Temporal organization of English clear and conversational speech[a)]

Rajka Smiljanić[b)] and Ann R. Bradlow[c)]

Department of Linguistics, Northwestern University, Evanston, Illinois 60208

This study investigated the effect of hyperarticulated, intelligibility-enhancing clear speech on temporal characteristics as reflected in number, durations, and variability of consonant and vowel intervals in sentence- and paragraph-length utterances. The results of sentence-in-noise listening tests showed a consistent clear speech intelligibility gain across the utterances of varying complexity indicating that the talkers successfully maintained clear speech articulatory modifications throughout longer stretches of speech. The acoustic analysis revealed that some temporal restructuring accompanied changes in speaking style. This temporal restructuring was observed in the insertion of consonant and vowel segments that were dropped or coarticulated in conversational speech and in an increase in the number of prosodic phrases for clear speech. Importantly, coefficients of variation (variation of consonantal and vocalic intervals normalized for changes in speaking rate) for both consonantal and vowel intervals remained stable in the two speaking styles. Overall, these results suggest that increased intelligibility of clear speech may be attributed to prosodic structure enhancement (increased phrasing and enhanced segmentability) and stable global temporal properties. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2990712]

I. INTRODUCTION

As speech unfolds in time, more information about its temporal organization is revealed. From the talker's perspective, control over the temporal dimension at the segmental level involves merging multiple influences on speech planning. From the listener's perspective, a crucial feature of speech comprehension involves parsing the temporal dimension at the phonetic level into a hierarchically ordered, complex structure. Thus, from both points of view, the temporal organization of speech crucially involves the integration of short time-scale durations (segments) with longer time-scale temporal modulation patterns (words, phrases, and paragraphs). The goal of the present study was to look at the temporal structure of English as reflected in segment-level durations under conditions that induce global temporal changes, such as variation in speaking style and utterance complexity. Specifically, we investigated the effect of the listener-oriented speaking style called "clear speech" on temporal characteristics of segmental intervals in sentence- and paragraph-length utterances. This study forms part of a broad project whose overarching goal is to understand the signal-dependent and linguistic principles that guide the conversational-to-clear speech transformation in production and that may underlie the enhanced intelligibility of clear speech over conversational speech.

Clear speech is an intelligibility-enhancing speaking style that talkers naturally and spontaneously adopt when listeners have perceptual difficulty due to, for instance, a hearing loss or a different native language. Clear speech is related to other goal-oriented modes of speech production, such as infant-directed speech, Lombard speech, and computer-directed speech, in which talkers "adjust" their output to meet the demands of their target audience or the communicative situation (Junqua, 1993; Skowronski and Harris, 2005; Kuhl *et al.*, 1997; Fernald, 2000; Liu *et al.*, 2004). Clear speech also shares characteristics with variation due to prosodic strengthening and fast-to-slow rate modifications (Cole *et al.*, 2007; de Jong, 1995; Cho, 2005; Hirata, 2004; Miller and co-workers, 1988, 1989). These modifications can all be viewed as involving changes from hypo- to hyperarticulation (H&H Theory, Lindblom, 1990) with respect to articulatory and acoustic features such as reduction of target undershoot and enhancement of phonemic contrasts as seen in, for instance, expanded vowel spaces and increased vowel length and consonant voicing contrasts. Articulatory studies of prominence and of prosodic structure showed that stressed and accented syllables and segments at phrase boundaries were produced with larger, less overlapped gestures and with larger linguapalatal contact (de Jong, 1995; Byrd and Saltzman, 2003; Byrd *et al.*, 2000, 2006; Fougeron and Keating, 1997; Cho, 2006). In a more direct investigation of clear speech articulatory strategies, Perkell *et al.* (2002) found that consonant-volwel-consonant (CVC) syllables in clear speech were produced with an increased articulatory effort (greater peak speed and larger articulatory movement distance) beyond increased segment durations for some talkers (intelligibility was not measured). These studies, then, lend the support for linking acoustic clear speech findings with articulatory hyperarticulation. An important difference between hyperarticulations induced by

---

prosodic structure and prominence and those due to clear speech changes may be that the former have been shown at the local (segment, syllable) level while clear speech may involve similar articulatory adjustments at a more global level, i.e., beyond stressed and accented syllables. Finally, while intelligibility of these hyperarticulated forms, such as stressed syllables and domain-initial segments, has not been investigated, conversational-to-clear speech modifications can provide us with a window into changes along the hypo- to hyperarticulation continuum with the overt purpose of enhancing intelligibility.

Previous work on English clear speech production and perception has established that naturally produced clear speech enhances intelligibility for various listener populations including adults with normal or impaired hearing (Picheny et al., 1986; Payton et al., 1994; Uchanski et al., 1996; Ferguson, 2004; Ferguson and Kewley-Port, 2002; Smiljanic and Bradlow, 2005), elderly adults (Schum, 1996; Helfer, 1998), both native and non-native listeners of the target language (Bradlow and Bent, 2002; Bradlow and Alexander, 2007; Smiljanic and Bradlow, 2007), and children with and without learning impairments (Bradlow et al., 2003). Acoustic comparisons of conversational and clear speech in English have shown that clear speech modifications typically involve enhancement of the overall acoustic salience of the speech signal by means of a decreased speaking rate, longer and more frequent pauses, an expanded pitch range, greater sound pressure levels, more salient stop releases, greater obstruent intensity, increased energy in the 1000–3000 Hz range of long-term spectra, and increased modulation depth of low frequency modulations of the intensity envelope (Picheny et al., 1986, 1989; Bradlow et al., 2003; Liu et al., 2004; Krause and Braida, 2004; Smiljanic and Bradlow, 2005). In addition, the vowel space is expanded in clear speech when compared to conversational speech (Picheny et al., 1986; Moon and Lindblom, 1994; Johnson et al., 1993; Ferguson and Kewley-Port, 2002, 2007). Vowel space expansion 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 despite the difference in vowel inventory size (more than ten in English but just five distinct vowel quality categories in Croatian). Thus, in addition to making the speech signal "stand out" more robustly from the masking effect of noise, clear speech production involves articulatory strategies that enhance vowel contrasts by making them spectrally more distinct from each other.

In the temporal domain, clear speech has been associated with segmental lengthening and an insertion of pauses and short segments, such as a schwa vowel at the ends of words or in voiced obstruent-semivowel clusters (Picheny et al., 1986). Importantly, previous work has demonstrated nonuniform segmental lengthening. For instance, tense vowels were lengthened more than lax vowels in clear speech (Picheny et al., 1986; Uchanski, 1988). Recently, in an in-depth, cross-language exploration of temporal changes for segmental contrasts in Croatian and English clear speech (relative to conversational speech), Smiljanic and Bradlow (2008) found similar asymmetrical patterns of lengthening

for Croatian phonemically long and short vowels, English vowels preceding voiced and voiceless coda stops, and Croatian and English voice onset time (VOT) for voiced versus voiceless stop categories. However, the duration ratio between the "long" and "short" members of the contrastive categories in both English and Croatian was found to be remarkably stable across the two speaking styles suggesting that, unlike the spectral contrasts, duration contrasts are stable rather than enhanced under conditions of hyperarticulation for clear speech.

Similar to clear speech, nonuniform increases in segment durations have been found with changes in speaking rate (Miller et al., 1986, 1988, 1989). A number of studies that examined the effect of speaking rate on temporal features 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.[1] 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 for various languages (Hirata, 2004; Hirata and Whiton, 2005; Kessinger and Blumstein, 1998; Pickett et al., 1999; Boucher, 2002). Some of these studies (e.g., Boucher, 2002) argue for the perceptual invariance as well as production invariance while others (Miller and Volaitis, 1989; Volaitis and Miller, 1992; Nagao and de Jong, 2007) show that the perceptual boundaries shift with changes in speaking rate. Importantly, all of these studies suggest that speakers and listeners rely on local timing relations when producing temporal contrasts and when judging category affiliation in a way that allows them to address changes in speaking rate.

In order to further investigate the nature of the temporal changes of clear speech relative to conversational speech, the present study examines durational characteristics more globally, i.e., we look at the vowel and consonant durations in terms of successive vowel and consonant intervals and their variability across whole utterances ranging from short sentences to a long paragraph (LP). This work builds on investigation of rhythmic characteristics of languages using similar measurements (Ramus et al., 1999; Low et al., 2000; Grabe and Low, 2002). Based on a set of duration measurements from the acoustic speech signal that presumably reflect the phonological properties of syllable structure and unstressed vowel reduction, these studies have made some progress toward linking impressionistically determined, typologically distinct rhythmic classes to durational characteristics of the speech signal. Stress-timed languages, such as English, tend to have a large number of syllable types as well as vowel reduction processes, which acoustically result in a smaller proportion of vocalic intervals within the sentence (%V) and greater variability in the vowel and consonant intervals ($\Delta V, \Delta C$) compared with syllable-timed languages, such as Spanish, which have fewer/simpler syllable types and no vowel reduction yielding higher %V and less variability in the intervocalic (consonantal) intervals.

The purpose of this study was to investigate in depth how the relative number, durations, and variability of successive segmental intervals (approximating closing and opening

gestures of the vocal tract) are affected by different speaking styles with an eye to linking these temporal changes to enhanced intelligibility. Moreover, we examined speech materials of various complexities ranging from isolated sentences to a lengthy and complex paragraph in order to observe temporal variability or stability under various discourse-level conditions. In our analyses, we avoid attributing segmental interval variability to any one feature of linguistic structure, such as overall rhythm class, prosodic structure, syllable complexity, phonological and inherent duration properties of segments, etc.; instead, we view segment durations as reflecting the conflation of multiple levels of linguistic structure into a single temporal dimension.

The specific aim of the present study was twofold. The first objective was to assess the ability of talkers to maintain conversational-to-clear speech articulatory modifications over the course of materials ranging from short sentences to a long and syntactically more complex paragraph (see Sec. II). The second objective of this study was to investigate whether/how conversational-to-clear speech modifications alter the durations of successive consonant and vowel intervals and their variability. In connection with our first goal, we were particularly interested in connecting previous clear speech work, which has focused mostly on intelligibility of isolated vowels, words, and short sentences (Picheny *et al.*, 1986; Gagne *et al.*, 2002; Bradlow and Bent, 2002; Ferguson and Kewley-Port, 2002; Krause and Braida, 2002; Smiljanic and Bradlow, 2005) to real world communicative situations that are more complex than those invoked in a laboratory and may require prolonged hyperarticulation on part of the talker. While the materials used in the present study, as in earlier studies, were read in a laboratory, they represent an attempt to move clear speech research toward more ecologically valid communication settings.

Our main prediction regarding the maintenance of clear speech across longer paragraphs is that, as with words and sentences, clear speech will be more intelligible than conversational speech. However, the overall intelligibility gain for clear speech paragraphs as a whole will not reveal whether enhanced intelligibility is present throughout the duration of the paragraph. With regard to this question, we envision three possible outcomes. Clear speech may require an effort on the part of the talker that is difficult to maintain during longer stretches of speech and as a result the clear speech intelligibility benefit may steadily decline throughout a paragraph. A second possibility is that the intelligibility gain for clear over conversational speech will remain quite constant across the paragraph. This would indicate a clear speech "resetting" that remains constant and adjusts to any changes in inherent intelligibility of the utterance due to the availability of contextual information, discourse and information structure, etc. A third possibility is that intelligibility could steadily increase toward the end of the paragraph indicating the combined beneficial effect of context and speaking style. In order to address the question of whether conversational-to-clear speech modifications and increased intelligibility can be maintained throughout the duration of a paragraph, we look at intelligibility scores for various portions of the paragraphs separately.

The second objective of this study was to investigate of whether/how durations of successive consonant and vowel intervals and their variability are affected by changes in speaking style). Previous research has established that slowing down in clear speech is realized as both the lengthening of segments and the insertion of more frequent pauses for monosyllabic words and short sentences (Picheny *et al.*, 1986; Krause and Braida, 2004; Ferguson and Kewley-Port, 2002; Bradlow *et al.*, 2003, Smiljanic and Bradlow, 2005, 2008). Furthermore, insertion of short segments was shown to characterize clear speech as well (Picheny *et al.*, 1986). We extend this previous work by providing a more global, typologically motivated set of measures including %C, %V, variability of C and V intervals, and coefficients of variability as well as phrasing for these intervals across utterances of various lengths and complexities. Based on the previous findings we expect that the durations of all segments will increase in clear speech. We also expect that the extent of variability of the short and long C and V intervals will change in clear speech compared to conversational speech. This could occur if, for instance, stressed syllables were lengthened more than unstressed syllables, thereby increasing the difference between short and long intervals. The increase in the interval variability would also be achieved through asymmetrical lengthening of various segmental duration contrasts as discussed above. Furthermore, a decrease in speaking rate and a reduction of coarticulation in clear speech could also result in the insertion of short C and/or V elements that were dropped or coarticulated with surrounding sounds in conversational speech (similar to the schwa insertion reported in Picheny *et al.*, 1986). These asymmetrical lengthening patterns along with the segmental insertion could lead to the improved intelligibility of clear speech through a more transparent reflection of prosodic structure and of underlying phonological structure (e.g., segment identification, word/syllable boundaries, and word-level stress patterns). Finally, it is possible that, although C and V variability increases in absolute numbers, proportional measures of this variability (normalizing for the changes in overall speaking rates) remain stable across two speaking styles (similar to the findings for the duration ratios for segmental contrasts found in Smiljanic and Bradlow, 2008).

This analysis presents a part of a larger effort to determine the consistent salient acoustic-phonetic correlates of increased speech intelligibility. However, it is important to note that a direct relationship between acoustic-phonetic variation due to changes in speaking style and variability in intelligibility has been difficult to establish (Uchanski *et al.*, 1996; Krause and Braida, 2002). Although numerous acoustic-phonetic features of conversational-to-clear style transformations have been identified, as noted above, it is not well understood yet if and how each of these modifications (including segmental lengthening and the overall temporal structure) affect intelligibility. Nevertheless, we hope to gain a better understanding of the acoustic-phonetic changes that talkers reliably and consistently produce in clear speech and that may underlie increased intelligibility. To this end, we provide a detailed analysis of temporal structure under various styles and discourse-level conditions.

II. METHODS

A. Participants

1. Production

Six native talkers of English (three female and three male) between the ages of 24 and 32 served as participants in the production study. They were graduate students in the Linguistics Department at Northwestern University and were all native speakers of general American English (GAE). 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. All participants were paid at the end of the recording session.

2. Perception

One hundred and twenty undergraduate students at Northwestern University, all native speakers of GAE, participated in sentence-in-noise perception tests. Half of the participants listened to the short sentences and half listened to the short and long paragraphs (SP and LP) (see Sec. II B). They received class credit for their participation in the listening test. Their ages ranged between 18 and 22 years. None of the listeners had any known speech or hearing impairment at the time of the test.

B. Stimuli

The stimuli consisted of three sets of materials: short sentences, one SP, and one LP. The 20 short sentences were taken from Ramus *et al.* (1999); they are short newslike declarative statements (e.g., *The next local elections will take place during the winter*). The mean number of syllables per sentence is 16 (range: 15–18). A SP, *The North Wind and the Sun*, is a standard text from phonetic research available in the Handbook of the International Phonetic Association (1999). Finally, a LP from contemporary literature (Sedaris, 2001) was included as well. The Sedaris paragraph is longer than the SP paragraph (212 versus 113 words) and has a more complex syntactic structure, including direct and reported speech, and, therefore, provides more opportunity for subjects to vary their prosody approximating spontaneous speech more closely than in the sentences or SP paragraph.

C. Procedure

1. Production

The subjects were recorded producing all three sets of materials in a sound-attenuated booth in the phonetics laboratory in the Department of Linguistics at Northwestern University. The participants first read the short sentences, which were written on index cards and randomized for each reading. Next, they read the SP followed by the LP. They read into a microphone and the speech was recorded directly to a disk at 24 bit accuracy using an Apogee PSX-100 analog-to-digital/digital-to-analog converter at a sampling rate of 16 kHz. Participants read all the materials first in conversational and then in clear speech. For the conversational style, the talkers were instructed to read as if they were talking to someone familiar with their voice and speech patterns. For the clear speaking style, the talkers were instructed to read as if they were talking to a listener with a hearing loss or a non-native speaker. The acoustic analyses of the recorded materials were done using PRAAT software for speech analysis (Boersma and Weenink, 2006).

2. Perception

After the recordings were made, the digital speech files were segmented into sentence-length files. For the two paragraphs, breaks were made at points of natural junctures (pauses, ends of phrases and sentences) and at points at which segmentation was possible. The North Wind and the Sun was segmented into 13 sentence-length files and the Sedaris (2001) paragraph was segmented into 31 files for each speaker. In order to obtain equivalent overall amplitude levels, all speech files were equated for rms amplitude[2] and then mixed with speech shaped noise at a −5 dB signal-to-noise ratio (SNR). The SNR level used in this experiment was chosen based on the results from previous studies with the goal of keeping the average intelligibility in the 45%–65% range (Smiljanic and Bradlow, 2005; 2008).

Each participant in the perception experiment heard either short sentences or the two paragraphs (SP and LP). In the short sentence condition, 60 listeners each heard 20 sentences (ten conversational and ten clear) produced by only one of the talkers (ten listeners per talker). Clear speech sentences always preceded conversational speech sentences so that any clear speech benefit obtained could not be explained by the subject's adaptation to the task or to the talker's speech patterns. For each talker, the ten sentences that were presented in the conversational style in one condition (in which half of the listeners participated) were presented in the clear speaking style in another condition (in which the other half of the listeners participated). This was done to guard against possible confounding effects of any particular sentence-style combination. The listeners never heard the same sentence twice.

Sixty new listeners participated in the paragraph condition. The clear speech paragraph always preceded the conversational style paragraph. Half of the listeners in the paragraph condition first heard the SP in clear speech followed by the LP in conversational speech. The other half of the listeners first heard the LP in clear speech followed by the SP in the conversational speaking style. Each listener heard both paragraphs produced by the same talker. In all conditions, each sentence/utterance in the listening test was preceded by a 400 ms leading silence and a 500 ms noise interval, and followed by a 500 ms noise interval.

The listeners were seated in front of a computer in a sound-attenuated booth in the phonetics laboratory in the Department of Linguistics at Northwestern University. Stimulus presentation was controlled by SUPERLAB PRO 2.01, a special-purpose experiment running software. Three practice sentences (from a different set of recordings by a different talker) were presented in both the short sentence and the paragraph conditions so that the subjects could get used to the nature of the stimuli mixed with noise and the procedure of advancing to the next trial. After each trial, the subject pressed the space bar on the keyboard to initiate the next trial. The listeners were instructed to write down every word

they heard. Each trial was presented only once but the duration of the pause between two trials was controlled by the subjects themselves.

D. Data analysis

1. Production

All the acoustic measurements were performed on the exact same sentences that were used in the sentence-in-noise perception tests. First, we marked consonant and vowel intervals for each sentence. Silences longer than 5 ms in duration excluding silent intervals preceding word-initial stop consonants were marked as pauses (following Smiljanic and Bradlow, 2005). All obstruents and sonorants that could be segmented out from the surrounding sounds were marked as belonging to consonantal intervals. If clear formant changes and amplitude differences were observed for glide-vowel sequences, for instance, glides were excluded from the vocalic portions. If no observable boundary could be found, the glides were included in the vocalic portion. All vowels were marked as vocalic intervals. In general, standard segmentation criteria were followed (Peterson and Lehiste, 1960; Lehiste, 1970). A contiguous sequence of one or more vowels (consonants) was marked as one vocalic (consonantal) interval even when it spanned syllable and word boundaries. If it was not possible to segment a consonant from the surrounding vocalic segments due to extensive coarticulation and lack of clear acoustic cues marking its beginning and end it was marked as a part of a vocalic interval. In this respect sonorants were more likely to be marked as vocalic although some obstruents were also highly coarticulated/weakened or completely dropped in productions. Such "segmentability" difficulties occurred more often in conversational speech than in clear speech (see Sec. III). Our segmentation process was, thus, based on acoustic rather than phonological criteria and in this respect resembles closely those of Ramus *et al.* (1999) and Grabe and Low (2002).

Once consonant and vowel intervals were labeled, the total number of consonant and vowel intervals and their durations were measured for each speaking style. For all measurements, subsequent calculations were done within prosodic phrases (PPs) and then averaged across PPs for different materials, speaking styles, and individual talkers. PPs were defined as stretches of speech delimited by pauses. We consider PPs to reflect phrasing implemented by talkers to emphasize the discourse structure of produced speech. In short sentence productions, PPs were mostly identical to the sentences themselves. In longer paragraphs, PPs sometimes corresponded with syntactic sentences and sometimes with intonation phrases (IPs) as defined by ToBI, a convention for transcribing the intonation and prosodic structure of spoken utterances (Silverman *et al.*, 1992). Some PPs encompassed several IPs.

The overall speech rate (vocalic intervals per second) was calculated by dividing the number of vocalic intervals in a PP by the total duration of that PP. Because we wanted to examine segmental lengthening independent of pauses in clear speech, pause duration was excluded from this calculation. Speaking rate was determined by the number of vocalic intervals measured rather than by the phonological number of syllables for each sentence. Although this measure may underestimate the number of phonologically defined syllables produced (due to the conflation of VV sequences into one V interval), it captures the variability in the articulation of syllables/segments (rate of syllable omission and the extent of coarticulation) as produced across individuals and speaking styles. The overall speaking rate for each speaking style and type of material was then determined by averaging the vocalic interval per second rate over all PPs produced by each talker.

Proportion of vowel and consonant intervals (%V,%C) and their variability (ΔV, ΔC) were also calculated %C and %V were calculated by dividing the total duration of C and V intervals, respectively, within a PP with the total duration of that PP. ΔV and ΔC are standard deviations of all vocalic and all consonant intervals within a PP. Finally, in order to take into account the overall slowing down of speech rate for clear versus conversational speech, we calculated coefficients of variation for vowel and consonant intervals as the standard deviation of the consonant or vowel intervals divided by the average duration of the consonant or vowel intervals for each PP. To help us compare variability in overall consonant and vowel interval duration for clear versus conversational speech without regard for the source of that variability (e.g., stress, vowel reduction, final lengthening, or inherent vowel duration), all vowel and consonant intervals within a PP were included in these calculations.

2. Perception

Each participant in the short sentence-in-noise perception test received a keyword correct score out of 107 for the 20 sentences they heard. Each participant in the paragraph-in-noise perception test received a keyword correct score out of 50 for the 13 sentences they heard in the SP and out of 109 for the 31 sentences they heard in the LP. For all materials, all content words counted as keywords. A keyword was counted as correct only if all morphemes of the target word were present and transcribed correctly, e.g., if the target word was "keeping," "keep," "keeps," or "kept" were scored as incorrect. Percentage correct scores were calculated and then converted to rationalized arcsine transform units (RAU) for statistical analyses (Studebaker, 1985). The transformed scores were then coded as RAU scores for −5 dB SNR conversational style and for −5 dB SNR clear style for each talker in the sentence and paragraph conditions.

III. RESULTS

A. Perception

The average perception scores (percentage correct keywords) and the average intelligibility gain (clear-conversational scores) for the short sentences, SP, and LP for each talker are given in Table I. The talkers are ordered by the amount of conversational-to-clear speech intelligibility gain in the short sentence condition from smallest to largest. Letters F or M designate a female or a male talker. The numbers refer to the recording order.

TABLE I. Average perception scores (percentage correct keywords) in conversational and clear speech and the average intelligibility gain (clear-conversational scores) for the short sentences, short paragraph (SP), and long paragraph (LP) for each talker.

| | Intelligibility (%) | | | | | | | | |
| | Short sentences | | | SP | | | LP | | |
Talker	Conv.	Clear	Clear-conv.	Conv.	Clear	Clear-conv.	Conv.	Clear	Clear-conv.
F6	47	51	4	34	35	2	32	44	12
F2	37	42	5	18	29	10	37	41	4
M1	38	53	15	26	25	−2	30	52	22
M5	40	62	22	44	84	40	45	71	26
M4	55	81	26	34	57	22	45	62	17
F3	39	74	35	21	60	39	39	53	14
Average	43	60	18	30	49	19	38	54	16

TABLE II. Average intelligibility scores (percentage keyword correct) for conversational and clear speech in four portions of the short paragraph (SP) and long paragraph (LP) for each talker. Each score is averaged across listeners ($n=60$).

Paragraph	Style	Talker	25%	50%	75%	100%
SP	Conv.	F6	17	37	35	45
		F2	7	32	13	20
		M1	13	34	17	42
		M5	27	55	25	66
		M4	25	40	27	46
		F3	2	20	12	49
		Average	15	36	21	45
SP	Clear	F6	20	40	38	42
		F2	7	35	38	34
		M1	3	40	23	32
		M5	75	95	77	88
		M4	47	69	48	62
		F3	42	74	62	62
		Average	32	59	48	53
LP	Conv.	F6	29	33	28	39
		F2	41	37	40	35
		M1	19	33	27	40
		M5	45	59	37	41
		M4	40	53	43	43
		F3	36	45	41	37
		Average	35	43	36	39
LP	Clear	F6	43	53	40	42
		F2	26	44	50	45
		M1	47	59	55	49
		M5	52	84	70	75
		M4	58	70	53	68
		F3	50	57	45	61
		Average	46	61	52	57

A repeated-measures analysis of variance (ANOVA) with material type (short sentence versus SP versus LP) and style (conversational versus clear) as within-subjects factors was performed on the RAU transformed percent correct scores (averaged for each listener in each style and material condition). There was a significant main effect of style on intelligibility score: $F(1,5)=15.745$, $p<0.05$. The effect of material type did not reach significance: $F(1,5)=5.034$, $p=0.075$. The two-way style by material type interaction was not significant either: $F(1,5)=0.085$, $p=0.783$. The results of the statistical analysis showed that intelligibility was not significantly different for short sentences and SP and LP, and, furthermore, that the clear speech intelligibility increase was equivalent for the three types of materials. Most talkers' conversational-to-clear speech modifications resulted in a moderate to large increase in intelligibility gain (range: 5%–40%). Intelligibility increase of less than 5% in clear speech was found for F6 for short sentences and SP, for F2 for LP and for M1 for SP.

In order to address the question of whether conversational-to-clear speech modifications and increased intelligibility are maintained throughout the duration of a paragraph, we divided each paragraph into four portions (0%–25%, 25%–50%, 50%–75% and 75%–100%) based on the total number of target keywords for each type of material. The average intelligibility scores (percentage keyword correct) for the two speaking styles in the four portions of each paragraph are given for each talker separately in Table II.

For most talkers in both paragraphs, clear speech was more intelligible than conversational speech in all four portions suggesting that talkers successfully maintained clear speech articulatory modifications throughout the duration of the paragraphs. For both paragraphs, listeners tended to have most difficulty in correctly identifying words in the initial portion (0%–25%). There was an increase in intelligibility after the initial portion (25%–50%), a slight decrease in the third quarter of the paragraph (50%–75%), and another increase during the last portion (75%–100%).

A three-way repeated-measures ANOVA with style (conversational versus clear), paragraph (SP versus LP), and portion (1, 2, 3, and 4) as within-subjects factors showed a significant main effect of style: $F(1,5)=13.616$, $p<0.05$. The portion factor was also significant: $F(3,15)=43.935$, $p<0.001$. However, the effect of the paragraph was not significant: $F(1,5)=4.939$, $p=0.077$. Only the paragraph by portion two-way interaction was significant: $F(3,15)=15.308$, $p<0.001$. The three-way interaction was not significant. Separate repeated-measures ANOVAs for the two paragraphs showed a main effect of portion [$F(3,15)=32.568$, $p<0.01$] but not of style [$F(1,5)=5.694$, $p=0.063$] only for SP. The two-way style by portion interaction was also significant: $F(3,15)=3.316$, $p=0.049$. Paired comparisons revealed that the effect of style was significant only for portions 2 and 3: $t(5)=2.492$, $p=0.05$; $t(5)=3.187$, $p<0.05$. Intelligibility in portions 1 and 4 was not significantly different for the two speaking styles: $t(5)=1.566$, $p=0.178$; $t(5)=1.77$, $p=0.137$. In contrast, both style and portion factors were significant for LP: [style: $F(1,5)=22.616$, $p<0.01$; portion: $F(3,15)=10.731$, $p<0.01$]. The two-way interaction was not significant: $F(3,15)=0.578$, $p=0.639$. For both paragraphs, then, overall intelligibility varied across the four portions. The clear speech effect was rather consistent throughout for the LP while it was not present in all portions for the SP.

Adding to the large body of work on factors that condition speech intelligibility, the results of the present study

pointed to a significant contribution of utterance-specific characteristics and speaking style to overall intelligibility. Portion-by-portion results revealed that intelligibility varied across paragraph-length utterances in both speaking styles (higher intelligibility in second and fourth portion compared to first and third portions) suggesting that context, choice of lexical items, introduction of a new topic, etc., all contribute to intelligibility significantly. Furthermore, different intelligibility results for the two paragraphs indicated that the nature of the text read may also affect speaking effort and accuracy, which in turn influence listening accuracy (SP is a familiar short fable while LP is a new unfamiliar conversationlike text, which may have caused a more animated/ hyperarticulated reading style). Additionally, LP could be easier perceptually than SP due to its lexical context and the nature of the material. Importantly, the intelligibility results illustrated that some of the perceptual difficulties that listeners face can be overcome by naturally produced clear speech in both sentence- and paragraph-length utterances. It is worth noting, though, that clear speech did not wipe out the detrimental effects of other factors, such as lack of contextual cues, lexical choices, or articulatory precision that varied within the course of the paragraphs. However, these factors will have to be addressed in future work, because the present materials were not designed to investigate them in a controlled manner. With regard to our initial question of whether clear speech can be maintained in paragraph-length utterances, these results demonstrated that clear speech can be maintained throughout longer stretches of speech. This expands on previous clear speech findings, which focused mostly on intelligibility of isolated vowels/syllables, words, and short sentences (Picheny *et al.*, 1986; Gagne *et al.*, 2002; Bradlow and Bent 2002; Ferguson and Kewley-Port, 2002; Krause and Braida, 2002; Smiljanic and Bradlow, 2005).

B. Production

The average results of the various duration measurements for each material in each speaking style for each talker are given in Table III. The results of the repeated-measures ANOVAs with style (conversational versus clear) and material type (sentence versus SP versus LP) as within-subjects factors on various dependent duration measures given in Table III.

C lengthening in clear speech was statistically significant: $[F(1,5)=26.118, p<0.01]$. The effect of the material type on C lengthening was significant as well: $[F(2,10)=19.091, p<0.001]$. The style by material type interaction was not significant: $[F(2,10)=1.6, p=0.250]$. V lengthening in clear speech was statistically significant as well: $[F(1,5)=13.412, p<0.05]$. The effect of the material type on V lengthening was not significant: $[F(2,10)=2.241, p=0.157]$. The style by material type interaction was not significant: $[F(2,10)=0.593, p=0.157]$. These results showed that talkers lengthened both C and V intervals significantly in clear speech. The effect of material type on C lengthening reflects a difference in the segmental makeup of the test materials.

Short sentences were typically produced as one phrase and were not conducive for pause insertion in either speaking style due to their length and relative syntactic simplicity. In contrast, syntactically more complex paragraphs showed an increase in the number of pauses in clear speech. More pauses, furthermore, indicated increased phrasing on average by three and eight phrases in SP and LP, respectively.

Speaking style had a significant effect on speaking rate: $[F(1,5)=43.5, p<0.001]$. The effect of the material type on speaking rate was significant as well: $[F(2,10)=7.263, p<0.05]$. The style by material type interaction was not significant: $[F(2,10)=0.035, p=0.121]$. The results showed that speaking rate was significantly decreased for all material types in clear speech. Short sentences spoken in the conversational speaking style had the fastest speaking rate while SP spoken in clear speech had the slowest speaking rate. The largest clear speech speaking rate decrease occurred for short sentences. These results suggest that the complexity and length of the planned speech affect the speaking rate. Furthermore, the amount of a speaking rate decrease in clear speech may be dependent on the baseline conversational speaking rate.

Next, we examined whether the proportional relation between C and V intervals changed in clear speech for short sentences and the two paragraphs. This measurement allowed us to examine whether segmental lengthening associated with a clear speech decrease in speaking rate affected C and V intervals equally or whether consonants and vowels were lengthened asymmetrically in clear speech. Since %C and %V are directly related (as one increases the other decreases) we discuss %C only. The results showed that consonantal intervals take up slightly more than half of the PP duration in both speaking styles. A repeated-measures ANOVA revealed that only material type had a significant effect on %C: $F(2,10)=16.076, p<0.05$. The effect of style was not significant: $[F(1,5)=0.486, p=0.157]$. The style by material interaction was not significant either: $[F(2,10)=0.020, p=0.981]$. These results showed that the average %C and %V differed for the three types of materials suggesting that syllable structure varied for short sentences, SP and LP with %C being highest for the SP. This is not surprising given that SP was constructed so that it contains all sounds of English as well as a wide variety of syllable types. The overall relationship between %C and %V, however, remained stable across two speaking styles for all material types, i.e., clear speech lengthening overall affected C and V intervals equally (%C is decreased in clear speech by only 0.42, 0.30, and 0.29 percentage points for short sentences, SP, and LP, respectively).

Even though the proportion of C and V intervals remained stable in the two speaking styles, the extent of variability of the C and V intervals may change in clear speech compared to conversational speech. The average variability of C and V intervals (standard deviations) is given in Table III. A repeated-measures ANOVA showed a significant main effect of only style on C variability: $F(1,5)=15.057, p<0.05$. The effect of material type was not significant: $[F(2,10)=0.797, p=0.477]$. The style by material interaction was not significant either: $[F(2,10)=2.647, p=0.0119]$. There was also a significant main effect of style:

TABLE III. Various C and V interval duration measurements in conversational and clear speech for short sentences and short paragraph (SP) and long paragraph (LP) for each talker: total number of C and V intervals and pauses, average speaking rate (vocalic interval/s), average C and V interval duration (s), average proportion of C and V intervals, average variability of C and V interval duration (standard deviations), and average coefficient of variance of C and V intervals. The averages were first calculated within PPs and then across PPs for each material type and individual talkers.

Material	Style	Talker	Total No. of C	Total No. of V	Total No. of pause	Avg. spkg. rate (voc. inters./s)	Avg. C dur. (s)	Avg. V dur. (s)	%C	%V	stdev C (s)	stdev V (s)	Coeff. of variance C	Coeff. of variance V
Short sent.	Conv.	F6	285	276	1	5.46	0.10	0.09	57.78	42.22	0.05	0.04	0.56	0.55
		F2	297	289	1	5.45	0.11	0.08	58.22	41.78	0.06	0.05	0.53	0.57
		M1	268	253	1	5.61	0.10	0.07	58.29	41.71	0.06	0.04	0.53	0.60
		M5	299	285	8	5.35	0.10	0.09	53.24	46.76	0.07	0.05	0.58	0.56
		M4	273	268	0	4.77	0.12	0.09	57.58	42.42	0.06	0.05	0.58	0.59
		F3	301	292	0	5.36	0.10	0.09	52.64	47.36	0.06	0.06	0.60	0.61
		Avg.	287.1	276.6	1.83	5.33	0.10	0.09	56.29	43.71	0.06	0.05	0.56	0.58
	Clear	F6	309	297	0	4.89	0.11	0.08	54.54	45.46	0.06	0.05	0.56	0.57
		F2	308	297	0	4.53	0.13	0.09	58.77	41.23	0.07	0.05	0.57	0.58
		M1	299	282	6	5.12	0.11	0.08	57.21	42.79	0.06	0.05	0.51	0.61
		M5	334	303	0	4.84	0.11	0.10	53.17	46.83	0.06	0.05	0.57	0.53
		M4	302	293	0	3.45	0.15	0.12	58.32	41.68	0.10	0.07	0.62	0.54
		F3	306	295	0	4.13	0.13	0.11	53.19	46.81	0.08	0.07	0.60	0.59
		Avg.	309.6	294.5	1.00	4.49	0.12	0.10	55.87	44.13	0.07	0.06	0.57	0.57
SP	Conv.	F6	136	131	7	5.15	0.11	0.08	58.06	41.94	0.06	0.04	0.56	0.55
		F2	134	131	7	4.75	0.12	0.08	58.96	41.04	0.07	0.05	0.54	0.53
		M1	135	130	7	5.47	0.10	0.07	59.76	40.24	0.05	0.04	0.51	0.58
		M5	135	129	8	5.29	0.11	0.08	58.74	41.26	0.05	0.04	0.51	0.50
		M4	133	124	13	4.04	0.14	0.10	57.06	42.94	0.07	0.06	0.51	0.58
		F3	135	129	7	5.09	0.11	0.08	56.98	43.02	0.06	0.04	0.55	0.53
		Avg.	134.6	129.0	8.17	4.97	0.11	0.08	58.26	41.74	0.06	0.05	0.53	0.55
	Clear	F6	139	134	9	4.98	0.11	0.09	56.55	43.45	0.06	0.05	0.57	0.53
		F2	139	133	12	3.92	0.15	0.10	59.94	40.06	0.08	0.06	0.53	0.57
		M1	134	127	10	4.90	0.12	0.08	58.71	41.29	0.06	0.05	0.52	0.59
		M5	139	135	9	4.20	0.13	0.10	57.60	42.40	0.07	0.05	0.55	0.51
		M4	146	137	14	3.65	0.15	0.11	57.69	42.31	0.09	0.05	0.58	0.45
		F3	138	132	12	3.85	0.14	0.11	57.31	42.69	0.08	0.06	0.59	0.54
		Avg.	139.1	133.0	11.00	4.25	0.13	0.10	57.96	42.04	0.07	0.05	0.56	0.53
LP	Conv.	F6	315	298	28	5.05	0.10	0.09	55.07	44.93	0.06	0.05	0.55	0.58
		F2	315	295	27	4.81	0.11	0.09	58.17	41.83	0.06	0.06	0.55	0.63
		M1	279	266	21	5.22	0.10	0.08	57.78	42.22	0.06	0.05	0.54	0.61
		M5	306	283	30	5.06	0.10	0.08	58.22	41.78	0.07	0.04	0.60	0.51
		M4	324	289	36	4.46	0.11	0.09	57.34	42.66	0.07	0.06	0.63	0.56
		F3	297	285	16	5.14	0.10	0.09	51.11	48.89	0.05	0.06	0.57	0.61
		Avg.	306.0	286.0	26.33	4.96	0.10	0.09	56.28	43.72	0.06	0.05	0.57	0.58
	Clear	F6	324	301	32	4.74	0.10	0.09	53.74	46.26	0.06	0.05	0.57	0.53
		F2	320	302	33	4.28	0.13	0.10	58.94	41.06	0.07	0.06	0.55	0.60
		M1	313	295	26	4.90	0.11	0.08	59.42	40.58	0.06	0.05	0.54	0.60
		M5	321	302	37	4.59	0.11	0.10	55.33	44.67	0.07	0.05	0.56	0.51
		M4	343	302	49	3.86	0.13	0.12	55.28	44.72	0.07	0.07	0.57	0.54
		F3	313	297	28	4.40	0.11	0.11	53.25	46.75	0.06	0.07	0.54	0.61
		Avg.	322.3	299.8	34.17	4.46	0.12	0.10	55.99	44.01	0.07	0.06	0.55	0.56

$F(1,5)=25.250$, $p<0.01$ and of material type: $F(2,10) =6.205$, $p<0.05$ on V variability. The style by material interaction was not significant: $[F(2,10)=0.736, \ p=0.503]$. The results revealed a significant increase in variability of C and V intervals for all three types of materials in clear speech (C variability increases in clear speech were 0.012, 0.013, and 0.005 s and V variability increases in clear speech were 0.009, 0.006, and 0.004 s for short sentences, SP, and LP, respectively). The overall V variability differed significantly

for the three types of materials (lowest for SP and highest for LP) due to the varied syllabic complexity across short sentences and the two paragraphs. This was already observed in the difference in %C and %V for the three different materials.

Next, we examine the number of V and C intervals in conversational and clear speech as well as the coefficients of C and V interval variation (normalizing for changes in speaking rate).

ANOVA results for the number of C intervals showed a main effect of material type $[F(2,10)=600.342, p<0.001]$ and of style $[F(1,5)=31.66, p<0.01]$. The two-way style by material type interaction was also significant: $F(2,10)=6.885, p<0.05$. Similar results were obtained for the number of V intervals: style $F(1,5)=26.295, p<0.01$; material type $F(2,10)=1777.389, p<0.001$; style by material type interaction $F(2,10)=152.028, p<0.05$. As expected, these results showed that the total number of C and V intervals varied for the short sentences, SP, and LP due to the difference in their length and complexity. Importantly, the results revealed a significant increase in the number of C: $[F(1,5)=31.66, p<0.01]$ and V: $[F(1,5)=26.295, p<0.01]$ intervals in clear speech compared with conversational speech. The number of C intervals was increased by 22, 4.5, and 16.33 and the number of V intervals was increased by 17.83, 4, and 13.83 for short sentences, SP, and LP, respectively.

ANOVA results revealed a significant effect of material type on the number of C intervals: $[F(2,10)=600.342, p<0.001]$ and V intervals: $[F(2,10)=1777.389, p<0.001]$. The style by material interaction was also significant for the number of C intervals: $[F(2,10)=6.885, p<0.05]$ and V intervals: $[F(2,10)=152.028, p<0.05]$. The significant two-way style by type of material interaction on the number of C and V intervals revealed that clear speech C and V increase differed for short sentences, SP, and LP. The smallest increase in the number of intervals was found for the SP. This paragraph was overall the shortest and provided the fewest opportunities for dropping of segments/syllables in conversational speech. The largest clear speech increase for both C and V intervals was found in short sentences. The fastest speaking rate (which occurred in the production of short sentences compared to SP and LP) induced the most dropping and coarticulation of segments, while the largest speaking rate decrease in clear speech (which also occurred for the short sentences) increased the number of inserted segments/syllables. Note that most previous work examined segment durations (speaking rate) and insertions in short sentences (Picheny *et al.*, 1986, Krause and Braida, 2002; Smiljanic and Bradlow, 2008). The present results indicate that some of the conversational-to-clear speech changes reported for short sentences in previous work may overestimate the presence of these features in longer utterances. It is, therefore, crucial to compare the production of short sentences with longer and more complex materials and their interaction with changes in speaking rate in an attempt to better understand speech communication in real-life situations. The results reported here also suggest that talkers may implement varied clear speech changes for different materials and different communicative demands.

In order to account for the differences in speaking rates (and the difference in the average C and V durations) between the two speaking styles, we examined the coefficient of variation for the two intervals. The coefficient of variation was calculated as a standard deviation of the C or V intervals divided by the average duration of the C or V intervals for each PP. The average results across PPs for each talker are given in Table III. ANOVA results revealed that style and type of material did not have a significant effect on C coef-

ficient of variation: style: $[F(1,5)=2.72, p=0.160]$; material: $[F(2,10)=3.186, p=0.085]$. The two-way interaction was not significant either: $[F(2,10)=3.328, p=0.078]$. There was a significant main effect of material type on V coefficient of variation: $F(2,10)=8.527, p<0.01$, but not of style: $[F(1,5)=1.943, p=0.222]$. The two-way interaction was not significant: $[F(2,10)=0.040, p=0.961]$. The results showed that the coefficient of variation differed for the three types of the materials (lowest for the SP and similar for the short sentences and the LP). However, the results for the coefficient of variation for C and V intervals showed that when speaking rate was taken into account variability for the two intervals remained rather stable across the two speaking styles, i.e., the increase in C and V variability (standard deviations) can largely be accounted for by the overall increase in the durations of the intervals.

We conclude this section with a note on individual talker data and intelligibility results. The overall goal of clear speech research is to find acoustic-articulatory changes that talkers consistently produce that may underlie enhanced intelligibility. While we have identified some acoustic-phonetic conversational-to-clear speaking style changes in the temporal domain, the data presented in this paper are limited in terms of providing a direct link between the articulatory changes and an increase in intelligibility. Our database does not contain enough talkers for meaningful correlational analyses between acoustic-phonetic variation and variability in intelligibility. Furthermore, given the nature of our materials (sentences and paragraphs), numerous other conversational-to-clear speech articulatory changes were implemented by our talkers (discussed in Smiljanic and Bradlow, 2005, 2008) and it would be difficult to assess which of these changes (or in which combination) contribute to increased intelligibility. Finally, we can only speculate at this point how the changes in the overall temporal organization, as discussed here, contribute to intelligibility. Nevertheless, we discuss some tendencies observed in our database, which will have to be explored in more detail in future work.

The results revealed a large amount of variability in individual results for various acoustic measurements and their relation to intelligibility. Nevertheless, some acoustic measurements seem correlated with intelligibility. For instance, larger lengthening of C intervals in short sentences and SP, and of V intervals in short sentences corresponds to increased intelligibility. A similar relationship with intelligibility holds for a speaking rate decrease in short sentences and SP, for an increase in the number of C intervals in short sentences and LP, and for V intervals in LP. However, these same measurements do not exhibit a systematic relation to intelligibility in other instances, i.e., with other materials. Larger lengthening of C intervals for LP and of V intervals for SP and LP, for instance, does not result in increased intelligibility. An insufficient number of data points does not allow for a more systematic analysis of these trends. Furthermore, on some acoustic measures, similar amounts of change affect intelligibility differently for various talkers. For instance, an increase of 5 and 11 C intervals in short sentences

results in an intelligibility gain of 35 and 5 percentage points, respectively, while an increase of 24 C intervals results in only 4% point intelligibility increase.

While it is difficult to ascertain the relationship between these acoustic cues and intelligibility due to the multidimensional clear speech changes in these data, nonetheless, it is likely that some of these clear speech adjustments may contribute to enhanced intelligibility. It seems also likely that a combination of different clear speech changes, rather than each one separately, result in increased intelligibility. Furthermore, clear speech probably varies for different materials and communicative demands, i.e., salient and systematic clear speech changes in short sentence production may not be implemented in longer and more complex materials or in spontaneous speech. Clear speech adjustments may also vary across the portions of longer utterances (similar to the intelligibility variability across different portions of the two paragraphs). Finally, even if a direct link between some acoustic cue (e.g., increased F2 for front vowels) and intelligibility is found to be reliable for all talkers and listeners, it still remains to be seen how this cue is used in running speech where linguistic processing and goals differ. Many of these issues remain to be explored in future work.

IV. SUMMARY AND DISCUSSION

This study investigated the effect of hyperarticulated, intelligibility-enhancing clear speech on temporal characteristics in short sentences and paragraph-length utterances. The major goals were to explore whether temporal restructuring at the level of segmental intervals accompanies changes in speaking style across materials of varying complexity. Moreover, we sought to extend previous findings on clear speech intelligibility by testing the maintenance of clear speech across paragraph-length utterances.

The results of sentence-in-noise listening tests showed a consistent clear speech intelligibility gain for short sentences and across the two paragraphs indicating that spontaneously produced clear speech enhances intelligibility for all material types. The data provided strong evidence that talkers successfully maintained clear speech articulatory modifications across longer stretches of speech such that intelligibility was increased in all portions of the paragraphs. Previous clear speech research largely focused on intelligibility of isolated vowels, words, and short sentences (Picheny *et al.*, 1986; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Ferguson, 2004; Ferguson and Kewley-Port, 2002). The current results expand on these findings by demonstrating that clear speech modifications can increase the intelligibility of longer and more complex utterances. This finding is an important step toward connecting laboratory-based research on variability in intelligibility to more naturalistic communicative settings where speech typically involves longer utterances with highly variable phrase types.

The present data also provide some insight into the underlying organizational framework for the now well-established decrease in speaking rate for naturally produced clear speech. The decrease in speaking rate was achieved through equal lengthening of consonant and vowel intervals in clear speech: %C and %V remained stable across speaking styles. Furthermore, clear speech slowing down resulted in the increased number of pauses (PPs) and of V and C intervals that were dropped or coarticulated in conversational speech. The results here replicate the findings that the overall speaking rate decrease in clear speech was achieved by a combination of individual segment lengthening and an increase in the number of segments and pauses (Picheny *et al.*, 1986; Krause and Braida, 2004; Ferguson and Kewley-Port, 2002; Smiljanic and Bradlow, 2005, 2008). They also expand the previous findings by looking at the variability of C and V intervals in a more systematic and detailed way across longer utterances. The overall effect of the conversational-to-clear speech style modification, thus, appears to involve some temporal restructuring in how speech unfolds in time as seen in the increased number of C and V intervals (and segments) and of pauses. This restructuring can be seen in the realizations of new articulatory targets that enhance segmental and prosodic compositions of the message. Alternatively, the restructuring could be viewed as a reduction in clear speech of articulatory target undershoot and target overlap present in conversational speech.

The connection between the temporal organization change from conversational to clear speech and their relation to increased intelligibility have yet to be fully established. Krause and Braida (2002), for instance, found that clear speech produced at normal/conversational speaking rates increased intelligibility, suggesting that the temporal changes are not necessary conditions for enhanced intelligibility. However, this does not mean that, when present, the temporal changes of slow clear speech are not important for the intelligibility benefit. Listeners may use the temporal cues when available either to further enhance intelligibility or to process the speech signal in a different manner (weighing available cues differently). The present data suggest several possible temporal-related mechanisms that may underlie the intelligibility increase. First, a decrease in speaking rate may be related to greater articulatory precision as seen, for instance, in an increased distance among the contrastive vowel categories (Smiljanic and Bradlow, 2005). Similarly, slower speaking rate allows for the lengthening and insertion of the short segments that were dropped or coarticulated with surrounding sounds in conversational speech. In addition to increasing the accuracy of the phoneme identification, these two effects conspire to make syllable and word structure more salient, presumably increasing accuracy in lexical access and word recognition.

Second, temporal restructuring through increased phrasing may further contribute to higher clear speech intelligibility compared to conversational speech intelligibility. In the present paper, PPs were crudely quantified as occurring between pauses. The results showed that the number of PPs was increased in clear speech. Moreover, informal observations showed an increase in the number of intermediate (ip) and IPs within PPs as well, i.e., a number of smaller PPs that did not coincide with the pauses were produced as well. These phrases are often marked by intonation patterns (F0), phrase-initial and -final lengthening and reduced coarticula-

tion between phonemes across boundaries (Wightman *et al.*, 1992; Keating *et al.*, 2003; Byrd *et al.*, 2000). There is evidence that the acoustic correlates of these phrases, as well as of word-level stress, aid listeners in lexical segmentation, i.e., in finding word boundaries and in resolving lexical competition (Cho *et al.*, 2007; Cutler and Otake, 1994; Fougeron, 1999; Christophe *et al.*, 2004). Christophe *et al.* (2004) found that monosyllabic words were accessed faster in a two-word sequence when the two words belonged to two separate phonological phrases than when they belonged to the same phrase. Recently, Cho *et al.* (2007) showed that listeners were better at resolving lexical ambiguity in two-word sequences when the initial syllable of the second word originated in a higher PP, i.e., intermediate phrase versus word phrase. The authors argued that under the conditions of lexical competition, domain-initial strengthening (longer and more extensive articulatory gestures) helped listeners derive the correct segmentation. These studies, among other, showed that prosodically driven phonetic detail helps listeners in segmenting speech and resolving lexical ambiguity. Similarly, the results of this study suggest that increased phrasing in addition to the increased segmentability (larger number of C and V intervals) provide contexts in which lexical segmentation and lexical access are improved for clear speech. Such detailed prosodic analyses were, however, beyond the scope of this study and will be pursued in future work.

Finally, an important finding of this study was that, in contrast to the increased duration variability as reflected in larger standard deviations, coefficients of variation for both C and V intervals remained unchanged across clear and conversational speech. Similar coefficient of variation stability for speech rhythm across speaking rates was found by Dellwo *et al.* (2004). This result echoes the findings in Smiljanic and Bradlow, 2008 that the proportional distances between the "short and long" vowel categories and between the voiced and voiceless stop categories remained unchanged across the two speaking styles cross-linguistically. In that study, we argued that duration was not a dimension of vowel and stop voicing category contrast enhancement. Rather, the language-specific pronunciation norms along this dimension were maintained in clear and conversational speech. While accumulated production and perception results from previous studies showed that fine grained phonetic detail of segmental duration encodes information from various linguistic levels and in turn governs speech comprehension (Cho *et al.*, 2007; Cutler and Otake, 1994; Fougeron, 1999; Christophe *et al.*, 2004), these data suggest that the same relative durations for these effects should be maintained across changes in speaking styles and rates. In other words, in order for listeners to successfully interpret the speech signal and derive information about linguistic structure (lexical access, segment identification, prosodic structure, etc.), they may rely on local timing relations, which can be measured as durational ratios. These results support the notion that the mechanism guiding articulatory timing may be inherently relational and that both speech production and perception rely on proportional stability in the temporal domain. Finally, these results suggest that increased intelligibility of clear speech may be attributed to prosodic structure enhancement (increased phrasing and enhanced segmentability) in combination with stable global temporal properties.

[1] It is important to note that fast-to-slow speaking rate changes, and conversational-to-clear speech modifications may not involve the same articulatory scaling mechanisms and the similarities and differences between the results of the various studies should, therefore, be interpreted with caution.

[2] Some pauses were eliminated due to the segmentation criteria (paragraphs were segmented into the sentence-length utterances in part where pauses occurred) prior to the leveling procedure. However, a few pauses remained within clear speech sentences, which could have affected their average rms amplitude (lower overall for clear than for the conversational sentences). Nevertheless, across all materials, the perceived subjective loudness remained constant (as determined by the author and the experimenters who tested the participants). Any existing difference may have affected the SNR during spoken portions of the sentences (better for clear speech), which could have artificially inflated the clear speech intelligibility benefit. However, this effect would have occurred only for a few sentences and the overall increase in clear speech intelligibility due to this factor is rather small.

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," Percept. Psychophys. **64**, 121–130.

Boersma, P., and Weenink, D. (**2006**). PRAAT: Doing phonetics by computer (Version 4.4.16). Retrieved from http://www.praat.org/ (Last viewd 12/27/2007).

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for learning-impaired children: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Bradlow, A. R., and Alexander, J. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.

Byrd, D., Kaun, A., Narayanan, S., and Saltzman, E. (**2000**). "Phrasal signatures in articulation," in *Lab. Phonology 5: Acquisition and the Lexicon*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press, Cambridge), pp. 70–87.

Byrd, D., and Saltzman, E. (**2003**). "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," J. Phonetics **31**, 149–180.

Byrd, D., Krivokapic, J., and Lee, S. (**2006**). "How far, how long: On the temporal scope of prosodic boundary effects," J. Acoust. Soc. Am. **120**, 1589–1599.

Cho, T. (**2005**). "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English," J. Acoust. Soc. Am. **117**, 3867–3878.

Cho, T. (**2006**). "Manifestation of prosodic structure in articulation: Evidence from lip movement kinematics in English," in *Laboratory Phonology 8: Varieties of Phonological Competence*, edited by L. Goldstein (Mouton de Gruyter, Berlin, New York), pp. 519–548.

Cho, T., McQueen, J. M., and Cox, E. (**2007**). "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," J. Phonetics **35**, 210–243.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (**2004**). "Phonological phrase boundaries constrain lexical access: I. Adult data," J. Mem. Lang. **51**, 523–547.

Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M. (**2007**). "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech," J. Phonetics **35.2**, 180–209.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for

language-specific listening," J. Mem. Lang. 33, 824–844.

Dellwo, V., Steiner, I., Aschenberner, B., Dankovičová, J., and Wagner, P. (2004). "The BonnTempo-Corpus & BonnTempo-Tools: A database for the study of speech rhythm and rate," in Proceedings of the Eighth ICSLP, Jeju Island, Korea.

de Jong, K. (1995). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," J. Acoust. Soc. Am. 97, 491–504.

Ferguson, S. H. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. 116, 2365–2373.

Ferguson, S. H., and Kewley-Port, D. (2002). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. 112, 259–271.

Ferguson, S. H., and Kewley-Port, D. (2007). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," J. Speech Lang. Hear. Res. 50, 1241–1255.

Fernald, A. (2000). "Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition," Phonetica 57, 242–254.

Fougeron, C. (1999). "Prosodically conditioned articulatory variations: A review," UCLA Working Papers in Phonetics 97, 1–74.

Fougeron, C., and Keating, P. A. (1997). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. 101, 3728–3739.

Gagne, J.-P., Rochette, A.-J., and Charest, M. (2002). "Auditory, visual and audiovisual clear speech," Speech Commun. 37, 213–230.

Grabe, E., and Low, E. L. (2002). "Durational variability in speech and the rhythm class hypothesis," Lab. Phonology 7, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 515–546.

Helfer, K. S. (1998). "Auditory and auditory-visual recognition of clear and conversational speech by older adults," J. Am. Acad. Audiol 8, 234–242.

Hirata, Y. (2004). "Effects of speaking rate on the vowel length distinction in Japanese," J. Phonetics 32, 565–589.

Hirata, Y., and Whiton, J. (2005). "Effects of speaking rate on the single/geminate stop distinction in Japanese," J. Acoust. Soc. Am. 118, 1647–1660.

Johnson, K., Flemming, E., and Wright, R. (1993). "The hyperspace effect: Phonetic targets are hyperarticulated," Language 69, 505–528.

Junqua, J. C. (1993). "The Lombard reflex and its role on human listeners and automatic speech recognizers," J. Acoust. Soc. Am. 93, 510–524.

Keating, P. A., Cho, T., Fougeron, C., and Hsu, C. (2003). "Domain-initial strengthening in four languages," in Lab. Phonology 6: Phonetic Interpretations, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 145–163.

Kessinger, R. H., and Blumstein, S. E. (1998). "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," J. Phonetics 26, 117–128.

Krause, J. C., and Braida, L. D. (2002). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. 112, 2165–2172.

Krause, J. C., and Braida, L. D. (2004). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. 115, 362–378.

Kuhl, P. K., Andruski, J. E., Chistovich, L., Chistovich, I., Kozhevnikova, E., Sundberg, U., and Lacerda, F. (1997). "Cross language analysis of phonetic units in language addressed to infants," Science 227, 684–686.

(2002). Handbook of the International Phonetic Association (Cambridge University Press, Cambridge).

Lehiste, I. (1970). Suprasegmentals (MIT, Cambridge, MA).

Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H&H theory," in Speech Production and Speech Modeling, edited by W. J. Hardcastle and A. Marchal (Kluwer, Netherlands).

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G. (2004). "Clear speech perception in acoustic and electrical hearing," J. Acoust. Soc. Am. 116, 2374–2383.

Low, E. L., Grabe, E., and Nolan, F. (2000). "Quantitative characterizations of speech rhythm: "Syllable-timing" in Singapore English," Lang Speech 43, 377–401.

Miller, J. L., Green, K. P., and 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., and Dexter, E. R. (1988). "Effects of speaking rate and lexical status on phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. 14, 369–378.

Miller, J. L., and Volaitis, L. E. (1989). "Effects of speaking rate on the perceived internal structure of phonetic categories," Percept. Psychophys. 46, 505–512.

Moon, S.-J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. 96, 40–55.

Nagao, K., and de Jong, K. J. (2007). "Perceptual rate normalization in naturally produced rate-varied speech," J. Acoust. Soc. Am. 120, 2882–2898.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. 95, 1581–1592.

Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (2002). "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," J. Acoust. Soc. Am. 112, 1627–1641.

Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. 32, 693–703.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. 29, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, 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," J. Speech Hear. Res. 32, 600–603.

Pickett, E. R., Blumstein, S. E., and Burton, M. W. (1999). "Effects of speaking rate on the singleton/geminate consonant contrast in Italian," Phonetica 56, 135–157.

Ramus, F., Nespor, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," Cognition 72, 265–292.

Schum, D. J. (1996). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol 7, 212–218.

Sedaris, D. (2001). Me Talk Pretty One Day, New Ed ed. (Abacus, New York, New York).

Silverman, K. E. A., Beckman, M., Pitrelli, J. F., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). "ToBI: A standard for labeling English Prosody," in Proceedings of the 1992 International Conference on Spoken Language Processing, Banff, Canada, pp. 867–870.

Skowronski, M. D., and Harris, J. G. (2005). "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments," Speech Commun. 48, 549–558.

Smiljanic, R., and Bradlow, A. R. (2005). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. 118, 1677–1688.

Smiljanic, R., and Bradlow, A. R. (2008). "Stability of temporal contrasts in conversational and clear speech," J. Phonetics 36.1, 91–113. http://www.sciencedirect.com/science/journal/00954470. Last viewed 1/03/08.

Smiljanic, R., and Bradlow, A. R. (2007). "Clear speech intelligibility: Listener and talker effects," Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrucken, Germany.

Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," J. Speech Hear. Res. 28, 455–462.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (1996). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. 39, 494–509.

Uchanski, R. M. (1988). "Spectral and temporal contributions to speech clarity for hearing impaired listeners," Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

Volaitis, L. E., and Miller, J. L. (1992). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," J. Acoust. Soc. Am. 92, 723–735.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). "Segmental durations in the vicinity of prosodic phrase boundaries," J. Acoust. Soc. Am. 91, 1707–1717.