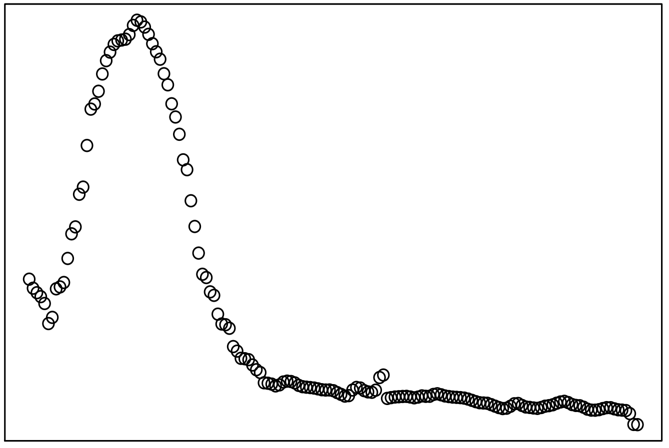
HISPhonCog 2019

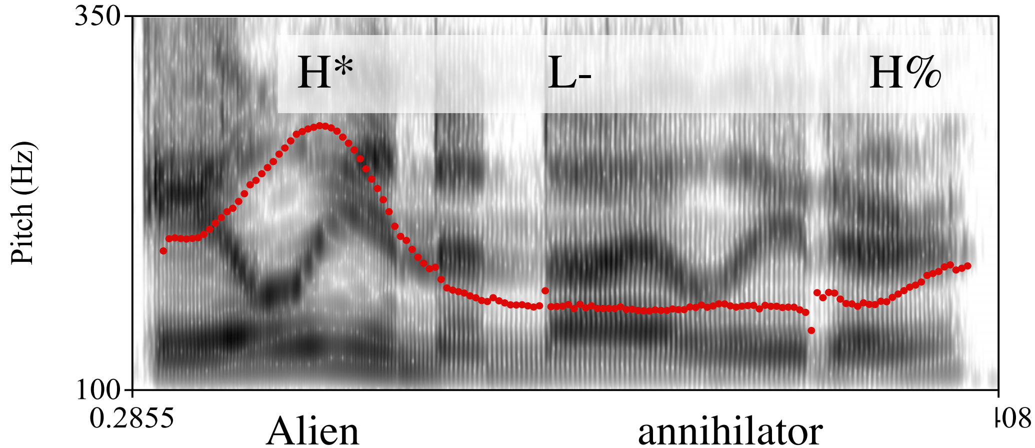
A Generative Phonetic Analysis of the Timing of L- Phrase Accents in English

Edward Flemming
MIT



The timing of L- phrase accents in English

- Pierrehumbert (1980) posits leftward spreading of L- in H*L-H% and H*L-L% tunes to explain why F0 does not interpolate from H* to the end of the phrase.



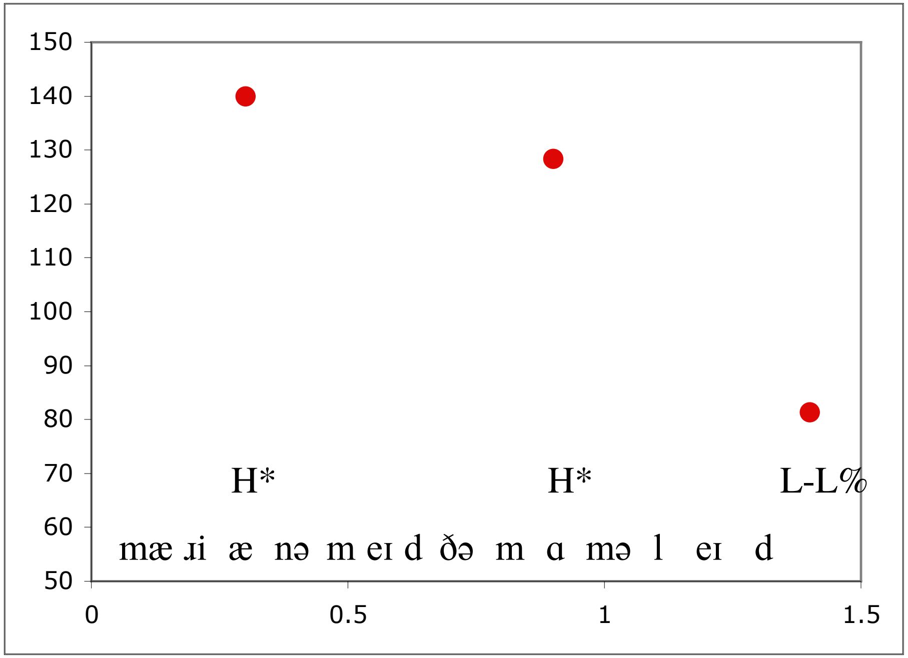
- Where does the low plateau associated with L- begin?
- Two hypotheses (Pierrehumbert 1980):
  - ➢ L- occurs at a fixed interval after H*
  - ➢ L- is aligned to the end of the nuclear-accented word.

Generative phonetics

- Many details of phonetic realization are language-specific
- So the outputs of a grammar must be phonetic representations
  - What does the phonetic component of grammar look like?
- Pioneering work on this topic concerns the phonetics of intonation (Pierrehumbert 1980, Pierrehumbert & Beckman 1988)

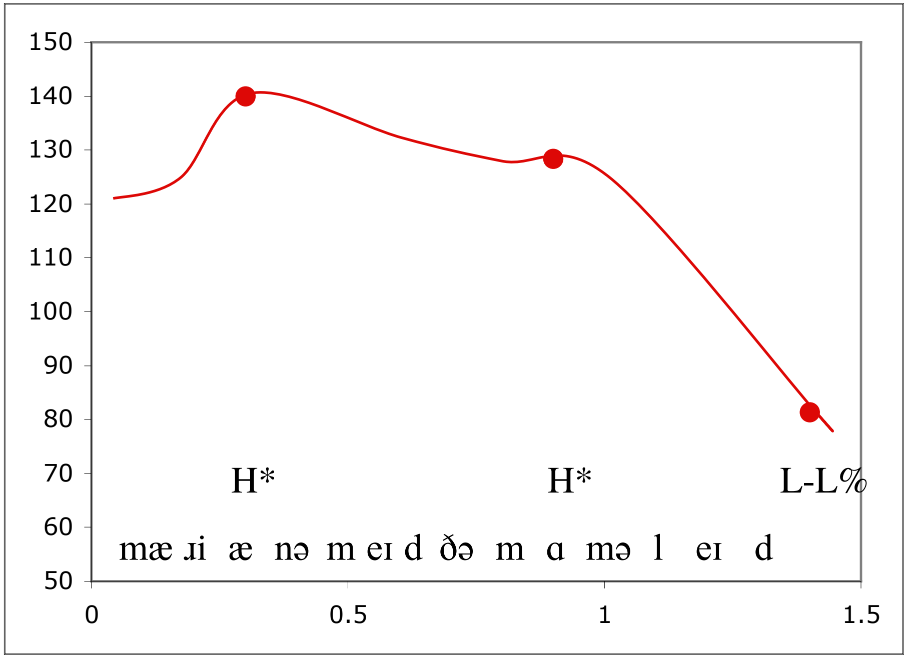
Generative phonetics of tone

- General scheme for implementation of tones (Pierrehumbert 1980):
  - ➢ Locate tone targets in time
  - ➢ Assign F0 values to targets
  - ➢ Interpolate between targets



Generative phonetics of tone

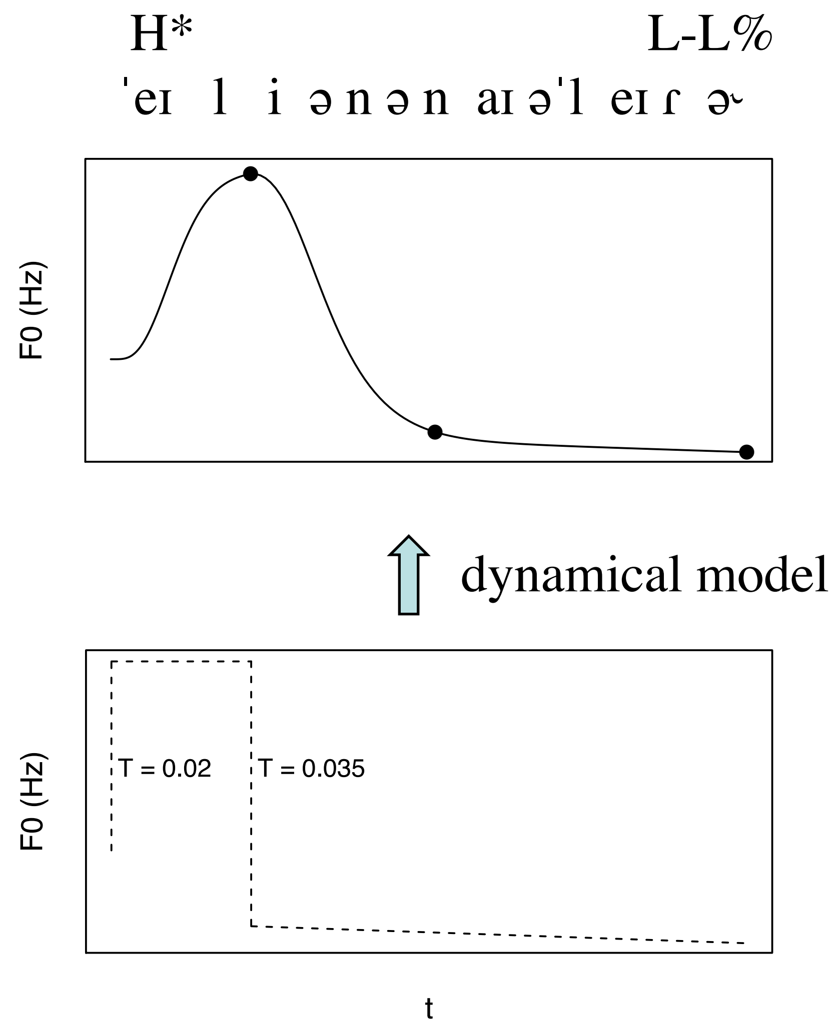
- General scheme for implementation of tones (Pierrehumbert 1980):
  - ➢ Locate tone targets in time
  - ➢ Assign F0 values to targets
  - ➢ Interpolate between targets



Generative phonetics of tone

Propose a revised organization for phonetic grammar of tone:

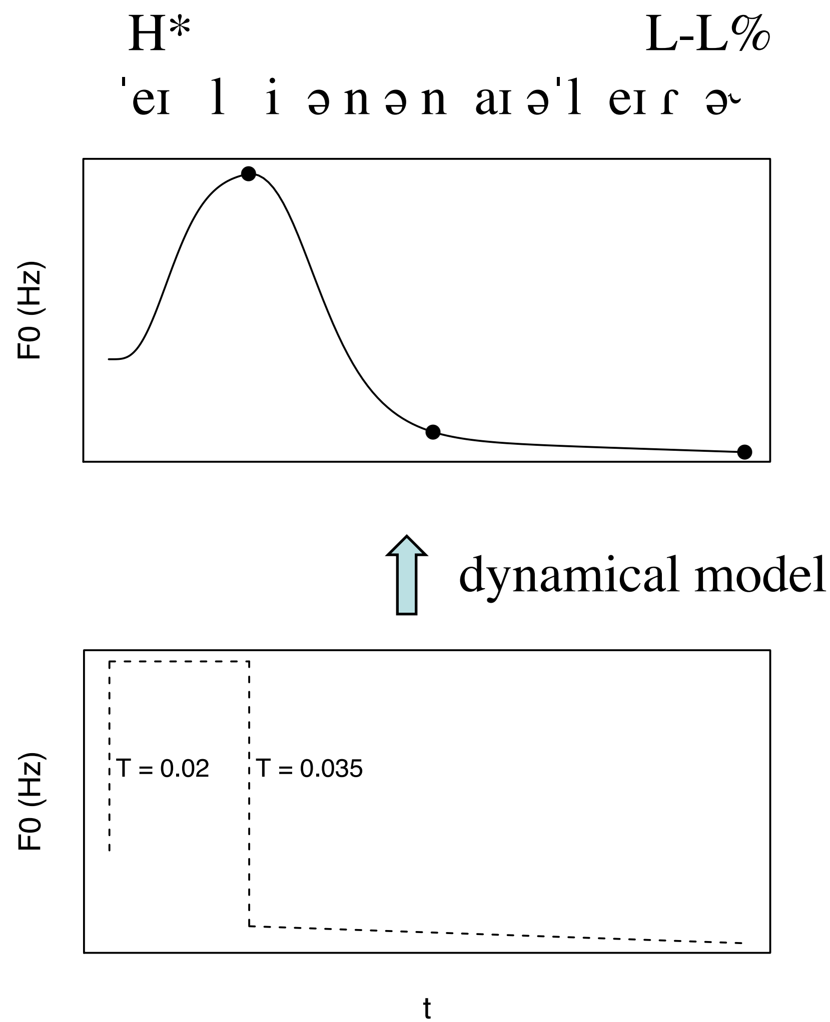
- F0 production is modeled as the response of a dynamical system to a control signal (cf. Fujisaki & Hirose 1984)
  - Control signal consists of step functions connected by linear ramps
- Select the control signal that generates the F0 trajectory that best satisfies constraints on:
  - realization of tone targets
  - alignment of tone targets
  - articulatory constraints



Generative phonetics of tone

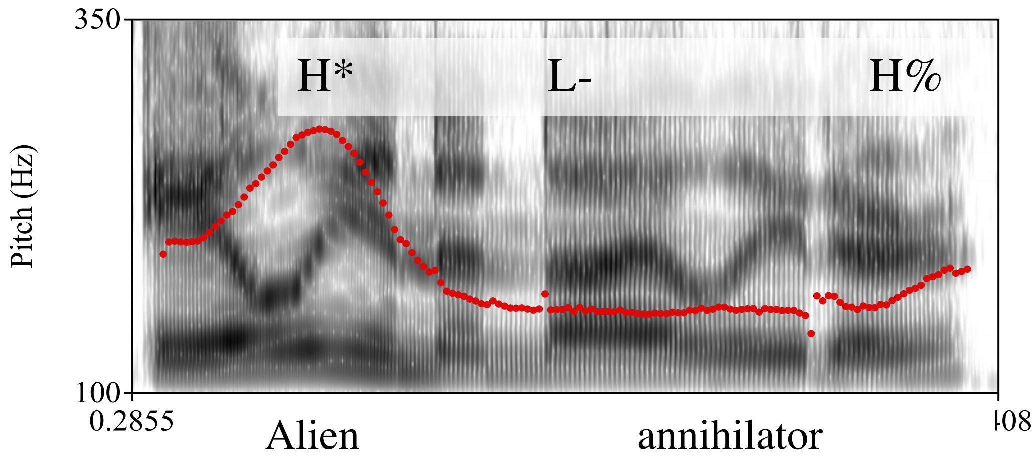
- Target assignment and interpolation are combined
- More explicit model of the interpolation function
- Mapping from tones to F0 trajectory is derived by optimization with respect to conflicting, violable constraints
  - cf. context sensitive realization rules

- The same approach is applicable to segmental realization, given an appropriate production model (cf. Flemming 2001).



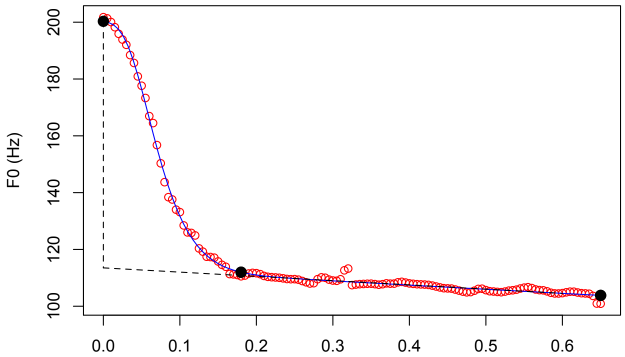
Identifying elbows

- To analyze the timing of L- we have to be able to locate it
  - The correlate of L- is an ‘elbow’ in the F0 trajectory
  - ‘it was very difficult to decide where the L- was located.’ (Pierrehumbert 1980:86)
- Many tones correspond to F0 elbows rather than F0 maxima or minima
  - e.g. rises and falls to/from F0 plateaus (L LH, HL L, LH H, etc) (Flemming & Cho 2017, D’Imperio 2000, Welby 2007 etc)
  - need to be able to locate these tonal targets in order to develop comprehensive models of tonal timing.



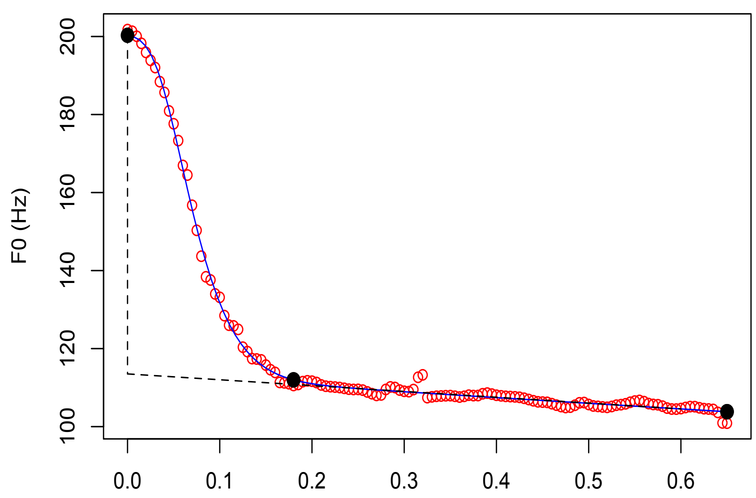
Identifying elbows

- Algorithms for locating F0 elbows have been proposed (e.g. del Giudice et al 2007, Reichel & Salveste 2015)
  - difficult to evaluate because there is no independent evidence concerning the true locations of elbows.
- The proper criterion for positing a target is that it accounts for the observed F0 contour.
- Analysis-by-synthesis using the model outlined above.
  - Assume:
    - ➢ H* target corresponds to the F0 peak
    - ➢ Transitions from H to L take a particular form (explicit model of the interpolation function)
  - Infer the location of L- by fitting the transition model to the H*L- trajectory.



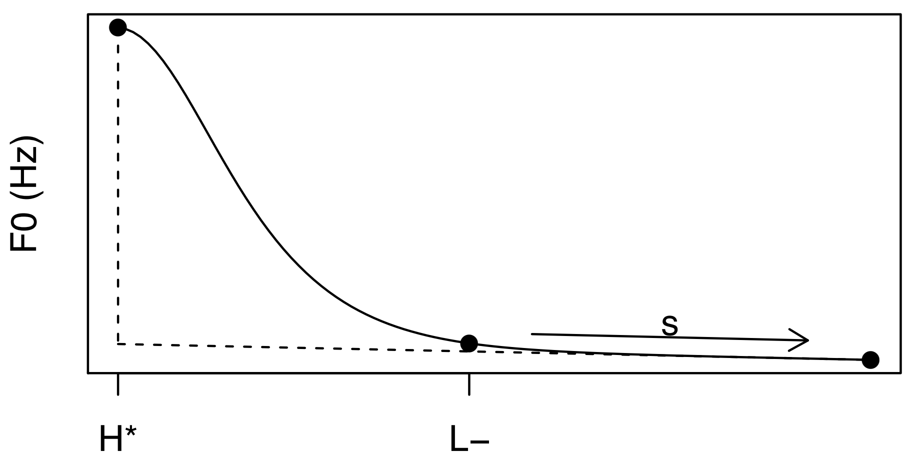
Model of H*L-T% production

- F0 falls are modeled as the response of a dynamical system to a step function input.
  - actually model log(F0)
- Transition from first L- target to the second is modeled as a linear transition.
- cf. Fujisaki & Hirose 1984, Anderson et al 1984



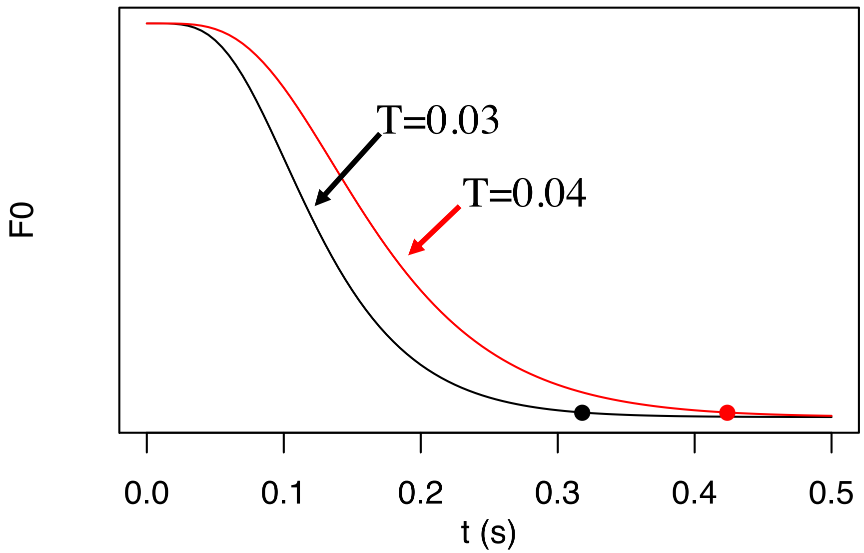
Model of H*L- production

- The dynamical model is a cascade of 4 first-order linear dynamical systems (Birkholz et al 2011).
  - provides a good fit to the contours
  - small number of parameters:
    - F0 targets for H* and L-
    - time constant ('stiffness')
    - initial acceleration
    - slope of plateau



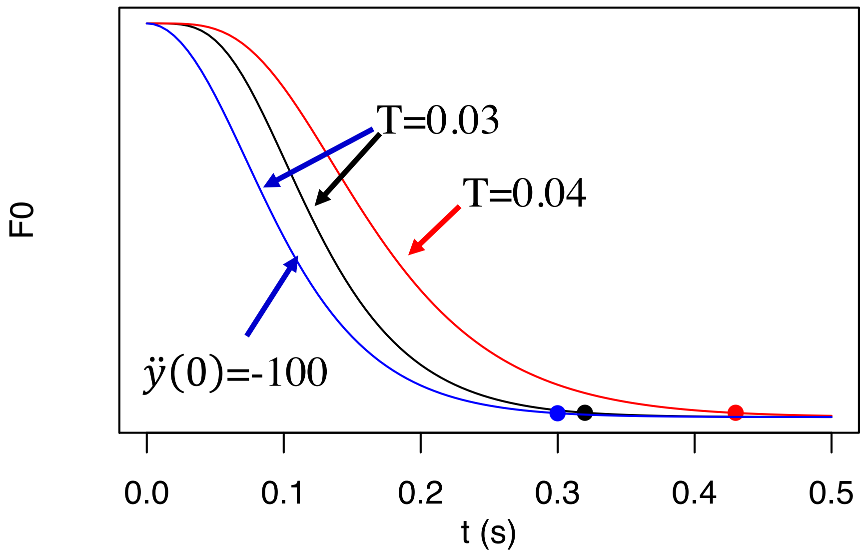
Model of H*L- production

- Parameters for the H*L- falling transition:
  - F0 targets for H* and L-
  - time constant (‘stiffness’)
  - initial acceleration



Model of H*L- production

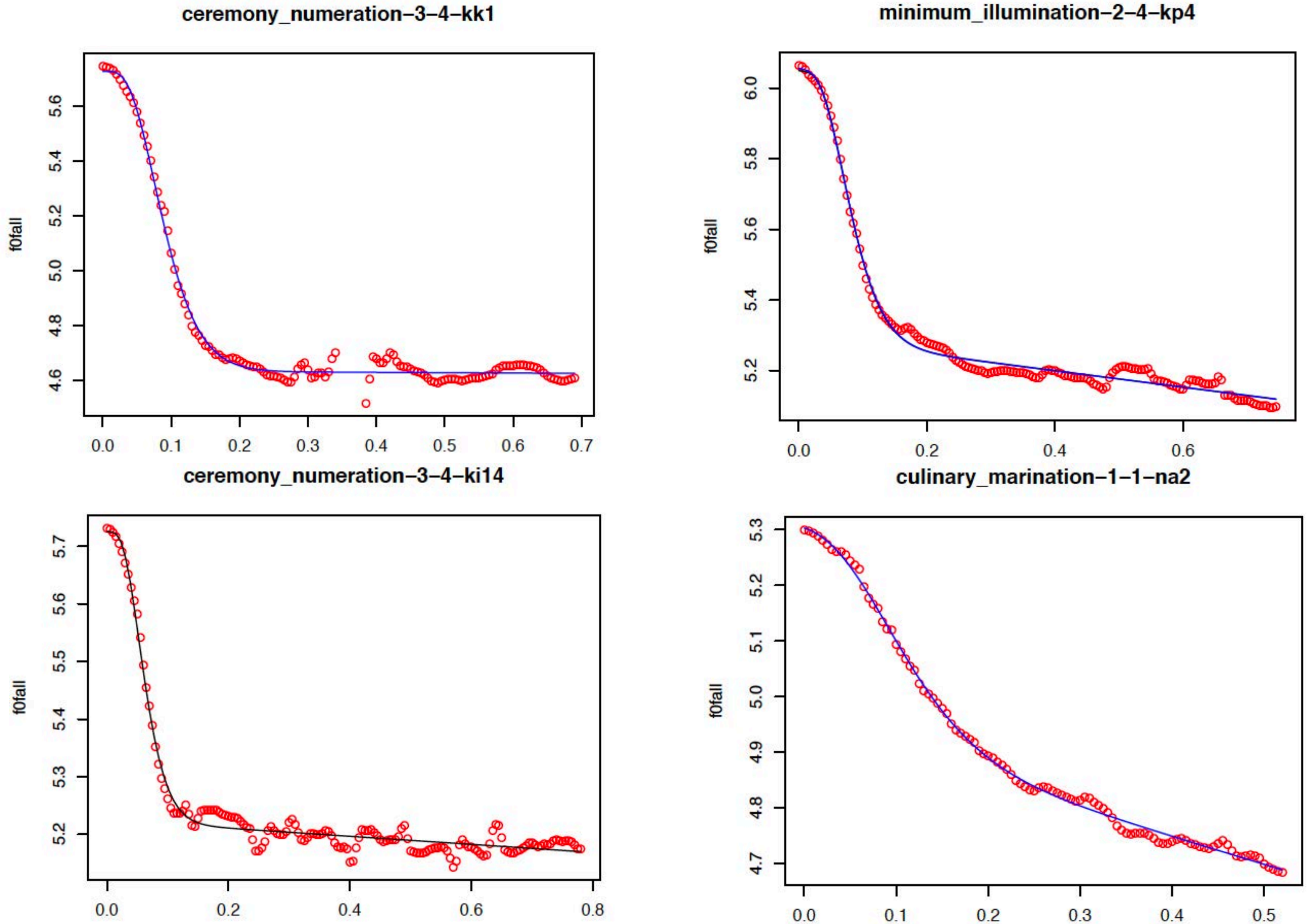
- Parameters for the H*L- falling transition:
  - F0 targets for H* and L-
  - time constant (‘stiffness’)
  - initial acceleration



- The movement never quite reaches its target
  - The end of the transition is taken to be within 2% of the target
  - For a given initial acceleration, the trajectory scales in time with T

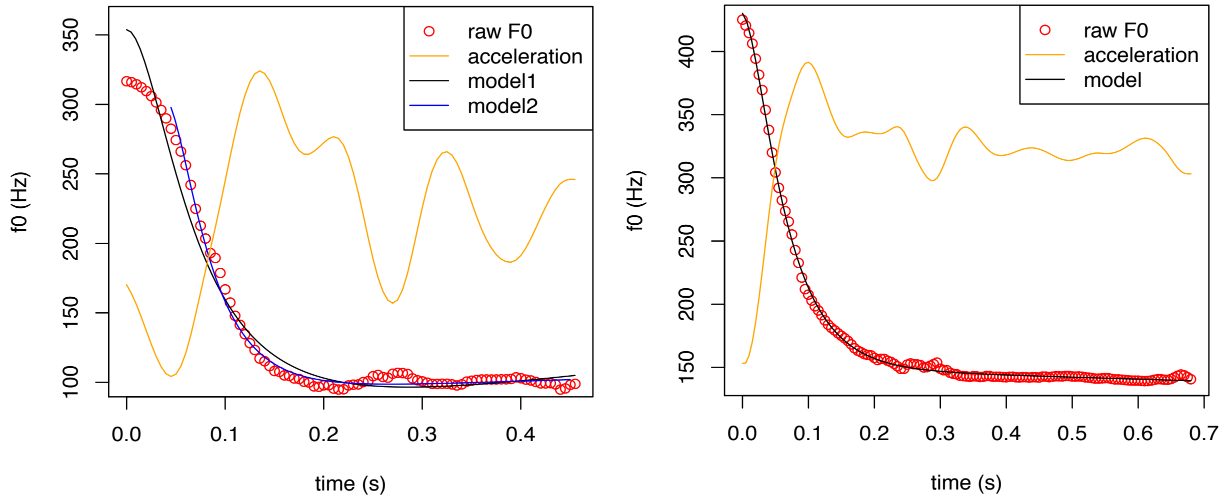
Model of H*L- production

– provides a good fit to the contours



Models of F0 production

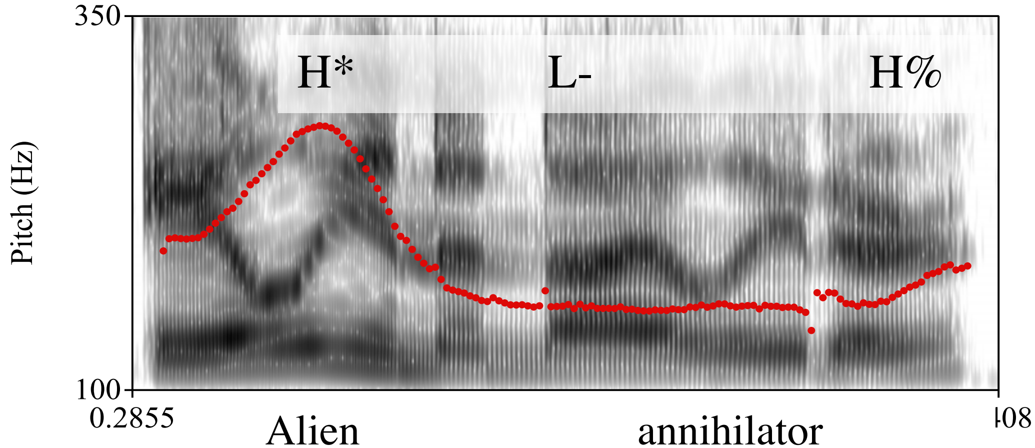
- Critically-damped second order (‘spring-mass’) systems are more familiar as models of F0 movements (Fujisaki & Hirose 1984) and articulator movements (Browman & Goldstein 1992).
- Damped second order system has peak acceleration at movement onset.
  - But this is not true of many F0 movements, resulting in poor fits for this kind of model



  - This problem is familiar from the study of other speech movements (e.g. Kröger et al 1995)
- Cascaded First Order system can accommodate variation in acceleration profiles of movements.

Testing hypotheses about the timing of L-

- Pierrehumbert (1980) posits leftward spreading of L- in H*L-H% and H*L-L% tunes to explain why F0 does not interpolate from H* to the end of the phrase.

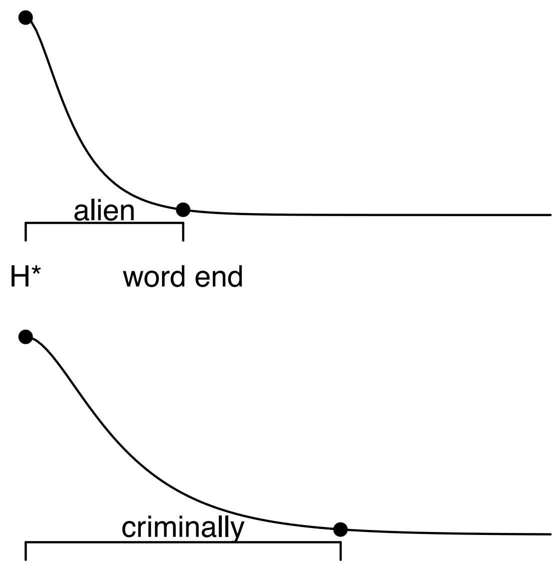


- Where does the low plateau associated with L- begin?
- Two hypotheses (Pierrehumbert 1980):
  - ➢ L- occurs at a fixed interval after H*
  - ➢ L- is aligned to the end of the nuclear-accented word.

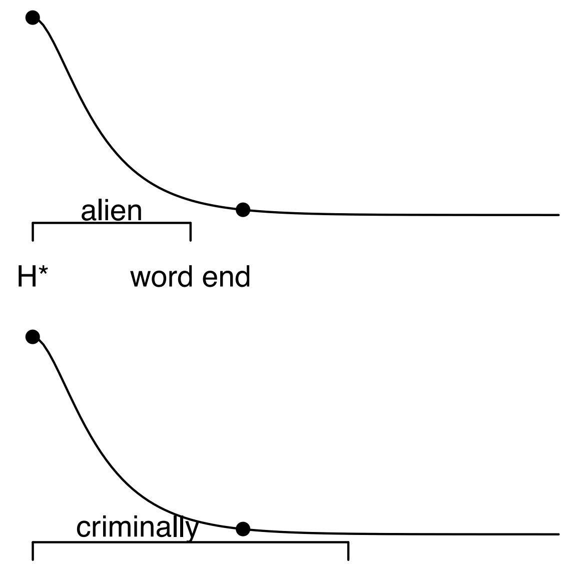
Predictions of the two hypotheses

- How should the time from H* to L- elbow change when the interval between H* and the end of the word changes?

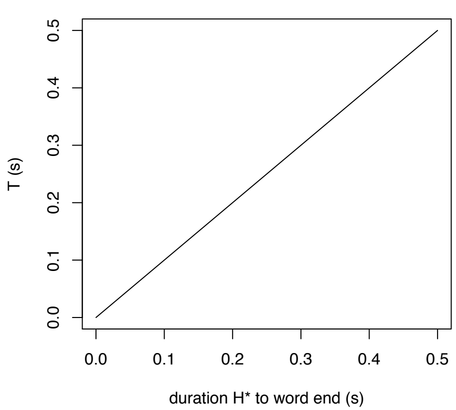
L– aligns to end of word

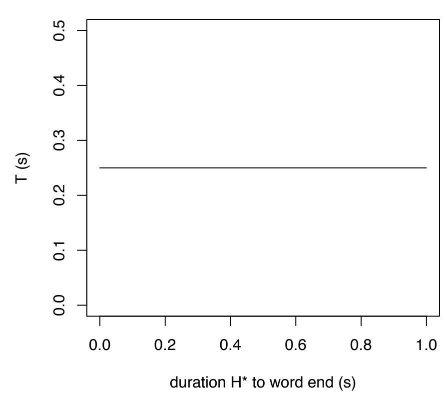


L– occurs at a fixed interval after H*



slope ≈ 1
intercept ≈ 0





slope = 0
intercept = fixed duration of fall

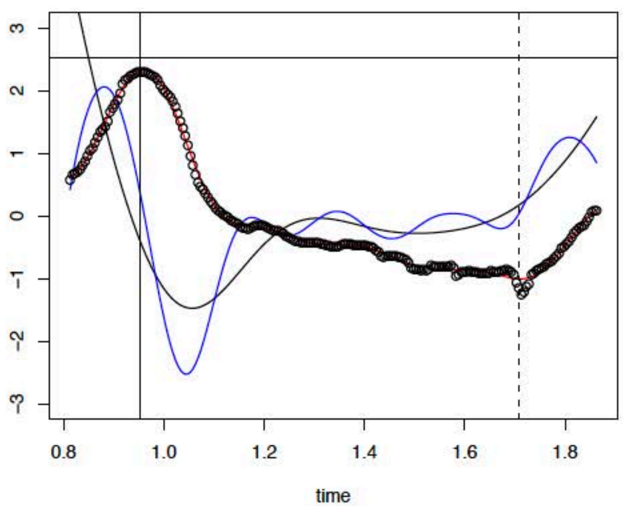
The Data

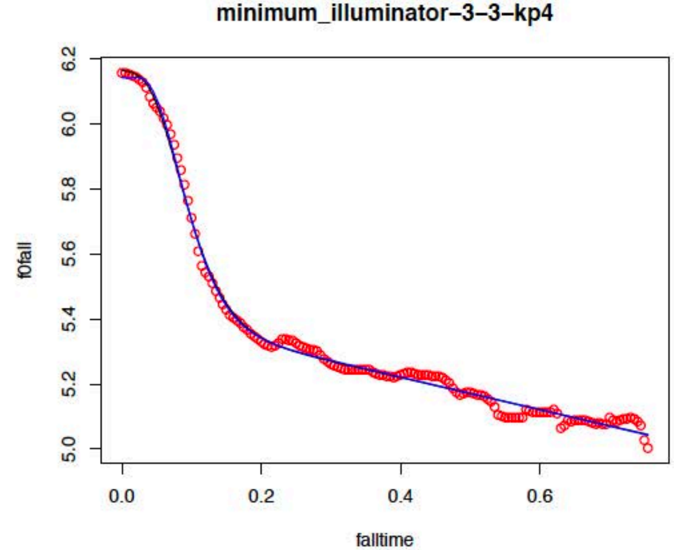
- Recordings from Barnes, Veilleux, Brugos & Shattuck-Hufnagel (2010)
- 25 two-word phrases in a context designed to elicit H* L- H% melody, with H* on the first word.
- First word: vary the number and length of syllables following primary stress
  - 2 álien, lánolin, Líllian, Márilyn, mínimum
  - 3 lúminary, pálimony, céremony, cúlinary, púlmonary
  - 3 críminally, sérially, términally, mínimally, nóminally

(1) Experimenter: George is a thoughtful sort of divorce lawyer—I go to him whenever I need a palimony ruminator.
Subject: A palimony ruminator???!!! (H* L-H%) I thought he was figuring out your plumbing problems!

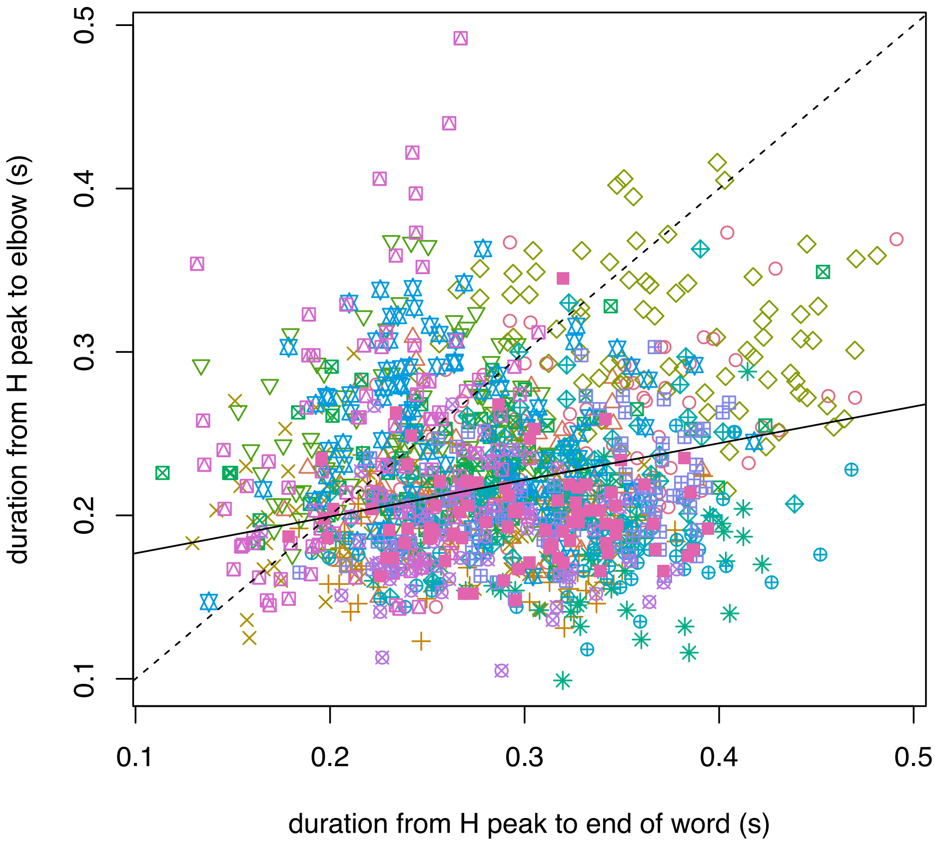
The Data

- 15 speakers (11 female), each produced 4 repetitions of the materials.
- 242 utterances excluded due to errors, disfluencies, pitch tracking problems, leaving 1258 observations.
- Tracked F0 with Praat (Boersma & Weenink 2018),
- Segmented the pitch contour from F0 peak (H*) to onset of the final rise (if any).
- Fitted the tone realization model by minimizing summed squared errors.
  - Identified L- 'elbow' target based on the fitted model





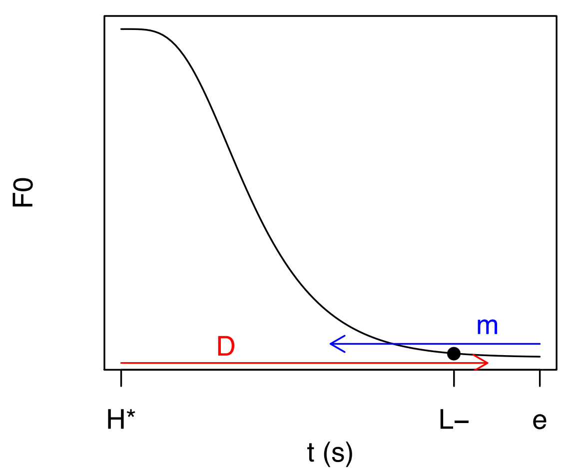
Results



- Neither hypothesis is supported (cf. Barnes et al 2010)
    - ➢ Time from H* to L- elbow is not fixed – it tends to increase as duration from H* to end of word increases ($\beta = 0.22$, $t = 5.5$)
    - ➢ But L- elbow does not track word end (dashed line)
- Substantial variation between and within (most) subjects

Analysis

- Timing of L- elbow is a result of compromise between two conflicting constraints :
  - ➢ L- should start at a fixed interval *D* after H*
    - preferred duration for an HL fall (cf. Cho 2010)
  - ➢ L- should start well before the end of the accented word
    - to keep H*L- distinct from tunes with a pitch accent on a following word (H* !H*, H* L*) (cf. Barnes et al 2010)
- These constraints conflict if the preferred fall duration would place L- too close to the end of the word
  - – The conflict is resolved by compromise (cf. Flemming 2001).

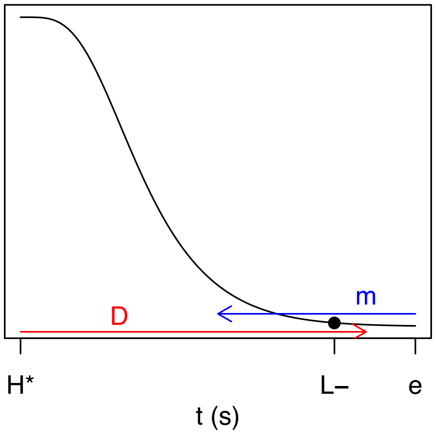


Analysis

- Time of H* = 0, L is the time of the initial target for L-
  - L- should start at a fixed interval D after H*

    $L = D$ penalty for violation: $(L\text{-}D)^2$
  - L- should start at least m seconds before the end of the accented word, e

    $L < e\text{-}m$ penalty for violation: $(L\text{-}(e\text{-}m))^2$



- Select L to minimize the weighted sum of constraint violations:

  $$w_D(L\text{-}D)^2 + w_e(L\text{-}(e\text{-}m))^2$$
- Optimal value of L is a weighted average of D and $e\text{-}m$

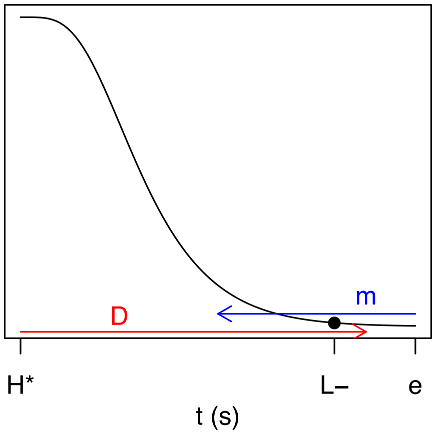
  $$L = \frac{w_D D + w_e(e - m)}{w_D + w_e}$$
  - if w_e is much lower than w_D, then the effect of word boundary is weak.

Analysis

- Time of H* = 0, L is the time of the initial target for L-
  - ➢ L- should start at a fixed interval D after H*

    $L = D$ penalty for violation: $(L\text{-}D)^2$
  - ➢ L- should start at least m seconds before the end of the accented word, e

    $L < e\text{-}m$ penalty for violation: $(L\text{-}(e\text{-}m))^2$



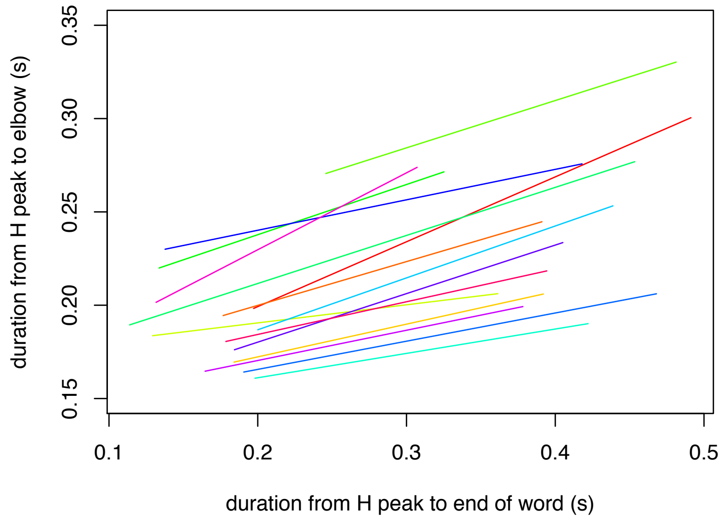
- Optimal value of L is a weighted average of D and e-m

$$L = \frac{w_D D + w_e(e - m)}{w_D + w_e}$$

- The observed relationship between L and time to word boundary is derived if $w_D = 1$, $w_e = 0.28$, $D = 0.25$s, $m = 0.19$s
  - – D and m are not uniquely identifiable, but these values are compatible with the data.

Variation between speakers

- Relationships between duration from H* peak to L- elbow and duration from H* peak to end of word, plotted by subject



  - Substantial variation in intercepts, and some variation in slope.
- Variation in intercept can be derived from variation in *D*
  - Suggests that *D* may be derived from compromise between effort minimization (high T) and a preference for a steep fall from H to L.
  - *D* variation then results from variation in constraint weights.
- Variation in slope follows from variation in ratio of w_D to w_e

Summary

- Study of timing of L- phrase accents in English motivates a model of tonal realization in which:
  - F0 trajectories are modelled as the response of a dynamical system to a control signal that consists of a sequence of step functions connected by linear ramps.
    - Specifying the mechanism that generates F0 trajectories allows for identification of F0 targets, including elbows. through analysis-by-synthesis.
  - The mapping from tones to F0 trajectory is derived by optimization with respect to conflicting constraints
    - The timing of L- is a compromise between a preferred duration for H*L- falls, and a weak preference to reach L- before the end of the accented word.
- The same structure should be applicable to the grammar of segmental phonetics (cf. Flemming 2001).

References

Anderson, M., Pierrehumbert, J.B., & Liberman, M. (1984). Synthesis by rule of English intonation patterns. *ICASSP '84*, vol. 9, 77-80.

Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2010). Turning points, tonal targets, and the English L- phrase accent. *Language and Cognitive Processes,* 25:7-9, 982-1023.

Birkholz, P., Kröger, B.J., & Neuschaefer-Rube, C. (2011). Model-based reproduction of articulatory trajectories for Consonant-Vowel sequences. *IEEE Transactions on Audio, Speech, and Language Processing,* 19, 1422-1433.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180.

del Giudice, A., Shosted, R., Davidson, K., Salihie, M., & Arvaniti, A. (2007). Comparing methods for locating pitch "elbows". *Proceedings of the 16th International Congress of Phonetic Sciences*, 1117-1120.

D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment*. PhD dissertation, Ohio State University.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7–44.

Flemming, E., & Cho, H. (2017). The phonetic specification of contour tones: Evidence from the Mandarin rising tone. *Phonology* 34, 1-40.

References

Fujisaki, H. & Hirose, K. (1984), Analysis of voice fundamental frequency contours for declarative sentences of Japanese, *Journal of the Acoustical Society of Japan*, 5, 233-242.

Kröger, B. J., Schröder, G., & Opgen-Rhein, C. (1995). A gesture-based dynamic model describing articulatory movement data. *Journal of the Acoustical Society of America,* 98(4), 1878–1889.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. Thesis, MIT, Cambridge, MA.

Pierrehumbert, J.B., & Beckman. M.E. (1988). *Japanese Tone Structure*. MIT Press, Cambridge, MA.

Reichel, U., & Salveste, N. (2015). Pitch elbow detection. *ESSV*, Katholische Universität Eichstätt-Ingolstadt, Germany, 25.–27. March 2015. Ed. Wirsching, G. Dresden: TUD Press, 143–149.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication, 49*(1), 28-48.