Is speech timing systematic?

What do speakers use duration for?

What is timed?

Is speech timing systematic?

- Strategy: Identify factors that affect speech timing in systematic ways.

- Factors that affect speech timing = Things that speakers signal with duration

Factors affecting speech timing

- Segmental identity
- Segmental context
- Overall speech rate
- Prosodic structure

Go through each one, asking

- What is timed
- Representations involved in speech timing
- Control mechanisms involved in speech timing
  - General
  - Speech-specific

- For today, assume that opening and closing movements in CVC is the affected domain (gross simplification)

Is speech timing systematic?

- Importance of controlled experiments
  - Small but systematic effects may be difficult to find in spontaneous speech corpora
    - E.g. Subtle effects of morphological boundaries observed at slow rates in controlled experiments
      - *bob* in *e.g. bobbing* 5-10% longer than *bob* in *bobbin* at a slow rate (Sugahara & Turk 2009)
  - Some contexts are underrepresented in spontaneous corpora, e.g.
    - Few 4-syllable words; difficult to determine durational effects of more than 2 syllables
    - Most phrase-final words are also phrasally-stressed (in English); difficult to determine effects of finality independent of phrasal stress.

Segment identity

- Primarily acoustic studies, e.g. Lehiste, 1972; Klatt 1976)
- Controlled carrier phrases, e.g.
  - Say *dad* for me.
  - Say *did* for me.
  - Say ___for me.
- Design controls
  - Preceding and following segmental context
  - Prosodic context
- Studies typically have
  - Multiple repetitions
  - Randomized order

Segment identity

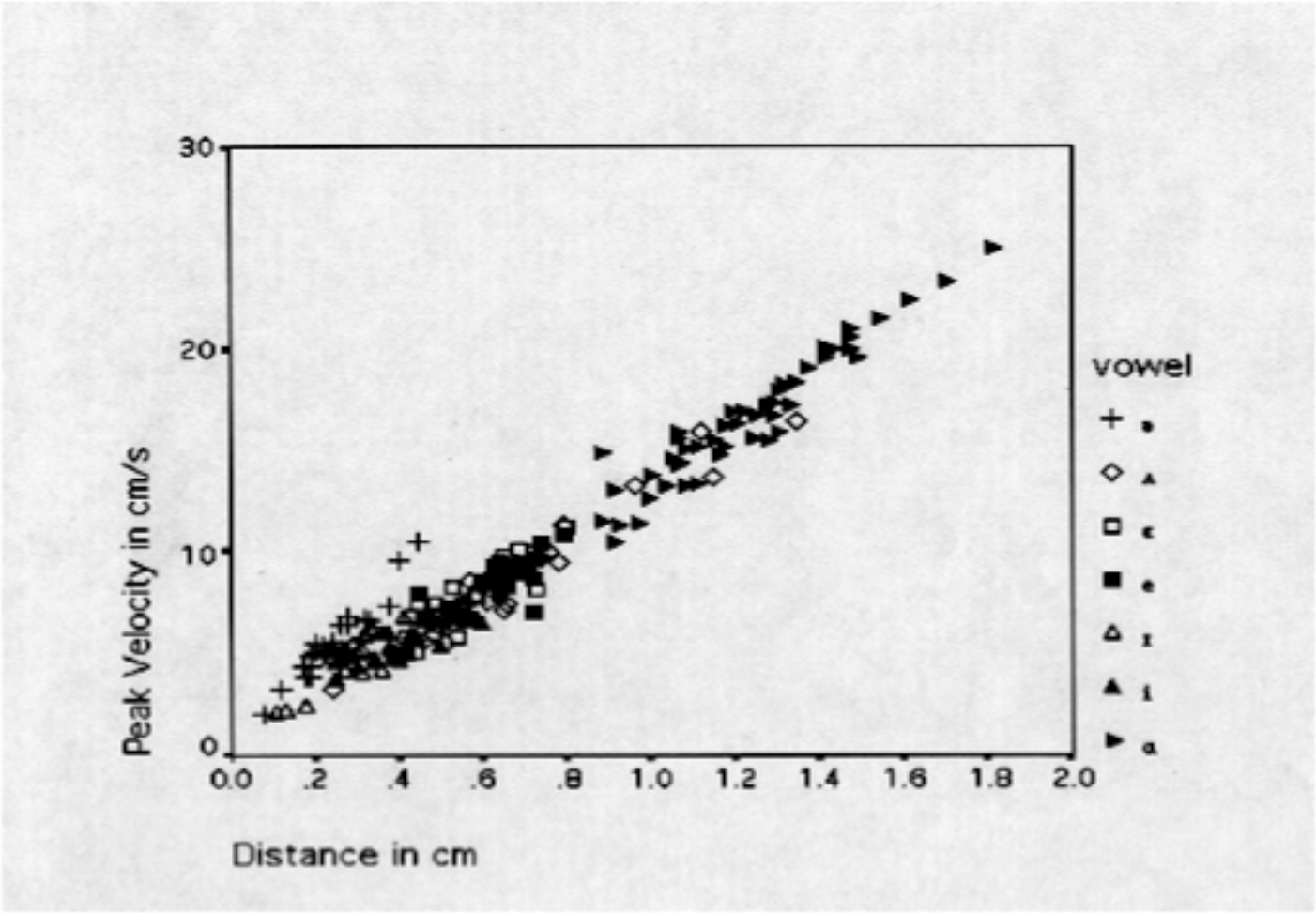
- For English (Klatt 1979), e.g.
  - High vowels shorter than low vowels
  - Front vowels shorter than back vowels
  - Diphthongs longer than monophthongs
  - Fricatives longer than stops
  - Etc.

Intrinsic segment durations

- Many co-occur with spatial differences
  - E.g. high vowels shorter than low vowels, monophthongs shorter than diphthongs
  - Duration differences co-occur with differences in quality

- To what extent are intrinsic segmental differences explicitly planned?
  - Could some of the differences be an artifact of the time it takes to reach different targets?

Relationship between peak tangential velocity and distance



Opening tongue tip movements for 6 full vowels and schwa, dVd frame; schwa in dVd again.

Are intrinsic segmental duration differences planned?

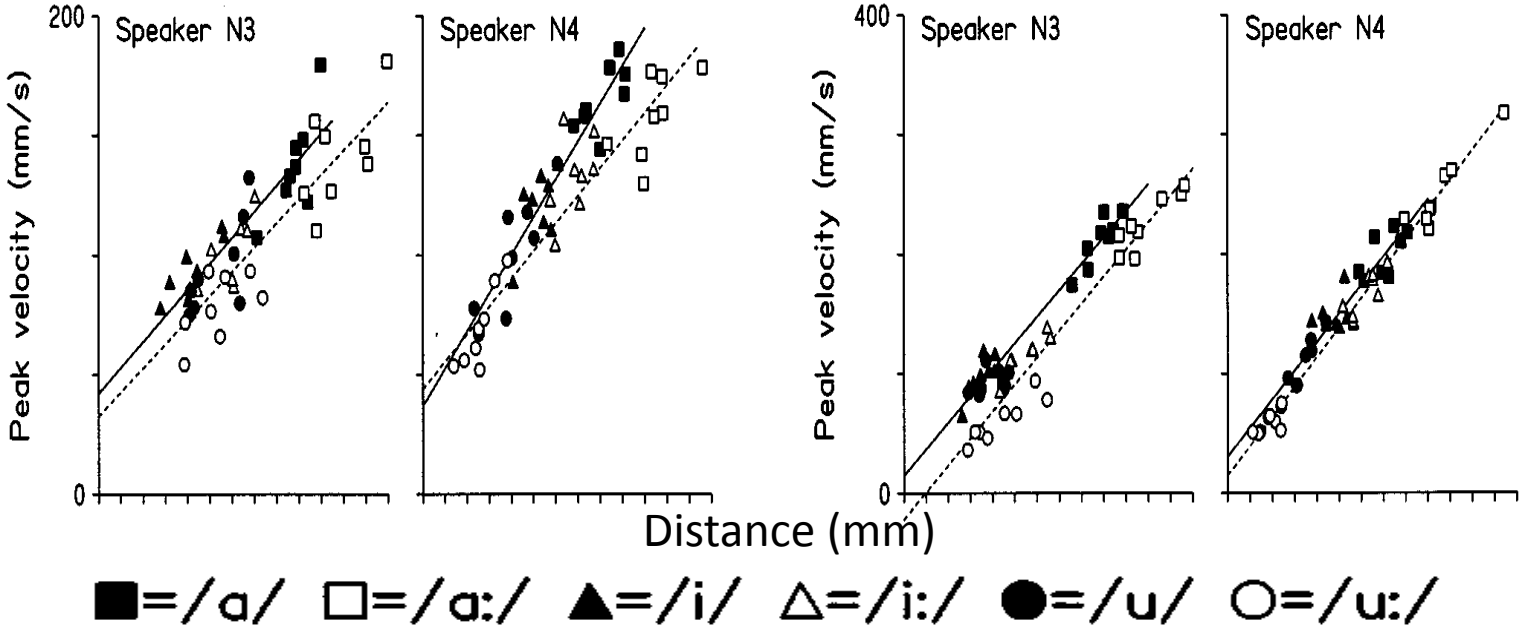
- Peak Velocity/Distance slopes are the same for all full vowels
- Peak velocities appear to be adjusted to compensate for distance traveled.
  - Planned to be the same duration?
  - Possibly. But surface durations are different.
- What general non-linguistic principles might account for systematic duration differences?
  - Precision of position requirements (high vowels may require less precision)
  - Spatial trajectories (curves take longer than straight lines)
  - For some consonants, aerodynamic requirements (e.g. shorter stop closures if closure must be completely voiced).
- General Question: To what extent can durational differences be explained by general non-linguistic principles?

Some durational differences are clearly planned

- Steeper Peak velocity/Distance slope (relatively faster for similar distance) for
  - Schwa as compared to full vowels in English
  - Phonologically short as compared to phonologically long vowels in German (Hertrich & Ackermann 1997)

Phonologically short vowels have faster opening movements for similar distance traveled (as compared to long vowels)—Hertrich & Ackermann 1997

- Speaker-dependent effects for closing movements.



What is timed?

- Interspeaker differences in articulatory strategy
  - Speakers differ in whether closing speeds are adjusted
  - Some vowel-dependent differences for some speakers (e.g. /u/ vs. /u:/ difference achieved via steady state manipulation for some speakers).
- Consistent with interval timing

Segmental context

- Best-studied example: Difference in English vowel duration before voiced vs. voiceless stops

- E.g. /u/ in *rude* longer than /u/ in *route*

General, non-linguistic accounts

- Production
  - closure for e.g. /d/ can be shorter than closure for /t/
  - Less coda overlap for e.g. /brood/ compared to /brute/
- Perception: Contrast effects (Kluender, Diehl & Wright 1988)
  - Longer preceding vowel duration will enhance the durational contrast between e.g. [u] and [d]
  - [d] will sound shorter (more /d/-like)

American English vowels before voiced stops have long steady states and, for some speakers, slower closing speeds (Summers 1987)

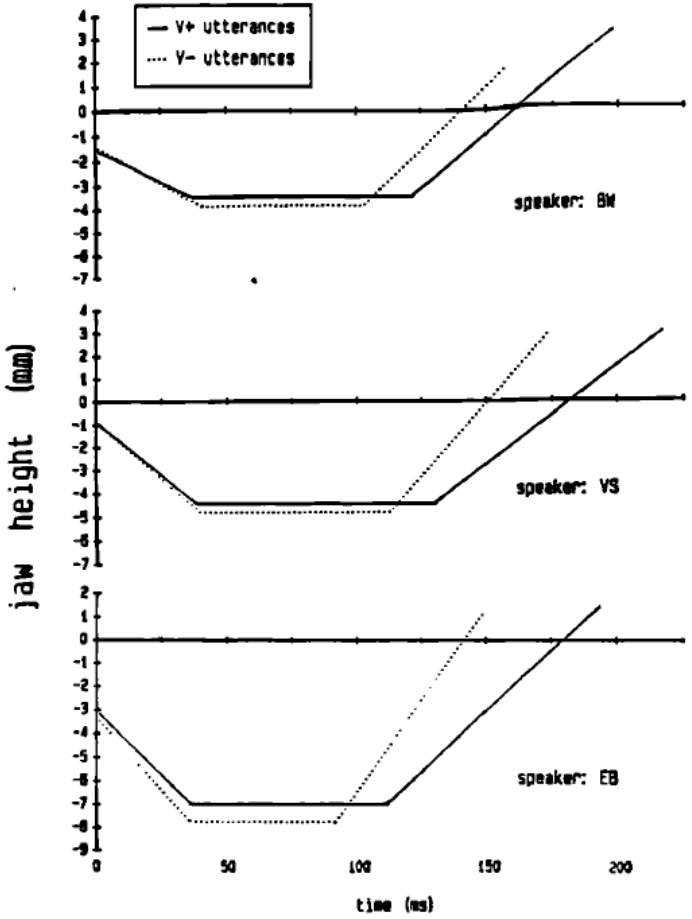


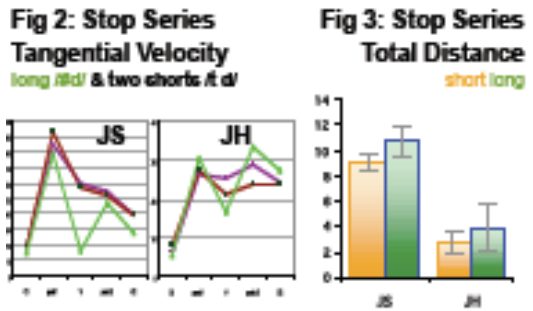
FIG. 6. Jaw position plots for utterances containing voiced ($V+$) versus voiceless ($V-$) final consonants, based on mean positions and mean durations listed in Table I.

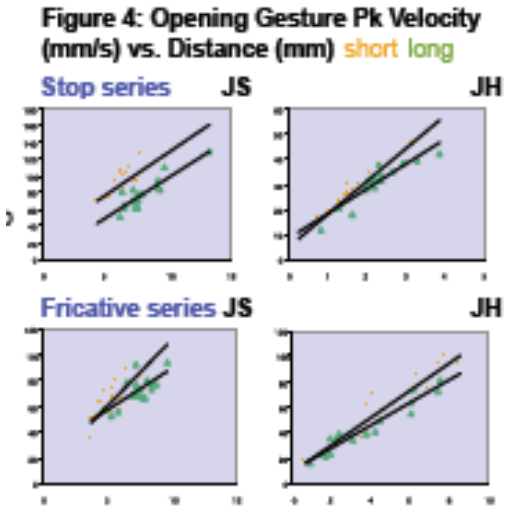
CVC; C = b,p,f,v
V = /ae, ɑ/

General accounts aren't the whole story

- Different varieties of English show different patterns
  - Scottish English (Scobbie et al.1999): Scottish Vowel Length Rule
  - Voicing effect for fricatives only:
    - *[u] in /rus/* short *(75-100 ms), [u] in /ruz/* long*(150-175 ms);*
    - [u] in *rude* and *route* both relatively short (ca 75-100 ms)
  - [u] before morpheme boundary is long
    - [u] in *ru+ed* long (before a morpheme boundary) (125-175 ms)

Opening movement pv/dist and steady states are affected





Cf. steady state and closing movement effects in Am. English (Summers 1987, de Jong)

Speech Rate

- Overall, global rate
  - What people refer to when they say someone speaks fast, or slow
  - Not to be confused with more local durational manipulations
- Often measured in syllables per second
- Some distinguish Speech rate (includes pause durations) vs. Articulation rate (rate without pause durations)
  - NB. Pause durations are difficult to measure

Effects of speech rate manipulations

- Fewer/shorter pauses at fast rates
- Vowel intervals are affected more than constriction intervals for obstruents
  - Vowels more “elastic” than consonants
- Rate manipulations are not just uniform re-scaling of the acoustics!

Articulatory strategies for manipulating rate

- For fast rates, articulatory movements can show
  - Smaller distances, and/or
  - Faster movements for a given distance, and/or
  - Greater articulatory overlap with other movements

  as compared to slow rates
- A combination of strategies suggests interval timing, not necessarily just a change of movement speed.

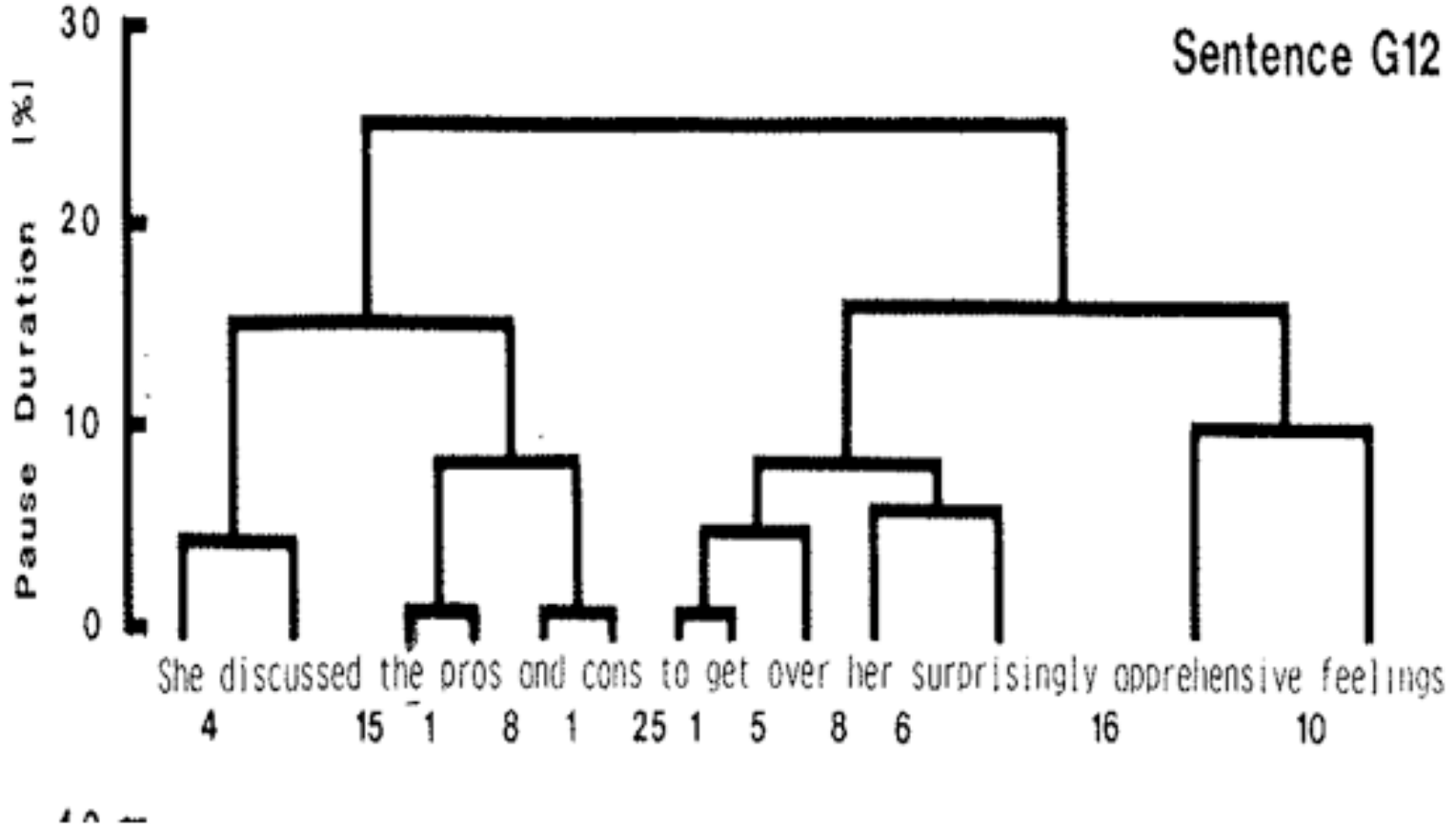
Speech rate: linguistic aspects

- Appears to interact with prosodic structure
  - fewer, less salient phrase breaks at fast rates (Caspers 1994, Strangert 2003)
  - May interact with prosodic prominence structure

Prosodic structure: Constituents and Prominences

- Gee & Grosjean's (1983) study of speech at very slow rates
  - Pause duration variation can be predicted by a hierarchical structure of constituents

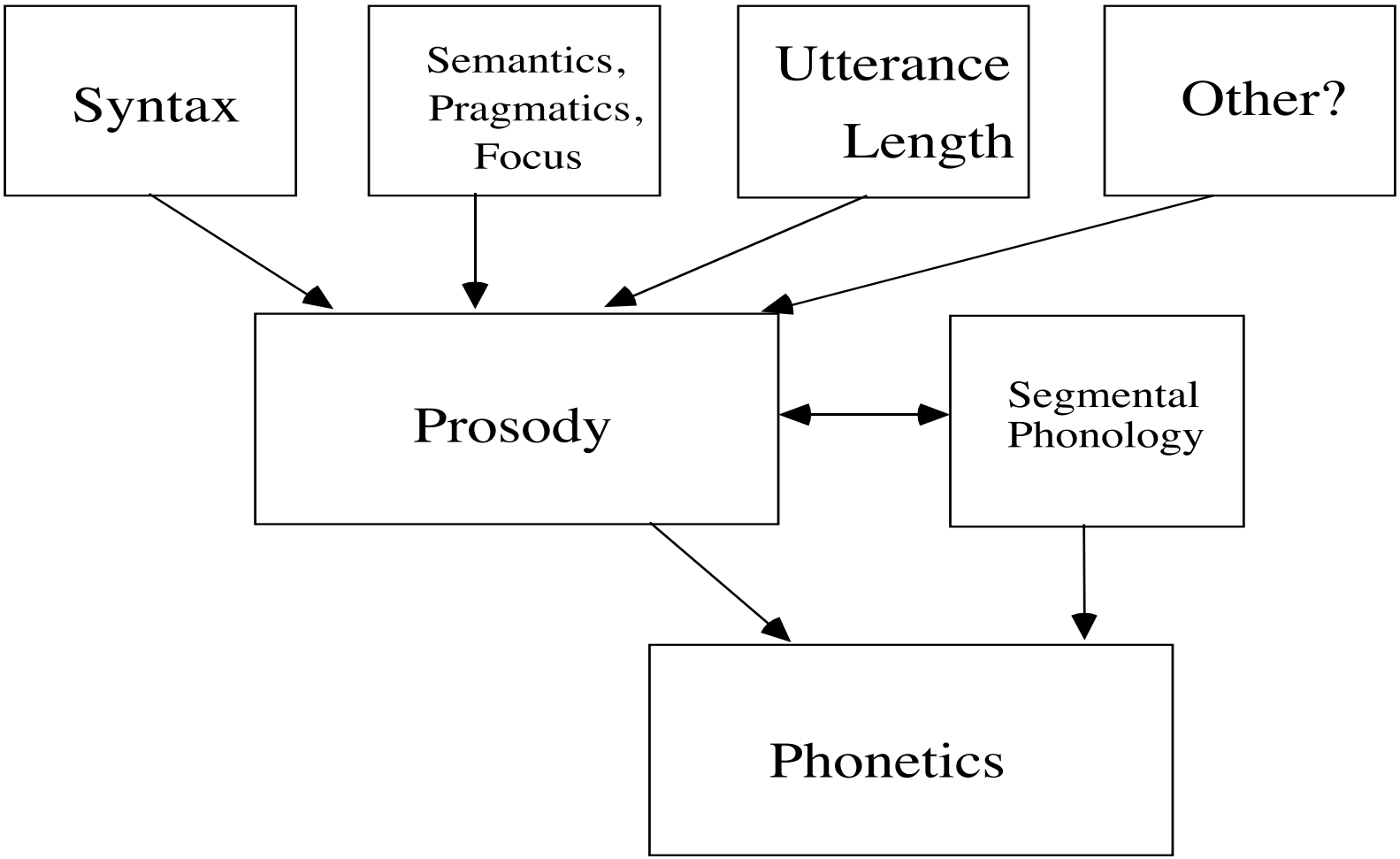
- Performance structures (= prosodic constituent structure)


Sentence G12
Pause Duration [%]
30
20
10
0
She discussed the pros and cons to get over her surprisingly apprehensive feelings
4 15 1 8 1 25 1 5 8 6 16 10

Prosodic constituent structure

- Related to syntax
  - (Old men) and women vs. Old men and women
- Not isomorphic with syntax
  - *Sesame Street is brought to you by…the Childrens' Television Workshop*
- Shows effects of factors other than syntax, e.g.
  - Pragmatic focus
  - Symmetry
  - Length

Prosodic structure and other components of grammar



Boundary strength can be transcribed impressionistically in running speech (Wightman et al. 1992)

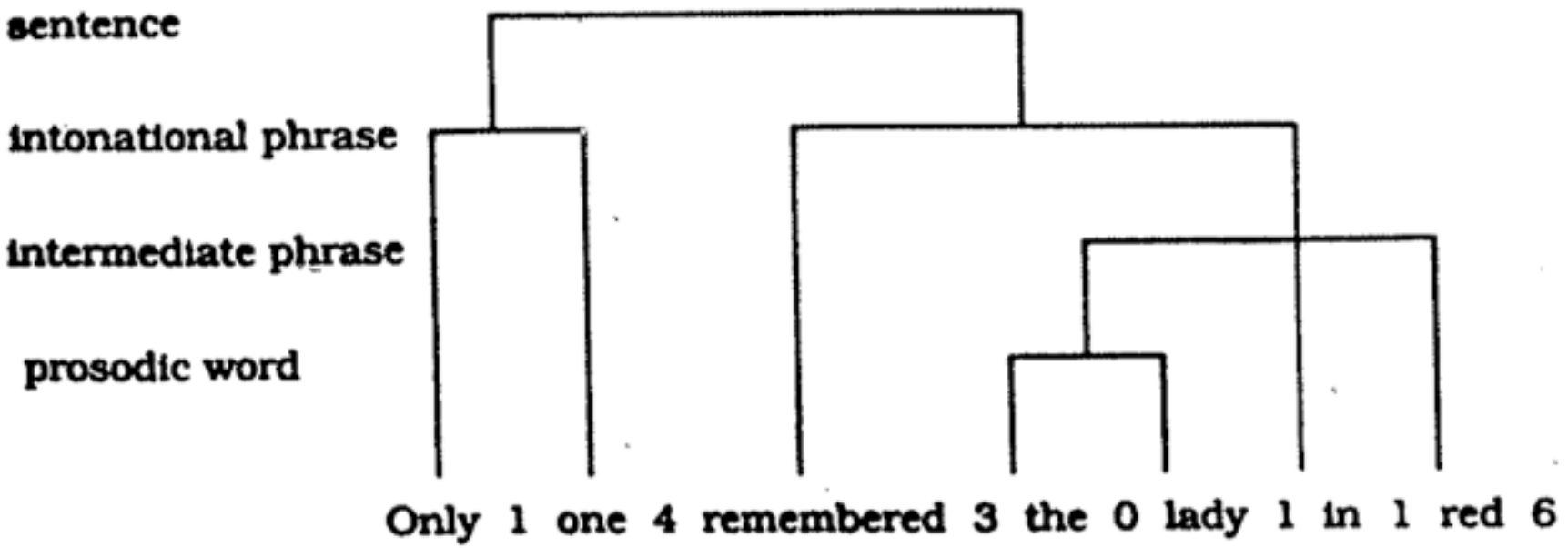


FIG. 1. A sample sentence from the corpus showing the prosodic labels (break indices) transcribed by human listeners. The prosodic structure implied by these labels is shown above the transcription.

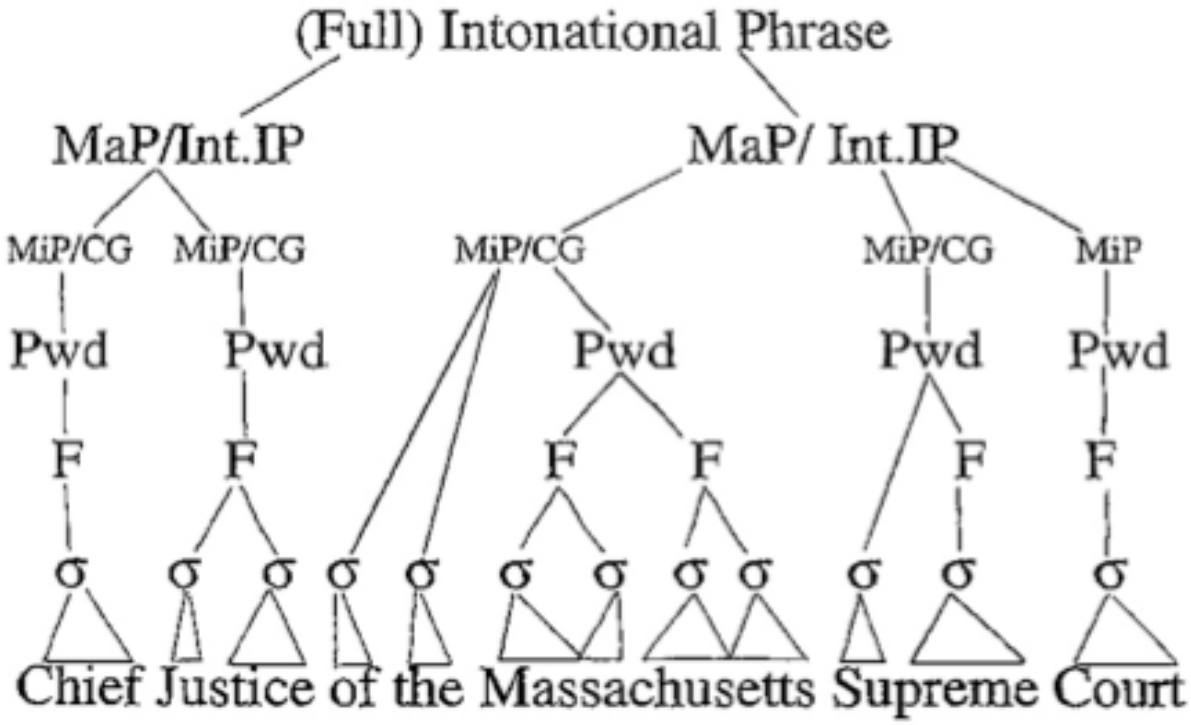


Figure from Shattuck-Hufnagel & Turk 1996

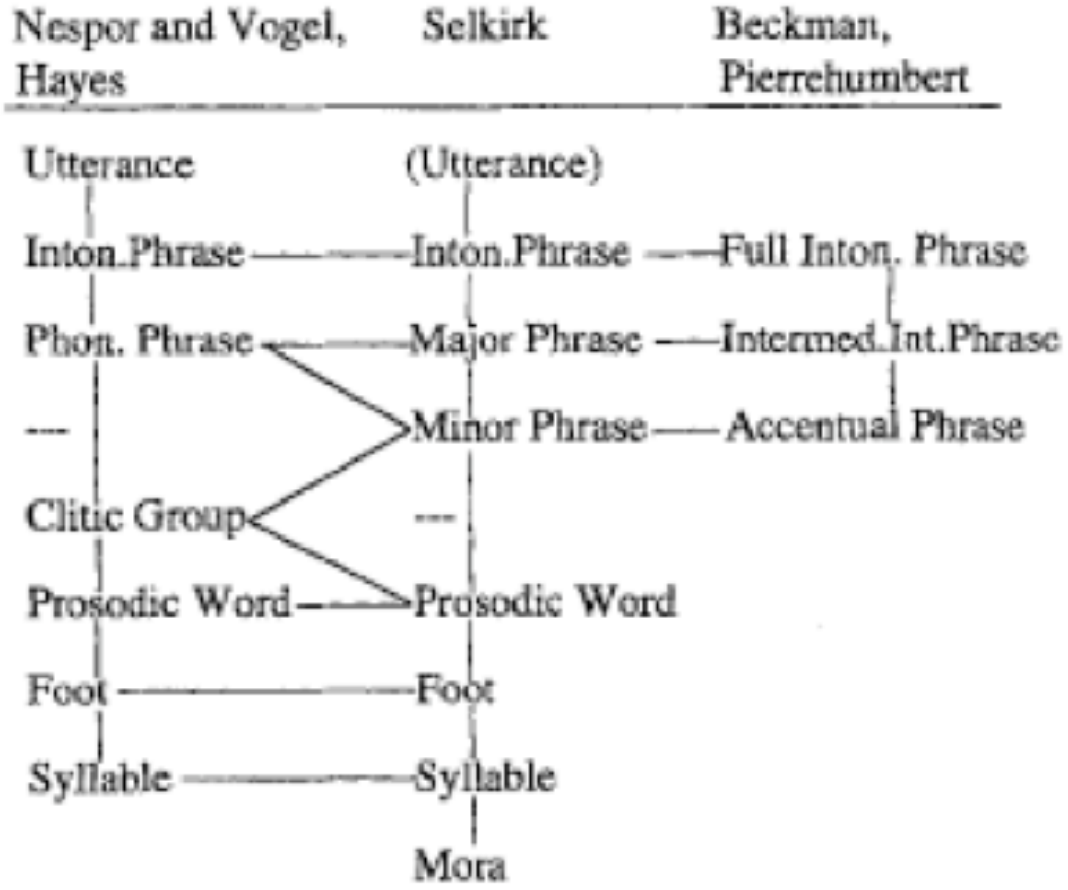


Fig. 2. Prosodic constituent hierarchies from the literature; additional important theories, such as those of Halle and Vergnaud (1987), Liberman (1975), Liberman and Prince (1977), Gussenhoven (1988) and others are discussed in the text.

Fig. from Shattuck-Hufnagel & Turk 1996

Hierarchical constituent structure

- Predicts more than just pausing behaviour
  - Breathing
  - Segmental sandhi
  - Intonational boundary tones
  - Final glottalization
  - Distribution of phrasal prominences (to some extent)
  - **Final lengthening**
  - **Initial strengthening/lengthening**
  - **Polysyllabic shortening**

Hierarchical constituent structure in the lab

Often elicited by manipulating syntax (Cambier-Langeveld 1997):

(1) Prosodic Word-boundary: *Piet wil die rare* ***rododendron****planten, gek als hij is.*
'Piet wants those strange rhododendron plants, crazy as he is.'
Phonological-Phrase boundary: *Piet wil die rare* ***rododendron*** *planten, gek als hij is.*
'Piet wants to plant that strange rododendron, crazy as he is.'
Intonational-Phrase-boundary: *Piet wil die rare* ***rododendron****, plantengek als hij is.*
'Piet wants that strange rododendron, plant-crazy as he is.'
Utterance –boundary: *Plantengek als hij is wil Piet die rare* ***rododendron****.*
'Plant-crazy as he is, Piet wants that strange rhododendron.'

Can also be elicited by manipulating length

- From Kainada (2009), Greek

[a'fu 'eçis em'fanisi **ci'notopi**]$_{Sub}$, [**ko'pela** ðen θa 'vɾis]$_{Main}$.

'Since you have a common appearance, you will not find a girl.'

- Vs.

[a'fu 'eçis ka'ta ɟeni'ci omolo'ɟia em'fanisi **ci'notopi**]$_{Sub}$, [**ko'pela** ðen θa 'vɾis xo'ɾis ɾizi'ci ala'ɟi]$_{Main}$.

'Since everyone agrees that you have a common appearance, you will not find a girl without a radical change.'

- Effects are qualitatively similar to effects based on syntactic manipulations (Kainada 2009, Astésano, Bard & Turk 2007, Watson & Gibson 2004)

Example: Breathing

- Modified when speaking.
- When not speaking, on average
  - We spend 40% of the time inhaling and 60% exhaling
  - We spend 10% of the time inhaling, and 90% exhaling (Perkins & Kent 1986)
- Breathing pauses tend to co-occur with prosodic boundaries (Grosjean & Collins 1979, Slifka 2000)
  - As do non-breathing pauses
- Breathing adapts to prosodic structure (rather than the other way around).

Breathing patterns: Rib cage movement (Grosjean & Collins 1979)

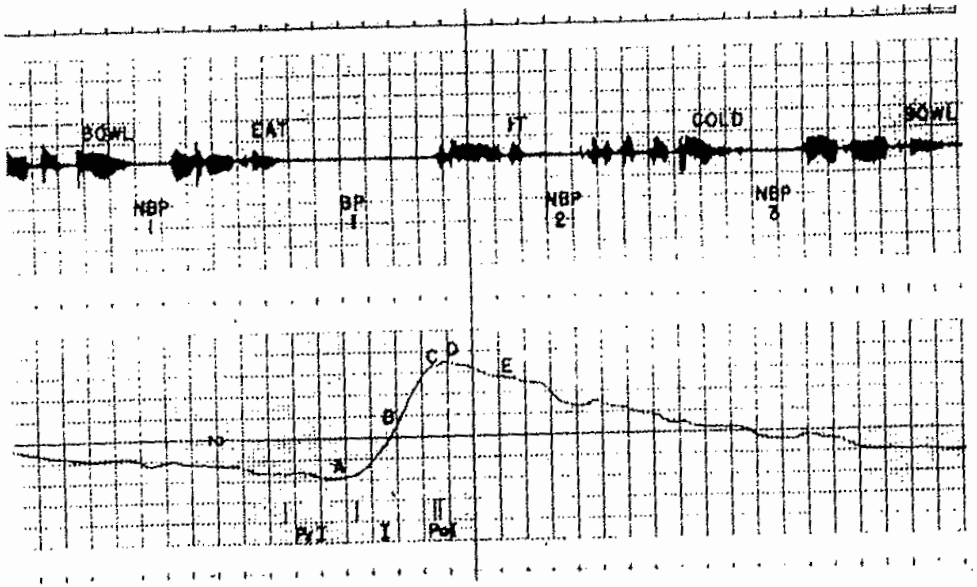
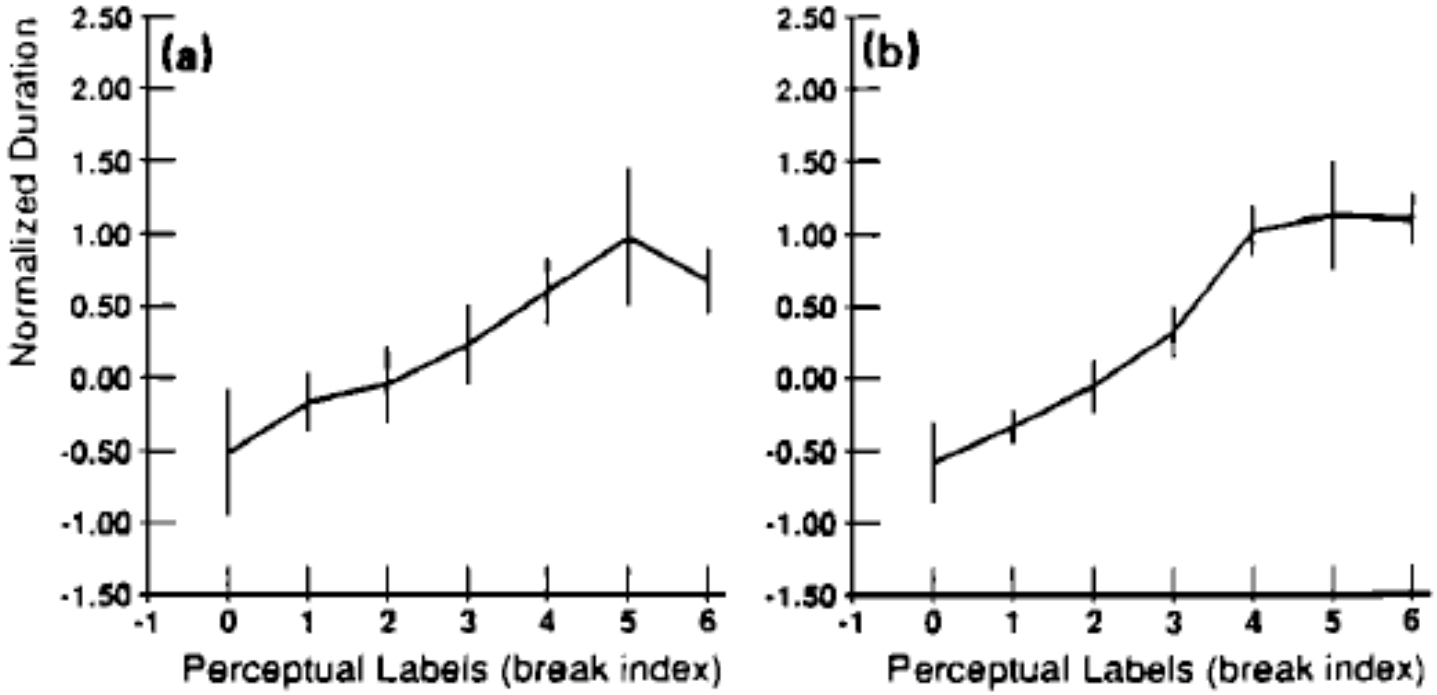


Fig. 1. A sample of the oscillographic recording of a subject's speech and breathing patterns. Three NBPs and one BP occur. The latter has been subdivided into its three components: preinspiration (PrI), inspiration (I), and postinspiration (PoI). The section of the passage that is of interest here is: '...saw a big bowl [NBP1] and began to eat [BP1]. She didn't like it [NBP2] because it was too cold [NBP3]. She went to the next bowl...'.

Final lengthening: Magnitude is proportional to boundary depth

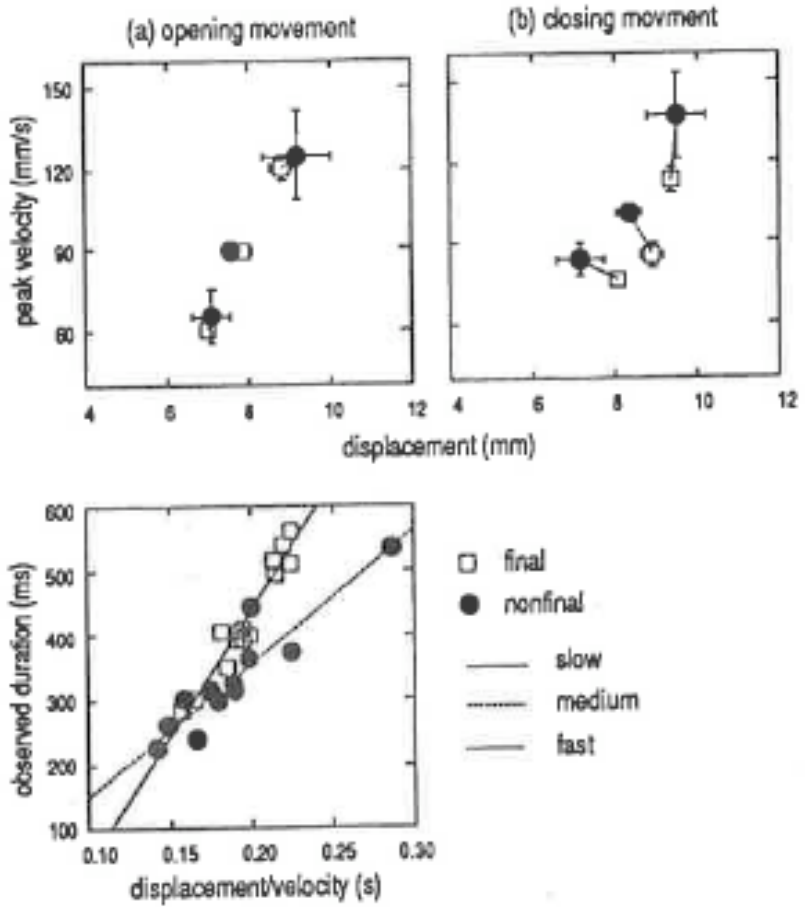
- From Wightman et al. 1992



Articulatory strategies for implementing final lengthening

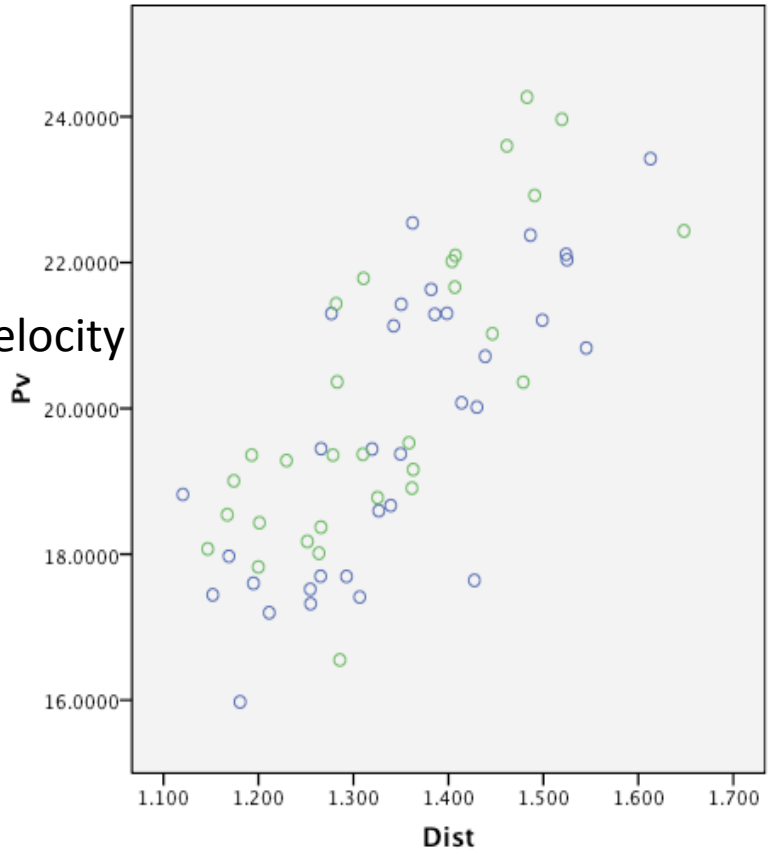
- Edwards, Beckman & Fletcher 1991
- Study of *pop* in medial and final positions, 4 speakers
  - At fast and normal rates
    - Slower movements towards final targets (Slower pv/dist relationship for closing movements)
    - Some differences in steady state duration
    - Differences in distance for closing movement
  - At a slow rate
    - No difference in peak velocity between medial and final tokens
    - Difference in steady state duration
- No single articulatory mechanism: Interval timing?

Articulatory strategies for implementing final lengthening (Edwards, Beckman & Fletcher 1991)

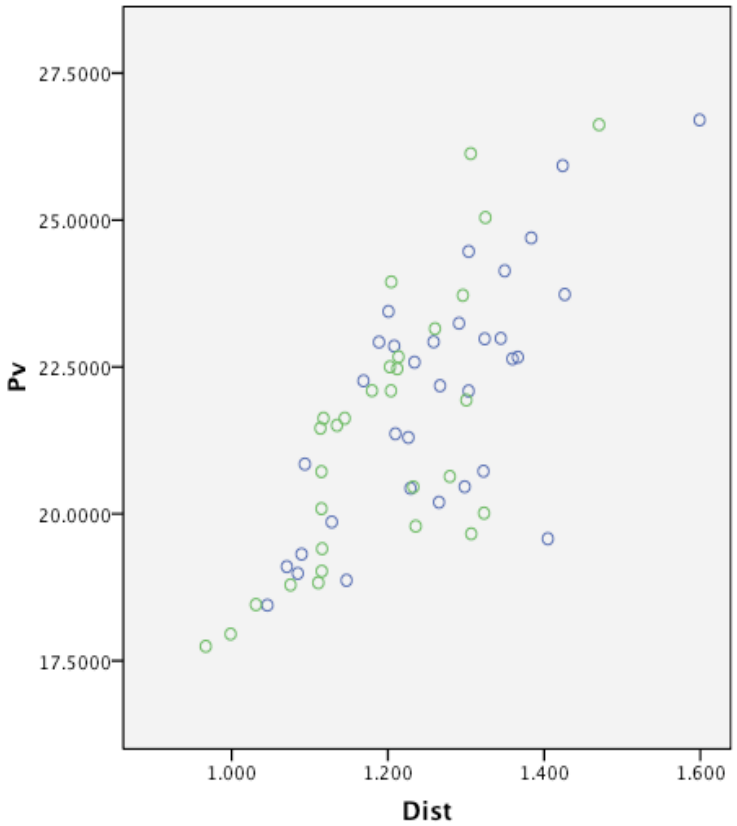


Please say a dad (blue circles--final) vs. Say a dad again (green circles-medial).

Opening



Closing

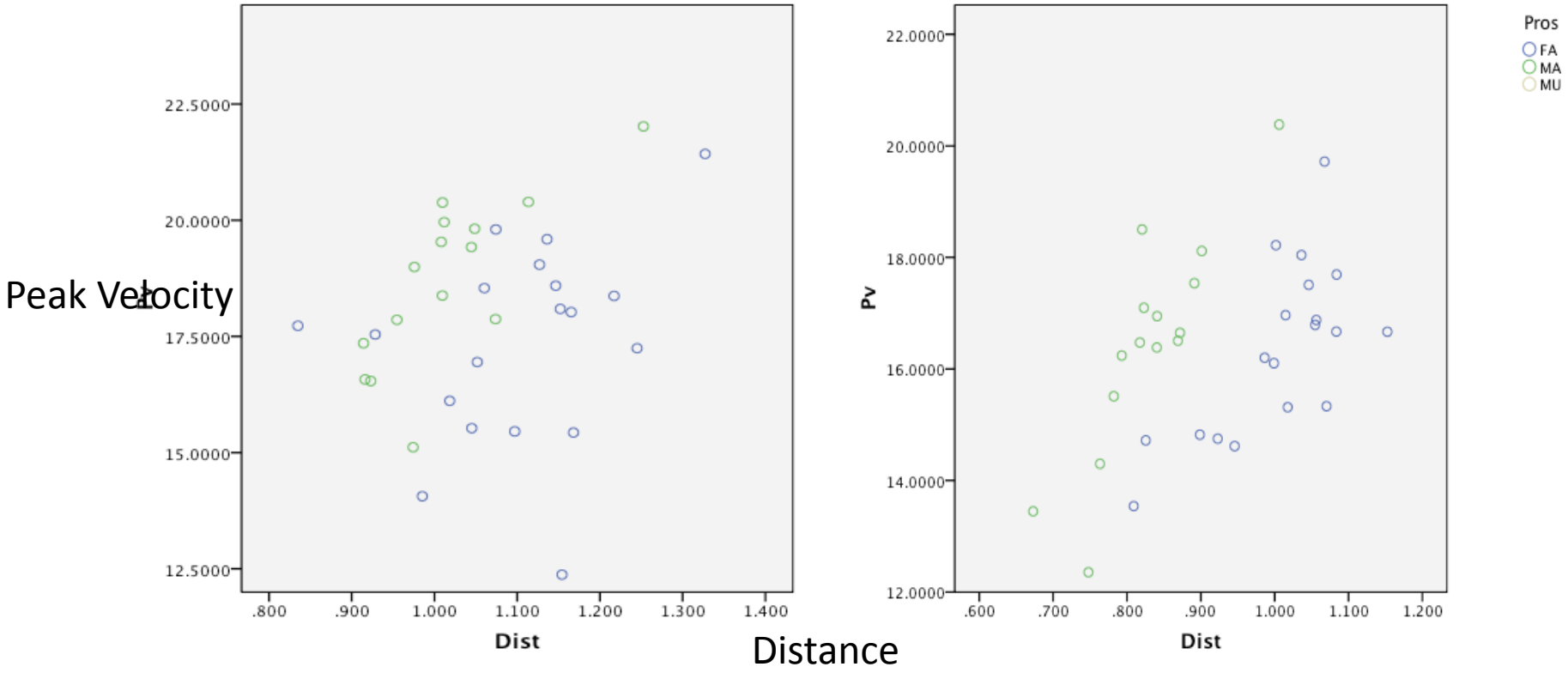


Peak Velocity

Distance

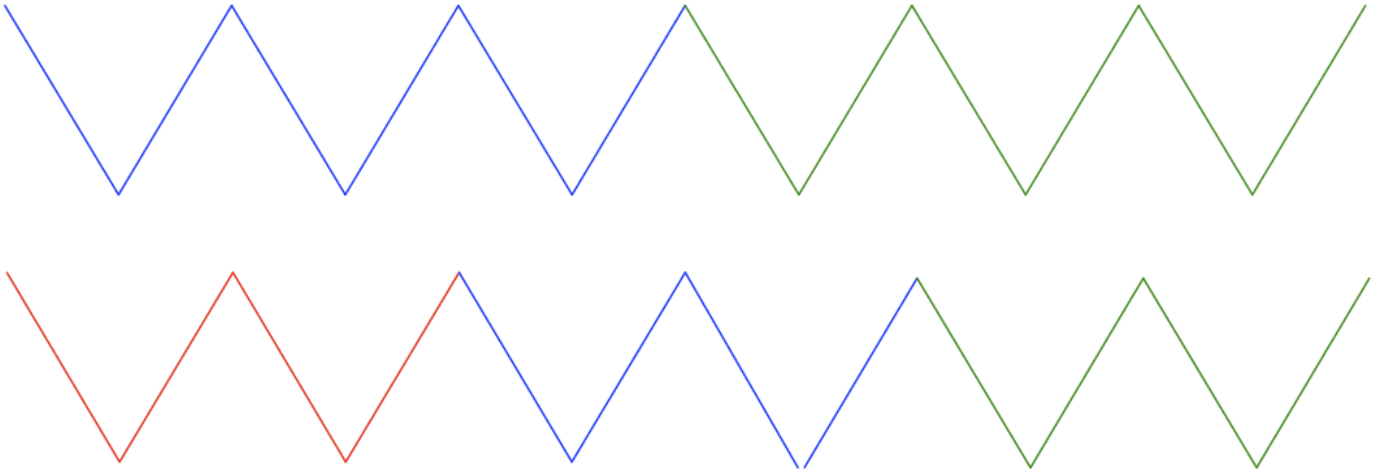
Please say a dead (blue circles—final) vs. *Say a dead again (green circles-medial)*.

- Peak velocity/dist relationship differs according to position



Final lengthening: General non-linguistic mechanisms?

- Shattuck-Hufnagel & Turk study of finger movements while tracing zigzags on paper



- Group-final lengthening for tracing zigzags?

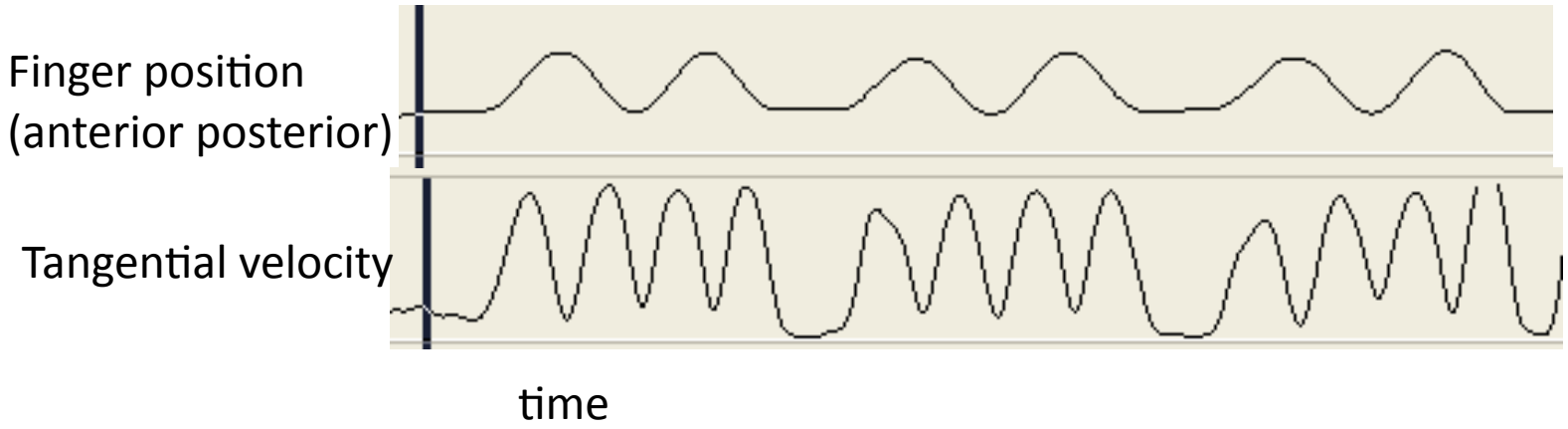
Final lengthening: General non-linguistic mechanisms?

2 tracers so far

No difference in peak tangential velocity

Lower group final minimum tang. Velocity

1 tracer measured in detail: Some subtle (ca. 10%) final lengthening because it takes slightly longer



This example suggests marginally lower peak velocities at group onset (cf. initial lenthening?)

Prosodic prominence structure

```
          X
X         X
X   X   X  X
```

Condensation

- Lexical stress on 1st and 3rd syllables of *condensation*
- Primary phrasal prominence associated with the syllable bearing primary lexical stress, optional "pre-nuclear" prominence on 1st syllable.

- *Did you say MORE condensation or LESS condensation*?
  - Lexical stress on 1st and 3rd syllables of *condensation;* distinction between stressed vs. unstressed is signaled phonetically via full vs. reduced vowel distinction
  - No phrasal prominence on *condensation*.

Phonetic correlates of prominence structure include

- Word-level stress:
  - **Duration**
  - Full vs. reduced vowels (in some languages)
  - Spectral tilt (understudied but cf. Sluijter & van Heuven 1996)
  - F0 (in some languages, e.g. Tunisian Arabic (Bouchhioua 2009)
  - Greater number of segmental contrasts (cf. Kingston, today)
- Phrasal prominence:
  - **Duration**
  - Phrasal pitch accents can be associated with stressed syllables (in some languages)
  - Spectral tilt
  - F0

Prominence: Articulatory strategies

- Reduced vs. full vowels: Differences in
  - Distance
  - Peak Velocity/Distance relationship

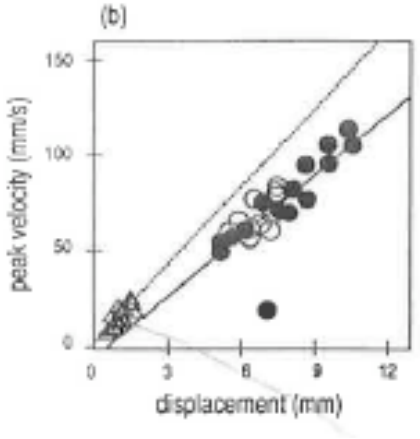


Figure from Edwards, Beckman & Fletcher 1991

Phrasal prominence

- Summers 1987
- CVC; C = /b,p,f,v/
- Steady state duration
- Distance (all speakers but 1)
- Pvelocity/Distance relationship of closing movements (speaker-specific)

Summers 1987

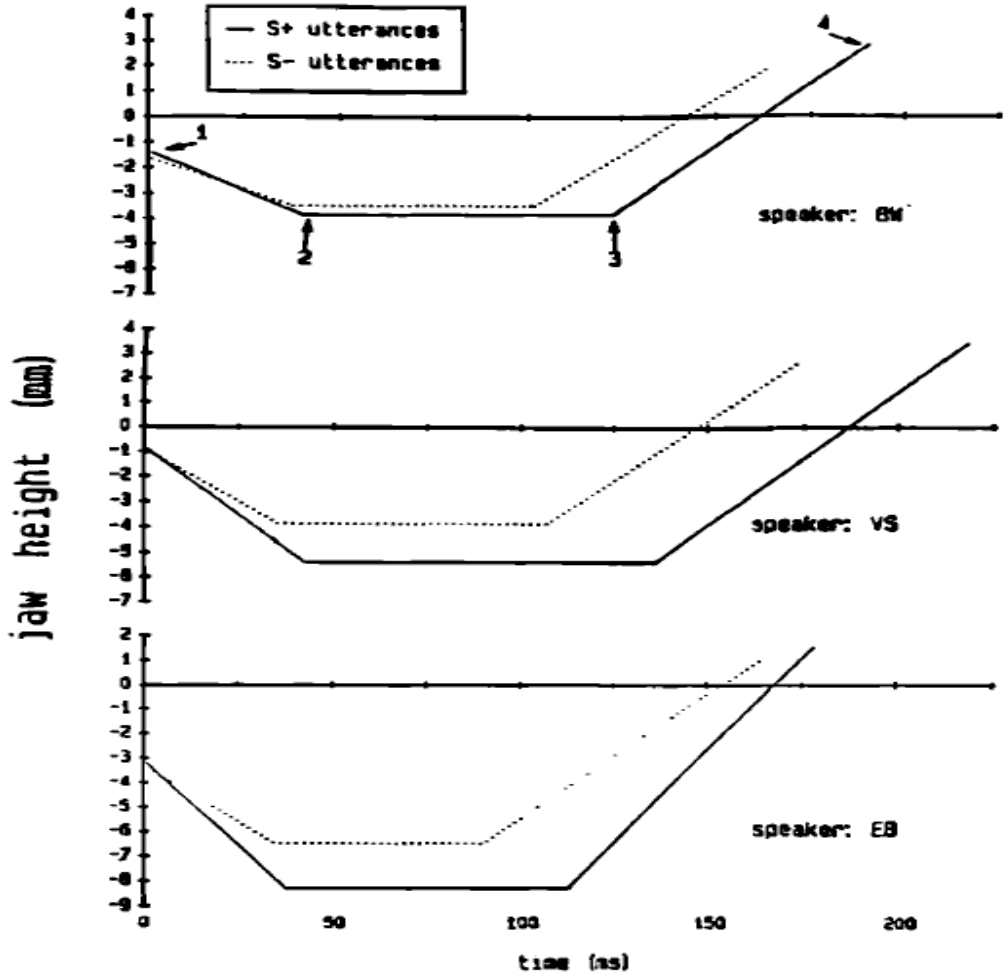


FIG. 5. Jaw position plots for stressed ($S+$) versus unstressed ($S-$) utterances, based on mean positions and mean durations listed in Table I.

Conclusion

- Speech shows systematic relationships between
  - Phonological representations and surface timing patterns
  - At multiple levels
    - Segmental
    - Prosodic

  Also Global effects of rate
  - Available evidence is consistent with interval timing, but studies and numbers of speakers are few.

References

• Beckman, M.E. & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. *Laboratory Phonology I, 152-178.*

• Beckman, M. E., & Edwards, J. (1992). Intonational categories and the articulatory control of duration. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure. Tokyo: OHM Publishing Co, 356-75.*

• Bouchhioua, N. (2008). The acoustic correlates of stress and accent in Tunisian Arabic: A comparative study with English. PhD Dissertation. Université de 7 Novembre, Carthage, Tunisia.

• Edwards, J., Beckman, M.E. & Fletcher J. (1991): The articulatory kinematics final lengthening. *Journal of the Acoustical Society of America, 89, 369-382.*

• Gee, J.P., Grosjean, F.E., 1983. Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15, 411–458*

• Grosjean, F. & Collins, M. (1979). Breathing, pausing and reading. *Phonetica 36,* 98-114.

• Hertrich, I., & Ackermann, H (1997). Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *Journal of the Acoustical Society of America,* 102, 523-536.

• Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *JASA 59,* 1208-1220.
• Klatt, D. H. (1979). Synthesis by rule of segmental durations in English sentences. In. B. Lindblom & S. Öhmann (eds.). *Frontiers of speech communication research. New York:*
• Academic Press, 287-300.

• Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018-2024.

• Ohala, J. J. 1997. Aerodynamics of phonology. *Proc. 4th Seoul International Conference on Linguistics [SICOL] 11-15 Aug 1997. 92-97*

• Perkins, W.H. & Kent, R.D. (1986). Functional anatomy of speech, language, and hearing: a primer. Allyn and Bacon. .

• Scobbie, J. & Turk, A. (2002). **An Articulatory (EMA) Study of English Voicing-Based Differential Vowel Duration. Poster presented at BAAP, Newcastle.**
• Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25, 193–247*
• *Slifka, J. (2000). Respiratory constraints on speech production at prosodic boundaries. PhD Dissertation. MIT..*
• Sugahara, M. & Turk, A. (2009). Durational correlates of sub-lexical constituent structure in English. *Phonology* 26. 477-524.

• Summers, W.V. (1987): Effects of stress and final consonant voicing on vowel production: Articulatory and acoustic analysis. *Journal of the Acoustical Society of America, 82, 847-863.*

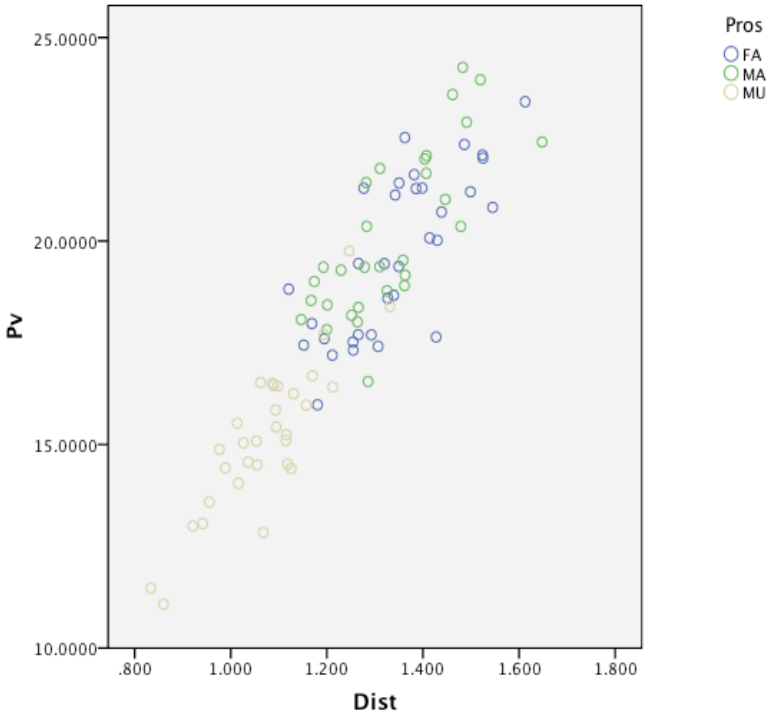
• Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91, 1707–1717.*

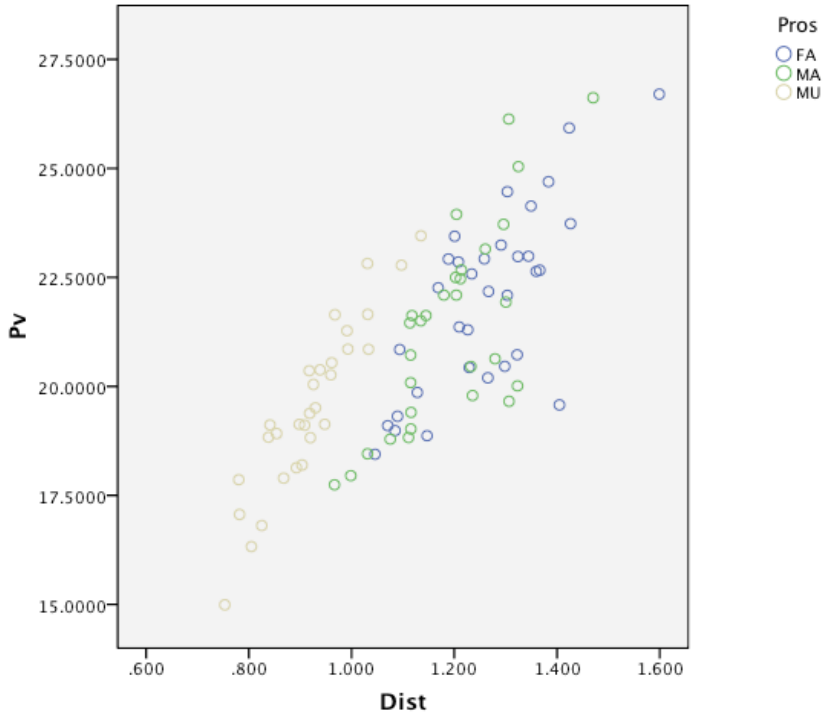
Extra--Different strategies for different functions?

- Beckman & Edwards' study suggests yes
- Effect of rate:
  - Peak velocity changed for little change in amplitude
- Effect of accent:
  - Some change in displacement
  - Change in phasing (longer steady state)—closing movement phased later for accented syls.
  - Unaccented syllables—closing movement truncates opening movement and thus causes change in displacement, as well as some change.
  - Schwa vs. full vowels--
- Effect of finality:
    - Fast and normal rates: change in peak velocity for closing movement.
    - Slow rate: No change in peak velocity, but a change in steady state duration (limits on how slow a movement can be—beyond this limit, need to change steady state)

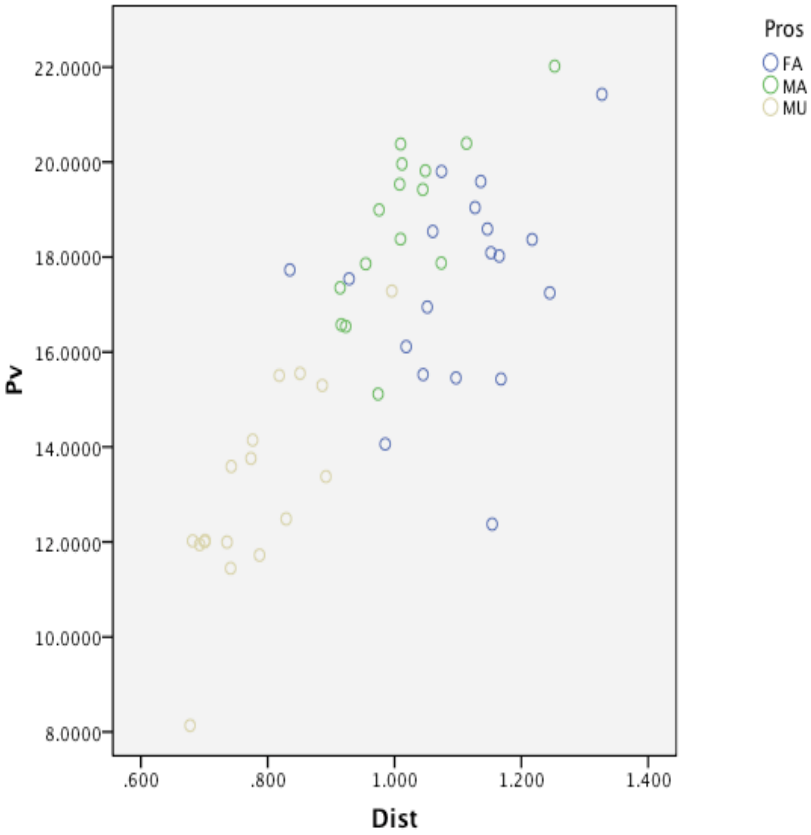
Extra--Different strategies for different functions?

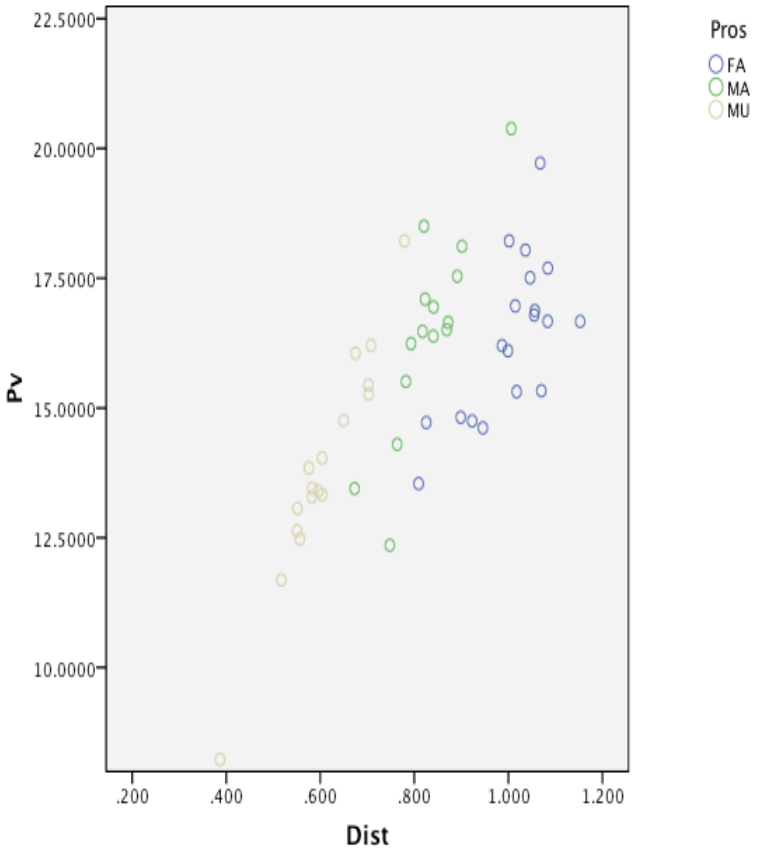
- Turk pilot data for one speaker—Tongue Tip
  - *dad* Utterance-final Accented (FA), Phrase-medial Accented (MA), Phrase-medial Unaccented (MU)





Extra--Dead same speaker





Extra-Different strategies for different functions?

- Possibly, but implementation is speaker-, and to some extent segment- specific.
- Consistent with interval timing
- Speakers may adopt particular strategies for implementing particular functions, but can adapt these as demands dictate

Extra--Is Speech Timing Systematic?

- Yes
- Systematic effects of many kinds
  - Segmental
    - Intrinsic
    - Contextual
  - Rate
  - Prosodic structure
    - Constituent structure
    - Prominence structure

Extra--Is speech timing systematic?

- What is timed?
  - Inter-speaker, inter-segmental variability suggests intervals
- What kinds of control mechanisms do we use?
  - Many are specific to language (functionally related)
  - Some regularities may be due to general principles, e.g.
    - Longer intervals for more precise movements
    - Longer intervals for curved trajectories
    - Some (but not all) aspects of final lengthening