Linguistic Portfolios

Volume 6 Article 10

2017

An Acoustic Phonetic Account of the Production of Word-Final /z/s in Central Minnesota English

Cassy Lundy
St. Cloud State University, casey.lundy@gmail.com

Ettien Koffi
St. Cloud State University, enkoffi@stcloudstate.edu

Linguistic Portfolios–ISSN 2472-5102 –Volume 6, 2017 | 109

AN ACOUSTIC PHONETIC ACCOUNT OF THE PRODUCTION OF WORD-FINAL /z/s IN CENTRAL MINNESOTA ENGLISH

ETTIEN KOFFI AND CASEY LUNDY[1]

ABSTRACT

Fromkin et al. (2014:278) hypothesize that word-final /z/s are devoiced when they are inflectional suffixes, but non-morphemic /z/s are not devoiced. To date, no known study has tested this hypothesis for any English dialect. The present study tests this hypothesis in Central Minnesota English (CMNE) by making use of five acoustic correlates: F0, center of gravity (CoG), intensity, duration, and the 40/60 Threshold. Nine participants, five males and four females, produced 17 words containing word-final /z/s. The findings reported in this study are based on 765 acoustic tokens (17 x 9 x 5). Overall, the acoustic analyses validate the first part of the hypothesis, but not the second part. Our findings are significant because of their relevance for the sociophonetic studies of language change and variation and for automatic speech recognition.

1.0 Introduction

In *An Introduction to Language*, Fromkin et al. (2014:278) put forth the following hypothesis:

> For many speakers of English, word-final /z/ is devoiced when the /z/[2] represents a separate morpheme. These speakers pronounce plurals such as *dogs, days,* and *dishes* as *[dɔgs], [des]*, and *[dɪʃəs]* instead of *[dɔgz]*, *[dez]*, and *[dɪʃəz]*. Furthermore, they pronounce possessives such as *Dan's, Jay's* and *Liz's* as *[dæ̃ns]*, *[dʒes]*, and *[lɪzəs]* instead of *[dæ̃nz]*, *[dʒez]*, and *[lɪzəz]*. Finally, they pronounce third-person singular verb forms such as *reads, goes,* and *fuses* as *[rids]*, *[gos]*, *[fʌses]* instead of *[ridz]*, *[goz]*, *[fʌsez]*. However, words such as *daze* and *Franz* are still pronounced *[dez]* and *[frænz]*, because the /z/ is not a separate morpheme.

[1] The first author assigned this project to the second author when he enrolled in his acoustic phonetic course. Thereafter, they met weekly to discuss the findings. The second author collected the data and did all the acoustic measurements in this paper for his capstone project for his BA in linguistics. The second author presented the preliminary findings in two venues: at Saint Cloud State University's Student Research Colloquium in 2015 and at the NCUR (National Conference on Undergraduate Research) at the University of North Carolina in 2016. The second author wrote a paper to fulfill an independent study requirement. The present version of the paper is substantially different from the one submitted by the second author. The first author has re-interpreted, re-analyzed, and expanded the original paper submitted by the second author.

[2] The following conventions are used throughout the paper. Phonemes are enclosed in slashes /…/, phones are in square brackets […], and graphemes are in angle brackets <…>.

The hypothesis, as stated, has two parts. We will investigate both to see whether or not they are validated by the speakers of the dialect of American English spoken in Central Minnesota.

2.0 Methodology, Equipment, Data Recording Procedures, and Participants

Five males and four females who are speakers of Central Minnesota English (CMNE) produced a corpus consisting of 17 words. Eleven of the words –<*Dan's, Liz's, Jay's, dogs, days, dishes, knees, goes, reads, fusses, is*>– are taken directly from Fromkin et al. (2014:278). We added six additional words <*fears, fierce, niece, knees, ease, is*> to investigate the acoustic differences between devoiced [z̥] and voiceless [s]. The complete list of test items is in Table 1. The first three columns deal with the morpheme /z/. The fourth column contains non-morphemic /z/ in syllable codas. The last column has words whose codas end in /s/.

Possessives	**Plurals**	**3rd Person Singular**	**Coda /z/**	**Coda /s/**
1-<Dan's>	4-<dogs>	9-<goes>	13-<daze>	16-<fierce>
2-<Liz's>	5-<days>	10-<reads>	14-<ease>	17-<niece>
2-<Jay's>	6-<dishes>	11-<fusses>	15-<Franz>	
	7-<knees>	12-<is>		
	8-<fears>			

Table 1: Data Set

The data was recorded on the second author's 2013 Macbook Pro laptop in quiet study rooms at St. Cloud State University and in a quiet conference room at Eich Motor Company, in St. Cloud, Minnesota. The acoustic correlates used to test the hypothesis are fundamental frequency (F0/pitch), intensity, Center of Gravity (CoG), duration, and the 40/60 Threshold. The total number of tokens investigated is 765 (17 x 5 x 9). The spectrograph in Figure 1 summarizes the relevant acoustic information discussed in this paper. All measurements have to do only with word-final /z/s and /s/s.[3] We did not include any measurement of the preceding vowels, as is done sometimes in other studies. Our focus is exclusively on the frication noise found in word-final alveolar fricatives.

[3] Every effort was made to annotate only the fricative portion of word-final /z/s and /z/s. But in some cases, tiny portions of the preceding vowels may have been included in the frication noise.

Linguistic Portfolios–ISSN 2472-5102 –Volume 6, 2017 | 111

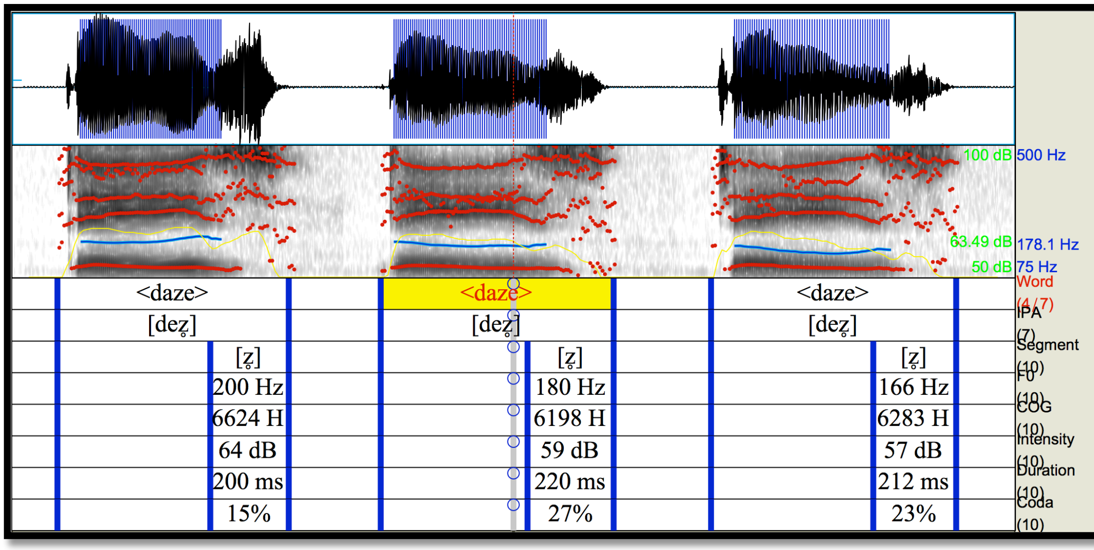

Figure 1: Annotation Sample

The data in Table 1 can be further subdivided into three classes. Eleven words – <*Dan's, Liz's, Jay's, dogs, days, dishes, knees, goes, reads, fusses, is*>– end with the inflectional suffix. Three words, <*daze, Franz, ease*>, have a non-morphemic /z/ in the coda. Two words, <*fierce, niece*>, end with the voiceless alveolar fricative /s/. If Fromkin et al.'s hypothesis is verified, the morphemic and non-morphemic /z/s would be acoustically different. In sections 3.0 to 7.0 we test the two parts of the hypothesis using five acoustic correlates.

3.0 Focus on F0

The first acoustic correlate used to test the hypothesis is fundamental frequency, otherwise known as F0 or pitch. Its measurements indicate the amount of times the vocal folds vibrate per second when a segment is produced. F0 helps classify segments into two broad categories. Segments that are produced when the vocal folds vibrate from onset to offset are said to be fully voiced. Those that are produced when the vocal folds do not vibrate at all or vibrate below a certain threshold are voiceless. The minimum F0 setting in Praat is 75 Hz and the maximum is 500 Hz (Boersma and Weenink 2010). The designers of Praat chose the minimum threshold of 75 Hz because most adults cannot hear frequencies below 75 Hz. Also, it is important to keep in mind that the lowest frequency that the vocal folds can produce when humans are speaking is 60 Hz (Fry 1979:68). When Praat labels a segment "undefined," we take it to mean that the segment is voiceless. It does not mean that the vocal folds do not vibrate at all, but that the vibrations are so slight that an average adult hearer cannot perceive them. We have assigned the numerical value of 74 Hz to all "undefined" segments in the data. There are 22 "undefined" alveolar fricatives out of the 153 attempted (14.79%). The undefined segments appear in bold in Tables 2A through 2D:

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1M	117	117	266	120	438	124	123	278	195	240
Speaker 2M	118	121	113	113	**74**	119	131	123	117	130
Speaker 3M	113	107	112	229	**74**	112	112	**74**	129	109
Speaker 4M	112	116	120	99	**74**	130	183	431	118	110
Speaker 5M	**74**	76	**74**	326	88	88	231	**74**	**74**	**74**
Mean	106	107	137	177	149	114	156	196	126	132
St. Deviation	18	18	74	97	161	16	50	155	43	63

Table 2A: F0 Male (Hz)

Words	Goes	Is	Jay's	Knees	Niece	Liz's	Reads
Speaker 1M	118	116	185	112	284	174	343
Speaker 2M	127	120	118	116	139	118	127
Speaker 3M	107	110	110	118	127	109	**74**
Speaker 4M	118	124	120	131	134	**74**	116
Speaker 5M	**74**	86	**74**	81	**74**	80	94
Mean	108	112	121	116	151	111	150
St. Deviation	20	15	40	18	78	39	109

Table 2B: F0 Male (Hz)

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1F	136	166	182	180	215	176	191	213	165	171
Speaker 2F	176	183	175	178	**74**	181	269	384	**74**	177
Speaker 3F	168	168	161	**74**	170	170	167	178	166	**74**
Speaker 4F	253	282	225	246	251	250	267	263	230	245
Mean	183	199	185	169	177	194	223	259	158	166
St. Deviation	49	55	27	71	76	37	52	90	64	70

Table 2C: F0 Female

Words	Goes	Is	Jay's	Knees	Niece	Liz's	Reads
Speaker 1F	349	198	174	160	177	175	160
Speaker 2F	184	180	182	177	177	**74**	**74**
Speaker 3F	315	166	158	126	**74**	130	**74**
Speaker 4F	217	224	220	219	242	282	273
Mean	266	192	183	170	167	165	145
St. Deviation	78	25	26	38	69	88	94

Table 2D: F0 Female

F0 cannot be reliably used to test the hypothesis because word-final /z/s are marked undefined regardless of their morphemic status. For example, Speaker 5M does not discriminate between the inflectional /z/ in <*Dan's*>, the non-morphemic /z/ in <*Franz*>, and the word-final /s/ in <*fierce*>. The mean F0 scores of the alveolar fricatives at the end of the 17 words in the data confirms what we already know about gender-based differences in pitch. The combined F0 measurements of word-final /z/s and /s/s by male talkers is 132 Hz, while that produced by females is 187 Hz. The difference of 55 Hz is perceptually significant. It confirms that there are anatomical differences in the larynxes of the male and female participants in our study. Stevens (2000:5, 9) reports that the vocal folds for male speakers is about 1.5 cm long, whereas that of female talkers is 1.0 cm long. The .5 cm difference results in higher F0 produced by female talkers compared to male talkers.

4.0 Focus on Intensity

A large number of studies, including Maddieson (1984:49-51), Ladefoged and Maddieson (1996:139), Thomas (2011:112), and Jogman et al. (2000), report that intensity is a robust cue for differentiating between sibilant and non-sibilant fricatives. However, it does not discriminate effectively among sibilant fricatives. For example, Jongman et al. (1998:201) report that /s/ and /z/ have exactly the same intensity, that is, 66.3 dB, but in their (2000) study, they report on page 1257 that there is an intensity difference of nearly 3 dB between them. The intensity of /z/ was 67.7 dB, whereas that of /s/ was 64.9 dB. Not every study finds /z/ to have a higher intensity than /s/. Frisch and Wright (2002:154) report that /s/ has a higher intensity than /z/. In pronunciations with 0 to 5% voicing, the intensity of /s/ was 56.3 dB, while that of /z/ was 49.1 dB. In pronunciation with 5 to 30% voicing, the intensity of /s/ was 57.1 dB, that of /z/ was 51 dB. Their findings confirm ours that the intensity of word-final /s/s is higher than that of word-final /z/s, as shown in Tables 3A through 3F:

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1M	56	59	60	57	61	58	58	63	58	57
Speaker 2M	48	51	52	48	48	52	52	55	49	52
Speaker 3M	58	59	57	57	59	58	59	61	56	55
Speaker 4M	65	70	68	69	64	69	63	70	67	68
Speaker 5M	61	67	66	64	68	67	67	68	64	65
Mean	58	61	60	59	60	61	60	63	59	59
St. Deviation	6	7	6	7	7	6	5	5	6	6

Table 3A: Intensity Male (dB)

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1M	57	56	58	54	60	58	59
Speaker 2M	53	52	51	52	58	52	48
Speaker 3M	54	57	57	56	60	57	56
Speaker 4M	65	68	66	64	71	64	68
Speaker 5M	62	66	66	63	66	66	67
Mean	58	60	59	58	63	59	60
St. Deviation	5	6	6	6	5	5	7

Table 3B: Intensity Male (dB)

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1F	58	60	60	60	56	60	59	64	55	54
Speaker 2F	53	59	58	60	60	56	59	61	51	57
Speaker 3F	58	63	62	61	62	60	60	65	60	61
Speaker 4F	56	55	58	54	55	57	60	61	58	55
Mean	56	59	60	60	59	59	59	63	55	57
St. Deviation	2	3	2	3	3	2	0	2	3	3

Table 3C: Intensity Female (dB)

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1F	54	58	56	56	62	56	55
Speaker 2F	59	57	57	56	56	57	56
Speaker 3F	60	62	62	59	64	59	62
Speaker 4F	54	59	55	58	61	57	55
Mean	58	59	58	57	61	57	58
St. Deviation	3	2	3	1	3	1	3

Table 3D: Intensity Female (dB)

Words	Dan's	Days	Daze	Dishes	dogs	Ease	Fears	Fierce	Franz	fusses
Mean	57	60	60	60	60	60	60	63	57	58
St. Deviation	5	6	5	6	6	5	4	4	6	5

Table 3E: Intensity Summary for all Speakers

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Mean	58	59	59	57	62	58	59
St. Deviation	4	5	5	4	4	4	6

Table 3F: Intensity Summary for all Speakers

Before determining whether or not intensity can be used to validate the hypothesis, we need to acquaint ourselves with intensity measurements and their relevance to speech perception. It is commonly noted that the smallest intensity difference that the human ear can perceive is 1 dB (Ladefoged 2003:90). This limen is known in the acoustic literature as Just Noticeable Difference (JND) in intensity. Even though the 1 dB threshold is technically accurate, it is valid only in controlled perception experiments in anechoic chambers (Burg et al. 2013:8). For intensity perception in everyday life, that is, for sounds or noises that occur outside laboratories, the JND of 3 dB is commonly used. Two signals are said to differ in intensity if there is a distance of 3 dB between them. Moore (2007:460) explains it as follows, "The smallest detectable change in intensity of a sound has been measured for many different types of stimuli by a variety of methods. … In everyday life, a change in level of 1dB would hardly be noticed, but a change in level of 3 dB would be fairly easily heard." In order to test the hypothesis, we will use 3 dB as an intensity threshold. If a difference of 3 dB is found to exist between the inflectional suffix /z/s and coda /z/s, then we would conclude that the two types of word-final /z/s are produced and perceived differently. However, if the difference between them is less than 3 dB, we would conclude that the participants in our study do not produce them differently.

The mean intensity score of the inflectional suffixes /z/ in <*Dan's, Liz's, Jay's, dogs, days, dishes, knees, goes, reads, fusses, is*> is 58.7 dB. The mean intensity score of the word-final /z/s in <*daze, ease, Franz*> is 59.66 dB. The intensity difference between the two types of /z/s is 0.96 dB. Since it is less than the required minimum JND of 3 dB, we conclude that the participants in our study do not pronounce word-final /z/s differently. If they devoice the inflectional suffix /z/, then they also devoice /z/ when it occurs in syllable codas, irrespective of whether it is a morpheme or not. In a follow-up study, speakers from Central Minnesota were asked to dictate <*daze of the week*> and

<*days of the week*> in Siri, Dragon Dictate, and Google Voice. In all instances, the outputs were the same. For all the participants, <*daze*> was perceived as <*days*>. The mean intensity scores for both words are the same, that is, 60 dB. Furthermore, eight of the nine participants produced the /z/ in <*days*> and the /z/ in <*daze*> identically. Only Speaker 4F produced them differently because the intensity difference between her /z/s is exactly 3 dB. Therefore, from the standpoint of intensity, there is no difference in the ways the participants produced morphemic and non-morphemic /z/s at the end of words.

5.0 Focus on Center of Gravity (CoG)

Center of Gravity (CoG) has more to do with place of articulation than voicing. This correlate helps to pinpoint articulatory differences between segments, usually fricatives, by showing the focal points of concentration of acoustic energy where they are produced (Gordon 2002). As a general rule of thumb, alveolar fricatives have higher CoGs than their non-alveolar counterparts. When using CoG to test Fromkin et al.'s hypothesis, we need to remember that frequency is perceived logarithmically, not arithmetically. The groundbreaking discoveries on how humans perceive frequencies were made by Fletcher (1940:50-51). The theoretical framework explaining his work is known as the Critical Bands Theory (CBT). We will not attempt to explain it here. Suffice it to say that different parts of the basilar membrane act as band filters for different frequencies. It takes only 1 Hz for people to perceive a difference between two speech signals on the F0 frequency band. However, on the F1 frequency band, the minimum distance required is 60 Hz. For F2, it is 200 Hz; for F3, it is 400 Hz; for F4, it is 630 Hz;[4] for F5, it is 800 Hz, and so on and so forth up to 20, 000 Hz. It is common knowledge that humans are capable of perceiving frequencies that range from 20 to 20,000 Hz. They cannot perceive frequencies lower than 20 Hz, nor can they perceive frequencies higher than 20,000 Hz.

For Fromkin et al.'s hypothesis to be validated, there should be a CoG difference of 630 Hz between the /z/s in <*Dan's, Liz's, Jay's, dogs, days, dishes, knees, goes, reads, fusses, is*> and those in <*daze, ease, Franz*>. Even before testing the hypothesis, we can surmise from Jongman et al. (2000:1257) that this is likely not the case. They found that the CoGs of /s/ and /z/ are identical in American English. We extrapolate from their findings that the participants in our study would not produce word-final /z/s differently. Let's examine our data to see if this prediction is borne out.

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1M	4962	5049	4791	4632	4739	4903	4461	4372	4695	4527
Speaker 2M	5542	5637	5542	5729	5098	5781	5184	5330	5386	5441
Speaker 3M	5580	5948	6583	5958	5585	6033	5782	5909	5699	5490
Speaker 4M	5753	5997	6218	6152	5764	6301	5598	5851	5980	5902
Speaker 5M	5066	4864	4935	5025	4714	5323	4681	4641	4890	5182
Mean	5381	5499	5614	5499	5180	5668	5141	5221	5330	5308
St. Deviation	346	518	782	645	480	559	569	696	538	507

Table 4A: Center of Gravity Male (Hz)

[4] The JND of 630 Hz is a compromise between the F4 of males, which is at 600 Hz, and that of females, which is at 700 Hz. See Stevens (2000:154, 300) for additional information.

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1M	4456	4751	4838	5084	4935	4946	4746
Speaker 2M	5183	5798	5564	5788	6189	5548	5541
Speaker 3M	5563	5909	5691	6420	6468	5356	5989
Speaker 4M	5830	6117	5872	5771	6227	5953	6138
Speaker 5M	4690	6416	5101	5066	5174	5151	5219
Mean	5144	5598	5413	5626	5670	5519	5527
St. Deviation	576	631	430	567	693	386	568

Table 4B: Center of Gravity Male (Hz)

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1F	6484	6858	6368	7164	7162	6505	5867	5939	6280	6274
Speaker 2F	8562	8585	8272	8252	7799	8041	6675	5224	8011	7708
Speaker 3F	8792	8076	8397	8393	8674	8718	7451	8829	8714	8738
Speaker 4F	4893	5557	8051	5933	6775	6851	6328	9395	7524	5894
Mean	7183	7269	7772	7435	7603	7529	6580	7347	7632	7153
St. Deviation	1846	1351	947	1142	830	1030	668	2072	1025	1314

Table 4C: Center of Gravity Female (Hz)

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1F	4900	6500	6392	6416	6630	6555	6457
Speaker 2F	6331	7994	9196	7011	7270	8844	8910
Speaker 3F	6797	8290	8492	8685	9102	8751	9103
Speaker 4F	4446	5606	3922	7798	9043	7635	5840
Mean	5619	7097	7001	7478	8011	7947	7577
St. Deviation	1124	1266	2373	984	1253	1078	1671

Table 4D: Center of Gravity Female (Hz)

Words	Dan's	Days	Daze	Dishes	dogs	Ease	Fears	Fierce	Franz	fusses
Mean	6282	6384	6693	6467	6391	6599	5861	6284	6481	6231
Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads			
Mean	5381	6348	6207	6552	6841	6733	6552			

Table 4E: Center of Gravity Summary for all Speakers

The mean CoG score of the inflectional suffix /z/ in <*Dan's, days, dishes, dogs, fears, fusses, goes, is, Jay's, knees, Liz's, reads*> is 6,282 Hz. The one in the coda of <*daze, ease, Franz*> is 6591 Hz. The difference between the two types of /z/s is 309 Hz, which is lower than the 630 Hz required for them to be perceived differently. In other words, the participants in our study produced both word-final /z/s identically. Moreover, since the mean CoG of the /s/ in <*fierce*> and <*niece*> is 6,562 Hz, and also since it is less than 630 Hz from either /z/s, we conclude that the participants in our study produce their alveolar fricatives similarly. This finding agrees with Jongman et al. (2000: 1257) who report that /s/ and /z/ are identical with respect to CoG. It is important to underscore that CoG is not a measure of voicing, but of place of articulation. Consequently, it is an ineffective correlate for testing the hypothesis.

6.0 Focus on Duration

Stevens et al. (1992:2979) report that listeners decide whether a fricative segment is voiced or voiceless on the basis of its duration alone. Their review of the literature is replete with sources that hold this view. They quote Crystal and House (1982:710) who found that in running speech, voiceless fricatives are longer than their voiced counterparts. On average, the former lasts 94.5 ms, while the length of the latter is 47.5 ms. Frisch and Wright (2002:154) report similar results for [s] and [z]. For voicing duration of 0 to 5%, they found that [s] was 173 ms long versus 133 ms for [z]. For voicing duration of 5 to 30%, [s] lasted 148 ms compared to [z] that lasted 127 ms. Everything being equal, voiceless fricatives are longer than their voiced counterparts. Jongman et al. (2000:1260) provide duration data that is in agreement with the aforementioned measurements. They report that the mean duration of [s] is 178 ms, while that of [z] is 118 ms.[5] Furthermore, Smith (1977:482) lists three durational measurements for /z/ that highlights the correlation between fricatives and voicing. She found that voiced /z/ was 70.5 ms long, while devoiced /z/ was 91 ms. All this leads to the following correlation between voicing and duration:

Voiceless fricatives >[6] devoiced fricatives > fully voiced fricatives

The explanation is that voiceless fricatives are longer than devoiced ones, which are also longer than voiced ones. Gradoville (2011:64) talks about a similar correlation, which he explains as follows:

> Duration, although strictly speaking not a measurement of voicing per se, may correlate with fricative voicing. … A second duration effect that the researchers found was that, as the frication became longer, so did the likelihood decrease that a voiced response would occur. … with longer durations believed to correspond with decreased in voicing.

If we find that the mean duration score of the /z/s in <*Dan’s, days, dishes, dogs, fears, fusses, goes, is, Jay’s, knees, Liz’s, reads*> is longer than those in <*daze, ease, Franz*>, then we would say that the duration correlate supports Fromkin et al.’s hypothesis. Let’s examine the data to see if the hypothesis is validated or invalidated.

Words	Dan’s	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1M	196	208	219	214	231	290	232	299	227	213
Speaker 2M	143	169	190	172	146	182	175	212	140	176
Speaker 3M	283	300	346	296	252	325	330	372	322	315
Speaker 4M	275	266	248	304	244	248	264	337	278	265
Speaker 5M	222	224	229	225	255	278	298	310	212	235
Mean	224	233	247	242	225	265	260	306	236	241
St. Deviation	58	51	59	56	45	54	60	60	69	53

Table 5A: Duration Male (ms)

[5] Their findings in this study contradict their earlier findings in Jongman et al.’s (1998:202), where they stated that “noise duration does not seem to be an important cue to fricative voicing.”

[6] The symbol “>” stands for “longer than.”

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1M	250	251	228	207	322	242	220
Speaker 2M	166	210	154	178	219	184	146
Speaker 3M	262	352	309	314	401	248	242
Speaker 4M	273	276	277	280	300	287	259
Speaker 5M	234	258	236	253	350	233	245
Mean	237	269	241	246	313	239	222
St. Deviation	42	52	58	55	67	37	45

Table 5B: Duration Male (ms)

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1F	146	149	211	234	215	176	191	213	181	212
Speaker 2F	228	278	269	241	234	279	241	356	204	279
Speaker 3F	200	232	279	233	203	280	255	368	223	259
Speaker 4F	132	158	170	180	162	187	190	264	156	174
Mean	176	204	232	222	191	240	225	315	191	231
St. Deviation	45	62	51	28	30	57	34	74	29	47

Table 5C: Duration Female (ms)

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1F	349	198	171	155	242	219	165
Speaker 2F	259	279	236	211	351	221	268
Speaker 3F	265	269	240	274	370	224	236
Speaker 4F	149	217	142	180	268	162	173
Mean	223	247	197	205	308	206	210
St. Deviation	82	39	48	51	62	30	50

Table 5D: Duration Female (ms)

Words	Dan's	Days	Daze	Dishes	dogs	Ease	Fears	Fierce	Franz	fusses
Mean	200	219	239	232	208	252	243	310	213	236
St. Deviation	55	54	55	45	39	55	52	62	57	47

Table 5E: Duration Summary for all Speakers

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Mean	230	258	219	223	313	223	216
St. Deviation	59	47	56	54	61	36	44

Table 5F: Duration Summary for all Speakers

Before using duration to test the hypothesis, let's first highlight how the duration cue is interpreted in acoustic phonetics studies. Since Hirsh (1959:765), studies have confirmed that the JND for duration is 10 ms. Better yet, if the duration distance between two acoustic signals is ≥ 17 ms, then they "perceived correctly" (Hirsh 1959:767). The most commonly used limen for duration is the 10 ms threshold. In other words, we will say that the inflectional suffix /z/s in <*Dan's, days, dishes, dogs, fears, fusses, goes, is, Jay's, knees, Liz's, reads*> are longer than those in <*daze, ease, Franz*> if there is at least 10 ms difference between them. The mean duration of the former is

225 ms, while that of the latter is 234 ms.[7] The difference between them is 9 ms. Since it falls below the JND in duration, we conclude that the participants in our study do not produce the two types of /z/s at the end of words differently. In other words, they devoice both the inflectional suffix /z/ and the non-morphemic /z/s in the coda of syllables.

7.0 Focus on Coda Voicing and the 40/60 Threshold

In order to determine the voicing status of word-final /z/s, we turn first to Smith's (1997) study on the devoicing of /z/ in American English, and then we consider the 40/60 Threshold proposed by Gradoville (2011). Smith makes the following statement on pages 478-9:

> The tokens of /z/ were divided into three categories according to the percentage of their duration during which there was voicing. The three categories were: 0-25% voicing = devoiced, 25-90% voicing = partially devoiced, 90-100% voicing = voiced. … There was not a very clear boundary between devoiced and partially voiced categories for any speaker, but the 0-25% division grouped together most of the tokens with less voicing.

Gradoville (2011:68) has proposed a simpler subdivision. He found that when 40% of a fricative segment is voiced, the participants in his study perceived the whole segment as voiced. Alternatively, when more than 60% was unvoiced, the segment was perceived as devoiced. When we combine the insights from Smith and Gradoville, a three-way distinction in voicing can be made:

1. If 10% or less of a segment is voiced, it is voiceless.
2. If 40% or more of a segment is voiced, it is voiced.
3. If 60% to 90% of a segment is unvoiced, it is devoiced.[8]

Praat makes it relatively easy to use these numerical parameters to determine if a segment is voiced, devoiced, or voiceless. All one needs to do is highlight the segment under consideration, click on the `Pulse' tab, select `Show Pulse', and click on `Voice Report'. Once these steps are completed, Praat displays percentages of voicing such as those reported in the tables below:

[7] The mean duration of the /s/s in <fierce> and <niece> is 311 ms.

[8] We fold Smith's "partially devoiced" and "devoiced" classification into one, since it proved to be inconsequential.

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1M	5	7	19	5	6	16	12	9	9	16
Speaker 2M	5	15	14	15	5	13	17	9	7	14
Speaker 3M	2	7	3	14	1	3	8	3	1	6
Speaker 4M	10	21	14	11	1	9	19	7	8	15
Speaker 5M	1	12	3	1	6	9	25	1	0	0
Mean	5	13	11	9	4	12	17	7	7	9
St. Deviation	3	6	7	6	3	5	6	4	4	7

Table 6A: Coda Voicing Male (%)

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1M	17	11	12	11	7	13	9
Speaker 2M	14	21	15	8	12	22	0
Speaker 3M	6	10	8	6	14	2	3
Speaker 4M	10	13	11	17	7	4	8
Speaker 5M	1	6	3	7	1	8	3
Mean	14	12	10	13	4	11	4
St. Deviation	6	5	4	4	5	8	4

Table 6B: Coda Voicing Male (%)

Words	Dan's	Days	Daze	Dishes	Dogs	Ease	Fears	Fierce	Franz	Fusses
Speaker 1F	12	20	22	21	16	17	22	9	11	10
Speaker 2F	1	6	6	3	0	11	22	16	6	2
Speaker 3F	0	3	0	2	1	6	13	1	0	3
Speaker 4F	13	13	17	16	2	18	17	6	19	17
Mean	7	10	11	10	5	13	19	8	9	8
St. Deviation	7	8	10	9	7	6	4	6	8	7

Table 6C: Coda Voicing Feale (%)

Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads
Speaker 1F	29	16	12	23	8	19	0
Speaker 2F	15	4	13	16	4	3	6
Speaker 3F	21	10	3	12	3	0	0
Speaker 4F	9	18	16	11	6	17	12
Mean	18	12	11	16	3	10	5
St. Deviation	8	6	6	5	2	10	6

Table 6D: Coda Voicing Male (%)

Words	Dan's	Days	Daze	Dishes	dogs	Ease	Fears	Fierce	Franz	fusses
Mean	6	12	11	10	4	12	17	7	7	9
Words	Goes	Is	Jay's	Knees	Niece	Liz's	reads			
Mean	14	12	10	13	4	11	4			

Table 6E: Coda Devoicing Summary for all Speakers

Let's examine the voice reports of the alveolar fricatives that occur at the end of the words in our data. The mean voice report of the /z/s in <*Dan's, days, dishes, dogs, fears, fusses, goes, is, Jay's, knees, Liz's, reads*> shows that 10.16% is voiced. In other words, about 90% of the duration of /z/ is unvoiced. Therefore, the inflectional suffix /z/ is devoiced. The portion of the word-final /z/s in <*ease, Franz, daze*> that is voiced is 10%, while 90% of it is unvoiced. The 40/60 Threshold confirms that the word-final /z/s in the data are produced identically, namely that they are both devoiced because their locally unvoiced portions are approximately 90% of their total frication noise.[9]

8.0 The Speech Recognition Test

If Fromkin et al.'s hypothesis is correct, word recognition systems such as Siri, Dragon Dictate, and Google Voice should discriminate between <*days*> versus <*daze*>. However, when some talkers from Central Minnesota are asked to dictate <*daze of the week*> into Siri, Dragon Dictate, and Google Voice the response that they get is <*days of the week*>. The word <*daze*> is consistently rendered as <*days*>. This confirms our findings that there the inflectional suffix /z/ and the /z/ that occurs in the coda of syllables are produced identically. Consequently, the second part of Fromkin et al.'s hypothesis is also not supported by automatic speech recognition data.

The speech recognition data also confirms the observation made in 3.0, namely that some speakers pronounce word-final /z/s, not as devoiced [z̥]s, but as voiceless [s]s. To test this claim further, we asked several CMNE speakers who were not part of the original experiments to dictate the sentence <*face your fierce fears day*>[10] into Siri, Dragon Dictate, and Google Voice. If <*fierce*> is confused with <*fears*>, then this would support the contention that the devoiced [z̥] in <*fears*> is produced as voiceless [s]. The following response was given <*face your fears fierce day*>. The substitution of <*fierce*> by <*fears*> demonstrates clearly that the word-final [z]s in these two words are produced identically. These two words are homophones in the speech of many talkers in Central Minnesota. Speakers 4M, 5M, 1F and 2F produced the [z]s and [s]s in <*knees*> and <*niece*> identically. Consequently, <*hurting knees*> sounds the same as <*hurting niece*>.

9.0 Implication for Variation Studies

There is a progressive change in the pronunciation of voiced alveolar fricatives, which cause them to be devoiced to [z̥]s. They, in turn, are changed into [s]s. This may be seen as the continuation of a process that began in the Middle English period. Fromkin et al. (2014:341-2) report that /v, z, ð/ did not exist in Old English as full-fledged phonemes, but were simply allophones of /f, s, θ/. The evidence presented in this paper shows that fricatives continue to undergo changes. It may be the case that in the dialects that Fromkin et al. had in mind for their hypothesis, there is a perceptual difference between the inflectional suffix /z/ which is devoiced to [z̥], and the coda /z/ that is fully voiced. However, as our measurements indicate, for the speakers of CMNE, all word-final /z/s are devoiced. In some instances, the devoicing is so strong that there is no perceptual difference between the devoiced [z̥] in the coda and voiceless [s]. Hennen and Koffi (2017:74, Figure 1) show that words such as <*these*> have variable pronunciations. Sometimes the /z/ in <*these*> is devoiced as [z̥]. In other cases, it is

[9] The mean voice report measurements for the /s/s in <*fierce, niece*> is 4.5% voicing and 95.5% unvoicing.

[10] October 18th is the "National Face Your Fears Day."

unvoiced as [s]. The unvoicing is strong among some speakers to the point that the /z/s at the end of <*please*> and <*cheese*> sound like and [s]. When this happens <*please*> and <*cheese*> are transcribed narrowly as [plis] and [tʃis].

10.0 Summary

We have used five acoustic measurements to test the hypothesis formulated by Fromkin et al (2014:278). The F0 correlate shows that 14.97 % of all word-final /z/s are marked "unidentified," which means that they are, for all practical purposes, strongly devoiced or voiceless. The intensity correlate does not show that the inflectional suffix /z/ is produced differently from the [z] in the coda of words because the intensity difference between them falls short of the 3 dB threshold. The CoG cue shows that alveolar fricatives are produced identically in CMNE. The frequency difference between word-final [z]s, [z̥]s, and [s]s is well below the minimum of 630 Hz difference need to discriminate between alveolar fricatives. The JND in duration also shows that all word-final /z/s in the data are produced the same. The durational distance between the two types of devoiced /z/ falls short of the minimum threshold of 10 ms. Finally, the voice report shows that all the word-final /z/s in the data are devoiced equally. Their unvoiced portions are approximately 90% of the entire duration of the frication noise. Consequently, they are perceptually identical. The voice report findings are robust cue because, according to Gradoville (2011:71), "The voice report most closely matches what the linguistically-trained participants perceived."

In light of the findings discussed in this paper, we conclude that the first part of the hypothesis formulated by Fromkin et al (2014:278) is fully validated. The inflectional suffix /z/ in <*Dan's, Liz's, Jay's, days, goes, dishes, dogs, fusses, goes, is, fears, knees, reads*> are devoiced to [z̥]. However, the second part of the hypothesis which claims that the /z/s in the coda of <*daze, ease, Franz*> are fully voiced is not supported by our measurements. The five acoustic correlates used to test this part of the hypothesis did not find that CMNE talkers produced word-final /z/s differently. They were all devoiced irrespective of their morphemic status. In the process of testing the Fromkin et al.'s hypothesis, we discovered rather accidentally that some Central Minnesota speakers devoiced some word-final /z/s so strongly that they become unvoiced. As a result, <niece> and <knees> on the one hand, and <fierce> and <fears> on the other are true homophones. Their unvoiced portions are ≥ 90% of the duration of frication noise.

ABOUT THE AUTHORS

Ettien Koffi, Ph.D. in linguistics from Indiana University, teaches linguistics at Saint Cloud State University, MN. Author of many peer-reviewed articles on various topics in linguistics and of four books: *Language Society in Biblical Times* (1996), *Paradigm Shift in Language Planning and Policy: Game Theoretic Solutions* (2012), *Applied English Syntax* (2010, 2015), and the *New Testament in Anyi Morofu* (2017), a task which took over 25 years. Specializing in acoustic phonetics, dialect variation, and emergent orthographies, his current research centers on speech acoustics of L2 English (within the Speech Intelligibility Framework), Central Minnesota English, and Anyi. He can be reached at enkoffi@stcloudstate.edu.

Cassy Lundy is a Storage Engineer with Dell Compellent Technologies and the Technical Director for the Pioneer Place Theater Company in St. Cloud, MN. He has bachelor's degrees in Linguistics and Spanish from St. Cloud State University and extensive background in audio/visual and information technology. Casey plans to attend grad school to study human and computer interaction from the perspective of linguistics and computer/information science. He can be reached via email at luca0901@stcloudstate.edu or at casey.lundy@gmail.com.

References

Boersma, Paul & Weenink, David (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.17, retrieved 21 April 2016 from http://www.praat.org/

Belvin, Juliette. 2003. The Independent Nature of Phonotactic Constraints: An Alternative to Syllable-Based Approaches, pp. 375-403. *The Syllable on Optimality Theory*, ed. By Caroline Fery and Ruben van de Vijver. Cambridge University Press: New York.

Burg, Jennifer, Jason Romney, and Eric Schwartz. 2008. Digital Sounds & Music: Sound Perception and Acoustics. Online textbook available at: http://csweb.cs.wfu.edu/~burg/CCLI/Templates/curriculum_index.php

Crystal, Thomas H. and Arthur S. House. 1982. Segmental Durations in Connected Speech Signals: Preliminary Results. *Journal of the Acoustical Society of America* 72 (3): 705-716.

Gradoville, Michael. 2011. Validity in Measurements of Fricative Voicing: Evidence from Argentine Spanish. Selected Proceedings of the *5th Conference on Laboratory Approaches to Romance Phonology*, ed. Scott M. Alvord, pp. 59-74. Somerville, MA: Cascadilla Proceedings Project.

Fletcher, Harvey. 1940. Auditory Patterns. *Reviews of Modern Physics*, Volume 12, pp. 47-65.

Fry, Dennis B. 1979. Fry, Dennis. B. 1979. The *Physics of Speech.* New York: Cambridge University Press.

Jongman, Allard, Ratree Wayland, and Serena Wong. 1998. Acoustic Characteristics of English Fricatives: I. State Cues. *Working Papers of Cornell Phonetics Laboratory*, Volume 12:195-205.

Jongman, Allard, Ratree Wayland, and Serena Wong. 2000. Acoustic Characteristics of English Fricatives. *Journal of the Acoustical Society of America* 108 (3): 1252-1263.

Hirsh, Ira J. 1959. Auditory Perception of Temporal Order. *Journal of the Acoustical Society of America* 31 (6): 759-767.

Hennen, Alex and Ettien Koffi. 2017. The Acoustics of Coda Devoicing in a Central Minnesota English Ideolect. *Linguistic Portfolios*, Volume 6: 71-81.

Kachru, Braj. 1997. Foreword, pp. iii-viii. New Englishes: A West African Perspectives, ed. By Ayo Bamgbose, Ayo Banjo, and Andres Thomas. African Wordld Press, Inc.: Trenton, Jew Jersey.

Koffi, Ettien. 2016. *Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility.* Manuscript. St. Cloud, MN.

Koffi, Ettien. 2016. The Acoustic Correlates of [±ATR] Vowels: An Analysis by Reference Levels of Anyi Vowels. *Linguistic Portfolios*, Volume 5:115-134.

Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Malden, MA: Blackwell Publishers Inc.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge Studies in Speech Science and Communication. New York: Cambridge University Press.

Moore, Brian C.J. (2007:460). Psychoacoustics, pp. 459-501. In *Springer Handbook of Acoustics*, ed. by T. D. Rossing. New York: Springer Science+Business, LLC.

Smith, Caroline L. 1997. The Devoicing of /z/ in American English: Effects of Local and Prosodic Context. *Journal of Phonetics* 25, 471-500.

Stefan A. and Richard Wright. 2002. The Phonetics of Phonological Errors: An Acoustic Analysis of Slips of the Tongue. *Journal of Phonetics* 30: 139-162.

Stevens, Kenneth, Sheila Blumstein E, Laura Glicksman, Martha Burton, and Kathleen Kurowski. 1992. Acoustic and Perceptual Characteristics of Voicing in Fricatives and Fricative Clusters. *Journal of the Acoustical Society of America* 91 (5): 2979-3000.

Stevens, Kenneth N. 2000. *Acoustic Phonetics.* The MIT Press: Cambridge, Massachusetts.

Thomas, Erik R. 2011. *Sociophonetics: An Introduction*. New York: Palgrave Macmillan.