

Dynamic hyperarticulation of coda voicing contrasts

Scott Seyfarth^{a)}

Department of Linguistics, University of California, San Diego, 9500 Gilman Drive,
La Jolla, California 92093-0108, USA
sseyfarth@ucsd.edu

Esteban Buz and T. Florian Jaeger^{b)}

Department of Brain and Cognitive Sciences, University of Rochester, Meliora Hall,
RC 270268, Rochester, New York 14627, USA
ebuz@bcs.rochester.edu, fjaeger@bcs.rochester.edu

Abstract: This study investigates the capacity for targeted hyperarticulation of contextually-relevant contrasts. Participants communicated target words with final /s/ or /z/ when a voicing minimal-pair (e.g., target *dose*, minimal-pair *doze*) either was or was not available as an alternative in the context. The results indicate that talkers enhance the durational cues associated with the word-final voicing contrast based on whether the context requires it, and that this can involve both elongation as well as shortening, depending on what enhances the contextually-relevant contrast. This suggests that talkers are capable of targeted, context-sensitive temporal enhancements.

© 2016 Acoustical Society of America

[AL]

Date Received: July 21, 2015 **Date Accepted:** February 10, 2016

1. Introduction

If a spoken word is likely to be misunderstood, a talker may enhance it to make it easier to identify. What strategies do speakers use, and what kinds of enhancements are available to them? Here, we study whether and how talkers make enhancements that target contextually-relevant contrasts.

In particular, we ask two questions. First, if an intended word (e.g., *dose*) is likely to be misunderstood in a particular speech context (e.g., because its minimal-pair *doze* is another word available in the context), do talkers selectively enhance those aspects of the signal that increase the contrast between the two words? Second, how are these enhancements realized phonetically? For example, can talkers dynamically elongate and shorten parts of the signal in order to enhance contrasts? Or, does targeted hyperarticulation involve only proportional elongation, as is typical of more global hyperarticulation (Smiljanić and Bradlow, 2008, 2009)? These questions inform broader debates about whether and how communicative goals influence articulation (Lindblom, 1990; Jaeger, 2013), and whether targeted hyperarticulation differs from more general modes of clear speech (Ohala, 1994).

1.1 Targeted enhancement of contextually-relevant contrasts

Previous research has shown that when only part of an utterance is misunderstood, talkers focus their hyperarticulation on the misunderstood word (Oviatt *et al.*, 1998; Stent *et al.*, 2008). Further, if only a single segment has been misunderstood (or is likely to be misunderstood), talkers may limit their enhancements to that segment. When a speaker needs to be clear that the intended word is aspirated *pat* and not unaspirated *bat*, they lengthen only the aspiration of the /p/ (Baese-Berk and Goldrick, 2009; Kirov and Wilson, 2012; Schertz, 2013; Buz *et al.*, 2014). Thus, it has been argued that talkers selectively enhance parts of the signal, based on the context and the needs of the communicative situation (Jaeger, 2013; Schertz, 2013; Buz *et al.*, 2014).

^{a)}Author to whom correspondence should be addressed.

^{b)}Also with Department of Computer Science and Department of Linguistics, University of Rochester, Meliora Hall, RC 270268, Rochester, NY 14627, USA.

Yet not all studies have found explicit support for targeted enhancements. While there is strong evidence for context-sensitive elongation of word-initial aspiration and prevoicing (see references above), and of vowel length contrasts (de Jong, 2004; Schertz, 2013), findings have been less straightforward for coda voicing contrasts. For example, de Jong (2004) finds that the difference between voiced and voiceless codas increases (as cued by vowel duration) when talkers put focus on the voicing contrast. However, in the same study, the voicing contrast is enhanced even more when talkers put focus on a vowel quality contrast instead. One interpretation of this result is that talkers are limited in their ability to dynamically adjust temporal relations in the rime to enhance relevant contrasts (see de Jong, 2001). Alternatively, when talkers focused on the coda voicing contrast, they may have been primarily targeting different cues for enhancement, as opposed to vowel duration (see Stevens *et al.*, 1992). This could explain why the vowel-duration effect was unexpectedly small during enhancement of the coda voicing contrast.

Thus, in the current study, we investigate enhancement of coda voicing contrasts. We measure a number of cues that may be targeted for enhancement. In particular, we evaluate temporal enhancement of the English word-final /s, z/ contrast (e.g., *dose* versus *doze*), which potentially involves at least three differences in the rime. Voiceless /s/ is longer than voiced /z/, but the nucleus vowel before coda /s/ is much shorter than before coda /z/. In addition, although voiced obstruents are typically partially or fully devoiced in this position, voicing may be maintained longer into /z/ than into /s/ (Derr and Massaro, 1980; Smith, 1997; see also Stevens *et al.*, 1992). Besides these temporal differences, spectral differences between /s/ and /z/ are illustrated in Maniwa *et al.* (2009), and spectral enhancement of sibilants is investigated by Silbert and de Jong (2008), Maniwa *et al.* (2009), Julien and Munson (2012), and Clayards and Knowles (2015).

1.2 Realization of enhancements

There are several ways that talkers might enhance final /s/ and /z/ to increase the contrast between the two. What strategies do they actually use for targeted enhancements? For global temporal enhancements—clear speech that is not targeted to a particular lexical competitor—talkers elongate segments in proportion to their durations in unenhanced, conversational speech, as well as inserting more and longer pauses (de Jong, 2001; Smiljanić and Bradlow, 2008, 2009). For targeted temporal enhancements, previous work has found mainly elongation of particular segments or acoustic cues (Oviatt *et al.*, 1998; Silbert and de Jong, 2008; Baese-Berk and Goldrick, 2009; Julien and Munson, 2012; Schertz, 2013).

This might suggest that talkers generally slow their speech rate when making enhancements, but just do so over a longer or shorter stretch of speech, depending on how much is relevant to the context. For example, Kirov and Wilson (2012) find that English /p/ aspiration is lengthened not only when contrasting target *peak* with *beak*—where longer aspiration is a direct cue to the difference—but also when contrasting target *peak* with *teak*, where longer aspiration is not a direct cue to the difference. While it is plausible that lengthening improves other cues to the place contrast (see also Kirov and Wilson, 2012), or that lengthening is a side-effect of a more direct enhancement of the place contrast, it remains possible that general elongation of a crucial segment is the default strategy for enhancing a contrast. Under this account, talkers would be predicted to elongate both coda /s/ and /z/ in our experiment when the coda contrast is contextually-relevant.

Alternatively, talkers might be able to make more dynamic temporal adjustments, beyond elongation of the intended segment. For example, they might shorten the short vowel before /s/ and lengthen the long /s/ coda in *dose*, but lengthen the vowel before /z/ and shorten the coda in *doze*. In this way, they would adjust the overall duration of voicing in the signal when the voicing contrast is relevant (Keyser and Stevens, 2006).

1.3 The current study

We asked experimental participants to produce /s, z/-final words in two contexts. In one context, participants had to be sure that a listener would not confuse a target word with its voicing minimal-pair (e.g., target *dose* must be differentiated from *doze*, or vice-versa). In the other context, there was no potential to confuse the target word with its voicing minimal-pair.

We measured vowel duration, coda duration, and the duration of voicing in the coda. Because coda voicelessness is cued by a short vowel but a long coda (and

vice versa for voicedness; Derr and Massaro, 1980), these stimuli allowed us to evaluate whether talkers invariably use elongation to make words easier to identify, or whether they might use a different strategy to enhance a contrast where across-the-board elongation would be less useful.

2. Methods

Forty participants were recruited through Amazon Mechanical Turk to play an online communication game with a partner. A browser-based Flash application captured audio at 22.05 kHz with 16-bit depth from participants' microphones. All participants reported being native English speakers.

The task was modeled after Baese-Berk and Goldrick (2009), using the software and paradigm developed in Buz et al. (2014). On each trial, a participant would see three words on their screen. After 1.5 s, one of the three words was highlighted by the computer. The participant was asked to verbally produce this word (the target) for a partner, who could also see the three words but did not know which of the three was the target. They were told that their partner would listen to the word that they said, and try to choose it from the three possibilities. The partner was simulated by the computer, and always chose the target word.

The critical targets were /s/- or /z/-final words. In *contrastive* trials, one of the two alternative words was the voicing-final minimal pair of the target. In *control* trials, neither of the two alternatives was the voicing-final minimal pair. For example, when the target was /s/-final *dose*, the control and contrastive conditions had the following words (with the target word underlined):

<i>Control</i>	bade	<u>dose</u>	maul
<i>Contrastive</i>	doze	<u>dose</u>	maul

When the target was /z/-final *doze*, the control and contrastive conditions had these words:

<i>Control</i>	mode	<u>doze</u>	maul
<i>Contrastive</i>	dose	<u>doze</u>	maul

Thus, in the contrastive condition, participants were aware that their partner must identify whether the target is voiceless-final *dose* or voiced-final *doze*. Based on previous work using this method (Baese-Berk and Goldrick, 2009; Kirov and Wilson, 2012; Buz et al., 2014), we expected that they would target the voicing contrast for enhancement. On the other hand, in the control condition, while participants may speak clearly, the coda voicing contrast does not need to be enhanced.

2.1 Items

The critical targets were 18 word-final /s, z/ minimal pairs, listed in Table 1. Half of the participants saw only the /s/-final words, and half saw only the /z/-final words. Matched pairs were chosen so that, with the exception of any contrastive enhancements, the vowel and coda durations would be as similar as possible.

Participants saw each target word in only one trial. Each participant saw half of their targets in a contrastive trial, and half in a control trial. In addition to the critical trials, participants saw 39 filler trials. In nine of these trials, the filler target was a minimal pair with one of the other words in the trial, to avoid drawing special attention to the /s, z/ voicing contrast. Fillers and critical trials were presented according to a pseudo-randomized list, with half of the participants seeing the list in backwards order (following Buz et al., 2014). Target presentation was balanced so that the highlighted targets appeared roughly equally often in all three positions on the screen (left, center, or right).

Table 1. Critical target words.

/s/	base	bus	cease	deuce	dice	dose	face	fuss	goose
/z/	bays	buzz	seize	dues	dies	doze	phase	fuzz	goos
/s/	hiss	mace	moose	niece	noose	pace	piece	spice	vice
/z/	his	maze	moos	knees	news	pays	peas	spies	vies

3. Analysis

Because participants recorded themselves with their own laptop or peripheral microphone, the recording quality was variable. Of the 40 included participants, 20 used a built-in laptop or desktop microphone, 15 used a head-mounted microphone, 4 used a peripheral desk-mounted microphone, and 1 participant declined to provide information. Because the experimental condition (control versus contrastive) was manipulated within-participant, uneven recording quality across participants could not have biased the results.

Prior to annotation, 7 participants (12.5%) were excluded because of a technical issue (such as failing to upload audio), excessively noisy recordings, or failing to follow directions. Nine additional participants (16.1%) were excluded after they wrote on a debriefing questionnaire that they suspected their partner was simulated. Discussion of the questionnaire, and of the believability of the simulated partner in this paradigm, is provided in [Buz *et al.* \(2014\)](#).

The experiment was run until reaching 20 included /s/ participants and 20 included /z/ participants, totaling 720 productions of critical targets (40 subjects * 18 items). All exclusion criteria were identical to [Buz *et al.* \(2014\)](#).

3.1 Annotation

Four annotators marked the vowel and coda segment boundaries. Annotators were native to the trial condition. Vowel onsets were marked at the onset of periodicity, or at the onset of dark formant bands if the preceding segment was voiced. Vowel offsets were marked at the onset of sibilant noise in the range above 3500 Hz. Coda segments were marked from the onset to the offset of sibilant noise above 3500 Hz immediately following the vowel.

To assess inter-annotator agreement, all annotators segmented productions from a test set of two /s/ participants and two /z/ participants. Pearson's r was calculated between each pair of annotators within each participant. For vowel durations, the mean pairwise r values were 0.87, 0.95, 0.99, and 0.99 for the 4 test participants. For coda durations, the mean pairwise r values were 0.65, 0.72, 0.89, and 0.96. Because of the lower agreement rates for coda durations, the coda segment results should be interpreted cautiously, as the noisier annotations inflate the Type II error rate.

Thirteen productions (1.8%) were removed from all analyses because of heavy audio clipping or another recording issue that prevented segmentation (such as a loud background noise at a segment boundary), or because the word was cut off in the recording. Twelve productions (1.7%) were removed from all analyses because the participant said the wrong word, no word, or more than one word. Five more productions (0.1%) were excluded from the vowel analyses because the vowel duration was more than 2.5 standard deviations from the participant's mean. Eleven productions (1.5%) were excluded from the codas analyses because the coda duration was more than 2.5 standard deviations from the participant's mean.

After annotating the vowel and coda boundaries, Praat was used to count the total number of voiced 10 ms frames in each coda fricative ([Boersma, 1993](#); [Boersma and Weenink, 2014](#)). Figure 1 shows vowel durations, coda durations, and coda voicing proportions for the /s/- and /z/-final target words.

3.2 Models and results

Vowel durations, coda durations, and coda voicing durations were analyzed in separate linear mixed-effects models. Fixed effects in all three models were phonological

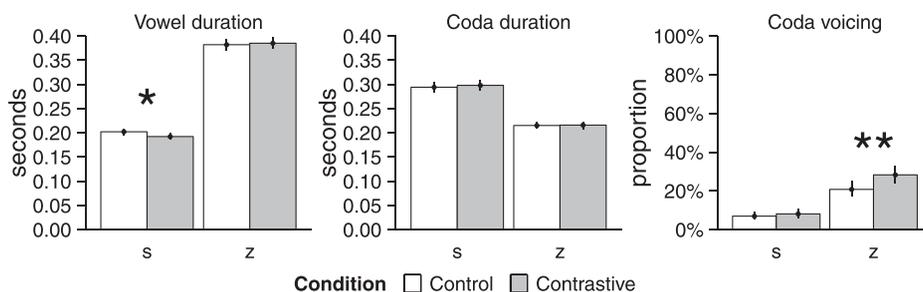


Fig. 1. Mean segment durations and coda voicing proportions, by experimental condition. Error bars show bootstrapped 95% confidence intervals of the means.

coda voicing (voiceless or voiced) and critical trial type (control or contrastive), plus the interaction. Models also included by-participant intercepts and slopes for critical trial type, and by-item intercepts and slopes for all three fixed effects. For the random groupings, a minimal pair (e.g., *dose* and *doze*) was treated as a single item, but all results were the same if /s/ and /z/ stimuli were modeled separately (without parameters for phonological voicing), or if they were modeled together but treated as different items. *p*-values were calculated using the Satterthwaite approximation for degrees of freedom. Results were qualitatively the same with and without log-transformation. Segment duration models were planned analyses; the coda voicing duration analysis was added *post hoc* in response to the low agreement rates for the coda offsets.

Vowel durations. Vowels were significantly shorter overall in /s/-final words compared to /z/-final words ($\hat{\beta} = -187$ ms, $t = 11.5$, $p < 0.0001$). Crucially, for /s/-final words, vowels were also significantly shorter in the contrastive condition, where the target word's voicing-final minimal pair was present, compared to the control condition ($\hat{\beta} = -9$ ms, $t = 2.3$, $p < 0.05$). For /z/-final words, vowel durations were not significantly different between the control and contrastive conditions ($p > 0.7$).

Coda durations. Coda /s/ was significantly longer overall than coda /z/ ($\hat{\beta} = 79$ ms, $t = 4.9$, $p < 0.0001$). There was no significant difference in coda durations between conditions for either /s/-final words ($p > 0.2$) or /z/-final words ($p > 0.9$).

Coda voicing durations. Coda voicing was maintained significantly longer in /z/ than in /s/ ($\hat{\beta} = 2.6$ frames, $t = 2.8$, $p < 0.01$). Crucially, for /z/ words, voicing was also maintained significantly longer in the contrastive condition compared to the control condition ($\hat{\beta} = 1.6$ frames, $t = 4.0$, $p < 0.001$; also significant after family-wise error correction). This model was fit without by-item slopes for phonological voicing, since it did not converge with slopes. For /s/-final words, coda voicing durations were not significantly different between the control and contrastive conditions ($p > 0.7$).

4. Discussion

Talkers produced relatively shorter vowels before voiceless /s/ and maintained voicing longer into voiced /z/ when the voicing contrast was contextually-relevant. Our first question was whether talkers selectively enhance aspects of the signal that increase a relevant contrast. Both of the timing changes that we found increase the contrast between the target word and its voicing-final minimal pair. This indicates that talkers make selective, context-specific enhancements in our contrastive condition, when targeting the coda voicing contrast. This extends similar findings on other contrasts, such as voice-onset time in word-initial plosives, the tense-lax English vowel distinction, and some spectral measures that distinguish fricatives (Maniwa *et al.*, 2009; Kirov and Wilson, 2012; Schertz, 2013; Buz *et al.*, 2014; Clayards and Knowles, 2015). Our second question was whether targeted hyperarticulation invariably uses the elongation processes that are typical of more global hyperarticulation. The results suggest that this is not the case: talkers are capable of dynamic temporal enhancements in particular, contexts where across-the-board or proportional elongation of a word or segment may be less helpful.

Both effect sizes are comparable to the enhancement of prevoicing and aspiration durations in tasks where talkers explicitly clarify and repeat words that have been misidentified by a listener (Schertz, 2013). This suggests that the effects observed here were plausibly intended to improve lexical identification. Buz *et al.* (2016) argue that such contrastive enhancement needs to be understood with regard to the avoidance of perceptually-ambiguous productions near phonetic category boundaries, where even smaller durational changes are known to affect comprehension (McMurray *et al.*, 2002). It is a question of ongoing research as to what types of enhancements serve to facilitate comprehension (Uchanski, 2008; Smiljanić and Bradlow, 2009).

One view of the two effects is that talkers may use different strategies to enhance coda voicelessness (vowel shortening) versus voicedness (longer phonetic voicing). Alternatively, it may be that the enhancement target is the overall voicing duration or the relative timing of the voicing offset within the word, rather than the duration of the vowel or coda individually (Keyser and Stevens, 2006; Choi *et al.*, 2015; see Massaro and Cohen, 1977; Stevens *et al.*, 1992 on perception).¹ This would help explain why previous work that investigates enhancement of the coda-voicing contrast has not found the expected effect on vowel duration alone. For example, Goldrick *et al.* (2013) found no lexically-mediated enhancement of vowel duration (after controlling for onset aspiration), de Jong (2004) found that the vowel duration contrast between voiced- and voiceless-final words was enhanced more weakly in the context of voicing-relevant competitors (contrasting *bat*–*bad*, *bet*–*bed*) than voicing-irrelevant ones

(contrasting *bat–bet*, *bed–bad*), and Choi *et al.* (2015) found no vowel duration enhancement in the context of voicing-relevant competitors.

More generally, our findings provide evidence that selective, context-specific enhancement is not limited to onsets, and point to ways in which talkers may use different phonetic strategies for targeted enhancements as compared to more global hyperarticulation.

Acknowledgments

We thank Anna MacDonald, Aishwarya Krishnamoorthy, Brian Leonard, and Lindsey Harris for help with segmentation, and Meghan Clayards, Marc Garellek, Matthew Goldrick, the phonetics and phonology group at UCSD, and audiences at the University of Geneva and CUNY 2015 for helpful discussion and comments. This work was supported by an NSF Graduate Research Fellowship to S.S., an NRSA Predoctoral Fellowship (No. F31HD083020) to E.B., and an NSF CAREER (No. IIS-1150028) to T.F.J. The views expressed here do not necessarily reflect those of the funding agencies.

References and links

- ¹To evaluate whether talkers initiate devoicing earlier before voiceless codas when they are contrasted with voiced ones (cf. Clayards and Knowles, 2015, on prominence effects), we conducted a *post hoc* analysis of voicing duration during the vowel. Voicing duration was shorter before /s/ in the contrastive condition relative to the control condition ($\beta = -9$ ms, $p < 0.05$). However, most vowel productions in our data were fully voiced, and the duration of vowel voicing was highly correlated with the duration of the vowel ($r = 0.96$), making it impossible to distinguish this effect from the effect on vowel duration reported in the text.
- Baese-Berk, M., and Goldrick, M. (2009). “Mechanisms of interaction in speech production,” *Lang. Cognit. Process.* **24**(4), 527–554.
- Boersma, P. (1993). “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,” in *Proceedings of the Institute of Phonetic Sciences*, University of Amsterdam, Vol. 17, pp. 97–110. Available online at http://www.fon.hum.uva.nl/Proceedings/Proc_17_Postscript/Proc_Boersma_1993.ps (Last viewed July 10, 2015).
- Boersma, P., and Weenink, D. (2014). “Praat: Doing phonetics by computer,” available online at <http://www.fon.hum.uva.nl/praat/> (Last viewed July 10, 2015).
- Buz, E., Jaeger, T. F., and Tanenhaus, M. K. (2014). “Contextual confusability leads to targeted hyper-articulation,” in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.
- Buz, E., Tanenhaus, M. K., and Jaeger, T. F. (2016). “Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers’ subsequent pronunciations,” *J. Memory Lang.*
- Choi, J., Cho, T., Kim, S., Baek, Y., and Jang, J. (2015). “Phonetic encoding of coda voicing contrast and its interaction with information structure in L1 and L2 speech,” in *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Clayards, M., and Knowles, T. (2015). “Prominence enhances voicelessness and not place distinction in English voiced sibilants,” in *Proceedings of the 18th International Congress of Phonetic Sciences*.
- de Jong, K. (2004). “Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration,” *J. Phonetics* **32**(4), 493–516.
- de Jong, K. J. (2001). “Effects of syllable affiliation and consonant voicing on temporal adjustment in a repetitive speech-production task,” *J. Speech, Lang., Hear. Res.* **44**(4), 826–840.
- Derr, M. A., and Massaro, D. W. (1980). “The contribution of vowel duration, f0 contour, and frication duration as cues to the /juz/-/jus/ distinction,” *Percept. Psychophys.* **27**(1), 51–59.
- Goldrick, M., Vaughn, C., and Murphy, A. (2013). “The effects of lexical neighbors on stop consonant articulation,” *J. Acoust. Soc. Am.* **134**(2), EL172–EL177.
- Jaeger, T. F. (2013). “Production preferences cannot be understood without reference to communication,” *Front. Psychol.* **4**, 1–4.
- Julien, H. M., and Munson, B. (2012). “Modifying speech to children based on their perceived phonetic accuracy,” *J. Speech, Lang., Hear. Res.* **55**(6), 1836–1849.
- Keyser, S. J., and Stevens, K. N. (2006). “Enhancement and overlap in the speech chain,” *Lang.* **82**(1), 33–63.
- Kirov, C., and Wilson, C. (2012). “The specificity of online variation in speech production,” in *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, Sapporo, Japan.
- Lindblom, B. (1990). “Explaining phonetic variation: A sketch of the H&H theory,” in *Speech Production and Speech Modeling*, edited by W. Hardcastle and A. Marchal (Springer, New York), pp. 403–439.
- Maniwa, K., Jongman, A., and Wade, T. (2009). “Acoustic characteristics of clearly spoken English fricatives,” *J. Acoust. Soc. Am.* **125**(6), 3962–3973.
- Massaro, D. W., and Cohen, M. M. (1977). “Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction,” *Percept. Psychophys.* **22**(4), 373–382.
- McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2002). “Gradient effects of within-category phonetic variation on lexical access,” *Cognition* **86**, B33–B42.
- Ohala, J. J. (1994). “Acoustic study of clear speech: A test of the contrastive hypothesis,” in *Proceedings of the International Symposium on Prosody*, Yokohama, Japan, pp. 75–89.

- Oviatt, S., Levow, G.-A., Moreton, E., and MacEachern, M. (1998). "Modeling global and focal hyperarticulation during human-computer error resolution," *J. Acoust. Soc. Am.* **104**(5), 3080–3098.
- Schertz, J. (2013). "Exaggeration of featural contrasts in clarifications of misheard speech in English," *J. Phonetics* **41**(3–4), 249–263.
- Silbert, N., and de Jong, K. (2008). "Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production," *J. Acoust. Soc. Am.* **123**(5), 2769–2779.
- Smiljanić, R., and Bradlow, A. R. (2008). "Stability of temporal contrasts across speaking styles in English and Croatian," *J. Phonetics* **36**(1), 91–113.
- Smiljanić, R., and Bradlow, A. R. (2009). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," *Lang. Linguist. Compass* **3**(1), 236–264.
- Smith, C. L. (1997). "The devoicing of /z/ in American English: Effects of local and prosodic context," *J. Phonetics* **25**(4), 471–500.
- Stent, A. J., Huffman, M. K., and Brennan, S. E. (2008). "Adapting speaking after evidence of misrecognition: Local and global hyperarticulation," *Speech Commun.* **50**(3), 163–178.
- Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (1992). "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters," *J. Acoust. Soc. Am.* **91**(5), 2979–3000.
- Uchanski, R. M. (2008). "Clear speech," in *The Handbook of Speech Perception*, edited by D. Pisoni and R. Remez (Blackwell, Malden, MA), pp. 207–235.