1997

Fricative devoicing: effects of prosodic context on a lenition process

Caroline L. Smith
University of Ottawa

Fricative devoicing: effects of prosodic context on a lenition process

Caroline L. Smith

University of Ottawa[1]

Prosodic structure organizes speech into feet, words and phrases which are characterized by stress, intonation and other properties. Stress and intonation distinguish one syllable or word from another by assigning them different degrees of prominence, so that one syllable or word is, for example, relatively louder, higher in pitch or longer than another. If prosodic structure also organizes speech on a smaller as well as a larger scale, then the relative prominence of individual speech sounds would be expected to vary as a function of their position in the prosodic structure of an utterance. Lenition processes provide an example of the kind of variability that can result from differences in prosodic structure. This kind of variability can be attributed to differences in the suprasegmental context of an utterance, much as differences in local context create variability in the realization of speech sounds. One common kind of variation in English, especially for voiced sibilants, is devoicing. I will argue that in certain contexts this can be a lenition process that is subject to prosodic influences similar to those governing some other forms of lenition.

Investigation of the effects of prosodic structure on speech production have revealed that articulatory movements are often larger in extent at the beginning of prosodic domains, and reduced in size at the end of prosodic units. This tendency has been investigated for prosodic domains varying in size from the syllable to the phonological phrase, suggesting that syllable-initial movements tend to be larger than syllable-final ones, and phrase-initial ones larger than phrase-final. (The question of whether such processes should in general be ascribed to initial strengthening or final weakening is discussed in detail by Fougeron & Keating 1996.) Examples of articulatory reduction in syllable-final position compared to syllable-initial position have been observed for all parts of the vocal tract. In the supralaryngeal vocal tract, there is reduced contact of the tongue to the palate in stops (Keating 1995; Fougeron & Keating 1996), loss of oral closure in syllable-final voiceless stops (Manuel 1991), and smaller movements in syllable-final stops (Browman & Goldstein 1995). For the nasal and laryngeal subsystems of the vocal tract, evidence of similar patterns has been observed as greater RMS magnitude (more glottal opening) in phrase-initial /h/ than in phrase-medial /h/ (Pierrehumbert & Talkin 1992), larger glottal

[1]This work was supported by a postdoctoral fellowship from NIH grant DC00008 to the UCLA Division of Head & Neck Surgery. Thanks also to the UCLA Phonetics Laboratory for the use of their facilities.

K. Kusumoto (ed.), NELS 27, 397-411

opening in syllable-initial aspirated stops (Cooper 1991), and as more velum raising in phrase-initial position than in phrase-final (Krakow, Bell-Berti & Wang 1995).

1. Fricative devoicing as an example of articulatory reduction

Articulatory reduction might be expected to be especially evident in the case of sounds that are difficult to produce, regardless of their context. If the requirements to produce a given sound are more demanding, the speaker may more frequently fall short of the intended production. Voiced fricatives are one class of sound that is often described as difficult to produce. This is because their production imposes two potentially competing requirements on the vocal tract. To produce voicing, the subglottal air pressure must be sufficiently higher than the air pressure in the oral cavity. But in order to produce the turbulent air flow that generates the frication noise, a relatively narrow oral constriction is necessary. This narrow oral constriction impedes the airflow through the mouth and hence increases the pressure in the oral cavity, making it difficult to sustain the necessary transglottal pressure difference which causes the vocal folds to vibrate. Calculations based on the cross-sectional areas of glottal and oral constrictions and the pressures in different parts of the vocal tract show that special maneuvers explicitly directed at prolonging glottal vibration are required to make voicing last for the entire duration of the oral constriction in a voiced fricative (Stevens et al 1992). This difficulty of producing voicing with frication has been cited as the reason why phonologically voiced fricatives are often devoiced (Ohala 1983), and it leads to a more general expectation that segments of this type will show variability in production. For these reasons, the fricative /z/ in American English was chosen as the target of investigation in the present study.

There are two ways in which the difficulty of producing voiced fricatives might be avoided. One way would be to modify the requirement that voicing be sustained, which would result in a devoiced fricative. The other way would be to form a less tight oral constriction, so that the oral pressure would be reduced. This allows a transglottal pressure difference that is adequate for voicing, but at the cost of failing to create the turbulent airflow that is necessary for frication noise. The outcome would be a voiced approximant rather than a fricative.

Of these two potential ways of simplifying voiced fricatives, the loss of voicing seems to be more common, at least for English (Haggard 1978, Docherty 1992). Speakers do not usually simplify voiced fricatives by widening the oral constriction and allowing oral pressure to fall. In particular, speakers of English seem to avoid the loss of frication in sibilants, although it does occur in non-sibilant fricatives. In the present study, the target fricative, /z/, was frequently devoiced but was never produced without frication. However, there seem to be two different sources of devoicing: one arises from assimilation to a voiceless environment adjacent to the fricative, and the other is a form of reduction that appears to be more likely in some prosodic contexts than others.

A phonologically voiced fricative is said to be devoiced if it lacks vocal fold vibration during all or part of its frication duration, defined as the time during which there is strong aperiodic noise. Note that there are several ways in which an absence of voicing can occur. An active glottal opening gesture might be substituted for the voicing configuration (which would neutralize a contrast between voiced and voiceless fricatives), or the maneuvers required to sustain voicing might be reduced or omitted. In the latter case, devoicing during phonologically voiced fricatives can be viewed as a lenition process because it is a relaxation towards a less-demanding articulation. (This may appear surprising, as lenition of consonants is often considered to involve a change from voiceless to voiced, especially in intervocalic position (Lass 1984).)

2. Amount and likelihood of devoicing

For any devoiced fricative, it is possible to measure the extent of devoicing in two different dimensions. One of these, AMOUNT of devoicing, is equal to the percentage of fricative duration that is devoiced, i.e., the time during which there is both voicing and frication divided by the total duration of frication. This percentage varies in different productions, even among those of a single speaker, so comparisons can be made of the Amount of devoicing observed in different productions of a fricative in any given context.

Fricatives produced in different contexts can also be compared as to the LIKELIHOOD of devoicing in these different contexts. Likelihood refers to how often a fricative in a particular context was devoiced, over all the productions in that context during the course of an experiment. This variable has received more attention in previous studies than the amount of devoicing. One reason for this emphasis may be because likelihood of devoicing in different contexts can be studied by making a binary determination (voicing/devoicing) for each individual token, but comparison of the amount of devoicing in different tokens requires that the entire time course of each token be examined. Previous studies that have investigated the frequency of occurrence of devoicing have found substantial variation among different phonological contexts. For British English, Haggard (1978) found that fricatives following a voiced stop were devoiced more often than intervocalic fricatives, and fricatives following a voiceless stop were even more likely to be devoiced. Docherty (1992) for British English, and Veatch (1989) and Stevens et al. (1992), for American, observed similar patterns.

3.1 Experimental method: materials

These earlier studies identified several contextual factors that may contribute to devoicing in English fricatives. The goal of the present experiment was to investigate the effects of both local and prosodic contexts on the devoicing of /z/, and to investigate differences in the production of voiced and devoiced /z/'s. This was done by recording native speakers of American English while they read sentences which contained /z/'s in a variety of contexts. Examples of /s/ were also collected to provide a basis for comparison with the devoiced /z/'s. In order to encourage the speakers to produce natural speech, the target fricatives were produced in (more or less) meaningful sentences. In these sentences, /s/ and /z/ occurred in contexts matched for type of neighboring sounds and position in word or phrase. For each context in which a /z/ was measured, an /s/ in a matching context was measured as well. While it was not always possible to match the contexts for paired /s/ and /z/ exactly, their contexts were as similar as possible with respect to the phonological factors. For example, the /z/ in sentence (1) and the /s/ in sentence (2) constituted one matched pair. Their contexts have the same value for the factors PROSODIC POSITION (word-final), and FOLLOWING SOUND (a voiced stop, /b/).

(1) /z/
There was a short pau<u>se</u> before she answered her boss.

(2) /s/
John's bo<u>ss</u> bemoaned his false pretenses for avoiding work.

The word or words that made up the immediate context for a target /s/ or /z/ will be referred to as an utterance, and one repetition of an utterance will be referred to as a token. Note that in this usage, an utterance consists of only one or two words and is much smaller than a sentence. The experiment included 20 utterances containing target /z/ and

20 containing target /s/. In order to reduce the duration of the experiment, two or three target fricatives were included in each sentence. However, the utterances for each pair of contexts matched for /s/ and /z/ occurred in different sentences. The complete set of sentences is given in Table I.

Table I. Sentences used in the experiment. The underlining has been added to mark the fricatives that were measured.

> Ms. Barnes observed him reading this book while he was eating dessert.
> Her husband wears a false beard that slides around when he sits down.
> John's boss bemoaned his false pretenses for avoiding work.
> The music paused for a long time after these bands finished playing.
> Pour that liquid into the red sink, and make sure the zinc closure fits tightly.
> We should replace broken glass from the earthquake before any more of it falls.
> The red zinc platter in the kitchen belongs to my housebound aunt.
> His boss asked him why he falls behind in his work so often.
> The jack in the box pops out very quickly.
> When Bob's out, the noise level falls perceptibly.
> The statement "Niagara Falls is in Vermont" is totally false.
> On a test question, choosing true or false is easier than multiple choice.
> There was a short pause before she answered her boss.
> Mary's boss laughed for five minutes without a pause.
> The long pause outraged impatient listeners in the Roseland concert hall.
> The hushed pause lengthened as the Mafia boss passed down Rossland Avenue.
> After the protester shouted obscene slogans at the palace guard, he escaped through an ingenious deceit.
> The pitcher's lengthy pause postponed the start of the Dodgers game.
> A lunar cycle recurs basically once every 28 days.

Factors that were varied in the experiment can be divided between those affecting the local context of the fricative and those relating to the phrasal context or prosodic position. Local context includes the identity of the following sound and of the preceding sound. The prosodic position of the fricative was either syllable-final, word-final, or sentence-final. Another prosodic factor that was varied was the presence or absence of stress on the syllable containing the fricative (see Table II). The experiment also included fricatives produced in additional contexts that did not specifically test the effects of the factors listed here.

Table II. Matched pairs of utterances that were used to test the effect of prosodic factors. Listed here are the immediate contexts for the target fricatives, which were spoken as part of the complete sentences listed in Table I. Utterances in the same row share the same value for the factor listed at left.

Prosodic context		
position in utterance	/s/	/z/
syllable-final	housebound	husband
word-final	boss bemoaned	pause before
sentence-final	boss.	pause.
stress pattern		
in unstressed syll	palace guard	Dodgers game
in stressed syll	replace broken	recurs basically

3.2 Experimental method: data collection

Since the purpose of this study was to investigate the occurrence of voicing and devoicing, it was essential to establish a reliable, consistent method for identifying times when there is or is not voicing during a fricative. Previous work on devoicing has mostly been based on acoustic measurements, but it is not always easy to identify unambiguously the exact onset and offset of voicing from a waveform and spectrogram. In this experiment, the technique known as electroglottography (EGG) was used because it is a measure that shows vocal fold vibration independently of the supralaryngeal vocal tract. EGG involves the speaker wearing a fabric collar with electrodes that are sensitive to variation in the level of a low current passed through the speaker's neck and larynx. The electrodes detect the presence or absence of contact between vocal folds, and thus provide the information necessary to make a determination of the presence or absence of voicing. The other physiological measure that was made was oral airflow. This was used to estimate the amount of opening of the vocal folds, and any variation in subglottal pressure, although it can be hard to separate the contributions of these different factors. Because physiological measures provide less ambiguous information about the state of the vocal folds than do acoustic measures, their use in investigating devoiced fricatives during running speech constitutes one of the major contributions of this study.

During the experiment, each speaker wore the EGG collar to measure vocal fold contact, and a pneumotachographic face mask to measure airflow. These signals and the acoustic signal from a head-mounted microphone were recorded directly to disk at an 8 kHz sampling rate. To permit acoustic analysis of the high frequencies present in fricatives, a tape-recording was made simultaneously and digitized at 20kHz. For further details of the experimental method, see Smith (1996).

Four speakers were recorded. They were young adults (20's and 30's) from the Midwest and Western United States. Speakers 1 and 3 were male, Speakers 2 and 4 were female. All speakers had previous experience using the experimental set-up and were capable of speaking in a comfortable and relaxed manner despite the presence of the face mask and EGG collar. Speaker 1 read through the set of sentences 5 times; the other speakers read them 6 times. Because of difficulties with equipment, several of speaker 1's sentences were not recorded properly. A few tokens from speakers 1, 2 and 3 had to be discarded because of speaker error. The total number of matched pairs of /z/ and /s/ that were analyzed for each speaker were:

Speaker 1	83	Speaker 3	117
Speaker 2	119	Speaker 4	120

3.3 Experimental method: measuring the amount of voicing

During a voiced sound, the EGG signal shows oscillations corresponding to the vibration of the vocal folds between contact and separation. Vocal fold vibration does not usually start or stop abruptly, but in order to facilitate comparisons among different tokens it was desirable to identify specific times corresponding to the beginning or end of voicing. The following algorithm was used to identify these times, as illustrated in Figure 1. The amplitude of one EGG cycle (maximum – minimum during one excursion) was measured at the time of maximum acoustic RMS energy in the vowel preceding the fricative. In utterances in which a stop preceded the fricative, the EGG amplitude was measured at the time of maximum RMS energy in the vowel preceding the stop. The EGG amplitude during the vowel was divided by 10 to get a criterion amplitude: the fricative was considered to be voiced during the portion of its duration that the amplitude of the EGG cycles exceeded the criterion amplitude. Voicing was considered to cease when the amplitude of an EGG cycle fell below this criterion. If the amplitude of several successive EGG cycles wavered between just above and just below the criterion, the offset of voicing was marked where the average amplitude of two successive cycles was below the criterion. This method permitted a single time to be identified as the onset or offset of voicing in a fricative that was partially voiced. After these times had been identified, for each token of /z/, the percentage of fricative duration with voicing was calculated by dividing the duration of frication with EGG amplitude exceeding the criterion by the total duration of acoustic frication.

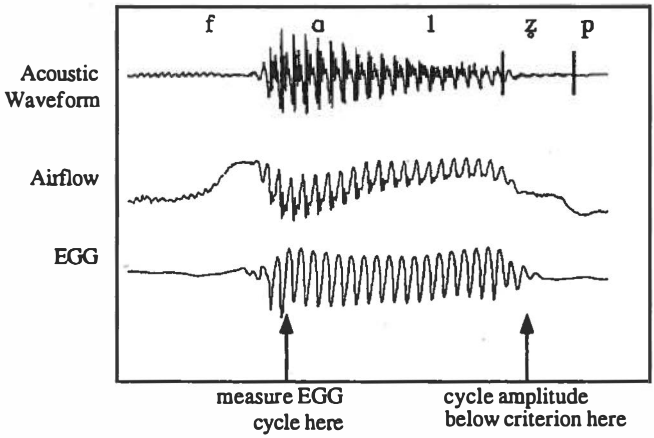

Figure 1. A partially devoiced token of "falls perceptibly", spoken by Speaker 1.

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

Each token of /z/ was categorized individually; therefore it was possible for the several tokens of a given utterance to fall into different voicing categories. These categories provide a measure of the AMOUNT of devoicing of each token. The LIKELIHOOD of devoicing was determined for each utterance by counting the number of tokens of the utterance that fell into each of these categories.

4. Results

The results will be presented as follows: first, general trends for likelihood of devoicing, then overall differences between /s/'s and /z/'s. Finally, the differences between /z/'s occurring in different prosodic environments will be discussed.

Likelihood of devoicing was examined by grouping the tokens of /z/'s into the three voicing categories described above. Although all speakers produced both voiced and devoiced tokens of /z/, they varied considerably as to how many of their tokens fell into each of the three voicing categories. Speaker 1 produced the most voiced tokens, while Speaker 2 produced the most devoiced. Figure 2 shows these differences.

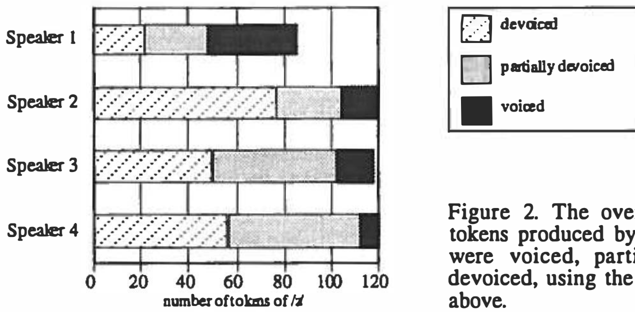

Figure 2. The overall percentage of tokens produced by each speaker that were voiced, partially devoiced or devoiced, using the grouping criterion above.

4.1 Results: differences between /s/'s and /z/'s

In order to compare the production of /z/'s and /s/'s, comparisons of acoustic duration and airflow were made between the pairs of /z/'s and the /s/'s that occurred in matched contexts. For example, the /z/ in "pause" produced by Speaker 3 in the second repetition of the sentence shown in (1) was matched with the /s/ in "boss" in the second repetition of the other sentence shown in (2). Since there was vocal fold vibration during 39% of the /z/ in this repetition of "pause", it was tallied in the "partially devoiced" category, and its matching /s/ is therefore also assigned to this category to provide the basis for comparison. This method ensures that the comparisons between groups of /s/ and /z/ are being made between tokens that occurred in matching contexts. However, the comparisons involve different numbers of tokens, because the different speakers produced different numbers of tokens that were devoiced, partially devoiced or voiced.

Within each voicing category, paired t-tests were used to compare the acoustic and aerodynamic measurements of the /s/ tokens with the measurements of the matching /z/ tokens. These comparisons are particularly important for the devoiced /z/'s: to the extent that the devoiced /z/'s have different durations and aerodynamic characteristics than the /s/'s that occurred in similar contexts, it can be concluded that the devoiced /z/'s are being produced differently from /s/'s. The results of the t-tests (see Table III) show that there were significant differences between /s/'s and /z/'s in the overwhelming majority of the comparisons, regardless of the voicing status of the /z/.

For all sets of comparisons, the acoustic duration of the frication noise was significantly shorter for /z/ than for matched /s/. In the comparisons involving devoiced and partially devoiced /z/'s, the preceding vowels were significantly longer than vowels preceding the matched /s/'s. (Comparisons of vowel duration included only those utterances where the fricative was immediately preceded by a vowel.) Note that comparison among the groups of /z/'s (or among the groups of /s/'s) is not meaningful because the three groups were composed of tokens occurring in different contexts, which gives rise to durational differences due to other factors, such as phrase-final lengthening.

Similar results were obtained for the measurements of airflow; /z/'s were characterized by lower mean and maximum airflow than /s/'s, with greater differences between the devoiced /z/'s and their matched /s/'s than between the voiced /z/'s and their matched /s/'s. (See Smith (1996) for numerical results.) These results indicate clearly that there is no neutralization between the devoiced /z/'s and the /s/'s. Given that the /z/'s seem to be characterized by reduced airflow, compared to their corresponding /s/ tokens, it seems unlikely that the devoicing is being accomplished by opening the glottis as much as for a genuine voiceless consonant like /s/. If it is the case that the devoiced /z/'s do not involve adding a glottal opening, then they have not been strengthened. Rather, the precise conditions for the /z/ have been relaxed, so the devoicing is more aptly viewed as a reduction process, not a strengthening process.

Table III. Results of t-tests comparing acoustic and aerodynamic measures of /s/ and /z/ grouped by voicing category of /z/ tokens. A * indicates that the difference was statistically significant with p<.05. *R* indicates results in the opposite direction of the usual pattern. Unmarked cases were in the predicted direction but did not reach significance.

Speaker	AMOUNT of voicing	Duration of fricative	Duration of prec. vowel	Mean airflow	Maximum airflow
	voiced	*	*	*	*
1	partially devoiced	*	*	*	*
	devoiced	*	*	*	*
	voiced	*		*R*	
2	partially devoiced	*	*	*	*
	devoiced	*	*	*	*
	voiced	*		*	*
3	partially devoiced	*	*	*	*
	devoiced	*	*	*	*
	voiced	*	*R*		*
4	partially devoiced	*	*	*	*
	devoiced	*	*		*

4.2 Results: effect of local context on Likelihood of devoicing

The results of the previous section establish that devoiced /z/'s are distinct from /s/'s: the acoustic durations of the devoiced /z/'s and their preceding vowels are typical of phonologically voiced sounds, and the devoiced /z/'s are produced with significantly less airflow than /s/'s in comparable contexts. Having established that devoicing is not resulting in neutralization, I turn next to the effect of different contexts on the likelihood of devoicing, that is, how frequently a speaker will devoice a /z/ over the course of several productions of /z/ in a given context. For each utterance containing a target /z/, likelihood was calculated for each speaker by counting the number of repetitions of that utterance that were classified in each of the three voicing categories. Based on previous work, it was expected that both local and prosodic context would play a role in determining the likelihood of devoicing.

The effect of local context can be seen in a comparison for a set of utterances all involving the word 'pause', with word-final /z/ followed by different classes of sound: a vowel, a sonorant consonant /l/, a voiced stop /b/ and a voiceless stop /p/. It was expected that the likelihood of devoicing would be greater for /z/ preceding sounds that are lower on the sonority hierarchy (Kenstowicz 1994) compared to /z/ preceding more sonorous sounds. The graph shows how many tokens of each of these utterances were produced as voiced, partially devoiced or devoiced.

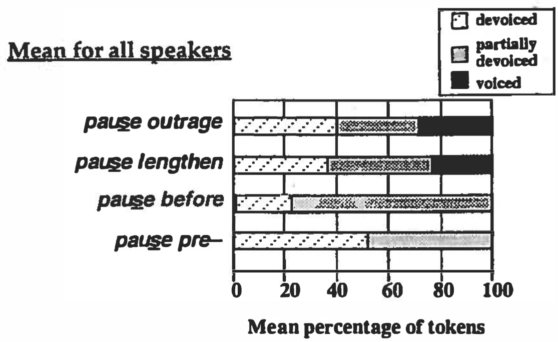

Figure 3. The number of tokens of /z/ in the word "pau**s**e" followed by a vowel, a sonorant consonant /l/, a voiced stop /b/ and a voiceless stop /p/ that the speakers produced as devoiced, partially devoiced, or fully voiced.

These data are averaged across all 4 speakers. When "pause" is followed by a vowel or sonorant consonant, there are many tokens with voicing or partial devoicing, but this is not true when the following sound is a stop. Individual speakers did not vary greatly in the rank order of Likelihood of devoicing among these different utterances. All speakers showed more devoicing before /p/ than before either /b/ or a vowel, and three of the four speakers showed more devoicing before /p/ than before /l/.

The pattern of data in this graph suggest that likelihood of devoicing is substantially influenced by the voicing state of the sound following the fricative, that in these contexts devoicing may be essentially an assimilatory process whereby the glottal gesture for the following consonant (voicing or glottal opening) is being anticipated during the /z/. Anticipation of voicing would mean the vocal folds are approximated during the fricative; anticipation of glottal opening would mean the folds were more separated during the fricative, which might result in higher airflow than during a fully voiced fricative. This possibility is discussed further in section 4.4 below.

4.3 Results: effect of prosodic context on Likelihood of devoicing

Turning from local to larger-scale contextual effects, the set of utterances given in Table II was used to compare /z/'s in different positions in the phrase: sentence-final, word-final not sentence-final, and syllable-final not word-final. All of these /z/'s were followed by /b/. From the graph in Figure 4, it can be seen that devoicing was more common when the fricative was followed by a larger boundary. All speakers consistently devoiced sentence-final /z/. In addition, speakers were more likely to devoice word-final /z/ than word-medial, syllable-final /z/. All speakers produced most of their tokens of word-final /z/ with partial devoicing; no speaker produced any fully voiced tokens of /z/ in this utterance. In contrast, for word-medial /z/, Speakers 1, 2 and 3 produced one or more fully voiced tokens. However, Speaker 4 produced more tokens of devoiced /z/ in word-medial position than in word-final position. Given that all of this speaker's word-medial /z/'s in "husband" were completely devoiced, the most likely interpretation is that she has adopted a pronunciation of "husband" with an /s/ rather than a /z/. Further evidence for this interpretation is that the mean airflow for the /z/ in "hu**s**band" was higher than the airflow for the matching /s/ in "hou**s**ebound." (Comparisons of airflow in different contexts are in section 4.4 below.) Overall, the speakers show a consistent tendency to produce more devoicing before larger prosodic boundaries.

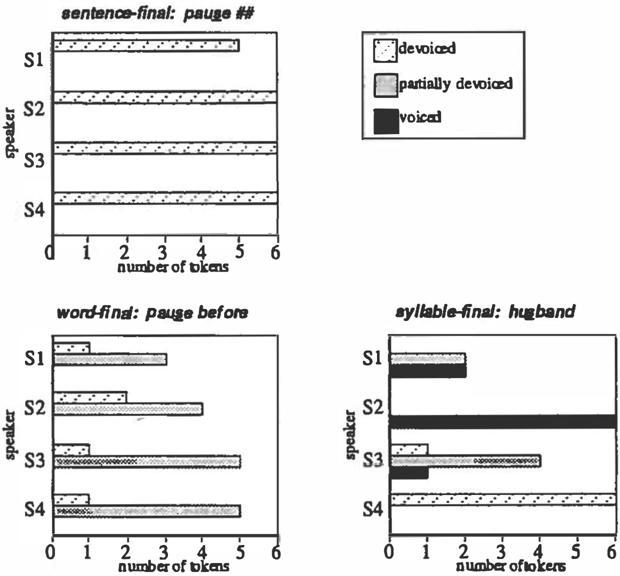

Figure 4. The number of tokens of sentence-, word- and syllable-final /z/ that the speakers produced as devoiced, partially devoiced, or fully voiced.

The effect of stress on the likelihood of devoicing can also be considered a reflection of prosodic organization. Two-syllable words with different stress patterns were compared to see whether the word-final /z/ was more likely to be devoiced at the end of stressed or unstressed syllables. It was expected that devoicing would be more frequent at the end of an unstressed syllable than at the end of an stressed syllable, since the lack of stress signals a prosodically weaker position. Speakers 1 and 3 fulfilled this prediction, as can be seen in Figure 5. Speaker 1 produced the stressed /z/ in "recúrs" either fully voiced or partially devoiced, and the unstressed /z/ in "Dódgers" as either partly or completely devoiced. Speaker 3 produced most tokens of stressed /z/ with partial devoicing, but always produced the unstressed /z/ as fully devoiced. Unlike these two speakers, Speaker 2 made no difference in the likelihood of voicing stressed and unstressed /z/. Only Speaker 4 showed a pattern contrary to the prediction: she produced equal numbers of partly and fully devoiced tokens of stressed /z/, but for unstressed /z/ she produced more partly devoiced tokens. Averaging across the 4 speakers there were 23% fewer tokens with partial or full voicing in the unstressed /z/ than in the stressed /z/, providing at least partial support for the hypothesis that devoicing is more likely in the weaker unstressed syllable.

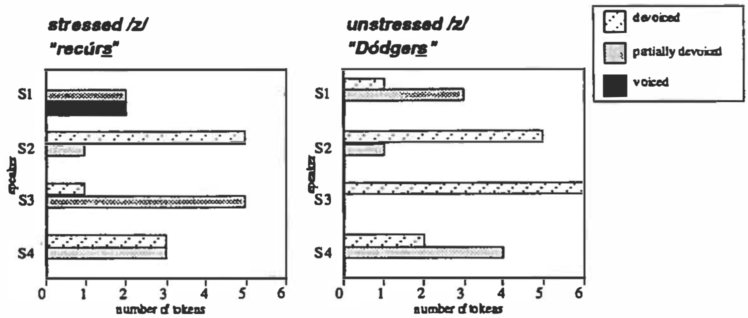

Figure 5. The number of tokens of /z/ at the end of a stressed or unstressed syllable that the speakers produced as devoiced, partially devoiced, or fully voiced.

The results presented so far suggest that devoicing is favored in prosodically "weaker" environments. The comparison of airflow for /s/ and /z/ in section 4.1 showed that in general devoiced /z/ is being produced with lower airflow than for voiceless /s/, suggesting that the cause of devoicing is not likely to be glottal opening (which is responsible for the voicelessness of /s/). Instead, the loss of voicing may be due to the aerodynamic conditions in the vocal tract diverging from the particular state that is necessary to produce simultaneous voicing and frication, possibly due to subglottal pressure falling below the level necessary for the maintenance of voicing. Lowered subglottal pressure would be consistent with the low airflow observed in the devoiced /z/'s. However, this explanation is not necessarily valid for the occurrences of devoicing that were hypothesized to result from assimilation of glottal position during the fricative to an adjacent voiceless context. In these contexts the glottis might be more open than during voicing, with a concomitant increase in airflow. Further comparison of airflow during the fricatives produced in different contexts could reveal whether devoicing is always accompanied by low airflow (in which case the explanation of reduced subglottal pressure seems most appropriate), or whether in some contexts devoicing may be accompanied by high airflow (in which case the glottis may be opening).

4.4 Results: comparison of airflow in different prosodic positions

To investigate further whether devoicing can generally be attributed to reduced subglottal pressure, or possibly lower aerodynamic effort directed at maintaining this pressure, comparisons were made of the airflow values for the utterances in different prosodic contexts grouped by phrasal context (syllable-final, word-final or sentence-final) rather than by voicing category. Thus this set of comparisons tests the effect of context on airflow, rather than the effect of different amounts of voicing.

The patterns for maximum flow and mean flow were similar; the data for maximum flow are given in Figure 6. For Speaker 1, the flow is lower for /z/ than for /s/ in each of the three contexts. Since all the sentence-final /z/'s were completely devoiced, but some of the syllable-final ones were fully voiced, it can be concluded that for Speaker 1, regardless of the amount of voicing in the /z/, the airflow for /z/'s is lower than that of /s/'s produced in the same context. For this speaker, it appears that devoicing is not the result of increased airflow through the glottis.

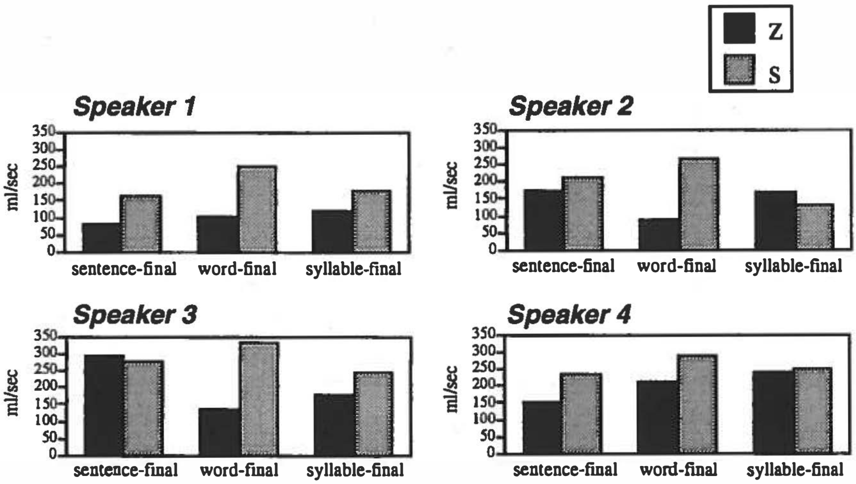

Figure 6. Maximum airflow during /z/ and /s/ in different prosodic positions: sentence-final "pau<u>s</u>e", word-final "pau<u>s</u>e before" and syllable-final "hu<u>s</u>band".

The other three speakers each show lower airflow for /z/ than for /s/ in two of the three contexts. Thus Speaker 3 shows lower airflow for /z/ than for /s/ in word- and syllable-final position, as expected, but surprisingly shows relatively high flow for both /z/ and /s/ in sentence-final position. This high air flow suggests that the speaker may be neutralizing the /z/-/s/ contrast in this one position. It may be that this speaker is opening the vocal folds at the end of the sentence in anticipation of the open position of the glottis that is typical of respiration during a pause after a sentence. The high airflow in the sentence-final position appears to be a special case of an assimilatory process, different from the examples of assimilation to adjacent sounds.

Speaker 4 shows a different exception to the general pattern in the flow data than does Speaker 3. For Speaker 4, airflow is lower, as expected, in the sentence-final and word-final /z/'s than in the corresponding /s/'s, but the airflow in the syllable-final /z/'s is almost as high as in the syllable-final /s/'s. Speaker 4 always completely devoiced these syllable-final /z/'s in the word 'hu<u>s</u>band'. The high airflow for these devoiced /z/'s suggest that, as noted in section 4.3, this speaker may have a different lexical form for this word, so that it contains an /s/ rather than the /z/ that might be expected. Such lexical variation is found in English in other words, such as the two pronunciations 'ab[s]urd' and 'ab[z]urd'. The airflow data for Speaker 2 were similar to that of Speaker 4, but the high airflow for syllable-final /z/ is difficult to explain for Speaker 2 since her /z/'s were fully voiced in this utterance.

5. Two mechanisms for devoicing

Two separate influences on the likelihood of devoicing of /z/ have been identified here: one that can be viewed as assimilation to a local context that lacks voicing, and one that can be ascribed to the position of the /z/ in phrasal or prosodic structure. The results demonstrate that while much of the variability in the likelihood of devoicing can be accounted for by these influences combined with speakers' overall differences for likelihood of devoicing, devoicing in these /z/'s is nonetheless a process best described in probabilistic terms – more or less likely rather than possible or impossible.

As discussed in the introduction, voiced fricatives present a particularly exigent set of demands on the vocal tract. Because they require precise conditions for successful production, it may be that a comparatively small divergence from these conditions is more likely to result in a salient difference from the "default" characteristics of a voiced fricative than would be the case for some other sound. In the present data, divergence from the canonical form of /z/ always showed up as devoicing, rather than loss of frication. The tendency to devoice can be explained in part by evidence suggesting that the glottis is always somewhat open during voiced fricatives – more open, at least, than for voiced stops. Such evidence comes from studies using transillumination to examine glottal opening (Lisker, Abramson, Cooper & Schvey 1969), as well as EMG data showing more suppression of the adductory interarytenoid and lateral cricoarytenoid muscles in word-medial voiced fricatives than voiced stops (Hirose & Ushijima 1978).

If the usual state of the glottis for voiced fricatives is somewhat open, maintaining sufficient subglottal pressure for vocal fold vibration will require greater airflow from the lungs than for voiced sounds produced with a more closed glottis, so vocal fold vibration may fail more often. In addition, just a small additional opening of the glottis could lead to devoicing. Laver (1994) argues that during devoiced sounds such as /z/ the glottis is probably in a state intermediate between voicing and voicelessness, like the state of the glottis that is used in whisper, with the glottis open but the folds very close together.

Based on the results of the experiment reported here, there appear to be two possible routes to devoicing: in one scenario the glottis opens more, increasing trans-glottal airflow and equalizing pressure above and below the glottis so that the vocal folds no longer vibrate. Alternatively, the configuration of the glottis remains the same, but sub-glottal pressure nonetheless becomes insufficient to maintain vibration of the folds. The second scenario is plausible in part because muscular action is necessary to maintain adequate subglottal pressure for speech (Ohala 1990). Thus if a speaker does not use enough effort in the abdominal muscles, subglottal pressure will fall below the level necessary to keep the vocal folds vibrating. This second scenario relies, therefore, on a reduction of effort by the speaker, whereas the first scenario (favored by Haggard 1978), relies on extra effort leading to additional opening of the glottis.

The second scenario seems appropriate to explain cases where the voicing for /z/ is lost because of reduced effort, such as seems typical of the prosodic positions where devoicing was found to be most common. The /z/'s in utterance-final or word-final position, or unstressed syllable codas, are in positions where articulator movement is often reduced (e.g. Manuel 1991; Krakow 1993, Byrd 1994, Keating 1995, Browman & Goldstein 1995). These considerations suggest that devoicing as lenition is a passive process, where voicing ceases because nothing active is being done to maintain it. This might be because of insufficient subglottal pressure compared to the oral pressure, which would result if the airflow from the lungs was lower. Even if the airflow were reduced, it couldn't fade away entirely or there would be no frication noise. So a lenited, devoiced /z/ should have lower airflow than an /s/, but still enough airflow to maintain the frication.

In contrast, the /z/'s that are devoiced because they are adjacent to a voiceless sound or pause may require a different explanation. For these, the first scenario, which proposes possible increased glottal opening as the mechanism for devoicing, seems more appropriate. The glottis "assimilates" to the adjoining sound's requirement for an open glottis. These /z/'s seem to be more malleable than the neighboring sounds to which they assimilate. As with the /z/'s whose devoicing is ascribed to their weak position in prosodic structure, this malleability may manifest itself because a small change to the articulatory position or aerodynamic conditions of a /z/ is more likely to have a discernible effect than a correspondingly small change to some other more robust articulation. Even if the "assimilated" /z/'s are devoiced because the glottis is more open, the additional opening cannot be very great because the airflow for these /z/'s is still low compared to /s/.

The results presented here show much variability in likelihood of devoicing, but a large part can be attributed to differences in context. Tolerance for such variability may be related to the constraints on the speaker, as well as the interaction between speaker and listener (Lindblom 1990). The pattern of reduction in weak prosodic environments suggests that in these environments there is less need for the speaker to produce maximal distinctions between /z/ and /s/. Indeed, lenition of /z/ may be a marker of prosodic structure, in that it occurs more often adjacent to stronger prosodic boundaries and at least for some speakers, in less salient unstressed syllables. Processes such as devoicing could be a way of making the prosodic boundaries more salient to the listener. Thus, devoicing provides an example of a gradient process that reflects phonological organization.

References

Browman, C. and L. Goldstein. 1995. Gestural syllable position effects in American English. In *Producing speech: contemporary issues: for Katherine Safford Harris*, ed. F. Bell-Berti and L. Raphael, 19-33. Woodbury, NY: AIP Press.

Byrd, D. 1994. Articulatory timing in English consonant sequences. Doctoral dissertation, University of California, Los Angeles. In *UCLA Working Papers in Phonetics* 86.

Cooper, A. 1991. Glottal gestures and aspiration in English. Doctoral dissertation, Yale University, New Haven, Connecticut.

Docherty, G.J. 1992. *The timing of voicing in British English obstruents. Netherlands Phonetics Archives* 9. Berlin: Foris.

Fougeron, C. and P. A. Keating. 1996. Articulatory strengthening in prosodic domain-initial position. *UCLA Working Papers in Phonetics* 92: 61-95.

Haggard, M. 1978. The devoicing of voiced fricatives. *Journal of Phonetics* 6: 95-102.

Hirose, H. and T. Ushijima. 1978. Laryngeal control for voicing distinction in Japanese consonant production. *Phonetica* 35: 1-10.

Keating, P. A. 1995. Effects of prosodic position on /t,d/ tongue/palate contact. In *Proceedings of the XIIIth international congress of phonetic sciences* 3, ed. K. Elenius and P. Branderud, 432-435. Stockholm: Stockholm University.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.

Krakow, R. A. 1993. Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate. In *Nasals, nasalization, and the velum*, ed. M. K. Huffman and R. A. Krakow, 87-116. San Diego: Academic Press.

Krakow, R. A., F. Bell-Berti and Q. Wang. 1995. Supralaryngeal declination: evidence from the velum. In *Producing speech: contemporary issues: for Katherine Safford Harris*, ed. F. Bell-Berti and L. Raphael, 333-353. Woodbury, NY: AIP Press.

Lass, R. 1984. *Phonology*. Cambridge: Cambridge University Press.

Laver, J. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H & H theory. In *Speech production and speech modelling*, ed. W. J. Hardcastle and A. Marchal, 403-439. Dordrecht: Kluwer.

Lisker, L., A. Abramson, F. Cooper, and M. Schvey. 1969. Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* 45: 1544-1546.

Manuel, S. 1991. Some phonetic bases for the relative malleability of syllable-final versus syllable-initial consonants. In *Actes du XIIème congrès international des sciences phonétiques* 5, 118-121. Aix-en-Provence: Université de Provence.

Ohala, J. J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, ed. P. MacNeilage, 189-216. New York: Springer-Verlag.

Ohala, J. J. 1990. Respiratory activity in speech. In *Speech production and speech modelling*, ed.W. J. Hardcastle and A. Marchal, 23-53. Dordrecht: Kluwer.

Pierrehumbert, J. and D. Talkin. 1992. Lenition of /h/ and glottal stop. In *Papers in laboratory phonology II: gesture, segment, prosody*, ed. G. J. Docherty and D. R. Ladd, 90-117. Cambridge: Cambridge University Press.

Smith, C. 1996. The devoicing of /z/ in American English: effects of local and prosodic context. *UCLA Working Papers in Phonetics* 92: 125-150.

Stevens, K., S. Blumstein, S. Glicksman, M. Burton, and K. Kurowski. 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America* 91: 2979-3000.

Veatch, T. 1989. Word-final devoicing of fricatives in English. Paper presented at the Linguistic Society of America meeting, Washington, DC.

Eloquent Technology, Inc.
2389 Triphammer Road
Ithaca, NY 14850

caroline@eloq.com

