The effect of accent on the acoustic cues to stop voicing in Radio News speech

Jennifer S. Cole, Hansook Choi, Heejin Kim, and Mark Hasegawa-Johnson
University of Illinois at Urbana-Champaign
E-mail: j-cole5@uiuc.edu, hchoi5@uiuc.edu, h-kim17@uiuc.edu, jhasegaw@uiuc.edu

ABSTRACT

Data from three acoustic cues to stop voicing is analyzed for the effect of phrasal accent in a corpus of read radio news speech from a single speaker of American English. The results show that VOT, Closure Duration and F0 are significant cues to voicing for stops in this corpus, though the acoustic patterns vary by place of articulation. There is a significant effect of phrasal accent on each cue, with increased values for both voiced and voiceless stops, described here as a pattern of syntagmatic strengthening. Patterns of paradigmatic strengthening, with enhanced voicing contrast under accent, are observed only in sporadic cases. The results indicate that accent effects in radio news speech are similar to those reported for "laboratory speech" from prior phonetics studies.

1. INTRODUCTION

There is increasing evidence for the significant effect of stress and accentual prominence on variation in speech production. Articulatory studies show that segments in stressed or accented syllables are more strongly articulated ---with gestures exhibiting greater magnitude, duration and peak velocity--- than segments in unaccented syllables (Cho 2001; deJong 1995; Edwards et al 1991; Beckman et al 1992, among others). Evidence of prominence-induced strengthening is also found in acoustic studies in the form of increased durations, increased energy and spectral features that indicate more peripheral constriction locations under stress and accentual prominence (Pierrehumbert & Talkin 199x; deJong 1998; Cooper 1991; Beckman & Edwards 1994).

The work cited above provides an important foundation for understanding the phonetic expression of prosodic structure, but it is limited in that the evidence is based exclusively on the analysis of "laboratory" speech --- speech produced by subjects reading scripts prepared by the experimenter, in the absence of a compelling, natural discourse context. An important question is how stress and accent affect variation in speech produced under more natural speaking conditions, where stress and accent convey information critical for speech understanding.

This paper reports on a study of the effects of phrasal accent on acoustic variation, based on evidence from acoustic patterns in the speech of an announcer from the Boston University Radio News corpus. [1] This report focuses on the acoustic features that mark the laryngeal articulation of stop consonant voicing. The analysis is based on speech that is read from a prepared script and produced for a genuine communicative function in which the announcer must use prosody to signal the structure and information content of the news story. The speech is not controlled by the experimenter for the purposes of the phonetic investigation. The findings presented here are part of our larger study of the patterns of acoustic variation conditioned by prosody in the Radio News speech corpus.

2. ACCENTUAL STRENGTHENING

Our analysis of accent-induced variation in patterns of stop voicing is based on measurements from three acoustic features: VOT, Closure Duration, and F0. [2] Voiced and voiceless stops in English have been found to differ in these features, with voiceless stops characteristically exhibiting greater VOT, longer closure duration, and higher F0 than their voiced counterparts at the same place of articulation (Laver 19xx).

The works cited above on the effect of stress and accentual prominence reveal two different effects that prosodic prominence can have on both articulatory and acoustic variation. *Paradigmatic strengthening* increases the phonetic distinctiveness of phonologically contrastive sounds, while *syntagmatic strengthening* involves an enhancement of articulatory and acoustic features that affects contrastive sounds similarly. With this distinction in mind, we hypothesize three possible effects of phrasal accent on the acoustic cues to stop voicing.

Hypothesis 1: Increased acoustic durations (syntagmatic strengthening). Accent conditions a general increase in the magnitude and duration of articulatory gestures, which are reflected in observations of increased acoustic duration of

[1] Distributed by the Linguistics Data Consortium.
[2] The duration of a preceding vowel is known to be another important cue to stop voicing in English, but was not studied due to the variability of syllable shapes in our corpus.

accented segments (de Jong 1998; Beckman & Edwards 1994). Based on these findings, accented stop consonants are predicted to exhibit a general increase in closure duration, and increased positive VOT values. Longer intervals of voicing during closure (lead voicing) might also be expected with voiced stops for some speakers, but an increase in this parameter will be independently limited by aerodynamic factors.

Hypothesis 2: Increased acoustic energy (syntagmatic strengthening). Accent typically induces an increase in acoustic energy that affects the noise portions of all segments in the accented syllable. Because an increase in acoustic energy may be accompanied by an increase in F0, accent may be found to condition an increase in the F0 measures of both voiced and voiceless stops.

Hypothesis 3: **Enhanced acoustic distinctions between voiced and voiceless stops** (paradigmatic strengthening). Accent conditions a general strengthening of supralaryngeal articulations in English, resulting in greater acoustic distinctions between contrastive sounds (Cho 2001; de Jong 1995). If contrast enhancement is a primary effect of accent, then voiced and voiceless stops should exhibit enhanced distinctions in some or all of the acoustic features that cue the voicing contrast. More specifically, VOT, F0, and Closure Duration are all expected to have increased values for voiceless stops and decreased values for voiced stops under the condition of phrasal accent.

3. METHODS

Stop consonant tokens were extracted from the speech of announcer F3 of the Boston University Radio News database. This database includes a transcription aligned at the word level with the acoustic signal, and a set of prosodic labels that identify phrase-level pitch accents (among other prosodic features) based on the ToBI labeling conventions (Beckman & Ayers 1993). 330 tokens of the stop consonants /p,t,k,b,d,g/ were extracted and subject to acoustic analysis. The target consonants are all taken from pre-vocalic, word-initial position (#**C**V) in two conditions of phrasal prominence: **Accented** and **Unaccented**. Consonants in the Accented condition are from initial stressed syllables that are marked with a pitch accent, while those in the Unaccented condition include tokens from both stressed and unstressed initial syllables all of which lack pitch accent. Table 1 summarizes the distribution of the analyzed stop tokens in Accented and Unaccented contexts, distinguishing also phrase-initial from phrase-medial positions.

Segmentation of the target consonants and labeling for acoustic measurements was done manually based on spectrogram, waveform and listening. Acoustic measurements of VOT, Closure Duration (CD), and F0 at vowel onset were taken from each token as follows**.** VOT measurements were taken from the release burst to the nearest zero-crossing preceding the onset of the second formant in the following vowel.[3] CD was measured from the left by the ending point of the second formant of a preceding vowel, and from the acoustic release of a preceding stop if observable. If no acoustic landmark of a release was observed in the context of a preceding consonant, no CD measurement was taken. The right edge of the CD was marked at the stop release for all tokens. F0 measurements were calculated manually as the mean duration of the first three periods of the vowel following the stop.

		p	b	t	d	k	g
Initial	Accented	1	1	1	2	5	1
	Unaccented	1	12	13	9	3	0
Medial	Accented	17	24	17	9	24	5
	Unaccented	19	40	59	21	32	14
	Total = 330	38	77	90	41	64	20

Table 1. Distribution of stops in Accented and Unaccented syllables.

3.2. RESULTS

Box plots of the results for each acoustic measure are shown in Figures 1-3. VOT (Fig. 1) is found to be consistently greater for voiceless stops than for voiced, and accent has the effect of raising VOT for all stops except /g/. CD (Fig. 2) differ for voiced and voiceless stops among the labials and alveolars, but it is the voiced stop that is found to have longer closures, contrary to our expectation. Accented stops have longer closure durations than unaccented stops for all stops except /g/. F0 (Fig. 3) is higher for voiceless stops among the labials and velars, and accent has a raising effect on F0 for all stops except /b/.

Statistical analysis was done separately for each acoustic measure using 3-way ANOVA with the independent variables of Voicing (voiced, voiceless), Accent (accented, unaccented), and Place of Articulation (labial, alveolar, velar). Results were first considered separately for data from stops in phrase-medial position, in order to see the effect of Accent independent of a possible effect of strengthening in initial position of the Intonational Phrase. In phrase-medial position, there were significant effects of both Accent and Voicing. Effects of Accent were found for all three acoustic measures: **VOT** [$F(1,174) = 11.308$, $p<.001$]; **F0** [$F(1,269) = 17.10$, $p<.001$]; **CD** [$F(1,205) = 20.17$, $p<.001$] Effects of Voicing were also found for all three acoustic measures: **VOT** [$F(1,174) = 40.521$, $p<.001$];

[3] Tokens of voiced stops with negative VOT (voice lead) were labeled as such, but are excluded from the results reported here.

F0 [F(1,269) = 29.050, p<.001]; and **CD** [F(1,205) = 12.073, p<.001]. There was a mildly significant effect of Place of Articulation on **VOT** [F(2,174) =3.410, p<.05] and a stronger effect on **CD** [F(2,205) = 9.555, p<.001]. There were no significant interactions between Voicing, Accent and Place.

Another 3-way ANOVA was performed for each acoustic measure pooling data from phrase-initial and phrase-medial positions, with independent variables of Accent, Voicing and Position in Intonational Phrase. Place of articulation was not considered in this analysis. Note that CD was not measured for phrase-initial tokens due to segmentation ambiguity. Just as with the data from phrase-medial position, the pooled data showed significant effects for both Accent and Voicing for all three acoustic measures. The results from Accent were: **VOT** [F(1,215) = 7.596, p<.005]; **F0** [F(1,322) = 12.930, p<.001]; **CD** [F(1,213)= 40.266, p<.001]. The results from Voicing were: **VOT** [F(1,215)= 38.554, p<.001]; **F0** [F(1,322) = 40.767, p<.001]; and **CD** [F(1,213)= 38.145, p<.001]. There were no significant effects of Position in Phrase. Weaker but ignificant effects of interaction were found for Accent*Voicing with **VOT** [F(1,215) = 4.286, p<.05], and with **F0** [F(1,322) =7.011, p<.05].

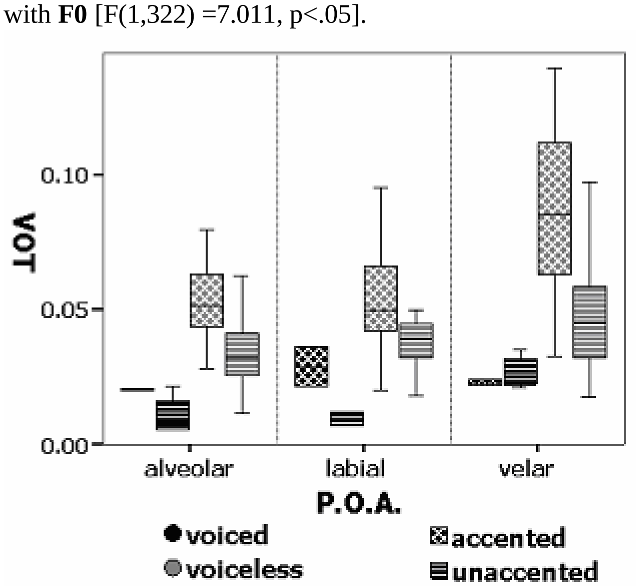

Figure 1. VOT values (in ms.) for voiced and voiceless stops under accented and unaccented conditions, grouped by place of articulation. The boxed region contains the inner two quartiles of the distribution.

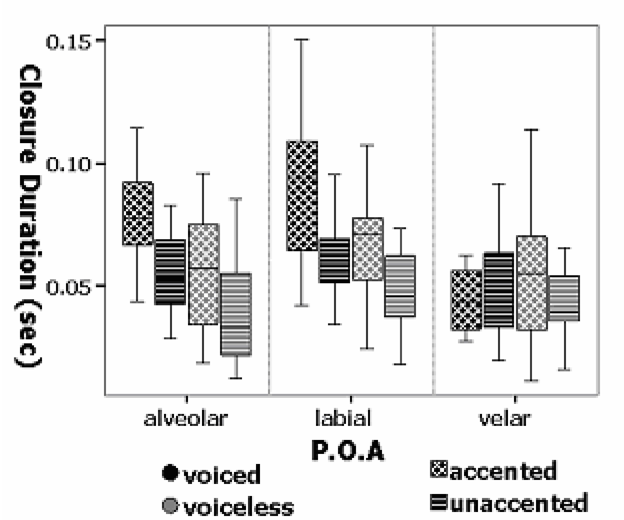

Figure 2. CD values (in seconds) for voiced and voiceless stops under accented and unaccented conditions.

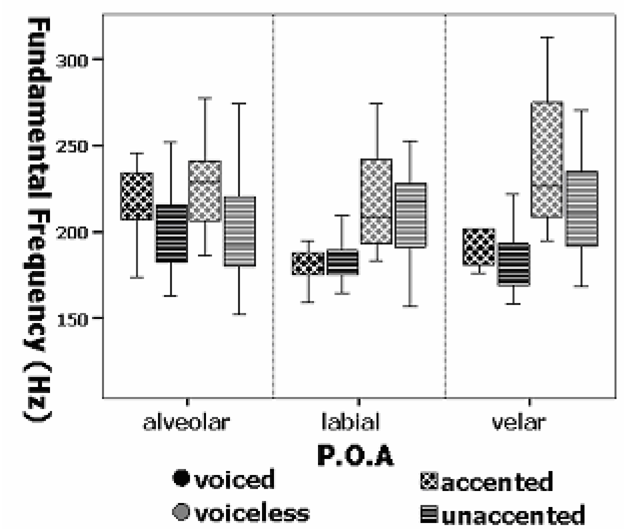

Figure 3. F0 values for voiced and voiceless stops under accented and unaccented conditions.

4. DISCUSSION

The results provide compelling evidence of the strengthening effect of phrasal accent on the acoustic cues to stop voicing. We note first that the three acoustic cues to voicing analyzed here are all found to differ significantly in mean values for voiced and voiceless stops, suggesting that these cues indeed play an important role in signaling the phonological voicing feature for speaker F3. The CD distinction was opposite to what we expected based on phonetic descriptions of English. In our data CD was longer for voiced stops than for voiceless, and this pattern was consistent under both accent conditions. We note that the #CV context from which our stop tokens are taken differs from the within-word [? VCV] contents in which

CD is reported as longer for voiceless stops (c.f., Lisker 1957). Our findings indicate that the way in which CD functions as cue for voicing varies as a function of syllable and stress context. Observing the distribution boxplots for each acoustic cue, we note substantial overlap in the distribution of acoustic values for voiced and voiceless stops within a place of articulation category. Thus, no single cue by itself provides an effective means of categorizing stops by the Voice feature across all the places of articulation. For this speaker, VOT appears to be the most salient distinction between voiced and voiceless stops.

The statistical analysis demonstrates a significant effect of accent on VOT, F0 and CD, and the distribution boxplots clearly indicate that the effect is one of strengthening: all measures show increased values under the accent condition for both voiced and voiceless stops at most places of articulation. Furthermore, the accent effect is the same in initial and medial position in the Intonational Phrase. No independent effect of Position in Phrase is found in these data, but we note that the number of stops in phrase-initial position was very low and their distribution was uneven across the three places of articulation. Additional data from initial position would be required to fully consider the interaction of accentual and positional effects on voicing cues.

The uniform direction of the accent effect (increased values for all measures) for both voiced and voiceless stops indicates that the effect is one of syntagmatic strengthening. The statistical results of data pooled from phrase-initial and phrase-medial indicate an interaction between Accent and Voicing. The boxplots of the VOT and F0 measures show that accent effects are larger and more consistent across places of articulation for voiceless stops than for voiced. The accent effect on CD is greater for the voiceless stop only in the case of the bilabial. This nonuniformity in accentual strengthening of voiced and voiceless stops results in paradigmatic strengthening of the voicing contrast under accent in sporadic cases: for velars with the VOT cue; for velars and labials with the F0 cue; for bilabials with the CD cue. The paradigmatic strengthening effect is, however, clearly secondary to the stronger and more consistent effect of syntagmatic strengthening.

The patterns of accentual strengthening demonstrated in these data provide clear evidence that strengthening affects the laryngeal gesture in ways comparable to the supralaryngeal strengthening shown in prior research. The observed increases in VOT suggest a longer duration for the laryngeal gesture of spread vocal folds. The observation of longer CD values under accent is further evidence of strengthening of the supralaryngeal gesture of the stop under accent. The fact that VOT increases under accent while CD is simultaneously increased indicates that the laryngeal gesture of spread vocal folds during stop closure is significantly lengthened under accent. Overall, the CD and VOT results indicate that accent production involves lengthening of both supralaryngeal and laryngeal gestures and a phasing between the two gestures that enhances the VOT distinction for voiced and voiceless stops. Finally, the increased F0 values observed under accent may result from increased acoustic energy, as stated earlier, but may also be related to the presence of a High tone in a Pitch Accent that marks the accented syllable. Analysis of the tonal type of the pitch accent could test this claim, but was not performed in the present study.

In conclusion, we find evidence of accentual strengthening of the acoustic cues for stop voicing in this corpus of Radio News speech. The acoustic effects of strengthening are similar to those reported in experimenter-controlled “laboratory” speech, suggesting that the mechanism for strengthening is similar, at least for read speech, under a range of communicative functions.

REMEMBER: change all F0 to mark subscript

(Cho 2001; deJong 1995; Edwards & Beckman 1988; Edwards et al 1991; Beckman et al 1992; Harrington et al 1995; Cooper 1991).