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
1 Rhythm metrics
2 Perception study
3 Production study
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)
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)
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 → multi-dimensional model of speech rhythm

How do different correlates of prominence interact?
Aims

- Quantify the influence of differences in mean f₀ 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 f₀ 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 \( k^{th} \) vocalic interval

\[
d'_k = d_k + d_k \times p \times (f_k - f_{k+1}); \quad (2)
\]

\[
d'_{k+1} = d_{k+1} + d_{k+1} \times p \times (f_{k+1} - f_k);
\]

where \( d_k \) is the acoustic and \( d'_k \) the perceived duration, and
\( f_k \) the mean fundamental frequency
of the \( k^{th} \) vocalic interval, and
\( p \) a constant that adjusts for the increase in perceived duration
2AFC experiment: utterances consisting of the syllable /pap/, 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
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 (Sirsia 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.
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) → perceived duration of focussed syllables sometimes longer and sometimes shorter than acoustic duration
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
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
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.


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.


References IV


References V


