24.964

Phonetic Realization

Models of phonetic realization

Rules and constraints

Readings for next week:

Phonetics-phonology interaction

- Steriade (2001) proposes that the P-map projects correspondence constraints and their ranking.
  - Phonetic similarity determines constraint rankings.
- If the P-map refers to phonological features like [+/-voice] (as opposed to [pː - bː], [pː - bp]), then the P-map and associated rankings must be language-specific (Kawahara 2006).
- How does this work?
  - Determine phonetic realizations of segments in context, independently of phonology.
  - Construct P-map based on these realizations, e.g. T-D ➢ T:-D:
  - Ranks correspondence constraints based on the P-map, e.g. Ident(+voi)$_{sing}$ >> Ident(+voi)$_{gem}$

P-map in acquisition?

- Could the effects of phonetics on constraint ranking arise in acquisition?
  - Child constructs P-map based on experience with language.
  - Uses P-map in ranking constraints.
- Can only explain typology if the child enforces the relationship between similarity and ranking of correspondence constraints.
  - What happens if a child is confronted with a language in which T-D/_[-son] ➢ T-D/_# but distribution indicates Ident[voice]/_# >> Ident[voice]/_[-son]?
  - If the typological generalization is a consequence of P-map projection being enforced in acquisition then the child must enforce P-map projection and fail to learn the language.
    - adjust phonetic realization
    - adjust output of phonology

Phonetics-Phonology correspondence

- So the core of the P-map proposal is the condition on the relation between phonetic realization and constraint ranking: correspondence constraints must be ranked according to phonetic similarity of correspondents.
  - principle of grammar, not (just) part of an acquisition procedure.
  - potentially neutral as to the directionality of effects (phonetics affects constraint ranking vs. constraint ranking affects phonetics).
    - model sketched above is unnecessarily procedural.
- There are possible phonologies (constraint rankings)
- There are possible phonetic realization components
  - resulting in different P-maps.
- P-map projection places conditions on pairings of phonology and phonetic implementation.
  - no need to specify a procedure for deriving rankings from phonetics (except perhaps in acquisition).
  - may also be consistent with ‘incomplete’ P-maps (e.g. lacking information about clicks): rankings must be consistent with the P-map, but do not all have to be dictated by the P-map.

Phonetics-Phonology correspondence

- Is this approach motivated or is just a way to salvage unmotivated assumptions about phonological representations?
  - alternative: phonetic detail in phonology.
- This ‘correspondence’ approach makes more sense if there are conditions on phonology that are independent of the details of phonetic realization and vice versa.
  - i.e. phonetic and phonological typology are substantially independent, although connected by P-map principles.
  - if typologies are more interdependent then an integrated model may be required.

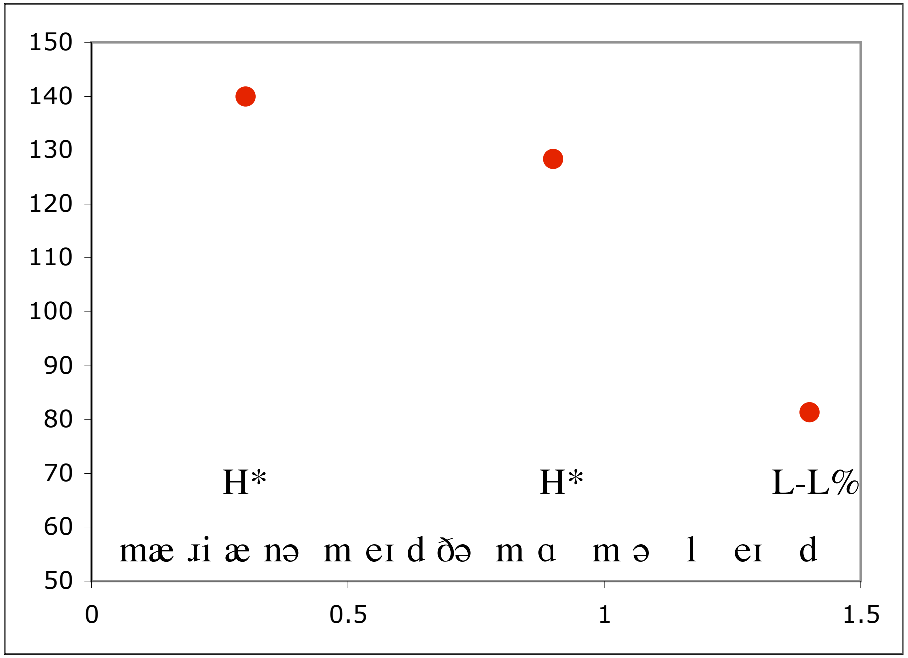
Models of phonetic realization

- The earliest explicit models of phonetic realization are largely rule-based: a series of rules map phonological representations on phonetic realizations.
  - e.g. Pierrehumbert 1980, Cohn 1990, 1993.
- General scheme: targets and interpolation

Targets and Interpolation - Pierrehumbert (1980)

General scheme:

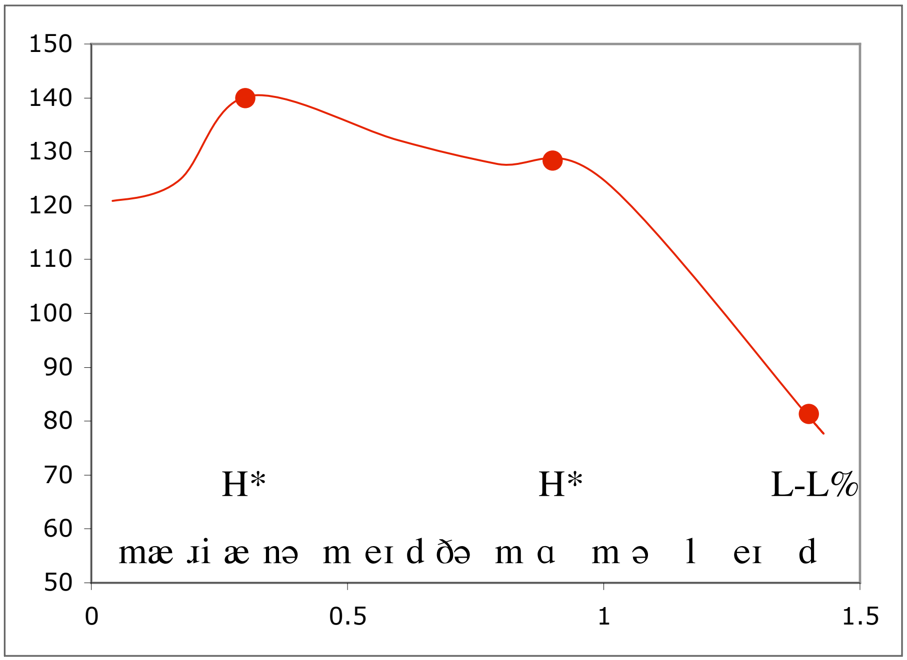
- Locate tone targets in time.
- Assign f0 values to targets
- Interpolate between targets



Targets and Interpolation - Pierrehumbert (1980)

General scheme:

- Locate tone targets in time.
- Rules assign f0 values to targets
- Interpolate between targets



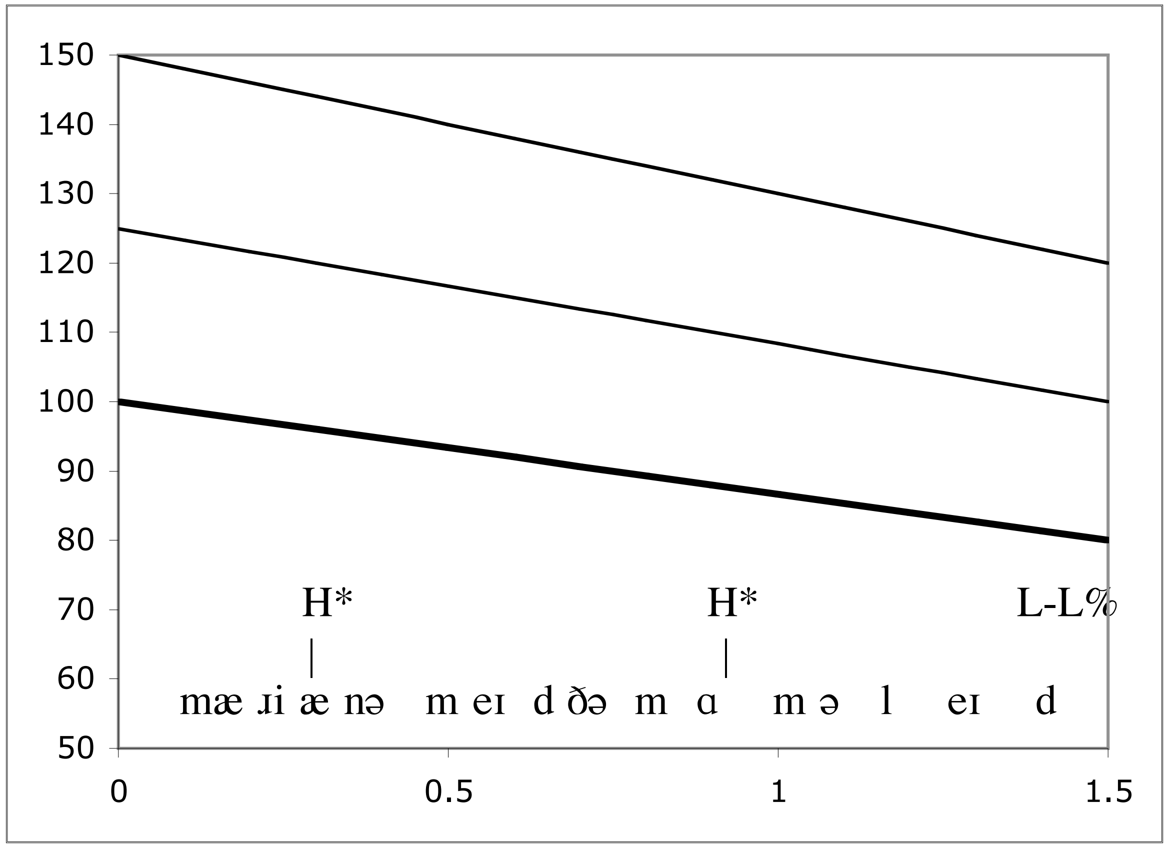
Some of P80's tone realization rules

'Each speaker has a declining F0 baseline…[that] represents the lowest F0 value that the speaker would be disposed to reach at any given point in the utterance.'

Rules formulated in terms of baseline units above baseline:

$$/T/ = \frac{T(Hz) - b_p}{b_p}$$

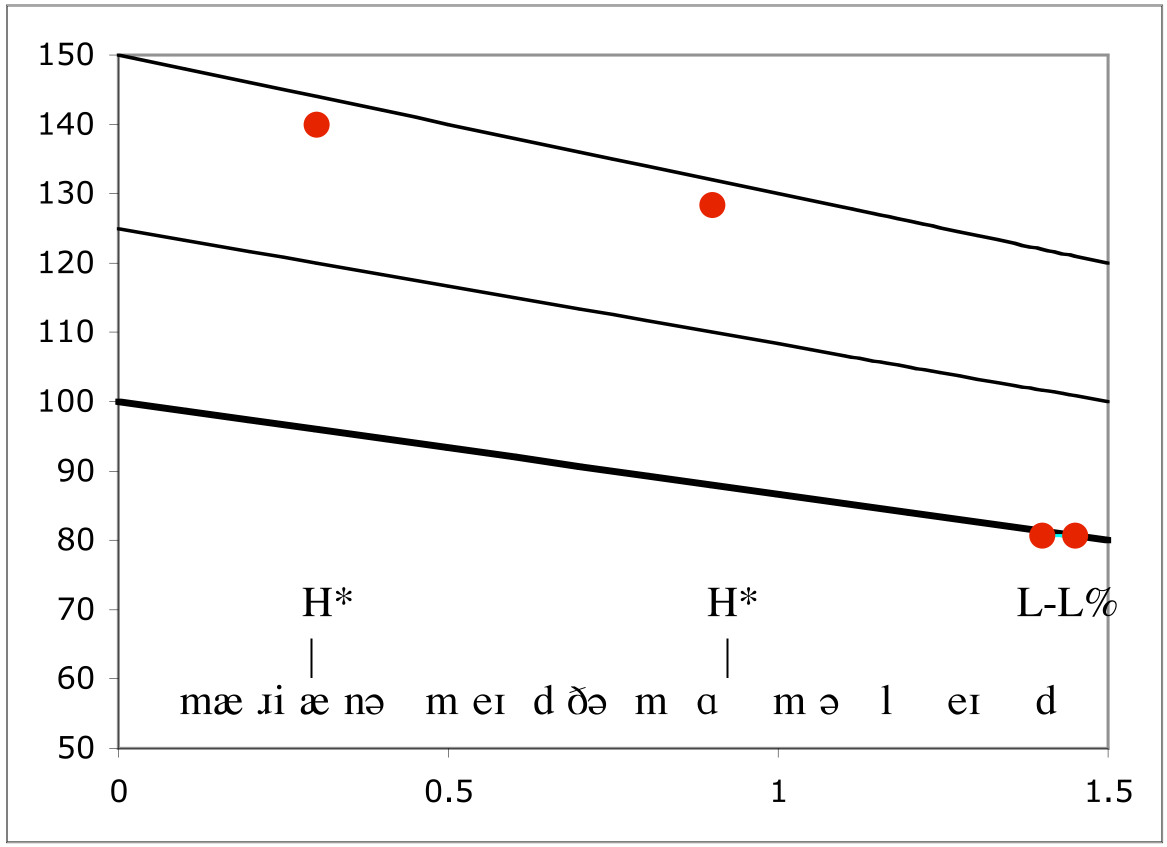
where b_p is the value of the baseline at the time of p.



Some of P80's tone realization rules

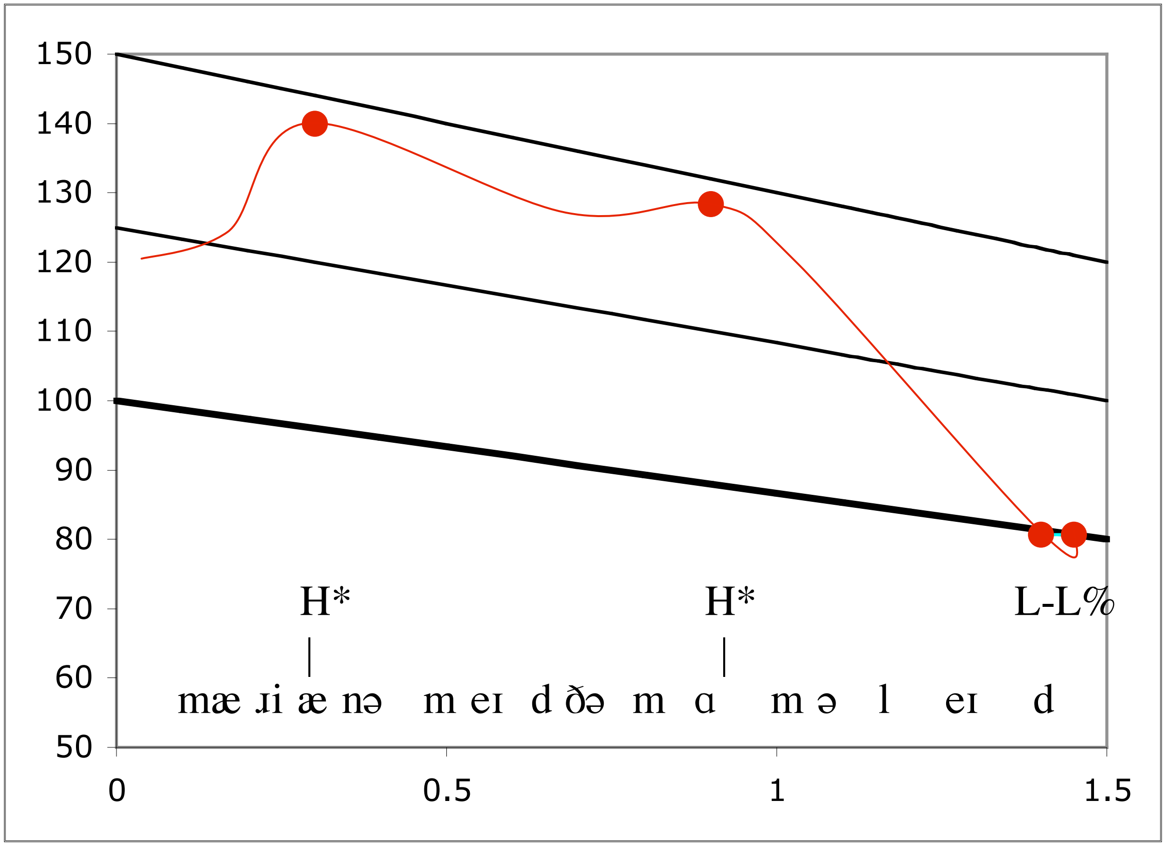
$$/H^*_{i+1}/ = /H^*_i/ . \frac{\text{Prominence}(H^*_{i+1})}{\text{Prominence}(H^*_i)} \qquad \text{in } H^*_i(+L)\ (L+)H^*_{i+1}$$

H-H scaling, L%

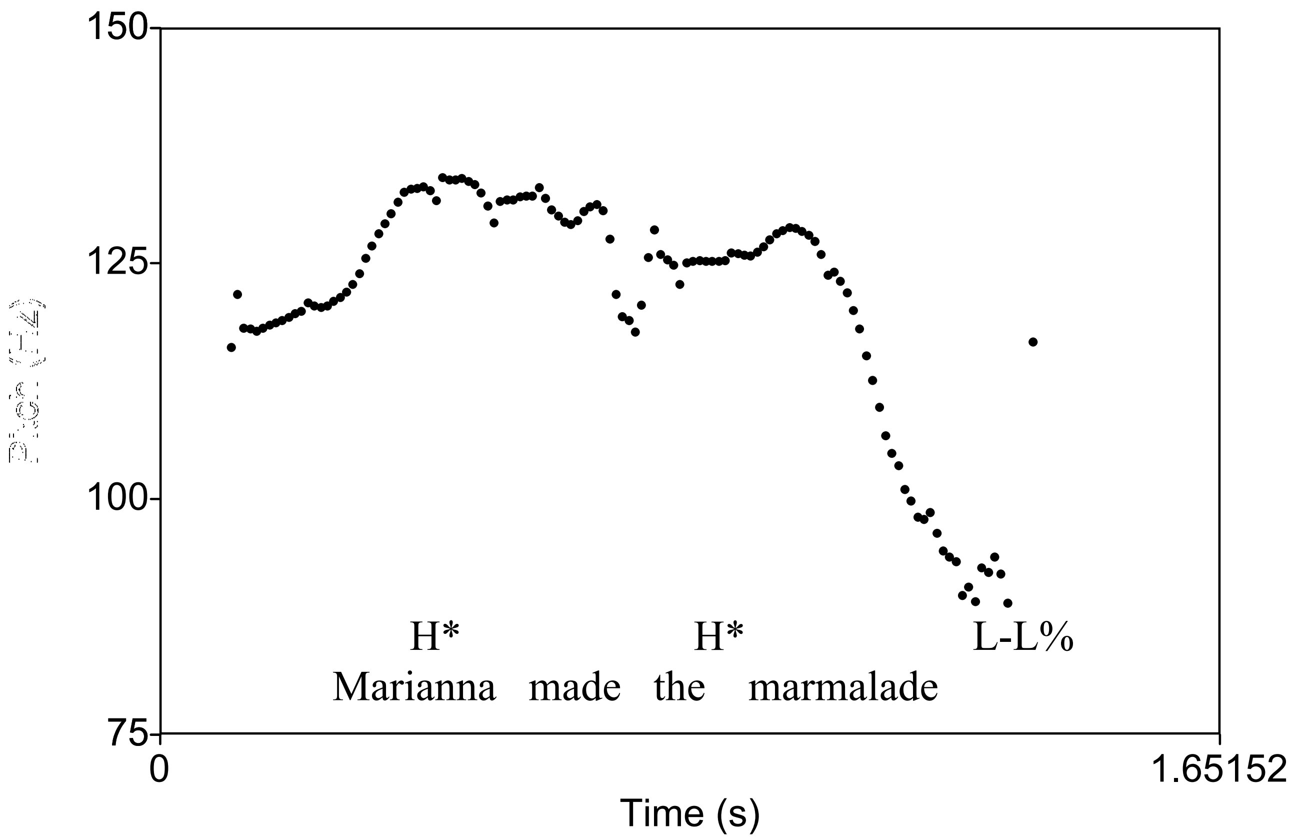


Some of P80's tone realization rules

sagging interpolation between H tones

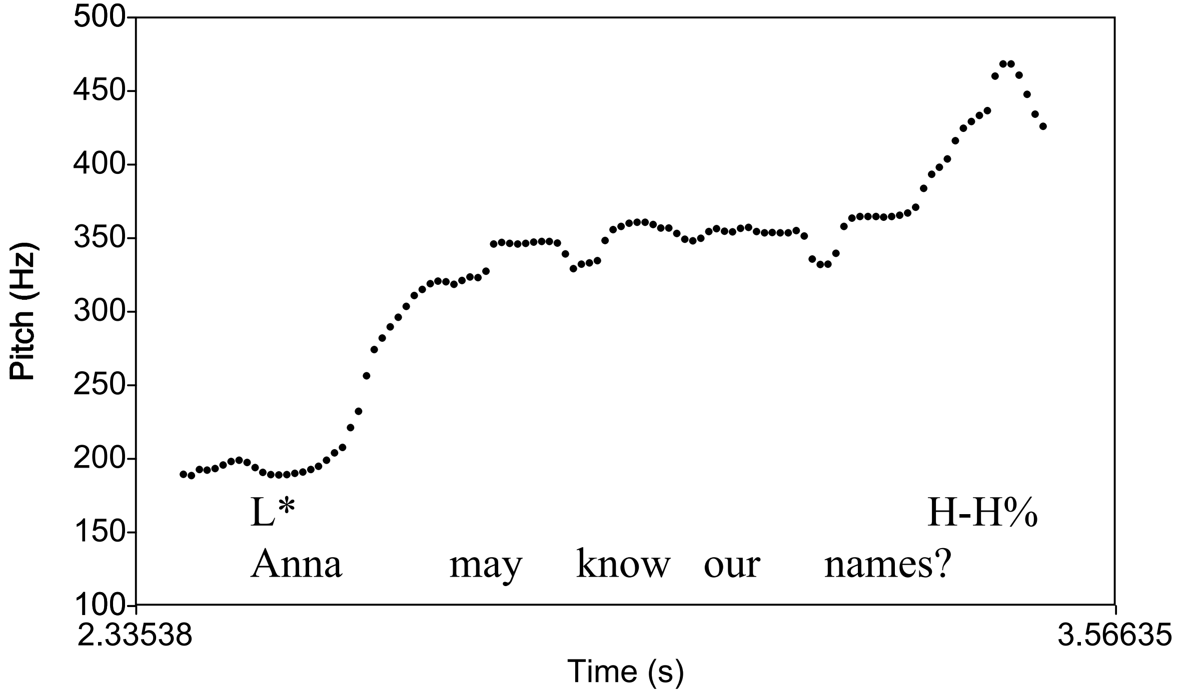


Some of P80's tone realization rules



Audio: 5_marianna.wav

Some of P80's tone realization rules



Audio: 5_anna.wav

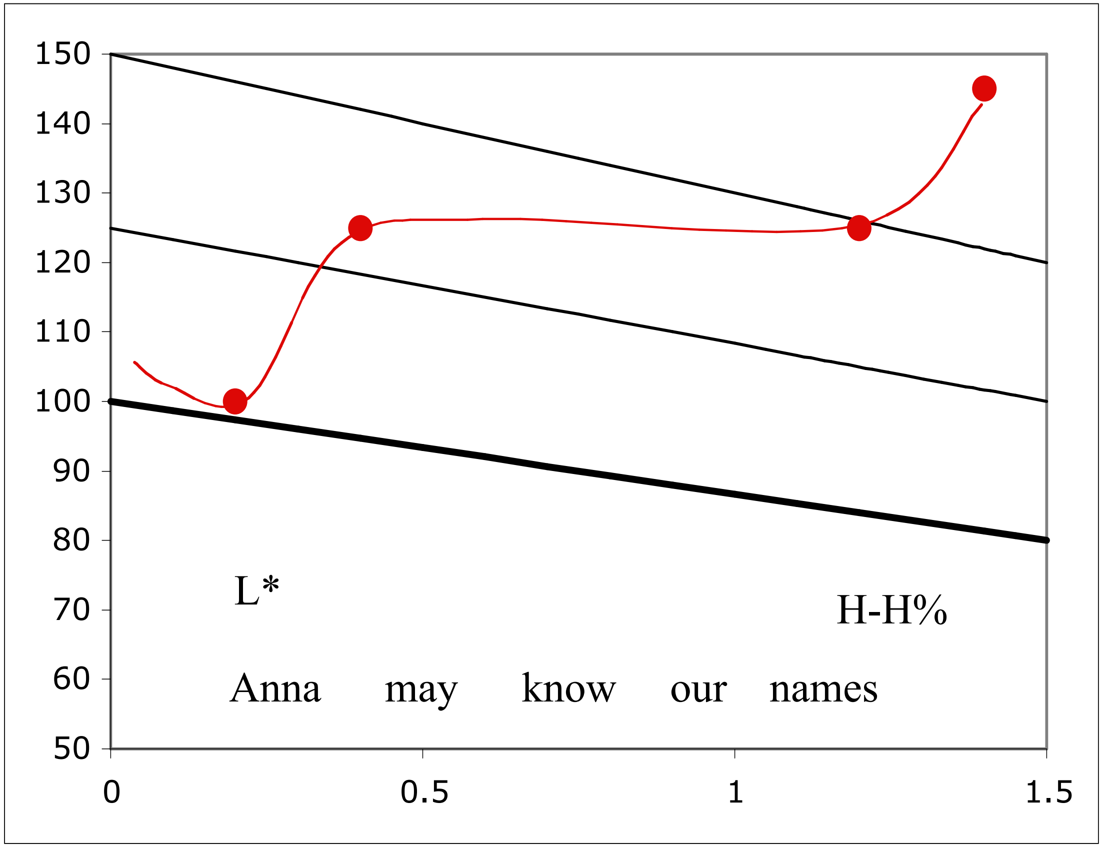
upstep

Audio file courtesy of the Ohio State University Research Foundation and the OSU ToBI Research Group. Used with permission.

Some of P80's tone realization rules

Upstep

in H-T%: /T/ = /H-/ + /T/



upstep

P80's tone realization rules

- Sharp distinction between phonetic and phonological representations.
  - Phonological: string of H, L tone associated to syllables, phrase boundaries.
  - Phonetic: f0 contour - f0 (Hz) as a function of time (s).
- Phonetic realization is simply the end product of applying a sequence of realization rules.
- What kinds of tone realization rules are there?
  - typology.

Phonetic realization of [nasal]
Cohn (1990, 1993)

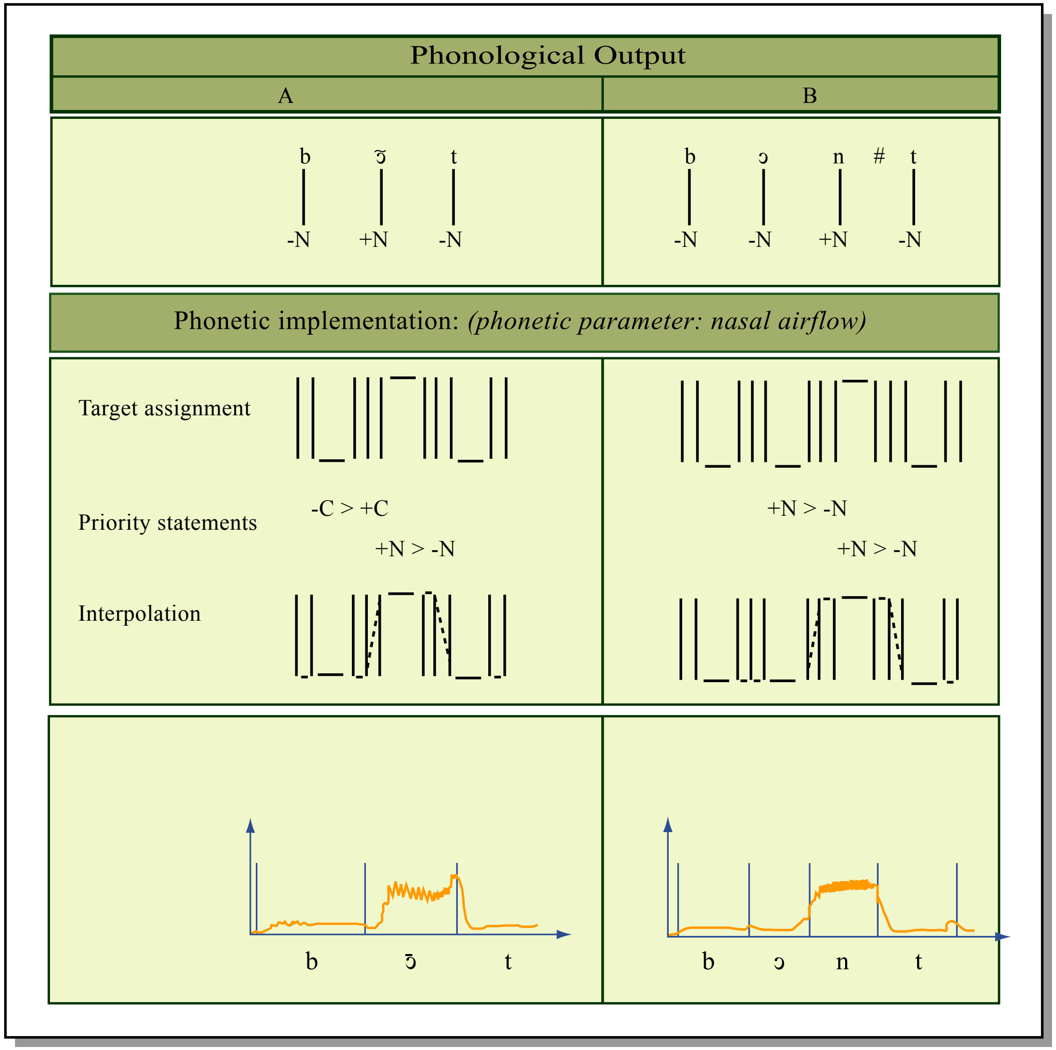


Image by MIT OpenCourseWare. Adapted from Abigail Cohn. "Phonetic and Phonological Rules of Nasalization." Ph.D. dissertation, University of California, Los Angeles, 1990. Also, from Cohn, A. "Nasalization in English: Phonology or Phonetics?" *Phonology* 10 (1993): 43-81.

Analyzing F2 transitions

- In a CV syllable, we generally observe F2 movement as a result of the articulatory movement from consonant to vowel.
- F2 is an important cue to vowel quality (front vs. back, rounding) and consonant place.
  - an important part of the phonetic realization of C and V.

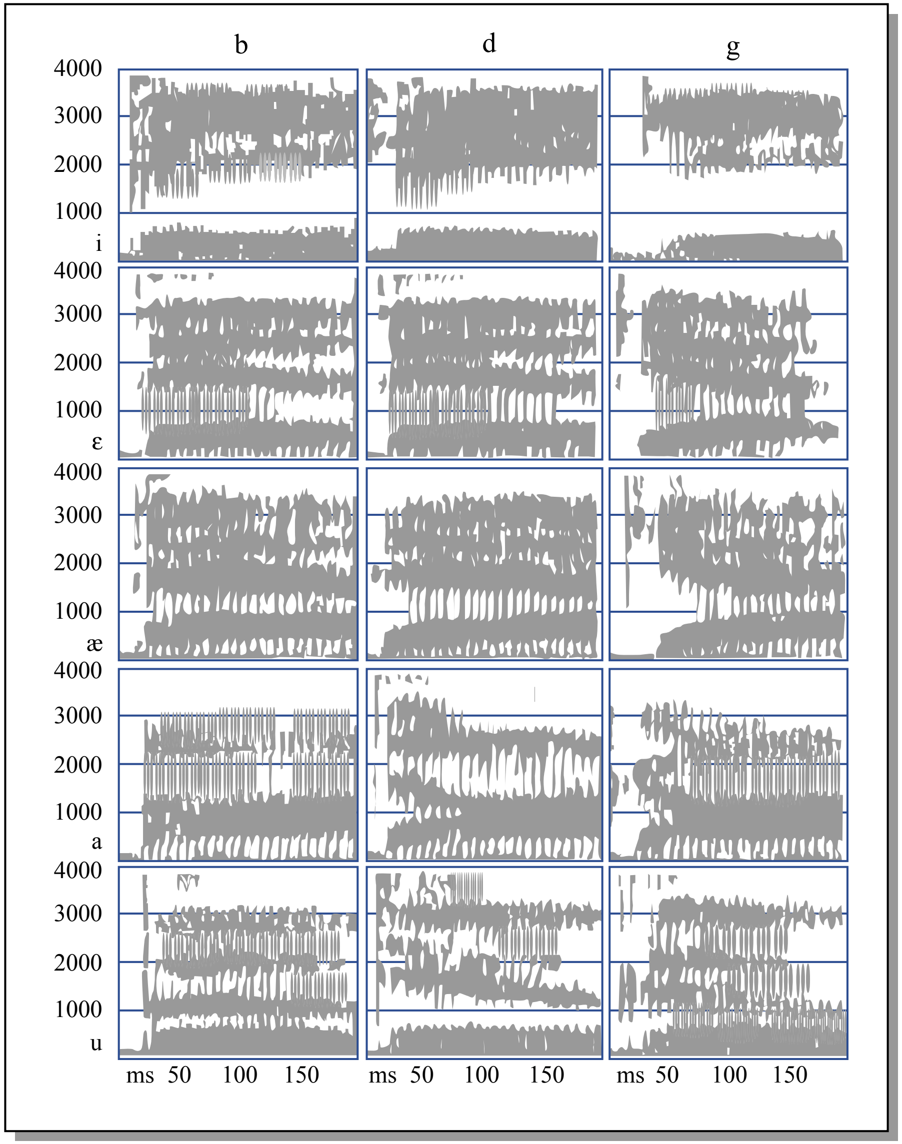
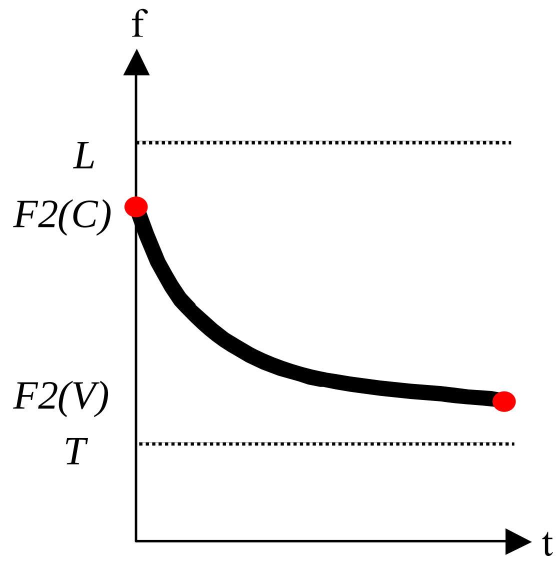
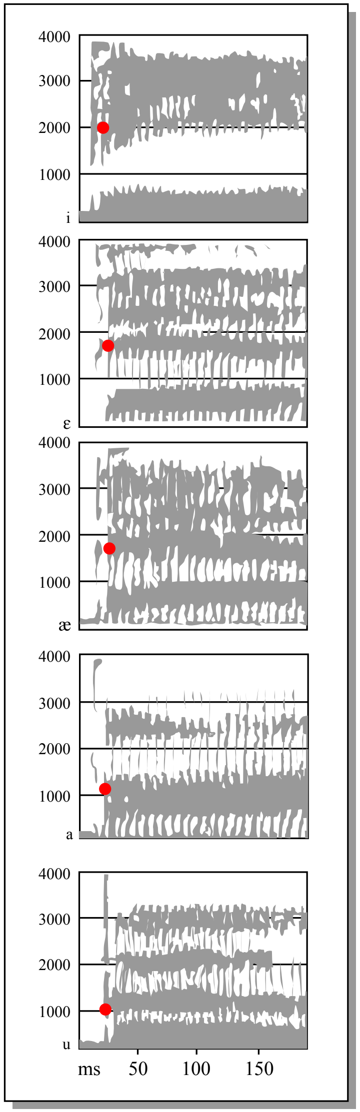


Image by MIT OpenCourseWare. Adapted from Ladefoged, Peter. Phonetic Data Analysis. Malden, MA: Blackwell, 2003.

Analyzing F2 transitions

- A targets-and-interpolation approach:
  - assign F2 targets to C and V
  - interpolate (decaying exponential?)
- Problem: F2 at stop release depends on the following vowel





- F2 at stop release depends on the following vowel
- So a context-dependent target assignment rule would be needed.

Image by MIT OpenCourseWare. Adapted from Ladefoged, Peter. *Phonetic Data Analysis.* Malden, MA: Blackwell, 2003.

- F2 in the vowel depends on the preceding consonant
  - effects are more apparent in CVC syllables.
- Again, context-sensitive target assignment would be required.

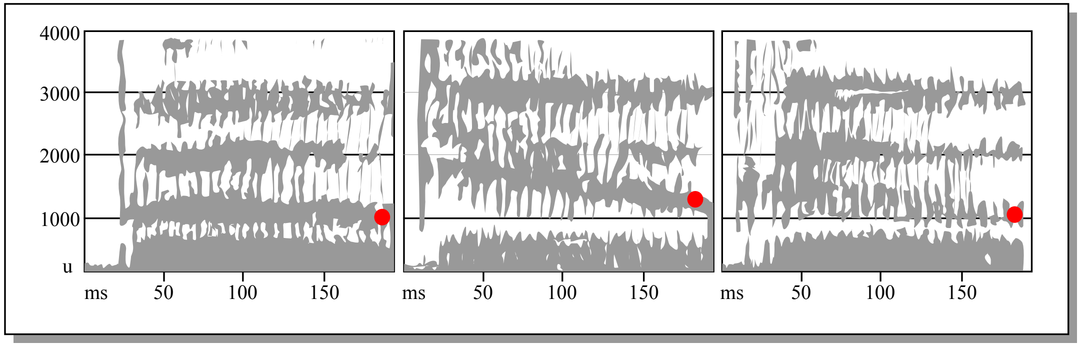


Image by MIT OpenCourseWare. Adpated from Ladefoged, Peter. *Phonetic Data Analysis*. Malden, MA: Blackwell, 2003.

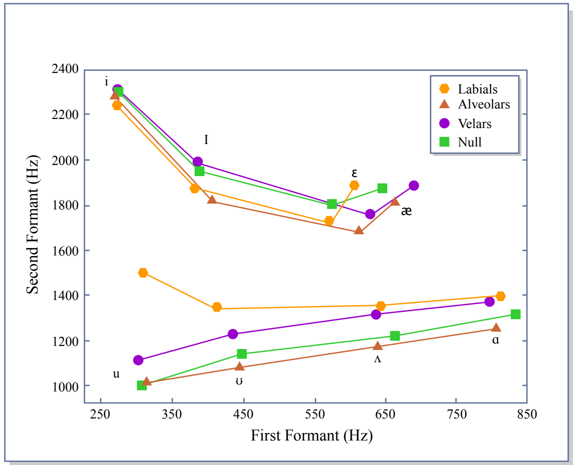
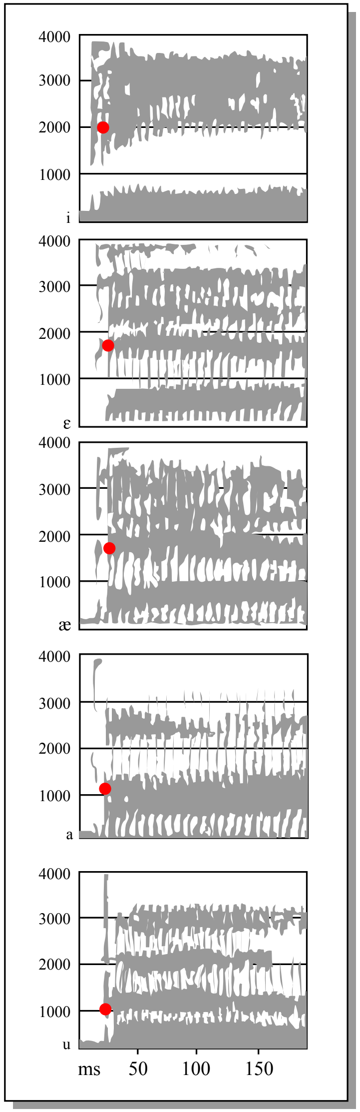


Image by MIT OpenCourseWare. Adapted from Hillenbrand, Clark, and Nearey. "Effects of consonant environment on vowel formant patterns." *The Journal of the Acoustical Society of America* 109, no. 2 (February 2001): 748-763.

Locus equations

- F2 at stop release depends on the following vowel
- Typically consonant F2 is a linear function of F2 at the midpoint of the adjacent vowel (Lindblom 1963, Sussman et al 1993, etc).
- The slope and intercept of this function depend on the consonant.



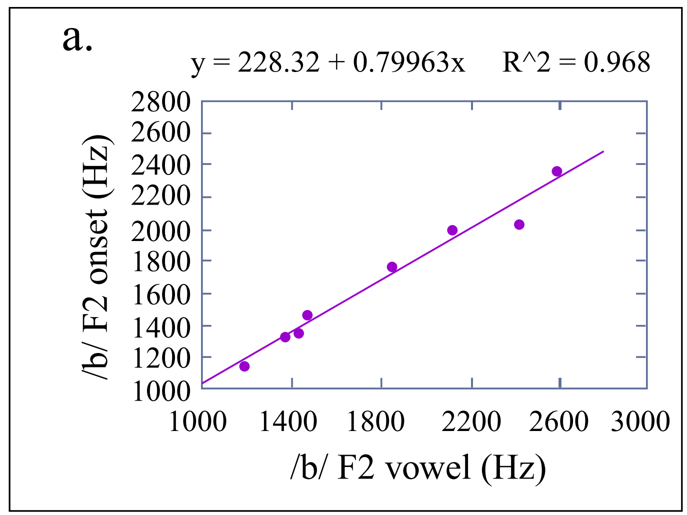


Image by MIT OpenCourseWare. Adapted from Fowler, C. A. "Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation." *Perception and Psychophysics* 55 (1994): 597-610.

Image by MIT OpenCourseWare. Adapted from Ladefoged, Peter. *Phonetic Data Analysis.* Malden, MA: Blackwell, 2003.

Locus equations

- The slope and intercept of this function depend on the consonant.

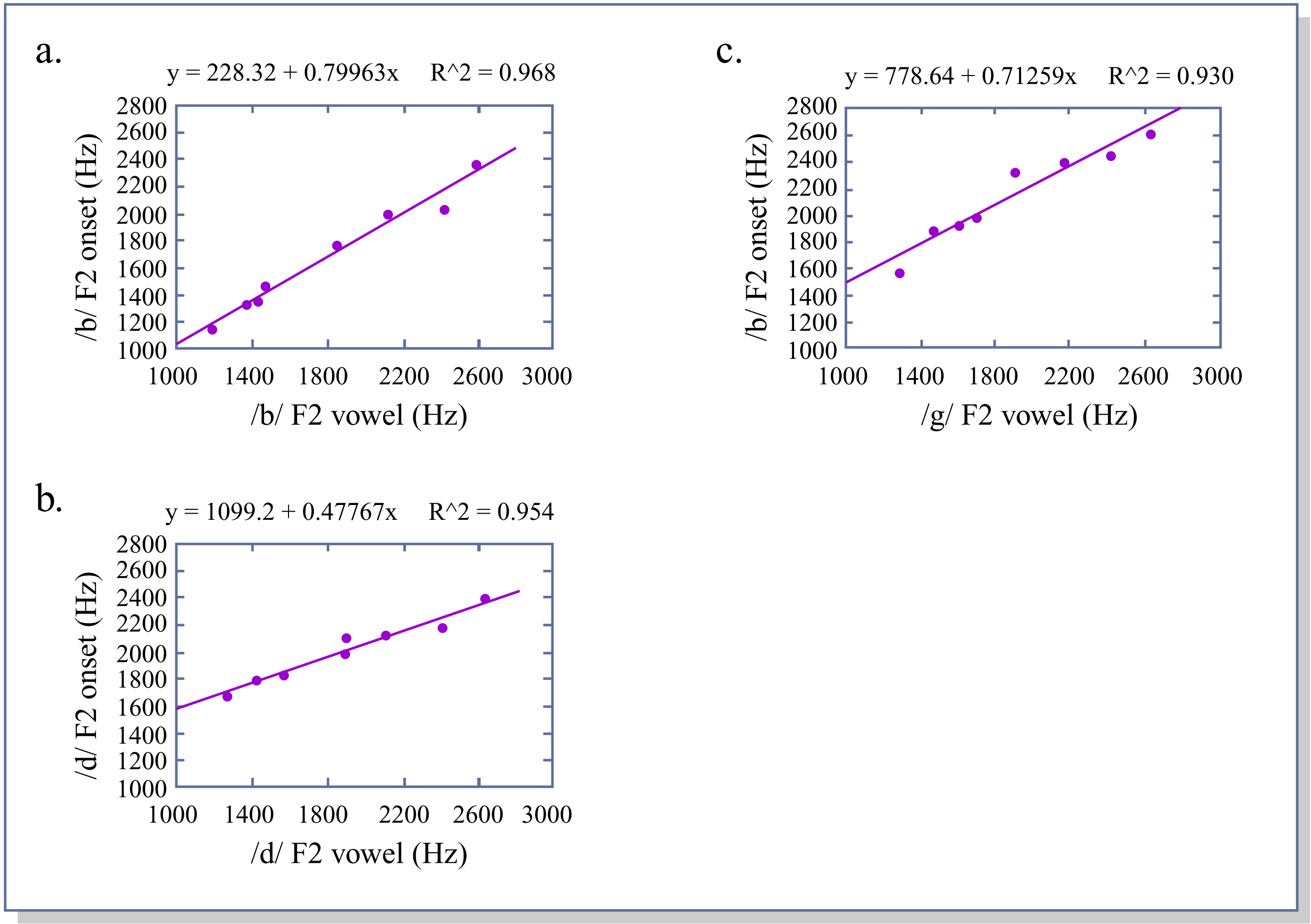


Image by MIT OpenCourseWare. Adapted from Oh, Eunjin. "Non-native Acquisition of Coarticulation: The Case of Consonant-Vowel Syllables." Ph.D. dissertation, Stanford, 2000.

Locus equations

- Linear relationship between F2 at release of consonant, *F2(C)*, and F2 at the vowel steady state or mid-point, *F2(V)*:
  - $F2(C) = k_1 F2(V) + c$ k_1, c depend on C
  - $F2(C) = k_1(F2(V)\text{-}L) + L$ $L = c/(1\text{-}k)$ (Klatt 1987)
    - interpretation: L = consonant target for F2 (‘locus’).
    - C assimilates to V: *F2(C)* deviates from L towards *F2(V)*
- This looks like a context-sensitive target assignment rule (cf. rewrite rule) but:
  - it depends on *phonetic* context: *F2(V)*
  - why does it take this form?

Vowel target undershoot

- Lindblom (1963) studied Swedish CVC sequences where $C_1=C_2$ (1 subject).
- Varied duration by manipulating stress and position in sentence.
- At long vowel durations, vowel formants are relatively independent of C context (targets).
- At shorter vowel durations, formants are progressively displaced towards values characteristic of the consonants (undershoot).
- Obtained reasonable fits with the following model:

$$F2(V) = k_2(F2(C)-T)e^{-\beta D}+T$$

Where: k_2, β depend on consonant
D is duration of the vowel.
T is the target F2 for the vowel.

- Similar results from Broad & Clermont (1987)

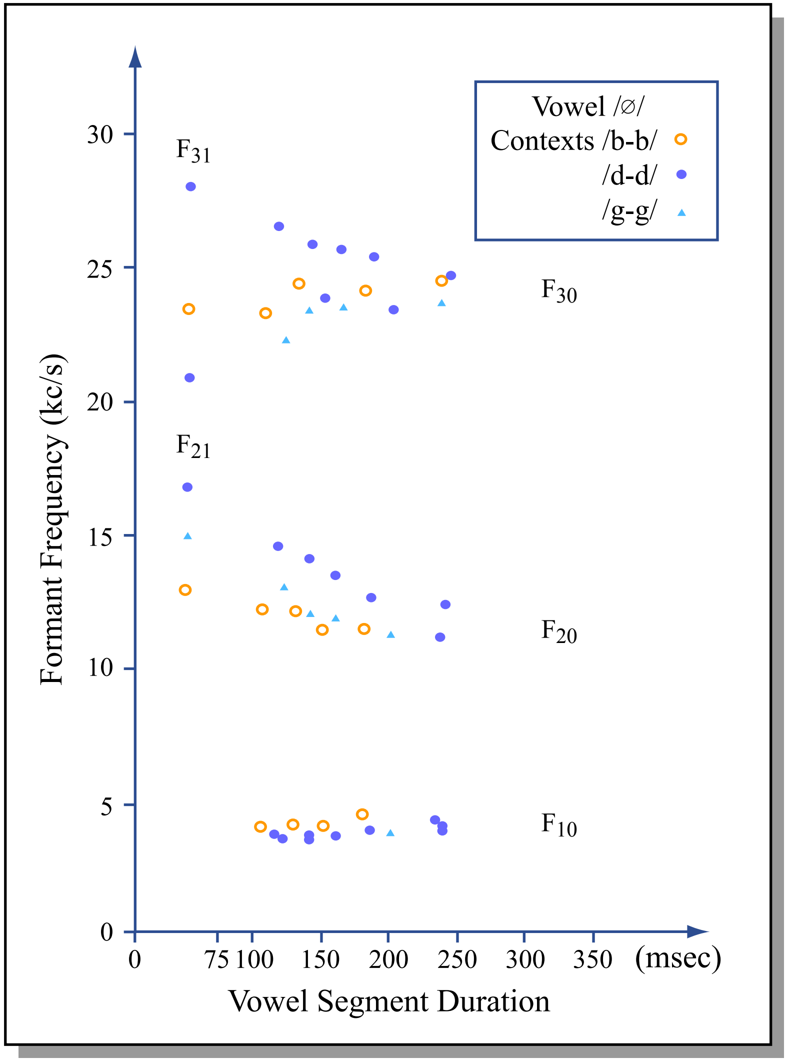
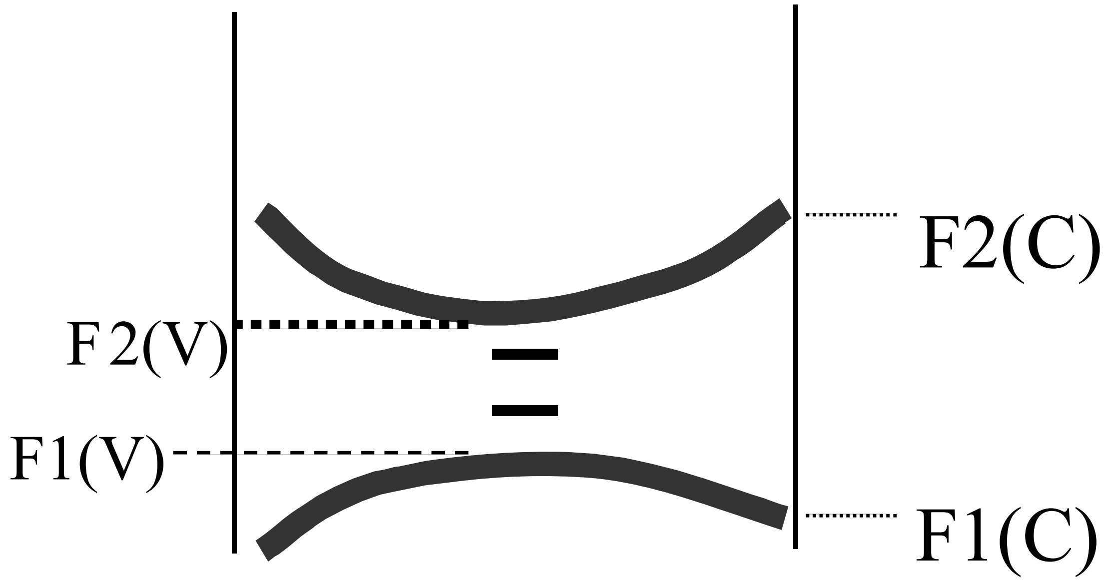


Image by MIT OpenCourseWare. Adapted from Lindblom, Björn. "Spectrographic Study of Vowel Reduction." *Journal of the Acoustical Society of America* 35 (1963): 1773-1781.

Undershoot as a consequence of effort minimization

- Faster movements are more effortful (Nelson 1983, Perkell et al 2002).
- In a CVC sequence, the articulators have to move to and from the position for the vowel.
- Undershoot results from avoiding fast movements.



Vowel target undershoot

$$F2(V) = k_2(F2(C)\text{-}T)e^{-\beta D}+T$$

- Given a fixed vowel duration, *d*, this implies a linear relationship between *F2(V)* and *F2(C)*, parallel to locus equation, where $k'_2 = k_2\, e^{-\beta d}$:

$$F2(V) = k'_2(F2(C)\text{-}T)+T$$

  - More undershoot the farther you have to move (bigger *F2(C)-T*).
- This relationship is not as well substantiated as locus equation relationship.
- Again, this looks like a context-sensitive target assignment rule, but it depends on phonetic context (*F2(C)*) and it is purely descriptive.

A constraint-based model

- F2 transitions are a compromise between
  - achieving the F2 targets for consonant and vowel
  - avoiding fast movement between the two
- Given *L*, *T*, select *F2(V)*, *F2(T)* so as to minimize violation of the following constraints:

	Constraint	Cost of violation
IDENT(C)	$F2(C) = L$	$w_c(F2(C) - L)^2$
IDENT(V)	$F2(V) = T$	$w_V(F2(V) - T)^2$
MINIMISE EFFORT	$F2(C) = F2(V)$	$w_e(F2(C) - F2(V))^2$

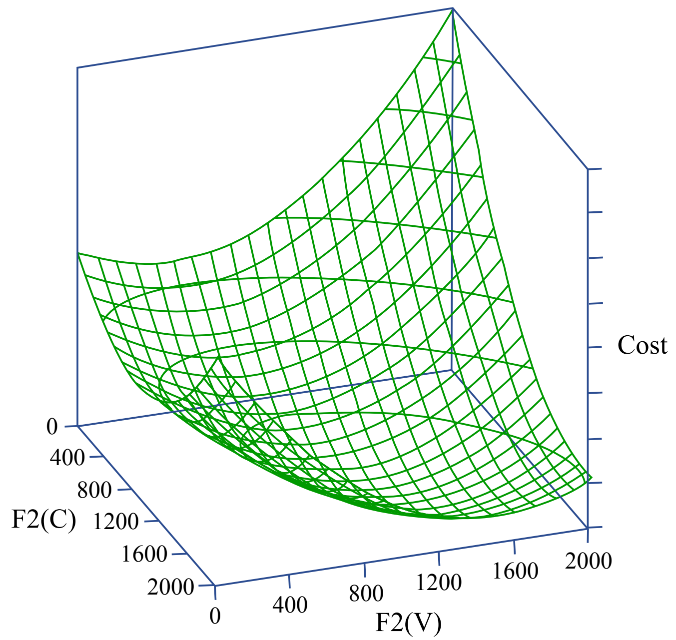
Image by MIT OpenCourseWare. Adapted from Flemming, Edward. "Scalar and Categorical Phenomena in a Unified Model of Phonetics and Phonology." *Phonology* 18, no. 1 (2001): 7-44.

  - w_i are positive weights.
- Resolving conflict: minimize summed constraint violations:

$$cost = w_c(F2(C) - L)^2 + w_v(F2(V) - T)^2 + w_e(F2(C) - F2(V))$$

Finding optimal values

- Given the form of the constraints, the cost function is smooth and convex.
  - optimum lies at the bottom of a 'bowl'.
- So optimum can be found using simple search algorithms (e.g. steepest descent).
- In this case cost function is simple enough to derive a closed form solution.



Cost plotted against *F2(C)* and *F2(V)*, with *L* = 1700 Hz, *T* = 1000 Hz, and all weights set to 1. The minimum is located at *F2(V)* = 1233 Hz, *F2(C)* = 1467 Hz.

Image by MIT OpenCourseWare. Adapted from Flemming, Edward. "Scalar and Categorical Phenomena in a Unified Model of Phonetics and Phonology." *Phonology* 18, no. 1 (2001): 7-44.

Finding optimal values

- Optimum is at the bottom of the bowl, where gradient in all directions is zero.
- Differentiate the cost function with respect to *F2(C)* and *F2(V)* to derive the gradient, then solve to find where the gradient is zero.
- Find the minimum along the *F2(C)* axis:

(i) $$\frac{\partial c}{\partial F2(C)} = 2w_c(F2(C) - L) + 2w_e(F2(C) - F2(V))$$

(ii) $$\frac{\partial c}{\partial F2(C)} = 0 \quad \text{when} \quad F2(C) = \frac{w_e}{w_c + w_e}(F2(V) - L) + L$$

(locus equation)

Finding optimal values

- Find the minimum along the *F2(V)* axis:

(iii) $$\frac{\partial c}{\partial F2(V)} = 2w_v(F2(V) - T) + 2w_e(F2(V) - F2(C))$$

(iv) $$\frac{\partial c}{\partial F2(V)} = 0 \quad \text{when} \quad F2(V) = \frac{w_e}{w_v + w_e}(F2(C) - T) + T$$

- Substitute (ii) into (iv) and vice versa, and rearrange to derive expressions for optimal values of *F2(C)* and *F2(V)* (NB there is a missing minus sign in the paper):

$$F2(C) = -u_c(L - T) + L \qquad u_c = \frac{w_e w_v}{w_e w_c + w_v w_c + w_e w_v}$$

$$F2(V) = u_v(L - T) + T \qquad u_v = \frac{w_e w_c}{w_e w_c + w_v w_c + w_e w_v}$$

The solution

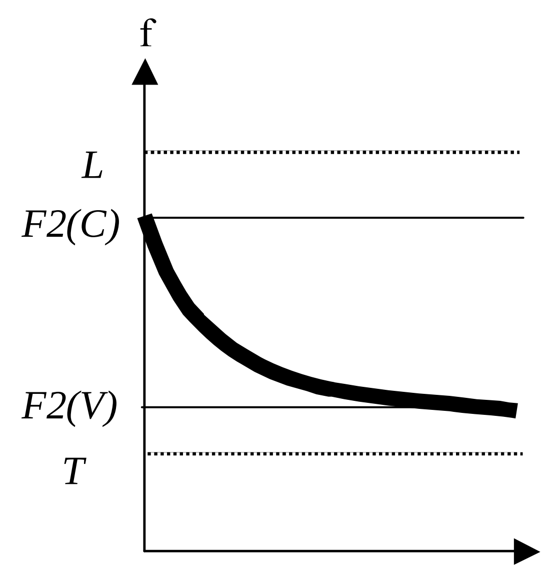
$F2(C) = -u_c(L - T) + L$

$$u_c = \frac{w_e w_v}{w_e w_c + w_v w_c + w_e w_v}$$

$F2(V) = u_v(L - T) + T$

$$u_v = \frac{w_e w_c}{w_e w_c + w_v w_c + w_e w_v}$$

- The interval between *L* and *T* is divided into three parts by *F2(C)* and *F2(V)*
  - C undershoot $w_e w_v$
  - V undershoot $w_e w_c$
  - transition $w_v w_c$
- In the proportions $w_e w_v$: $w_e w_c$: $w_v w_c$



Advantages of the constraint-based analysis

- The solution gives us two expressions that could be used as context-sensitive target assignment rules. What are the advantages of deriving these from constraints?
- Constraints are more interpretable, plausibly universal, and could be generalized to other cases, not involving formants (e.g. tone).
  - i.e. patterns have been analyzed in terms of compromise between minimizing articulatory effort and realizing perceptual targets.
  - a step towards typology in phonetic realization.
- There are few (interesting) phonetic universals but many tendencies (e.g. V shortening before voiceless obstruents). Universal but violable constraints provide a basis for analyzing these patterns.
  - e.g. conflict between pre-voiceless V shortening (duration compensation?) and realization of vowel duration contrasts (e.g. Czech)

Concerns about the analysis of F2 transitions

- The proposed analysis gets the ‘right’ answers (e.g. linear locus equations), but it is simplified in crucial respects. Can we keep the right answers as we develop the analysis?
- Effort is measured in formant space, but it should be measured articulatorily. In general acoustics are non-linearly related to articulation, so would a more realistic model of effort fail to derive linear locus equations?
  - perhaps the relationship between articulatory position and formant frequencies is relatively linear in the relevant region.
  - there are some indications that velar locus equation is not linear.
- Deviation from targets is measured in Hz, not auditory units. This is also essential to deriving the linear relationships.
  - For vowels: perhaps vowel undershoot is not linear.
  - For stops: burst is important as well as F2 at release, and the two are not independent. Perhaps the F2 target is also standing in for a burst target.

Weighted constraints vs. ranked constraints

- The data we have analyzed are analogous to allophonic variation (with lots of allophones).
- The constraints are essentially gradient OT constraints.
- Could we analyze the patterns in standard OT with constraints in a strict dominance hierarchy?
  - cf. Zhang (2001), Boersma (1998), Kirchner (1997).
- Can't rank Minimize Effort, Ident(C), Ident(V) because this would only allow for total assimilation and/or total faithfulness, whereas the observed pattern is a compromise between the three constraints, with some violation of each.
  - Minimizing the summed weighted violations yields compromise (at least given quadratic constraints).

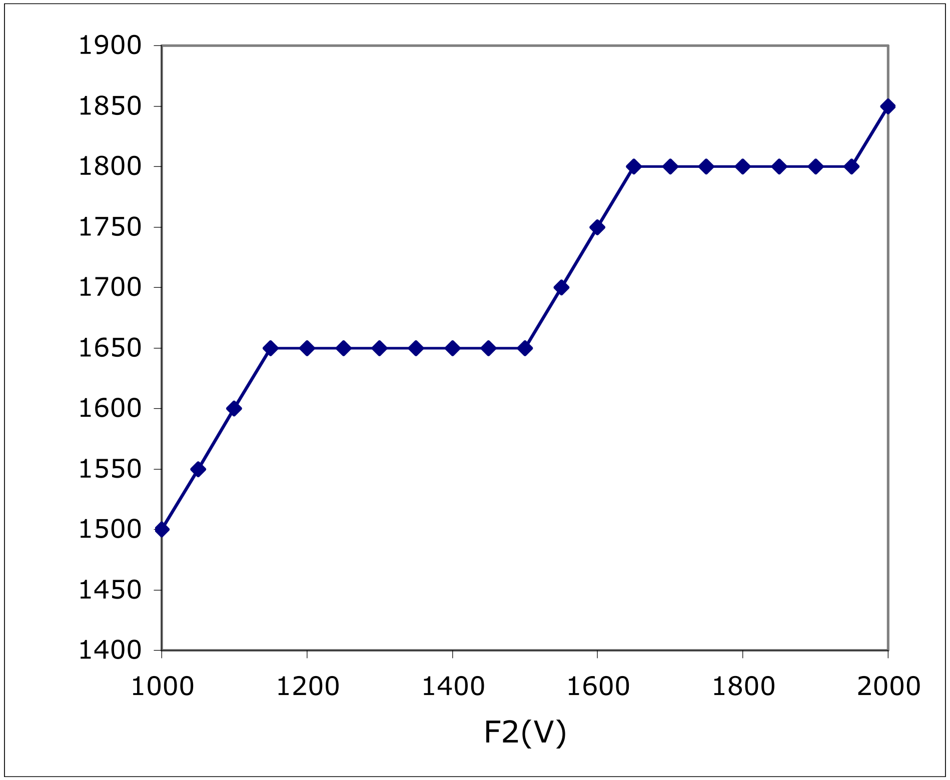
Weighted constraints vs. ranked constraints

- To derive a trade-off between constraints with strict ranking it is necessary to decompose the basic constraints into a hierarchy of sub-constraints that can be interleaved.
  - e.g. $|F2(C)\text{-}L| < 150 \gg |F2(C)\text{-}L| < 100 \gg |F2(C)\text{-}L| < 50$ etc.
- Many constraints needed!
- This makes it possible to derive all kinds of non-linear relationships between *F2(V)* and *F2(C)* whereas the attested relationships are all linear (although varying in slope).

Weighted constraints vs. ranked constraints

$\text{IDENT}_V \leq 0$
$\text{IDENT}_C \leq 300$
EFFORT≤500
$\text{IDENT}_C \leq 250$
$\text{IDENT}_C \leq 200$
$\text{IDENT}_C \leq 150$
EFFORT≤450
EFFORT≤400
EFFORT≤350
EFFORT≤300
EFFORT≤250
EFFORT≤200
EFFORT≤150
$\text{IDENT}_C \leq 100$
$\text{IDENT}_C \leq 50$
$\text{IDENT}_C \leq 0$
EFFORT≤100
EFFORT≤50
EFFORT ≤ 0

- $\text{Ident}_v \leq 0$ is undominated.
- L=1800 Hz
- Don't deviate from L unless effort >150
  - *F2(V)*<1650
- Deviate from L by up to 150 to avoid effort >150.
- Do not deviate any further unless effort >500
  - *F2(V)*<1150 Hz



Patterns of nasal coarticulation in French (Cohn 1990)

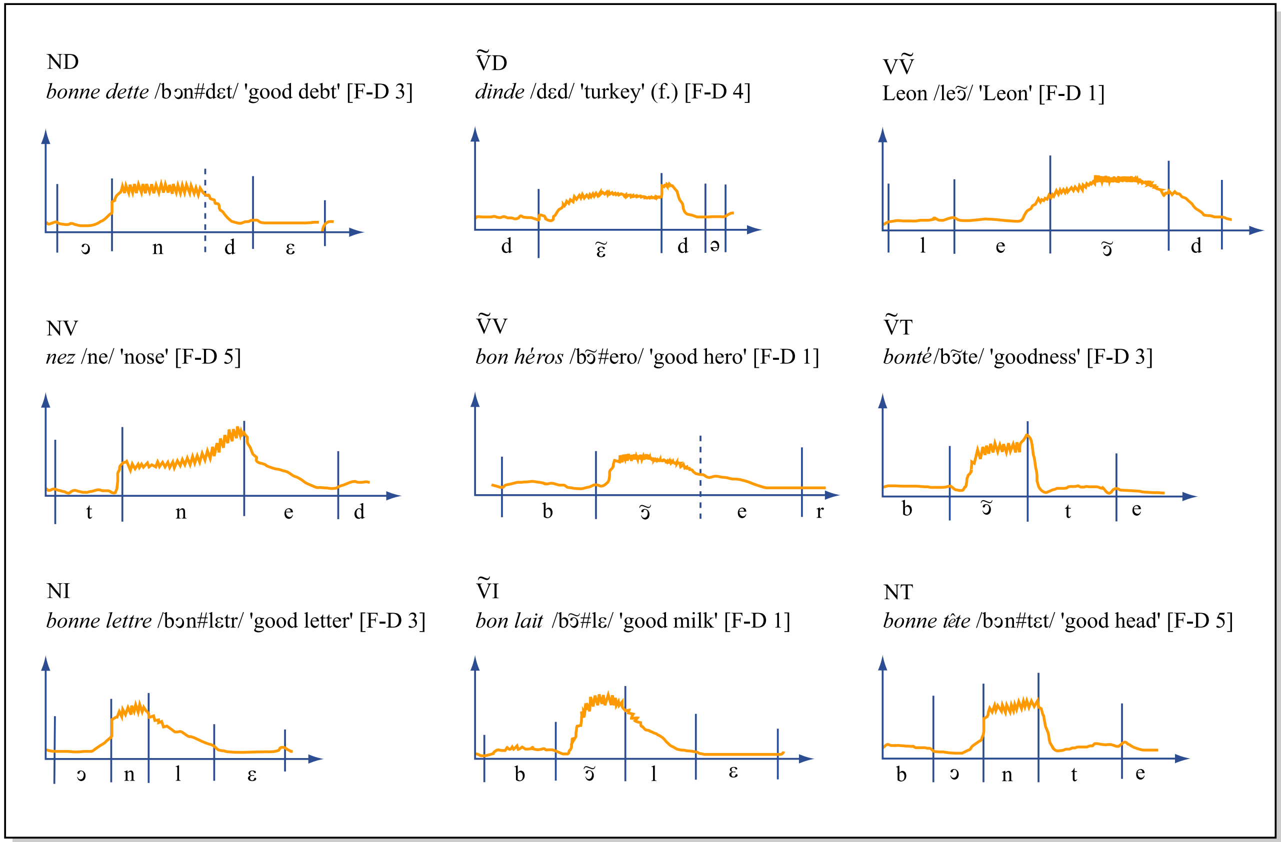


Image by MIT OpenCourseWare. Adapted from Abigail Cohn. "Phonetic and Phonological Rules of Nasalization." Ph.D. dissertation, University of California, Los Angeles, 1990. *UCLA Working Papers in Phonetics* 76.

References

- Boersma, Paul (1998). *Functional Phonology*. PhD dissertation, University of Amsterdam.
- Broad, D.J., and F. Clermont (1987). A methodology for modeling vowel formant contours in CVC context. *Journal of the Acoustical Society of America* 81, 155-165.
- Cohn, A. (1990) Phonetic and Phonological Rules of Nasalization. UCLA Working Papers in Phonetics 76.
- Cohn, A. (1993). Nasalization in English: Phonology or phonetics? Phonology 10, 43-81.
- Flemming, E. (2001) Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7-44.
- Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics*, 55(6), 597-610.
- Hillenbrand, Clark & Nearey (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America* 109:2, 748-763.
- Kawahara, S. (2006). A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82.3.
- Kirchner, Robert (1997). Contrastiveness and faithfulness. *Phonology* 14, 83-111.
- Ladefoged, P.N. (2003). *Phonetic Data Analysis*. Blackwell.
- Lindblom, Björn (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.

References

- Nelson, W.L. (1983). Physical principles for economies of skilled movement. *Biological Cybernetics* 46, 135-147.
- Perkell, Joseph S., Majid Zandipour, Melanie L. Matthies, and Harlan Lane (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America* 112, 1627-41.
- Pierrehumbert, Janet B. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.
- Steriade, Donca. 2001. Directional asymmetries in assimilation. In E. Hume & K. Johnson (eds.), The Role of Perception in Phonology. New York: Academic Press. 219-278.
- Zhang, J. (2001). *The effects of duration and sonority on contour tone distribution*. PhD dissertation, UCLA.

24.964 Topics in Phonology: Phonetic Realization
Fall 2006