least much of the community members (Janda & Joseph, 2003, p. 13). The present study suggests that innovation in the above sense occurs all the time within a stable speech community. As discussed in Chapter 1, it will take adoption of Language is an "object possessing orderly heterogeneity" (Weinreich, Labov, & Herzog, 1968, p. 100). Studying individual variation in language use is a useful approach to study language change.

Another implication is that linguistic phenomena must be studied through the interaction of multiple components of language, as assumed in the Complex Adaptive System approach (Beckner et al., 2007). One level of interaction is between the individual and the community. Language exists both in individuals (as idiolect) and in the speech community (as communal language), and each idiolect emerges from an individual's unique experience of language use with other individuals in the communal language (Beckner, et al., 2007, p. 12). Any linguistic behavior observed from individuals must be in some aspect in accord with the structure found in the communal language. Another level of interaction is between speech perception and speech production over time. Studying a consequence of one component over the other (not in a sense of the linear causation, but in the sense of the circular causation) will provide much more illuminating results than studying each component of the language in isolation.

## 6.5 Open Questions and Future Research

### 6.5.1 Phonologization of Allophones

As mentioned in Chapter 3, one of the long-standing questions for a theory of sound change is an issue of phonologization: when does coarticulatory allophonic variation become phonemic? This question highlights the importance of the one often-neglected question in synchronic phonology: are all allophones equal? Within the logic of the standard phonological theory,

allophones are the variants that are treated by the speakers in two ways at once—as belonging to the same phonological category and a single perceptual object, but as distinct production objects (Whalen, Best & Irwin, 1997). A question is: Are they all perceptually equivalent?

An example that fits the above definition is two contextual allophones of voiceless stops. Word initially, voiceless stops are aspirated except after /s/, where these stops are unaspirated. Word medially, these stops are aspirated before stressed vowels and unaspirated before unstressed vowels (Lisker & Abramson, 1967). Lisker and Abramson's production data and numerous subsequent production studies have shown that these allophones are produced as clearly distinct acoustic patterns. Regarding the tacit knowledge of the speaker, a previous study using a concept formation test showed that native speakers of American English treated these two types of allophones as belonging to the same category (Jaeger, 1986). A study on the perception and production of word medial allophones also found that these allophones were perceived by American listeners as members of a single category (Whalen et al., 1997). These results support the hypothesis that allophones are organized by speakers as systematic and distinct realizations of a single underlying phonological category.

Although the present study did not directly test the perception of naturally produced allophones of /u/ in alveolar and bilabial contexts, results from the perception study (Chapter 4) that speakers of American English judged the /u/-like vowels differently depending on the contexts and with different across-listener convergence between the contexts suggest that these listeners perceive fronted and non-fronted allophones of /u/ as distinct perceptual objects. Taken together, these results suggest that all allophones are distinct articulatory patterns of a single lexical phoneme, but not all allophones are treated as a single category in perception and cognition.

All allophones are, by definition, phonologically distinct. A theory of phonologization, therefore, has to have a tool to classify allophones into those that are only allophonically distinct and others that are fully distinct. Different perceptual distinctness among allophones may be a useful explanatory factor, because it allows us to distinguish phonologizable allophones from un-phonologizable ones. More studies on the perceptual distinctness and the cognitive status of allophones are needed so that the difference among allophones can be accounted for in the phonological theory.

## 6.5.2 Sensitivity toward Fronted /u/

One issue that has been deliberately omitted in the Perception study was the issue regarding the effect of language experience on the listener's sensitivity to fine phonetic detail. While the most (if not all) of the language effect on speech perception reviewed in Chapter 4 (§4.2.2) may be explained in terms of listener knowledge of statistical properties (e.g. distribution, co-occurrence, etc.) of speech sounds, one phenomenon that cannot be explained in terms of knowledge is "categorical perception" (Liberman, Harris, Hoffman & Griffith, 1957), which is characterized by category effect on discrimination sensitivity. Categorical perception has been replicated in many studies involving listeners of various linguistic backgrounds (Repp, 1984), suggesting that a listener's ability to discriminate speech sounds is closely related to the

phonological system that the listener has acquired. Categorical phonological status of fronted /u/ in alveolar context needs to be tested with a discrimination task.

### 6.5.3 Encoding Sub-allophonic Variation

Another question that has emerged from the Vowel Repetition experiment was about the potential factors that influence relative use of continuous acoustic representation and categorical linguistic representation. The proposed model of speech perception (Figure 6.1) assumes that even the finest representation needs to be granular and thus phonetic (i.e. phone), so the above question is re-phrased as the relative use of the finest continuous-like phonetic representation and more abstract and categorical-like representation. In addition to the various factors such as class of sounds and stimulus duration (see Pisoni, 1973a for review), the present study suggested two more potential factors. One was the amount of attention paid to fine phonetic detail. It would be of interest to test whether "the law of lesser effort" (Grammont, 1939, p. 176) applies to speech perception.

The other was the similarity of speech input and the listener's own production range. Assuming that this speculation is true, it would be of interest to conduct another vowel repetition study using the similarity between the subject's own production and the acoustic range of stimuli as an independent factor should yield useful data about perceptual categorization. Under the condition that stimuli have a matching acoustic range with the subject's natural production, listeners might respond to the stimuli with multiple pronunciation categories, guided, presumably, by the listener's knowledge of other languages and/or other dialects, or simply guided by the listener's perceptual sensitivity. Such a study would provide the much needed information about language users' ability to learn sub-phonemic and sub-allophonic variations, helping us to understand how listeners adapt to idiosyncratic speech habit of a particular speaker, various dialects, and foreign accented speeches, and in extension, how innovative pronunciation may arise from a pool of synchronic variation.

# References

Adank, P. M. (2003). Vowel normalization: a perceptual-acoustic study of Dutch vowels. (Unpublished doctoral dissertation). Katholieke Universiteit Nijmegen.

Adank, P. M., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5), 3099-3107.

Alivuotila, L., Hakokari, J., Savela, J., Happonen, R-P., & Aaltonen, O. (2007). Perception and imitation of Finnish open vowels among children, naïve adults, and trained phoneticians. *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken,* pp. 361-364.

Amos, J. (2007). Wadda boo'iful place: an analysis of the variables (ju) and (t) in Mersea Island English. (Unpublished M.A. thesis). University of Essex.

Barnes, J. (2006). *Strength and weakness at the interface: Positional neu-tralization in phonetics and phonology*. Berlin: Mouton de Gruyter.

Baudouin de Courtenay, J. (1970). Phonetic laws. In E. Stankiewics (Ed. & Trans.), *A Baudouin de Courtenay Anthology: The Beginning of Structural Linguistic*s (pp. 260-277). Indiana University Press. (Translated from French summary "Les Lois phonétiques" (pp. 57-82) of "O 'prawach głosowych,'" Rocznik slawistyczny, 3 (pp. 1-57). Original published in 1910.)

Beckman, M. E., & Pierrehumbert, J. B. (2004). Interpreting 'phonetic interpretation' over the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology VI* (pp. 13-38). Cambridge University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, 59: Suppl. 1, 1-26.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language,* 85, 785-821.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics,* 30, 591-627.

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America,* 106, 2868-2887.

Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: a study of nasal vowel height. *Phonology Yearbook,* 3, 197-217.

Bell, A. (1984). Language style as audience design. *Language in Society,* 13, 145-204.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Blevins, J., & Garrett, A. (2004). The evolution of metathesis. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 117-156). Cambridge University Press.

Bloomfield, L. (1933). *Language*. University of Chicago Press.

Blumstein, S. E., & Stevens, K. N. (1981). Phonetic features and acoustic invariance in speech. *Cognition,* 10, 25-32.

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.5.15) [computer program]. Retrieved February 18, 2007, from http://www.praat.org.

Bradlow, A. R., Kraus, N., & Heyes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research,* 46, 80-97.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, 74, 1-15.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.

Browman, C. P., & Goldstein, L. M. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155-180.

Browman, C. P., & Goldstein, L. M. (1995). Dynamics and articulatory phonology. In T. van Gelder & B. Port (Eds.), *Minds as motion* (pp. 175-193). Boston, MA: MIT Press.

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow, & S. Kemmer (Eds.), *Usage-based models of language.* Stanford, CA: CSLI Publications.

Bybee, J. (2001). Frequency effects on French liaison. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 337-359). Amsterdam: John Benjamins.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change,* 14, 261-290.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language,* 82, 711-733.

Cheng, L. L.-S. (1991). Feature geometry of vowels and co-occurrence restrictions in Cantonese. In A. Halpern (Ed.), *Proceedings of WCCFL9: The 9th West Coast Conference on Formal Linguistics* (pp. 107-124). Stanford, CA: CSLI Publications.

Chistovich, L., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. (1966). Mimicking of synthetic vowels. *Speech Transmission Laboratory-Quarterly Progress and Status Report,* 7(2), 1-18.

Cohn, A. C. (1993). Nasalisation in English: Phonology or phonetics. *Phonology,* 10, 43-81.

Cohn, A. C. (2006). Is there gradient phonology? In G. Faneslow, C. Féry, R. Vogel, & M. Schlesewsky (Eds.), *Gradience in grammar: Generative perspectives* (pp. 25-44). Oxford University Press.

Connine, C. M., & Clifton, C. Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* 13, 291-299.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 19, 81-94.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 24, 597-606.

Coupland, N. (1984). Accommpdation at work: some phonological data and their implications *International Journal of the Sociology of Language*, 46, 49-70.

Cowan, N., & Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, 79, 500-507.

Crowder, R. G. (1982). Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Human Learning and Memory*, 8, 153-162.

Cutler, A. (2010). Abstraction-based efficiency in the lexicon. *Laboratory Phonology*, 1, 301-318.

Dahan, D., & J. S. Magnuson (2006). Spoken word recognition. In M. J. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 249-283). Amsterdam: Academic Press.

De Charmes, R. C., & Zador, A. (2000). Neural representation and the cortical code. *Annual Review of Neuroscience*, 23, 613-647.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 27:769-774.

Denes, P. B., & Pinson, E. N. (1973). The speech chain. New York, NY: Knopf Doubleday.

Divenyi, P. (2009). Perception of complete and incomplete formant transitions in vowels. *Journal of the Acoustical Society of America*, 126, 1427-1439.

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. Language *Variation and Change*, 1, 245-267.

Eckert, P. (2008). Variation and the Indexical field. *Journal of Sociolinguistics*, 12, 453-476.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143-165.

Farnetani, E., & Recasens, D. (1993). Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Language and Speech*, 36, 279-302.

Fernand, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mother's and father's speech to preverbal infants. *Journal of Child Language*, 16, 47-501.

Fischer, J. L. (1958). Social Influences in the choice of a linguistic variant. *Word*, 14, 47-56.

Flemming, E. (2001). Scalar and categorical phenomena in a united model of phonetics and phonology. *Phonology,* 18, 7-44.

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. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics,* 68, 161-177.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, 48, 559-570.

Fowler, C. A., & Saltzman, E. L. (1993). Coordination and coarticulation in speech production. *Language and Speech,* 36, 171-195.

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance,* 26, 877-888.

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

Fry, C. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech,* 5, 171-189.

Fujimura, O. (1971). Remarks on stop consonants: Synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), *Form and substance: Phonetic and linguistic papers presented to Eli Fischer-Jørgensen* (pp. 221-232). Copenhagen: Akademisk Forlag.

Fujimura, O., Macchi, M. J., & Streeter, L. A. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech,* 21, 337-346.

Fujisaki, H. (1979). On the models and mechanisms of speech perception—Analysis and interpretation of categorical effects in discrimination. In B. Lindblom, & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 177-189). London: Academic Press.

Fujisaki, H. (1980). Some remarks on recent issues in speech-perception research. *Language and Speech*, 23, 75-90.

Ganong, W. F. III. (1980). Phonetic categorization in auditory word perception, *Journal of Experimental Psychology*, 6, 110-125.

Garrett, A. (in press). The historical syntax problem: Reanalysis and directionality. In D. Jonas, J. Whitman, & A. Garrett (Eds.), *Grammatical change: Origins, nature, outcomes.* Oxford University Press.

Garrett, A., & Johnson, K. (in press). Phonetic bias in sound change. In A. C.-L. Yu (Ed.), Origins of sound change: Approaches to phonologization. Oxford University Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22, 1166-1183.

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

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics,* 31, 305-320.

Goldstein, L. M., Byrd, D., & Saltzman, E. L. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action of language via the mirror neuron system* (pp. 215-249). New York, NY: Cambridge University Press.

Grammont, M. (1939). *Traité de phonétique.* Paris: Delagrave.

Grant, K.W., & Braida, L. D. (1991). Evaluating the articualtion index for auditory-visual input. *Journal of the Acoustical Society of America,* 89, 2952-2960.

Grice, P. (1989). Logic and conversation. In H. P. Grice (Ed.), *Studies in the way of words* (pp. 22-40). Cambridge, MA: Harvard University Press. (Reprinted from D. Davidson & G. Harman (Eds), *The Logic of Grammar* (pp. 64-75), 1975, Encino, CA: Dickenson.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39, 350-365.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America,* 100, 1111-1121.

Guion, S. G. (1996) Velar palatalization: Coarticulation, perception, and sound change. (Doctoral dissertation). University of Texas at Austin.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America,* 102, 655-658.

Hale, M. (2003). Neogrammarian sound change. In B. D. Joseph, & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 343-368). Malden, MA: Blackwell Publishing.

Hansson, G. Ó. (2008). Diachronic explanations of sound patterns. *Language & Linguistic Compass,* 2, 859-893.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the queen's Christmas broadcasts. *Journal of Phonetics*, 34, 439-457.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America,* 123, 2825-2835.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics,* 31, 373-405.

Hawkins, S. (2004). Puzzles and patterns in 50 years of research on speech perception. In J. Slifka, S. Manuel & M. Matties (Eds.), *From Sound to Sense: 50+ Years of Discovery in Speech Communication* (pp. B223-B246). Cambridge, MA, MIT.

Hawkins, S., & Smith, R. (2001). Polysp: A polysystemic, phonetically rich approach to speech understanding. *Rivista di Linguistica,* 13, 99-188.

Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). Oprah and /ay/: lexical frequency, referee design and style. *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco,* CA, pp. 1389-1392.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of phonetics,* 34, 458-484.

Hayes, B. & Steriade, D. (2004). Introduction: the phonetic bases of phonological Markedness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 1-33). Cambridge University Press.

Hillenbrand, J., Getty, L. A., Clark, M. J., & and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America,* 97, 3099-3111.

Hock, H. H. (1991). *Principles of historical linguistics*. Berlin: Mouton de Gruyter.

Hock, H. H., & Joseph, B. D. (1996). *Language history, language change, and language relationship: An introduction to historical and comparative linguistics.* New York, NY: Mouton de Gruyter.

Holt, L. L., & Kluender, K. R. (2000). General auditory processes contribute to perceptual accommodation of coarticulation. *Phonetica,* 57, 170-180.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America,* 108, 710-722.

Holt, L. L., Stephens, J. D., & Lotto, A. J. (2005). A critical evaluation of visually-moderated phonetic context effects. *Perception & Psychophysics*, 67, 1102-1112.

Hombert, J.-M. (1974). Towards a theory of tonogenesis: an empirical, physiologically and perceptually-based account of the development of tonal contrasts in language. (Doctoral dissertation). University of California at Berkeley.

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

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. M. Christie, Jr. (Ed.), *Current progress in historical linguistics* (pp. 96-105). Amsterdam: North-Holland.

Houde, J. F., Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research,* 45, 295-310.

House A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America,* 25, 105-113.

Hume, E. (1996). Coronal consonants, front vowel parallels in Maltese. *Natural Language and Linguistic Theory, 14,* 163-203.

Hyman, L. M. (1972). Nasals and nasalization in Kwa. *Studies in African Linguistics*, 3, 167-206.

Hyman, L. M. (1975). *Phonology: Theory and analysis*. New York, NY: Holt, Rinehart & Winston.

Hyman, L. M. (1976). Phonologization. In A. Juilland (in collaboration with A. M. Devine, & L. D. Stephens) (Ed.), *Linguistic studies presented to Joseph H. Greenberg* (Vol. 4, pp. 407-418). Saratoga, CA: Anma Libri.

Hyman, L. M. (2008). Enlarging the scope of phonologization. *UC Berkeley Phonology Lab Annual Report 2008* (p. 382-409). University of California Phonology Lab

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America,* 97, 553-562.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America,* 99, 1130-1140.

Jaeger, J. (1986) Concept formation as a tool for linguistic research. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 211-237). Orlando, FL: Academic Press.

Janda, R. D., & Joseph, B. D. (2003). On language, change, and language change—Or, of history, linguistics, and historical linguistics. In B. D. Joseph, & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 3-180). Malden, MA: Blackwell Publishing.

Johnson, K. (1990). The role of perceived speaker identity in *F0* normalization of vowels. *Journal of the Acoustical Society of America,* 88, 642-654.

Johnson, K. (1991). Differential effects of speaker and vowel variability on fricative perception. *Language and Speech*, 34, 265-279.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-166). San Diego, CA: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of phonetics,* 34, 485-499.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of phonetics,* 27, 359-384.

Jonasson, J. (1971). Perceptual similarity and articulatory reinterpretation as a source of phonological innovation. *Speech Transmission Laboratory-Quarterly Progress and Status Report,* 12(1), 30-42.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Jusczyk, P. W. (1993). From general to language-specific capacities—the WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3-28.

Kataoka, R. (2009). A Study on Perceptual Compensation for /u/-fronting in American English. In *Proceedings of the 35th annual meeting of the Berkeley Linguistics Society* (pp. 156-167). Berkeley, CA: Berkeley Linguistics Society.

Katseff, S. (2010). Linguistic constraints on compensation for altered auditory feedback. (Doctoral dissertation). University of California at Berkeley.

Kawasaki, H. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In. J. J. Ohala, & Jaeger, J. J. (Eds.), *Experimental phonology* (pp. 239-252). Orlando, FL: Academic Press.

Keating, P. A. (1998). Phonetic representations in a generative grammar. *Journal of Phonetics,* 18, 321-334.

Kent, R. D. (1973). The imitation of synthetic vowels and some implications for speech memory. *Phonetica,* 28, 1-25.

Kent, R. D. (1974). Auditory-motor formant tracking: A study of speech imitation. *Journal of Speech and Hearing Research,* 17, 203-222.

Kent, R. D. (1979). Imitation of synthesized English and nonEnglish vowels by children and adults. *Journal of Psycholinguistic Research,* 8, 43-60.

Kent, R. D., & Moll, K. L. (1972). Cinefluorographic analysis of selected lingual consonants. *Journal of Speech and Hearing Research,* 15, 453-473.

Kiparsky, P. (1965). Phonological change. (Doctoral dissertation). Massachusetts Institute of Technology.

Kiparsky, P. (1988). Phonological change. In F. Newmwyer (Ed.), *Linguistics: The Cambridge survey* (pp. 363-415). Cambridge University Press.

Kiparsky, P. (2003). The phonological basis of sound change. In B. D. Joseph, & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 313-342). Malden, MA: Blackwell.

Kiparsky, P. (2006). The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics*, 32, 217-236.

Kiparsky, P. (2008). Universals constrain change; change results in typological generalizations. In J. Good (Ed.), *Linguistic universals and language change* (pp. 23-53). New York, NY: Oxford University Press.

Kiritani, S. (1986). X-ray microbeam method for measurement of articulatory dynamics-techniques and results. *Speech Communication,* 5, 119-140.

Kiritani, S., Itoh, K., Hirose, H., & Sawashima, M. (1977). Coordination of the consonant and vowel articulations X-ray microbeam study on Japanese and English. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* (Vol. 11, pp. 11-21). Tokyo University.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology,* 51, 141-178.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory & Language,* 56, 1-15.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 19, 332-338.

Koops, C. (2010). /u/-fronting is not monolithic: two types of fronted /u/ in Houston Anglos. *University of Pennsylvania Working Papers in Linguistics, 16 (2)*, 113-122.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics,* 50, 93-107.

Kuhl, P. K., Andruski, J. E., Chistovich, L., Chistovich, I., Kozhevnikova, E., Sundberg, U., & Lacerda, F. (1997). Cross language analysis of phonetic units in language addressed to infants, *Science,* 227, 684–6.

Labov, W. (1963). The social motivation of a sound change. *Word*, 19, 273-309.

Labov, W. (1666). *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.

Labov, W. (1994). *Principles of linguistic change, vol. 1: Internal factors.* (Language in Society 20) Cambridge, MA: Blackwell.

Labov, W. (2001). *Principles of linguistic change, vol. 2: Social factors.* (Language in Society 29) Malden, MA: Blackwell.

Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, 34, 500-515.

Labov, W. (2007). Transmission and diffusion. *Language*, 83, 344-387.

Labov, W. (2010). *Principles of linguistic change, vol. 3: Cognitive and cultural factors.* (Language in Society 39) Malden, MA: Wiley-Blackwell.

Labov, W. Ash, S., Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology, and sound change.* Berlin: Mouton de Gruyter.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America,* 29, 98-104.

Lahiri, A. & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 637-676). Berlin: Mouton de Gruyter.

Lehiste, I., & Peterson, G. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America,* 33, 419-425.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America,* 29, 117-123.

Liberman, A. M., Harris, K. S., Eimas, P., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech,* 4, 175-195.

Liberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology,* 54, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology,* 61, 379-388.

Liberman, A. M., & Mattingly, G. (1985). The Motor Theory of speech perception revised. *Cognition,* 21, 1-36.

Liberman, A. M. & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences,* 4, 187-196.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America,* 35, 1773-1781.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217-245). New York, NY: Springer.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W, J. Hardcastle, & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403-439). Dordrecht: Kluwer.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica,* 7, 5-37.

Lindblom, B., & Studdert-Kennedy, M. (1967). On the rôle of formant transitions in vowel recognition. *Journal of the Acoustical Society of America,* 42, 830-843.

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

Lotto, A. J., Kluender, K. R. (1998). General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics,* 60, 602-619.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America* 102, 1134-1140.

MacNeilage, P. F., & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America,* 45, 1217-1233.

Magnuson, J.S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cognitive Science,* 27, 285-298.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407-412.

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition,* 24, 169-196.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 28, 213-228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America,* 69, 548-558.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition,* 25, 71-102.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology,* 10, 29-63.

Martinet, A. (1952). Function, structure and sound change. *Word,* 8, 1-32.

Martinet, A. (1962). *A functional view of language.* Oxford University Press.

Matisoff, J. A. (1973). Tonogenesis in southeast Asia. In L. M. Hyman (Ed.), *Consonant types and tone* (pp. 71-95), Los Angeles, CA: University of Southern California.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology,* 2, 131-157.

Maye, J., Aslin, R. N., Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science,* 32, 543-562.

Mazaudon, M., & Michailovsky, B. (1989). Lost syllables and tone contour in Dzongkha (Bhutan). In D. Bradley, E. J. A. Henderson, & M. Mazaudon (Eds.), *Prosodic analysis and Asian linguistics: to honour R. K. Sprigg* (pp. 115-136). Canberra: Australian National University, Research School of Pacific Studies.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature,* 264, 746-748.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29, 4, 539–553.

Michailovsky, B. (1975). On some Tibeto-Burman sound changes. In *Proceedings of the first annual meeting of the Berkeley Linguistics Society* (pp. 322-331). Berkeley, CA: Berkeley Linguistics Society.

Miller, J. L. (2001). Mapping from acoustic signal to phonetic category: Internal category structure, context effects and speeded categorization. *Language and Cognitive Processes,* 16, 683-690.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics,* 25, 457-465.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of phonetic category. *Perception & Psychophysics,* 46, 505-512.

Mitterer, H. (2006). On the cause of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics*, 68, 1227-1240.

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

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

Nearey, T. M. (1978). *Phonetic feature systems for vowels*. Indiana University Linguistic Club.

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: effects of temporal distance. *Perception & Psychophysics,* 58, 540-560.

Nguyen, N. (to appear). Representations of speech sound patterns in the speaker's brain: Insights from perception studies. In A. Cohn, C. Fougeron & M. Huffman (Eds.), *Handbook of laboratory phonology*. Oxford University Press.

Nguyen, N., Wauquier, S., & Tuller, B. (2009). The dynamical approach to speech perception: From the phonetic detail to abstract phonological categories. In F. Pellegrino, E. Marsico, I. Chitoran & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 191-218). Berlin: Walter de Gruyter.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences,* 23, 299-370.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology,* 47, 204-238.

Nowak, P. M. (2006). Vowel reduction in Polish (Doctoral dissertation). University of California at Berkeley.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from The 17th Regional Meeting of the Chicago Linguistics Society: Parasession on Language and Behavior* (pp. 178-203). Chicago, IL: Chicago Linguistics Society.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik, & E. H. Jahr (Eds.), *Trends in Linguistics, Studies and Monographs Language change: No. 43. Contributions to the study of its causes*: (pp. 173-198). Berlin: Mouton de Gruyter.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: problems and perspectives* (pp. 237-278). London: Longman.

Ohala, J. J. (2008). Understanding variability in speech: A brief survey over 2.5 millennia. *UC Berkeley Phonology Lab Annual Report 2008,* 366-373.

Ohala, J. J., & Busa, M. G. (1995). Nasal loss before voiceless fricatives: A perceptually-based sound change. In C. Fowler (Ed.) *Special issue on the phonetic basis of sound change. Rivista di Linguistica,* 7, 125-144.

Ohala, J. J., & Feder, D. (1994). Listeners' identification of speech sounds is influenced by adjacent "restored" phonemes. *Phonetica,* 51, 111-118.

Ohala, J. J., & Shriberg, E. E. (1990). Hyper-correction in speech perception. *Proceedings of ICSLP 90: First International Conference on Spoken Language Processing, Kobe,* Vol. 1, pp. 405-408.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America,* 75, 224-230.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America,* 39, 151-168.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America,* 41, 310-320.

Oshthoff, H. & Brugmann, K. (1967). Preface to morphological investigations in the sphere of the Indo-European languages I. In W. P. Lehmann (Trans. & Ed.), *A reader in nineteenth century historical Indo-European linguistics* (pp. 197-209). Bloomington, IN: Indiana University Press. (Translated and reprinted from *Morphologische untersuchungen auf dem gebiete der indogermanischen sprachen I*, 1878, Leipzig, Germany: S. Hirzel).

Oudeyer, P-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology,* 233, 435-449.

Paul, H. (1970). *Principles of the history of language.* (H. A. Strong, Trans.). London. UK: Swan Sonnenschein, Lowrey, & Co. (Translated and reprinted from *Prinzipien der sprachgeschichte* (2nd ed.)., 1888, Halle (Saale): Max Niemeyer).

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America,* 24, 175-184.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language,* 60, 320-342.

Phillips, B. S. (2001). Lexical diffusion, lexical frequency, and lexical analysis. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 123-136). Amsterdam: John Benjamins.

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

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101-139). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech,* 46, 115-154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics,* 34, 516-530.

Pisoni, D. B. (1973a). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics,* 13, 253-260.

Pisoni, D. B. (1973b). The role of auditory short-term memory in vowel perception. *Haskins Laboratories Status Report,* SR-34, 89-117.

Pisoni, D. B., & Levi, S. V. (2005). Some observations on representations and representational specificity in speech perception and spoken word recognition. *Research on Spoken Language Processing Progress Report* (No. 27, pp. 3-26). Speech Research Laboratory, Indiana University.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics,* 15, 285-290.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language,* 61, 19-36.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language,* 39, 347-370.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and metaanalytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance,* 19, 699-725.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWinney (Ed.), *The emergence of language* (pp, 381-415). Mahwah, NJ: Erlbaum.

Poeppel, D., Idsardi, W. J., & Van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Phylosophical Transactions of the Royal Society B: Biological Sciences,* 363, 1071-1086.

Pols, L. C. W. & Van Son, R. J. J. H. (1993). Acoustics and perception of dynamic vowel segments. *Speech Communication*, 13, 135-147.

Recasens, D. (1991). An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics,* 19, 179-192.

Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America,* 125, 2288-2298.

Recasens, D., Pallarés, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America,* 102, 544-561.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics,* 30, 217-227.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol 10, pp. 243-335). New York, NY: Academic Press.

Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America,* 69, 1154-1163.

Repp, B. H., & Williams, D. R. (1985). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. *Speech Communication,* 4, 105-120.

Repp, B. H., & Williams, D. R. (1987). Categorical tendencies in imitating self-produced isolated vowels. *Speech Communication,* 6, 1-14.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology,* 1, 333-382.

Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology,* 62, 49-72.

Samuel, A. G., & Kraljic, T. (2009). Tutorial review: Perceptual learning for speech. *Attention, Perception & Psychophysics*, 71, 1207-1218.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language,* 48, 416-434.

Sancier, M. L., Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics,* 25, 421-436.

Schilling-Estes, N. (2002). Investigating stylistic variation. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.) *The handbook of language variation and change* (pp. 375-401). Malden, MA: Blackwell.

Schouten, M. E. H. (1977). Imitation of synthetic vowels by bilinguals. *Journal of phonetics,* 5, 273-283.

Schouten, M. E. H. & Van Hessen, A. J. (1992). Modeling phoneme perception, I: Categorical perception. *Journal of the Acoustical Society of America,* 92, 1841-1855.

Schuchardt, Hugo. (1972). On sound laws: Against the neogrammarians. In T. Vennemann & T. H. Wilbur (Trans. & Eds.), *Schuchardt, the neogrammarians, and the transformational theory of phonological change: Four essays* (pp. 39-72). Frankfurt, Germany: Athenäum Verlag. (Translated and reprinted from *Über die Lautgesetze. Gegen die Junggrammatiker*. 1885, Berlin: Robert Oppenheim.

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

Silverman, D. (2006). The diachrony of labiality in Trique, and the functional relevenace of gradience and variation. In L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology 8* (pp. 133-154). Berlin: Mouton de Gruyter.

Smith, C. (2007). Prosodic accommodation by French speakers to a non-native interlocutor. *Proceedings of the XVIth International Congress of Phonetic Sciences*. Saarbrucken, Germany.

Solé, M. J. (1992). Phoentic and phonological processes: The case of nasalization. *Language and Speech*, 35, 29-43.

Solé, M. J., & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 607-655). Berlin: Mouton de Gruyter.

Sonderegger, M., & Yu, A. C.-L. (2010). A rational account of perceptual compensation for coarticulation. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 375-380). Austin, TX: Cognitive Science Society.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America,* 111, 1872-1891.

Stevens, K. N. (2004). Invariance and variability in speech: Interpreting acoustic evidence. In J. Slifka, S. Manuel & M. Matties (Eds.), *From Sound to Sense: 50+ Years of Discovery in Speech Communication* (pp. B77-B85). Cambridge, MA: MIT.

Stevens, K. N., Andrade, A., & Viana, M. C. (1987). Perception of vowel nasalization in VC contexts: a cross-language study. *Journal of the Acoustical Society of America*, 82(S1), S119 (A).

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 64, 1358-1368.

Steves, K. N., & House, A. S. (1963). Pertubation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research,* 6, 111-128.

Steves, K. N., House, A. S., & Paul, A. P. (1966). Acoustical description of syllable nuclei: An interpretation in terms of a dynamic model of articulation. *Journal of the Acoustical Society of America,* 40, 123-132.

Strand, E. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology,* 18, 86-100.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, 36, 314-331.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd, & R. Campbell (Eds.), *Hearing by eye: the psychology of lip-reading* (pp. 3-51). London: Lawrence Erlbaum.

Sumner, M., & Samuel, A. (2009). The role of experience in the processing of cross-dialectal variation. *Journal of Memory and Language*, 60, 487-501.

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America,* 90, 1309-1325.

Sweet, H. (1888). *A history of English sounds from the earliest period, with full word-lists*. Oxford: Clarendon Press.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423, 866-869.

Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge University Press.

Uther, M., Knoll, M., & Burnham, D. (2007). Do you speak E-N-G-L-I-S-H? A comparison of foreigner- and infan-directed speech. *Speech Communication*, 49, 2-7.

Vallabha, G. K., & Tuller, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *Journal of the Acoustical Society of America*, 116, 1184-1197.

van Alphen, P. M., & McQueen, J. M. (2006). The effect of Voice Onset Time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 178-196.

van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing, *Journal of Phonetics,* 32, 455-491.

Van Hessen, A. J., & Schouten, M. E. H. (1992). Modeling phoneme perception. II: A model of stop consonant discrimination. *Journal of the Acoustical Society of America,* 92, 1856-1868.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America,* 92, 723-735.

Vroomen, J., & De Gelder, B. (2001). Lipreading and the compensation for coarticualtion mechanism. *Language and Cognitive Processes*, 2001, 661-672.

Wayland, S. C., Miller, J. L. & Volaitis, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America,* 95, 2694-2701.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science,* 167, 392-393.

Wedel, A. (2006). Exemplar models, evolution and language change. *Linguistic Review,* 23, 247-274.

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann, & Y. Malkiel (Eds.), *Directions for historical linguistics: A symposium* (pp. 95-188). University of Texas Press.

Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25, 501-528.

Wheeler, B. I. (1901). The cause of uniformity in phonetic change. In *Transactions and Proceedings of the American Philological Association* (pp. 5-15). Boston, MS: Ginn.

Yoneyama, K. (2007). *Contributions Towards Research and Education of Language: Vol. 14 [語学教育フォーラム　第14号]. Phonological neighborhoods and phonetic similarity in Japanese word recognition.* Tokyo: Institute for the Research and Education of Language, Daito Bunka University.

Yu, A. C.-L. (2004). Explaining final obstruent voicing in Lezgian: Phonetics and history, *Language*, 80, 73-97.

Yu, A. C.-L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE*, 5(8), e11950.

Yu, A. C.-L. (in press) Individual difference in socio-cognitive processing and the actuation of sound change. In A. C.-L. Yu (Ed.), *Approaches to phonologization*. Oxford University Press.

Appendix A—Chapter 3: Median F1 and F2 (Hz) of vowels, measured at the temporal midpoint of a vowel, in reference words *heed, hid, head, had, hot, HUD, hood,* and *who'd*. Data of subject #18 and #30 were excluded (*N* = 30: 18 females & 14 males).

| | | *heed* /hid/ | | *hid* /hɪd/ | | *head* /hɛd/ | | *had* /hæd/ | |
|---|---|---|---|---|---|---|---|---|---|
| Sbj. | Sex | F1 | F2 | F1 | F2 | F1 | F2 | F1 | F2 |
| 2 | F | 273 | 2840 | 466 | 2236 | 691 | 1935 | 866 | 1835 |
| 3 | F | 279 | 2621 | 480 | 2230 | 665 | 2025 | 977 | 1768 |
| 4 | F | 237 | 2632 | 373 | 2052 | 648 | 2014 | 847 | 1857 |
| 5 | F | 315 | 2737 | 508 | 2180 | 645 | 2009 | 833 | 1829 |
| 6 | F | 366 | 2888 | 547 | 2315 | 670 | 2144 | 928 | 1883 |
| 8 | F | 359 | 2750 | 560 | 2231 | 694 | 1967 | 939 | 1756 |
| 9 | F | 369 | 2905 | 563 | 2399 | 809 | 1843 | 959 | 1353 |
| 13 | F | 367 | 3001 | 518 | 2626 | 764 | 2456 | 958 | 2242 |
| 14 | F | 318 | 3017 | 466 | 2415 | 704 | 2186 | 995 | 1887 |
| 15 | F | 330 | 2830 | 540 | 2286 | 712 | 2217 | 990 | 2038 |
| 16 | F | 259 | 2829 | 519 | 2283 | 705 | 2050 | 964 | 1839 |
| 17 | F | 343 | 2675 | 541 | 2253 | 723 | 2025 | 897 | 1880 |
| 21 | F | 397 | 2800 | 569 | 2101 | 781 | 1958 | 962 | 1875 |
| 23 | F | 318 | 2759 | 494 | 2239 | 679 | 2101 | 920 | 1931 |
| 24 | F | 325 | 2895 | 476 | 2355 | 683 | 2075 | 856 | 1819 |
| 25 | F | 393 | 2886 | 587 | 2422 | 774 | 2212 | 1005 | 1851 |
| 26 | F | 306 | 2887 | 467 | 2283 | 744 | 2045 | 988 | 1906 |
| 27 | F | 357 | 2648 | 456 | 2189 | 695 | 2011 | 879 | 1765 |
| F-Average | | 328 | 2811 | 507 | 2283 | 710 | 2071 | 931 | 1851 |
| 1 | M | 301 | 2399 | 444 | 1891 | 563 | 1740 | 783 | 1573 |
| 7 | M | 192 | 2216 | 342 | 1847 | 444 | 1677 | 557 | 1610 |
| 10 | M | 147 | 2547 | 340 | 2098 | 414 | 2083 | 640 | 1770 |
| 11 | M | 222 | 2443 | 373 | 1883 | 441 | 1687 | 548 | 1592 |
| 12 | M | 276 | 2323 | 367 | 1877 | 571 | 1663 | 679 | 1476 |
| 19 | M | 292 | 2139 | 451 | 1745 | 536 | 1599 | 653 | 1541 |
| 20 | M | 283 | 2529 | 407 | 1943 | 572 | 1691 | 781 | 1501 |
| 22 | M | 269 | 1958 | 381 | 1790 | 496 | 1640 | 649 | 1561 |
| 28 | M | 358 | 1768 | 411 | 1730 | 522 | 1603 | 620 | 1571 |
| 29 | M | 286 | 2178 | 410 | 2113 | 560 | 1896 | 688 | 1691 |
| 31 | M | 292 | 2394 | 445 | 2049 | 619 | 1793 | 752 | 1609 |
| 32 | M | 275 | 2025 | 401 | 1815 | 529 | 1615 | 675 | 1549 |
| M-Average | | 266 | 2243 | 398 | 1898 | 522 | 1724 | 669 | 1587 |

Appendix A—Chapter 3 (Continued): Median F1 and F2 (Hz) of vowels, measured at the temporal midpoint of a vowel, in reference words *heed, hid, head, had, hot, HUD, hood,* and *who'd*. Data of subject #18 and #30 were excluded (*N* = 30: 18 females & 14 males).

| | | *hot* /hɑt/ | | *HUD* /hʌd/ | | *hood* /hʊd/ | | *who'd* /hud/ | |
|---|---|---|---|---|---|---|---|---|---|
| Sbj. | Sex | F1 | F2 | F1 | F2 | F1 | F2 | F1 | F2 |
| 2 | F | 935 | 1549 | 675 | 1637 | 416 | 1679 | 247 | 1038 |
| 3 | F | 935 | 1426 | 637 | 1713 | 500 | 1604 | 350 | 1423 |
| 4 | F | 861 | 1186 | 654 | 1727 | 482 | 1478 | 272 | 1217 |
| 5 | F | 848 | 1409 | 645 | 1711 | 514 | 1687 | 343 | 1425 |
| 6 | F | 922 | 1445 | 686 | 1826 | 532 | 1896 | 397 | 1446 |
| 8 | F | 890 | 1581 | 698 | 1843 | 510 | 1803 | 409 | 1808 |
| 9 | F | 915 | 1500 | 801 | 1814 | 598 | 1890 | 354 | 1624 |
| 13 | F | 949 | 1752 | 780 | 2092 | 583 | 2125 | 355 | 1124 |
| 14 | F | 957 | 1440 | 709 | 1839 | 546 | 1721 | 295 | 1510 |
| 15 | F | 986 | 1519 | 756 | 1901 | 532 | 2042 | 341 | 1484 |
| 16 | F | 860 | 1372 | 718 | 1779 | 554 | 1733 | 393 | 1454 |
| 17 | F | 873 | 1447 | 716 | 1686 | 525 | 1773 | 368 | 1670 |
| 21 | F | 992 | 1489 | 721 | 1781 | 656 | 1822 | 437 | 2031 |
| 23 | F | 880 | 1338 | 698 | 1592 | 493 | 1471 | 317 | 1189 |
| 24 | F | 966 | 1486 | 692 | 1809 | 478 | 1771 | 327 | 1398 |
| 25 | F | 944 | 1394 | 839 | 1826 | 679 | 1789 | 421 | 1564 |
| 26 | F | 1068 | 1492 | 780 | 1746 | 568 | 1741 | 356 | 1654 |
| 27 | F | 841 | 1395 | 727 | 1801 | 507 | 1829 | 402 | 1986 |
| F-Average | | 923 | 1457 | 718 | 1785 | 537 | 1770 | 355 | 1503 |
| 1 | M | 700 | 1109 | 593 | 1408 | 476 | 1367 | 359 | 1174 |
| 7 | M | 538 | 1101 | 458 | 1371 | 355 | 1504 | 261 | 1143 |
| 10 | M | 588 | 1121 | 440 | 1631 | 314 | 1576 | 331 | 1545 |
| 11 | M | 671 | 1331 | 475 | 1517 | 364 | 1575 | 234 | 1360 |
| 12 | M | 593 | 1272 | 538 | 1313 | 441 | 1441 | 278 | 1145 |
| 19 | M | 675 | 1270 | 539 | 1422 | 469 | 1389 | 309 | 1362 |
| 20 | M | 704 | 1155 | 586 | 1405 | 444 | 1368 | 362 | 1364 |
| 22 | M | 666 | 1170 | 553 | 1353 | 433 | 1341 | 366 | 1075 |
| 28 | M | 619 | 1097 | 575 | 1303 | 438 | 1545 | 370 | 1167 |
| 29 | M | 657 | 1106 | 663 | 1370 | 331 | 1500 | 316 | 1502 |
| 31 | M | 718 | 1184 | 594 | 1391 | 505 | 1356 | 349 | 1227 |
| 32 | M | 668 | 1095 | 591 | 1464 | 438 | 1560 | 311 | 1292 |
| M-Average | | 650 | 1168 | 550 | 1412 | 417 | 1460 | 321 | 1280 |

Appendix B—Chapter 3: Mean of median F1 and F2 (Hz) and mean duration (ms) of vowels calculated from 4-6 tokens of each test words (*dude*, *dune*, noon, toot, tune, Seuss, and zoos (i.e. /D_D/ contexts) in fast, medium, and slow speech. F1 and F2 were measured at the point of F2 minimum for each token. Data of subject #18 and #30 were excluded (*N* = 30: 18 females & 14 males).

| | | fast | | | medium | | | slow | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sbj. | Sex | F1 | F2 | Dur. | F1 | F2 | Dur. | F1 | F2 | Dur. |
| 2 | F | 198 | 1976 | 129 | 204 | 1695 | 182 | 131 | 1484 | 287 |
| 3 | F | 251 | 1982 | 103 | 271 | 1892 | 157 | 242 | 1650 | 288 |
| 4 | F | 243 | 2015 | 107 | 174 | 1797 | 101 | 213 | 2041 | 150 |
| 5 | F | 337 | 1768 | 113 | 336 | 1729 | 130 | 329 | 1679 | 161 |
| 6 | F | 313 | 2130 | 102 | 299 | 2100 | 133 | 288 | 2038 | 141 |
| 8 | F | 363 | 1900 | 112 | 359 | 1654 | 162 | 360 | 1782 | 174 |
| 9 | F | 345 | 1961 | 112 | 345 | 2014 | 129 | 313 | 1915 | 149 |
| 13 | F | 266 | 1975 | 104 | 265 | 1932 | 108 | 255 | 1674 | 124 |
| 14 | F | 253 | 2338 | 120 | 244 | 2290 | 141 | 236 | 2033 | 214 |
| 15 | F | 237 | 2107 | 136 | 237 | 2079 | 166 | 230 | 2049 | 219 |
| 16 | F | 296 | 1814 | 126 | 328 | 1753 | 159 | 305 | 1608 | 225 |
| 17 | F | 350 | 2005 | 94 | 327 | 2096 | 129 | 322 | 2031 | 151 |
| 21 | F | 361 | 2022 | 162 | 383 | 2028 | 234 | 362 | 2085 | 199 |
| 23 | F | 316 | 1668 | 176 | 318 | 1561 | 245 | 316 | 1551 | 301 |
| 24 | F | 309 | 2001 | 143 | 299 | 1916 | 183 | 298 | 1589 | 318 |
| 25 | F | 369 | 2090 | 144 | 365 | 1997 | 189 | 355 | 2009 | 226 |
| 26 | F | 294 | 2056 | 122 | 295 | 1990 | 146 | 287 | 1960 | 239 |
| 27 | F | 326 | 1993 | 118 | 314 | 2035 | 132 | 310 | 1919 | 157 |
| F-Average | | 302 | 1989 | 124 | 298 | 1920 | 157 | 286 | 1839 | 207 |
| 1 | M | 307 | 1616 | 86 | 312 | 1566 | 121 | 309 | 1435 | 205 |
| 7 | M | 265 | 1581 | 127 | 234 | 1591 | 191 | 256 | 1555 | 260 |
| 10 | M | 207 | 1494 | 122 | 144 | 1667 | 151 | 181 | 1712 | 167 |
| 11 | M | 232 | 1517 | 130 | 185 | 1517 | 156 | 226 | 1335 | 223 |
| 12 | M | 277 | 1697 | 116 | 278 | 1650 | 150 | 262 | 1606 | 207 |
| 19 | M | 335 | 1554 | 96 | 311 | 1584 | 149 | 325 | 1557 | 216 |
| 20 | M | 322 | 1570 | 89 | 323 | 1631 | 119 | 319 | 1371 | 232 |
| 22 | M | 303 | 1632 | 108 | 294 | 1627 | 162 | 272 | 1468 | 208 |
| 28 | M | 315 | 1660 | 135 | 299 | 1645 | 177 | 305 | 1591 | 275 |
| 29 | M | 302 | 1762 | 111 | 324 | 1773 | 147 | 295 | 1690 | 232 |
| 31 | M | 327 | 1662 | 101 | 314 | 1638 | 150 | 303 | 1529 | 248 |
| 32 | M | 239 | 1780 | 110 | 256 | 1793 | 125 | 250 | 1693 | 221 |
| M-Average | | 286 | 1627 | 111 | 273 | 1640 | 150 | 275 | 1545 | 225 |

Appendix C—Chapter 3: Median F1 and F2 (Hz) and mean duration (ms) of vowels calculated from 4-6 tokens of the control word *booed* (/bud/) in fast, medium, and slow speech. F1 and F2 were measured at the temporal midpoint of a vowel for each token. Data of subject #18 and #30 were excluded (*N* = 30: 18 females & 14 males).

| | | fast | | | medium | | | slow | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sbj. | Sex | F1 | F2 | Dur. | F1 | F2 | Dur. | F1 | F2 | Dur. |
| 2 | F | 179 | 1092 | 112 | 171 | 1102 | 217 | 158 | 1045 | 264 |
| 3 | F | 213 | 1277 | 98 | 300 | 1471 | 158 | 232 | 1275 | 348 |
| 4 | F | 218 | 1350 | 101 | 177 | 1364 | 135 | 214 | 1252 | 172 |
| 5 | F | 318 | 1384 | 110 | 325 | 1260 | 145 | 324 | 1227 | 182 |
| 6 | F | 341 | 1295 | 116 | 257 | 1427 | 149 | 341 | 1358 | 164 |
| 8 | F | 395 | 1464 | 100 | 401 | 1474 | 189 | 390 | 1484 | 187 |
| 9 | F | 326 | 1308 | 109 | 315 | 1658 | 161 | 314 | 1460 | 162 |
| 13 | F | 268 | 1048 | 115 | 248 | 1007 | 127 | 253 | 889 | 140 |
| 14 | F | 278 | 1873 | 130 | 173 | 1745 | 159 | 229 | 1469 | 216 |
| 15 | F | 261 | 1793 | 190 | 231 | 1742 | 144 | 216 | 1697 | 232 |
| 16 | F | 232 | 1437 | 112 | 309 | 1490 | 181 | 299 | 1318 | 280 |
| 17 | F | 361 | 1538 | 98 | 343 | 1819 | 142 | 347 | 1747 | 178 |
| 21 | F | 383 | 1867 | 166 | 375 | 1904 | 242 | 388 | 1941 | 239 |
| 23 | F | 308 | 1334 | 185 | 304 | 1333 | 232 | 321 | 1241 | 265 |
| 24 | F | 311 | 1434 | 128 | 288 | 1297 | 229 | 266 | 1073 | 383 |
| 25 | F | 374 | 1670 | 158 | 383 | 1676 | 178 | 346 | 1529 | 242 |
| 26 | F | 232 | 1486 | 161 | 295 | 1527 | 186 | 261 | 1352 | 309 |
| 27 | F | 317 | 1456 | 93 | 321 | 1324 | 144 | 280 | 1511 | 202 |
| F-Average | | 295 | 1450 | 127 | 290 | 1479 | 173 | 288 | 1382 | 231 |
| 1 | M | 315 | 1029 | 84 | 336 | 1115 | 121 | 326 | 1026 | 247 |
| 7 | M | 278 | 1279 | 116 | 272 | 1249 | 191 | 220 | 1262 | 187 |
| 10 | M | 242 | 877 | 123 | 159 | 818 | 151 | 174 | 898 | 135 |
| 11 | M | 244 | 1196 | 119 | 155 | 1281 | 156 | 236 | 1214 | 224 |
| 12 | M | 286 | 1287 | 127 | 284 | 1305 | 150 | 264 | 1316 | 245 |
| 19 | M | 321 | 1333 | 78 | 296 | 1312 | 149 | 317 | 1309 | 267 |
| 20 | M | 345 | 1138 | 108 | 353 | 1296 | 119 | 311 | 1082 | 248 |
| 22 | M | 289 | 1474 | 94 | 310 | 1511 | 162 | 270 | 862 | 285 |
| 28 | M | 300 | 1308 | 165 | 304 | 1345 | 177 | 296 | 1223 | 288 |
| 29 | M | 307 | 1440 | 110 | 334 | 1485 | 147 | 298 | 1372 | 244 |
| 31 | M | 314 | 1002 | 105 | 308 | 1085 | 150 | 300 | 1172 | 259 |
| 32 | M | 278 | 1318 | 118 | 281 | 1390 | 125 | 240 | 1372 | 273 |
| M-Average | | 293 | 1223 | 112 | 283 | 1266 | 150 | 271 | 1176 | 242 |

Appendix D—Chapter 3: Median F1 and F2 (Hz) and mean duration (ms) of vowels calculated from 4-6 tokens of the control word *who'd* (/hud/) in fast, medium, and slow speech. F1 and F2 were measured at the temporal midpoint of a vowel for each token. Data of subject #18 and #30 were excluded (*N* = 30: 18 females & 14 males).

| | | fast | | | medium | | | slow | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sbj. | Sex | F1 | F2 | Dur. | F1 | F2 | Dur. | F1 | F2 | Dur. |
| 2 | F | 219 | 1007 | 97 | 222 | 972 | 147 | 155 | 981 | 256 |
| 3 | F | 250 | 1824 | 109 | 350 | 1423 | 117 | 320 | 1340 | 299 |
| 4 | F | 255 | 1321 | 113 | 253 | 982 | 78 | 214 | 1271 | 165 |
| 5 | F | 315 | 1537 | 94 | 343 | 1425 | 138 | 283 | 1050 | 147 |
| 6 | F | 333 | 1328 | 118 | 195 | 1201 | 103 | 321 | 1138 | 163 |
| 8 | F | 391 | 1729 | 71 | 409 | 1808 | 120 | 337 | 1691 | 160 |
| 9 | F | 342 | 1450 | 118 | 322 | 1560 | 124 | 325 | 1510 | 118 |
| 13 | F | 296 | 1240 | 94 | 257 | 931 | 104 | 250 | 863 | 118 |
| 14 | F | 298 | 1694 | 87 | 295 | 1510 | 114 | 231 | 1476 | 186 |
| 15 | F | 278 | 1643 | 115 | 273 | 1484 | 142 | 221 | 1494 | 190 |
| 16 | F | 386 | 1563 | 121 | 393 | 1454 | 137 | 353 | 1321 | 184 |
| 17 | F | 339 | 1593 | 82 | 368 | 1670 | 103 | 350 | 1546 | 136 |
| 21 | F | 376 | 2047 | 123 | 437 | 2031 | 203 | 404 | 2005 | 212 |
| 23 | F | 292 | 1271 | 146 | 317 | 1189 | 228 | 331 | 1217 | 234 |
| 24 | F | 267 | 1557 | 118 | 294 | 1300 | 199 | 259 | 1087 | 344 |
| 25 | F | 381 | 1679 | 134 | 421 | 1564 | 155 | 366 | 1445 | 219 |
| 26 | F | 310 | 1968 | 112 | 311 | 1654 | 157 | 308 | 1586 | 256 |
| 27 | F | 304 | 1678 | 83 | 402 | 1986 | 84 | 285 | 1602 | 130 |
| F-Average | | 313 | 1563 | 108 | 326 | 1452 | 136 | 295 | 1368 | 195 |
| 1 | M | 334 | 1124 | 64 | 325 | 1135 | 106 | 332 | 904 | 195 |
| 7 | M | 224 | 1326 | 120 | 228 | 1089 | 183 | 231 | 1080 | 259 |
| 10 | M | 335 | 1373 | 91 | 331 | 1545 | 122 | 286 | 950 | 175 |
| 11 | M | 241 | 1412 | 92 | 192 | 1328 | 119 | 230 | 1228 | 209 |
| 12 | M | 281 | 1451 | 75 | 278 | 1145 | 126 | 267 | 1129 | 145 |
| 19 | M | 337 | 1071 | 76 | 309 | 1362 | 152 | 301 | 1403 | 199 |
| 20 | M | 313 | 1185 | 74 | 353 | 1357 | 101 | 345 | 991 | 183 |
| 22 | M | 300 | 1228 | 104 | 284 | 1200 | 152 | 270 | 986 | 263 |
| 28 | M | 328 | 1560 | 144 | 339 | 1109 | 228 | 289 | 1206 | 250 |
| 29 | M | 302 | 1164 | 78 | 316 | 1502 | 103 | 280 | 1269 | 204 |
| 31 | M | 342 | 1341 | 75 | 332 | 1130 | 123 | 297 | 1040 | 213 |
| 32 | M | 258 | 1124 | 77 | 294 | 1261 | 107 | 262 | 1319 | 220 |
| M-Average | | 300 | 1280 | 89 | 298 | 1264 | 135 | 283 | 1125 | 210 |

Appendix E—Chapter 5: F2 (Hz) in each of the four repeated vowels from /dVt/ stimuli (step #1-10). Data of subject #15 were excluded ($N$ = 29: 15 females & 14 males).

| | | Female Subject # | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| V | Rep. | 1 | 3 | 4 | 5 | 7 | 8 | 9 | 10 | 12 |
| 1 | 1 | 2956 | 2222 | 2927 | 2947 | 2444 | 2038 | 2914 | 2821 | 2124 |
| 1 | 2 | 3057 | 2434 | 2816 | 2962 | 2332 | 2481 | 2842 | 3042 | 2603 |
| 1 | 3 | 3103 | 2609 | 3066 | 2983 | 2256 | 2287 | 3020 | 2888 | 2478 |
| 1 | 4 | 3048 | 2386 | 3013 | 2942 | 2140 | 2548 | 2845 | 3148 | 2375 |
| 2 | 1 | 2842 | 2321 | 2713 | 3038 | 2197 | 2037 | 2975 | 2475 | 2292 |
| 2 | 2 | 2987 | 2423 | 2938 | 2970 | 2221 | 1946 | 2907 | 2582 | 2281 |
| 2 | 3 | 2974 | 2092 | 2533 | 2981 | 1986 | 2051 | 2920 | 2895 | 2482 |
| 2 | 4 | 3035 | 2524 | 2973 | 2944 | 2229 | 2020 | 2895 | 2302 | 2659 |
| 3 | 1 | 2852 | 2677 | 2708 | 2983 | 2157 | 2021 | 2872 | 2767 | 2552 |
| 3 | 2 | 2949 | 2310 | 2405 | 3122 | 2135 | 2041 | 2800 | 2415 | 2137 |
| 3 | 3 | 2875 | 2067 | 2581 | 2986 | 2208 | 1978 | 2834 | 2574 | 2600 |
| 3 | 4 | 2790 | 2131 | 2347 | 3013 | 2291 | 2071 | 2872 | 3073 | 2597 |
| 4 | 1 | 2652 | 2097 | 2305 | 2297 | 2102 | 2098 | 1863 | 2271 | 1698 |
| 4 | 2 | 2747 | 2070 | 2567 | 2464 | 2191 | 1882 | 2250 | 2507 | 1705 |
| 4 | 3 | 2736 | 2021 | 2445 | 2361 | 2161 | 1981 | 1895 | 2628 | 1695 |
| 4 | 4 | 2376 | 1981 | 2311 | 2431 | 2107 | 2119 | 1931 | 1897 | 2461 |
| 5 | 1 | 2715 | 2094 | 2218 | 1780 | 2200 | 1936 | 1969 | 2223 | 1727 |
| 5 | 2 | 2790 | 1996 | 2499 | 1916 | 2288 | 1961 | 1903 | 1545 | 1520 |
| 5 | 3 | 2267 | 1304 | 2411 | 2292 | 2301 | 2011 | 1755 | 1817 | 1730 |
| 5 | 4 | 3024 | 876 | 2427 | 2386 | 2201 | 1803 | 2007 | 1754 | 2442 |
| 6 | 1 | 2072 | 1435 | 2284 | 1830 | 1932 | 1569 | 2069 | 1541 | 1317 |
| 6 | 2 | 2271 | 1136 | 2212 | 1847 | 1718 | 1482 | 1939 | 1181 | 1279 |
| 6 | 3 | 1929 | 1407 | 2560 | 1904 | 1667 | 2011 | 1969 | 1397 | 1128 |
| 6 | 4 | 2435 | 1353 | 2356 | 1870 | 1691 | 1681 | 1882 | 1090 | 1148 |
| 7 | 1 | 1725 | 1166 | 2117 | 1893 | 1804 | 1598 | 1863 | 1286 | 1320 |
| 7 | 2 | 2274 | 795 | 1962 | 1906 | 1635 | 1530 | 1805 | 1609 | 1736 |
| 7 | 3 | 1951 | 812 | 1527 | 2000 | 1548 | 1331 | 1771 | 1392 | 1906 |
| 7 | 4 | 1969 | 950 | 1868 | 1916 | 1594 | 1432 | 1885 | 1349 | 1485 |
| 8 | 1 | 2110 | 1086 | 2025 | 1869 | 1497 | 1592 | 1527 | 1386 | 1524 |
| 8 | 2 | 1823 | 1145 | 1491 | 1481 | 1646 | 1563 | 1799 | 1246 | 1559 |
| 8 | 3 | 1708 | 1123 | 1800 | 1763 | 1715 | 1562 | 1872 | 1580 | 1406 |
| 8 | 4 | 1792 | 791 | 1804 | 1839 | 1779 | 1521 | 1675 | 1344 | 1299 |
| 9 | 1 | 1905 | 716 | 1041 | 1893 | 1543 | 1574 | 1382 | 1515 | 1259 |
| 9 | 2 | 1872 | 969 | 1107 | 1614 | 1677 | 1774 | 1485 | 1295 | 1141 |
| 9 | 3 | 1920 | 1062 | 2070 | 1872 | 1748 | 1289 | 1216 | 1164 | 1136 |
| 9 | 4 | 1808 | 1045 | 1583 | 1960 | 1667 | 1482 | 1171 | 1204 | 1313 |
| 10 | 1 | 1890 | 1014 | 1572 | 1942 | 1576 | 1656 | 1385 | 1532 | 1136 |
| 10 | 2 | 1832 | 1090 | 1847 | 1785 | 1685 | 1440 | 1102 | 1306 | 1195 |
| 10 | 3 | 1801 | 774 | 1696 | 1688 | 1415 | 1640 | 1173 | 1253 | 1299 |
| 10 | 4 | 1969 | 736 | 1975 | 1836 | 1870 | 1556 | 1118 | 1195 | 1222 |

Appendix E—Chapter 5 (Continued): F2 (Hz) in each of the four repeated vowels from /dVt/ stimuli (step #1-10). Data of subject #15 were excluded (*N* = 29: 15 females & 14 males).

| | | Female subject # | | | | | |
|---|---|---|---|---|---|---|---|
| V | Rep. | 13 | 17 | 18 | 24 | 27 | 30 |
| 1 | 1 | 3373 | 2149 | 1975 | 2740 | 2797 | 2847 |
| 1 | 2 | 3325 | 2116 | 2487 | 2844 | 2355 | 2741 |
| 1 | 3 | 3285 | 2574 | 2634 | 2541 | 2879 | 2749 |
| 1 | 4 | 3260 | 2646 | 2555 | 2457 | 2867 | 2850 |
| 2 | 1 | 3386 | 2151 | 2683 | 2727 | 2829 | 2792 |
| 2 | 2 | 3386 | 2411 | 2533 | 2727 | 2810 | 2744 |
| 2 | 3 | 3263 | 2554 | 2684 | 2710 | 2870 | 2810 |
| 2 | 4 | 3356 | 2156 | 2528 | 2467 | 2599 | 2428 |
| 3 | 1 | 3364 | 2180 | 2654 | 2694 | 2695 | 2855 |
| 3 | 2 | 3088 | 2159 | 2415 | 2625 | 2578 | 2446 |
| 3 | 3 | 3452 | 2627 | 2464 | 2518 | 2932 | 2984 |
| 3 | 4 | 3269 | 2613 | 2700 | 2521 | 2717 | 2262 |
| 4 | 1 | 1856 | 2736 | 2669 | 2500 | 2250 | 1859 |
| 4 | 2 | 1883 | 2109 | 2520 | 2571 | 2392 | 1964 |
| 4 | 3 | 1711 | 2572 | 2568 | 2487 | 2779 | 2156 |
| 4 | 4 | 2596 | 2630 | 2679 | 2491 | 2751 | 2410 |
| 5 | 1 | 1762 | 2027 | 1397 | 2371 | 2308 | 2087 |
| 5 | 2 | 1975 | 2058 | 2579 | 2121 | 2121 | 2145 |
| 5 | 3 | 1812 | 2018 | 1558 | 2569 | 2286 | 2406 |
| 5 | 4 | 1919 | 2607 | 1352 | 2289 | 2279 | 1893 |
| 6 | 1 | 1787 | 1403 | 1251 | 2193 | 1298 | 1523 |
| 6 | 2 | 1889 | 1530 | 1389 | 2263 | 1601 | 1501 |
| 6 | 3 | 1983 | 1542 | 1513 | 1878 | 1194 | 1524 |
| 6 | 4 | 1926 | 1372 | 1349 | 1906 | 1504 | 1750 |
| 7 | 1 | 1720 | 1355 | 1413 | 2022 | 1284 | 1595 |
| 7 | 2 | 1976 | 1515 | 1388 | 1710 | 1358 | 1395 |
| 7 | 3 | 1555 | 1493 | 1479 | 1555 | 1423 | 1335 |
| 7 | 4 | 2031 | 1415 | 1523 | 1883 | 1359 | 1663 |
| 8 | 1 | 1990 | 2391 | 1493 | 1363 | 1195 | 1305 |
| 8 | 2 | 1626 | 1219 | 1343 | 1742 | 1521 | 1187 |
| 8 | 3 | 1772 | 1271 | 1535 | 2047 | 1134 | 1185 |
| 8 | 4 | 1928 | 1397 | 1412 | 1664 | 1186 | 1616 |
| 9 | 1 | 1527 | 1420 | 1154 | 1665 | 1251 | 1023 |
| 9 | 2 | 2036 | 1526 | 1345 | 1495 | 1345 | 1205 |
| 9 | 3 | 1925 | 1592 | 1364 | 1748 | 1207 | 1051 |
| 9 | 4 | 1740 | 1548 | 1345 | 1497 | 1307 | 1091 |
| 10 | 1 | 2051 | 1293 | 1362 | 2000 | 1223 | 1304 |
| 10 | 2 | 2091 | 1355 | 1234 | 1901 | 1305 | 1226 |
| 10 | 3 | 1917 | 1379 | 1282 | 1418 | 1377 | 1251 |
| 10 | 4 | 1686 | 1477 | 1287 | 1471 | 1124 | 1152 |

Appendix E—Chapter 5 (Continued): F2 (Hz) in each of the four repeated vowels from /dVt/ stimuli (step #1-10). Data of subject #15 were excluded ($N$ = 29: 15 females & 14 males).

| | | Male Subject # | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| V | Rep. | 2 | 6 | 11 | 14 | 16 | 19 | 20 | 21 | 22 |
| 1 | 1 | 2034 | 2388 | 2055 | 2256 | 2356 | 2269 | 2439 | 2198 | 1783 |
| 1 | 2 | 1824 | 2352 | 2218 | 2182 | 2337 | 2274 | 2450 | 2180 | 2226 |
| 1 | 3 | 2065 | 2396 | 2195 | 2144 | 2196 | 2171 | 2377 | 2117 | 2064 |
| 1 | 4 | 1935 | 2398 | 2118 | 2121 | 2435 | 2145 | 2517 | 2085 | 2347 |
| 2 | 1 | 1969 | 2343 | 2026 | 2188 | 2206 | 2374 | 2163 | 2164 | 1661 |
| 2 | 2 | 1875 | 2356 | 2093 | 2029 | 2289 | 2241 | 2410 | 2188 | 1863 |
| 2 | 3 | 1970 | 2358 | 2204 | 2169 | 2243 | 2271 | 2423 | 2239 | 2358 |
| 2 | 4 | 1916 | 2391 | 2173 | 2166 | 2294 | 2098 | 2462 | 2130 | 1709 |
| 3 | 1 | 2086 | 2318 | 2170 | 1864 | 2385 | 2310 | 2467 | 2200 | 1856 |
| 3 | 2 | 1942 | 2333 | 2103 | 1953 | 2259 | 2117 | 2318 | 2105 | 2134 |
| 3 | 3 | 1760 | 2228 | 2098 | 2064 | 2123 | 2153 | 2421 | 2056 | 2231 |
| 3 | 4 | 1920 | 2021 | 2103 | 2171 | 2436 | 2155 | 2456 | 2041 | 2274 |
| 4 | 1 | 1842 | 2103 | 1960 | 1878 | 2347 | 2255 | 1600 | 1935 | 1725 |
| 4 | 2 | 1536 | 2113 | 2116 | 1899 | 1974 | 2173 | 2390 | 2180 | 1881 |
| 4 | 3 | 1812 | 2054 | 1932 | 1962 | 2198 | 1265 | 2350 | 2127 | 1963 |
| 4 | 4 | 1636 | 1885 | 1811 | 1853 | 2209 | 1261 | 2297 | 1020 | 2296 |
| 5 | 1 | 1205 | 1760 | 2091 | 1780 | 1970 | 1182 | 2145 | 2241 | 1432 |
| 5 | 2 | 1222 | 1558 | 1800 | 1945 | 2175 | 1321 | 1721 | 977 | 1779 |
| 5 | 3 | 1690 | 1793 | 1912 | 1853 | 2089 | 1204 | 1665 | 1121 | 1399 |
| 5 | 4 | 1045 | 1836 | 2110 | 1795 | 2176 | 1256 | 1630 | 2236 | 1113 |
| 6 | 1 | 1101 | 1404 | 1114 | 1596 | 1739 | 1215 | 1561 | 1428 | 1300 |
| 6 | 2 | 1113 | 1437 | 1688 | 1590 | 1844 | 1250 | 1469 | 1053 | 1246 |
| 6 | 3 | 1208 | 1473 | 1164 | 1447 | 1665 | 1207 | 1622 | 804 | 1160 |
| 6 | 4 | 999 | 1245 | 1262 | 1852 | 1855 | 1306 | 1578 | 1185 | 1165 |
| 7 | 1 | 1331 | 1404 | 1673 | 1381 | 1748 | 1190 | 1426 | 1267 | 1137 |
| 7 | 2 | 1373 | 1375 | 1222 | 1656 | 1880 | 1219 | 1609 | 996 | 1172 |
| 7 | 3 | 1209 | 1235 | 1038 | 1516 | 1467 | 1177 | 1538 | 862 | 1172 |
| 7 | 4 | 1218 | 1065 | 893 | 1585 | 1752 | 1301 | 1560 | 1190 | 1130 |
| 8 | 1 | 1316 | 960 | 933 | 1332 | 1813 | 1287 | 1589 | 1158 | 1274 |
| 8 | 2 | 1208 | 1031 | 1122 | 1399 | 1793 | 1271 | 1457 | 1073 | 1221 |
| 8 | 3 | 1153 | 765 | 1749 | 1442 | 1801 | 1153 | 1388 | 927 | 1236 |
| 8 | 4 | 1226 | 1179 | 1592 | 1479 | 1710 | 1175 | 1350 | 1093 | 1192 |
| 9 | 1 | 1275 | 1100 | 933 | 1374 | 1747 | 1167 | 1576 | 1084 | 1078 |
| 9 | 2 | 1111 | 870 | 1237 | 1399 | 1805 | 1105 | 1485 | 905 | 1100 |
| 9 | 3 | 1291 | 679 | 1037 | 1568 | 1776 | 1131 | 1489 | 976 | 1235 |
| 9 | 4 | 1169 | 742 | 1112 | 1509 | 1737 | 1376 | 1510 | 938 | 1080 |
| 10 | 1 | 1333 | 1221 | 1634 | 1171 | 1753 | 1144 | 1589 | 1073 | 1297 |
| 10 | 2 | 1239 | 770 | 1092 | 1451 | 1749 | 1191 | 1518 | 1135 | 1154 |
| 10 | 3 | 1027 | 862 | 1538 | 1613 | 1613 | 1201 | 1607 | 1283 | 1138 |
| 10 | 4 | 1119 | 779 | 960 | 1125 | 1767 | 1210 | 1559 | 931 | 1116 |

Appendix E—Chapter 5 (Continued): F2 (Hz) in each of the four repeated vowels from /dVt/ stimuli (step #1-10). Data of subject #15 were excluded ($N$ = 29: 15 females & 14 males).

| | | Male Subject # | | | | |
|---|---|---|---|---|---|---|
| V | Rep. | 23 | 25 | 26 | 28 | 29 |
| 1 | 1 | 2669 | 2424 | 2655 | 1815 | 2148 |
| 1 | 2 | 2270 | 2352 | 2657 | 1870 | 2129 |
| 1 | 3 | 2351 | 2520 | 2616 | 1789 | 1965 |
| 1 | 4 | 2341 | 2539 | 2649 | 2046 | 2111 |
| 2 | 1 | 2453 | 2360 | 2656 | 1890 | 2132 |
| 2 | 2 | 2331 | 2519 | 2532 | 1730 | 2165 |
| 2 | 3 | 2187 | 2368 | 2625 | 1831 | 1906 |
| 2 | 4 | 2239 | 2416 | 2264 | 1740 | 2092 |
| 3 | 1 | 2210 | 2123 | 2236 | 1673 | 2009 |
| 3 | 2 | 2128 | 2213 | 2349 | 1695 | 1986 |
| 3 | 3 | 2189 | 2288 | 2474 | 1649 | 1737 |
| 3 | 4 | 2353 | 2315 | 2320 | 1873 | 1985 |
| 4 | 1 | 2077 | 2158 | 2311 | 1764 | 1767 |
| 4 | 2 | 2162 | 2233 | 1767 | 1752 | 1942 |
| 4 | 3 | 2183 | 2084 | 2170 | 1723 | 1828 |
| 4 | 4 | 2100 | 2387 | 1770 | 1807 | 1747 |
| 5 | 1 | 2167 | 1762 | 1539 | 1747 | 1593 |
| 5 | 2 | 2008 | 2069 | 1454 | 1596 | 1793 |
| 5 | 3 | 2067 | 2111 | 1477 | 1548 | 2021 |
| 5 | 4 | 2056 | 2103 | 1334 | 1584 | 2049 |
| 6 | 1 | 1909 | 1387 | 1584 | 1541 | 1663 |
| 6 | 2 | 2111 | 1382 | 1484 | 1426 | 1708 |
| 6 | 3 | 1703 | 1506 | 1654 | 1531 | 1411 |
| 6 | 4 | 1946 | 1037 | 1479 | 1547 | 1507 |
| 7 | 1 | 1642 | 1136 | 1446 | 1591 | 1577 |
| 7 | 2 | 1400 | 1119 | 1374 | 926 | 1590 |
| 7 | 3 | 1713 | 958 | 1271 | 1551 | 1392 |
| 7 | 4 | 1910 | 932 | 1423 | 1524 | 1446 |
| 8 | 1 | 1380 | 1154 | 1305 | 1554 | 1597 |
| 8 | 2 | 1418 | 1093 | 1518 | 1016 | 1148 |
| 8 | 3 | 1434 | 1047 | 1364 | 1532 | 1489 |
| 8 | 4 | 1166 | 967 | 1468 | 868 | 1167 |
| 9 | 1 | 1631 | 1396 | 1353 | 1532 | 1797 |
| 9 | 2 | 1609 | 1045 | 1344 | 1473 | 1194 |
| 9 | 3 | 1099 | 1045 | 1228 | 1117 | 1321 |
| 9 | 4 | 1044 | 965 | 1214 | 1489 | 1347 |
| 10 | 1 | 1249 | 946 | 1448 | 1512 | 1520 |
| 10 | 2 | 1508 | 1077 | 1394 | 967 | 1635 |
| 10 | 3 | 1464 | 1342 | 1484 | 934 | 1474 |
| 10 | 4 | 983 | 1062 | 1542 | 1114 | 1469 |

Appendix F—Chapter 5: F2 (Hz) in each of the four repeated vowels from /bVp/ stimuli (step #1-10). Data of subject #15 were excluded ($N$ = 29: 15 females & 14 males).

| | | Female Subject # | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| V | Rep. | 1 | 3 | 4 | 5 | 7 | 8 | 9 | 10 | 12 |
| 1 | 1 | 2842 | 2322 | 3050 | 2999 | 2168 | 2384 | 2948 | 2508 | 2483 |
| 1 | 2 | 3015 | 2475 | 2615 | 2280 | 2116 | 2570 | 2929 | 2287 | 2366 |
| 1 | 3 | 2756 | 2308 | 2971 | 2809 | 2098 | 2478 | 2912 | 2925 | 2594 |
| 1 | 4 | 2872 | 2184 | 2953 | 2885 | 2426 | 2550 | 2950 | 2863 | 2633 |
| 2 | 1 | 2886 | 2172 | 2806 | 3036 | 2317 | 2551 | 2913 | 2737 | 2580 |
| 2 | 2 | 2800 | 2448 | 2664 | 2948 | 1802 | 2069 | 3006 | 2390 | 2540 |
| 2 | 3 | 2830 | 2452 | 3132 | 2938 | 2344 | 2227 | 2969 | 2393 | 2669 |
| 2 | 4 | 2898 | 2529 | 2692 | 2906 | 2107 | 2022 | 2859 | 2713 | 2633 |
| 3 | 1 | 2596 | 2621 | 2579 | 2977 | 2075 | 2002 | 2919 | 2397 | 2381 |
| 3 | 2 | 2689 | 2479 | 2780 | 2941 | 2148 | 1923 | 2889 | 2785 | 2375 |
| 3 | 3 | 2861 | 2337 | 2665 | 2851 | 2243 | 2076 | 2920 | 2752 | 2608 |
| 3 | 4 | 2945 | 2336 | 2669 | 2953 | 2196 | 2064 | 2858 | 2772 | 2627 |
| 4 | 1 | 2756 | 2617 | 2372 | 2886 | 2281 | 2035 | 2918 | 1577 | 1868 |
| 4 | 2 | 2906 | 2055 | 2485 | 2411 | 2089 | 1947 | 2843 | 2501 | 1786 |
| 4 | 3 | 2764 | 2365 | 2455 | 2538 | 1703 | 1914 | 2911 | 2599 | 2016 |
| 4 | 4 | 2587 | 1904 | 2283 | 2425 | 2126 | 2028 | 2939 | 2494 | 1881 |
| 5 | 1 | 2500 | 1824 | 2066 | 2318 | 2048 | 2428 | 2889 | 2325 | 2075 |
| 5 | 2 | 2149 | 2058 | 2194 | 2256 | 1798 | 2011 | 2948 | 1427 | 1903 |
| 5 | 3 | 2432 | 1193 | 2435 | 2409 | 2290 | 1916 | 2915 | 2146 | 1312 |
| 5 | 4 | 2151 | 901 | 2204 | 2339 | 2234 | 2029 | 1888 | 2318 | 2660 |
| 6 | 1 | 2006 | 2156 | 2123 | 1740 | 1726 | 1748 | 1912 | 2237 | 1720 |
| 6 | 2 | 2110 | 760 | 2096 | 1842 | 2073 | 1856 | 1858 | 1408 | 1307 |
| 6 | 3 | 1978 | 2047 | 2204 | 1906 | 1523 | 1726 | 1980 | 1348 | 2183 |
| 6 | 4 | 2737 | 1360 | 1782 | 1862 | 1621 | 1730 | 1801 | 1444 | 2519 |
| 7 | 1 | 1878 | 1614 | 1839 | 2051 | 1485 | 1724 | 1837 | 1570 | 1483 |
| 7 | 2 | 1952 | 1467 | 1884 | 1882 | 2037 | 2019 | 1967 | 1433 | 1332 |
| 7 | 3 | 1870 | 925 | 2234 | 1901 | 1621 | 1513 | 1934 | 1162 | 1117 |
| 7 | 4 | 2182 | 1186 | 1449 | 1893 | 2062 | 1929 | 1844 | 1404 | 1306 |
| 8 | 1 | 1948 | 985 | 1213 | 1834 | 1555 | 1596 | 1755 | 1361 | 1359 |
| 8 | 2 | 2115 | 1520 | 1563 | 1856 | 1362 | 1479 | 1673 | 1358 | 1393 |
| 8 | 3 | 2233 | 799 | 1882 | 1811 | 1836 | 1960 | 1900 | 1190 | 1364 |
| 8 | 4 | 1706 | 1411 | 1549 | 1847 | 1529 | 1588 | 1696 | 1086 | 1288 |
| 9 | 1 | 1686 | 1488 | 1550 | 1840 | 1641 | 1561 | 1834 | 1265 | 1334 |
| 9 | 2 | 1770 | 1231 | 2041 | 1993 | 1609 | 1265 | 1648 | 1218 | 1242 |
| 9 | 3 | 1842 | 1280 | 2256 | 1889 | 1558 | 1978 | 1779 | 1353 | 1188 |
| 9 | 4 | 2129 | 1343 | 1094 | 1797 | 1559 | 1771 | 1736 | 1191 | 989 |
| 10 | 1 | 2037 | 1129 | 1361 | 1854 | 1263 | 1726 | 1418 | 1474 | 834 |
| 10 | 2 | 1789 | 1469 | 1980 | 1888 | 1746 | 1164 | 1754 | 1229 | 1398 |
| 10 | 3 | 1630 | 825 | 1939 | 1861 | 1440 | 1569 | 1893 | 1318 | 881 |
| 10 | 4 | 1736 | 664 | 1557 | 1968 | 1607 | 1545 | 1789 | 1409 | 1306 |

Appendix F—Chapter 5 (Continued): F2 (Hz) in each of the four repeated vowels from /bVp/ stimuli (step #1-10). Data of subject #15 were excluded ($N$ = 29: 15 females & 14 males).

| | | Female subject # | | | | | |
|---|---|---|---|---|---|---|---|
| V | Rep. | 13 | 17 | 18 | 24 | 27 | 30 |
| 1 | 1 | 3422 | 2440 | 2775 | 2844 | 2795 | 2691 |
| 1 | 2 | 3223 | 2116 | 2514 | 2802 | 2892 | 2687 |
| 1 | 3 | 3352 | 2461 | 1939 | 2734 | 2858 | 2865 |
| 1 | 4 | 3314 | 2444 | 2642 | 2520 | 2911 | 2840 |
| 2 | 1 | 3352 | 2559 | 2795 | 2740 | 2745 | 2709 |
| 2 | 2 | 3212 | 2616 | 2679 | 2580 | 2860 | 2315 |
| 2 | 3 | 3292 | 1709 | 2654 | 2578 | 2829 | 2738 |
| 2 | 4 | 3271 | 2593 | 2624 | 2580 | 2420 | 2848 |
| 3 | 1 | 3340 | 2721 | 2640 | 2738 | 2742 | 2890 |
| 3 | 2 | 3289 | 2365 | 2786 | 2571 | 2483 | 2836 |
| 3 | 3 | 3247 | 2582 | 2640 | 2578 | 2828 | 2792 |
| 3 | 4 | 3144 | 2585 | 2596 | 2719 | 2223 | 2787 |
| 4 | 1 | 3059 | 2094 | 1359 | 2551 | 2680 | 2748 |
| 4 | 2 | 3143 | 1577 | 2577 | 2687 | 2527 | 2375 |
| 4 | 3 | 3316 | 1587 | 2619 | 2486 | 1401 | 2552 |
| 4 | 4 | 2664 | 1362 | 2587 | 2379 | 2589 | 2770 |
| 5 | 1 | 1999 | 2323 | 1201 | 2352 | 2846 | 2465 |
| 5 | 2 | 2219 | 1382 | 2689 | 2379 | 2325 | 2792 |
| 5 | 3 | 1923 | 1213 | 1448 | 2619 | 2523 | 2482 |
| 5 | 4 | 2067 | 1438 | 1817 | 2548 | 2119 | 2046 |
| 6 | 1 | 2058 | 1545 | 1285 | 1733 | 2747 | 2444 |
| 6 | 2 | 1954 | 1565 | 1439 | 2007 | 2588 | 1803 |
| 6 | 3 | 1692 | 1262 | 1368 | 1785 | 2743 | 1519 |
| 6 | 4 | 2161 | 1610 | 1436 | 1541 | 2602 | 2098 |
| 7 | 1 | 1704 | 1072 | 1394 | 1816 | 1112 | 1496 |
| 7 | 2 | 1515 | 1161 | 1515 | 2015 | 1310 | 1638 |
| 7 | 3 | 1989 | 1409 | 1195 | 1490 | 1375 | 1668 |
| 7 | 4 | 1947 | 1336 | 1490 | 1539 | 1515 | 945 |
| 8 | 1 | 1520 | 1456 | 1375 | 1506 | 1450 | 1110 |
| 8 | 2 | 2270 | 1510 | 1390 | 1980 | 1219 | 972 |
| 8 | 3 | 1800 | 1444 | 1438 | 1455 | 1518 | 1394 |
| 8 | 4 | 1900 | 1318 | 1338 | 1513 | 1346 | 1025 |
| 9 | 1 | 1868 | 1393 | 1411 | 1776 | 1143 | 1277 |
| 9 | 2 | 1390 | 1337 | 1333 | 1817 | 1293 | 1077 |
| 9 | 3 | 2068 | 1474 | 1388 | 1394 | 1392 | 1585 |
| 9 | 4 | 1893 | 1351 | 1042 | 1946 | 1397 | 1667 |
| 10 | 1 | 1930 | 1229 | 1398 | 1417 | 1390 | 1146 |
| 10 | 2 | 2135 | 1486 | 1390 | 1463 | 1345 | 1159 |
| 10 | 3 | 1801 | 1674 | 1442 | 1487 | 1291 | 1466 |
| 10 | 4 | 1802 | 1297 | 1650 | 1455 | 1248 | 1136 |

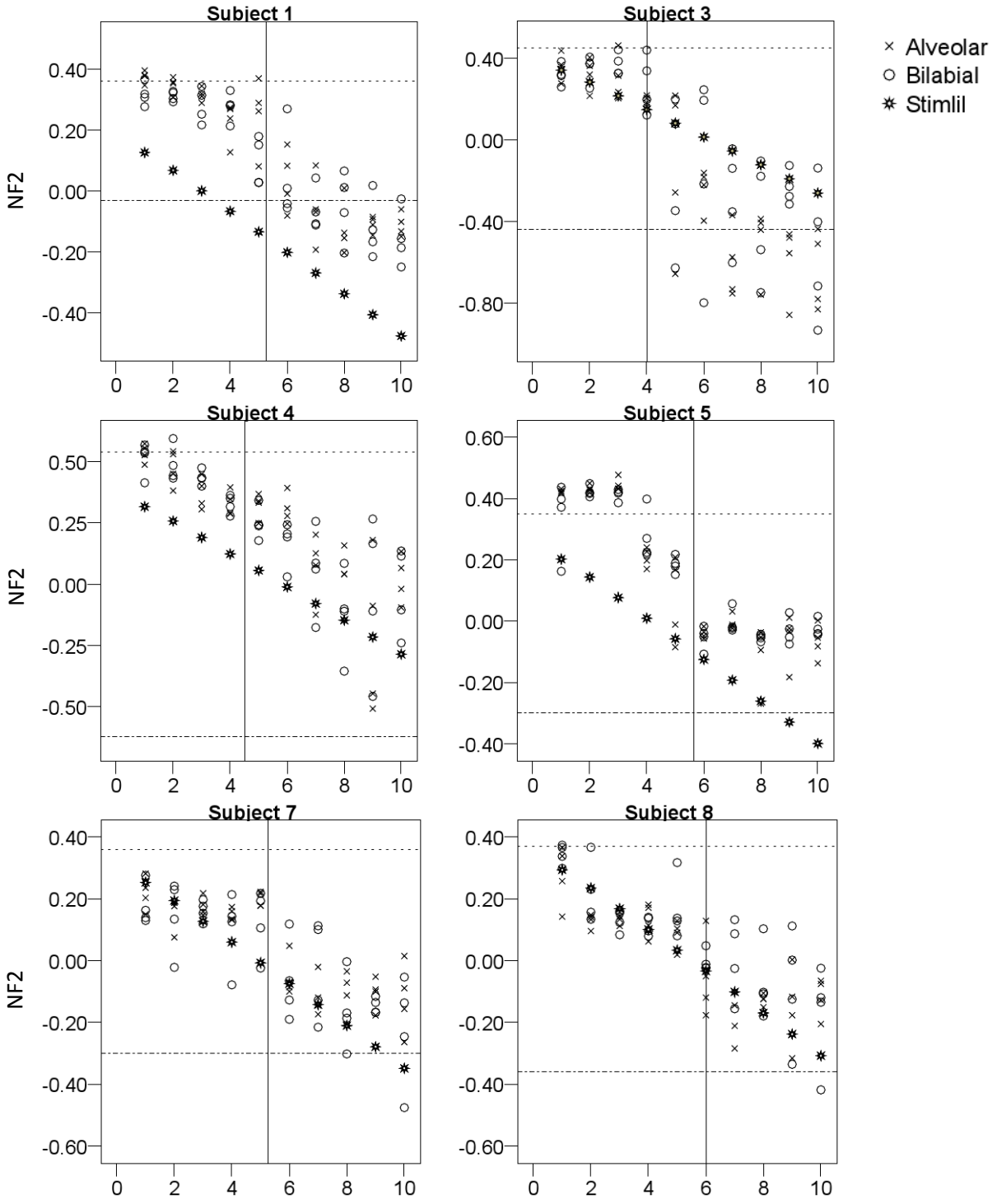
Appendix F—Chapter 5 (Continued): F2 (Hz) in each of the four repeated vowels from /bVp/ stimuli (step #1-10). Data of subject #15 were excluded ($N$ = 29: 15 females & 14 males).

| | | Male Subject # | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| V | Rep. | 2 | 6 | 11 | 14 | 16 | 19 | 20 | 21 | 22 |
| 1 | 1 | 2146 | 2462 | 2236 | 2300 | 2417 | 2366 | 2406 | 2227 | 2072 |
| 1 | 2 | 2229 | 2400 | 2181 | 2298 | 2466 | 2342 | 2436 | 2172 | 2195 |
| 1 | 3 | 2256 | 2328 | 2172 | 2126 | 2319 | 2219 | 2438 | 2192 | 2279 |
| 1 | 4 | 2115 | 2417 | 2175 | 2259 | 2346 | 2176 | 2449 | 1978 | 2430 |
| 2 | 1 | 2155 | 2341 | 2165 | 2129 | 2348 | 2270 | 2456 | 2217 | 2328 |
| 2 | 2 | 2029 | 2403 | 2170 | 2366 | 2193 | 2233 | 2438 | 2252 | 1960 |
| 2 | 3 | 2072 | 2355 | 2157 | 2178 | 2423 | 2182 | 2433 | 2110 | 2250 |
| 2 | 4 | 2075 | 2270 | 2103 | 2219 | 2318 | 2205 | 2416 | 2093 | 2044 |
| 3 | 1 | 2006 | 2342 | 2095 | 1902 | 2266 | 2273 | 2476 | 2201 | 2008 |
| 3 | 2 | 2109 | 2287 | 2120 | 2253 | 2349 | 2164 | 2434 | 2199 | 2254 |
| 3 | 3 | 2166 | 2335 | 2130 | 2200 | 2476 | 2071 | 2457 | 2111 | 2258 |
| 3 | 4 | 2084 | 2475 | 2137 | 2204 | 2192 | 2296 | 2408 | 2171 | 2113 |
| 4 | 1 | 1974 | 2044 | 1894 | 1920 | 1830 | 2262 | 2279 | 2210 | 1917 |
| 4 | 2 | 1865 | 1705 | 1937 | 2023 | 1812 | 2161 | 2279 | 2188 | 1158 |
| 4 | 3 | 1718 | 1707 | 2111 | 1971 | 2116 | 2274 | 2402 | 2059 | 1786 |
| 4 | 4 | 1896 | 1965 | 1497 | 1842 | 2321 | 2146 | 2244 | 1353 | 1559 |
| 5 | 1 | 1976 | 1928 | 2015 | 1880 | 1845 | 1287 | 1555 | 911 | 1873 |
| 5 | 2 | 1490 | 1782 | 1936 | 1985 | 1755 | 1217 | 1475 | 855 | 1263 |
| 5 | 3 | 1511 | 1943 | 1969 | 1885 | 1852 | 1331 | 1608 | 1021 | 1219 |
| 5 | 4 | 1895 | 1805 | 2055 | 1812 | 1994 | 1270 | 1629 | 1022 | 2122 |
| 6 | 1 | 1614 | 1590 | 1533 | 1695 | 1832 | 1344 | 1612 | 980 | 1297 |
| 6 | 2 | 1466 | 1584 | 2150 | 1710 | 1756 | 1290 | 1593 | 889 | 1409 |
| 6 | 3 | 1660 | 1578 | 1247 | 1656 | 1893 | 1228 | 1505 | 827 | 1167 |
| 6 | 4 | 1273 | 1411 | 1421 | 1720 | 1710 | 1332 | 1423 | 1339 | 1290 |
| 7 | 1 | 1226 | 1384 | 1673 | 1398 | 1977 | 1217 | 1512 | 2182 | 1347 |
| 7 | 2 | 1173 | 1336 | 1146 | 1381 | 1937 | 1294 | 1590 | 1276 | 1353 |
| 7 | 3 | 1282 | 1396 | 1824 | 1565 | 1897 | 1149 | 1529 | 1098 | 1322 |
| 7 | 4 | 1191 | 1309 | 1361 | 1492 | 1829 | 1164 | 1666 | 904 | 1553 |
| 8 | 1 | 1315 | 1339 | 1421 | 1244 | 1965 | 1187 | 1603 | 880 | 1380 |
| 8 | 2 | 1173 | 1277 | 1020 | 1118 | 1799 | 1170 | 1423 | 1198 | 1075 |
| 8 | 3 | 1156 | 1109 | 1177 | 1365 | 1874 | 1123 | 1379 | 979 | 1261 |
| 8 | 4 | 1272 | 1394 | 1000 | 1570 | 1786 | 1411 | 1320 | 935 | 1154 |
| 9 | 1 | 1310 | 1359 | 1073 | 1380 | 1809 | 1273 | 1398 | 991 | 1320 |
| 9 | 2 | 1262 | 1421 | 925 | 1086 | 1927 | 1199 | 1447 | 1082 | 1219 |
| 9 | 3 | 1275 | 1413 | 1030 | 1533 | 1595 | 1223 | 1633 | 1056 | 1239 |
| 9 | 4 | 1125 | 1345 | 1495 | 1514 | 1806 | 1174 | 1241 | 961 | 1189 |
| 10 | 1 | 1143 | 859 | 1327 | 1149 | 1961 | 1188 | 1550 | 1055 | 1247 |
| 10 | 2 | 1148 | 1274 | 1187 | 1239 | 1809 | 1153 | 1594 | 912 | 1104 |
| 10 | 3 | 1206 | 1342 | 974 | 1215 | 1718 | 1293 | 1359 | 1066 | 1286 |
| 10 | 4 | 1055 | 1363 | 1013 | 1573 | 1411 | 1198 | 1441 | 1085 | 1265 |

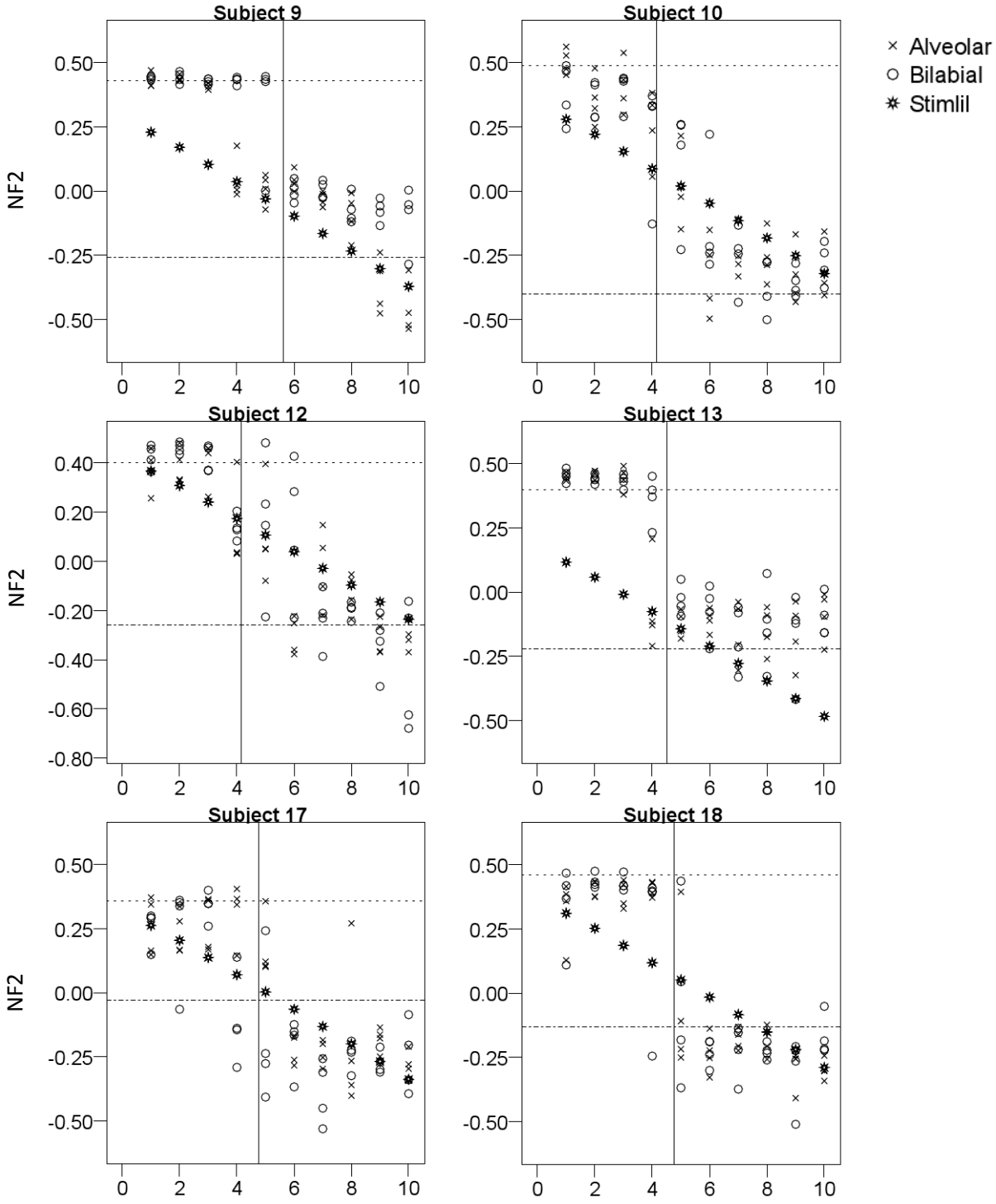
Appendix F—Chapter 5 (Continued): F2 (Hz) in each of the four repeated vowels from /bVp/ stimuli (step #1-10). Data of subject #15 were excluded ($N$ = 29: 15 females & 14 males).

| | | Male Subject # | | | | |
|---|---|---|---|---|---|---|
| V | Rep. | 23 | 25 | 26 | 28 | 29 |
| 1 | 1 | 2242 | 2426 | 2598 | 1870 | 2223 |
| 1 | 2 | 2504 | 2473 | 2651 | 1863 | 2189 |
| 1 | 3 | 2374 | 2645 | 2606 | 1860 | 2148 |
| 1 | 4 | 2127 | 2363 | 2602 | 1983 | 2123 |
| 2 | 1 | 2273 | 2458 | 2612 | 1801 | 2122 |
| 2 | 2 | 2330 | 2509 | 2478 | 1892 | 2198 |
| 2 | 3 | 2097 | 2705 | 2595 | 1957 | 2164 |
| 2 | 4 | 2276 | 2658 | 2620 | 1875 | 2118 |
| 3 | 1 | 2087 | 2312 | 2607 | 1703 | 2047 |
| 3 | 2 | 2275 | 1982 | 2337 | 2049 | 2116 |
| 3 | 3 | 2185 | 2491 | 2542 | 1858 | 2076 |
| 3 | 4 | 2155 | 2146 | 2567 | 1773 | 1802 |
| 4 | 1 | 2218 | 2232 | 1867 | 2000 | 1900 |
| 4 | 2 | 2087 | 2081 | 2026 | 1758 | 1810 |
| 4 | 3 | 2168 | 1941 | 1515 | 1812 | 2005 |
| 4 | 4 | 2134 | 2592 | 2434 | 1842 | 1916 |
| 5 | 1 | 2239 | 2330 | 2361 | 1776 | 1841 |
| 5 | 2 | 2162 | 2180 | 1652 | 1860 | 1898 |
| 5 | 3 | 1908 | 1941 | 1624 | 1589 | 2100 |
| 5 | 4 | 2098 | 1940 | 1522 | 1709 | 1864 |
| 6 | 1 | 2039 | 1822 | 1532 | 1564 | 1817 |
| 6 | 2 | 2066 | 1588 | 1602 | 1554 | 1874 |
| 6 | 3 | 2015 | 2123 | 1523 | 1520 | 1826 |
| 6 | 4 | 2077 | 1612 | 1459 | 1587 | 1784 |
| 7 | 1 | 2070 | 1225 | 1416 | 1390 | 1403 |
| 7 | 2 | 1797 | 1638 | 1677 | 1579 | 1237 |
| 7 | 3 | 2065 | 1273 | 1573 | 1545 | 1571 |
| 7 | 4 | 1833 | 1253 | 1392 | 1548 | 1542 |
| 8 | 1 | 1563 | 928 | 1514 | 1189 | 1368 |
| 8 | 2 | 2060 | 863 | 1277 | 1438 | 1221 |
| 8 | 3 | 1970 | 1010 | 1579 | 916 | 1739 |
| 8 | 4 | 1718 | 1052 | 1408 | 1492 | 1244 |
| 9 | 1 | 1639 | 843 | 1374 | 1359 | 1362 |
| 9 | 2 | 1508 | 1093 | 1164 | 1022 | 1463 |
| 9 | 3 | 1528 | 1078 | 1391 | 1025 | 1237 |
| 9 | 4 | 1820 | 1077 | 1236 | 1401 | 1257 |
| 10 | 1 | 1502 | 876 | 1556 | 1002 | 1152 |
| 10 | 2 | 1538 | 992 | 1532 | 964 | 1116 |
| 10 | 3 | 1432 | 1157 | 1277 | 1055 | 1233 |
| 10 | 4 | 2217 | 853 | 1424 | 990 | 1620 |

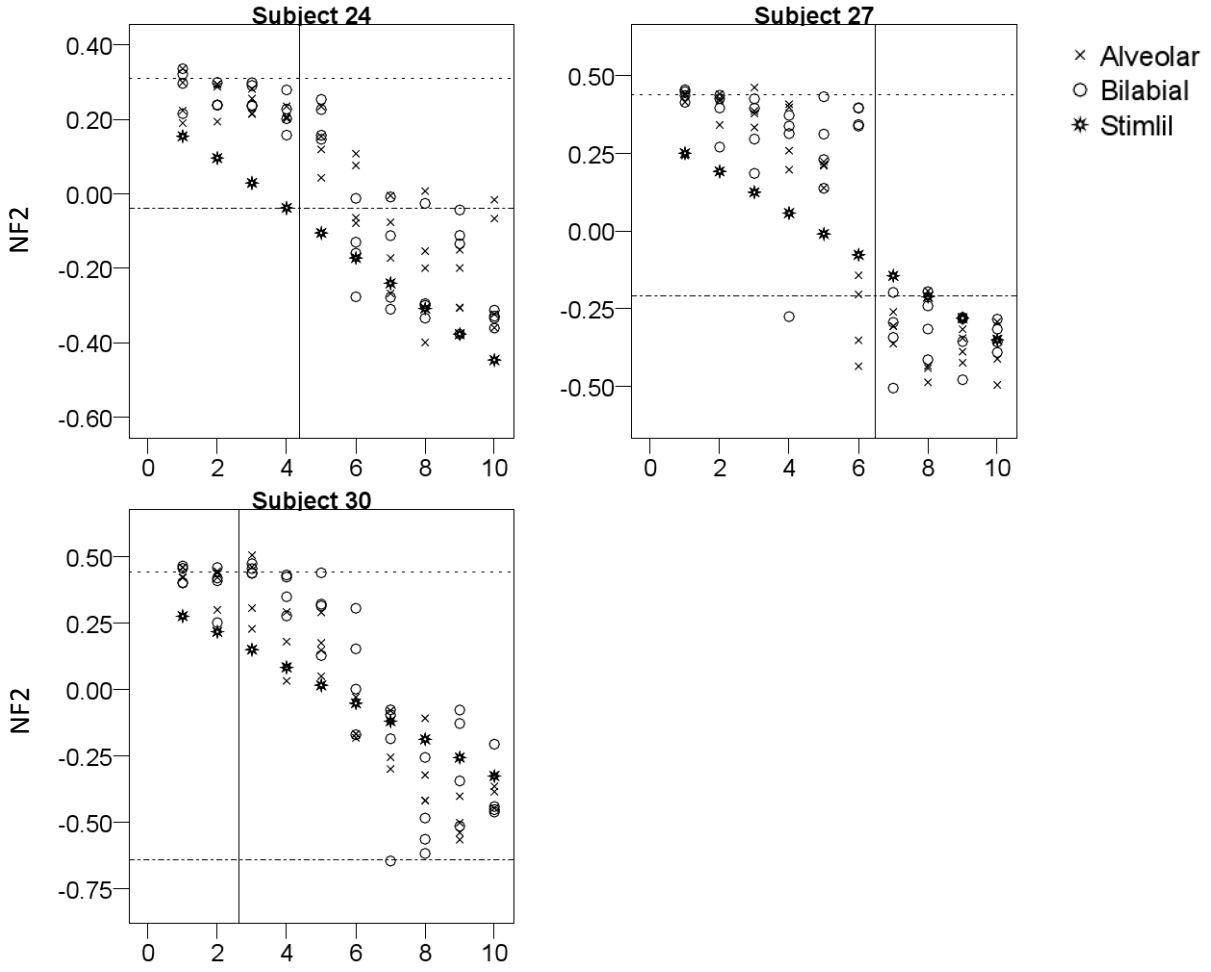
Appendix G—Chapter 5: Results from fifteen female subjects: NF2 of response vowels as a function of stimulus number (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli (a uniform set of stimulus F2 was transformed into unique set of NF2 for each subject). A vertical line indicates Boundary, and two horizontal dotted lines indicate NF2 in *heed* (upper) and NF2 in *who'd* (lower) for each subject.



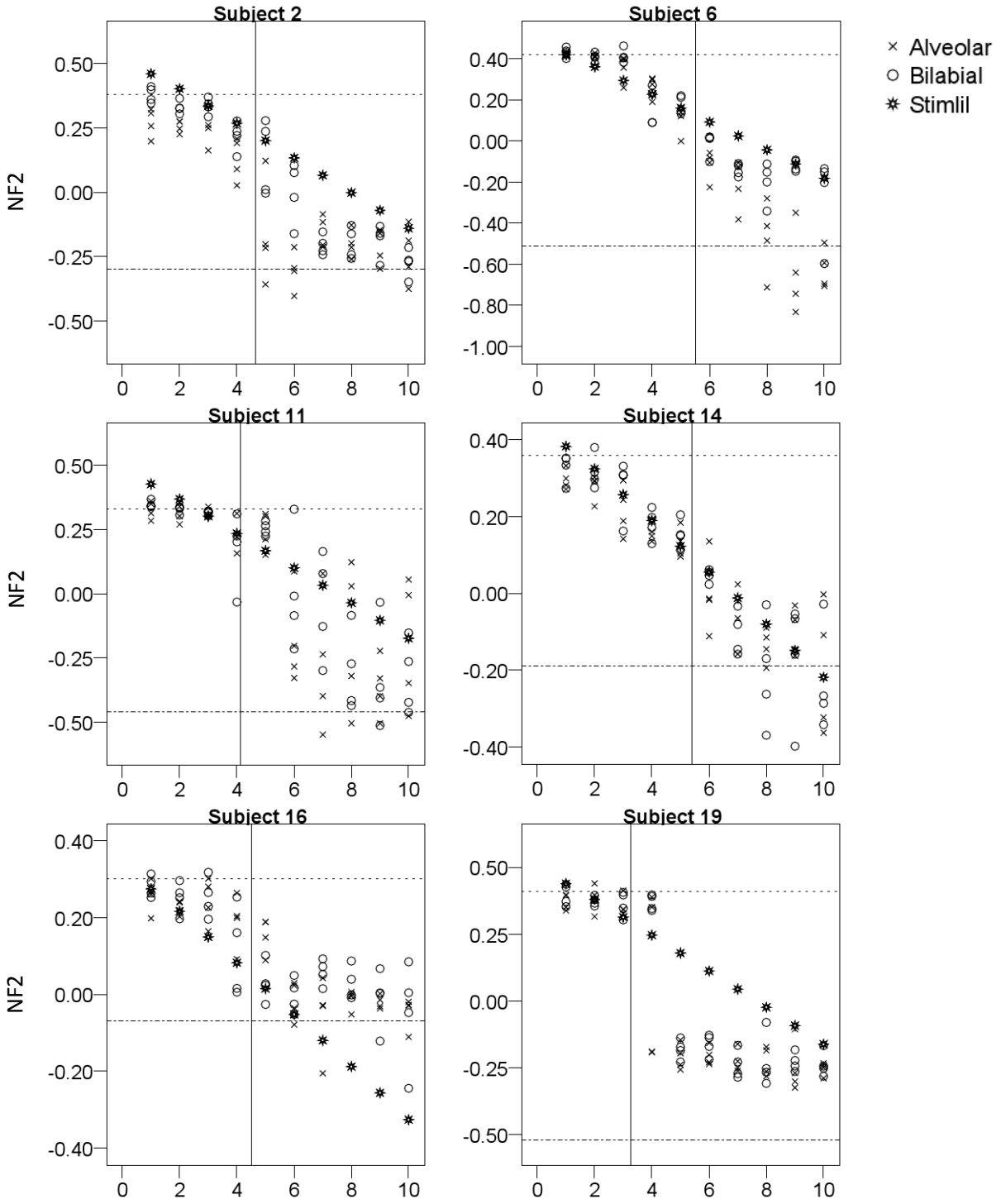
Appendix G—Chapter 5 (Continued): Results from fifteen female subjects: NF2 of response vowels as a function of stimulus number (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli.



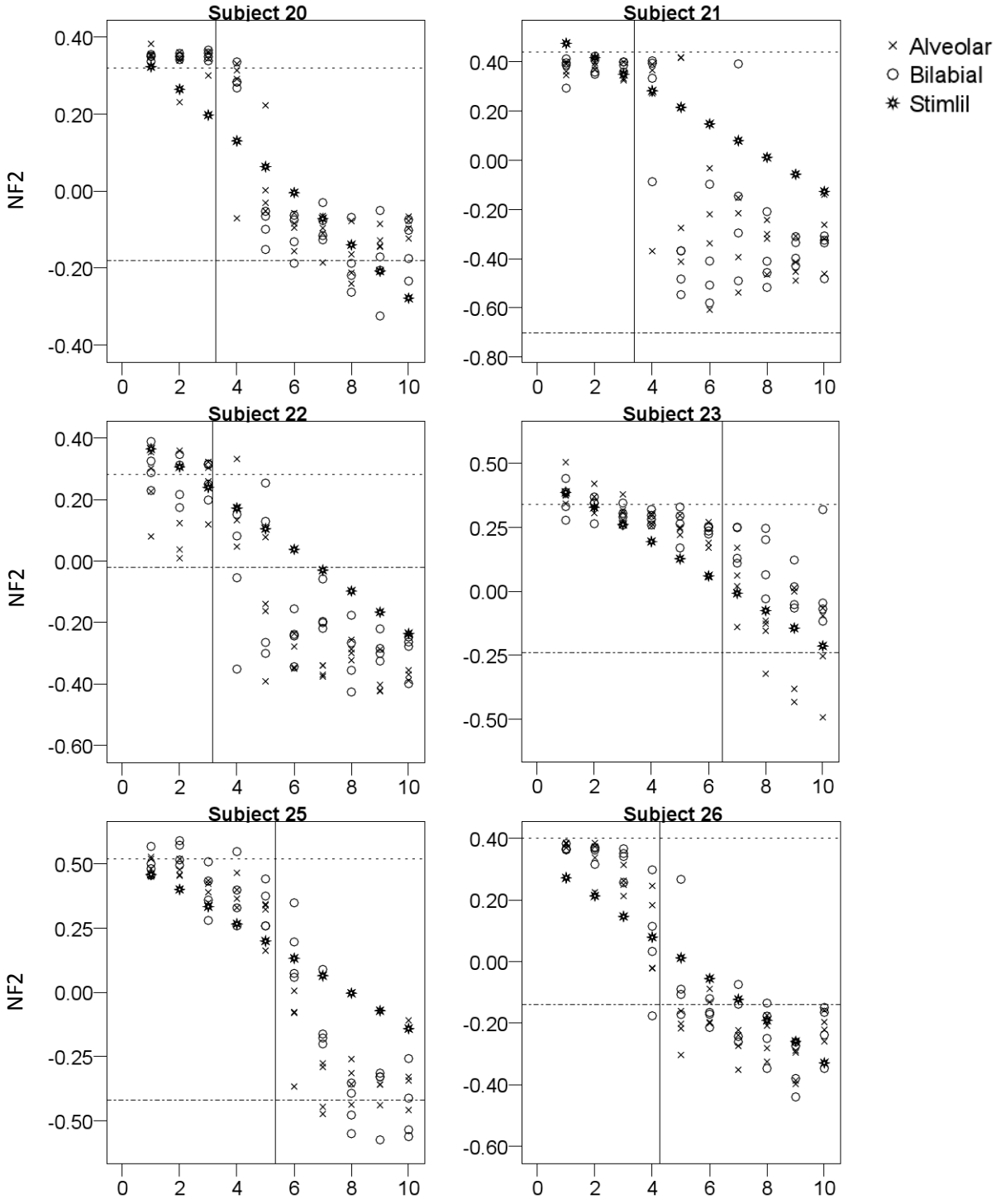
Appendix G—Chapter 5 (Continued): Results from fifteen female subjects: NF2 of response vowels as a function of stimulus number (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli.



Appendix H—Chapter 5: Results from fourteen male subjects: NF2 of response vowels as a function of stimulus number (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli (a uniform set of stimulus F2 was transformed into unique set of NF2 for each subject). A vertical line indicates Boundary, and two horizontal dotted lines indicate NF2 in *heed* (upper) and NF2 in *who'd* (lower) for each subject.



Appendix H—Chapter 5 (Continued): Results from fourteen male subjects: NF2 of response vowels as a function of stimulus number (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli.



Appendix H—Chapter 5 (Continued): Results from fourteen male subjects: NF2 of response vowels as a function of stimulus number (#1-10) and context (Alveolar or Bilabial), with NF2 of stimuli.