

**Production and perception of prosodically varying
inter-gestural timing in American English laterals.**

by
Susan Sychi Lin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2011

Doctoral Committee:

Professor Patrice Speeter Beddor, Chair
Professor J. Brian Fowlkes
Associate Professor Robin M. Queen
Assistant Professor Andries W. Coetzee

© Susan Sychi Lin 2011
All Rights Reserved

For my father, who loves science,
and for my mother, who loves words.

ACKNOWLEDGEMENTS

Thank you, Pam Beddor, my amazing advisor. This dissertation is as much my work as it is yours, and I don't just mean because of the typos and nonsense you've helped ferret out. You have an unparalleled ability to find just the right words to calm me during my many crises of various magnitudes. Your support and encouragement have been instrumental to my development as a scholar. I can't believe how far I've come in the last few years, and I owe you a debt that I worry can't be repaid.

Thanks, too, to my exceptional committee members, Andries Coetzee, Brian Fowlkes and Robin Queen, for your valuable contributions to this dissertation, and also for being exceptionally accommodating during the defense scheduling process. I am also very grateful for the support I received from many other faculty here at the University of Michigan, especially to Debby Keller-Cohen and Sally Thomason, for kind words during hard times.

There have been many exceptional scholars whom I have shared space with during my graduate school career. But I owe a very special thanks to those I have shared lab space with during the last three years, especially to Kevin McGowan. I hope you agree with me that we've grown tremendously together, as scholars and as human beings. I hope you will eventually forgive me for distracting you from your work on a daily basis. Also, to Anna Babel and Damon Tutunjian, my small but delightful cohort: finally, we're all finished!

To Jordan, thank you for everything you've done to help, especially over the last four years, and most especially during the last three months. I would be much poorer without

you, reminding me to eat and bathe, listening to me complain, and giving my life purpose beyond my work.

And finally, to my parents, thank you for your unwavering support throughout the years. I will never forget how much you've done to give me an edge, from the after-school math lessons to the long weekly drives to Chinese school, and piano lessons, clarinet lessons, marching band practice, and all of the other things I once thought were ridiculous and unnecessary. Now I know better. This is for you.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	x
LIST OF APPENDICES	xiii
 CHAPTER	
I. Introduction	1
1.1 Syllabic and Prosodic Effects on Articulator Movement	2
1.1.1 Syllable Effects	2
1.1.2 Prosodic Effects	6
1.2 Models & Predictions	7
1.2.1 Perceptual motivations	8
1.2.2 Speaker-based models	10
1.3 Summary	15
 II. Experiment 1: Production	 17
2.1 Design	17
2.2 Equipment and Techniques	18
2.2.1 Equipment	18
2.2.2 Analysis Techniques	22
2.2.3 Measuring Tongue Movement	25
2.3 Results and Discussion	29
2.3.1 Speaker Exclusion	29
2.3.2 Statistical Procedures	30
2.3.3 Results - Overall Lag	31
2.3.4 Results - Coda /l/	31
2.3.5 Results - Initial /l/	38
2.3.6 Discussion	44
 III. Experiment 2: Perception	 49
3.1 Design	50
3.1.1 Stimuli	50
3.1.2 Task	60

3.1.3	Participants	63
3.2	Analysis & Results	63
3.2.1	Measurements	64
3.2.2	Results	68
3.3	Conclusions	74
IV.	Summary and Conclusion	76
4.1	Perception	76
4.2	Production	77
4.3	Overall Strength of the Hypotheses	82
4.3.1	Perceptual & jaw-cycle hypotheses	83
4.3.2	π -gesture	84
4.4	Conclusion	86
APPENDICES		87
BIBLIOGRAPHY		132

LIST OF FIGURES

Figure

1.1	Schematization of the effect of jaw opening on extent of articulatory movement required for target achievement.	10
1.2	Example gestural score, for production of “palm,” pronounced [p ^h am], adapted from Browman and Goldstein (1992)	12
1.3	Schematizations of in-phase (top), anti-phase (middle) and other unstable (bottom) relationship between two coupled oscillators.	14
2.1	Equipment used in Experiment 1, inside (right) and outside (left) the sound-attenuated booth at the University of Michigan.	19
2.2	Ultrasound Stabilization Helmet, in use.	21
2.3	A series of ultrasound frames during production of /k/ from “ok,” as produced by Speaker S04, showing movement of the tongue back towards closure (frames 354 and 355), during closure (frame 356), and after release (frames 357 to 359). Tongue tip is located to the right, and tongue root to the left.	23
2.4	Locating the acoustic correlate of release in the velar stop in “ok;” Sample taken from speaker S04, the same /k/ closure for which ultrasound images are shown in Figure 2.3 . . .	23
2.5	Locating the boundaries of phones in the target syllables in “...(Cam)mie. Lap...”; Sample taken from speaker S04.	24
2.6	Example of poor EdgeTrak output (left) and corresponding hand-correction (right); Sample taken from speaker S04.	25
2.7	Example of aggregate tongue contours from all frames for a single /CVIVC/ production (solid lines) “regions of interest” (solid segments) along the palate trace (dashed line) for tongue dorsum target (left), tongue tip target (right), for speaker S04.	26
2.8	Tongue dorsum (top) and tongue tip (bottom) aperture, measured in pixels, against time during production of “...male. Em(barrassing)...” by speaker S13. Vertical lines delineate acoustic boundaries of the segments, with the acoustic duration of /l/ shaded in gray. SIL denotes silence. The first frame measured (at 0 ms along the x-axis) is five frames prior to the acoustic onset of C ₁ in the target C ₁ V ₁ IV ₂ C ₂ sequence.	27

2.9	Gestural lag in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.	34
2.10	Lateral duration in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.	35
2.11	Duration of tongue dorsum gesture in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, significance at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.	36
2.12	Duration of tongue tip gesture in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.	37
2.13	Gestural lag in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.	40
2.14	Lateral duration in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.	41
2.15	Duration of tongue dorsum gesture in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.	42
2.16	Duration of tongue tip gesture in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.	43
2.17	Schematization of results from Experiment 1.	47
3.1	Gestural score and generated audio for “lap” with -60 ms gestural lag. See text for details.	52
3.2	Gestural score and generated audio for “lap” with 0 ms gestural lag. See text for details.	53
3.3	Gestural score and generated audio for “lap” with 80 ms gestural lag. See text for details.	54
3.4	F1, F2, and F3 plotted against time for target stimuli.	58
3.5	First three formants taken at point of tongue dorsum achievement from selected tokens produced by speakers during Experiment 1.	61
3.6	Histogram of reaction times from Experiment 2, from all participants. Bins are 50 ms in size.	65
3.7	D' by gestural lag, all speakers, and quadratic best fit.	69
3.8	Reaction time (normalized) by gestural lag, all speakers, and quadratic best fit.	70
3.9	D' by gestural lag for high-accuracy listeners and mid-accuracy listeners.	72

3.10	Normalized reaction time by gestural lag for high-accuracy listeners (top) and mid-accuracy listeners (bottom).	73
4.1	Schematization of the hypotheses.	78
4.2	Schematization of the results from Experiment 1.	81
A.1	Duration of coda lateral production (ms) by prosodic context, for speakers S01-S09	91
A.2	Duration of gestural lag (ms) in coda /l/s by prosodic context, for speakers S01-S09	93
A.3	Duration of tongue tip gesture (ms) in coda /l/s by prosodic context, for speakers S01-S09 .	95
A.4	Duration of tongue dorsum gesture (ms) in coda /l/s by prosodic context, for speakers S01-S09	97
A.5	Duration of onset lateral production (ms) by prosodic context, for speakers S01-S09	98
A.6	Duration of gestural lag (ms) in onset /l/s by prosodic context, for speakers S01-S09	99
A.7	Duration of tongue tip gesture (ms) in onset /l/s by prosodic context, for speakers S01-S09	100
A.8	Duration of tongue dorsum gesture (ms) in onset /l/s by prosodic context, for speakers S01-S09	101
C.1	Spectrogram information for “lap” with -60 ms gestural lag.	128
C.2	Spectrogram information for “lap” with -40 ms gestural lag.	128
C.3	Spectrogram information for “lap” with -20 ms gestural lag.	129
C.4	Spectrogram information for “lap” with 0 ms gestural lag.	129
C.5	Spectrogram information for “lap” with +20 ms gestural lag.	130
C.6	Spectrogram information for “lap” with +40 ms gestural lag.	130
C.7	Spectrogram information for “lap” with +60 ms gestural lag.	131
C.8	Spectrogram information for “lap” with +80 ms gestural lag.	131

LIST OF TABLES

Table

1.1	Summary of inter-gestural timing relationship findings	5
2.1	Target words for Experiment 1	17
2.2	Target sentences for Experiment 1	19
2.3	Summary of participants in Experiment 1	30
2.4	Factors and dependent variables in this study.	30
2.5	Coda lateral results; variables affected significantly by prosodic position as predicted are marked with “X”.	44
2.6	Initial laterals: variables with significantly greater values in sentence-initial context than in word-initial context marked by “X,” variables with significantly lower values in sentence-initial context marked by “O.”	46
3.1	Selected text from a gestural score (.G) file, from “lap” with 20 ms gestural lag.	51
3.2	Sample text from a HLsyn (.HL) file, from “lap” with 20 ms gestural lag. This set of variables provides information for timestamp 220 ms.	51
3.3	Stimuli for Experiment 2.	56
3.4	Responses from post-experimental stimulus identification for participant S05.	63
3.5	Sample raw data from listener S02.	66
3.6	Number of false alarms by stimulus and listener out of 120 possible, with overall false alarm rates (%).	67
3.7	Number of correct “I” responses out of 15 possible, by listener and target item.	67
3.8	Average d' scores by speaker.	68
3.9	Mean normalized reaction time (ms), by group and target gestural lag.	75
3.10	Mean d' , by group and target gestural lag.	75

A.1	Mean lateral duration values (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$	90
A.2	Mean gestural lag (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$	92
A.3	Mean tongue tip gesture durations (ms) for coda laterals by speaker, for S01-S09. bolded data are significant are $p < 0.05$	94
A.4	Mean tongue dorsum gesture durations (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$	96
A.5	Mean lateral durations (ms) for onset laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$	97
A.6	Mean gestural lag (ms) for onset laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$	98
A.7	Mean tongue tip gesture durations (ms) for onset laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$	99
A.8	Mean tongue dorsum gesture durations (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$	100
B.1	Measurements and statistics for results for coda laterals from Experiment 1, speaker S01. Bolded data are significant at $p < 0.05$	103
B.2	Measurements and statistics for results for coda laterals from Experiment 1, speaker S02. Bolded data are significant at $p < 0.05$	103
B.3	Measurements and statistics for results for coda laterals from Experiment 1, speaker S04. Bolded data are significant at $p < 0.05$	104
B.4	Measurements and statistics for results for coda laterals from Experiment 1, speaker S06. Bolded data are significant at $p < 0.05$	105
B.5	Measurements and statistics for results for coda laterals from Experiment 1, speaker S07. Bolded data are significant at $p < 0.05$	105
B.6	Measurements and statistics for results for coda laterals from Experiment 1, speaker S08. Bolded data are significant at $p < 0.05$	106
B.7	Measurements and statistics for results for coda laterals from Experiment 1, speaker S09. Bolded data are significant at $p < 0.05$	106
B.8	Measurements and statistics for results for coda laterals from Experiment 1, speaker S10. Bolded data are significant at $p < 0.05$	107
B.9	Measurements and statistics for results from for coda laterals Experiment 1, speaker S12. Bolded data are significant at $p < 0.05$	108
B.10	Measurements and statistics for results for coda laterals from Experiment 1, speaker S13. Bolded data are significant at $p < 0.05$	109

B.11	Measurements and statistics for results for coda laterals from Experiment 1, speaker S16. Bolded data are significant at $p < 0.05$.	110
B.12	Measurements and statistics for results for onset laterals from Experiment 1, speaker S01. Bolded data are significant at $p < 0.05$.	111
B.13	Measurements and statistics for results for onset laterals from Experiment 1, speaker S02. Bolded data are significant at $p < 0.05$.	112
B.14	Measurements and statistics for results for onset laterals from Experiment 1, speaker S04. Bolded data are significant at $p < 0.05$.	112
B.15	Measurements and statistics for results for onset laterals from Experiment 1, speaker S06. Bolded data are significant at $p < 0.05$.	113
B.16	Measurements and statistics for results for onset laterals from Experiment 1, speaker S07. Bolded data are significant at $p < 0.05$.	113
B.17	Measurements and statistics for results for onset laterals from Experiment 1, speaker S08. Bolded data are significant at $p < 0.05$.	114
B.18	Measurements and statistics for results for onset laterals from Experiment 1, speaker S09. Bolded data are significant at $p < 0.05$.	114
B.19	Measurements and statistics for results for onset laterals from Experiment 1, speaker S10. Bolded data are significant at $p < 0.05$.	115
B.20	Measurements and statistics for results for onset laterals from Experiment 1, speaker S12. Bolded data are significant at $p < 0.05$.	116
B.21	Measurements and statistics for results for onset laterals from Experiment 1, speaker S13. Bolded data are significant at $p < 0.05$.	117
B.22	Measurements and statistics for results for onset laterals from Experiment 1, speaker S16. Bolded data are significant at $p < 0.05$.	118
C.1	First four formants, for stimulus with -60 ms gestural lag, taken from the HLSyn (.HL) file.	120
C.2	First four formants, for stimulus with -40 ms gestural lag, taken from the HLSyn (.HL) file.	121
C.3	First four formants, for stimulus with -20 ms gestural lag, taken from the HLSyn (.HL) file.	122
C.4	First four formants, for stimulus with 0 ms gestural lag, taken from the HLSyn (.HL) file.	123
C.5	First four formants, for stimulus with +20 ms gestural lag, taken from the HLSyn (.HL) file.	124
C.6	First four formants, for stimulus with +40 ms gestural lag, taken from the HLSyn (.HL) file.	125
C.7	First four formants, for stimulus with +60 ms gestural lag, taken from the HLSyn (.HL) file.	126
C.8	First four formants, for stimulus with +80 ms gestural lag, taken from the HLSyn (.HL) file.	127

LIST OF APPENDICES

Appendix

A.	Experiment 1, 3 Prosodic Contexts, S01 - S09	88
	A.1 Coda Laterals	89
	A.2 Initial Laterals	89
B.	Experiment 1, 2 Prosodic Contexts, All Speakers	102
	B.1 Final laterals	103
	B.2 Initial Laterals	111
C.	Experiment 2: Synthetic Stimuli	119

CHAPTER I

Introduction

Individual language users are both producers and perceivers of speech. Current theories of speech perception and production differ in the extent to which these capacities are modeled as being intrinsically linked or as being independent of each other. One of the key issues in this ongoing discussion is which components of a speakers' actions during speech production are primarily a consequence of the physical and mechanical constraints imposed by the speech organs, and which components are influenced primarily by speakers' knowledge of how speech is perceived. That is, given a specific action (or set of actions) during speech production, is it more likely to have arisen due to articulatory constraints of the speaker, or for the perceptual needs of the listener? This dissertation contributes to the discussion in two ways. First, it attempts to tease apart the contributions of production and perception to speakers' actions during production of the American English lateral /l/, expanding on previous research (e.g., Browman and Goldstein, 1995; Sproat and Fujimura, 1993) by manipulating the prosodic position of /l/. Second, it attempts to substantiate claims (e.g., Gick et al., 2006) that timing relations between articulators directly affect perceptibility of the resulting acoustics.

1.1 Syllabic and Prosodic Effects on Articulator Movement

It is well established that, in many languages (likely all languages, in fact), speakers show consistent and reliable differences in the production of segments that depend on the position of the segment within the syllable, and within larger prosodic units. These differences appear both in the magnitude and duration of movement of a single articulator and in the relative timing between articulators in multi-articulator segments.

1.1.1 Syllable Effects

Considerable articulatory evidence exists showing that the movement of some articulators is quantifiably “stronger” or more “extreme” in syllable onset position than in coda position. Both Giles and Moll (1975) and Browman and Goldstein (1995) found weaker tongue tip “excursion” (defined as maximum tongue height and percent lingual-palatal contact, respectively) during coda /l/ compared to onset /l/ in English speakers’ productions. Additionally, Browman and Goldstein (1995) found that for the voiceless oral stops /p/, /t/ and /k/, speakers achieved tighter constrictions in initial compared to final stops (defined as higher tongue position for initial /t/s and /k/s, and tighter lip closure in initial /p/s). In contrast, Krakow (1989, 1999) and Vaissière (1989) showed that during production of nasal consonants, speakers tend to produce large velum displacement as well as a long velum plateau in syllable coda nasals, in comparison to syllable onset nasals. This outcome may initially appear to be an exception to the above generalization that articulations are stronger in syllable-initial position. However, Krakow (1999) suggests an alternative interpretation, that tighter constrictions (e.g. oral constriction in stops, tongue tip raising in laterals) are more typical of onset consonants than coda consonants, while more open configurations (e.g. velum lowering in nasal stops) are more typical of coda consonants than onset consonants. Similarly, Sproat and Fujimura (1993) suggested that speech ges-

tures can be classified as either “consonantal,” typified by tight constriction, or “vocalic,” typified by open constriction. The notion of articulatory “strengthening”, then, applies to both onsets and codas, and must be defined by gesture type, so that the tighter “consonantal” gestures are enhanced in onset consonants, while the more open “vocalic” gestures are enhanced in coda consonants. Thus, the oral constriction in an onset nasal may be tighter than that of a coda nasal, while the velum opening gesture in a coda nasal may be more extensive than in a corresponding onset nasal.

In addition to influencing magnitude and duration of articulator motion, syllable position has been shown to have an effect on the temporal relationship between articulators in multi-articulator segments. For instance, as illustrated in row 1 of Table 1.1, Krakow (1989) and Byrd et al. (2009) demonstrated that, in relation to the oral closure, velum lowering occurred substantially earlier in coda nasals than in onset nasals. Moreover, in onset nasals, the end of velum lowering was generally aligned with the end of lower-lip raising, while in coda nasals, the end of velum lowering was aligned with the onset of lower-lip raising. Krakow attributed the general finding in the coarticulation literature of greater anticipatory coarticulation (CVN) than carry-over coarticulation (NVC) in English to this asymmetry.

Sproat and Fujimura (1993), Browman and Goldstein (1995) and Gick (2003) examined syllable effects on English laterals, the production of which involves tongue tip raising and tongue dorsum retraction. As was found for nasals, all three studies found that the two gestures were temporally separated in coda laterals, with the tighter constriction of the two (in the case of laterals, tongue tip raising) occurring further from the syllable nucleus (see rows 2-4 in Table 1.1). However, in the case of onset laterals, while Browman and Goldstein (1995) found that tongue tip and dorsum movements are “roughly synchronous,” paralleling results for nasal consonants, data from Sproat and Fujimura (1993) and Gick

(2003) show that the two gestures in onset laterals were not synchronous, with tongue tip raising preceding tongue dorsum backing. Additionally, Sproat and Fujimura (1993) showed that the degree of gestural lag (the temporal distance between two gestures) in coda laterals was correlated with the duration of the containing rime. That is, the longer the rime containing /l/, the greater the temporal distance between tongue dorsum movement and tongue tip movement peaks. They further suggested that consonantal gestures are “attracted” to syllable boundaries while vocalic gestures are “attracted” to syllable nuclei.

In addition to laterals, Gick (2003) examined inter-articulator coordination in the labio-velar approximant /w/ syllable-initially and -finally, and found that achievement of lip constriction in onset /w/ occurred on average 40-50 ms before achievement of tongue dorsum retraction. He noted that this result is more consistent with an analysis in which the labial constriction is classified as consonantal and the dorsum retraction gesture is classified as vocalic, despite the fact that both are relatively open constrictions. Indeed, Gick (2003) also showed that the labial gesture in coda /w/ was reduced compared to onset /w/s, a behavior consistent with consonantal gestures in other segments. Gick (2003) thus suggested that constriction degree may not be the sole distinction between consonantal and vocalic gestures, going so far as to suggest that these two gesture categories are “language-specific, phonologically specified categories.” Interestingly, two of the three speakers recorded by Gick (2003) produced coda /w/ with minimal asynchrony, such that lip constriction and tongue dorsum retraction were generally achieved nearly simultaneously. This outcome (illustrated in row 5 of Table 1.1) differs substantially from the results discussed above for nasals and laterals, in which greater synchrony was found in syllable-initial than in syllable-final position, but why this should be the case was not explored.

Table 1.1: Summary of inter-gestural timing relationship findings

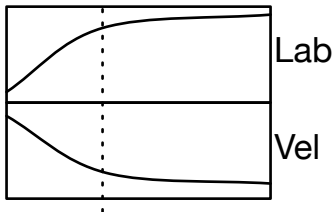
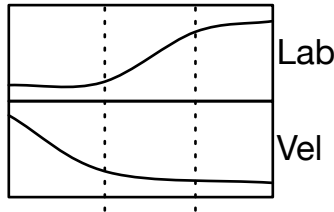
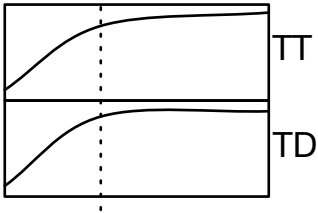
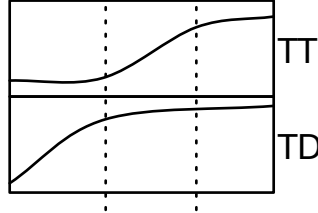
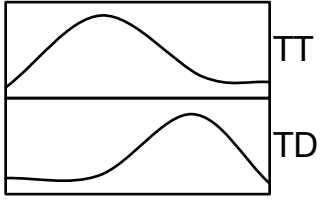
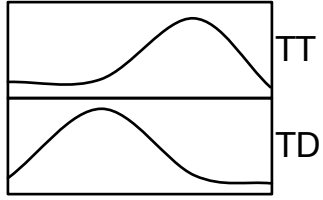
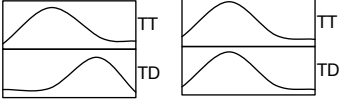
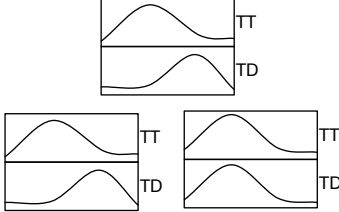
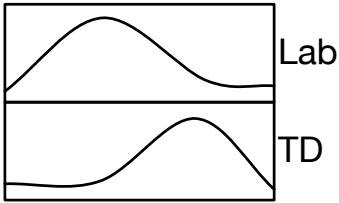
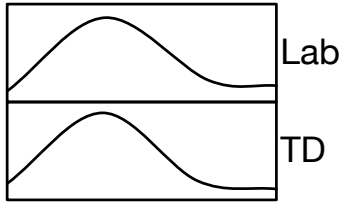
	Author / Method	Syllable-Initial	Syllable-Final
nasal stops (m, n)	Krakow (1989) – Velotrace (velum movement) and Selspot (lower lip movement); Byrd et al. (2009) – MRI	<p>Velum and labial movement roughly co-occur; the end of velum lowering aligns with the end of bottom lip raising</p> 	<p>Velum movement precedes labial; the end of velum lowering aligns with beginning of bottom lip raising</p> 
laterals (l)	Browman and Goldstein (1995) – X-ray microbeam pellet	<p>Tongue dorsum and tongue tip movement roughly co-occur; the end of dorsum retraction aligns with the end of tip raising</p> 	<p>Tongue dorsum movement precedes tongue tip; the end of dorsum retraction aligns with the beginning of tip raising</p> 

Table 1.1 continued on next page...

... Table 1.1 continued from previous page

	Author / Method	Syllable-Initial	Syllable-Final
laterals (l)	Sproat and Fujimura (1993) – X-ray microbeam pellet	Tongue tip movement precedes tongue dorsum 	Tongue dorsum movement precedes tongue tip; amount of “tip delay” increases with duration of rime increases 
laterals (l)	Gick (2003), Gick et al. (2006) – ultrasound imaging	Tongue tip movement precedes tongue dorsum (W. Canadian and American English; left), or gestures co-occur (Serbo-Croatian, Squamish Salish; right) 	Tongue dorsum movement precedes tongue tip generally (bottom left); reversed in Korean (top), and co-occurring in Serbo-Croatian (bottom right) 
labio-velar glide (w)	Gick (2003)	Lip constriction movement precedes tongue dorsum movement 	Lip constriction and dorsum retraction co-occur 

1.1.2 Prosodic Effects

In addition to identifying the effects of syllable position on articulators, researchers have shown that the increase in articulatory strength in syllable-initial position discussed in §1.1.1 interacts with prosodic organizational units. Specifically, onset consonants in strong prosodic positions (e.g. utterance onsets, intonational phrase onsets) are generally produced with stronger gestures than onset consonants in weaker prosodic positions (e.g.

phrase-medial word-initial). In particular, the strengthened articulations tend to be longer, and they tend to reach a more extreme position. This tendency, termed domain-initial strengthening by Keating (2006), has been found not only for the tongue tip gesture during production of the nasal consonant /n/ in American English speakers (Fougeron and Keating, 1997), but also for segments spoken by speakers of several other languages, including French and Korean (Keating et al., 2003; Cho and Keating, 2001; Fougeron, 2001).

This prosodic effect on duration holds only for gesture duration; it does not necessarily hold for acoustic duration. Instead, as reported by Fougeron and Keating (1997), acoustic duration of utterance-initial nasals was typically shorter than that of nasals at the beginning of utterance-medial intonational phrases. It is likely that tongue tip movement for utterance-initial /n/ production typically preceded vocalization, given that duration of tongue tip movement in utterance-initial /n/s was longer than in utterance-medial /n/s (Keating, 2006). What is unknown is whether tongue tip constriction achievement¹ precedes, coincides with, or follows vocalization onset. This is noteworthy. Except in the case of labials, interdental and some dentals, speech perceivers are primarily only privy to information transmitted by sound. Any potential information encoded in the relative timing between tongue tip raising and velum lowering for /n/s, or between any two gestures during production of sonorants prior to voicing, is lost to the perceiver.

1.2 Models & Predictions

Table 1.1 summarizes the inter-gestural timing relations discussed in §1.1.1. While there are several differences (and even some contradictions) across the findings, there are also striking commonalities. In each case discussed, articulator movement for syllable-

¹“Achievement” is defined henceforth as the moment that a gesture reaches its target. Precisely how gestural targets are defined has been largely study-dependent: Krakow (1989) and Browman and Goldstein (1995) define target achievement to occur when the articulator has raised to a certain percentage of maximum articulator height, while Gick et al. (2006) define target achievement as the point at which an articulator has reached its maximum height. The differences between the studies cited here are largely due to differences in experimental methods. The methods used by Gick et al. (2006) resulted in lower temporal resolution than Krakow (1989) and Browman and Goldstein (1995).

initial consonants is either synchronous or offset such that the movement towards tighter constrictions precedes movement towards more open constrictions. On the other hand, articulator movement in syllable-final consonants is generally offset in the opposite direction, such that the more open constrictions tend to occur before the more closed constrictions. Synchrony between gestures in multi-articulator coda segments appears to be rare.

That these phenomena are widely reported, in a variety of segments and in a variety of languages, suggests that they may be generalizable². The purpose of this dissertation is to investigate what, ultimately, the sources of these phenomena may be. Several models of speech production are of theoretical interest, with respect to these phenomena. These are summarized below.

1.2.1 Perceptual motivations

Listener-motivated hypotheses of speech production are focused primarily on explaining how the needs of listeners affect speakers' productions. Work by Keating and colleagues (Fougeron and Keating, 1997; Fougeron, 2001; Keating et al., 2003; Keating, 2006) establishes that consonants at the onset of large prosodic structures tend to be strengthened. Keating (2006) suggests that this strengthening occurs in order to facilitate listeners' ability to perceive the segment with the greatest information load, the position of greatest uncertainty. That is, because segments immediately following strong prosodic boundaries are least predictable, speakers are arguably more likely to try to improve perceptibility at these positions. Cho et al. (2007) sought to investigate the perceptual effects of low-level phonetic detail resulting from domain-initial strengthening. They found that

²Certainly, there are known exceptions: American English /w/ production (Gick, 2003), discussed above exhibits inter-gestural synchrony in coda position; tongue dorsum movement precedes tongue tip movement during Georgian onset /l/ production (Berry and Archangelli, 2010); inter-gestural timing has been reported for American English rhotics (Gick and Goldstein, 2002), though interpretation of these findings within the framework established by Krakow (1999) is unclear, on account of uncertainty regarding the constriction degrees of the gestures involved, relative to one another.

segments with the acoustic characteristics of domain-initial strengthening aided listeners during the reconstruction of prosodic structure, though the effect was weak.

Extrapolating from Keating (2006), I suggest that, when producing speech intended to be more easily perceived, speakers are likely to focus on onsets, especially at higher prosodic boundaries, and are likely to be concerned with the speed at which their listeners receive gestural information. Thus, speakers should initiate gestures as early as possible, in order to allow listeners to begin the process of lexical retrieval quickly. The resulting productions of multi-articulator sonorants, such as laterals, should tend towards simultaneity. This argument only holds for sonorants, however. Consider obstruent sequences, such as the onset stop-stop consonant clusters in Georgian, as studied in Chitoran et al. (2002). In this study, Chitoran et al. (2002) hypothesized that onset stop-stop sequences should be produced with asynchronous gestures, for perceptual reasons. That is, because cues to the identity of stop consonants are generally contained in the audible transitions into and out of the consonants, synchrony between gestures in such sequences would be likely to mask these cues. The perceptual hypothesis suggests that *when possible*, synchrony in gestures should increase perceptibility of onset consonants. This claim has yet to be substantiated, and it is the focus of the experiment in Chapter III to test this hypothesis.

Thus, the predictions of the listener-oriented model of speech production are as follows:

1. Tongue tip gesture should follow tongue dorsum gesture in coda laterals; tongue tip and tongue dorsum gestures should be simultaneous in onset laterals.
2. Gestures in utterance-initial laterals (i.e. at a major prosodic boundary) should exhibit even greater gestural synchrony than word-initial laterals.

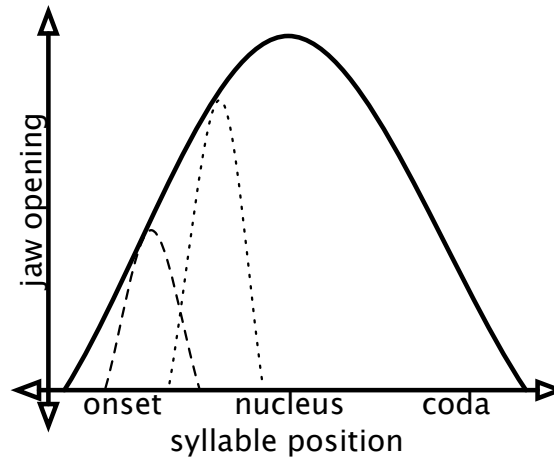


Figure 1.1: Schematization of the effect of jaw opening on extent of articulatory movement required for target achievement.

1.2.2 Speaker-based models

Proponents of speaker-based models of speech production are primarily interested in explaining speech production in light of factors such as the physical constraints on the speech organs, the efficiency of speaker effort, and other biomechanical and physiological aspects of speech production.

The jaw-cycle hypothesis

One speaker-based hypothesis is the “jaw-cycle” hypothesis, which assumes that syllables correlate roughly with a jaw opening-closing cycle, where the jaw is closed at syllable boundaries and open during the syllable nucleus (Lindblom, 1983; Redford, 1999). According to this assumption, all else being equal, a gesture that forms a tight constriction (e.g. lip closure during a /p/) will require less time to reach its target if the gesture is located close to the syllable edge, while the jaw is relatively closed, as illustrated by the dashed line for the tighter constriction in Figure 1.1. On the other hand, gestures forming relatively open constrictions (dotted line) would be more likely to be found closer to the syllable nucleus. Additionally, as anterior gestures (those closer to the lips) are more af-

fected than posterior gestures by the degree of jaw opening, all else being equal, anterior gestures are more likely to be produced when the jaw is relatively closed, or close to the syllable boundaries.

As discussed previously, the American English laterals examined in this dissertation are typically produced with two gestures, a tongue tip and a tongue dorsum gesture. The jaw-cycle hypothesis makes the following predictions for these laterals:

1. In coda laterals, the tongue tip gesture should follow the tongue dorsum gesture, and in onset laterals, the tongue tip gesture should precede the tongue dorsum gesture.
2. At the end of large prosodic structures, speech slows, lengthening the duration of jaw closing. Thus, magnitude of gestural lag should increase at stronger prosodic boundaries.

The findings of Gick (2003) and Gick et al. (2006) are consistent with these predictions. On the other hand, the gestural synchrony found in onset laterals by Browman and Goldstein (1995) suggests that the jaw cycle may not be the only factor affecting gestural timing.

Articulatory Phonology

Many of the observations of the ways in which syllable position affects production of speech segments, as outlined in Section §1.1.1, have been made in support of Articulatory Phonology. Articulatory Phonology stipulates that gestures³ are the primitives of phonology, and that a speaker's phonology plans the coordination and timing of these gestures within an utterance in a "gestural score" (Browman and Goldstein, 1986, 1992; Chitoran et al., 2002; Goldstein et al., 2006). Figure 1.2 shows an example of a gestural score for the word "palm," pronounced [p^ham] from Browman and Goldstein (1992).

³A "gesture" is defined in Articulatory Phonology literature as a goal-oriented movement of a speech organ, regardless of whether or not the motion produces its intended target.

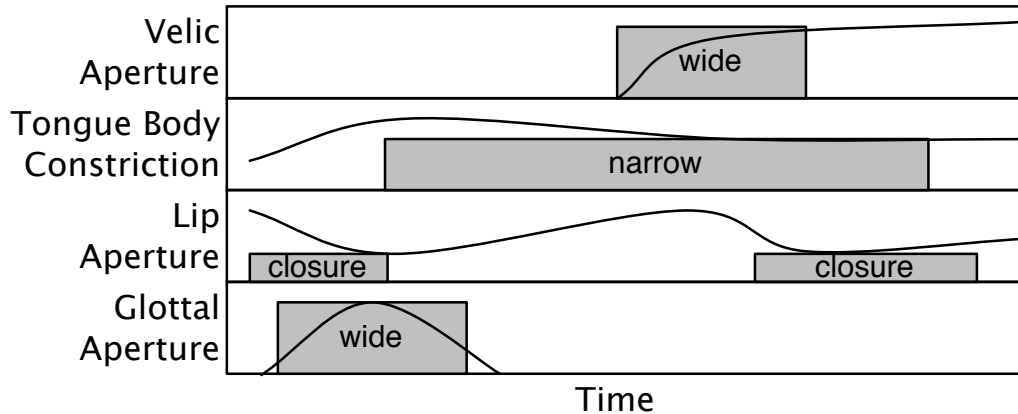


Figure 1.2: Example gestural score, for production of “palm,” pronounced [pʰam], adapted from Browman and Goldstein (1992)

The π -gesture hypothesis

As a component of a more comprehensive model of Articulatory Phonology (that is, one which addresses higher-level phonological structure such as prosody), Byrd and Saltzman (2003) proposed a prosodic gesture, the π -gesture, which encompasses entire prosodic structures. Production of these gestures involves local slowing, towards prosodic boundaries, and speeding up away from boundaries. This proposal helps explain the domain-initial strengthening effects, such as those discussed by (Keating, 2006): as a consequence of boundary slowing, gestures of speech organs may be more capable of achieving their target, or task. Recently, Krivokapić and Byrd (2010) suggested that any effects caused by prosodic boundary adjacency will have the greatest impact on the gestures closest to that boundary. They also suggested that, since gestures of multiple consecutive onset consonants (such as in “spot”) are typically planned to be anti-phase with each other, gestures in mutli-articulator consonants such as nasals and laterals should be no different. The predictions of this hypothesis are:

1. Tongue tip and tongue dorsum gestures should be asynchronous in all laterals regardless of syllabic position.

2. Amount of gestural lag between the two gestures should be greater at stronger prosodic boundaries.

The task-dynamic model of speech production

Articulatory phonologists have adopted Task Dynamics as the mechanism by which gestural scores are implemented. The task dynamic model of speech production assumes that there exists an underlying physiological dynamical system which governs not only motor coordination on the large scale (e.g. arm motion during extension), but also the coordination of speech organs. Within the task dynamic approach to speech production, gesture activation is modeled as a dampened spring, which oscillates in a sinusoidal curve. Coupled oscillators tend to synchronize (Goldstein, 2010), with the most stable temporal relationship between two coupled oscillators being complete synchrony, or “in-phase,” and the second most stable relationship being “anti-phase,” or a phasing difference of 180° . Schematizations of in-phase, anti-phase, and an unstable phasing in oscillators are shown in Figure 1.3. To extend this analogy to speech gestures, the nature of coupled oscillators suggest that synchrony between gestures is the most stable inter-gestural temporal configuration, while asynchrony such that activation of one gesture coincides with achievement of another gesture, is the next most stable configuration (Goldstein et al., 2006). Articulatory phonologists largely attribute the patterns outlined in Section §1.1.1 as arising from in-phase coordination being the dominant gestural configuration at syllable onsets, and anti-phase coordination being dominant in syllable rimes. Thus, for laterals, the task dynamic model predicts that the tongue tip and tongue dorsum gestures should be coordinated in-phase in onset laterals, and anti-phase in coda laterals. It should be noted though, that the task dynamic model of speech production predicts that synchrony between gestures occurs at gestural initiation, or onset. This is a departure from earlier articulatory phonology literature (Browman and Goldstein, 1995), in which synchrony occurs at the

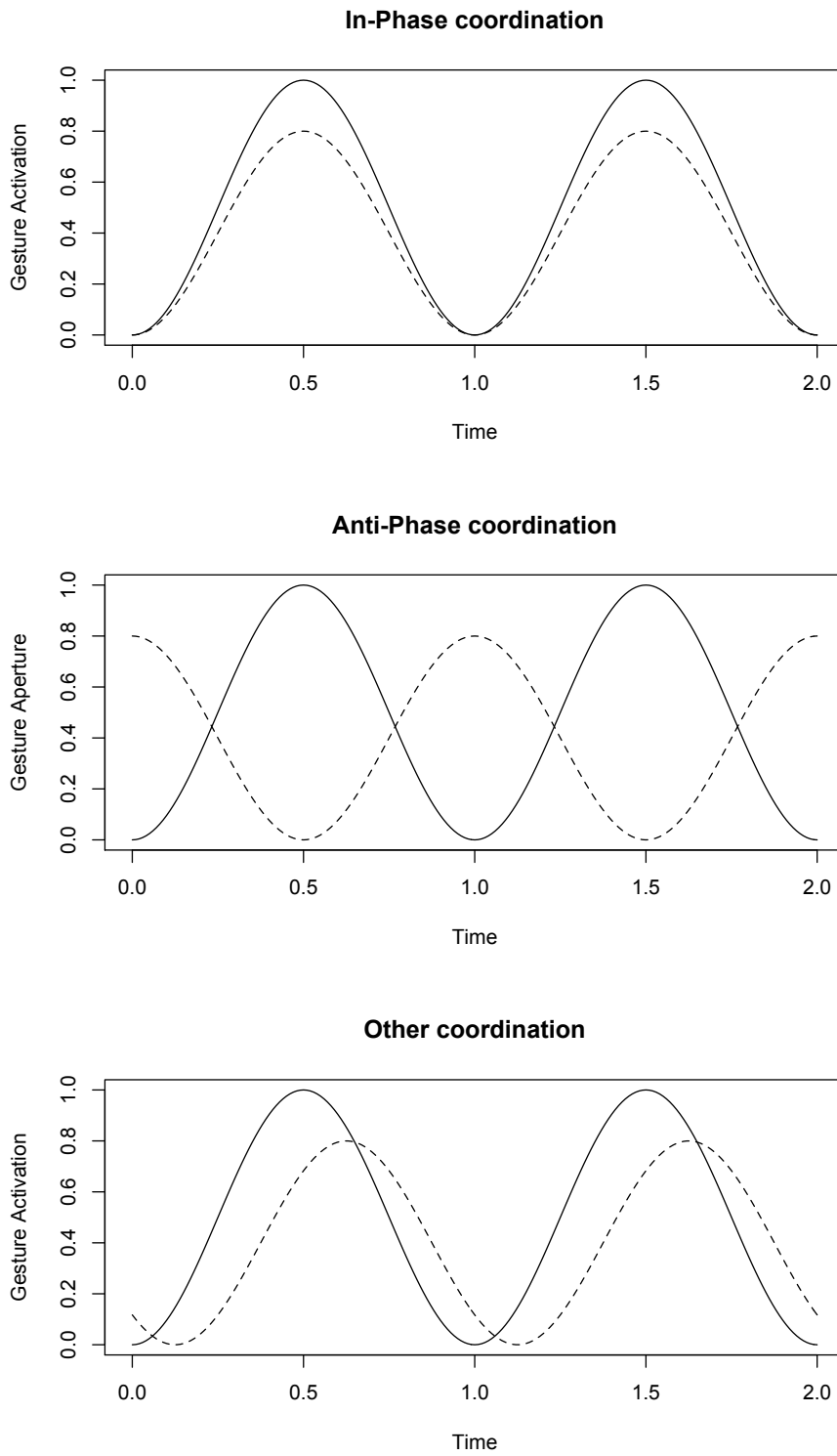


Figure 1.3: Schematizations of in-phase (top), anti-phase (middle) and other unstable (bottom) relationship between two coupled oscillators.

point of gestural achievement. Because this study primarily investigates the relative timing of gestural achievement, the task dynamic predictions are not included in the list of hypotheses below. However, given that the tongue tip is typically a faster moving speech organ than the tongue body, it is most likely the case that if tongue tip and tongue dorsum gestures are initiated simultaneously, tongue tip achievement will precede tongue dorsum achievement in this model.

1.3 Summary

The hypotheses of this dissertation are summarized below:

- During production:

Hypothesis 1 (control; coda lateral condition) In coda /l/s, tongue tip movement should follow tongue dorsum movement, and final laterals preceding stronger prosodic boundaries should exhibit greater gestural lag than final laterals preceding weaker prosodic boundaries. (This will serve in part as a control, to validate the methodology used in Experiment 1.)

Hypothesis 2 In initial /l/s:

Hypothesis 2a (Perceptual) If speakers are motivated primarily by perceptual factors, tongue tip movement and tongue dorsum movement should be roughly simultaneous, and this effect of gestural synchrony may be stronger in higher prosodic positions.

Hypothesis 2b (Jaw-Cycle) If the jaw-cycle dominates, tongue tip movement should occur prior to tongue dorsum movement and, analogous to final /l/s, there should be greater gestural lag in laterals following strong prosodic boundaries than laterals following weak boundaries.

Hypothesis 2c (π -Gesture) Under a π -gesture extension of Articulatory Phonology, tongue tip gestures should precede tongue dorsum gestures in onset laterals, gestural lag in sentence-initial laterals should be greater than in word-initial laterals, and effects of the prosodic boundary (i.e. lengthening, in this study) should be greatest for the tongue tip gesture due to its relative proximity to the π -gesture boundary.

- During perception:

Hypothesis 3 Following from the motivations for Hypothesis 2a, listeners will most quickly and accurately identify initial laterals when tongue tip and tongue dorsum gestures are synchronous.

The principal goals of this dissertation are first, to examine the systematic differences in articulatory coordination in productions of laterals in American English in different syllabic and prosodic contexts, and second, to explore the extent to which these differences are motivated perceptually or physiologically determined. In Chapter II, I present a speech production experiment utilizing ultrasound that manipulates both the syllabic and prosodic positions of /l/, and addresses Hypotheses 1 and 2. Towards substantiating the theoretical claims that lie at the heart of Hypothesis 2, Chapter III presents a perception experiment utilizing synthetic speech, addressing the issues of perceptibility in Hypothesis 3. Chapter IV discusses the implications of the findings in these two experiments, in light of the literature and theoretical framework outlined in this chapter.

CHAPTER II

Experiment 1: Production

Experiment 1 examines inter-articulator timing in syllable-initial and syllable-final laterals across different prosodic positions. Data on syllable-final laterals were collected to serve as a control, and are compared to results reported by Sproat and Fujimura (1993) to verify that the experimental protocol worked as intended.

2.1 Design

	[i]	[e]	[æ]
initial-/l/	leap	lame	lap
final-/l/	peel	male	pal

Table 2.1: Target words for Experiment 1

The targets chosen for this experiment were monosyllabic words of the form /lVC/ and /CVl/, where V was one of the three vowels [i], [e] and [æ]. Front vowels were chosen because, during a pilot version of the study, it was found to be extremely difficult, and in many cases impossible, to distinguish between tongue dorsum movement for back vowels and tongue dorsum movement for the lateral. The design includes multiple vowels in order to assess generalizability, as well as to provide some variety for speakers, with the hopes of avoiding repetition-related confounds during production experiments. The non-lateral consonants (C) were chosen to be bilabials, and not alveolars or velars, in order to minimize any coarticulatory effects on the size and timing of tongue tip and tongue

dorsum gestures in the targeted /l/. When possible, this consonant was the voiceless stop [p]. This was not possible in the /IVC/ [e] context (“*lape”), so [m] was chosen instead for both /IVC/ (“lame”) and /CVI/ (“male”) contexts. The target words are given in Table 2.1.

For speakers S01 - S09, the target words were placed in three prosodic conditions: sentence-initial / -final, phrase-initial / -final, and word-initial / -final, and speakers produced 10 repetitions of each utterance. These stimuli are listed in Table 2.2, arranged by syllable-position, then by prosodic position, then by vowel. For all measurements taken, most speakers produced a significant contrast between word and sentence adjacent (i.e. initial and final) contexts, while measurements from phrase adjacent contexts typically fell between them. (See Appendix A for full details.) Therefore, for subsequent participants, phrase-initial and -final contexts were dropped in favor of increased repetitions, and thus greater statistical power, for utterance- and word-initial and -final contexts. Thus, for speakers S10 - S16, each syllabic condition was matched with two prosodic position conditions: sentence-initial / -final and word-initial / -final, but speakers produced 16 repetitions of each sentence, instead of only 10. The stimuli themselves were identical to those used for speakers S01 - S09.

2.2 Equipment and Techniques

2.2.1 Equipment

The equipment used in this experiment is shown in Figure 2.1. The ultrasound engine used to collect data was a portable *z.one mini* ultrasound system, commercially available from Zonare Medical Systems, using a P4-1c phased array transducer. This transducer is often used for cardiac ultrasound imaging, making it an excellent transducer for tongue imaging due to its high theoretical maximal scan rate and appropriate imaging depth (6-

Sentence-Initial	i	The dog sees Amy. Leaping across the river, he runs towards her.
	e	The wounded fox sees Amy. Lamely, he stumbles away.
	æ	Amy fed Cammie. Lapping up the milk, she purred.
Phrase-Initial	i	The dog, seeing Amy, leaps across the river.
	e	The wounded fox, seeing Amy, lamely stumbles away.
	æ	Cammie, purring at Amy, laps up the milk.
Word-Initial	i	The dog wants to see Amy leap across the river.
	e	A wounded Amy lamely made her way to the hospital.
	æ	Amy wants to see Cammie lap up the milk.
Sentence-Final	i	I dropped an orange peel. Emptying the box, I found it.
	e	My friend Jane is male. Embarrassing as it can be, he likes it.
	æ	Amy, drive my pal. Embarking from the airport is grueling.
Phrase-Final	i	Add an orange peel, immersing it in the cider.
	e	Whether female or male, employees must wear slacks.
	æ	Amy and my pal, embarking on their trip, made one final stop.
Word-Final	i	An orange peel emerged from the bubbling cider.
	e	Even the female employees are required to wear ties.
	æ	My brother's pal embarked on her trip.

Table 2.2: Target sentences for Experiment 1

10 cm). The ultrasound engine itself was set outside of the sound-attenuated recording booth at the University of Michigan in which the speakers sat, while the transducer was passed through a portal in the booth padded with sound-dampening foam. This setup was designed to allow for high-quality audio recordings during ultrasound imaging sessions.

One of the issues that most speech researchers using ultrasound imaging must deal with is correcting for or eliminating head movement in relation to the transducer during

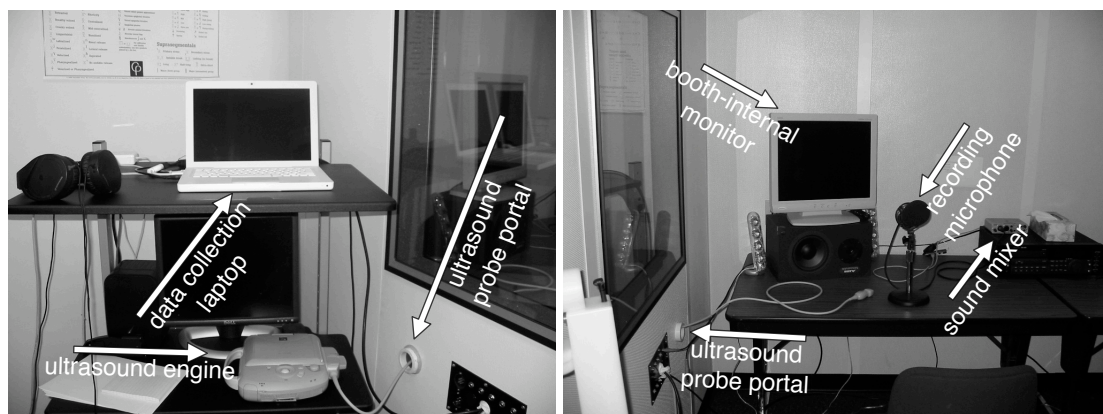


Figure 2.1: Equipment used in Experiment 1, inside (right) and outside (left) the sound-attenuated booth at the University of Michigan.

data collection. There are a variety of methods for solving this problem. Some systems (e.g. Palatron; Mielke et al., 2005) utilize a light-weight head-mounted device (such as glasses) and an external video recorder that captures the movement of specific points on the head-mounted device. These systems perform a post-collection analysis of the external video data to maintain a fixed axis in the ultrasound video. Other systems, such as the HATS system (Stone and Davis, 1995), include an ultrasound probe holding stand, usually placed to the side of the talker, with an apparatus to hold the head still during the course of the experiment. The current experiment uses an Ultrasound Stabilization Helmet, pictured in Figure 2.2¹, manufactured by Articulate Instruments Ltd., created specifically for ultrasound speech research. Like the HATS system, it maintains a fixed position of the ultrasound transducer in relation to the speaker's chin. The primary advantage of using a headset instead of a person-external stand is that the speaker is able to move his or her head without disturbing the relative positioning of the probe. Additionally, though irrelevant for this study, the headset is portable, which helps makes ultrasound image collection in the field reasonable.

Ultrasound video was captured by digitally streaming the output from the engine onto a laptop through an Epiphan USB2DVI frame grabber, which operated at a rate of 29.97 Hz (33 ms between frames)². The video was captured by QuickTime Pro on a Apple MacBook laptop. The scan rate of the ultrasound engine given the settings used for this experiment was 34 Hz (29.4 ms between scans) in b-mode operation³. This presents a mathematically non-trivial issue: four frames a second are dropped between the ultrasound engine and the video capture device. More explicitly, it is impossible to know whether two frames adjacent in the video captured by the laptop are truly 29.4 ms apart or (if a frame was

¹Credit to Andries Coetzee for the photograph, and to Joseph Tyler for modeling.

²Attempts were made to capture data on the ultrasound engine itself, but the amount of data that the engine was capable of storing at one time (100 frames, or 3 seconds at 34 Hz) was too limited.

³M-mode operation would allow for higher temporal resolution, at the expense of only being able to measure lingual movement in one dimension. Since it is necessary in this study to be measuring lingual movement in two places, b-mode was used.



Figure 2.2: Ultrasound Stabilization Helmet, in use.

dropped between the two) 58.8 ms apart. However, the probability that any two given frames are actually one frame apart is $p(t = 29.4ms) = 0.8824$, while the probability that they are two frames apart is $p(t = 58.8ms) = 0.1176$, and the estimated actual time between any two adjacent frames in the streamed output is $E(t) = 29.4 * 0.8824 + 58.8 * 0.1176$, or 32.9 ms. In fact, for as few as $N = 10$ tokens, it is very probable (90% likely) that the mean actual time between two adjacent frames is between 29.4 ms and 35.3 ms, with the mean value being $t = 32.8ms$. In other words, with a large enough set of tokens, it is highly likely that the average actual time between two frames is 33 ms. It is for this reason, I believe, that researchers who capture ultrasound video at the standard NTSC rate of 29.97 Hz assume an underlying 30 Hz for their analyses. This study follows this procedure.

Video from the collection laptop outside the sound booth was relayed to a monitor

inside the booth so that the speaker was capable of seeing his or her oral cavity when learning to produce a water bolus. During the speaking task, this monitor was used to display the stimuli to the speakers, and speakers were no longer capable of viewing the ultrasound images. The audio signal was recorded to the laptop using an AKG microphone connected to the laptop via an Edirol USB sound mixer.

2.2.2 Analysis Techniques

Following data collection, the audio was segmented by stimulus and matched with its corresponding ultrasound video clip. Ultrasound video was then extracted to frames, at a rate of 29.97 Hz, and video frames were aligned to relevant audio landmarks.

In this experiment, video and audio were aligned by having speakers produce the word “ok” before each stimulus item. This provided a clear acoustic landmark (stop burst) corresponding to a clear articulatory landmark (velar closure release). The stop burst in the acoustic signal was defined as the onset of the first visible aperiodic noise following the closure, and marked in a TextGrid in Praat (Boersma and Weenink, 2008). The ultrasound video frames were visually analyzed for the corresponding velar release, which was defined to be the first frame after maximal tongue body retraction or raising for the velar stop, as in frame 357 of Figure 2.3. The frame number corresponding to the release in the video signal was entered in the TextGrid, as shown in Figure 2.4.

Following signal alignment, the acoustic signal was analyzed for the target $/C_1V_1|V_2C_2/$ syllables, as listed in Table 2.2, and marked for the onsets and offsets for each phone.⁴ This is illustrated for a sample token in Figure 2.5. EdgeTrak was used to convert the ultrasound images from the target syllables into quantifiable data. EdgeTrak is a snake-algorithm based edge detection program developed for speech research (Li et al., 2005) to calculate tongue traces for each of the frames. This is a semi-automated process in which

⁴In some speakers’ productions of $/CVl\#om/$, the vowel $/ə/$ was not acoustically present due to severe reduction or deletion, and therefore no V_2 was recorded on the TextGrid.

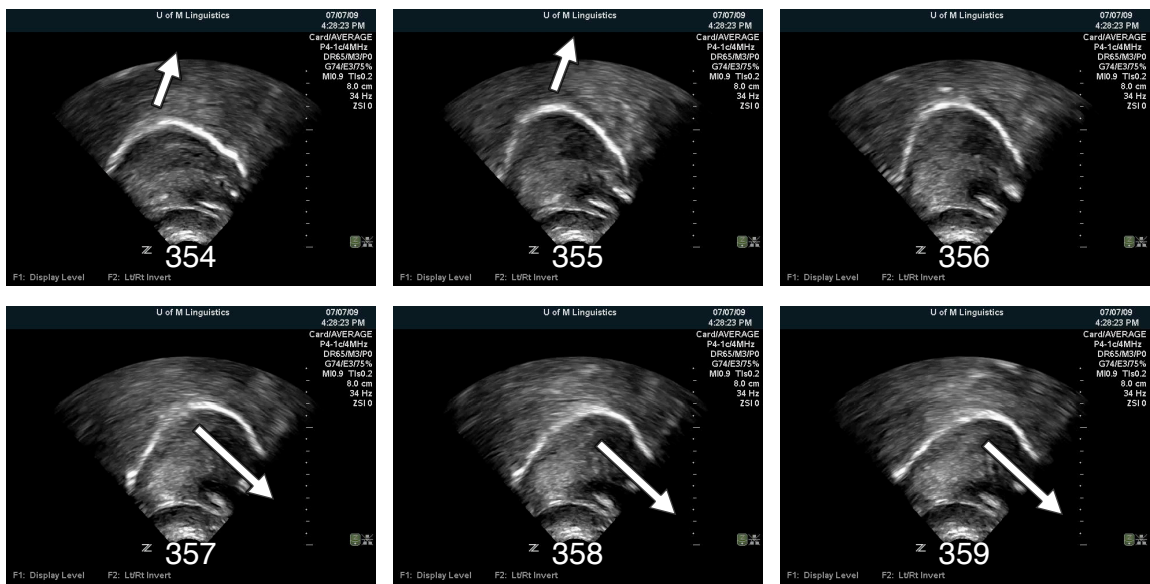


Figure 2.3: A series of ultrasound frames during production of /k/ from “ok,” as produced by Speaker S04, showing movement of the tongue back towards closure (frames 354 and 355), during closure (frame 356), and after release (frames 357 to 359). Tongue tip is located to the right, and tongue root to the left.

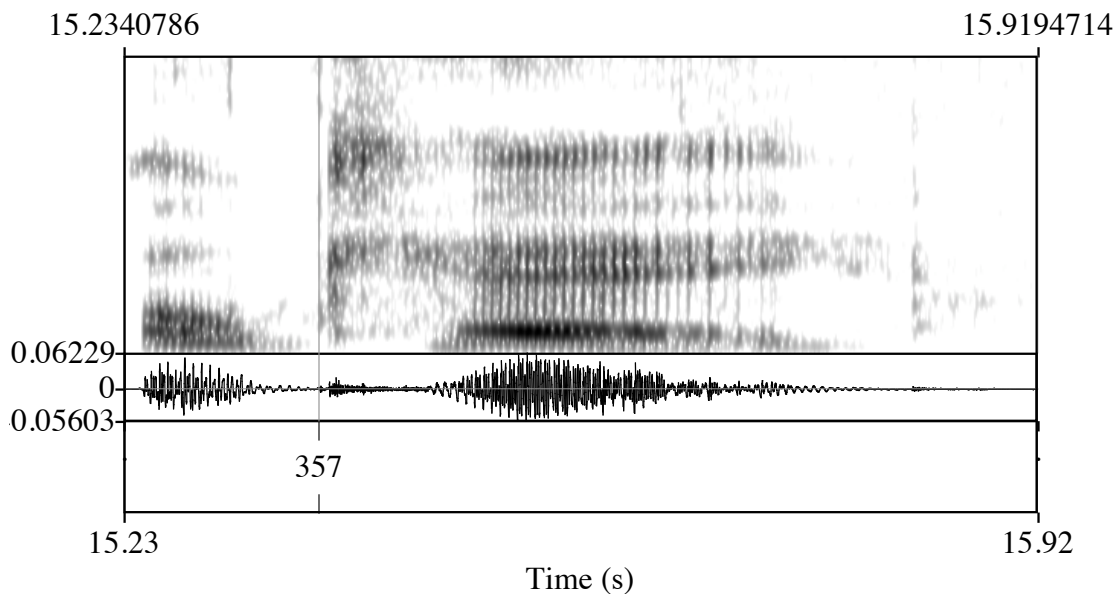


Figure 2.4: Locating the acoustic correlate of release in the velar stop in “ok;” Sample taken from speaker S04, the same /k/ closure for which ultrasound images are shown in Figure 2.3

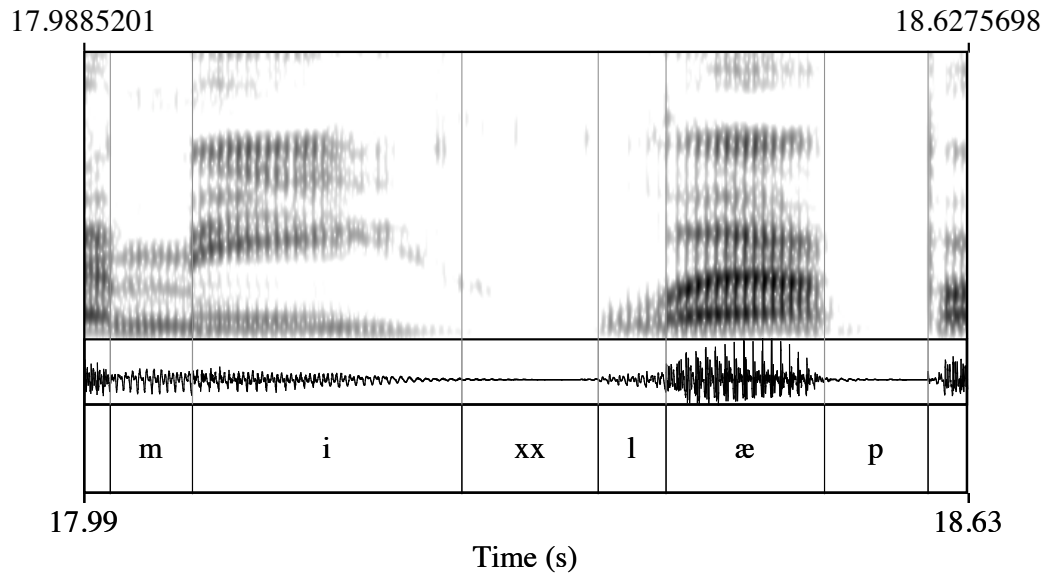


Figure 2.5: Locating the boundaries of phones in the target syllables in “...(Cam)mie. Lap...”; Sample taken from speaker S04.

the user fits a contour to the first image in the series. The software uses a modified active contour model (snake algorithm) to produce best-fit contours for each subsequent image, given the previous image. In my analysis, I cycled through the frames to validate the edge detection, identifying poor edge detection decisions (as in Figure 2.6 (left)) and hand-correcting them (Figure 2.6 (right)). The percentage of frames requiring hand-correction varied widely from speaker to speaker, likely due to tongue movement speed, and ranged from roughly 2% to 15% of all frames analyzed. The output that EdgeTrak produces on a given series of images is of the form:

$$\begin{array}{cccccc}
 X_{1,1} & Y_{1,1} & X_{1,2} & Y_{1,2} & X_{1,3} & Y_{1,3} & \dots \\
 X_{2,1} & Y_{2,1} & X_{2,2} & Y_{2,2} & X_{2,3} & Y_{2,3} & \dots \\
 \dots & & & & & &
 \end{array}$$

where $X_{a,b}$ is the X-coordinate of the a^{th} point along the contour of the b^{th} image. For instance, $X_{2,5}$ is the X-coordinate of the fifth point along the contour of the second image, and $Y_{2,5}$ is the corresponding Y-coordinate.

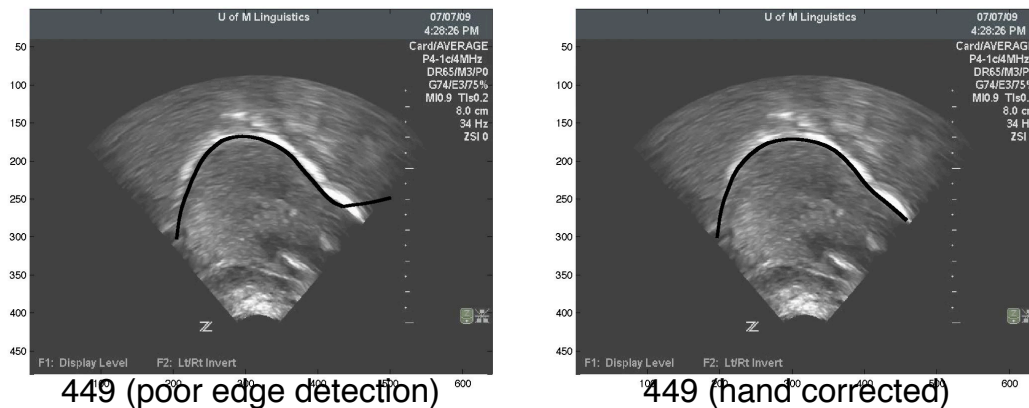


Figure 2.6: Example of poor EdgeTrak output (left) and corresponding hand-correction (right); Sample taken from speaker S04.

The edge detection process was also used to obtain palate traces for each speaker. EdgeTrak was run on images of a water bolus at various stages of swallowing, and the output was averaged across all images from a given speaker to produce a single palate trace for that speaker.

2.2.3 Measuring Tongue Movement

Tongue movement was measured as aperture over time, where aperture is defined as the least distance from the tongue trace to the palate trace in particular “regions of interest” (ROI). This method is a modification of that used by Byrd et al. (2009) to measure velum and tongue tip aperture in the nasal stop /n/ in MRI images. In this study, two ROIs were chosen along the palate trace, such as is shown in Figure 2.7, representing the speaker’s gestural targets along the palate for the places of articulation in question – one for tongue tip movement and one for tongue dorsum movement. To find the regions of interest, I plotted the tongue contours of several randomly selected /CVVC/ productions and the palate trace for a given speaker. Then, based on the tongue contours with significant tongue dorsum retraction or tongue tip raising, I determined a region along the palate

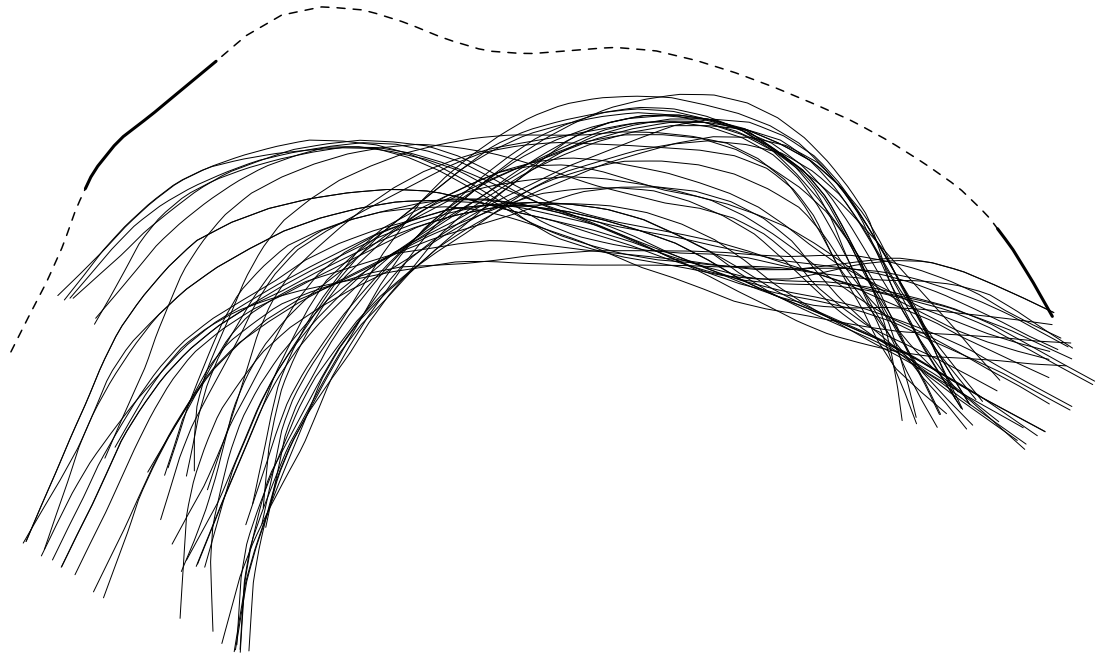


Figure 2.7: Example of aggregate tongue contours from all frames for a single /CVVC/ production (solid lines) “regions of interest” (solid segments) along the palate trace (dashed line) for tongue dorsum target (left), tongue tip target (right), for speaker S04.

corresponding to the apparent gestural targets. ROIs were determined independently for each of the speakers in the study. Within these regions, the minimum distance between the tongue and the palate traces was found. This method produces an inverted graph, such as that in Figure 2.8, as smaller aperture measures represent greater closure, with zero representing full closure.

Based on these trajectory graphs, the following values were calculated:

Lag time (ms) Lag⁵ is defined as time of achievement of minimum tongue dorsum aperture minus time of achievement of minimum tongue tip aperture. Use of “achievement” here is adapted from Gick et al. (2006): achievement frame is defined to be the first frame showing aperture within 3 pixels of the minimum aperture. Thus, a

⁵In this dissertation, I am adopting the convention of using the term “lag” as short hand for “time between achievement of two gestures.” I do not prefer this term, as I feel it implies that gestures are “meant to be” synchronous with one another, but I adopt it here for clarity.

Example Aperture Plot (S13)

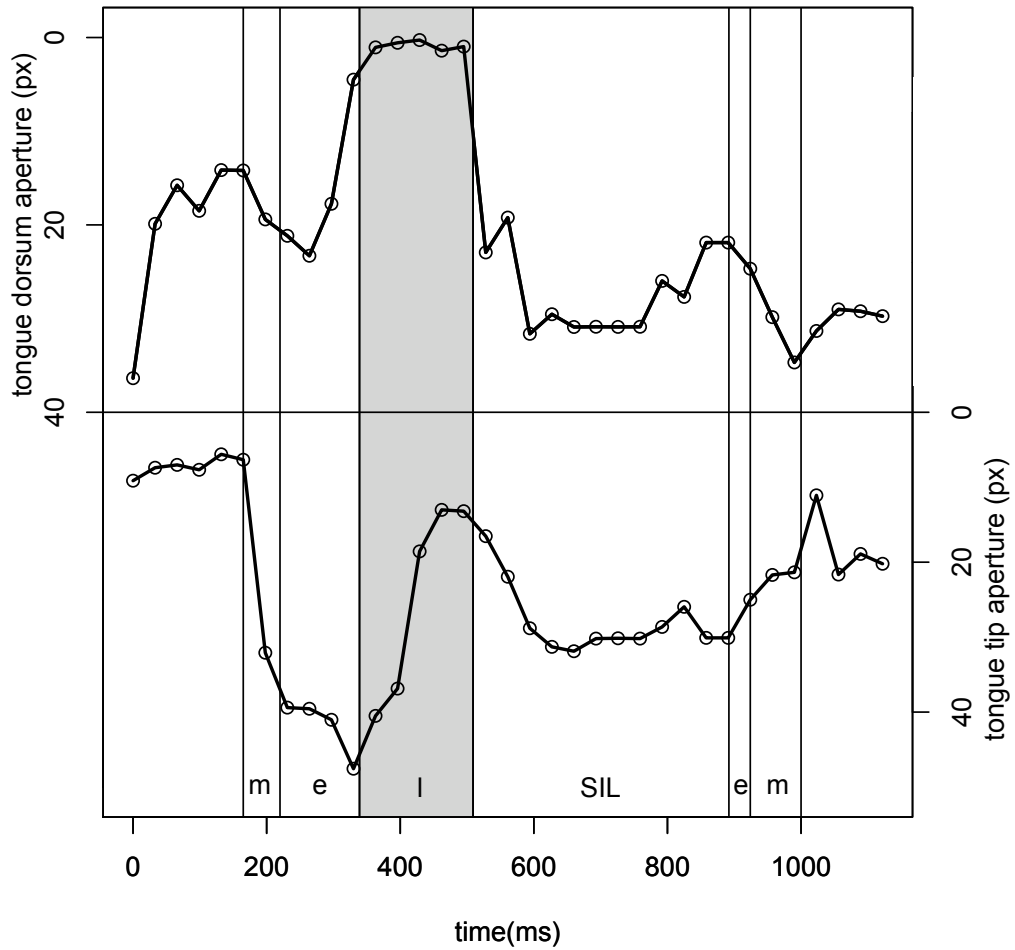


Figure 2.8: Tongue dorsum (top) and tongue tip (bottom) aperture, measured in pixels, against time during production of “...male. Em(barrasing)...” by speaker S13. Vertical lines delineate acoustic boundaries of the segments, with the acoustic duration of /l/ shaded in gray. SIL denotes silence. The first frame measured (at 0 ms along the x-axis) is five frames prior to the acoustic onset of C_1 in the target $C_1V_1IV_2C_2$ sequence.

positive lag indicates that tongue tip raising preceded tongue dorsum backing, while a negative lag indicates that tongue tip raising follows tongue dorsum backing. Note that this is opposite the convention used by Sproat and Fujimura (1993), Browman and Goldstein (1995) and Gick et al. (2006), but conforms to the convention used by Byrd et al. (2009) and Krivokapić and Byrd (2010).

Tongue tip gesture duration (ms) This is defined to be the duration, from onset of movement towards the gestural target to the offset of movement away from the peak, of a gesture. Gestural onset is defined as the point of maximal tongue tip aperture between the acoustic boundaries of the lateral, or immediately preceding the left acoustic boundary. Gestural offset is defined as the point of maximal tongue tip aperture after the onset.

Tongue dorsum gesture duration (ms) This is identical to the duration of the tongue tip gesture, but for the tongue dorsum aperture.

Total lateral duration (ms) This is defined to be the duration, from earliest onset of movement to latest offset of movement for the lateral, where onset corresponds to either the onset of the tongue tip gesture or onset of the tongue dorsum gesture, whichever occurs first. Similarly, offset corresponds to either offset of tongue tip or tongue dorsum gesture, whichever occurs last.

In ultrasound images generated by phased array transducers, such as the one used in this study, the ultrasound signals emanating from the transducer are not produced all at once. The transducer used in this experiment has 39 zones, each of which fires independently of each other. This is potentially problematic for this study because the primary variable is temporal distance between two different lingual targets, which are found in different zones. Thus, the tongue tip and tongue dorsum gesture may appear to be simultaneous when in

fact they are offset by as much as 29 ms. For speakers S10 through S16, this issue was circumvented by reversing the orientation of the transducer in half of the collected data⁶. Thus, when averaged together, the offset in lag values will cancel each other out. Data for speakers S01 through S09 were collected before I was aware of the issue. However, post-collection corrections were made for these speakers. Under the ultrasound settings used in this study, each zone fires 200 μ s apart. Thus, all data in the image is obtained in 7.8 ms from left to right (DeBusschere, personal communication, Dec. 11, 2009). To perform the correction, the number of zones between the tongue tip and tongue dorsum regions of interest were found, this was multiplied by 0.2 ms, and added to total lag values.

2.3 Results and Discussion

2.3.1 Speaker Exclusion

Sixteen native speakers of American English participated in this experiment. Table 2.3 lists the 16 speakers, their genders, and a brief language background. While female speakers were not specifically targeted over male speakers, the overwhelming majority of individuals responding to the recruiting materials were female. Of these participants, data from five were not used in the analysis. Data from speakers S03, S05 and S11 were excluded because, in many contexts, critical portions of their tongues were outside of the viewable area of the ultrasound. Data from speaker S15 was excluded due to poor ultrasound image quality. Finally, data from speaker S14 was excluded because this speaker's tongue body gestures for the target front vowels were extremely front, and in many cases obscured tongue tip movement for adjacent laterals.

Speaker	Gender	Language Background	Data Used?
S01	female	English monolingual (Michigan)	YES
S02	female	English monolingual (Michigan)	YES
S03	female	English monolingual (Illinois)	NO: tongue outside of viewable area
S04	female	English monolingual (Michigan)	YES
S05	male	English monolingual (Michigan)	NO: tongue outside of viewable area
S06	female	English monolingual (Michigan)	YES
S07	female	English monolingual (Michigan)	YES
S08	female	English monolingual (Michigan)	YES
S09	female	English monolingual (California)	YES
S10	female	English monolingual (Michigan)	YES
S11	male	English monolingual (Michigan)	NO: tongue outside of viewable area
S12	female	English monolingual (Michigan)	YES
S13	female	English monolingual (Michigan)	YES
S14	female	English monolingual (Michigan)	NO: tongue tip movement obscured
S15	male	English monolingual (Wisconsin)	NO: poor image quality
S16	male	English monolingual (Michigan)	YES

Table 2.3: Summary of participants in Experiment 1

Factors	prosodic context, vowel context
Dependent Variables	gestural lag, lateral duration, tongue tip gesture duration, tongue dorsum gesture duration

Table 2.4: Factors and dependent variables in this study.

2.3.2 Statistical Procedures

To determine whether productions of coda and onset laterals in the collected data exhibited positive lag, negative lag, or lag not significantly different from zero, I performed a two-layer mixed linear model. To further analyze the data, I performed a series of two-way ANOVAs (prosodic context x vowel context) for each speaker in the study, to determine whether the dependent variables, listed in Table 2.4, are affected by the independent factors. For consistency across the data, only utterance- and word-boundary prosodic contexts were used in these analyses. For full details regarding the three-way utterance-, phrasal-, and word-boundary contrasts for speakers S01-S09, see §A.1 in the appendix. Further, in the following analyses, data points are separated by syllable position, and discussed independently.

⁶Thanks to Maureen Stone for first pointing out this problem, and then suggesting this solution.

2.3.3 Results - Overall Lag

The results of the two-layer mixed linear model show that, as expected, coda laterals produced by speakers in this study typically had negative gestural lag, indicating that the tongue dorsum gesture preceded the tongue tip gesture ($F(1, 10.21) = 99.94, p < 0.0001$), with an estimate of -92.82 ms, and a standard error of 9.285 ms. Additionally, it shows that onset laterals were generally produced with positive gestural lag, indicating that the tongue tip gesture preceded the tongue dorsum gesture ($F(1, 10.30) = 24.35, p = 0.0001$), with an estimate of 39.02 ms, and a standard error of 7.907 ms. Thus, overall, the predictions of the speaker-oriented hypotheses were upheld.

2.3.4 Results - Coda /l/

Recall from section §1.3, the hypothesis (H1), based on Sproat and Fujimura (1993) and Browman and Goldstein (1995):

Hypothesis 1 In coda /l/s, tongue tip movement should follow tongue dorsum movement, and final laterals preceding stronger prosodic boundaries should exhibit greater gestural lag than final laterals preceding weaker prosodic boundaries.

Thus, the expectation is that coda laterals will show negative lag, and that utterance-final laterals should show greater negative lag than word-final laterals, due to the general articulatory slowing that precedes strong prosodic boundaries. Indeed, this is borne out for all speakers analyzed, as can be seen in Figure 2.9. A repeated measures ANOVA, with speaker as the error factor, showed that there is a significant main effect of prosodic position on gestural lag in coda laterals, in the direction expected ($F(1, 6) = 50.60, p = 0.0004$). Overall, as expected, no main effect of vowel quality was found ($F(2, 6) = 1.14, p = 0.3800$) and no interaction between vowel quality and prosodic context was found ($F(2, 6) = 1.90, p = 0.2300$).

There is, as should be expected, some speaker variation. To evaluate results from individual speakers, individual ANOVAs were performed on data from each speaker. Full statistical analyses are reported in appendix section §B.1, but barplots of the duration measures are presented here in Figures 2.9-2.12.

As shown in Figure 2.9, all eleven speakers produced, as expected, syllable-final laterals with negative lag, or lag not significantly different from zero. This indicates that tongue dorsum movement typically preceded tongue tip movement, and when it did not, gestures were close to simultaneous. Additionally, most of the speakers showed the same effects of prosodic structure, in the expected direction: both tongue tip (Figure 2.11) and dorsum gestures (Figure 2.12) were longer in sentence-final position than word-final position, as was the total duration of lateral articulation (Figure 2.10). Finally, as predicted, gestural lag was greater in sentence-final contexts than in word-final contexts (Figure 2.9). The productions of three speakers deviated, and are discussed briefly below.

Speaker S01

In the productions of speaker S01, both tongue tip and tongue dorsum gestures, as well as the total duration, were longer in sentence-final laterals than in word-final laterals, as expected. However, gestural lag was not found to be significantly greater in sentence-final laterals than in word-final laterals, though the difference trended in the predicted direction ($F(1, 42) = 3.43, p = 0.0712$).

Speaker S08

Prosodic position had no significant effect on any of the four measures taken in coda laterals, for speaker S08. This speaker produces all coda laterals similarly, across all contexts, both prosodic and vocalic.

Speaker S09

For speaker S09's productions, tongue tip gestures and total lateral articulation were both longer in sentence-final laterals than in word-final laterals. However, tongue dorsum gestures in sentence-final laterals were not significantly longer than in word-final laterals. This result is consistent with predictions made by Krivokapić and Byrd (2010), that prosodic boundaries should have the strongest lengthening effect on gestures closest to them. Thus, tongue dorsum gestures in coda laterals should be less affected by adjacent prosodic boundaries than their tongue tip counterparts.

Summary

Generally speaking, both tongue tip and tongue dorsum gestures were longer in sentence-final laterals than in word-final sentence-medial laterals, as was total duration of lateral articulation. Correspondingly, and consistent with Hypothesis 1, gestural lag in sentence-final laterals was found to be greater than in word-final laterals.

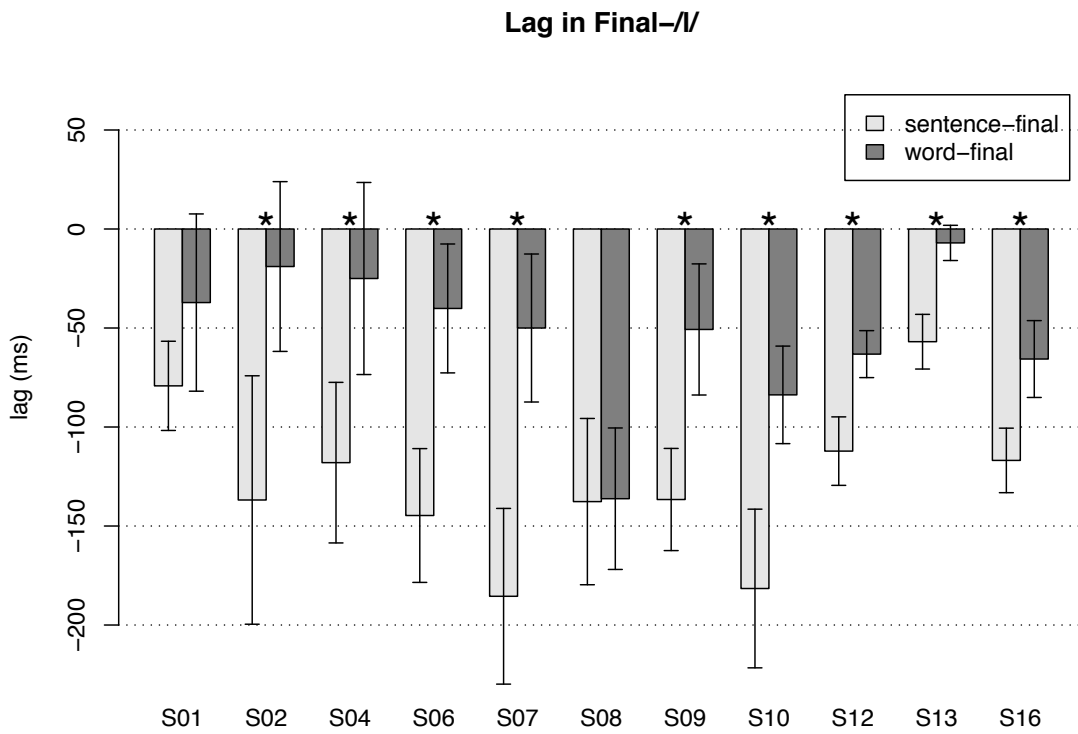


Figure 2.9: Gestural lag in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.

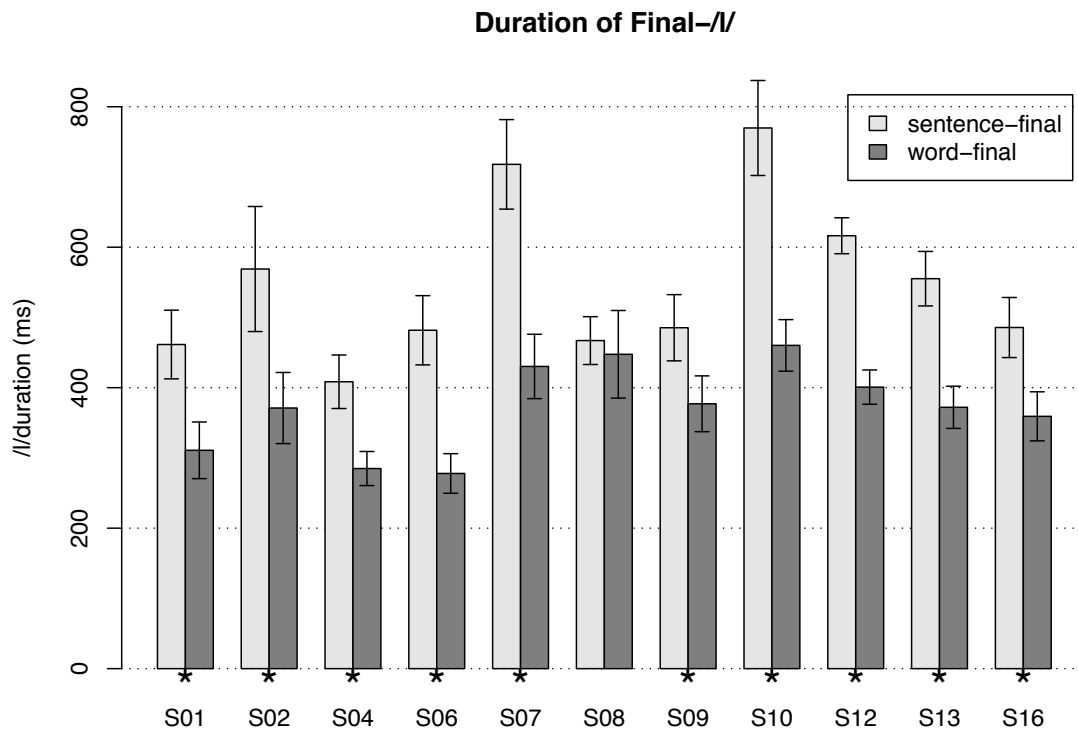


Figure 2.10: Lateral duration in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.

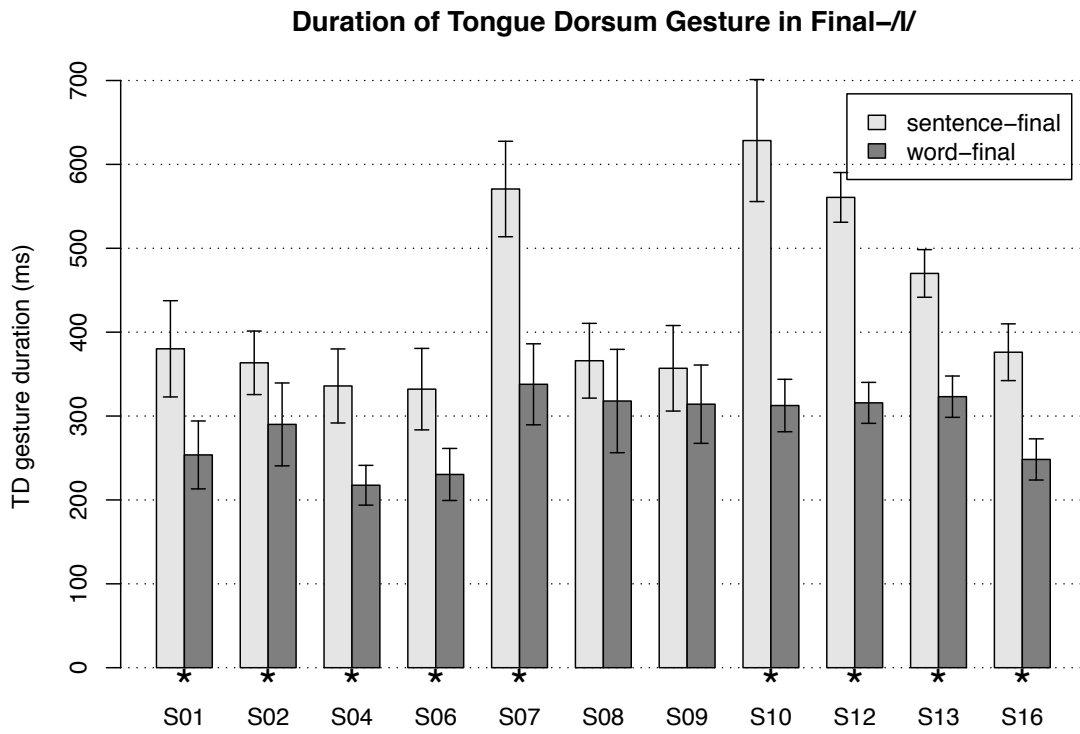


Figure 2.11: Duration of tongue dorsum gesture in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, significance at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.

Duration of Tongue Tip Gesture in Final-/l/

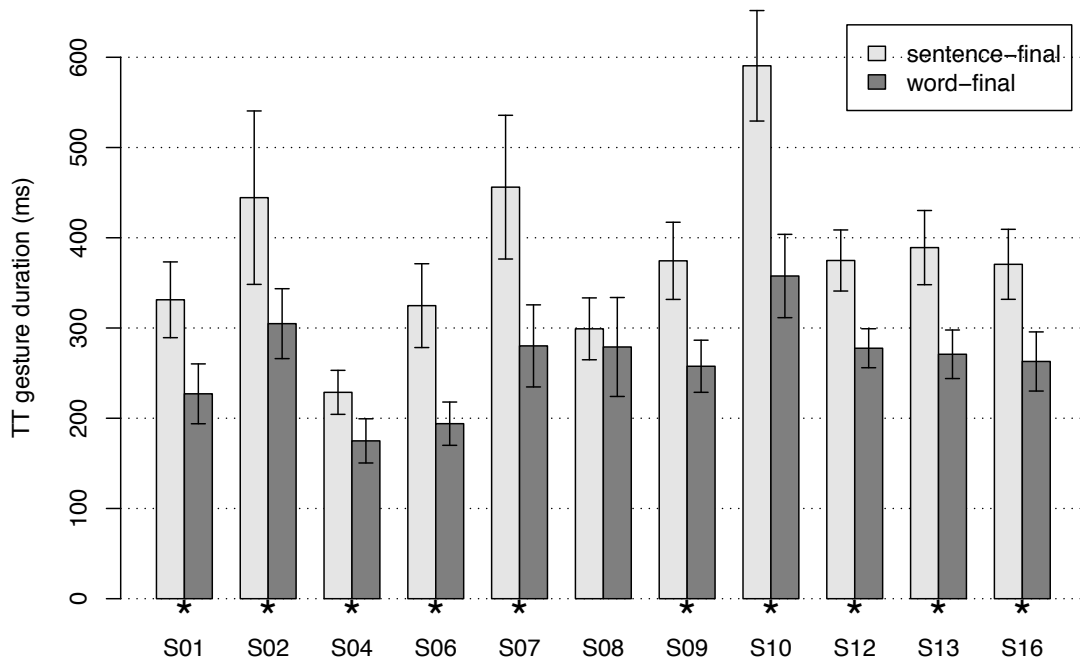


Figure 2.12: Duration of tongue tip gesture in syllable-final position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Negative lag indicates TD gesture precedes TT gesture.

2.3.5 Results - Initial /l/

Figures 2.13 - 2.16 summarize the results for initial laterals. Full statistical analyses for individual speakers are located in the appendix, section §B.2. Recall, from section §2.3.5, the hypotheses regarding gestural lag in onset laterals.

Hypothesis 2 In initial /l/s:

Hypothesis 2a (Perceptual) If speakers are motivated primarily by perceptual factors, tongue tip movement and tongue dorsum movement should be roughly simultaneous, and this effect of gestural synchrony may be stronger in higher prosodic positions.

Hypothesis 2b (Jaw-Cycle) If the jaw-cycle dominates, tongue tip movement should occur prior to tongue dorsum movement and, analogous to final /l/s, there should be greater gestural lag in laterals following strong prosodic boundaries than laterals following weak boundaries.

Hypothesis 2c (π -Gesture) Under a π -gesture extension of Articulatory Phonology, tongue tip gestures should precede tongue dorsum gestures in onset laterals, gestural lag in sentence-initial laterals should be greater than in word-initial laterals, and effects of the prosodic boundary (i.e. lengthening, in this study) should be greatest for the tongue tip gesture due to its relative proximity to the π -gesture boundary.

Consistent with findings from previous studies (Sproat and Fujimura, 1993; Browman and Goldstein, 1995; Gick, 2003; Gick et al., 2006; Krivokapić and Byrd, 2010), and with Hypotheses 2a-2c, speakers are expected to produce initial laterals with tongue tip movement preceding tongue dorsum movement, or with synchronous gestures. This was

found to be true for the speakers in this study. As can be seen in Figure 2.13, gestural offset was always positive or close to zero⁷ for all speakers.

A repeated ANOVA, with speaker as the error factor, showed that there was no significant overall main effect of prosodic position on gestural lag in onset laterals ($F(1, 6) = 1.97, p = 0.2096$), no significant main effect of vowel ($F(2, 6) = 0.44, p = 0.6627$), nor was there a significant interaction ($F(2, 6) = 0.54, p = 0.6094$). As was found for coda laterals, the onset lateral productions varied substantially from speaker to speaker. The remainder of this section refers to ANOVAs performed on individual speakers.

With respect to the effects of prosodic structure on gestural lag, prosodic position had a significant effect on lag for only three of the eleven speakers (S10, S12, and S13) analyzed in this study⁸. These data are shown in Figure 2.13. The direction of the effect for these three speakers was such that gestural lag was greater in word-initial laterals than in sentence-initial laterals. In particular, this result is opposite the predictions of the jaw-cycle and the π -gesture hypotheses (2b and 2c, respectively), but consistent with the predictions of the perceptual (2a) hypothesis.

Also unexpected, as Figure 2.14 illustrates, is that, while duration of sentence-initial laterals was longer than word-initial laterals for all speakers, the difference was significant only for four speakers (S04, S07, S08 and S13). Finally, Figures 2.15 and 2.16 show tongue dorsum and tongue tip gesture durations. In general, prosodic position had a significant lengthening effect on the duration of tongue dorsum gesture (8 of the 11 speakers), but not on tongue tip gesture (2 of 11 speakers). This is of particular note, since Krivokapić and Byrd (2010) suggest that prosodic boundaries should have the most lengthening effect on gestures directly adjacent to the boundary, here, the tongue tip gesture. In sentence-

⁷“Close to zero” here was arbitrarily defined to be within 33 ms, or one frame length, of zero.

⁸These three speakers are all among the four speakers for whom 16 repetitions of each stimulus was collected, rather than 10. It is possible that 30 data points, at most, per prosodic condition simply did not provide enough power to evidence a main effect of prosodic context, for onset laterals.

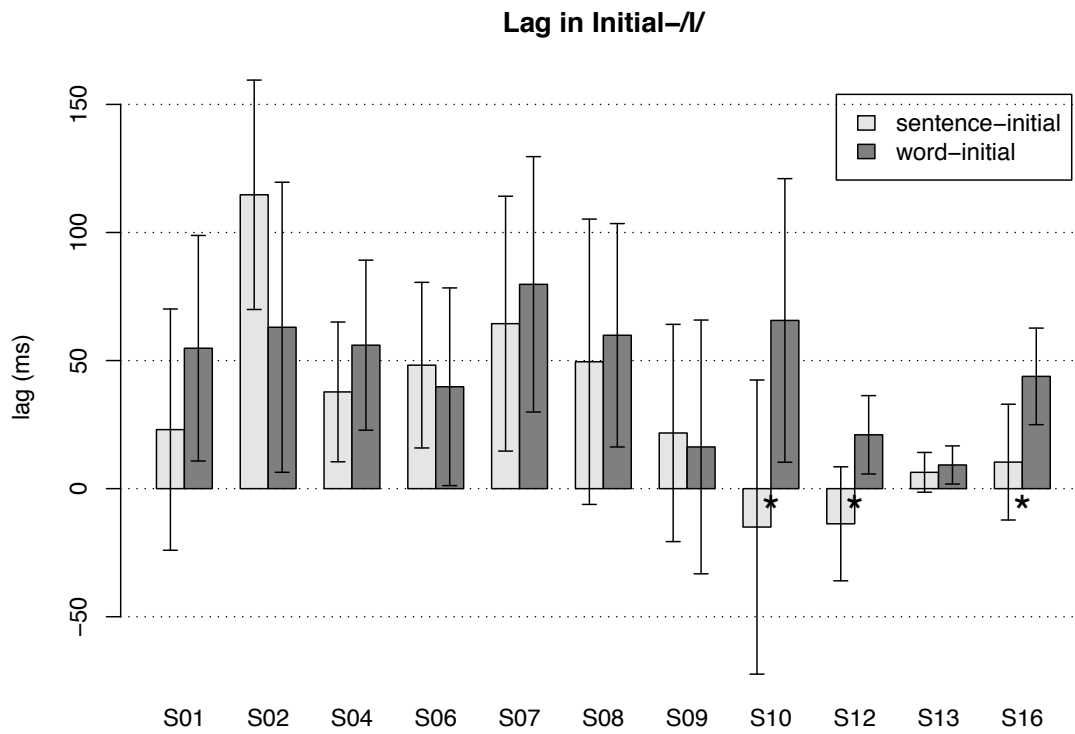


Figure 2.13: Gestural lag in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.

initial laterals then, tongue tip gestures should be lengthened more than tongue dorsum gestures. This will be addressed below in further detail.

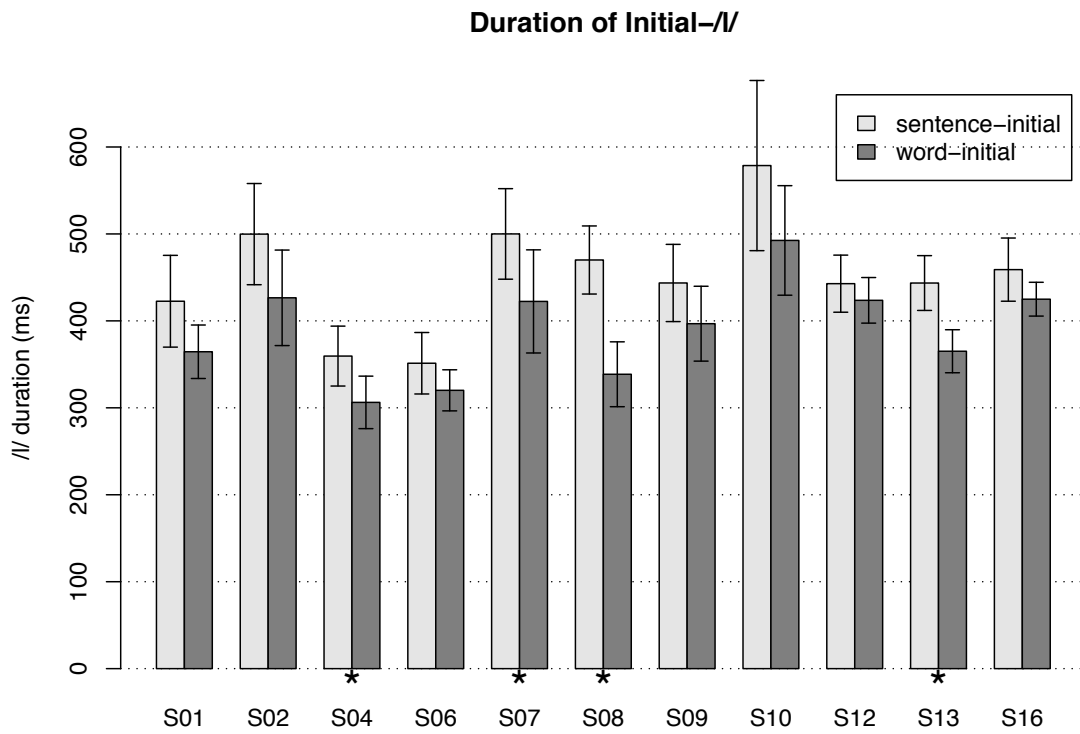


Figure 2.14: Lateral duration in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.

Duration of Tongue Dorsum Gesture in Initial-/l/

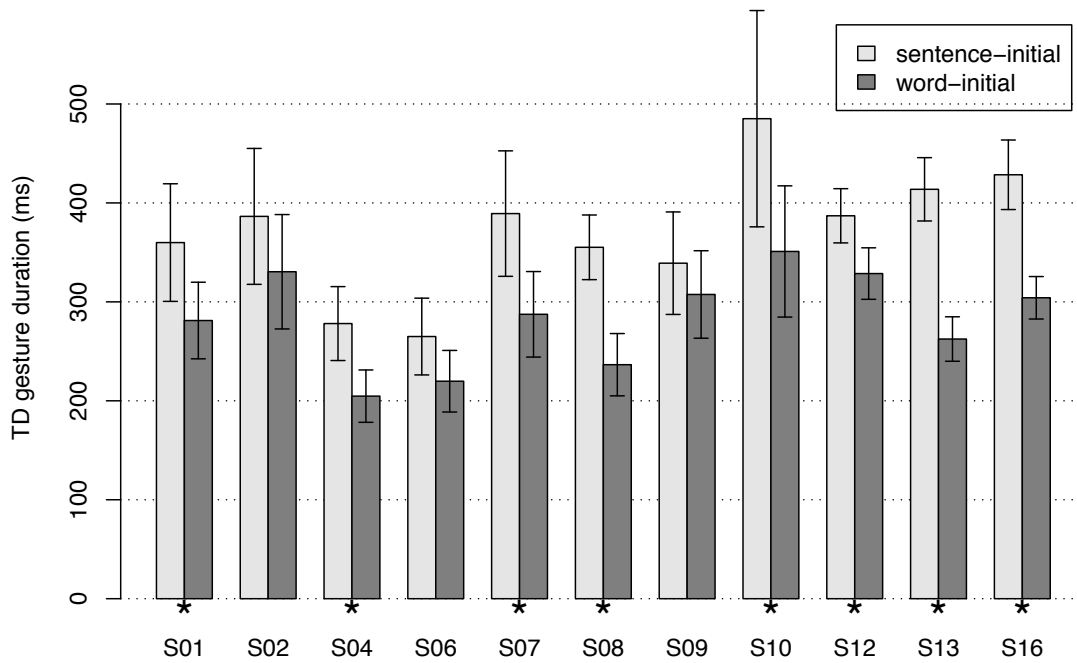


Figure 2.15: Duration of tongue dorsum gesture in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.

Duration of Tongue Tip Gesture in Initial-/l/

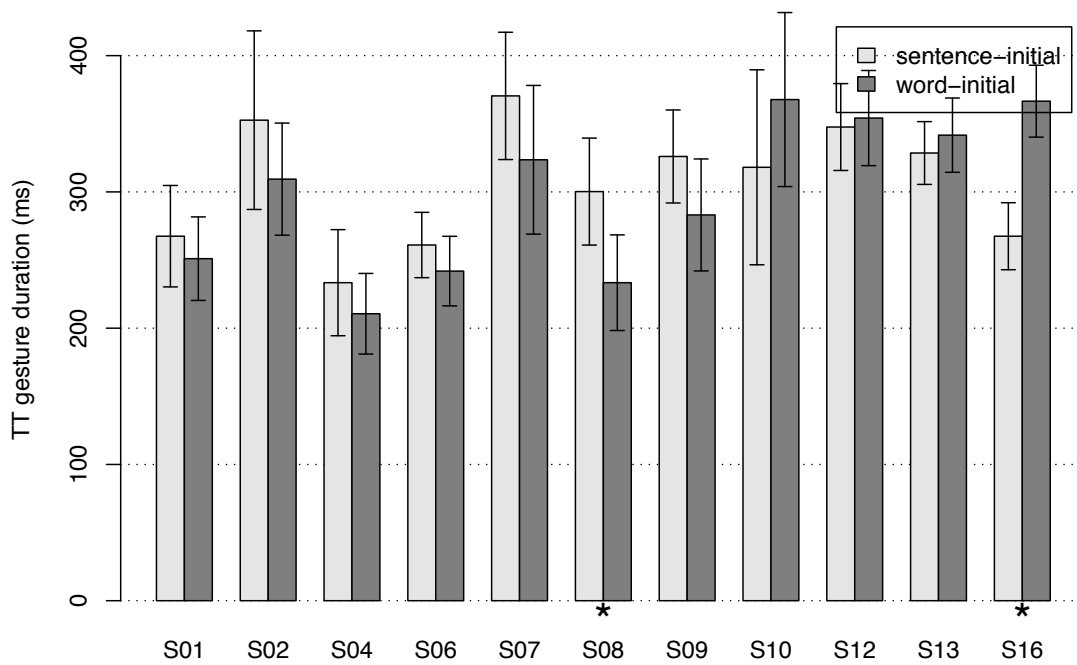


Figure 2.16: Duration of tongue tip gesture in syllable-initial position, by prosodic position and speaker; asterisks signify a significant main effect of prosody, at $p < 0.05$. Positive lag indicates TT gesture precedes TD gesture.

2.3.6 Discussion

Coda laterals

Final laterals were included in this study to verify that the experimental procedures would yield findings compatible with the existing literature. As discussed above in section §1.1, several previous studies (Sproat and Fujimura, 1993; Browman and Goldstein, 1995; Gick et al., 2006) have shown that speakers produce American English coda laterals with tongue tip gestures following tongue dorsum gestures, and that the gestural lag is greater when adjacent to stronger prosodic boundaries compared to weaker prosodic boundaries. This expectation was generally upheld by the productions of speakers in this study. Additionally, most speakers produced longer gestures, and had longer overall articulatory motion during lateral production, in sentence-final laterals than in word-final laterals. These results are summarized below in Table 2.5.

Speaker ID	Duration	Lag	TD	TT
S02, S04, S06, S07, S10, S12, S13, S16	X	X	X	X
S09	X	X		X
S01	X		X	X
S08				

Table 2.5: Coda lateral results; variables affected significantly by prosodic position as predicted are marked with “X”.

There are two items of particular note in the results from coda laterals. First, data from speaker S08 showed no significant effect of prosodic position on any of the four measurements taken. Second, for data from speaker S09, tongue tip gestures were longer during sentence-final laterals than word-final laterals, but tongue dorsum gestures were not significantly affected. This particular result is consistent with the prediction by Krivokapić and Byrd (2010) that prosodic boundaries will have a stronger affect on the most adjacent gestures though, notably, other speakers do not show this pattern. In the case of laterals, their approach predicts that if only one of the two gestures is affected by a sentence-final

prosodic boundary, that gesture should be the tongue tip gesture rather than the vowel-adjacent tongue dorsum gesture.

Onset laterals

Based on the findings of previous researchers, and on the predictions of all variations of Hypothesis 2, speakers are expected to produce initial laterals with tongue tip gesture preceding tongue dorsum gesture, or with synchronous gestures. This was largely found to be true for the speakers in this experiment. As can be seen in Figure 2.13, gestural offset was always positive or close to 0 for all speakers.

Recall that a speaker-motivated model of gestural coordination (the jaw-cycle and π -gesture hypotheses) predicts that sentence-initial laterals should be produced with greater gestural asynchrony than word-initial laterals. On the other hand, a listener-oriented model predicts that speakers will strive towards gestural synchrony in onsets, especially at informationally important locations in the utterance, such as at the onsets of higher prosodic levels. As shown in Table 2.6, speakers in this study can be roughly divided into three groups, with respect to the effects of prosody on onset laterals.

- Group 1 (S04, S07, S08, S13): Sentence-initial laterals were longer than word-initial laterals, but prosodic position had no effect on gestural lag. Note however that, as can be seen in Figure 2.13, and as is schematized in Figure 2.17, the onset laterals produced by speaker S13 exhibited lag not significantly different from zero (one-sample t-test; $t(44) = 1.46, p = 0.1530$), unlike the other speakers in this group. This distinction will be discussed further in Chapter IV.
- Group 2 (S10, S12, S16): Sentence-initial laterals were not longer than word-initial laterals, but prosodic position did have an effect on gestural lag, such that lag in word-initial laterals was greater than lag in sentence-initial laterals.

- Group 3 (S01, S02, S06, S09): Neither duration nor gestural lag was affected by prosodic position.

Group	Speaker ID	Duration	Lag	TD	TT
Group 1	S04, S07, S13	X		X	
	S08	X		X	X
Group 2	S10, S12		O	X	
	S16		O	X	O
Group 3	S02, S06, S09				
	S01			X	

Table 2.6: Initial laterals: variables with significantly greater values in sentence-initial context than in word-initial context marked by “X,” variables with significantly lower values in sentence-initial context marked by “O.”

Speakers from Group 3 simply may not produce sentence-initial laterals differently than word-initial laterals. Since neither gestural offset nor segmental duration was affected by prosodic position, the results from this group show that, at least in this experiment setting, prosodic context does not have a significant effect on durational aspects of onset laterals in the productions of all speakers.

Results from speakers in Group 1 don’t fit the listener-oriented predictions, as only speaker S13 shows true inter-gestural synchrony. For the other three speakers in this group, mean offset in initial laterals is greater than 33 ms. These speakers did show a significant effect of prosodic structure on the total duration of lateral articulation, which was due primarily to lengthening of the tongue dorsum gesture coupled with a lack of change in either tongue tip gesture duration or the relative timing between the gestures. Thus, the tongue dorsum gesture lengthens in the direction of the following vowel, as shown in Figure 2.17.

Keeping in mind that the production of all speakers support the prediction of both speaker- and listener-motivated models that gestural lag in initial laterals will be either zero or positive, the effects of prosodic structure on the productions of Group 2 speakers support the perceptual hypothesis. These results suggest that these speakers may have

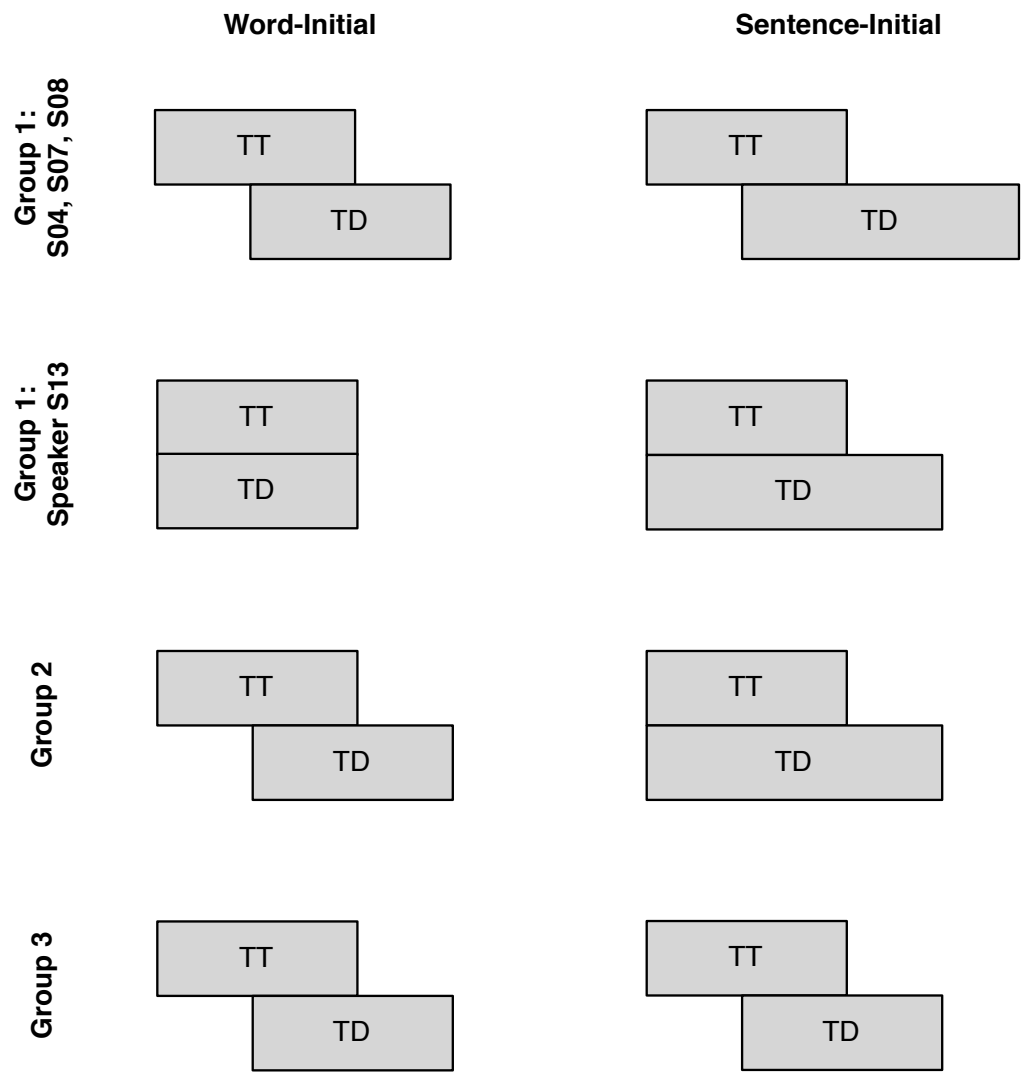


Figure 2.17: Schematization of results from Experiment 1.

different motivations at different prosodic positions: speaker-motivated in word-initial laterals, but listener-oriented in sentence-initial laterals. This is consistent with the model of domain-initial strengthening discussed in section §1.1.2, as posited by Keating (2006), in which speakers “strengthen” segments at the onset of higher prosodic domains to increase perceptibility. Under this model, speakers should be more likely to exhibit listener-oriented behavior at the onsets of higher prosodic boundaries. As shown in Figure 2.17, speakers appear to be achieving this effect by extending the tongue dorsum gesture towards the onset of the syllable, rather than its nucleus.

These results and their interpretation under the hypotheses presented in this dissertation will be discussed in further detail in Chapter IV.

CHAPTER III

Experiment 2: Perception

The goal of Experiment 2 is to address the claim proposed by Gick et al. (2006), and extended here, that gestural synchrony in onset laterals is useful for listeners during perception. In order to test this claim, this study presents listeners with a modified phoneme monitoring task, intended to test their ability to recognize onset laterals and to measure the speed at which they are registered as laterals. What might synchrony between the tongue tip and tongue dorsum gesture in onset laterals contribute to the acoustics, and in what way is it beneficial to listeners? The primary claim of the perceptual hypothesis, as outlined in section §1.3, is that if listeners receive perceptual information about the component gestures of a segment early, they should be able to identify the segment or lexical item earlier as well. Such perceptual information must be encoded in the acoustics, and the main acoustic correlates of onset laterals in American English include a relatively low F2 frequency, a sharp F1 rise at the transition out of the lateral into the following vowel, and an antiformant between F2 and F3 (Stevens, 1998; Johnson, 2003). Additionally, English laterals also typically have a relatively high F3, which helps differentiate /l/ from the rhotic /ɫ/, which is characterized by F1 and F2 frequencies similar to that found in /l/, but a low F3. The accessibility of these acoustic cues in the stimuli used in this experiment will be discussed in more detail below, in section §3.1.1.

3.1 Design

3.1.1 Stimuli

All stimuli were generated using a Matlab based articulatory synthesizer, Task-Dynamic Application (TADA) (Nam et al., 2010), developed at Haskins Laboratories. This system creates audio files based on the acoustic consequences of vocal tract configurations over time, allowing for control over the temporal and spatial characteristics of gestures in 10 millisecond increments. The user may input to the system an English word, in English orthography. The synthesizer then generates a phonetic transcription, based on an English dictionary, and generates a gesture file containing a list of articulators (e.g. tongue tip, lips, etc.) and movement details. An example of text from a file for “lap” is given here, in Table 3.1. In this file, lines (2) and (3) specify that the articulator ‘LA,’ lip aperture, changes its state from closed (-2 mm aperture¹), to open (11 mm aperture). In other words, this is a complete lip closure gesture, including labial release. The onset and offset of the gestures is encoded in the second numerical value of each line, in 10 ms increments. Thus, onset of this lip gesture will occur at 420 ms from the start of the stimulus, and the onset of the release will occur 80 ms later, at 500 ms from the start of the stimulus. In the stimuli created for this experiment, variation across members of the continuum occurs exclusively in lines 13-16, which specify the tongue tip gesture².

The synthesizer uses the gestural score file to generate variables required for the Hlsyn, a “quasiarticulatory synthesizer,” which uses a set of 13 physiological variables to control a Klatt synthesizer. Table 3.2 shows these 13 variables, and sample values from “lap” with 20 ms gestural lag. Of particular interest to this study are the variables “f1,” “f2,” and “f3,” which refer to the first, second and third formants, respectively. From this information, an

¹Negative values of aperture are used by the TADA system for full closure, to reflect the malleability of most vocal tract organs. The more negative the value, the “stronger” the closure, and the longer it will take for the closure release to overcome the inertia of closure.

²The ‘TTCL’ variable stands for “tongue tip constriction location,” and defines place of constriction. The ‘TTCD’ variable stands for “tongue tip constriction degree,” and defines the strength of the closure. The previously discussed labial closure involves only one variable (‘LA’) because “constriction location” is not a valid variable for labial closure.


```

2) 'LA' 0 42 50 0 -2 8 1 JA=8,UH=5,LH=1 100 0.01
3) 'LA' 0 50 53 0 11 8 1 JA=8,UH=5,LH=1 1 1
4) 'TBCL' 0 10 19 0 125 8 1 JA=10,CL=1,CA=1 10 0.1
5) 'TBCL' 0 10 40 0 190 4 1 JA=1,CL=1,CA=1 1 1
6) 'TBCL' 0 35 44 0 190 8 1 JA=10,CL=1,CA=1 100 0.01
7) 'TBCL' 0 10 19 0 4 8 1 JA=10,CL=1,CA=1 100 0.01
8) 'TBCL' 0 10 40 0 14 4 1 JA=1,CL=1,CA=1 1 1
9) 'TBCL' 0 35 44 0 14 8 1 JA=10,CL=1,CA=1 100 0.01
11) 'GLO' 0 44 53 0 0.4 16 1 GW=1 0 0
12) 'GLO' 0 0 13 0 0.4 16 1 GW=1 0 0
13) 'TTCL' 0 8 17 0 56 8 1 JA=32,CL=32,CA=32,TL=1,TA=1 1 1
14) 'TTCL' 0 17 21 0 24 8 1 JA=512,CL=512,CA=512,TL=1,TA=1 1 1
15) 'TTCD' 0 17 21 0 11 8 1 JA=512,CL=512,CA=512,TL=1,TA=1 1 1
16) 'TTCD' 0 8 17 0 2 8 1 JA=32,CL=32,CA=32,TL=1,TA=1 1 1

```

Table 3.1: Selected text from a gestural score (.G) file, from “lap” with 20 ms gestural lag.

```

ag 220 4
al 220 100
ab 220 100
an 220 0
ue 220 0
f0 220 1684.1455
f1 220 474.7173
f2 220 1521.3382
f3 220 2842.9034
f4 220 3331.1693
ps 220 8
dc 220 0
ap 220 0

```

Table 3.2: Sample text from a Hlsyn (.HL) file, from “lap” with 20 ms gestural lag. This set of variables provides information for timestamp 220 ms.

audio file is generated by an implementation of the Klatt synthesizer.

Figures 3.1-3.3 show the gestural score from TADA for tongue body and tongue tip motion for three of the target stimuli in this study, as well as spectrograms from the generated audio. The three vocal tract configurations are taken (from left to right) from the onset of voicing, the beginning of the tongue dorsum plateau³, and the end of the tongue dorsum plateau. Note that the gestural score is scored as aperture, so smaller values indicate a tighter constriction.

³In the case of “lap” with synchronous gestures, this coincides with the achievement of the tongue tip gesture. While it appears that the tongue tip aperture is still decreasing at this point according to the gestural score, the vocal tract configurations beyond this point reveal no discernible motion of the tongue tip. Notice, too, that this is the point at which the slope of the tongue tip gestural score changes from steep to shallow.

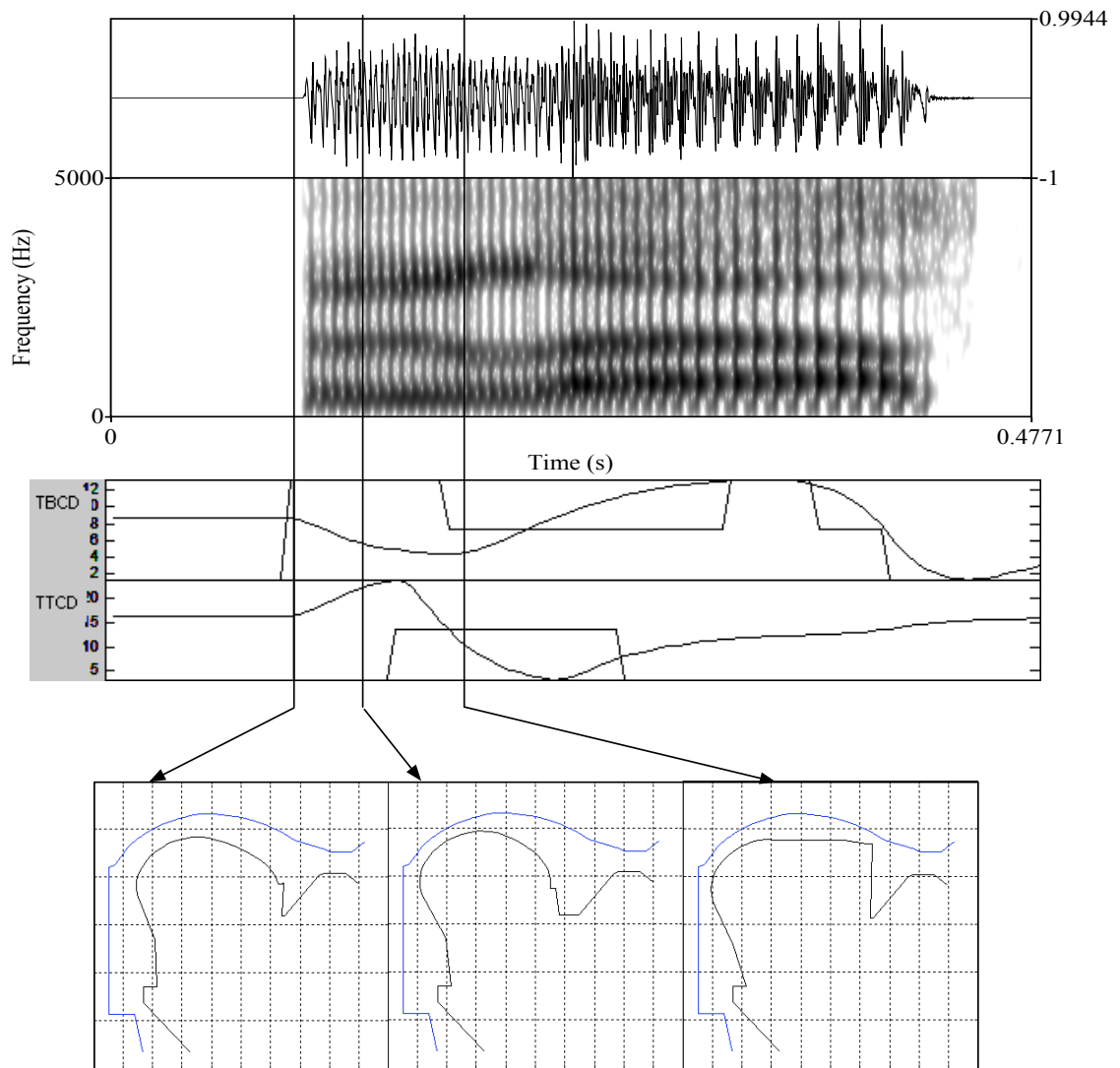


Figure 3.1: Gestural score and generated audio for "lap" with -60 ms gestural lag. See text for details.

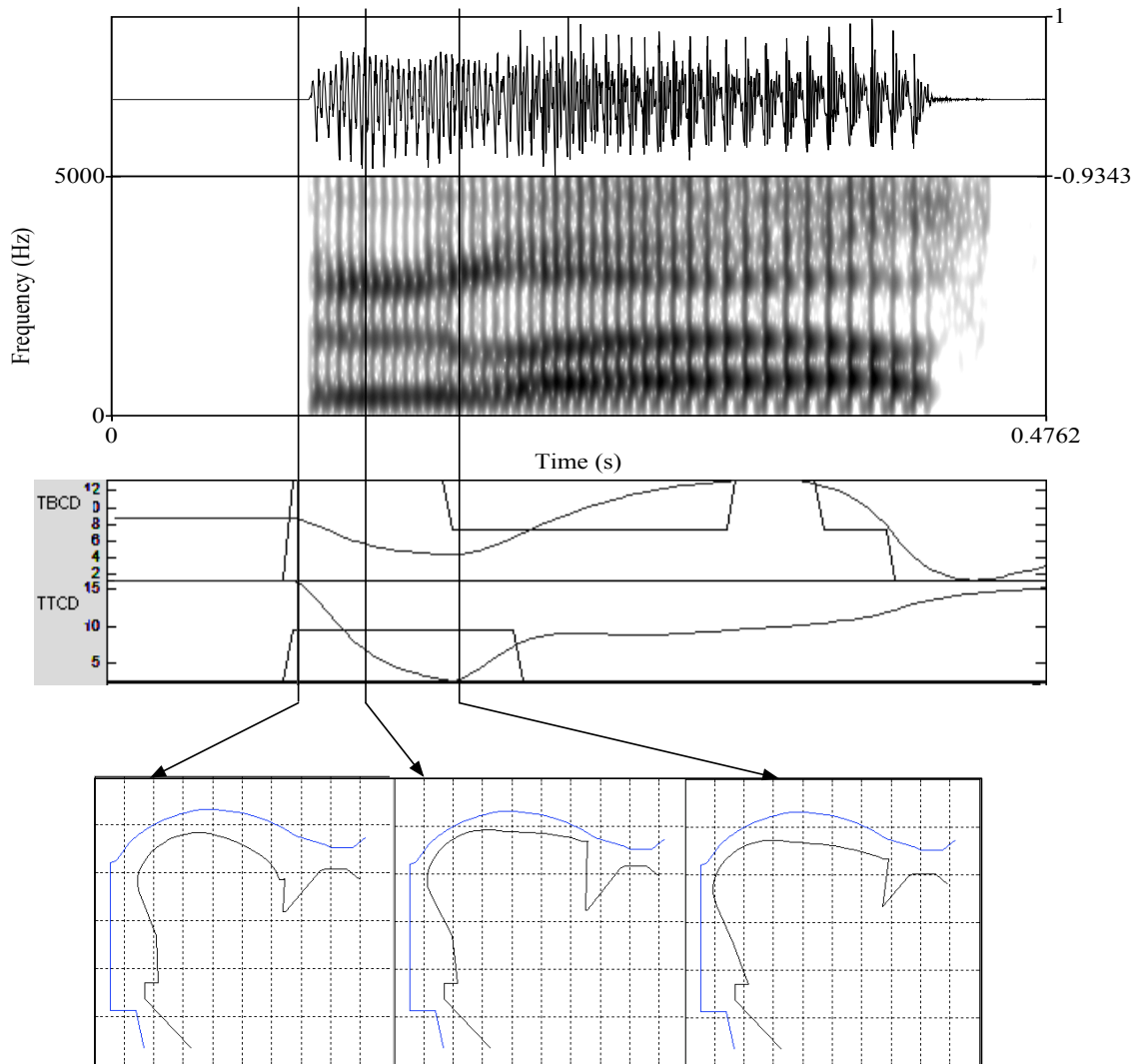


Figure 3.2: Gestural score and generated audio for “lap” with 0 ms gestural lag. See text for details.

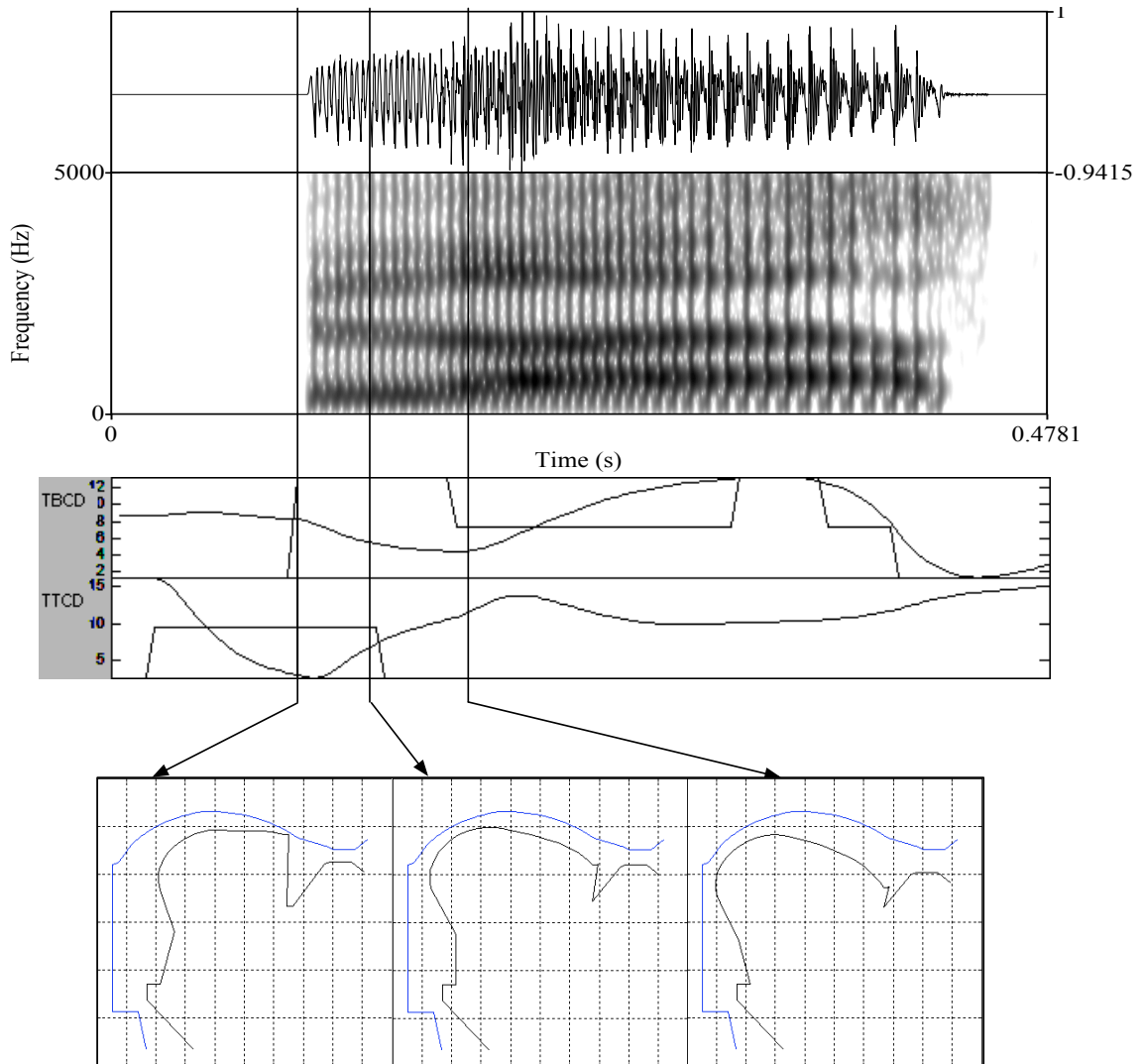


Figure 3.3: Gestural score and generated audio for “lap” with 80 ms gestural lag. See text for details.

TADA has some notable limitations for synthesizing laterals. In particular, gestural targets for most articulators can be specified in only two dimensions, on the mid-sagittal plane of a dummy speaker. Information about constrictions that vary along a coronal slice are not able to be encoded or used by the system. Thus, it is impossible to encode lateral tongue tip gestures as being closed with one side of the tongue depressed, to allow for airflow laterally across the tongue. Instead, laterals must be represented in the gestural score with a narrow, but not closed tongue tip constriction.

Targets and Fillers

Target stimuli for this experiment were eight versions of the word “lap.” The eight versions of the target stimulus were designed to have gestural lag between -60 ms and 80 ms, in 20 ms intervals. As shown in Figure 2.13 (section §2.3.5), in general, gestural lag in the onset laterals produced by speakers in Experiment 1 from this study fell between -50⁴ and 150 ms. Creating laterals in TADA with extreme positive gestural lag while maintaining a fixed relationship between voicing and the onset of the tongue dorsum gesture was not possible without the entirety of the tongue tip gesture, from onset of tongue tip raising to offset of tongue tip lowering, occurring prior to voicing. At 80 ms lag, tongue tip achievement preceded both the onset of tongue dorsum gesture and the onset of voicing by approximately 20 ms. This was chosen to be the positive endpoint of stimuli for this study. Similarly, at -60 ms lag, the onset of the tongue tip gesture coincides with the achievement of the tongue dorsum gesture, and this was chosen as the negative end point for stimuli. The 20 ms step size was chosen on the basis of pilot testing, in which listeners were very accurate in distinguishing between TADA generated “lap” stimuli with 40 ms lag differences, and were moderately accurate in distinguishing between stimuli with 20 ms lag differences. At 10 ms step sizes, all listeners performed at chance. The step

⁴In the negative lag cases, voicing did not begin before initiation of the tongue tip gesture.

Targets	lap
Competing Fillers	rap nap map
Non-competing Fillers	sap snap cap tap

Table 3.3: Stimuli for Experiment 2.

size of 20 ms was chosen as it was the smallest step size possible while retaining some discriminability.

Differences in inter-gestural timing were achieved by modification of the gestural information file. To create a stimulus representing a lateral with synchronous gestures, the tongue tip and tongue dorsum gestures both began simultaneously and achieved their target in the gestural score at the same time. (In TADA, default tongue tip and tongue dorsum gestures in laterals have the same onset-to-achievement time.) Voicing onset was set to the onset of the tongue dorsum gesture⁵. To create stimuli representing laterals with varying gestural lag, the tongue tip gesture was adjusted away from or towards the vowel, in 20 ms increments. Filler stimuli were seven phonologically similar words, listed in Table 3.1.1. The /VC/ portions of the gestural score for these filler stimuli were identical to that of the target stimuli. Three phonological competitors were chosen, with the same rime as the target stimuli and different sonorant onsets. These fillers were intended to make the task modestly more difficult, and it was expected that most false alarms would occur in response to these competing filler stimuli. Four obstruent-initial non-competitors were also chosen.

Acoustics of synthesized stimuli

The original design of this experiment included an acoustic analysis of the audio data collected during the production experiment, and selection of natural stimuli from these recordings as representatives of varying degrees of gestural lag. However, an initial sur-

⁵The default gestural configuration for “lap” generated by TADA consists of synchronous tongue tip and tongue dorsum gestures with the onset of voicing 20 ms prior to the achievement of both gestures. In order to retain the information about the gestures during motion towards their targets, the onset of voicing was adjusted to be earlier.

vey and analysis showed this to be unreasonable for two reasons. First, as a study that focuses on very fine adjustments to the speech organs during production, it is necessary for the stimuli used in the perception experiment to be controlled as much as possible. Changes in pitch, intensity, and other acoustic factors that do not change the identity of the speech sounds may still influence listeners' perception of the stimuli. Additionally, the majority of sentence-initial laterals produced by speakers in Experiment 1 were produced with voicing onset occurring after achievement of the tongue tip gesture, meaning that, for many productions of sentence-initial laterals, the acoustic consequences of tongue tip raising were lost. However, voicing rarely began after tongue tip lowering began. Therefore, in naturally produced sentence-initial laterals, the audible signal typically began with the acoustic consequences of a raised tongue tip. Thus, from the point of view of the listener, the natural productions differed primarily in whether the acoustic effects of the tongue dorsum retraction gesture occurred after or at the same time as the acoustic consequences of the raised tongue tip.

Spectrograms for all target stimuli can be found in Appendix C. Figure 3.4 displays, for all target stimuli, F1, F2 and F3 during the first 200 ms of audible stimulus, as generated in the Hlsyn files (Table 3.2). In this figure, dashed lines represent stimuli with positive lag, dotted lines represent stimuli with negative lag, and the solid line represents the stimulus with zero gestural lag. As in Figures 3.1-3.3, the left-most vertical line in Figure 3.4 marks the onset of voicing⁶, and the remaining vertical lines denote the “plateau” of the tongue dorsum gesture. For the stimulus with synchronous gestures (solid line), the second and third line also delineate the plateau of the tongue tip gesture.

As mentioned earlier in this chapter, among the main acoustic correlates of onset laterals in American English are a relatively low F2 frequency, a sharp F1 rise at the transition

⁶Hlsyn files contain projected formant information regardless of whether or not voicing has begun, but any formants left of the first vertical line are lost due to lack of voicing.

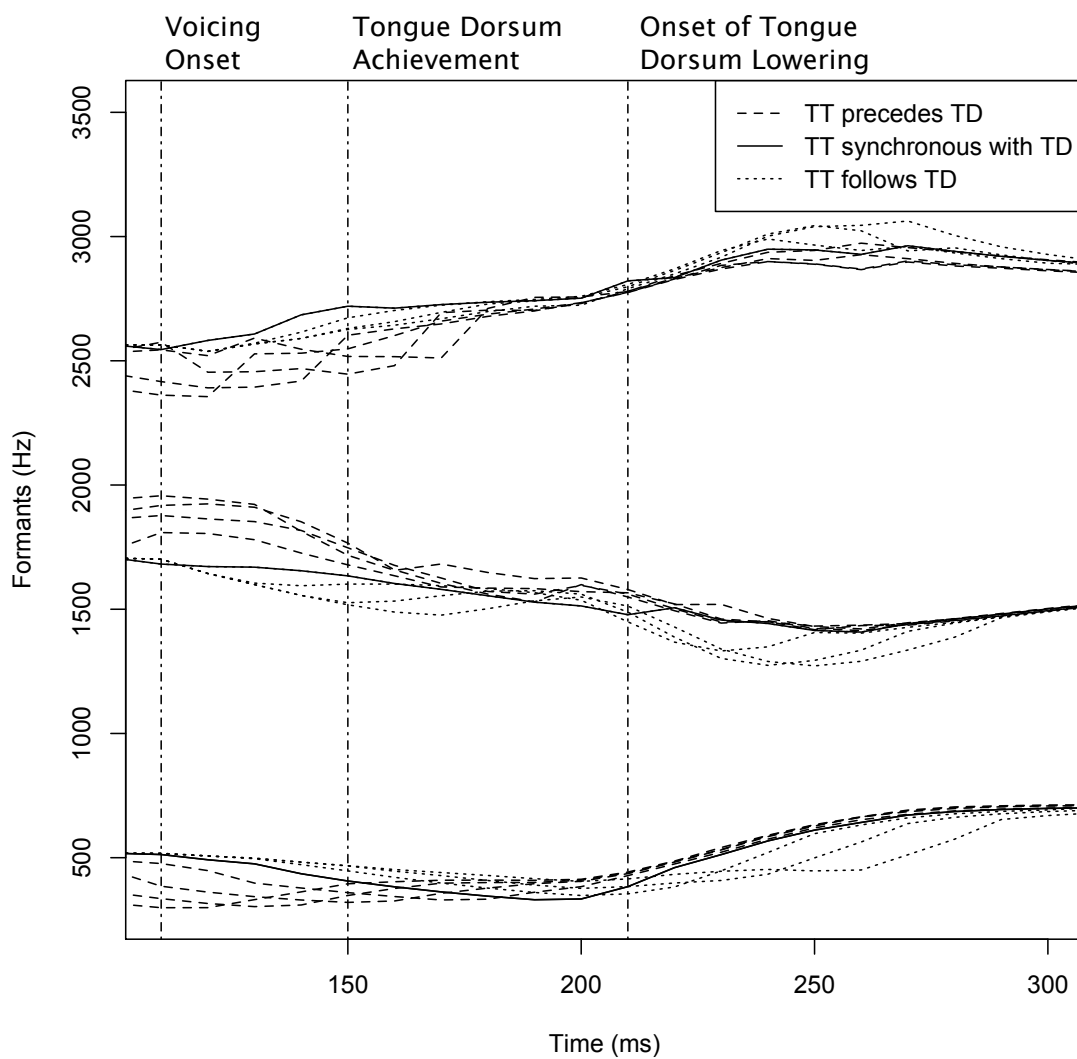


Figure 3.4: F1, F2, and F3 plotted against time for target stimuli.

out of the lateral into the following vowel, and an antiformant between F2 and F3 (Stevens, 1998; Johnson, 2003). The antiformant, which is the result of the side cavity created by the laterality of airflow around the tongue tip, was not replicable by the TADA synthesizer, owing to the limitations described in section §3.1.1. However, both the low F2 frequency and the F1 rise into the following vowel are present in the formants for the stimulus with synchronous gestures, as seen in Figure 3.4 (solid line). The primary way in which the stimuli with positive lag (dashed lines in Figure 3.1.1) diverge from the synchronous stimuli is that F2 frequency began higher and F3 frequency began lower. On the other hand, the stimuli with negative lag (dotted lines) diverge from the synchronous stimuli in the lateness of the F1 rise. Additionally, F2 frequency in negative lag stimuli falls, then rises again within the first 100 ms of the audible signal, which may confuse listeners if low-frequency F2 is a critical indicator of an American English lateral. Thus, for these synthetic stimuli, the acoustic cues typical of American English laterals appear earliest and most clearly in the stimulus with synchronous gestures.

Comparison with naturally produced onset laterals

In light of the artificial nature of the stimuli, it is worth examining the similarity between the onsets of the synthesized stimuli used in this experiment and the naturally produced onset laterals recorded in Experiment 1. One unavoidable difference between the synthetic stimuli and the naturally produced tokens is that the onset laterals in the natural production data were all preceded by the vowel /i/, contributing coarticulatory information to the acoustics of the naturally produced laterals that is not present in the synthetic stimuli, which were created without a preceding vowel context. These coarticulatory effects were absent or minimal in the natural speech onset laterals that were preceded by silence, which was especially common for sentence-initial laterals. However, for this subset of natural stimuli, the onset of voicing, as mentioned previously, often fell between

tongue tip achievement and tongue dorsum achievement, making direct comparisons of the acoustics at lateral onset impossible. For these reasons, the following acoustic comparisons are made at tongue dorsum achievement for onset laterals preceded by silence in which achievement of the tongue dorsum gesture occurred during voicing.

As shown in Figure 3.4, in the synthetic stimuli, at the point of tongue dorsum achievement, F1 frequencies decrease from negative to positive lag, F2 values increase, and F3 values decrease (except that the stimulus with zero lag has the highest F3 frequency at this point). Figure 3.5 displays the first three formants taken at the point of tongue dorsum achievement from the subset of onset laterals identified above produced by speakers participating in Experiment 1. The selected onset laterals were those that were preceded by silence, and in which achievement of the tongue dorsum gesture occurred during voicing. The formant values were then plotted against gestural lag, as measured for that token, and linear regression models were fit to F1, F2 and F3 formants against lag values. The F1 values were not well modeled linearly ($p = 0.8597$), but F2 and F3 values were. As can be seen in Figure 3.5, F2 values in natural productions did increase from negative to positive lag, and the linear model was significant ($p = 0.0028$; $r^2 = 0.0511$). Additionally, F3 values in the naturally produced tokens decreased from negative to positive lag, although the model was not significant ($p = 0.06133$; $r^2 = 0.0164$). Though these regressions produced relatively low r^2 values, which may be in large part due to inter-speaker variation. In fact, when speaker ID was added as a factor to the linear regressions, r^2 values substantially increased for all three formants: F1 ($p < 0.0001$; $r^2 = 0.1732$), F2 ($p < 0.0001$; $r^2 = 0.2194$) and F3 ($p < 0.0001$; $r^2 = 0.2823$).

3.1.2 Task

The task presented to the listeners was a single target multiple location phoneme monitoring task. During phoneme monitoring tasks, listeners listen to a speech stream consist-

F1, F2, F3 vs. Gestural Lag

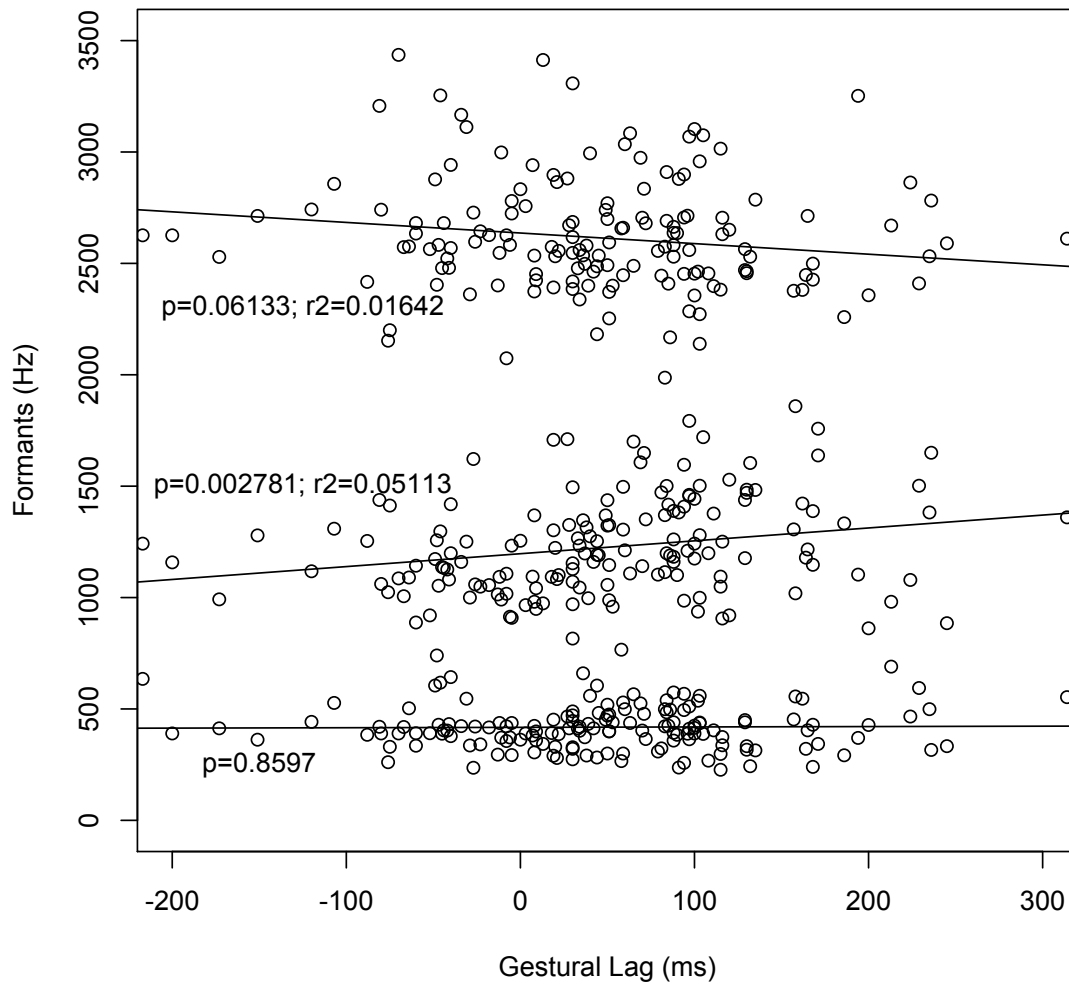


Figure 3.5: First three formants taken at point of tongue dorsum achievement from selected tokens produced by speakers during Experiment 1.

ing of words and sometimes non-words, and respond when they hear a word with a target phoneme, typically on a response box or keyboard (Connine and Titone, 1996). A single target, multiple location phoneme monitoring task is one in which only one phoneme is being listened for, and that phoneme can occur multiple times during the task. In this experiment, each of the eight target stimuli was presented to the listener 15 times, and each filler stimulus was presented 120 times. Thus, each listener was presented with 120 instances of each filler item, and 120 total target stimuli, for a total of $120 \times 8 = 960$ trial items. Participants were presented stimuli over headphones, from a laptop running SuperLab, in a sound-attenuated booth. Responses were recorded through SuperLab, using a response box. Stimuli were blocked and pseudo-randomized, such that at least two filler stimuli separated any two target stimuli. Each stimulus .wav file had a 60 ms lead time, and the audible portion of each stimulus file was 350 ms long. Stimuli onsets were separated by 600 ms, giving listeners 250 ms of silence between stimuli. This stimulus presentation rate was selected to be rapid enough to force listeners to make quick decisions, but not so rapid that listeners were incapable of keeping up with the task. Several pilot versions of this experiment with varying inter-stimulus intervals (ISI) were run. Results from the pilots suggested that ISIs of 100 ms or 150 ms were too short, resulting in uniformly low accuracy rates and participants reporting that they felt they performed poorly. ISIs greater than 400 ms were too slow, resulting in uniformly high accuracy and no reported sense of “being rushed” by the participants.

Listeners were informed prior to the experiment that they would be hearing a very fast sequence of similar computer generated English monosyllabic words, some of which may sound more natural than others. They were instructed to listen for the sound “l” anywhere in the word except at the end⁷, and to hit any button on the response box as soon as they

⁷During a pilot version of this experiment, listeners were instructed to listen for “l” at the beginning of the word. Several listeners reported hearing “l” preceded by a short vowel or other sonorant in many of the target stimuli, and therefore did not respond to them.

Stimulus Item	lap	rap	map	nap	sap	snap	cap	tap
Response	a wag	a rat	mat	nap	sap	snap	tap	jab

Table 3.4: Responses from post-experimental stimulus identification for participant S05.

registered it. During the experiment, listeners were given two breaks, and were encouraged to stand and stretch. Each block of stimuli ran 3.5 minutes, and listeners were allowed 2-5 minutes for each break. The task ran between 15 and 20 minutes.

The main task was not preceded by a familiarization task, in order to prevent listeners from acquiring a bias towards any one of the eight target stimuli. Following the task, listeners were presented with each of the filler stimuli and the target stimulus with synchronous gestures, and were prompted to type in a free-response box what English word they thought they heard. Listeners were instructed to guess, if they were not sure. Results from this task were used to disqualify listeners who were unable to identify the intended stimulus onsets.

3.1.3 Participants

The participants in this experiment were 15 monolingual⁸ speakers of American English with no known speech or hearing difficulties. The participants were all undergraduate students from the University of Michigan, and were each paid \$10 for their participation.

3.2 Analysis & Results

Of the fifteen participants, data from one listener (S05) were discarded due to inability to reliably identify onset consonants in the synthesized stimuli. Table 3.4 shows this speakers' responses during the free-response survey following the experiment. Data from two other speakers were discarded because of confusion regarding the instructions, resulting in artificially low accuracy. Thus, data from 12 participants were used in the following analysis.

⁸Some listeners had taken foreign language courses at the high school and college level, but none claimed fluency.

Recall from Chapter I, section §1.3, Hypothesis 3:

Hypothesis 3 Initial laterals are most quickly identified when tongue tip and tongue dorsum movement are synchronous.

This hypothesis should manifest in a reduction in reaction time: if onset laterals with synchronous gestures are more easily distinguishable from other onset segments, we should expect a reduction of reaction time for those laterals when compared with laterals with positive or negative gestural lag. Additionally, due to the nature of the task, stimuli towards the endpoints of the continuum (i.e. stimuli with gestural lag) may also be identified less accurately than stimuli with synchronous or near-synchronous gestures. If onset laterals with synchronous or near-synchronous gestures are more easily distinguishable from other onset segments, we should expect a peak in identification or accuracy in response to these laterals compared to laterals with positive or negative gestural lag. Given that gestural synchrony falls roughly in the center of the continuum, we would expect a second degree polynomial (a parabola) to provide a good fit for the relationship between gestural lag and reaction time or accuracy.

3.2.1 Measurements

SuperLab outputs text files like that given in Table 3.5⁹. Response times were recorded from the onset of the stimulus and, when no response was registered, a 0 was recorded.

Responses recorded were processed in the following manner:

1. If a response was registered and the response time registered was less than 360 ms (300 ms from acoustic onset of the stimulus), it was assumed that the button press registered was in response to the previous stimulus. Botwinick and Thompson (1966) showed that motor time, the amount of time between muscle activation and button

⁹Some uninformative columns were removed for clarity.

Histogram of Reaction Time (all participants)

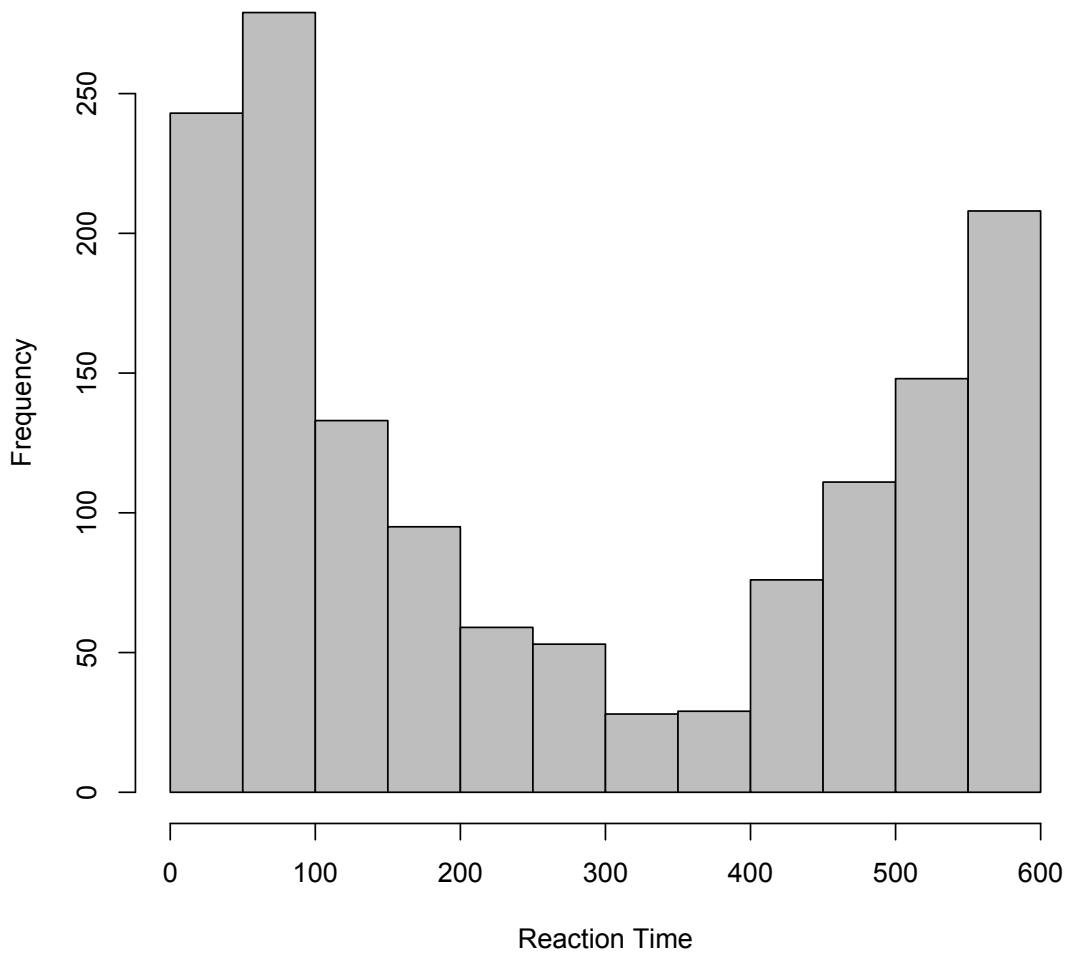


Figure 3.6: Histogram of reaction times from Experiment 2, from all participants. Bins are 50 ms in size.

Listener	Block	Item	Button Pressed	Response Time
S2	1	map	(no response)	0
S2	1	sap	(no response)	0
S2	1	lap-60	Button3	469
S2	1	nap	(no response)	0
S2	1	cap	(no response)	0
S2	1	map	(no response)	0
S2	1	cap	(no response)	0
S2	1	lap-40	(no response)	0
S2	1	map	Button 3	155
S2	1	snap	(no response)	0
S2	3	map	(no response)	0
S2	3	cap	(no response)	0
S2	3	lap40	(no response)	0
S2	3	map	Button3	567
S2	3	tap	(no response)	0

Table 3.5: Sample raw data from listener S02.

press, to be approximately 40 ms and constant, while premotor time (time between onset of stimulus and muscle activation) was found to be variable, but consistently between 200 and 250 ms. Creating a histogram of raw reaction time values, as shown in Figure 3.6¹⁰, reveals substantially fewer responses in the 300-400 ms range. I inferred that typical listener reaction times were between 400 and 900 (600 + 300) ms.

2. If a response is registered and the reaction time registered was greater than or equal to 360 ms, it was assumed that the button press registered was in response to the current stimulus. Thus, the button presses in Table 3.5 top and middle were counted for stimuli “lap -60” and “lap -40,” respectively, while the button press at the bottom was counted for “map,” rather than “lap 40.”

Reaction time

Reaction time, as reported by SuperLab, was measured from the onset of stimulus presentation. As described previously, each stimulus file began with 60 ms of silence, which was subtracted from the reaction time values. Because the differences in reaction

¹⁰The histogram uses 50 ms bins, resulting in a total of 12 bars

Listener	rap	map	nap	sap	snap	cap	tap	Total	False Alarm %
S02	1	0	3	0	0	0	0	4	0.48
S03	3	0	0	0	0	0	0	3	0.36
S04	6	4	6	0	0	0	2	18	2.14
S06	2	4	4	1	1	0	4	16	1.90
S07	5	14	10	4	4	0	0	37	4.40
S08	11	13	14	5	1	0	0	34	4.05
S10	1	0	1	0	0	0	0	2	0.24
S11	4	15	9	0	0	0	0	28	3.33
S12	18	12	21	6	13	5	6	81	9.64
S13	12	2	7	0	0	0	0	21	2.50
S14	9	4	9	0	1	1	2	26	3.10
S15	13	5	12	1	1	0	0	32	3.81
Overall	85	73	96	17	20	6	14	302	2.996
False Alarm %	5.90	5.07	6.67	1.18	1.39	0.42	0.97	2.996	

Table 3.6: Number of false alarms by stimulus and listener out of 120 possible, with overall false alarm rates (%).

Listener	-60	-40	-20	0	20	40	60	80	Total	Correct %
S02	13	13	14	14	15	13	10	12	104	86.67
S03	15	15	14	14	14	15	14	15	116	96.67
S04	9	8	10	10	13	8	4	7	69	57.50
S06	6	8	8	7	9	9	7	4	58	48.33
S07	15	13	12	14	13	14	15	15	111	92.50
S08	10	10	13	13	13	15	10	13	97	80.83
S10	10	9	11	11	11	12	10	10	84	70.00
S11	13	14	13	12	15	14	13	14	108	90.00
S12	13	13	13	14	15	12	13	12	105	87.50
S13	11	10	11	12	13	11	10	10	88	73.33
S14	9	10	11	12	10	7	9	9	77	64.17
S15	10	13	14	11	13	13	11	14	99	82.50
Overall	134	136	144	144	154	143	126	135	1116	77.50
Correct %	74.44	75.56	80.00	80.00	85.56	79.44	70.00	75.00	77.50	

Table 3.7: Number of correct “I” responses out of 15 possible, by listener and target item.

time in this experiment are expected to be small, and may therefore be lost in noise due to inter-listener latency variation, mean reaction time was calculated for all speakers, and reaction time values were normalized for each listener by subtracting the given speaker’s mean.

Accuracy & D'

In order to obtain a reliable measure of accuracy, participant response bias needs to be taken into account. Within signal detection theory (Macmillan and Creelman, 2005), the

Listener	Correct %	False Alarm %	d'	Group
S02	86.67	0.48	3.701	high
S03	96.67	0.36	4.522	high
S07	92.50	4.40	3.146	high
S10	70.00	0.24	3.345	high
S11	90.00	3.33	3.116	high
S04	57.50	2.14	2.220	mid
S06	48.33	1.90	2.033	mid
S08	80.83	4.05	2.616	mid
S12	87.50	9.64	2.453	mid
S13	73.33	2.50	2.583	mid
S14	64.17	3.10	2.229	mid
S15	82.50	3.81	2.707	mid

Table 3.8: Average d' scores by speaker.

measure d' is a measure of accuracy considered to be independent of the listeners' decision criterion. This value is calculated as $d' = z(H) - z(F)$, where H is hit percentage and F is false alarm percentage. For the data in this experiment, the false alarm rate was defined to be the percentage of filler tokens that listeners reported hearing as //l/. Table 3.6 lists the number of false alarm button presses that the participants in this study produced in response to each filler, while Table 3.7 lists the hits, or correct button presses. As shown in Table 3.6, the filler stimuli designed to be competitors with the target stimuli (“rap,” “map,” and “nap”) were in fact more often confused with laterals than non-competitor stimuli. However, none of the competitor stimuli stands out as exhibiting unusually high false alarm rates. For each listener, d' was calculated separately for each set of target stimuli; the mean d' scores are listed in Table 3.8. The split between speakers into “high” and “mid” accuracy categories is discussed below.

3.2.2 Results

Figure 3.7 shows d' values against gestural lag in target stimuli, for all speakers. The dotted line represents the best quadratic fit for these data. A repeated measures ANOVA, with listener as the Error factor, showed that gestural lag is not a significant predictor for accuracy, either modeled linearly ($F(1, 11) = 0.47; p = 0.5077$) or by a quadratic

D' values vs. Gestural lag (all subjects)

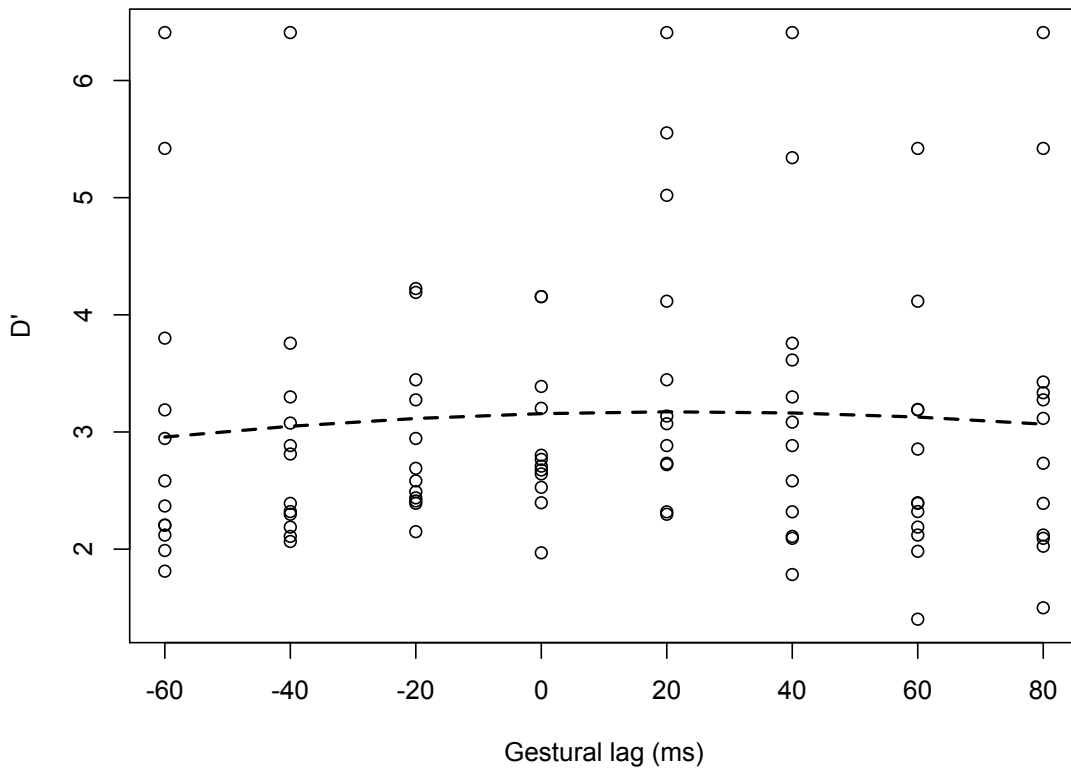


Figure 3.7: D' by gestural lag, all speakers, and quadratic best fit.

($F(1, 71^{11}) = 0.45; p = 0.5055$). Overall, when taking all listeners into consideration, gestural synchrony did not improve accuracy during the task.

Figure 3.8 shows normalized reaction time against gestural lag in target stimuli, for all speakers. Unlike the d' figures, however, a repeated measures ANOVA with listener as Error factor showed that gestural lag is a significant predictor for accuracy, under a quadratic model ($F(1, 1039) = 6.34; p = 0.0120$). A quadratic regression confirmed that the best quadratic fit (shown in Figure 3.8 as a dotted line) is significant but slight ($F(2, 1061) = 3.46; p = 0.0317; r^2 = 0.0046$).

Closer examination of the data suggested that there may be a natural division between

¹¹Here, and elsewhere, when examining a parabolic fit, denominator degrees of freedom is exceptionally high due to squaring: $\sqrt{71} = 8.43$.

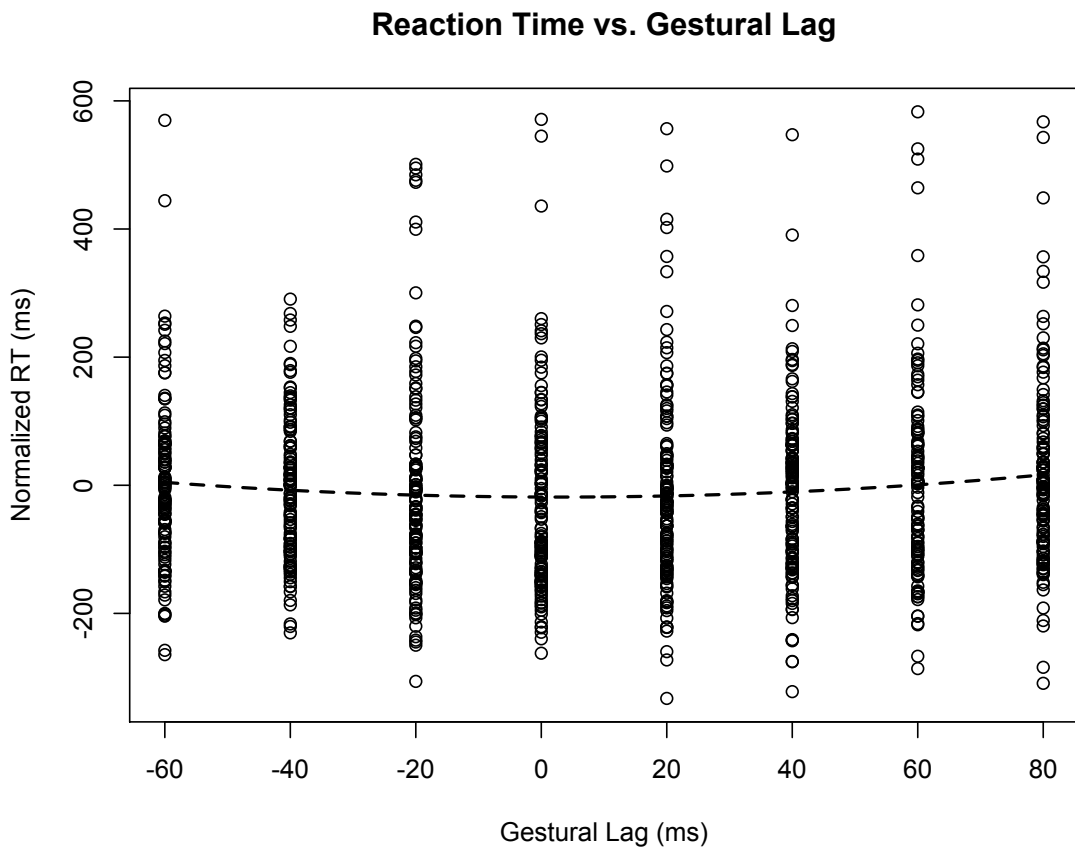


Figure 3.8: Reaction time (normalized) by gestural lag, all speakers, and quadratic best fit.

listeners, based on their accuracy, or d' values. “High” accuracy listeners were defined to be those with an average d' score of above 3, while those with average d' scores between 1.5 and 3 were defined to be “mid” accuracy listeners¹² (See Table 3.8). Inclusion of “group” as a second factor vastly improves the fit for the quadratic model, for both accuracy and reaction time.

Quadratic regression for d' scores against lag, with group as a factor, revealed an increase in fit for the model ($r^2 = 0.3921$) which was significant ($F(5, 90) = 13.26; p < 0.0001$). This analysis also revealed a significant effect of group on accuracy ($F(1, 90) = 62.20; p < 0.0001$), as well as a significant interaction between group and the square of gestural lag ($F(1, 90) = 4.22; p = 0.0429$). This outcome indicates that listeners in the two groups behave differently with respect to the effects of gestural lag on accuracy. Figure 3.9 plots d' scores against gestural lag, for both high- and mid-accuracy groups. For high-accuracy listeners, the quadratic model is not significant ($F(2, 53) = 0.18; p = 0.8400$), but it is significant for mid-accuracy listeners ($F(2, 37) = 4.83; p = 0.0137; r^2 = 0.1642$).

Similarly, using groups as a factor in the quadratic regression for reaction time against lag showed an improvement in significance of the model ($F(5, 1058) = 3.51; p = 0.0038$) as well as an improvement in predictability ($r^2 = 0.0117$). As before, group was found to have a significant effect on reaction time ($F(1, 1038) = 6.38; p = 0.0117$), and a significant interaction between group and the square of gestural lag was found ($F(1, 1038) = 7.22; p = 0.0073$), again indicating that listeners belonging to the two groups behave differently. Here, data from the high-accuracy listeners did significantly conform to a quadratic model ($F(2, 692) = 8.98; p = 0.0001; r^2 = 0.0224$), while data from the mid-accuracy listeners did not ($F(2, 366) = 0.61; p = 0.5418$). Figure 3.10 shows reaction time values against gestural lag for high- and mid-accuracy listeners (top

¹²“Low” accuracy listeners were defined to be those with d' scores below 1.5, though these listeners were also rejected due to misunderstandings of the experimental instructions revealed during post-experimental interview.

D' values by Gestural lag

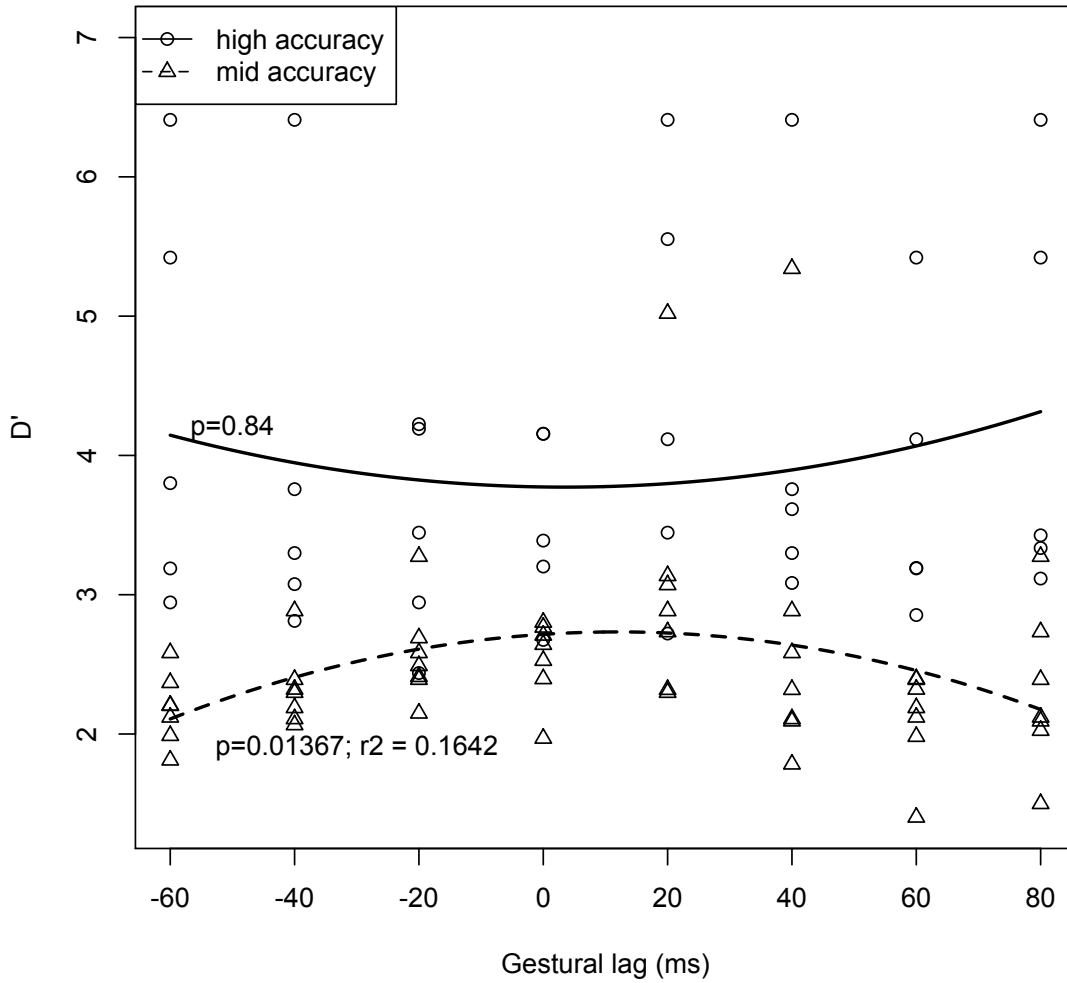
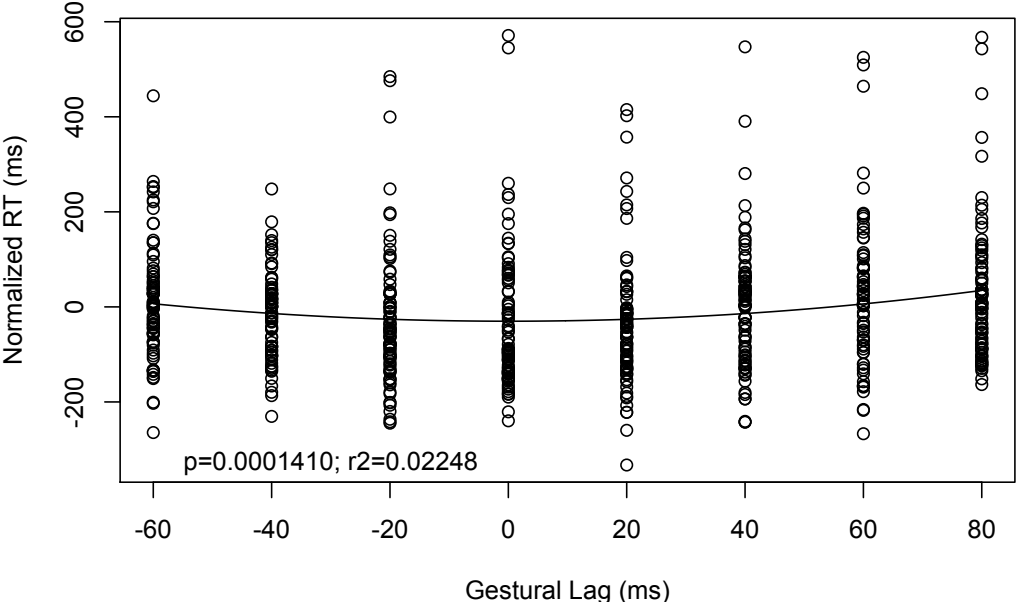


Figure 3.9: D' by gestural lag for high-accuracy listeners and mid-accuracy listeners.

Reaction Time vs. Gestural Lag (High-Accuracy)



Reaction Time vs. Gestural Lag (Mid-Accuracy)

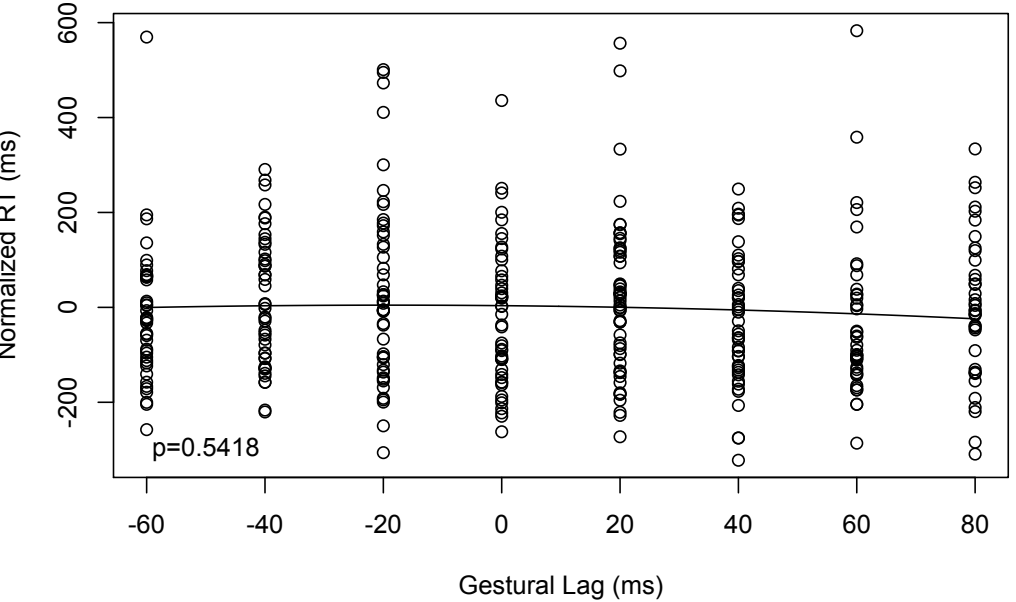


Figure 3.10: Normalized reaction time by gestural lag for high-accuracy listeners (top) and mid-accuracy listeners (bottom).

and bottom, respectively). The best quadratic fits are plotted with the data, in dotted lines.

Thus, responses of high-accuracy listeners have uniformly high accuracy for all stimuli, but lower reaction time for stimuli with gestural lag equal to or close to zero, while responses of mid-accuracy listeners have uniformity in reaction time, but an increase in accuracy for stimuli with gestural lag close to zero. Listeners in the different groups appear to utilize different response strategies, beyond what signal detection theory can correct for. Specifically, mid-accuracy listeners may be more careful with their responses, taking uniformly more time to decide if a stimulus is lateral-like enough to warrant a response or not. Indeed, a post-hoc Welch's *t*-test showed that non-normed reaction time was significantly greater for mid-accuracy listeners than for high-accuracy listeners ($t = -4.037; df = 645.7; p < 0.0001$). These listeners then perceive stimuli with synchronous gestures more frequently as laterals than those with gestural lag. That several of these speakers still made numerous mistakes should not detract from this point. These speakers may, in fact, have been more careful because they were less confident in their judgments. In contrast, high-accuracy listeners may be more confident in their judgments, reacting more quickly and accurately. These listeners appeared also to detect those stimuli with simultaneous or near-simultaneous gestures faster than those with more asynchronous gestures.

3.3 Conclusions

This experiment was intended primarily to substantiate claims that gestural synchrony in onset sonorants, such as laterals, aids their identification. As summarized in Tables 3.9 and 3.10, results from this experiment indicate that synchronous gestures do facilitate perception under the conditions presented to listeners in this experiment, and that this increase in perceptibility is manifested in a reduction of reaction time for some listeners

Group	-60	-40	-20	0	20	40	60	80
High-Accuracy	12.35	-13.38	-33.93	-34.28	-33.85	-0.9316	23.55	20.95
Mid-Accuracy	-19.56	14.28	33.61	-17.84	17.54	-31.42	-32.00	-0.2493

Table 3.9: Mean normalized reaction time (ms), by group and target gestural lag.

Group	-60	-40	-20	0	20	40	60	80
High-Accuracy	4.353	3.871	3.449	3.515	4.449	4.033	3.754	4.341
Mid-Accuracy	2.183	2.322	2.570	2.544	3.066	2.730	2.114	2.305

Table 3.10: Mean d' , by group and target gestural lag.

and an increase in accuracy for others.

What contributes to this increase in listeners' performance? As discussed in section §3.1.1, the primary acoustic correlate of gestural synchrony in onset laterals appears to be comparatively earlier onset of formant structure typical of laterals in American English, in particular, low F2 frequency at lateral onset and a sharp rise in F1 frequency in the lateral-vowel transition. Thus, the results from this experiment indicate that under controlled listening contexts, synchrony between the tongue tip and tongue dorsum gestures in laterals leads to faster or more accurate identification of onset laterals. Furthermore, examination of the acoustic characteristics of the stimuli suggests that the improvement in identification speed and accuracy is due to earlier onset and better match for listeners' expectations of F1 and F2 properties for laterals with gestural synchrony than for those with gestural lag.

CHAPTER IV

Summary and Conclusion

In this chapter, I return to the hypotheses outlined at the end of Chapter I, and investigate how the results from this dissertation may be interpreted within the given models.

4.1 Perception

As was just discussed in section §3.3, listeners in Experiment 2 were capable of correctly identifying synthetic stimuli, and were faster or more accurate at the task when the stimuli were designed with minimal lag between tongue tip and tongue dorsum gestures. These stimuli corresponded to acoustic signals in which the primary perceptual cues for the American English lateral were present earlier in the signal.

In Chapter I, I argued that this may be motivated by listener-oriented productions of the speaker aimed at providing information about both gestures as early as possible, to facilitate perception. The primary acoustic features which distinguish the stimuli with negative lag from the stimulus with synchronous gestures are a delay in the rise in F1 frequency, in increments that equal the step size of the gestural lag continuum, and an unstable F2. The primary distinguishing feature for stimuli with positive lag is a high F2 frequency, followed by a relatively sharp F2 fall as the tongue dorsum retracts. Thus, I propose that the interaction of the somewhat opposing effects of dorsum retraction and tip raising on F2 frequency when the gestures are synchronous serves to create a steadily

low F2. In this way, I argue that producing onset laterals with synchronous gestures is consistent, though not exclusively so (see Footnote 1 on page 79), with the hypothesis of perceptually motivated production.

4.2 Production

Having established that gestural synchrony is perceptually useful in synthetic onset laterals, here I examine how the results from Experiment 1 can be interpreted within the articulatory and perceptual models presented in Chapter I. In this experiment, speakers read a set of sentences, with target /l/-onset words in two prosodic positions. Speakers' tongues during speech were imaged with ultrasound, and tongue tip and tongue dorsum apertures were measured over time. I begin with a restatement of the hypotheses, from Chapter I. Figure 4.1 provides schematic representations of these hypotheses.

Perceptual If speakers are motivated primarily by the perceptual needs of the listener, tongue tip movement and tongue dorsum movement should be roughly simultaneous in onset laterals, and this effect of gestural synchrony may be stronger in higher prosodic positions.

Jaw-Cycle If the jaw cycle governs gestural coordination, tongue tip movement should occur prior to tongue dorsum movement in onset laterals and, analogous to coda /l/s, there should be greater gestural lag following strong prosodic boundaries than laterals following weak boundaries.

π -Gesture Under a π -gesture extension of Articulatory Phonology, tongue tip gestures should precede tongue dorsum gestures in onset laterals, gestural lag in sentence-initial laterals should be greater than in word-initial laterals, and effects of the prosodic boundary (i.e. lengthening, in this study) should be greatest for the tongue tip gesture due to its relative proximity to the π -gesture boundary.

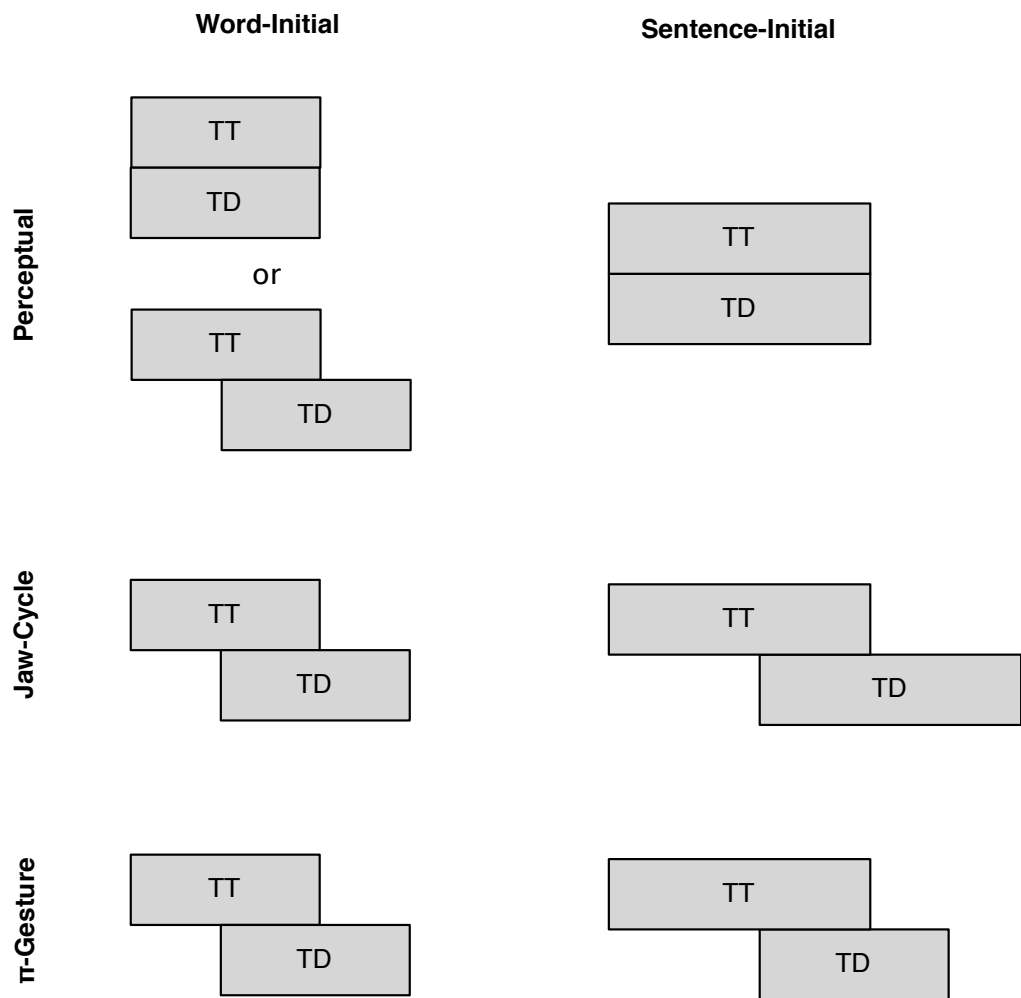


Figure 4.1: Schematization of the hypotheses.

Note that two of the three models of speech production discussed in this dissertation are bio-mechanically motivated: the jaw-cycle hypothesis relies on constraints imposed by the vocal tract, while the π -gesture hypothesis relies on the assumption that activation of speech gestures is governed by the same physical forces that govern activation of other physical gestures. These two models have in common the prediction that, in general, tongue tip gesture achievement should precede tongue dorsum gesture achievement in onset laterals, though the mechanism by which the models arrive at this prediction differs. The perceptual model is the only model considered here that predicts synchronous gestures under either context¹. As shown in the schema, the perceptual model predicts that, if speakers' productions do exhibit gestural lag, it is more likely to occur between gestures in word-initial laterals than in sentence-initial laterals.

Next, I summarize the results from Experiment 1. As is typical for production experiments, not all speakers behaved the same, and the speakers in this study fell into three groups, based on the significance of prosodic position as a main effect on total articulatory duration and gestural lag (schematizations in Figure 4.2 reproduced from Chapter II):

- Group 1 (S04, S07, S08, S13): For these four speakers' productions, sentence-initial laterals were longer than word-initial laterals, but prosodic position had no effect on gestural lag, which was positive for three speakers' productions, but not significantly different from zero in speaker S13's productions.
- Group 2 (S10, S12, S16): Sentence-initial laterals produced by these three speakers were not longer than word-initial laterals, but prosodic position did have an effect on gestural lag.
- Group 3 (S01, S02, S06, S09): In the productions of these four speakers, gestural lag

¹The Task Dynamics model of speech production also predicts gestural synchrony, though it predicts synchrony of gestural onsets rather than achievements. Due to this distinction, the analyses reported in this dissertation cannot be interpreted within a Task Dynamic model.

was always positive, and neither duration nor gestural lag was affected by prosodic position.

For all speakers, when gestural lag was significantly different from zero, it was positive, indicating that the physically-informed hypotheses are reasonably successful at predicting the basic behavior of speakers in this task. However, both the jaw-cycle and the π -gesture hypotheses predict that the amount of gestural lag should increase in sentence-initial position, contrary to the results from this experiment: gestural lag in sentence-initial laterals produced by speakers in Groups 1 and 3, including those produced by speaker S13, was not significantly different from gestural lag in word-initial laterals, and gestural lag in sentence-initial laterals produced by speakers in Group 2 was significantly less than lag in word-initial position.

How well does the perceptual hypothesis fit the results? The productions from Speaker S13 are most consistent with the perceptual hypothesis: in both word- and sentence-initial laterals, the gestures were synchronous, and only the perceptual hypothesis predicts synchrony of gestural achievement. Thus, the results from this speaker are consistent with this speaker producing listener-oriented speech in both prosodic contexts. The other speakers in Group 1 produced sentence-initial laterals that were longer than word-initial laterals, but laterals in both contexts exhibited positive gestural lag. Though not perfectly consistent with the physically-based hypotheses, as both predict greater gestural lag in sentence-initial laterals, these findings are even less consistent with the predictions of the perceptual hypothesis.

Speakers in Group 2 produced sentence-initial laterals with gestural synchrony, but word-initial laterals with positive gestural lag. This result is consistent with a model of speech production in which speakers produce physically-motivated speech in one context but this motivation may be overridden by listener-oriented considerations in other contexts.

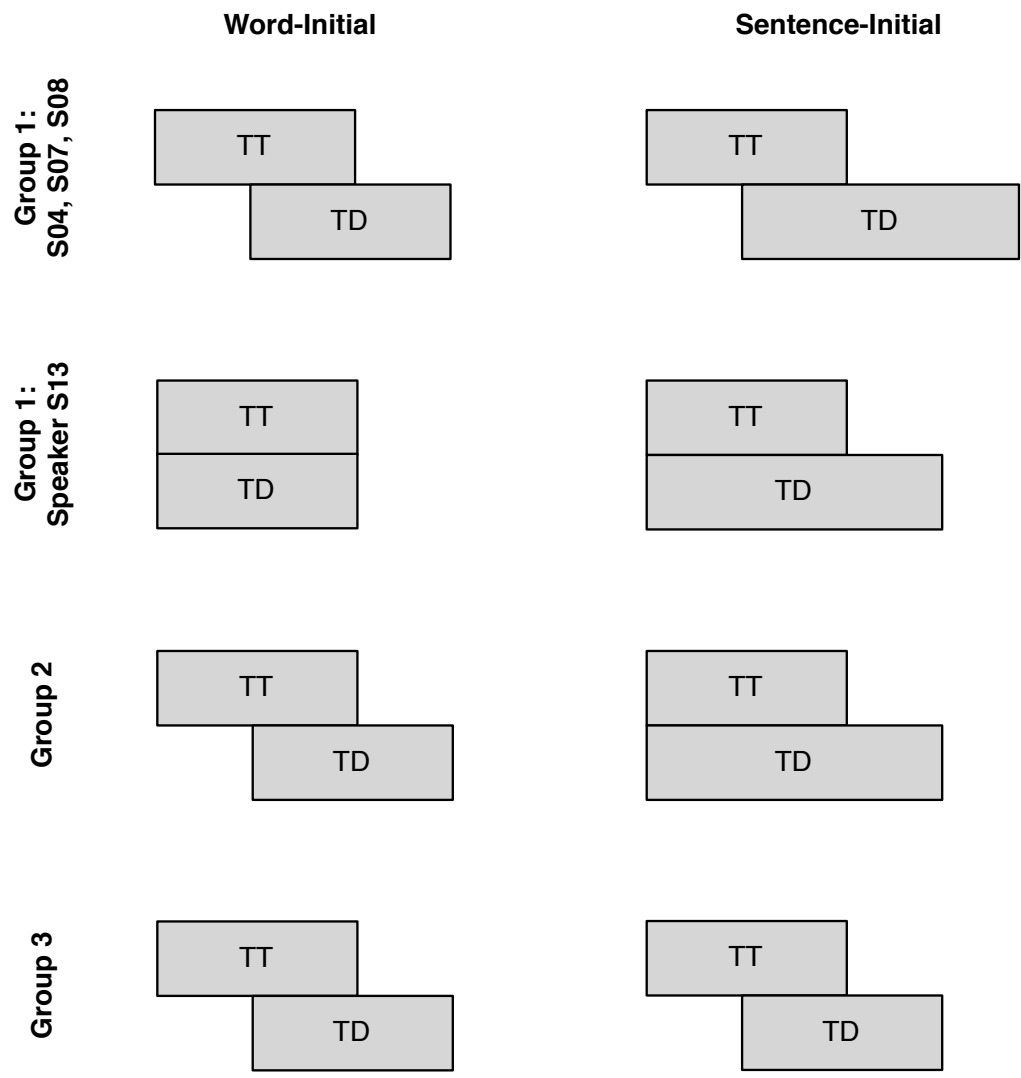


Figure 4.2: Schematization of the results from Experiment 1.

Recall that Keating (2006) proposes that the articulatory strengthening that occurs at the onset of large prosodic domains occurs for the benefit of the listener. A corollary to this hypothesis is that word-initial onset consonants do not require much listener-motivated attention. As discussed previously, positive lag in onset laterals is associated with the physically-motivated hypotheses of speech production in this study. Thus, the results from the speakers in this group are highly consistent with a model of speech production that allows speakers to switch between speaker- or listener-directed modes of communication.

Finally, speakers in Group 3 did not produce sentence-initial laterals that were significantly distinct from word-initial laterals, with respect to any of the measurements taken in this study. Specifically, neither gesture lengthened in sentence-initial laterals, and there was no significant change in gestural lag. The results from these speakers are not strongly consistent with any of the hypotheses. It is possible that, when reading the stimuli, some speakers simply read them, without performing them as if they were naturally spoken sentences. This is an issue not uncommon to speech production experiments in laboratory settings. This cannot be precisely what is happening with these speakers, since they do produce sentence-final laterals differently from word-final laterals. However, the effect of prosodic boundary on coda laterals is generally attributed to a slowing of the speech rate, while the effect of prosodic boundary on onset laterals may be due to the perceptual needs of the listener. Thus, it is possible that these speakers are performing one set of effects that characterize strong prosodic boundaries, but not those that may be attributed to perceptual needs, due to the absence of an active listener.

4.3 Overall Strength of the Hypotheses

While the previous section examined the data in light of the hypotheses set forth, this section will examine each hypothesis, and discuss its viability in light of the data.

4.3.1 Perceptual & jaw-cycle hypotheses

Of the three hypotheses discussed in this dissertation, a combination of the perceptual and jaw-cycle hypotheses appear to be the best supported by the results from this study. Given the domain-initial strengthening hypothesis, it is not surprising that a complete explanation of the data collected in Experiment 1 calls upon both the listener-oriented and jaw-cycle (speaker-motivated) hypotheses. Such a pairing suggests that there may be speakers who rarely employ listener-oriented strategies, those who frequently produce listener-motivated speech, and speakers who vary their strategy in a predictable and context-dependent manner. Importantly, when speakers did vary gestural coordination due to prosodic context, it was always in the direction of favoring perception in the stronger prosodic context; no speaker exhibited greater gestural lag in sentence-initial laterals than in word-initial laterals. Under this analysis, speaker S13 belongs to the group of individuals who employ a listener-motivated pattern of coordination independent of context, while the production patterns of the remaining speakers in Group 1 reveal exhibit speaker-motivated gestural lag patterns, across the board. Speakers in Group 2, however, varied gestural coordination, in a manner consistent with a perceptually-based model of domain-initial strengthening.

This notion, that speakers are capable of choosing from different speech production strategies based on the listeners' needs (or perceived needs), is at the core of Lindblom's (1983; 1990) theory of Hypo- and Hyper-speech (H&H Theory). This theory hypothesizes that speakers are continuously balancing speaker-motivated forces (hypo-speech) with listeners' needs for perceptual distinctiveness (hyper-speech), and it takes the strong and controversial stance that these two forces are in opposition with one another: to create perceptually clear speech necessitates greater speaker effort, and speech that is articulatorily efficient will be reduced or otherwise less distinct and consequently more difficult

for listeners to understand.

The proposal put forth in this dissertation shares with H&H the assumption that speakers' phonetic knowledge includes knowledge of the perceptual effectiveness of their speech, and that speakers are capable of adapting their speech to take advantage of this awareness. I do not assume, contra Lindblom (1990), that speaker-oriented speech is necessarily less perceptible. While the results of Experiment 2 show that onset laterals with synchronous gestures were identified more accurately and more quickly, the stimuli in this experiment were isolated, lacking any preceding phonetic context. In a phonetically appropriate context (e.g. sentence-medial word-initial position), onset laterals with asynchronous gestures might not be less perceptible. Under the testing conditions used here, it can only be concluded that, when devoid of preceding phonetic context, gestural synchrony in onset laterals results in improved perception.

4.3.2 π -gesture

While the π -gesture hypothesis, like the jaw-cycle hypothesis, correctly predicted that, in general, onset laterals would be produced with the tongue tip gesture preceding the tongue tip gesture, the data also present some problems for the hypothesis as laid out by Krivokapić and Byrd (2010). They suggest that onset laterals should exhibit gestural lag, and that the amount of lag should increase at higher prosodic positions. Additionally, their approach predicts that the gesture closest to the prosodic boundary should be most affected by the π -gesture. In the case of onset laterals, that means that the tongue tip gesture should lengthen more than the tongue dorsum gesture. However, for the speakers in Experiment 1, whenever only one of the two gestures was significantly longer in sentence-initial laterals than word-initial laterals, it was the tongue dorsum gesture rather than the tongue tip gesture, an outcome that contrasts with the predictions of this hypothesis.

It is possible that the tongue tip gestures in sentence-initial laterals are longer than

word-initial laterals, but the magnitude of the gestural lengthening is such that the difference is not significant. Figure 2.16 on page 43 suggests that this may be true for at least half of the speakers in Experiment 1. In this case, a more appropriate way to measure gestural lengthening may be to consider the proportion of tongue tip gesture duration to tongue dorsum gesture duration. If the π -gesture lengthens both the tongue tip and the tongue dorsum gesture proportionally, this proportion should be the same in word- and sentence-initial laterals. If its effect is proportionally larger for the tongue tip than the tongue dorsum gesture, this proportion should be greater in sentence-initial laterals than in word-initial laterals. Either outcome would be consistent with the π -gesture hypothesis. However, a repeated-measures ANOVA using speaker as an error factor, show that this proportion is significantly smaller in sentence-initial laterals than in word-initial laterals ($F(1, 10) = 9.26, p = 0.0124$).

Furthermore, the proposal that speakers can switch between listener- and physically-motivated speech depending on context hinges on the ability of the listener-directed and physically-motivated hypotheses to coexist. While the sentence-initial predictions of the jaw-cycle hypothesis are at odds with those of the perceptual hypothesis, the suggestion that a speaker's knowledge of listeners' needs can overcome speaker inertia to produce physically-motivated speech is not unreasonable, and in fact is, as discussed above, at the core of H&H theory (Lindblom, 1983, 1990). In contrast, the predictions of the π -gesture hypothesis fall directly out of a theory of prosodic structure in which prosodic gestures encompass large prosodic structures, and the slowing and lengthening of both sentence-initial and sentence-final segments is attributable the same source: their proximity to the boundary of the prosodic gesture (Byrd and Saltzman, 2003). It is thus less compatible with the perceptual hypothesis proposed in this dissertation, which is based on a contrasting theory of prosodic structure which suggests that the slowing and lengthening of

sentence-initial and -final segments may be due to different sources (Fougeron and Keating, 1997; Keating, 2006).

4.4 Conclusion

The data from onset laterals produced by most speakers in this study are consistent with speaker-based patterns of gestural coordination. When prosodic context does prove to have a significant effect on onset lateral production, the data are most consistent with a model of speech production in which speakers are capable of modifying their production strategies based the needs of their listeners. Specifically, the data suggest that some speakers produce sentence-initial laterals, that are particularly important from an informational load perspective, in a manner that provides their listeners with acoustic information pertinent to identifying a lateral as early as possible. Thus, as an examination of the speaker-as-listener relationship, the overall results of this study are most consistent with a model of speech production in which gestural coordination is largely determined by physical and bio-mechanical factors which may be partially overridden, by some speakers, in ways that take advantage of speakers' access to sophisticated phonetic knowledge about speech perception.

APPENDICES

APPENDIX A

Experiment 1, 3 Prosodic Contexts, S01 - S09

The tables and figures in this appendix summarize the results from Experiment 1 from speakers S01 through S09¹, including all three prosodic contexts (word, phrase and sentence). All statistics reported below are for individual speakers, and are based on 2-way ANOVAs (variables: prosodic position and vowel), examining effects of prosodic position and vowel quality on lateral duration, gestural lag, and tongue tip and tongue dorsum gestures. Post-hoc Tukey tests were performed when appropriate, to test for pair-wise differences. Measurements reported in this section are: lateral duration, gestural lag, tongue tip gesture duration, and tongue dorsum gesture duration. These measurements are as defined in Section §2.2.3. The expectation is that all duration measures for sentence-initial and sentence-final laterals should be greater than word-initial and word-final laterals, respectively. Additionally, the amount of gestural lag is expected to be more negative in sentence-final laterals compared to word-final laterals. In general, measurements for phrase-final and phrase-initial laterals were intermediate between sentence and word contexts. Based on these results, Experiment 1 was modified for subsequent speakers to exclude phrase-initial and -final contexts, in favor of increasing the number of repetitions for each context.

¹Data from only seven of these speakers was used in this study, for the reasons presented in Table 2.3.

A.1 Coda Laterals

In general, sentence-final laterals were significantly longer than word-final and phrase-final laterals. Additionally, sentence-final laterals exhibited significantly more negative gestural lag than word-final laterals. Both tongue tip and tongue dorsum gestures were typically longer in sentence-final laterals than word-final laterals. Values for phrase-final laterals were intermediate in nearly all contexts.

A.2 Initial Laterals

Generally, the effect of prosodic context on onset laterals was substantially weaker than on coda laterals. There was a significant effect of prosodic context on duration of onset laterals in the productions of three speakers, and on gestural lag for only one speaker. Similarly, there was a significant effect of prosodic context on duration of tongue dorsum gestures for three speakers, and on duration of tongue tip gestures for one. In the cases for which prosodic context had a significant effect on measurements, values in phrase-initial laterals were intermediate, between sentence- and word-initial.

Speaker	Word -Final	Phrase -Final	Sentence -Final	ANOVA, Tukey HSD
S01	310.9	424.1	461.5	$F(2, 60) = 11.78, p < 0.0001$ word v. phrase ($p = 0.003819$) word v. sentence ($p = 0.0000360$)
S02	371.1	463.8	569.0	$F(2, 70) = 7.622, p = 0.00101$ sentence v. phrase ($p = 0.06394$) word v. sentence ($p = 0.0007253$)
S04	284.9	300.8	408.4	$F(2, 78) = 21.90, p < 0.0001$ word v. sentence ($p < 0.0001$) sentence v. phrase ($p < 0.0001$)
S06	277.9	364.6	481.8	$F(2, 74) = 29.00, p < 0.0001$ word v. phrase ($p = 0.005524$) word v. sentence ($p < 0.0001$) sentence v. phrase ($p < 0.0001$)
S07	430.2	530.8	718.0	$F(2, 76) = 32.85, p < 0.0001$ word v. phrase ($p = 0.01790$) word v. sentence ($p < 0.0001$) sentence v. phrase ($p < 0.0001$)
S08	447.5	416.2	467.1	$F(2, 63) = 1.677, p = 0.1953$
S09	377.1	417.9	485.3	$F(2, 61) = 6.341, p = 0.03148$ sentence v. phrase ($p = 0.08251$) word v. sentence ($p = 0.002477$)

Table A.1: Mean lateral duration values (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$.

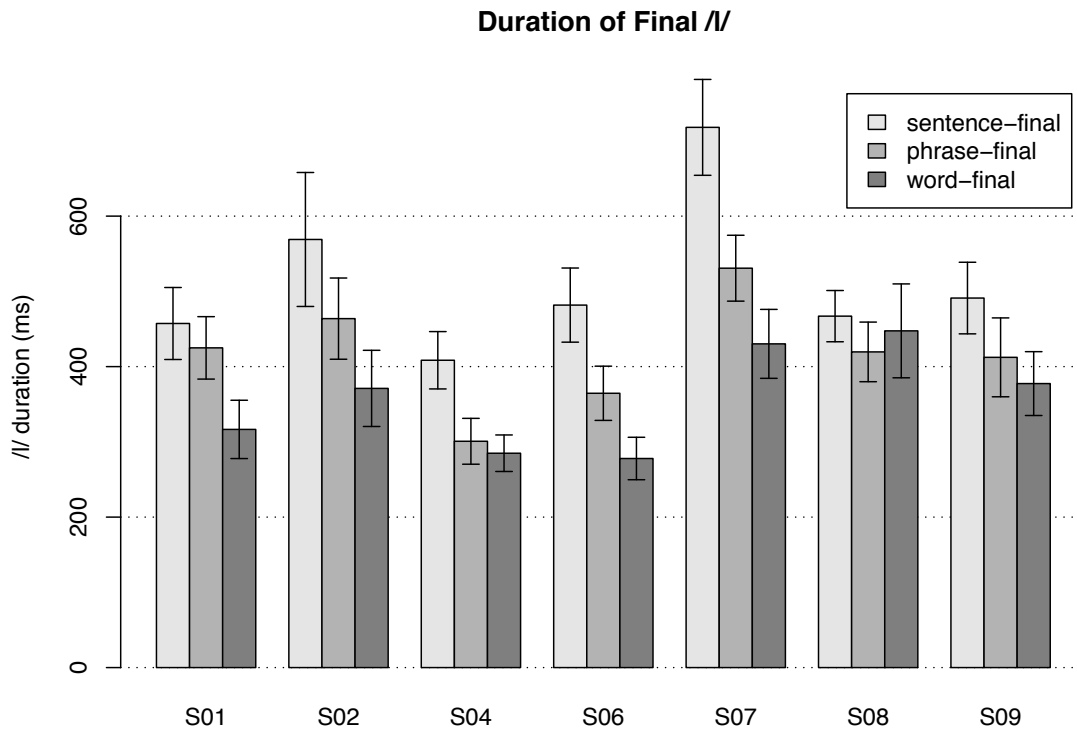


Figure A.1: Duration of coda lateral production (ms) by prosodic context, for speakers S01-S09

Speaker	Word	Phrase	Sentence	ANOVA, Tukey HSD
S01	-37.15	-59.38	-79.21	$F(2, 60) = 1.607, p = 0.2089$
S02	-18.95	-129.7	-138.9	$F(2, 70) = 4.335, p = 0.01680$ word v. phrase ($p = 0.03403$) word v. sentence ($p = 0.02335$)
S04	-25.00	-63.00	-118.0	$F(2, 78) = 5.995, p = 0.003786$ word v. sentence ($p = 0.002678$) sentence v. phrase ($p = 0.1030$)
S06	-40.12	-97.55	-114.7	$F(2, 74) = 10.77, p < 0.0001$ word v. phrase ($p = 0.03410$) word v. sentence ($p < 0.0001$) sentence v. phrase ($p = 0.08249$)
S07	-50.00	-148.9	-185.5	$F(2, 76) = 15.89, p < 0.0001$ word v. phrase ($p = 0.0003811$) word v. sentence ($p < 0.0001$)
S08	-136.2	-60.31	-137.7	$F(2, 63) = 4.032, p = 0.02250$ sentence v. phrase ($p = 0.0344$) word v. phrase ($p = 0.06564$)
S09	-50.73	-73.68	-136.6	$F(2, 61) = 8.189, p = 0.0007077$ sentence v. phrase ($p = 0.01692$) word v. sentence ($p = 0.0007804$)

Table A.2: Mean gestural lag (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$.

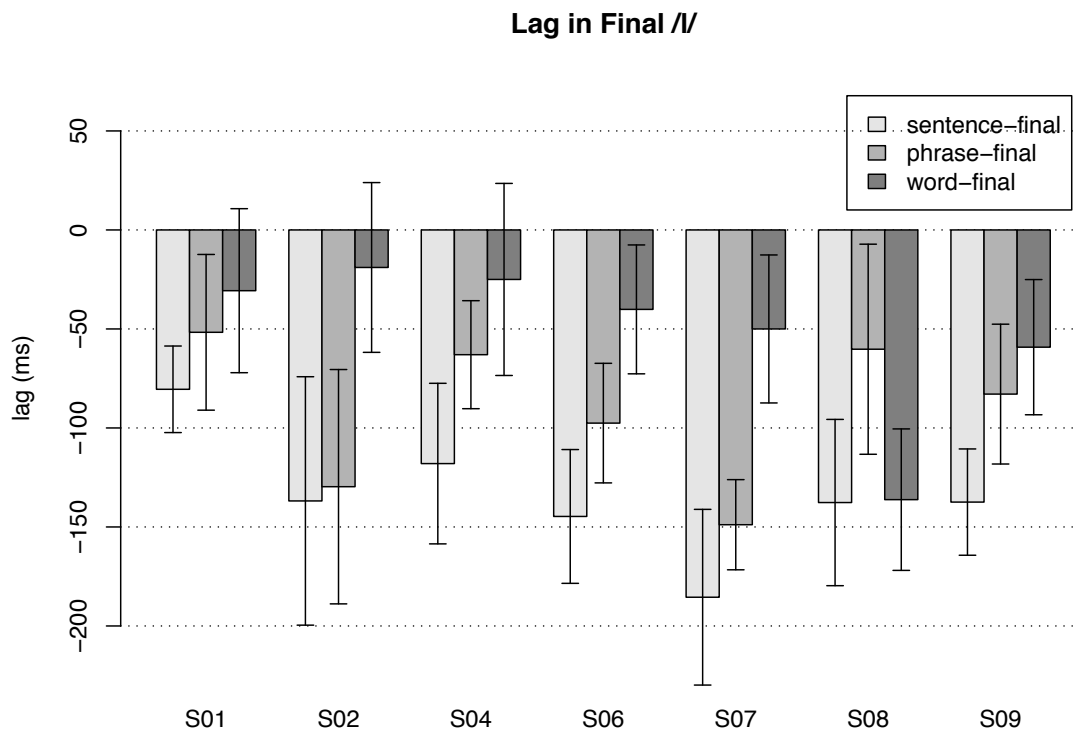


Figure A.2: Duration of gestural lag (ms) in coda //s by prosodic context, for speakers S01-S09

Speaker	Word -Final	Phrase -Final	Sentence -Final	ANOVA, Tukey HSD
S01	227.1	287.0	331.3	$F(2, 60) = 7.418, p = 0.001322$ word v. phrase ($p = 0.1034$) word v. sentence ($p = 0.0008303$)
S02	304.8	341.1	444.4	$F(2, 70) = 4.2088, p = 0.01879$ sentence v. phrase ($p = 0.07596$) word v. sentence ($p = 0.02468$)
S04	174.9	191.2	228.7	$F(2, 78) = 7.764, p = 0.0008415$ word v. sentence ($p = 0.0007434$) sentence v. phrase ($p = 0.02210$)
S06	194.0	250.3	324.8	$F(2, 74) = 12.85, p < 0.0001$ word v. phrase ($p = 0.08368$) word v. sentence ($p < 0.0001$) sentence v. phrase ($p = 0.01090$)
S07	280.2	374.7	456.1	$F(2, 76) = 9.639, p = 0.0001859$ word v. phrase ($p = 0.05156$) word v. sentence ($p = 0.0001060$) sentence v. phrase ($p = 0.09515$)
S08	278.9	242.1	299.1	$F(2, 63) = 2.441, p = 0.09529$
S09	257.5	325.5	374.4	$F(2, 61) = 9.370, p = 0.0002829$ word v. phrase ($p = 0.04869$) word v. sentence ($p = 0.0001682$)

Table A.3: Mean tongue tip gesture durations (ms) for coda laterals by speaker, for S01-S09. bolded data are significant are $p < 0.05$.

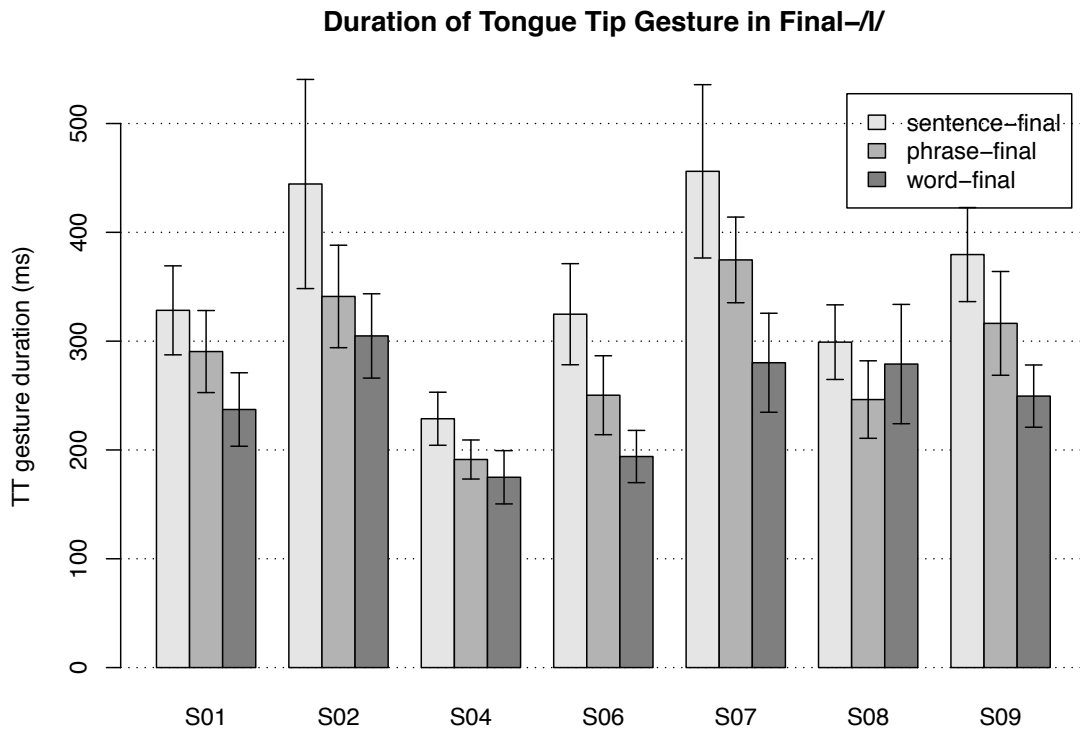


Figure A.3: Duration of tongue tip gesture (ms) in coda /l/s by prosodic context, for speakers S01-S09

Speaker	Word -Final	Phrase -Final	Sentence -Final	ANOVA, Tukey HSD
S01	253.6	362.7	380.1	$F(2, 60) = 7.033, p = 0.001803$ word v. phrase ($p = 0.01453$) word v. sentence ($p = 0.001959$)
S02	290.1	319.0	363.4	$F(2, 70) = 3.252, p = 0.04462$ word v. sentence ($p = 0.04114$)
S04	217.5	234.3	335.9	$F(2, 78) = 15.19, p < 0.0001$ word v. sentence ($p = 0.0007434$) sentence v. phrase ($p = 0.02210$)
S06	230.4	258.3	332.1	$F(2, 74) = 9.739, p = 0.0001759$ word v. sentence ($p < 0.0001$) sentence v. phrase ($p < 0.0001$)
S07	337.8	443.2	570.7	$F(2, 76) = 20.55, p < 0.0001$ word v. phrase ($p = 0.01287$) word v. sentence ($p < 0.0001$) sentence v. phrase ($p = 0.001493$)
S08	317.9	318.8	365.9	$F(2, 63) = 1.586, p = 0.2128$
S09	314.1	302.4	356.9	$F(2, 61) = 1.696, p = 0.1919$

Table A.4: Mean tongue dorsum gesture durations (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$.

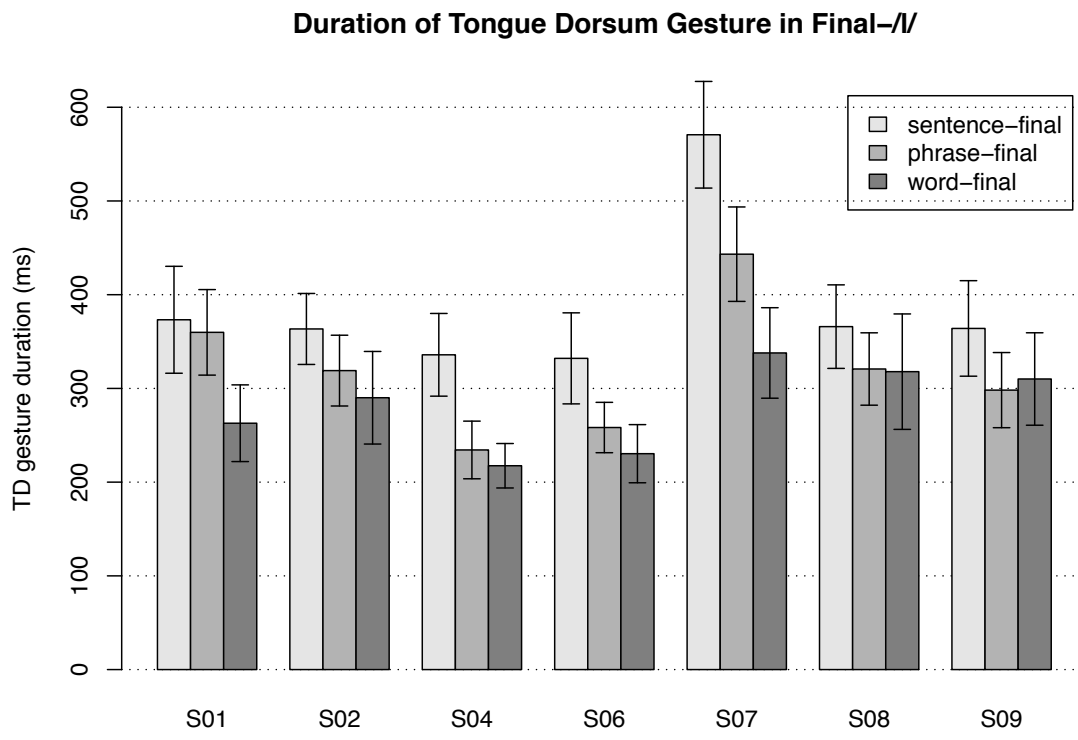


Figure A.4: Duration of tongue dorsum gesture (ms) in coda /l/s by prosodic context, for speakers S01-S09

Speaker	Word -Final	Phrase -Final	Sentence -Final	ANOVA, Tukey HSD
S01	364.5	344.7	422.6	$F(2, 64) = 3.931, p = 0.02454$ sentence v. phrase ($p = 0.02593$) word v. sentence ($p = 0.09953$)
S02	426.5	438.5	499.8	$F(2, 72) = 2.291, p = 0.1085$
S04	306.2	333.9	359.5	$F(2, 77) = 3.366, p = 0.03970$ sentence v. word ($p = 0.0302$)
S06	320.1	312.3	351.3	$F(2, 72) = 1.920, p = 0.1540$
S07	422.4	454.4	500.0	$F(2, 64) = 2.154, p = 0.1243$
S08	338.6	407.9	470.0	$F(2, 66) = 11.93, p < 0.0001$ word v. phrase ($p = 0.05471$) word v. sentence ($p < 0.0001$) sentence v. phrase ($p = 0.07506$)
S09	396.8	407.5	443.6	$F(2, 50) = 1.260, p = 0.2926$

Table A.5: Mean lateral durations (ms) for onset laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$.

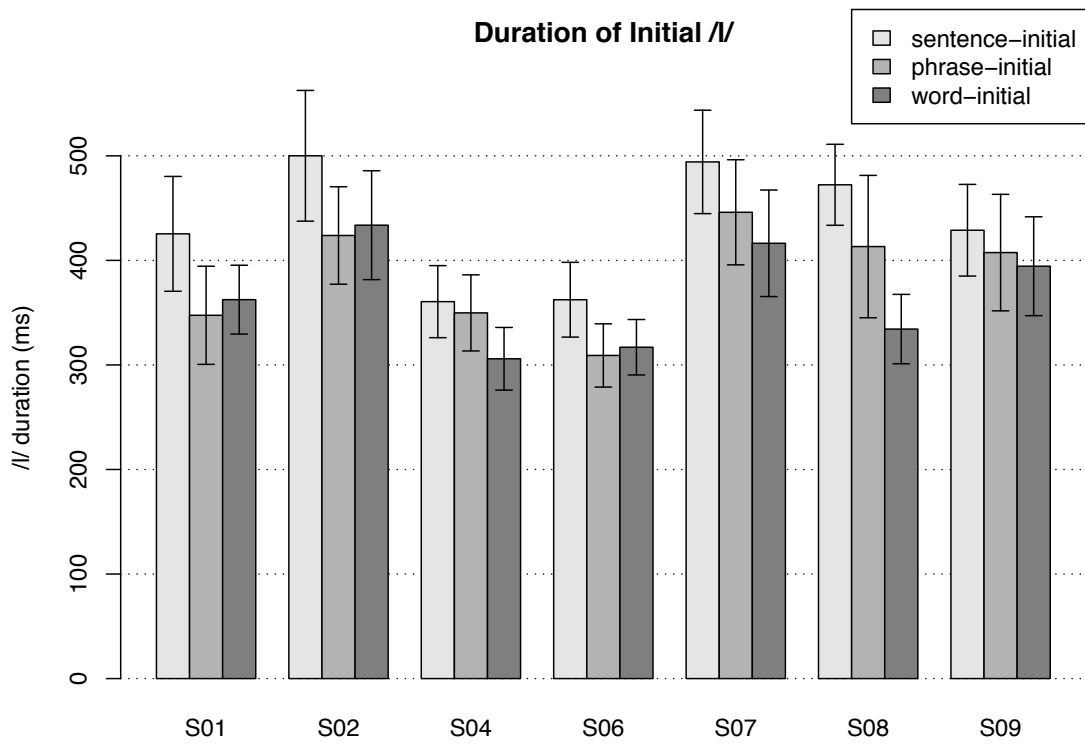


Figure A.5: Duration of onset lateral production (ms) by prosodic context, for speakers S01-S09

Speaker	Word -Final	Phrase -Final	Sentence -Final	ANOVA, Tukey HSD
S01	54.81	3.500	23.04	$F(2, 64) = 1.435, p = 0.2457$
S02	63.00	82.92	114.7	$F(2, 72) = 1.275, p = 0.2857$
S04	56.00	69.57	37.76	$F(2, 77) = 0.9439, p = 0.3936$
S06	39.77	34.85	48.21	$F(2, 72) = 0.1758, p = 0.8391$
S07	79.76	35.88	64.51	$F(2, 64) = 0.9342, p = 0.3982$
S08	59.88	130.6	49.53	$F(2, 66) = 3.329, p = 0.04194$ sentence v. phrase ($p = 0.04384$)
S09	16.29	46.80	21.73	$F(2, 50) = 0.4588, p = 0.6347$

Table A.6: Mean gestural lag (ms) for onset laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$.

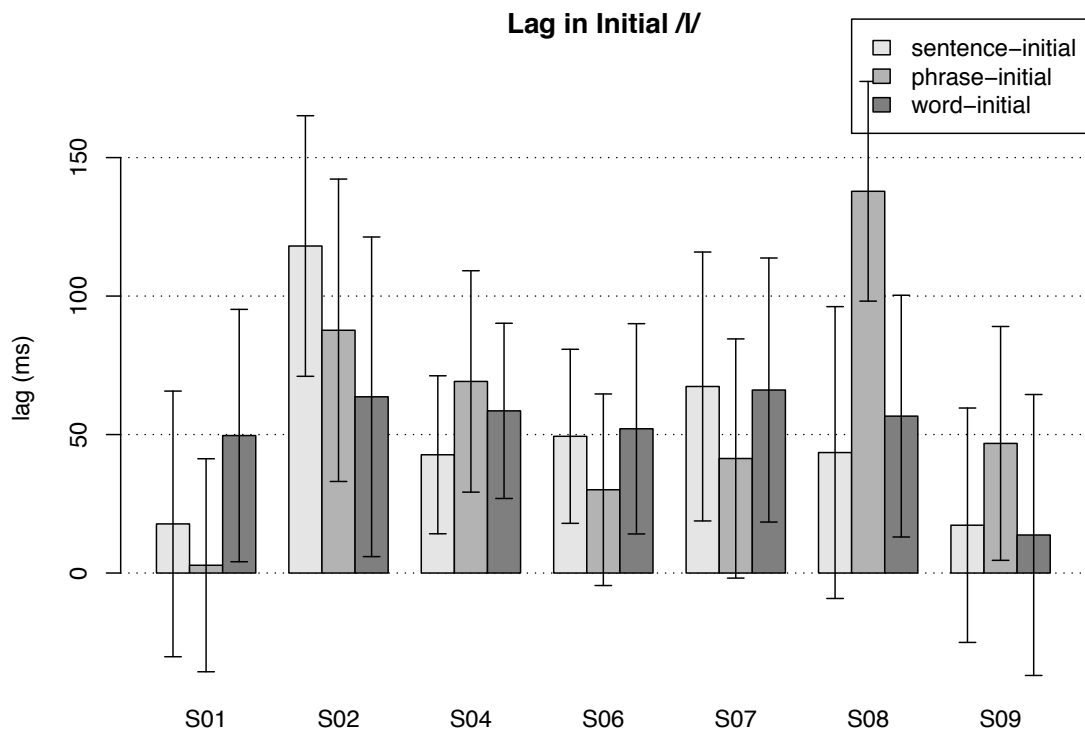


Figure A.6: Duration of gestural lag (ms) in onset /l/s by prosodic context, for speakers S01-S09

Speaker	Word -Final	Phrase -Final	Sentence -Final	ANOVA, Tukey HSD
S01	251.0	235.8	267.5	$F(2, 64) = 1.082, p = 0.3451$
S02	309.4	302.4	352.6	$F(2, 72) = 1.320, p = 0.2735$
S04	210.6	213.9	233.4	$F(2, 77) = 0.7464, p = 0.4775$
S06	241.9	249.3	261.0	$F(2, 72) = 0.5815, p = 0.5616$
S07	323.6	344.9	370.4	$F(2, 64) = 0.8796, p = 0.4199$
S08	233.4	313.2	313.2	$F(2, 66) = 5.079, p = 0.008880$ word v. phrase ($p = 0.01444$) sentence v. word ($p = 0.02790$)
S09	283.1	307.7	326.0	$F(2, 50) = 0.9386, p = 0.3980$

Table A.7: Mean tongue tip gesture durations (ms) for onset laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$.

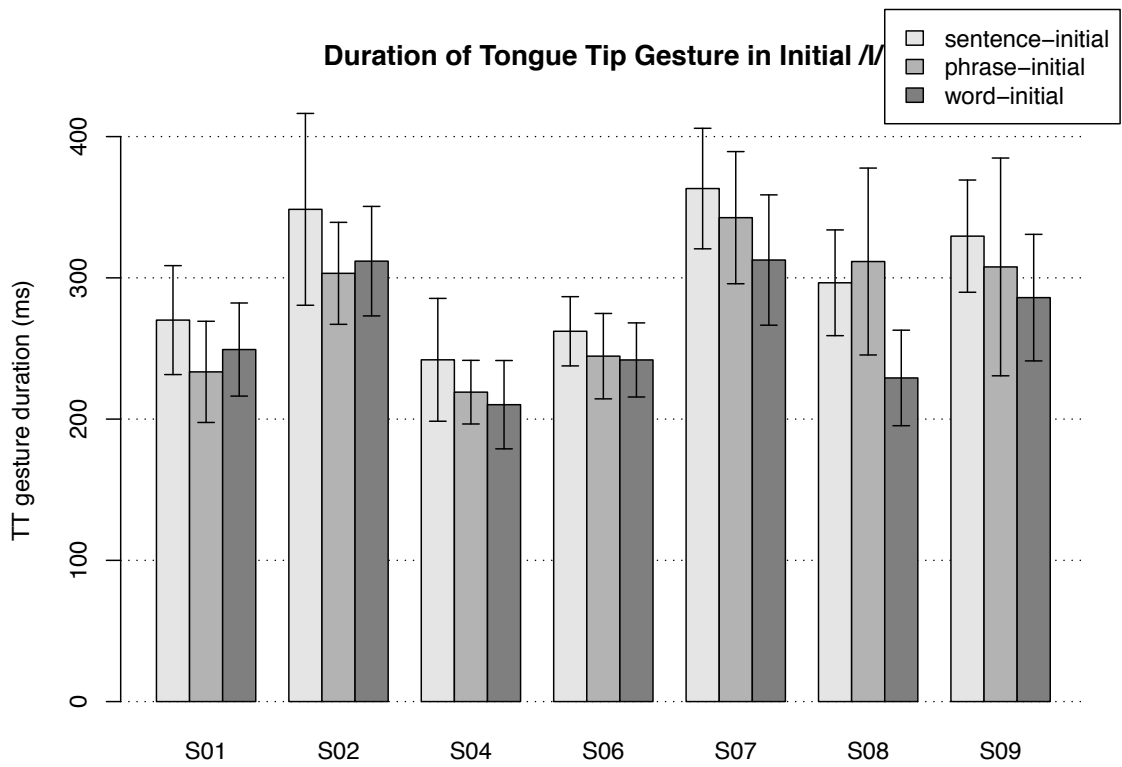


Figure A.7: Duration of tongue tip gesture (ms) in onset /l/s by prosodic context, for speakers S01-S09

Speaker	Word -Final	Phrase -Final	Sentence -Final	ANOVA, Tukey HSD
S01	281.1	301.0	359.9	$F(2, 64) = 3.447, p = 0.03788$ word v. sentence ($p = 0.3446$)
S02	330.4	369.9	386.4	$F(2, 72) = 0.9605, p = 0.38755$
S04	204.7	256.3	278.1	$F(2, 77) = 6.631, p = 0.002203$ word v. phrase ($p = 0.04083$) word v. sentence ($p = 0.001908$)
S06	219.8	250.8	265.0	$F(2, 72) = 1.755, p = 0.1802$
S07	287.4	338.8	389.2	$F(2, 64) = 3.214, p = 0.4675$
S08	236.5	291.9	355.1	$F(2, 66) = 15.12, p < 0.0001$ word v. phrase ($p = 0.05681$) word v. sentence ($p < 0.0001$) sentence v. word ($p = 0.01774$)
S09	307.5	318.3	339.1	$F(2, 50) = 0.5330, p = 0.5901$

Table A.8: Mean tongue dorsum gesture durations (ms) for coda laterals by speaker, for S01-S09. Bolded data are significant at $p < 0.05$.

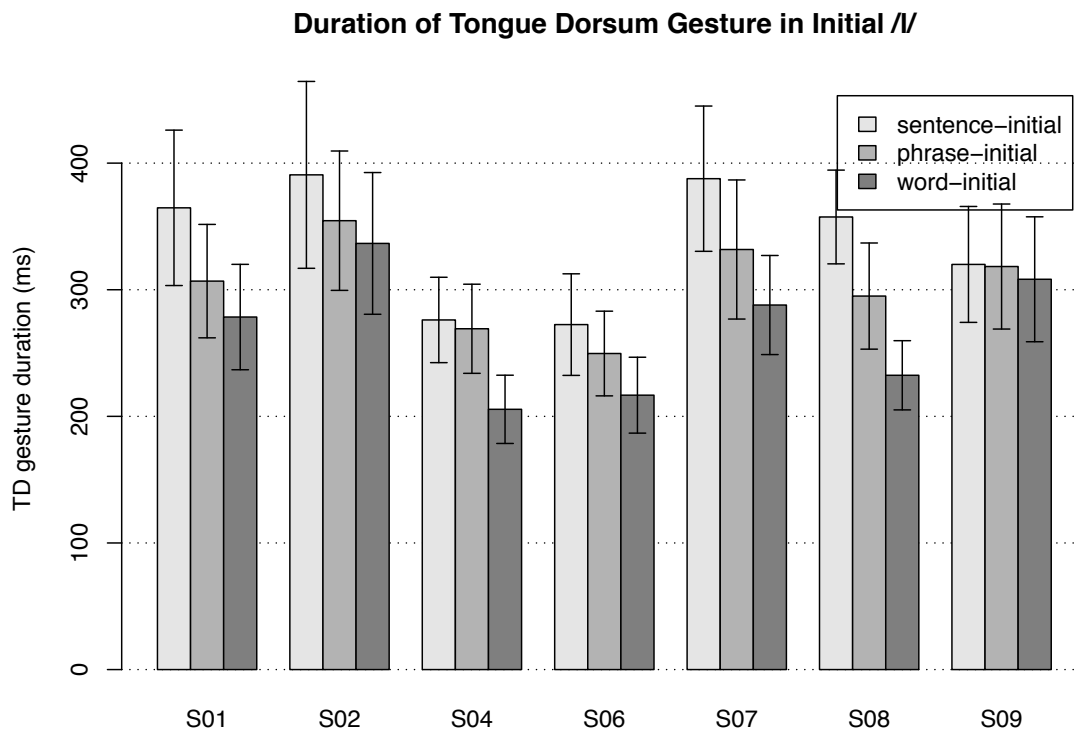


Figure A.8: Duration of tongue dorsum gesture (ms) in onset /l/s by prosodic context, for speakers S01-S09

APPENDIX B

Experiment 1, 2 Prosodic Contexts, All Speakers

The tables in this section provide, for each of the eleven speakers, the measurements and relevant statistics for Experiment 1 for final and initial laterals in sentence- and word-adjacent contexts. All reported statistics are from 2-way ANOVAs (variables: prosodic position and vowel), and post-hoc Tukey tests were performed when appropriate. Measurements reported in this section are: lateral duration, gestural lag, tongue tip gesture duration, and tongue dorsum gesture duration. These measurements are as defined in Section §2.2.3. The expectation is that all duration measures for sentence-initial and sentence-final laterals should be greater than word-final and word-final laterals, respectively. Additionally, the amount of gestural lag is expected to be more negative in sentence-final laterals compared to word-initial laterals. These results are displayed and discussed at greater length in Sections §2.3.4, §2.3.5, and §2.3.6. Although there are several significant vowel effects and interactions in this data, very little is consistent from speaker to speaker, suggesting that these significant effects are likely to be speaker-specific. One relatively consistent significant effect, however, is that gestural durations in the coda lateral in “pal” tended to be longer than those in either “male” or “peel.”

B.1 Final laterals

S01		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 42) = 21.29, p < 0.0001$
	Vowel Quality	$F(2, 42) = 1.558, p = 0.2224$
	Interaction	$F(2, 42) = 0.3578, p = 0.7013$
Lag	Prosodic Position	$F(1, 42) = 3.427, p = 0.07116$
	Vowel Quality	$F(2, 42) = 0.3176, p = 0.7296$
	Interaction	$F(2, 42) = 0.6780, p = 0.5131$
TT	Prosodic Position	$F(1, 42) = 14.39, p = 0.0004693$
	Vowel Quality	$F(2, 42) = 1.719, p = 0.1916$
	Interaction	$F(2, 42) = 0.6823, p = 0.5109$
TD	Prosodic Position	$F(1, 42) = 11.35, p = 0.001625$
	Vowel Quality	$F(2, 42) = 0.7663, p = 0.4711$
	Interaction	$F(2, 42) = 0.5161, p = 0.6006$

Table B.1: Measurements and statistics for results for coda laterals from Experiment 1, speaker S01. Bolded data are significant at $p < 0.05$.

S02		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 43) = 12.12, p = 0.001160$
	Vowel Quality	$F(2, 43) = 1.135, p = 0.3310$
	Interaction	$F(2, 43) = 0.3077, p = 0.7368$
Lag	Prosodic Position	$F(1, 43) = 7.960, p = 0.007209$
	Vowel Quality	$F(2, 43) = 0.3617, p = 0.6986$
	Interaction	$F(2, 43) = 0.4103, p = 0.6660$
TT	Prosodic Position	$F(1, 43) = 5.418, p = 0.02471$
	Vowel Quality	$F(2, 43) = 0.6886, p = 0.5078$
	Interaction	$F(2, 43) = 0.3234, p = 0.7254$
TD	Prosodic Position	$F(1, 43) = 6.196, p = 0.01675$
	Vowel Quality	$F(2, 43) = 2.055, p = 0.1404$
	Interaction	$F(2, 43) = 0.2068, p = 0.8140$

Table B.2: Measurements and statistics for results for coda laterals from Experiment 1, speaker S02. Bolded data are significant at $p < 0.05$.

S04		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 51) = 37.76, p < 0.0001$
	Vowel Quality	$F(2, 51) = 0.6416, p = 0.5307$
	Interaction	$F(2, 51) = 7.4105, p = 0.001495$
Lag	Prosodic Position	$F(1, 51) = 9.418, p = 0.003438$
	Vowel Quality	$F(2, 51) = 0.2996, p = 0.7424$
	Interaction	$F(2, 51) = 2.498, p = 0.09229$
TT	Prosodic Position	$F(1, 51) = 13.66, p = 0.0005348$
	Vowel Quality	$F(2, 51) = 2.213, p = 0.1198$
	Interaction	$F(2, 51) = 9.186, p = 0.0003917$
TD	Prosodic Position	$F(1, 51) = 23.67, p < 0.0001$
	Vowel Quality	$F(2, 51) = 0.1975, p = 0.8214$
	Interaction	$F(2, 51) = 2.631, p = 0.08178$

S04 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
Duration	i	325.4	334.8	$p = 0.9998$
	e	263.4	441.1	$p < 0.0001$
	æ	268.2	442.1	$p = 0.0001055$
TT	i	210.7	172.0	$p = 0.6700$
	e	144.5	236.8	$p = 0.005607$
	æ	172.9	271.6	$p = 0.003523$

Table B.3: Measurements and statistics for results for coda laterals from Experiment 1, speaker S04. Bolded data are significant at $p < 0.05$.

S06				
Measure	Variable		ANOVA	
Duration	Prosodic Position		$F(1, 48) = 55.14, p < 0.0001$	
	Vowel Quality		$F(2, 48) = 1.364, p = 0.2653$	
	Interaction		$F(2, 48) = 3.372, p = 0.04263$	
Lag	Prosodic Position		$F(1, 48) = 20.51, p < 0.0001$	
	Vowel Quality		$F(2, 48) = 0.05370, p = 0.9478$	
	Interaction		$F(2, 48) = 1.7919, p = 0.1776$	
TT	Prosodic Position		$F(1, 48) = 26.31, p < 0.0001$	
	Vowel Quality		$F(2, 48) = 1.305, p = 0.2805$	
	Interaction		$F(2, 48) = 3.251, p = 0.04742$	
TD	Prosodic Position		$F(1, 48) = 14.68, p = 0.00037$	
	Vowel Quality		$F(2, 48) = 3.490, p = 0.03845$	
	Interaction		$F(2, 48) = 3.787, p = 0.02946$	
S06 Effect of Vowel Quality (ms)				
Measure	Vowel		Tukey	
TD	æ	325.8	e v. æ, $p = 0.3210$	
	e	279.8	i v. æ, $p = 0.03012$	
	i	242.6	i v. e, $p = 0.4363$	
S06 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
Duration	i	279.4	458.8	$p = 0.01105$
	e	295.6	427.7	$p = 0.06508$
	æ	259.0	556.5	$p < 0.0001$
TT	i	169.0	347.4	$p = 0.005322$
	e	215.1	257.8	$p = 0.9175$
	æ	192.2	371.3	$p = 0.001558$
TD	i	224.6	256.7	$p = 0.6700$
	e	245.2	311.0	$p = 0.005607$
	æ	220.0	421.0	$p = 0.003523$

Table B.4: Measurements and statistics for results for coda laterals from Experiment 1, speaker S06. Bolded data are significant at $p < 0.05$.

S07		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 49) = 56.36, p < 0.0001$
	Vowel Quality	$F(2, 49) = 0.8524, p = 0.4326$
	Interaction	$F(2, 49) = 2.135, p = 0.1292$
Lag	Prosodic Position	$F(1, 49) = 22.94, p < 0.0001$
	Vowel Quality	$F(2, 49) = 1.028, p = 0.3653$
	Interaction	$F(2, 49) = 1.586, p = 0.2150$
TT	Prosodic Position	$F(1, 49) = 15.34, p = 0.0002779$
	Vowel Quality	$F(2, 49) = 2.515, p = 0.0913$
	Interaction	$F(2, 49) = 0.9089, p = 0.4097$
TD	Prosodic Position	$F(1, 49) = 42.34, p < 0.0001$
	Vowel Quality	$F(2, 49) = 0.5101, p = 0.6036$
	Interaction	$F(2, 49) = 3.015, p = 0.05825$

Table B.5: Measurements and statistics for results for coda laterals from Experiment 1, speaker S07. Bolded data are significant at $p < 0.05$.

S08		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 40) = 0.3789, p = 0.5412$
	Vowel Quality	$F(2, 40) = 0.9708, p = 0.3875$
	Interaction	$F(2, 40) = 1.054, p = 0.3582$
Lag	Prosodic Position	$F(1, 40) = 0.0025, p = 0.9600$
	Vowel Quality	$F(2, 40) = 0.4556, p = 0.6373$
	Interaction	$F(2, 40) = 0.6174, p = 0.5444$
TT	Prosodic Position	$F(1, 40) = 0.4746, p = 0.4948$
	Vowel Quality	$F(2, 40) = 0.4961, p = 0.6126$
	Interaction	$F(2, 40) = 2.003, p = 0.1482$
TD	Prosodic Position	$F(1, 40) = 1.796, p = 0.1878$
	Vowel Quality	$F(2, 40) = 1.699, p = 0.1958$
	Interaction	$F(2, 40) = 0.0735, p = 0.9293$

Table B.6: Measurements and statistics for results for coda laterals from Experiment 1, speaker S08. Bolded data are significant at $p < 0.05$.

S09				
Measure	Variable		ANOVA	
Duration	Prosodic Position		$F(1, 42) = 13.47, p = 0.0006783$	
	Vowel Quality		$F(2, 42) = 0.2590, p = 0.7731$	
	Interaction		$F(2, 42) = 3.283, p = 0.04733$	
Lag	Prosodic Position		$F(1, 42) = 17.73, p = 0.0001316$	
	Vowel Quality		$F(2, 42) = 1.043, p = 0.3614$	
	Interaction		$F(2, 42) = 0.04480, p = 0.9563$	
TT	Prosodic Position		$F(1, 42) = 19.35, p < 0.0001$	
	Vowel Quality		$F(2, 42) = 0.4148, p = 0.6631$	
	Interaction		$F(2, 42) = 0.5604, p = 0.5752$	
TD	Prosodic Position		$F(1, 42) = 1.617, p = 0.2104$	
	Vowel Quality		$F(2, 42) = 1.746, p = 0.1868$	
	Interaction		$F(2, 42) = 0.6468, p = 0.5289$	
S09 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
Duration	i	318.3	544.3	$p = 0.002052$
	e	419.9	474.4	$p = 0.9035$
	æ	391.1	455.8	$p = 0.7458$

Table B.7: Measurements and statistics for results for coda laterals from Experiment 1, speaker S09. Bolded data are significant at $p < 0.05$.

S10		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 49) = 104.0, p < 0.0001$
	Vowel Quality	$F(2, 49) = 7.021, p = 0.002083$
	Interaction	$F(2, 49) = 8.034, p = 0.000960$
Lag	Prosodic Position	$F(1, 49) = 21.28, p < 0.0001$
	Vowel Quality	$F(2, 49) = 3.857, p = 0.02782$
	Interaction	$F(2, 49) = 2.057, p = 0.1388$
TT	Prosodic Position	$F(1, 49) = 54.64, p < 0.0001$
	Vowel Quality	$F(2, 49) = 6.681, p = 0.002719$
	Interaction	$F(2, 49) = 5.642, p = 0.006236$
TD	Prosodic Position	$F(1, 49) = 76.12, p < 0.0001$
	Vowel Quality	$F(2, 49) = 3.222, p = 0.04845$
	Interaction	$F(2, 49) = 1.528, p = 0.2272$

S10 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Duration	æ	666.4	e v. æ, $p = 0.04576$
	i	538.3	i v. æ, $p = 0.002549$
	e	637.2	i v. e, $p = 0.9367$
Lag	æ	-157.5	e v. æ, $p = 0.02800$
	i	-116.3	i v. æ, $p = 0.2421$
	e	-102.5	i v. e, $p = 0.3826$
TT	æ	523.9	e v. æ, $p = 0.3748$
	e	520.8	i v. æ, $p = 0.001790$
	i	389.4	i v. e, $p = 0.2909$
TD	æ	507.4	e v. æ, $p = 0.4249$
	e	520.1	i v. æ, $p = 0.03981$
	i	397.1	i v. e, $p = 0.7519$

S10 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
Duration	i	456.6	647.2	$p = 0.4367$
	e	475.7	706.4	$p = 0.01279$
	æ	460.0	910.3	$p < 0.0001$
TT	i	348.3	444.2	$p = 0.001125$
	e	324.3	605.0	$p = 0.06223$
	æ	373.9	701.1	$p < 0.0001$

Table B.8: Measurements and statistics for results for coda laterals from Experiment 1, speaker S10. Bolded data are significant at $p < 0.05$.

S12		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 82) = 179.1, p < 0.0001$
	Vowel Quality	$F(2, 82) = 8.761, p = 0.0003563$
	Interaction	$F(2, 82) = 1.709, p = 0.1874$
Lag	Prosodic Position	$F(1, 82) = 21.73, p < 0.0001$
	Vowel Quality	$F(2, 82) = 0.9553, p = 0.3889$
	Interaction	$F(2, 82) = 1.775, p = 0.1759$
TT	Prosodic Position	$F(1, 82) = 186.5, p < 0.0001$
	Vowel Quality	$F(2, 82) = 3.831, p = 0.0257$
	Interaction	$F(2, 82) = 4.716, p = 0.01152$
TD	Prosodic Position	$F(1, 82) = 26.99, p < 0.0001$
	Vowel Quality	$F(2, 82) = 9.477, p = 0.0001983$
	Interaction	$F(2, 82) = 0.3371, p = 0.7148$

S12 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Duration	æ	552.8	e v. æ, $p = 0.001147$
	i	483.7	i v. æ, $p = 0.002513$
	e	500.0	i v. e, $p = 0.9029$
TT	æ	378.9	e v. æ, $p = 0.0003886$
	i	303.3	i v. æ, $p = 0.003462$
	e	294.8	i v. e, $p = 0.6790$
TD	æ	460.6	e v. æ, $p = 0.02337$
	e	424.8	i v. æ, $p = 0.7576$
	i	441.6	i v. e, $p = 0.1178$

S12 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
Lag	i	-43.00	-118.9	$p = 0.0006787$
	e	-70.00	-102.5	$p = 0.5888$
	æ	-79.13	-114.9	$p = 0.3199$
TT	i	264.1	345.1	$p = 0.1165$
	e	229.4	338.4	$p = 0.03572$
	æ	321.1	436.8	$p = 0.004598$

Table B.9: Measurements and statistics for results from for coda laterals Experiment 1, speaker S12. Bolded data are significant at $p < 0.05$.

S13		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 83) = 67.77, p < 0.0001$
	Vowel Quality	$F(2, 83) = 12.09, p < 0.0001$
	Interaction	$F(2, 83) = 0.3021, p = 0.7400$
Lag	Prosodic Position	$F(1, 83) = 36.85, p < 0.0001$
	Vowel Quality	$F(2, 83) = 3.205, p = 0.04563$
	Interaction	$F(2, 83) = 0.3886, p = 0.6793$
TT	Prosodic Position	$F(1, 83) = 26.55, p < 0.0001$
	Vowel Quality	$F(2, 83) = 9.321, p = 0.0002230$
	Interaction	$F(2, 83) = 0.9033, p = 0.4092$
TD	Prosodic Position	$F(1, 83) = 61.41, p < 0.0001$
	Vowel Quality	$F(2, 83) = 1.083, p = 0.3435$
	Interaction	$F(2, 83) = 1.431, p = 0.2450$

S13 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Duration	æ	533.2	e v. æ, $p < 0.0001$
	i	439.9	i v. æ, $p = 0.002783$
	e	424.0	i v. e, $p = 0.3264$
Lag	æ	-45.53	e v. æ, $p = 0.0550$
	i	-24.90	i v. æ, $p = 0.1368$
	e	-27.31	i v. e, $p = 0.8710$
TT	æ	395.1	e v. æ, $p = 0.0005308$
	e	297.0	i v. æ, $p = 0.002718$
	i	299.9	i v. e, $p = 0.7967$

Table B.10: Measurements and statistics for results for coda laterals from Experiment 1, speaker S13. Bolded data are significant at $p < 0.05$.

S16		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 69) = 24.97, p < 0.0001$
	Vowel Quality	$F(2, 69) = 14.06, p < 0.0001$
	Interaction	$F(2, 69) = 2.742, p = 0.07147$
Lag	Prosodic Position	$F(1, 69) = 17.91, p < 0.0001$
	Vowel Quality	$F(2, 69) = 5.006, p = 0.009333$
	Interaction	$F(2, 69) = 0.4997, p = 0.6089$
TT	Prosodic Position	$F(1, 69) = 21.67, p < 0.0001$
	Vowel Quality	$F(2, 69) = 16.33, p < 0.0001$
	Interaction	$F(2, 69) = 0.4407, p = 0.6454$
TD	Prosodic Position	$F(1, 69) = 32.10, p < 0.0001$
	Vowel Quality	$F(2, 69) = 3.160, p = 0.04863$
	Interaction	$F(2, 69) = 1.020, p = 0.3661$

S16 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Duration	æ	508.8	e v. æ, $p = 0.004710$
	i	360.4	i v. æ, $p < 0.0001$
	e	439.6	i v. e, $p = 0.4059$
Lag	æ	-116.6	e v. æ, $p = 0.1405$
	e	-101.6	i v. æ, $p = 0.007883$
	i	-73.93	i v. e, $p = 0.6998$
TT	æ	401.8	e v. æ, $p = 0.00103$
	i	258.0	i v. æ, $p < 0.0001$
	e	322.0	i v. e, $p = 0.5262$
TD	æ	350.7	e v. æ, $p = 0.4531$
	e	352.2	i v. æ, $p = 0.03784$
	i	285.4	i v. e, $p = 0.5776$

Table B.11: Measurements and statistics for results for coda laterals from Experiment 1, speaker S16. Bolded data are significant at $p < 0.05$.

B.2 Initial Laterals

S01				
Measure	Variable		ANOVA	
Duration	Prosodic Position		$F(1, 45) = 4.525, p = 0.03892$	
	Vowel Quality		$F(2, 45) = 1.198, p = 0.3112$	
	Interaction		$F(2, 45) = 3.541, p = 0.03731$	
Lag	Prosodic Position		$F(1, 45) = 0.9650, p = 0.3312$	
	Vowel Quality		$F(2, 45) = 0.3438, p = 0.7109$	
	Interaction		$F(2, 45) = 0.0273, p = 0.9731$	
TT	Prosodic Position		$F(1, 45) = 0.7094, p = 0.4041$	
	Vowel Quality		$F(2, 45) = 3.499, p = 0.03868$	
	Interaction		$F(2, 45) = 8.318, p = 0.0008431$	
TD	Prosodic Position		$F(1, 45) = 5.965, p = 0.01858$	
	Vowel Quality		$F(2, 45) = 0.5003, p = 0.6097$	
	Interaction		$F(2, 45) = 3.657, p = 0.03377$	

S01 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
TT	æ	289.7	e v. æ, $p = 0.3167$
	e	257.4	i v. æ, $p = 0.03024$
	i	225.4	i v. e, $p = 0.4099$

S01 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
Duration	i	312.8	412.4	$p = 0.3705$
	e	423.1	388.7	$p = 0.9713$
	æ	353.1	481.1	$p = 0.1020$
TT	i	183.0	273.9	$p = 0.1417$
	e	301.6	217.7	$p = 0.1149$
	æ	259.9	332.3	$p = 0.3040$
TD	i	263.3	358.6	$p = 0.6015$
	e	358.4	329.7	$p = 0.9939$
	æ	225.9	404.4	$p = 0.03227$

Table B.12: Measurements and statistics for results for onset laterals from Experiment 1, speaker S01. Bolded data are significant at $p < 0.05$.

S02		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 49) = 3.551, p = 0.06546$
	Vowel Quality	$F(2, 49) = 1.607, p = 0.2109$
	Interaction	$F(2, 49) = 0.5838, p = 0.5616$
Lag	Prosodic Position	$F(1, 49) = 2.321, p = 0.1341$
	Vowel Quality	$F(2, 49) = 2.660, p = 0.08003$
	Interaction	$F(2, 49) = 1.016, p = 0.3696$
TT	Prosodic Position	$F(1, 49) = 1.334, p = 0.2536$
	Vowel Quality	$F(2, 49) = 2.451, p = 0.09674$
	Interaction	$F(2, 49) = 0.3210, p = 0.7269$
TD	Prosodic Position	$F(1, 49) = 1.665, p = 0.2030$
	Vowel Quality	$F(2, 49) = 2.614, p = 0.08340$
	Interaction	$F(2, 49) = 0.0042, p = 0.9958$

Table B.13: Measurements and statistics for results for onset laterals from Experiment 1, speaker S02. Bolded data are significant at $p < 0.05$.

S04				
Measure	Variable		ANOVA	
Duration	Prosodic Position		$F(1, 52) = 6.373, p = 0.01468$	
	Vowel Quality		$F(2, 52) = 2.870, p = 0.06572$	
	Interaction		$F(2, 52) = 2.578, p = 0.08564$	
Lag	Prosodic Position		$F(1, 52) = 0.7930, p = 0.3773$	
	Vowel Quality		$F(2, 52) = 2.072, p = 0.1362$	
	Interaction		$F(2, 52) = 1.278, p = 0.2871$	
TT	Prosodic Position		$F(1, 52) = 1.088, p = 0.3018$	
	Vowel Quality		$F(2, 52) = 2.728, p = 0.07474$	
	Interaction		$F(2, 52) = 4.672, p = 0.01362$	
TD	Prosodic Position		$F(1, 52) = 11.41, p = 0.001390$	
	Vowel Quality		$F(2, 52) = 0.8221, p = 0.4451$	
	Interaction		$F(2, 52) = 0.06803, p = 0.03377$	
S04 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
TT	i	196.1	253.3	$p = 0.3151$
	e	257.5	269.0	$p = 0.06457$
	æ	172.0	309.4	$p = 0.7215$

Table B.14: Measurements and statistics for results for onset laterals from Experiment 1, speaker S04. Bolded data are significant at $p < 0.05$.

S06		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 49) = 2.098, p = 0.1538$
	Vowel Quality	$F(2, 49) = 1.015, p = 0.3700$
	Interaction	$F(2, 49) = 0.2816, p = 0.7558$
Lag	Prosodic Position	$F(1, 49) = 0.1206, p = 0.7299$
	Vowel Quality	$F(2, 49) = 2.004, p = 0.1457$
	Interaction	$F(2, 49) = 0.0926, p = 0.9117$
TT	Prosodic Position	$F(1, 49) = 1.323, p = 0.2556$
	Vowel Quality	$F(2, 49) = 1.668, p = 0.1992$
	Interaction	$F(2, 49) = 1.594, p = 0.2135$
TD	Prosodic Position	$F(1, 49) = 3.174, p = 0.08101$
	Vowel Quality	$F(2, 49) = 0.4330, p = 0.6510$
	Interaction	$F(2, 49) = 0.0268, p = 0.9736$

Table B.15: Measurements and statistics for results for onset laterals from Experiment 1, speaker S06. Bolded data are significant at $p < 0.05$.

S07			
Measure	Variable	ANOVA	
Duration	Prosodic Position	$F(1, 42) = 4.308, p = 0.04408$	
	Vowel Quality	$F(2, 42) = 1.620, p = 0.2101$	
	Interaction	$F(2, 42) = 1.288, p = 0.2865$	
Lag	Prosodic Position	$F(1, 42) = 0.2443, p = 0.6244$	
	Vowel Quality	$F(2, 42) = 3.401, p = 0.04275$	
	Interaction	$F(2, 42) = 4.037, p = 0.02491$	
TT	Prosodic Position	$F(1, 42) = 1.777, p = 0.1897$	
	Vowel Quality	$F(2, 42) = 0.9783, p = 0.3844$	
	Interaction	$F(2, 42) = 0.2711, p = 0.7639$	
TD	Prosodic Position	$F(1, 42) = 6.8037, p = 0.01255$	
	Vowel Quality	$F(2, 42) = 1.605, p = 0.2129$	
	Interaction	$F(2, 42) = 0.9705, p = 0.3872$	

S07 Effect of Vowel Quality (ms)			
Measure	Vowel	Tukey	
Lag	æ	54.14	e v. æ, $p = 0.1912$
	i	32.82	i v. æ, $p = 0.8096$
	e	123.4	i v. e, $p = 0.04076$

S07 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
Lag	i	56.22	6.500	$p = 0.9289$
	e	69.57	161.1	$p = 0.5166$
	æ	136.4	8.444	$p = 0.2852$

Table B.16: Measurements and statistics for results for onset laterals from Experiment 1, speaker S07. Bolded data are significant at $p < 0.05$.

S08		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 48) = 30.64, p < 0.0001$
	Vowel Quality	$F(2, 48) = 8.971, p = 0.0004895$
	Interaction	$F(2, 48) = 0.1665, p = 0.8471$
Lag	Prosodic Position	$F(1, 48) = 0.0858, p = 0.7709$
	Vowel Quality	$F(2, 48) = 1.0653, p = 0.3526$
	Interaction	$F(2, 48) = 1.710, p = 0.1918$
TT	Prosodic Position	$F(1, 48) = 8.106, p = 0.006475$
	Vowel Quality	$F(2, 48) = 7.164, p = 0.001894$
	Interaction	$F(2, 48) = 1.426, p = 0.2503$
TD	Prosodic Position	$F(1, 48) = 30.27, p < 0.0001$
	Vowel Quality	$F(2, 48) = 4.026, p = 0.02419$
	Interaction	$F(2, 48) = 0.1961, p = 0.8226$

S08 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Duration	æ	480.2	e v. æ, $p = 0.01239$
	e	397.2	i v. æ, $p = 0.0004637$
	i	359.5	i v. e, $p = 0.5684$
TT	æ	320.4	e v. æ, $p = 0.001294$
	i	274.3	i v. æ, $p = 0.2700$
	e	213.4	i v. e, $p = 0.06957$
TD	æ	339.2	e v. æ, $p = 0.3993$
	e	308.3	i v. æ, $p = 0.01809$
	i	262.3	i v. e, $p = 0.3143$

Table B.17: Measurements and statistics for results for onset laterals from Experiment 1, speaker S08. Bolded data are significant at $p < 0.05$.

S09		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 37) = 2.381, p = 0.1314$
	Vowel Quality	$F(2, 37) = 0.4140, p = 0.6640$
	Interaction	$F(2, 37) = 0.7624, p = 0.4737$
Lag	Prosodic Position	$F(1, 37) = 0.0292, p = 0.8653$
	Vowel Quality	$F(2, 37) = 0.4529, p = 0.6393$
	Interaction	$F(2, 37) = 0.7279, p = 0.4897$
TT	Prosodic Position	$F(1, 37) = 2.719, p = 0.1076$
	Vowel Quality	$F(2, 37) = 1.0935, p = 0.3456$
	Interaction	$F(2, 37) = 0.1559, p = 0.8562$
TD	Prosodic Position	$F(1, 37) = 0.9378, p = 0.3391$
	Vowel Quality	$F(2, 37) = 1.5830, p = 0.2189$
	Interaction	$F(2, 37) = 0.6716, p = 0.5170$

Table B.18: Measurements and statistics for results for onset laterals from Experiment 1, speaker S09. Bolded data are significant at $p < 0.05$.

S10		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 31) = 3.020, p = 0.09216$
	Vowel Quality	$F(2, 31) = 1.909, p = 0.1653$
	Interaction	$F(2, 31) = 2.529, p = 0.09609$
Lag	Prosodic Position	$F(1, 31) = 5.218, p = 0.02936$
	Vowel Quality	$F(2, 31) = 3.011, p = 0.06380$
	Interaction	$F(2, 31) = 2.231, p = 0.1244$
TT	Prosodic Position	$F(1, 31) = 1.086, p = 0.3054$
	Vowel Quality	$F(2, 31) = 0.1737, p = 0.8414$
	Interaction	$F(2, 31) = 0.3981, p = 0.6750$
TD	Prosodic Position	$F(1, 31) = 6.727, p = 0.01437$
	Vowel Quality	$F(2, 31) = 3.733, p = 0.03526$
	Interaction	$F(2, 31) = 2.484, p = 0.09988$

S10 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
TD	æ	364.0	e v. æ, $p = 0.9999$
	e	358.5	i v. æ, $p = 0.08382$
	i	488.0	i v. e, $p = 0.05830$

Table B.19: Measurements and statistics for results for onset laterals from Experiment 1, speaker S10. Bolded data are significant at $p < 0.05$.

S12		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 77) = 1.008, p = 0.3185$
	Vowel Quality	$F(2, 77) = 8.668, p = 0.0004027$
	Interaction	$F(2, 77) = 0.2139, p = 0.8079$
Lag	Prosodic Position	$F(1, 77) = 7.779, p = 0.006656$
	Vowel Quality	$F(2, 77) = 6.712, p = 0.002056$
	Interaction	$F(2, 77) = 0.2683, p = 0.7654$
TT	Prosodic Position	$F(1, 77) = 0.1003, p = 0.7524$
	Vowel Quality	$F(2, 77) = 7.386, p = 0.001163$
	Interaction	$F(2, 77) = 5.836, p = 0.004365$
TD	Prosodic Position	$F(1, 77) = 12.24, p = 0.007812$
	Vowel Quality	$F(2, 77) = 4.558, p = 0.01346$
	Interaction	$F(2, 77) = 7.752, p = 0.0008569$

S12 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Duration	æ	423.81	e v. æ, $p = 0.2624$
	e	387.0	i v. æ, $p = 0.02222$
	i	485.8	i v. e, $p = 0.0002987$
Lag	æ	9.871	e v. æ, $p = 0.3802$
	e	30.36	i v. æ, $p = 0.04560$
	i	-26.30	i v. e, $p = 0.001630$
TT	æ	353.8	e v. æ, $p = 0.07222$
	i	397.9	i v. æ, $p = 0.1873$
	e	297.0	i v. e, $p = 0.0007318$
TD	æ	350.0	e v. æ, $p = 0.5501$
	e	328.4	i v. æ, $p = 0.1087$
	i	390.7	i v. e, $p = 0.01154$

S12 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
TT	i	364.2	434.1	$p = 0.4007$
	e	287.2	307.6	$p = 0.9945$
	æ	399.8	304.7	$p = 0.06887$
TD	i	400.0	380.7	$p = 0.9860$
	e	306.9	351.7	$p = 0.6836$
	æ	283.8	420.7	$p = 0.0000480$

Table B.20: Measurements and statistics for results for onset laterals from Experiment 1, speaker S12. Bolded data are significant at $p < 0.05$.

S13		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 87) = 18.18, p < 0.0001$
	Vowel Quality	$F(2, 87) = 8.884, p = 0.0003084$
	Interaction	$F(2, 87) = 1.778, p = 0.1751$
Lag	Prosodic Position	$F(1, 87) = 0.2973, p = 0.5870$
	Vowel Quality	$F(2, 87) = 1.830, p = 0.1665$
	Interaction	$F(2, 87) = 1.690, p = 0.1905$
TT	Prosodic Position	$F(1, 87) = 0.6479, p = 0.4230$
	Vowel Quality	$F(2, 87) = 6.572, p = 0.002198$
	Interaction	$F(2, 87) = 3.429, p = 0.03686$
TD	Prosodic Position	$F(1, 87) = 68.28, p < 0.0001$
	Vowel Quality	$F(2, 87) = 0.3546, p = 0.7025$
	Interaction	$F(2, 87) = 8.578, p = 0.000398$

S13 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Duration	æ	450.8	e v. æ, $p = 0.0002208$
	i	414.5	i v. æ, $p = 0.2796$
	e	354.5	i v. e, $p = 0.02691$
TT	æ	367.9	e v. æ, $p = 0.001662$
	i	341.9	i v. æ, $p = 0.3949$
	e	297.1	i v. e, $p = 0.06487$

S13 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
TT	i	327.9	355.0	$p = 0.9283$
	e	296.8	297.5	$p > 0.9999$
	æ	407.6	333.1	$p = 0.1090$
TD	i	256.0	425.6	$p = 0.0000103$
	e	301.3	354.4	$p = 0.5337$
	æ	225.0	461.3	$p < 0.0001$

Table B.21: Measurements and statistics for results for onset laterals from Experiment 1, speaker S13. Bolded data are significant at $p < 0.05$.

S16		
Measure	Variable	ANOVA
Duration	Prosodic Position	$F(1, 82) = 2.922, p = 0.09116$
	Vowel Quality	$F(2, 82) = 0.0303, p = 0.9701$
	Interaction	$F(2, 82) = 2.119, p = 0.1266$
Lag	Prosodic Position	$F(1, 82) = 6.632, p = 0.01181$
	Vowel Quality	$F(2, 82) = 9.943, p = 0.0001361$
	Interaction	$F(2, 82) = 2.556, p = 0.08380$
TT	Prosodic Position	$F(1, 82) = 31.86, p < 0.0001$
	Vowel Quality	$F(2, 82) = 2.188, p = 0.1187$
	Interaction	$F(2, 82) = 1.897, p = 0.1566$
TD	Prosodic Position	$F(1, 82) = 41.51, p < 0.0001$
	Vowel Quality	$F(2, 82) = 1.609, p = 0.2063$
	Interaction	$F(2, 82) = 3.645, p = 0.03045$

S16 Effect of Vowel Quality (ms)			
Measure	Vowel		Tukey
Lag	æ	51.90	e v. æ, $p = 0.9868$
	e	48.26	i v. æ, $p = 0.0005178$
	i	-11.16	i v. e, $p = 0.001142$

S16 Interactions (ms)				
Measure	Vowel	Word-Final	Sentence-Final	Tukey
TD	i	279.4	418.4	$p = 0.0005413$
	e	364.3	414.6	$p = 0.6995$
	æ	276.3	454.6	$p < 0.0001$

Table B.22: Measurements and statistics for results for onset laterals from Experiment 1, speaker S16. Bolded data are significant at $p < 0.05$.

APPENDIX C

Experiment 2: Synthetic Stimuli

This appendix lists the acoustic characteristics of the synthetic target stimuli used in the perception experiment. Details include values of the first four formants at 10 ms intervals, as taken from Hlsyn files, and resulting spectrograms.

Time	F1	F2	F3	F4
110	517.4873	1701.4227	2563.7787	3285.2719
120	507.0538	1643.7519	2537.9223	3181.917
130	497.9312	1599.0966	2567.7786	3196.3765
140	482.206	1555.5858	2589.9031	3153.508
150	467.8402	1517.954	2625.9909	3167.4858
160	452.3129	1488.6246	2643.2216	3128.2099
170	440.2557	1476.2894	2669.298	3138.0813
180	428.4355	1502.0727	2698.0935	3169.1358
190	415.6446	1533.9061	2719.6596	3170.8135
200	404.5556	1549.3128	2723.7482	3066.3464
210	414.6335	1512.7881	2789.9284	3125.4965
220	433.0477	1428.4766	2849.7951	3100.7877
230	445.3548	1339.8496	2916.6736	3113.9085
240	452.8522	1289.8635	3001.5489	3234.0725
250	446.8412	1272.0001	3039.9727	3239.5036
260	450.5813	1290.5274	3045.1461	3223.2247
270	511.774	1335.8634	3062.5625	3604.282
280	574.3914	1387.4063	3004.6507	3527.7733
290	653.5997	1465.7426	2958.7998	3550.5139
300	669.6439	1489.8677	2927.8589	3557.9402
310	680.1411	1510.8507	2903.1976	3564.1126
320	687.2333	1528.6705	2883.4634	3568.6389
330	692.1755	1542.8512	2868.6175	3570.9578
340	695.9185	1554.2178	2856.7798	3572.6211
350	699.133	1562.8022	2847.4704	3574.5497
360	702.4634	1568.3921	2840.7479	3577.7161
370	707.145	1570.1505	2836.8566	3586.2148
380	712.9485	1568.0467	2834.7116	3599.2459
390	718.8514	1560.9065	2838.2689	3612.8013
400	727.7194	1548.2508	2839.5738	3637.14
410	742.2402	1530.4792	2836.1864	3659.1374
420	756.7692	1497.6778	2831.4186	3709.5787
430	782.9789	1455.1131	2870.2482	3866.5289
440	777.3972	1375.2405	2859.3267	3951.3428
450	665.2063	1293.4222	2770.852	3899.595
460	574.802	1353.9211	2838.8897	3961.3148

Table C.1: First four formants, for stimulus with -60 ms gestural lag, taken from the HLsyn (.HL) file.

Time	F1	F2	F3	F4
110	517.4873	1701.4227	2563.7787	3285.2719
120	507.0538	1643.7519	2537.9223	3181.917
130	497.9312	1599.0966	2567.7786	3196.3765
140	482.206	1555.5858	2589.9031	3153.508
150	466.7346	1526.2502	2630.2008	3170.725
160	445.7477	1532.772	2657.5857	3131.5278
170	427.9565	1554.3184	2694.7163	3144.2758
180	408.2351	1572.8878	2728.5938	3194.8627
190	389.6923	1574.5673	2741.0644	3110.6227
200	376.5572	1560.8054	2754.0496	3035.922
210	383.6785	1488.4242	2798.0348	3009.3891
220	398.1791	1387.7108	2863.4517	3034.2029
230	408.3566	1302.1252	2935.0257	3081.4522
240	433.4382	1274.1197	3010.0757	3234.1572
250	499.2588	1294.1005	3044.04	3615.7555
260	563.2216	1335.2795	3023.1642	3510.3161
270	637.4014	1411.1791	2942.1985	3410.6314
280	662.5488	1444.4537	2954.8286	3525.1978
290	676.709	1468.6723	2926.3069	3538.7843
300	685.7656	1490.9401	2903.3023	3549.5463
310	691.5427	1510.6884	2884.5071	3557.4373
320	695.2003	1527.7395	2869.2032	3562.8468
330	697.6016	1541.5717	2857.4828	3565.8296
340	699.4643	1552.7236	2848.0171	3568.2683
350	701.2761	1561.1965	2840.5621	3570.9361
360	703.5501	1566.7667	2835.3167	3574.8174
370	707.4303	1568.5707	2832.6269	3584.0167
380	712.598	1566.5381	2831.4883	3597.7517
390	718.0301	1559.5153	2835.8875	3611.9393
400	726.5624	1547.0203	2837.8576	3636.7284
410	741.0862	1529.4982	2834.5433	3658.4033
420	755.2338	1496.7879	2830.3868	3709.4684
430	780.7698	1454.1952	2870.2547	3867.2332
440	775.3739	1374.7097	2860.0927	3952.2923
450	661.3274	1292.9323	2771.1536	3900.4051
460	571.1261	1353.6016	2839.8557	3962.3336

Table C.2: First four formants, for stimulus with -40 ms gestural lag, taken from the HLsyn (.HL) file.

Time	F1	F2	F3	F4
110	517.4873	1701.4227	2563.7787	3285.2719
120	507.0538	1643.7519	2537.9223	3181.917
130	495.9862	1605.2207	2570.7851	3190.7335
140	471.9952	1594.5624	2615.3025	3165.9006
150	446.9156	1601.2188	2672.8999	3215.6964
160	421.0753	1601.3387	2702.567	3166.7196
170	397.9253	1594.8124	2723.7818	3094.0013
180	377.6709	1583.0905	2737.8321	3023.9515
190	359.9995	1566.7605	2747.0203	2968.9182
200	347.8966	1536.2222	2759.4131	2934.4909
210	354.4882	1449.2398	2807.1646	2950.756
220	382.2689	1367.6201	2870.0652	3016.4636
230	447.1318	1332.0555	2943.7079	3294.2288
240	520.63	1348.647	2990.1968	3524.6379
250	596.7688	1406.3909	2965.8012	3381.5405
260	631.8093	1403.7646	2943.9097	3391.8814
270	660.4996	1427.2624	2960.4327	3499.7161
280	676.5582	1447.7415	2934.0397	3516.7031
290	686.4321	1470.2077	2911.0652	3533.1283
300	692.3109	1491.4107	2892.6685	3545.6133
310	701.229	1518.7133	2887.4662	3564.5761
320	703.103	1535.0828	2873.8441	3569.6259
330	704.1321	1548.4169	2863.2622	3572.3415
340	704.9225	1559.1502	2854.5599	3574.5296
350	705.895	1567.2475	2847.562	3576.9026
360	707.5203	1572.4282	2842.5588	3580.4255
370	710.9468	1573.7901	2839.8876	3589.0869
380	715.769	1571.2642	2838.5812	3602.0822
390	720.8524	1563.6684	2842.5744	3615.31
400	732.7544	1552.2769	2838.356	3629.1559
410	743.0835	1532.2523	2839.8925	3659.4725
420	753.2242	1496.8848	2835.6744	3713.0961
430	778.9126	1454.2881	2875.5524	3871.8822
440	774.0784	1374.9633	2865.0019	3955.1698
450	660.9451	1293.3616	2775.1498	3900.3347
460	571.1249	1354.1178	2843.5908	3963.2115

Table C.3: First four formants, for stimulus with -20 ms gestural lag, taken from the HLsyn (.HL) file.

Time	F1	F2	F3	F4
110	512.754	1681.8289	2545.0611	3231.4504
120	491.6994	1671.7952	2581.8842	3213.3875
130	475.2788	1669.3203	2607.8405	3193.9499
140	435.394	1654.9285	2684.9157	3257.4546
150	406.1656	1634.271	2719.6594	3232.5415
160	382.8221	1603.706	2711.8537	2991.7047
170	362.3572	1579.9267	2726.1936	2940.5229
180	345.1332	1554.3582	2735.6758	2904.1713
190	330.8039	1528.5141	2742.0174	2879.5365
200	333.8418	1513.2632	2752.0177	2888.3751
210	383.5	1477.946	2821.6282	3233.0263
220	460.0005	1506.0725	2835.0343	3159.7767
230	512.4329	1456.5767	2905.0421	3316.0303
240	567.474	1442.6427	2949.177	3375.7575
250	611.0447	1414.6018	2946.1403	3378.701
260	641.9233	1408.0233	2928.3695	3390.2968
270	671.7164	1441.2304	2962.9533	3511.0989
280	685.7173	1459.3954	2940.3839	3527.9873
290	693.9212	1480.5316	2920.3524	3544.434
300	698.4534	1500.9326	2903.9895	3556.9413
310	700.859	1519.6718	2890.0496	3565.8815
320	702.0485	1536.2266	2878.1144	3572.0323
330	702.6689	1549.7958	2868.7557	3575.5849
340	703.2711	1560.7761	2860.9075	3578.3576
350	704.2108	1569.1084	2854.4903	3581.0992
360	705.9161	1574.4838	2849.8351	3584.7912
370	709.5	1575.9703	2847.3559	3593.3747
380	714.5522	1573.5168	2846.1221	3606.0734
390	719.938	1565.9642	2850.0983	3618.8092
400	728.4345	1552.7387	2851.5179	3641.9396
410	738.36	1532.5484	2854.2809	3673.3144
420	758.663	1501.627	2840.6807	3705.7014
430	775.8063	1455.353	2893.2883	3890.1832
440	773.7563	1376.0179	2874.0177	3960.8887
450	673.3182	1297.1218	2813.2846	3972.704
460	574.8953	1355.4233	2849.5978	3963.5039

Table C.4: First four formants, for stimulus with 0 ms gestural lag, taken from the HLsyn (.HL) file.

Time	F1	F2	F3	F4
110	476.4424	1808.5614	2543.9572	3266.7056
120	448.0246	1804.852	2520.6448	3177.2938
130	397.7447	1779.804	2591.3094	3167.0653
140	378.0247	1726.1133	2546.5655	2942.4116
150	359.4698	1678.5993	2517.4412	2840.9619
160	343.6392	1634.7052	2516.6225	2808.7037
170	329.841	1589.7394	2511.2291	2797.6665
180	333.0529	1585.0386	2710.7924	2871.2721
190	360.9716	1583.0255	2755.2784	3197.1748
200	383.478	1570.7732	2756.9172	3248.3264
210	427.8952	1566.2218	2778.6946	3209.2061
220	474.7173	1521.3382	2842.9034	3331.1693
230	522.0441	1460.5292	2891.9666	3332.6431
240	574.6524	1449.1957	2936.6226	3372.7206
250	619.7618	1432.2216	2944.0534	3386.1731
260	652.8483	1436.0606	2973.3541	3487.0881
270	674.7878	1442.5509	2955.4947	3505.1503
280	688.4663	1462.1617	2939.2076	3529.2043
290	695.739	1482.8351	2921.3638	3546.9681
300	699.5989	1503.1012	2906.4574	3560.0715
310	701.5185	1521.8373	2893.4689	3569.2563
320	702.3457	1538.3856	2882.1845	3575.4797
330	702.6894	1551.9161	2873.2043	3579.0125
340	703.0594	1562.8062	2865.5177	3581.6428
350	703.7925	1570.9852	2859.0835	3584.1829
360	705.3056	1576.124	2854.322	3587.6368
370	708.729	1577.3453	2851.6461	3595.9414
380	713.6142	1574.5877	2850.145	3608.321
390	718.529	1566.4323	2853.6005	3620.7413
400	727.12	1553.1735	2854.9232	3643.5172
410	737.121	1532.9287	2857.5649	3674.5681
420	753.1959	1500.3275	2851.18	3717.8935
430	777.9773	1456.832	2895.2421	3886.7852
440	772.3942	1376.1559	2879.1309	3965.1144
450	662.3219	1294.7353	2782.7214	3897.4056
460	574.5163	1356.2811	2852.9342	3963.851

Table C.5: First four formants, for stimulus with +20 ms gestural lag, taken from the HLsyn (.HL) file.

Time	F1	F2	F3	F4
110	385.2496	1877.2249	2574.6708	3191.0425
120	361.4223	1863.4627	2453.9501	3033.7144
130	343.2059	1852.7056	2456.0634	2960.4251
140	329.7627	1817.0831	2468.6381	2871.3787
150	320.0964	1747.2768	2446.0723	2795.634
160	325.8568	1676.6348	2480.4767	2789.1914
170	356.681	1625.6984	2693.6557	3089.5293
180	377.7502	1571.5925	2703.8014	3179.4149
190	395.4518	1560.7688	2702.7048	3196.6097
200	404.9815	1595.944	2730.7264	3269.5497
210	435.881	1560.953	2780.0961	3364.6812
220	482.3837	1501.7759	2831.7631	3347.0826
230	531.2379	1448.9316	2875.0929	3328.3453
240	584.8609	1449.7499	2911.3225	3354.404
250	627.2824	1417.6199	2903.9876	3360.9019
260	664.4996	1434.222	2927.0249	3439.3544
270	685.2603	1435.9225	2910.9841	3472.3733
280	698.0263	1452.9799	2892.1924	3496.5992
290	704.9209	1473.745	2877.5603	3517.797
300	708.3485	1494.4675	2866.5631	3533.6701
310	709.5632	1513.4404	2857.0632	3545.1938
320	710.9302	1530.792	2843.1866	3543.8303
330	710.4495	1544.4503	2837.0887	3549.1107
340	708.4408	1554.811	2837.4257	3563.529
350	708.3145	1563.1711	2833.2284	3567.8168
360	709.0427	1568.6296	2830.5197	3573.0104
370	711.8288	1570.3805	2829.8495	3583.3258
380	716.1386	1568.2451	2830.1897	3597.8273
390	720.7672	1560.9749	2835.5508	3612.4029
400	728.5425	1548.1317	2838.1473	3637.395
410	741.8307	1530.0138	2835.9691	3660.4324
420	754.7027	1496.6884	2832.6368	3712.5875
430	769.0471	1445.5765	2820.2537	3766.1849
440	772.4703	1373.6813	2859.1219	3951.7791
450	658.4173	1292.3784	2764.5645	3893.1282
460	570.5502	1354.5229	2844.071	3963.2183

Table C.6: First four formants, for stimulus with +40 ms gestural lag, taken from the HLsyn (.HL) file.

Time	F1	F2	F3	F4
110	335.2297	1917.9083	2416.1794	3021.611
120	315.6358	1924.0271	2391.0583	2985.5103
130	303.0808	1910.5196	2394.0242	2919.6062
140	309.6279	1852.1138	2418.5543	2856.8367
150	347.8854	1766.1135	2602.7624	3080.2359
160	376.0786	1658.2706	2628.5651	3174.4447
170	397.5846	1604.1655	2648.2894	3201.1596
180	400.0093	1561.0928	2678.7556	3239.7849
190	400.4575	1529.7174	2700.7494	3268.0251
200	409.503	1599.1284	2733.537	3388.9647
210	438.8102	1549.2504	2774.8402	3378.0002
220	484.1193	1494.0715	2827.2535	3347.3225
230	533.2431	1444.2971	2869.2213	3323.0875
240	588.2689	1453.9609	2899.1601	3335.5198
250	631.0609	1422.3938	2888.6436	3343.8392
260	662.8404	1414.0971	2866.821	3363.8414
270	689.1038	1438.9164	2897.5588	3457.7346
280	701.5861	1455.1841	2881.5421	3485.629
290	705.9334	1472.765	2872.5167	3514.7099
300	708.9569	1493.2585	2862.1446	3530.9885
310	709.9239	1512.2537	2853.2251	3542.8408
320	709.8659	1529.2235	2845.7117	3551.6326
330	709.2681	1543.0598	2840.1004	3557.4379
340	708.6964	1554.214	2835.3815	3562.1946
350	708.6471	1562.8376	2831.721	3566.8008
360	709.4922	1568.5699	2829.4648	3572.2825
370	712.324	1570.462	2829.102	3582.7922
380	716.6763	1568.4323	2829.6879	3597.4409
390	721.3833	1561.2774	2835.2487	3612.115
400	732.6984	1550.1375	2832.7524	3628.054
410	742.7933	1530.4744	2835.3597	3659.3683
420	755.6125	1497.0822	2832.1264	3711.6382
430	769.8747	1445.8987	2819.786	3765.2604
440	773.1804	1373.901	2858.9833	3951.3355
450	661.7844	1293.3879	2774.6292	3901.0831
460	571.4041	1354.5437	2843.8626	3962.9104

Table C.7: First four formants, for stimulus with +60 ms gestural lag, taken from the HLsyn (.HL) file.

Time	F1	F2	F3	F4
110	298.6957	1956.8359	2361.7742	2992.575
120	298.9167	1943.2028	2355.6731	2982.2827
130	331.1761	1922.0407	2527.8253	3124.4373
140	362.4103	1813.8733	2530.5635	3198.5965
150	395.0401	1717.111	2547.962	3230.4232
160	403.2555	1656.498	2601.5094	3255.3236
170	409.3751	1682.1166	2659.6257	3293.8405
180	409.7919	1647.858	2687.6809	3309.2655
190	408.9828	1623.2363	2708.2452	3321.9044
200	413.1431	1625.7911	2734.0619	3411.7277
210	442.2063	1579.3252	2775.7876	3393.2497
220	486.5339	1518.8631	2829.3966	3360.0119
230	540.9636	1519.5497	2884.0346	3355.612
240	590.0663	1463.222	2899.3679	3334.993
250	632.9844	1430.2363	2890.0651	3343.9874
260	664.9228	1420.7786	2869.2937	3364.1248
270	691.6048	1445.6744	2902.1335	3460.6142
280	704.3327	1461.3405	2886.577	3488.5231
290	708.8555	1478.414	2877.6338	3517.6448
300	711.9982	1498.6211	2867.2987	3533.8496
310	713.0463	1517.396	2858.3567	3545.6175
320	712.9955	1534.1148	2850.7049	3554.2558
330	712.451	1547.8312	2845.0025	3559.9612
340	711.9767	1558.9389	2840.2329	3564.639
350	714.738	1569.9561	2834.0414	3561.7251
360	715.7003	1575.5077	2831.579	3566.8362
370	718.7288	1577.129	2830.909	3576.8172
380	723.3324	1574.7828	2831.1461	3590.8766
390	728.3159	1567.3137	2836.2811	3605.0408
400	768.5168	1589.5484	2912.1304	3740.0266
410	780.8591	1567.1656	2917.4696	3772.4553
420	796.5466	1529.9523	2917.6382	3830.4906
430	810.8466	1473.1562	2911.2083	3897.1175
440	776.8993	1375.3724	2861.9894	3952.785
450	671.8878	1295.1773	2778.4111	3899.4496
460	580.4106	1355.9163	2846.1796	3962.2349

Table C.8: First four formants, for stimulus with +80 ms gestural lag, taken from the HLsyn (.HL) file.

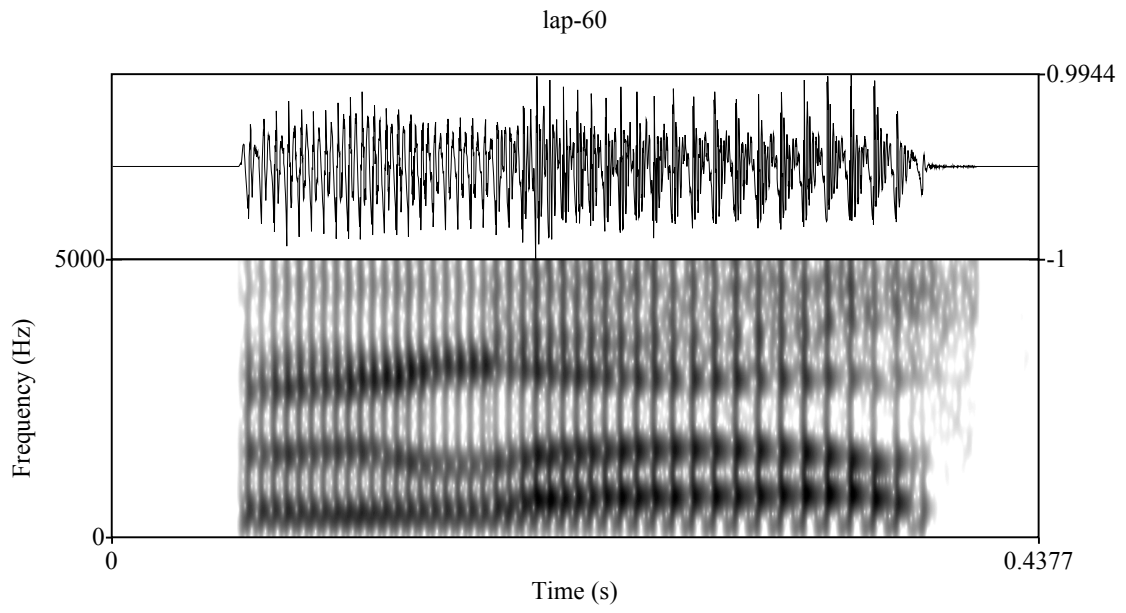


Figure C.1: Spectrogram information for “lap” with -60 ms gestural lag.

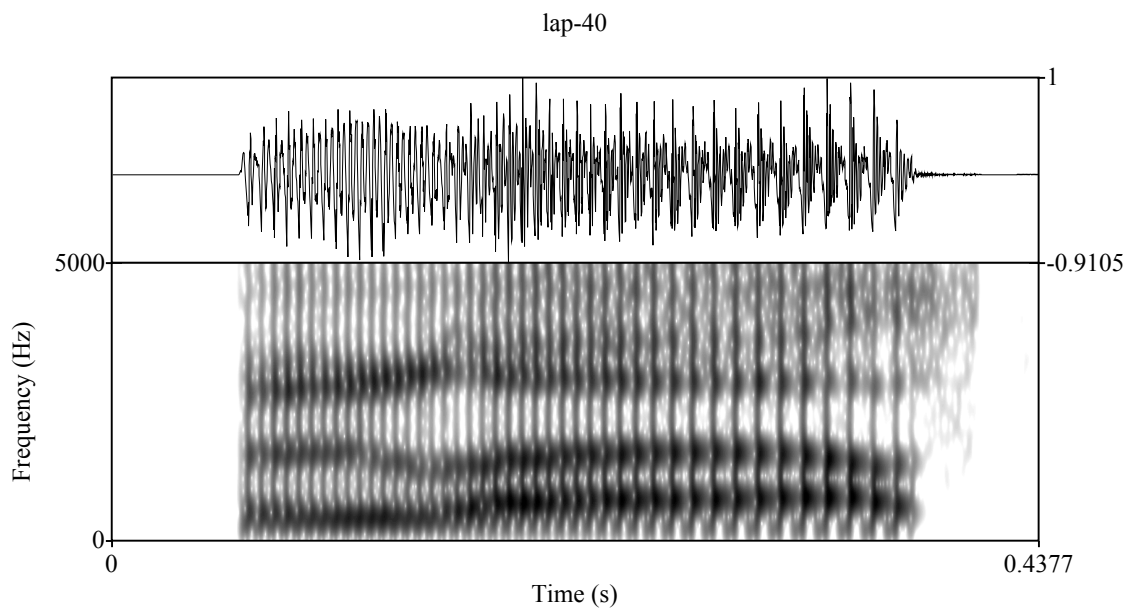


Figure C.2: Spectrogram information for “lap” with -40 ms gestural lag.

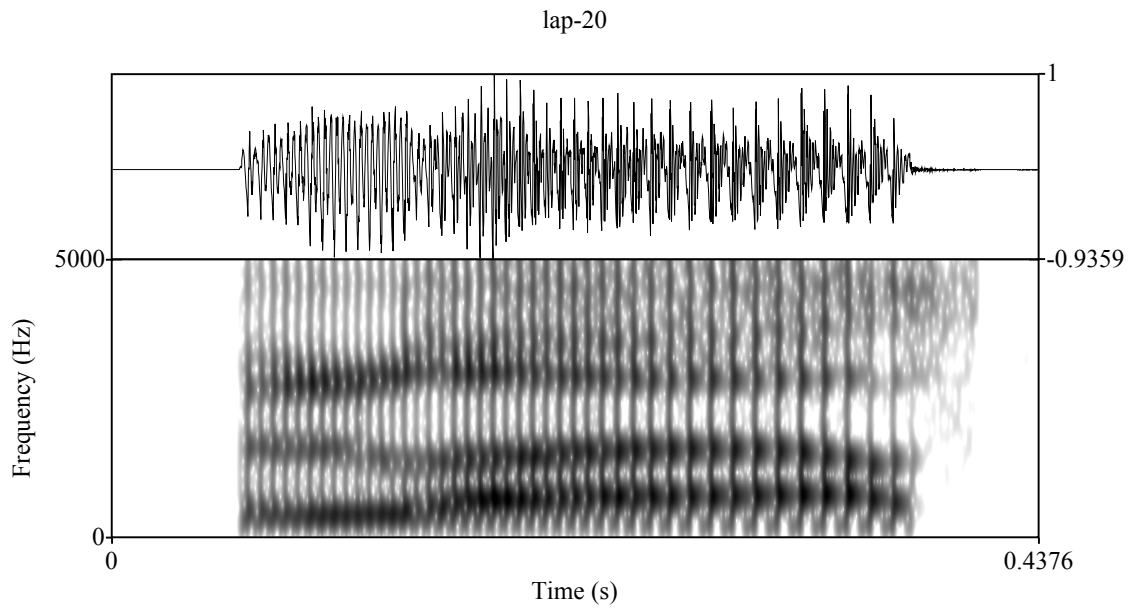


Figure C.3: Spectrogram information for “lap” with -20 ms gestural lag.

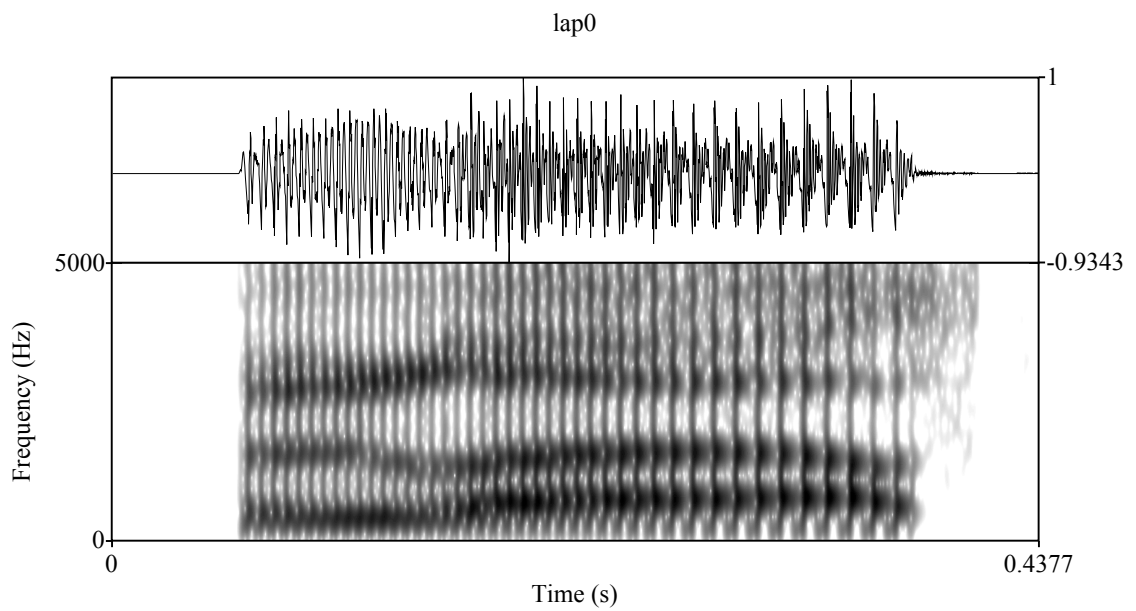


Figure C.4: Spectrogram information for “lap” with 0 ms gestural lag.

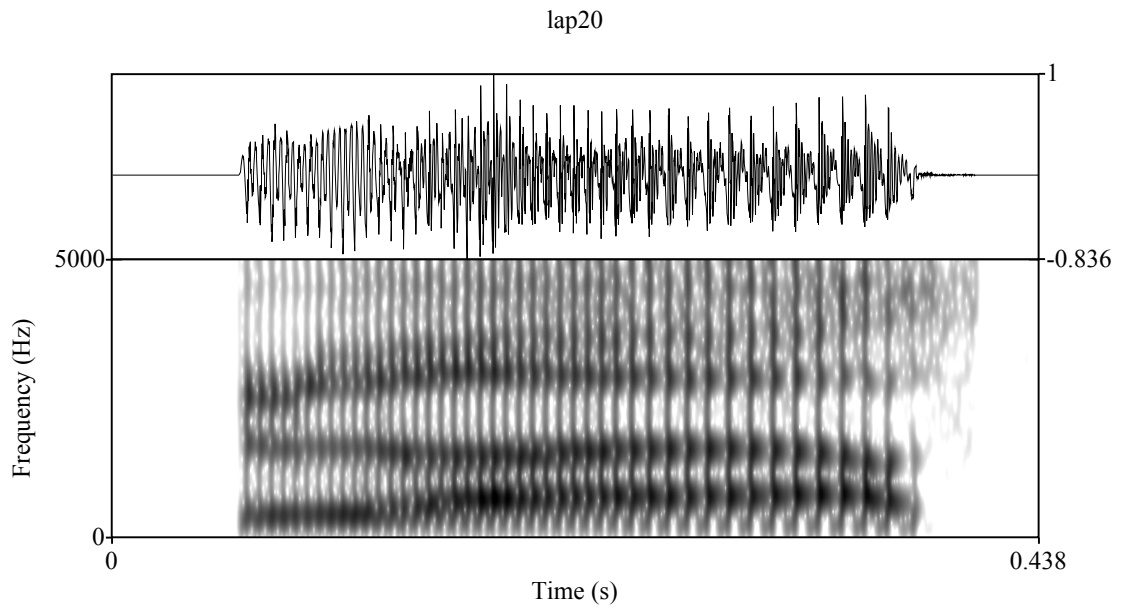


Figure C.5: Spectrogram information for “lap” with +20 ms gestural lag.

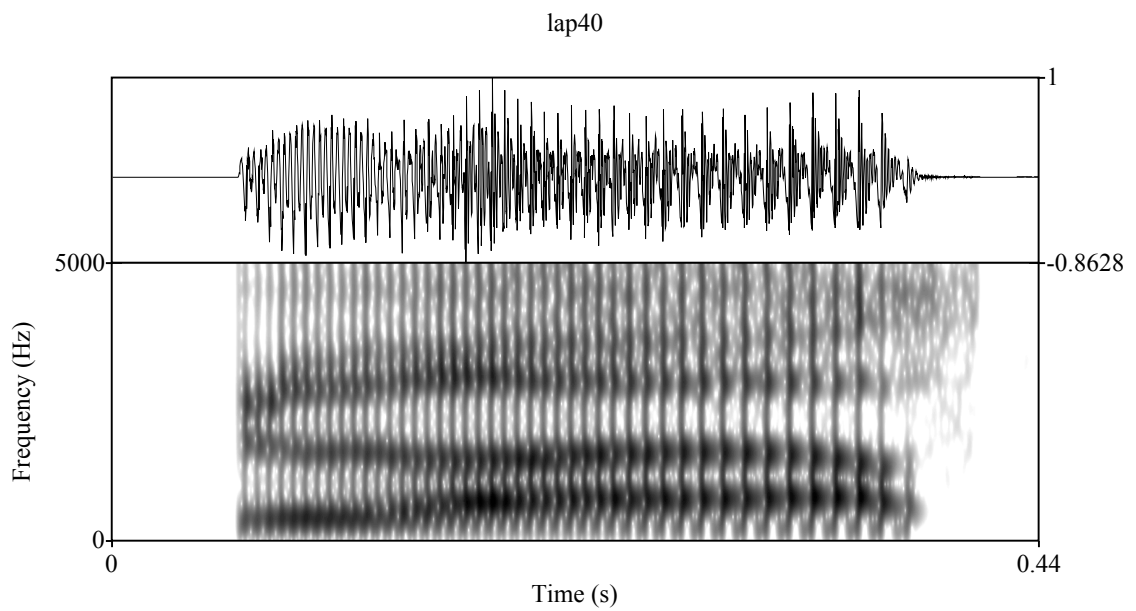


Figure C.6: Spectrogram information for “lap” with +40 ms gestural lag.

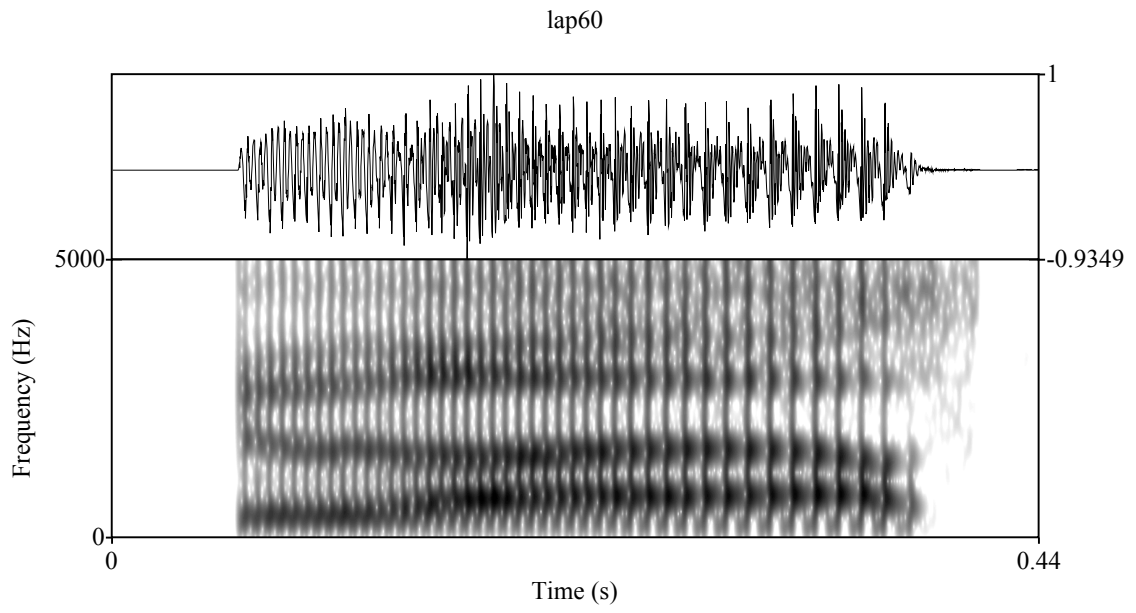


Figure C.7: Spectrogram information for “lap” with +60 ms gestural lag.

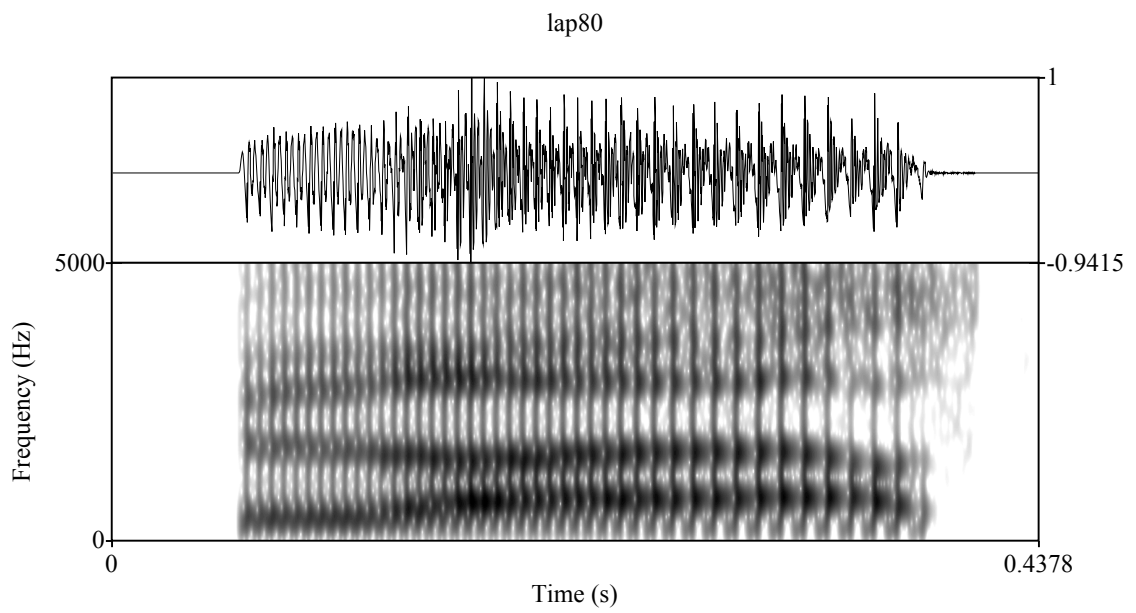


Figure C.8: Spectrogram information for “lap” with +80 ms gestural lag.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Berry, J. and Archangelli, D. Different light and dark /l/s: English and Georgian. Poster presentation, Ultrafest V, 2010.
- Boersma, P. and Weenink, D. Praat: doing phonetics by computer [Computer Program]. Version 5.0.14. Retrieved March 14, 2008 from www.praat.org, 2008.
- Botwinick, J. and Thompson, L. W. Premotor and motor components of reaction time. Journal of Experimental Psychology, 1966.
- Browman, C. and Goldstein, L. Towards an articulatory phonology. Phonology Yearbook, 3:219–252, 1986.
- Browman, C. and Goldstein, L. Articulatory phonology: an overview. Phonetica, 49:155–180, 1992.
- Browman, C. and Goldstein, L. Gestural syllable position effects. In Bell-Berti, F. and Raphael, L., editors, Producing Speech: Contemporary Issues, pages 19–33. AIP Press, New York, 1995.
- Byrd, D. and Saltzman, E. The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. Journal of Phonetics, 31:149–180, 2003.
- Byrd, D., Tobin, S., Bresch, E. and Narayanan, S. Timing effects of syllable structure and stress on nasals: A real-time MRI examination. Journal of Phonetics, 2009.
- Chitoran, I., Goldstein, L. and Byrd, D. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Papers in Laboratory Phonology VII: Phonology and Phonetics, pages 419–447. 2002.
- Cho, T. and Keating, P. Articulatory strengthening at the onset of prosodic domains in Korean. Journal of Phonetics, 28:155–190, 2001.
- Cho, T., McQueen, J. M. and Cox, E. A. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. Journal of Phonetics, 35:210–243, 2007.
- Connine, C. and Titone, D. Phoneme monitoring. Language and Cognitive Processes, 11(6):635–645, 1996.
- Fougeron, C. Articulatory properties of initial segments in several prosodic constituents in French. Journal of Phonetics, 2:109–135, 2001.
- Fougeron, C. and Keating, P. Articulatory strengthening at edges of prosodic domains. Journal of the Acoustical Society of America, 101(6):3728–3740, 1997.
- Gick, B. Articulatory correlates of ambisyllabicity in English glides and laterals. In Local, J., Ogden, R. and Temple, R., editors, Papers in Laboratory Phonology VI: Phonetic Interpretation, pages 222–236. Cambridge University Press, Cambridge, 2003.

- Gick, B., Campbell, F., Oh, S. and Tamburri-Watt, L. Towards universals in the gestural organization of syllables: A cross-linguistic study of liquids. Journal of Phonetics, 34:49–72, 2006.
- Gick, B. and Goldstein, L. Relative timing of the three gestures of North American English /r/. Journal of the Acoustical Society of America, 111:2481–2481, 2002.
- Giles, S. and Moll, K. Cinefluorographic study of selected allophones of English /l/. Phonetica, 31:206–227, 1975.
- Goldstein, L. Dynamical stability in speech production and sound change. Talk presented at the Workshop on Sound Change, Barcelona, 2010.
- Goldstein, L., Byrd, D. and Saltzman, E. The role of vocal tract gestural action units in understanding the evolution of phonology. In Arbib, M., editor, Action to Language via the Mirror Neuron System, pages 215–249. Cambridge University Press, Cambridge, 2006.
- Johnson, K. Acoustic and Auditory Phonetics. Blackwell, Oxford, 2nd edition, 2003.
- Keating, P. Phonetic encoding of prosodic structure. In Harrington, J. and Tabain, M., editors, Speech Production: Models, phonetic processes, and techniques, pages 167–186. Psychology Press, 2006.
- Keating, P., Cho, T., Fougeron, C. and Hsu, C. Domain-initial articulatory strengthening in four languages. In Local, J., Ogden, R. and Temple, R., editors, Papers in Laboratory Phonology VI: Phonetic Interpretation, pages 143–161. Cambridge University Press, Cambridge, 2003.
- Krakow, R. The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures. Ph.D. thesis, Yale University, 1989.
- Krakow, R. Physiological organization of syllables: a review. Journal of Phonetics, 27:23–54, 1999.
- Krivokapić, J. and Byrd, D. Prosodic boundaries and what they tell us about the nature of gestural goals in production. Poster presentation, Labphon 12, 2010.
- Li, M., Akgul, Y. and Kambhamettu, C. Edge Trak [Computer Program]. Version 1.0.0.4. Retrieved October 3, 2008, from vims.cis.udel.edu/EdgeTrak/, 2005.
- Lindblom, B. Economy of speech gestures. In MacNeilage, P., editor, The Production of Speech, pages 217–246. Springer-Verlag, New York, 1983.
- Lindblom, B. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W. J. and Marchal, A., editors, Speech Production and Speech Modelling, pages 403–439. Kluwer Academic, 1990.
- Macmillan, N. A. and Creelman, C. D. Detection theory: a user's guide. Lawrence Erlbaum Associates, Inc, 2005.
- Mielke, J., Baker, A. and Archangelli, D. Palatron: a technique for aligning ultrasound images of the tongue and palate. Presented at Ultrafest III in Tucson, AZ, 2005.
- Nam, H., Browman, C., Goldstein, L., Proctor, M., Rubin, P. and Saltzman, E. TADA: TAsk Dynamic Application [Computer Program]. Retrieved March 28, 2010, from Nam, H., 2010.
- Redford, M. The mandibular cycle and reversed-sonority onset clusters in Russian. In Proceedings of the XIVth International Congress of Phonetic Sciences. International Congress of Phonetic Sciences, 1999.
- Sproat, R. and Fujimura, O. Allophonic variation in English /l/ and its implications for phonetic implementation. Journal of Phonetics, 21:291–311, 1993.
- Stevens, K. N. Acoustic Phonetics. The MIT Press, 1998.

Stone, M. and Davis, E. A head and transducer support system for making ultrasound images of tongue/jaw movement. Journal of the Acoustical Society of America, 98(6):3107–3112, 1995.

Vaissière, J. Prediction of articulatory movement of the velum from phonetic input. Phonetica, 45:122–139, 1989.