

Parsing speech for grouping and prominence, and the typology of rhythm

Michael Wagner, Alvaro Iturralde Zurita, and Sijia Zhang
Interspeech Brno 2021

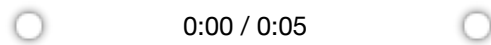
The iambic trochaic law

Alternating duration: iambs

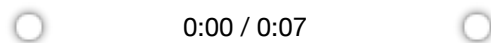
- When one plays alternating sounds that differ in duration, English listeners tend to hear a sequence of **iambs**

(cf. Bolton (1894), Woodrow (1909), Rice (1992), Hay and Diehl (2007), Iversen, Patel, and Ohgushi (2008), i.a.)

... ● — ● — ● — ● — ... → ... (● —)(● —)(● —)(● —) ...



- The same is true in speech (cf. Hay and Diehl (2007), Bhatara et al. (2013), i.a.):



Alternating loudness: trochees

- When one alternates sounds differing in intensity, English listeners typically hear a sequence of **trochees**:

(cf. Bolton (1894), Woodrow (1909), Rice (1992), Hay and Diehl (2007), Iversen, Patel, and Ohgushi (2008), i.a.)

... ● ● ● ● ● ● ● ● ● ● ... → ... ●)(● ●)(● ●)(● ●)(● ●)(● ...



0:00 / 0:04



- Again, the same is true in speech (Hay and Diehl (2007), Bhatara et al. (2013), i.a.):



0:00 / 0:06



Crosslinguistic ITL differences

The ITL is not universal. Earlier proposals attributed cross-linguistic differences to how languages vary in...

- **grouping:** Iversen, Patel, and Ohgushi (2008) argue that Japanese differs from English because of differences in word order, or...
- **prominence:** Bhatara et al. (2013) argue that French differs from German because of differences in word stress

These studies, however, didn't tease apart these two dimensions:

- They used a binary forced choice task
- However, there are (at least) **four** relevant outcomes depending on prominence and grouping, not **two**:
For the speech sequence: *BAga, baGA, GAba, gaBA*

Two dimensions

Two-dimensional parsing

Wagner (2021):

Once we look at the sequences as structured along the orthogonal dimensions of prominence and grouping, the ITL emerges given the cue distribution in English:

- **Stressed and final syllables are longer** →
so when every other syllable is sufficiently long, it will be perceived as stressed and final, resulting in the percept of iambs
- **Stressed and initial syllables are louder** →
so when every other syllable is sufficiently loud, it will be perceived as stressed and initial, resulting in the percept of trochees

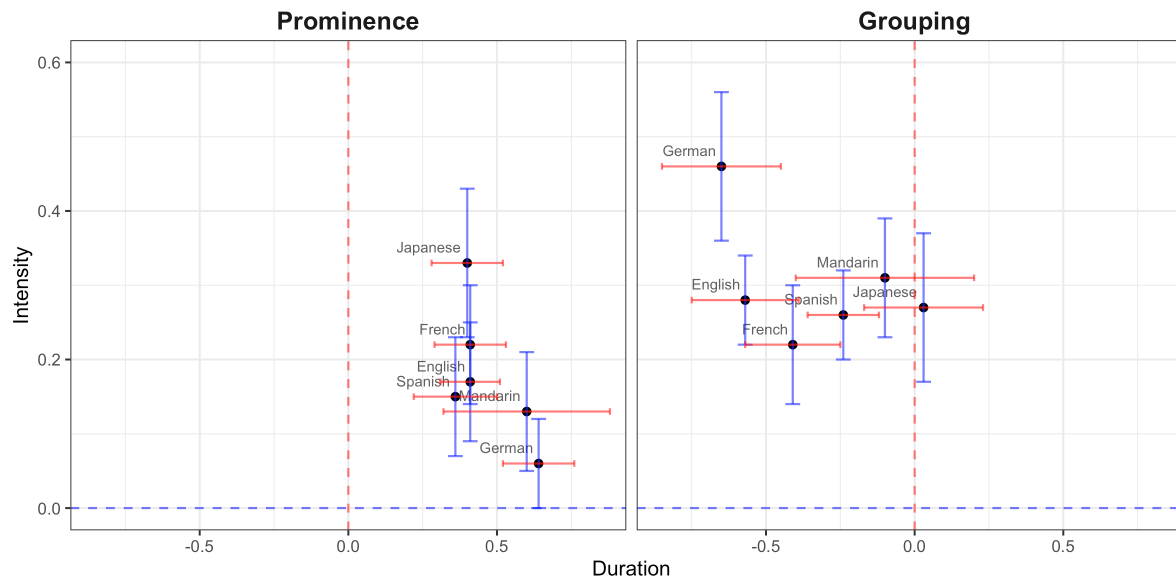
No reference to iambs or trochees is necessary to explain the ITL!

This study:

**A typology of rhythm based on the
cues for prominence and grouping**

A map of the parsing typology

- Our experiments, with stimuli crossing intensity and duration manipulations, consisted of two binary forced choice tasks: prominence (Was *ba* or *ga* prominent?), grouping (did you hear *baga* or *gaba*?)
- Below is a map based on the effect-sizes of duration and intensity on the two decisions



- Differences between languages are small and mostly non-significant for **duration** when looking at **prominence**
- Differences between languages are small and mostly non-significant for **intensity** when looking at **grouping**
- This could explain how we bootstrap into the signal when first exposed to language
- **duration** gives a relatively robust cue for **prominence**, and **intensity** for **grouping**

Rhythm typology

- Grouping and prominence capture an essential part of what we intuitively perceive as 'rhythm'
- Our proposed rhythm typology based on prominence and grouping cues might better capture intuitive rhythm differences than the so-called syllable-timed/stress-timed dichotomy, or more recent gradient acoustic measures aimed at capturing it, such as the %V-measure (Ramus, Nespor, and Mehler (1999)) or the PVI (Low, Grabe, and Nolan (2000))
- We are currently exploring whether our perception maps mirror the cue distribution in production in these languages

Michael Wagner, McGill Linguistics



Alvaro Iturralde Zurita, McGill SCSD



Sijia Zhang, McGill Linguistics



Funded through NSERC Grant RGPIN-2018-06153:
[Three dimensions of sentence prosody \(https://osf.io/6539t/\)](https://osf.io/6539t/)

References

Selected references

- Bhatara, Anjali, Natalie Boll-Avetisyan, Annika Unger, Thierry Nazzi, and Barbara Höhle. 2013. "Native Language Affects Rhythmic Grouping of Speech." *The Journal of the Acoustical Society of America* 134 (5): 3828–43.
- Bolton, Thaddeus L. 1894. "Rhythm." *The American Journal of Psychology* 6 (2): 145–238.
- Hay, Jessica S. F., and Randy L. Diehl. 2007. "Perception of Rhythmic Grouping: Testing the Iambic/Trochaic Law." *Perception and Psychophysics* 69 (1): 113–22.
- Iversen, John R, Aniruddh D Patel, and Kengo Ohgushi. 2008. "Perception of Rhythmic Grouping Depends on Auditory Experience." *The Journal of the Acoustical Society of America* 124 (4): 2263–71.
- Low, Ee Ling, Esther Grabe, and Francis Nolan. 2000. "Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English." *Language and Speech* 43 (4): 377–401.
- Ramus, F., M. Nespore, and J. Mehler. 1999. "Correlates of Linguistic Rhythm in the Speech Signal." *Cognition* 73 (3): 265–92.