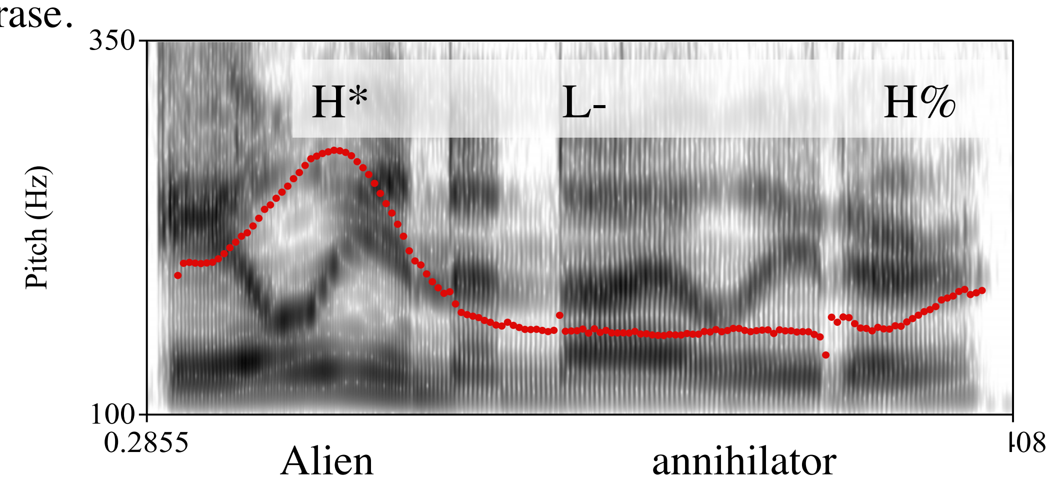
ETAP4@UMass

Modeling 'Elbows' in F0 Contours: Phrase Accents in English

Edward Flemming (flemming@mit.edu)
Massachusetts Institute of Technology

The problem with Elbows: an example

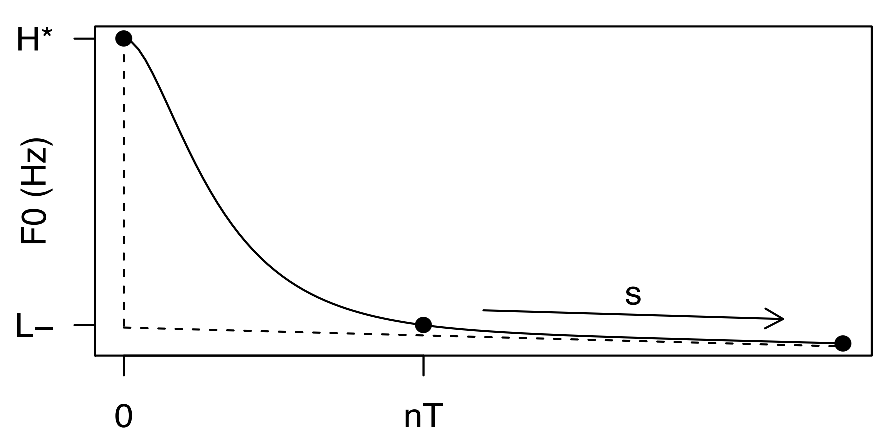
- Pierrehumbert (1980) posits leftward spreading of L- in H*L-H% and H*L-L% tunes to explain why F0 does not interpolate from H* to the end of the phrase.



- Two hypotheses concerning the timing of the onset of L- (Pierrehumbert 1980):
  - ➢ L- occurs at a fixed interval after H*
  - ➢ L- is aligned to the end of the nuclear-accented word.
- To test these hypotheses we have to locate L-
- – The correlate of L- is an 'elbow' or inflection in the F0 trajectory
- – 'it was very difficult to decide where the L- was located.' (Pierrehumbert 1980:86)

Identifying elbows through analysis-by-synthesis

- Analysis-by-synthesis of F0 trajectories:
  - – Rather than identifying elbows using general-purpose algorithms, then modeling the results (e.g. del Giudice et al 2007, Reichel & Salveste 2015),
  - – 'Elbow' targets should be inferred in the process of modeling F0 trajectories.
- Model of H*L-(T%) production:
  - – The transition from H* to the first L- target is realized as the response of a critically-damped linear second order system ('spring-mass system') to a step input.
  - – The transition from the first L- target to the second is the response of the same system to an input linear transition between the two targets.
  - – Cf. Fujisaki & Hirose (1984), Anderson et al (1984)



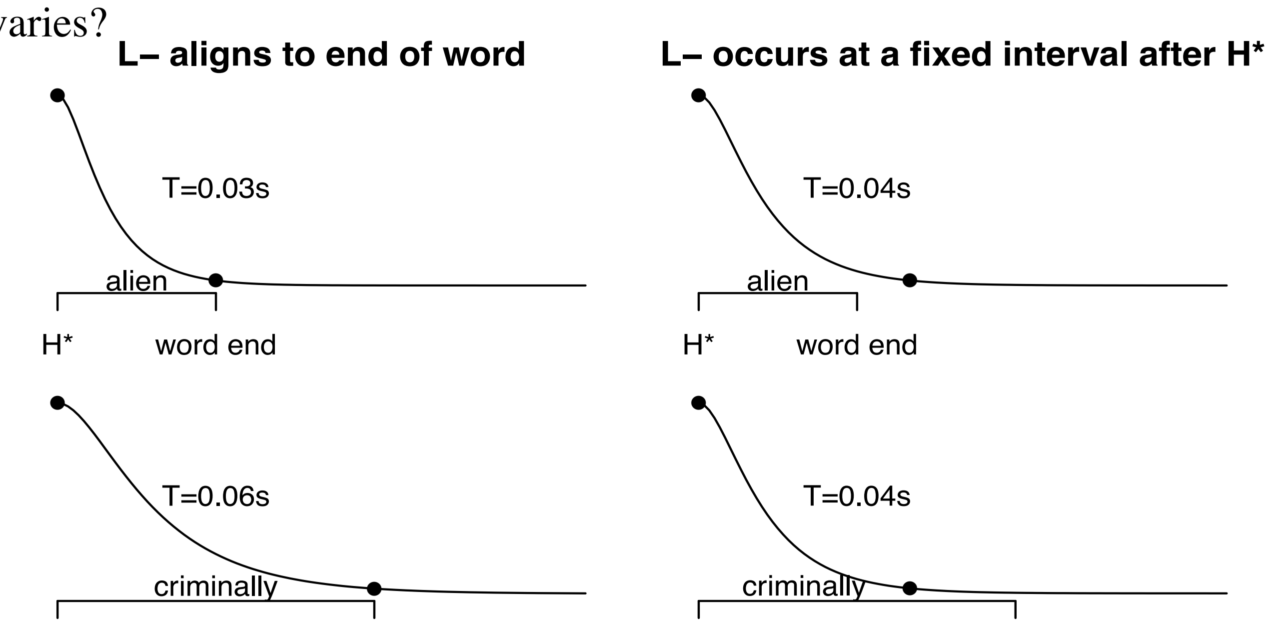
- The form of this trajectory if initial velocity = 0:

$$\log(F_0) = L + (H-L)\left(1+\frac{t}{T}\right)e^{-\frac{t}{T}} + sT\left(\frac{t}{T} - 2 + \left(2+\frac{t}{T}\right)e^{-\frac{t}{T}}\right)$$

- Fit this model to the observed F0 trajectories to obtain estimates of F0 targets.
- Since a critically damped movement strictly never reaches its target, this model does not directly specify the time of the L- target
- The timing of the effective target is specified in multiples of T
  - – e.g. 99% of the movement is completed in about $5T$

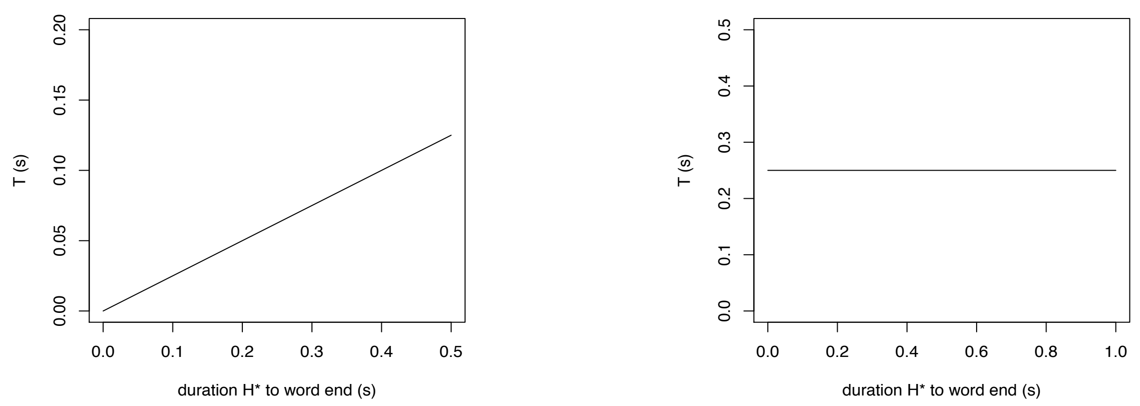
Testing hypotheses about the timing of L-

- Predictions of the two hypotheses: What should happen to the time constant T when the duration of the interval between H* and the end of the word varies?



- T is predicted to be a linear function of H*-to-word-end duration, with intercept ≈ 0
- Slope of the line is $1/n$, where n is the number of time constants to reach the L- target.

- T doesn't vary with interval duration.
- slope = 0, intercept = fixed T.



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!

- 15 speakers (11 female), each produced 4 repetitions of the materials.
- 239 utterances excluded due to errors, disfluencies, pitch tracking problems.
- Tracked F0 with Praat (Boersma & Weenink 2018), segmented the pitch contour from F0 peak (H*) to onset of the final rise, and fitted the tone realization model using non-linear least squares (`nls` (R Core Team 2016)).

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

Anderson, M., J.B. Pierrehumbert & M.Y. Liberman. 1984. Synthesis by rule of English intonation patterns. *ICASSP '84*.

del Giudice, A., R. Shosted, K. Davidson, M. Salihie, & A. Arvaniti. 2007. Comparing methods for locating pitch "elbows". *ICPHS XVI*, 1117-1120.

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.

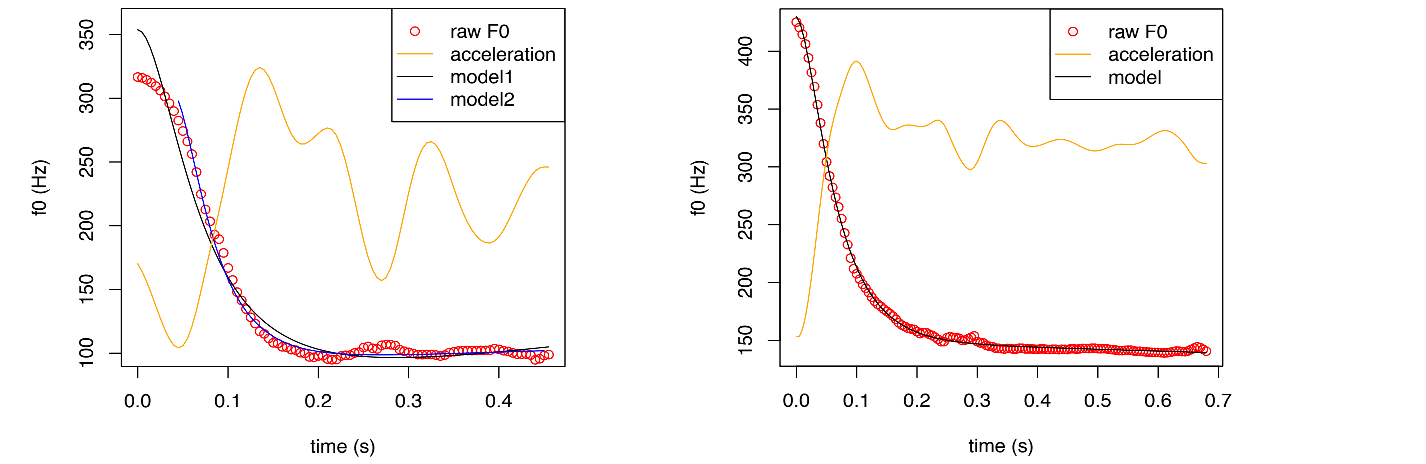
Kröger, B.J., Schröder, G., Opgen-Rhein, C. 1995. A gesture-based dynamic model describing articulatory movement data. *JASA* 98, 1878-1889

Pierrehumbert, J.B. 1980. *The Phonology and Phonetics of English Intonation*. PhD Thesis, MIT.

Reichel, U., & N. Salveste. 2015. Pitch elbow detection. *ESSV*, G. Dresden: TUD Press, 143–149.

Results I – production model

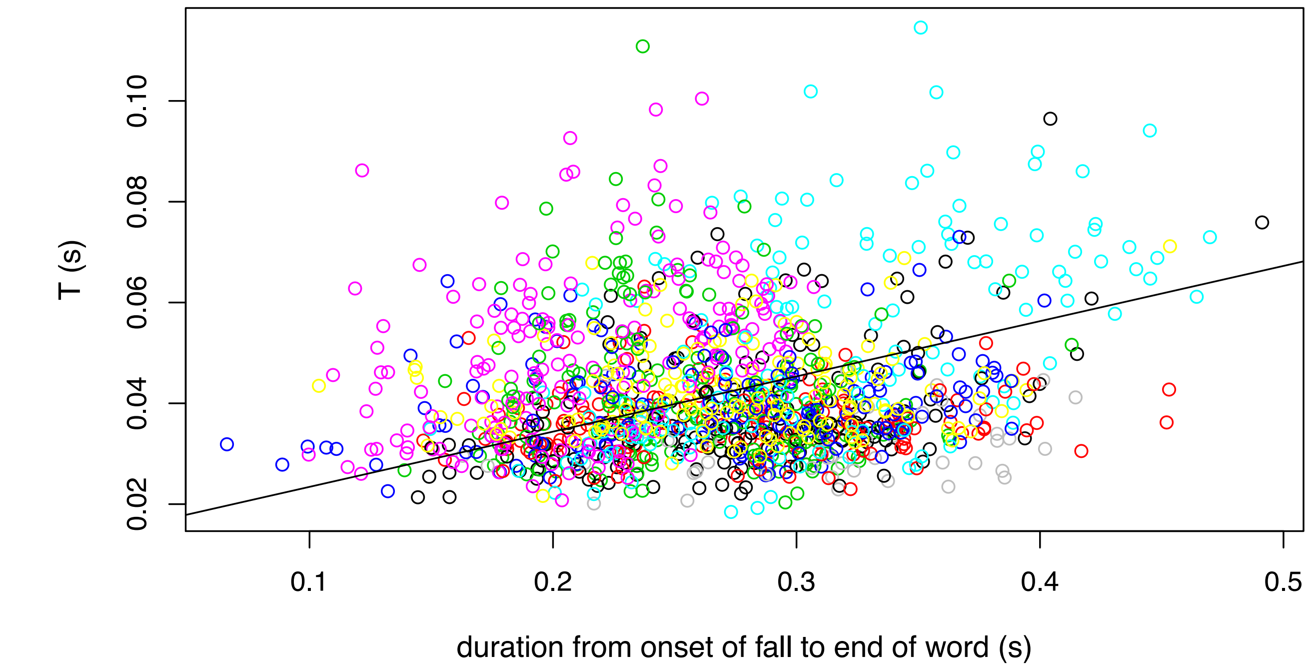
- The critically-damped model does not fit all speakers/utterances well
  - – Problem: damped 'spring-mass' models have peak acceleration at movement onset, but this is not true of all H*L- transitions



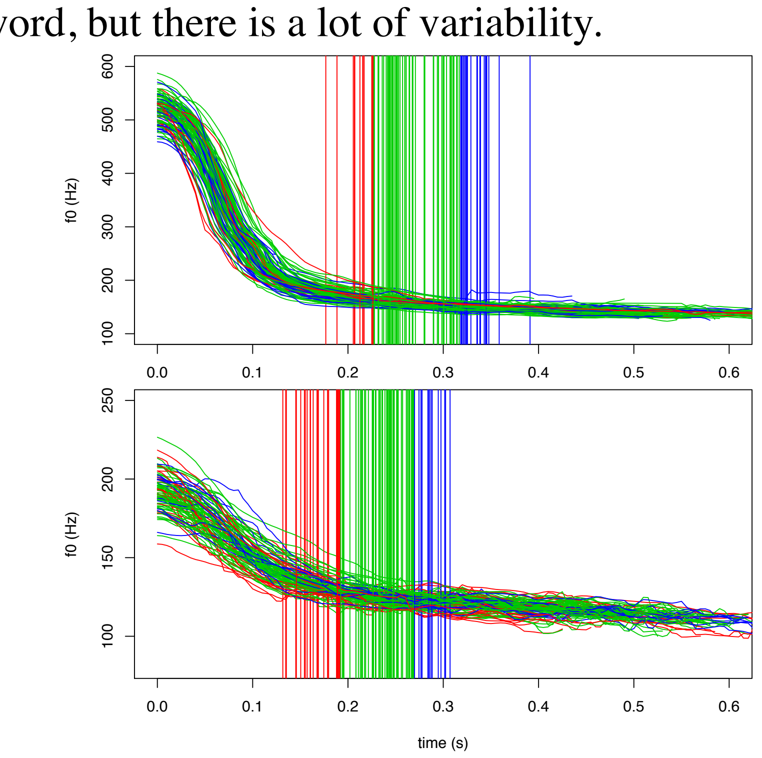
  - – This problem is familiar from the study of other speech movements (e.g.Kröger et al 1995)
  - – Current solution: Model the H*L- transition with two step functions, starting the second from the acceleration minimum, with estimated initial velocity.

$$\log(F_0) = L + (H-L)\left(1+\left(\frac{v_0}{(H-L)}+\frac{1}{T}\right)t\right)e^{-\frac{t}{T}} + sT\left(\frac{t}{T} - 2 + \left(2+\frac{t}{T}\right)e^{-\frac{t}{T}}\right)$$

Results II – timing of L-



- T tends to increase as duration from H* to end of word increases ($\beta = 0.11, t = 6.5$)
- So the interval between H* and L- is not fixed, but L- does not track word end either
  - – $\beta = 0.11$ would imply that target is achieved at $9T$, also intercept > 0 ($\beta = 0.012, t = 3.4$)
  - – This pattern could represent a compromise between a preferred value for T and a preference to keep L- within the accented word, but there is a lot of variability.



- F0 trajectories from two subjects, aligned on H* peak
- Vertical lines mark word end
- Color codes duration from H*-to-word-end

download this poster: web.mit.edu/~flemming/www/paper/elbows.pdf

Thanks to Jon Barnes, Alejna Brugos, Stefanie Shattuck-Hufnagel and Nanette Veilleux for sharing their recordings, and to Run Chen for research assistance