UC Berkeley
UC Berkeley PhonLab Annual Report

Title
A Production Study on Controlled Coarticulation: A Case of Contextual Fronting of /u/ in American English

Author
Kataoka, Reiko

A production study on controlled coarticulation: a case of contextual fronting of /u/ in American English[*]

Reiko Kataoka

Department of Linguistics, University of California at Berkeley,
1203 Dwinelle Hall, Berkeley, California 94720-2650
kataoka@berkeley.edu

1. Introduction

The study reported in this paper addresses the question of whether in American English fronting of the high back vowel /u/ in alveolar contexts is the result of physical and physiological constraints alone or it is under the speakers' 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 the vowel /u/, then such control requires context-specific target specification for the vowel. What the question asks, then, is this: do speakers maintain two separate articulation targets for fronted and non-fronted variants of /u/?

A larger question that motivates the present study is an issue of phonologization (Hyman, 1972, 1975, 1976). Phonologization is a process whereby a speech sound acquires/loses phonetic feature, based on physical and physiological constraints in a given phonetic environment, which is exaggerated to the degree that the feature (or lack of feature) is no longer perceived as induced by the phonetic context but rather independently controlled as a distinctive specification of the sound. It is generally considered that main mechanism of assimilatory sound changes is phonologization of acoustic perturbations that originate in coarticualtion.

One assumption underlying the concepts of phonologization is that context-specific speech variation can become a production goal in its own right and thus mentally represented as such. However, exactly what types of coarticulatory variations should be considered as phonologized variations are still open questions. Hyman (2008) discusses two stages of phonologization, as in (1).

(1)   a.   b.   c.

    universal phonetics   >   Language-specific phonetics   >   phonology

      ("automatic")   ("speaker-controlled")   ("structured")

In the stage (b), where coarticulation is implemented in language-specific way which exceeds the degree of mechanical interaction, the result is still phonetic if it is gradient rather than categorical. Coarticulation enters the domain of phonology when, as in the stage (c), the controlled property becomes structured and categorical (p. 385). Since the above event could take place gradually it is difficult to know where phonetic details become phonological pattern. One might consider gradient, scalar, or probabilistic phonology (Cohn, 2006; Flemming, 2001; Silverman, 2006).

Instead of determining whether contextual /u/-fronting in American English should be considered as phonological or phonetic, this study focuses on obtaining sufficient evidence for whether /u/-fronting is purely automatic coarticulation due to production constraints or it has controlled component. This study also provides some acoustic descriptions of the fronted and non-fronted /u/s. In the situation where there is no generally accepted set of criteria for determining whether phonologization has or has not occurred, description of coarticulatory variations is much needed and useful for the development of the theory of phonologization.

The rest of the paper is organized as follows. First, I will present some attested cases that the allophonic split of the high back vowel in fronting and non-fronting contexts has become phonemic (Section 2). Next, I will discuss phonetic bases of contextual /u/-fronting by reviewing previous articulatory and acoustic studies (Section 3). Then I will present a research hypothesis (Section 4) and justify the method of testing this hypothesis (Section 5). I will then report the experimental study and its results (Section 6). Finally, I will discuss the implications of the findings for the theory of control mechanism of coarticulation and theory on the role of coarticulation in phonology/phonologization (Section 7). The paper ends with prospectus for future research on the issue of mental representation of coarticulation.

2.   Attestation

The effects of coarticulatory fronting of /u/ adjacent to alveolar consonant can be observed in historical sound changes as well as synchronic sound patterns. Relevant diachronic cases have been reported from comparison of Written Tibetan (WT), which was established in about the eighth century, with its modern descendants Lhasa (Michailovsky, 1975) and Dzongkha (Mazaudon and Michailovsky, 1988), the national language of Bhutan. As shown in Table 1, the modern reflexes of WT /aT/, /oT/, and /uT/ (T = /d/, /s/, /n/, or /l/) show vowel fronting in Lhasa, and the modern reflexes of WT /Vd/ (V = /a/, /o/, or /u/) show vowel fronting in Dzongkha.

Cognates in Western Grassfields Bantu languages of the Ring (Nkom) subgroup spoken in Cameroon suggest a similar phonemic split as shown in Table 2[1]. 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 source of the Ring languages split into /u/ and /y/ in Babanki as conditioned by the coronal onset.

Synchronically, the co-occurrence restriction of the vowels and coronal consonants is observed in Cantonese (Cheng, 1991). Cantonese has both front and back non-low rounded

---

[1] My sincere thanks to Larry Hyman for kindly sharing the Ring language data with me.

Table 1 Sound changes from WT (8th C.) to Lhasa Tibetan (Michailovsky, 1975:323) and to Dzongkha (Mazaudon and Michailovsky, 1988:126). IPA symbols are normalized to conform standard convention, and markers for tone and voice quality are omotted.

(1)	WT	Lhasa Tibetan	gloss
	sk**a**d	q**ɛ**:	'language'
	b**a**l	ph**ɛ**:	'wool'
	b**o**d	ph**ø**:	'Tibet'
	st**o**n	t**ø̃**:	'autumn'
	l**u**s	l**y**:	'body'

(2)	WT	Dzongkha	gloss
	sk**a**d	k**e**:	'noise, speech'
	brgy**a**d	g**ɛ**:	'eight'
	khy**o**d	ʧ**oe**:	'you (sg.)'
	dr**o**d	dhr**oe**:	'heat, fever
	l**u**d	l**ue**:	'manure'

Table 2 Cognates of the Ring languages of Western Grassfields, Cameroon. Markers for tone are omitted. (Data from Ring Language Database)

	Babanki	Aghen	Isu	Kom	gloss
(1)	ə.k**u**	kɨ.k**u**	ke.k**a**	a.k**u**	'forest'
	m**uu**	m**uu**	mw**i**	ə.m**u**	'water'
	k**u**	i.f**uo**	n**i**	f**u**	'to give'
	f**u**	i.k**uo**	kw**ɔ**	k**u**	'to snore'
(2)	ə.l**y**	tɨ.z**u**	tə.z**u**	ə.l**u**	'honey'
	ʒ**y**	i.z**u**	z**u**	ʒv**ʊ**	'to skin'
	ʃ**y**	i.s**u**	s**u**ʔ	s**u**	'to wash'

Table 3 Co-occurrence restrictions for /u/ and /o/ in coronal contexts in Cantonese (Cheng, 1991:110)

	Cantonese	gloss
(1)	p**u**t	'to wipe out'
	m**u**n	'board'
	k**o**n	'dry'
	k**o**t	'to cut'
(2)	t'ok	'to support'
	tsok	'to create'

* COR o COR
* COR u

vowels (/ü/, /u/, /ö/, /o/), but the high back rounded vowel /u/ does not occur with coronal onset and the mid back rounded vowel /o/ does not occur between coronal onset and coda as shown in Table 3. Data in (1) show that these vowels can occur with coronal coda as long as the onset is non-coronal. Data in (2) show that /o/ can occur with coronal onset if the coda is non-coronal. Cheng proposed that these patterns arise due to assimilatory fronting of /u/ and /o/ in the coronal context (p. 121). Her analysis implies that there was an allophonic split of non-low back rounded vowels into front and back variants before the current sound patterns had established.

These attestations illustrate the pervasiveness of the effect of assimilatory fronting of the back vowels on the synchronic and diachronic patterning of the sounds. Further, that these attestations come from unrelated languages indicates that the fronting of back vowels in coronal contexts is likely to originate in universal phonetic constraints. The next section examines these constraints from articulatory and acoustic studies on /u/-fronting.

3.   Observations from Articulatory and Acoustic studies

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. Öhman, 1966, 1967; MacNeilage and DeClerk, 1969; Kent and Moll, 1972; Kiritani et al., 1977; Kiritani, 1986; Recasens, 1991; Farnetani and Recasens, 1993; Recasens et al., 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 review). For example, Kiritani, Itoh, Hirose and Sawashima's (1977) x-ray microbeam study on the 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/). 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 Kiritani et al's (1977) data. Another is that there are observable differences in the articulation of a vowel in /b_b/ context, which is a 'neutral' context for tongue articulation for a vowel (p. 1218), and in other symmetrical /C_C/ contexts. And finally, in either C_1V or VC_2 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) (Figure 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
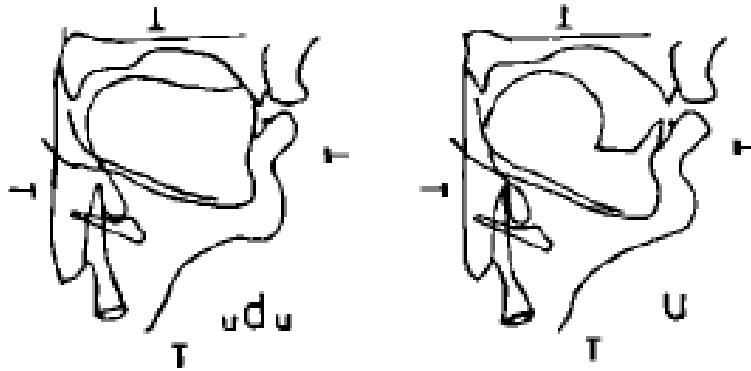
Figure 1   Contour tracings from x-ray motion pictures of /udu/ (left) and /u/ (right) uttered by a male Swedish speaker. The edges of the hyoid bone, the mandible, and the epiglottis are shown. (Öhman, 1966:166)

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

In sum, 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.

3.2   *Acoustic properties of /u/ in fronting and non-fronting contexts*

The coarticulatory influence of adjacent consonants on the vocal tract shape of a given vowel results in altered acoustic properties of the vowel both in high-low and front-back dimensions, as often revealed by measurements of F1 and F2 values (e.g., Lindblom, 1963; Stevens and House, 1963; Öhman, 1966, 1967; Recasens, 1985; Farnetani and Recasens, 1993). On the effect of alveolar consonants on the back vowel /u/, previous studies unanimously report a raising effect on F2 of the vowel. For example, Öhman (1966) reported that in /udu/ utterances, the vowel's F2 increases 490 Hz at the VC juncture and 690 Hz the CV juncture compared with the point where F2 is steady. This type of dynamic change in F2 is expected, as the tongue gradually moves from a vowel configuration to a consonant configuration (or from a consonant to a vowel), as observed in MacNeilage and DeClerk (1969). Stevens and House (1963) measure 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 (Figure 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,
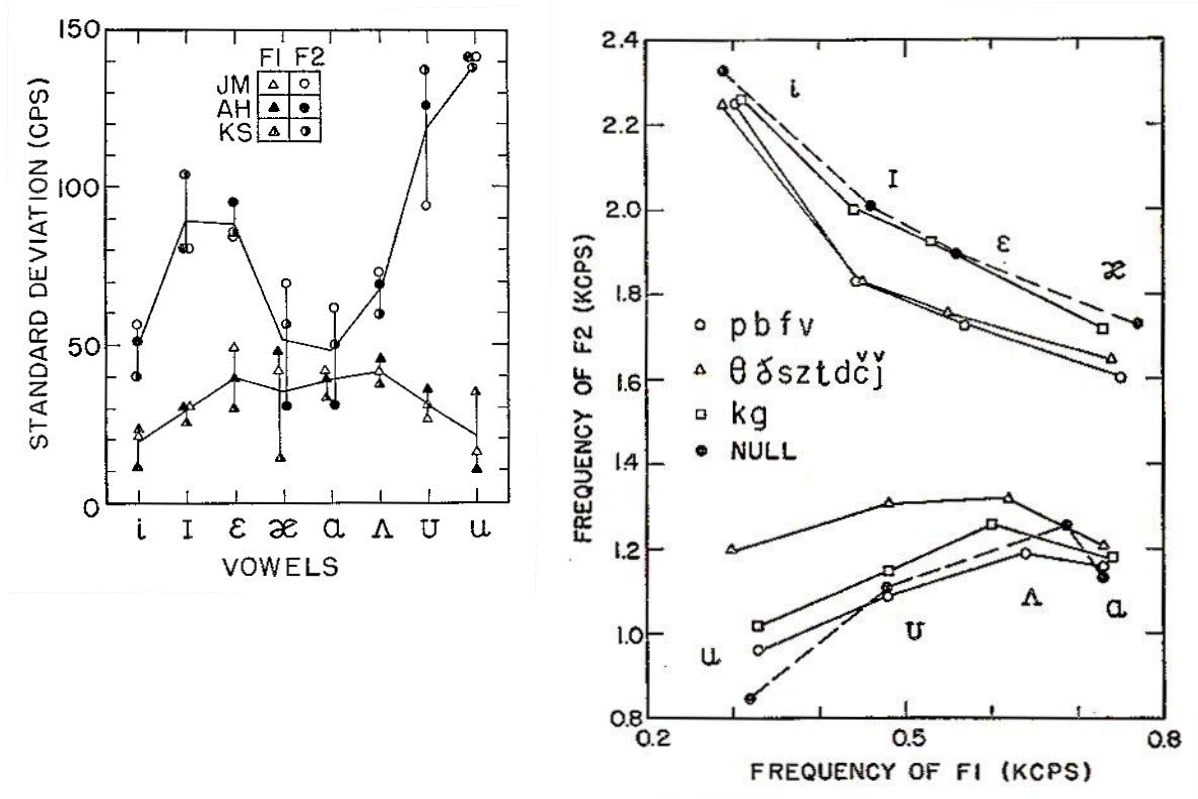
Figure 2   Stevens and House (1963) data showing the extent of variability of F1 and F2 in the 14 consonantal contexts (left) and the effect of the place of articulation of the consonants on vowels' F2 (right).

these observations point to a particular vulnerability of F2 in the high back rounded vowels in the context of alveolar consonants. In 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, 1981:180).

4.   Hypotheses

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 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 bases to hypothesize that this is the case. A rational support comes from the analysis on control mechanism 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.
>
> <div align="right">(Öhman, 1967:310)</div>

By extending Öhman's account, one might expect that it is not possible to associate invariant-target VT shapes with the back vowel /u/ because the constriction gesture of the apical articulator of an alveolar consonant is superimposed onto the tongue body gesture for an adjacent /u/ and the resulting VT shape would be uniquely different from the VT shape for /u/ in 'null' environments.

Empirical support comes from numerous previous studies showing that phonetic implementations of speech signals consist of both mechanical components and controlled components and this controlled component shapes phonetic output in language-specific ways as often observed as cross-linguistic differences in coarticulation. In a 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 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, and 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 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) presents evidence that there are both 'automatic' types and 'controlled' types of coarticulation (see Section 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?

5. Methodology

Lindblom (1963) employs vowel manipulation method to test whether vowel reduction involves either or both of centralization and coarticulatory assimilation. His data show 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/) reduce as the vowels' duration increases: the formant frequencies of each of the vowels approach asymptotic values as the duration increases. Further, derived regression models for each vowel's formant values are generally successful without including centralization as a predictive factor in the model. From these results Lindblom makes the following claims: (1) vowel reductions are due to assimilation, 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, 2007) and phonological vowel reduction (Barnes, 2006).

Solé (1992) uses the same method to test cross-linguistic variation in temporal extent of vowel nasalization in vowel-coda nasal sequences. Her results show that in American English the duration of nasalization during the vowel is 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 remains constant regardless of overall vowel duration. With these data, Solé claims 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 and Ohala, 2010).

The studies discussed in the above paragraphs demonstrate the usefulness of duration manipulation in investigating 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. A conceptual hypothesis is: "In American English, contextual fronting of the back vowel /u/ in alveolar contexts have distinct production goal separately from that of canonical /u/." 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.

6. Experiment

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.

6.2 Materials

A list of English test words, a control word, and reference words was created (Table 4). 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/, where we expect the vowel to have a comparable articulatory configuration as the vowel in isolation, the 'null' context (Stevens and House, 1963:116). The purpose of eliciting reference vowels was to construct speaker-specific vowel spaces in which to calculate the degree of /u/-fronting for each speaker.

Table 4   Words elicited in the 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]

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") four to 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)

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. 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 (Figure 3, upper panel). For the test vowel
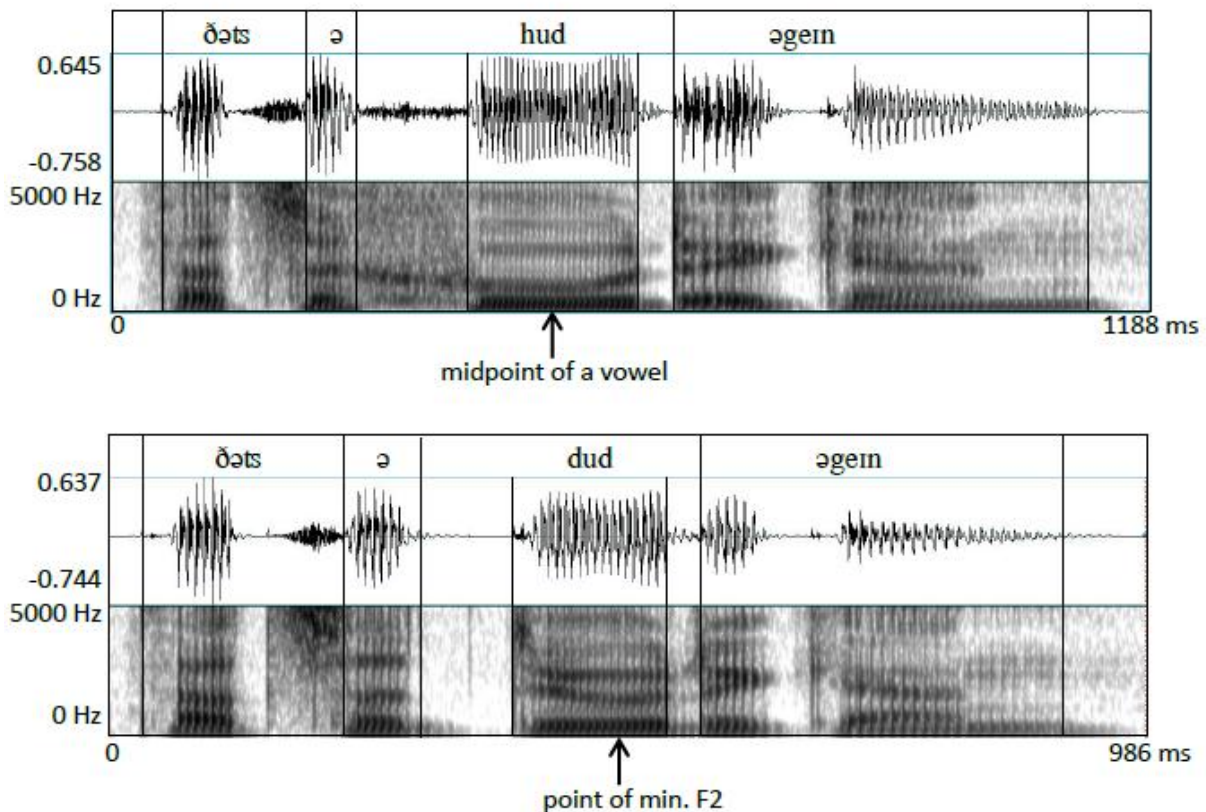
Fig. 3   Examples of a waveform and a spectrogram of a reference word (upper) and a test word (lower) embedded in a carrier sentence. Each example shows demarcation for formant measurements: If the onset was a fricative, the beginning was set at the beginning of the vowel (upper); if the onset was a stop, the beginning was set to the onset release (lower). Arrows indicate the points from where F1 and F2 were measured.

/u/, F1 and F2 values were measured from the temporal point where F2 reaches its minimum (Figure 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). 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 a vowel, often near the very end of the vowel.

Formant measurement was done with Praat (Boersma and Weenink, 2005) by using a script that measures and logs 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 a 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 data 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.

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, cited in Adank, 2003) individual log-mean method was employed. This method is based on the assumption that the ratios of formant frequencies are more relevant for vowel recognition than the actual frequencies of those formants. Each speaker's vowels are located in a logarithmic F1-F2 space in relation to a single reference point, the log mean (Adank, 2003:22). The choice of this normalization method was motivated by Adank (2003) and Adank, Smits, and van Hout (2004), which show 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 be also maintained after normalization.
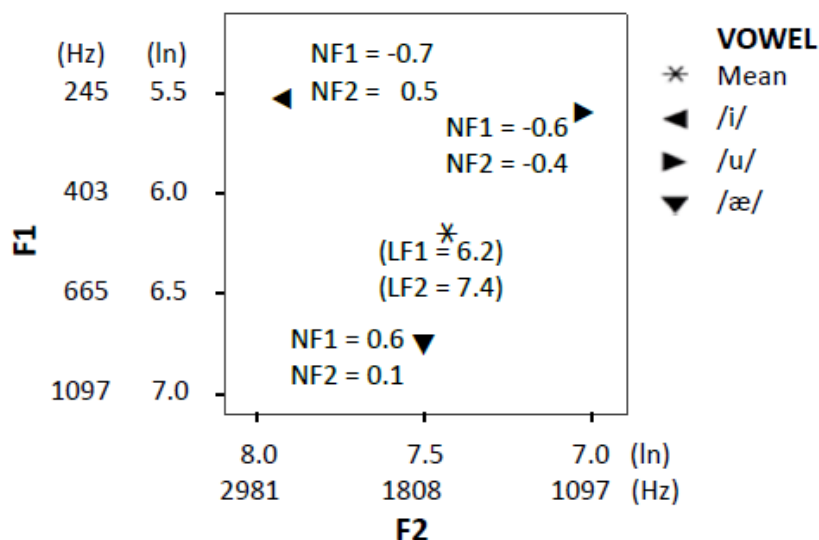
Fig. 4 A sample of vowel normalization for speaker #7 (female). Log mean was 6.2 for F1 and 7.4 for F2. Normalized F1 and F2 for /i/, /u/, and /æ/ are provided in the F1-F2 plots.

Figure 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 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 each gender 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 gender groups speakers vary considerably in physical size, and presumably also in vocal tract size. As a result of gender differences and individual differences in the formant values the plots exhibited considerable

overlaps of the vowel categories. Figure 6 shows the distribution of the normalized formant frequencies (NF1 and NF2) of the same vowels. As the data were normalized for each speaker, gender differences of the formant values were greatly reduced and more distinctive 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 gender differences still remain. The trend
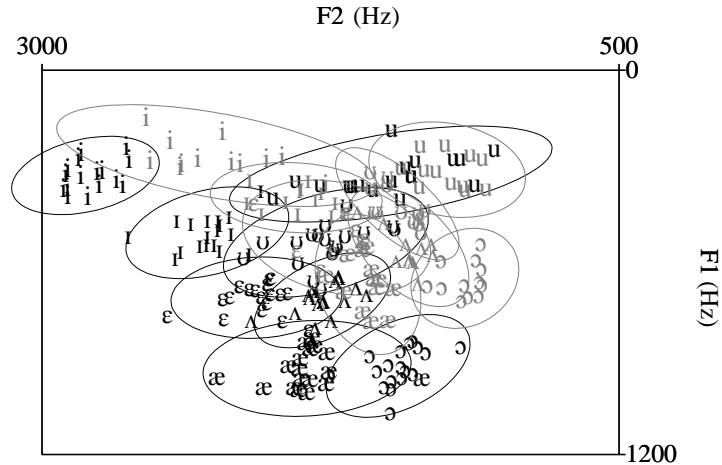
Figure 5   Mean F1 and F2 values (Hz) of each reference vowel for 18 females (black) and 12 males (gray) with 95 % confidence ellipses (N = 240: 8 vowels x 30 speakers).
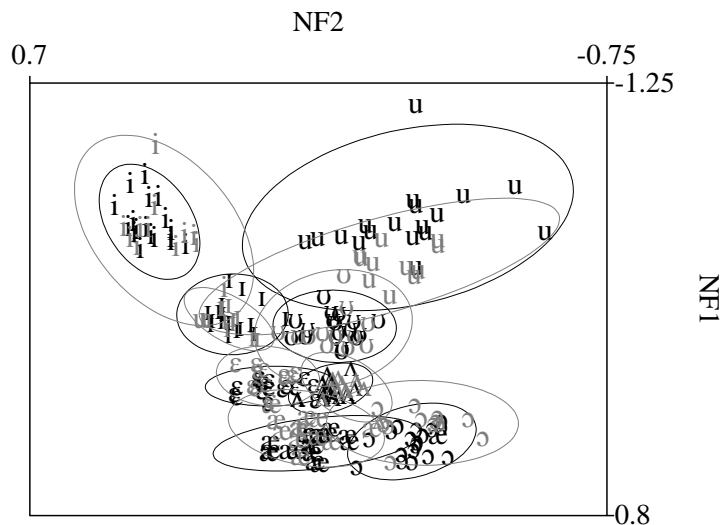
Figure 6   Mean NF1 and NF2 values (ln) of each reference vowel for 18 females (black) and 12 males (gray) with 95 % confidence ellipses (N = 240: 8 vowels x 30 speakers). Lower panel shows only the ellipses with group means.

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/.

6.6   Analyses and Results

6.6.1 Vowel duration

Figure 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
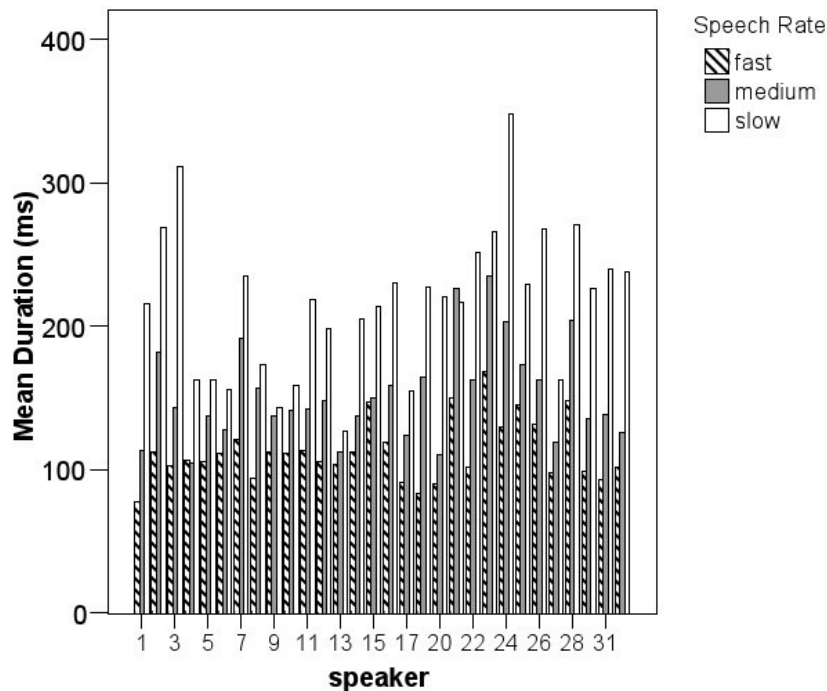
Figure 7   Mean vowel duration per speech rate per subject (subjects #18 and #30 excluded (see §3.5.3)).

Table 5   Summary for vowel duration (ms) by rate (N = 270).

Rate	N	Mean	SD	Min.	Max.
Fast	90	113.18	25.59	64.46	190.29
Medium	90	152.60	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

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 5.

6.6.2 Variation of /u/

Before examining the effect of vowel duration on the degree of fronting, some observations on the acoustic properties of /u/ in fronting and non-fronting contexts were made. Figure 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 segments were defined as intervals between the stop release and vowel offset for the stop-vowel-stop words. With the stop release included in the interval, formant measurements were reliable only from the second point. With the same reason, only the measurements taken from the last four points were used for 'toot' and 'tune'.

Several patterns emerge from the plots. First, the vowel /u/ exhibits distinct NF2 trajectories in the test words than in the control or the reference word. The difference persists even at the point where NF2 of test words reaches its minimum. Second, all test words have very similar NF2 trajectories in the later portion of the segment, and NF2 seem to converge to a
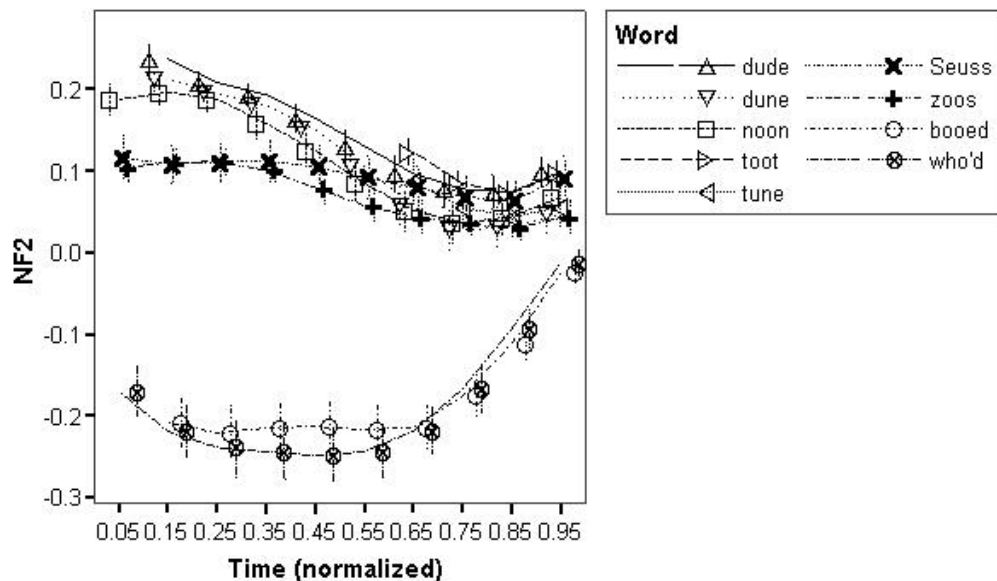

Figure 8   Averaged and time-normalized NF2 trajectories with 95% confidence intervals based on the measurements taken from the ten equally-distanced points of the mid 90% of /u/ in each test word ('dude', 'dune', etc.), control word ('booed') and reference word ('who'd'). The trajectories are aligned 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 (see Section 6.4).

common value at the vowel offset. This is not surprising given all words share the same vowel-coda sequence. Third, at vowel onset, 'Seuss' and 'zoos' have lower NF2 than other test words. One possible interpretation of this pattern is that this is an artifact of the segmentation: for these words the vowel segments started at vowel onset, excluding release noise. This means that for given time point, articulatory events are probably not comparable 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/,
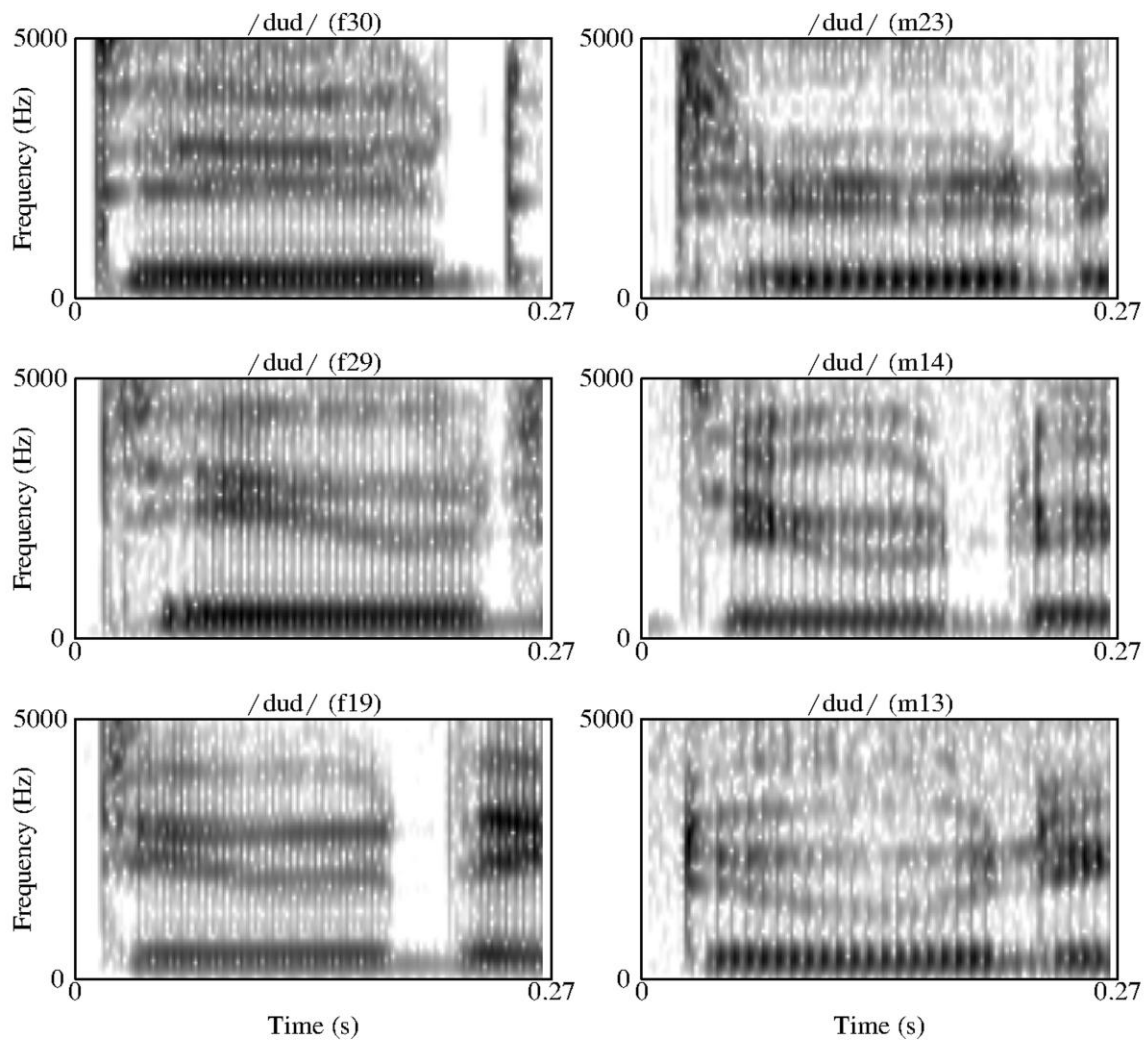
Figure 9 Spectrograms of the American English speakers' production of a word 'dude' (/dud/) and a part of following 'a' (/ə/) in a carrier sentence "That's a ___ again." The spectrograms in the left and the right column represent utterances of female (#30, #29, #19) and male (#23, #14, #13) speakers, respectively, and the spectrograms in the top, the middle, and the bottom row represents 'flat', 'up-and-down', and 'U-shape' trajectories, respectively.

respectively. Finally, despite antagonistic relationship between alveolar onset and the vowel /u/, the vowel's NF2 does not fall immediately after vowel onset. Rather, F2 seems to remain in its initial level or even rise for a short period of time after vowel onset. Indeed many of the test word tokens exhibit rising-falling F2 contour, similar to the F2 contour in a sequence of palatal glide and /u/ as in words like 'human', 'youth', and etc.

Further observation revealed that F2 trajectories for stop-vowel-stop words vary across speakers, but generally fall in one of the following three types. For one type, though there were not many tokens that fell in this type, F2 starts high at CV juncture, immediately starts falling to reach its minimum toward the end of the vowel, and rises again toward the release of the coda stops. This would be called as a 'U-shape' trajectory. For another type, F2 makes considerable rise before it falls, as described in the previous paragraph. This type would be called as an 'up-and-down' trajectory. For the last type, F2 makes relatively flat trajectory. This would be called as a 'flat' trajectory. A sample spectrogram illustrating each type of F2 trajectory from female speech and male utterance of 'dude' is presented in Figure 9. All three types of F2 trajectories were observed from both male and female speakers' production; however, there was an interesting gender difference in that majority of the male speakers produced 'flat' trajectories, while majority of the female speakers produced 'up-and-down' trajectories.

6.6.3 Distribution of /u/ in NF1-NF2 space

Figure 10 shows NF1-NF2 plots of test vowels ("d") and the control vowel ("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/. For the test vowels each data point represents the mean NF1 and mean NF2 of all of the seven words ('dude', 'toot', 'zoos', 'Sues', 'noon', 'dune', '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 nearly the same as the distribution of the reference vowel /u/. The vowel /u/ in the fronting context, on the other hand, occupied an entirely different space, in the open region right next to the vowel /i/. 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. 11). This trend suggests that shifts in NF2 in the fronting context are not a result of physiological constraints; instead, speakers aim at different acoustic patterns, which are more narrowly defined than their null-context counterparts.

6.6.4 F2 as a function of vowel duration

The acoustic properties of the fronted and non-fronted /u/ were further examined in terms of their response to the vowel duration manipulation. Figure 12 shows the plots of NF2 of /u/ in /D_D/,
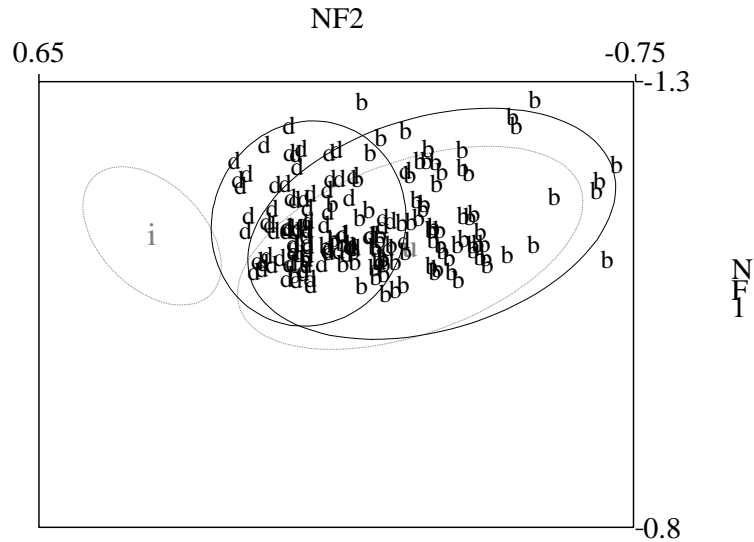
Figure 10 95% confidence ellipses for NF1 and NF2 of reference vowels /i/ and /u/ (gray symbols with dotted lines) and NF1-NF2 plots and 95% confident ellipses of test vowels (d) and the control vowel (b) spoken in all three speech rates (black symbols and solid lines) (N = 180: 2 vowels x 3 rates x 30 speakers). For test words, each data point represents the mean of seven test words ('dude', 'toot', 'zoos', 'Sues', 'noon', 'dune', 'tune').
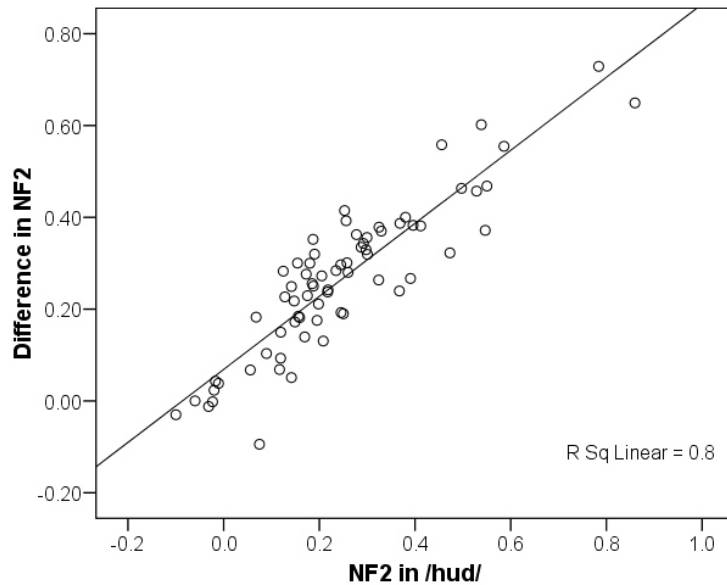

Figure 11 Scatter plots of NF2 of /u/ in the null context (/h_d/) spoken in fast, medium, and slow rates, and the difference in NF2 of /u/ between the null context and the fronting contexts (/D_D/) (n = 90: 3 rates x 30 speakers). A linear regression line was added.
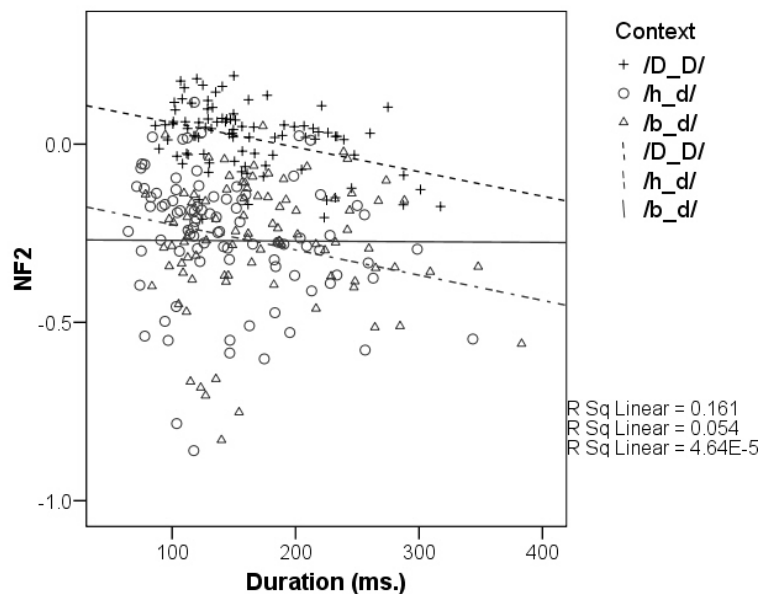
Figure 12   Scatter plots of NF2 values of /u/ as a function of segment duration. Each data point represents mean NF2 values of /u/ calculated from all test words (/DuD/), reference words (/hud/) and control words (/bud/) for each speech rate (fast, medium, and slow) for each speaker. (n = 270; 3 contexts x 3 rates x 30 speakers). Linear regression line was added for each context.

/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 (30 speakers x 3 contexts x 3 rates = 270 data points). Linear regression lines for each context were added to the plots.

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

    Whether the degree of fronting of /u/ in the fronting context persists 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 are 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 is tested to see whether the y-intercepts of the regression lines are all the same. For the present study rejection of the second null hypothesis supports the

hypothesis that NF2 of the vowels in fronting and non-fronting contexts are significantly different. However, ANCOVA is not appropriate for the present data because the assumptions of independent observation and constant variance are violated: the same subjects repeated vowel productions for different contexts and different speech rates, and NF2 for /hud/ and /bud/ is 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. Vowel duration was the covariate, the 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:

$$NF2_{ij} = (b_0 + u_{0j}) + (b_1 + u_{1j}) (Context)_{ij} + b_2 Duration_{ij} + \varepsilon_{ij} \qquad (1)$$

In the equation (1), the js represent levels of the random variable (i.e. subjects, in this case); the is represent the levels of the fixed variable (i.e., /D_D/, /b_d/, /h_d/); b_0 is fixed intercept; u_{0j} reflects variability in intercepts; b_1 is fixed slope for 'context'; u_{1j} reflects variability in separate slopes; b_2 is fixed slope for 'duration'; and finally ε represents error. Note that the variable 'rate' was included in the model to accurately reflect clustering of the data, but this variable was not tested as a predictor; therefore, in the results of the regression analyses the effect of 'context' reflects the effect across all three speech rates.

The results show that contexts are significantly associated with NF2 after the effects of vowel duration are controlled (Table 6). Estimated mean NF2 values for /bud/, /hud/, and /DuD/ at the average value of vowel durations (161 ms) are shown in Table 7. Mean NF2 was highest for /DuD/, and means for the other two contexts were very similar to each other. Thus, the estimated parameter value for the /bud/ context, which represents the difference of NF2 between the /b_d/ context and the reference context (i.e. /h_d/), is very small (0.015) and not significant (t (21) = 0.885, $p = 0.39$), while the parameter value for the /D_D/ context was significantly larger (t (52) = 13.84, $p < 0.01$) than for the /h_d/ context (Table 8). From these results I conclude that the vowels produced in the fronted contexts are qualitatively distinct sounds no matter how slowly the vowels are produced.

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 is it 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/ remains across differences in vowel duration. This result is further confirmed by statistically significant NF2 differences between fronted and canonical /u/. These results indicate that although vowel duration has an effect on NF2, longer vowel duration does not make these contextual variants more similar to each other. Rather, the effect of longer vowel duration is applied equally for /u/ in both fronting and non-fronting contexts. These results imply that the

Table 6   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 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 factors.

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

Table 8   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

fronted and non-fronted variants of /u/ are distinct acoustic patterns, and speakers do not aim to make these vowels as the same sounds even in slow speech when the articulator has more time to approximate the intended target sound. Second, fronted /u/ does 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/ exhibits. This result implies production constraints for fronted /u/ are greater than for canonical /u/; while speakers seem to have a considerable degree of freedom in articulating canonical /u/, they hit a more narrowly specified articulatory target for /u/ in alveolar contexts. From these observations, 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, how do we (1) 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 of the early models is an inertia-based undershoot model (Lindblom, 1963; Stevens and House, 1963; Stevens, House, and Paul, 1966; Moon and Lindblom, 1994). This model assumes that vowel phonemes have invariant acoustic targets, and predicts vowel undershoot to occur when the articulator does not have sufficient time to hit the target. The undershoot model is clearly incompatible with the present results. Although NF2 values become lower as vowel duration increases, the acoustic distance between fronted /u/ and canonical /u/ remain the same in terms of NF2. 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 is Öhman's (1966, 1967) model, which is similar to the undershoot model in that it also assumes invariance, but in Öhman's model 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, which does not lower completely during the following vowel because the tip does not receive a direct command for the vowel and thus is free to carry over coarticulatory effects from the previous consonant. However, the results obtained in the current study 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 and 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 and Munhall, 1989), and coarticulation is modeled as an overlap between gestures (Browman and Goldstein, 1992; Fowler and 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 coarticualtory effect. For example, by using the gestural activation wave (Fowler and 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, and Saltzman, 2006) and coupling between alveolar consonants and /u/ that is tighter than other kinds of CV coupling.

Viewing /u/-fronting as a case of gestural constellation has the merit of the capturing greater acoustic effects of alveolar consonants on /u/ than other types of CV coarticulations (as discussed in Section 2) and lesser variability of fronted /u/ compared with canonical /u/ that are observed in the present study. The articulatory phonology model fits the data nicely, but this model has an underlying assumption that gestural constellations emerge online as natural consequences of gestural coordination (Fowler and Saltzman 1993; Browman and Goldstein 1995). A speaker may 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 (Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2001, 2002; Bybee, 2001, 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 2005:28). 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 experience. 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)

One type of evidence for exemplar-based grammar is word frequency effects on phonetic reduction (Bybee, 2001; Pierrehumbert, 2001), on sound change (Schuchardt, 1985; Bybee, 2001) and on word recognition (Broadbent, 1967; Connine, Titone and Wang, 1993). Another type of evidence is the effect of assumed talker identity on speech perception (Johnson, 1997; Hay, Warren, and Drager, 2006). One claim of the theory is that "phonology is represented in phonetic detail rather than in featural abstraction" (Johnson 2005:28); however, it is quite reasonable to assume that repeatedly experienced generalizations over phonetically detailed exemplars receive structurally different status (i.e. as a category node) rather than raw exemplars in multiple levels of representation (Pierrehumbert 2003). Within this framework, therefore, one would expect that general image of repeatedly experienced fronted variants of /u/ to be mentally represented either as a phonetically distinct sound category, as a distinct articulatory category, or as both.

An implication of a multi-layered and exemplar-based memory 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 these sound patterns are repeatedly experienced by a language user. In this approach the fronting of /u/ in alveolar contexts in American English would be a good 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

using a production pattern stored in memory or by on-line planning for a strongly coupled CV sequence. Predictions from these two models can be tested, for example, by using priming tasks with fronted and non-fronted variants of /u/ as primes. If certain word forms are primed only by fronted or non-fronted /u/, then this would constitute evidence for sub-phonemic variants being mentally represented. The present study does not fully address the question of the mental representation of subphonemic variations, which remains a topic for future research.

References

Adank, P., 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.

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.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, 74(1), 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* (p. 341-376). Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155-180.

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

Bybee, J. L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. L. (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. L. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82(4), 711-733.

Cheng, L. L.-S. (1991). Feature geometry of vowels and co-occurrence restrictions in Cantonese. *Proceedings of the 9th West Coast Conference on Formal Linguistics* (p. 107-124).

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* (p. 25-44). Oxford: Oxford University Press.

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(1), 81-94.

Farnetani, E., & Recasens, D. (1993). Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Language and Speech*, 36(2, 3), 279-302.

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

Fowler, C. A., & Saltzman, E. L. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2, 3), 171-195.

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

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* (p. 215-249). Cambridge University Press.

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.

Hyman, L. M. (1972). Nasals and nasalization in Kwa. *Studies in African Linguistics*, 3(2), 167-206.

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

Hyman, L. M. (1976). Phonologization. In Alphonse Juilland (Ed.), *Linguistic studies presented to Joseph H. Greenberg* (Vol. 4, p. 407-418). Saratoga, CA: Anma Libri.

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

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (p. 145-165). San Diego: Academic Press.

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

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 Research Institute of Logopedics and Phoniatrics No. 11 (p. 11-21). University of Tokyo.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35(11), 1773-1781.

MacNeilage, P. F., & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America*, 45(5), 1217-1233.

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* (p. 115-136). Canberra: Australian National University, Research School of Pacific Studies.

Michailovsky, B. (1975). On some Tibeto-Burman sound changes. *Proceedings of the first annual meeting of the Berkeley Linguisitcs Society* (p. 322-331).

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(1), 40-55.

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 Parasession on Language and Behavior* (Vol. 17). Chicago: Chicago Linguistic Society.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39(1), 151-168.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America*, 41(2), 310-320.

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

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology VII* (p. 101-139). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2-3), 115-154.

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

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(1), 544-561.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333-382.

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.), *Papers in Laboratory Phonology VIII* (p. 133-154). Mouton de Gruyter.

Solé, M. J. (1992). Phonetic and phonological processes: the case of nasalization. *Language and Speech*, 35(1, 2), 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.), *Papers in Laboratory Phonology 10* (p. 607-655). Berlin: Mouton de Gruyter.

Steves, K. N., & House, A. S. (1963). Pertubation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 6(2), 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(1), 123-132.