Prosody-induced acoustic variation in English stop consonants

Hansook Choi
University of Illinois at Urbana-Champaign, U.S.A.
E-mail: hchoi5@uiuc.edu

ABSTRACT

This work compares the acoustic patterns of English stop consonants (here /p/ vs. /b/) in three distinct prosodic contexts: accented position, domain-initial position, and domain-final position. Target consonants are analyzed from utterance (U) or Intonational Phrase (IP)-initial position, U/IP-final position, and U/IP-medial position, with and without contrastive focal accent. VOT and F0 at the onset of the following vowel are measured as the main acoustic correlates of voicing in English from the speech of six male American English speakers. The statistical results indicate that the accentual effect is greater and more consistent than the boundary effect on acoustic patterns. The distribution of acoustic measures also shows a greater distinction between the two contrastive stops when accented. Positional variation is not very distinctive, and the enhanced contrast marking in domain-initial position predicted from articulatory studies is not observed in the present setting.

1. INTRODUCTION

Articulatory studies report distinctive patterns in prosodically prominent contexts, such as initial or final in the prosodic domain, or stressed/accented syllables. The distinctive articulations in such positions are understood to anifest 'hyperarticulated' [1] or 'strengthened' [2] gestural movement. The articulatory patterns in accented syllables, however, are not identical to those in domain-boundary positions. Both the onset and vowel of an accented syllable are produced with longer, faster, and larger articulatory movements. The effects of a prosodic boundary on articulation are local to the segment at the absolute domain edge, and condition movements that are longer but not necessarily larger and faster [2, 3, 4]. In addition, the domain-initial position conditions different speech patterns from the domain-final position in that the domain final position shows lengthening in the rhyme preceding the boundary [5] while domain initial gestural strengthening is reported mostly for the onset consonant after the boundary [2, 3, 6].

These discrepancies lead to the expectation that the acoustic variation in domain boundary positions will display different patterns from the variation in accented positions. We further expect that prosodic effects will be more robust in accented segments and domain-initial consonants than in consonants in domain-final CV syllables when comparison is restricted to the patterns of consonants in syllable-onset.

These prosodically driven articulatory patterns have been further understood to enhance linguistic contrast. Generally, the articulatorily distinctive movement in the prosodically prominent positions is considered to enhance linguistic contrast syntagmatically. Consonants show greater constriction and vowels show greater opening gestures, and thus the contrast between consonants and vowels is enhanced in accented or boundary contexts [2, 7]. Prosodic prominence also facilitates a better paradigmatic contrast within categories. Acoustic studies reported that the distance among contrastive stop consonants in acoustic space increases when they are in prosodically prominent positions [8, 9].

The present study is designed to explore the acoustic correlates of the reported prosodically-driven articulatory variation. Specifically, it examines acoustic variation in boundary position (here utterance (U) or Intonational phrase (IP)-initial, medial and Final) and accented position (here contrastive focal accent) with two acoustic measures (Voice Onset Time (VOT) and Fundamental Frequency (F0)) which signal voicing in English stop consonants [10]. Based on these acoustic measures, the voicing contrast among English stops in three prosodically prominent positions will be compared.

2. METHODS

A. Material

Target consonants were analyzed from the initial CV syllable of a set of 8 English words (with the target syllables underlined): <u>po</u>ttery, <u>bo</u>tany, <u>pe</u>ter, <u>bea</u>ter, <u>pe</u>tter, <u>be</u>ttor, <u>pah</u>, <u>bah</u>. The first 6 words were presented in two prosodic contexts, located in phrase-initial and phrase-medial positions of a carrier sentence, and the last 2 words were used only in phrase-final position of a carrier sentence. The IP and utterance boundaries coincided in the carrier sentences. Two sets of dialogs were constructed for each carrier sentence: one with contrastive focus on the target word, and another with contrastive focus on another word in the sentence.

B. Subjects and Procedure

Speech data were collected from 6 male American English speakers from Chicago area, all undergraduate students in University of Illinois at Urbana-Champaign. The subjects had no phonetic training or knowledge and reported no speaking impairment.

 ISBN 1-876346-48-5 © 2003 UAB

To control for focal and positional conditions, sets of dialogue were designed with the target sentence as an answer to a question represented as in Table I. Subjects were supposed to take the role of the answerer. Test sentences were presented in a quasi-random order after a short rehearsal session. Sentences were separated on the basis of the prosodic condition of the target word, with the conditions listed in Table 1. The four sentence groups were presented in separate blocks in order to sustain reasonable prosody, and each block was repeated 5 times over the course of the experiment, varying within-block sentence order across repetitions, for a total of 840 tokens. The recording procedure was done in a sound-attenuated booth in the phonetics laboratory in University of Illinois at Urbana-Champaign. All sound stimuli were recorded with a head-mounted microphone and a Tascam DAT recorder.

i) *INITIAL-FOCUSED*

Q: Botany is the main subject of your book?

T: Pottery is the main subject of my book.

ii) *INITIAL-NONFOCUSED*

Q: Pottery is the title of your book?

T: Pottery is the main subject of my book.

iii) *MEDIAL-FOCUSED*

Q: A yellow botany book was on the table?

T: A yellow pottery book was on the table.

iv) *MEDIAL-NONFOCUSED*

Q: A yellow pottery book was on the desk?

T: A yellow pottery book was on the table.

TABLE I. Samples of the speech corpus. In a dialogue with a question (Q) and the target sentence (T) as an answer, subjects took the role of an answerer. The CV target syllables are underlined.

C. Measurements

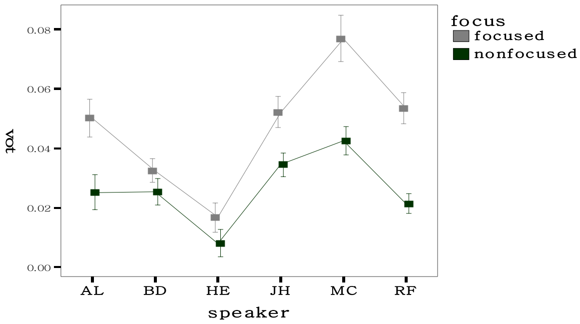
The recorded sounds were transferred to a PC and analyzed with the Praat program [11]. The duration from the stop release to the onset of the second formant in the following vowel were measured as VOT. Fundamental frequency values were manually calculated from the mean period over the initial 3 periodic cycles of the wave forms after the stop release. The calculated values were compared with results from the autocorrelation pitch analysis function in Praat, which showed similar values but reported some missing values.

3. RESULTS

A. Acoustic variation by prosody

Results from 3-way ANOVAs (Position x focus x CV syllable) performed for each subject show that focus is a significant factor for subjects in both VOT and F0 measurements except for two subjects' F0 (AL ($F(1,112)= 76.739$, $p<.0001$ for VOT; $F(1,112)=238.64$, $p<.0001$ for F0), BD ($F(1,112)=4.835$, $p<.05$ for VOT; $F(1,112) =38.742$, $p<.0001$ for F0), HE ($F(1,112)=33.085$, $p<.0001$ for VOT; $F(1,112)=98.171$, $p<.0001$ for F0), JH ($F(1,112)=131.559$, $p <.0001$ for VOT; $F(1,112)= 122.234$, $p<.0001$ for F0), MC ($F(1,112)=146.648$, $p<.0001$ for VOT), and RF ($F(1,112) = 239.299$, $p<.0001$ for F0)).

(a) Voice Onset Time (Sec) in two focal conditions



(b) Fundamental Frequency (Hz) in two focal conditions

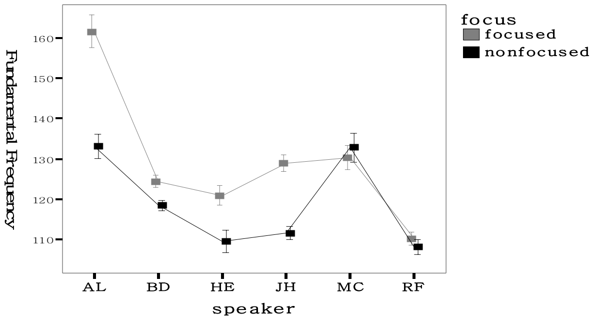
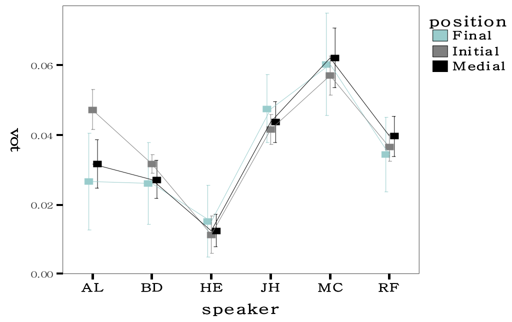


Figure I. Six speakers' VOT and F0 distributions in two different focal conditions. Error bars show Mean +/-1.0 Standard Error of Mean, and boxes show Means.

On the other hand, position is a significant factor for one speaker's VOT and three speaker's F0 (AL ($F(1,112)=18.928$, $p<.0001$ for VOT), BD ($F(1,112)= 25.939$, $p<.0001$ for F0), HE ($F(1,112) = 197.762$, $p<.0001$ for F0) and MC ($F(1,112)= 17.725$, $p<.0001$ for F0)).

In general, VOT and F0 show higher values under focus for the speakers in the present study as in Figure I. When focused, VOT values show an increase for all speakers with a clear separation from those for nonfocused pairs as in Figure I (a). Similarly, Figure I (b) shows that F0 is higher under focus except two speakers (MC and RF) for whom VOT values in focused and in nonfocused conditions are not very distinct. In contrast, the difference due to positional variation is not very clear compared with focal variation.

(a) Voice Onset Time (under positional variation)



(b) Fundamental Frequency (under positional variation)

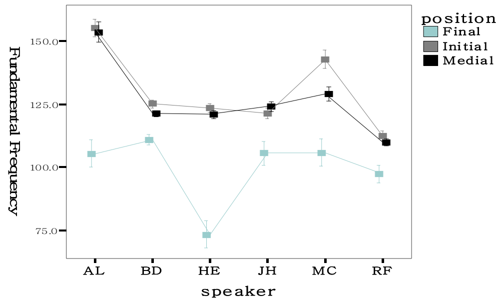


Figure II. Six speakers' VOT and F0 distributions in initial, medial and final positions. Error bars show Mean +/-1.0 Standard Error of Mean, and boxes show Means.

The VOT distribution in Figure II (a) does not show any consistent discrepancies based on positional difference. The initial condition displays the highest values for AL and BD, whereas the medial condition is highest for MC and RF, and final condition is highest for HE and JH. However, figure II (b) shows that F0 patterns are conditioned by positional variation. Turkey post hoc comparison shows that the final position displays significantly lower F0 values than the initial and medial positions. Yet the increased F0 in the initial position is not found to be statistically different from F0 in medial and final positions, as predicted by articulatory findings. The difference between initial and medial condition is not significant. Furthermore, the final position is not expected to be less prominent than the medial position. Therefore, lower F0 in final position might be understood as a reflection of declined intonation at the end of a phrase. The same factor may account for the minor increase in in initial position from some speakers.

B. Contrast marking

The increased acoustic values under focus lead to an enhanced contrast between voiced and voiceless stops in the 2-dimensional space of VOT and F0, as seen in Figure III. Except for RF, both voiced and voiceless labial stops show an increase in the two acoustic measurements under focus. This can be understood as a syntagmatic strengthening of the stop category. Nonetheless, all speakers show a greater separation between voiced and voiceless stops when focused. The reverse movement of RF also leads to a better contrast marking under focus.

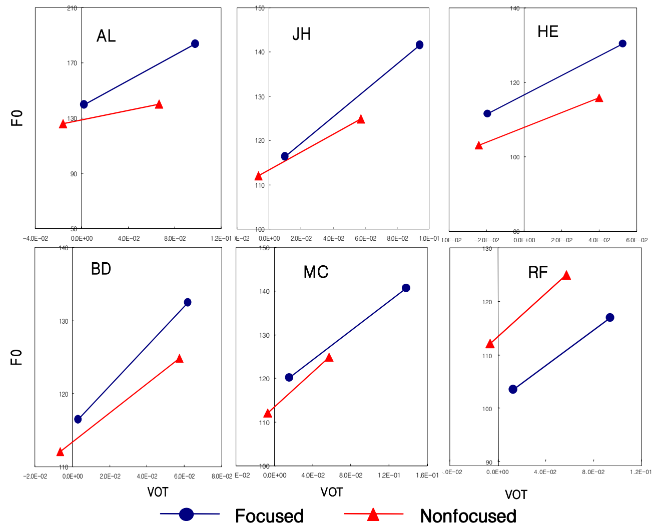


Figure III. The stop voicing contrast in 2-dimensional space of VOT and F0 in focused and nonfocused condition. Acoustic measures in medial position were used to establish focal variation without an added positional effect. Each bar connects means values of a voicing pair, voiced stops and voiceless stops in each condition, and diagrams mark coordinates of mean VOT and F0.

Contrary to the enhanced voicing contrast under focus, the contrast in a relatively prominent position (namely domain initial position) is not always better marked. The stop contrast in UP/IP-initial in Figure IV, which is marked with circles, does not show a greater distance between voiced and voiceless stops compared to the medial or final position. The contrast in the final position, though predicted to be relatively less prominent, is as clear as that in initial position for most speakers. Some speakers show rather a decreased contrast in the initial position compared with the medial and final positions, which is also against the prediction.

In summary, accentual effects are generally more obvious and consistent than positional effects in the present acoustic measurements across all the speakers. VOT shows a clear increase under focus for all subjects (especially for voiceless stops) and F0 shows higher values under focus for some speakers. A positional difference is not very noticeable in the two acoustic measures. Furthermore, the expected discrepancy in initial and final position was not confirmed in the present study. In contrast marking, focus induces a better separation between the two contrastive

ISBN 1-876346-48-5 © 2003 UAB

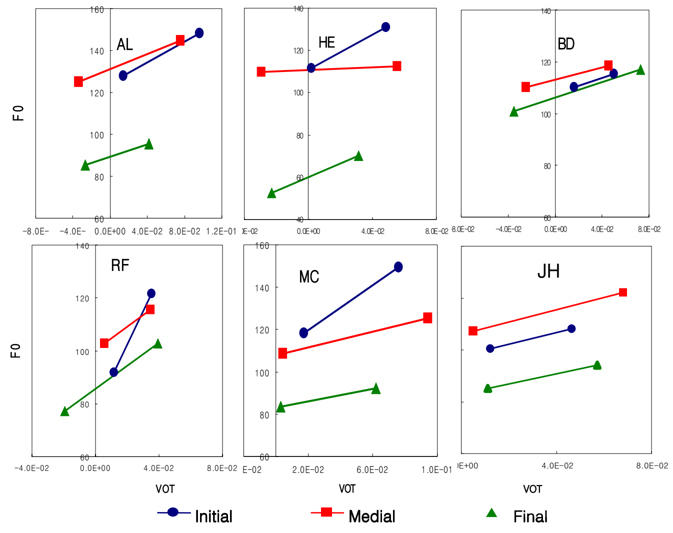


Figure IV. Stop voicing contrast in 2-dimensional space by VOT and F0 in the initial, medial and final position. Acoustic measures from nonfocused tokens were used to note positional variation without focal effect.

stops as well as an increase in acoustic measures in both stops. The expected positional variation is not visible in the contrast marking by the two acoustic measures.

4. DISCUSSION

The focus condition is well marked by VOT and partly by F0. The durational cue, VOT, shows more accentual variation compared to F0. This discrepancy can be explained partly by the fact that F0 is an important feature to express intonational features as well as a cue for the voicing contrast in English. The F0 perturbation, therefore, would be more restricted according to the intonational and accentual patterns.

In contrast to the effect of focus, acoustic variation due to phrasal position is not very clear in the present setting. The present experiment studied word-initial CV syllables which varied only in the position within IP or U. This result might indicate that the main acoustic variation lies in the positional difference in domains lower than IP. If that is the case, the utterance and IP boundary would not involve remarkable variation. Further research is required since there are inconsistent reports on acoustic and articulatory patterns in IP and Utterance from diverse languages [2, 6, 8, 9, 12, 13].

The present study reveals that the acoustic cues to stop voicing are enhanced under focus both for voiced and voiceless stops, and that the contrast between voiced and voiceless stops is increased under focus. Therefore, focused tokens have stronger cues both for manner (stop) and for voicing than unfocused tokens. Position in phrase does not condition any variation in the cues to the voicing contrast. The prediction of asymmetry between initial and final edges of prosodic domains in patterns of acoustic strengthening is not supported.

REFERENCES

[1] De Jong, K., "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *Journal of Acoustic Society of America (JASA)*. 97, pp. 491-504, 1995.

[2] Fougeron, P. & Keating, P., "Articulatory strengthening at edges of prosodic domains," *JASA* 101(6), pp. 3728-3740, 1997.

[3] Cho, T., *Effects of Prosody on Articulation in English*, PhD dissertation. UCLA, 2001.

[4] Edwards et al., "The articulatory kinematics of final lengthening," *JASA* 89(1), pp. 369-382, 1991.

[5] Wightman et al., "Segmental durations in the vicinity of prosodic phrase boundaries," *JASA* 91(3), pp. 1707-1717, 1992.

[6] Keating, P. et al., "Domain-initial strengthening in four languages," *UCLA working papers in Phonetics* 97, pp. 139-151, 1999.

[7] Edwards, J. & Beckman, M., "Articulatory timing and the prosodic interpretation of syllable duration," *Phonetica* 45, pp. 156-174, 1988.

[8] Cho, T. & Jun, S., "Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean," *Chicago Linguistics Society*, 36(1), pp. 31-44, 2000.

[9] Hsu, C & Jun, S., "Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic?" *UCLA working papers in Phonetics* 96, pp. 69-89, 1998.

[10] Whalen, et al., "F0 gives voicing information even with unambiguous voice onset time," *JASA* 93(4), pp. 2152-2159, 1993.

[11] Boersma, P. & Weenink, D., *The Praat program*, Version 4.0.13, University of Amsterdam, 2002.

[12] Jun, S., *The Phonetics and Phonology of Korean Prosody*, PhD dissertation, Ohio State University, 1993.

[13] White, L., *English Speech Timing*, PhD dissertation, The University of Edinburgh, 2002.