# Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English

*Robert Fuchs*[1]

[1] Englisches Seminar, Westfälische Wilhelms-Universität Münster

robert.fuchs@uni-muenster.de

## Abstract

Most research on speech rhythm has focussed on duration. For example, [1] suggested the normalised pairwise variability index for vocalic intervals (nPVI-V) in order to measure the variability of vocalic durations. This paper argues that speech rhythm research should also take into account other correlates of prominence as well as their interaction. The duration-based nPVI, or nPVI-V(dur), is supplemented by an nPVI that measures variability in average loudness, nPVI-V(avgLoud). These two metrics account for variability in duration and loudness, but cannot measure if loudness and duration reinforce each other by varying simultaneously in the same direction. This simultaneous variability is accounted for by the combined nPVI-V(dur+avgLoud), which is higher than the average of the other two measures if vocalic intervals that are longer than average are also louder than average. The three metrics are subsequently applied to recordings of a reading task performed by 20 speakers of Indian English (IndE) and 10 speakers of British English (BrE). Results indicate that IndE has less variability in duration and less variability in loudness than BrE. In addition, IndE has less simultaneous variability in duration and loudness than BrE. This indicates that duration and loudness are less often used together as cues to prominence in IndE compared to BrE.

**Index Terms**: speech rhythm, duration, loudness, Indian English, British English

## 1. Introduction

Early definitions of speech rhythm held that languages fall into discrete rhythm classes: English, for example, was said to be stress-timed [2], as opposed to the syllable-timed rhythm of French [3]. However, a more nuanced view of varieties of English around the world resulted in descriptions of only some, mainly more established varieties such as British English (BrE) and American English, as stress-timed. By contrast, many of the younger national varieties of English (such as Indian, Nigerian and Singapore English), which are used in administration, education and as a national link language in their respective countries, have been classified as syllable-timed. This is often explained with transfer from the local syllable-timed languages [4–6].

The categorical view of rhythm with two separate classes of languages has in the last years given way to a gradient analysis. Accordingly, languages can be placed at any point on a continuum between a prototypically stress-timed and a prototypically syllable-timed pole. This analysis is supported by quantifications of rhythm based on the durations of vocalic intervals, consonantal intervals and syllables, and their variability, as suggested by [1,7–15]. One widely used rhythm metric is the normalised pairwise variability index for vocalic intervals, or nPVI-V [1]. It is computed by calculating the mean of the differences between successive vocalic intervals divided by their sum, multiplied by 100:

$$nPVI-V = 100 \times \frac{\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right|}{m-1}; \qquad (1)$$

where $m$ is the number of vocalic intervals

and $d_k$ is the duration of the k$^{th}$ vocalic interval

Some of these rhythm metrics have been shown to predict language discrimination in perception experiments [16]. Comparisons of vocalic and consonantal metrics by [17, 18] suggest that vocalic, speech rate normalised metrics are the most reliable measures of speech rhythm. However, even these metrics are influenced by variation between speakers, texts and transcribers. It is therefore recommended to use phonetically balanced or large samples from a greater number of speakers, and adhere to a clearly defined set of transcription rules. How crucial these suggestions are is demonstrated by [19], who was unable to replicate the results of [17, 18], perhaps because (1) she relied on samples from each language that were not representative but specifically designed to elicit sentences that differed from each other in rhythm as much as possible, and (2) because segmentation criteria and the treatment of hesitations were not specified and possibly not controlled for.

However, an approach that measures rhythm on the basis of durational variability alone may present a lop-sided account of speech rhythm. Most studies on rhythm in the last 15 years have concentrated exclusively on duration, which means that they considered only one correlate of prominence. Exceptions are [20–23], who suggested rhythm metrics based on variability in intensity, $f_0$, and sonority. Together, these metrics may form a multidimensional account of rhythm, as demanded by [24–26]. A comparison of two languages might show that one has less variability in all these correlates of prominence and is therefore more syllable-timed than the second language on all these dimensions. On the other hand, the first language might also turn out to have less variability in intensity, but more in duration than the second language. Any language might therefore have multiple co-existing and different rhythms [26], which contribute in different ways to a succession of elements relatively similar in prominence (syllable-timing) or relatively different in prominence (stress-timing).



10.21437/SpeechProsody.2014-45

## 2. Integrating variability in loudness and duration

While a consideration of multiple correlates of prominence contributes to a multidimensional analysis of rhythm, the metrics suggested so far rely exclusively on acoustic correlates, not perceptual correlates of rhythm. This is particularly important for the relation of intenstiy, an acoustic property, to loudness, a perceptual property. Sounds with equal intensity but different frequencies have been shown to differ in loudness [27, 28]. Therefore, this paper suggests that loudness instead of intensity should be considered as a perceptual correlate of rhythm.

A second shortcoming of a multidimensional analysis of rhythm as outlined above is that different acoustic or perceptual correlates are considered separately. In addition to an individual analysis of these correlates, their interaction should also be taken into account. For example, if stressed vowels are both longer and louder than unstressed vowels, then variability in duration and loudness may be said to reinforce each other. By contrast, if stressed vowels are only longer but not usually louder than unstressed vowels, there is no reinforcement between these two cues of prominence.

In order to derive a measure of loudness, the computer programme Praat can be used [29]. After extracting the relevant vocalic interval from the recording, first its spectrum, then excitation, and finally maximum loudness in Sone can be derived (a Praat script is available from the author upon request). As a measure of variability in loudness, a PVI can be computed by entering these loudness values (instead of duration values) into formula (1). This metric will be called nPVI-V(avgLoud), and the duration measure will be referred to as nPVI-V(dur).

The next step, accounting for simultaneous variability in loudness and duration necessitates a change to the PVI formula. For every pair of successive vocalic intervals, the difference in duration and in loudness are computed separately and divided by their sum. These relative differences in duration and loudness are then added and squared. Squaring this sum of relative differences in duration and loudness gives an advantage (higher values) to simultaneous in- or decreases in both correlates of prominence, and a disadvantage to cases where duration in- but loudness decreases (or the other way around).

$$nPVI-V(dur+avgLoud) = \quad (2)$$
$$100 \times \frac{\sum_{k=1}^{m-1} 2 \times \left( \frac{d_k - d_{k+1}}{d_k + d_{k+1}} + \frac{l_k - l_{k+1}}{l_k + l_{k+1}} \right)^2}{m-1};$$

where $m$ is the number of vocalic intervals,

$d_k$ is the duration of the k$^{th}$ vocalic interval

and $l_k$ is the loudness of the k$^{th}$ vocalic interval

For example, if the second vocalic interval is twice as long and twice as loud as the first, then the square of the differences is $(\frac{1-2}{1+2} + \frac{1-2}{1+2})^2 = 0.444$. By contrast, if the second vocalic interval is twice as long as the first, but equal in loudness to the first, then the square of the differences is lower: $(\frac{1-2}{1+2} + \frac{1-1}{1+1})^2 = 0.111$. And if the second vocalic interval is twice as long as the first, but half as loud as the first, the square of the differences is zero: $(\frac{1-2}{1+2} + \frac{1-0.5}{1+0.5})^2 = 0$. The measure therefore accounts for whether loudness and duration increase (or decrease) simultaneously, thereby reinforcing each other in the generation of prominence, or whether one decreases and the other increases, which causes them to offset each other in the generation of prominence.

## 3. Indian English

IndE is a postcolonial variety of English used mainly in public contexts such as education, administration, business and politics, but also by Indians who travel or reside in a region whose local language they do not speak. English is the primary domestic language for only a small minority, although many others use it at home when discussing topics belonging to the public domain, as for example when a parent asks their child what happened at school that day. Around 23 % of the population of India have at least basic knowledge of English, and 4 % are fluent [30]. Based on the 2011 census [31], this means there are 50 million fluent speakers.

Although standard IndE still lacks full official recognition, it is the de facto standard taught in schools and universities is standard IndE [32, 33]. Nevertheless, this standard has not yet been fully codified, and in such a context the kind of language used by educated speakers can be used as a yardstick for what is considered acceptable by the speech community [34–36]. The phonology of IndE differs in several respects from that of BrE [37–42], likely due to transfer from Indian languages [43]. While they belong to several different language families, these languages have converged over time in several respects and form a *sprachbund* [44–46].

IndE has been suggested to be syllable-timed or more syllable-timed than BrE [32, 43, 47–52]. On the basis of these descriptions, the present paper hypothesises that IndE has lower variability of vocalic durations, lower variability in average loudness in vowels, and lower simultaneous variability in duration and loudness in IndE compared to BrE. This suggestes that nPVI-V(dur), nPVI-V(avgLoud) and nPVI-V(dur+avgLoud) are all smaller in the IndE group than in the BrE group. Additionally, it is expected that duration and loudness as correlates of prominence also reinforce each other less often in IndE than in BrE, even when lower levels of variability in loudness and duration (considered separately) are taken account of. This suggests that the difference between the group means for IndE and BrE in nPVI-V(dur+avgLoud) is greater than both the differences between the group means in nPVI-V(dur) and nPVI-V(avgLoud).

## 4. Data

Recordings of a text read by 10 speakers or Standard Southern BrE and 20 speakers of IndE were used. The BrE data was taken from the DyViS database [53]. The IndE speakers were recorded by the present author reading the same text. All speakers were university students at the time of recording. The IndE speakers were equallly divided between four different L1 groups, and had either Hindi or Bengali (both Indo-Aryan languages), or Telugu or Malayalam (both Dravidian languages) as first languages. L1 was determined on the basis of a sociolinguistic interview involving questions on when speakers first started using what language. They were between 20 and 28 years of age, and, with the exception of one speaker each, had exclusively attended English-medium schools and universities, and had not resided outside of India.[1]

[1]Only one speaker did not attend English-medium schools throughout, but went to a Telugu-medium primary and English-medium secondary school. Another speaker spent several years in an Arab country, but had only interaction with the local South Asian expat community in

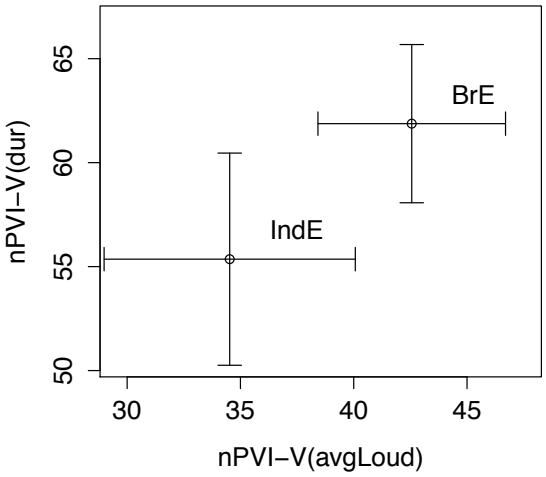


Figure 1: Group averages and standard deviations for variability in average loudness (x-axis) and in duration (y-axis).

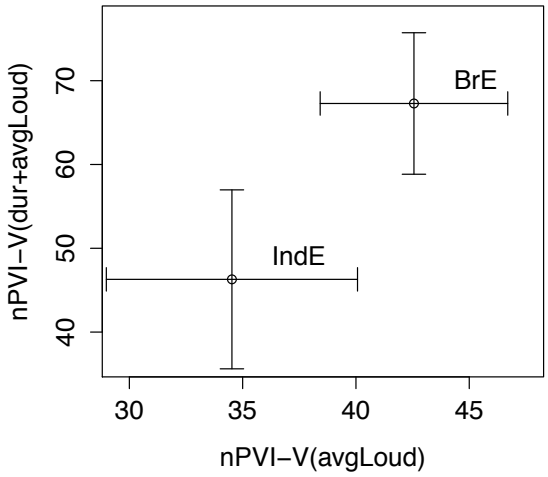


Figure 2: Group averages and standard deviations for variability in average loudness (x-axis) and simultaneous variability in duration and average loudness (y-axis).

Two thirds of the reading passage, or 392 words, were segmented according to the recommendations provided by [17, 18, 54]. nPVI-V(dur), nPVI-V(avgLoud) and nPVI-V(dur+avgLoud) were computed individually for utterances comprising at least three vocalic intervals, excluding the final interval. Next, for each individual speaker, the median of the rhythm scores of all utterances produced by this speaker was computed.

## 5. Results

Variability in duration was significantly lower in the IndE group (55.4) than in the BrE group (61.9, $p<0.0001$, t-test). Average variability in duration was 10.5 % lower in IndE than in BrE. Variability in average loudness was also signifcantly lower in the IndE group (34.5) than in the BrE group (42.7, $p<0.0001$). Average variability in loudness was 19.0 % lower in IndE than in BrE. The average values for variability in loudness and duration for both groups are shown in figure 1.

Scores for individual speakers are shown in figure 3, with crosses for IndE and circles for BrE speakers. Using both metrics, separation between the two groups is relatively good, although there is some overlap between the two groups. Two BrE speakers have less variability in duration and loudness (nPVI-V(dur) and nPVI-V(avgLoud)) than the other BrE speakers and are close to some of the IndE speakers. Likewise, there are two IndE speakers who have high values for variability in duration and loudness so that they are similar in this respect to the BrE group.

Combined variability in duration and average loudness was significantly lower in IndE (46.3) than in BrE (67.3, $p<0.0001$). The average combined variability in duration and loudness was 31.2 % lower in IndE than in BrE, which means that the difference in combined variability in duration and loudness (nPVI-V(dur+avgLoud)) was higher than the variability in either variability in duration (nPVI-V(dur)) or average loudness (nPVI-V(avgLoud)). The average values for variability in loudness and simultaneous variability in duration and loudness for both groups are shown in figure 2.

Scores for individual speakers are shown in figure 4. Considering simultaneous variability in loudness and duration, there are again two IndE speakers (crosses) with values for variability similar to the BrE group. There is also one BrE speakers (circles) with a variability slightly lower than the IndE speakers with the highest variability.

## 6. Discussion

The results have shown that IndE has

- significantly less variability in duration
- significantly less variability in loudness and
- significantly less simultaneous variability in duration and loudness

than BrE. Crucially, the difference in simultaneous variability in duration and loudness between IndE and BrE was higher than either the difference in variability in loudness or duration. This result shows that a measure of simultaneous variability in duration and loudness, the nPVI-V(dur+avgLoud) suggested in this paper, captures an important aspect of variability in prominence between successive vocalic intervals. If variability in duration on the one hand, and variability in average loudness on the other hand, were both randomly distributed, one would expect that the difference between IndE and BrE in the simultaneous variability in loudness and duration were equal to the average of the difference between the two varieties in loudness, and in duration (i.e. 14.75 %). However, this was not the case. Instead, the difference in combined variability in loudness and duration turned out to be much higher than the variability in either loudness or duration taken separately. This result also implies that in BrE, vocalic intervals that are longer than average also tend to

her daily routine.

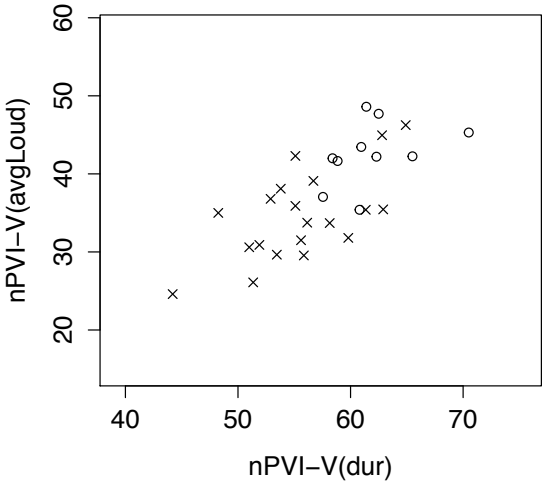


Figure 3: Variability in average loudness (x-axis) and in duration (y-axis) for individual speakers of IndE (crosses) and BrE (circles).

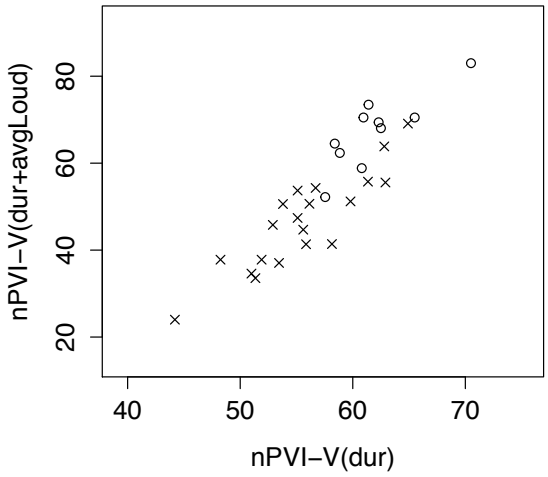


Figure 4: Variability in average loudness (x-axis) and simultaneous variability in duration and average loudness (y-axis) for individual speakers of IndE (crosses) and BrE (circles).

be louder than average, and vice versa. By contrast, in IndE, vocalic intervals that are longer than average are not as frequently and to the same extent also louder than average, and vice versa, as in BrE.

This comparison of variability in duration, loudness and simultaneous variability in loudness and duration also contributes to a better understanding of how speech rhythm is realised in different dimensions, using different acoustic and perceptual correlates of prominence. Duration is but one of these correlates of prominence. Loudness is another, and the simultaneous in- or decrease of loudness and duration is a third, and separate, aspect that needs to be considered within a multidimensional model of speech rhythm.

The results also substantiate previous descriptions of IndE as more syllable-timed than BrE [32,43,47–52]. Such a description, using a relative expression ("more syllable-timed") appears to be adequate since syllable- and stress-timing are probabaly better regarded as poles of a continuum, and perhaps as ideals that are rarely or never realised in actual speech.

Furthermore, a small number of outliers with values for variability similar to the other group occured. This shows that differences in variability in duration, loudness and simultaneous variability in loudness and duration do not constitute a categorical contrast between IndE and BrE. However, the differences are highly significant. Taking into account that a relatively long text of 392 words was used, it is very likely that the results can be generalised to other speakers of BrE and educated IndE. Considering the background of the IndE speakers, who were all engaged in university studies, it is likely that they speak a variety of IndE that is a good example of the emerging standard of IndE. This standard appears to involve a rhythm that is more syllable-timed than BrE speech rhythm.

## 7. References

[1] E. L. Low, E. Grabe, and F. Nolan, "Quantitative characterization of speech rhythm: Syllable-timing in singapore english," *Language and Speech*, vol. 43, no. 4, pp. 377–401, 2000.

[2] K. L. Pike, *The Intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

[3] D. Abercrombie, *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, 1967.

[4] T. Platt, H. Weber, and M. L. Ho, *The New Englishes*. London/ Melbourne: Routledge, 1984.

[5] D. Crystal, "Documenting rhythmical change," in *Studies in General and English Phonetics: Essays in Honour of Professor J D O'Connor*, J. W. Lewis, Ed. London: Routledge, 1995, pp. 174–9.

[6] R. Mesthrie, "Synopsis: the phonology of english in africa and south and southeast asia," in *Varieties of English. Africa, South and Southeast Asia*, R. Mesthrie, Ed. Berlin: de Gruyter, 2008, pp. 307–319.

[7] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, pp. 265–92, 1999.

[8] V. Dellwo, "Rhythm and speech rate: A variation coefficient for deltac," in *Language and Language-Processing. Proceedings of the 38th Linguistics Colloquium*, P. l. Karnowski and I. Szigeti, Eds. Frankfurt am Main: Peter Lang, 2006, pp. 231–241.

[9] D. Gibbon and U. Gut, "Measuring speech rhythm," in *Proceedings of Eurospeech 2001*, Aalborg, Denmark, 2001, pp. 91–94.

[10] D. Deterding, "The rhythm of singapore english," in *Proceedings of the Fifth Australian International Conference on Speech Science and Technology*, R. Togneri, Ed. Perth: Uniprint, 1994, pp. 316–321.

[11] ——, "The measurement of rhythm: a comparison of singapore and british english," *Journal of Phonetics*, vol. 29, pp. 217–230, 2001.

[12] P. Wagner and V. Dellwo, "Introducing yard (yet another rhythm determination) and re-introducing isochrony to rhythm research," in *Proceedings of Speech Prosody 2004*. ISCA, 2004, pp. 227–230.

[13] U. Gut, "Non-native speech rhythm in german," in *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, M.-J. S. Daniel Recasens and J. Romero, Eds. Barcelona: Universitat Autónoma de Barcelona, 2003, pp. 2437–2440.

[14] P. M. Bertinetto and C. Bertini, "On modeling the rhythm of natural languages," in *Proceedings of Speech Prosody 2008, Campinas, Brazil*, P. A. Barbosa, S. Madureira, and C. Reis, Eds. ISCA Archive, 2008, pp. 427–30. [Online]. Available: http://www.isca-speech.org/archive/sp2008

[15] V. Dellwo, A. Fourcin, and E. Abberton, "Rhythmical classification of languages based on voice parameters," in *Proceedings of ICPhS XVI*, J. Trouvain and W. J. Barry, Eds. Dudweiler: Pirrot, 2007, pp. 1129–1132.

[16] L. White, S. L. Mattys, L. Series, and S. Gage, "Rhythm metrics predict rhythmic discrimination," in *Proceedings of the 16th International Congress of Phonetic Sciences*, 2007, pp. 1009–1012.

[17] L. White and S. L. Mattys, "Calibrating rhythm: First language and second language studies," *Journal of Phonetics*, vol. 35, no. 4, pp. 501–522, 2007.

[18] ——, "Rhythmic typology and variation in first and second languages," *Segmental and Prosodic Issues in Romance Phonology*, vol. 282, pp. 237–257, 2007.

[19] A. Arvaniti, "The usefulness of metrics in the quantification of speech rhythm," *Journal of Phonetics*, vol. 40, pp. 351–373, 2012.

[20] E. L. Low, "Prosodic prominence in singapore english," Ph.D. dissertation, University of Cambridge, 1998.

[21] L. He, "Syllabic intensity variations as quantification of speech rhythm: Evidence from both l1 and l2," in *Speech Prosody 2012, 6th International Conference*, Shanghai, 2012. [Online]. Available: http://speechprosody2010.illinois.edu/papers/100039.pdf

[22] R. E. Cumming, "The language-specific integration of pitch and duration," Ph.D. dissertation, University of Cambridge, 2010.

[23] A. Galves, J. Garcia, D. Duarte, and C. Galves, "Sonority as a basis for rhythmic class discrimination," in *Proceedings of Speech Prosody 2002*, 2002, pp. 323–326.

[24] D. Stojanovic, "Issues in the quantitative approach to speech rhythm comparisons," *Working Papers in Linguistics (University of Hawai'i at Mānoa)*, vol. 40, no. 9, pp. 1–20, 2009.

[25] A. Loukina, G. Kochanski, B. Rosner, E. Keane, and C. Shih, "Rhythm measures and dimensions of durational variation in speech," *Journal of the Acoustical Society of America*, vol. 129, no. 5, pp. 3258–3270, 2011.

[26] F. Nolan and E. L. Asu, "The pairwise variability index and coexisting rhythms in language," *Phonetica*, vol. 66, pp. 64–77, 2009.

[27] S. S. Stevens, "The relation of pitch to intensity," *Journal of the Acoustical Society of America*, vol. 6, p. 150, 1935.

[28] M. E. Beckman, *Stress and Non-Stress Accent*. Dordrecht/Riverton: Foris, 1986.

[29] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer (Computer Program). Version 5.3.04*, Std. [Online]. Available: www.praat.org

[30] S. B. Desai, A. Dubey, B. L. Joshi, M. Sen, A. Shariff, and R. Vanneman, *Humand Development in India*. Oxford: Oxford University Press, 2010.

[31] C. Chandramouli, *Census of India 2011. Provisional Population Totals*. New Delhi: Government of India, 2011.

[32] C. P. Masica, *The Sound System of Indian English*. Hyderabad: Central Institute of English and Foreign Languages, 1972.

[33] J. Mukherjee, "Steady states in the evolution of new englishes: Present-day indian english as an equilibrium," *Journal of English Linguistics*, vol. 35, no. 2, pp. 157–187, 2007.

[34] P. Trudgill, "Standard english: What it isn't," in *Standard English: The Widening Debate*. Routledge, 1999, pp. 177–128.

[35] A. Deumert and W. Vandenbusche, "Research directions in the study of language standardization," in *Germanic Standardizations: Past to Present*, A. Deumert and W. Vandenbusche, Eds. Amsterdam: John Benjamins, 2003, pp. 455–467.

[36] J. Edwards, *Language and Identity*. Cambridge: Cambridge University Press, 2009.

[37] O. Maxwell and J. Fletcher, "The acoustic characteristics of diphthongs in indian english," *World Englishes*, vol. 29, no. 1, pp. 27–44, 2010.

[38] ——, "Acoustic and durational properties of indian english vowels," *World Englishes*, vol. 28, no. 1, pp. 52–69, 2009.

[39] O. Maxwell, "Marking of focus in indian english of l1 bengali speakers," in *Proceedings of Speech Sciene and Technology 2010*. Australasian Speech Science and Technology Association, 2010. [Online]. Available: http://assta.org/sst/SST-10/SST2010/PDF/AUTHOR/ST100046.PDF

[40] L. Pickering and C. Wiltshire, "Pitch accent in indian-english teaching discourse," *World Englishes*, vol. 19, no. 2, pp. 173–183, 2000.

[41] C. Wiltshire and R. Moon, "Phonetic stress in indian english vs. american english," *World Englishes*, vol. 22, no. 3, pp. 291–303, 2003.

[42] C. R. Wiltshire and J. D. Harnsberger, "The influence of gujarati and tamil l1s on indian english: a preliminary study," *World Englishes*, vol. 25, no. 1, pp. 91–104, 2006.

[43] P. Sailaja, "Indian english: Features and sociolinguistic aspects," *Language and Linguistics Compass*, vol. 6, no. 6, pp. 359–370, 2012.

[44] M. B. Emeneau, "India as a linguistic area," *Language*, vol. 32, no. 1, pp. 3–16, 1956.

[45] ——, *Language and Linguistic Area*, A. S. Dil, Ed. Stanford: Standford University Press, 1980.

[46] C. P. Masica, *Defining a Linguistic Area*. Chicago, etc.: University of Chicago Press, 1976.

[47] P. Trudgill and J. Hannah, *International English*, 4th ed. London: Arnold, 2002.

[48] R. Gargesh, "Indian english: Phonology," in *A Handbook of Varieties of English*, E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton, Eds. Berlin: Mouton de Gruyter, 2004, vol. 1, pp. 992–1002.

[49] R. Hickey, "South asian englishes," in *Legacies of Colonial English: Studies in Transported Dialects*. Cambridge: Cambridge University Press, 2004, pp. 536–558.

[50] C. Lange, "Review of pingali sailaja *Indian English*," *Annual Review of South Asian Languages and Linguistics*, vol. 3, pp. 213–216, 2009.

[51] P. Sailaja, *Indian English*. Edinburgh: Edinburgh University Press, 2009.

[52] ——, "The standard, (non-)rhoticity and rhythm: A response to lange," *Annual Review of South Asian Languages and Linguistics*, vol. 4, pp. 183–186, 2010.

[53] F. Nolan, K. McDougall, G. de Jong, and T. Hudson, "A forensic phonetic study of dynamic sources of variability in speech: The dyvis project," in *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, P. Warren and C. Watson, Eds., 2006, pp. 13–18.

[54] P. Machač and R. Skarnitzl, *Principles of Phonetic Segmentation*. Prague: Epocha, 2009.