

What is Rhythm?
Is Speech Rhythmic?

Definitions/Properties of rhythmicity

- Regular recurrence of events (surface isochrony)
- Perceived isochrony of events or “beats” (Lehiste 1979)
- Listeners can move to (dance to, sing with) rhythmic sounds (Cummins 2009 among others)
- Timing of upcoming events is predictable from previous events (Kochanski & Shih 2010)
- Shows compression of units as more occur within a rhythmic interval
- Systematic relationship between structure and surface timing patterns

Speech can be rhythmic...But...

- In many (most?) situations, speech is not rhythmic in the way that music is.
- Nevertheless, the rhythm label has been applied to speech
 - Rhythm classes (Pike 1946, Abercrombie 1967)
 - Stress-timed
 - Syllable-timed
 - Mora-timed (Bloch 1942, Han 1962, Beckman 1982)
- Based on impressions of interval isochrony

The North Wind and the Sun, as parsed by David Abercrombie

Abercrombian Feet

|¹The |²north wind and the |³sun were dis |⁴puting
|⁵which was the |⁶stronger, when a |⁷traveller came a
|⁸long |⁹wrapped in a |¹⁰warm |¹¹cloak. |¹²They a
|¹³greed that the |¹⁴one who |¹⁵first suc |¹⁶ceeded in
|¹⁷making the |¹⁸traveller |¹⁹take his |²⁰cloak off should be
con |²¹sidered |²²stronger than the |²³other . . . |²⁴Then
the |²⁵north wind |²⁶blew as |²⁷hard as he |²⁸could,
but the |²⁹more he |³⁰blew the more |³¹closely did the
|³²traveller |³³fold his |³⁴cloak a |³⁵round him, and at
|³⁶last the |³⁷north wind |³⁸gave up the at |³⁹tempt . . .
|⁴⁰Then the |⁴¹sun |⁴²shone out |⁴³warmly, and im-
|⁴⁴mediately the |⁴⁵traveller took |⁴⁶off his |⁴⁷cloak,
|⁴⁸and so the |⁴⁹north wind was o |⁵⁰bliged to con-
|⁵¹fess that the |⁵²sun was the |⁵³stronger of the
|⁵⁴two.

Prominence-based constituents: Abercrombian feet

- Can include word-fragments
 - Take **Greater** London.
 - Take **Grey to** London.
 - Take **Grey to**-morrow.
- Assumed to be delimited by phrasal prominences
 - Abercrombie 1973:

(32) | Know then thy- | -self, pre- | -sume not | God to | scan | ^ |,

Surface isochrony doesn't exist

- Inter-stress intervals spoken by David Abercrombie vary drastically (from less than 200 ms to ca. 700 ms, measured by E. Uldall 1971, 1978)
- Inter-stress interval duration depends on number of syllables in the interval (see also Dauer 1983)
 - Additional syllables add 100 ms or more
 - Much more than the just noticeable difference for duration perception

Polysyllabic shortening—Evidence for Abercrombian feet?

- Add more weak syllables after a strong syllable --> shorten the strong syllable
 - *sleep* longer than *sleepy*
 - *sleepy* longer than *sleepiness*

(Lehiste 1971)

Polysyllabic Shortening: Do rhythmic units predict its occurrence? Shattuck-Hufnagel & Turk (in prep.)

- Does it **respect word boundaries**, as in Prosodic Words, or Clitic Groups?
 - *bak-* expected to be shorter in
 - *Baking apples, Bake us apples, Bake us an apple*
than in *bake apples*, but not shorter in e.g. *Bake avo-cados*
- Does it ignore word boundaries, as in Abercrombian Feet?
 - *bak-* expected to be shorter in
 - *Bake e-lixirs, Bake avo-cados*
than in *bake apples*

Polysyllabic Shortening

- Evidence that this duration adjustment process is influenced by word boundaries
 - **tuna** choir (respects words)
 - **tune a**-cquire (fragments words)
- Duration of [tun]:
 - [tun-] in **tuna** less than [tun-] in **tune a**-
(Turk and Shattuck-Hufnagel 2000)

Polysyllabic Shortening

- Missing a critical comparison:
 - Duration of [tun] in two-syllable strings:
 - *Tuna* choir < *Tune* a-cquire
 - Are these both shorter than [tun] in 1-syl string?
 - *Tune* choirs
 - If so, the Abercrombian foot would be suggested as an additional unit which influences polysyllabic shortening.

Hints/Ambiguities in the Literature

- Polysyllabic shortening across word fragments
 - *sticky* ~ ...*stick in*-(creased)
 - *stickiness* ~ ...*stick was dis*-(carded) (Lehiste 1971)
 - *Chees(es) (a)bound(ed) (ab)out* (Huggins 1973)
 - NP-VP: *Cheese a*bounded vs. *cheese* bounded
 - No difference
 - VP: *Bound a*bout vs *Bound* out
 - Polysyllabic shortening within VP
- Phrasal prosody not specified

Large-scale studies don't distinguish all possible unit types

- Williams & Hiller (1994)
 - Strong syllable shorter if followed by more wk syllables
 - Largely within-word & word combo units? (*bake us an*)
 - No separate test of Abercrombian Foot (*bake avo-*)
- Kim and Cole (2005)
 - Strong syllable shorter if followed by more wk syllables
 - True for both within-word and word combination units
 - *baking, bake us, bake an, bake us an*
 - No separate test of Abercrombian Foot (word fragments)

Rhythmic vs. word-based constituents

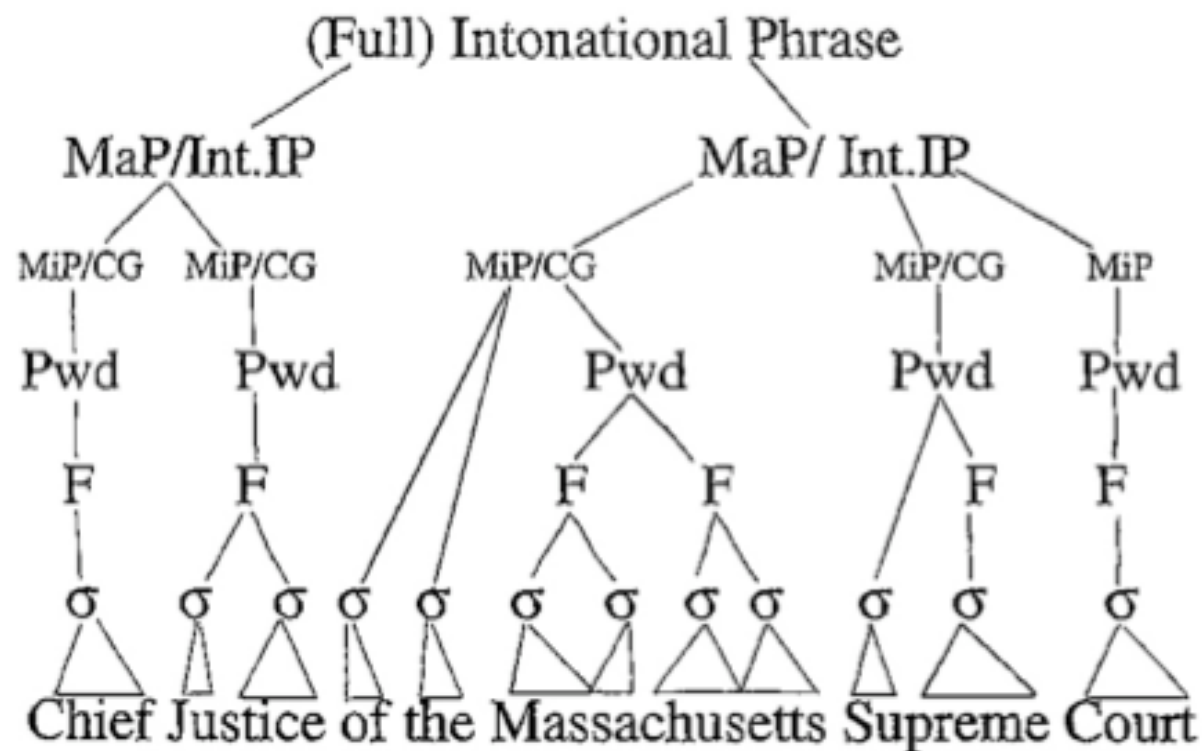


Figure from Shattuck-Hufnagel & Turk 1996

Word-based vs. prominence-based constituents (Abercrombian feet)

Word-based

- Bake] [apples
- Baking] [apples
- Bake us] [apples
- Bake] [avocados
- Bake] [elixirs

Prominence-based, rhythmic

- Bake] [apples
- Baking] [apples
- Bake us] [apples
- Bake avo-] [-cados
- Bake e-] [-lixirs

Elicitation Method: limericks as a way to encourage prominence based, rhythmic constituents

Shattuck-Hufnagel & Turk in prep.

- 10 monosyllabic verbs
 - *bake, pick, cook, tab, bag, stop, track, grab, crib, catch*
- Embedded in 4th line of limerick

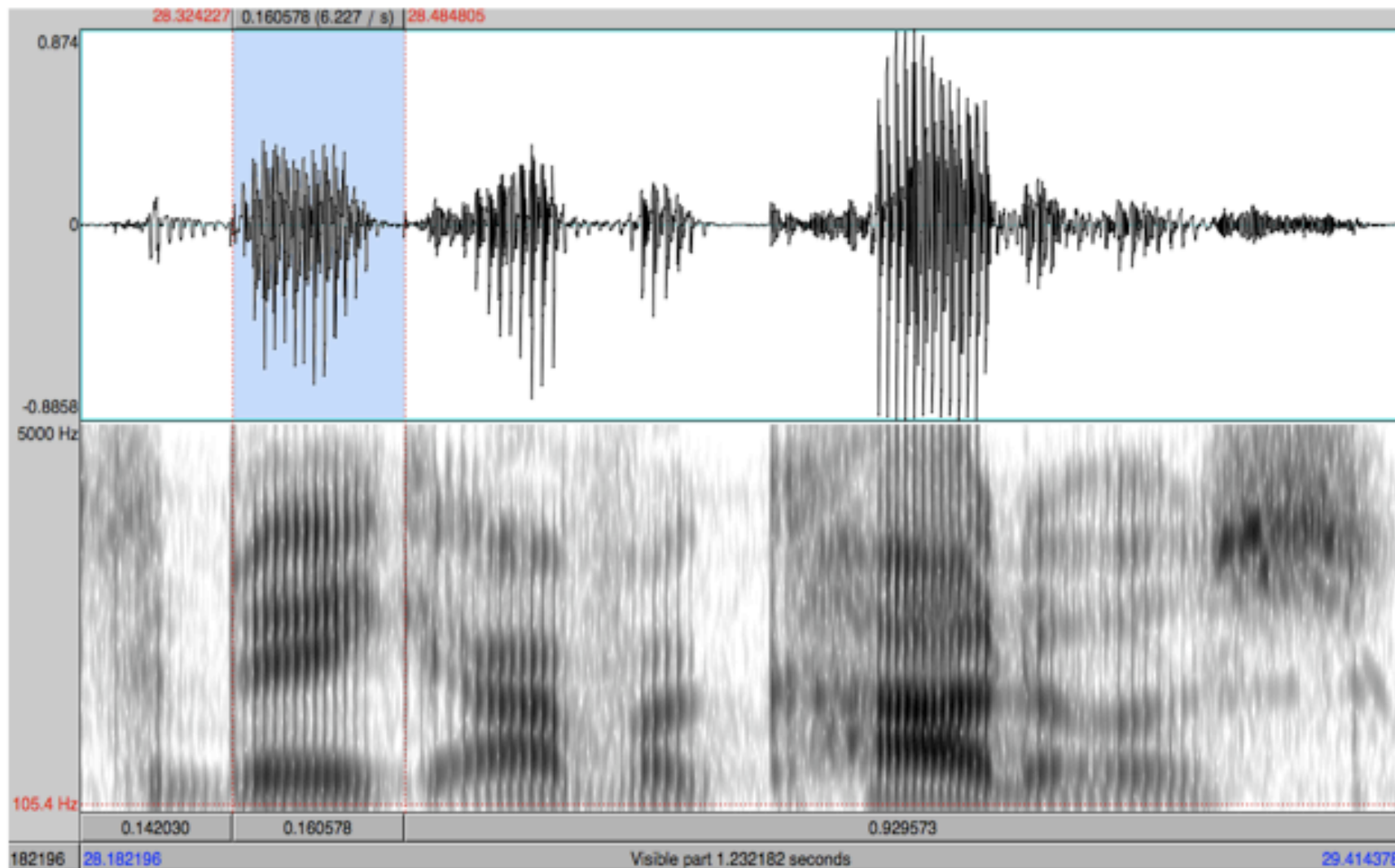
*There once was a boy from St. Paul
Who loved to bake fruit in the fall
He'd give up his Snapple
To **bake us an** apple
With butter and sugar and all.*
- 7 speakers, 3 reported here

Acoustic Analyses

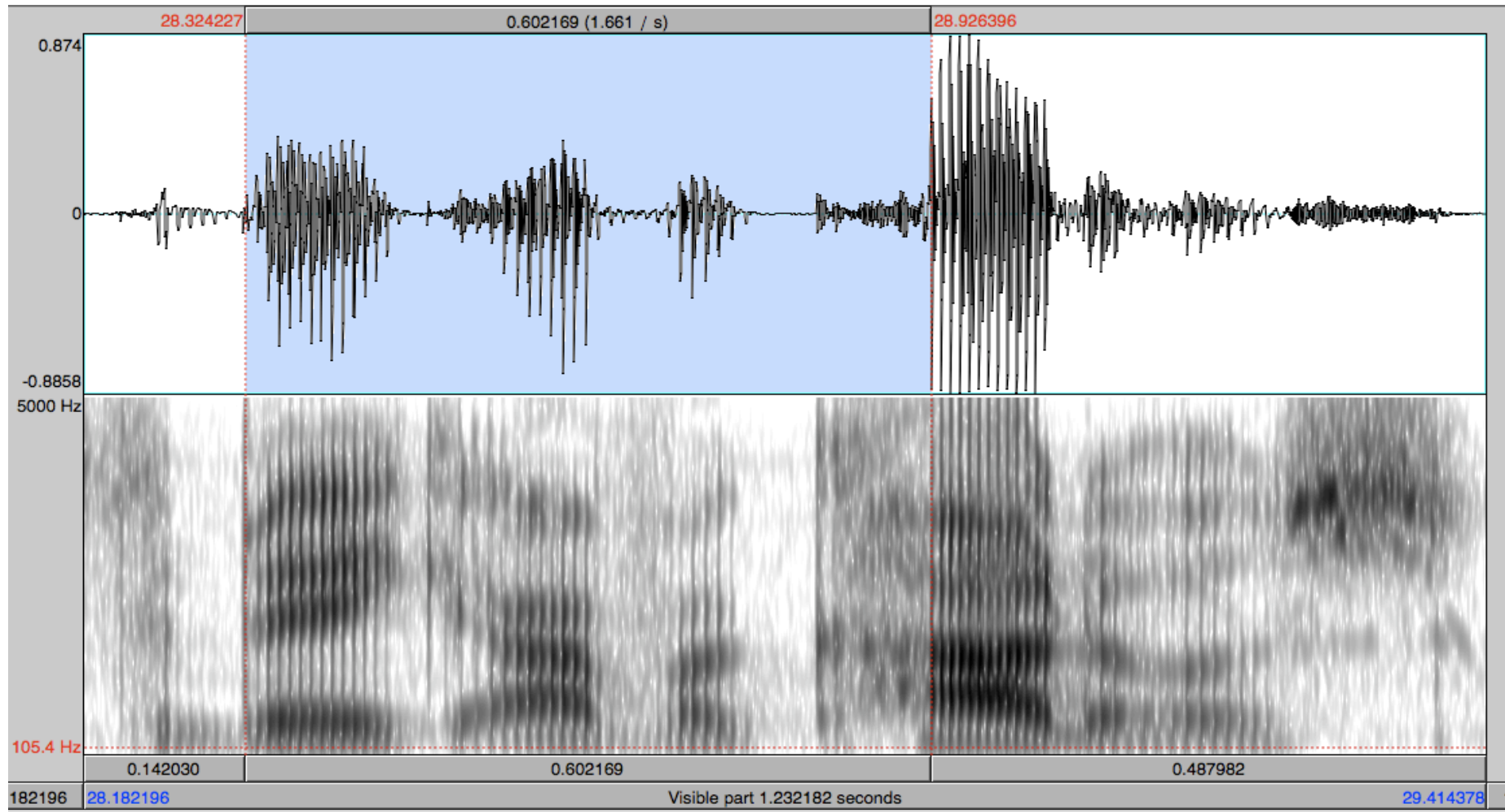
- Rhyme duration of base verb
- Duration of inter-prominence interval
- Presence of boundary markers
 - Pause, irregular glottalized pulses at word-onset V



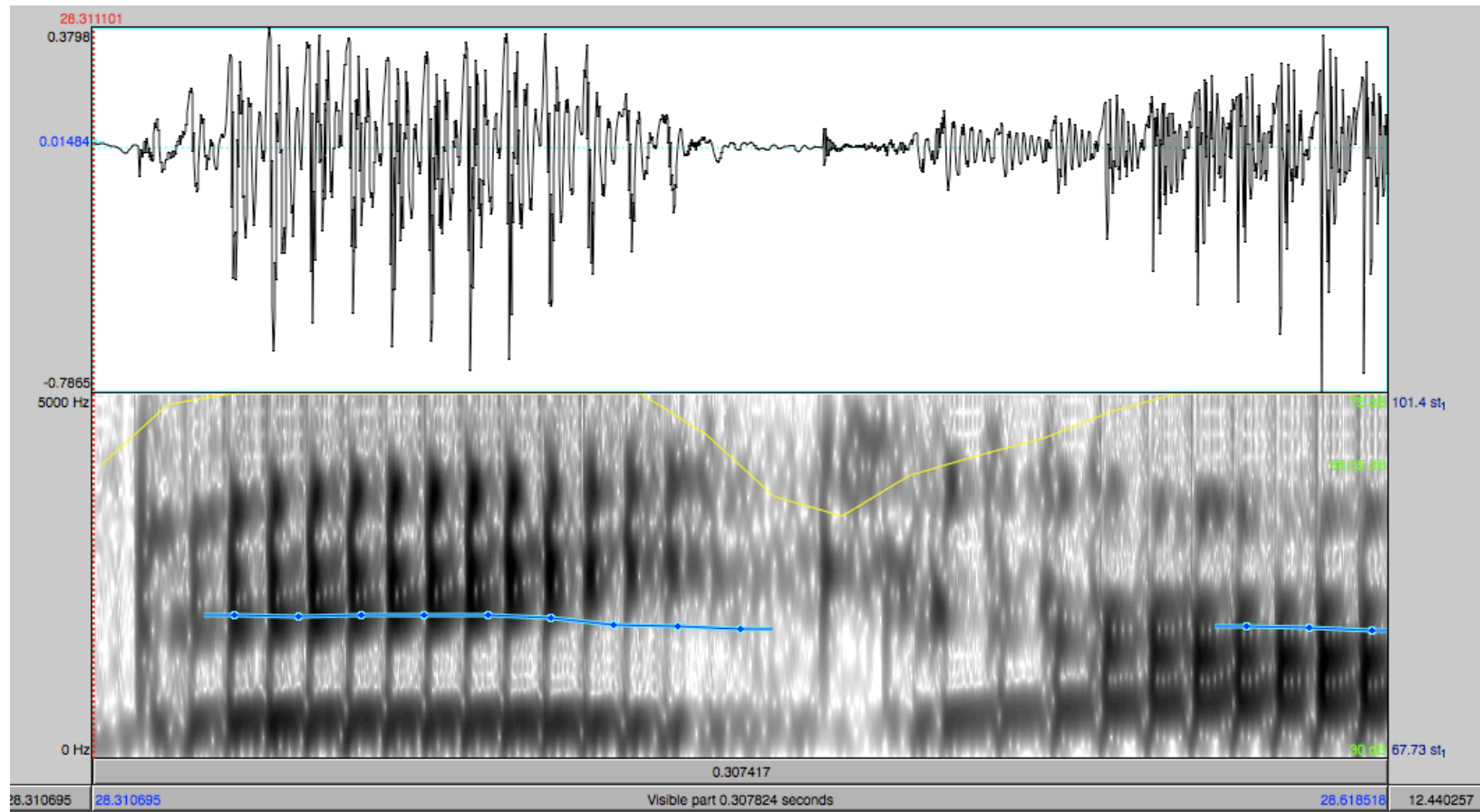
Duration of Base Verb Rhyme



Duration of InterProminence Interval



Insertion of Boundary Cues

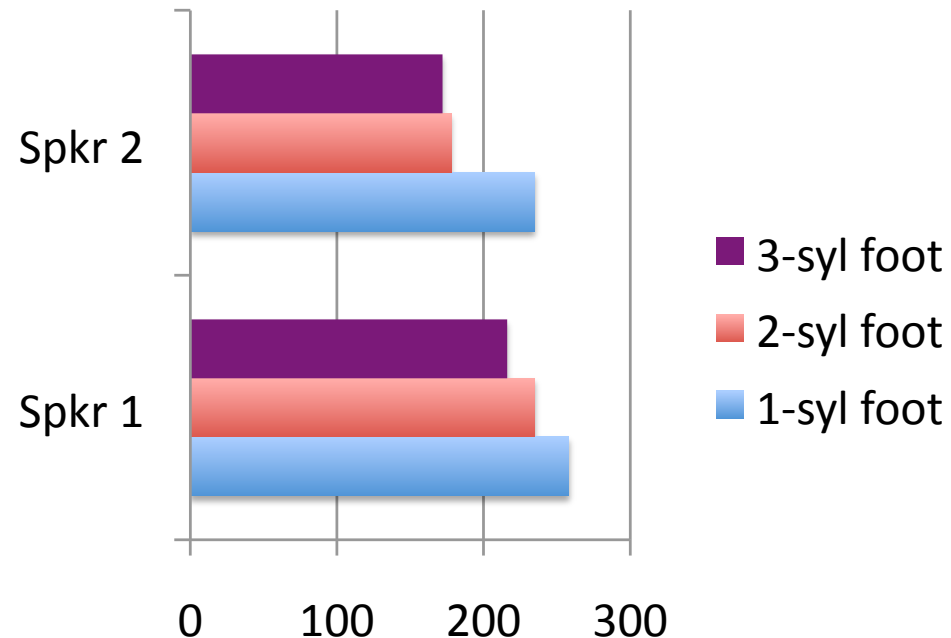


Motivation for boundary cue evaluation

- Silence and/or irregular pitch periods often occur at word-onset vowels of larger prosodic constituents
 - Full Intonational Phrase (Pierrehumbert & Talkin 1992)
 - Intermediate Intonational Phrase
(Dilley, Shattuck-Hufnagel and Ostendorf 1996)
- Do these cues also occur at the word-onset vowels of certain weak syllables in these S-w strings?
 - If so, may suggest a constituent boundary
 - *bake_us* apples vs. *bake_an* apple
 - *bake_e*-lixirs, *bake_avo*-cados

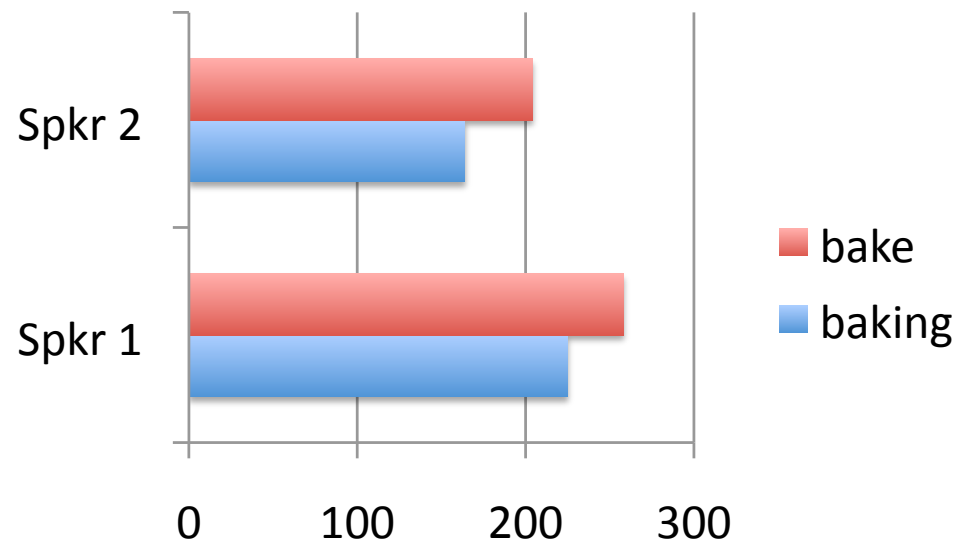
Results (1) Evidence of polysyllabic shortening for two of three speakers

- All foot types pooled
 - 1 syl foot: bake apples
 - 2 syl foot: baking, bake us, bake an, bake e-
 - 3 syl foot: bake us an, bake avo-
- Mean rhyme duration in ms



Results (2): Polysyllabic shortening occurs within words

- For two of three speakers
- *bak-* shorter in *baking* than in *bake*



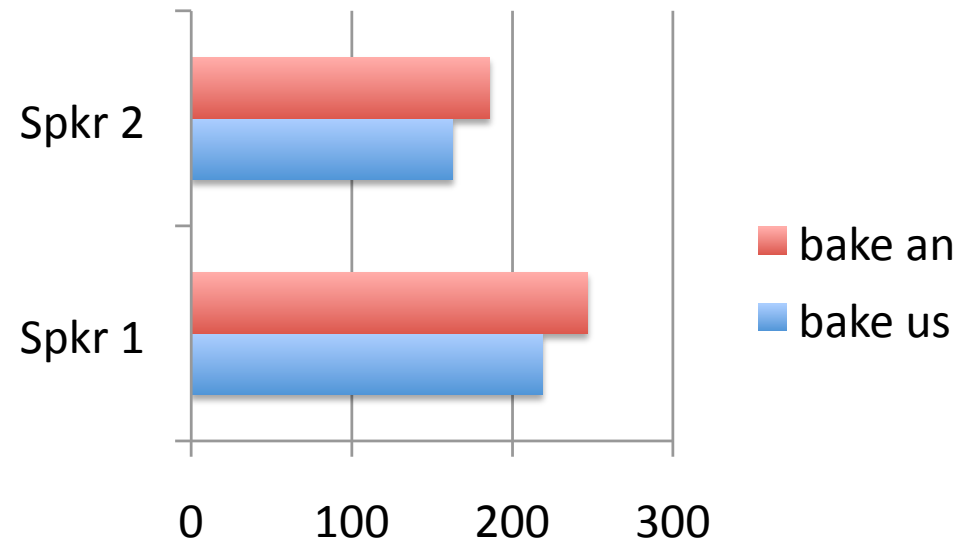
Results (3) Polysyllabic shortening occurs within word combinations

- e.g. Clitic Groups

	<i>bak-</i> shortened in <i>bake us</i> , as in <i>baking</i> ?
Spkr 1	yes (no difference)
Spkr 2	yes (no difference)
Spkr 3	(no difference)

Results (4): Syntactic groupings also influence polysyllabic shortening

- For two of three speakers
- *bak-* shorter in *bake us* than in *bake an*



Results (5)

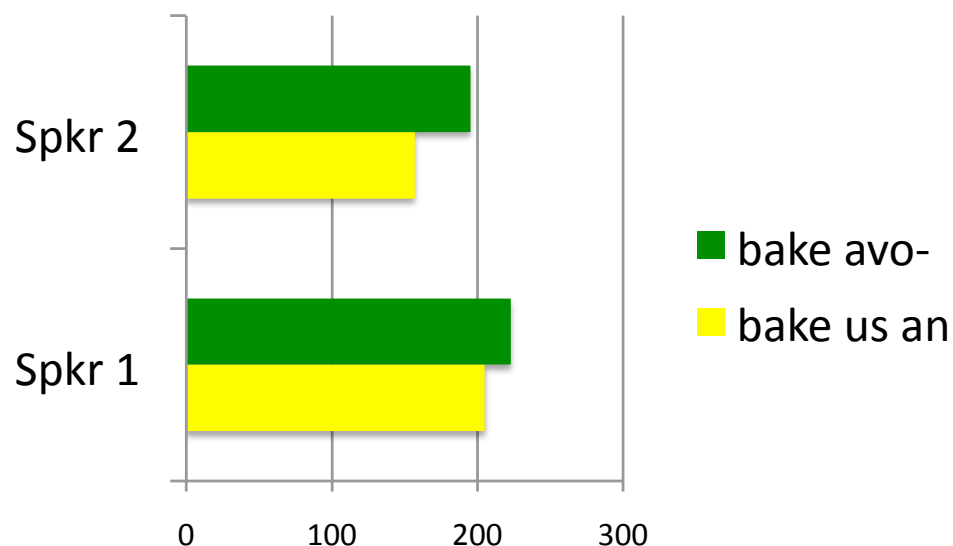
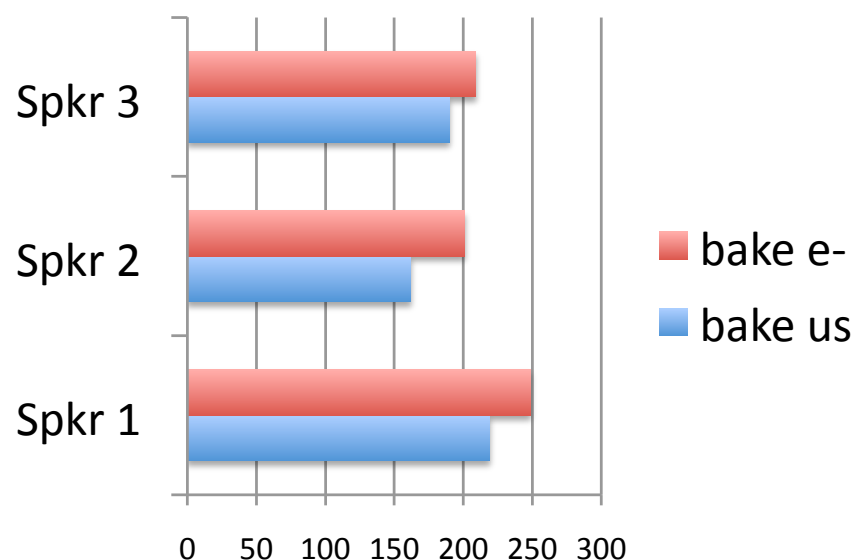
- Does polysyllabic shortening occur in word-fragment combinations? (Abercrombian Feet):

	<i>bak-</i> shorter in <i>bake</i> e- than in <i>bake</i> ?	<i>bak-</i> shorter in <i>bake</i> <i>avo-</i> than in <i>bake</i> ?
Spkr 1	n.s.	yes* (223 vs 258)
Spkr 2	n.s.	n.s.
Spkr 3	wrong direx	n.s.

Results (6) Polysyllabic shortening in word-based constituents is stronger than in comparable Abercrombian feet

bak- in *bake us* shorter than in *bake e-*

bak- in *bake us an* shorter than in *bake avo-*



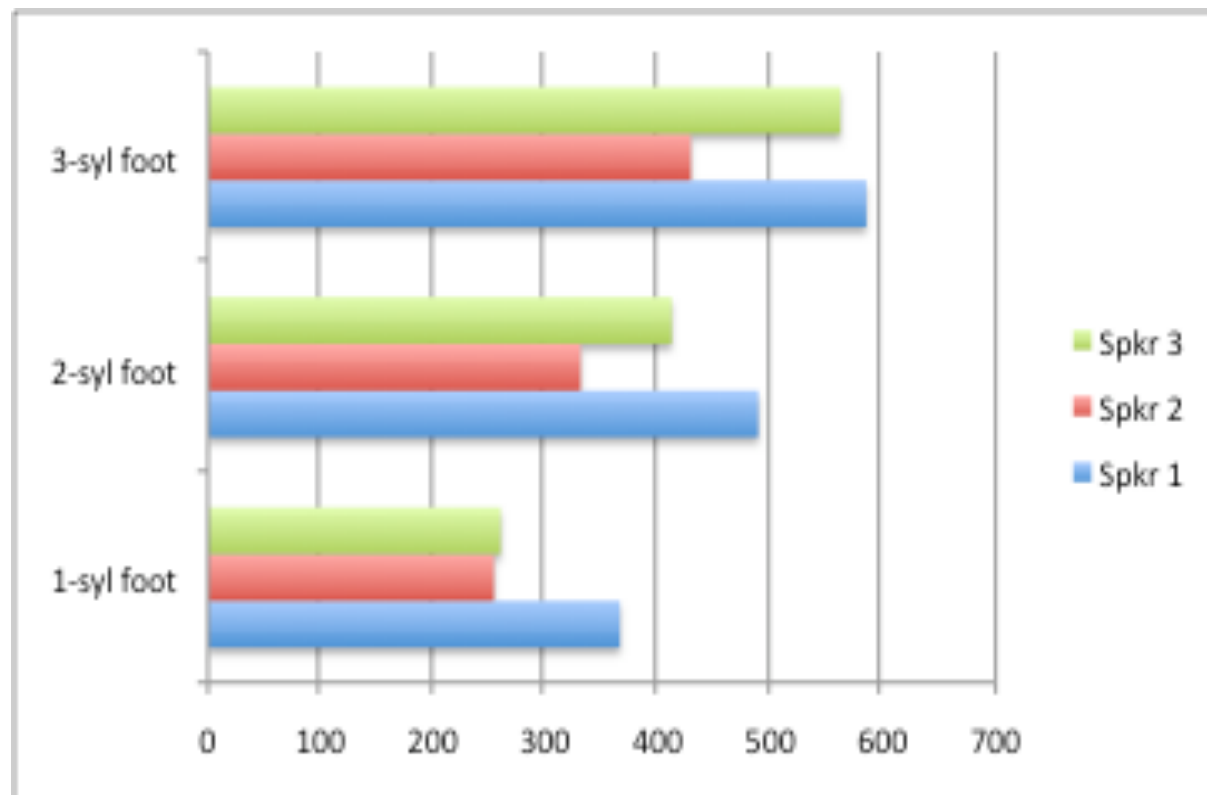
Results (7)

- Do boundary cues occur at word onsets within Abercrombian Feet?

	<i>bake _e-(lixirs), bake _avo-(cados)</i>
Spkr 1	15/20
Spkr 2	12/20
Spkr 3	16/20

Results (8) Speakers don't achieve isochrony

- Duration of Inter-Prominence Interval (in ms) increases with number of syllables



Summary

- Analysis of more speakers in progress
- For speakers showing polysyllabic shortening
- Evidence for polysyllabic shortening by word-based weak syllables is strong
- Evidence for polysyllabic shortening by word-fragment weak syllables is weak, but exists
 - some speakers insert boundary markers within Abercrombian Foot candidates

- Speakers may simultaneously build word-based and prominence-based constituents
 - But word-based constituents are more influential in determining acoustic segment durations
 - Even in highly rhythmic speech (internally generated rhythm)

Word-based constituents

- Also constrain rhythmicity in phonology, e.g.
 - Alternating stress patterns operate **within words** (cf. Hayes)
 - *'Apa'lachi'cola*
 - stress clash avoidance mechanisms operate within phrases (e.g. Chi'nese vs. 'Chinese vase), (Shattuck-Hufnagel, Ostendorf & Ross 1994)
 - Symmetric parsing of high level prosodic constituents into lower word-based constituents (Gee & Grosjean 1983)

Back to polysyllabic shortening— normal, non-limerick speech

- Can we conclude that polysyllabic shortening is a rhythmic principle that operates within word-based units?
- Not in an obvious way--compression does not fully compensate for added syllables within units
- [ek] 27 ms shorter in *baking* than in *bake*
- But *baking* 119 ms longer than *bake*
- May be possible to implement this subtle *tendency* towards word-based-unit isochrony as competition in a coupled oscillator model, with word-based units and syllables as levels (proposed by Saltzman, Nam, Krivokapic & Goldstein 2008)

However, the story is more complicated...

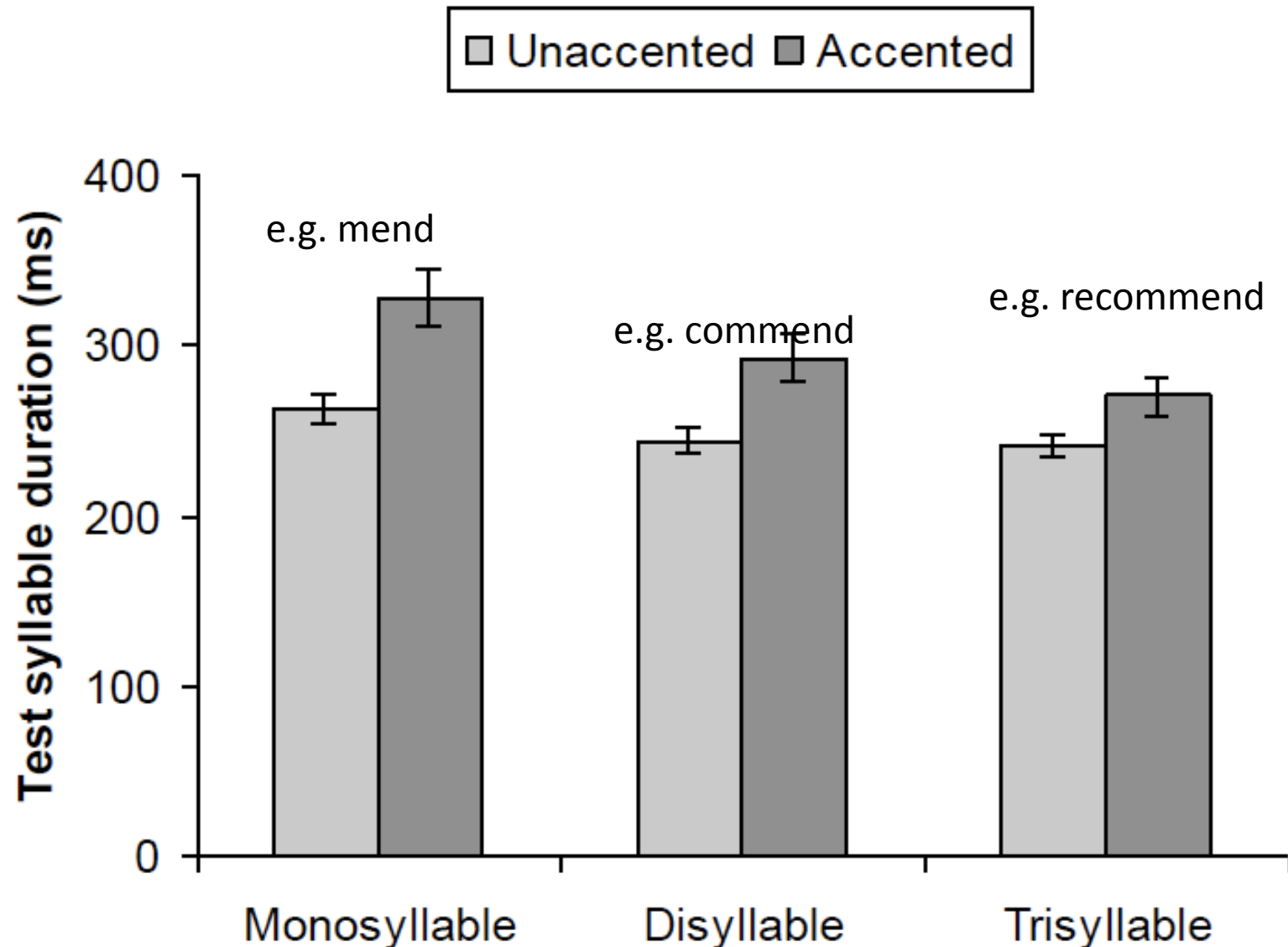
- Polysyllabic shortening effects are often difficult to distinguish from **edge effects**
- i.e. [bek] may be longer in *bake* than in *baking* because of final lengthening
- But White & Turk (in press) found e.g. –*mend* longer in *com****mend*** than in *recom****mend*** in phrasally-stressed contexts.
- Suggests polysyllabic shortening is real: *mend* is final in both cases, but polysyllabic shortening effects still occur.

Polysyllabic shortening is not obligatory

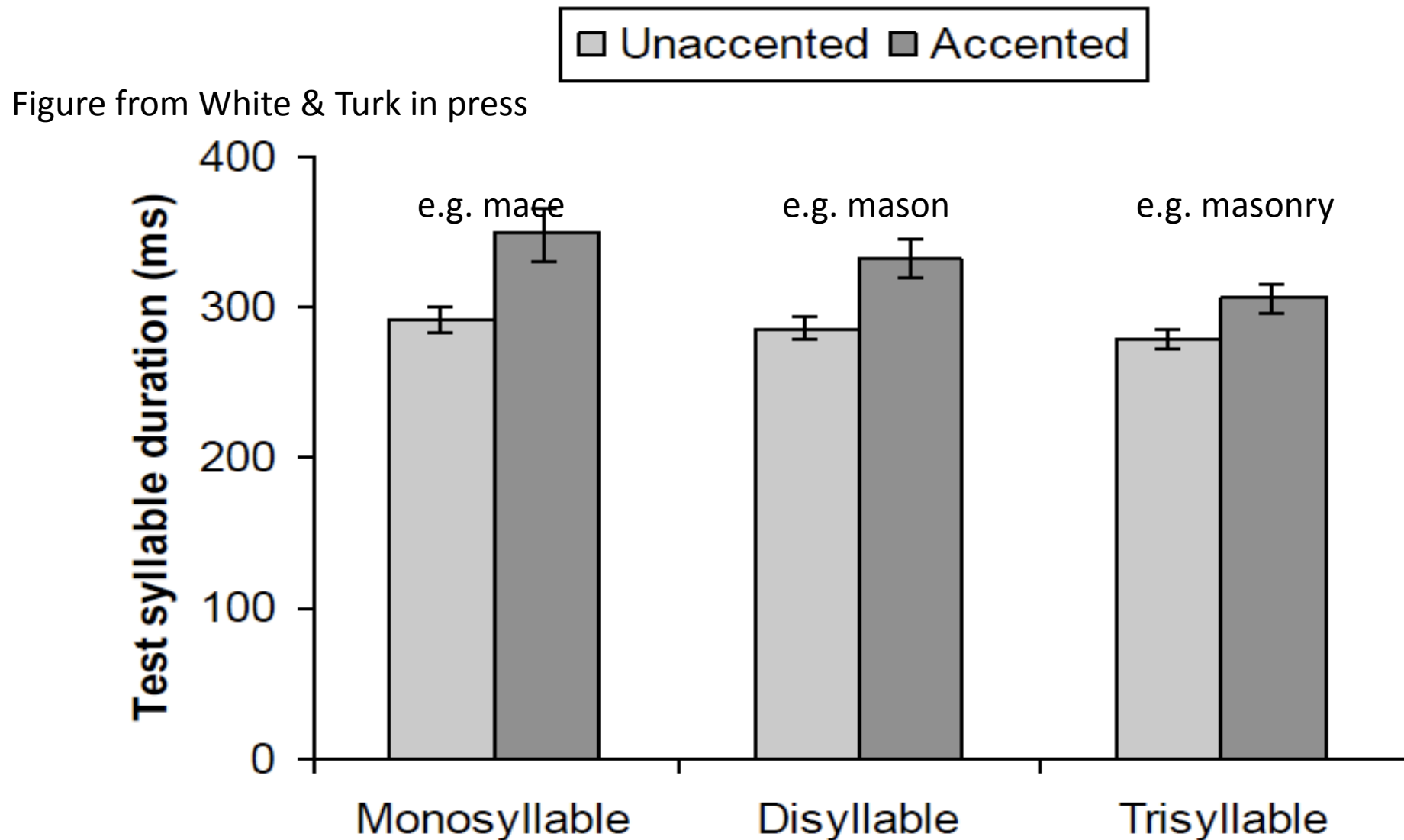
- No polysyllabic shortening in Finnish (language with fixed word-initial stress), even in phrasally prominent contexts. (Suomi 2008)
- Effects in English can be much reduced and even absent in non-phrasally stressed contexts. Turk & Shattuck-Hufnagel 2000 and White & Turk (in press)

-

No difference in e.g. *mend* in *commend* or *recommend* when not phrasally stressed



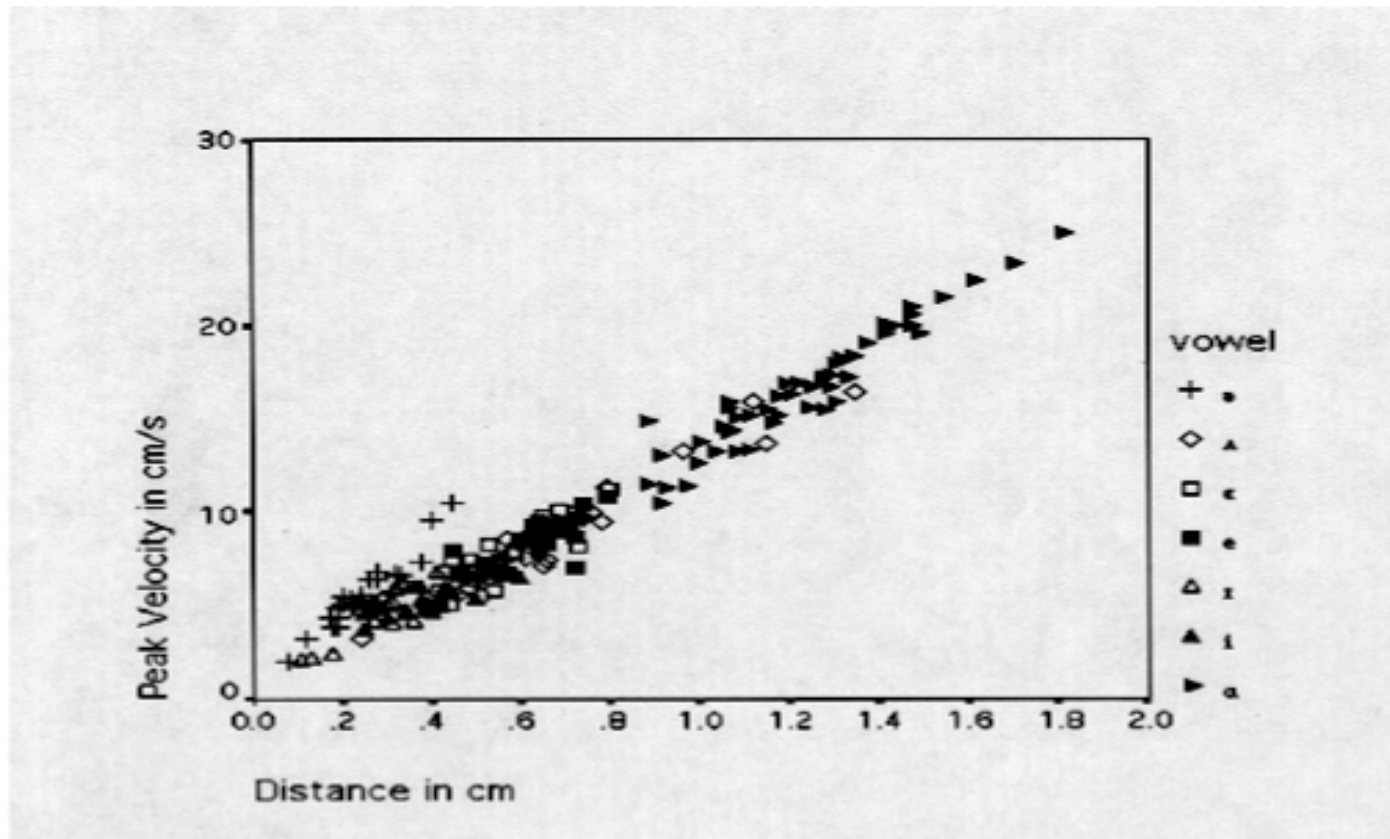
The difference between e.g. [mes] in *mace* vs. *mason* and *masonry* was reduced when not phrasally stressed (as compared to phrasally stressed versions).



Polysyllabic shortening is not obligatory

- Maybe more likely to be used in prominent contexts where signaling word boundaries is more important?
- Not used in Finnish because word boundary locations are already signaled by fixed word-initial stress?

Compression: higher peak velocities for larger distances



Doesn't result in surface isochrony (cf. intrinsic segment duration differences)

Other types of Compression effects:

Polysegmental shortening

- Munhall, Fowler, Hawkins & Saltzman 1992 (e.g. *baps* vs. *bap*).
- coda [ps] 97 ms longer than coda [p]
- Shorter vowel by only 13 ms
- Compression occurred but was very weak

Summary so far

- Evidence for rhythmic principles in normal speech production is weak
 - Rhythmic principles in phonology exist, but operate within **words** and **word-based constituents** (and therefore may not be purely rhythmic)
 - Compression effects exist, but
 - Operate preferentially in word-based constituents
 - Don't result in surface isochrony
 - Don't occur in all prominence contexts
 - Are optional in some languages

Recent renewed interest in rhythm classes

- Development of surface “rhythm” measures that differentiate e.g. English vs. Spanish vs. Japanese
 - Proportion of vocalic intervals (%V)
 - Standard deviation in duration of vocalic intervals (ΔV)
 - Standard deviation in duration of consonantal intervals (ΔC)
 - Pairwise variability measures (average difference/ratio between successive vocalic or intervocalic intervals)
- (Low & Grabe 1995, Low, Grabe & Nolan 2000, Ramus, Nespor & Mehler 1999)

Rhythm typology is problematic...

- Measures suggest a continuum of languages, rather than 3 distinct classes
- Measures are to some extent speaker, material, and rate dependent
- Many of the surface timing characteristics can be attributed to differences that may have little to do with rhythm per se (and everything to do with systematic relationships between
 - abstract cognitive factors and surface measures
 - physical factors and surface measures

Surface timing differences among languages are likely to be due to

- Differences in phonological representations, phonotactics, phonetic implementation conventions.
- These result in surface timing characteristics that are different for different languages (but are not *rhythmic* per se)
- Daur 1983

So is normal speech rhythmic?

- Not if rhythmicity refers to the regular recurrence of surface events.
 - Can't move to normal speech like we can move to music
- Speech does show systematic relationships between
 - Cognitive structures and surface timing patterns
 - General physical and cognitive principles and surface timing patterns
- Purpose of these systematic relationships appears to be to signal *meaning* via the word-based constituents of the prosodic hierarchy.
 - Cf. purpose of timing patterns in music is different: so others can synchronize with it.

General summary

- Surface timing patterns relate to segmental effects, segmental context effects, rate of speech effects, prosodic effects. Some of these effects may be due to general, non-linguistic principles, but many are linguistic in nature—i.e. related to linguistic structure and may be produced in speech- and language- specific ways.
- Some evidence of “motor equivalence” among timing implementation strategies.
- Prosodic effects show up in special places in the signal: Constituent edges and on prominent syllables. Available evidence consistent with the view that targeted stretches of speech are best described in traditional structural terms.
- Also some evidence for compression effects within constituents (but only in certain contexts, optional cross-linguistically).
- Timing effects related to prosodic structure appear to be part of a “bag of tricks” that speakers use to signal meaning via word-based constituents.
- Speakers can use timing skills to achieve surface rhythmicity, but often do this in a way that does not destroy the timing patterns that signal word-based constituency.

Thanks

Rhythmic principles in phonetics?

- In Phonetics
 - Systematic relationship between structure and surface patterns at many levels
 - E.g. final lengthening, intrinsic vowel durations, etc....
 - None result in surface isochrony.
 - Compression as more material occurs within larger constituents
 - Higher peak velocities for longer distances
 - Polysegmental shortening
 - Polysyllabic shortening
 - These do not result in surface isochrony

Is there any sense in which speech can be described as rhythmic?

	Music	Normal Speech
Regular recurrence of events	Yes	No
Listeners perceive isochrony	Yes	?
Listeners can synchronize with it	Yes	No
Timing of upcoming events is predictable from previous events	Yes	No
Systematic relationship between structure and surface timing patterns	Yes	Yes
Compression as more units occur within constituents	Yes	Yes
