24.963 Linguistic Phonetics
Fall 2005

24.963

Linguistic Phonetics

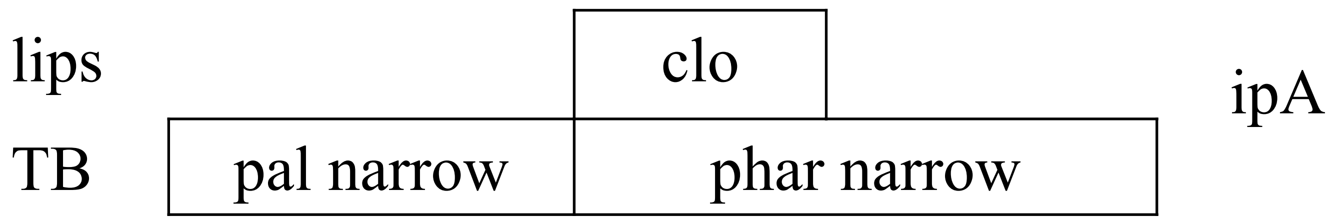
Speech Production II

Reading for week 13: Ladefoged and Broadbent.

Assignment: Write up experiment 2 - Voicing effects on formants.

Modeling V-to-V coarticulation: Articulatory Phonology

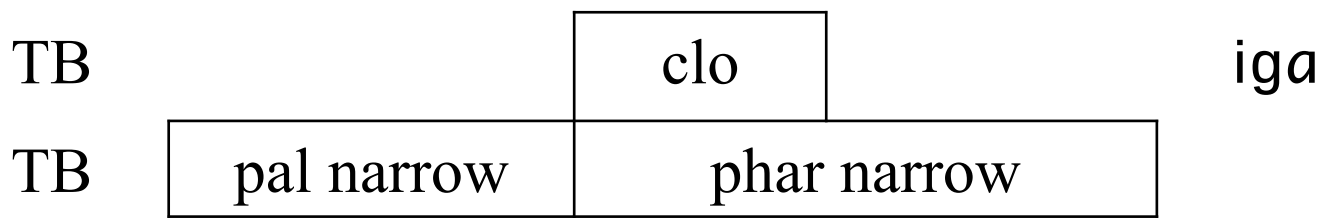
- overlap:



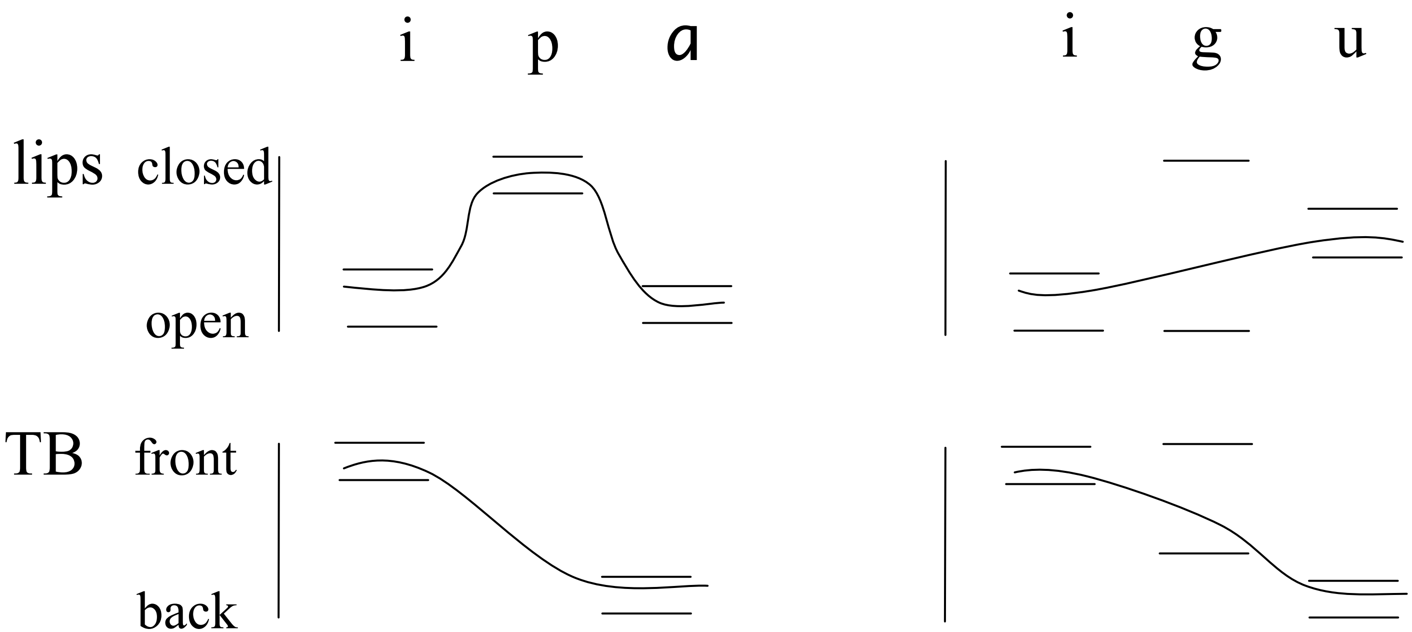
- movement to V2 starts at the same time as movement to C - anticipatory coarticulation.
  – may not be sufficient: V2 can affect the mid-point of V1. Magen (1997) found some effects of V3 at onset of V1 in bV_1bəbV_3b words.
- movement to V2 starts from position of V1 - carryover coarticulation.

Modeling V-to-V coarticulation: Articulatory Phonology

- blending:



Modeling V-to-V coarticulation: Windows model



- Non-specification or wide windows during C allows interpolation between Vs on some dimensions.

Distinguishing the models: Duration of anticipatory coarticulation

- Gestural/coproduction models tend to predict fixed duration of anticipatory coarticulation relative to onset of segment, whereas targets-and-interpolation models tend to predict interpolation across any number of unspecified segments.
- Studies have yielded mixed results (e.g. above).
- Perkell and Matthies (1992) argue for a ‘hybrid’ model on the basis of data on anticpatory labial coarticulation in English.

Distinguishing the models: Duration of anticipatory coarticulation

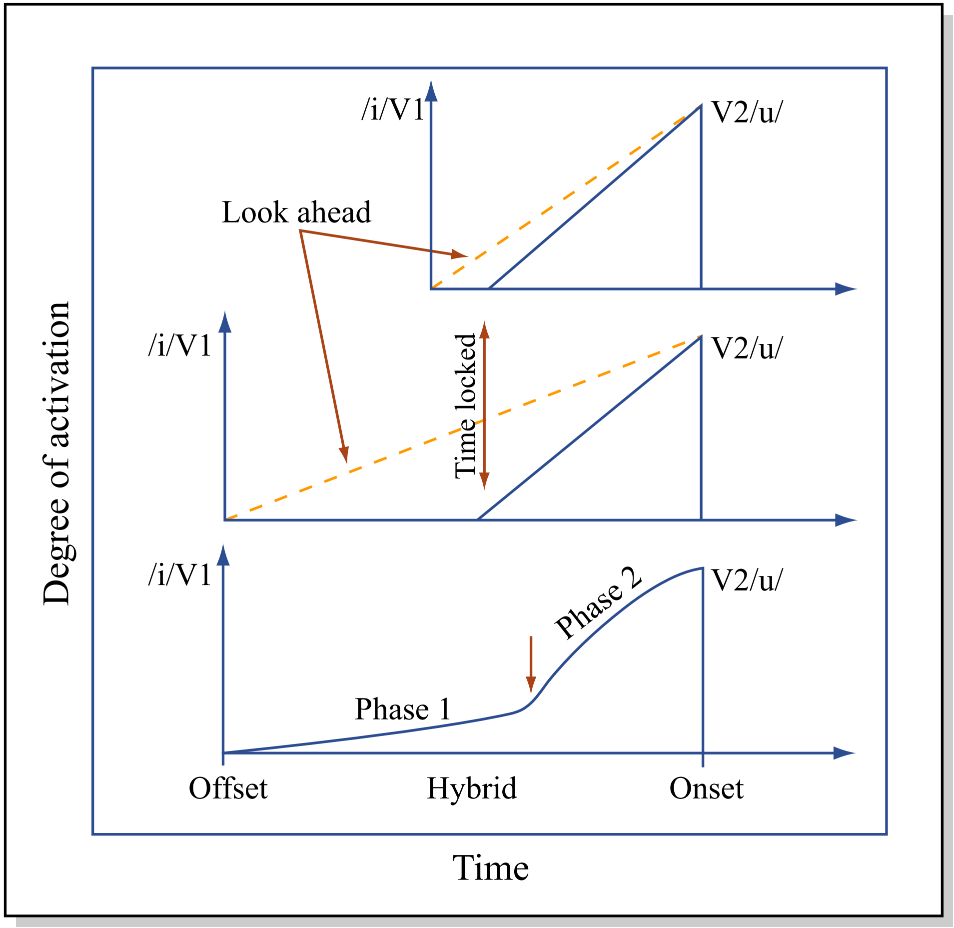


Image by MIT OpenCourseWare. Adapted from Perkell, J. S., and L. M. Matthies. "Temporal Measures of Anticipatory Labial Coarticulation for the Vowel /u/: Within- and Cross-subject Variability." *The Journal of the Acoustical Society of America* 91, no. 5 (1992): 2911-2925.

Distinguishing the models: Duration of anticipatory coarticulation

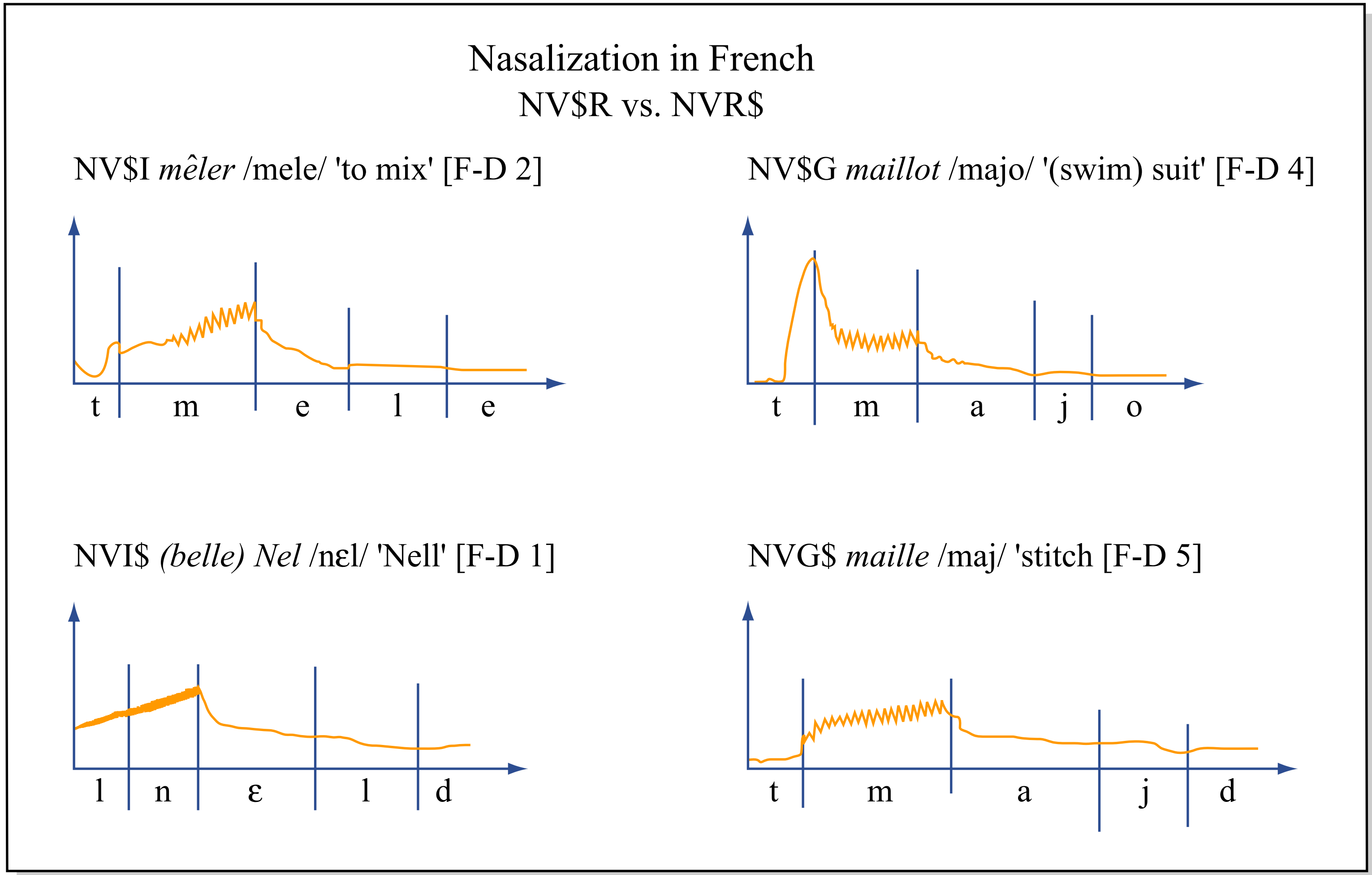
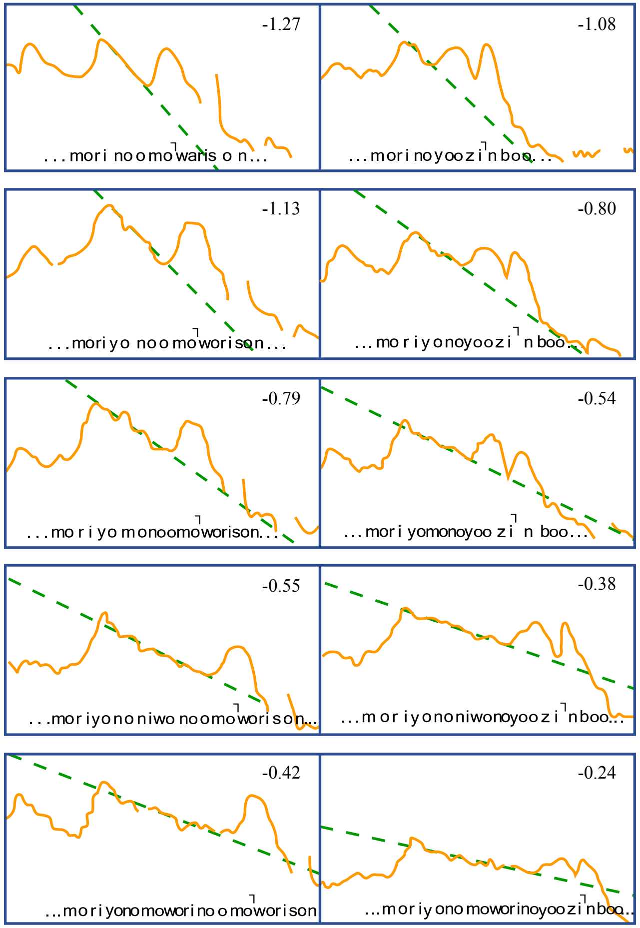


Image by MIT OpenCourseWare. Adapted from Cohn, Abigail. "Phonetic and Phonological Rules of Nasalization." Ph.D. dissertation, University of California, Los Angeles, 1990.

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



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.

Timing and coordination

- Two questions:
  - How is the coordination and timing of articulatory movements achieved on-line?
  - What kinds of timing relations do we observe in languages?

Timing and coordination

- In Articulatory Phonology, coordination is specified in terms of the cycle of an abstract undamped spring-mass system with the same stiffness as the actual critically damped gesture.
- The onset of a gesture is 0 , the target is taken to be achieved at 240 , and the release at 290 .
- In Browman and Goldstein (1990, 1995), coordination is assumed to be achieved by rules specifying simultaneity of particular points in the cycles of two gestures.
- It is assumed that only 3 points in a gesture are referred to in phasing rules: 0 , 240 , 290 .
    - e.g. in -C_1C_2- cluster 0 in C_2 is aligned to 240 in C_1.
- So timing is specified in terms of coordination of landmarks internal to gestures, not via specified durations and an external clock.

Timing and coordination

- In early work on Articulatory Phonology/ Task Dynamics, it was suggested that phasing of gestures was invariant under variation in rate and stress (e.g. Tuller and Kelso 1984).
- Subsequent work shows that coordination is in fact highly variable depending on the nature of the segments involved, segmental context, stress, prosodic context, speech rate, speaker etc (e.g. Nittrouer et al 1988, Byrd 1996).
- One promising line of investigation analyzes articulatory timing in terms of its acoustic consequences rather than focusing exclusively on the articulations.

Chitoran et al (2002): ‘Gestural overlap and recoverability’

- Studied stop-stop clusters in Georgian (S. Caucasian).
- Tested two hypotheses derived from the idea that gestures are coordinated so as to ensure the availability of acoustic cues to the presence and nature of the gestures:
  - Word-initial stop-stop sequences will be less overlapped than like word-internal sequences.
  - Stop-stop sequences with a back-to-front order of constriction location (coronal-labial, dorsal-labial, dorsal-coronal) will evidence less gestural overlap than stop-stop sequences with a front-to-back order.
- Two speakers.

Chitoran et al (2002)

- Materials (frame: sit'q'va__gamoit^{h}k^{h}mis ordʒer '*word _ is pronounced twice*').

Consonants C1	C2	Word-initial Sequences		Word-internal Sequences	
Front-to-back					
b	g	bgera	'sound'	abga	'saddle bag'
p^h	t^h	p^{h}t^{h}ila	'hair lock'	ap^{h}t^{h}ar-i	'hyena'
d	g	dg-eb-a	's/he stands up'	a-dg-eb-a	's/he will stand up'
Back-to-front					
g	b	g-ber-av-s	's/he is inflating you'	da-gbera	'to say the sounds'
t^h	b	t^{h}b-eb-a	'it is warming up'	ga-t^{h}b-a	'it has become warm'
g	d	gd-eb-a	'to be thrown'	a-gd-eb-a	'to throw smth. in the air'

Image by MIT OpenCourseWare. Adapted from Chitoran, I., L. Goldstein, and D. Byrd. "Gestural Overlap and Recoverability: Articulatory Evidence from Georgian." In *Laboratory Phonology 7*. Edited by C. Gussenhoven and N. Warner. New York, NY: Mouton de Gruyter, 2002, pp. 419-447.

Chitoran et al (2002)

- Collected movement data using EMMA (ElectroMagnetic Mid-saggital Articulometer) with pellets near tongue tip, back of tongue body, and an intermediate point.
- Quantified overlap in terms of the 'percentage of the interval between target achievement and release for first stop at which movement onset for the second stop is initiated'

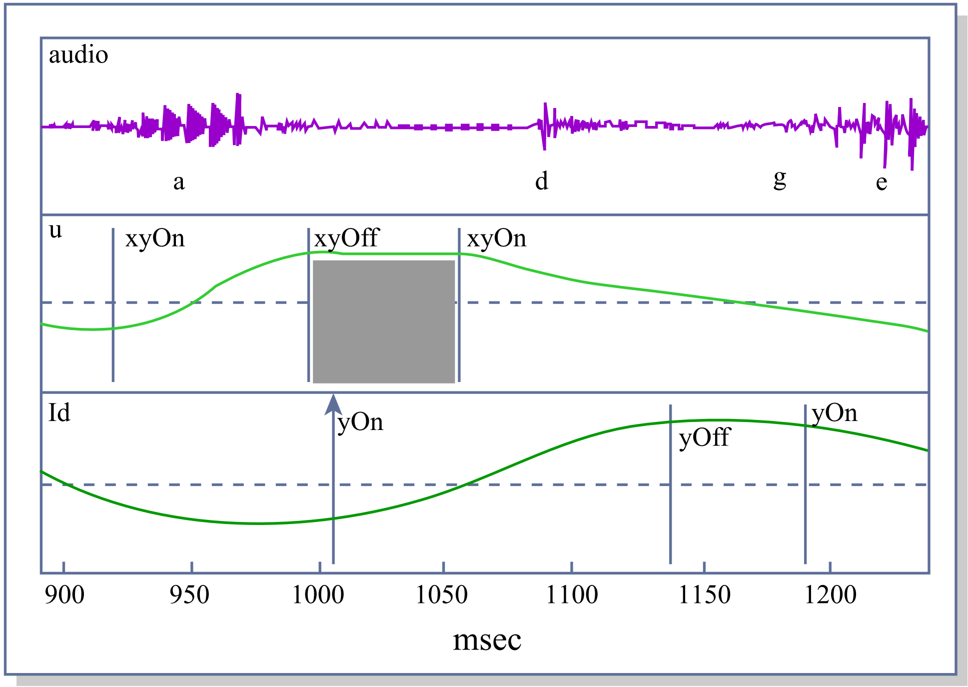


Image by MIT OpenCourseWare. Adapted from Chitoran, I., L. Goldstein, and D. Byrd. "Gestural Overlap and Recoverability: Articulatory Evidence from Georgian." In *Laboratory Phonology 7*. Edited by C. Gussenhoven and N. Warner. New York, NY: Mouton de Gruyter, 2002, pp. 419-447.

Chitoran et al (2002)

- More overlap medially than initially.
- More overlap in front-to-back clusters in initial position (both speakers).
- Also medially for speaker 1.
- Labial+coronal [$p^h t^h$] are least overlapped. Related to aspiration?

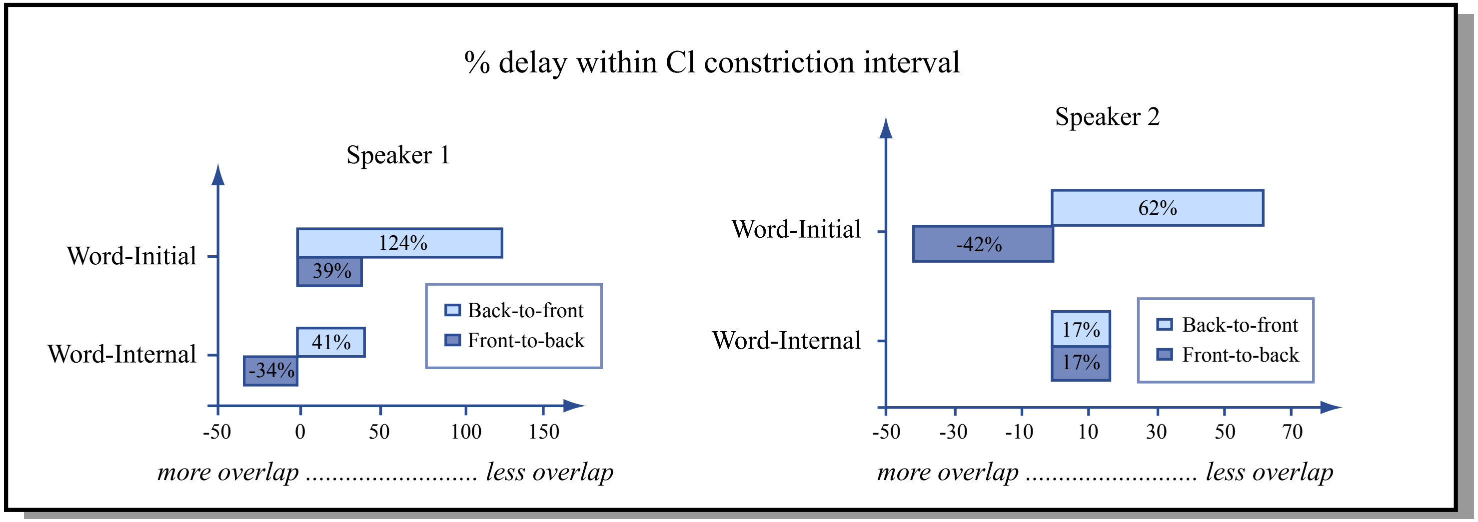


Image by MIT OpenCourseWare. Adapted from Chitoran, I., L. Goldstein, and D. Byrd. "Gestural Overlap and Recoverability: Articulatory Evidence from Georgian." In *Laboratory Phonology 7*. Edited by C. Gussenhoven and N. Warner. New York, NY: Mouton de Gruyter. 2002, pp. 419-447.

Chitoran et al (2002)

- Analyze positional variation in overlap in terms of two potentially conflicting constraints:
  - Maximize rate of transmission of information (favors overlap).
  - Recoverability: Maximize cues to the presence and identity of gestures (cf. maximize the distinctiveness of contrasts).

Patterns of nasal coarticulation in French (Cohn 1990)

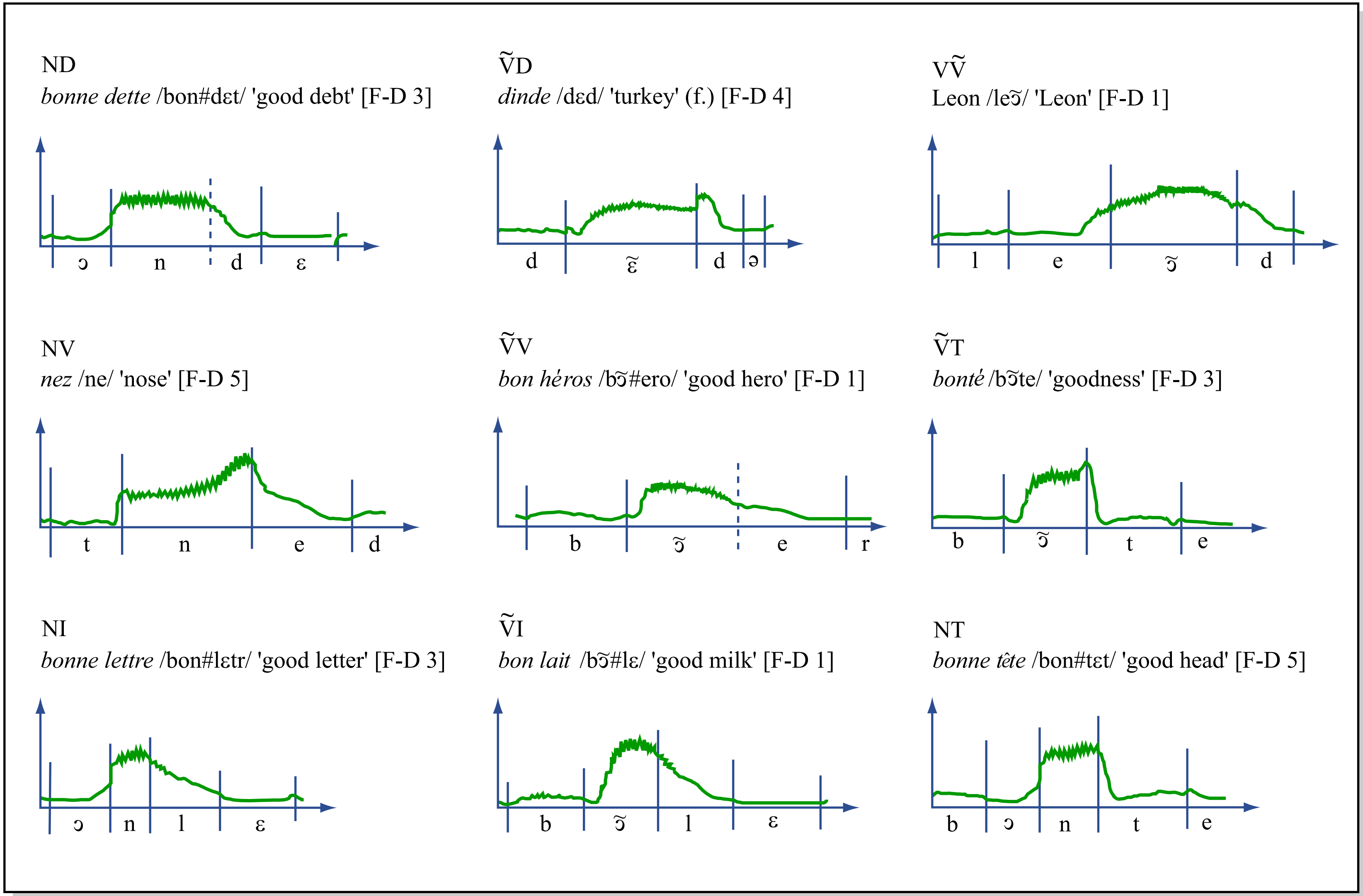


Image by MIT OpenCourseWare. Adapted from Cohn, A. "Phonetic and Phonological Rules of Nasalization." *UCLA Working Papers in Phonetics* 76 (1990). Ph.D. dissertation, University of California, Los Angeles.

Auditory/acoustic targets in speech production

- It has been proposed that the primary goals in speecg production are defined in acoustic (or auditory) terms.
- E.g. Ladefoged et al (1972), Johnson et al (1993), Guenther (1998).
- Motor plan is constructed so as to realize acoustic targets (and satisfy motor efficiency constraints).
- On-line or off-line?

Auditory/acoustic targets in speech production

- Might also help to explain complexities of articulatory adjustments with changes in speech rate. E.g. Tsou (Wright 1996):
- Tsou allows a wide variety of initial and medial CC clusters, including stop-stop clusters [tpihi, tʃ^tʃ^].
- C1 is always audibly released in initial stop-stop clusters, but only in 2/3 of medial clusters.
- In initial position, release is likely to provide the only audible cues to place of C1, medially closure transition is available.
- Nasal C1 in nasal-stop clusters are released less consistently (35% initially, 17% medially).
- Initial stop-stop clusters show little decrease in duration with increased speech rate (-> little increase in gestural overlap?). Nasal-stop and fricative-stop initial clusters show substantial shortening of C1 with increased speech rate.

Reference frames in speech production

- Guenther (2002, etc) postulates multiple reference frames in the planning of speech:
  - Muscle length reference frame: lengths and shortening velocities of muscles.
  - Articulator reference frame: ‘coordinates roughly correspond to the primary movement degrees of freedom of the speech articulators’
  - Tactile reference frame: ‘states of pressure receptors...on the surfaces of the speech articulators’
  - Constriction reference frame: ‘locations of key constrictions in the vocal tract’ cf. tract variables in articulatory phonology.
  - Acoustic reference frame: ‘e.g. formant frequencies’
  - Auditory/perceptual reference frame: representation in auditory cortical areas.
- Goals are formulated primarily at the auditory/perceptual level (closest to the listener), and then translated into other reference frames.
- Could we place goals even closer to the listener, e.g. intended word percept?

Listener-oriented behavior

- Do speakers take the needs of the listener into account in speech production?
- The speak more loudly in the presence of background noise (Lombard 1911).
- Some evidence that they speak more clearly where context provides less assistance in the interpretation of an utterance: Lieberman (1963), Hunnicutt (1985) found intelligibility differences between words excised from high and low redundancy contexts.

Wright (2004), Munson (2004)

- Evidence that speakers selectively deploy hyperarticulation where it is needed more:
  - ‘Hard’ words: low frequency, many high frequency neighbors.
  - ‘Easy’ words: high frequency, few and low frequency neighbors.
- In a reading task, the vowels of hard words are more dispersed in F1-F2 space than the vowels of easy words.
- Clearer speech where there is more likely to be perceptual difficulty.

Wright (2004), Munson (2004)

- Could this be a speaker-related effect?
- Hypothesis (e.g. Jurafsky et al 2003):
  - Lexical access in production is more difficult for hard words.
  - Speakers tend to hyperarticulate words when they have difficulty with lexical access.
- Munson (2004): extended Wright's experiments:
  - Crossed high and low neighborhood density, high and low frequency (four categories of words).
  - Two conditions:
    1. Subjects read words immediately after presentation.
    2. Subjects were forced to wait 1000ms before producing words.
  - Difficulties with lexical access should not affect production with a 1000 ms delay.

Munson (2004)

- Results:
  - Vowels in words from dense neighborhoods were more dispersed in both conditions.
  - In the undelayed condition, vowels in lower frequency words were more dispersed.
  - In the delayed condition the effect of frequency disappeared.
- Frequency effect may represent speaker difficulty, but the hyperarticulation of words from dense neighborhoods is for the benefit of the listener.
- Speakers selectively deploy clear speech where it is needed more

Prosody and speech production

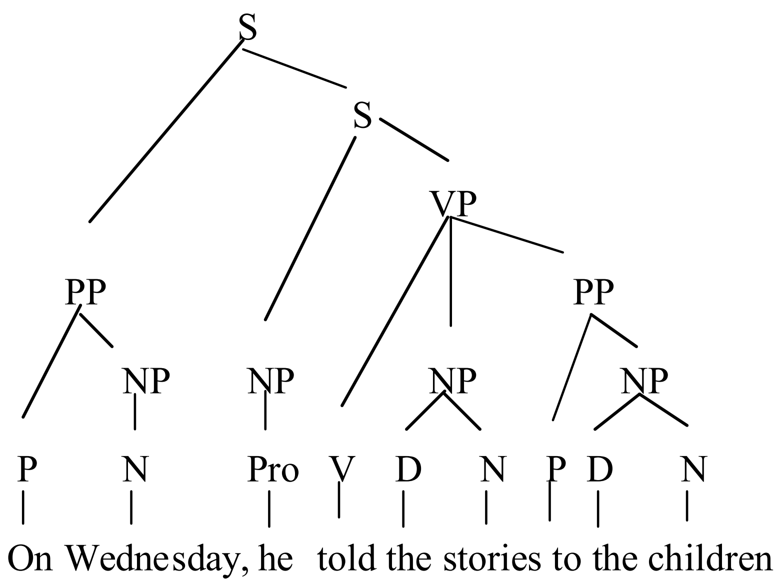
- The realization of a segment depends on its prosodic position as well as on its segmental context.
- Aspects of prosody:
  - Stress, accent
  - Phrasing

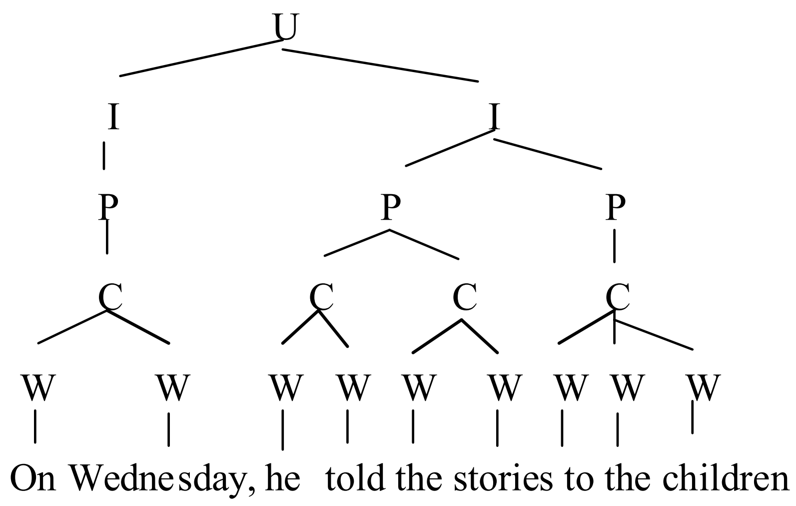
Prosody and speech production

Central ideas of prosodic phonology (Selkirk 1980, 1984, Nespor and Vogel 1982, 1986):

- ‘utterances are PHRASED, in the same sense that musical passages are phrased’ (Hayes 1989:201).
- Prosodic phrasing is hierarchical.
- Prosodic phrasing is related to, but distinct from, syntactic phrasing.

Prosodic structure





Phonetic correlates of phrasing

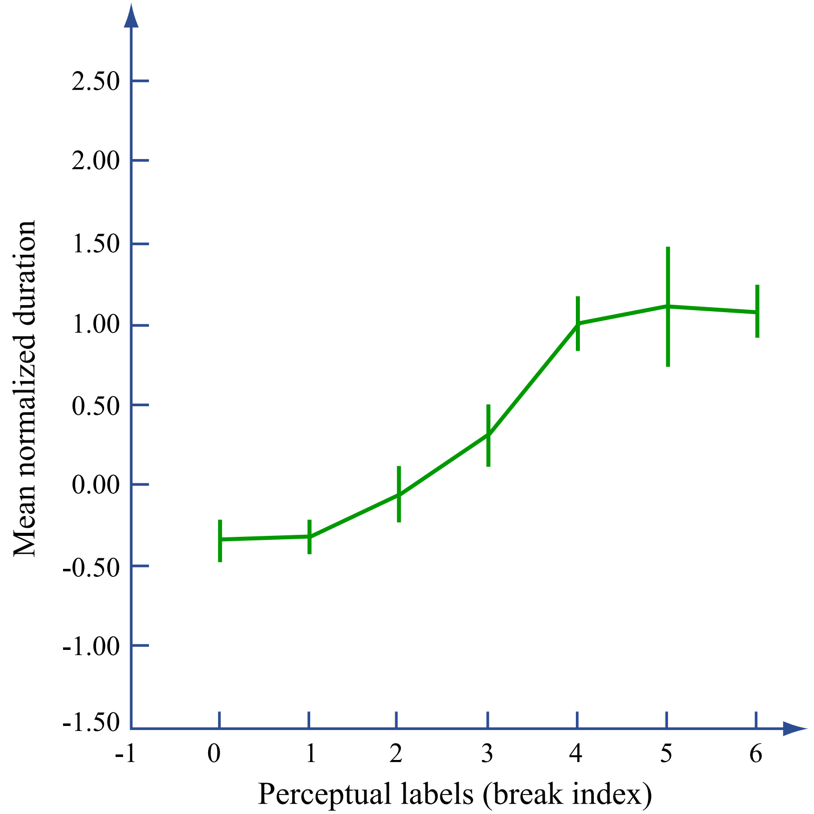
- In many languages, segments are lengthened before constituent boundaries.
- Pauses typically occur at intonational phrase boundaries.

Wightman et al (1992)

- Studied a corpus of sentences read by professional news announcers.
- Prosodic boundaries marked using a 7-point scale of break indices
    - 0 No prosodic break (cliticization)
    - 1 Prosodic word boundary
    - 2 ‘accentual phrase’
    - 3 Intermediate phrase
    - 4 Intonation phrase
    - 5 Superior major tone group
    - 6 Sentence boundary
- Lengthening measured in terms of normalized segment durations (standard deviations from the mean for that segment, adjusted for estimated speech rate of each sentence).

Phonetic correlates of phrasing

- Lengthening before constituent boundaries, localized to the final syllable rhyme (VC).
- Degree of final lengthening increases with size of prosodic constituent.



Mean normalized duration of the word-final vowel as a function of the break index following the word. The vertical bars correspond to confidence intervals (95% protection level).

Image by MIT OpenCourseWare. Adapted from Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. "Segmental Durations in the Vicinity of Prosodic Phrase Boundaries." *Journal of the Acoustical Society of America* 91 (1992): 1707-1717.

Phonetic correlates of phrasing

- Initial strengthening (Keating et al 2003)
  - Domain-initial consonants show greater tongue-palate contact and longer duration than domain medial consonants. Progressive increase with constituent size (English, French, Korean).

Image removed due to copyright restrictions.
See Fougeron, C., and P. A. Keating. "Articulatory Strengthening at Edges of Prosodic Domains." *Journal of the Acoustical Society America* 101 (1997): 3728-3740.

Phonetic correlates of phrasing

- Initial strengthening (Keating et al 2003)
  - Domain-initial consonants show greater tongue-palate contact and longer duration than domain medial consonants. Progressive increase with constituent size (English, French, Korean).

Graph removed due to copyright restrictions.

Please see Figure X.2 in Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu. "Domain-initial articulatory strengthening in four languages." In *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Edited J. Local, R. Ogden, and R. Temple. New York, NY: Cambridge University Press, 2003, pp. 143-161.

Phonetic correlates of phrasing

- Initial strengthening: VOT in Korean (Jun 1993).

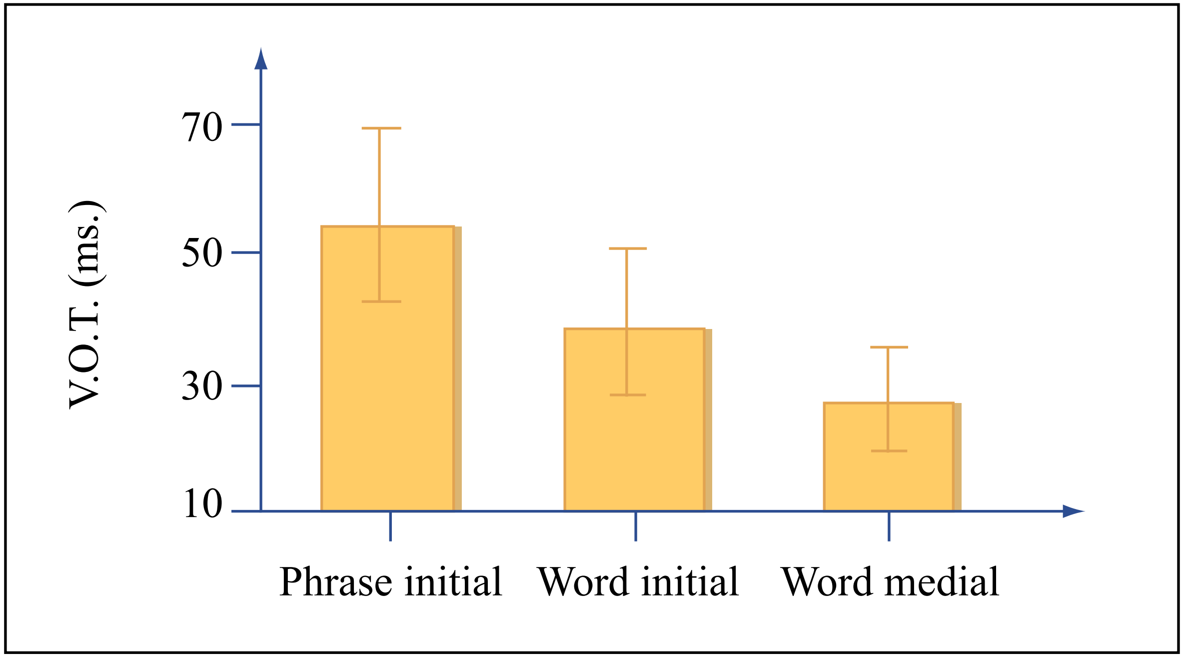


Image by MIT OpenCourseWare. Adapted from Jun, Sun-Ah. "The Phonetics and Phonology of Korean Prosody." Ph.D. dissertation, OhioState University, 1993. Published as *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure*. New York, NY: Garland, 1996.