**Title**

A Study on Perceptual Compensation for /u/-fronting in American English

**Author**

Kataoka, Reiko

# A Study on Perceptual Compensation for /u/-fronting in American English

REIKO KATAOKA
*University of California, Berkeley*

## 0.    Introduction

There is an undeniable link between speech production and speech perception. In running speech, the phonetic form of consonants and vowels varies due to the overlapping of adjacent articulations (coarticulation). However, the listener hears them as if there were no coarticulatory distortions on the segments. The latter 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 current study. This phenomenon is similar to constancy in visual perception. Perception of an object's color, for example, remains constant across illumination types; that is, visual information about object's color is "corrected" for irrelevant viewing conditions (Yantis 2001: 7).

This is an important area of inquiry because of its contributions to the linguistic theories of sound change and theories of speech perception in general. Sound change refers to change in pronunciation norms over time in a speech community, and one major cause for common sound changes is listeners' misperception of contextually perturbed speech sounds (Ohala 1981, 1993). Since compensation prevents this particular type of misperception, it is an essential component of the theory of sound change. Although much work has been devoted to this topic (see, e.g., Repp 1982; Diehl, Lotto, & Holt 2004 for reviews), exactly how the human auditory system achieves compensation is yet to be fully understood, and more studies are needed to this end. This paper will report one such study, an experimental study on compensation for /u/-fronting in an alveolar context.

## 1.    Background
### 1.1.    Hypo-Correction

Ohala (1981, 1993) proposed a theory of sound change due to the listener's misperception. The theory acknowledges considerable variations in the phonetic form of functionally equivalent speech units, and states that when encountering coarticulatory speech variation, listeners either: 1) perceptually compensate for predict-

able variations and arrive at the pronunciation target intended by the speaker, or 2) fail to compensate for coarticulation and assume that the coarticulated form is the intended pronunciation. The former scenario describes what happens in normal speech perception; the latter, what happens in the type of misperception[1] termed 'hypo-correction'. According to Ohala, hypo-correction is the underlying mechanism for many assimilatory sound changes.

The key concept of hypo-correction is that 'contextually induced' perturbation is interpreted by a listener as an 'intended' feature of the speech sound. In this way, hypo-correction has the potential to alter the listener's phonological grammar by what Hyman (1976: 408) called 'phonologization', a process whereby intrinsic/automatic variation becomes extrinsic/controlled. Today, many researchers analyze sound change as a result of phonologization (e.g., Barnes 2002, Blevins 2004, Blevins & Garrett 1998, Yu 2004), underscoring the theoretical importance of hypo-correction as a precondition for sound change via phonologization.

## 1.2.    Compensation for Coarticulation

Listeners' ability to normalize coarticulation has been demonstrated in many studies. For example, Lindblom and Studdert-Kennedy (1967) examined listeners' recognition of a vowel in a series of [jVj] and [wVw] syllables varying perceptually from /CɪC/ to /CʊC/, and found that ambiguous vowel stimuli were more often heard as /ʊ/ in the [j_j] context as opposed to the [w_w] context. Their results indicate that a listener's categorization boundary on a vowel continuum shifts toward the /ɪ/-end in the [j_j] context. Similar results were obtained by Ohala and Feder (1994), where their American listeners judged a vowel stimulus ambiguous between /i/ and /u/ more frequently as /u/ in an alveolar context than in a bilabial context. The listener's response in these studies mirrors coarticulatory fronting of a high back vowel in palatal and alveolar contexts.

Listeners are also capable of taking non-segmental contexts into account in judging speech sounds. In the above mentioned study, Lindblom and Studdert-Kennedy found a greater boundary shift in shorter speech stimuli than in longer ones, indicating that their listeners employed greater compensation in response to faster speech (as the listener would measure speech rate from the duration of the vowel). Ladefoged and Broadbent (1957) tested, among other things, listeners' identification of a vowel stimulus from an /i/-to-/ɛ/ continuum after a precursor phrase, F1 of which was shifted up or down. They showed that ambiguous vowels were more often heard as /i/ (with lower F1) when the precursor phrase had high F1 than low F1, presumably because their listeners took the overall low- or high-frequency context into account when judging the height of the target vowel.

---

[1] The theory acknowledges two systems of 'hearing' sounds—the peripheral system that detects change in acoustic energy (sensation) and the central system that 'interprets' input from the peripheral system (perception). By explicitly stating that what listeners 'hear' may or may not be attributed to acoustic properties alone, it highlights the role of expectation at higher level of processing.

These findings suggest that compensation and other contrastive context effects are closely related phenomena, and that compensation is achieved by a dynamic process, involving not only a local-level adjustment (i.e. adjusting the interpretation of a target acoustic signal relative to the immediate context), but also a larger-level adjustment or some sort of transformation of perceptual scale that relates to time normalization as well as adjustment over larger time windows.

### 1.3. Cause of Hypo-Correction

Ohala points out that hypo-correction occurs when a listener fails to employ compensation, or more specifically, when the listener lacks experience with various contextual variations that enables him/her to do such correction, or fails to detect the conditioning environment due to various reasons such as noise and filtering associated with communication channels (1993:246-7).

However, later studies suggest that there could be yet another reason for hypo-correction to occur—namely, 'variation' in the compensation. For example, Beddor and Krakow (1999) tested American listeners' nasality judgments on the nasalized vowel [ɛ̃]/[õ] between nasal consonants ([mṼn]), on oral vowels [ɛ]/[o] between oral consonants ([bVd]), and on the same oral vowels in isolation ([#V#]), and found that 25% of [Ṽ] in nasal contexts were heard as more nasal than [V] in oral contexts, showing that compensation was not complete or uniform.

Later, Harrington et al. (2008) demonstrated systematic variation in compensation between young and old listeners. They compared the two groups' identification of a vowel from an /i/-to-/u/ continuum in palatal ([j_st]) and labial ([sw_p]) contexts. 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 closer to the boundary on the palatal continuum, indicating less compensation than the older group. According to the authors, these results reflected a difference in the listeners' own speech production: younger speakers' /u/ was generally more fronted than older speakers'.

These findings suggest that listeners compensate for only as much coarticulation as is expected in their own grammar, and that this 'compensation grammar' is, just as other components of grammar, shaped by the listener's previous linguistic experience. Following from this, one might then add to Ohala's list of causes of hypo-correction differences in the coarticulation/compensation norm (grammar) between a speaker and a listener, which could result in occasions where a listener employs compensation and still fails to extract from a heavily coarticulated speech segment 'the same pronunciation target intended by the speaker'.

### 1.4. Hypotheses and Research Questions

One aim of the current study was to replicate and elaborate three findings from previous work. First, it aims to replicate Ohala and Feder's (1993) findings of perceptual compensation for /u/-fronting in an alveolar context and that the effect can be induced not only by acoustic context but also by assumed, or 'restored',

context. Second, it aims to replicate Lindblom and Studdert-Kennedy's (1967) findings of speech rate effects on compensation. Third, it aims to elaborate the findings of Harrington et al. (2008) by testing for systematic differences in category boundaries across listeners in a single speech community. Following Harrington et al., compensation is operationally defined as shift in category boundary on the stimulus continuum as a function of phonetic context. This effect is tested by using an /i/-/u/ continuum in [b_p] and [d_t] contexts. Thus, the following four hypotheses were formulated:

H1:    The /i/-/u/ category boundary will be shifted towards the /i/-end (more stimuli 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 with 'acoustic' context).

H2:    The /i/-/u/ category boundary will be shifted towards the /i/-end when the vowel is heard in the 'restored' alveolar context as compared to the 'restored' bilabial context (i.e. positive compensation effect with 'restored' 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:    The listeners vary systematically in terms of the category boundary; that is, a group of listeners whose category boundary is closer to /i/-end than the other group in one condition systematically exhibits the same difference in other conditions.

In addition to testing these hypotheses, the study addresses an issue of exactly how context alters perception of a target sound. One particularly heated debate concerns whether compensation is achieved by gestural perception (e.g., Liberman & Mattingly 1985, Fowler 1986) or by spectral contrast (e.g., Lotto, Kluender & Holt 1997). The spectral contrast view has strong support such as the finding that both speech and non-speech contexts induce comparable compensation effects (Holt & Kluender 2000), dismissing the need to access the representations for speech production. However, there have been ample demonstrations that compensation can be mediated by non-acoustic cues such as visual information (Fowler 2000) and lexical status of the context (Elman & McClelland 1988), suggesting that spectral contrast alone does not account for the full range of effects.

The current study aims to contribute to this debate by investigating the effect of a natural precursor phrase on the degree of compensation and by examining the reaction time (RT). Will an additional precursor induce greater degree of compensation, by possibly encouraging the listeners to engage in a speech mode of processing? Does the precursor provide facilitative or impeding effects on phoneme identification that can be observed in RT data? These are the additional questions asked in the current study.

## 2.    Methods

A series of experiments were conducted to test the four hypotheses: 1) compensation for [u]-fronting in an alveolar context, 2) compensation for [u]-fronting in a 'restored' alveolar context 3) an increasing degree of compensation with an increase in speech rate of the stimuli, and 4) systematic individual variation in compensation. Data on the effect of a precursor on compensation and RTs were also collected. The basic experimental design follows.

### 2.1.    Participants

Thirty-two native speakers of American English (18 female, 14 male), aged between 19 and 45 years, participated as listeners. Of these 32 participants, 27 evaluated themselves as a 'speaker' of one or more foreign languages with varying proficiency. For example, the number of subjects who indicated having above 'good' speaking ability of foreign language on the 5-level scale (poor, fair, good, excellent, native-like) are: seven for Spanish, three for French, one for Italian, one for Korean, and one for Chinese (dialect unspecified). The participants were paid $10 upon completion of the experiments.

### 2.2.    Stimuli

Two sets of ten-step CVC continua ranging between minimal pairs *beep-boop* (/bip/-/bup/) and *deet-doot*[2] (/dit/-/dut/) were created in a fast, medium, and slow speech rate by concatenating a natural onset stop burst, a re-synthesized steady-state vowel without formant transitions, and a natural coda stop burst by using Praat (Boersma & Weenink 2007). In addition, another two sets of CVC continua where Cs were replaced by white noise were created in the medium speech rate. Rather than being fully synthesized, the vowels were re-synthesized by using the speaker's voice source so that the stimuli would sound natural when played after a precursor spoken by the same speaker.

The process of vowel re-synthesis was as follows. First, a 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 and iterated to obtain a vowel of 80 ms. From this vowel, 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 a pre-emphasis frequency of 50 Hz; and applying inverse filtering of the LPC object on the original sound. Next, the source signal was applied to a filter, which was specified by five center frequency values and corresponding bandwidth values, to create a steady-state vowel. For example, peak frequencies (and bandwidths, in Hz) for the /i/-end of the continuum are: F1=375 (50), F2=1200 (100), F3=2319 (150), F4=3500 (200), and F5=4500 (250). The nine other vowels were made by apply-

---

[2] The listener's judgment might be biased toward real words (i.e. towards 'beep'/'deet' vs. 'boop'/'doot'). However, this bias should be neutralized after the practice session, where equal numbers of the end stimuli were heard.

ing nine different filters that had identical specifications except the F2 and F3. These values are given in Table 1.

**Table 1.** F2 and F3 values on the ten-step vowel continuum ranging 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 |

To these vowels, a smooth amplitude contour was added by applying a Hamming window to the first and the last 15 ms. Then, an F0 contour (130 Hz at onset and 90 Hz at offset) was added to obtain natural-sounding vowels. 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).

Another set of ten-step continuum was created by replacing the onset and the coda bursts with 20 ms of white noise (NVN stimuli). The white noise had the steady amplitude envelope and its amplitude matched to the peak amplitude of the vowel.

Two more sets of /bip/-/bup/ and /dit/-/dut/ continua, one with 100 ms of vowel duration (CVC duration = 190 ms) and the other with 120 ms of vowel duration (CVC duration = 210 ms), were created by using comparable methods used for creating abovementioned CVC stimuli.

Three kinds of precursor phrase were created by altering the duration of the phrase "I guess the word is", spoken by the same speaker. The final durations of the 'fast', 'medium', and 'slow' precursors were 800 ms, 1000 ms, and 1200 ms. These values were determined by trial-and-error as listening to the precursor followed by each of the fast, medium, and slow CVC stimuli for naturalness.

### 2.3. Procedure

The experiment consisted of four blocks. The first block tested the 'acoustic' context and 'restored' context conditions. In the 'acoustic' context condition, each of the two sets of medium (190 ms) CVC stimuli was presented in isolation (i.e. no precursor). The two sets of NVN stimuli were presented in isolation in the 'restored context condition'. In the remaining blocks, each of the 'fast', 'medium', and 'slow' rate conditions was tested in random order by using the 'fast', 'medium', and 'slow' CVC stimuli. Within a block, [dVt] and [bVp] stimuli were tested in separate sub-blocks in counterbalanced order. In each sub-block, each of the ten stimuli from the continuum was presented four times in random order for two-alternative forced-choice (/CiC/ or /CuC/) tasks. In the first block, each of the ten CVC stimuli and NVN stimuli was presented in the same sub-block in random

order. There were 160 trials in this block—10 stimuli x 4 trials x 2 contexts (alveolar or bilabial) x 2 types (CVC or NVN). In the subsequent three blocks, there were 80 trials in each block (10 stimuli x 4 trials x 2 contexts). The procedure was the same for all blocks except that in the last three blocks, each stimulus was played after the precursor phrase of matching speech rate. Each block was preceded by a short practice block to familiarize the listeners with the task and stimuli.

The manner of stimuli presentation and response logging was identical across blocks. The listener was sitting in front of a computer monitor and a five-button response box. The 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 to each stimulus over headphones carefully and to enter a response as quickly as possible.

The written instruction on the computer monitor serves the purpose of leading the listener to believe, during the trials in the first block, that each of the NVN stimuli is identical to its CVC counterpart (i.e., either [dVt] or [bVp] stimulus depending on the sub-block in which the stimuli are presented) except that onset and coda bursts were masked with the white noise, inducing 'restored phoneme' effect.

## 2.4.    Data Analysis

A repeated-measures ANOVA was used to test for effects of context (alveolar vs. bilabial, tested separately for each block and further for 'acoustic' and 'restored' condition for the first block), precursor phrase (with precursor vs. without precursor), and speech rate (fast vs. medium vs. slow). The effect of precursor was tested by comparing the response to the CVC stimuli in the first block (CVC duration = 170 ms) and that to the 'medium' rate block, where the identical CVC stimuli were presented after precursor phrase. For the context effect, dependent variables were category boundary obtained on the /dit/-/dut/ and /bip/-/bup/ continua (e.g., see Fig. 1, panel C) separately for each listener.  For the precursor effect and the speech rate effect, dependent variables were 'distance' between the boundaries on the two continua (e.g., compare Fig. 1, panel C and D) obtained separately for each listener. Category boundary was defined as the stimulus number (1-10) for which responses with /i/ or /u/ were at 50%. Following Harrington et al. (2008) and Lotto et al. (1996), the 50% boundary was calculated using probit analysis.

In order to test for systematic individual variation in category boundary, the listener was classified as a "Fronter" or a "Backer" based on the results in the first block, only from trials in the 'acoustic context' condition, and then the difference in category boundary between the two groups was tested for the other three blocks. Those who had mean category boundaries (midway between the boundaries on alveolar and bilabial continua) below 4.5 were classified as Fronters, the others as Backers. Each group had 16 listeners.

Reaction time was measured as the time between stimulus onset and the moment when the button press was made. When there was no response, there was no RT data. Out of 12800 total observations (32 listeners x 400 trials per listener), there were 71 (0.55%) missing responses.

## 3.    Results

Figure 1 presents the percentage of /u/-responses for the [dVt] and [bVp] stimuli and mean 50% boundary locations on each continuum in the 'acoustic' context condition (A), 'restored' context condition (B), 'fast' rate condition (C), 'medium' rate condition (D), and 'slow' rate condition (E). Significant context effects (i.e. compensation) were observed in the 'medium' rate [$F(1, 31) = 4.98$; $p < 0.05$] and the 'fast' rate [$F(1, 31) = 18.27$; $p < 0.01$] conditions (panel D & panel C). Speech rate had a significant effect on compensation[3] [$F(2, 62) = 7.15$; $p < 0.01$] (panel C vs. panel D vs. panel E). Although there was a discernible increase of the degree of compensation in the 'with precursor' condition when compared with the 'no precursor' condition (panel C vs. panel A), this difference was not significant [$F(1, 31) = 1.69$; $p = 0.20$].

Figure 2 shows the distribution of category boundary on /dVt/ and /bVp/ continua in four conditions (no-precursor, 'slow' rate, 'medium' rate, and 'fast' rate) by the Fronters and the Backers. As mentioned earlier, the listeners were classified as a Fronter or a Backer based on the results in the no-precursor condition, and then the difference between the group's mean boundaries was tested for each of the other three conditions.  There is a systematic pattern where 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: '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' [$t(30) = -3.61$; $p < 0.01$], 'fast' [$t(30) = -2.79$; $p < 0.01$].

The mean RT from /u/-responses to the /dVt/ and /bVp/ stimuli (#4 and above) in all but 'restored' context condition are presented in Figure 3. The RT data obtained in the 'restored' context condition was not considered because the purpose of this measure was to test the effect of presence or absence of precursor and speech rate and all other condition needed to be kept constant.  Also, the RT data obtained from stimuli #1 to #3 are not considered because the percentages of /u/-responses were very low (6% for stimulus #3 in the 'slow' condition, less than 3% for all other stimuli), indicating that most of these responses were errors.

---

[3] Overall category boundary shifted toward the /u/-end as speech rate increased. This was an unexpected result, whose explanation will be sought in the future. Nevertheless, that relative /u/-bias became stronger as speech rate increased is taken as evidence for greater compensation caused by speech rate manipulation.

**Figure 1.** 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. 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).
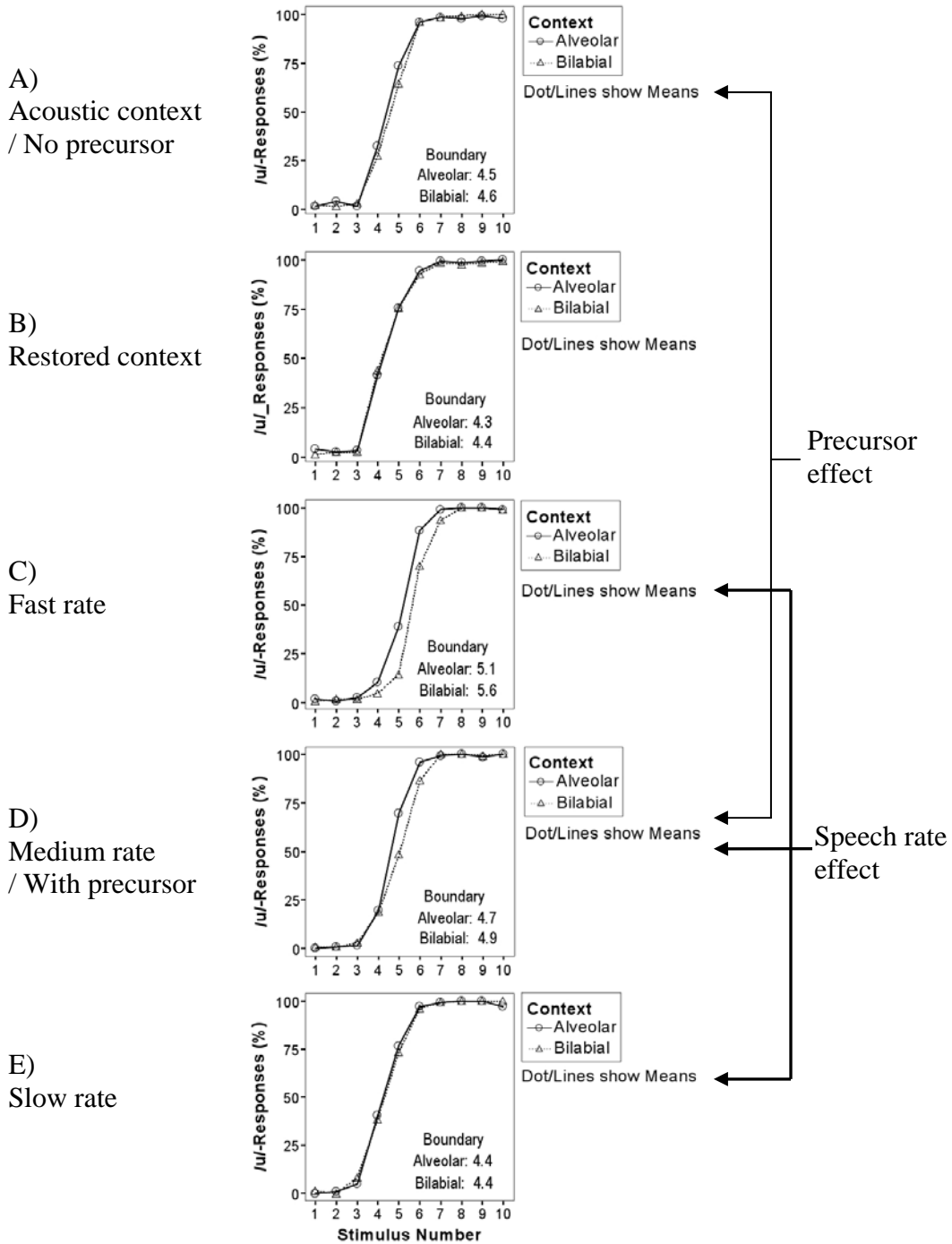


A)
Acoustic context
/ No precursor

B)
Restored context

C)
Fast rate

D)
Medium rate
/ With precursor

E)
Slow rate

Precursor
effect

Speech rate
effect

**Figure 2.** Distribution of category boundary by Backers (white) and Fronters (striped) on a /dVt/ continuum (left two plots in a given condition) and on a /bVp/ continuum (right two plots in a given condition), 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.
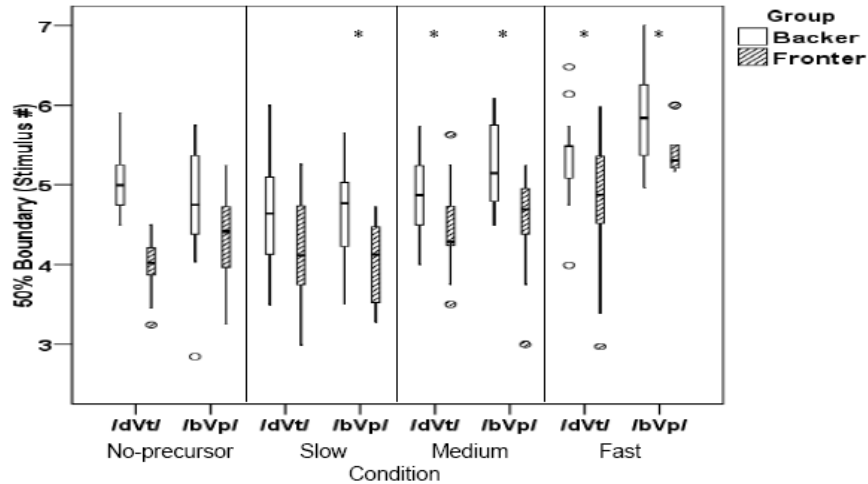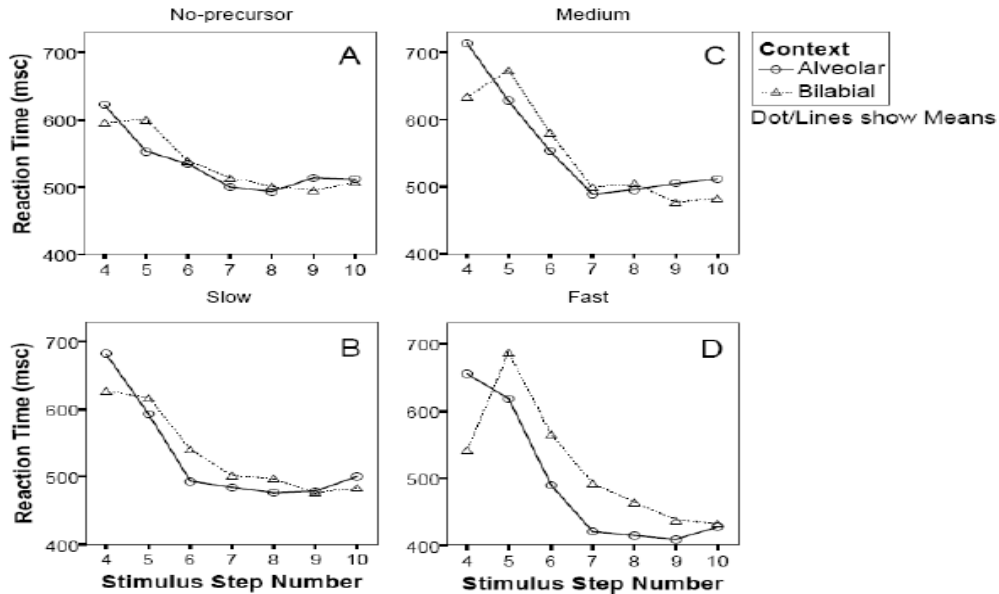


**Figure 3.** Mean RT as a function of stimulus number (increasing number corresponds to decreasing F2 and F3) on a /dVt/ continuum (solid) and a /bVp/ continuum (dotted) in four conditions: A) no precursor (i.e. acoustic context); B) slow rate; C) medium rate; and D) fast rate.

Some patterns emerge from these results. For stimuli near the /u/-end, RT tends to be shorter for /dVt/ stimuli than for /bVp/ stimuli. The pattern is somewhat inconsistent for the stimuli near the /i/-end, but in 'medium' and 'fast' rate conditions, RT is generally shorter in [bVp] stimuli than in [dVt] stimuli. The RT data for stimuli #6 to #10, where within-condition RTs are relatively invariant across stimuli, show that mean RTs are markedly shorter in the 'fast' rate condition than in other conditions. Finally, the RT data show much smaller across-stimuli variation in the 'no precursor' condition as compared to the 'medium' rate condition; another way to interpret these data is that for stimuli near the category boundary, RTs are much shorter in the 'no precursor' condition than in other conditions. These results are interesting since the target CVC stimuli were identical in duration in these conditions.

## 4.     Discussion

The results generally supported the hypotheses. The hypothesis (H1) that listeners compensate for the fronting of a high back vowel in an alveolar context was supported in the 'fast' and 'medium' speech rate conditions. These results confirm the robustness of the compensation effect on phoneme identification. The hypothesis (H2) that the compensation, as defined by the boundary shift, and speech rate are positively correlated was supported by the results showing that the degree of boundary shift increased monotonically from the 'slow' to 'medium' and 'fast' rate conditions. The hypothesis (H3) that there is systematic individual variation in category boundary was supported in five out of six comparisons: the Fronters who had the category boundary closer to the /i/-end than the Backers in the 'no precursor' condition consistently had it this way in other conditions as well.

Now, how we can explain these results? As mentioned earlier, the mechanism of compensation is still a matter of debate. A general auditory approach explains the effect in terms of spectral contrast (Holt & Kluender 2000), while gestural approaches explain the effect in terms of the listener's ability to recover from the speech signal either 'intended gestures' through analysis-by-synthesis using an 'innate vocal-tract synthesizer' (Liberman & Mattingly 1985) or actual gestures directly from the acoustic signal (Fowler 1986). Since the current study does not offer decisive evidence in support of one approach over the other, the rest of the discussion considers how the results can be explained in terms of either approach and considers implications for future research.

The speech rate effect can be explained in several ways. One is in terms of the listener's knowledge about speech production, which enables the listener, as discussed by Lindblom and Studdert-Kennedy (1967: 839), to 'predict' the degree of coarticulation from the perceived speech rate. This explanation is compatible with the analysis-by-synthesis approach. Short RTs for /u/-responses in the alveolar context in the 'fast' condition might be taken as support for this analysis: in a context where a strong fronting effect is expected, low-frequency prominence might be mapped onto a back vowel more quickly than in other contexts.

Lindblom and Studdert-Kennedy (1967: 840) also discussed another possibility, which is more compatible with the general auditory approach—the tendency to overshoot or extrapolate the formant values for short stimuli with rapidly changing spectra. Although the vowels had steady-state formants, the spectral peak in the preceding stop burst and the beginning of vowel formants might provide sufficient dynamism to cause perceptual extrapolation so that the vowels are perceived as having lower resonant frequencies than they actually have. This scenario predicts both a stronger compensation in shorter stimuli and a null effect in longer stimuli, where there is a sufficient steady-state region so that as the analysis proceeds the extrapolated resonant frequency would match the actual frequency.

Yet another possible explanation, which also accounts for the null effect in the 'no precursor' condition, is that a listener actively or passively varied the unit of analysis across conditions. Since the decision-making in the current task ultimately depended on vowel identification, the listener would pay more attention to the vowel rather than entire CVC stimuli if doing so was possible. A longer segment would be more easily isolated than a shorter one, making it easier to pay attention selectively to that segment. The more the vowel was dissociated from the context, the less of an effect the context would have on phoneme identification.

In this scenario, the effect of the presence of a precursor would be to bias the listener to process the auditory information in larger chunks—syllables or words—since this is how listeners parse acoustic events in natural communication situations (speech mode of parsing). As RT indicates task difficulty, relatively constant RTs in the 'no precursor' condition might support the idea that the absence of a precursor indeed enables listeners to isolate the vowel from the context.

If this scenario is true, then it would be of interest to further investigate whether the same effect is obtained with a non-speech precursor that has speech-like prosody, such as periodic amplitude modulation repeating itself with a syllable-sized period. Positive results from such experiment would suggest the possibility for a general priming effect to account for variation in the unit of parsing, eliminating the necessity of a speech mode of parsing.

The results showing gradually increasing RT toward the category boundary strongly support the idea that the phonemic category is not simply an abstract unit; it has well-defined internal structure, with category members varying in 'category goodness' (Miller 1994). Further, the fact that listeners are all native speakers of American English and yet systematically vary in phoneme categorization suggests that knowledge of phonemic entities is personal, being acquired through linguistic experience unique to each individual. This is a micro-level counterpart of what Harrington et al. (2008) found in their age-group comparison. A listener's compensation grammar varies across individuals, and one possible source of this state of affair is the individual difference in previous linguistic experience.

These findings have significant implications for a model of sound change. Imagine a situation where a group of listeners hear a word *dude* with heavily fronted /u/. Some hear the vowel as an instance of ordinary /u/, while others hear it as a sound different from ordinary /u/ (say, /y/, for convenience); that is, some listen-

ers pick up an 'accent' from such utterance. Even if these listeners compensate for fronting on another occasion and pick up /dud/ from that instance of the same word, these listeners would possess the mental representations /dyd/ and /dud/ as synonymous forms of *dude*. Such listeners, when they turn into speakers, would be in a position to utter this word either as /dud/ or /dyd/, whichever sounds better to the ears of the speaker. In this way, even when a listener employs perceptual compensation, speech variation may plant a seed of sound change in the auditory field of the listener.

## References

Boersma, Paul, and David Weenink. 2007. Praat: doing phonetics by computer (Version 4.5.15) [computer program]. Retrieved February 18, 2007, from http://www.praat.org.

Barnes, Jonathan. 2006. *Strength and Weakness at the Interface: Positional Neutralization in Phonetics and Phonology*. Berlin: Mouton de Gruyter.

Beddor, Patrice S., and Rene A. Krakow. 1999. Perception of Coarticulatory Nasalization by Speakers of English and Thai: Evidence for Partial Compensation. *Journal of the Acoustical Society of America* 106(5):2868-2887.

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette, and Andrew Garrett. 1998. The Origins of Consonant-Vowel Metathesis. *Language* 74(3):508-556.

Diehl, Randy L., Andrew J. Lotto, and Lori L. Holt. 2004. Speech Perception. *Annual Review of Psychology* 55(1):149-179.

Elman, Jeffrey L., and James L. McClelland. 1988. Cognitive Penetration of the Mechanisms of Perception: Compensation for Coarticulation of Lexically Restored Phonemes. *Journal of Memory and Language* 27:143-165.

Fowler, Carol A. 1986. An Event Approach to the Study of Speech Perception from a Direct Realist Perspective. *Journal of Phonetics* 14:3-28.

Fowler, Carol A., Julie M. Brown, and Virginia A. Mann. 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.

Harrington, Jonathan, Felicitas Kleber, and Ulrich Reubold. 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(5):2825-2835.

Holt, Lori L., and Keith R. Kluender. 2000. General Auditory Processes Contribute to Perceptual Accommodation of Coarticulation. *Phonetica* 57:170-180.

Hyman, Larry M. 1976. Phonologization. In Alphonse Juilland (ed.), *Linguistic Studies Offered to Joseph Greenberg, Second Volume: Phonology. Studia linguistica et philologica* 4:407-418. Saratoga: Anma Libri.

Ladefoged, P., and D. E. Broadbent. 1957. Information Conveyed by Vowels. *Journal of the Acoustical Society of America* 29(1):98-104.

Liberman, Alvin M., and Ignatius G. Mattingly. 1985. The Motor Theory of Speech Perception Revised. *Cognition* 21:1-36.

Lindblom, Björn, and Michael Studdert-Kennedy. 1967. On the Role of Formant Transitions in Vowel Recognition. *Journal of the Acoustical Society of America* 42(4):830-843.

Lotto, Andrew J., Keith R. Kluender, and Lorri L. Holt. 1997. Perceptual Compensation for Coarticulation by Japanese Quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America* 102(2):1134-40.

Miller, Joanne L. 1994. On the Internal Structure of Phonetic Categories: A Progress Report. *Cognition* 50:271-285.

Ohala, John J. 1981. The Listener as a Source of Sound Change. In C. S. Masek, R. A. Hendrick, and M. F. Miller (eds.), *Papers from the 17th Regional Meeting of the Chicago Linguistic Society: Parasession on Language and Behavior*, 178-203. Chicago: Chicago Linguistic Society.

Ohala, John J. 1993. The Phonetics of Sound Change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*, 237-278. London: Longman.

Ohala, John J., and Deborah Feder. 1994. Listeners' Identification of Speech Sounds is Influenced by Adjacent "Restored" Phonemes. *Phonetica* 51:111-118.

Repp, Bruno H. 1982. Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception. *Psychological Bulletin* 92:81-110.

Yantis, Steven. 2001. Visual Perception: An Overview. In Steven Yantis (ed.), Visual Perception: Essential Readings, 1-18. Philadelphia: Psychology Press.

Yu, Alan C. L. 2004. Explaining Final Obstruent Voicing in Lezgian: Phonetics and History. *Language* 80(1):73-97.