

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Journal of Phonetics 36 (2008) 91–113

---



---

**Journal of  
Phonetics**


---



---

[www.elsevier.com/locate/phonetics](http://www.elsevier.com/locate/phonetics)

# Stability of temporal contrasts across speaking styles in English and Croatian

Rajka Smiljanic\*, Ann R. Bradlow

*Linguistics, Northwestern University, Evanston, IL 60208, USA*

Received 6 December 2005; received in revised form 10 January 2007; accepted 12 February 2007

---

**Abstract**

In this paper, we investigate the effect of clear speech, a distinct, listener-oriented, intelligibility-enhancing mode of speech production, on vowel and stop consonant contrasts along the temporal dimension in English and Croatian. Our previous work has shown that, in addition to enhancing the overall acoustic salience of the speech signal through a decrease in speaking rate and expansion of pitch range, clear speech modifications increased the spectral distances between vowel categories in both languages despite the different sizes of their vowel inventories (+ 10 in English, five in Croatian). Here, we examine how clear speech affects the duration of English tense (long) vs. lax (short) vowels, English vowels preceding voiced (long) vs. voiceless (short) coda stops, Croatian long vs. short vowels and Croatian and English voice onset time (VOT) duration for voiced and voiceless stops. Overall, the results showed that the proportional distance between the 'short' and 'long' vowel categories and between the voiced and voiceless stop categories was remarkably stable across the two speaking styles in both languages. These results suggest that, in combination with the spectral enhancement of vowel contrasts, language-specific pronunciation norms along the temporal dimension are maintained in clear and conversational speech.

© 2007 Elsevier Ltd. All rights reserved.

---

**1. Introduction**

This paper presents a cross-linguistic study of the effect of 'clear speech,' a listener-oriented, intra-speaker variation in speaking mode for the express purpose of enhancing speech intelligibility, on production of temporal contrasts in English and Croatian. Specifically, we examine whether clear speech modifications enhance vowel and stop voicing contrasts along the single-acoustic-phonetic dimension of duration or whether the temporal contrasts remain stable across speaking styles in the two languages. Additionally, we explore the interaction of clear speech enhancement effects with those of prosodic position for these vowel and stop durational contrasts.

Smiljanic and Bradlow (2005) showed that spontaneously produced clear speech enhanced intelligibility for both English and Croatian listeners. They also found that in addition to enhancing the overall acoustic salience of the speech signal through a decrease in speaking rate and expansion of pitch range, clear speech

---

\*Corresponding author. Tel.: +1 847 491 3847; fax: +1 847 491 3770.

E-mail addresses: [rajka@northwestern.edu](mailto:rajka@northwestern.edu) (R. Smiljanic), [abradlow@northwestern.edu](mailto:abradlow@northwestern.edu) (A.R. Bradlow).

modifications expanded the  $F1 \times F2$  vowel space. This vowel space expansion increased the distances between the contrastive vowel categories in both languages regardless of the different sizes of their vowel inventories (five distinct vowel quality categories in Croatian vs. 10+ in English). The purpose of this paper is to further explore whether clear speech enhances the distinctiveness of phonemic categories by focusing on aspects of sound structure that differ across the two languages. In particular, we focus on vowel and stop consonant contrasts along the temporal dimension, i.e., vowel duration and voice onset time (VOT).

We chose these two aspects of the contrasting systems of English and Croatian because they are timing-based contrasts that differ in status across the two phonologies and in their phonetic implementation. Croatian has a phonemic vowel length contrast (five long and five short vowels) that is characterized largely by duration independently of vowel quality (Lehiste, 1970; Lehiste & Ivić, 1986; Skarić, 1991). While English phonology has no ‘pure’ phonemic vowel length contrast, duration plays a role in the tense vs. lax vowel contrast. Specifically, while English tense and lax vowels differ primarily in their spectral characteristics, tense vowels tend to be longer than lax vowels (Hillenbrand, Clark, & Nearey, 2001; Hillenbrand, Getty, Clark, & Wheeler, 1995; Klatt, 1976; Peterson & Barney, 1952). However, these duration differences are not entirely systematic, for instance, /æ/ is classified as a lax vowel yet in many American English dialects it is longer than /ɛ/, its closest lax neighbor (Hillenbrand et al., 1995; Peterson & Lehiste, 1960). A second English vowel duration contrast that we examine here is the case of vowels preceding a voiced vs. a voiceless coda consonant. Vowels before voiceless obstruents are predictably shorter than those before voiced obstruents (Klatt, 1976). Although the duration cue is redundant with voicing, it seems to be linguistically specified in English in that it is larger than in other languages and is highly relevant in perception of obstruent voicing (Chen, 1970; Zimmerman & Sapon, 1958). Furthermore, the ‘voice induced lengthening’ is enhanced under stress and focus in English suggesting that it may be encoded in the linguistic system and subject to similar enhancement in other hyperarticulation situations such as clear speech (de Jong, 2004).

In the present study, we examine how clear speech affects the duration of Croatian long vs. short vowels, English tense vs. lax vowels and English vowels preceding voiced vs. voiceless coda stops. We expect that some of these duration contrasts will be enlarged through larger clear speech lengthening of the longer member of the contrasting pair. In particular, we expect that the extent to which duration contrast will be enhanced in clear speech will depend on the status of the duration cue within the linguistic system (phonemic status vs. allophonic status vs. secondary cue status). This prediction is consistent with our overarching hypothesis that clear speech production reflects the interaction of universal, auditory–perceptual factors, which serve to enhance the overall acoustic salience of the speech signal, and phonological, structural factors, which serve to enhance the acoustic ‘distance’ between contrasting phonological categories.

In addition to investigating whether clear speech enhances vowel duration contrasts in ways that maximize phonological contrasts, we explore the effect of sentence prosody on the extent of any observed contrast enhancement. Prosodic factors have been shown to affect the realization of sounds in the temporal domain in ways that result in enhancement of durational contrasts (Cho & Keating, 2001; Cole, Kim, Choi, & Hasegawa-Johnson, 2007; de Jong, 1995; Edwards, Beckman, & Fletcher, 1991; Fougeron, 2001; Fougeron & Keating, 1997; Klatt, 1976; Lehiste, 1970; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Accordingly, we look at the durational contrast between English vowels preceding voiced vs. voiceless coda stops and of its enlargement in clear speech in two prosodic conditions, namely sentence-medial vs. sentence-final position.

Similar to the effect of clear speech on vowel duration, we explore the effect of speaking style on the temporal domain with respect to stop consonant production in the two languages. Specifically, we examine closure and VOT durations for voiced and voiceless stops in conversational and clear speaking styles. Although both languages have two phonological voicing categories, their phonetic realization along the VOT continuum is different. English ‘voiced’ and ‘voiceless’ stops are characterized by short and long lag VOT respectively (Lisker & Abramson, 1964), whereas, Croatian ‘voiced’ and ‘voiceless’ stops are characterized by prevoicing and short lag, respectively. The same phonetic realization of short lag VOT is, thus, represented as phonologically ‘voiced’ in English and as ‘voiceless’ in Croatian. Different phonetic realizations of the voicing contrast in the two languages will allow us to investigate whether the two categories are affected differently by changes in speaking style, as well as to explore whether clear speech enhances the voicing contrast in two languages. Here too, we explore whether prosody (word-initial vs. word-medial

position) interacts with speaking style in affecting the magnitude of this duration-based consonant contrast in English and Croatian.

### 1.1. Background

Clear speech is a distinct listener-oriented speaking style that speakers adopt when, for example, the listener has a hearing deficit or comes from a different language background. Conversational-to-clear speech modifications, therefore, involve hypo-to-hyper articulation changes (Lindblom, 1990) that are operationally similar to rate and focus modifications in that they all involve articulatory adjustments which are presumably aimed at achieving less target undershoot (e.g., de Jong & Zawaydeh, 2002; Hirata, 2004; Picheny, Durlach, & Braida, 1986; Uchanski, 1988, 1992). Clear speech differs from rate changes in that it is produced with the specific purpose of enhancing intelligibility and it differs from focus modifications in that it affects speech globally, i.e., the entire utterance is affected, rather than changes being localized on particular words under focus. The accumulated results of numerous studies demonstrate that English clear speech significantly enhances intelligibility for various listener populations under a variety of degraded listening conditions (Bradlow & Bent, 2002; Bradlow, Kraus, & Hayes, 2003; Ferguson, 2004; Payton, Uchanski, & Braida, 1994; Picheny et al., 1986; Uchanski, Choi, Braida, Reed, & Durlach, 1996). Investigation of the acoustic–phonetic features of clear speech has revealed that clear speech production typically involves insertion of longer and more frequent pauses, decreased speaking rate, wider dynamic pitch range, greater sound pressure levels, more salient stop releases and greater obstruent intensity, all of which increase the overall salience of the speech signal (Bradlow et al., 2003; Liu, Del Rio, Bradlow, & Zeng, 2004; Moon & Lindblom, 1994; Picheny, Durlach, & Braida, 1989; Picheny et al., 1986; Smiljanic & Bradlow, 2005).

A relatively small number of studies have pursued the idea that clear speech is guided by a principle of phoneme contrast enhancement. Several studies have shown that the English vowel space is expanded in clear speech when compared to conversational speech (Bradlow, 2002; Ferguson & Kewley-Port, 2002; Johnson, Flemming, & Wright, 1993; Moon & Lindblom, 1994; Picheny et al., 1986). More recently, expansion of the vowel space in clear speech was demonstrated in a cross-linguistic study by Smiljanic and Bradlow (2005) who showed equivalent vowel space expansion in English and Croatian regardless of the difference in vowel inventory size (more than 10 but just five distinct vowel quality categories in English and Croatian, respectively). Thus, clear speech production strategies generally make vowel categories spectrally more distinct from each other. Clear speech modifications of temporal patterns have also been shown to increase the distance between contrastive categories. For example, Uchanski (1988, 1992) found that the duration contrast between tense and lax vowels was enhanced in English clear speech by lengthening the tense vowels to a greater extent than the lax vowels. These vowel space and vowel duration results suggest that clear speech is in part guided by a principle of maximizing the distance between language-specific contrasting sound categories. The asymmetric lengthening of ‘short’ and ‘long’ vowels and the vowel expansion in the absence of a speaking rate decrease, as reported by Krause and Braida (2004), in clear speech suggest that such contrast enhancement strategies cannot be attributed entirely to the principle of uniform slowing down and the concomitant more extreme articulatory gestures. Rather, these conversational-to-clear speech modifications reflect a contrast-enhancing nature of hyperarticulated clear speech.

Other hypo- to hyperarticulation modifications (i.e., articulatory modifications that are likely to reduce target undershoot), such as broad to narrow focus and fast to slow speaking rates, seem to be guided by similar language-specific contrast enhancement principles. Studies by Heldner and Strangert (2001), de Jong and Zawaydeh (2002) and Smiljanic (2004, 2006) have investigated the amount of focal lengthening in Swedish, Arabic and Serbian and Croatian, respectively, all of which have lexically contrastive vowel length. In all the languages, lengthening long vowels more than short vowels in hyperarticulated speech increased the durational distinction that marks the phonological vowel length contrast.<sup>1</sup> In addition, Hirata (2004) and Pind (1995, 1999) found that speaking rate had a similar effect on the Japanese and Icelandic vowel length distinctions. As the rate decreased, long vowels increased in duration more than short vowels resulting in a

<sup>1</sup>de Jong and Zawaydeh (2002) explored the effect of stress and segmental focus on vowel duration. Segmental focus differs from narrow focus in its scope (the contrast is on one segment, *bear* vs. *pair*, rather than on the whole word, *John made dinner*, not *Bob*).

greater duration difference between the two categories in slow compared to normal and fast speech. In related work, [Smith \(2000, 2002\)](#) found differential effects of a speaking rate *increase* on various temporal patterns of English. Specifically, while an increase in speaking rate decreased the extent of vowel lengthening before voiced obstruents, it increased the extent of prepausal vowel lengthening (in both relative and absolute terms). Smith suggested that this interesting pattern may be due to differential effects of rate changes (and by extension, hypo–hyper articulation changes) on the ‘local’ (segmental) and ‘global’ (prosodic) levels.

Similar increase in durational distinction between the contrastive categories has been obtained in studies of the production of voicing. [Summerfield \(1981\)](#) and [Miller, Green, and Reeves \(1986\)](#) showed that speaking rate affects VOT duration differently for voiced and voiceless stops in English. As speaking rate is slowed, the VOT of voiceless stops lengthens while VOT of voiced stops is changed much less making the two voice categories more distinct. Furthermore, [Kessinger and Blumstein \(1997\)](#) found language-specific effects of speaking rate on VOT duration in English, French and Thai. English and French each have a two-way contrast: short and long lag in English and prevoiced and short lag in French. Thai has a three-way contrast among prevoiced, short lag and long lag categories. In all three languages speaking rate had an asymmetrical effect on the voicing categories. In Thai and French, prevoiced categories were affected while in English and Thai, long lag categories were affected. In all three languages, the VOT distribution of the short lag category remained stable across changes in speaking rate regardless of whether it was ‘bounded’ by another voicing category or not. These studies suggest that speaking rate modifications increase the distance among the voicing stop contrasts cross-linguistically.

In addition to style and speaking rate effects, numerous articulation and acoustic studies have demonstrated the effects of prosodic structure on realization of sounds in temporal domain ([Beckman, Edwards, & Fletcher, 1992](#); [Cho & Keating, 2001](#); [Cho & McQueen, 2005](#); [de Jong, 1995](#); [Edwards et al., 1991](#); [Fougeron, 2001](#); [Fougeron & Keating, 1997](#); [Klatt, 1976](#); [Lehiste, 1970](#); [Wightman et al., 1992](#)). Furthermore, similar to the style effects, prosodic structure has been shown to maximize phonological contrasts ([Cho, 2002, 2005](#); [Cho & Jun, 2000](#); [de Jong, 1995](#); [Hsu & Jun, 1998](#)). For instance, in Korean, which has a three-way voicing contrast among fortis, lenis and aspirated voiceless stops, a prosodic influence was seen in lengthening of lenis and aspirated stops but not of fortis stops [Cho and Jun \(2000\)](#). They argued that an increase in VOT for the fortis stops would reduce the contrast with the other two stops. All of the above results combined suggest that temporal modifications due to style, speaking rate and prosodic structure are not uniformly subject to lengthening/strengthening, but rather that these effects are limited by the system of language-specific contrasts in such a way that the paradigmatic contrasts are maximized.

Although there is some supporting evidence from the few studies that have been carried out that hyperarticulated speech in general is guided by contrast enhancement strategies, there is some counter-evidence that casts doubt on this principle. [Ohala \(1995\)](#), in a study of consonant production in English clear speech, failed to find evidence of temporal enhancement of VOT differences that distinguish stop consonant categories. The author argued that the established pronunciation norms are maintained in both speaking styles and these norms, rather than contrast enhancement principle, are guiding clear speech production. Similarly, a number of studies that examined the effect of speaking rate on temporal dimensions of speech, such as VOT, short/long vowel duration and single/geminate stop duration, argue for relational invariance in the production of these contrasts across speaking rates. Although they found that the duration difference is enlarged between the two members of the contrasting pair when expressed in absolute measures, proportional measures exhibited stability across speaking rates ([Boucher, 2002](#); [Hirata, 2004](#); [Hirata & Whiton, 2005](#); [Kessinger & Blumstein, 1998](#); [Pind, 1986, 1995, 1999](#); [Pickett, Blumstein, & Burton, 1999](#)). For instance, [Boucher \(2002\)](#) argued that the ratio of VOT to syllable duration for American English /t/ and /d/ is maintained across speaking rates which suggests that the contrast is not increased when rate is decreased. Instead, when viewed in relational terms, there is a rather stable /t-/d/ VOT boundary across rate variability. Furthermore, this study demonstrated that this stable boundary constitutes a perceptual criterion by which listeners judge the category affiliation of a given VOT. Although these various studies adopted different proportional measures (e.g., relative to syllable or rhyme duration) to best capture the relational invariant for quantity and VOT contrasts across speaking rates, they all show that relative temporal patterns are stable in both production and perception across speaking rates and indicate that speakers and listeners rely on rather local timing relations to allow for rate-independent speech processing.

In this paper, then, we set out to explore whether clear speech, which is distinguished from other forms of hyperarticulated speech in its connection to overall intelligibility enhancement for the listener, is driven by language-specific contrast enhancement strategies in the temporal domain of vowel and stop consonant production in English and Croatian. In addition, we wanted to explore whether durational contrasts are further enlarged in prosodically ‘strong’ positions (stressed vs. unstressed syllable and sentence-medial vs. final position). Finally, the goal of this paper is to investigate whether the same contrast maximization principle holds true in production for both absolute and proportional measures.

## 2. Experiment

### 2.1. Participants

Five native speakers of Croatian (two female and three male) served as participants in the experiment (henceforth referred to as CF1, CF2, CM1, CM2 and CM3). Their ages ranged between 18 and 25 years. They were undergraduate students at Northwestern University. They had all moved to the United States within the previous 5 years and were from the same region on the coast of Croatia.<sup>2</sup> Five native speakers of English (three female and two male) provided the English recordings (henceforth referred to as EF1, EF2, EF3, EM1 and EM2). All were graduate students in the Linguistics Department at Northwestern University. Their age range was between 28 and 48 years. None of the talkers had any known speech or hearing impairment at the time of recording. They were not aware of the purpose of the recordings. They were paid for their participation.

### 2.2. Stimuli

Twenty sentences were designed in each language to investigate the effect of clear speech production and perception in Croatian and English. In order to minimize the signal-independent contextual cues available to listeners in the perception tests, we constructed semantically anomalous sentences. Example sentences are given in (1) for each language.<sup>3</sup> Target vowels are italicized and target stops are underlined:

- a) Croatian: Nada æ dobiti tri dokaza i puni mjeseć.  
 ‘Nada will get three proofs and a full moon.’
- b) English: Your tedious beacon lifted our cab.

The particular words used in all sentences were selected to allow for measurement of language-specific phonological contrasts of each language. English sentences contained words with three instances of the following ten target tense and lax vowels /i, e, a, o, u, ɪ, ε, æ, ʌ, ʊ/. The target vowels were always in initial, open, stressed syllables embedded between two stop consonants. The majority of the target words were bisyllabic. Vowels in the tense and lax series were followed by a roughly equal number of voiced and voiceless stops so as to minimize the effect of the coda voice on vowel length (seven lax vowels were followed by voiced stops and eight lax vowels were followed by voiceless stops; six tense vowels were followed by voiced stops and nine tense vowels were followed by voiceless stops). Three minimal pairs (*cab, cap, bet, bed, pick* and *pig*) were used for examination of the effect of coda voicing on English vowel duration. Croatian sentences contained target words with three tokens of all five short and long vowels: /a, e, i, o, u/. All target vowels were stressed, in open syllables and embedded between two stop consonants. Most target vowels were in the word-initial syllable of either two or three syllabic words.

<sup>2</sup>Croatian talkers started studying English in elementary school as part of the standard foreign language curriculum in Croatia. They are fairly fluent in English as confirmed by the General Record Examination (GRE) and Test of English as a Foreign Language (TOEFL) scores required for admission to Northwestern University. They all continue to use Croatian daily (in communicating with each other on campus or with their families in Croatia). We do not, therefore, expect that their fluency in English affected their Croatian clear speech productions.

<sup>3</sup>A full set of stimuli is available alongside the electronic version of the published article at Supplementary data.

In addition, English sentences contained three tokens of word-initial /b, d, p, t/ and word-medial /b, g, p, k/. Croatian sentences contained three tokens of word-initial /b, d, p, t/ and word-medial /b, d, g, p, t, k/.<sup>4</sup> All word-initial stops were onsets of stressed syllables and were preceded by either a vowel or a consonant other than a stop (e.g., fricative, nasal) which ended the previous word. Word-initial stops were always followed by a vowel. Word-medial stops were onsets of unstressed syllables and were in all instances in intervocalic positions. Some of the word-initial and word-medial tokens came from the same word.<sup>5</sup> Croatian and English sentences were of similar length: the mean number of syllables was 12.8 (range 10–16) and 11.7 (range 9–14) in Croatian and English, respectively.

In order to explore the effect of sentence level prosody on the extent of contrast enhancement, a position-within-sentence (medial vs. final) manipulation was introduced for English. An additional 20 sentences were constructed in which sentence-final monosyllables (included to investigate the effect of coda voicing on English vowel duration) were moved to a non-final position. The rest of the sentences in the second set were not manipulated, i.e., the two sets of English sentences differed only in the position of the monosyllabic minimal pairs.

### 2.3. Procedure

All English and Croatian speakers were recorded producing all 20 sentences in their native language in a sound-attenuated booth in the phonetics laboratory in the Department of Linguistics at Northwestern University. The participants read the sentences written on index cards into a microphone. The speech was recorded directly to disk at 24 bit accuracy using an Apogee PSX-100 A/D D/A converter at a sampling rate of 16 kHz. Participants read the sentences once in conversational and once in clear speech. English speakers recorded two sets of sentences (with monosyllabic minimal pairs in final and in non-final sentence positions) in two sessions 6 weeks apart. For the conversational style, the speakers were instructed to read as if they were talking to someone familiar with their voice and speech patterns. For the clear speaking style, the speakers were instructed to read as if they were talking to a listener with a hearing loss or a non-native speaker. Sentences were randomized for each reading. This yielded 40 sentences per speaker (200 total) for Croatian and 80 sentences per speaker (400 total) for English.

It is important to note that both ‘conversational’ and ‘clear’ speech terms used throughout this paper refer to read laboratory speech elicited by specific instructions given to talkers rather than in a more natural setting. Although they may not designate the speech elicited here most accurately, we continue to use these labels in keeping with previous work on clear speech (Uchanski, 2005 and references therein) and to provide consistency with our companion paper (Smiljanic & Bradlow, 2005).

### 2.4. Measurements and data analysis

In order to investigate the articulatory modifications that speakers adopted in clear speech production, we performed a series of comparable acoustic analyses in both languages. The specific acoustic–phonetic parameters measured were vowel duration (long vs. short vowels in Croatian, tense vs. lax vowels in English and vowels preceding voiced vs. voiceless coda stops in sentence-final and medial positions in English) and closure duration and VOT of voiced and voiceless stops in the word-initial and medial positions in both languages. All vowels were embedded between two stops and their durations were measured from the onset of voicing and periodicity as seen in the spectrogram and the waveform to the beginning of the closure for the following stop as marked by the cessation of regular periodicity and a substantial decrease in the amplitude of the waveform. The offglides of the English long vowels /e/ and /o/ were included in the duration of the vowel.

<sup>4</sup>Word-medial English /t, d/ stops were not included in the VOT and closure duration analyses due to the expected /t/ flapping intervocalically. There were not sufficient word-initial English /k, g/ tokens to include in the analyses since our initial cross-style comparison was intended for word-medial tokens only. The same pair was excluded from Croatian word-initial analyses so that the VOT duration measurements would be more comparable across the two languages.

<sup>5</sup>Although we did not control for all possible factors that affect vowel and VOT duration (e.g., number of segments and syllables in the word), we expect to find the same influences and amount of variability in clear as in conversational speaking styles since the same words/sentences were used in both styles.

VOT was measured as the time between the onset of the release burst and the onset of the periodic energy of the following vowel. In the case of prevoicing, VOT was measured between the onset of the stop release burst and the beginning of voicing during closure (yielding negative VOT numbers). Closure duration was measured from the offset of the preceding vowel, nasal or a fricative to the beginning of the stop burst. All measurements were made from inspection of waveforms and wideband spectrograms and by listening using PRAAT software for speech analysis (Boersma & Weenink, 2006). The total number of English tense and lax vowels analyzed was 300 (10 vowels  $\times$  3 tokens  $\times$  2 styles  $\times$  5 speakers) and of vowels preceding voiced vs. voiceless stops was 120 (6 vowels  $\times$  2 styles  $\times$  2 positions  $\times$  5 speakers). The total number of Croatian short and long vowels analyzed was 300 (5 vowels  $\times$  2 lengths  $\times$  3 tokens  $\times$  2 styles  $\times$  5 speakers). The total number of English voiced and voiceless stops analyzed was 240 (4 stops  $\times$  3 tokens  $\times$  2 positions  $\times$  2 styles  $\times$  5 speakers). The total number of Croatian word-initial stops analyzed was 120 (4 stops  $\times$  3 tokens  $\times$  2 styles  $\times$  5 speakers). Croatian word-medial stops were not analyzed due to the extensive lenition in this position (see Results below).

### 3. Results

#### 3.1. Vowel duration

The goal of this section is to investigate the effect of clear speech on vowel duration variability across ‘long-short’ vowel pairs in the two languages and of sentence prosody on English vowel duration contrasts before voiced vs. voiceless stops by comparing the effect of position-in-sentence (medial vs. final) on the magnitude of the contrast. In order to explore the possibility of relational invariance for the temporal contrasts under investigation here, we examined proportional vowel lengthening in clear speech as well.

##### 3.1.1. English: tense vs. lax vowel duration

Average tense and lax vowel durations for English in conversational and clear speaking styles are shown in Fig. 1. The average values for each vowel separately and for the average tense and lax series are given in Table A1 in Appendix A.

Repeated-measures analysis of variance (ANOVA) for the effect of style (conversational vs. clear), length (tense vs. lax) and vowel pair (/a-æ, e-ɛ, i-ɪ, o-ʌ, u-ʊ/) on vowel duration was conducted. The vowels were paired for height and tenseness (so that each pair contains one phonologically tense and one lax member). The vowel pair factor was introduced to allow the investigation of the possible vowel by length interaction (e.g., lax /æ/ is often longer than other lax vowels). The results revealed that all three main effects were significant: (style:  $F(1,4) = 13.946, p < .05$ ; length:  $F(1,4) = 98.379, p < .001$ ; vowel pair:  $F(4,16) = 39.808, p < .001$ ). Only the length by vowel pair two-way interaction was significant: (length by vowel pair:  $F(4,16) = 41.968, p < .001$ ; style by length:  $F(1,4) = 3.626, p = .130$  ns; style by vowel pair:  $F(4,16) = 1.650, p = .211$  ns). The three-way interaction was not significant:  $F(4,16) = .790, p = .548$  ns. These results showed that all vowels were lengthened in clear speech. Furthermore, the duration cue distinguished tense from lax vowels. Finally, the five vowel pairs differed in their overall duration. The absence of a significant style by length interaction indicated that both lax and tense vowels were lengthened to a similar degree and that the tense/lax duration contrast

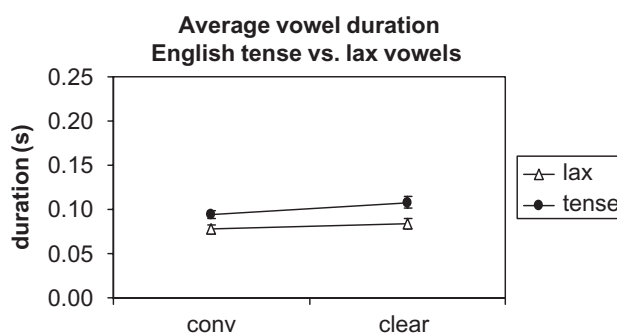


Fig. 1. Average lax and tense vowel duration (s) in conversational and clear speaking styles. The scale on the y-axis is kept the same as in Figs. 2 and 3 for ease of comparison. The error bars in this and subsequent vowel plots represent standard errors of the mean.

remained equivalent in clear speech for most of the speakers tested here. However, the significant two-way length by vowel pair interaction suggested a difference in durational distinction between tense and lax vowels for the five vowel pairs. We conducted separate paired *t*-tests for each vowel pair (collapsed over the two speaking styles since there was no significant two- or three-way interaction with style) to explore whether the durational relationship between the two members of each pair was significantly different, i.e., whether tense vowels were always longer than lax vowels for each pair. The durational distinction was significant for all pairs at  $p < .01$  except for /ɑ-æ/ pair: (/e-ɛ/:  $t(4) = 18.500$ ; /i-ɪ/:  $t(4) = 4.491$ ; /o-ʌ/:  $t(4) = 8.232$ ; /u-ʊ/:  $t(4) = 18.174$ ; /ɑ-æ/:  $t(4) = -1.037$ ,  $p = .358$  ns.). It can be seen in Table A1 that for four pairs the tense member of the pair was longer than the lax member. In contrast, /æ/, a lax vowel, was longer than the tense /ɑ/ and much longer than the other lax vowels. These results for tense/lax vowel duration confirm earlier findings of the tense/lax durational contrast (Crystal & House, 1988; Hillenbrand et al., 1995, 2001; Peterson & Lehiste, 1960).

Since we were interested in investigating whether the durational contrast for tense and lax vowels was enlarged in clear speech, and it was possible that the inclusion of the much longer /æ/ as a lax vowel, masked a significant style by length interaction, we ran an additional ANOVA with the /ɑ-æ/ pair excluded. The results of the repeated-measures ANOVA for the effect of style (conversational vs. clear) and length (tense vs. lax) on vowel duration revealed a similar pattern to the one obtained when vowels /ɑ/ and /æ/ were included in the analyses. The effect of style and length were both significant: (style:  $F(1,4) = 8.534$ ,  $p < .05$ ; length:  $F(1,4) = 420.464$ ,  $p < .001$ ). The two-way style by length interaction was not significant:  $F(1,4) = 2.980$ ,  $p = .159$ . These results showed that even when vowels /ɑ/ and /æ/ are excluded, the effect of style was equivalent for both tense and lax vowels and the durational contrast remained stable across two speaking styles.

Percentage vowel lengthening relative to their conversational duration (duration of vowels in clear speech minus duration of vowels in conversational speech divided by duration of vowels in conversational speech (henceforth clear-conv/conv)) of lax and tense vowels in clear speech was 9% and 14% (7% and 15% without /ɑ-æ/) respectively. Pairwise comparison indicates that overall proportional clear speech lengthening for tense and lax vowels was not significantly different ( $t(4) = 1.195$ ,  $p = .298$  with /ɑ-æ/;  $t(4) = 1.101$ ;  $p = .333$ ). The average tense/lax duration ratios were 1.22 and 1.28 (1.33 and 1.43 without /ɑ-æ/) in conversational and clear speech, respectively (see Table A1 in Appendix A). Combined, the results for both absolute and proportional measures suggest that the tense vs. lax vowel duration contrast remained unchanged in clear speech for most speakers.

### 3.1.2. English: vowel duration before voiced vs. voiceless coda consonant in sentence-medial and final positions

Average durations of vowels preceding voiced and voiceless stops for English in conversational and clear speaking styles in sentence-medial (left) and sentence-final (right) positions are shown in Fig. 2. The corresponding numerical values are given in Table A2 in Appendix A.

Repeated-measures ANOVA results for the within-subjects factors of style (conversational vs. clear), position (sentence-medial vs. sentence-final) and voice (voiced coda consonant vs. voiceless coda consonant)

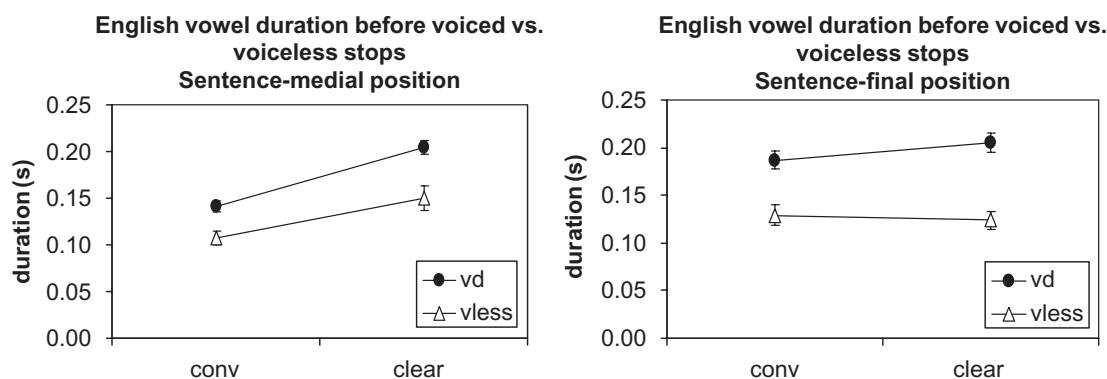


Fig. 2. Average vowel duration (s) before voiced vs. voiceless coda stop in conversational and clear speaking styles for sentence-medial (left) and sentence-final (right) positions.

on vowel duration revealed a main effect of style ( $F(1,4) = 89.580, p < .001$ ) and of voice ( $F(1,4) = 75.076, p < .001$ ), but not of position ( $F(1,4) = 2.211, p = .211$ ). All two-way interactions were significant as well: style by position ( $F(1,4) = 42.079, p < .05$ ); style by voice ( $F(1,4) = 17.437, p < .05$ ); position by voice ( $F(1,4) = 53.221, p < .05$ ). The three-way style by position by voice interaction was not significant ( $F(1,4) = .075, p = .798$ ). These results reveal that vowel duration varied systematically for voiced and voiceless coda consonants, i.e., vowels were longer before a voiced stop than before a voiceless stop. Furthermore, vowels were lengthened in clear speech. The significant two-way interactions indicate that the effect of style on vowel duration differed for voiced and voiceless coda consonant in the two positions within a sentence. Separate ANOVAs for each sentence position revealed that while voice is a significant factor in both sentence positions (sentence-medial: ( $F(1,4) = 73.851, p < .001$ ); sentence-final: ( $F(1,4) = 73.497, p < .001$ )) style was significant only in sentence-medial position (sentence-medial: ( $F(1,4) = 104.779, p < .001$ ); sentence-final: ( $F(1,4) = 2.364, p = .199$ )). The two-way style by voice interaction was significant in sentence-final position ( $F(1,4) = 17.226, p < .05$ ) but not in sentence-medial position ( $F(1,4) = 7.140, p = .056$ ). The effect of style, thus, was different in two sentence positions. As seen in Fig. 2, in sentence-medial position, vowels in both voicing contexts were lengthened in clear speech. In sentence-final position, on the other hand, vowels before voiceless stops were slightly shortened while vowels before voiced stops were lengthened in clear speech. The duration contrast between vowels in the context of voiced vs. voiceless coda consonants was, thus, only marginally increased for some speakers in clear speech in sentence-medial position while it was increased significantly in clear speech in sentence-final position.

Average proportional lengthening (expressed as percentage) for vowels preceding voiceless and voiced coda stops in clear speech relative to the vowel duration in conversational speech (clear-conv/conv) in sentence-medial position were 41% and 46%, respectively. In sentence-final position, average percentage lengthening were 11% for vowels before voiced codas and average percentage *shortening* was 3% for vowels preceding voiceless codas. ANOVA results for the factor of position (sentence-medial vs. sentence-final) and voice (voiced coda vs. voiceless coda) on percentage vowel lengthening reveal that there was a significant main effect of position ( $F(1,4) = 42.549, p < .05$ ) but not of voice ( $F(1,4) = 3.686, p = .127$ ). The two-way interaction of position and voice was not significant either ( $F(1,4) = 1.129, p = .348$ ).

These results indicate that vowels before voiced and voiceless codas showed similar proportional lengthening for clear speech relative to conversational speech although the lengthening was overall larger in sentence-medial than in sentence-final position. In other words, the duration contrast for vowels before voiced and voiceless codas was rather stable across two speaking styles for most speakers in both sentence positions. The average ratio for vowels before voiced/voiceless codas in sentence-medial position was 1.31 and 1.36 in conversational and clear speech, respectively. The average ratio for vowels before voiced/voiceless codas in sentence-final position was 1.45 and 1.66 in conversational and clear speech, respectively (see Table A2, Appendix A). The relative duration results, thus, differ from the results for vowel lengthening expressed in absolute measures in that the increase in the proportional duration contrast was of equal magnitude in both prosodic positions whereas the absolute duration contrast was increased in sentence-final position.

### 3.1.3. Croatian: short vs. long vowels

Average short and long vowel durations for Croatian in the two speaking styles are shown in Fig. 3. The average values for each vowel separately and for the average long and short series are given in Table A3 in Appendix A.

Repeated-measures ANOVA results for the effect of style (conversational vs. clear), length (long vs. short) and vowel pair (/a:-a/, /e:-e/, /i:-i/, /o:-o/ and /u:-u/ where : indicates length) on vowel duration showed a significant main effect of all three factors: (style:  $F(1,4) = 18.241, p < .05$ ; length:  $F(1,4) = 31.638, p < .01$ ; vowel pair:  $F(4,16) = 28.090, p < .001$ ). The style by length and length by vowel pair two-way interactions were significant: (style by length:  $F(1,4) = 9.535, p < .05$ ; length by vowel pair:  $F(4,16) = 13.672, p < .001$ ; style by vowel pair:  $F(4,16) = 1.767, p = .185$  ns.). The three-way interaction was not significant:  $F(4,16) = .522, p = .721$  ns. The statistical analyses confirmed that long and short vowels in Croatian significantly differed in duration. Furthermore, all vowels were longer in clear speech. Finally, the durational distinction varied for the different vowel pairs.

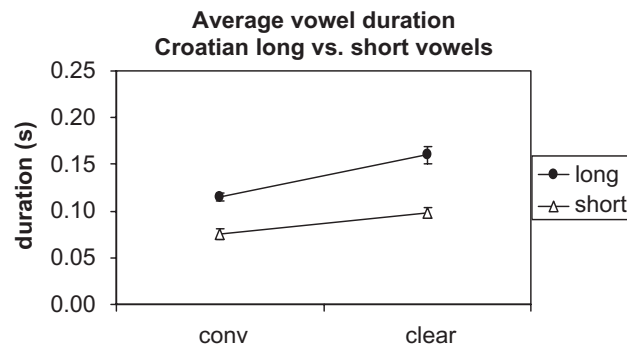


Fig. 3. Average long and short vowel duration (s) in Croatian in conversational and clear speaking styles.

The significant two-way style by length interaction revealed that the effect of style differed for the two vowel categories, i.e., short vowels were lengthened less in clear speech than long vowels. This asymmetrical pattern of lengthening enlarges the contrast between the two vowel categories in hyperarticulated clear speech. The significant two-way length by vowel pair interaction revealed a difference in durational distinction between long and short vowels for the five vowel pairs. We conducted separate paired *t*-tests for each vowel pair (collapsed over the two speaking styles since there was no significant three-way interaction with style) to explore whether long vowels are always significantly longer than short vowels. For four vowel pairs (except /i:-i/) the long member was significantly longer than the short pair at  $p < .01$ : (/a:-a/:  $t(4) = 10.119$ ; /e:-e/:  $t(4) = 5.872$ ; /o:-o/:  $t(4) = 4.426$ ; /u:-u/:  $t(4) = 4.274$ ; /i:-i/:  $t(4) = 2.400$ ,  $p = .074$  ns.). Even for /i:-i/ pair the long vowel was longer than the short vowel. However, the long and short /i/ were shorter than the other long and short vowels respectively and the durational distinction between /i:-i/ was smaller when compared to other vowel pairs (see Table A3). Nevertheless, the overall distinction between short and long series was rather robust.

Average percentage lengthening of short and long vowels relative to their conversational duration (clear-conv/conv) were 29% and 39%, respectively. For four out of five speakers, long vowels were proportionally lengthened more than short vowels. Such asymmetrical lengthening increased the distance between the two vowel categories. However, pairwise comparison of proportional lengthening indicated that long vowels were not lengthened significantly more than the short vowels in clear speech ( $t(4) = 1.929$ ,  $p = .126$ ). The average long/short duration ratios in clear and conversational speech were 1.52 and 1.63, respectively (see Table A3, Appendix A). In Croatian, then, similar to the vowel duration distinction before voiced vs. voiceless stops in English, the contrast was enhanced only when expressed in absolute measures but remained stable when expressed proportionally.

### 3.2. Summary and discussion

Considering absolute measures first, the results revealed a difference between the English tense/lax vowels on the one hand, and the English vowels preceding voiced and voiceless codas and Croatian short and long vowels on the other in terms of the magnitude of the contrast and their durational modification in clear speech relative to conversational speech. The tense/lax durational distinction, albeit systematic, was of smaller magnitude than either of the other two vowel duration distinctions. Furthermore, both tense and lax vowels were lengthened equally in clear speech, i.e., the durational contrast was of similar magnitude in both speaking styles. Duration, thus, appears to play the role of a secondary cue in the production of the tense/lax vowel contrast with the primary role being played by spectral cues (Hillenbrand et al., 1995, 2001; Klatt, 1976; Peterson & Barney, 1952). The secondary role of duration in distinguishing lax from tense vowels was demonstrated in perception by Hillenbrand, Clark, and Houde (2000) who showed that in vowel perception duration had a measurable but rather modest overall effect on tense vs. lax vowel identification. They argued that although vowels such as /i:-i/ and /u:-u/ showed systematic durational differences in production their spectral features were sufficiently distinct so that duration played a rather limited role in the perception of their quality. Contrary to this conclusion, de Jong (2004) found that stress and focus systematically expanded

the durational, but not spectral, difference between /æ/ and /ɛ/. Although both vowels belong to the lax series, they were largely distinguished by duration which was enlarged under stress and focus. In the present study, we conducted additional analyses with /æ-ɑ/ pair excluded, so that the much longer lax /æ/ vowel would not obscure the overall durational patterns and possible enhancement of the contrast between lax and tense vowels. However, there was no interaction between style and length. These results suggest that the magnitude of the tense/lax duration contrast was maintained in both speaking styles.

The durational distinction of English vowels preceding voiced vs. voiceless codas and of Croatian long and short vowels was rather comparable (English sentence-medial conv: 33 ms, clear: 55 ms; English sentence-final conv: 58 ms, clear: 82 ms; Croatian conv: 39 ms, clear: 62 ms) and much larger than for the English tense/lax distinction (conv: 17 ms, clear: 24 ms). Both English vowels preceding voiced vs. voiceless coda and Croatian short vs. long vowels lengthened asymmetrically, i.e., the long member of the pair was lengthened more than the short member. The absolute durational contrast was, thus, enhanced in both languages. In English, the difference between the two vowel series increased by 21 and 24 ms in sentence-medial and sentence-final positions respectively and in Croatian by 23 ms (for English tense/lax vowels the distinction increased by only 7 ms in clear speech). Although the contrast enhancement did not reach statistical significance in the case of English vowels before voiced vs. voiceless codas in sentence-medial position, the overall more similar patterning of this English vowel duration contrast to that of Croatian long vs. short vowels suggests that this duration contrast is a learned feature of the English obstruent voicing contrast and has a different status in the linguistic system from that of the duration cue in the English tense/lax contrast. Similar enhancement of the voiced/voiceless duration contrast under stress and focus was found for English but not for Arabic where presumably vowel duration is not linguistically specified as a part of the voicing contrast (de Jong, 2004; de Jong & Zawaydeh, 2002). Note also that this pattern is consistent with the findings reported by Smith (2000, 2002) regarding the reduction of this English duration contrast in fast speech. In other words, the English vowel duration contrast before voiced vs. voiceless codas appears to increase systematically along the hypo-to-hyper articulation continuum whether examined with respect to conversational-to-clear enhancement (present study) or with respect to stress/focus enhancement (de Jong, 2004) or with respect to medium-to-fast rate reduction (Smith, 2000, 2002).

Vowel duration before voiced vs. voiceless codas was affected by sentence prosody (sentence-medial vs. final position) in addition to speaking style. Although the duration contrast between the two vowel series was larger in clear speech in sentence-final position compared to sentence-medial position, the overall effect of clear speech was smaller in sentence-final compared to sentence-medial position. First, in sentence-final position, clear speech did not affect vowel duration before voiceless stops while it did so in the sentence-medial position. Second, clear speech lengthening of vowels before voiced stops in sentence-medial position was larger than in sentence-final position (64 vs. 19 ms). This was presumably due to the combined effect of domain-final lengthening and clear speech. Since vowels were overall longer in sentence-final position, lengthening them in clear speech to the same extent as in sentence-medial position could possibly have yielded unnaturally long vowels. Therefore, in order to make the contrast between vowels preceding voiced vs. voiceless stops more salient, the duration of vowels before voiceless stops remained unchanged or was even shortened for some speakers. In this way, the vowel duration contrast was enhanced in a prosodically more prominent sentence-final position despite a somewhat reduced lengthening effect of clear speech. These results suggest that final lengthening and clear speech effects were not additive, i.e., there was not a constant amount of lengthening that accompanied both effects which was then added when both factors influenced vowel realization. There was a limit to the amount of combined prosodic and style-related lengthening.

In contrast with the absolute durations, the ratios of vowel lengthening in clear speech relative to the vowel duration in conversational speech were less affected by changes in speaking style in both languages. In other words, the relational measures of both 'short and long' members of all three vowel series discussed above were affected equally by hyperarticulated clear speech. Similar stability of relational measures across speaking rates was reported by Hirata (2004), Hirata and Whiton (2005), Boucher (2002) and Pickett et al. (1999) among others. For instance, Hirata (2004) found that although the Japanese phonemically short and long vowels were affected by speaking rate in a way that increased the duration contrast in absolute measures, the ratios of long-to-short vowels in real words were not affected at normal and slow speaking rate (although they were smaller in fast speech). The author argued for the existence of 'relational' acoustic invariance which is stable across

speaking rates. Although we considered relative lengthening across speaking styles rather than long-to-short vowel ratios within speaking styles (since our stimuli did not contain minimal pairs), our results suggest that the relative relationship between short and long vowels was maintained in both speaking styles due to equal lengthening. Furthermore, average long-to-short vowel ratio across speaking styles showed similar consistency to that found in Hirata (2004). In the present study, the English tense/lax vowel ratio was 1.22 and 1.28 in conversational and clear speech, respectively. The English vowel before voiced/voiceless ratio in sentence-medial position was 1.31 and 1.36 in conversational and clear speech, respectively and in sentence-final position it was 1.45 and 1.66 in conversational and clear speech, respectively. Finally, the Croatian long/short vowel ratio was 1.52 and 1.63 in conversational and clear speech, respectively. In all considered cases, both the relative lengthening and the long-to-short vowel ratios exhibited remarkable stability across the two speaking styles.

These results suggest that the temporal dimension is at best a rather limited dimension of vowel contrast enhancement regardless of the status of the duration cue in the linguistic system, i.e., Croatian phonemically long and short vowels showed the same relational invariance as did both English vowel contrasts examined here (i.e., tense vs. lax vowels and vowels before voiced vs. voiceless codas). It thus appears that language-specific pronunciation norms in terms of relative timing of vowel duration contrasts are maintained in clear and conversational speech.

### 3.3. Closure duration and VOT

In addition to the effect of clear speech on vowel duration contrasts, we investigated the effect of speaking style and of prosodic position within the word (word-initial/stressed vs. word-medial/unstressed) on VOT contrasts and on closure duration.<sup>6</sup> In order to take into account changes in speaking rate, we examined proportional VOT lengthening as well (ratios of aspiration (for voiceless stops) and of voicing (for voiced stops) to total stop duration).<sup>7</sup>

#### 3.3.1. English closure duration and VOT in word-initial and medial stops

Closure duration measurement results for English stops in word-initial/stressed and word-medial/unstressed positions in two speaking styles are shown in Fig. 4. As seen, the overall closure duration was longer in clear speech than in conversational speech. This was especially true for stops in the word-initial position when compared with the word-medial position.

Repeated-measures ANOVA results for the effect of style (conversational vs. clear), voice (voiced vs. voiceless) and position (word-initial vs. word-medial) on stop closure duration showed a significant main effect of style ( $F(1,4) = 34.296, p < .001$ ) and of position ( $F(1,4) = 145.634, p < .001$ ), and a marginal effect of voice ( $F(1,4) = 8.155, p = .046$ ). The two-way style by position and voice by position interactions were significant: (style by position ( $F(1,4) = 9.465, p < .05$ ); voice by position ( $F(1,4) = 21.840, p < .05$ ). The style by voice interaction was not significant ( $F(1,4) = .012, p = .917$ ). Finally, the three-way style by voice by position interaction was significant ( $F(1,4) = 19.318, p < .05$ ).

The results of the statistical analyses, therefore, confirmed the general observation that the overall stop closure duration was longer in clear than in conversational speech, as well as in the word-initial when compared to the word-medial position. The overall stop closure duration differed slightly for voiced and

<sup>6</sup>Our data do not allow us to disentangle the contribution of stress and word-initial position in stop consonant ‘strengthening’. In addition, most of the words containing target stops bear accents. All of these factors have been shown to affect the realization of consonants (Cho & Keating, 2001; de Jong, 1995; Edwards et al., 1991; Fougeron, 2001; Fougeron & Keating, 1997; Wightman et al., 1992, etc.). In this study, however, we compare the ‘strongest’ and ‘weakest’ prosodic effects, i.e., we collapse stress and word-initial position and compare this with the effect of unstressed, word-medial position on stop consonant production.

<sup>7</sup>We looked at VOT as a percentage of the stop duration rather than of the syllable duration (as was done by Boucher, 2002) because it was often impossible to measure the duration of the unstressed syllable in the conversational speaking style due to extensive coarticulation with the following word. Furthermore, word-initial syllables all had CV structure while word-medial ones had varying structure (due to the nature of our corpus/task since we used all real words). Boucher found that the same stable boundary can be found when VOT ratios are calculated by referring to the duration of C1V, VC2 and C1VC2 (where VOT of C1 is measured). Therefore, it is possible that VOT relative to the smallest unit (stop duration) will exhibit the same regularity.

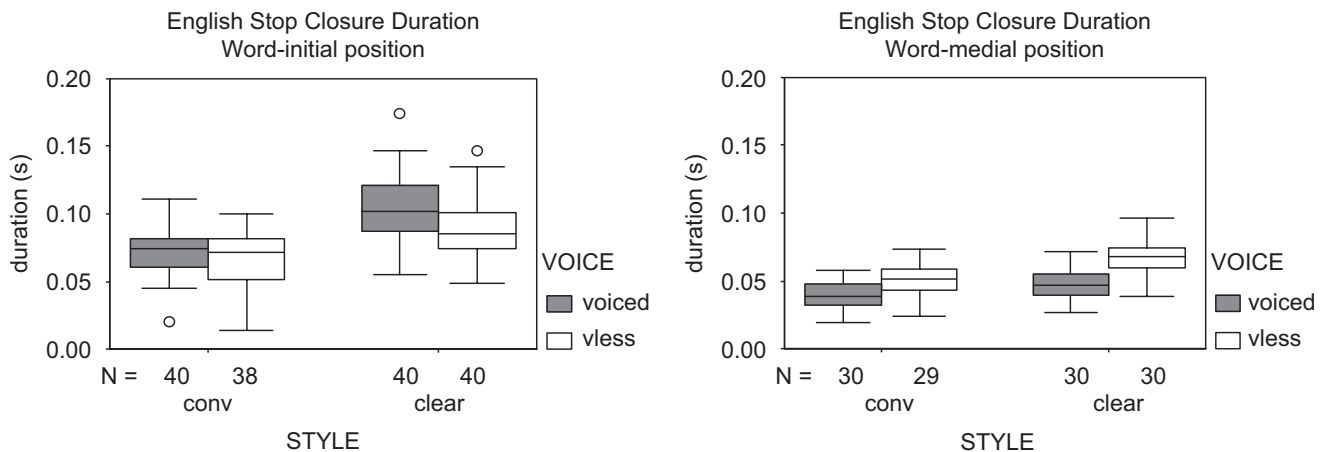


Fig. 4. Box plots of closure durations for word-initial and word-medial stops in two speaking styles. In this and subsequent box plots, the boxes represent the interquartile range, which contains the middle 50% of values. The whiskers are lines that extend from the box to the highest and lowest values, excluding outliers (given as circles). A line across the box indicates the median.

voiceless stops. The significant interactions indicated that the effect of clear speech on closure duration of voiced vs. voiceless stops differs for different within-word positions. To investigate these interactions more closely, we conducted separate ANOVAs for word-initial and word-medial positions. The results for the effect of style (clear vs. conversational) and voice (voiced vs. voiceless) on stop closure duration in the word-initial position showed a main effect of style ( $F(1,4) = 22.994, p < .001$ ), and of voice ( $F(1,4) = 18.342, p < .05$ ). The style by voice interaction was marginally significant ( $F(1,4) = 7.857, p = .049$ ). For the word-medial stops, there was a significant main effect of both style and voice on stop closure duration (style: ( $F(1,4) = 63.186, p < .05$ ); voice: ( $F(1,4) = 22.767, p < .05$ ). The style by voice interaction was also significant ( $F(1,4) = 22.185, p < .05$ ). The ANOVAs for word-initial stops revealed that there was a difference in the closure duration between voiced and voiceless stops. All stop closures were lengthened in clear speech, however, the significant two-way style by voice interaction indicated that voiced stop closures were lengthened slightly more in clear speech compared to voiceless stop closures (a similar effect of accent on closure duration was reported in Cole et al., 2007). Closure duration for the word-medial voiceless stops, on the other hand, was lengthened more in clear speech than closure duration of voiced stops.

Stop closure lengthening in clear speech is consistent with the overall slowing down in clear speech which is reflected both in the production of fewer syllables per second and in the insertion of more pauses (as reported in Smiljanic and Bradlow (2005) for these specific materials). The increase in the closure duration distinction between voiced and voiceless stops suggests an increase in the voicing distinction between the two categories. However, the opposite effect for word-initial and word-medial stops (larger lengthening of voiced vs. voiceless stops in the two positions) shows that this enhancement was not systematic (Crystal and House (1982) report no differences in closure duration as a function of voicing). The effect of position-within-word on stop closure duration shows that stops in the onsets of word-initial stressed syllables, i.e., in prosodically 'strong' positions, were longer than stops in the onsets of word-medial unstressed syllables, i.e., in prosodically 'weak' positions. Furthermore, word-initial closures were lengthened more in clear speech than word-medial closures. Similar 'strengthening' of obstruents in more prominent prosodic positions (stressed vs. unstressed, accented vs. unaccented syllables; word- and phrase-initial vs. word- and phrase-medial positions, etc.) has been demonstrated by Lisker and Abramson (1967), Pierrehumbert and Talkin (1992), Fougerson and Keating (1997), Keating (1984), Cole et al. (2007), Cho and Keating (2001), among others. This asymmetry in lengthening patterns suggests that stressed syllables are lengthened more than unstressed syllables in hyperarticulated speech. A decrease in speaking rate, thus, is not uniformly applied to all segments.

VOT durations for English word-initial and word-medial stops in the two speaking styles are given in Fig. 5. Average VOT durations for voiced and voiceless stops are given in Table A4 in Appendix A.

Repeated-measures ANOVA results for the effect of style (conversational vs. clear), voice (voiced vs. voiceless) and position (word-initial vs. word-medial) on VOT duration showed a significant main effect of

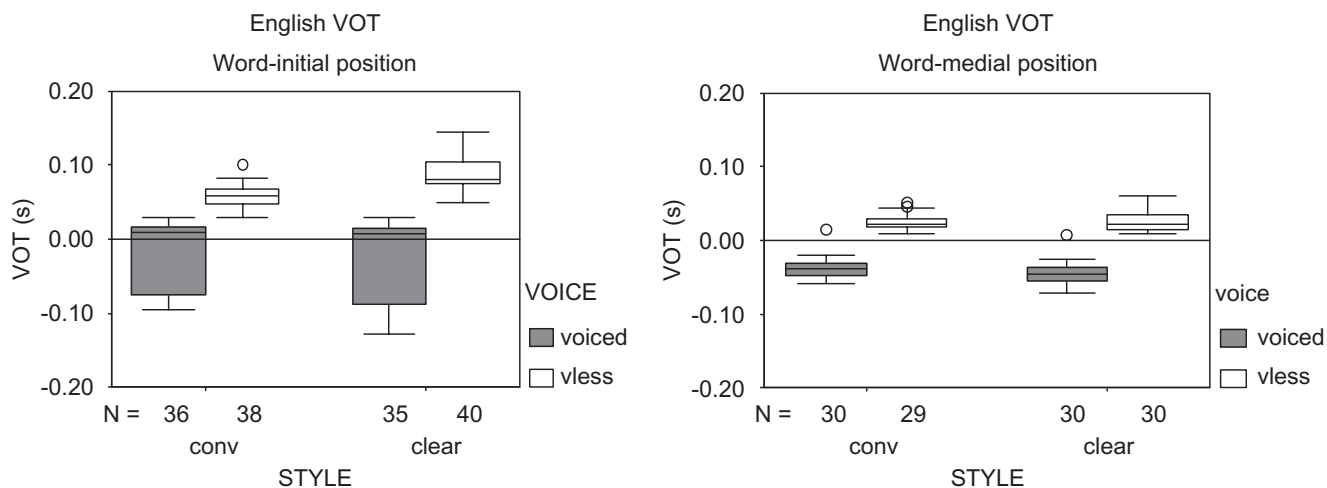


Fig. 5. Box plots of VOT durations for voiced and voiceless stops in word-initial and word-medial positions in two speaking styles. Line at 0 VOT (s) represents stop closure release.

voice ( $F(1,4) = 114.121$ ,  $p < .001$ ) and of position ( $F(1,4) = 12.294$ ,  $p < .05$ ), but not of style ( $F(1,4) = .028$ ,  $p = .874$ ). The two-way style by voice interaction was significant ( $F(1,4) = 34.735$ ,  $p < .05$ ). The style by position and voice by position interactions were not significant: ( $F(1,4) = 3.514$ ,  $p = .134$ ; ( $F(1,4) = 6.012$ ,  $p = .070$ ). Finally, the three-way style by voice by position interaction was not significant ( $F(1,4) = 4.690$ ,  $p = .096$ ). A separate ANOVA for VOT in the word-initial position showed a significant main effect of voice but not of style (voice ( $F(1,4) = 56.337$ ,  $p < .05$ ); style ( $F(1,4) = 1.042$ ,  $p = .365$ ). The style by voice interaction was significant: ( $F(1,4) = 13.158$ ,  $p < .05$ ). ANOVA results for the word-medial position revealed a significant main effect of voice and style on VOT duration: (voice:  $F(1,4) = 148.821$ ,  $p < .001$ ; style ( $F(1,4) = 16.664$ ,  $p < .05$ ). The style by voice interaction was also significant:  $F(1,4) = 9.128$ ,  $p < .05$ ). As expected, the results showed that VOT duration differed significantly for the two voice categories in both positions. Surprisingly, speaking style affected the overall VOT duration significantly only in word-medial position. The significant interactions for the word-initial position revealed that the effect of speaking style differed for the two voicing stop series. i.e., the long lag category was lengthened more than the short lag category which remained rather stable across two speaking styles. In word-medial position, VOT duration of the voiced stops was affected more than of voiceless stops.

The results demonstrated that VOT duration was largely unaffected by changes in speaking style. However, asymmetrical VOT lengthening of just one stop series in clear speech increased the distance between the voiced and voiceless categories. VOT of word-initial voiceless stops was extended in clear speech in the positive direction (longer aspiration) while VOT of voiced stops remained stable across the two speaking styles. In this way, the average distance between the two categories was indeed increased in clear speech in the word-initial position (the difference was 78 ms in conversational speech vs. 120 ms in clear speech, an increase of approximately 42 ms). Similar VOT lengthening of voiceless stops in prosodically 'strong' positions (lexically stressed, phrase accented and domain-initial syllables) in English was reported in Lisker and Abramson (1967), Pierrehumbert and Talkin (1992) and Cole et al. (2007). Furthermore, asymmetrical patterns of VOT manipulation through larger VOT lengthening of voiceless stops when compared to voiced stops in slow when compared to fast speaking rates was reported by e.g., Kessinger and Blumstein (1997), Summerfield (1981) and Miller et al. (1986).

In contrast to the word-initial results, word-medial VOT duration for voiced stop was lengthened while VOT duration of voiceless stops remained stable in clear speech. The distance between the two categories was also increased in this way. However, the increase was rather small (the VOT duration difference was 63 ms in conversational speech vs. 71 ms in clear speech, an increase of approximately 9 ms). Furthermore, the VOT lengthening of voiced stops was likely due to the overall lengthening of stop closures in clear speech (which are voiced throughout) rather than to a different strategy for increasing the VOT contrast when compared with the word-initial position where VOT duration of voiceless stops was increased. Combined, these results

suggest that in terms of absolute VOT measurements, the voicing contrast was increased in hyperarticulated speech in the prosodically ‘strong’ position, i.e., in the word-initial, stressed syllable.

We now turn to an analysis of these VOT measurements in terms of a proportional measure. Expressed as a percentage of stop duration (aspiration/stop duration where stop duration = closure + aspiration), the aspiration portion of word-initial voiceless stops was 47% and 50% in conversational and clear speaking styles respectively averaged across speakers. The aspiration portion of word-medial voiceless stops was 33% and 28% in conversational and clear speaking styles respectively. Repeated-measures ANOVA results for the effect of style (conversational vs. clear) and position (word-initial vs. word-medial) showed a significant main effect of position ( $F(1,4) = 18.504$ ,  $p < .05$ ) but not of style ( $F(1,4) = 1.605$ ,  $p = .274$ ) on this proportional VOT measure for voiceless stops. The two-way style by position interaction was marginally significant ( $F(1,4) = 7.644$ ,  $p = .051$ ). The two-way interaction comes from the fact that in the word-initial position percentage aspiration was slightly increased in clear speech while it was decreased in word-medial stops. Pairwise comparisons revealed that the effect of style, in fact, was not significant in either position: word-initial ( $t(4) = -2.034$ ,  $p = .112$ ); word-medial ( $t(4) = 2.570$ ,  $p = .062$ ). The results showed that even for the word-initial stops, which in absolute terms had longer VOT in clear speech, the amount of aspiration relative to the whole stop duration was stable across the two speaking styles.

As reported above, in absolute numbers, VOT of voiced stops was not affected by the speaking style in word-initial position, i.e., VOT was not significantly longer in clear speech when compared with conversational speech. VOT of voiced stops in word-medial position was affected by the speaking style but it was argued that this was a consequence of the effect of style on the closure duration. Expressed in relative terms, the percentage of stop duration that was voiced (voicing/stop duration) was fairly stable in both positions. In word-initial position, 36% and 37% of the entire stop duration was voiced in conversational and clear speaking styles, respectively. In word-medial position, 94% and 91% of the entire stop duration was voiced in conversational and clear speaking styles, respectively.<sup>8</sup> Repeated-measures ANOVA results for the effect of style (conversational vs. clear) and position (word-initial vs. word-medial) on percentage voicing in voiced stops showed a significant main effect of position ( $F(1,4) = 26.702$ ,  $p < .05$ ) but not of style ( $F(1,4) = .022$ ,  $p = .889$ ). The two-way style by position interaction was not significant ( $F(1,4) = .077$ ,  $p = .795$ ). The results show that, similar to aspiration, the relative amount of voicing was stable across the two speaking styles.

It is important to note that the word-initial voiced stops were either voiced throughout or produced with short lag. The voicing percentages (36% and 37%) mentioned above are, therefore, somewhat misleading. If we consider only prevoiced tokens, the percentage of voicing was 87% and 89% in conversational and clear speaking styles, respectively.<sup>9</sup> Short lag voiced tokens were aspirated 16% and 12% in conversational and clear speaking styles. Even when considered separately, the effect of style was not significant for either of the voiced stop groups: (prevoiced: ( $t(4) = .072$ ,  $p = .947$ ); short lag: ( $t(4) = 2.235$ ,  $p = .089$ )). Overall, then, the results show that the ratios of VOT to stop duration for this sample of Midwest American English voiced and voiceless stops were rather stable across speaking styles.

### 3.3.2. Croatian closure duration and VOT

Croatian stop closure and VOT durations for word-initial stops are shown in Fig. 6. Word-medial stops are not included in the present analyses because a large number of voiced /b, d, g/ stops (38/60) underwent lenition in the onset of unstressed syllables, i.e., they were realized in a less obstruent-like manner. These lenited or ‘weakened’ stops were characterized by the absence of a detectable closure and by the presence of more acoustic energy than is found in voiced stops in the word-initial position. Oftentimes, it was hard to tell the lenited stops from the flanking vowels.

<sup>8</sup>The reason why word-medial stops are not 100% voiced throughout is twofold. First, two tokens were produced with short lag. Second, a few prevoiced tokens which in addition to being voiced throughout the closure duration had a few irregular periods following the release and these periods differ from the vowel periods (possibly due to the aerodynamic factors causing a delay in the complete vowel opening). This portion is considered part of the stop duration rather than a part of the vowel.

<sup>9</sup>Again, the reason why it is not 100%, although the closure is voiced throughout, is that there are some tokens which have a few irregular pulses after the release before the vowel begins. This portion is included in the total stop duration.

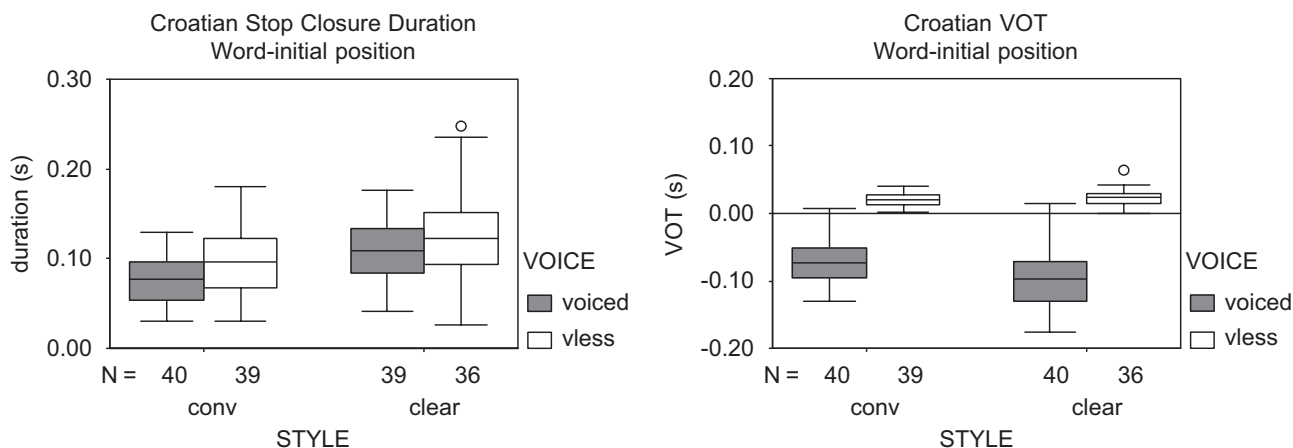


Fig. 6. Box plots of stop closure (left panel) and VOT (right panel) durations for Croatian word-initial stops in two speaking styles. Line at 0 VOT (s) represents stop closure release.

Repeated-measures ANOVA results for the effect of style (conversational vs. clear) and voice (voiced vs. voiceless) on stop closure duration showed a significant main effect of style ( $F(1,4) = 7.471, p < .05$ ) but not of voice ( $F(1,4) = 5.045, p = .057$ ). The two-way style by voice interaction was not significant either ( $F(1,4) = .090, p = .779$ ). Similar to English, Croatian results reveal that the stop closure duration was lengthened in clear speech due to the overall decrease in speaking rate. Furthermore, the effect of style affected both voiced and voiceless stop closure duration in a similar way.

Nearly all Croatian voiced stops were produced with negative VOT in both speaking styles, i.e., they were prevoiced. All prevoiced stops were voiced throughout the closure. Voiceless stops were produced with positive VOT, i.e., with short lag. As seen in Fig. 6, clear speech affected VOT duration of voiced (prevoiced) stops more than of voiceless (short lag) stops in Croatian. Recall that in English, the main effect of style was on word-initial voiceless (long lag) stops. Repeated-measures ANOVA results for the effect of style (conversational vs. clear) and voice (voiced vs. voiceless) on Croatian VOT duration showed a significant main effect of voice ( $F(1,4) = 657.911, p < .001$ ) but not of style ( $F(1,4) = 4.892, p = .091$ ). The two-way style by voice interaction was not significant ( $F(1,4) = 5.607, p = .077$ ). These results showed that the VOT duration remained stable across the two speaking styles despite the tendency of voiced stops to lengthen slightly more when compared with voiceless stops in clear speech.

Relative measures reveal a similar pattern. The ratio of aspiration to total stop duration (i.e., closure + aspiration if VOT is positive) for voiceless stops was 19% and 18% in conversational and clear speaking styles, respectively (these values approximate the ratio of short lag 'voiced' stops in English). Pairwise comparison showed that the effect of style on aspiration ratio was not significant ( $t(4) = .854, p = .441$ ). The ratio of voicing to stop duration (four tokens with positive VOT were excluded from this calculation) was 96% and 94% in conversational and clear speaking styles, respectively. Pairwise comparison showed that the effect of style on voicing ratio was not significant ( $t(4) = 1.023, p = .364$ ). The results showed that lengthening of VOT was not a strategy for enhancing the voicing contrast in hyperarticulated clear speech in Croatian either in absolute numbers or when expressed as ratios relative to total stop duration.

### 3.4. Summary and discussion

The results for absolute VOT durations revealed that the effect of clear speech differed in the two languages. In English, VOT of voiceless stops was lengthened more than VOT of voiced stops. Such asymmetrical VOT lengthening increased the distinction between the two voicing categories in word-initial stops. English clear speech, in this respect, patterns with slow speech and VOT lengthening in prosodically 'strong' positions, such as stressed and accented syllables, domain-initial position, etc. (Cole et al., 2007; Lisker & Abramson, 1967; Miller et al., 1986; Summerfield, 1981, etc.).

The Croatian voicing contrast was not significantly increased in clear speech although VOT of voiced stops was affected more by clear speech (it was extended to include longer prevoiced tokens) than VOT of voiceless stops which remained stable across speaking styles (similar to other cross-language studies of languages which have prevoiced stops in their inventories (Cho & McQueen, 2005; Kessinger & Blumstein, 1997)). However, the fact that VOT lengthening of the Croatian voiced stops did not reach statistical significance suggests that the voicing contrast might not be universally enhanced in hyperarticulated speech (Ohala, 1995, also failed to find contrast enhancement in English clear speech).

Larger VOT lengthening of the voiceless (long lag) category in English and the tendency for the reverse pattern in Croatian suggest language-specific strategies in clear speech implementation as a direct consequence of language-specific patterns of phonetic implementation of phonological voicing categories in conversational speech. Moreover, these results point to a language-universal strategy in that in both languages, the unaffected category is the short lag one regardless of its phonological representation as voiced (as in English) or voiceless (as in Croatian) (see Kessinger and Blumstein (1997) for similar cross-language results and for an articulatory explanation for such cross-linguistic constancy).

In contrast to absolute measures which showed some degree of contrast enhancement, relative VOT measures revealed that the amount of aspiration and voicing relative to total stop duration was stable across the two speaking styles in both languages mirroring the vowel findings. Similar stability of the voicing contrasts across speaking rates when expressed proportionally has been shown by Summerfield (1981), Kessinger and Blumstein (1998) and Boucher (2002). For instance, Boucher (2002), demonstrated that VOT of a stop in the VIC1V2C2 sequence (e.g., *a tap*) relative to C1V2, as well as to V2C2 and C1V2C2, presents a stable value across speaking rates. This suggests that VOT may be processed relative to units as large as a syllable. Studies examining the effect of speaking rate on perception of the voicing contrast suggest that listeners process the affiliation of the voicing exemplars in a rate-dependent way, i.e., in relation to the duration of the surrounding segments (Miller et al., 1986; Miller & Volaitis, 1989; Summerfield, 1981; Volaitis & Miller, 1992). As syllable duration is increased in slower speech, the perceptual boundary between voiced and voiceless stop categories shifts to longer VOT values. Other properties of the acoustic signal show similar rate-dependent processing characteristics. Miller and Liberman (1979) showed that the duration of the initial formant transition that perceptually differentiates a stop consonant from a semivowel (/b/-/w/) is not absolute. Listeners adjusted for changes in speaking rate and treated the duration of the formant transitions in relation to overall syllable duration when making phonetic judgments. In other words, as syllable duration decreased (indicating a decrease in speaking rate) the duration of the transition that distinguished /b/ and /w/ increased. These rate-dependent perceptual adjustments suggest that VOT and formant transitions for differentiating stops from approximants are processed in relation to the surrounding segments. Boucher (2002) further showed that listeners do not make voicing identification based on a shifting category boundary as absolute values would suggest but that their decisions are based on a stable ratio across speaking rates. These stable boundaries specify a perceptual criterion by which listeners make phonetic judgments. Combining all of these results suggests that listeners attend to timing relations in speech that refer to a number of local events including CVs and CVCs and possibly the stop itself. The present comparison across conversational and clear speech is in accord with this general pattern and suggests that language-specific pronunciation norms along the temporal dimension are maintained in clear and conversational speech.

It is important to note that throughout this paper we use the term stability to refer to relational invariance for both vowel and VOT duration. Although we use the absence of a statistical significance for the effect of style on the proportional temporal patterns as an indication of such stability, we are aware that this approach is overly simplistic. The limitations of such approach are confounded by a relatively small number of speakers and by other sources of temporal variation that were not taken into account. To truly establish temporal stability, in future work, we should gather large amount of data under many sources of temporal variation and from more speakers and we should develop more appropriate tests of stability. However, the data reported here are highly suggestive of relational invariance and, furthermore, similar results reported in other studies described above give us confidence that our interpretation is valid.

Although relative VOT results show that the voicing contrast was stable across speaking styles, our results revealed that clear speech may enhance syntagmatic contrasts between vowels and consonants as well as the prosodic structure of a language. Clear speech affected stressed and unstressed syllables differently.

The overall closure duration was longer in both languages in clear speech indicating a decrease in speaking rate, however, the effect of position-within-word on closure duration in English indicates that slowing down was not uniformly applied to all segments. Word-initial stops at the onsets of stressed syllables were lengthened more than word-medial/unstressed stops in clear speech relative to conversational speech. This word-initial/stressed stop strengthening serves to enhance the distinction between consonants and vowels in clear speech. Such syntagmatic strengthening is characterized by longer constriction duration (and strength) of consonants and by larger and longer opening gesture of vowels (Beckman et al., 1992; de Jong, 1995; Pierrehumbert & Talkin, 1992). This varied strengthening of vowels and consonants increases their distinctiveness and facilitates sound segmentation in clear speech. Similarly, in Croatian clear speech, stops were strengthened in word-initial position and were rather distinct from the surrounding stressed vowels while lenited stops in unstressed/word-medial position were more approximant-like and were often difficult to differentiate from the surrounding vowels. These results, along with the vowel duration results and vowel space expansion results (Smiljanic & Bradlow, 2005), suggest that, in clear speech, stressed syllables are strengthened in relation to the unstressed syllable. Although we do not have the results for the unstressed vowels, it is plausible to hypothesize that, in clear speech, unstressed vowels similar to the unstressed stops will not be lengthened and expanded in  $F1 \times F2$  space to the same extent as the stressed vowels. The disproportionate clear speech strengthening of stressed syllables compared to unstressed syllables enhances the prosodic structure and further facilitates (word) segmentation which is also aided by the sentence-final vowel lengthening. Combined, our results suggest that a major feature of clear speech production (and a source of its increased intelligibility) is at the level of prosodic structure.

#### 4. Conclusions

In this paper, we set out to investigate the effect of clear speech on vowel and stop consonants contrasts along temporal dimension in English and Croatian. Overall our results showed that the proportional distances between the 'short and long' vowel categories and between the voiced and voiceless stop categories were rather stable across the two speaking styles in both languages. These results suggest that vowel and VOT duration is not a dimension of category contrast enhancement. Rather, the language-specific pronunciation norms along this dimension are maintained in clear and conversational speech.

Within H&H theory (Lindblom, 1990), clear speech is a listener-oriented intelligibility-enhancing mode of speech production that incorporates all of the essential elements of the hyperarticulation end of the hypo–hyper articulation continuum. Therefore, clear speech is expected to involve phonological contrast enhancement at the acoustic–phonetic level. Indeed, the present findings along with several others (e.g., de Jong & Zawaydeh, 2002; Hay, Sato, Coren, Moran, & Diehl, 2006; Heldner & Strangert, 2001; Uchanski, 1988, 1992, etc.) have demonstrated increased duration contrasts for various phonological temporal contrasts when viewed in absolute terms. The present findings differ from these previous studies in that what seems to be an increase in the durational contrasts in absolute measures presents a rather stable proportional distance between the contrastive categories across the conversational and clear speaking styles. While the speaking style and focus studies mentioned above did not consider proportional measurements, other studies of speaking rate modifications found similar stable boundaries between contrastive categories (Boucher, 2002; Hirata, 2004; Hirata & Whiton, 2005; Kessinger & Blumstein, 1998; Pickett et al., 1999; Pind, 1986, 1999). They have, furthermore, demonstrated that the listeners use these relational boundaries for judging category affiliation (Boucher, 2002). In other words, listeners may rely on local timing relations which can be measured as durational ratios. Proportional stability of durational contrasts may, thus, be a key to intelligibility.

These findings are important for theories of speech production and perception including the H&H theory, suggesting that the notion of 'contrast enhancement' and the notion of 'maintenance of pronunciation norms' need to be considered as parallel principles that guide speech production under conditions that favor hyperarticulation. Specifically, it may be that temporal and spectral dimensions of articulation are independently controlled during the production of intelligibility-enhancing clear speech with the temporal dimension being governed by the principle of maintenance of pronunciation norms and the spectral dimension being governed by the principle of contrast enhancement. At this point, we can only speculate as to the reason that these opposing hyperarticulation principles coexist. One possibility is that, in fact, the dominating

principle is the principle of maintenance of pronunciation norms and that the apparent enhancement of spectrally cued contrasts, such as vowel quality contrasts, resemble durational contrasts in relational stability but the appropriate relational measure has yet to be discovered. An alternative possibility is that the observed ‘split’ between the principles guiding hyperarticulation in the temporal and spectral domains reflects differences in the mechanisms that control articulatory timing vs. articulatory precision. Specifically, while timing is inherently relational, articulatory target approximation, as for example in the case of achieving a particular constriction location (place of articulation), has a more absolute interpretation.

## Acknowledgments

We are grateful to Irena Zovko for help in recruiting Croatian subjects, and to Ken Konopka for assistance in data collection. We thank Robert Daland for help with Praat scripts. We also thank Gerry Docherty and three anonymous reviewers for helpful suggestions concerning the research reported here. This research was supported by Grant No. NIH-R01-DC005794 from NIH-NIDCD and by The Hugh Knowles Center at Northwestern University.

## Appendix A. Results

See Tables A1–A5 for further details.

Table A1

Grand average durations across five speakers (average of three tokens for each English vowel category for each of the five speakers) in seconds and standard deviations (in parentheses) in conversational and clear speaking styles and for the tense and lax vowel groups in both styles

English	ɑ	e	i	o	u	æ	ɛ	ɪ	ʌ	ʊ
Conv.	.113 (.020)	.103 (.015)	.067 (.011)	.100 (.012)	.090 (.012)	.117 (.014)	.070 (.010)	.061 (.008)	.079 (.024)	.060 (.010)
Clear	.127 (.021)	.110 (.016)	.082 (.012)	.110 (.017)	.109 (.013)	.131 (.015)	.070 (.011)	.065 (.009)	.086 (.027)	.067 (.017)
Diff.	.014	.007	.015	.010	.018	.015	.000	.004	.007	.007
English	Tense		Lax		Diff.		Tense/lax ratio			
Conv.	.095 (.010)		.077 (.012)		.017		1.22			
Clear	.108 (.015)		.084 (.014)		.024		1.28			
Diff.	.013		.007		.007		.06			

Table A2

Grand average durations across five speakers (average of three vowels for each of five speakers) in seconds and standard deviations (in parentheses) for English vowels before voiced (vd) and voiceless (vless) codas in two sentence positions in conversational and clear speaking styles. Negative sign indicates shortening of vowels before a voiceless stop in clear speech, sentence-final position when compared with conversational speech

English	Medial				Final			
	vd coda	vless coda	Diff.	vd/vless ratio	vd coda	vless coda	Diff.	vd/vless ratio
Conv.	.141 (.013)	.107 (.018)	.033	1.31	.187 (.021)	.129 (.037)	.058	1.45
Clear	.205 (.015)	.150 (.027)	.055	1.36	.206 (.013)	.124 (.032)	.082	1.66
Diff.	.064	.043	.021	.05	.019	−.005	.024	.21

Table A3

Grand average durations across five speakers (average of three tokens for each Croatian vowel category for each of the five speakers) in seconds and standard deviations (in parentheses) in conversational and clear speaking styles and for the tense and lax vowel groups in both styles

Croatian	Long					Short				
	a	e	i	o	u	a	e	i	o	u
Conv.	.151 (.016)	.128 (.008)	.095 (.009)	.113 (.013)	.106 (.022)	.084 (.018)	.077 (.019)	.060 (.007)	.087 (.015)	.071 (.013)
Clear	.199 (.016)	.175 (.022)	.117 (.040)	.153 (.011)	.156 (.024)	.106 (.011)	.097 (.016)	.073 (.022)	.108 (.022)	.106 (.011)
Diff.	.048	.047	.022	.040	.050	.022	.020	.013	.021	.035
Croatian	Long			Short		Diff.		Long/short ratio		
Conv.	.115 (.009)			.076 (.012)		.039		1.52		
Clear	.160 (.021)			.098 (.013)		.062		1.63		
Diff.	.045			.022		.023		.12		

Table A4

Grand average VOT durations across five speakers (average of six stop tokens for each speaker) in seconds and standard deviations for English voiced and voiceless (vless) stops in word-initial and word-medial positions

English	Word-initial			Word-medial		
	Voiced	vless	Diff.	Voiced	vless	Diff.
Conv.	-.019 (.026)	.059 (.010)	.078	-.037 (.006)	.026 (.004)	.063
Clear	-.032 (.041)	.088 (.017)	.120	-.045 (.009)	.027 (.005)	.071
Diff.	-.013	.029	.042	-.008	.001	.009

Table A5

Grand average VOT durations across five speakers (average of six stop tokens for each speaker) in seconds and standard deviations for Croatian voiced and voiceless (vless) stops in word-initial position

Croatian	Word-initial		
	Voiced	vless	diff.
Conv.	-.073 (.015)	.021 (.004)	.094
Clear	-.098 (.022)	.023 (.007)	.121
Diff.	-.025	.002	.027

## Appendix B. Supplementary materials

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.wocn.2007.02.002](https://doi.org/10.1016/j.wocn.2007.02.002).

## References

- Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, Prosody* (pp. 68–86). Cambridge: Cambridge University Press.

- Boersma, P., & Weenink, D. (2006). *Praat: Doing phonetics by computer (Version 4.4.16) [Computer program]*. Retrieved from <<http://www.praat.org/>>.
- Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception and Psychophysics*, 64(1), 121–130.
- Bradlow, A. R. (2002). Confluent talker- and listener-related forces in clear speech production. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 241–273). Berlin & New York: Mouton de Gruyter.
- Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272–284.
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for learning-impaired children: Sentence perception in noise. *Journal of Speech, Language and Hearing Research*, 46, 80–97.
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22, 129–159.
- Cho, T. (2002). *The effects of prosody on articulation in English*. New York: Routledge.
- Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America*, 117(6), 3867–3878.
- Cho, T., & Jun, S. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistic Society*, 36, 31–44.
- Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190.
- Cho, T., & McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33, 121–157.
- Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech. *Journal of Phonetics*, 35, 180–209.
- Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America*, 72, 705–716.
- Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, 83(4), 1553–1573.
- de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97, 491–504.
- de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32, 493–516.
- de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, 30, 53–75.
- Edwards, J. E., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89, 369–382.
- Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, 116, 2365–2373.
- Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 112, 259–271.
- Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29, 109–135.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3728–3739.
- Hay, J., Sato, M., Coren, A., Moran, C., & Diehl, R. (2006). Enhanced contrast for vowels in utterance focus: A cross-language study. *Journal of the Acoustical Society of America*, 119(5), 3022–3033.
- Heldner, M., & Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics*, 29(3), 329–361.
- Hillenbrand, J., Clark, M., & Nearey, T. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109(2), 748–763.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099–3111.
- Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, 108(6), 3013–3022.
- Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, 32, 565–589.
- Hirata, Y., & Whiton, J. (2005). Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of the Acoustical Society of America*, 118(3, Part 1), 1647–1660.
- Hsu, C.-S., & Jun, S.-A. (1998). Prosodic strengthening in Taiwanese: Syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics*, 96, 69–89.
- Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 69, 505–528.
- Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60, 286–319.
- Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25, 143–168.
- Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26, 117–128.

- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59(5), 1208–1221.
- Krause, J. C., & Braid, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America*, 115(1), 362–378.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lehiste, I., & Ivić, P. (1986). *Word and sentence prosody in Serbocroatian*. Cambridge, MA: MIT Press.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Netherlands: Kluwer Academic.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- Lisker, L., & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1–28.
- Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. G. (2004). Clear speech perception in acoustic and electrical hearing. *Journal of the Acoustical Society of America*, 116, 2374–2383.
- Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43, 106–115.
- Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics*, 25, 457–465.
- Miller, J. L., & Volaitis, L. E. (1989). Effects of speaking rate on the perceived internal structure of phonetic categories. *Perception and Psychophysics*, 46, 505–512.
- Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, 96, 40–55.
- Ohala, J. J. (1995). Clear speech does not exaggerate phonemic contrast. In *Proceedings of the fourth European conference on speech communication and technology, Eurospeech'95*, pp. 1323–1325.
- Payton, K. L., Uchanski, R. M., & Braid, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America*, 95, 1581–1592.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693–703.
- Picheny, M. A., Durlach, N. I., & Braid, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29, 434–446.
- Picheny, M. A., Durlach, N. I., & Braid, L. D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, 32, 600–603.
- Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica*, 56, 135–157.
- Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–127). Cambridge, England: Cambridge University Press.
- Pind, J. (1986). The perception of quantity in Icelandic. *Phonetica*, 43, 116–139.
- Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higher order invariants for two Icelandic speech cues. *Perception and Psychophysics*, 57(3), 291–304.
- Pind, J. (1999). Speech segment durations and quantity in Icelandic. *Journal of the Acoustical Society of America*, 106(2), 1045–1053.
- Skarić, I. (1991). Fonetika Hrvatskog Knjizevnog Jezika. In R. Katicic (Ed.), *Povijesni pregled, glasovi i oblici hrvatskoga knjizevnog jezika* (pp. 61–379). Zagreb: Globus, Nakladni zavod.
- Smiljanic, R. (2004). *Lexical, pragmatic, and positional effects on prosody in two dialects of Croatian and Serbian: An acoustic study*. Outstanding Dissertations in Linguistics. New York: Routledge.
- Smiljanic, R. (2006). Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian. In L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Papers in laboratory phonology 8*. Berlin/New York: Mouton de Gruyter.
- Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America*, 118(3, Part 1), 1677–1688.
- Smith, B. L. (2000). Variation in temporal patterns of speech production among speakers of English. *Journal of the Acoustical Society of America*, 108, 2438–2442.
- Smith, B. L. (2002). Effects of speaking rate on temporal patterns of English. *Phonetica*, 59, 232–244.
- Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5), 1074–1095.
- Uchanski, R. M. (1988). *Spectral and temporal contributions to speech clarity for hearing impaired listeners*. Unpublished Doctoral Dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
- Uchanski, R. M. (1992). *Segment durations in conversational and clear speech*. Unpublished manuscript.
- Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni, & R. Remez (Eds.), *The handbook of speech perception* (pp. 207–235). Malden/Oxford: Blackwell.

- Uchanski, R. M., Choi, S. S., Braid, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research*, 39, 494–509.
- Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, 92(2, Part 1), 723–735.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707–1717.
- Zimmerman, S. A., & Sapon, S. M. (1958). Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America*, 30, 152–153.