Sound change and coarticulatory variability involving English /ɹ/

Bridget J. Smith
North Carolina State University
Campus Box 8105, Raleigh, NC
27695-8105 USA
bsmith8@ncsu.edu

Jeff Mielke
North Carolina State University
jimielke@ncsu.edu

Lyra Magloughlin
University of Ottawa
lyra@uottawa.ca

Eric Wilbanks
University of California at Berkeley
wilbanks_eric@berkeley.edu

1 Token counts for auditory coding

	Raleigh conversation			Lab read speech		
variable	total	range	median	total	range	median
word-initial /stɹ/	1502	0-33	9	644	12-33	23
word-medial /stɹ/	818	0-34	5	604	12-28	23
/tɹ/	3094	0-71	20	756	16-36	29
/dɹ/	1639	0-60	9	635	12-26	23
/s#ɹ/	766	0-30	4	222	2-10	9
/z#ɹ/	1269	0-39	8	228	5-11	8
/ɹs/	1252	0-59	6	227	2-10	9
/ɹz/	2514	0-56	17	223	1-12	9
/ɹ#s/	1229	0-28	8	235	3-10	9
/ɹ#z/	3	0-1	0	225	2-10	9

Table 1: Token counts of each variable in the Raleigh corpus and laboratory data, total counts and summary of token counts per talker.

.

2 Auditory coding details

Two of the authors and a third linguist listened to and rated recorded clips containing each of the variables from the Raleigh Corpus and from the May corpus. Each automatically segmented word was padded by 100 ms on either side and extracted from the larger sound file. A Praat Multiple Forced Choice script was created for each variable from each corpus, so each variable was in a single block. Two of the listeners heard each token in a random order, while one listener (rater B) heard the tokens grouped by talker. For /tɹ/ and /dɹ/ , the choices were "affricated", "ambiguous", "not affricated", and "NA (wrong sound or other problem)". For the other variables, the choices included "retracted" and "not retracted" instead of "affricated" and "not affricated". For these tasks, "affricated" was taken to mean more like /tʃ dʒ/, and "not affricated" was taken to mean more like /t d/. "Retracted" was taken to mean more /ʃ ʒ/-like, and "not retracted" was taken to mean more /s z/-like. "NA" tokens, which included laughter, noises, or incorrectly labeled intervals, were removed from analysis.

Using this 3-way coding scheme, inter-rater agreement for non-NA tokens was, on average, 62% for affrication. The intermediate category of "ambiguous" was used differently by the different raters, based on their threshold for what they each considered "affricated", and whether they felt confident in their ability to discern partial affrication from phonological affrication, or a noisy stop from a partially affricated one. Rater A almost never used the "ambiguous" category, retaining its use only for instances where the rater felt it was impossible to tell partial affrication from phonological affrication. Rater A thus grouped a great many partially affricated tokens under the label "unaffricated". Rater B, on the other hand, used the "ambiguous" category extensively, to denote partially affricated tokens that did not sound like they contained a phonological /tʃ/, reserving the "unaffricated" label for stops with little or no frication phase. Rater C also rarely used the "ambiguous" category, but had a much lower threshold for what sounded affricated, grouping partially affricated tokens with the "affricated" category. In order to to bring the judgments into closer alignment with each other, we decided to create two categories of "less affricated" and "more affricated", in which we grouped the "ambiguous" tokens with the endpoints based on each rater's bias. Thus, the ambiguous tokens were assigned to the "more affricated" category for the higher threshold coder, and to the "less affricated" category for the lower threshold coders, which brought the agreement rate up to 81% each for /tɹ/

and /dɹ/. Agreement on retraction was higher than affrication, on average 85%, but in order to bring the coding into alignment with the 0 and 1 scheme for the affrication judgments, the "ambiguous" categories for these were recoded for all raters as "retracted", so that the retraction category included partially retracted tokens. With these changes, the average agreement rate was 90% for /s/ and /z/ retraction and 86% across all categories. Inter-rater statistics for each variable can be found in Table 2.

Tokens with preceding or following segments that are known or suspected to cause retraction or affrication (such as palatals or fricatives) were excluded from analysis. Additionally, twelve of the Raleigh interview recordings included a reading passage and a word list containing 25 tokens of /stɹ/, five tokens of /tɹ/, and one token of /dɹ/ used for a previous study of /stɹ/ retraction (Piergallini 2011). These were excluded from the statistical analysis and graphs, but it is noteworthy that talkers who had a phonologically affricated target (>50%) exhibited a higher degree of /tɹ/ and /dɹ/ affrication in read speech than in conversational speech, and talkers who had a phonologically unaffricated target (< 50%) exhibited a lower degree of /tɹ/ and /dɹ/ affrication in read speech than in conversational speech; and Raleigh talkers who did not have a phonologically retracted target in /stɹ/ generally exhibited a lower degree of /s/ retraction in read speech than in conversational speech.

3 Inter-rater statistics

Table 2 shows inter-rater statistics for the auditory coding analysis.

variant	avg. ambiguous judgments (original)				avg. affricated judgments (recoded)			
	rater			agreement (3 categories)	rater			agreement (2 categories)
	A	B	C		A	B	C	
/tɹ/	0.339	0.296	0.017	0.58	0.452	0.643	0.915	0.81
/dɹ/	0.185	0.253	0.013	0.66	0.459	0.502	0.894	0.81
variant	avg. ambiguous judgments (original)				avg. retracted judgments (recoded)			
	rater			agreement (3 categories)	rater			agreement (2 categories)
	A	B	C		A	B	C	
/#stɹ/	0.090	0.178	0.030	0.77	0.214	0.261	0.433	0.86
/stɹ/	0.095	0.198	0.009	0.79	0.343	0.321	0.028	0.85
/s#ɹ/	0.163	0.281	0.031	0.84	0.220	0.340	0.058	0.87
/z#ɹ/	0.307	0.367	0.032	0.80	0.359	0.410	0.042	0.84
/ɹs/	0.123	0.112	0.124	0.91	0.175	0.149	0.032	0.93
/ɹz/	0.165	0.090	0.010	0.91	0.203	0.099	0.011	0.92
/ɹ#s/	0.157	0.156	0.018	0.91	0.182	0.131	0.021	0.87
/ɹ#z/	0	0	0	1.0	0	0	0	1.0

Table 2: Inter-rater statistics for auditory coding.

.

4 Participant details

Participant	Sex	Birth Year	Birth Place
02	Male	1990	North Carolina
03	Female	1996	North Carolina
04	Female	1994	North Carolina
05	Female	1995	North Carolina
06	Female	1996	Wisconsin
07	Female	1956	North Carolina
09	Male	1995	Delaware
10	Female	1995	North Carolina
11	Female	1991	North Carolina
13	Female	1997	North Carolina
14	Female	1996	North Carolina
15	Male	1988	North Carolina
16	Female	1995	Texas
17	Female	1993	New Jersey
18	Male	1995	North Carolina
19	Female	1996	North Carolina
20	Female	1997	North Carolina
21	Female	1970	New York
22	Male	1996	North Carolina
23	Male	1996	North Carolina
24	Female	1985	North Carolina
25	Male	1993	Georgia
26	Male	1979	Tennessee
27	Male	1994	Virginia
28	Male	1994	North Carolina
29	Female	1994	North Carolina
30	Female	1991	North Carolina
31	Male	1996	North Carolina
33	Female	1992	North Carolina

Table 3: Participant details.

.

5 Auditory coding statistical output

medial /stɹ/: `lmer(score~(poly1+poly2)*birthyear*gender+(1|talker) +(1|word))+(1|rater)`

medial /stɹ/	Est.	SE	*z*	*p*
Intercept	-2.6033	1.0209	-2.550	0.0107*
birthyear (linear)	26.5248	21.5216	1.232	0.2177
birthyear (quadratic)	101.9637	28.1391	3.624	0.0003*
gender (male)	0.3871	0.2497	1.550	0.1210
birth linear:gender male	-12.6601	29.5322	-0.429	0.6682
birth quadr:gender male	-12.7101	28.5911	-0.445	0.6566

Table 4: Model output for medial /stɹ/ .

.

/#stɹ/: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/#stɹ/	Est.	SE	*z*	*p*
Intercept	-1.494401	0.381751	-3.915	0.0000*
birthyear (centered)	0.011270	0.008415	1.339	0.1805
gender (male)	0.468308	0.174424	2.685	0.0073*
birthyear:gender male	-0.021244	0.011078	-1.918	0.0551(.)

Table 5: Model output for initial /#stɹ/.

.

/tɹ/: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/tɹ/	Est.	SE	*z*	*p*
Intercept	0.859664	0.788274	1.091	0.2755
birthyear (centered)	0.019871	0.007588	2.619	0.0088*
gender (male)	0.405465	0.137081	2.958	0.0031*
birthyear:gender male	-0.003366	0.010150	-0.332	0.7402

Table 6: Model output for /tɹ/.

.

/dɹ/: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/dɹ/	Est.	SE	*z*	*p*
Intercept	0.791609	0.848768	0.933	0.3510
birthyear (centered)	0.018625	0.009033	2.062	0.0392*
gender (male)	0.369784	0.165276	2.237	0.0253*
birthyear:gender male	0.005651	0.011462	0.493	0.6220

Table 7: Model output for /dɹ/.

.

/z#ɹ/: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/z#ɹ/	Est.	SE	*z*	*p*
Intercept	-2.255323	0.908856	-2.481	0.0131*
birthyear (centered)	-0.027861	0.008782	-3.173	0.0015*
gender (male)	0.547482	0.177428	3.086	0.0020*
birthyear:gender male	-0.018821	0.012344	-1.525	0.12735

Table 8: Model output for /z#ɹ/.

.

/s#ɹ: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/s#ɹ/	Est.	SE	*z*	*p*
Intercept	-2.375144	0.680776	-3.489	0.0004*
birthyear (centered)	-0.040333	0.010472	-3.852	0.0001*
gender (male)	0.756176	0.236018	3.204	0.0013*
birthyear:gender male	-0.009511	0.014007	-0.679	0.4971

Table 9: Model output for /s#ɹ/.

.

/ɹs/: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/ɹs/	Est.	SE	*z*	*p*
Intercept	-3.766282	0.616360	-6.111	0.0000*
birthyear (centered)	-0.005368	0.012177	-0.441	0.6594
gender (male)	0.449842	0.212077	2.121	0.0339*
birthyear:gender male	-0.007649	0.016305	-0.469	0.6390

Table 10: Model output for /ɹs/.

.

/ɹz/: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/ɹz/	Est.	SE	*z*	*p*
Intercept	-3.447958	0.822856	-4.190	0.0000*
birthyear (centered)	0.003928	0.008450	0.465	0.642
gender (male)	0.160802	0.165781	0.970	0.332
birthyear:gender male	-0.005384	0.011846	-0.455	0.649

Table 11: Model output for /ɹz/.

.

/ɹ#s/: `lmer(score~birthyear*gender+(1|talker)+(1|word))+(1|rater)`

/ɹ#s/	Est.	SE	*z*	*p*
Intercept	-3.403365	0.706512	-4.817	0.0000*
birthyear	-0.011470	0.011049	-1.038	0.2992
sex male	0.689578	0.221860	3.108	0.0019*
birthyear:sexmale	0.001036	0.014355	0.072	0.94247

Table 12: Model output for /ɹ#s/.

.

6 Automatic classification model outputs

Table 13 shows the /tɹ/ A-ratio model, and Table 14 shows the /dɹ/ A-ratio model.

/tɹ/	Est.	SE	df	*t*	*p*
(Intercept)	0.5821	0.0439	134.910	13.274	0.000*
birthyear	0.0115	0.0026	122.590	4.413	0.000*
sex male	-0.169582	0.0624	119.70	-2.717	0.005*
birthyear: sexmale	-0.004828	0.003780	121.130	-1.277	0.204

Table 13: A-ratio results for /tɹ/.
.

/dɹ/	Est.	SE	df	*t*	*p*
(Intercept)	0.693408	0.042636	107.39	16.263	0.000
birthyear	0.015377	0.002546	103.76	6.039	0.000
sexmale	-0.240680	0.061544	103.77	-3.911	0.000
birthyear sexmale	-0.008678	0.003697	102.02	-2.348	0.020

Table 14: A-ratio results for /dɹ/.
.

7 Statistics for acoustic analysis of /stɹ/

Table 15 shows the /stɹ/ COG ratio model.

/stɹ/	Est.	SE	df	*t*	*p*
(Intercept)	-0.9249	0.0774	179	-11.950	0.000
birthyear	-0.0298	0.0215	266	-1.390	0.166
typeS	-0.1656	0.0260	3178	-6.360	0.000
typeSTR	0.3965	0.0488	3516	8.120	0.000
typeSH	1.1680	0.0347	2784	33.67	0.000
sexmale	0.0009	0.0302	256	0.032	0.974
positionmedial	0.1651	0.0305	5195	5.419	0.000
prev_manneraffricate	0.2038	0.0712	78340	2.862	0.004
prev_mannerfricative	-0.0889	0.0129	79120	-6.904	0.000
prev_mannerglide	0.0429	0.2897	77140	0.148	0.882
prev_mannerliquid	0.0962	0.0138	70110	6.979	0.000
prev_mannernasal	-0.1003	0.0102	40380	-9.848	0.000
prev_mannernon_speech	0.0174	0.0089	78450	1.943	0.052
prev_mannerstop	-0.1285	0.0084	61670	-15.337	0.000
log(duration)	-0.3849	0.0304	143	-12.679	0.000
birthyear:typeS	0.0935	0.0133	78840	7.046	0.000
birthyear:typeSTR	0.0963	0.0268	78860	3.593	0.000
birthyear:typeSH	0.0078	0.0163	78880	0.478	0.633
birthyear:sexmale	-0.0048	0.0312	258	-0.153	0.878
typeS:sexmale	0.0831	0.0182	78730	4.573	0.000
typeSTR:sexmale	-0.0583	0.0398	78890	-1.465	0.143
typeSH:sexmale	-0.0179	0.0233	78620	-0.770	0.441
birthyear:positionmedial	-0.0141	0.0197	78620	-0.717	0.473
typeS:positionmedial	0.0184	0.0350	4813	0.526	0.599
typeSTR:positionmedial	-0.0503	0.0654	5662	-0.769	0.442
typeSH:positionmedial	-0.0999	0.0427	4116	-2.339	0.019
sexmale:positionmedial	-0.0143	0.0274	79040	-0.522	0.602
birthyear:typeS:sexmale	0.0064	0.0189	78610	0.336	0.737
birthyear:typeSTR:sexmale	-0.1169	0.0412	78800	-2.837	0.004
birthyear:typeSH:sexmale	0.0152	0.0244	78670	0.624	0.532
birthyear:typeS:positionmedial	-0.0326	0.0223	78420	-1.463	0.144
birthyear:typeSTR:positionmedial	0.1445	0.0466	76130	3.099	0.002
birthyear:typeSH:positionmedial	-0.0003	0.0258	78550	-0.012	0.991
birthyear:sexmale:positionmedial	0.0389	0.0285	78580	1.369	0.171
typeS:sexmale:positionmedial	0.0595	0.0310	78800	1.916	0.055
typeSTR:sexmale:positionmedial	-0.0176	0.0665	76430	-0.265	0.791
typeSH:sexmale:positionmedial	0.0489	0.0364	79050	1.346	0.178
birthyear:typeS:sexmale:positionmedial	-0.0623	0.0321	78470	-1.937	0.053
birthyear:typeSTR:sexmale:positionmedial	-0.1516	0.0691	77700	-2.194	0.028
birthyear:typeSH:sexmale:positionmedial	-0.0994	0.0378	78810	-2.628	0.008

Table 15: Statistical output for modeling /stɹ/ COG ratio across birthyear.

.

8 Articulatory methods details

In order to facilitate quantitative analysis of tongue position changes over time, we performed Eigentongues decomposition (principal component analysis (PCA) of pixel intensities in filtered and downsampled ultrasound images Hueber et al. 2007; Hoole & Pouplier 2017). This analysis was performed as implemented by Carignan (2014) and described in Mielke, Carignan & Thomas (2017: 336). All ultrasound images within one second of segmented speech were retained for analysis. They were rotated to make the occlusal plane horizontal, and the images were filtered using anisotropic speckle reduction (edge-sensitive noise reduction, Yu & Acton 2002; Hueber et al. 2007), median filtering (localized noise reduction), Gaussian filtering (global noise reduction), and Laplacian filtering (edge contrast enhancement). A hand-selected region of interest mask was then applied to the images from each session, and then all the images were downsampled in order to limit the dimensionality of the input data for the principal component analysis. The main difference between our application of Carignan's procedure and the application used by Mielke et al. (2017) is that our ultrasound images started out at a resolution of 640×480 instead of 320×240, and we reduced the resolution of the filtered images to 20% of the original resolution via bicubic interpolation (instead of 30%).

A similar PCA technique was used to analyze the video images, which were first cropped to isolate the mouth, reducing resolution from 640×480 to between 360×220 and 450×320. The images were not filtered (since the filters described above were selected specifically to enhance ultrasound images), and no further region of interest mask was applied (after the initial cropping). Resolution was still reduced to 20%, yielding smaller images than in the case of the ultrasound data, because the starting images were smaller after cropping. For both ultrasound and video images, the first 50 principal components were retained for analysis. Previous work with similar data has shown that this includes a sufficient portion of the variance in the images (Mielke et al. 2017: 337).

The result of each PCA is a 50-dimensional vector representing each ultrasound and video frame. We performed Linear Discriminant Analysis (LDA) of the PCA output in order to quantify the articulatory similarity of coarticulated or assimilated consonants to word-initial /ɹ/ (representing the local coarticulation source) and word-initial postalveolar consonants /ʃ tʃ dʒ/ (which they may be more likely to resemble in a phonologized assimilation pattern) (Hoole & Pouplier 2017; Strycharczuk & Scobbie 2017).

LDA models of pixel intensities in ultrasound and lip video images were generated for prevocalic /ɹ/, /ʃ tʃ dʒ/, and /s z t d/, vs. each other and vs. all other phones, and were used to examine the similarity between target sounds and these reference sounds.

9 Production Stimuli

The stimuli for the laboratory study are shown in Tables 16-19. Only stimuli listed in Tables 17-18 were reported here. Other stimuli provided data for related projects.

	/stɹ/	/s/	/st/	/ʃ/
/i/	street	seep	steep	sheep
	streaky			
/ɪ/	strip	sip	stick	ship
	strictly			
/ʌ/	struck	sub	stuck	shut
	struggle			
/ɑ/	strong	sob	stop	shop
	strawberry			

Table 16: Initial /stɹ/ and comparison stimuli.

.

	/tɹ/	/t/	/tʃ/	/dɹ/	/d/	/dʒ/
/i/	tree	tea	cheap	Dreep	deep	jeep
	treat	teach	cheek	dream	deem	jeans
/ɪ/	trick	tick	chick	drift	dibs	Jim
	trip	tip	chip	drip	dip	jib
/ʌ/	trust	tub	chub	drum	dumb	jug
	truck	tuck	chuck	drug	duck	jump
/ɑ/	trot	talk	chalk	draw	dog	job
	Tron	top	chop	drop	dock	jock

Table 17: /dɹ/, /tɹ/, and comparison stimuli.

.

	/ɹ#s/	/ɹ#z/	/ɹs/	/ɹz/	/s#ɹ/	/z#ɹ/	/ə#ɹ/
/i/	beer sip	beer zip	pierce	beers	kiss reed	fizz reed	a reed
/o/	bore sip	bore zip	Morse	bores	kiss road	fizz road	a road
/ɑ/	bar sip	bar zip	parse	bars	kiss rock	fizz rock	a rock

Table 18: Stimuli with /s z/ next to /ɹ/.

.

administration, astronomy, Australia, babe, back, bad, bade, bag, bail, bam, ban, bane, bang, bathrobe, bathroom, bead, bean, beb, bed, bell, Ben, bid, bin, bod, bode, bon, bone, bony, both, bud, destruction, distributed, fabulous Tron, heap, heft, hip, home, hoop, hope, illustrated, ingenious trick, leap, left, lip, loop, oath, only, peel, peep, pill, pip, pole, pool, poop, pull, reed, ref, registration, rip, road, rock, rub, rude, south, sumptuous treat, these, thirty, thorough, through, three, threw, throw, windowless truck

Table 19: Additional stimuli.

.

References

Carignan, Christopher. 2014. *TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) software suite*. http://christophercarignan.github.io/TRACTUS.

Hoole, Philip & Marianne Pouplier. 2017. Öhman returns: New horizons in the collection and analysis of imaging data in speech production research. *Computer Speech & Language* 45. 253–277. https://doi.org/10.1016/j.csl.2017.03.002.

Hueber, Thomas, Guido Aversano, Gérard Chollet, Bruce Denby, Gérard Dreyfus, Yacine Oussar, Pierre Roussel & Maureen Stone. 2007. Eigentongue feature extraction for an ultrasound-based silent speech interface. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. 1245–1248. Honolulu, HI: Cascadilla Press. https://doi.org/10.1109/ICASSP.2007.366140.

Mielke, Jeff, Christopher Carignan & Erik R. Thomas. 2017. The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *Journal of the Acoustical Society of America* 142(1). 332–349. https://doi.org/10.1121/1.4991348.

Piergallini, Mario. 2011. /str/ retraction in North Carolina. North Carolina State University M.A. capstone paper.

Strycharczuk, Patrycja & James Scobbie. 2017. Whence the fuzziness? Morphological effects in interacting sound changes in Southern British English. *Laboratory Phonology* 8(1). 7, 1–21. https://doi.org/10.5334/labphon.24.

Yu, Yongjian & Scott T. Acton. 2002. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing* 11(11). 1260–1270. https://doi.org/10.1109/TIP.2002.804276.