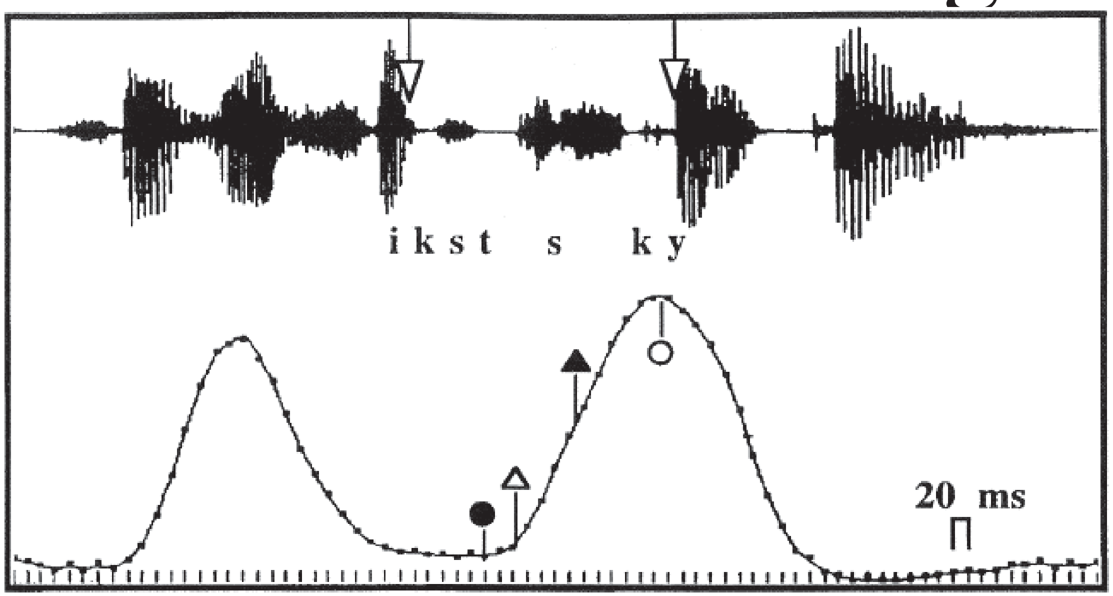
24.963

Linguistic Phonetics

Coarticulation: Gestures and timing



© Université Stendhal-Grenoble. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Noiray, Aude. "L'anticipation d'arrondissement vocalique et le Modèle d'Expansion du Mouvement Deux extensions: du français à l'anglais et de l'adulte à l'enfant." PhD diss., Université Stendhal-Grenoble III, 2007.

- Reading: Villacorta et al (2007) on real-time modified auditory feedback.
- Assignment: Write a short report outlining your final project, and progress so far (due 11/24).

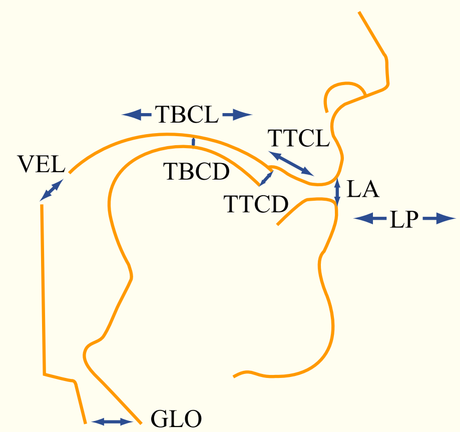
Articulatory Phonology

- Theory developed by Browman and Goldstein (1986, 1987, 1989 etc).
- Not a theory of phonology.
- The basic unit of articulatory control is the **gesture**.
- A gesture specifies the formation of a linguistically significant constriction.
- Defined within the framework of Task Dynamics (Saltzmann and Munhall 1989).

Articulatory Phonology

- A gesture specifies the formation of a linguistically significant constriction.
- The goals of gestures are defined in terms of <u>tract variables</u> (e.g. lip aperture).
- Movement towards a particular value of a tract variable is typically achieved by a set of articulators.
- A gesture takes a tract variable from its current value towards the target value.

Tract variable		Articulators involved
LP	lip protrusion	upper and lower lips, jaw
LA	lip aperture	upper and lower lips, jaw
TTCL	tongue-tip constriction location	tongue-tip, tongue-body, jaw
TTCD	tongue-tip constriction degree	tongue-tip, tongue-body, jaw
TBCL	tongue-body constriction location	tongue-body, jaw
TBCD	tongue-body constriction degree	tongue-body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis



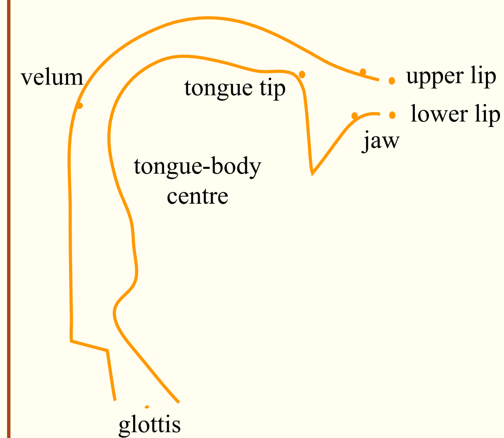
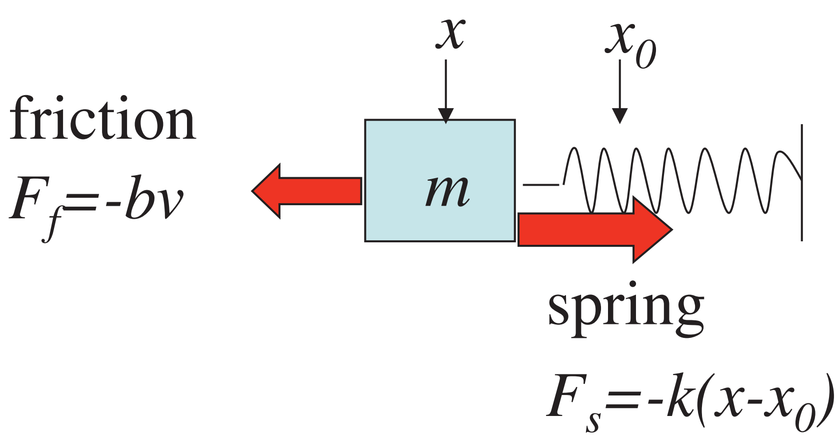


Image courtesy of MIT OCW.
Adapted from Browman and Goldstein. *Journal of Phonetics* 18 (1990): 299-320.

Articulatory Phonology

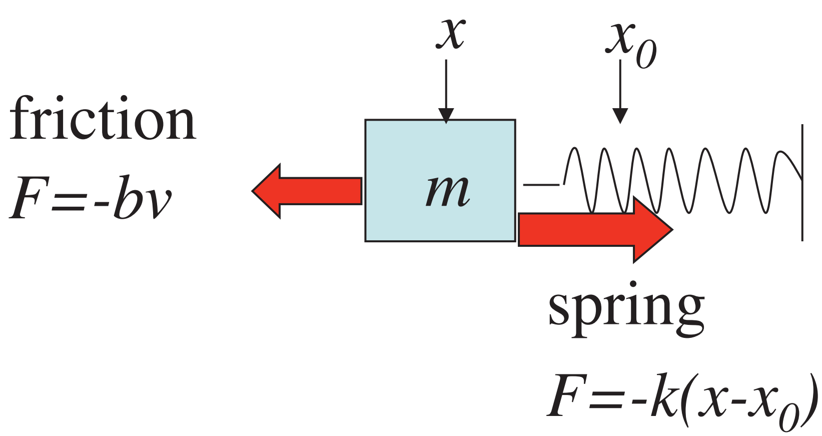
- Since a gesture involves the formation of a constriction it is usually specified by:
  - constriction degree
  - (constriction location)
  - (constriction shape)
  - stiffness
- In the Task Dynamic model, movement along a tract variable is modeled as a spring-mass system.
- In Browman and Goldstein's model critical damping is assumed, so articulators move towards the target position on the tract variable in a non-linear, assymptoting motion.

Damped mass-spring model



- Hooke's Law (linear spring): $$F_s = -k(x - x_0)$$
- Friction: $$F_f = -bv = -b\dot{x}$$
- Newton's 2nd Law: $$F = ma = m\ddot{x}$$
- Equate: $$m\ddot{x} = -b\dot{x} - k(x - x_0)$$ $$m\ddot{x} + b\dot{x} + k(x - x_0) = 0$$

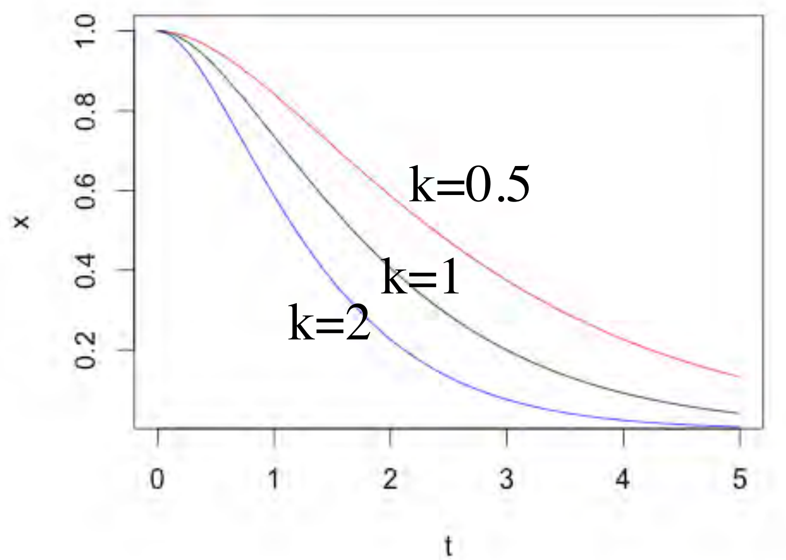
Damped mass-spring model



- If there's no damping ($b = 0$), then the solution is sinusoidal oscillation.
- B&G assume critical damping (no oscillation):

$$x(t) = (A + Bt)e^{-\sqrt{\frac{k}{m}}t}$$

Damped mass-spring model



$$x(t) = \left(x_0 + \left(\dot{x}_0 + \sqrt{\frac{k}{m}} x_0 t\right)\right) e^{-\sqrt{\frac{k}{m}} t}$$

where initial position, $x_0 = 1$ and initial velocity = 0:

$$x(t) = \left(1 + \sqrt{\frac{k}{m}} t\right) e^{-\sqrt{\frac{k}{m}} t}$$

- Gesture moves towards its target along an exponential trajectory, never quite reaching the target.
- If stiffness, k, is higher, tract variable changes faster.
- So a gesture specifies a movement from current tract variable values towards target values, following an exponential trajectory.
- Speech movements do show characteristics of being generated by a second order dynamical system (a damped 'mass-spring' system)

Articulatory Phonology

- Gestures are coordinated together to produce utterances (represented in the 'gestural score' format).

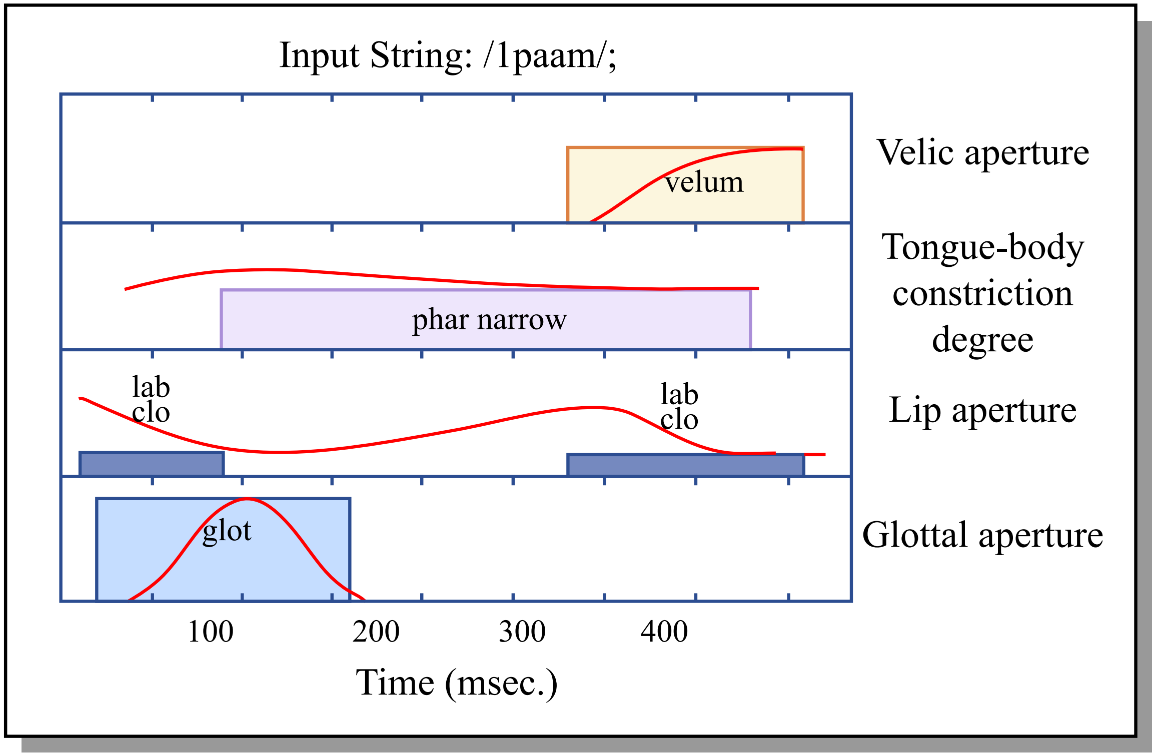


Image by MIT OCW.
Adapted from Browman and Goldstein. *Journal of Phonetics* 18 (1990): 299-320.

Articulatory Phonology

- Control parameters: articulator movements are derived from control of a limited set of tract variables and stiffness parameters.
- Gestures specify dynamic movements, but are defined in terms of static parameters.

Modeling coarticulation

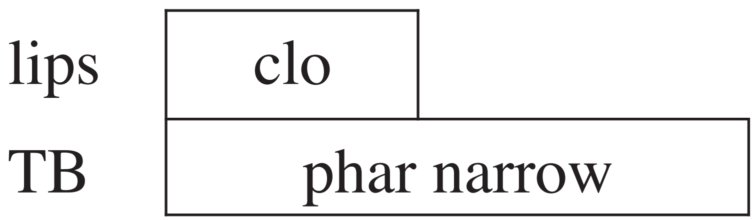
- The aspects of Articulatory Phonology presented so far could be used as part of the implementation of a constraint-based model of coarticulation.
- In these terms, the constraint-based model outlined here would serve to generate a gestural score: targets, coordination and stiffness of gestures vary to derive the preferred magnitude and timing of transitions.
  - Comparable to the 'Hybrid' model discussed by Rubin et al (1996), Saltzman (2004).
- However, most work in Articulatory Phonology and related models has analyzed coarticulation in terms of constraints on the coordination of gestures, and a mechanism of gestural blending.

Modeling coarticulation

- Gestural overlap is argued to be the basic mechanism for modeling coarticulation - coarticulation as coproduction (Fowler 1980).
  - E.g. vowel gestures will typically overlap with consonant gestures.
- When two gestures involve the same tract variables (e.g. vowels and velars, two vowels), blending results (a compromise between the demands of the two simultaneously active gestures).
- Coarticulatory effects will also result from the fact that gestures specify movement from the current location to form a particular constriction, so the articulator movements resulting from a given gesture will depend on the initial state of the articulators.

Modeling C-V coarticulation: Articulatory Phonology

- overlap:



[pa]

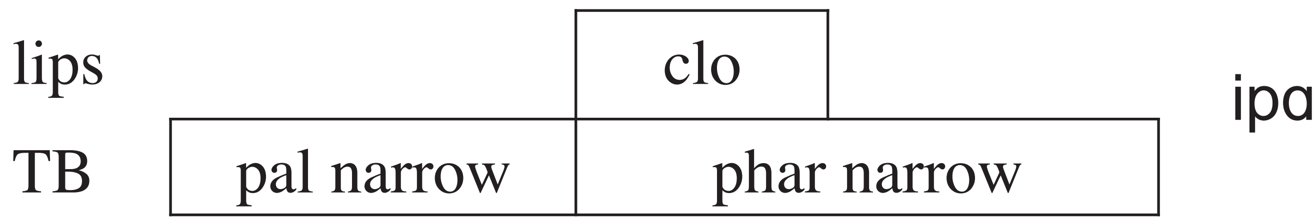
- V and C gestures in a CV begin simultaneously (Goldstein et al 2006).
- C gesture is completely overlapped by the tongue body gesture associated with the following V.
- Hence 'anticipation' of V tongue body position during labials, coronals.
  - Resistance of coronals to coarticulation does not follow directly.
- Blending of velar TB gesture with vowel TB gesture.

Modeling C-V coarticulation

- Vowel undershoot cannot be analyzed in terms of overlap because the midpoint of a vowel is not overlapped by the preceding consonant gestures.
- In principle target undershoot can be derived in this model: if a gesture is too short relative to its stiffness then it can fall short of its target.
- In a CVC sequence, the vowel can be prevented from reaching its target if a conflicting C gesture begins before the V target is reached.

Modeling V-to-V coarticulation: Articulatory Phonology

- overlap:



- movement to V2 starts at the same time as movement to C - anticipatory coarticulation.
  - may not be sufficient: V2 can affect the mid-point of V1. Magen (1997) found some effects of V3 at offset of V1 in $bV_1 bəbV_3 b$ words.
- movement to V2 starts from position of V1 - carryover coarticulation.

Modeling V-to-V coarticulation: Articulatory Phonology

- blending:

TB		clo	yɡɑ
TB	pal narrow	phar narrow	

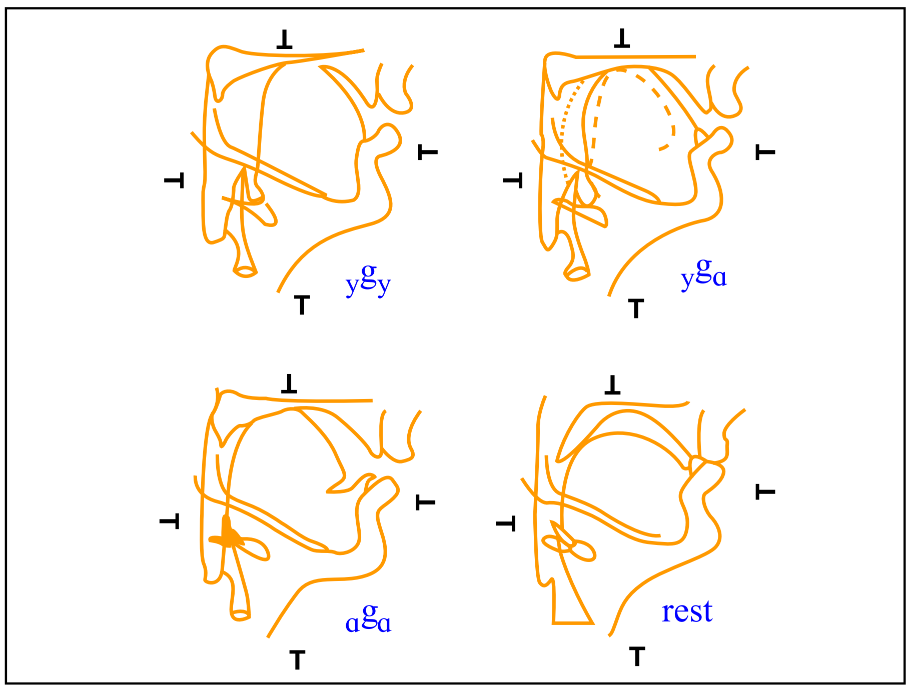


Image by MIT OCW.
Adapted from Öhman, S.E.G. "Coarticulation in VCV utterances: Spectrographic measurements." *Journal of the Acoustical Society of America* 39 (1966): 151–168.

The temporal extent of coarticulation

- Analyzing coarticulation in terms of these kinds of constraints on the coordination of gestures leads to the prediction that anticipatory coarticulation should be relatively limited.
- Cf. Bell-Berti & Harris's (1981) 'frame model' of coarticulation: movement towards the targets for a segment begin at a fixed duration before that segment.
  - E.g. lip-rounding begins at a fixed duration preceding the acoustic onset of a rounded vowel.
  - Although this anticipation can be overriden by conflicting requirements of a preceding segment.

The temporal extent of coarticulation

- Contrast Benguerel & Cowan's (1974) 'look-ahead' model of lip rounding coarticulation:
  - Movement towards a target begins as soon as the preceding target has been realized.
  - Consonants lack targets for lip rounding.
  - Movement towards the lip position for a rounded vowel begins at the onset of the preceding consonant cluster, no matter how long.
  - 'une sinistre structure' [istrstry]

The temporal extent of coarticulation

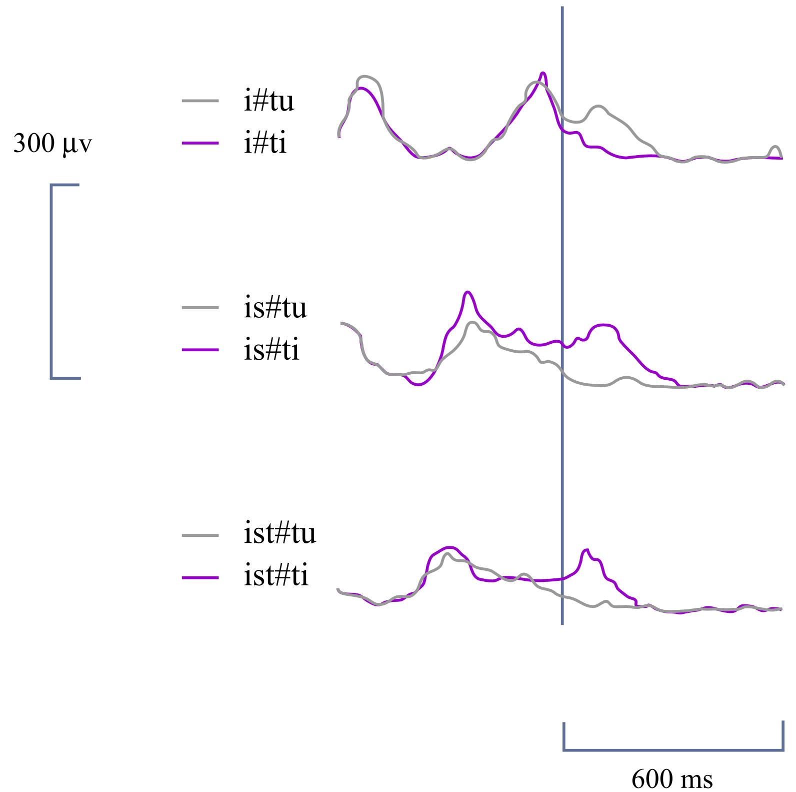
- Models where:
  - Segments can lack targets on some dimensions (phonetic underspecification), or where targets are violable/windows and
  - constraints favor minimization of peak velocities
- predict variable extent of anticipatory coarticulation depending on context.
- Unlike Benguerel & Cowan, they predict that earlier onset should be accompanied by slower movement (or reduced undershoot).

The temporal extent of anticipatory coarticulation

- Tests of the extent of anticipatory coarticulation have yielded mixed results.
- E.g. Benguerel & Cowan (1974) on French: upper lip protrusion in $C_1 \ldots C_n V$ sequences where V is rounded begins 'most frequently on or before C_1'
- But Boyce et al (1990) found that lip-rounding precedes rounded vowel onset by a relatively fixed duration in English.
  - electromyography
  - <u>lower</u> lip muscle (OOI)

Coarticulation between non-adjacent segments

- Boyce et al (1990)



Ensemble-averaged EMG activity for 15 to 20 tokens of 3 pairs of utterances with increasing numbers of "neutral" consonants. The vertical mark indicates the lineup point for averaging tokens, at the acoustic onset of the second vowel.

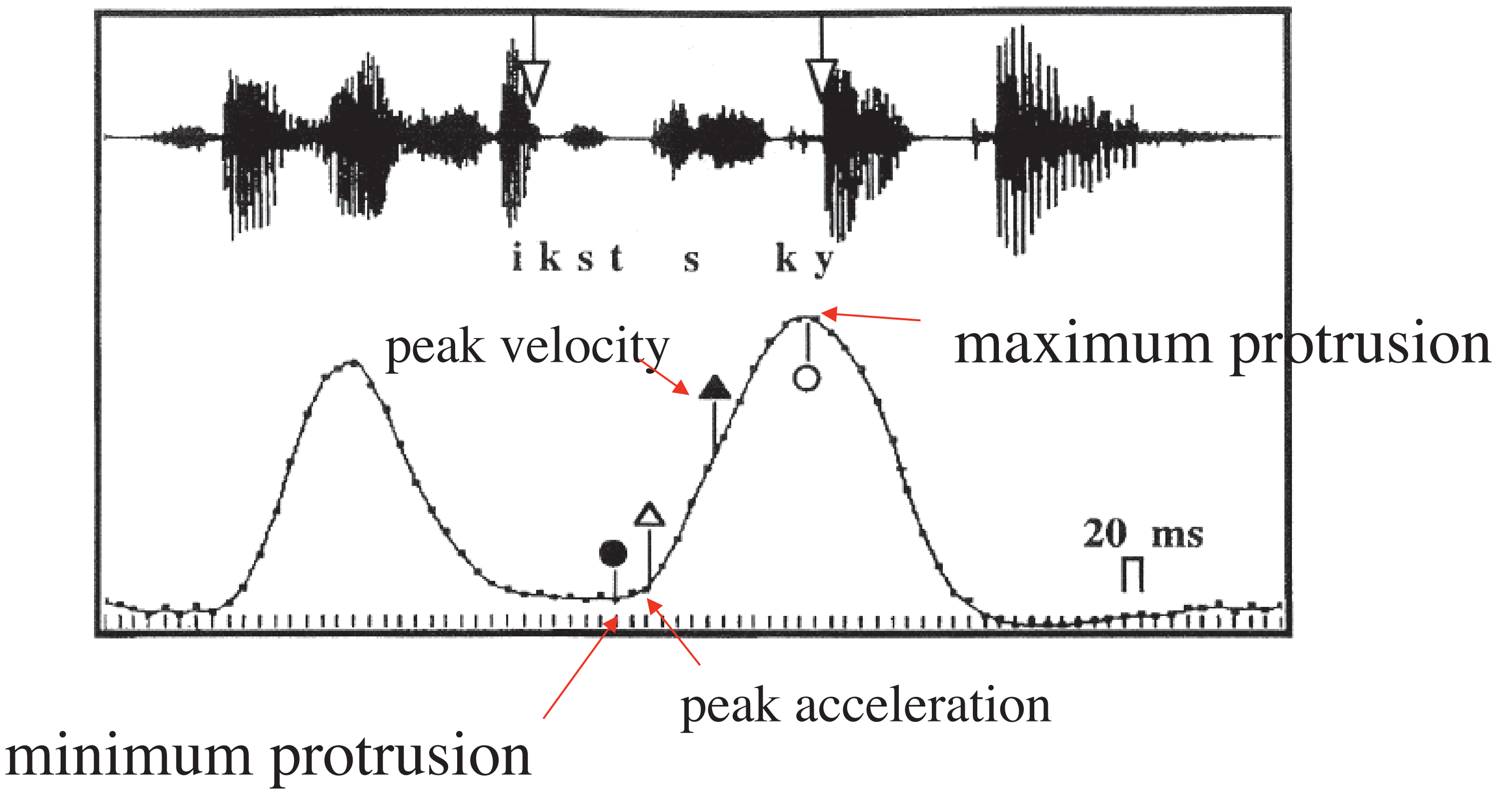
Image by MIT OCW.
Adapted from Boyce, S.E., R.A. Krakow, F. Bell-Berti, and C.E. Gelfer. "Converging sources of evidence for dissecting articulatory movements into core gestures." *Journal of Phonetics* 18 (1990): 173-188.

The temporal extent of anticipatory coarticulation

- More recent studies provide strong evidence for partial look-ahead in French and English (Abry & Lallouache 1995, Noiray et al 2011).
- Abry & Lallouache studied upper lip protrusion in French [iC_ny] sequences, n = 0-5.
  - e.g. Sixtes sculptèrent [-ikstsky-]
  - speakers wore blue lipstick, recorded with two video cameras (side, front).
  - lip position was tracked from the videos (50 Hz).
  - trajectory was interpolated between samples.

Abry & Lallouache 1995

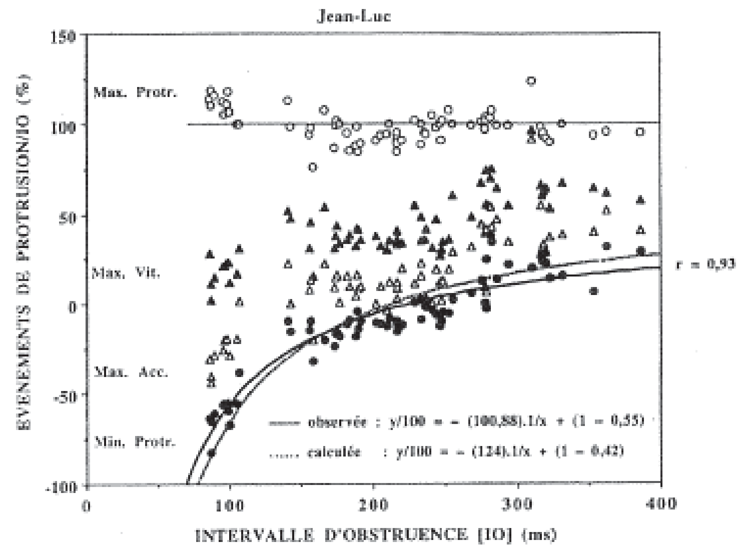
C cluster duration

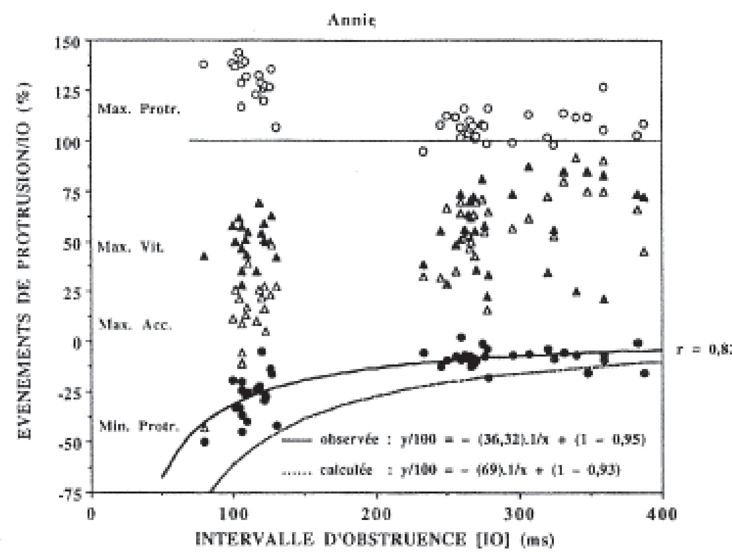


© Institut de la Communication Parlée. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.

Source: Abry, Christian, and Tahar Lallouache. "Le MEM: un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français." Les Cahiers de l'ICP. Bulletin de la communication parlée 3 (1995): 85-99.

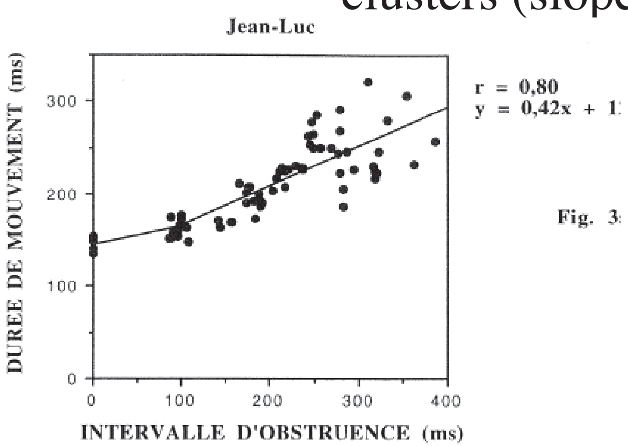
- Timing of lip movement events as a percentage of cluster duration.
- Maximum protrusion occurs at, or slightly after, vowel onset.
- Protrusion movement occupies a smaller proportion of the cluster duration for longer clusters.



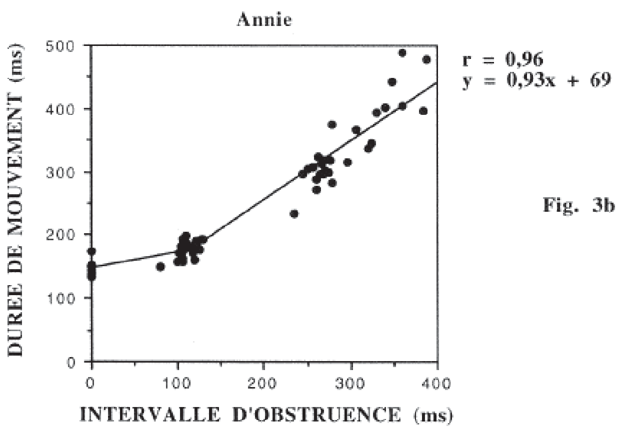


© Institut de la Communication Parlée. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/. Source: Abry, Christian, and Tahar Lallouache. "Le MEM: un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français." Les Cahiers de l'ICP. Bulletin de la communication parlée 3 (1995): 85-99.

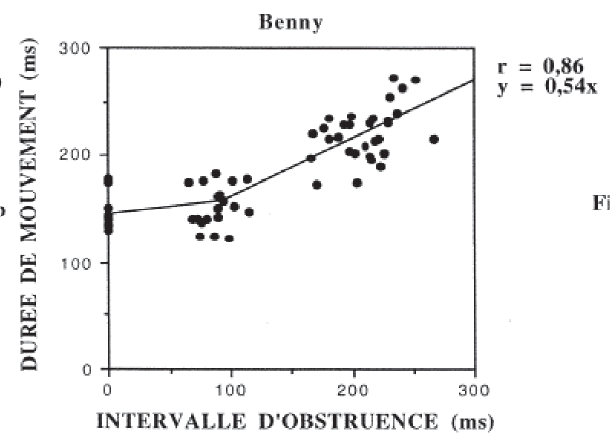
- As consonant clusters lengthen from a minimum of ~100 ms, the duration of the lip protrusion movement (min to max) increases approximately linearly.
  - Slope less than 1, range from 0.42 to 0.93.
- Lip movement cannot be shorter than ~140 ms, as observed from [iy] sequences.
- Not frame behavior - that would yield a slope of 0.
- Not simple look-ahead:
  - Lip protrusion can begin over 50ms before the onset of the C interval with single consonants (duration ~100 ms).
  - Lip protrusion begins after C cluster onset with longer clusters (slope <1)



$y = 0.42x + 124$



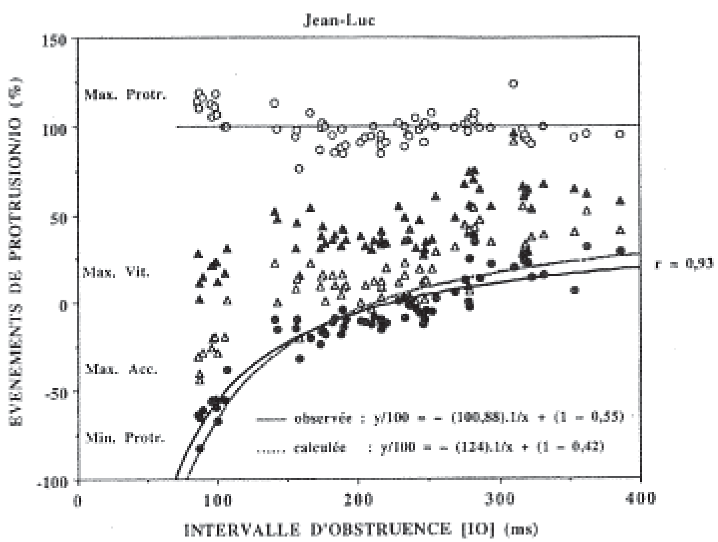
$y = 0.93x + 69$



$y = 0.54x + 105$

© Institut de la Communication Parlée. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/. Source: Abry, Christian, and Tahar Lallouache. "Le MEM: un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français." Les Cahiers de l'ICP. Bulletin de la communication parlée 3 (1995): 85-99.

- Not simple look-ahead:
  - Early onset of movement correlates with a longer (presumably slower) movement, not early attainment of target (cf. Benguerel & Cowan's look-ahead model).
- Maximum protrusion occurs a little later at shortest cluster durations.
  - Not predicted by frame model or simple look-ahead models.



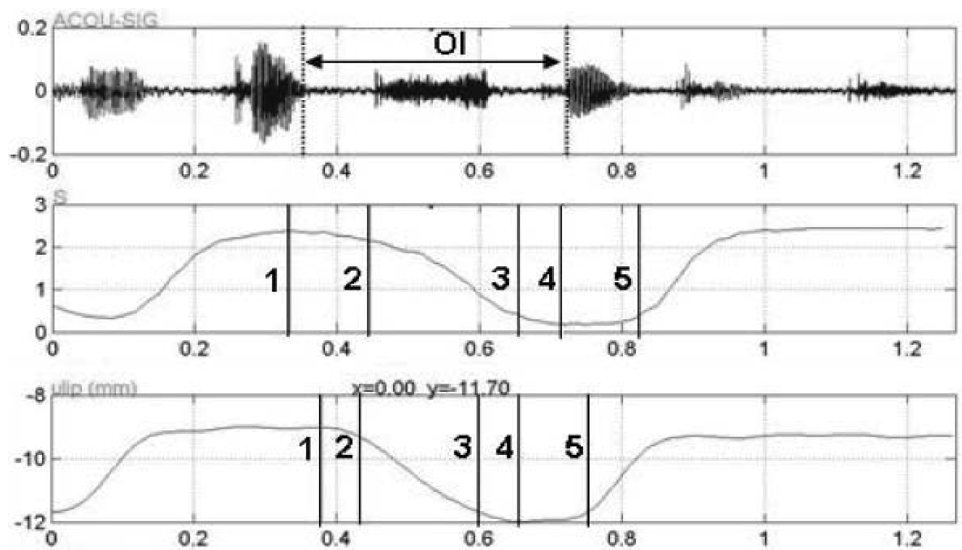
© Institut de la Communication Parlée. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/. Source: Abry, Christian, and Tahar Lallouache. "Le MEM: un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français." Les Cahiers de l'ICP. Bulletin de la communication parlée 3 (1995): 85-99.

- Perkell & Matthies (1993) found similar, but much noisier, patterns of lip protrusion anticipation in English.
- Noiray et al (2011) using similar methods to Abry & Lallouache, found that Canadian English speakers generally had too little lip protrusion in [u] for reliable measurement of movement onset.
- Measured lip area instead.
  – More acoustically relevant than lip protrusion
  – but can be affected by jaw height as well as lip movement, so Noiray et al also recorded a bite-block condition: speakers talk with a block clenched between the teeth to fix jaw height.
- Recorded [iC$_n$u] sequences, n = 0-4.

lip area

lip protrusion

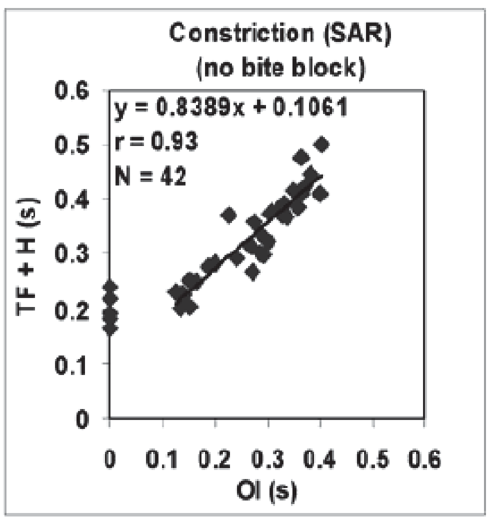
(inverted)

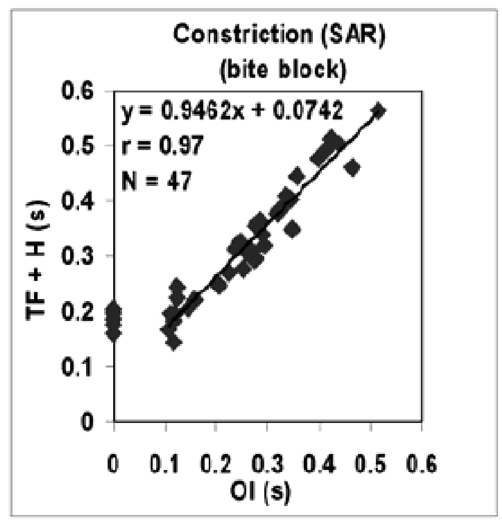


© Université Stendhal-Grenoble. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Noiray, Aude. "L'anticipation d'arrondissement vocalique et le Modèle d'Expansion du Mouvement Deux extensions: du français à l'anglais et de l'adulte à l'enfant." PhD diss., Université Stendhal-Grenoble III, 2007.

Noiray et al 2011 - English

- Plot duration of lip movement+constriction (closure+hold) against consonant cluster duration (OI 'Obstruent Interval').
- Regression line for *n* = 1-4.
- Slopes from 0.84-1.12 without bite-block, 0.75-1.08 with bite-block.
- No information about the timing of onset/offset of movement relative to acoustic onset of the vowel.

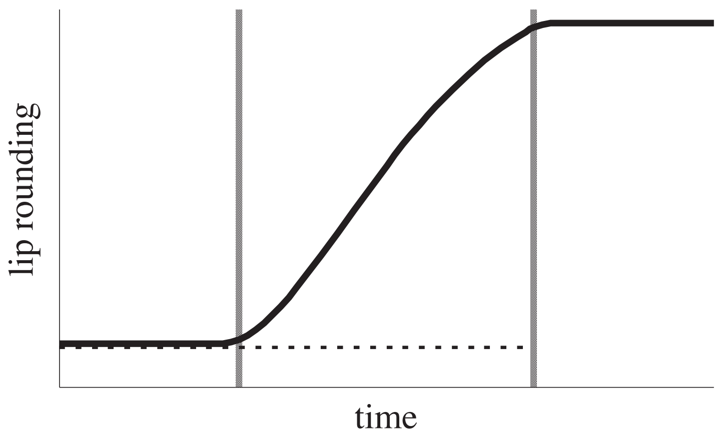


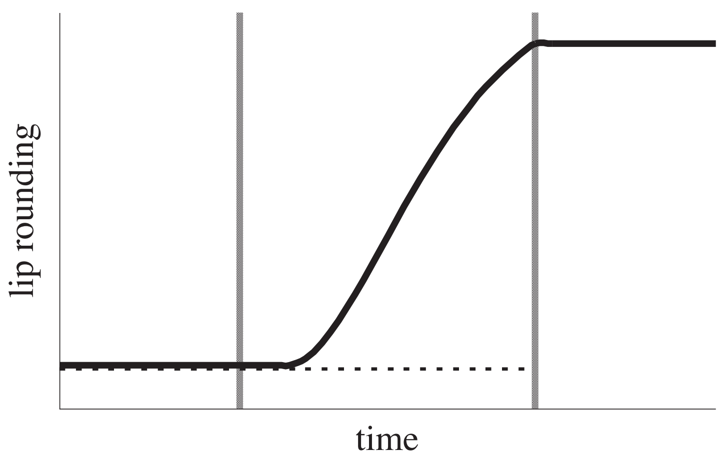


© Université Stendhal-Grenoble. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Noiray, Aude. "L'anticipation d'arrondissement vocalique et le Modèle d'Expansion du Mouvement Deux extensions: du français à l'anglais et de l'adulte à l'enfant." PhD diss., Université Stendhal-Grenoble III, 2007.

A constraint-based analysis

- These qualitative patterns of anticipatory coarticulation can be understood in terms of a compromise between effort minimization and faithfulness to lip rounding targets (or their perceptual equivalents).
- Assume rounded vowels have a target for lip-rounding extending through the duration of the vowel, unrounded vowels require spread lips throughout the vowel.
- So in an [iy] sequence any transition from unrounded to rounded incurs violations of faithfulness to targets - favors minimizing transition duration.
- Minimization of peak velocity favors maximizing the duration of the rounding movement.
- Minimum movement duration derives from the optimal trade-off between these two conflicting constraints.
- If Cs lack rounding targets, then transition during Cs incurs no cost, so transitions should be at least as long as the consonant cluster in [iC_ny] (full look-ahead, with a minimum duration).
- If Cs have low weight targets for unrounded lips, then we can derive partial look-ahead: extent of look-ahead depends on relative weights of effort and faithfulness to C targets.





- Effort minimization (minimizing peak velocity) favors a longer slower transition.
- Faithfulness to the low weighted C targets favors a shorter faster transition.
- Partial look-ahead can result from a compromise between these two constraints.

A constraint-based analysis

- It is plausible that many consonants have targets pertaining to lip position
  - Most consonants prefer unrounded lips, e.g. [t]: to ensure high frequency burst, high F2 onset (i.e. not directly a lip target).
  - [s] produced with some lower lip protrusion, perhaps to open lips with high jaw, to allow radiation of frication noise.
  - [ʃ] is produced with lip protrusion to lower peak frequency in frication noise.
- This line of analysis can explain the fact that maximum protrusion occurs later at the shortest consonant durations: this is another way to lengthen the movement, minimizing effort, at the cost of violating faithfulness to the [y] target.
- This arises at the shortest C durations because the rounding movement cannot be completed during a short consonant, so rounding has to begin during the unrounded vowel. Shifting the offset of the movement trades a small violation of faithfulness to [y] for a reduction in violation of faithfulness to [i].

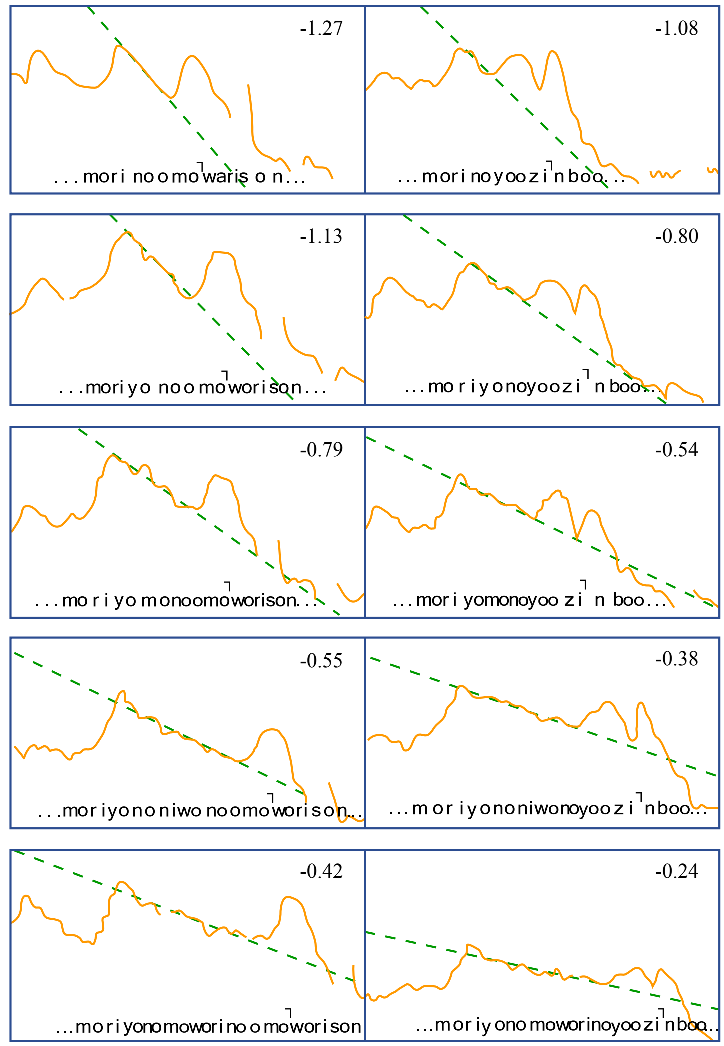
Predictions of the analysis

- 'Frame' behavior is expected where targets for all segments have relatively high weights.
  - High weights -> transition duration is minimized.
  - If weights are high but vary substantially from segment to segment we would expect the timing of the transition to vary to minimize transition during targets with highest weights (cf. tones).
- Look-ahead is expected across segments with relatively low weight targets.
  - prime example is fundamental frequency in intonation languages.
- Earlier onset of transitions ('look-ahead') should generally be accompanied by longer, slower transitions rather than early target attainment (cf. Benguerel & Cowan's look-ahead model).

Clear 'look-ahead' in intonation:

Pierrehumbert and Beckman (1988) - Fundamental frequency in Tokyo Japanese.

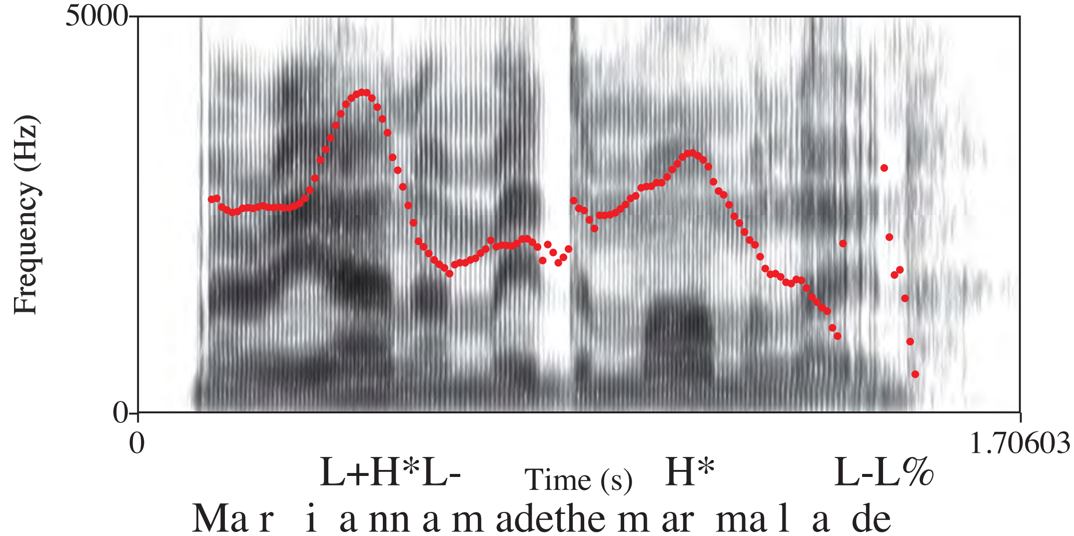
- Accentual phrase is marked by initial LH rise.
- Phrase-final L%.
- With unaccented words, F_0 interpolates from H to L.



Unaccented phrases with 3, 4, 5, 7, and 8 morae before accentual-phrase boundaries with two different allophones of L%. The dashed line in each panel is a regression line fitted to all f_0 values between the peak for the phrasal **H** in the first phrase and the minimum for the interphrasal L%. The number in the upper right-hand corner is the slope of the regression line. Here a right corner in the transcription indicates the location of an accent.

Image by MIT OCW.
Adapted from Pierrehumbert, J., and Mary E. Beckman. *Japanese Tone Structure*. Cambridge MA: MIT Press, 1988. ISBN: 9780262161091.

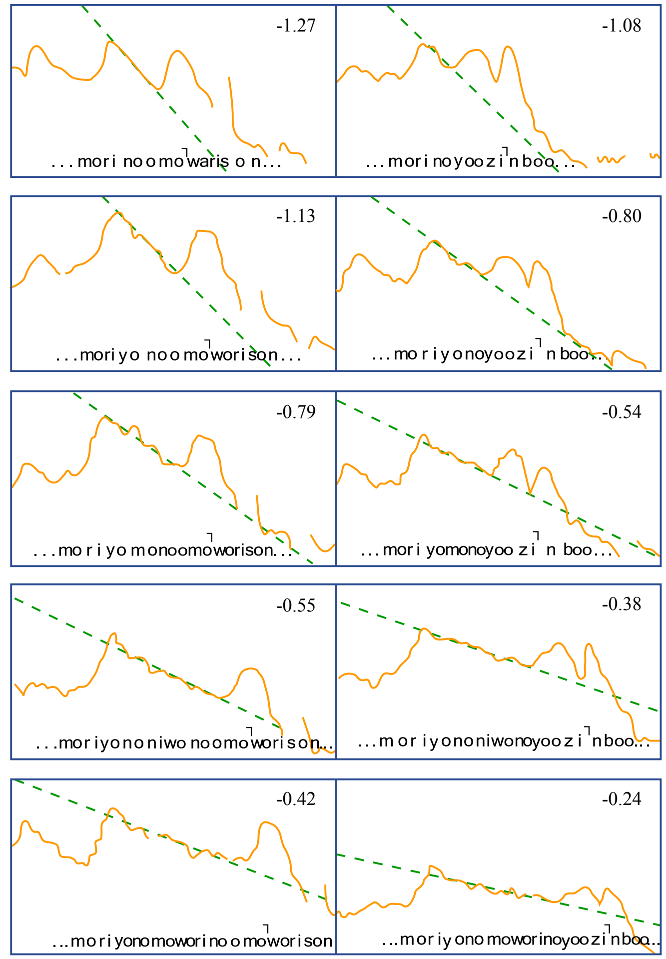
Tonal anticipation in intonation



- Near-linear interpolation from L- to H* across [meɪdðə].

Interpolation in intonation

- Cs and Vs impose weak constraints on F_0 - F_0 is perceptually and articulatorily relatively independent from most segmental contrasts.
  - True ‘wide windows’
- Where tonal specifications are widely spaced, minimum effort transitions are possible.
- Distinctiveness of tonal contrasts may also play a role here - other forms of transition between tones could be confusable with the presence of a pitch accent.
- Other examples of weak constraints?



Unaccented phrases with 3, 4, 5, 7, and 8 morae before accentual-phrase boundaries with two different allophones of L%. The dashed line in each panel is a regression line fitted to all f_0 values between the peak for the phrasal **H** in the first phrase and the minimum for the interphrasal L%. The number in the upper right-hand corner is the slope of the regression line. Here a right corner in the transcription indicates the location of an accent.

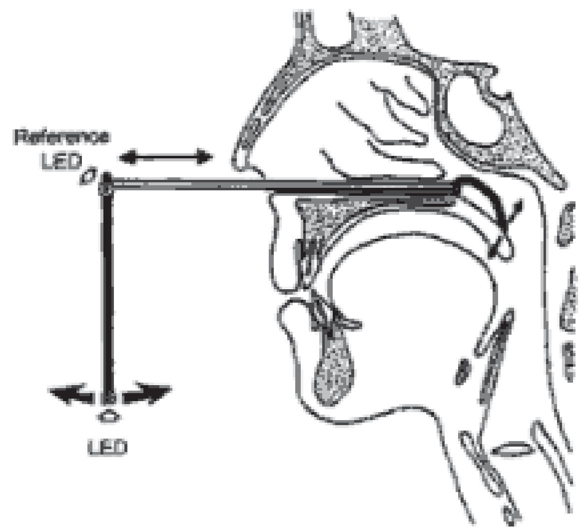
Image by MIT OCW.
Adapted from Pierrehumbert, J., and Mary E. Beckman.
Japanese Tone Structure. Cambridge MA: MIT Press, 1988. ISBN: 9780262161091.

Weighted targets vs. windows

- This line of analysis requires violable target constraints rather than windows:
- [k] can be produced with substantial lip-protrusion in [ku].
  - Windows: [k] has a wide window for lip protrusion.
  - Weighted constraints: [k] constraint on lip protrusion are weak compared to [u] constraint.
- Tendency to minimize anticipatory lip protrusion even with long C clusters (-ikt#ku-) is attributed to dispreference for lip-rounding during first [k].
  - Windows: incompatible with wide window motivated above.
  - Weighted constraints: [k] has a target that requires unrounded lips, but lip-rounding results in low-cost deviation from this target.

Coarticulation as perceptual cue

- It is unlikely that all coarticulation is motivated by articulatory limits and effort minimization.
- E.g. nasal coarticulation in English
- Krakow (1989) : Velum movement tracked via velotrace + Selspot.



© Cencage Learning. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/.
Source: Baken, Ronald J., and Robert F. Orlikoff. Clinical measurement of speech and voice. Cengage Learning, 2000.

from 'Clinical Measurement of Speech and Voice', Robert F. Orlikoff

Coarticulation as perceptual cue

- In word-final [m], velum reaches plateau ~200ms before labial closure (vertical line).
- Anticipatory nasalization serves as a cue to nasal.
- Should still be constrained by conflicting targets.

Figure removed due to copyright restrictions.
Source: Figure 1, Krakow, Rena A. "Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate." Haskins Laboratories Status Report on Speech Research SR-117/118 (1994): 31-48.

Summary

- Effects to be accounted for:
  - target variation/target undershoot
  - coarticulatory effects between non-adjacent segments
  - partial look-ahead in some contexts
- Mechanisms:
  - violable targets: allow for undershoot.
  - effort constraints: motivate undershoot, look-ahead.
  - gestural coordination constraints?

MIT OpenCourseWare
https://ocw.mit.edu

24.915 / 24.963 Linguistic Phonetics
Fall 2015

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.