"Ok, whatever!": phonetic variability of stops and flaps in spontaneous and careful speech

Natasha Warner

(Dept. of Linguistics, Univ. of Arizona, Tucson)

Benjamin V. Tucker

(Dept. of Linguistics, Univ. of Alberta, Edmonton, Canada)

Submission date: October 29, 2010

Running Title: Phonetic variability of stops and flaps

Corresponding Author:

Natasha Warner

PO Box 210028

Department of Linguistics

University of Arizona

Tucson, AZ 85721-0028

U.S.A.

E-mail: nwarner@u.arizona.edu

Abstract:

Variability is perhaps the most notable characteristic of speech, and it is particularly noticeable in spontaneous conversational speech. The current research examines how speakers realize the American English stops /p, k, b, g/ and flaps ([ɾ] from /t, d/), in casual conversation and in careful speech. Target consonants appear after stressed syllables (e.g. 'lo**bb**y') or between unstressed syllables (e.g. 'humani**t**y'), in one of six segmental/word-boundary environments. This work documents the degree and types of variability listeners encounter and must parse. Findings show greater reduction in connected and spontaneous speech, greater reduction in high frequency phrases (but not within high frequency words), and greater reduction between unstressed syllables than after a stress. Although highly reduced productions of stops and flaps occur often, with approximant-like tokens even in careful speech, reduction does not lead to a large amount of overlap between phonological categories. Approximant-like realizations of expected stops and flaps in some conditions constitute the majority of tokens. This shows that reduced speech is something that listeners encounter, and must perceive, in a large proportion of the speech they hear.

PACS: 43.70Fq

I. INTRODUCTION

One of the hallmarks of human speech is variability, especially in casual speech (Koopmans-van Beinum, 1980; Dalby, 1986; Greenberg, 1999; Johnson, 2004; Shattuck-Hufnagel & Veilleux, 2007), but even within careful speech (e.g. Richter, 1930; Blumstein & Stevens, 1979; Lisker, 1985; Sussman et al., 1991). Even for something as seemingly basic as stop consonants, when we examine connected, spontaneous, casual speech, we encounter surprising variability. Where we expect a stop or a flap, we may find an approximant, perhaps with a clear dip in amplitude and weakening of formants, or perhaps with only a slight trace of an amplitude dip. Occasionally, the consonant may be completely deleted (Greenberg, 1999). For example, 'that is' may have just a slight weakening of the vocalic material where one expects /t/ (Figure 1B below), or 'yesterday' may be pronounced as [jɛʃɛʲ] with the /d/ and several other segments deleted. Yet, these highly reduced consonants are often perceptible, and very reduced speech in context leads to good recognition (Arai, 1999; Ernestus et al., 2002). This paper asks how much of normal speech is made up of such reductions: are reduced sounds the normal way for humans to communicate, or are they a relatively uncommon exception, perhaps mostly limited to very casual speech? Furthermore, given that the use of speech for communication requires making distinctions, is the variability in spontaneous speech strong enough to (partially) neutralize phonological distinctions?

Barry & Andreeva (2001) document intervocalic stops in several languages being realized as approximants, Horna (1998) confirms this for American English flaps, and the authors have noticed themselves sometimes producing approximants instead of flaps in conversation. Flaps may be particularly subject to reduction because even a clear flap has a very short closure, so that even a little target undershoot may produce an approximant. Crystal & House (1982, 1988a, b) point out that approximately half of the stops in their read connected speech lacked either a closure or a burst. How often are expected stops not stops or even flaps at all? If reduction of consonants is the norm rather than the exception, listeners must be parsing these reduced consonants, which do not provide the same acoustic cues as careful productions. The current work provides detailed information on how speech sounds are realized, both in casual conversation and careful speech, to determine whether reduction is primarily a casual speech phenomenon. It also investigates preceding stress, segmental environment, word boundaries, and word frequency. Much past literature asks whether and where speakers flap (de Jong, 1998; Patterson & Connine, 2001; Riehl, 2003; Fukaya & Byrd, 2005; Eddington & Elzinga, 2008; Herd et al., 2010; etc.). The present study asks instead what the range of acoustic variability is for expected stops and flaps. It also provides the production complement to perception data in Tucker (2007, in press) and Warner et al. (2009).

Work on speech styles, including spontaneous and reduced speech, has increased recently (e.g. Kohler, 2001; Johnson, 2004; Keune et al., 2005; Pluymaekers et al., 2005a, 2005b; see Warner, 2011 for literature), but methodology for obtaining and measuring spontaneous speech is still developing (Warner, in press). Since realizations are highly variable, measurements must be flexible enough to be meaningful both for a true voiceless stop and for an approximant, even to measure just the stop phonemes. One cannot focus on VOT, for example, if many tokens lack a burst. Thus, most literature on reduced speech either uses duration as the main measure (as exemplified by Pluymaekers et al., 2005a), or compares a phonetic transcription of the speech to a lexical listing, and uses number of transcription symbols altered or deleted as the measure (as exemplified by Greenberg, 1999; Johnson, 2004). These methods supply useful information, but they do not tell us how reduction is realized acoustically. What are the possibilities for how intended stops or flaps can be pronounced, and how often do these possibilities occur?

This paper has both descriptive and theoretical purposes. The descriptive purpose is to document how often speakers produce reduced forms, focusing on intervocalic stops and flaps (which we will refer to for clarity as stops and flaps based on phonological categories, regardless of how they are realized in a particular token). We use a variety of preceding and following sounds, to determine whether identity of neighboring phonemes affects speech reduction. These descriptive data have implications for second language teaching and other applications of phonetics, but are also important for basic phonetic knowledge. Phonetic science is sufficiently well established that we might expect that we know what the acoustic characteristics of stops and flaps are (Port, 1977; Zue & Laferriere, 1979; Crystal & House 1988a, b; de Jong, 1998; Fukaya & Byrd, 2005; Warner et al., 2009). However, examining waveforms and spectrograms of spontaneous conversation, one is often surprised at what realizations are possible, and even quite intelligible (examples at http://www.u.arizona.edu/~nwarner/reduction_examples.html, and cf. Figure 1 below). The current work gives us a broader understanding of variability.

This work has several theoretical goals. The first is to determine the scale of low-level phonetic variability relative to phonological distinctions. Since a single segment /d/ can be produced as [d̥, d, ɾ, ɾ̥, ɾ̞, ɨ], it is possible that when the full range from casual to careful speech is considered, the variability of phonetic reduction could cause substantial overlap among distinct segments. The second theoretical goal is to determine what factors affect speech planning and articulation. Several authors have made headway in determining whether word frequency, predictability, dialect, and gestural overlap affect degree of reduction (Jurafsky et al., 2002; Keune et al., 2005; Pluymaekers, 2005 a, 2005b; Aylett & Turk, 2006; Mitterer & Ernestus, 2006; Gahl, 2008). The current work extends this investigation to sub-segmental acoustics, and focuses on the role of speech style. Using three speech styles from each speaker (casual conversation, connected reading, and isolated word-lists) gives a picture of speech style that neither most corpora nor most controlled experiments, using a single speech style, can provide. The latter two speech styles are chosen as the ones most commonly used in phonetics research, and the former as a style common in daily life. The final goal is to determine what acoustic information listeners have available in reduced speech for word recognition and sentence processing, in order to model recognition of reduced speech.

II. METHODS

A. Speakers

Recordings of 13 speakers (2 male, 11 female) were analyzed for this study. (Additional male volunteers were excluded because of divergent language backgrounds.) All were native speakers of American English with no reported speech or hearing problems, and most were from the Southwest. Most were monolingual English speakers without any other language spoken in their environment during childhood. Two had some childhood exposure to another language (Canadian French, Chinese), but were not fluent in the other language. Speakers with any Spanish exposure in the home were excluded because of approximation of intervocalic /b, d, g/ in Spanish (Hualde, 2005). All speakers were young adults recruited from an introductory linguistics course at the University of Arizona, and they received a small amount of extra credit.

B. **Procedures and materials**

Speakers were recorded in a sound-protected booth holding a phone conversation with a friend or family member, then reading stories containing target words, and then reading a list of isolated words. Speakers wore a high-quality head-mounted microphone on the opposite side from where they habitually hold a telephone, and were recorded at 44,100 Hz on a digital CD recorder. The speakers used either their cell phones or a landline to call a friend or family member (except one speaker who called the second author), and conversed for approximately 10 minutes on any topic they wished. Speakers were asked to avoid large amounts of backchanneling, so that enough of their speech would be recorded. The recording was made through the head-mounted microphone, not the telephone, and the telephone served only to create a casual speech situation with a well-known interlocutor. Interlocutors were not consented subjects, and were not recorded.

This method allowed us to obtain high-quality soundbooth recordings of casual spontaneous speech (Warner, in press). The topics ranged from extremely informal (e.g. boyfriends, drinking games) to moderately serious (e.g. a grandmother's medical decisions, course registration, child-raising). After the conversation, the experimenter asked the speakers to read five short stories out loud (approximately one paragraph each), and then to read a list of individual words. The words were not placed in frame sentences (but see below for the word-boundary condition). The tasks were ordered from most casual to most careful so that having just read a word list could not influence speakers to think about their pronunciation during the conversation. Only one repetition of the stories and wordlists was recorded for each speaker, to allow for a large number of items, which provide more independent data than multiple repetitions of fewer items.

For the word-list, 10 English words were chosen in each condition (Table 1), to the extent that familiar words exist in the lexicon. Words were selected using the NewDic lexicon (http://dingo.sbs.arizona.edu/~hammond/lsasummer03/newdic.txt). The conditions were defined by stress (following a stressed syllable, 'lo**bb**y' vs. between unstressed syllables, 'humani**t**y'), phoneme (p, t, k, b, d, g; as in 'ha**pp**en,' 'quo**t**a,' 'bro**cc**oli,' 'ha**b**itat,' 'free**d**om,' 'yo**g**a'), and segmental/word boundary environment (six conditions representing the common flapping environments for American English, Table 1). The stress factor has only the levels post-stress and inter-unstressed, because the primary interest was in flapping environment, and flapping occurs before unstressed syllables (Kahn, 1976; Patterson & Connine, 2001; Fukaya & Byrd, 2005).

INSERT TABLE 1 APPROXIMATELY HERE

Most of the word boundary items were placed into short phrases (e.g. 'wri**t**e a letter') to make them plausible. The phrasal (word-boundary condition) items were randomized and grouped at the end of the list, and the rest of the list (isolated words) was also randomized. A few words were read and measured twice, and some speakers missed a few words, but because of the overall size of the dataset, this is not a concern.

The post-/r/ environment supersedes others, in that a target consonant following /r/ is not allowed to appear in the other environments. This is the only post-consonantal flapping environment, so it might show different patterns of reduction. The consonantal nature of syllabic /l/ and the differences in tongue position among [i, ɚ, ə] could also impact consonant reduction, as could the presence of a word boundary. It was not possible to locate 10 items for every condition, and some conditions do not occur. From each condition, 1-5 items were incorporated into short, semi-formal stories (Table 1).

The dataset comprises over 9200 measured stop/flap tokens, an average of more than 700/speaker, providing a detailed picture of variability in these sounds. Together, the three tasks provide samples of casual conversation, connected but read speech, and isolated read speech from each speaker. Conversations do not contain the same words as the two read speech styles, nor the same words (or even conditions) across speakers. However, this method does allow us to obtain a large number of consonants in similar environments in all three speech styles from all speakers. See Warner (in press) for discussion of methods for reduced speech, and of item-matching and casualness.

C. Measurements

To obtain acoustic information about consonant realizations, we measured several aspects of duration, intensity, voicing, and formant structure, using Praat (Boersma & Weenink, 2010). For each target consonant, one of the authors or an assistant labeled the onset and offset of the consonant and the peak of the intensity contour during the preceding and following vowels, and labeled whether voicing ceased, whether there was a burst, and whether the second and/or third formants remained strong throughout the consonant ("strong formants") or remained present but weak ("weak formants").

The criteria reflect flexibility for labeling segments on a continuum from true voiceless stops to weak approximants (Warner, in press), and are exemplified in Figure 1. Onset of the consonant was considered to be at cessation of clear F2 if that occurred, or at the point where F2 most suddenly lost amplitude (i.e. gray instead of a clear black band in the spectrogram). In the absence of any sudden weakening of F2, if the intensity curve showed a sudden drop, that point was used. When lacking even that, the point half-way through the fall in the intensity curve was chosen as onset of the consonant. Tokens lacking any dip in intensity and any weakening of formants did not have visible consonants, so were coded as too deleted to be measured. For the offset of the consonant (onset of following vowel), the same criteria were applied in reverse, i.e. onset of clear F2. This method should give results similar to the method Hualde et al. (2010) describe based on maximum slope of intensity, but with human checking of all tokens.

INSERT FIGURE 1 APPROXIMATELY HERE

Cessation of voicing was determined from the waveform, including even low-amplitude closure-voicing as voiced. Presence of a burst was identified as a sudden broadband noise in the spectrogram, even if it was faint. To label intensity maxima, the intensity calculation window was set to a short 21.33 ms (minimum pitch of 150 Hz in Praat) to prevent obscuring flaps' brief intensity changes (Boersma & Weenink, 2010). The intensity minimum between the two maxima was located automatically.

Judgment of formant strength was categorical, as a supplement to the continuous measure of intensity drop. If F2 or F3 had clear, dark vertical striations visible in the spectrogram at all time points during the consonant, the token was coded as having "strong formants." If it did not, but had either light F2 or F3 present at all time points during the consonant, then it was coded as having "weak formants," but this distinction was later found to be less reliable than other measures, so is not reported. Tokens with a simultaneous complete gap at some point in F2 and F3 were classified as "no formants."

From these measurements, five dependent variables were calculated: the duration of the consonant, the intensity difference between the average of the preceding and following vowels' intensity maxima and the consonant's minimum intensity, cessation/continuation of voicing, presence/absence of a burst, and presence/absence of any formants (weak or strong). The first two are continuous variables, the rest categorical. Remeasurement of approximately .5% of the data by the first author (originally labeled by others) confirmed reliability: average absolute value of error for consonant duration was 4 ms, average absolute value of error for intensity difference was .04 dB, and errors in even the categorical measures constituted 2.1-8.3% of the data.

If speakers misread a target word in a way that altered the consonant's condition (e.g. [ˈfoʷ**ʤ**i] for 'fo**g**ey,' [əˈ**b**æ**k**əs] for 'a**b**a**c**us'), the token was excluded. Other mispronunciations (e.g.[jukə] for 'yu**cc**a,' still a pre-schwa /k/) were accepted. If a neighboring vowel was deleted, the token was used if material surrounding the target was sonorant (e.g. [bɹa**k**li] for 'bro**cc**oli' included, 'typi**c**al' with [p**k**] excluded). If the speaker paused at the consonant (with or without a glottal stop), produced clear secondary stress on the following vowel, or if it was not possible to tell whether the target consonant was the deleted one of two consonants (e.g. [pɹa**b**li] 'pro**b**ably'), the token was excluded.

For the conversational data, we located all words where a consonant matching the conditions above would be expected in clear speech. Slightly broader segmental environments were used: any full following vowel, rather than only /i/, was allowed for the pre-full-vowel condition (e.g. 'Fri**d**ay,' 'pho**t**os'). For the word-boundary condition, the following vowel had to be unstressed but did not have to be schwa (e.g. 'tha**t** I,' 'ge**t** it'). The same exclusions as for reading data were applied. Stress environment was assessed from perceived stress, not lexical stress, because of varying intonation.

III. RESULTS

In spontaneous speech one obtains unequal numbers of tokens, with no tokens in some conditions. Even in read speech, lexical gaps create this problem. Therefore, the analysis requires examining subsets of data for each question and averaging over some factors. Below, we explain which data can be used to answer each question.

A. Segmental and word-boundary environment

We performed a three-factor ANOVA for each of the six dependent variables using the factors speech style (story- and list-reading only), segmental/boundary environment (all six levels), and phoneme (/b, d, g, p, t, k/), for the post-stress data only. However, six speakers did not produce any measurable words in a few story-reading conditions. Therefore, the remaining seven speakers were used for the 3-factor design. For follow-up analyses of list-reading data only, all 13 speakers were used, to increase power. For simple effects tests within the story data, all speakers with data for that set of conditions were used. Figure 2 shows the data for intensity difference, graphing the seven speakers with full datasets for story-reading and all speakers for list-reading. For all ANOVAs, the data were averaged over items within each condition (by-subjects analysis).

INSERT FIGURE 2 APPROXIMATELY HERE

The ANOVA for intensity drop showed significant main effects of all factors (segmental environment: $F(5,30)=5.41$, $p<.005$ $\eta^2_p=.44$, style: $F(1,6)=63.64$, $p<.001$ $\eta^2_p=.91$, phoneme: $F(5,30)=96.05$, $p<.001$ $\eta^2_p=.94$) as well as a significant three-way interaction ($F(25,150)=2.12$, $p<.005$ $\eta^2_p=.26$) and one significant two-way interaction (phoneme by environment: $F(25,150)=3.05$, $p<.001$ $\eta^2_p=.34$). (We use η^2_p, partial eta squared, as a measure of effect size.) Therefore, the two-factor design of phoneme by segmental/boundary environment was tested for each speech style separately. Both styles had a significant two-way interaction (story-reading: $F(25,150)=2.46$, $p<.001$, list-reading: $F(25,300)=4.51$, $p<.001$), so the simple effect of environment was tested for each phoneme in each speech style separately. Environment affected how strongly intensity dips, and thus how stop-like the consonantal portion of the signal appears, for all phonemes in the word-list style (/b/: $F(5,60)=4.14$, $p<.005$, /d/: $F(5,60)=5.11$, $p<.005$, /g/: $F(5,60)=8.94$, $p<.001$, /p/: $F(5,60)=4.33$, $p<.005$, /t/: $F(5,60)=3.91$, $p<.005$, /k/: $F(5,60)=6.51$, $p<.001$) and for three in the story-reading style (/b/: $F(5,45)=3.65$, $p<.01$, /d/: $F(5,60)=3.60$, $p<.01$, /p/: $F(5,60)=3.06$, $p<.02$). Figure 2 shows that this effect may partially reflect a smaller dip in intensity (hence greater consonant reduction) before schwa. However, it is otherwise difficult to determine specific, consistent patterns that account for the effects. There is some suggestion that consonants may be less reduced before [ɚ], but this effect is inconsistent across conditions.

The ANOVA for consonant duration (data not graphed here), the other continuous dependent variable, showed a similar statistical outcome of significant main effects (environment: F(5,30)=13.85, p<.001 η^2_p=.70, style: F(1,6)=59.12, p<.001 η^2_p=.91, phoneme: F(5,30)=288.89, p<.001 η^2_p=.98), significant two-way interactions, and a significant three-way interaction for the overall ANOVA (F(25,150)=3.17, p<.001) η^2_p=.35,. There were significant two-way interactions for both follow-up tests of environment by phoneme (story-reading: F(25,150)=4.01, p<.005, list-reading: F(25,300)=15.03, p<.001); and significant simple effects of environment, in this case for every phoneme in each speech style (story: /b/ F(5,45)=4.52, p<.001, /d/ F(5,60)=6.50, p<.001, /g/ F(5,45)=5.21, p<.005, /p/ F(5,60)=8.04, p<.001, /t/ F(5,45)=4.02, p<.005, /k/ F(5,55)=2.92, p<.03; list: /b/ F(5,60)=13.74, p<.001, /d/ F(5,60)=15.44, p<.001, /g/ F(5,60)=20.45, p<.001, /p/ F(5,60)=19.53, p<.001, /t/ F(5,60)=14.46, p<.001, /k/ F(5,60)=48.10, p<.001). However, these effects are often not in the same direction as for intensity drop, although the greater reduction before schwa is often apparent.

We do not present the statistical results for the three categorical dependent variables, but they are similar in showing interactions, simple effects of segmental environment in some conditions, and a lack of consistent direction of the environment effect. However, the greater reduction before schwa is relatively consistent across measures, phonemes, and reading styles.

Because of the varied patterns of the simple effects of segmental/boundary environment, and the potential for familywise error in a thorough analysis of this six-level factor, we do not attempt to test pairwise comparisons within this factor. We note that consonants may reduce more before schwa than in most other environments, reflecting a more approximant-like pronunciation. In order to allow testing of the other factors of interest in the experiment, we will collapse over the segmental/phrasal environment factor in subsequent analyses. All results, averaged first over items within a condition, then over speakers and then segmental/boundary environments in order to weight each environment equally, appear in Table 1. While it is difficult to obtain conversational tokens of each phoneme before syllabic /l/ between a stressed and an unstressed syllable, for example, it is possible to obtain conversational tokens of most phonemes post-stress in *some* segmental/boundary environment, so averaging over the environment factor allows the further analyses. This provides very stable data from many data points for all of the post-stress read speech conditions and many others.

B. Effect of preceding stress

Some literature claims that flapping is more likely in a stressed-unstressed syllable sequence such as "sta**t**us" than between unstressed syllables as in "capi**t**alist" (Zue & Laferriere 1979). Others state that only the lack of stress on the following vowel matters (Kahn 1976, see discussion in Riehl 2003). If flapping is more likely post-stress, we might expect greater phonetic consonant reduction in that environment as well. Articulatory work (Gick et al., 2006; Byrd et al., 2009; and references therein) suggests a strong influence of syllable structure on timing of consonantal gestures, so reduction might differ after stressed vs. unstressed syllables. However, the past studies do not extend readily to make specific predictions about post-stress vs. inter-unstressed flaps.

Results for the stress factor (for the two read speech styles, which have complete data) appear in Figure 3. (See Table 2 for additional measures.) To test the effect of stress, we used three-factor within-subjects ANOVAs, with the factors style (story, list), phoneme (the 6 phonemes), and stress (post-stress, inter-unstressed), for each dependent variable. Consonant duration had significant main effects of style ($F(1,12)=44.34$, $p<.001$ $\eta^2_p=.79$) and phoneme ($F(5,60)=586.71$, $p<.001$ $\eta^2_p=.98$), of their interaction ($F(5,60)=22.28$, $p<.001$ $\eta^2_p=.65$) and the phoneme by stress interaction ($F(5,60)=11.45$, $p<.001$ $\eta^2_p=.49$), and a significant three-way interaction ($F(5,60)=3.25$, $p<.02$ $\eta^2_p=.21$). The main effect of stress was not significant ($F(1,12)=2.89$, $p>.1$ $\eta^2_p=.19$). Because of the interaction, phoneme by stress was tested for each speech style, and each showed a two-way interaction (story: $F(5,60)=2.67$, $p<.03$, list: $F(5,60)=22.32$, $p<.001$). Thus, the simple effects of stress were tested. /b, d, t/ were significantly longer between unstressed syllables than after stress in both styles (story: /b/ $F(1,12)=9.29$, $p<.02$, /d/ $F(1,12)=5.17$, $p<.05$, /t/ $F(1,12)=16.05$, $p<.005$; list: /b/ $F(1,12)=24.35$, $p<.001$; /d/ $F(1,12)=18.94$, $p<.005$, /t/ $F(1,12)=21.58$, $p<.005$). However, the effect was significant in the opposite direction for list-reading /g, k/ (/g/: $F(1,12)=6.61$, $p<.03$; /k/: $F(1,12)=13.70$, $p<.005$). The remaining conditions had no significant effect of stress on consonant duration.

INSERT FIGURE 3 AND TABLE 2 APPROXIMATELY HERE

Longer consonant durations generally indicate clearer, less reduced consonants, and there is a strong positive correlation (across all data $r=.83$, $p<.001$) between consonant duration and intensity dip. Thus, this result seems to indicate that there is usually greater reduction post-stress, as one might extrapolate from earlier literature on flapping. The inconsistency of /g, k/ could have to do with velar articulation, or could be chance.

The intensity difference data (Figure 3B), however, show a different pattern. Again, all factors participated in at least one significant interaction in the overall ANOVA (main effects of style: $F(1,12)=71.80$, $p<.001$ $\eta^2_p=.86$, phoneme: $F(5,60)=199.76$, $p<.001$ $\eta^2_p=.94$, stress: $F(1,12)=9.60$, $p<.01$ $\eta^2_p=.44$; style by phoneme: $F(5,60)=3.16$, $p<.02$ $\eta^2_p=.21$, phoneme by stress: $F(5,60)=7.91$, $p<.001$ $\eta^2_p=.40$, other interactions not significant). Again, there were significant two-way interactions for each style (story: $F(5,60)=3.94$, $p<.005$; list: $F(5,60)=10.04$, $p<.001$). However, the intensity difference was smaller (more approximant-like) in inter-unstressed environment than in post-stress for all five significant simple effects (story /d/ $F(1,12)=7.10$, $p<.03$, /p/ $F(1,12)=12.24$, $p<.005$, /k/ $F(1,12)=6.01$, $p<.05$; list /p/ $F(1,12)=59.69$, $p<.001$, /k/ $F(1,12)=48.74$, $p<.001$). Thus, the consonants with significant effects of stress showed greater reduction in inter-unstressed position as measured by intensity but the opposite as measured by consonant duration (for six of the eight significant conditions).

As continuous measures, consonant duration and intensity difference provide the most sensitive data. Two of the additional measures, voicelessness and cessation of formants, had a significant three-way interaction in the overall ANOVA (style by phoneme by stress) or had all factors participating in a significant two-way interaction. In follow-up tests for each speech style separately, voicelessness had significant two-way interactions of phoneme by stress for each of the read speech styles, and the formant measure had this interaction only for list-reading. In each of those cases, the simple effect of stress was tested for each phoneme, as above. There was greater reduction for inter-unstressed tokens in all significant cases (voicelessness: story /b/ $F(1,12)=6.64$, $p<.03$ and /p/ $F(1,12)=27.01$, $p<.001$, list /p/ $F(1,12)=12.71$, $p<.005$ and /k/ ($F(1,12)=33.51$, $p<.001$; lack of formants: list /p/ $F(1,12)=8.79$, $p<.02$). When interactions did not motivate tests of simple effects, there also was no significant main effect of stress. All significant effects on the categorical measures confirmed what the intensity measure showed: greater reduction between unstressed syllables, counter to what consonant duration seems to indicate. Thus with regard to stress, rather than all five measures reflecting various aspects of a single continuum of reduction, consonant duration reflects something different. We return to this below.

C. Speech style

To analyze the speech style factor, the data were collapsed over the segmental/word-boundary factor as above, and all post-stress conditions other than the phoneme /g/ were included. Five speakers used no words containing /g/ in appropriate environments in their conversation, and one speaker used no /p/ tokens, so analyzing speech style requires either excluding six speakers or excluding the phoneme /g/ and one speaker, which excludes fewer data-points. The analyses below use 12 speakers for ANOVAs that include /p/, and use all 13 speakers for simple effects tests for the other phonemes.

Results for two dependent variables appear in Figure 4, and all measures appear in Table 2 (including all 13 speakers' data). The data were analyzed using two-factor within-subjects designs of speech style (conversation, story-reading, list-reading) by phoneme (/b, d, p, t, k/). Four dependent variables showed a significant interaction between style and phoneme (consonant duration: $F(8, 88)=8.11$, $p<.001$ $\eta^2_p=.42$; intensity difference $F(8,88)=5.66$, $p<.001$ $\eta^2_p=.34$; voicelessness: $F(8,88)=2.92$, $p<.01$ $\eta^2_p=.21$; cessation of formants: $F(8,88)=5.92$, $p<.001$ $\eta^2_p=.35$). All measures showed significant main effects of both factors, with greater reduction for the more casual styles (consonant duration style: $F(2,22)=42.16$, $p<.001$ $\eta^2_p=.79$, phoneme: $F(4,44)=646.47$, $p<.001$ $\eta^2_p=.98$; intensity style: $F(2,22)=18.35$, $p<.001$ $\eta^2_p=.63$, phoneme: $F(4,44)=345.07$ $\eta^2_p=.97$; voicelessness style: $F(2,22)=4.39$, $p<.03$ $\eta^2_p=.29$, phoneme: $F(4,44)=290.37$ $\eta^2_p=.96$; burst style: $F(2,22)=19.14$, $p<.001$ $\eta^2_p=.64$, phoneme: $F(4,44)=73.30$ $\eta^2_p=.87$; cessation of formants style: $F(2,22)=25.37$, $p<.001$ $\eta^2_p=.70$, phoneme: $F(4,44)=250.91$, $p<.001$ $\eta^2_p=.96$). For the measures with significant interactions, the simple effect of style was tested for each phoneme separately, and demonstrated significant effects on consonant duration and intensity dip for almost all phonemes (duration: /b/ $F(2,24)=14.63$, $p<.001$, /d/ n.s., p $F(2,22)=11.18$, $p<.001$, /t/ $F(2,24)=7.45$, $p<.005$, /k/ $F(2,24)=34.82$, $p<.001$; intensity: /b/ $F(2,24)=13.68$, $p<.001$, /d/ $F(2,24)=7.30$, $p<.005$, /p/ $F(2,22)=8.896$, $p<.005$, /t/ $F(2,24)=15.75$, $p<.001$, /k/ $F(2,24)=13.96$, $p<.001$), as well as on voicelessness and cessation of formants for some phonemes (voicelessness: /b/ $F(2.24)=6.50$, $p<.01$; cessation of formants: /b/ $F(2,24)=12.43$, $p<.001$, /k/ $F(2,24)=9.52$, $p<.005$). Thus, all dependent variables show greater reduction in more casual speech.

INSERT FIGURE 4 APPROXIMATELY HERE

The pattern among the three styles is often a stair-step pattern, with story-reading intermediate to the other styles (Figure 4). However, in some conditions the story-reading groups with the conversational speech (e.g. for /t/ in both graphs). However, story-reading sometimes groups with list-reading instead: for two phonemes the intensity measure showed a difference between the conversation and story-reading but not between the two read styles (pairwise comparisons, /b/: conversation vs. story $F(1,12)=11.56$, $p<.01$, story vs. list $F(1,12)=4.73$, $p=.050$; /d/: conversation vs. story $F(1,12)=9.63$, $p<.01$, story vs. list $F(1,12)=1.39$, $p>.25$). Furthermore, in two-factor interaction comparisons on the lack of formants measure (Figure 4B), testing phoneme by style (the two connected styles in one analysis and the two read styles in a separate analysis), the main effect of style was much larger when only conversation and story-reading were included ($F(1,11)=17.05$, $p<.005$, $\eta^2_p=.61$) than it was when only the two read-speech styles were included ($F(1,12)=8.23$, $p<.02$, $\eta^2_p=.41$). This shows that story-reading differs more from conversation than from list-reading when averaged across phonemes. Thus, while it initially appears that the story-reading is most similar to the conversational speech, in some cases story-reading is more similar to list-reading, so it is more accurate to view story-reading as simply showing an intermediate level or frequency of reduction.

Another way to examine the effect of style is through how many consonants were so reduced that we were unable to locate them to measure them (e.g. Figure 1A). Out of more than 9,200 tokens, there were only 147 such deletions. This is not because reduction to near-deletion is rare, but because we measured even extremely reduced consonants, as long as a dip in the intensity curve could be located. In conversation, more than 10% of all expected /t/ and /b/ phonemes in target environments were deleted, and nearly 10% were for /g/. Even in word-list reading, approximately 1% of all /t/ tokens were deleted. Story-reading showed an intermediate frequency of deletion. Far more tokens were measurable but very nearly deleted (e.g. Figure 1B). We do not perform statistical analysis of these data because there were no deletions in many conditions.

D. Word frequency

Several past researchers have found that higher frequency words display more reduction (Pluymaekers et al., 2005a; Gahl, 2008), although this finding is not uniform (Greenberg, 1999; Kuperman et al., 2007; and cf. discussion in Gahl, 2008). We test whether word frequency correlates with reduction at the sub-segmental level. We test not just duration, which could reflect overall faster speech, but also how consonantal the stop or flap consonant is. Furthermore, we test whether the effect of word frequency is the same in spontaneous conversation vs. in word-list reading, and whether word frequency affects reduction of consonants within a word (e.g. high-frequency 'par**t**y' vs. low-frequency 'bar**t**er') vs. those spanning a word boundary (e.g. 'wai**t** a' vs. 'orbi**t** a').

For this analysis, word frequencies were obtained from the Corpus of Contemporary American English (COCA: Davies, 2008), using the total frequency in the corpus. (Frequencies for the spoken (newscasting) portion were also analyzed, but these correlate strongly with total frequency.) With COCA, one can obtain frequencies from the same corpus for single words and the word-boundary two-word phrases (e.g. 'wai**t** a'). Since these frequencies are not comparable, analyses on word frequency are performed for all within-word segmental environments (pre-schwa, post-/ɹ/, etc.) together, and separately for the word-boundary condition. To avoid the problem of high frequency outliers, we took the logarithm after adding 1 to each frequency (removing 0 values).

The effect of word frequency was analyzed through linear multiple regression analyses, using intensity difference and the logarithm of consonant duration (the two continuous measures) as dependent variables in separate analyses. (Converting consonant duration to its logarithm avoids excessive influence of outlier very long consonants, but results are similar when the analyses are performed on raw consonant duration.) Data were first averaged over speakers, in order to avoid clusters of dependent data within the regression and artificially inflated power. Although the log of word frequency is the primary factor of interest, we also included stress (post-stress, inter-unstressed) in the analysis and included recoded factors representing the speech style and the properties of groups of phonemes, since both are categorical factors with more than two levels. Speech style was recoded into a variable connectedness, which separates the list-reading (coefficient of 2) from the two connected speech styles (coefficients -1), and into an additional variable which separates the conversation from the story-reading (with list-reading coded as 0, and conversation and story-reading codes of 1 and -1 respectively divided by the total number of data points in the category so that codes sum to 0 (Stockburger, 1996)). The phonemes were recoded as a variable that distinguishes /p, k/ (the two phonemes that are usually phonetically voiceless, coded as 2) from the other phonemes (each coded as -1), and a second variable that distinguishes flapping phonemes (/t, d/, codes of 1 divided by the number of data points in that category) from /b, g/ (voiced stops, coded as -1 divided by the number of data points, with /p, k/ coded as 0).

The regression analysis of consonants at the word boundary (two-word phrase frequencies) showed a significant effect of phrase frequency on consonant duration ($t=-4.43$, $p<.001$), as well as of both phoneme factors (/p, k/ vs. rest: $t=32.02$, $p<.001$;

flap vs. /b, d/: t=-6.08, p<.001) (Figure 5A). For intensity difference, these same factors were significant (frequency: t=-3.13, p<.005, /p, k/ vs. rest: t=32.21, p<.001, flap vs. /b, g/: t=-4.07, p<.001), as were the effect of stress (t=2.73, p<.01) and the difference between list-reading vs. connected speech (t=2.81, p<.01). Thus, both measures show that consonants at a word boundary are more reduced in higher frequency two-word phrases. However, this effect is not consistent across all consonants (Figure 5A): the regression line for /g/ runs in the opposite direction. This anomalous result could stem from the relative paucity of /g/ tokens at word boundary.

INSERT FIGURE 5 APPROXIMATELY HERE

For consonants within a word (not at word boundary), however, the results were quite different (Figure 5B). All factors except word frequency and the difference between connected speech styles had significant effects on consonant duration (stress: t=-.2.50, p<.02, /p, k/ vs. rest: t=56.63, p<.001, flap vs. /b, g/: t=-18.99, p<.001, wordlist vs. connected: t=10.49, p<.001), but word frequency did not (t=-1.31, p>.15). For intensity difference, word frequency was the only factor lacking a significant effect (stress: t=2.98, p<.005, /p, k/ vs. rest: t=59.66, p<.001, flap vs. /b, g/: t=-13.54, p<.001, wordlist vs. connected: t=11.21, p<.001, story vs. conversation: t=-2.59, p<.02, frequency: t=0.64, p>.5). These results use the "enter" variable selection method of multiple regression (all variables entered into the model at once), but the stepwise method showed similar results. Thus, greater phrasal frequency correlated with greater reduction of consonants at the word boundary (e.g. 'wha**t** if'), but frequency had no significant effect within a word.

Frequency could affect reduction more, or differently, in conversation than in careful isolated speech. Therefore, we also performed multiple regressions for the conversation data and the word-list reading data separately (for the phrasal and the non-phrasal environments, as above). Stress, the two factors coding consonant type, and the frequency (as above) were included as factors, with intensity difference and log of consonant duration as dependent variables. The results were entirely consistent with those above using all speech styles: for each of conversation and list reading, frequency significantly affected reduction for consonants at the word boundary (log duration: conversation $t=-2.49$, $p<.02$, list $t=-3.28$, $p<.005$), but had no effect on word-internal consonants (log duration: conversation $t=.67$, list $t=.06$, both p's$>.10$). (Results for intensity difference are similar.) Thus, consonants at word boundaries are more reduced if the phrase is high frequency, regardless of whether the speech planning process involves managing a conversational interaction or retrieving syllables while reading.

IV. DISCUSSION

To summarize, the results show greater reduction in more casual speech, greater reduction at the word boundary in higher frequency phrases but not internal to higher frequency words, and greater reduction between two unstressed syllables than after a stressed syllable (e.g. 'limi**t**ed' vs. 'sta**t**us'). The effect of speech style is gradient, with conversational speech the most reduced, connected read speech the next most reduced, and words read in isolation least reduced.

There are some detailed patterns of reduction in particular segmental environments, such as speakers reducing more before schwa, e.g. 'status,' than in most other environments. Overall in the current results, segmental environments do not demonstrate many clear or consistent effects on reduction. That is, consonants reduce to approximately the same extent in approximately the same way before a syllabic /l/ as before [ɚ] or after /ɹ/. There may be further systematic patterns within the segmental environments, but it is at least clear that they are not large effects, and the η^2_p confirms that segmental environment has the smallest effect of any factor in the statistical analysis. The effects of stress, word frequency, and speech style are larger and more consistent.

The effect of stress (post-stress vs. inter-unstressed) at first seems mysterious: post-stress environment has more reduction as measured by duration, but less as measured by all other significant measures (intensity dip, cessation of formants, voicing). This is the only result in this work with consistent opposite directions for the various measures. We propose that this is because factors other than reduction affect duration: the post-stress consonants are shortened as part of the adjustments to timing for the stressed preceding syllable, presumably because the alternating stress pattern of English leads to lengthening of stressed and shortening of unstressed syllables. This does not necessarily mean the post-stress consonants are reduced. The non-duration measures thus show the reduction of consonants, with greater reduction in the middle of a string of unstressed syllables, while the duration measure in this case reflects stress effects instead. This indicates that duration may not be adequate as a single measure of reduction (Tucker, in press).

Although speech style has a clear effect, it might seem surprising that its effect is not larger. In all ANOVAs, style has a larger η^2_p than segmental environment or stress, but a consistently smaller one than the factor phoneme. Reduction could be something that happens mostly when one is talking very casually, something that does not happen in normal or careful speech, in which case the style effect would be larger. Crystal & House (1988a) show that the difference between slow- and fast-talking speakers can be as large as the difference between speech styles. In the current results, speech style certainly does not cause a categorical shift in reduction, from 0% to 100% on any measure. The phoneme /b/, which has the largest effect for cessation of formants (Figure 4B), has at least some F2 or F3 visible throughout the consonant in over 50% of isolated read tokens and over 80% of conversation tokens. The fact that the story-reading style shows intermediate reduction suggests that reduction is influenced both by whether words are read vs. chosen while speaking, and by planning at the phrasal level (vs. isolation). There is greater reduction in casual speech, but reduction occurs in all styles.

We believe the explanation lies in two important facts about speech: first, phonological distinctions usually cause larger effects than phonetic variability. Second, acoustic reduction is not a "sloppy" exception characteristic of casual conversations, instead it is the normal way to communicate. We will address these points in turn.

In Figures 2-5, one can see that the effect of the identity of the phoneme is far larger in terms of simple differences in means than any other factor tested (stress, style, etc.), and η^2_p confirms this. The large effect of phoneme is most obvious in comparing /p, k/ to the other phonemes, but /b, g/ are also consistently more consonantal (less approximant-like) than /t, d/. That is, the phonologically voiceless stops (/p, k/) are the most consonantal across all measures, the expected voiced stops (/b, g/) are much more approximant-like (have smaller intensity dips, are more likely to have formants and to maintain voicing, less likely to have bursts), and the expected flaps (/t, d/) are the most approximant-like. This suggests that phonological distinctions and allophonic flapping determine the approximate range in which the consonant is likely to be realized, and phonetic variability such as reduction leads to a value usually within this range.

This is similar to Keating's window model of articulations (1990), where tokens of each allophone fall within a window (which could consist of a mean and standard deviation instead). Keating's model was developed for coarticulation, but it could apply to other within-category variability, such as reduction. The range of possible realizations of a segment (e.g. Figure 1) is so broad that such windows might seem too large to be meaningful. However, if windows contain distribution information (e.g. mean and standard deviation), one could model the fact that a few tokens of /k/ have strong formants while not making the range of /k/ mostly overlap with the range of [ɾ]. Such a window/distribution model might allow the production system to generate a full range of reduced forms without storing multiple variants in the lexicon.

The current data show that the phonology of the language (both distinctions and allophonic alternations) creates larger differences in consonants than phonetic variability within a category does. This is true even for a wide range of realizations, including high-frequency function words spoken in a casual conversation with a close friend (e.g. "i**t** is," frequency greater than 300,000 in COCA) and low-frequency content words (e.g. "matriar**ch**y," frequency 57) read in isolation. Thus, the difference among phonemes (i.e. /p, k/ vs. /b, g/ vs. flapped /t, d/) creates far larger acoustic differences than the variability of reduction does: this is in the nature of making linguistic distinctions.

This leads to the second general feature of speech: reduction is a normal part of all speech, even careful list reading. This is clearest in the categorical measures. According to the definition of stops, none of /p, k, b, g/ should have the second or third formants continuing throughout the consonant; all should have a gap in the spectral structure during the closure. However, Table 2 shows that 49-60% of /b, g/ tokens in list-reading had at least weak F2 or F3 throughout the consonant, as did 65-77% of story-reading /b, g/ tokens. Many had clear, dark formants. Thus, many intervocalic /b, g/ consonants, even in isolated read speech, are realized more like an approximant than a stop. These results support findings that intervocalic /d/ and flaps can be realized as approximants (Zue & Laferreire 1979, Barry & Andreeva 2001, Horna 2005), and provide quantitative detail. Even flapped /t, d/ are not expected to have F2 and F3 throughout the flap (e.g. Ladefoged & Johnson, 2010) but Table 2 shows that the proportion that do is quite high. 80-89% of flap tokens in the two reading styles had at least some F2 or F3 throughout the consonant, indicating that many "flaps" were approximant-like.

Results for the proportion of consonants lacking a burst (Table 2) confirm this finding: while /p, k/ and /b, g/ would be expected to have bursts according to their careful speech definition, a large proportion do not (even though we included very weak bursts). The results for the proportion of consonants that maintain voicing support this conclusion as well: /p, k/ would be expected to have voicing cease, but even for these most stop-like of the six consonants, it continued in 6-15% of list-reading tokens and 10-33% of connected speech tokens. (The fact that /b, g/ sometimes lacked voicing probably reflects the aerodynamic problem of maintaining voicing during closure (Ohala 1983). The fact that /g/ devoiced more often than /b/ supports this conclusion (Ohala 1983).)

The categorical measures show voicing, lack of a burst, and continuing formants in a large proportion of tokens, even in careful speech. Greenberg (1999) and Johnson (2004) point out how common reduction is based on the percentage of expected segments that are altered to a different transcription or deleted in conversational speech. The current work, instead, shows how these reduced segments are produced, even if the differences in them cannot be transcribed readily in IPA. Several of the cases in Figure 1, for example, are quite difficult to transcribe, but there is substantial reduction, although some trace of the consonant is audible. The current results also augment Greenberg's and Johnson's findings by demonstrating how common acoustic reduction is even in careful read speech, where complete deletions are relatively uncommon and the transcription-based method might fail to detect much reduction. Because we are so familiar with spectrograms of very carefully articulated speech, where consonants largely match the textbook expectations, it is easy to think that a highly reduced token of a consonant (Figure 1A-C) is an exception or even a speech error. One might think that reduction in relatively formal speech is only a feature of young speakers, or perhaps a dialectal feature (e.g. "Valley Girl" speech). However, this seems unlikely. We have heard obvious reductions in the speech of a university president giving a formal lecture and in formal addresses by politicians, and the authors notice reductions in their own speech. Reduction affects a large proportion of the consonants speakers produce. This finding suggests that it might be helpful to include reduced speech in foreign language teaching, to train learners to recognize the range of frequently occurring variants of sounds, rather than training them entirely on learner-directed exceptionally careful speech.

V. CONCLUSIONS

This work documents pervasive speech reduction in both careful read speech and casual conversation, as it affects stops and flaps between vowels and/or sonorants. This reduction leads to expected voiceless stops (/p, k/) being produced sometimes with voicing, and occasionally as approximants. It leads to expected voiced stops (/b, g/), and even more so expected flaps (/t, d/), being realized as approximants or vowel-like sounds.

The current data also show that consonants are more reduced between two unstressed syllables than after a stressed syllable. This suggests that the long string of unstressed segments in sequential unstressed syllables leads speakers to reduce over a larger time span. Gestural planning at the syllabic or higher level (cf. Gick et al. 2006, Byrd et al. 2006) thus affects reduction. The current results furthermore demonstrate that word frequency affects segmental reduction, but only by affecting how speakers coordinate their gestures across word boundaries, with no effect of greater reduction for consonants within higher frequency words, although further research may show such effects within the word. Overall, this work contributes to our knowledge of the variability speakers produce, in natural daily life speech and in careful speech. In conclusion, listeners encounter reduction in the majority of the speech they hear.

ACKNOWLEDGMENTS

We wish to thank Mike Hammond, Holger Mitterer, Mirjam Ernestus, Miquel Simonet, and two anonymous reviewers for helpful feedback on this work. We are also very grateful to the many research assistants who measured this data. Any errors or misinterpretations, of course, are our own.

REFERENCES

Arai, T. (1999). “A case study of spontaneous speech in Japanese,” in Proceedings of the International Congress of Phonetic Sciences, San Francisco, 1, 615–618.

Aylett, M., and Turk, A. (2006). “Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei,” J. Acoust. Soc. Am. 119(5), 3048–3058.

Barry, W., and Andreeva, B. (2001). “Cross–Language Similarities and Differences in Spontaneous Speech Patterns,” J. Int. Phonetic Assoc. 31(1), 51–66.

Blumstein, S. E., and Stevens, K. N. (1979). “Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants,” J. Acoust. Soc. Am. 66(4), 1001–1017.

Boersma, P. and Weenink, D. (2010). “Praat: doing phonetics by computer [Computer program],” Version 5.1.43, retrieved 4 August 2010 from http://www.praat.org/

Byrd, D., Tobin, S., Bresch, E., Narayanan, S. (2009). “Timing effects of syllable structure and stress on nasals: a real-time MRI examination,” J. Phonetics 37, 97-110.

Byrd, D., Krivokapić, J., and Lee, S. (2006). “How far, how long: On the temporal scope of prosodic boundary effects,” J. Acoust. Soc. Am. 120(3), 1589-1599.

Crystal, T., and House, A. (1982). "Segmental durations in connected speech signals: Preliminary results." J. Acoust. Soc. Am. 72, 705-716.

Crystal, T., and House, A. (1988a). "Segmental durations in running speech: Current results." J. Acoust. Soc. Am. 83, 1553-1573.

Crystal, T., and House, A. (1988b). "The duration of American English stop consonants: An overview." J. Phonet. 16, 285-294.

Dalby, J. (1986). *Phonetic structure of fast speech in American English*. Ph.D. dissertation, Indiana University.

Davies, M. (2008). "The Corpus of Contemporary American English (COCA): 410+ million words, 1990–present," Available online at <http://www.americancorpus.org.>

de Jong, K. (1998). "Stress–related variation in the articulation of coda alveolar stops: flapping revisited," J. Phonetics 26(3), 283–310.

Eddington, D., and Elzinga, D. (2008). "The Phonetic Context of American English Flapping: Quantitative Evidence," Lang. Speech 51(3), 245 –266.

Ernestus, M., Baayen, R. H., and Schreuder, R. (2002). "The recognition of reduced word forms," Brain Lang. 81, 162–173.

Fukaya, T., and Byrd, D. (2005). "An Articulatory Examination of Word–Final Flapping at Phrase Edges and Interiors," J. Int. Phonetic Assoc. 35(1), 45–58.

Gahl, S. (2008). "*Time* and *Thyme* Are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech," Language 84(3), 474–496.

Gick, B., Campbell, F., Oh, S., and Tamburri–Watt, L. (2006). "Toward universals in the gestural organization of syllables: A cross–linguistic study of liquids," J. Phonetics 34(1), 49–72.

Greenberg, S. (1999). "Speaking in shorthand – A syllable–centric perspective for understanding pronunciation variation," Speech Commun. 29, 159–176.

Herd, W., Jongman, A., and Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. J. Phonetics, doi:10.1016/j.wocn.2010.06.003

Horna, J. (1998). "An Investigation into the Acoustics of American English Flaps, with a Secondary Emphasis on Spanish Flaps, in Fluent Speech," Ph.D. thesis, New York

University.

Hualde, J. I. (2005). *The sounds of Spanish*. (Cambridge University Press, Cambridge).

Hualde, J. I., Nadeu, M., and Simonet, M. (2010). “Lenition and phonemic contrast in Majorcan Catalan. Romance Linguistics 2009,” in *Selected Proceedings of the Linguistic Symposium on Romance Languages 39*, edited by S. Colina, A. Olarrea and A.M. Carvalho. (John Benjamins, Amsterdam), pp. 63-79.

Johnson, K. (2004). “Massive reduction in conversational American English,” In *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, edited by K. Yoneyama and K. Maekawa. (The National International Institute for Japanese Language, Tokyo, Japan), pp. 29–54.

Jurafsky, D., Bell, A., and Gyrand, C. (2002). “The Role of the Lemma in Form Variation,” in *Papers in Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner. (Mouton de Gruyter, Berlin/New York), pp. 1–34

Kahn, D. (1976). “Syllable–based generalizations in English phonology,” Ph.D. thesis Massachusetts Institute of Technology.

Keating, P. A. (1990). “The window model of coarticulation: articulatory evidence,” in *Papers in Laboratory Phonology I*, edited by J. Kingston and M. Beckman. (Cambridge University Press, Cambridge), pp. 451–470.

Keune, K., Ernestus, M., Van Hout, R., and Baayen, R. H. (2005). “Social, geographical, and register variation in Dutch: From written ‘mogelijk’ to spoken ‘mok’,” Corpus Ling. and Ling. Theory 1, 183–223.

Kohler, K. J. (2001). "Articulatory dynamics of vowels and consonants in speech communication," J. Int'l. Phonet. Assoc. 31, 1-16.

Koopmans-van Beinum, F. J. (1980). *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions*. (Academische Pers B.V., Amsterdam).

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, H. (2007). "Morphological predictability and acoustic duration of interfixes in Dutch compounds," J. Acoust. Soc. Am. 121(4), 2261–2271.

Ladefoged, P., and Johnson, K. (2010). *A Course in Phonetics*. (Cengage Learning).

Lisker, L. (1985). "The pursuit of invariance in speech signals," J. Acoust. Soc. Am. 77(3), 1199–1202.

Mitterer, H., and Ernestus, M. (2006). "Listeners recover /t/s that speakers reduce: Evidence from /t/–lenition in Dutch," J. Phonetics 34(1), 73–103.

Ohala, J. J. (1997). "Aerodynamics of phonology," in Proceedings of the Seoul International Conference on Linguistics, pp. 92–97.

Patterson, D., and Connine, C. M. (2001). "Variant Frequency in Flap Production A Corpus Analysis of Variant Frequency in American English Flap Production," Phonetica *58*, 254–275.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005a). "Frequency and acoustic length: the case of derivational affixes in Dutch," J. Acoust. Soc. Am. 118, 2561–2569.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005b). "Articulatory planning is continuous and sensitive to informational redundancy," Phonetica 62, 146–159.

Port, R. F. (1977). “The influence of tempo on stop closure duration as a cue for voicing and place,” in Haskins Laboratories Status Report on Speech Research, SR–39/40, 81, 51/52, 59–73.

Richter, E. (1930). "Beobachtungen über Anglitt und Abglitt an Sprachkurven und umgekehrt laufenden Phonogrammplatten," in *Berichte über die I. Tagung der Internationalen Gesellschaft für experimentelle Phonetik*, edited by P. Menzerath, (Univ.-Buchdr. Scheur.: Bonn), pp. 87-90.

Riehl, A. K. (2003). “American English Flapping: Perceptual and Acoustic Evidence Against Paradigm Uniformity with Phonetic Features,” Working Papers of the Cornell Phonetics Laboratory 15, 271–337.

Shattuck–Hufnagel, S., and Veilleux, N. M. (2007). “Robustness of Acoustic Landmarks in Spontaneously–Spoken American English,” in Proceedings of the 16th International Congress of Phonetic Sciences, pp. 925–928.

Stockburger, D. W. (1998). *Multivariate statistics: concepts, models and applications Web version 1.0.* <http://www.psychstat.missouristate.edu/multibook/mlt00.htm>

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (1991). “An investigation of locus equations as a source of relational invariance for stop place categorization,” J. Acoust. Soc. Am., 90(3), 1309–1325.

Tucker, B. V. (2007). “Spoken word recognition of the reduced American English Flap,” Ph.D. thesis University of Arizona.

Tucker, B. V. (in press). “The effect of reduction on the processing of flaps and /g/ in isolated words,” J. Phonet.

Warner, N. (2011). “Reduction,” in *The Blackwell Companion to Phonology,* edited by

M. van Oostendorp, C. Ewen, E. Hume, and K. Rice (Wiley, Oxford).

Warner, N. (in press). “Methods for studying spontaneous speech,” in *Handbook of Laboratory Phonology*, edited by A. Cohn, C. Fougeron, and M. Huffman (Oxford University Press, Oxford)

Warner, N., Fountain, A., and Tucker, B. V. (2009). “Cues to perception of reduced flaps,” J. Acoust. Soc. Am. 125(5), 3317–3327.

Zue, V. W., and Laferriere, M. (1979). “Acoustic study of medial /t,d/ in American English,” J. Acoust. Soc. Am. 66(4), 1039–1050.

TABLE I. Sample item for each condition used, and example from the story-reading task. Target consonants in bold.

Post-str.	b	d	g	p	t	k
pre full-V	a**bb**ey	alrea**d**y	ba**gg**y	cho**pp**y	beau**t**y	chee**k**y
pre-l	a**b**le	cra**d**le	ea**g**le	a**pp**le	ba**tt**le	buc**k**le
pre-ə	a**b**acus	aca**d**emy	do**gg**ed	a**pp**etite	au**t**omatic	bazoo**k**a
pre-r	cup**b**oard	ce**d**ar	be**gg**ar	Coo**p**er	be**tt**er	a**c**re
post-r	ar**b**or	bor**d**er	bur**g**undy	car**p**enter	bar**t**er	tur**k**ey
phrasal	ca**b** again	woul**d** a	ba**g** of	ma**p** a	wha**t** a	ta**k**e a
Inter-unstressed						
pre full-V		anybo**d**y		cano**p**y	abili**t**y	fini**ck**y
pre-l	availa**b**le			multi**p**le	capi**t**ol	arti**c**le
pre-ə	hali**b**ut	acci**d**ent	esopha**g**us		capi**t**alist	Veroni**c**a
pre-r		corri**d**or	vine**g**ar	juni**p**er	administra**t**or	
post-r			hambur**g**er		comfor**t**er	anar**ch**y
phrasal		lande**d** a		develo**p** a	visi**t** a	atti**c** above

Excerpt from story-reading, from a fictitious piece on menu planning:
"For the more adult palate, tur**k**ey with papri**k**a, ha**dd**ock, or grilled hali**b**ut is pro**b**ably preferable. (Fini**ck**y children, of course, do not consi**d**er such fare e**d**ible.) Some of these recipes are not ela**b**orate, but can be made readily out of things in most peo**p**le's cup**b**oards, such as flour, hambur**g**er, basic spices, vine**g**ar, milk, etc."

TABLE II. Average (left) and standard deviation (right) for each dependent measure (consonant duration (sec.), intensity drop, proportion with a voiceless portion during the consonant, proportion with a burst, proportion with cessation in F2 and F3), and number of tokens. S: post-stress, U: inter-unstressed.

Conver-sation	CDur (s)		Int. Diff.		Prop. Vless.		Prop. w/Burst		Prop. No Formant		n
b-S	.045	.02	11.73	4.3	.02	.09	.21	.35	.21	.37	79
d-S	.031	.01	9.91	4.1	.03	.15	.25	.33	.12	.28	163
g-S	.049	.02	15.33	6.8	.04	.14	.25	.40	.38	.46	19
p-S	.087	.02	30.16	6.2	.90	.26	.76	.34	.95	.16	78
t-S	.028	.01	9.07	4.0	.05	.16	.20	.29	.12	.24	360
k-S	.088	.01	26.54	5.1	.81	.31	.78	.28	.82	.25	200
b-U	.066	.03	16.68	5.7	.07	.19	.29	.49	.50	.50	8
d-U	.032	.01	9.90	4.4	.02	.11	.29	.42	.06	.22	45
g-U	.075	.03	17.78	6.6	1.00	.0	.50	.71	1.00	.0	2
p-U	.096	.01	33.17	17.6	.75	.50	.38	.48	1.00	.0	5
t-U	.031	.01	7.97	3.3	.00	.02	.13	.30	.03	.10	124
k-U	.081	.02	24.45	8.4	.80	.40	.78	.43	.71	.47	25
Story reading	CDur (s)		Int. Diff.		Vless.		Burst		No Form.		n
b-S	.051	.01	15.50	6.2	.15	.25	.17	.28	.35	.36	190
d-S	.032	.01	11.22	3.1	.04	.13	.28	.28	.15	.24	208
g-S	.053	.02	16.03	6.4	.21	.36	.44	.44	.32	.42	136
p-S	.086	.01	31.01	6.1	.88	.23	.63	.37	.95	.19	215
t-S	.029	.01	8.80	4.0	.02	.09	.19	.29	.11	.25	197
k-S	.092	.01	29.85	5.5	.79	.32	.83	.30	.92	.23	167
b-U	.061	.01	14.93	4.4	.08	.23	.23	.38	.32	.43	39
d-U	.035	.01	10.28	4.5	.04	.17	.24	.36	.20	.34	102
g-U	.051	.02	17.41	7.9	.26	.45	.37	.49	.23	.43	38
p-U	.088	.02	28.30	7.2	.67	.32	.58	.35	.93	.20	137
t-U	.035	.01	9.21	4.5	.02	.13	.20	.33	.12	.26	171
k-U	.091	.02	27.58	6.6	.78	.33	.84	.25	.93	.19	152
List reading	CDur (s)		Int. Diff.		Vless.		Burst		No Form.		n
b-S	.062	.01	17.20	5.1	.15	.23	.38	.26	.45	.27	779
d-S	.034	.01	11.76	3.4	.04	.11	.37	.22	.17	.19	780
g-S	.064	.01	18.43	5.2	.24	.28	.61	.26	.51	.26	664
p-S	.103	.01	35.38	3.9	.94	.10	.77	.23	.99	.03	779
t-S	.033	.01	11.21	3.2	.03	.06	.30	.22	.16	.17	786
k-S	.110	.02	33.61	4.5	.91	.14	.94	.10	.98	.05	741
b-U	.072	.02	16.91	6.0	.11	.27	.36	.36	.42	.35	163
d-U	.038	.01	11.73	2.9	.04	.10	.37	.29	.20	.22	404
g-U	.059	.02	17.54	6.5	.28	.44	.58	.45	.40	.48	52
p-U	.104	.02	31.33	5.2	.84	.24	.78	.28	.96	.11	221
t-U	.036	.01	11.16	2.9	.03	.08	.31	.25	.17	.18	613
k-U	.111	.02	30.63	4.9	.84	.22	.90	.15	.97	.09	397

Figure captions

Figure 1. Waveforms, spectrograms, and intensity curves (overlaid on spectrogram) demonstrating a range of reduction in flaps from conversation. Time marked at thirds of the figure. Except for A, the portion between vertical lines marks the consonant duration. A: "digi**t**al," /t/ (auditorily the marked portion) unmeasurable because there is no intensity dip or formant weakening. B: "tha**t** is," /t/ nearly deleted, strong formants throughout, identified based on weakening of F2 and F3. C: "fraterni**t**y," approximant realization, strong formants but clear dip in intensity. D: "wha**t** are you," slightly reduced flap with weak formants and voicing throughout, but with a burst. E: "(ex)ci**t**ed," clear flap with voicing but no formants, with a burst and a large drop in intensity.

Figure 2. Intensity difference for each segmental/boundary environment in story-reading (A) and list-reading (B), post-stressed consonants only. Averages include only the seven speakers with data in all conditions for story-reading, but all thirteen speakers for list-reading. Error bars represent 95% confidence intervals.

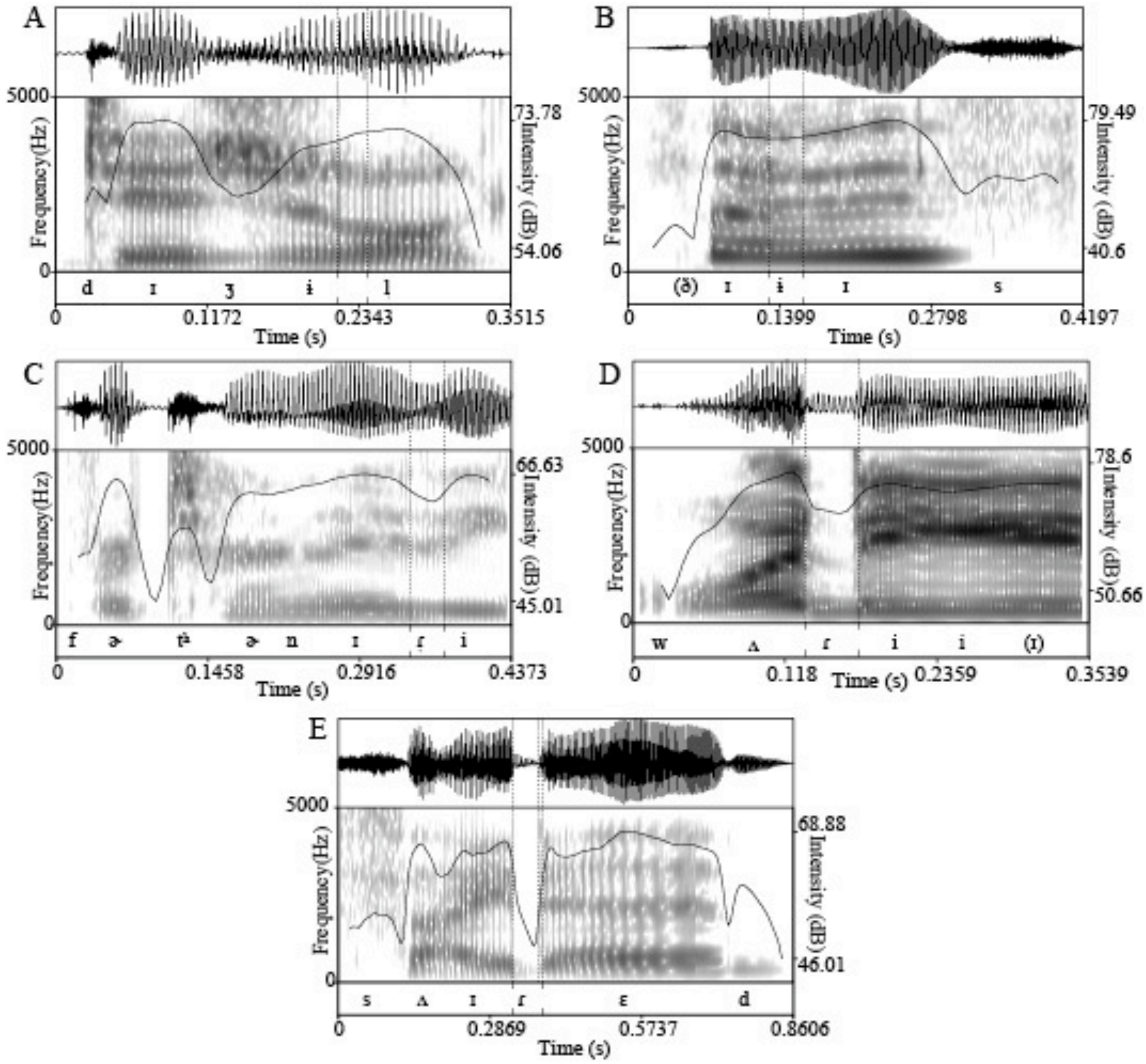
Figure 3. Consonant duration (A) and intensity difference (B) by stress, read speech only, averaged over segmental environments. Error bars show 95% confidence intervals.

Figure 4. Intensity difference (A) and proportion with cessation of formants (B), for the three speech styles. Data are averaged over segmental environments, but are post-stress only. Error bars show 95% confidence intervals. * indicates a significant difference

(p<.05) between neighboring bars for the pairwise comparisons, or significance of the main effect of a factor in the interaction comparisons when marked in the key.

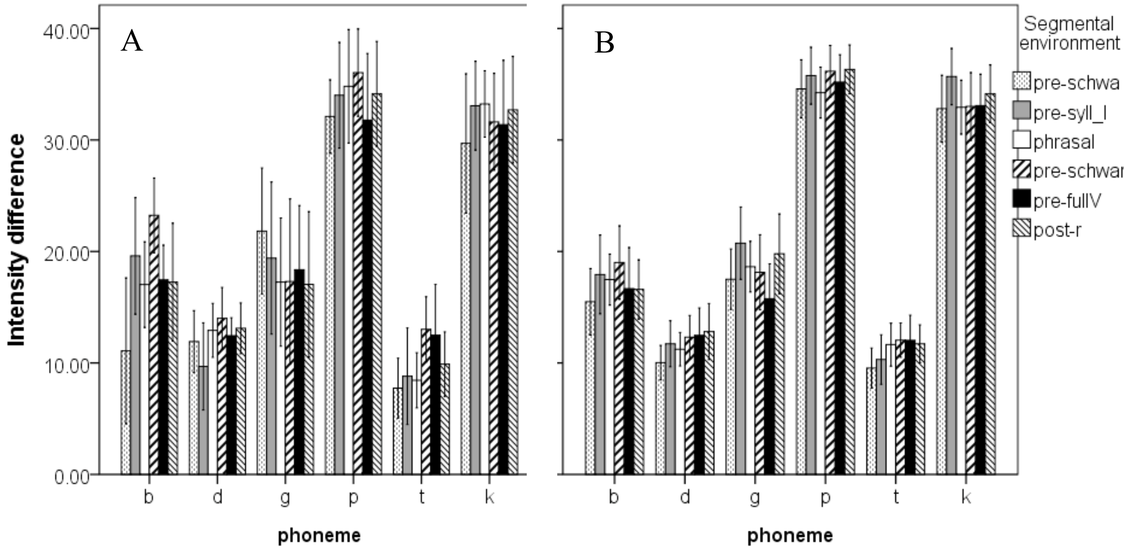
Figure 5. Scatter plots of intensity difference and log of frequency plus 1, at a word boundary (A) and within the word (all other segmental environments, B).

Figure 1



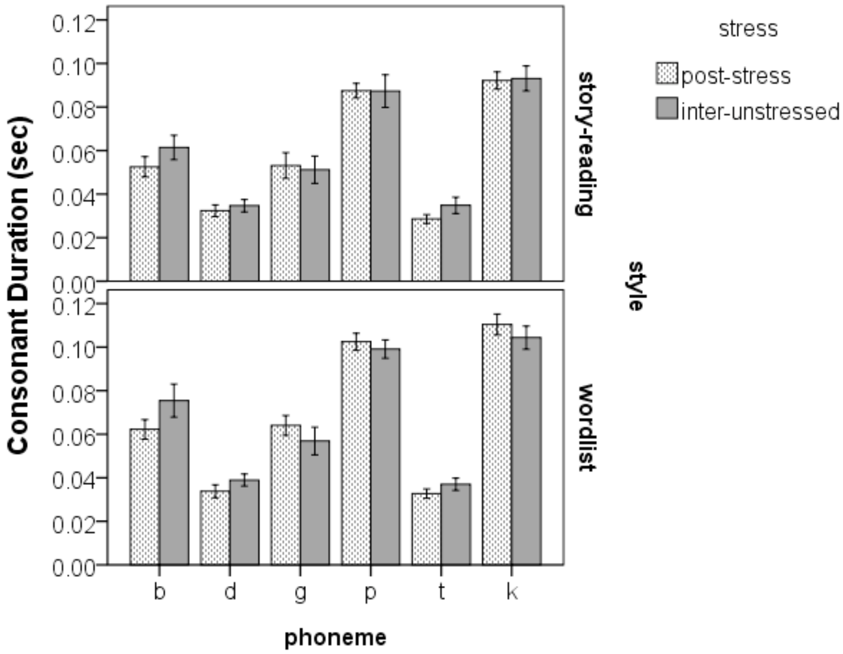
A
5000
0
Frequency(Hz)
73.78
54.06
Intensity (dB)
0
0.1172
0.2343
0.3515
Time (s)
B
5000
0
Frequency(Hz)
79.49
40.6
Intensity (dB)
0
0.1399
0.2798
0.4197
Time (s)
C
5000
0
Frequency(Hz)
66.63
45.01
Intensity (dB)
0
0.1458
0.2916
0.4373
Time (s)
D
5000
0
Frequency(Hz)
78.6
50.66
Intensity (dB)
0
0.118
0.2359
0.3539
Time (s)
E
5000
0
Frequency(Hz)
68.88
46.01
Intensity (dB)
0
0.2869
0.5737
0.8606
Time (s)

Figure 2

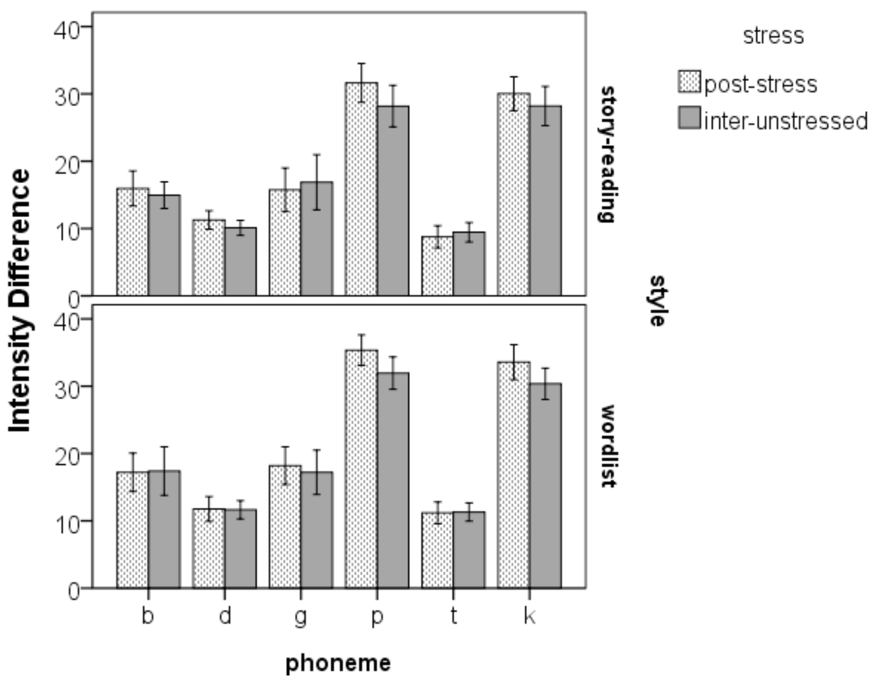


A
B
Intensity difference
40.00
30.00
20.00
10.00
0.00
b
d
g
p
t
k
phoneme
phoneme
Segmental environment
pre-schwa
pre-syll_l
phrasal
pre-schwar
pre-fullV
post-r

Figure 3


Consonant Duration (sec)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
story-reading
wordlist
style
stress
post-stress
inter-unstressed
b
d
g
p
t
k
phoneme

A


Intensity Difference
40
30
20
10
0
story-reading
wordlist
style
stress
post-stress
inter-unstressed
b
d
g
p
t
k
phoneme

B

Figure 4

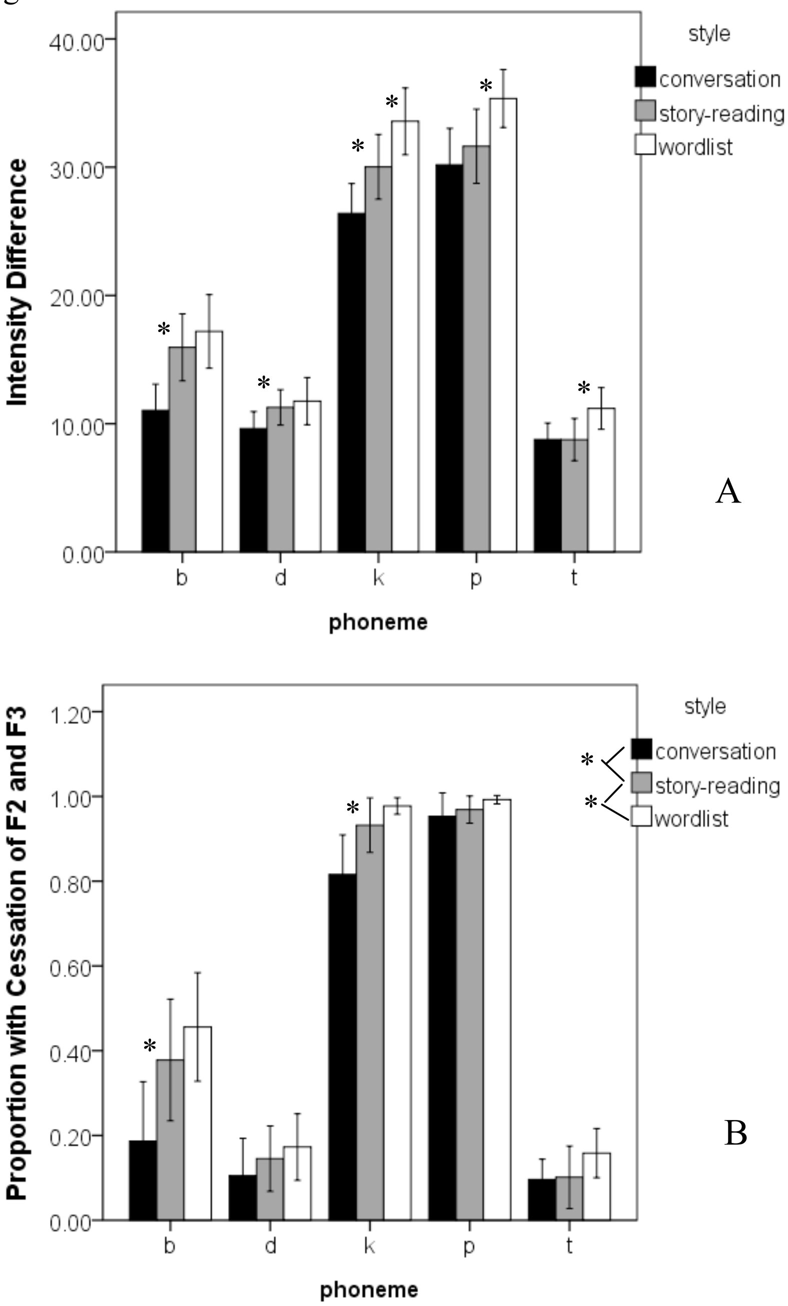
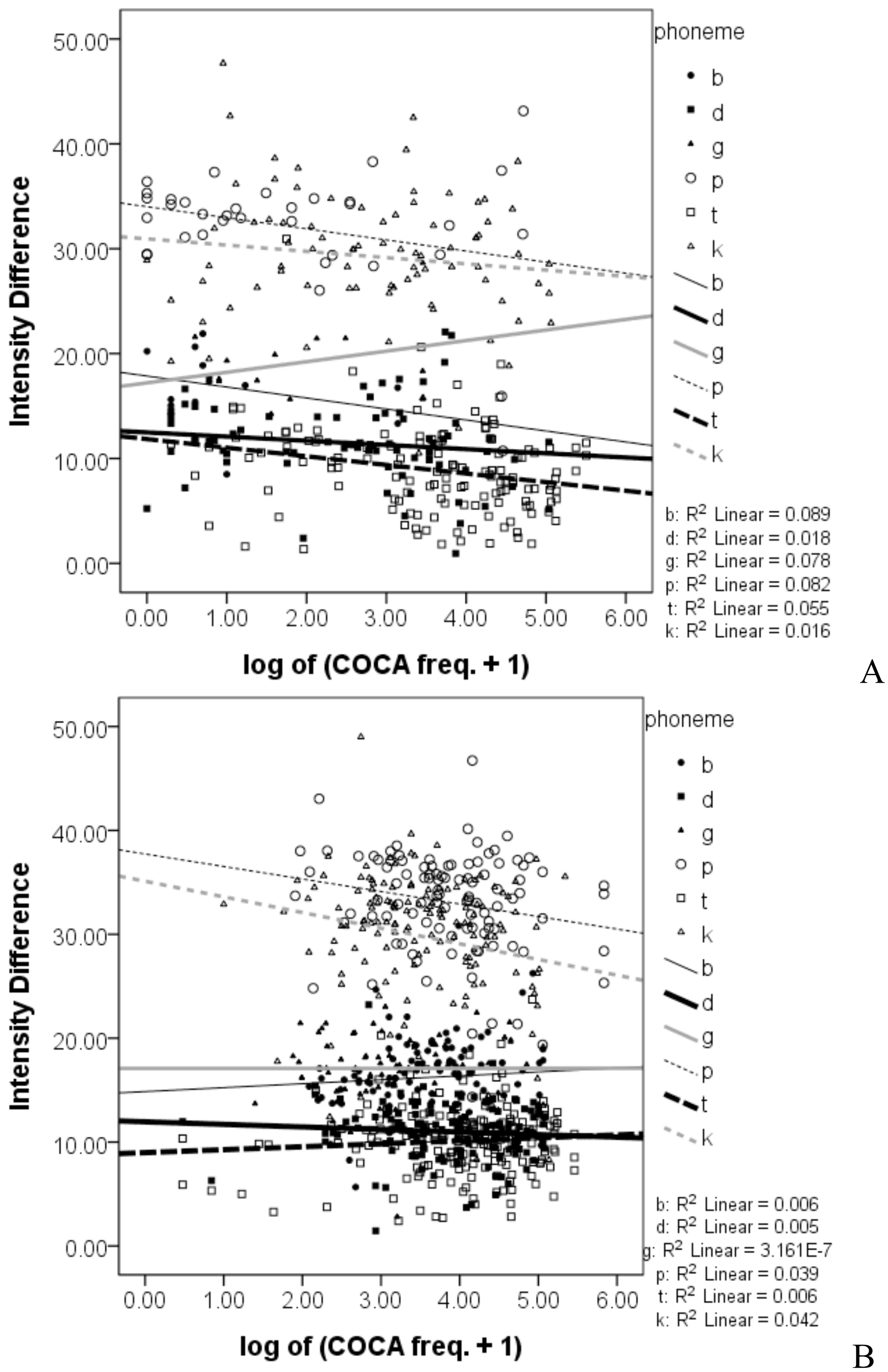

Intensity Difference
40.00
30.00
20.00
10.00
0.00
b
d
k
p
t
phoneme
style
conversation
story-reading
wordlist
A
Proportion with Cessation of F2 and F3
1.20
1.00
0.80
0.60
0.40
0.20
0.00
b
d
k
p
t
phoneme
style
conversation
story-reading
wordlist
B

Figure 5


Intensity Difference
50.00
40.00
30.00
20.00
10.00
0.00
0.00 1.00 2.00 3.00 4.00 5.00 6.00
log of (COCA freq. + 1)
phoneme
b
d
g
p
t
k
b
d
g
p
t
k
b: R2 Linear = 0.089
d: R2 Linear = 0.018
g: R2 Linear = 0.078
p: R2 Linear = 0.082
t: R2 Linear = 0.055
k: R2 Linear = 0.016
A
Intensity Difference
50.00
40.00
30.00
20.00
10.00
0.00
0.00 1.00 2.00 3.00 4.00 5.00 6.00
log of (COCA freq. + 1)
phoneme
b
d
g
p
t
k
b
d
g
p
t
k
b: R2 Linear = 0.006
d: R2 Linear = 0.005
g: R2 Linear = 3.161E-7
p: R2 Linear = 0.039
t: R2 Linear = 0.006
k: R2 Linear = 0.042
B