Exaggeration of featural contrasts in clarifications of misheard speech in English

Jessamyn Schertz

University of Arizona
Douglass 200E
Tucson, AZ 85721
Tel: (812) 390-2033
Email: jschertz@email.arizona.edu

Abstract

This study investigates the extent to which speakers manipulate featural distinctions when trying to clarify misunderstood speech, focusing on voicing contrasts in stops and height and backness (represented by F1 and F2) and durational contrasts in vowels. Participants interacted with a simulated speech recognizer, repeating words when they were "guessed" incorrectly. Both phonemically voiced and voiceless stops showed more extreme VOT values when elicited by an incorrect guess in which the consonant was a minimal pair in voicing with the target consonant (e.g. subject reads "bit", computer guesses "pit"), but not when elicited by an open-ended request for repetition (e.g. subject reads "bit", computer guesses "What did you say?"). A follow-up study showed that the change in VOT between the two repetitions was only present when the incorrect guess contrasted in voicing, but not when it contrasted in place or manner. In contrast, for vowels, the amount and direction of formant change in the F1-F2 space was not significantly different from zero for either type of incorrect guess. However, when there was a durational component to the vowel contrast (/i/ vs. /ɪ/), speakers exaggerated the durational differences between the segments, as opposed to when there was not a durational contrast (e.g. /i/ vs. /u/). The results show that speakers perform local, systematic, and phonologically informed manipulations of temporal contrasts online when clarifying phonetic segments.

Keywords: Voice onset time, Clarification of misheard speech, Hyperarticulation, Phonetic variation, Contrast enhancement

1. Introduction

Speakers manipulate segmental features systematically in order to adjust to different communicative demands. For instance, studies comparing clear and conversational speech report phonetic differences between the two speech styles, including vowel space expansion and changes in the realization of consonants (see review in Smiljanić and Bradlow 2009), some of which could be driven by strategies for featural enhancement. Similar effects have been found in prosodic strengthening environments (Cho 2005; Cho and McQueen 2005; Choi 2003; de Jong 1995), in "Lombard speech" (the speech produced by speakers in noisy conditions, e.g. Hazan and Baker 2011; Cooke and Lu 2010; Garnier et al. 2010), and in communicative tasks involving conveying information to a listener (Baese-Berk and Goldrick 2009). The present study explores the nature and specificity of phonetic featural enhancement in stops and vowels by examining how speakers manipulate phonetic characteristics of these segments when clarifying misheard words.

Much of the work examining how speakers manipulate speech to increase clarity is situated within the framework of Lindblom's (1990) "H&H Theory," which considers speech on a continuum from hypo- to hyper-articulated speech styles, resulting from competing system- and output-oriented constraints. Under this view, hyperarticulation is speech in which phonetic contrasts are exaggerated in order to maintain maximum clarity (an "output-oriented constraint"), potentially sacrificing gestural economy (a "system-oriented constraint"). Variation in the speech signal is therefore attributable to speakers striving for "sufficient discriminability," the output of which will be different in different communicative situations.

The term "hyperarticulation" is used quite loosely throughout the literature, generally referring to any speech which is intentionally or unintentionally clearer for either communicative or structural reasons. However, despite the intuitive validity and theoretical simplicity of a single-dimensional "conversational-to-clear" speech continuum, this metric fails to account for the large amounts of variation which may be orthogonal to the hypo- vs. hyper-articulated axis (see Warner (2011) for discussion), and studies focusing explicitly on elicitation of various clear speech styles have found acoustic differences between them. For example, Uther et al. (2007) found that pitch and emotional affect (positive vs. negative, as rated by naive listen-

ers) were different in foreigner- vs. infant-directed speech, while Hazan and Baker (2011) found that speakers made different types of phonetic changes to aid intelligibility in different adverse listening conditions. Similar complexity is found in the domain of prosodic strengthening, in which segments are articulated at prosodic boundaries or in prosodically prominent (stressed or accented) domains (see Cho and Keating (2009) and Shattuck-Hufnagel and Turk (1996) for reviews): Cho and Keating (2009) showed that English speakers make different modifications at prosodic boundaries than in prosodically prominent domains, and Cho et al. (2011) further showed (with Korean speakers) that these two types of strengthening are encoded differently than communicatively-driven hyperarticulation (i.e. clear speech). Such findings suggest that speakers are attuned to the details of different speaking situations and manipulate phonetic parameters accordingly, and that hyperarticulation is better considered as a cluster of adaptation strategies than a stable mode of speaking.

1.1. Phonetic contrast enhancement in hyperarticulation

Despite differences found in the various types of hyperarticulation, similar patterns which may be attributable to phonetic contrast enhancement have been found across studies examining different types of hyperarticulation. Several studies have examined the realization of the stop voicing contrast during hyperarticulation. Evidence from studies showing that phonemically voiceless stops have longer voice onset times (VOT) in clear than in conversational speech suggests that clear speech may be contrast-enhancing and that VOT is one way this contrast enhancement is realized (English: Picheny et al. 1986; Korean: Kang 2009; Cho et al. 2011). However, it is difficult to separate potential effects of contrast enhancement from effects of speaking rate. Smiljanić and Bradlow (2008a), who found a similar pattern in English and Croatian, show that while the raw VOT for voiceless stops increases for clear speech, the proportion of the stop which is aspirated relative to the total stop length remains constant for both voiced and voiceless stops: in other words, the results could be interpreted as simply a decrease in speaking rate, while the temporal pronunciation norms remain the same.

Cross-linguistic work focusing on the realization of the voicing contrast in different speech rates (Kessinger and Blumstein 1997) and prosodic positions (Cho and Keating 2001; Cho and McQueen 2005; Cho et al. 2011; Kuzla and Ernestus 2011) has shown that language-specific phonological properties can have an effect on how VOT is modified in these environments. For example,

VOTs of voiceless stops in prosodically stronger positions are longer in English (Lisker and Abramson 1964; Choi 2003; Cho and Keating 2009), but shorter in Dutch (Cho and McQueen 2005). These results demonstrate that the modification of VOT is governed at least in part by phonological properties specific to a given language, as opposed to a global decrease in speaking rate. Further evidence for language-specific modifications comes from work by Granlund et al. (2012), who examined differences in VOT between Finnish /p/ and English /b/, both of which are phonetically voiceless, in late bilingual Finnish-English speakers. They found that there was more of a decrease in VOT between conversational and clear speech in English than in Finnish. They hypothesized that the effect arises from speakers trying to distance the English /b/ from its counterpart /p/, whereas there is not a comparable two-way contrast in Finnish.

Turning to work on vowels, a large number of studies have reported vowel space expansion in clear speech when compared with less formal speech or more spontaneous styles (e.g. Ferguson et al. 2010; Ferguson and Kewley-Port 2007; Smiljanić and Bradlow 2005; Erickson 2002; Bradlow 2002; Koopmans-van Beinum 1980; Picheny et al. 1986; Krause and Braida 2004; Bradlow et al. 1996), and in stronger prosodic positions (Cho 2005; de Jong 1995). However, findings are not totally consistent: Ladefoged et al. (1976) failed to find variation in formants due to a change in speech style, and Krause and Braida (2004) found no expansion for clear speech spoken at a normal (as opposed to slower) speaking rate. Granlund et al. (2012) found a smaller distance on the F1-F2 plane between /i/ and /ɪ/ in clear speech than in casual speech, as opposed to the expected expansion in spectral difference.

The fact that vowel space expansion seems to be found relatively consistently is notable; however, it is important to keep in mind that the expansion in all of these styles is relative to a more casual speech task (normally referred to as "conversational speech," despite the fact that the speech is elicited by asking subjects to read sentences in a "natural" or "conversational" style). While more extreme formant values in the clear speech condition are generally interpreted as expanding the vowel space from the "baseline" of the conversational style, it is also possible that the extreme values are actually the targets and the less extreme "baseline" values result from reduction. It is therefore difficult to tease apart any potential effects of expansion from known effects of reduction.

The language-specific phenomena influencing the variation in the realization of stop consonants in hyperarticulation have not been attested for vowel

space expansion. If vowel space expansion is driven by maximizing contrasts, it might be expected that languages with more crowded vowel spaces might show more expansion, whereas languages with smaller inventories would show less, as there is less risk of confusing the different vowel categories. However, Bradlow (2002) found comparable amounts of expansion in languages with differently sized vowel inventories (Spanish and English). Similarly, Cho et al. (2011) found vowel space expansion for Korean, a language with a relatively small vowel inventory. These results suggest that the expansion effect may not be dependent on language-specific phonological contrasts, but rather that it may results from global hyperarticulation strategies.

Durational increases in hyperarticulated vowels are well-documented (e.g. Perkell et al. 2002). As in the above discussion of VOT lengthening, these durational increases can be difficult to separate from effects of slower speaking rate; however, work by de Jong and Zawaydeh (2002) and de Jong (2004) shows that the durational increases found in vowels under contrastive focus are not solely attributable to global lengthening, but that they are dependent on the phonological contrasts of the language. In particular, although there are durational differences in vowels preceding voiced vs. voiceless stops in both English and Arabic, the difference is exaggerated in stressed syllables in English, but not in Arabic, presumably because the durational difference is used in a phonologically contrastive way in English but not in Arabic. Work by Smiljanić and Bradlow (2008a) also revealed cross-linguistic differences: speakers exaggerated the Croatian long vs. short vowel contrast more than English speakers exaggerated the tense-lax duration distinction. The authors argue that this difference reflects the greater importance of duration in the Croatian contrast than in the English contrast, which also has a quality distinction. However, again the mapping is not always straightforward: Granlund et al. (2012) examined the tense/lax contrast in bilingual speakers of English and Finnish in conversational and clear speech, expecting to find greater spectral enhancement of the English contrast and greater durational enhancement of the Finnish contrast in clear speech, reflecting the relative importance of those cues in each language. However, there was no significant difference between enhancement strategies in the two languages, even though speakers did use distinct cues to produce the vowels in each language.

In sum, cross-linguistic work has revealed that hyperarticulation elicits some language-specific patterns of phonetic modification, particularly in voice onset time in stops and durational properties of vowels. This work suggests that hyperarticulation in these domains is defined, at least in part, by

the exaggeration of language-specific featural contrasts. Furthermore, there is evidence that these modifications are augmented when a word with an explicit featural contrast is present in the communicative context. Kang and Guion (2008) and Baese-Berk and Goldrick (2009) found that Korean and English speakers, respectively, exaggerated VOT during read speech when a minimal pair in stop voicing was present. Kirov and Wilson (2012) replicated this effect and found that VOT was also exaggerated when a minimal pair in place of articulation was present, but not when there was a manner of articulation contrast. These results show that speakers are capable of specific phonetic enhancement in certain communicative situations. This study investigates which features are enhanced, as well as how specific and local this enhancement is, in the clarification of misheard speech, a communicative context which would be expected to maximize contrastive effects, should they exist.

1.2. Clarification of misheard speech

Several studies have documented phonetic modifications made by talkers clarifying apparently misheard speech. Oviatt et al. (1998a) found that speakers globally lengthen speech segments and pauses and exaggerate intonational contours when correcting errors made by a simulated speech recognition system. Using a similar paradigm, Ohala (1994) also found durational increases in both vowels and voiceless stop consonants. Ohala (1994) further examined whether these increases were larger in the specific segment that had been incorrectly guessed; for example, whether subjects would increase the VOT of the word-initial voiceless stop in 'pat' more if the program had incorrectly guessed a minimal pair in stop voicing ('bat') than if it had incorrectly guessed a minimal pair in the vowel ('pot'). Ohala found no significant difference in VOT between the repetitions elicited by a stop contrast and those elicited by a vowel contrast, and found no difference in vowel formant values between the two repetitions. These results suggest that speakers do not heighten contrastive cues in order to emphasize contrasts, nor is there local enhancement of misheard parts of words, but rather that speakers generally maintain pronunciation norms when making clarifications.

However, later work using similar paradigms has revealed differences between "global" and "focal" hyperarticulation in error correction. Oviatt et al. (1998b) analyzed productions of speakers correcting a simulated speech recognizer which had two types of errors: general recognition failure (system

responded "???"), or substitution (e.g. system guessed "International graphics" for spoken "National oceanographic"). Durational increases (such as number and length of pauses) were found for both global and focal corrections, but effects were larger during focal error repairs. Levow (1999) found similar differences in durational patterns using a corpus drawn from field trials of an actual recognition system. Work by Stent et al. (2008) showed that speakers make use of local hyperarticulation: subjects spoke more slowly when making repairs, and also produced more "clear forms" of consonants (e.g. mid-word /t/ instead of /ɾ/, released instead of unreleased word-final /t/), and these modifications were greater during the part of the utterance that had been misunderstood than during the rest of the utterance. A post-hoc analysis of vowels in clarifications suggested that the majority of speakers were more likely to have produce front vowels further front (i.e. higher F2) in clarifications; however, this did not hold for all speakers, and was not analyzed statistically. Finally, Maniwa et al. (2009) used a similar paradigm to that of Ohala (1994) to examine enhancement of place and voicing contrasts in fricatives. Subjects read words containing target fricatives and were asked to repeat the words when they were incorrectly guessed as a minimal pair in either place or voicing by a simulated speech recognizer. Using a constellation of 14 acoustic measurements, including both temporal and spectral features, Maniwa et al. found that acoustic modifications were produced in a direction that enhanced the relevant contrast.

In sum, although it appears that speakers make local modifications to misheard speech, the details of these phonetic modifications are not yet well understood. In particular, VOT and vowel formant manipulation, which have been robustly documented in other types of hyperarticulation such as clear speech, have only been explored systematically by Ohala (1994), whose extremely small dataset may have masked more subtle effects. The current study therefore revisits the question of how speakers modify stop consonants and vowels when making clarifications, and in particular, whether speakers specifically enhance phonetic contrasts online, or whether instead the modifications reflect more general strategies for hyperarticulation.

1.3. Goals of the study

This study investigates the scope and phonetic specificity of speakers' clarifications of misheard speech. Experiment 1 uses the general paradigm of Ohala (1994) with a modified procedure and larger variety of items in an aim to address three specific questions. First, what phonetic modifications do

speakers make to stop consonants and vowels when clarifying words that have been misheard? Second, what is the scope of these modifications: do they target the specific segment that has been misheard, or are all segments in the misheard word targeted to an equal extent? Finally, are the modifications influenced by the nature of the phonetic contrast between the target and misheard segment (the "contrastive hypothesis") or are they driven by global strategies for hyperarticulation, regardless of the misheard segment type? For stops, the contrastive hypothesis predicts that speakers should exaggerate VOT durations when confronted with a voicing contrast; that is, they should have longer positive VOT when repeating a voiceless stop that has been misheard as voiced, and shorter (or longer negative) VOT when repeating a voiced stop that has been misheard as voiceless. For vowels, the direction of movement in formant values predicted by the contrastive hypothesis should depend on the quality of the vowel that has been misheard. For example, a token of /u/ elicited by an incorrect guess of /i/ would be expected to be further back (lower F2), while a /u/ elicited by an incorrect guess of /o/ would be expected to be higher (lower F1). Amplitude and durational measures were also taken to determine whether a switch to a globally clearer speech style might account for any phonetic modifications.

The specificity and locality of consonantal manipulations found in Experiment 1, accompanied by a lack of comparable manipulations for vowels, led to two follow-up studies which further explore questions raised by the results of the first experiment. Experiment 2 focuses on consonant contrasts, addressing the specificity of the VOT exaggeration found in Experiment 1. In particular, the second experiment examines whether speakers exaggerate VOT whenever the consonant has been misheard, regardless of segment was substituted in the incorrect guess, or whether instead the VOT exaggeration is specific to disambiguating the voicing contrast, which is primarily distinguished by VOT, but not other contrasts (e.g. place of articulation) which do not use VOT as the primary contrastive feature. Experiment 3 replicates the vowel study of Experiment 1 using stimuli containing vowels that are closer together in the vowel space (the tense-lax /i/-/ɪ/ contrast) and thus potentially more likely to elicit featural exaggeration. Since the English tense-lax distinction is realized by a durational as well as a spectral contrast (e.g. Hillenbrand et al. 2000), Experiment 3 additionally tests whether speakers exaggerate this durational aspect of the contrast in addition to (or instead of) the spectral aspect of the contrast.

Together, these experiments compare the phonetic modifications made by

speakers clarifying misheard stops and vowels. They allow for comparison with previous work documenting featural enhancement in other contexts eliciting hyperarticulation (e.g. clear speech and stronger prosodic positions), and further explore the variability present in these modifications, both in terms of their scope and specificity (i.e. which segments are targets for enhancement) and in terms of the differences in how different featural contrasts (spectral vs. temporal) are enhanced.

2. Methodology

2.1. Participants

Subjects in all studies were undergraduate students at the University of Arizona who received course credit for participation. 12 subjects (2 males, 10 females) participated in Experiment 1, 15 in Experiment 2 (6 males, 10 females), and 15 in Experiment 3 (4 males, 11 females). Speakers ranged in age from 18 to 25 years, and most grew up in Arizona (with the exception of five from Illinois, five from California, one from New York, and one from Washington) and were monolingual in English until at least high school. All reported that they used only English regularly. No subject participated in more than one experiment.

2.2. Procedures

Subjects were recorded in a soundproof recording booth in the Douglass Phonetics Laboratory at the University of Arizona. A high quality head-mounted microphone and a CD recorder (sampled at 44.1 kHz) were used for the recordings. The participants were instructed that they were going to interact with a computer program. They were asked to read words which appeared on the computer screen, and after reading the word, the computer would guess what they said by displaying a written guess (e.g. "Did you say 'beat'?"). If the computer guessed correctly, the subject was instructed to say "yes," and the program would move on to the next word. If the guess was incorrect, the subject was asked to repeat the original word in isolation[1].

[1]A more natural response to the computer's incorrect guess would be something like "No, I said 'bit'". However, in order to keep the two repetitions as similar as possible, subjects were explicitly told to repeat only the word itself (the purported reason being that the computer program could only recognize words in isolation). None of the subjects had difficulty doing this.

Instead of guessing the words, the computer was actually pre-programmed to provide specific incorrect guesses for approximately one-third of the words in order to elicit repetitions of target words from the subjects. In Experiment 1, the incorrect guesses were of two types: either the computer incorrectly guessed a word that differed by a minimal pair in either initial consonant voicing or vowel quality (Contrastive condition, *e.g.* subject reads "bit," computer responds "Did you say 'beat'?"), or it simply asked for a repetition (Open Response condition, *e.g.* subject reads "bit," computer responds "What did you say?"). The sentences containing the target stimuli were randomized such that all conditions (Open Response vs. Contrastive and Vowel vs. Consonant) were intermingled. In Experiments 2 and 3, only the Contrastive condition was used.

In all experiments, a practice block demonstrating correct and both types of incorrect guesses was presented first to ensure familiarity with the experimental task. This was followed by three test blocks, which were randomized such that each participant saw the stimuli in a different order. Subjects could take a break between each block. Fillers were correct guesses; these made up approximately 2/3 of the trials in each experiment[2]. Subjects were told that the computer program was not very good at this stage in the developmental process; nevertheless they were instructed to speak naturally, purportedly because the computer program was being trained to recognize natural speech. Each experiment took between 20 and 35 minutes.

2.3. Stimuli

Target items were monosyllabic, CVC words. A summary of the stimuli used for each condition is given in Table 1, while the complete set of stimuli for all experiments are given in Appendices A-C. For the consonant conditions, target items consisted of minimal pairs differing in the initial stop (*e.g.* target: "dime", incorrect guess contrasting in voicing: "time"). For the vowel conditions, target items consisted of minimal pairs in vowels (*e.g.* target: "beat", incorrect guess contrasting in height: "bait", incorrect guess

[2]Target words were never guessed correctly on the first try, in order to avoid over-representation of target words with respect to filler words (subjects already were hearing each target word at least twice in incorrect guess conditions, whereas filler items only occurred once). However, the "correct guess" fillers did include phonologically similar words containing target sounds (e.g. word-initial stops), so it was not the case that the target sounds were *always* guessed incorrectly.

Table 1: Summary of conditions and stimuli for all experiments. The columns "Target" and "Guess" provide an example target word (which is read by the subject) and an example guess made by the computer for each condition. After an incorrect guess, the subject was asked to repeat the target word, resulting in two repetitions of each target per subject.

Experiment	Condition	Target	Guess	Number of items
Experiment 1	Contrastive (C voicing)	"bit"	"pit"	/p/-/b/, /b/-/p/, /t/-/d/, /d/-/t/, /k/-/g/, /g/-/k/ (9 each = 54 total)
	Contrastive (V height/backness)	"dean"	"dune"	/i/-/u/, /u/-/i/, /i/-/e/, /e/-/i/, /e/-/o/,/o/-/e/, /o/-/u/, /u/-/o/ (5 each = 40 total)
	Open Response	"dean"	"???"	68 (all items from Contrastive condition)
	Correct fillers	"meat"	"meat"	261
Experiment 2	C voicing	"dime"	"time"	/t/-/d/, /d/-/t/ (14 each = 28 total)
	C place	"deem"	"beam"	/t/-/p/, /t/-/k/, /d/-/b/, /d/-/g/ (7 each = 28 total)
	C manner	"dash"	"sash"	/t/-/z/, /t/-/n/, /d/-/s/ (7 each = 28 total)
	Incorrect fillers	"mate"	"meet"	84
	Correct fillers	"mint"	"mint"	255
Experiment 3	/i/-/ɪ/	"beet"	"bit"	/i/-/ɪ/ (12)
	/ɪ/-/i/	"bit"	"beet"	/ɪ/-/i/ (12)
	/i/-/u/	"beet"	"boot"	/i/-/u/ (12)
	/ɪ/-/ɛ/	"bit"	"bet"	/ɪ/-/ɛ/ (12)
	Incorrect fillers	"date"	"gate"	86
	Correct fillers	"nice"	"nice"	269

contrasting in backness: "boot"). Fillers (also CVC words) were chosen to maximize phonetic diversity in the words read by the subject by including sounds underrepresented in the target words (e.g. low vowels). For Experiment 1, all fillers were correct guesses because the target sounds which were guessed incorrectly were quite diverse (all 6 stops and 4 vowels). For Experiments 2 and 3, there was a smaller set of target sounds, so incorrect guesses were also added as fillers in order to keep participants from fixating on the target sounds. The stimuli were randomized and divided into three blocks, using a counterbalanced Latin square design.

2.4. Acoustic measurements

All acoustic measurements were performed with Praat (Boersma and Weenink 2011). If a subject mispronounced a word in either the first or second repetition, both repetitions of that word were omitted from the analysis (less than three percent of trials were omitted). Initial consonants and vowels were labeled for both repetitions of each target word.

2.4.1. Global measures of clear speech

Vowel duration and peak intensity were measured to determine whether speakers increased duration and/or amplitude when making clarifications. These two measures were chosen because they have routinely been found to change in studies of clear speech (amplitude: Picheny et al. 1986; Granlund et al. 2012; duration: Moon and Lindblom 1994; Ferguson and Kewley-Port 2007; Bradlow 2002; Granlund et al. 2012).

Vowel duration: The marker for the beginning of the vowel was placed at the first zero crossing in the waveform after the beginning of periodicity of the vowel; the marker for the end was placed at the end of the visible second formant in the spectrogram before the following consonant (which was always a stop or a fricative).

Intensity: Intensity contours were generated by Praat's "To Intensity..." function. Peak intensity was calculated as the RMS intensity measured during a 32 ms window around the intensity peak of the vowel.

2.4.2. Local measures

Stops - Voice onset time: If there was prevoicing, VOT was labeled from the start of visible voicing in the waveform to the beginning of the consonant

burst in the waveform[3]. If there was no prevoicing, VOT was measured from the beginning of the consonant burst in the waveform to the first zero crossing in the waveform following the onset of periodicity in the following vowel.

Vowels - Formant measures: The first and second formants of all target vowels were extracted automatically using the Burg algorithm in Praat. Prior to extraction, individual formant ceilings for each speaker and each vowel were chosen manually after visual inspection of the data, following the procedure in Escudero et al. (2009), in order to minimize error. Since some target vowels were diphthongized (in particular /e/), the possibility of formant change needed to be accounted for in the measurements. After manual inspection of the data, the 1/3 and 2/3 points were determined to be the best point to measure the formants in order to catch the endpoints of the diphthongs without running into formant transitions from flanking consonants. The formant values were then manually checked and corrected when necessary.

F1 and F2 values were used to compute two metrics of vowel shift between repetitions. $\Delta F1F2$ was measured as the Euclidean distance in F1-F2 space between the first and second repetitions of each target word. The change in F1 and F2 between the two repetitions ($\Delta F1$ and $\Delta F2$) were also computed in order to examine changes on each dimension separately.

2.5. Statistical Analyses

Statistical analyses were done using within-subjects ANOVAs. Unless otherwise noted, the dependent variable was the difference between the two repetitions in each target word (e.g. for vowel duration, the dependent variable was the difference between the vowel duration of the clarification and the vowel duration of the read version of a given word by a given speaker). The ANOVAs were used to determine whether the independent variables were correlated with differing degrees of effects; t-tests were then used to determine whether these differences were significantly different than zero. Phonemic Voicing (phonemically voiceless /p t k/ vs. phonemically voiced

[3]During the review process it was suggested that this method of measurement does not capture the difference between voicing that continues all the way through the closure and prevoicing which is followed by a period of silence before the stop burst. This difference could reflect how strongly a speaker is emphasizing prevoicing. The current data included only a few tokens with cessation of prevoicing, but this measurement method could slightly exaggerate the strength of prevoicing.

/b d g/) and Guess Type (Contrastive or Open Response) were used as independent variables throughout the analyses[4]. For all models, all significant interactions were tested, and for all analyses, p-values less than 0.05 were considered significant.

3. Results: Global Measures

It was expected that the experimental paradigm would induce participants to produce globally clearer speech in the clarification than in the read token of each word. Relative amplitude and vowel duration were measured as possible phonetic correlates of globally clearer speech, since these factors have been shown to increase in most studies of clear speech (amplitude: Picheny et al. 1986; duration: Picheny et al. 1986; Moon and Lindblom 1994; Ferguson and Kewley-Port 2007; Bradlow 2002 (though see Perkell et al. 2002 for a lack of effect of amplitude in clear speech). If the experiment elicited two different speech styles comparable to previous work, changes in these two measures would be expected. The global measures are drawn from all tokens from Experiment 1 (1598 tokens over 12 speakers)

Intensity: A one-way, within-subjects ANOVA showed no effect of Guess Type on change in intensity between the two repetitions (ΔInt). The effect of Guess Type was not significant ($F(1, 11) = 3.84, p > .05$). The average difference in intensity between the first and second repetitions of each word (-0.15 dB) was not significantly different than zero between speakers ($t(11) = -.85, p > .05$).

Duration: A one-way, within-subjects ANOVA showed no effect of Guess Type on change in vowel duration between the two repetitions (ΔDur). The effect of Guess Type was not significant ($F < 1$). The average difference in vowel duration between the first and second repetitions of each word, over both Guess Types (1.8 ms) was not significantly greater than zero between speakers ($t(11) = .66, p > .05$).

In summary, results from intensity and vowel duration measurements suggest that there was not a significant increase in factors normally associated

[4]Block was originally included in the ANOVAs to check for changes in speaker behavior over time in the experiment; no practice effects were found. Place was also included as a factor and failed to show significant main effects or participate in interactions. As there was no effect for either of these factors, both were removed in the final analyses to avoid loss of statistical power.

with clearer speech between the two repetitions. This lack of difference most likely results from the fact that speakers were speaking very clearly even on the first repetition. The fact that speakers did not adopt a globally clearer speech style in the second repetition of words suggests that the phonetic modifications made by the speakers are not the result of global "clear speech" adjustments, but rather a result of the adjustments made specifically to clarify misheard speech.

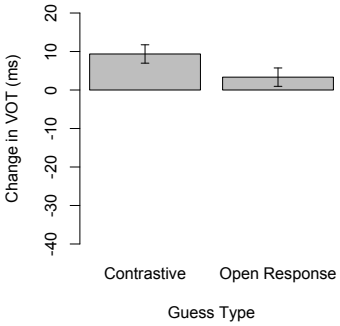
4. Results and Discussion: Consonants

4.1. Experiment 1: Effect of Guess Type on VOT manipulation

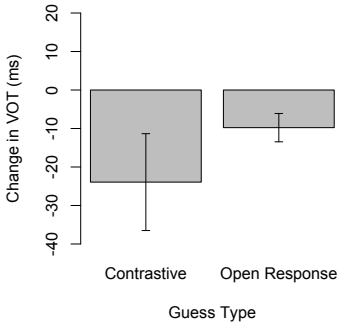
Results for VOT are shown in Figure 1. Data were analyzed in a two-factor within-subjects ANOVA, with a dependent variable of difference in VOT duration between the clarification and read token of each target word

Figure 1: Change in Voice Onset Time (ΔVOT) by Voicing and Guess Type. In this as well as all subsequent figures showing within-subjects effects, error bars represent the 95% confidence intervals based on the normalized means for each speaker.

(a) **Voiceless stops:** ΔVOT is significantly greater in the Contrastive than in the Open Response condition



(b) **Voiced stops:** ΔVOT has a significantly greater negative value in the Contrastive than in the Open Response condition



(ΔVOT). The independent variables were Guess Type (Contrastive or Open Response), and Phonemic Voicing (phonemically voiced or voiceless). The main effect of Voicing was significant ($F(1, 11) = 5.55, p < .05$), as was the two-way interaction of Guess Type by Phonemic Voicing ($F(1, 11) = 37.02, p < .001$). The main effect of Guess Type was not significant ($F = 2.27, p > .05$). Since the two-way interaction indicated that the effect of Guess Type was different at the different levels of Voicing, the effect of Guess Type on voiced and voiceless stops was analyzed separately in two one-factor ANOVAs. For phonemically voiceless stops, there was a significant main effect of Guess Type ($F(1, 11) = 6.11, p < .05$). Voiceless stops in the Contrastive condition had significantly greater ΔVOT than those in the Open Response condition (clarifications were on average 9 ms longer than read tokens in the Contrastive condition, compared to 3 ms longer in the Open Response condition). Furthermore, ΔVOT was significantly greater than zero in the Contrastive condition ($t(11) = 2.97, p < .05$), but not in the Open Response condition ($t(11) = 0.88, p > .05$). Voiced stops also showed a main effect of Guess Type ($F(1, 11) = 14.21, p < .005$), showing that ΔVOT for voiced stops is significantly larger in the Contrastive condition than in the Open Response condition (-24 ms change in the Contrastive condition vs. -10 ms change in the Open Response condition). Again, ΔVOT was significantly different from zero in the Contrastive condition ($t(11) = -2.81, p < .05$) but not in the Open Response condition ($t(11) = -1.60, p > .05$).

4.1.1. Variation in phonetic voicing of phonemically voiced tokens

One potential issue with the measurement of voicing is that prevoiced (negative VOT) and voiceless unaspirated stops were grouped together for analysis. Although this does capture the general pattern that speakers are making voiced stops "more voiced," it is not clear from this analysis whether speakers are prevoicing longer, switching from voiceless to prevoiced, or both. Three subjects did not prevoice any of their stops; they are excluded from the following discussion. Table 2 shows the frequency of each possible combination across repetitions for phonemically voiced target words. The majority of words were produced as voiceless unaspirated for both repetitions (306 tokens), while the next most common was for both repetitions to be prevoiced (149 tokens). Changing from one repetition to the other was less common, but switching from a voiceless token in the first reading to voiced in the clarification was more common than switching from voiced to voiceless (44 vs. 13 tokens).

Table 2: Voicing patterns in phonemically voiced tokens across repetitions

Read	**Clarification**	**Number of tokens**
Prevoiced	Prevoiced	149
Prevoiced	Voiceless	13
Voiceless	Prevoiced	44
Voiceless	Voiceless	306

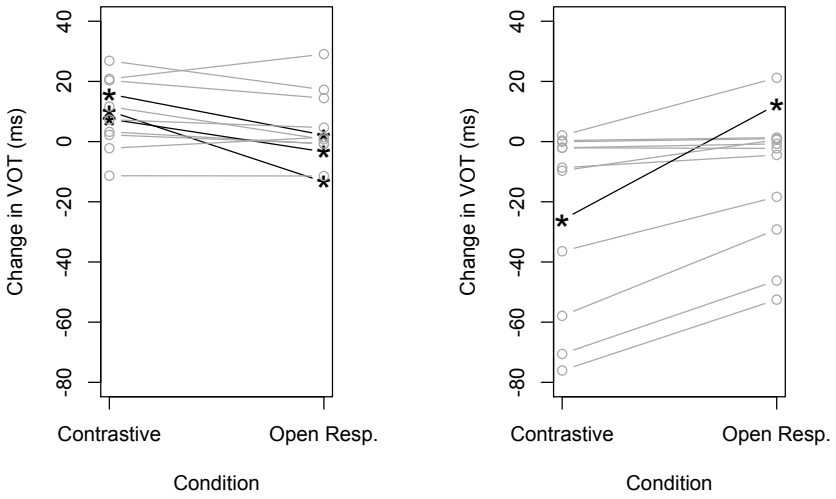
When both repetitions are prevoiced, it is expected that the duration of prevoicing in the second repetition relative to the first repetition would be longer in the Contrastive than in the Open Response condition. This is the case for the subset of the data in which both repetitions are prevoiced: the mean difference in voicing is -39 ms for the Contrastive Condition and -22 ms for the Open Response condition, a difference which is significant based on an ANOVA of prevoicing duration difference by Guess Type and Segment (/b/, /d/, /g/). Guess Type was significant ($F(1, 42) = 5.03, p < .05$), while Segment was not a significant predictor of difference, nor was there an interaction of Guess Type by Segment. For cases in which one repetition was prevoiced and the other was not, it is expected that a change from voiceless to voiced would occur more in the Contrastive Condition. This was again the case in the relevant subset of data: in the Contrastive Condition, a switch from voiceless to voiced was 3 times more frequent than a switch from voiced to voiceless, while it was only 1.2 times more frequent in the Open Response condition. For those tokens in which both repetitions were voiceless, the mean difference in VOT between repetitions was less than 1 ms, and a two-way ANOVA analyzing VOT difference by Guess Type and Place confirmed that there were no significant differences in the duration between the Guess Types or Places ($F < 1$).

4.2. Individual differences in VOT manipulation

Beyond the statistically significant group effects, individual performance was examined: a graph of speakers' individual means of ΔVOT in the Contrastive vs. Open Response conditions is shown in Figure 2, while numerical means and statistical results are shown in Appendix D. One-way ANOVAs were used to analyze whether the difference between the means in the two conditions for a given speaker were significantly different, taking voiced and

Figure 2: Individual speaker means (in ms) for ΔVOT in the Contrastive and Open Response conditions. Asterisks indicate a statistically significant difference between the two conditions for that speaker.

(a) **Voiceless stops** (b) **Voiced stops**



voiceless stops separately. The means are based on 25-26 tokens per condition (27 was the full set; several tokens had to be removed because of e.g. speaker error, as described in the methodology). Only 3 speakers showed statistically significant differences between the conditions for voiceless stops, and only 1 speaker did for voiced stops. All significant effects were in the same direction as the group effect. The three subjects who did not prevoice any stops (Subjects 1, 3, and 11), along with one subject who prevoiced only two tokens (Subject 7) had the smallest mean difference for voiced stops. Although the majority of speakers did not reach significance for either voiced or voiceless stops, the differences are for the most part in the same direction of the group effects.

4.3. Effect of Misheard Segment Type on VOT manipulation

It is possible that the difference between the Contrastive and Open Response conditions was not due to the explicit voicing distinction between the target and incorrect guess, but rather one of the other factors that was different in the two conditions; for example, it is possible that speakers perform differently when presented with a specific incorrect guess than with an open response, regardless of what that incorrect guess is. In order to test this, consonantal tokens from the vowel condition were analyzed (e.g. comparing the read token and clarification of 'boat' when the computer incorrectly guessed 'bait'). These consisted of 24 voiced and 24 voiceless tokens for each speaker (all stop-initial words listed in the vowel contrast section of Appendix A). Data were analyzed, as in the previous analysis, in a two-way ANOVA with a dependent variable of ΔVOT and independent variables of Phonemic Voicing and Misheard Segment Type (consonant or vowel). The main effect of Phonemic Voicing was significant ($F(1, 11) = 8.07, p < .05$), showing that voiceless stops have a longer VOT than voiced stops, as was the interaction between Voicing and Misheard Segment Type ($F(1, 11) = 6.38, p < .05$). The main effect of Misheard Segment Type was not significant ($F(1, 11) = 3.30, p > .05$). Following the significant interaction, a separate one-way ANOVA was run for voiced and voiceless stops. For voiceless stops, the effect of Misheard Guess Type was significant ($F(1, 11) = 10.72, p < .05$). ΔVOT was greater when the misheard segment was a consonant ($\Delta VOT = 9ms$) than when it was a vowel ($\Delta VOT = 3ms$). As shown above, ΔVOT was significantly different than zero in the consonant condition; however, when the misheard segment was a vowel, ΔVOT was not significantly different from zero ($t(11) = 1.33, p > .05$). The effect was parallel for voiced stops: the effect of Misheard Guess Type was significant ($F(1, 11) = 4.97, p < .05$), showing that the ΔVOT found for voiced stops in the consonant condition was greater than that found in the vowel condition (-24 ms when the misheard segment was a consonant vs. 0 ms when the misheard segment was a vowel). Again, ΔVOT was not significantly different from zero when the misheard segment was a vowel ($t(11) = -.042, p > .05$). The effect of VOT manipulation was therefore local, occurring only when the consonant itself was misheard.

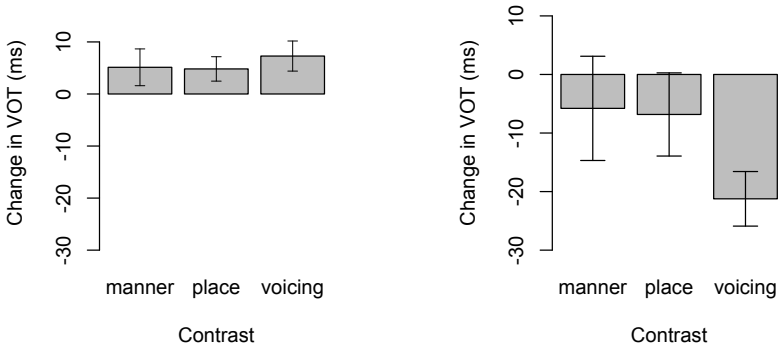
4.4. Experiment 2: Effects of non-voicing contrasts on VOT manipulation

A follow-up experiment tested the effects of non-voicing contrasts on VOT manipulation to determine whether *any* consonantal contrast, not only a

Figure 3: Change in Voice Onset Time ΔVOT by Contrast.

(a) **Voiceless stops** ΔVOT shows a trend in which it is greater in the Voicing than in the Place and Manner conditions, though this difference is not significant ($p < .1$)

(b) **Voiced stops:** ΔVOT has a significantly greater negative value in the Voicing than in the Place and Manner conditions



minimal pair in voicing, would elicit the same patterns in clarifications. The distribution of stimuli and conditions for Experiment 2 is shown in Table 1, and the complete set of stimuli is given in Appendix B. Whereas in the first study, all of the incorrect guesses in the consonant condition differed in word-initial voicing, in this study, there were three different contrasts, all minimal pairs differing in the initial consonant: different voicing but same manner and place (e.g. target: "dime", incorrect guess "time"), different place but same manner and voicing (e.g. target "deem", incorrect guess "beam"), and different manner and voicing but same place (e.g. target: "dash", incorrect guess "sash"). For a more controlled data set, only alveolar consonants (/t/ and /d/) were used as target consonants.

Results for VOT are shown in Figure 3. Data were analyzed in a two-factor ANOVA, with a dependent variable of ΔVOT. The independent variables were Contrast type (voicing, place, or manner), and Phonemic Voicing (phonemically voiced or voiceless). The main effect of Phonemic Voicing

was significant ($F(1, 14) = 5.62, p < .05$), with longer VOTs for voiceless than for voiced stops, as was the interaction between Voicing and Contrast ($F(1, 11) = 5.34, p < .05$). Following this significant interaction, the effect of Contrast was examined separately in one-way ANOVAs for each level of Voicing. For voiceless stops, the effect of Contrast was not significant ($F < 1$). For voiced stops, the effect of Contrast was significant ($F(1, 14) = 6.05, p < .05$), showing that ΔVOT was significantly different for the different contrast types. As shown in the boxplot below, the ΔVOT is much larger for the voicing contrast (mean -21 ms) than for the manner (-6 ms) or place (-6 ms) contrasts, and a one-way ANOVA comparing the Voicing condition to the other two conditions together showed that this effect is significant ($F(1, 14) = 17.47, p < .01$). T-tests were performed to determine whether ΔVOT was significantly different from zero in each condition type (these were done on both types of stops, despite the fact that the main effect for Contrast was not significant for voiceless stops, because the theoretical question of interest involves individual differences between the condition. However, these results must be interpreted with caution because of the lack of main effect). For both voiced and voiceless tokens, speakers produced changes in VOT that were significantly different from zero in the Voicing condition (voiceless stops: $t(14) = 2.18, p < .05$; voiced stops: $t(14) = -2.94, p < .05$), but not in the Place or Manner conditions (Place (voiceless): $t(14) = 1.59, p > .05$; Place (voiced): $t(14) = -.98, p > .05$; Manner (voiceless): $t(14) = 1.43, p > .05$; Manner (voiced): $t(14) = -.83, p > .05$).

4.5. Consonants: summary and discussion

Experiment 1 tested whether speakers manipulate VOTs of word-initial consonants when making a clarification, and whether the degree of manipulation differs depending on whether they think that a specific segment has been misheard (Contrastive condition) or whether the whole word has been misheard (Open Response). For both voiced and voiceless tokens, VOTs were exaggerated (longer positive values for voiceless stops; longer negative values for voiced stops) in the Contrastive condition but not the Open Response condition[5]. For phonemically voiced tokens, tokens which were prevoiced

[5]The effect was larger for voiced than for voiceless stops, in contrast to previous work on conversational vs. clear speech (Smiljanić and Bradlow 2008a) and prosodic enhancement (Choi 2003). This may be because this data set contained a relatively large number of prevoiced tokens, as discussed above, and the effect is carried by these tokens.

in the initial production were more likely to have longer prevoicing in the clarification, and voiceless unaspirated tokens were more likely to 'acquire' voicing in the clarification in the Contrastive condition as compared to the Open Response condition. Furthermore, these effects only occurred when the consonant itself has been guessed incorrectly; when the incorrect guess was a minimal pair differing in the vowel, there was no effect. These results show that speakers are able to manipulate VOT contrasts in a systematic way.

The results from Experiment 1 show that the VOT effects only occurred when the consonant itself was misheard, but since all of the incorrect guesses were minimal pairs in voicing with the target consonant, the results left open the possibility that *any* incorrect consonant guess would elicit the same effects, regardless of featural content. Experiment 2 addressed this possibility by including incorrect guesses differing in place or manner and showed that only a minimal contrast in voicing elicits a significant change in VOT, while a minimal contrast in place does not elicit exaggeration of VOT. This effect is clear for voiced tokens, but must be interpreted with caution for voiceless tokens, since the difference in effect the two conditions was not statistically significant. The prediction that follows from this is that place contrasts are enhanced in the Place condition (but not the voicing condition). Although the current data set is not sufficient to test this prediction, it is a topic for future work.

5. Results and Discussion: Vowels

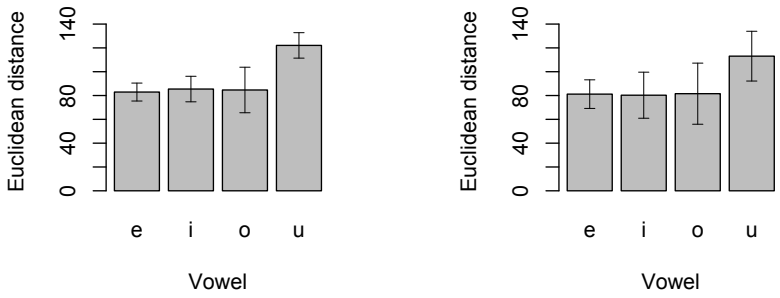
5.1. Experiment 1: Effect of Guess Type on vowel shift

Magnitude of vowel shift: Results for the magnitude of vowel shift between repetitions on the F1-F2 plane are shown in Figure 4. Data were analyzed in a two-factor within-subject ANOVA, with a dependent variable of $\Delta F1F2$ (the Euclidean distance between the vowels in the two repetitions)[6]. The independent variables were Guess Type (Contrastive or Open Response) and Segment (/i/, /e/, /o/, or /u/). The main effect of Segment was significant ($F(1, 11) = 9.36, p < .05$). Neither the main effect of Guess Type nor the interaction of Segment by Guess Type was significant (Guess Type: $F(1, 11) = 1.46, p > .05$; Segment by Guess Type: $F < 1$). The

[6]Measurements were taken at the 1/3 and 2/3 point of the vowel, and analyses were done on both. The significance of main effects and interactions was the same for both measurement points; results from the 1/3 point are reported here.

Figure 4: Euclidean distance (on F1-F2 plane) between first and second repetitions of target words. Differences between the Contrastive and Open Response conditions were not significant.

(a) **Contrastive condition** (b) **Open Response condition**



amount of shift by segment is shown in Figure 4. /u/ has a larger shift (99 Hz) than the other segments (82 Hz for /e/, 83 Hz for /i/ and /o/). This was confirmed by a one-way ANOVA comparing the /u/ to the averages of the other vowels ($F(1, 11) = 16.36, p < .01$).

Direction of shift: Results for direction of vowel shift on the F1 and F2 dimensions are shown in Table 3. Although the magnitude of the vowel shift was not significantly different between the two conditions, the direction of movement still may reveal interesting patterns. Because speakers might be expected to modify the vowel in different directions depending on the quality of the misheard segment, a one-factor ANOVA was run for each segment in the Contrastive condition in order to predict whether there was a significant difference in $\Delta F1$ or $\Delta F2$ based on the incorrect guess. None of the effects reached significance. T-tests were performed to see if each group was significantly different from zero, and all of these tests also failed to reach significance.

Table 3: Experiment 1: Mean differences on F1 and F2 dimensions, sorted by incorrect guess type. "???" indicates the Open Response condition. ANOVAs tested the difference in $\Delta F1$ and $\Delta F2$ predicted by different incorrect guesses. T-tests tested whether the mean for each vowel was different than zero. All t-tests were not significant, $p > .05$.

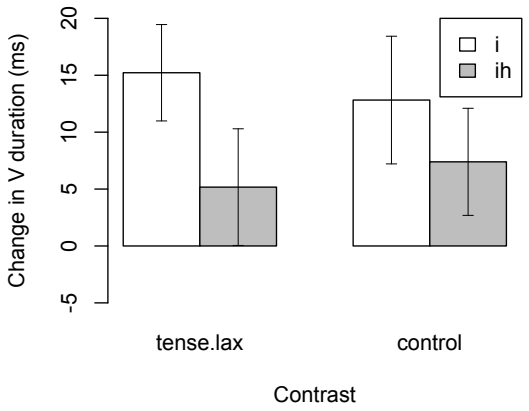
		F1			**F2**		
Target	**Guess**	ΔHz	**ANOVA**	**t-test**	ΔHz	**ANOVA**	**t-test**
/e/	???	4	$F = 1.48$,	$t = 1.40$	0	$F < 1$	$t = -.82$
/e/	/i/	3	$p > .05$		-9		
/e/	/o/	12			-6		
/i/	???	-1	$F < 1$	$t = -.31$	-3	$F = 1.12$,	$t = 1.61$
/i/	/e/	-3			35	$p > .05$	
/i/	/u/	1			6		
/o/	???	4	$F = 2.75$,	$t = -1.41$	7	$F < 1$	$t = 1.08$
/o/	/e/	-8	$p > .05$		-14		
/o/	/u/	-8			-20		
/u/	???	6	$F < 1$	$t = 1.10$	-11	$F < 1$	$t = -1.62$
/u/	/i/	8			-50		
/u/	/o/	1			-6		

5.2. Experiment 3: /i/-/ɪ/ vowel contrasts

Since it is possible that the vowel contrasts used in Experiment 1 were not "minimal" enough to elicit contrast, a follow-up experiment tested the effects of clarification on the /i/-/ɪ/ distinction. The distribution of stimuli and conditions for Experiment 3 is shown in Table 1, and the complete set of stimuli is given in Appendix C. Target words consisted of tokens with vowels /i/ or /ɪ/. There were two types of contrasts: tense-lax (/i/-/ɪ/ (e.g. target "beat", incorrect guess "bit") and /ɪ/-/i/ (e.g. target "bit", incorrect guess "beat")), and two control contexts in which the incorrect guesses are phonologically minimal pairs with the target segments but which have the same tense/lax values (/i/-/u/ (e.g. target "beat", incorrect guess "boot"), and /ɪ/-/ɛ/ (e.g. target "bit", incorrect guess "bet")). There were 12 tokens for each contrast, for a total of 48 target tokens. The computer always guessed either a correct or incorrect word; there was no Open Response condition.

Duration: Results for vowel duration are shown in Figure 5. A within-subjects ANOVA analyzing the effect of Segment (/i/ or /ɪ/) and Contrast

Figure 5: Change in vowel duration between the first and second repetition, by Segment and Contrast Type (Tense-lax = /i/-/ɪ/; Control = /i/-/u/ or /ɪ/-/ɛ/).



Type (tense/lax or control) on the change in duration between the first and second repetitions (ΔDur) revealed a significant interaction between the two conditions ($F(1, 14) = 6.17, p < .05$), with no significant main effect in either condition (Segment: $F(1, 14) = 3.14, p > .05$, Contrast: $F < 1$). Following the interaction, the effect of Segment was tested in a one-way ANOVA at each level of Contrast. For the tense/lax condition, the effect of Segment was significant ($F(1, 14) = 5.69, p < .05$), while for the control condition, the effect of Segment was not significant ($F(1, 14) = 4.42, p > .05$). Speakers increased the duration of /i/ more than /ɪ/ between the two repetitions in the tense/lax condition ($\Delta Dur = 17$ ms for /i/ and 6 ms for /ɪ/), but there was no significant difference between ΔDur of the two segments in the control condition.

Since there were only 12 tokens per condition for each speaker, there was not enough power to do individual statistics; however, individual means are shown in Appendix D. As is evident from the means, there is considerable

variability between speakers. While the majority (12 out of 15 speakers) show the group pattern of a larger change in /i/ than in /ɪ/ when presented with the tense-lax contrast, it is implemented differently, with some speakers lengthening both but lengthening /i/ more, while others shorten /ɪ/.

Magnitude of vowel shift: Data were analyzed as in Experiment 1 above, in a two-factor within-subject ANOVA, with a dependent variable of $\Delta F1F2$, the Euclidean distance between the vowels in the two repetitions and independent variables of Segment and Contrast. Neither of the main effects were significant, nor was the interaction (Segment: $F(1, 14) = 1.49, p > .05$; Contrast: $F < 1$; Segment by Contrast: $F < 1$).

Direction of shift: ANOVAs examining the effect of Segment and Contrast on $\Delta F1$ and $\Delta F2$ separately showed no significant main effects or interactions (F1: Segment: $F(1, 14) = 1.62, p > .05$; Contrast: $F < 1$; Segment by Contrast: $F < 1$; F2: Segment: $F(1, 14) = 1.65, p > .05$; Contrast: $F < 1$; Segment by Contrast: $F(1, 14) = 1.09, p > .05$).

5.3. Vowels: summary and discussion

Experiment 1 tested whether speakers manipulate formant values when clarifying vowels, and whether the degree and direction of manipulation differs depending on how they think the target sound has been misheard. For the sounds /e/, /i/, /o/, and /u/, the only significant effect on $\Delta F1F2$ was that there was a larger Euclidean difference between the repetitions for /u/ than for the other segments. A possible explanation for this difference is that /u/ is often fronted in this dialect of English, leaving a lot of room for it to move back (i.e. lower on the F2 dimension). Indeed, the means for $\Delta F2$ for /u/ were negative in all conditions; however, t-tests showed that these differences were not significantly different than zero (nor were the differences in any of the other vowels). There was no significant difference in $\Delta F1$ or $\Delta F2$ for any of the vowels.

A follow-up experiment (Experiment 3) tested what was thought to be a "more minimal" contrast, with the prediction that it would be more likely to elicit contrastive manipulations. However, the effect of clarification on F1 and F2 again failed to reach significance. On the other hand, there was a significant effect of clarification on duration for the tense-lax contrast: /i/ was lengthened more than /ɪ/, contrasting with the control condition in which there was no significant difference in amount of lengthening between the two segments. This finding reflects the results from consonants showing that

speakers are not globally increasing the duration of segments regardless of contrast, but rather specifically exaggerating contrastive features.

In sum, subjects did not make systematic changes in vowel quality when repeating a word elicited by an incorrect guess which included a minimal vowel contrast. This is surprising in light of the above results for consonants, in which there is a rather robust effect of contrast. However, when the vowel contrast included a temporal element as well, as in the /i/-/ɪ/ distinction, subjects did manipulate duration in order to exaggerate the phonetic contrast.

6. General Discussion and Conclusion

6.1. Summary of main findings

This study has shown that speakers perform fine-grained manipulation of temporal elements of phonetic contrasts when trying to clarify misheard sounds. VOTs for word-initial voiceless stops have longer aspiration in a second repetition in which the word is being differentiated from a minimal pair beginning with a phonemically voiced stop. Similarly, voiced stops become 'more voiced' when they are being contrasted with a voiceless stop; stops which are prevoiced in the first repetition show a longer period of prevoicing in the second repetition, and phonemically voiced stops which are realized as voiceless unaspirated in the first repetition are more likely to be voiced in the second repetition than *vice versa*. Speakers made these modifications only in the Contrastive condition (when clarifying a word they thought had been misheard as a different word contrasting minimally in one sound), and not in the Open Response condition (when clarifying a word after the computer responded "What did you say?"). Furthermore, these modifications occurred only when the misheard segment contrasted in voicing with the target segment, and not in place or manner (though this effect must be interpreted with caution for voiceless stops, which did not show a significant interaction in the difference between conditions). For vowels, when clarifying sounds in the tense-lax pair /i/-/ɪ/, speakers lengthen /i/ more than /ɪ/, presumably in order to enhance the durational contrast between the two segments.

These results run counter to Ohala (1994), who found no contrastive effects in a similar paradigm, and since the experimental design was quite similar, one might wonder why the effects were not found in his study as well. One likely explanation is that a much more restricted set of stimuli was used in his study (five minimal pairs, all showing a /p/-/b/ contrast, not all of

which were real words (e.g. 'pid')), so a lack of power may have contributed to the null effect. The effects mirror more recent results from Maniwa et al. (2009), in which similar contrastive effects were found for fricatives, and extend them to stops and vowels.

On the other hand, there was no significant effect of clarification on spectral contrast (as measured in shift on the F1-F2 plane), either in the first study examining contrasts between /i e o u/ or in a follow-up study focusing on /i/ vs. /ɪ/. The only significant effect found for formants was a greater Euclidean difference between repetitions for /u/ than for the other vowels; however, the direction of this shift was not consistent between speakers. There was no significant change on either the F1 or F2 dimension for any of the vowels.

6.2. Global hyperarticulation vs. featural enhancement

The paradigm used in the current study did not elicit global "clear speech" effects: clarifications were neither significantly longer in duration nor higher in amplitude than the original read tokens, replicating results of Van Heuven (1994), who found a similar lack of effect of intensity or duration when speakers were clarifying specific segments of words. This was likely due to the fact that all tokens were read in a very clear speaking style, and speakers may have already reached their clear speech "ceiling" on the first repetition. At the same time, speakers *did* make contrast-specific phonetic modifications. The lack of global clear speech effects, taken together with the lack of effects in the Open Response condition, therefore provide evidence for the specific and local nature of the manipulations found in this study. Talkers are able to manipulate phonetic contrasts for purposes of clarity independently of global clear speech adjustments. These modifications are local to the contrast being targeted; they occur only when trying to clarify a minimal pair with the relevant featural contrast. Furthermore, the modifications are robust enough to be seen in a context of hyper-clear laboratory speech. Together, the findings suggest that the contrastive effects of clarification summarized above are just that: specific modifications made to enhance phonetic featural contrasts, as opposed to effects resulting from switching to a globally clearer speech style.

Although the types of VOT changes made by the speakers in this study (longer for voiceless and longer negative for voiced) were similar to those found in previous work examining hyperarticulation of English stops in clear speech (Picheny et al. 1986; Smiljanić and Bradlow 2008b) and in prosodic

strengthening environments (Choi 2003; Cho and Keating 2009), the process governing these modifications in this study may be different. While language-specific phonological inventories modulate variation in both global hyperarticulation and prosodic strengthening environments (e.g. Cho et al. 2011), the variation shown in the current data appears to be governed by an even narrower domain: the phonetic specification of the relevant contrast. This featural enhancement hypothesis predicts that speakers will clarify different contrasts in different ways; for example, when clarifying a /p/ that has been misheard as a /t/, speakers should exaggerate the features which are perceptually informative about place of articulation (e.g. formant transitions), a prediction which could be tested in future work. Supported by similar results from Maniwa et al. (2009), in which speakers modify acoustic features of fricatives differently depending on how the fricative has been misheard, the featural enhancement hypothesis seems to be the most plausible explanation for the current results. Further support comes from follow-up work using the same paradigm that found language-specific effects of VOT clarifications: Spanish speakers enhance VOT of voiceless stops differently than English speakers when clarifying the voicing contrast, by decreasing, instead of increasing, VOT, reflecting the differences in featural specifications of the voicing contrast in the different languages (Schertz 2012).

Although the language-specific effects found in different strengthening environments in previous work cited above provide evidence that the phonological system of a language governs the types of manipulations, they leave open the question of what level of awareness speakers have of these contrasts; is the method of hyperarticulation simply implicitly governed by the phonological system, without speakers' being aware of the contrasts, such that speakers no longer need to maintain a reference to the contrast itself? The current results contribute to work demonstrating that sub-phonemic representations are in fact available to, and used by, speech production mechanisms (e.g. Nielsen 2011). The fact that speakers selectively make phonetic modifications informed by the featural specification of a contrast suggests that speakers do indeed have access to the featural specification of the contrast online, and use it when it is necessary for a particular communicative situation.

6.3. Lack of spectral manipulations in vowels

The manipulations found for temporal contrasts in stops and vowels are easily described in terms of Lindblom's (1990) H&H theory: speakers increase discriminability in this particular communicative task by exaggerating the

phonetic differences in a given sound contrast. However, under this account, the lack of spectral manipulation of vowels is unexpected. This lack of effect is particularly striking given that English speakers and listeners rely more heavily on spectral information than durational information for the tense-lax distinction (e.g. Ainsworth 1972; Hillenbrand et al. 2000); if speakers exaggerate contrastive features when making clarifications, more enhancement of spectral than durational differences would be predicted. Instead, the speakers are doing the opposite: exaggerating the temporal, but not the spectral, contrast. The fact that speakers *are* modifying the durational aspect of the vowel contrast suggests that it is not the case that speakers are globally unwilling to to hyperarticulate vowels in clarifications, as may have been suggested by the results of Experiment 1, but that they did not perform the spectral manipulations which were expected based on our understanding of the nature of the vowel contrast.

Furthermore, although in the current work, the temporal features showed effects while spectral features did not, it does not appear to be driven by an inherent difference between the way speakers are able to enhance spectral vs. temporal features, since results from Maniwa et al. (2009) show spectral (as well as temporal) enhancement effects for fricatives. Therefore, the current results raise the question of why speakers do not enhance the first two formant frequencies, whereas listeners make use of these spectral variations when identifying vowel contrasts.

Adding further to the puzzle is the fact that vowel space expansion is found relatively consistently (Picheny et al. 1986; Bradlow et al. 2003; Erickson 2002; Whalen et al. 2004), suggesting that speakers are not generally reluctant to manipulate formant values for communicative purposes. However, this expansion appears to be a reflex of global hyperarticulation strategies, as opposed to local enhancement strategies found in the current work; furthermore, expansion usually occurs with a slower speech rate, so it could be attributed to the fact that speakers have more time to reach targets (the vowel space expansion found in the clear speech of the subjects in work by Krause and Braida (2004) disappeared when the speakers were asked to speak clearly but at a normal, i.e. faster, speaking rate). Furthermore, in contrast to the robust language-specific effects found for prosodic strengthening or clear speech effects in consonants, vowels appear to be spectrally enhanced in a similar manner in different languages, regardless of inventory size, suggesting that the enhancement does not have to do with how confusable the vowels are (Bradlow 2002; Smiljanić and Bradlow 2005; Cho et al. 2011), and

it appears that bilinguals use the same strategies to enhance vowel contrasts in both languages, even when the bilinguals have distinct language-specific perceptual weights defining the contrasts (Wassink et al. 2007; Granlund et al. 2012). In general, finding language- or contrast-specific formant manipulations has proven more elusive than temporal contrasts like VOT or vowel duration manipulations.

Under the assumption that F1 and F2 are in fact the relevant spectral features defining the vowel contrasts, one possibility for the lack of formant manipulation is that speakers have already reached the boundaries of the F1-F2 range for each vowel in the first reading of the word and are not willing and/or able to cross this boundary, whereas speakers will lengthen durations beyond the already hyperarticulated first reading. This difference could be driven by the fact that crossing the extremes of the F1-F2 space for a given vowel could be either articulatorily impossible at the edges of the vowel space (e.g. it may not be possible to lower the F1 of /i/ without producing a consonantal constriction) or perceptually suboptimal inside the vowel space (e.g. /ɪ/ might risk running into /ɛ/-space). On the other hand, a speaker exaggerating the temporal contrasts elicited in the current study would not encounter these issues: manipulating durations is not a problem from an articulatory point of view, nor does enhancing the durational contrast in vowels risk pushing them into a different sound category. Under this hypothesis, speakers would be expected not to enhance temporal contrasts if enhancement would move a segment closer to another category. This prediction could be tested with a more complex temporal contrast (e.g. the three-way vowel length contrast such as that found in Estonian).

Another possibility is that different speakers make different kinds of spectral manipulations in clarifications of vowels. Work by Smiljanić and Bradlow (2005), among others, suggests that individual talkers use different strategies to achieve vowel space expansion in clear speech, so the same might be true for contrast enhancement. However, if it is the case that speakers manipulate formants in order to enhance contrast in the same situation that they manipulate temporal properties of sounds, there should have still been a group effect of Condition (Contrastive vs. Open Response) on the Euclidean distance between the two repetitions, with a larger distance for the Contrastive than for the Open Response condition. Nevertheless, it would be worthwhile to look more closely at individual patterns of vowel clarification with a larger data set which would allow for statistical analysis of individual speakers' results.

6.4. Intelligibility of clarifications

Although the phonetic modifications were made in response to a very specific communicative context which would not be used in everyday speech, they provide information about the structural contrasts implicit in a speaker's phonological system and show that speakers *can* and *do* access this structure in specific situations. The current results leave open the question of why speakers make these modifications, and whether the modifications actually improve intelligibility. Most work looking at the intelligibility of "clear" vs. "casual" speech has found the clear version to be more intelligible in adverse communicative situations and/or in different listener populations; however, there has been less success in determining how and whether specific acoustic-phonetic features of clear speech map directly onto benefits in intelligibility (see Smiljanić and Bradlow (2009) for a review). For instance, while Ferguson and Kewley-Port (2007) found that speakers who produced an intelligibility benefit had more vowel space expansion than those who did not, there were large individual differences within the groups (see also Schum 1996; Perkell et al. 2002; Krause and Braida 2004). Furthermore, different features may provide intelligibility benefits in different sorts of adverse listening situations (Liu and Zeng 2006). Documentation of specific clarification strategies, such as those outlined in the current work, can provide a baseline for work testing the mapping between the phonetic modifications used by speakers to clarify speech and the intelligibility benefit (or lack thereof) of these modifications.

6.5. Conclusion

In sum, this study has shown that speakers perform specific, consistent, and local manipulation of temporal characteristics of English stops and vowels when clarifying misheard words, while spectral elements of vowel contrasts are not similarly enhanced. The extent and the nature of phonetic modification found in temporal contrasts depends on how the word has been misheard. Phonetic modifications to a given sound only occur when the sound itself had been misheard (as opposed to a different sound in the word being misheard, or a global misunderstanding of the word), showing that the scope of clarification can be local to a segment. Furthermore, contrastive features of segments are only manipulated when the relevant contrast is elicited by the incorrect guess, showing that speakers make use of subphonemic information online during production. The current findings therefore add to the growing body of work showing that speakers have the flexibility to modify acoustic-phonetic features online to respond to very specific communicative

tasks, and that these modifications are systematically constrained by their language-specific phonology.

Acknowledgements

I would like to thank Diana Archangeli, Edwin Maas, Francisco Torreira, and Natasha Warner for helpful discussion and feedback. Thanks also to the Associate Editor Taehong Cho and three anonymous reviewers for their comments and suggestions.

Appendix A. Stimuli for Experiment 1

The complete set of target words for Experiment 1 is listed below. Each target word was presented once (for consonants) or twice (for vowels) in the Contrastive condition, and once in the Open Response condition, and target words from both conditions were randomly distributed with filler items. Although the target stimuli skewed toward a small set of vowels (/i e o u/), fillers were designed to provide an equal number of different vowels throughout the experiment.

Consonant minimal pairs (words with voiced initial consonants are all listed as targets; the same pairs were used with the voiceless consonant token as the target):

Target	**Guess**	**Target**	**Guess**	**Target**	**Guess**
bade	paid	dale	tale	gale	kale
bane	pain	dame	tame	game	came
base	pace	dean	teen	gape	cape
baste	paste	deem	team	gave	cave
beak	peak	dime	time	gill	kill
beat	Pete	doom	tomb	goal	coal
beep	peep	dote	tote	goat	coat
bees	peas	doze	toes	goon	coon
		dune	tune	goop	coop

Vowel minimal triplets:

Target	**Guess**	**Guess**	**Target**	**Guess**	**Guess**
boot	beat	boat	boat	bait	boot
coop	keep	cope	cope	cape	coop
doom	deem	dome	dome	dame	doom
moon	mean	moan	moan	main	moon
tomb	team	tome	tome	tame	tomb
beat	bait	boot	bait	beat	boat
deem	dame	doom	cape	keep	cope
keep	cape	coop	dame	deem	dome
mean	main	moon	main	mean	moan
team	tame	tomb	pays	peas	pose

Appendix B. Stimuli for experiment 2

Voicing		Place		Manner	
Target	**Guess**	**Target**	**Guess**	**Target**	**Guess**
dale	tale	deem	beam	dale	sail
deal	teal	dean	bean	dale	hail
dean	teen	debt	bet	dash	sash
deem	team	dote	boat	dean	seen
deer	tier	dot	bot	debt	set
dime	time	dole	bowl	deem	seem
dire	tire	dies	buys	deep	seep
dock	tock	dot	got	deep	seep
dole	toll	dale	gale	dies	sighs
dote	tote	debt	get	dire	sire
doze	toes	dole	goal	dot	sot
duck	tuck	dote	goat	dues	sues
dune	tune	doze	goes	dumb	sum
dusk	tusk	dune	goon	dune	soon
tale	dale	tick	pick	tag	zag
teal	deal	ties	pies	tale	nail
teen	dean	tin	pin	tame	name
team	deem	toll	pole	tap	zap
tier	deer	tool	pool	tape	nape
time	dime	top	pop	test	zest
tire	dire	toes	pose	tick	nick
tock	dock	take	cake	tight	night
toll	dole	tape	cape	tine	nine
tote	dote	toll	coal	ting	zing
toes	doze	tote	coat	tip	zip
tuck	duck	tune	coon	tone	zone
tune	dune	top	cop	tune	noon
tusk	dusk	tan	can	two	zoo

Appendix C. Stimuli for experiment 3

/i/-/ɪ/		/ɪ/-/i/		/i/-/u/		/ɪ/-/ɛ/	
Target	**Guess**	**Target**	**Guess**	**Target**	**Guess**	**Target**	**Guess**
deep	dip	dip	deep	beast	boost	pick	peck
beat	bit	bit	beat	beat	boot	chick	check
seat	sit	sit	seat	heat	hoot	flick	fleck
peak	pick	pick	peak	deep	dupe	trick	track
heap	hip	hip	heap	heap	hoop	bit	bet
meet	mitt	mitt	meet	keep	coop	sit	set
cheap	chip	chip	cheap	seep	soup	mitt	met
seek	sick	sick	seek	speak	spook	itch	etch
seep	sip	sip	seep	steep	stoop	sits	sets
feet	fit	fit	feet	sheet	shoot	fist	fest
sheep	ship	ship	sheep	meet	moot	miss	mess
cheek	chick	chick	cheek	geese	sail	disk	desk

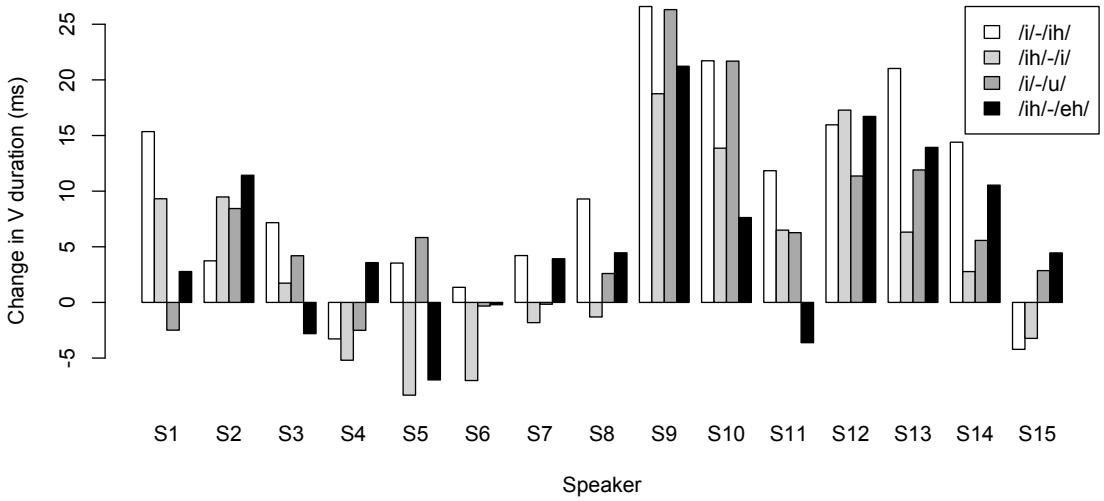
Appendix D. Individual speaker results for Experiment 1

Table D.4: Means and results of one-way ANOVAs comparing mean differences in ΔVOT between the Contrastive and Open Response conditions for each speaker individually. Asterisks indicate statistically significant differences.

	Voiceless		**Voiced**	
Speaker	**Mean difference (ms)**	**ANOVA results**	**Mean difference (ms)**	**ANOVA results**
1	13	$^*F(1,24) = 2.68, p < .01$	-1	$F < 1$
2	6	$F < 1$	-18	$F < 1$
3	24	$^*F(1,25) = 2.79, p < .01$	-1	$F < 1$
4	4	$F(1,25) = 1.11, p > .05$	-4	$F < 1$
5	11	$F(1,25) = 1.90, p < .1$	-10	$F < 1$
6	0	$F < 1$	-38	$^*F(1,25) = 2.05, p < .05$
7	11	$^*F(1,25) = 3.12, p < .01$	0	$F < 1$
8	3	$F < 1$	-29	$F(1,25) = 1.20, p > .05$
9	10	$F(1,25) = 1.16, p > .05$	-24	$F(1,24) = 1.06, p > .05$
10	-8	$F < 1$	-24	$F(1,24) = 1.08, p > .05$
11	-3	$F(1,25) = 1.04, p > .05$	-1	$F < 1$
12	3	$F < 1$	-19	$F(1,25) = 1.05, p > .05$

Appendix E. Individual speaker results for Experiment 3

Figure E.6: Results from individual speakers on vowel duration changes in Experiment 3. The four results for each speaker are the two tense-lax pairs (target /i/ with incorrect guess /ɪ/ and target /ɪ/ with incorrect guess /i/) and the two control pairs (target /i/ with incorrect guess /u/ and target /ɪ/ with incorrect guess /ɛ/).



Ainsworth, W., 1972. Duration as a cue in the recognition of synthetic vowels. The Journal of the Acoustical Society of America 51 (2B), 648–651.

Baese-Berk, M., Goldrick, M., 2009. Mechanisms of interaction in speech production. Language and Cognitive Processes 24 (4), 527–554.

Boersma, P., Weenink, D., 2011. Praat: doing Phonetics by computer, version 5.3: http://www.praat.org.

Bradlow, A., 2002. Confluent talker- and listener-oriented forces in clear speech production. In: Gussenhoven, C., Warner, N. (Eds.), Laboratory Phonology 7. Mouton de Gruyter, Berlin/New York, pp. 241–273.

Bradlow, A., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. Journal of Speech, Language, and Hearing Research 46 (1), 80–97.

Bradlow, A., Torretta, G. M., Pisoni, D. B., 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Communication 20 (3), 255–272.

Cho, T., 2005. Prosodic strengthening and featural enhancement: evidence from acoustic and articulatory realizations of /a, i/ in English. Journal of the Acoustical Society of America 117 (6), 3867–3878.

Cho, T., Keating, P., 2001. Articulatory and acoustic studies of domain-initial strengthening in Korean. Journal of Phonetics 29 (2), 155–190.

Cho, T., Keating, P., 2009. Effects of initial position versus prominence in English. Journal of Phonetics 37 (4), 466–485.

Cho, T., Lee, Y., Kim, S., 2011. Communicatively driven versus prosodically driven hyper-articulation in Korean. Journal of Phonetics 39 (3), 344–361.

Cho, T., McQueen, J., 2005. Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress. Journal of Phonetics 33 (2), 121–157.

Choi, H., 2003. Prosody-induced acoustic variation in English stop consonants. Proceedings of the 15th International Congress of Phonetic Sciences, 2661–2664.

Cooke, M., Lu, Y., 2010. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. Journal of the Acoustical Society of America 128 (4), 2059–2069.

de Jong, K., 1995. On the status of redundant features. In: Connell, B., Arvaniti, A. (Eds.), Laboratory Phonology 4: Phonology and Phonetic Evidence. pp. 68–86.

de Jong, K., 2004. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. Journal of Phonetics 32 (4), 493–516.

de Jong, K., Zawaydeh, B. A., 2002. Comparing stress, lexical focus, and segmental focus: Patterns of varation in Arabic vowel duration. Journal of Phonetics 30 (1), 53–75.

Erickson, D., 2002. Articulation of extreme formant patterns for emphasized vowels. Phonetica 59 (2-3), 134–149.

Escudero, P., Boersma, P., Rauber, A., Bion, R., 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. Journal of the Acoustical Society of America 126 (3), 1379–1393.

Ferguson, S., Kewley-Port, D., 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. Journal of Speech, Language, and Hearing Research 50 (5), 1241–1255.

Ferguson, S., Poore, M., Shrivastav, R., Kendrick, A., McGinnis, M., Perigoe, C., 2010. Acoustic correlates of reported clear speech strategies. Journal of the Academy of Rehabilitative Audiology 43, 45–64.

Garnier, M., Henrich, N., Dubois, D., 2010. Influence of sound immersion and communicative interaction on the Lombard effect. Journal of Speech, Language, and Hearing Research 53 (3), 588–608.

Granlund, S., Hazan, V., Baker, R., 2012. An acoustic-phonetic comparison of the clear speaking styles of Finnish-English late bilinguals. Journal of Phonetics 40 (3), 509–520.

Hazan, V., Baker, R., 2011. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. Journal of the Acoustical Society of America 130 (4), 2139–2152.

Hillenbrand, J. M., Clark, M. J., Houde, R. A., 2000. Some effects of duration on vowel recognition. The Journal of the Acoustical Society of America 108 (6), 3013–3022.

Kang, K.-H., 2009. Clear speech production and perception of Korean stops and the sound change in Korean stops. Ph.D. thesis, University of Oregon.

Kang, K.-H., Guion, S., 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. Journal of the Acoustical Society of America 124 (6), 3909–3917.

Kessinger, R., Blumstein, S., 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. Journal of Phonetics 25 (2), 143–168.

Kirov, C., Wilson, C., 2012. Specificity of online variation in speech production, poster presentation at the Annual Meeting of the Linguistic Society of America, Portland, OR.

Koopmans-van Beinum, F. J., 1980. Vowel contrast reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions. Ph.D. thesis, University of Amsterdam.

Krause, J. C., Braida, L. D., 2004. Acoustic properties of naturally produced clear speech at normal speaking rates. Journal of the Acoustical Society of America 115 (1), 362–378.

Kuzla, C., Ernestus, M., 2011. Prosodic conditioning of phonetic detail in German plosives. Journal of Phonetics 39 (2), 143–155.

Ladefoged, P., Kameny, I., Brackenridge, W., 1976. Acoustic effects of style of speech. Journal of the Acoustical Society of America 59 (1), 228–231.

Levow, G., 1999. Understanding recognition failurers in spoken corrections in human–computer dialogue. In: Proceedings of the ESCA Workshop on Dialogue and Prosody. Eindhoven, The Netherlands, pp. 736–742.

Lindblom, B., 1990. The status of phonetic gestures. In: Mattingly, I. G., Studdert-Kennedy, M. (Eds.), Modularity and the motor theory of speech perception. Psychology Press, pp. 7–31.

Lisker, L., Abramson, A., 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 384–422.

Liu, S., Zeng, F., 2006. Temporal properties in clear speech perception. The Journal of the Acoustical Society of America 120 (1), 424–432.

Maniwa, K., Jongman, A., Wade, T., 2009. Acoustic characteristics of clearly spoken English fricatives. Journal of the Acoustical Society of America 125 (6), 3962–3973.

Moon, S.-J., Lindblom, B., 1994. Interaction between duration, context, and speaking style in English stressed vowels. Journal of the Acoustical Society of America 96 (1), 40–55.

Nielsen, K., 2011. Specificity and abstractness of VOT imitation. Journal of Phonetics 39 (2), 132–142.

Ohala, J., 1994. Acoustic study of clear speech: A test of the contrastive hypothesis. International Symposium on Prosody.

Oviatt, S., Bernard, J., Levow, G.-A., 1998a. Linguistic adaptations during spoken and multimodal error resolution. Language and Speech 41, 419–442.

Oviatt, S., Levow, G., Moreton, E., MacEachern, M., 1998b. Modeling global and focal hyperarticulation during human–computer error resolution. The Journal of the Acoustical Society of America 104 (5), 3080–3098.

Perkell, J. S., Zandipour, M., Matthies, M. L., Lane, H., 2002. Economy of effort in different speaking conditions 1: A preliminary study of intersubject differences and modeling issues. Psychology Faculty Publications.

Picheny, M. A., Durlach, N. I., Braida, L., 1986. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. Journal of Speech, Language, and Hearing Research 29, 434–446.

Schertz, J., 2012. Featural enhancement in clarification of misheard stops in Spanish. Journal of the Acoustical Society of America 132 (3), 1937.

Schum, D. J., 1996. Intelligibility of clear and conversational speech of young and elderly talkers. Journal of the American Academy of Audiology 7 (3), 212–218.

Shattuck-Hufnagel, S., Turk, A., 1996. A prosody tutorial for investigators of auditory sentence processing. Journal of Psycholinguistic Research 25, 193–247.

Smiljanić, R., Bradlow, A., 2005. Production and perception of clear speech in Croatian and English. Journal of the Acoustical Society of America 118 (3), 1677–1688.

Smiljanić, R., Bradlow, A., 2008a. Stability of temporal contrasts across speaking styles in English and Croatian. Journal of Phonetics 36 (1), 91–113.

Smiljanić, R., Bradlow, A., 2008b. Temporal organization of English clear and conversational speech. Journal of the Acoustical Society of America 124 (5), 3171–3182.

Smiljanić, R., Bradlow, A., 2009. Speaking and hearing clearly: talker and listener factors in speaking style changes. Language and Linguistics Compass 3 (1), 236–264.

Stent, A., Huffman, M., Brennan, S., 2008. Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. Speech Communication 50 (3), 163–178.

Uther, M., Knoll, M. A., Burnham, D., 2007. Do you speak E-NG-L-I-SH? A comparison of foreigner-directed and infant-directed speech. Speech Communication 49 (1), 2–7.

Van Heuven, V. J., 1994. What is the smallest prosodic domain? In: Keating, P. (Ed.), Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III. Cambridge University Press, pp. 76–97.

Warner, N., 2011. Reduction. In: van Oostendorp, M., Ewen, C. J., Hume, E., Rice, K. (Eds.), The Blackwell Companion to Phonology. Blackwell, pp. 1866–1891.

Wassink, A., Wright, R., Franklin, A., 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. Journal of Phonetics 35, 363–379.

Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., Iskarous, K., 2004. Vowel production and perception: Hyperarticulation without a hyperspace effect. Language and Speech 47 (2), 155–174.