Mining speech corpora for gestural timing differences as a precursor to metathesis

Tyler Lau

University of California at Berkeley
tylerlau@berkeley.edu

42nd Penn Linguistics Conference
University of Pennsylvania (Philadelphia, PA)

March 25, 2018

What is metathesis?

- The reordering of phonological segments in a word (Grammont, 1933; Hock, 1985)

$$wasp \rightarrow waps \text{ (Steriade, 2001)}$$

- While factors in both perception and production have been advanced for metathesis, recent literature and experimental approaches have primarily focused on testing the role of perception

Questions

- Theoretical: What is the role of gestural timing in metathesis?
- Methodological: How can we use speech corpora to study precursors of metathesis?

Perceptual Explanations

Two main perceptual theories for metathesis

1. Ambiguity-Attestation Hypothesis (Hume, 2004)
   - There must be ambiguity in the phonetic signal
     (also Blevins and Garrett, 1998, 2004)
   - Metathesized sequence must be licit in language–metathesis is structure-preserving
     (but cf. Blevins and Garrett (1998) for counterargument)
   - Built off Ohala's listener-based model of sound change
     Ohala (1981, 1993)
2. Perceptual Optimization (Hume, 1998; Steriade, 2001)
   - Metathesis occurs to improve perceptibility
   - Can involve enhancement of a cue when accompanying cue is hard to hear or has disappeared

Masking and Cues: Sibilants and Stops

- Fricatives have strong internal cues in general
- Stops have weak place cues

	Manner	Place
Stops	silence (internal)	bursts, transitions (contextual)
Fricatives	noise (internal)	spectrum, amplitude (internal)

Cues for stops and fricatives

- The noisiness of sibilants also masks acoustics of a neighboring stop (Mielke, 2001)
- Auditory decoupling: high frequencies decoupled from speech stream in perception, so sibilants can cause confusion in sequential order (Bregman and Campbell, 1971; Bregman, 1990; Blevins and Garrett, 2004)

Masking and Cues

- Stops need a vowel transition
- Bias for CV over VC transitions (Fujimura et al., 1978)
- Diachronic patterns
  1. #TSV → #STV: Greek *p^hsyk^{h}e:* > *sp^{h}yk^{h}e:*
  2. VST# → VTS#: Southern American English *wasp* → *waps*
  3. VSTC → VTSC: Lithuanian *dresk-ti:* → *dreks-ti*
  4. VTSV → VSTV: Rural Latin *ipse* → *ispe*

Perception

- Perceived order in nonce aSTa/aTSa (Graff and Scontras, 2012)
  - More likely to hear [aksa] as [aska] than vice versa
  - Removal of burst makes this percept even *more* likely
  - Bias for stops in prevocalic position
- Perceived order of ST/TS in Hebrew by English speakers (Jones, 2016)
  - Fricatives and sibilants in cluster lead to significantly higher reaction times in determining order
  - Tendency to misinterpret [dz] as [zd], despite higher phonotactic frequency of [dz] in English
- Sibilant noise is key factor in ST/TS metathesis

Production & Gestural Timing

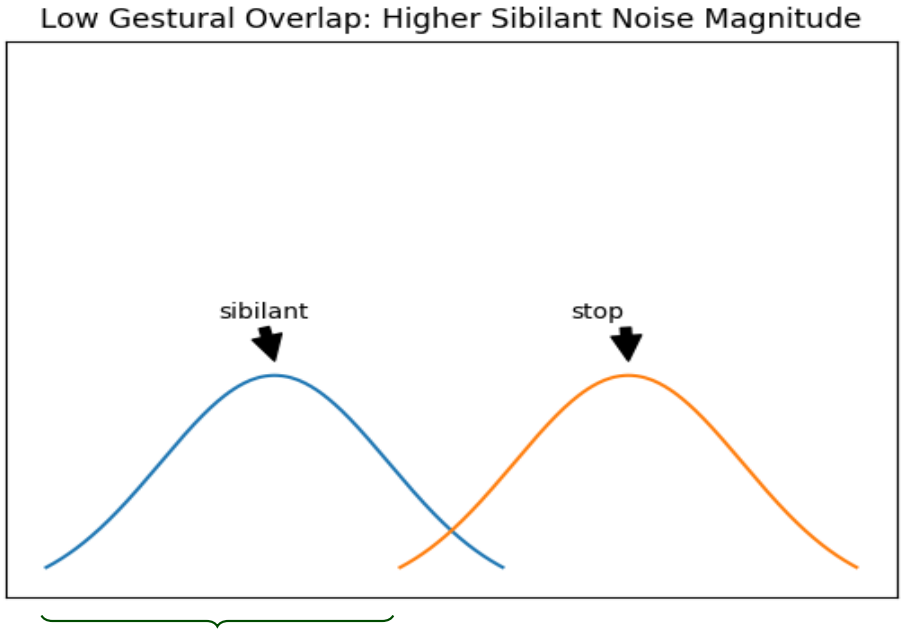
- Position in the word
  - More gestural overlap & variability in medial vs. onset position: Georgian (Chitoran et al., 2002); Hebrew (Yanagawa, 2003)
  - More gestural overlap and variability in coda vs. onset position in English (Byrd, 1996)
- Existence of morpheme boundary
  - More gestural variability hetero- vs. tautomorphemically: Korean (Cho, 2001); Hebrew (Yanagawa, 2003)
  - Non-morphemic *-s* significantly longer than morphemic *-s* if voiceless (Plag et al., 2017)
  - Non-morphemic *-s* significantly shorter than morphemic *-s* (Walsh and Parker, 1983; Losiewicz, 1995; Seyfarth et al., 2018)
- Yanagawa claims that metathesis in Hebrew occurs in *hi**t-p**a'el* binyan with TS sequences because they are *word-medial, heteromorphemic*

Questions

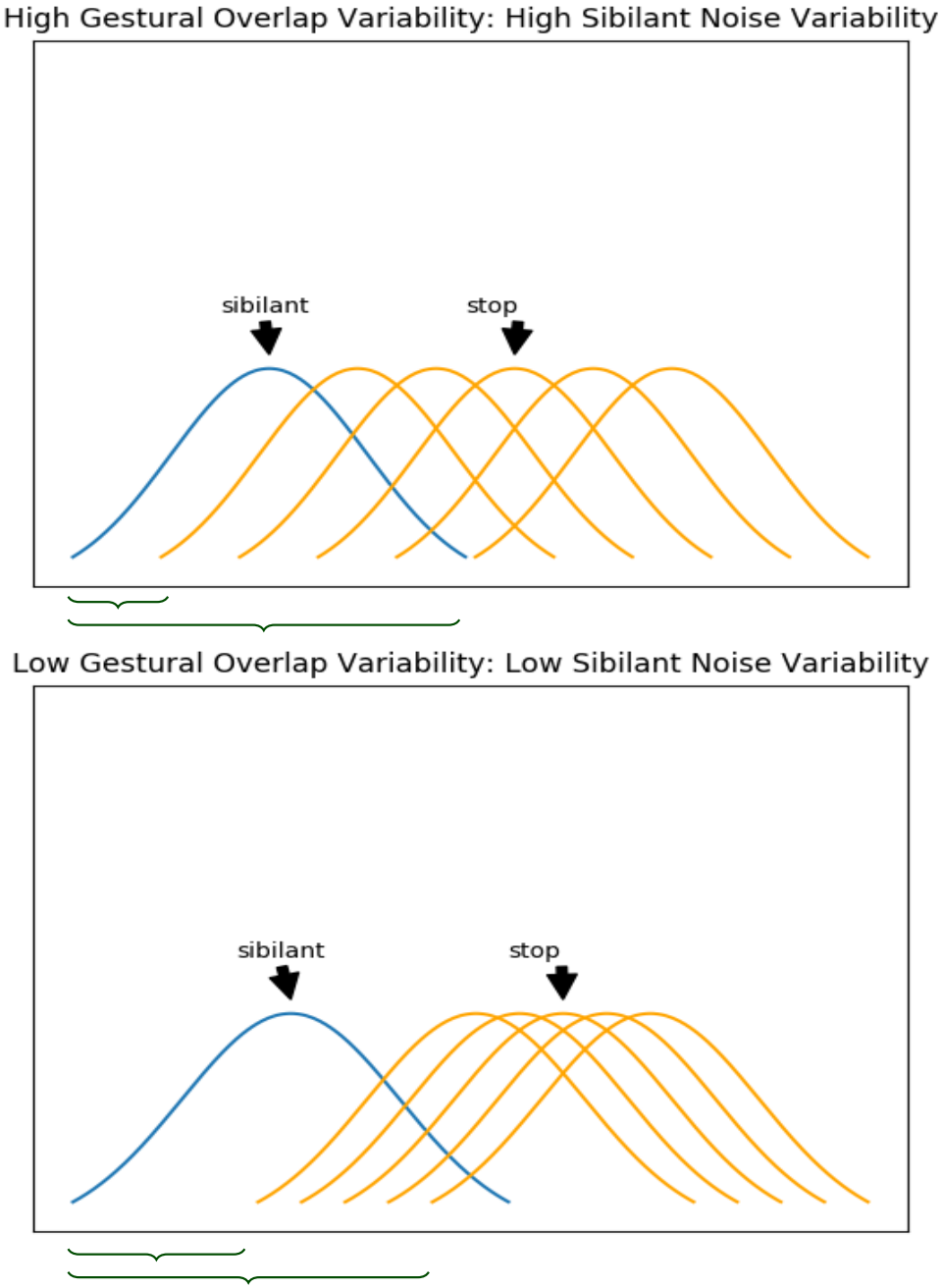
- Are articulatory findings reproduceable in acoustic data?
- Articulatory studies have few speakers:

  5 (Byrd, 1996), 3 (Cho, 2001), 2 (Chitoran et al., 2002), 1 (Yanagawa, 2003)
- Are experimental findings reproduceable in corpus data?
- Do magnitude and variability of gestural overlap provide an ambiguous signal that can lead to metathesis?
  - Do we see longer sibilant noise where we expect to see metathesis?
  - Do we see greater variability in sibilant noise where we expect to see metathesis?

Gestural Magnitude



Gestural Variability



The Buckeye Corpus

- Pulled data from Buckeye Corpus of Conversational Speech: recordings of interviews with 40 speakers from Columbus, OH (Pitt et al., 2007)
- Extract tokens with phonetic ST/TS clusters
- n = 13,975

Factors

- Independent variables
  - Position in word: Onset, Medial, Final
  - Morpheme existence (coded one of two different ways)
    - Binary: Yes, No
    - Ternary: None, -s, -ed
  - Cluster type: Stop is Prevocalic (STV), Postvocalic (VTS)
- Dependent variable: log ratio of sibilant to stop duration
- Other factors available (not significant)
  - Speaker gender
  - Speaker age
  - Interviewer gender

Hypotheses

- Word position effects
  - Gestural overlap greater and more variable in non-initial than initial position
  - Magnitude of sibilant duration: non-initial < initial
  - Variance of sibilant duration: non-initial > initial
- Morpheme boundary effects
  - Gestural overlap more variable in heteromorphemic than tautomorphemic clusters
  - Magnitude: conflicting
  - Variance: heteromorphemic > tautomorphemic
- Cluster type effects
  - Gestural overlap: ?
  - Magnitude: VTS > STV
  - Variability: VTS > STV

Model Comparison

- Ran LMERs with *lme4* (Bates et al., 2015) in R (R Core Team, 2017)
- Position in word, morpheme, cluster type as fixed effects
  - Morpheme coded either as binary (yes/no) or ternary morpheme type (none, -s, ed)
  - Interaction between position in word and cluster type considered
- Random intercepts for subjects and by-subject random slopes for fixed effects
- Best model
  - Binary morpheme coding
  - Interaction of Word Position x Cluster Type

Results

- Results reported without outliers (= 268 tokens, 1.92%)
- Before removal, Type (VTS) was *not significant*

Fixed Effect	Est.	SE	t		
Intercept	0.8814	0.014	63.07	***	
WP (Medial)	-0.0642	0.0097	-6.63	***	n.s.
WP (End)	-0.0513	0.0136	-3.78	***	
Morph (Yes)	0.0334	0.0155	2.16	*	
Type (VTS)	-0.0978	0.0405	-2.42	*	
WP (Medial) : Type (VTS)	0.2457	0.043	5.69	***	***
WP (End) : Type (VTS)	0.1653	0.048	3.44	**	

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

- Medial and final sibilants significantly shorter than onset
- Heteromorphemic sibilants significantly longer
- VTS sibilants significantly shorter than STV in *onset*
- Non-initial VTS sibilants significantly *longer* than STV


Model Estimates for Log Ratio by Position & Cluster Type
Fitted Log Ratio
1.00
0.95
0.90
0.85
0.80
onset
medial
end
Word Position
type
STV
VTS

Discussion

- Non-initial cluster sibilants longer than initial
  - Corroborates Byrd (1996); Chitoran et al. (2002); Yanagawa (2003)
  - Effect is weaker for final position
  - Final STV not significantly different from medial, but final VTS is
- Heteromorphemic cluster sibilants longer
  - Corroborates Seyfarth et al. (2018), counters Plag et al. (2017)
  - Latter used Buckeye as well, but
    - Coding differences
    - Did not check effect of voicing
- (Non-initial) VTS sibilants longer than STV ones
  - Greater sibilant noise can cause confusion in order (Bregman and Campbell, 1971; Bregman, 1990; Graff and Scontras, 2012; Jones, 2016)
  - Supports auditory metathesis (Blevins and Garrett, 2004)

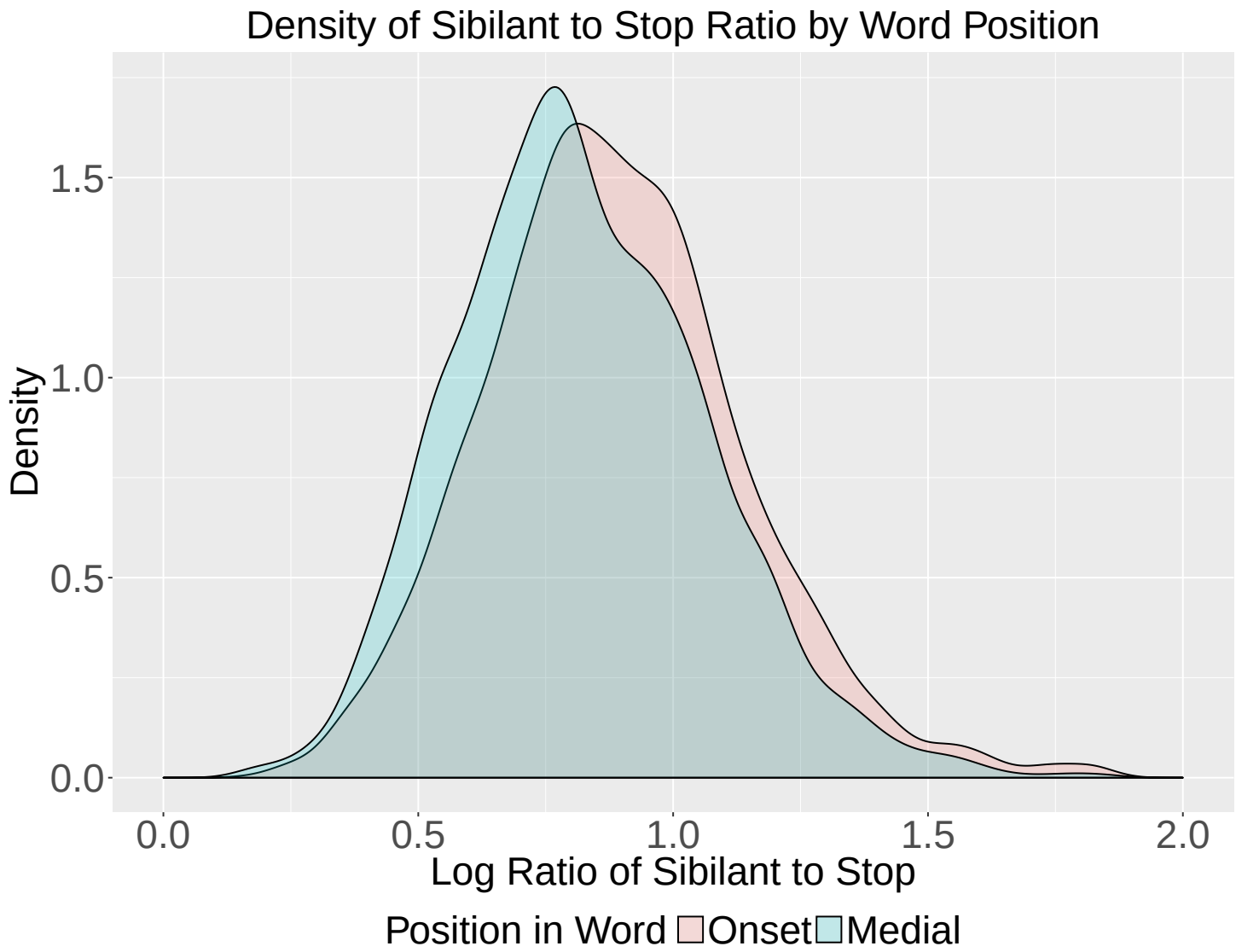
Measuring Variance Differences: Methodology

- Variance is heavily influenced by sample size

Speaker	Onset x STV	Medial x STV
1	144	132
2	172	75
...	...	...

- Limited to speaker x condition cells in which n >= 30
  - Leaves 26 speakers
  - Word position hypothesis: Onset x STV vs. Medial x STV
  - Cluster type hypothesis: STV x Medial vs. VTS x Medial
  - Could not test morpheme hypothesis
- n between cells also uneven (30 < n < 206)
- Ran 2 different analyses using LMER
  - Raw variance for each speaker x condition
  - Sampled 30 from each cell, ran LMER 5000 times, calculated % of cases in which LRT reveals p < .05

Position in Word (STV) and Variance



Onset: 3814 (σ^2 = 0.0676), Medial: 2511 (σ^2 = 0.0612)

Levene's Test: F(1,6323) = 2.324, p = 0.13

Position in Word (STV) & Variance

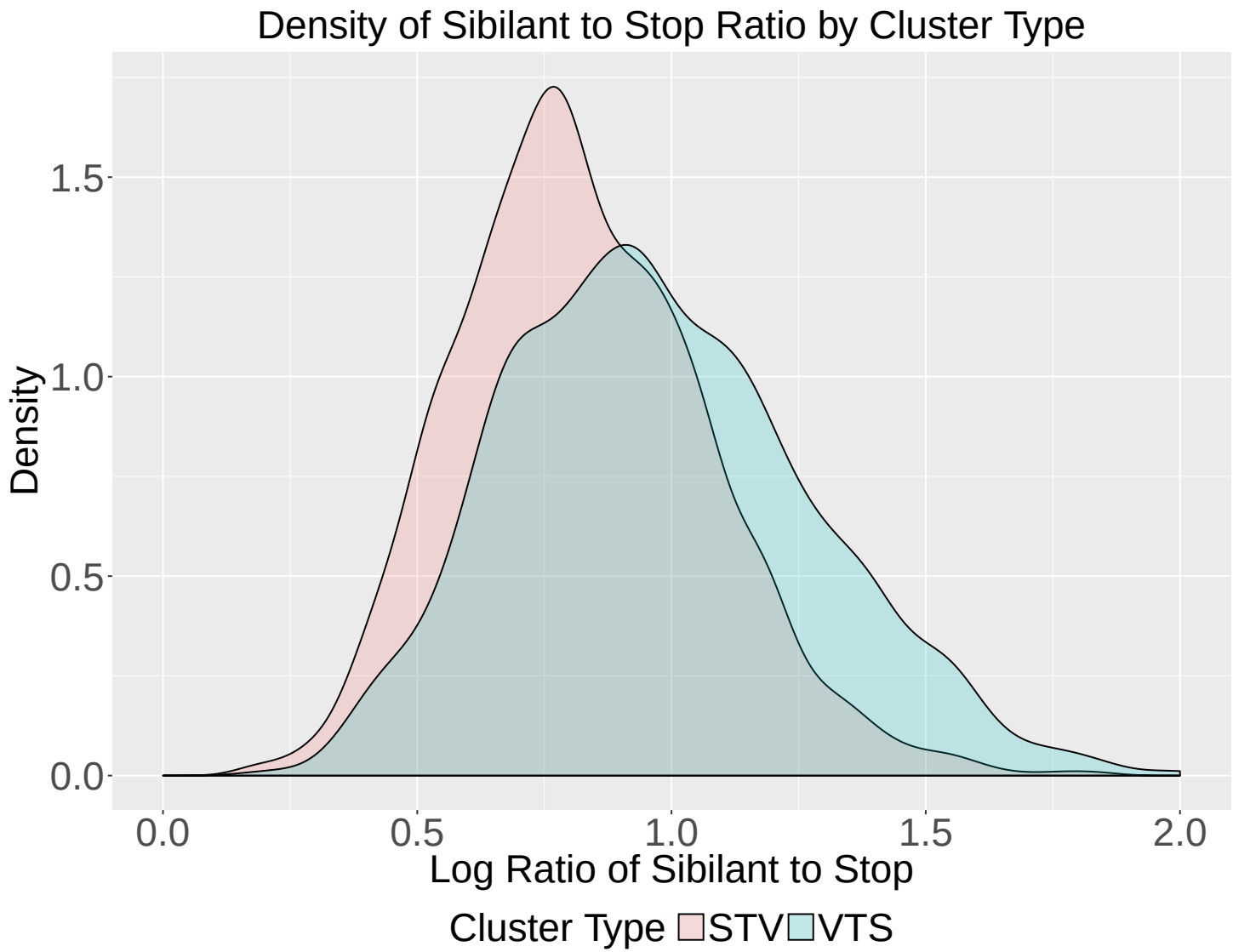
- Position in word as fixed effect, speaker as random effect
- Raw variance: one outlier speaker, removal reduces p-value from .036 to .002

Fixed Effect	Est.	SE	t	
Intercept	0.0598	0.0031	19.375	***
WP (Medial)	-0.0124	0.0036	-3.449	**

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

- Simulation without replacement: in 30.4% of simulations, model with word position significantly more predictive
- Simulation with replacement: in 31.5% of simulations, model with word position significantly more predictive

Cluster Type (Medial Position) & Variance



STV: 2511 (σ^2 = 0.0612), VTS: 1934 (σ^2 = 0.0948)

Levene's Test: F(1,4443) = 90.861, p < .001

Cluster Type (Medial Position) & Variance

- Cluster type as fixed effect, speaker as random effect
- Raw variance: same outlier speaker, removal makes no significant difference

Fixed Effect	Est.	SE	t	
Intercept	0.0474	0.0041	11.703	***
Type (VTS)	0.0263	0.0057	4.586	***

$^{*}p < .05 \quad ^{**}p < .01 \quad ^{***}p < .001$

- Simulation without replacement: in 96.98% of simulations, model with cluster type significantly more predictive
- Simulation with replacement: in 89.47% of simulations, model with cluster type significantly more predictive

Discussion

- Medial cluster sibilants less variable than initial
  - But this effect is weak
  - Counter to Chitoran et al. (2002); Yanagawa (2003)
  - However, Chitoran et al. only look at stop+stop sequences and Yanagawa looks at stop+stop/fricative
- VTS sibilants much more variable than STV
  - Gestural variability is a possible factor in the preferred directionality of VTSV to VSTV
  - Supports Yanagawa's hypothesis of gestural variability
- Morpheme numbers too few and imbalanced to make any claims about variance across morpheme boundaries

Hypothesis Review

Word position effects

- Gestural overlap greater and more variable in non-initial than initial position ~✓
- Magnitude of sibilant duration non-initial < initial ✓
- Variance of sibilant duration non-initial > initial ✗

Morpheme boundary effects

- Gestural overlap more variable in heteromorphemic than tautomorphemic clusters ?
- Magnitude heteromorph. > tautomorph. ← conflicting
- Variance heteromorph. > tautomorph. ?

Cluster type effects

- Gestural overlap less & more variable in VTS? ← ?
- Magnitude VTS > STV ✓
- Variability VTS > STV ✓

Summary

- Magnitude of sibilant noise
  - Supports previous findings except Plag et al. who found shorter -s
  - Higher magnitude in VTS possibly leads to segmental confusion
- Variance of sibilant noise
  - Lower in non-initial position than onset
  - Greater in VTS than STV
- Greater magnitude and variance of sibilant noise may conspire to pressure VTSV > VSTV metathesis
- Type effect attenuated in final position—perceptual optimization? Stop more salient in VTS# than VST#
- Acoustic corpus findings may corroborate articulatory experimental findings

Directions

- Articulatory corpus
  - Running analysis with articulatory X-ray microbeam database (Westbury, 1994)
  - Validate whether acoustic findings indeed match gestural overlap
- Experiments
  - Production experiment controlling for current factors
  - Perception experiment
    - Is perceptual metathesis more likely with longer sibilant?
    - Interaction with word position and/or morpheme status?
- Explore corpora of other languages

Thansk!

tylerlau@berkeley.edu

Many thanks to Susan Lin, Susanne Gahl, Andrew Garrett, Keith Johnson, Ron Sprouse, my undergraduate research apprentice Anstonia Ma, and my fellow classmates in LING201 for help at various stages in this research.

References I

Douglas Bates and Martin Mächler and Ben Bolker and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Blevins, Juliette and Garrett, Andrew. 1998. The Origins of Consonant-Vowel Metathesis. *Language* 74:508–556.

Blevins, Juliette and Garrett, Andrew. 2004. The evolution of metathesis. In *Phonetically based phonology*, ed. Bruce Hayes, Robert Kirchner, and Donca Steriade, 117–56. Cambridge University Press.

Bregman, Albert S. 1990. *Auditory Scene Analysis*. Massachusetts: MIT Press.

Bregman, Albert S and Campbell, Jeffrey. 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of experimental psychology* 89:244.

Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of phonetics* 24:209–244.

Chitoran, Ioana and Goldstein, Louis and Byrd, Dani. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. *Laboratory Phonology* 7:419–447.

Cho, Taehong. 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica* 58:129–62.

Fujimura, Osamu and Macchi, M. J. and Street, L. A. 1978. Perception of Stop Consonants with Conflicting Transitional Cues: A Cross-Linguistic Study. *Language and Speech* 21:337–346.

References II

Graff, Peter and Scontras, Gregory. 2012. Metathesis as Asymmetric Perceptual Realignment. In *WCCFL XXVIII*.

Grammont, Maurice. 1933. *Traité de phonétique*. Paris: Delagrave.

Hock, Hans Heinrich. 1985. Regular metathesis. *Linguistics* 23:529–546.

Hume, Elizabeth. 1998. The role of perceptibility in Consonant/Consonant metathesis. In *WCCFL XVII Proceedings*, 293–307. Stanford: CSLI.

Hume, Elizabeth V. 2004. The indeterminacy/attestation model of metathesis. *Language* 80:203–237.

Jones, Kyle. 2016. The perception of stop/sibilant clusters in Modern Hebrew. In *Poster presented at LabPhon 15: Speech Dynamics and Phonological Representations*. Ithaca, NY: Cornell University.

Ladefoged, Peter. 2001. *Vowels and consonants: An introduction ot the sounds of languages*. Malden: Blackwell Publishing.

Losiewicz, Beth L. 1995. Word frequency effects on the acoustic duration of morphemes. *The Journal of the Acoustical Society of America* 97:3243–3243.

Mielke, Jeff. 2001. Turkish /h/ deletion: evidence for the interplay of speech perception and phonology. In *Proceedings of NELS 32*.

Ohala, John J. 1981. The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior*, ed. Carrie S. Masek, Robert A. Hendrick, and Mary Frances Miller, 178–203. Chicago: Chicago Linguistics Society.

References III

Ohala, John J. 1993. The phonetics of sound change. In *Historical linguistics: Problems and perspectives*, ed. Charles Jones, 237–278. London: Longman.

Pitt, Mark and Dilley, Laura and Johnson, Keith and Kiesling, Scott and Raymond, William and Hume, Elizabeth and Fosler-Lussier, Eric. 2007. *Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]*. Columbus, OH: Department of Psychology, Ohio State University.

Plag, Ingo and Homann, Julia and Kuntner, Gero. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53:181–216.

R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Seyfarth, Scott and Garellek, Marc and Gillingham, Gwendolyn and Ackerman, Farrell and Malouf, Robert. 2018. Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33:32–49. doi: 10.1080/23273798.2017.1359634.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: a perceptual account. In *The role of speech perception in phonology*, ed. Elizabeth Hume and Keith Johnson, 219–50. Brill Academic Pub.

Walsh, Thomas and Parker, Frank. 1983. The duration of morphemic and non-morphemic [s] in English. *Journal of Phonetics* 11:201–206.

References IV

Westbury, John R. 1994. *X-ray microbeam speech production database user's handbook, version 1.0*. http://www.medsch.wisc.edu/ milenkvc/pdf/ubdbman.pdf, Madison, WI.

Yanagawa, Mariko. 2003. Metathesis in Modern Hebrew: An Analysis in Articulatory Phonology. In *Proceedings of the 15th International Congress of the Phonetic Sciences*, 1671–1674. Rundle Mall: Causal Productions.