Towards a perceptual model of speech rhythm: Integrating the influence of f0 on perceived duration

Robert Fuchs
University of Münster
robert.fuchs@uni-muenster.de

Interspeech 2014 Singapore

Rhythm and variability

- Stress-timed languages (English, German) vs. syllable-timed languages (Spanish, French)
- Also: varieties - British/American English vs. Indian, Singapore, Nigerian English
- Simple syllable structure and durations of syllables/vowels more similar to each other, and more in stress-timed languages/varieties
- How can durational variability be measured?
  - Standard deviation of durations of vocalic intervals
  - Standard deviation normalised for speech rate
  - Compare consecutive elements (long-short-long-short is more of a difference than long-long-short-short):
    Pairwise variability index' -> PVI
  - Normalised for speech rate: nPVI (Low et al. 2000)

Beyond duration

- Comparison of duration-based rhythm metrics showed nPVI-V is reliable (White and Mattys 2007a,b; White et al. 2007)
- Restrict definition of speech rhythm to duration?
- More inclusive: alternation of prominent and non-prominent units (such as syllables)

Correlates of prominence

- Rhythm metrics based on variability in intensity, f_0 and sonority (Cumming 2010, 2011; Fuchs 2014; Galves et al. 2002; He 2012; Low 1998)
- Different co-existing rhythms in a language $\rightarrow$ multi-dimensional model of speech rhythm
- How do different correlates of prominence interact?

Aims

- Quantify the influence of differences in mean f_0 on perceived duration (in contrast to acoustic duration)
- Integrate this influence into a quantitative measure of speech rhythm, by modifying the nPVI-V formula
- Apply the new nPVI-V(dur*f0) to data from (stress-timed) BrE and (more syllable-timed) Indian English (IndE), in order to determine whether the effect of differences in mean f_0 on perceived duration contributes to rhythm differences between the varieties

Influence of f_0 on speech rhythm

- Interaction of variability in duration and fundamental frequency:
- (1) Both duration and f_0 are, independently from each other, prominence-lending (Cumming 2010, 2011)
- (2) In addition, f_0 may have some influence on perceived duration

f_0 and perceived duration

- Tone height: Two syllables of the same duration, a syllable with a high tone may be perceived as longer than one with a low tone
- Slope size: A syllable with a large rise as longer than one with a small rise (vocalic intervals with a more extreme rise or drop in f_0 are perceived as longer) on perceived duration (Lehiste 1976; Lehnert-LeHouillier 2007; Pisoni 2005; Rosen 1976; Wang et al. 1976; Yu 2007)
- Aim: Quantify the influence of tone height

nPVI-V(dur*f0)

$$nPVI-V(dur*f0) = 100 \times \frac{\sum_{k=1}^{m-1} \left| \frac{d'_k - d'_{k+1}}{(d'_k + d'_{k+1})/2} \right|}{m-1}; \quad (1)$$

where m is the number of vocalic intervals

and d'_k is the perceived duration of the kth vocalic interval

$$d'_k = d_k + d_k \times p \times (f_k - f_{k+1}); \quad (2)$$
$$d'_{k+1} = d_{k+1} + d_{k+1} \times p \times (f_{k+1} - f_{k1});$$

where d_k is the acoustic and d'_k the perceived duration, and

f_k the mean fundamental frequency

of the kth vocalic interval, and

p a constant that adjusts for the increase in perceived duration

Perception experiment

- 2AFC experiment: utterances consisting of the syllable /pɑp/, generated with the en1 MBROLA voice (Dutoit et al. 1996)
- Vowel of the first syllable always 200 ms, vowel of the second syllable was varied in 18 steps from 40 to 300 ms.
- f_0 was varied in both syllables on three levels (85, 115, and 145 Hz), resulting in 6 f_0 combinations x 18 durations = 108 different stimulus pairs
- Participants: 31 native German speakers (21 f., 10 m., median age 24, range 19–62)

Results

- Binomial regression: DIFFERENCE_DUR and DIFFERENCE_F0 (both $p<0.0001$) had a significant effect on the perception of the second syllable as longer or shorter.
- Second syllable had same duration as the first, and f_0 60 Hz higher: chance of 59.3% to be classified as longer than the first.
- On average, an f_0 difference of 100 Hz (second syllable 100 Hz higher) adds 13 ms (6.5 %) in perceived duration $\rightarrow$ factor can be integrated into nPVI-V formula

Indian English

- Apply new nPVI-V(dur*f0) to data from Indian English (IndE) and British English (BrE)
- IndE is a postcolonial variety of English often spoken as a second language, its phonology differs in many respects from BrE (Sailaja 2012)
- Educated IndE is relatively homogeneous, independently of L1 (Sirsa and Redford 2013)
- Educated IndE is more syllable-timed than BrE, based on various acoustic correlates of prominence (Fuchs 2013, 2014)

Research question

- Does the influence of f_0 on perceived duration contribute to rhythm differences between IndE and BrE?
- This could be the case, if, in BrE, vocalic intervals with a long acoustic duration are also marked by high f_0, further increasing their perceived duration.

Reasoning

- Seems plausible: Focussed syllables (which have a longer acoustic duration than unfocussed syllables) often marked by an H*(L) pitch accent in BrE (Nolan 2006)
- Speakers of IndE tend to use both L*(H) and H*(L) pitch accents for focussed syllables (Maxwell 2010, 2014) $\rightarrow$ perceived duration of focussed syllables sometimes longer and sometimes shorter than acoustic duration

Hypothesis

- In BrE, variability in perceived duration is higher than in acoustic duration, while they are similar in IndE
- This would cause the difference in speech rhythm between the varieties to be greater when measured by perceived duration (nPVI-v(dur*f0)) than by acoustic duration (nPVI-V(dur)).

Data

- BrE: 10 speakers from the 'Dynamic variability in Speech' project (Nolan et al. 2006)
- Male speakers from Southern England
- IndE: 20 speakers, university students, went to English-medium schools
- L1: Hindi, Bengali (Indo-European), Telugu or Malayalam (Dravidian)
- Police interview yielded spontaneous yet similar data and detracted attention from monitoring one's speech.

Methods

- Text: appr. 400 words; Interview: at least five minutes
- Phonemic forced alignment and manual correction of phoneme boundaries
- Compute metrics for each utterance in R

Results

- IndE has virtually identical variability in acoustic and perceived duration (nPVI-V(dur) and nPVI-V(dur*f0), both 55.6)
- BrE has significantly more variability in perceived duration (64.0) than in acoustic duration (61.9; paired t-test $p<0.05$).
- Difference in variability between IndE and BrE is higher in perceived duration (difference between means 8.4) than in acoustic duration (difference between means 6.4)
- Hypothesis confirmed

Summary

- Quantify (by means of a perception experiment) the influence of mean f_0 on perceived duration
- Integrate it into the measurement of speech rhythm
- Case study showed how the new nPVI-V(dur*f0) can be applied
- Stress-timed rhythm of BrE is further reinforced by these perceptual effects
- By contrast, in educated IndE, the variability of perceived durations is similar to the variability of acoustic durations

Conclusions

- Results support an analysis of speech rhythm as an, at least partially, perceptual phenomenon (Couper-Kuhlen 1993).
- Influence of f_0 on perceived duration needs to be taken into account in the measurement of speech rhythm.
- Future research should investigate how other factors, such as the influence of dynamic f_0 being perceived as longer than level f_0, contribute to the perception of speech rhythm.
- This would further contribute to a multi-dimensional model of speech rhythm that takes into account various acoustic correlates of rhythm as well as their interaction.

References I

- Couper-Kuhlen, Elizabeth (1993). *English Speech Rhythm. Form and Function in Everyday Verbal Interaction*. Amsterdam: Benjamins.
- Cumming, Ruth E. (2010). "The Language-specific Integration of Pitch and Duration". PhD thesis. University of Cambridge.
- – (2011). "Perceptually informed quantification of speech rhythm in Pairwise Variability Indices". In: *Phonetica* 68.4, pp. 256–277.
- Dutoit, Thierry, Vincent Pagel, Nicolas Pierret, François Bataille, and Olivier van der Vrecken (1996). "The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes". In: *Proceedings of ICSLP 96 - Fourth International Conference on Spoken Language Processing*. Vol. 3, pp. 1393–1396.
- Fuchs, Robert (2013). "Speech Rhythm in Educated Indian English and British English". (available upon request from the author). PhD thesis. University of Münster.

References II

Fuchs, Robert (2014). “Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from Indian English and British English”. In: *Proceedings of Speech Prosody 7, Dublin*. Ed. by Nick Campbell, Dafydd Gibbon, and Daniel Hirst, pp. 290–294.

Galves, Antonio, Jesus Garcia, Denise Duarte, and Charlotte Galves (2002). “Sonority as a basis for rhythmic class discrimination”. In: *Proceedings of Speech Prosody 2002*, pp. 323–326.

He, Lei (2012). “Syllabic intensity variations as quantification of speech rhythm: Evidence from both L1 and L2”. In: *Speech Prosody 2012, 6th International Conference*. Shanghai.

Lehiste, Ilse (1976). “Influence of fundamental frequency pattern on the perception of duration”. In: *Journal of Phonetics* 4, pp. 113–117.

References III

Lehnert-LeHouillier, Heike (2007). "The influence of dynamic f0 on the perception of vowel duration: cross-linguistic evidence". In: *Proceedings of ICPhS XVI*. Ed. by Jürgen Trouvain and William J. Barry. Dudweiler: Pirrot, pp. 757–760.

Low, Ee Ling (1998). "Prosodic Prominence in Singapore English". PhD thesis. University of Cambridge.

Low, Ee Ling, Esther Grabe, and Francis Nolan (2000). "Quantitative characterization of speech rhythm: Syllable-timing in Singapore English". In: *Language and Speech* 43.4, pp. 377–401.

Maxwell, Olga (2010). "Marking of focus in Indian English of L1 Bengali speakers". In: *Proceedings of Speech Sciene and Technology 2010*. Australasian Speech Science and Technology Association.

References IV

Maxwell, Olga (2014). “The Intonational Phonology of Indian English. An Autosegmental-Metrical Analysis Based on Bengali and Kannada English”. PhD thesis. University of Melbourne.

Nolan, Francis (2006). “Intonation”. In: *The Handbook of English Linguistics*. Ed. by Bas Aarts and April McMahon. Malden etc.: Blackwell, pp. 433–458.

Nolan, Francis, Kirsty McDougall, Gia de Jong, and Toby Hudson (2006). “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*. Ed. by P. Warren and C.I. Watson, pp. 13–18.

Pisoni, David B. (2005). “Fundamental frequency and perceived vowel duration”. In: *Journal of the Acoustical Society of America* 59.S1, S39.

References V

Rosen, Stuart M. (1976). “Linear FM pitch sweeps and perceptual duration in speech and non-speech”. In: *Speech Transmission Laboratory Quarterly Progress and Status Reports* 17.4, pp. 1–12.

Sailaja, Pingali (2012). “Indian English: Features and sociolinguistic aspects”. In: *Language and Linguistics Compass* 6.6, pp. 359–370.

Sirsa, Hema and Melissa A. Redford (2013). “The effects of native language on Indian English sounds and timing patterns”. In: *Journal of Phonetics* 41.6, pp. 393–406.

Wang, William, Ilse Lehiste, Chin-Kuang Chuang, and Nancy Darnovsky (1976). “Perception of vowel duration”. In: *Journal of the Acoustical Society of America* 60.S1, S92.

White, Laurence and Sven L. Mattys (2007a). “Calibrating rhythm: First language and second language studies”. In: *Journal of Phonetics* 35.4, pp. 501–522.

References VI

White, Laurence and Sven L. Mattys (2007b). “Rhythmic typology and variation in first and second languages”. In: *Segmental and Prosodic Issues in Romance Phonology* 282, pp. 237–257.

White, Laurence, Sven L. Mattys, Lucy Series, and Suzi Gage (2007). “Rhythm metrics predict rhythmic discrimination”. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 1009–1012.

Yu, Alan C. L. (2007). “Tonal effects on perceived vowel duration”. In: *Laboratory Phonology 10*. Berlin: Mouton de Gruyter.