

# Rhythm in Language

Patel Chapter 3.3

# Introduction

- In many languages, it is normal to produce the syllables of an utterance with differing degrees of prominence.

She wrote her novels with a blue pen she inherited from her aunt.

- The most important physical correlates of prominence are duration, pitch movement, vowel quality and loudness.
- Most sections below will treat prominence as a binary quantity referred to as “stress.”
- Not all languages have lexical stress.
  - Tone languages with stress (Mandarin) and without (Cantonese)
  - Pitch accent languages with stress (Swedish) and without (Japanese)

# Periodicity and typology

- Kenneth Pike 1945
  - Syllable-timed languages: e.g. Spanish
    - Syllables mark off roughly equal temporal intervals
  - Stress-timed languages: e.g. English
    - Roughly equal temporal intervals between stresses.

                  x                                  x                                  x                                  x  
The teacher is interested in buying some books.

          x      x                          x                          x  
Big battles are fought daily.

- To achieve evenly timed feet, speakers would stretch or compress syllables to fit into the typical foot duration.

# Periodicity and typology cont.

- Abercrombie 1967
  - Each syllable is associated with a contraction of muscles associated with exhalation, and some contractions are especially strong (chest pulses, stress pulses).
  - In each language one or the other occurred rhythmically, and rhythm is equated with periodicity.
    - Stress-timed languages: English, Russian, Arabic.
    - Syllable-timed languages: French, Telegu, Yoruba
- Mora-timed languages: Japanese
  - A unit smaller than a syllable, usually consisting of a consonant and vowel, but sometimes containing only a consonant.

# Periodicity and typology cont.

- Empirical measurements of speech have failed to provide any support for isochrony of syllables or stresses.
  - Dauer 1983
    - English stress feet grow in duration with increasing number of syllables.
  - Roach 1982
    - Compared English, Russian and Arabic with French, Telegu and Yoruba and found they could not be discriminated.
  - Beckman 1982
    - Morae are not of equal duration in Japanese.
- Beckman 1992
  - The tripartite scheme correctly groups together languages that are perceived as rhythmically similar, even if the physical basis for this grouping is not well understood (and is not isochrony).

# Phonology and typology

- Research can move forward if one thinks of rhythm as systematic timing, accentuation and grouping patterns in a language that may have nothing to do with isochrony.
- Phonological approach:
  - The rhythm of a language is the *product* of its linguistic structure, not an organizational *principle* such as stress or syllable isochrony.
  - Languages are rhythmically different because they differ in phonological properties that influence how they are organized as patterns in time.

# Phonology and typology cont.

- Factors affecting speech rhythm (Dauer 1983, 1987).
  - Diversity of syllable structures
    - Dutch, English: V ('a') ... CCCVCC ('strength')
    - Japanese, Hawaiian: CV mostly
    - French, Spanish: intermediate
  - Vowel reduction
    - English: Unstressed syllables have reduced vowels: schwa
    - Spanish: unstressed syllables are rarely reduced.
  - Influence of stress on vowel duration
    - English: vowels in stressed syllables are 60% longer than in unstressed syllables.
    - Spanish: Stress does not condition vowel duration to the same degree.

# Phonology and typology cont.

- Dauer:
  - Languages traditionally classified as stress-timed vs. syllable-timed differ in these phonological features.
- The properties do not always co-occur:
  - Polish has complex syllable structure but no vowel reduction.
  - Catalan has simple syllable structure but has vowel reduction.



# Phonology and typology cont.

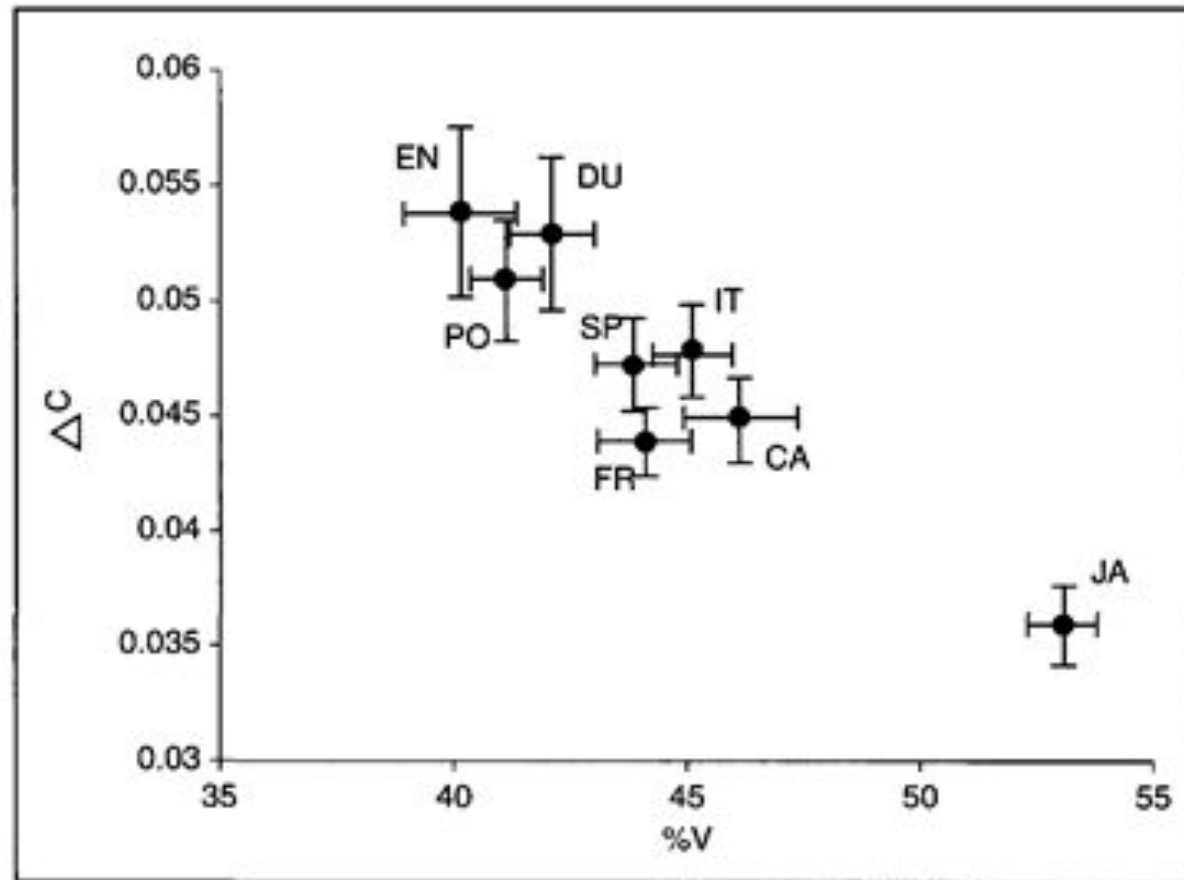
- Bolinger 1981
  - Divides English vowels into two classes: full and reduced (not just schwa, but also 'ih', 'uh', 'oh')
  - Syllables containing full and reduced vowels tend to alternate and long syllables borrow time from short ones.
    - Gets out dirt plain soap can't reach.  
L L L L L L L
    - Takes away the dirt that com-mon soaps can ne-ver reach.  
L- S L- S L- S L- S L- S L- S L-
- Patel:
  - Perhaps stress-timed languages have more contrast in adjacent vowel durations than syllable-timed languages, and have lengthening rules.

# Phonology and typology cont.

- Bolinger also suggested that there are two kinds of rhythm:
  - The rough alternation of long and short syllables and a lengthening rule.
  - A tendency to separate pitch accents so that they do not occur too closely together in time.
    - e.g. “stress shift” (Liberman & Prince 1977):
    - *thirtéen* vs. *thirteen mén*

# Duration and typology

- Ramus et al. 1999
  - Languages that use a greater variety of syllable types are likely to have relatively less time devoted to vowels than languages dominated by simple syllables: %V
  - The durational variability of consonantal intervals should be greater for languages with more diverse syllable structures:  $\Delta C$ .
    - CV.CCCVC.CV.CV.CVCC “stress-timed languages”
    - CV.CV.CVC.CV.CV.CVC.CV “syllable-timed languages”



**Figure 3.7** Percentage of sentence duration occupied by vowels versus the standard deviation of consonantal intervals within sentences for 8 languages. (CA = Catalan, DU = Dutch, EN = English, FR = French, IT = Italian, JA = Japanese, PO = Polish, SP = Spanish.) Error bars show  $\pm 1$  standard error. From Ramus, Nespor, & Mehler, 1999.

# Duration and typology (cont.)

- Frota & Vigário 2001
  - European Portuguese has been classified as stress-timed, and Brazilian Portuguese as syllable-timed.
  - European Portuguese has a significantly higher  $\Delta C$  and lower %V than Brazilian Portuguese.
- Mehler et al. 1996
  - Rhythmic class detection helps bootstrap language acquisition.
  - %V and  $\Delta C$  require minimal knowledge about linguistic units, and thus might do the trick.

# Duration and typology (cont.)

- Patel: Perhaps %V and  $\Delta C$  are not directly relevant to the perception of speech rhythm, but are correlated with variability in syllable duration.
- Patel measured syllable duration for all English and French sentences in the Ramus database, and found that on average English sentences have more variable syllable durations than French sentences do.

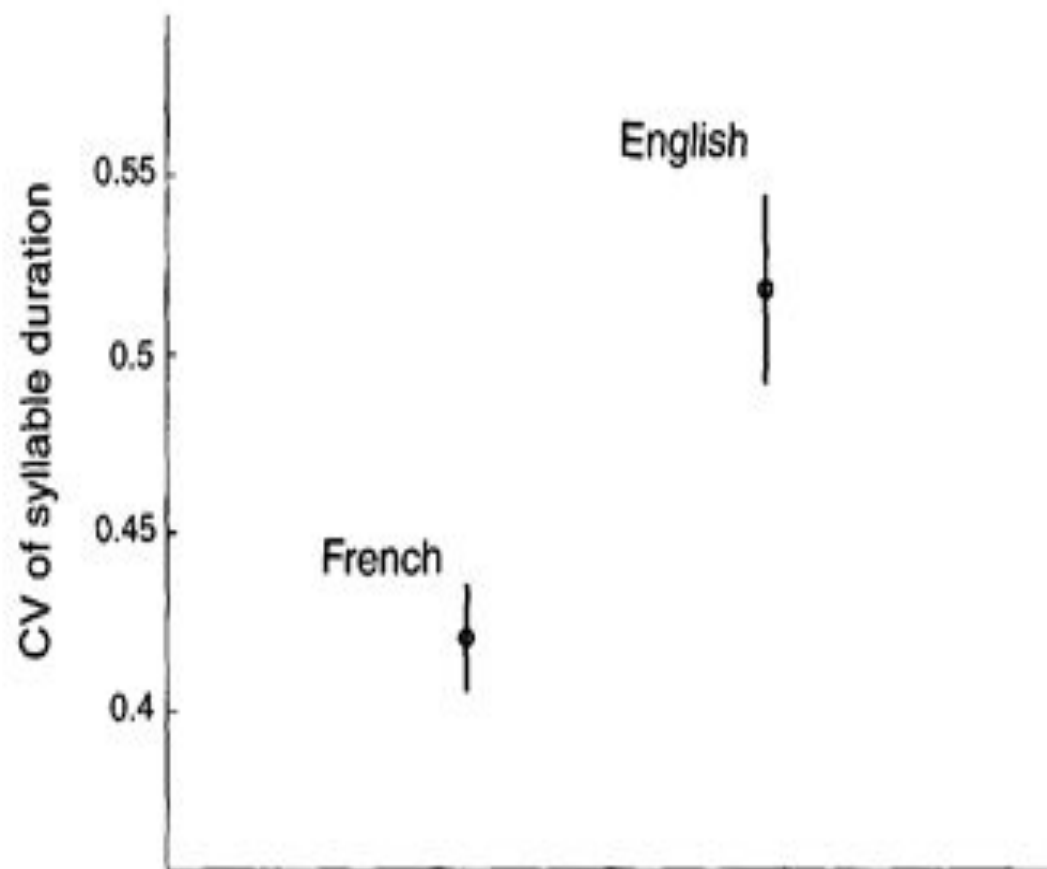


Figure 3.8c The coefficient of variation (CV) of syllable duration in 20 English and 20 French sentences. Error bars show  $\pm 1$  standard error.

# Duration and typology (cont.)

- Low, Grabe, & Nolan 2000 explored the idea that vowel reduction contributes to the idea of stress-timing via its impact on vowel duration variability.
- British English (stress-timed) vs. Singapore English (syllable-timed).
- Normalized pairwise variability index (nPVI) measures the degree of contrast between successive durations in an utterance.

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

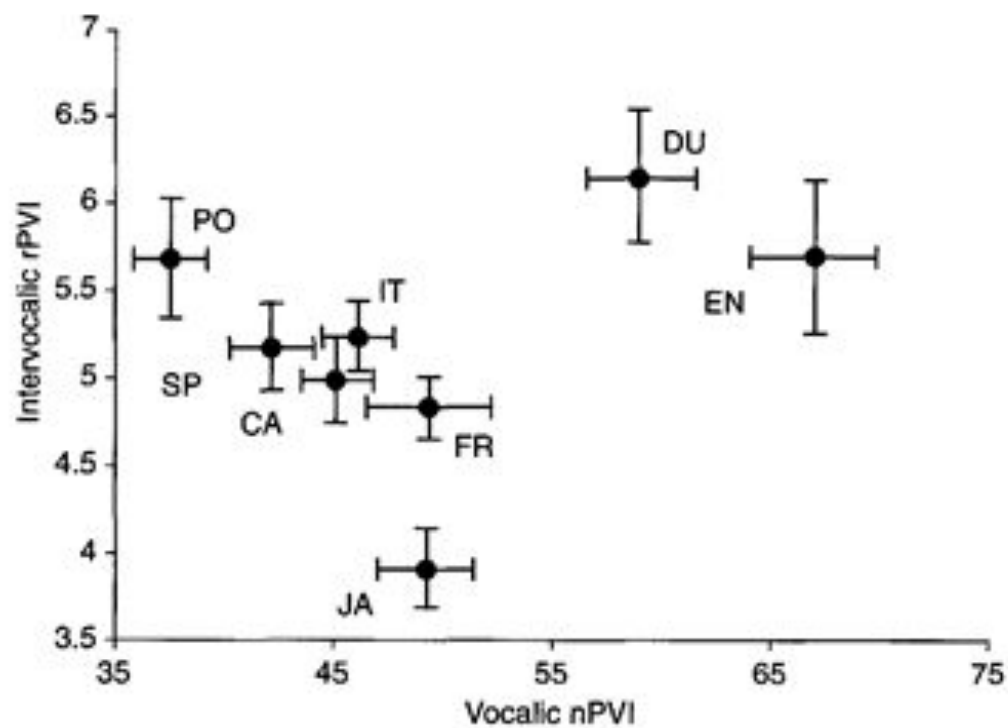




**Figure 3.9** Schematic of sequences of events with varying duration, to illustrate the nPVI (longer bars = longer durations). See text for details.

# Duration and typology (cont.)

- Ramus 2002 measured the vowel nPVI for all eight languages in his database.
- He plotted vowel nPVI against consonantal rPVI (Raw pairwise variability index (without the normalization term in the denominator)).
- nPVI separates English and Dutch from Spanish, Italian and French, and Polish is now far from the stress-timed languages (lacks vowel reduction).
- rPVI segregates Japanese, which has very low durational contrast between successive consonantal intervals.
- So at least two phonetic dimensions may be needed to capture differences between rhythmic classes.



**Figure 3.10** Vocalic nPVI versus Consonantal (intervocalic) rPVI for sentences in eight languages. (CA = Catalan, DU = Dutch, EN = English, FR = French, IT = Italian, JA = Japanese, PO = Polish, SP = Spanish.) Error bars show  $\pm 1$  standard error. From Ramus, 2002a.

# Perception and typology

- All typological theories of language rhythm are ultimately rooted in perception.
- The old classification scheme has its shortcomings.
  - Some languages straddle different categories (Polish, Catalan).
  - Many languages don't fit neatly into any of the existing categories.
- Any new science of rhythmic categories must have a body of perceptual data as its foundation.
  - Perceptual work on the rhythmic differences between languages has already begun.

# Perception and typology cont.

- Ramus & Mehler 1999 proposed that if a listener can tell two languages apart when the only cues are rhythmic, then the languages belong to distinct rhythmic classes.
- Speech resynthesis techniques were used to selectively remove various phonetic differences between languages and focus attention on rhythm.
- English vs. Japanese. [Examples](#).

# Perception and typology cont.

- Four versions of a sentence:
  - Original
  - Saltanaj
    - All fricatives are replaced by /s/, all vowels by /a/, liquids by /l/, plosives by /t/, nasals by /n/ and glides by /ai/.
  - Sasasa
    - All consonants are replaced by /s/ and all vowels by /a/.
  - Flat sasasa
    - Pitch is flattened into a monotone.
- French adults could discriminate between English and Japanese in all three conditions.
  - Supports the hypothesis that the rhythms of English and Japanese are indeed perceptually distinct.

# Perception and typology cont.

- Ramus et al. 2003
  - Focused on flat sasasa transformations.
  - French adults could discriminate Polish from English, Spanish and Catalan, but could not discriminate Catalan from Spanish.
  - Polish can be separated from the other languages in the earlier study by the variability of vowel duration ( $\Delta V$ ) because Polish has very low vowel duration variability.
- Thus, at least four typological classes must be distinguished.

# Perception and typology cont.

- Nazzi et al. 1998 tested newborn babies' rhythm perception
  - Used low-pass filtered speech, which removes most phonetic information but preserves syllable, stress and pitch patterns.
  - French newborns can discriminate English from Japanese.
  - They could discriminate English and Dutch from Italian and Spanish, but not English and Spanish from Dutch and Italian.
  - Thus babies can discriminate languages only when they belong to different rhythm classes.



# Principles governing the rhythmic shape of words and utterances

- For comparing rhythm in language and music, metrical phonology is important.
- Rhythmic prominence is treated as hierarchical.
  - Prominence is incrementally assigned at each level of the prosodic hierarchy according to systematic principles.
  - Prominence is not simply a binary feature called “stress” that syllables either have or don’t have.
  - Rather, prominence is a function of the syllable’s relation to the whole hierarchy of the utterance.
  - A metrical grid is associated with each utterance, similar to the metrical grid of Lerdahl & Jackendoff for music.
    - Prominence patterns are considered without regard to their exact timing.



Figure 3.11 A metrical grid for a sentence of English. From Selkirk, 1984.

- The second level marks stressed syllables, and is the level of the basic “beat,” in analogy to the tactus in music.
- The third level marks the primary lexical stress of each word, and the fourth level marks the main accent of each phrase.

# Principles governing the rhythmic shape of words and utterances cont.

- The text-to-grid assignment of beats provides the input on which rhythmic principles operate:
  - A “principle of rhythmic alternation” says that strong beats at any level should be separated by no more than two weak beats at that level.
  - A rule of “beat addition” might add a beat at the second level to avoid a long series of unstressed syllables.
  - At the third level, a rule of “beat movement” might shift the primary accent of a word to avoid the adjacency of primary accents.
- Metrical phonology derives the prominence pattern of a sentence using ideas directly inspired by theories of musical meter.
  - Both theories concern the patterning of time intervals at several timescales.

# Principles governing the rhythmic shape of words and utterances cont.

- Differences between linguistic and musical metrical grids.
  - Temporal periodicity in musical meter is much stricter than anything found in speech.
  - The regular periodicities of music allow meter to serve as a mental framework for sound perception.
    - An event can be perceived as metrically prominent even if it is physically quite weak, as in syncopation.
    - The prominences of language are not regular enough to allow for anything as abstract as syncopation.
  - Linguistic grids are not abstract periodic mental patterns but are simply maps of heard prominences, full of temporal irregularities.

# Principles governing the rhythmic shape of words and utterances cont.

- Maybe speech and music share a tendency to alternate weak and strong elements.
  - Kelly & Bock 1988: the tendency to stress the first syllable in English is weaker when a nonsense syllable is preceded by a stressed syllable:
    - The full teplez decreased.
    - Throw the teplez badly.
  - Cutler 1980: speech errors of omitting a syllable serve to shorten a long run of unstressed syllables:
    - x           x           x       x  
Next we have this bicentennial rug.
    - x           x                   x       x  
Next we have this bicentennial rug.

# Principles governing the rhythmic shape of words and utterances cont.

- But instead, these findings may reflect the action of nonrhythmic forces that seek to keep prominences at a comfortable distance from each other.
  - The observed tendencies are rather weak.
  - Stress clash avoidance may be due to the mechanics of articulation:
  - Avoid crowding large jaw movements together when speaking at a fast rate.

# The perception of speech rhythm

- The perception of isochrony in speech?
  - Lehiste 1977
    - Listeners had more difficulty identifying the shortest and longest ISI (interstress interval) in speech than in nonspeech analogs of the sentences containing clicks and noise.
    - Perhaps they are *perceiving* speech stresses as isochronic.
  - Donovan & Darwin 1979
    - When imitating speech stresses by tapping, subjects tapped with less temporal variability than the actual stressed syllables. This didn't happen when tapping to noise analogs.
- But there is no evidence that speech is perceived as isochronous under ordinary circumstances.

# The perception of speech rhythm cont.

- The role of rhythmic predictability in speech perception.
  - Cutler & Darwin 1981 found that listeners can predict an upcoming contrastive stress from preceding phonetic information:
    - Spliced a neutral version of a target word into a high and low emphasis sentence
      - She managed to remove the dirt from the rug, but not the grass stain.
      - She managed to remove the dirt from the *rug*, but not from their clothes.
    - Subjects were faster at detecting a phoneme (/d/) in the high emphasis sentence, as if the prediction of stress was influencing speech processing.



# The perception of speech rhythm cont.

- But this effect was not replicated in sentences where lexical stress could be predicted syntactically.
  - Pitt & Samuel 1990
    - The guard asked the visitor if she had a pérmit to enter the building.
    - The waiter decided he could not permít anyone else in the restaurant.
- In music, rhythmic predictability enhances beat perception and guides the coordination of ensemble performance and dance.
- It isn't clear that rhythmic predictability would have an adaptive role in speech.

# The perception of speech rhythm cont.

- The role of rhythm in segmenting connected speech.
  - English listeners expect strong syllables to be word initial.
    - Cutler & Butterfield 1992: when listeners missegment speech they tend to place word boundaries before stressed syllables:
      - by loose analogy → by Luce and Allergy
  - French and Spanish speakers segment syllabically, and Japanese speakers segment moraically.
  - The native strategies are used even when listening to a foreign language.
  - Does experience with a native language influence how one segments non-linguistic rhythmic information?

# The perception of speech rhythm cont.

- The role of rhythm in the perception of nonnative accents.
  - White & Mattys 2007 examined Spanish speakers of English.
    - The greater their vowel duration variability, the more native-sounding they were rated by English speakers.
    - Reflects the extent to which they have learned vowel reduction.
    - The greater their vowel nPVI, the more native-sounding they sounded, but this factor was less predictive than vowel duration variability.
    - Vowel duration variability may be more perceptually relevant for speech rhythm than durational contrastiveness.

# Final comments on speech rhythm: moving beyond isochrony

- The case for periodicity in speech is extremely weak.
  - The rhythms of speech are the by-product of phonological phenomena such as syllable structure, vowel reduction, lexical prominence location, stress clash avoidance, and prosodic phrasing in sentences.
  - In music, patterns of timing and accent are a focus of conscious design.
  - In music, but not speech, rhythm conveys a sense of motion to the listener.
- Still, we will see that rhythm in music and speech can be fruitfully compared.