Main Findings
Models of prosodic phrasing trained on multiple high-resource languages are used to identify boundaries in an unseen low-resource language.
- While pause is the most important feature for predicting phrase boundaries in all languages, the annotation of pause varies.
- The relative importance of other features varies by language.
- Different acoustic correlates of prosodic boundaries characterize different languages. In some, the relative importance of features is silence > pitch > intensity > duration, while for other languages intensity is more important than pitch.

Motivation
Uses of prosodic event detection:
Part-of-speech tagging, syntactic disambiguation, reducing language model perplexity, salience detection, distinguishing between given and new information, identifying turn-taking behavior and dialogue acts. Typically requires substantial hand-labeled data; not available for most languages.

Corpora
- BDC
- DIRNDL
- DUR
- Italian

Phrase Boundary Detection
- Pause features: whether the end of word precedes a silence, and duration of that pause.
- Duration features: the duration of the word and the difference of the duration of the current and following words.
- Intensity (dB) and Pitch (log Hz) contour features: raw and speaker-normalized signals at different level of aggregations (mean, maximum, minimum and standard deviation). Speaker normalization is performed by z-score normalization.

Cross-Language Phrase Boundary Detection

Comparing Feature Distributions

Future Work
- Cross-language adaptation
- Additional languages
- Examine which features of a language predict good cross-language performance

SpeechLab@Columbia: http://www.cs.columbia.edu/speech/lab.cgi
SpeechLab@QC CUNY: http://speech.cs.qc.cuny.edu/