

Jelena Krivokapić

# Rhythm and convergence between speakers of American and Indian English

**Abstract:** The study examines rhythmic convergence between speakers of American and Indian English. Previous research has shown that American English shows tendencies towards stress-timing, and Indian English has been claimed to be syllable-timed (Crystal 1994). Starting from the view that languages differ in their rhythmic tendencies, rather than that they have categorically different rhythmic properties, we examine in an acoustic study the rhythmic tendencies of the two languages, and whether these tendencies can change in the course of an interaction. The focus is on temporal properties (specifically, the duration of stressed syllables and of feet). The results show evidence of mixed rhythmic properties for both languages, with Indian English being more syllable-timed than American English. American speakers show a trend towards changes in foot duration that can be interpreted as accommodation in speech rate or as convergence towards a more syllable-timed foot duration pattern. One Indian English speaker converges in both examined properties towards a more stress-timing pattern. The results are discussed within a dynamical model of rhythmic structure (Saltzman, Nam, Krivokapić, and Goldstein 2008). It is suggested that rhythmic convergence can arise via a tuning between speakers of the prosodic interoscillator coupling function that is proposed in that model.

---

**Jelena Krivokapić:** Yale University. E-mail: jelena.krivokapic@yale.edu

## 1 Introduction

### 1.1 Convergence

In the process of an interaction, speakers can converge to each other, becoming more similar in their language (Pardo 2006; Babel 2009; Lewandowski 2012). Convergence has been found to occur at the phonetic, prosodic, syntactic, and lexical level. Thus speakers' VOTs and vowels can become more similar in their properties to the productions of their co-speakers (e.g., Sancier and Fowler 1997; Nielsen 2011; Babel 2012), and speakers use lexical items and syntactic

constructions that have been used by their co-speakers (Garrod and Anderson 1987; Pickering and Garrod 2004). At a more global level, speakers can also converge in speech rate, pause duration, fundamental frequency, pitch accent placement, and intensity (e.g., Natale 1975; Street 1984; Zvonik and Cummins 2002; Kim and Nam 2009; Krivokapić 2011; Levitan and Hirschberg 2011).

An aspect of convergence that has not been examined so far is whether speakers converge in their rhythmic properties. However, it has been argued that speakers are sensitive to the rhythmic properties (in terms of isochrony of intervals between pitch accents) of their co-speaker's speech and that such rhythmic properties inform turn-taking behavior (Couper-Kuhlen 1993; Auer, Couper-Kuhlen, and Müller 1999). Empirical support for this view is mixed (see also Bull and Aylett [1998] for a number of other factors affecting turn-taking). Włodarczak, Šimko, and Wagner (2012) show evidence that turn-taking is sensitive to regularity in syllable duration (see also Wilson and Wilson [2005] for arguments that turn-taking is based on convergence in speech rate at the level of the syllable). However, in a later work Włodarczak et al. (in press) suggest, based on evidence from Finnish, that it might be the alternation between vowels and consonants which determines the onset of turn-taking. Beňuš, Gravano, and Hirschberg (2011) show evidence of turn-taking being guided by rhythmic structure (in terms of isochrony of intervals between pitch accents) and suggest that it is this rhythmic structure that affords, in the sense of Gibson (1979), turn-taking in a manner appropriate to a conversational setting. From this point of view, we might expect speakers' rhythmic properties to converge as well, possibly even, if these form such a crucial component in interactions, to be the first to converge. It is unclear, however, whether rhythmic structure guides turn-taking when co-speakers are from dialects that differ in their rhythmic properties. A study on turn-taking between speakers of British English (a stress-timed language) and Singapore English (a syllable-timed language) found limited evidence of rhythmic integration between the speakers (Szczepek Reed 2010).

## 1.2 Rhythm

Temporal properties of speech rhythm have been extensively researched. A focus of a large number of studies has been the attempt to classify languages as stress-timed and syllable-timed. However, despite this effort and the advances made in the study of rhythm, a conclusion as to the existence of these rhythmic classes and as to the properties of stress-timed and the properties of syllable-timed languages has not been reached. Starting with the work by Pike (1945) and Abercrombie (1967), a large body of research has examined the idea that rhythm is

based on durational properties and that it involves regularly occurring intervals of equal duration, such as feet and syllables, either in the acoustic signal or as perceived by listeners (e.g., Bolinger 1965; Huggins 1972; Lehiste 1972, 1977; Nakatani, O'Connor, and Aston 1981). Evidence against a strict view of isochrony, based on durational characteristics, is overwhelming (see the overview in Arvaniti 2012). Dauer (1983, 1987) suggests that the impression of rhythmic structure arises from a number of converging properties in a language that are of a more qualitative nature, such as the variation of syllable structure, the existence of vowel reduction, and phonetic properties of prominence marking. Dauer (1983) argues that, rather than being divided into rhythmic classes, languages have more or less stress-timed language properties. A number of rhythm metrics has tried to capture this view of rhythm as a gradient property, while still, contrary to Dauer (1983), keeping the basic distinction between stress-timed and syllable-timed languages (e.g., Ramus, Nespore, and Mehler 1999; Grabe and Low 2002). Arvaniti (2009, 2012), however, examines different rhythm metrics and shows that they do not provide a reliable indicator of rhythmic classes. Based on a discussion of the nature of rhythm, she further argues that rhythm and timing cannot be equated, and that speech rhythm research should move away from examining only durational properties (Arvaniti 2009).

Despite the arguments against a categorical division of languages into stress-timed and syllable-timed, there is consensus that aspects of linguistic timing are relevant for rhythm. Evidence for the relevance of durational properties is seen in polysyllabic shortening – the phenomena that stressed syllables shorten with the addition of unstressed syllables to the word or the foot (e.g., Lehiste 1972; Klatt 1973; Port 1981; Rakerd, Sennett, and Fowler 1987; Kim and Cole 2005; White and Turk [2010] for pitch accented but not for de-accented words; and Shattuck-Hufnagel and Turk [2011] for polysyllabic shortening within words but generally not across words). For example, in a corpus study of American English, Kim and Cole (2005) find that the duration of the foot (where foot includes both cross-word feet and within-word feet) increases with the addition of unstressed syllables, while at the same time the stressed syllable shortens. The unstressed syllables do not shorten, and shortening does not occur if the foot spans a prosodic phrase boundary. By somewhat constraining the duration of the foot, polysyllabic shortening allows for a tendency towards isochrony of the foot to arise.

Saltzman et al. (2008) model these findings within the task-dynamics model of speech production (Saltzman and Munhall 1989) and a coupled oscillator model of intergestural timing (Saltzman and Byrd 2000; Nam and Saltzman 2003). Saltzman et al. (2008) show that rhythmic properties of American English can emerge through the coupling of a foot oscillator, syllable oscillator, and a

temporal modulation gesture  $\mu$  (see also Byrd and Saltzman [2003] for the related  $\pi$ -gesture for boundary lengthening). Within this model, a  $\mu$ -gesture slows the gestural activation during the stressed syllables. This leads to the lengthening of the gestures active during this period, resulting in the longer duration of stressed in comparison to unstressed syllables. The foot oscillator aims to keep the duration of the foot constant by squeezing the syllables in it, thus this oscillator is a force towards foot synchrony. The syllable oscillator aims to keep the duration of the syllable constant, and is thus a force towards syllable isochrony. The foot oscillator and the syllable oscillator are in competition, which is captured in the interoscillator coupling function. Interoscillator coupling functions are bi-directionally specified, and each oscillator has its own weighting, defining how strong the influence of the oscillator in the function is. The stronger the weighting of the foot oscillator, the more dominant the foot will be, and the stronger the tendency towards foot isochrony. This foot-dominance leads to polysyllabic shortening (see also O'Dell and Nieminen [1999] for such an approach). However, foot dominance leads to *all* syllables becoming shorter with the addition of more syllables to the foot. To achieve the shortening of the stressed but not of the unstressed syllables (as observed in Kim and Cole 2005), the foot oscillator's weight in the interoscillator coupling function is weaker during the unstressed than during the stressed syllables. In this way, the foot oscillator's squeezing is weaker during the unstressed than during the stressed syllables. The model of Saltzman et al. (2008; see also Nam, Saltzman, Krivokapić, and Goldstein 2008; O'Dell and Nieminen 1999) allows the *tendencies* towards isochrony to be captured, thus allowing both the intuitions and the evidence from the data to be accommodated, while at the same time not predicting strict isochrony. This is the view of rhythm adopted in this study, i.e., rhythm is seen as a *tendency* towards temporal regularity in the occurrence of a certain element that arises through the temporal regularity of oscillators and is mediated through the interaction with other prosodic oscillators (and, of course, through the interaction with other properties of speech such as, for example, segmental properties, prosodic boundaries [see, e.g., Bolinger 1965; Kim and Cole 2005], or speech planning).

### 1.3 Goals

The first aim of the present study is to compare the rhythmic properties of American English and Indian English (i.e., the English spoken by speakers who grew up in India and learned English as an L2), two languages that have been claimed to differ in their rhythmic properties. American English is claimed to be a stress-timed language, and shows, as mentioned above, evidence of a tendency towards

foot isochrony and the shortening of stressed syllables with the addition of unstressed syllables to the foot. Indian English has been claimed to be a syllable-timed language (Wells 1982; Crystal 1994; Trudgill and Hannah 2008). We are aware of only two studies examining rhythmic properties of Indian English. Sailaja (2009) reports a study by Prabhakar Babu (1971) finding that Indian English is “neither stress-timed nor syllable-timed”, but no further details of the study are provided. Fuchs (2012) finds a tendency towards lower variability in the duration of vocalic intervals in Indian English compared to British English, indicating a tendency towards syllable timing in Indian English as compared to British English.

The second aim is to investigate whether rhythmic convergence occurs in an interaction between speakers of these languages, and how to characterize convergence if it occurs. An acoustic experiment was conducted to examine these questions. The study used solo readings to examine the rhythmic properties of the languages. To examine convergence we used the synchronous speech paradigm (Cummins 2002) in which two speakers read a text simultaneously, prompted by the experimenter. This paradigm minimizes individual, non-linguistic variation without introducing artificial temporal properties to speech (Cummins 2004; Zvonik and Cummins 2002, 2003). It has been suggested that synchronization relies on speakers’ shared knowledge of linguistic timing (Cummins 2004). We use this paradigm in our study with the idea that the task itself might facilitate convergence between speakers.

## 2 Methods

### 2.1 Subjects

Eight speakers (four dyads) were recorded. Four subjects were native speakers of American English from the East Coast (AE1, female, 43 years old; AE2, female, 20 years old; AE3, male, 23 years old; AE4, male, 18 years old). Four subjects were life-time English speakers born in India (IE1, female, 35 years old; IE2, female, 19 years old; IE3, male, 19 years old; IE4, male, 29 years old). The Indian English speakers had different L1 backgrounds: Participant IE1 is a speaker of Hindi and Marathi, and she started learning English when she was 5. She lived in the U.S. for 8 years prior to the experiment. Participant IE2 is a native speaker of Hindi and a heritage Kashmiri speaker. She started learning English when she attended kindergarten in India at about 3 years of age. She had lived in the U.S. for 1 year at the time of the experiment. IE3 is a speaker of Hindi who grew up speaking both

English and Hindi at home, although Hindi was the dominant language. He considers Hindi his first language. He had lived in the U.S. for 6 months prior to the experiment. IE4 is a native speaker of Hindi and Kumaoni. He learned English at school, and he had lived in the US for 6 months at the time of the experiment. Dyad 1 consisted of speakers AE1 and IE1, Dyad 2 of speakers AE2 and IE2, and so on. The dyads were matched in gender and as close as possible in age, although for Dyad 4 the age difference was quite high. Dyads 1 and 2 were female, Dyads 3 and 4 male. All participants were undergraduate or graduate students, or residents of New Haven. They were paid for their participation and were naïve as to the purpose of the experiment.

## 2.2 Materials and recording

The materials consisted of a short story (given in the Appendix) that was read eight times (the story was constructed by the author, based on Honorof, McCullough, & Somerville [2000] and Wells' [1982] lexical sets). Between each repetition, subjects read 53 filler sentences that were part of another experiment. The filler sentences were randomized for each repetition.

In the first part of the experiment, participants read each sentence at the verbal prompt of the experimenter, and the story with only one prompt at the start of the story. The second part of the experiment took place about a week later and was conducted using the synchronous speech paradigm (Cummins 2002, 2004). Two participants (one speaker of American English and one speaker of Indian English) were paired up in each dyad. They were seated facing each other and they read the materials (same as in the first part of the experiment) simultaneously. As in the solo condition, the experimenter prompted the participants for each sentence, and once for the beginning of the story. In both the solo and the synchronous condition, prior to the recording the subjects familiarized themselves with the experiment materials by reading the story and the sentences aloud. During the experiment, they were asked to read the materials as if reading a story to someone. In case of errors, subjects were asked to repeat the sentence. Errors were rare.

Subjects were recorded on a DAT recorder using Shure head mounted uni-directional microphones. In the synchronous condition, the participants were recorded on separate channels, thus avoiding interference. The recordings were then transferred to a PC onto the right and left channels of a stereo file.

For the study, the eighth repetition of the story was used for both the solo and the synchronous condition. This repetition was chosen as it was assumed that the speakers will, over the course of the experiment, become more fluent and in that

way more rhythmic in their productions, and that they will be establishing their speech pattern throughout the experiment. Thus the last repetition seemed to be the best choice to examine the rhythm of the individual languages, whether rhythmic convergence occurs, and what the properties of convergence are.

## 2.3 Syllabification and prosodic annotation

The data were prosodically annotated and the duration of the syllables and feet labeled. For American English, the prosodic annotation was conducted by the author, marking Intermediate Phrases and Intonation Phrases as specified by the ToBI guidelines (Beckman and Ayers Elam 1997). Primary and secondary stress was also marked on lexical words. For function words, stress was marked if the words were pitch accented (following Selkirk 1984; Kim and Cole 2005). For syllabification, the 17<sup>th</sup> edition of the *Cambridge English Pronunciation Dictionary* (Roach, Hartman, Setter, and Jones 2006) was used. The syllabification principles in Roach et al. (2006) are as follows: the Maximal Onset Principle was followed, whereby intervocalic consonants were marked as onsets, unless this would yield phonotactic violations. In compounds, the component word boundaries are preserved.

For Indian English the situation was more complex, given the different language backgrounds of the speakers involved in the study, and given that accepted guidelines for prosodic annotation and for syllabification were not available. Two consultants, native speakers of Hindi, annotated stress and syllable structure. They agreed on stress, and in all but six cases they agreed on syllable structure. The instances of disagreement were all cases in which one of the consultants was not certain about the syllabification. A third Indian English speaker was asked about these cases. Function words were marked as stressed if they were perceived as prominent. The author annotated prominence (any word that was prominent in a sentence, regardless of what type of prominence it might be, was marked as prominent) and phrase boundaries. This was done based on the examination of the pitch track and based on perception. The 53 sentences that were collected in addition to the story contained stimuli designed to elicit various prosodic structures. These were used as a guide for the annotation. In addition, one consultant also annotated the data of all Indian English speakers for prosodic boundaries and for prominence. Except in a few cases, her labeling corresponded to the markings the author conducted for the data. In cases of disagreement, the consultant's labeling was chosen. As for American English, for Indian English prominence was only used to determine when to assign stress to function words.

The foot was taken to be from the onset of one stressed syllable to the onset of the next stressed syllable (where both primary and secondary stress counts, as in Kim and Cole [2005], and, as in Kim and Cole [2005], both cross-word and within-word feet were taken into account). To exclude the effect of boundary adjacent lengthening, feet (and the syllables in them) adjacent to an Intermediate or Intonation Phrase boundary were not included in the analysis. There were no pauses apart from pauses at prosodic boundaries.

## 2.4 Acoustic measurements and synchronization

Syllable and foot duration were examined. The measurements were conducted in Praat (Boersma and Weenink 2012). The acoustic segmentation into syllables was conducted in the same manner for both languages. The main guidelines were decided with the goal of keeping the measurements consistent, and segmentation criteria were based to a large extent on the criteria outlined in Klatt (1976), Turk and Shattuck-Hufnagel (2000, 2007), Turk, Nakai, and Sugahara (2006), and White and Turk (2010). Vowels following a stop were labeled as starting from the release burst, and the end of vowels preceding a stop was marked at the significant drop in amplitude indicating a closure. Creaky voicing was included in the duration of the vowel. The nasal-vowel boundary was marked by a sharp drop in amplitude. Nasals were marked by the beginning/end of formant structure. Fricatives were labeled as starting/ending with the onset/end of frication. To label vowels preceding or following ‘r’, the point was taken where the formants began/ended changing rapidly for the ‘r’, often occurring together with a decrease in amplitude in higher formants or an overall drop in amplitude. The vowel to /w, l, j/ boundaries were measured at the end of the rapid change in formant transitions for F1 and F2, which also typically occurred at the same time as a sharp decrease in amplitude.

As a measure of synchrony between speakers, for each pause-delimited phrase, the temporal difference between the beginning of a phrase for one speaker of the dyad and the beginning of the same phrase for the other speaker of the dyad was calculated (following similar procedures reported in Cummins 2003; Krivokapić 2007b). The same analysis was done for phrase ends. The average difference for Dyad 1 for phrase beginning is 86 ms ( $SD = 69$  ms) and for phrase end 43 ms ( $SD = 31$  ms), for Dyad 2 for phrase beginning 55 ms ( $SD = 37$  ms) and for phrase end 48 ms ( $SD = 92$  ms), for Dyad 3 for phrase beginning 69 ms ( $SD = 43$  ms) and for phrase end 39 ms ( $SD = 31$  ms), and for Dyad 4 for phrase beginning 81 ms ( $SD = 55$  ms) and for phrase end 53 ms ( $SD = 83$  ms). This level of synchronization is a bit less than usually reported (Cummins 2003; Krivokapić



2007b).<sup>1</sup> Importantly however, for all dyads it was never the case that one speaker was consistently leading and the other consistently following; instead, sometimes one speaker and sometimes the other speaker would start first, and which speaker started first did not determine which speaker ended first. Like in previous studies, this indicates that speakers did indeed synchronize, rather than one speaker leading the other.

## 2.5 Statistical analysis

A two-factor ANOVA on stressed syllable duration and on foot duration was performed, testing for each speaker separately the effects of (1) the number of syllables in the foot (with two levels: one and multiple, where “multiple” means two or three syllables per foot) and (2) the speaking condition (with two levels: solo or synchronous). The analysis was conducted on z-scores calculated for each speaker separately for stressed syllables and for feet. The criterion for significant difference was  $p < .05$ . All and only significant results are reported.

The rhythmic properties of the languages were examined in the solo condition. To examine convergence, the changes that occur in the duration of the stressed syllables and the feet in the synchronous compared to the solo task were evaluated.

The same number of mono- and multisyllabic feet was taken across comparisons. All monosyllabic feet were included in the analysis, and the number of disyllabic and trisyllabic feet was adjusted so that the sum of disyllabic and trisyllabic feet together matched that of the monosyllabic feet. The feet that were not included in the analysis were removed semi-randomly throughout the experiment (e.g., by removing every  $x^{\text{th}}$  disyllabic foot). The ratio of disyllabic and trisyllabic feet in the story matched the ratio of the disyllabic and trisyllabic feet that were kept in the analysis. By removing disyllabic and trisyllabic feet, the number of mono- and multisyllabic feet in both the solo and synchronous condition was the same. By extension, this also means that the number of stressed syllables in all conditions was equal. The total number of feet in the analysis for each speaker is as follows (as a reminder, AE are American English speakers and IE are Indian English speakers): AE1 = 43, AE2 = 39, AE3 = 39, AE4 = 44, IE1 = 40, IE2 = 39, IE3 = 39, IE4 = 47.

---

<sup>1</sup> For various trials, Cummins (2003) reports mean differences of around 61 ms ( $SD = 38$ ) to 63 ms ( $SD = 34$ ) phrase initially and 40 ms ( $SD = 24$ ) to 44 ms ( $SD = 25$ ) phrase finally. Krivokapić (2007b) reports mean differences of 79 ms ( $SD = 64$  ms) phrase initially, and 38 ms ( $SD = 48$  ms) phrase finally.

## 2.6 Predictions

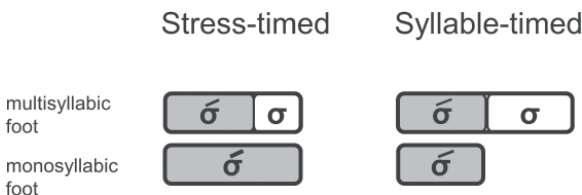
As discussed above, in stress-timed languages, the foot constrains the duration of the syllables, giving rise to a tendency towards foot isochrony and to polysyllabic shortening. In syllable-timed languages, the syllable is more dominant, giving rise to a tendency towards syllable isochrony. Thus in the syllable-timed languages the duration of the foot tends to increase with the number of syllables in it, and the syllables tend to not shorten with an increase in the number of syllables in the foot.

The predictions are then as follows: For American English stressed syllables will be longer in monosyllabic than in multisyllabic feet, while for Indian English there will not be a difference in the duration of the stressed syllables in mono- compared to multisyllabic feet (as schematically shown in Figure 1). For the foot, the prediction is that for American English the duration of the foot will show a tendency to be the same in mono- and in multisyllabic feet. For the syllable-timed Indian English, the prediction is that the duration of the foot will increase with the number of syllables in it (as schematically shown in Figure 2). Note that the predictions are tendencies only, i.e., the prediction is that the languages will show tendencies towards these extremes.

A difference in the behavior of a variable in the synchronous compared to the solo condition will indicate convergence (or divergence, i.e., speakers becoming more different from one another). For American English speakers, evidence of



**Fig. 1:** Schematized predictions for the duration of the stressed syllable.



**Fig. 2:** Schematized predictions for foot duration.

convergence could be seen in (1) less polysyllabic shortening in the synchronous compared to the solo condition or (2) more increase in the duration of the foot depending on the number of syllables in it in the synchronous compared to the solo condition. For Indian English speakers, evidence of convergence could be seen in (1) more polysyllabic shortening in the synchronous compared to the solo condition or (2) less increase in the duration of the foot depending on the number of syllables in it in the synchronous compared to the solo condition.

## 3 Results

### 3.1 Stressed syllable duration

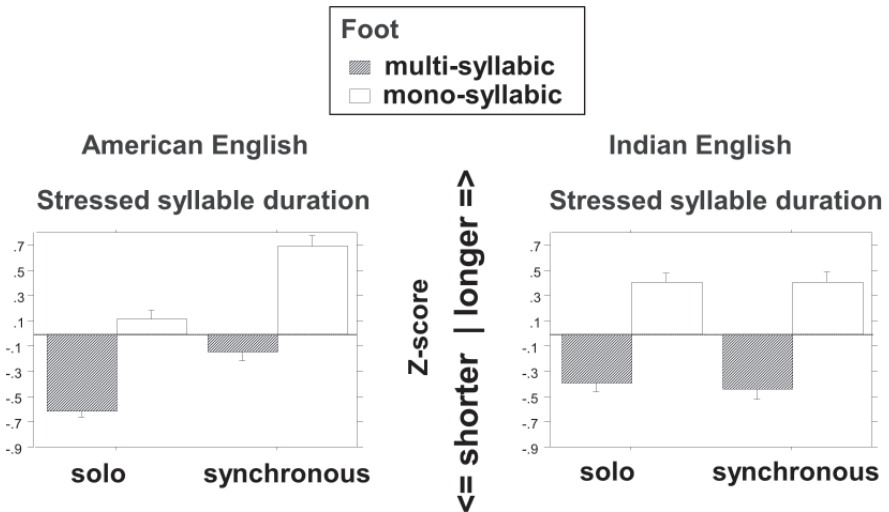
#### 3.1.1 American English speakers

For American English speakers, the *z*-scored duration of the stressed syllable in mono- and multisyllabic feet, by condition, is given in Table 1, together with the statistically significant results of the analysis. As can be seen, all speakers show an effect of speaking condition, such that stressed syllables are longer in the synchronous than in the solo condition. All speakers also show an effect of the number of syllables in the foot, such that the stressed syllables are longer in the mono- compared to the multisyllabic condition. There are no significant interactions.

Since all speakers patterned in the same manner, the results of the individual speakers were pooled. The results are shown in Figure 3 (left). Note that the figures are in *z*-scores. Values above zero represent syllable duration above the subjects' average syllable duration, and values below zero represent syllable duration below the subjects' average syllable duration. The pooled data show an effect of speaking condition ( $F(1, 656) = 60.038, p < .0001$ ), such that stressed syllables are longer in the synchronous than in the solo condition. There is also an effect of the number of syllables in the foot ( $F(1, 656) = 131.376, p < .0001$ ), such that stressed syllables are shorter in mono- compared to multisyllabic feet. Thus both for the individual and for the pooled speakers, the duration of the stressed syllable decreases with the addition of unstressed syllables to the foot, as is predicted for stress-timed languages. As there is no interaction between the two tested factors, we can assume there is no evidence of “rhythmic convergence”, either for individual speakers or for the speakers pooled.

**Table 1:** American English speakers. Average stressed syllable duration (z-scores) with standard deviations (in parenthesis) for the number of syllables for each condition. Significant results for the effect of number of syllables in the foot, speaking condition, and interactions are given in the second row of the table.

	AE1	AE2	AE3	AE4
<b>ANOVA</b>				
Speaking condition	$F(1, 168) = 18.678$ $p < .0001$	$F(1, 152) = 4.398$ $p = .0376$	$F(1, 152) = 23.536$ $p < .0001$	$F(1, 172) = 17.229$ $p < .0001$
Foot syllable number	$F(1, 168) = 15.226$ $p = .0001$	$F(1, 152) = 56.226$ $p < .0001$	$F(1, 152) = 35.301$ $p < .0001$	$F(1, 172) = 33.624$ $p < .0001$
Solo, monosyllabic	-.105 (.851)	.369 (.914)	.073 (.890)	.138 (.967)
Solo, multisyllabic	-.521 (.531)	-.615 (.728)	-.663 (.631)	-.682 (.546)
Synchronous, monosyllabic	.616 (1.185)	.703 (.995)	.849 (.998)	.645 (.983)
Synchronous, multisyllabic	-.047 (.935)	-.373 (.767)	-.079 (.935)	-.081 (.964)



**Fig. 3:** Shortening of stressed syllables.

### 3.1.2 Indian English speakers

The results for Indian English speakers are shown in Table 2. All speakers show an effect of the number of syllables in the foot, such that the stressed syllable is longer in mono- compared to multisyllabic feet. One speaker (IE1) shows an interaction of speaking condition and number of syllables in the foot, shown in Figure 4. The interaction is such that in the synchronous condition the stressed syllable shortens more than in the solo condition.

The three speakers who showed the same pattern (i.e., all but IE 1) were pooled together for further analysis. The pooled results are shown in Figure 3 (right). The results show an effect of the number of syllables in the foot on stressed syllable duration, such that, as for the individual speakers, the stressed syllable is longer in mono- compared to multisyllabic feet ( $F(1, 496) = 102.994, p < .0001$ ). The effect of the number of syllables in the foot on the shortening of the stressed syllable is comparable for American and Indian English speakers, as can be seen when the solo conditions are compared. The difference, in *z*-scores, between the stressed syllables in the mono- and multisyllabic feet is .789 in Indian English compared to .733 in American English (see also Figure 3, left).

**Table 2:** Indian English speakers. Average stressed syllable duration (*z*-scores) with standard deviations (in parenthesis) for the number of syllables for each condition. Significant results for the effect of number of syllables in the foot, speaking condition, and interactions are given in the second row of the table.

	IE1	IE2	IE3	IE4
ANOVA				
Foot syllable number:	$F(1, 156) = 28.786$ $p < .0001$	$F(1, 152) = 26.462$ $p < .0001$	$F(1, 152) = 28.016$ $p < .0001$	$F(1, 184) = 49.681$ $p < .0001$
Speaking condition * Foot syllable number	$F(1, 156) = 4.221$ $p = .0416$			
Solo, monosyllabic	.163 (.982)	.418 (1.052)	.289 (.918)	.489 (.694)
Solo, multisyllabic	-.308 (.777)	-.456 (.879)	-.458 (.703)	-.269 (.970)
Synchronous, monosyllabic	.590 (1.093)	.312 (.997)	.443 (1.114)	.426 (1.074)
Synchronous, multisyllabic	-.466 (.694)	-.342 (.756)	-.358 (.880)	-.643 (.764)

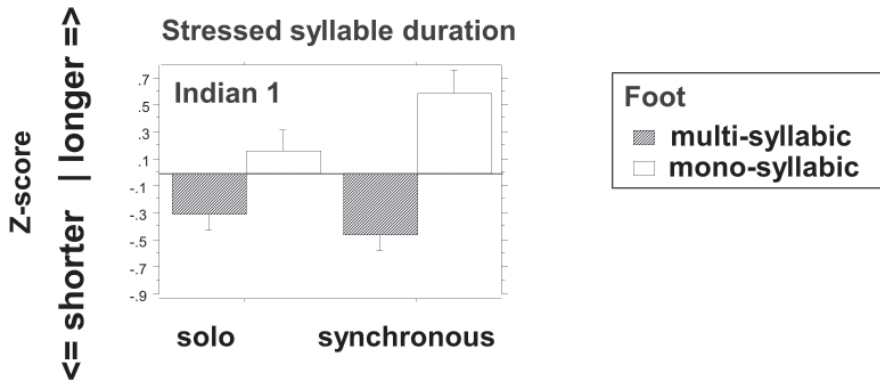


Fig. 4: Stressed syllable shortening for Indian English speaker 1.

To summarize, for Indian English the stressed syllable is longer in mono- than in multisyllabic feet. This is consistent with a stress-timing pattern, rather than with the expected syllable-timing pattern. For one speaker there is evidence of convergence, i.e., in our interpretation, the speaker becomes more stress-timed when co-speaking with an American English speaker.

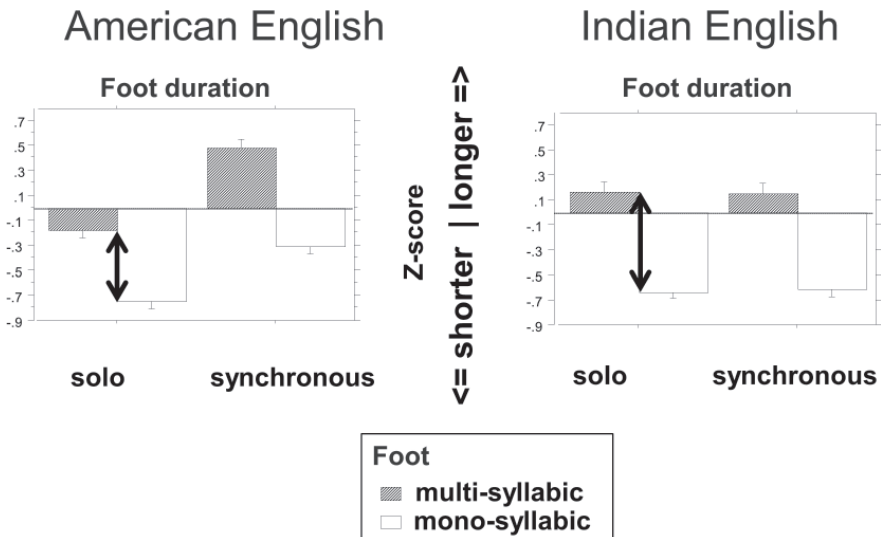
## 3.2 Foot duration

### 3.2.1 American English speakers

The *z*-scored duration of the foot for American English subjects is given in Table 3, together with the significant results of the analysis. For all subjects the effect of the number of syllables in the foot is significant, and the effect is such that multi-syllabic feet are longer than monosyllabic feet. There is also an effect of speaking condition for all speakers, such that foot duration is longer in the synchronous compared to the solo condition. No statistically significant interactions were found for the individual speakers. Since all subjects showed the same pattern, the speakers' data were pooled together. The pooled data show an effect of number of syllables in the foot ( $F(1, 656) = 79.779, p < .0001$ ), such that multisyllabic feet are longer than monosyllabic feet. There is also an effect of speaking condition ( $F(1, 656) = 122.495, p < .0001$ ), such that the duration of the feet are overall longer in the synchronous than in the solo condition. There is a tendency towards a significant interaction ( $F(1, 656) = 3.282, p = .0705$ ) shown in Figure 5. As can be seen from the figure, the interaction is such that in the synchronous condition, the effect of the number of syllables on the duration of the foot is stronger than in

**Table 3:** American English speakers. Average foot duration (z-scores) with standard deviations (in parenthesis) for the number of syllables in each condition. Results for the effect of number of syllables in the foot, speaking condition, and interactions are given in the second row of the table.

	AE1	AE2	AE3	AE4
ANOVA				
Speaking condition	$F(1, 168) = 20.285$ $p < .0001$	$F(1, 152) = 6.293$ $p = .0132$	$F(1, 152) = 34.398$ $p < .0001$	$F(1, 172) = 23.912$ $p < .0001$
Foot syllable number	$F(1, 168) = 50.777$ $p < .0001$	$F(1, 152) = 30.915$ $p < .0001$	$F(1, 152) = 17.828$ $p < .0001$	$F(1, 172) = 27.575$ $p < .0001$ ;
Solo, monosyllabic	-.865 (.615)	-.676 (.666)	-.725 (.707)	-.722 (.673)
Solo, multisyllabic	-.026 (.719)	-.051 (.817)	-.314 (.703)	-.328 (.661)
Synchronous, monosyllabic	-.343 (.857)	-.427 (.722)	-.097 (.799)	-.369 (.684)
Synchronous, multisyllabic	.543 (.944)	.317 (.856)	.610 (1.049)	.445 (.987)



**Fig. 5:** Foot duration. The black arrows show the qualitatively stronger effect that the increase of number of syllables has on the duration of the foot in Indian English compared to American English.

the solo condition. Thus American English speakers show evidence of a syllable-timing pattern in the duration of the foot: the duration of the foot depends on the number of syllables in it, such that the foot is longer when it contains two or three syllables than when it contains only one syllable. The *tendency* towards an interaction can be interpreted as the speakers becoming more syllable-timed when co-speaking with an Indian English speaker, as the foot seems to reduce its constraining effect on the duration of the syllables (an alternative interpretation will be discussed in Section 4).

### 3.2.2 Indian English speakers

The results of the analysis for the Indian English speakers are given in Table 4. For all speakers there was an effect of the number of syllables in the foot, such that multisyllabic feet were longer than monosyllabic feet. For one speaker, IE1, there is also an interaction effect of speaking condition and number of syllables in the foot, shown in Figure 6. The effect is such that the difference in duration between

**Table 4:** Indian English speakers. Average foot duration (z-scores) with standard deviations (in parenthesis) for the number of syllables in each condition. Results for the effect of number of syllables in the foot, speaking condition, and interactions are given in the second row of the table.

	IE1	IE2	IE3	IE4
ANOVA				
Foot syllable number:	$F(1, 156) = 38.879$ $p < .0001$	$F(1, 152) = 45.290$ $p < .0001$	$F(1, 152) = 53.470$ $p < .0001$	$F(1, 184) = 39.780$ $p < .0001$ ;
Speaking condition *	$F(1, 156) = 5.224$ $p = .0236$			
Foot syllable number:				
Solo, monosyllabic	-.747 (.746)	-.602 (.711)	-.730 (.622)	-.587 (.464)
Solo, multisyllabic	.344 (.855)	.070 (.810)	.149 (.762)	.265 (.890)
Synchronous, monosyllabic	-.423 (.830)	-.674 (.674)	-.623 (.746)	-.557 (.728)
Synchronous, multisyllabic	.083 (.771)	.246 (.753)	.222 (.803)	.020 (.935)



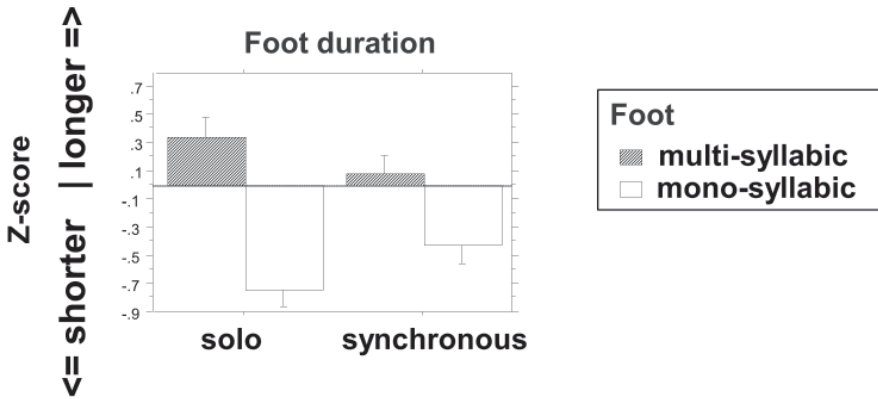


Fig. 6: Foot duration for Indian English speaker 1.

mono- and multisyllabic feet is larger in the solo than in the synchronous condition. The remaining three speakers were pooled for group analysis (since IE1 is the only speaker with a significant interaction, IE1 was not included in the group analysis). The results (see Figure 5) show an effect of the number of syllables in the foot ( $F(1, 496) = 137.134, p < .0001$ ), such that with an increase of number of syllables in the foot the duration of the foot increases. These results show a pattern consistent with syllable-timing. Note also that this pattern is more pronounced for Indian English than for American English. As indicated by the black arrows in Figure 5, the difference between mono- and multisyllabic foot duration in the solo condition is qualitatively stronger in Indian English (the difference, in z-scores, is .825) compared to American English (.569).

### 3.3 Speech rate

One notable effect is that all American English speakers were slower in the synchronous compared to the solo condition. Kim and Nam (2009), in their work on Mandarin Chinese, and O'Dell, Nieminen, and Mustanoja (2010) for Finnish show that speakers read slower in the synchronous compared to the solo condition. The slowing down thus might be an effect of the experimental paradigm used in this study, although the fact that it only occurred in American English speakers argues against that view. To examine speech rate further, we compare the difference in non-normalized means between speakers of a dyad (given in Tables 5 and 6). We examine the difference in absolute values, i.e., ignoring which speaker is faster. The comparison shows that in all but one case for Dyad 3 (given in italics,

the monosyllabic synchronous condition) the American English speakers' means are closer to those of their co-speakers in the synchronous condition than in the solo condition. We take this to indicate a degree of convergence in speaking rate.

**Table 5:** Average foot duration (ms) with standard deviations (in parenthesis) for each condition, given by speakers in a dyad. The difference between dyads is given as an absolute value. Italics indicate that the difference is larger in the synchronous than in the solo condition.

	Dyad 1		Dyad diff.	Dyad 2		Dyad diff.	Dyad 3		Dyad diff.	Dyad 4		Dyad diff.
	AE1	IE1		AE2	IE2		AE3	IE3		AE4	IE4	
Solo, mono	239 (82)	281 (94)	42	244 (69)	262 (80)	18	245 (88)	273 (76)	28	231 (82)	287 (56)	56
Solo, multi	350 (95)	419 (112)	69	308 (84)	337 (91)	29	296 (88)	380 (93)	84	279 (81)	390 (108)	111
Synch, mono	308 (114)	322 (105)	14	269 (74)	254 (75)	15	324 (100)	286 (91)	<i>38</i>	274 (83)	290 (88)	16
Synch, multi	425 (125)	386 (98)	39	346 (88)	357 (84)	11	412 (131)	389 (98)	23	374 (120)	360 (113)	14

**Table 6:** Average stressed syllable duration (ms) with standard deviations (in parenthesis) for each condition, given by speakers in a dyad. The difference between dyads is given as an absolute value. Italics indicate that the difference is larger in the synchronous than in the solo condition.

	Dyad 1		Dyad diff.	Dyad 2		Dyad diff.	Dyad 3		Dyad diff.	Dyad 4		Dyad diff.
	AE1	IE1		AE2	IE2		AE3	IE3		AE4	IE4	
Solo, mono	239 (82)	281 (94)	42	244 (69)	262 (80)	18	245 (88)	273 (76)	28	231 (82)	287 (56)	56
Solo, multi	199 (51)	236 (75)	37	171 (54)	196 (66)	25	172 (63)	213 (57)	41	162 (46)	225 (79)	63
Synch, mono	308 (114)	322 (105)	14	269 (74)	254 (75)	15	324 (100)	286 (91)	<i>38</i>	274 (83)	290 (88)	16
Synch, multi	244 (90)	221 (67)	23	189 (57)	204 (57)	15	231 (94)	221 (72)	10	213 (82)	195 (62)	18

## 4 Discussion

In terms of general rhythmic properties, American and Indian English show evidence of both stress- and syllable-timing. The results from the solo speech show that in both American and Indian English stressed syllables are longer in monosyllabic compared to multisyllabic feet. Thus both languages show shortening of the stressed syllable, a property that is indicative of a stress-timing pattern. The results from the solo speech for the duration of the foot for both languages show that it increases with the number of syllables in it; thus both languages show a pattern associated with syllable-timing. However, they show this property to a different degree, in that in Indian English, the foot duration increase seems larger than in American English. The two languages are thus quite similar in their rhythmic properties, although qualitatively the syllable-timed tendencies are stronger in Indian English. This indicates that in Indian English the foot “squeezes” the syllables less than it does in American English.

There is some evidence of convergence. For one speaker of Indian English (IE1), stressed syllables shorten more in the synchronous than in the solo condition. For the same speaker, the duration of the foot also becomes less affected by the number of syllables in the foot in the synchronous compared to the solo condition. Thus this speaker converges towards a more American English stress-timing pattern in both of the examined properties. American English speakers converge in speech rate to Indian English speakers. Furthermore, when pooled, American English speakers show a tendency, although it does not reach significance, for the duration of the foot to be more affected by the number of syllables in it, in the sense that the duration of the foot increases more in the synchronous compared to the solo condition. Thus there is some evidence that the foot squeezes the syllables less, becoming more syllable-timed.

Before examining in more detail how the results of the study can be accounted for, two points need to be raised. The first question that needs to be addressed is whether the observed changes are the result of rhythmic convergence or of speech rate effects. It was mentioned in Section 3.3. that American English speakers converge with Indian English speakers in speech rate by becoming slower in their speech. Thus the observed longer foot duration (Section 3.2.1.) could be an effect of global slowing down of American English speakers, rather than being a sign of rhythmic convergence. However, the fact that there is an interaction between the number of syllables in the foot and speaking condition on the duration of the foot indicates that the effect cannot be solely due to global speech rate changes. Another indicator that the effect is not solely due to global speech rate effects is the good synchronization between the subjects. As was shown in 2.4., subjects synchronized well, and there was no evidence of one sub-

ject trailing another. This suggests that the observed effects are driven by factors other than just global speech rate changes. However, the possibility that the observed changes in American English speakers are due to general speech rate changes cannot be excluded at this point.

For the Indian English speaker the speech rate explanation is not possible: For this speaker, as shown in Figure 4, in synchronous speech the stressed syllable in monosyllabic, but not in multisyllabic, feet is longer than in the solo condition. Moreover, as shown in Figure 6, multisyllabic feet become shorter in synchronous compared to solo speech, whereas monosyllabic feet become longer. The fact that the effects of the synchronous speech condition are not uniform indicates that the observed rhythmic effects cannot be accounted for by global speech rate change.

The second point to be made is that the speaker who showed clear rhythmic convergence (IE1) has been living in the U.S. for the longest period of time. This could be an indicator that familiarity with American rhythmic properties made it easier for this speaker to converge. However, a small study of convergence in segmental properties between speakers of British and American English (Krivokapić 2010) showed exactly the opposite effect, in that the British speakers who arrived recently to the U.S. converged more than speakers who had been in the U.S. longer. It is also well-known that a large number of factors influence if and to what extent convergence occurs (for overviews see Babel and Munson to appear; Lewandowski 2012); thus, the convergence by IE1 might be driven by any number of other factors.

In the remainder of this section we examine how the observed properties of rhythm and rhythmic convergence can be accounted for. The findings regarding the rhythmic properties of the two languages, as established in the solo condition, indicate that the difference between Indian English and American English is in the role of the foot. In the model of Saltzman et al. (2008) the role of the foot oscillator is to keep the duration of the foot constant, while the role of the syllable oscillator is to keep the duration of the syllable constant. The dominance of the foot oscillator in the coupling function leads to a tendency towards isochrony of the foot, and the dominance of the syllable oscillator leads to a tendency towards syllable isochrony. The results for the Indian English speakers in the solo condition suggest an oscillator network in which the foot oscillator is active but exerts less force on the syllable than in American English.

There was only limited evidence of rhythmic convergence. This could be modeled by adapting the coupling force between syllable and foot oscillator for both the American and the Indian English speakers. American English speakers showed a tendency toward foot properties that are more associated with a syllable-timing pattern when paired up with Indian English speakers (if the

observed effect for American speakers is due to rhythmic convergence rather than to global speech rate changes). That is, for American English speakers, the properties of the foot seem to change such that it exhibits less of a syllable-squeezing pattern. In this way, the duration of the foot becomes more dependent on the number of syllables in it, and less isochronous. This is only a tendency in the current experiment, but the prediction is that if this result becomes more prominent, a further property would surface, namely that the stressed syllable shortens less. The opposite pattern is observed for IE1. This speaker exhibits the stronger squeezing pattern associated with American English when paired up with an American English speaker. As a consequence, for IE1 the foot has a more constant duration and the stressed syllable shortens more in the synchronous compared to the solo condition. In both cases the results could be modeled by a change in coupling strengths. For American English, the weight of the foot oscillator in the coupling function would decrease, and for the Indian English speakers it would increase. Thus convergence would be modeled by adapting the coupling strengths so that the foot-to-syllable dominance of the speakers of Indian English and American English become more similar to each other.

This finding introduces the possibility that convergence can be modeled via a mutual tuning between speakers of the *coupling function(s)* that combine individual oscillators (here the foot and syllable oscillator) into a cohesive network. Coupling relations are known to contribute significantly in determining relative timing among articulatory gestures (e.g., Goldstein, Byrd, and Saltzman 2006; Nam, Goldstein, and Saltzman 2009). The work presented here suggests a new avenue for exploring linguistic convergence, which can be viewed not only as the entrainment of particular types of linguistic oscillators between speakers, but also as the convergence of coupling relations among the oscillators that are the primitives of gestural timing.

A final point needs to be made regarding rhythmic classifications of languages. Evidence of a qualitative difference in the temporal properties between the two examined languages is noticeable in that the foot seemed to be less constraining in Indian English than in American English (see Figure 5). This is particularly noteworthy given that the text used is identical, and that, due to the fact that the experiment was conducted in the U.S., the speakers of Indian English might have already adapted somewhat to their linguistic environment. This means that duration is a property well worth examining when investigating rhythm. At the same time, it is well known that durational properties alone will not suffice to characterize rhythm (see discussion in the introduction). There is also evidence for the relevance of  $F_0$  in the perception of rhythm (e.g., Dilley and Shattuck-Hufnagel 1999; Dilley and McAuley 2008; Barry, Andreeva, and Koreman 2009). Furthermore, recent work has shown evidence suggesting coordina-

tion between prosodic boundaries, in that boundaries within an utterance affect each other (e.g., Schafer 1997; Carlson, Clifton, and Frazier 2001; Clifton, Carlson, and Frazier 2002; Jun 2003; Frazier, Clifton, and Carlson 2004; Krivokapić 2007a). Such coordination could give rise to a rhythmic structure at a higher prosodic level (as suggested in Krivokapić 2007a; see also Jun [2012] for a suggested prosodic macro-rhythmic typology based on tonal properties of pitch accents and prosodic boundaries). The challenge for future work will be to understand how temporal and  $F_0$  properties combine to create rhythm and to examine if and how rhythm arises from larger prosodic units.

To summarize, the study finds evidence of stress-timing and syllable-timing in both Indian and American English, with Indian English showing qualitatively stronger properties of syllable-timing than American English. There is evidence of convergence for one speaker of Indian English towards a more stress-timed pattern and a tendency towards convergence towards a more syllable-timed pattern for American English speakers, although the results for the American speakers could possibly also be interpreted as a global speech rate effect. We suggested that the difference between the two languages lies in the squeezing properties of the foot, and that convergence can be understood as arising from the tuning of the prosodic interoscillator coupling function. The findings demonstrate the relevance of linguistic timing when exploring rhythmic properties of languages. They further suggest a novel way to think about convergence, namely in terms of entrainment of not only linguistic oscillators but also of their coupling functions.

## Acknowledgments

I would like to thank Dani Byrd, Jessica Hsieh, Maria Kouneli, Christine Mooshammer, Hosung Nam, Susanne Fuchs, Radhika Koul, Venkatesh Upadhayay, Stefano Vegnaduzzo, Chandra Sharma, Ashwini Deo, the participants of the study, and the Yale Phonetics Laboratory for their help with various parts of the project. I would also like to thank two anonymous reviewers whose comments significantly improved the manuscript.

## References

- Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Arvaniti, Amalia. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66. 46–63.
- Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40. 351–373.

- Auer, Peter, Elizabeth Couper-Kuhlen, & Frank Müller. 1999. *Language in time. The Rhythm and Tempo of Spoken Interaction*. Oxford: Oxford University Press.
- Babel, Molly. 2009. Phonetic and social selectivity in phonetic accommodation. Ph.D. dissertation, University of California, Berkeley, CA.
- Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40. 177–189.
- Babel, Molly, & Benjamin Munson. To appear. Producing socially meaningful linguistic variation. In Victor Ferreira, Matt Goldrick, & Michele Miozzo (eds.), *The Oxford Handbook of Language Production*, 1–48.
- Barry, William J., Bistra Andreeva, & Jacques Koreman. 2009. Do rhythm measures reflect perceived rhythm? *Phonetica* 66. 78–94.
- Beckman, Mary E., & Gayle Ayers Elam. 1997. Guidelines for ToBI labelling, version 3.0. Unpublished manuscript. (available online at: [http://www.ling.ohiostate.edu/~tobi/ame\\_tobi/labelling\\_guide\\_v3.pdf](http://www.ling.ohiostate.edu/~tobi/ame_tobi/labelling_guide_v3.pdf)).
- Beňuš, Štefan, Agustín Gravano, & Julia Hirschberg. 2011. Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics* 43. 3001–3027.
- Boersma, Paul, & David Weenink. 2012. Praat: doing phonetics by computer [Computer program]. Version 5.3.18, last retrieved 20 June 2012 from <http://www.praat.org/>.
- Bolinger, Dwight L. 1965. Pitch accent and sentence rhythm. In Isamu Abe & Tetsuya Kanekiyo (eds.), *Forms of English: Accent, Morpheme, Order*, 139–180. Cambridge MA: Harvard University Press.
- Bull, Matthew, & Matthew Aylett. 1998. An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. In Robert H. Mannell, & J. Robert-Ribes (eds.), *Proceedings of International Conference on Spoken Language Processing 98*, Vol. 4, 1175–1178. Australia, Australian Speech Science and Technology Association (ASSTA), Sydney.
- Byrd, Dani, & Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31. 149–180.
- Carlson, Katy, Charles Clifton, Jr., & Lyn Frazier. 2001. Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45. 58–81.
- Clifton, Jr., Charles, Katy Carlson, & Lyn Frazier. 2002. Informative prosodic boundaries. *Language and Speech* 45. 87–114.
- Couper-Kuhlen, Elizabeth. 1993. *English Speech Rhythm. Form and Function in Everyday Verbal Interaction*. Amsterdam: John Benjamins.
- Crystal, David. 1994. Documenting rhythmical change. In Jack Windsor Lewis (ed.), *Studies in General and English Phonetics*, 174–179. London: Routledge.
- Cummins, Fred. 2002. On synchronous speech. *Acoustic Research Letter Online*. 3. 7–11.
- Cummins, Fred. 2003. Practice and performance in speech produced synchronously. *Journal of Phonetics* 31. 139–148.
- Cummins, Fred. 2004. Synchronization among speakers reduces macroscopic temporal variability. In Kenneth Forbus, Dedre Gentner, & Terry Regier (eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 304–309.
- Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11. 51–69.
- Dauer, R. M. 1987. Phonetics and phonological components of language rhythm. *Proceedings of the XIth International Congress of Phonetic Sciences*, Vol. 5, 447–450. Tallinn, Estonia.
- Dilley, Laura, & J. Devin McAuley. 2008. Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language* 59. 294–311.

- Dilley, Laura, & Stefanie Shattuck-Hufnagel. 1999. Effects of repeated intonation patterns on perceived word-level organization. *Proceedings of the 14th International Congress of Phonetic Sciences*, Vol. 1, 1487–1490, San Francisco, USA.
- Frazier, Lyn, Charles Clifton, Jr., & Katy Carlson. 2004. Don't break, or do: prosodic boundary preferences. *Lingua* 114. 3–27.
- Fuchs, Robert. 2012. A duration-based account of speech rhythm in Indian English. Poster presented at *Laboratory Phonology 13*, Stuttgart, Germany.
- Garrod, Simon, & Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27. 181–218.
- Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Goldstein, Louis, Dani Byrd, & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael Arbib (ed.), *Action to Language via the Mirror Neuron System*, 215–249. New York: Cambridge University Press.
- Grabe, Esther, & Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. In Carlos Gussenhoven, & Natasha Warner (eds.), *Laboratory Phonology 7*, 515–546. Berlin: Mouton de Gruyter.
- Honorof, Douglas, Jill McCullough, & Barbara Somerville. 2000. Comma gets a cure. Available at: <http://web.ku.edu/~idea/readings/comma.htm>
- Huggins, A. W. F. 1972. On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America* 51. 1279–1290.
- Jun, Sun-Ah. 2003. Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research* 32. 219–249.
- Jun, Sun-Ah. 2012. Prosodic typology revisited: Adding macro-rhythm. *Proceedings of the 6th International Conference on Speech Prosody 2012*, Shanghai, China.
- Kim, Heejin, & Jennifer Cole. 2005. The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. *Proceedings of Interspeech 2005*, 2365–2368. Lisbon, Portugal.
- Kim, Miran, & Hosung Nam. 2009. Pitch accommodation in synchronous speech. *Journal of the Acoustical Society of America* 125. 2575.
- Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54. 1102–1104.
- Klatt, Dennis H. 1976. Linguistics uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59. 1208–1221.
- Krivokapić, Jelena. 2007a. The planning, production, and perception of prosodic structure. Ph.D. dissertation, University of Southern California, Los Angeles, CA.
- Krivokapić, Jelena. 2007b. Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics* 35. 162–179.
- Krivokapić, Jelena. 2010. Prosodic interaction between speakers of American and British English. *Journal of the Acoustical Society of America* 127. 1851.
- Krivokapić, Jelena. 2011. Prosodic and segmental conversion between speakers of different dialects. *Journal of the Acoustical Society of America* 129. 2658.
- Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51. 2018–2024.
- Lehiste, Ilse. 1977. Isochrony reconsidered. *Journal of Phonetics* 5. 253–263.
- Levitan, Rivka, & Julia Hirschberg. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech 2011*, Florence, Italy.



- Lewandowski, Natalie. 2012. Individual differences in convergence in native-nonnative dialogs. Ph.D. dissertation, University of Stuttgart.
- Nakatani, Lloyd H., Kathleen O'Connor, & Carletta H. Aston. 1981. Prosodic aspects of American English speech rhythm. *Phonetica* 38. 84–106.
- Nam Hosung, Louis Goldstein, & Elliot Saltzman. 2009. Self-organization of syllable structure: a coupled oscillator model. In François Pellegrino, Egidio Marisco, & Ioana Chitoran (eds.), *Approaches to Phonological Complexity*, 299–328. Berlin/New York: Mouton de Gruyter.
- Nam, Hosung, & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. *Proceedings of the XIth International Congress of Phonetic Sciences*, 2253–2256. Barcelona, Spain.
- Nam, Hosung, Elliot Saltzman, Jelena Krivokapić, & Louis Goldstein. 2008. Modeling the durational difference of stressed vs. unstressed syllables. *Proceedings of the 8th Phonetic Conference of China (PCC 2008)*, Beijing, China.
- Natale, Michael. 1975. Convergence of mean vocal intensity in dyadic communications as a function of social desirability. *Journal of Personality and Social Psychology* 32. 790–804.
- Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39. 132–142.
- O'Dell, Michael L., & Tommi Nieminen. 1999. Coupled oscillator model of speech rhythm. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville, & Ashlee C. Bailey (eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1075–1078.
- O'Dell, Michael L., Tommi Nieminen, & Liisa Mustanoja. 2010. Assessing rhythmic differences with synchronous speech. *Proceedings of Speech Prosody 2010*, 100141:1–4.
- Pardo, Jennifer. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119. 2382–2393.
- Pickering, Martin J., & Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavior & Brain Sciences* 27. 169–226.
- Pike, Kenneth. 1945. *The Intonation of American English*. University of Michigan Publications in Linguistics 1. Ann Arbor: University of Michigan Press.
- Port, Robert. 1981. Linguistics timing factors in combination. *Journal of the Acoustical Society of America* 69. 262–274.
- Prabhakar Babu, B. A. 1971. Prosodic features in Indian English: stress, rhythm and intonation. *CIEFL Bulletin* 8. 33–39.
- Rakerd, Brad, Sennett, William, & Carol A. Fowler. 1987. Domain-final lengthening and foot level shortening in spoken English. *Phonetica* 44. 147–155.
- Ramus, Franck, Marina Nespou, & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73. 265–292.
- Roach, Peter, James Hartman, Jane Setter, & David Jones. 2006. *The Cambridge English Pronouncing Dictionary* (17<sup>th</sup> ed.). Cambridge: Cambridge University Press.
- Sailaja, Pingali. 2009. *Indian English*. Edinburgh: Edinburgh University Press.
- Saltzman, Elliot, & Dani Byrd. 2000. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19. 499–526.
- Saltzman, Elliot L., & Kevin G. Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1. 333–382.
- Saltzman, Elliot L., Hosung Nam, Jelena Krivokapić, & Louis Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proceedings of the 4th International Conference on Speech Prosody*, 175–184. Campinas, Brazil.

- Sancier, Michele L., & Carol A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25. 421–436.
- Schafer, Amy J. 1997. Prosodic parsing: The role of prosody in sentence comprehension. Ph.D. dissertation, University of Massachusetts, Amherst, MA.
- Selkirk, Elisabeth. 1984. *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA: MIT Press.
- Shattuck-Hufnagel, Stefanie A., & Alice Turk. 2011. Durational evidence for word-based vs. prominence-based constituent structure in limerick speech. *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Sciences, 1806–1809*. Hong Kong, China.
- Street, Richard L. 1984. Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research* 11. 149–169.
- Szcepek Reed, Beatrice. 2010. Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics* 42. 1037–1059.
- Trudgill, Peter, & Jean Hannah. 2008. *International English: A Guide to Varieties of Standard English* (5<sup>th</sup> ed.). London: Hodder Education.
- Turk, Alice, Satsuki Nakai, & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: a practical guide. In Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter, & Johannes Schließer (eds.), *Methods in Empirical Prosody Research*, 1–28. Berlin, New York: De Gruyter.
- Turk, Alice, & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related durational patterns in English. *Journal of Phonetics* 28. 397–440.
- Turk, Alice, & Stefanie Shattuck-Hufnagel. 2007. Phrase-final lengthening in American English. *Journal of Phonetics* 35. 445–472.
- Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.
- White, Laurence, & Alice E. Turk. 2010. English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics* 38. 459–471.
- Wilson, Margaret, & Thomas P. Wilson. 2005. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review* 12. 957–968.
- Włodarczak, Marcin, Juraj Šimko, & Petra Wagner. 2012. Temporal entrainment in overlapped speech: Cross-linguistic study. *Proceedings of Interspeech 2012*, Portland, OR.
- Włodarczak, Marcin, Juraj Šimko, Petra Wagner, Michael O'Dell, Mietta Lennes, & Tommi Nieminen. In press. Finnish rhythmic structure and entrainment in overlapped speech. In *Nordic Prosody. Proceedings of XIth Conference, Tartu 2012*. Frankfurt am Main: Peter Lang.
- Zvonik, Elena, & Fred Cummins. 2002. Pause duration and variability in read texts. In *Proceedings of the 2002 International Conference on Spoken Language Processing (ICSLP '02)*, 1109–1112. Denver, Colorado.
- Zvonik, Elena, & Fred Cummins. 2003. The effect of surrounding phrase lengths on pause duration. In *Proceedings of EUROSPEECH 2003*, 777–780. Geneva, Switzerland.

## Appendix

Stimuli (constructed based in part on Honorof et al. [2000] and using Wells' [1982] lexical sets).

The nurse took a cold bath when she woke up. Then she put on a plain yellow dress and a fleece jacket, picked up a pear and her goose Nico, and headed north to work. But a strange thing happened when she opened the door: she saw a goat standing near the big garage where she kept her boat and tools. That reminded her of a story where a dictator had a square hat and always talked about caffeine and boats at press conferences. She was a young adult when she first heard the story from a minister. In those days, her favorite meal on weekends was tuna with parsley. That was a long time ago. Last time she ate a tuna sandwich was when her sister bought it as a surprise for her, about a year ago. Nowadays, she ate either hotdogs, or, if it was hot outside, bananas with ice cream.

As she was thinking about the goat and the dictator, rain started falling. Her clothes and her face got all wet, and she hurried to work. She worked with her father, doing research for a company that produced cloth for cleaning crystal. The job, with its busy schedule and interesting projects, suited her, and she was very successful. The laboratory was also surprisingly famous.

She arrived at the lab a few minutes past eight. She had just put her purse on the big oak table, when suddenly, she noticed another strange thing: a spunky cat was sitting next to a large lamp writing a long letter. There was not a single comma, but every word was spelled correctly, without any mistakes. It was a mystery to the nurse how this could be possible. The cat was scratching its back all the while. The nurse considered introducing her goose, but she didn't know the cat's name. She was also not quite sure which verb to use.

She laughed at the thought and started preparing for the day. In the evening, as she was listening to her favorite record, she thought about her day, which had been full of unusual events. She was going to write about it in the local magazine if she could only find their address.

She was hoping that tomorrow she would encounter some other mysterious animal. Maybe a deer in her mother's kitchen? Or a dancing panda? Maybe even a skunk taking a bath?

She went to sleep early. She had had a long, strange day, and by 6 pm, her sleepiness was making her see things (as if she hadn't seen enough today!).



Copyright of Laboratory Phonology is the property of De Gruyter and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.