Acoustic Correlates of Reported Clear Speech Strategies

Sarah Hargus Ferguson and Meredith A. Poore
University of Kansas

Rahul Shrivastav
University of Florida

This study examined whether clear speech acoustic modifications agreed with talkers' descriptions of strategies they used when speaking clearly during recording of the Ferguson Clear Speech Database (Ferguson, 2004). For the strategies "spoke more loudly" and "put extra stress on content words," acoustic differences between clear and conversational speech differed significantly among talkers who did and did not report the strategy. No differences were observed, however, between talkers who did and did not report that they "enunciated more carefully" or "spoke more slowly." Results are discussed in terms of talker awareness and control of specific aspects of speech production as well as implications for clear speech training.

Clear speech, the speech a talker produces when he or she is told to speak as though their communication partner has difficulty understanding, is usually more intelligible than ordinary conversational speech. A clear speech intelligibility benefit has been observed for sentences presented to listeners with normal hearing (Bradlow & Bent, 2002; Gagné, Querengesser, Folkeard, Munhall, & Masterson, 1995; Krause & Braida, 2002; Liu, Del Rio, Bradlow, & Zeng, 2004; Payton, Uchanski, & Braida, 1994; Smiljanic & Bradlow, 2005; Uchanski, Choi,

Sarah Hargus Ferguson and Meredith A. Poore, Department of Speech-Language-Hearing: Sciences and Disorders, University of Kansas; Rahul Shrivastav, Department of Speech, Language, and Hearing Sciences, University of Florida.

Sarah Hargus Ferguson is now at the Department of Communication Sciences and Disorders, University of Utah.

Correspondence concerning this article should be addressed to Sarah Hargus Ferguson, Department of Communication Sciences and Disorders, 390 South 1530 East, Salt Lake City, Utah, 84112. Telephone: 801-585-6776. Fax: 801-585-7995. E-mail: sarah.ferguson@hsc.utah.edu

Braida, & Durlach, 1996), listeners with hearing loss (Helfer, 1998; Payton et al., 1994; Picheny, Durlach, & Braida, 1985; Schum, 1996; Uchanski et al., 1996), cochlear implant users (Liu et al., 2004), and listeners with auditory neuropathy (Zeng & Liu, 2006). Results have been more mixed for vowel materials. For example, Ferguson and Kewley-Port (2002) found a significant clear speech vowel intelligibility benefit for young listeners with normal hearing but not for older listeners with hearing loss, while Ferguson and Lee (2006) found a clear speech vowel intelligibility benefit for just one of three cochlear implant users. Finally, the clear speech benefit has been shown to vary significantly among talkers (Ferguson, 2004; Gagné, Masterson, Munhall, Bilida, & Querengesser, 1994).

Clear speech differs from conversational speech on a wide array of acoustic dimensions. Perhaps the most robust finding has been that clear speech is slower than conversational speech (Bradlow, Kraus, & Hayes, 2003; Ferguson & Kewley-Port, 2002, 2007; Krause & Braida, 2004; Liu et al., 2004; Picheny, Durlach, & Braida, 1986; Smiljanic & Bradlow, 2005). Clear speech also has a wider fundamental frequency (F0) range, fewer unreleased consonant bursts (Bradlow et al., 2003; Krause & Braida, 2004; Picheny et al., 1986), and a higher temporal modulation index (Krause & Braida, 2004; Liu et al., 2004) than conversational speech. Finally, vowels in clear speech occupy a greater overall vowel space (Bradlow et al., 2003; Ferguson & Kewley-Port, 2002, 2007; Krause & Braida, 2004; Liu et al., 2004; Picheny et al., 1986) and have a greater amount of dynamic formant movement (Ferguson & Kewley-Port, 2002, 2007) than vowels in conversational speech.

Several studies have sought to determine which clear speech acoustic changes are actually responsible for the superior intelligibility of clear speech. One approach has been to modify conversational speech acoustically to approximate clear speech on some dimension. Such studies have demonstrated that increasing the consonant-vowel intensity ratio leads to improved intelligibility for consonant-vowel syllables (Gordon-Salant, 1986, 1987), while slowing conversational speech does not improve sentence intelligibility (Liu & Zeng, 2006; Nejime & Moore, 1998; Picheny, Durlach, & Braida, 1989; Uchanski et al., 1996). Another approach is to collect both perceptual and acoustic data from naturally-produced clear and conversational speech. For example, Bradlow et al. (2003) compared two talkers in terms of clear speech intelligibility and acoustic changes and found that the talker who made the greatest speaking rate reduction showed the greatest clear speech intelligibility benefit. Ferguson and Kewley-Port (2007) examined clear and conversational vowels produced by 12 talkers – 6 for whom listeners with normal hearing had shown a large clear speech vowel intelligibility benefit, and 6 who produced no vowel intelligibility benefit when speaking clearly. Their results suggested that a greater vowel space expansion and greater vowel duration increases were associated with a greater vowel intelligibility benefit in clear speech.

The question of "what makes clear speech clear" is interesting for a number of reasons. Picheny and his colleagues (1985) hoped to identify clear speech acoustic characteristics that could be incorporated into hearing aid design in order to improve speech intelligibility. Another application of clear speech research is to improve training programs for the frequent communication partners of individuals with hearing loss. Although Schum (1996) concluded that these partners can, "without much instruction or practice, . . . be expected to be able to improve their intelligibility at will" (p. 216), Gagné et al. (1994) and Ferguson (2004) showed that talkers vary widely in their ability to produce effective clear speech when given general instructions. Ferguson in particular noted that certain talkers may require more specific instructions. Knowing which acoustic changes are actually beneficial to listeners with hearing loss is crucial for the development of these specific instructions.

The idea of developing specific clear speech instructions for the communication partners of individuals with hearing loss hinges upon the assumption that talkers are conscious of what they are doing during speech production and are able to make both global and fine-grained acoustic changes to their speech. This assumption lies at the heart of Lindblom's H&H (Hyper- and Hypospeech) theory (1990), which claims that talkers control the degree to which output constraints (assuring speech is discriminable for the listener) or system constraints (tending toward motor economy) dominate speech production. There is ample evidence that talkers are able to change at least some global aspects of their speech when asked to do so. Studies have shown that both normal and disordered talkers produce reliable voice amplitude and speaking rate changes when instructed to speak more loudly and/or more slowly (Fourakis, 1991; Kleinow, Smith, & Ramig, 2001; Tjaden & Wilding, 2004; Turner & Weismer, 1993). However, there are theoretical reasons to suspect that talkers would not have conscious awareness of the details of their speech production. Since speech production is a learned motor skill, it can be considered procedural knowledge which is not available for introspection (e.g., Towell, Hawkins, & Bazergui, 1996). Fluent speech production has also been described as an automatic process as opposed to a controlled one (e.g., McLaughlin, Rossman, & McLeod, 1983), again arguing against awareness and conscious control. Support for these ideas comes from Watson and Hixon (1985), who found that trained opera singers' descriptions of their respiratory behaviors while singing bore little resemblance to what they actually did.

To explore the extent to which talkers are aware of what they are doing during the production of clear speech, the present study analyzed the results of talker interviews performed at the time the Ferguson Clear Speech Database (Ferguson, 2004) was recorded. Among other questions, talkers were asked about how the speech they produced during the clear speech recording session differed from their everyday conversational speech. These strategies were tallied and acoustic analyses were performed to assess the degree to which talkers were accurate in their descriptions of their speaking behaviors. Specifically, the current study asked two questions: What do talkers think they are doing when they speak clearly? And can we find acoustic evidence that talkers did or did not use these strategies?

METHOD

The Ferguson Clear Speech Database

The database consists of clear and conversational speech produced by 41 talkers. The talkers, aged 18 to 45 years, reported normal hearing, speech, and language. All were from the South Midland region of the United States (Labov, Ash, & Boberg, 2006); none had any training in phonetics or diction or any professional speaking experience.

Each talker produced 188 sentences in each speaking style. Fourteen sentences in each style (a different list for each style) were selected from the Central Institute for the Deaf (CID) Everyday Sentences Test (Davis & Silverman, 1978). The remaining sentences were constructed using a list of 16 neutral sentence frames with keywords inserted. In /bVd/ sentences, the keywords were the 10 vowels (/i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/) in /bVd/ context, with 7 unique sentence frames for each word. In consonant-vowel-consonant (CVC) sentences, the keywords for conversational and clear speech were lists 3 and 4, respectively, of the Northwestern University Auditory Test No. 6 (Tillman & Carhart, 1966), plus two additional words per style containing the vowel /ʊ/. Each CVC word was recorded in two unique sentence frames.

Recordings were completed in two sessions at least 1 day apart. In the first session, all talkers were instructed to read the test sentences in a manner approximating the way they spoke in everyday conversation. In the second session, they were instructed to say the test sentences as they would if they were talking to a person with hearing loss. Instructions were given orally and also printed on a card placed prominently in the recording booth. The text of the instructions for the conversational speech session read as follows:

> In this experiment, I am looking at how people speak. For the first part, I want you to say the sentences as you would in everyday, normal conversation. You'll have a chance to practice this normal conversational style before we actually start recording. It is important that your speech be as much like your normal conversational style as possible. (Ferguson, 2002, p. 136)

For the clear speech session, the instructions were

> In this part of the experiment, I am looking at how people adjust their speech when they talk to a hard-of-hearing person. For this second part, I want you to say the sentences as you would if you were talking to a person who has a hearing loss. You'll have a chance to practice this speaking style before we actually

> start recording. It is important that you speak clearly, so that a hearing-impaired person would be able to understand you. (p. 137)

For both sessions, the text shown above was followed by additional details about the recording procedures, as well as guidance for pronouncing the non-word "bood" (/bʊd/). Talkers were given a list of 16 practice sentences (see Appendix A) for rehearsing the particular speaking style prior to the recordings. This list contained all 16 sentence frames and an example of each of the vowel keywords. In the conversational speech session, talkers received feedback during rehearsal. This feedback typically consisted of comments about how conversational the speech sounded, along with instructions to repeat the practice list until the experimenter judged the speech to be sufficiently conversational. No such feedback nor any additional instructions about how to produce clear speech were given in the clear speech session.

Clear Speech Strategies

Upon completion of the clear speech recordings, talkers were interviewed briefly. One of the questions was, "Did you do anything different with your speech in this session, compared to the first session?" All talkers answered affirmatively, and each was asked to describe the specific changes that he or she made. First, the talkers' descriptions of what they did during the clear speech recording session were examined carefully and the number of talkers reporting various strategies was tallied. The strategies are reported below. Next, for the four most frequently-reported strategies, talkers were divided into two groups: those who had reported the strategy (reporters) and those who had not (non-reporters). The reporters and non-reporters of each strategy were then compared in terms of acoustic measures appropriate to the given strategy. The results of each acoustic analysis were submitted to a two-way repeated measures analysis of variance (ANOVA) with one within-subjects factor (speaking style: clear vs. conversational) and one between-subjects factor (reporting status: did or did not report). Acoustic analysis methods will be described below as the results for each specific clear speech strategy are presented.

RESULTS

Clear Speech Strategies

Though specific descriptions varied, the clear speech strategies that the talkers reported fell into 11 broad categories. Eight of these strategies were reported by 2 or more talkers; three were each reported by only 1 talker. The strategies and the number of times each was reported are shown in Table 1. All but 1 of the talkers reported two or more strategies; the largest number of reported strategies was five. As seen in Table 1, the most frequently-reported strategy was pronouncing or enunciating the sentences more carefully ($n = 29$). Also frequently-

Table 1

Reported Clear Speech Strategies

Strategy	# of talkers reporting
Enunciated more carefully	29
Spoke more loudly	20
Spoke more slowly	20
Placed extra stress on content words	15
Tried to be easier to lipread	9
Was careful to enunciate consonants	8
Used more pauses between words	6
Kept more even pitch and volume	2
Placed extra stress on pronouns	1
Imagined preceding the sentence with, "I *said*"	1
Used a higher voice pitch	1

reported were speaking more loudly (n = 20), speaking more slowly (n = 20), and placing extra stress on content words (n = 15). Eight talkers reported being careful to enunciate consonants, particularly final consonants. Six talkers reported using more pauses in clear speech than in conversational speech. Nine talkers reported making an effort to be easier to lipread in the clear speech condition.

Acoustic Correlates of Reported Strategies

Enunciated More Carefully

Two measures, vowel space perimeter and final consonant release, were used to compare talkers who did (n = 29) and did not (n = 12) report enunciating more carefully in clear speech.

Vowel space perimeter. As part of an ongoing acoustic study of the Ferguson database, vowel space perimeter was determined for each talker in each style using steady-state first and second formant frequency (F1 and F2) values extracted from linear predictive coding formant tracks generated for each vowel token using WaveSurfer (Sjölander & Beskow, 2006). The values for each vowel in each style were determined for each talker by averaging over two tokens. Vowel space perimeter was then calculated as the sum of the Euclidean distances between the vowels /i/ and /æ/, /æ/ and /a/, /a/ and /u/, and /u/ and /i/ in Barks (Traunmüller, 1990). A total of 82 perimeter values, one for each talker in each style, were the dependent variables for the two-way ANOVA.

The main effect of speaking style was significant, $F(1, 39) = 27.77$, $p < .01$, with an average vowel space perimeter of 12.9 Barks in conversational speech and 13.9 Barks in clear speech. Neither the main effect of reporting status, $F(1, 39) = 1.4$, $p = .24$, nor the interaction between speaking style and reporting status, $F(1, 39) = 1.24$, $p = .27$, was significant. Though talkers who reported enunciating more carefully in clear speech showed a larger average increase in the size of the vowel space (1.2 Barks) than those who did not report this strategy (0.82 Barks), the variability within each group of talkers was extremely large, with standard deviations of over 1.1 Barks for the difference between clear and conversational speech.

Final consonant release. The second measure, final consonant release, was determined by examining the final /d/ of the /bVd/ words produced by the talkers in each style in one of six sentence contexts:

1. His/her mother calls him/her ______ at home.
2. Jean bought a ______ at the store.
3. I think the word ______ is hard for kids to say.
4. He looked for a ______ but couldn't find one.
5. Her/his friends called her/him ______ back in school.
6. He said ______ but he meant box.

These sentence frames were selected from the larger set of 16 frames based on the phoneme following the /bVd/ keyword. In the first three sentences, the keyword is followed by a non-nasalized vowel, a context which favors /d/ being released. In the last three, the keyword is followed by a stop, which tends not to favor /d/ release. Final consonant release was extremely difficult to judge in the other 10 sentence frames due to coarticulation with the following phoneme, which in many cases was a vowel followed by /n/ (see Appendix A). A total of 28 /bVd/ words, of which half were followed by stops and half by vowels, were assessed for each talker in each style.

For each token, the final /d/ was judged as either a strong stop or a weak stop based on listening to the token as well as viewing its waveform and spectrogram. A final /d/ was judged as strong (or released) if a clear /d/ burst was both audible and visible, and weak if the /d/ was either unreleased or flapped and no burst was visible. Judgments were performed by two independent judges. First, each judge assessed all the tokens for a given talker and style separately. The first author compared their judgments and made a list of tokens on which the judges disagreed. The two judges then re-judged these items separately. The first author again compared the judgments and made a final decision for any remaining tokens for which disagreement persisted. Finally, the percentage of tokens in which the final /d/ was judged a strong stop was determined for each talker in each style and submitted to the two-way ANOVA.

The main effect of speaking style was highly significant, $F(1, 39) = 100.2$, $p < .001$. Talkers produced a strong stop for the final /d/ consonant in 75% of the tokens analyzed in clear speech and just 28% of the tokens analyzed in conversational speech. Neither the main effect of reporting status nor the speaking style by reporting status interaction were significant, $F(1, 39) = 3.45$, $p = .07$ and F(1,

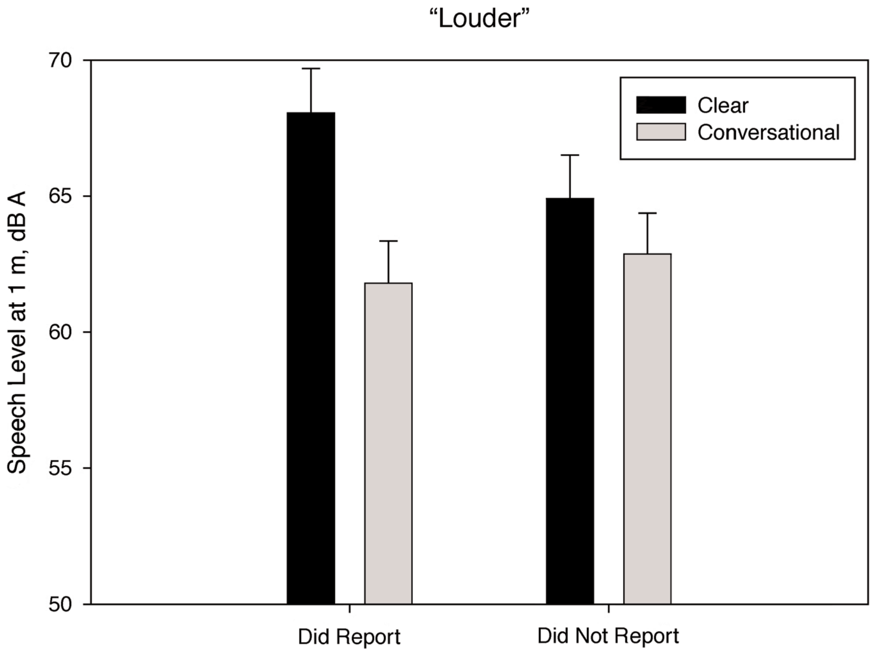

Figure 1. Average speech level at 1 m in clear and conversational speech for talkers who did and did not report speaking more loudly as a clear speech strategy. Error bars indicate 95% confidence intervals.

39) = 0.27, $p = .6$, respectively. Talkers who reported enunciating more carefully increased the percentage of strong stops from 34% in conversational speech to 83% in clear speech; talkers who did not report this strategy increased their percentage of strong stops from 23% to 67%.

Spoke More Loudly

Level measurements made during the Ferguson database recording were used to assess whether talkers who reported speaking more loudly in clear speech ($n = 20$) made greater level increases than talkers who did not report this strategy ($n = 21$). About halfway through each recording session, the first author entered the recording booth and instructed the talker to read the list of practice sentences using the exact same manner and level he/she had been using for the test sentences. As the talker read the list, the experimenter estimated the average overall speech intensity in dB A using a Realistic sound-level meter (No. 33-2050) held 1 m from the talker's lips. The speech levels recorded for each talker in each speaking style were used as the dependent variable in the two-way ANOVA.

The main effect of speaking style on speech level was significant, $F(1, 39)$

= 71.50, $p < .01$. Talkers spoke with greater intensity in clear speech (66 dB A) than in conversational speech (62 dB A). Although the main effect of reporting group was not significant, $F(1, 39) = 1.13$, $p = .29$, the interaction between reporting group and speaking style was, $F(1, 39) = 18.34$, $p < .01$. This interaction is illustrated in Figure 1, which shows that the clear speech level effect (difference in level between clear and conversational speech) was greater for talkers who reported speaking more loudly in clear speech (6.25 dB) than for talkers who did not report this strategy (2 dB).

Spoke More Slowly

To assess the extent to which talkers who did and did not report speaking more slowly in clear speech ($n = 20$ and 21, respectively) actually did so, speaking rate was measured using the CID Everyday Sentences. The lists of sentences that were recorded in clear and conversational speech are shown in Appendices B and C, respectively. Speaking rate was determined for each talker in each speaking style by dividing the number of words in each list of 14 sentences (73 in clear and 72 in conversational) by the duration of the sentence list (including pauses between the sentences) and submitted to the two-way repeated-measures ANOVA. The main effect of speaking style was significant, $F(1, 39) = 100.18$, $p < .01$, with talkers using a lower speaking rate in clear speech (111 words per minute) than in conversational speech (155 words per minute). The main effect of reporting group was not significant, $F(1, 39) = 1.49$, $p = .23$, nor was the interaction between speaking style and reporting group, $F(1, 39) = 0.41$, $p = .53$. The talkers who reported speaking more slowly in clear speech reduced their speaking rate by 29%; non-reporters reduced their speaking rate by 27%.

Placed Extra Stress on Content Words

This strategy (reported by 15 talkers; not reported by 26) was examined by measuring acoustic correlates of word-level stress for content words relative to the sentence containing them. Specifically, duration, amplitude, and pitch measures were carried out on three sentences for each talker in each speaking style. Measures were designed with the assumption that the magnitude of each measure would be greater in clear speech than in conversational speech, and that the difference between the speaking styles would be greater for talkers who reported placing extra stress on content words than it would be for talkers not reporting this strategy. Three sentences were selected for this analysis based on their similar overall prosodic pattern (main content words are underlined):

1. Vera put the bed near the table.
2. Jean bought a bed at the store.
3. Please put the bed next to the dresser.

Duration. One way to increase the stress level of a word within a sentence is

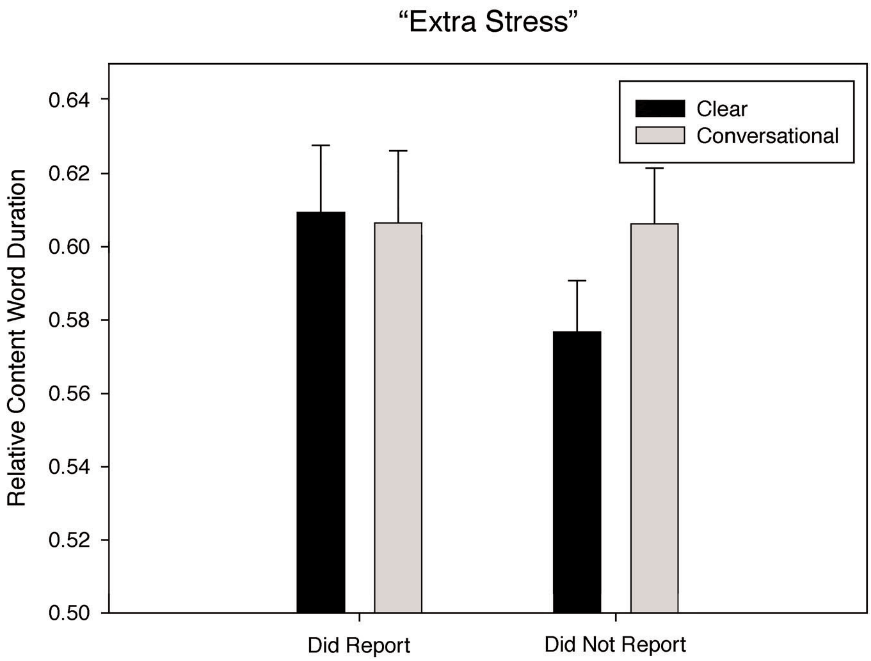

Figure 2. Duration of three main content words relative to total sentence duration in clear and conversational speech for talkers who did and did not report placing extra stress on content words as a clear speech strategy. Error bars indicate 95% confidence intervals.

to increase the duration of that word. For each sentence, both the sentence duration and the duration of each of the three main content words (e.g., "Vera," "bed," and "table" in Sentence 1) were measured. We assumed that content words receiving extra stress would be longer than content words not receiving extra stress, but also noted that content words would be longer in clear speech than in conversational speech. To determine whether content words were lengthened preferentially, therefore, we calculated relative content word duration by summing the durations of the three content words and dividing this sum by the duration of the sentence. If content words were lengthened preferentially, then the relative content word duration should be greater in clear speech than in conversational speech. If such preferential lengthening did not occur, the relative content word duration should be the same in the two speaking styles.

Relative content word duration was determined for each of the three sentences for each talker in each style and submitted to a two-way repeated measures ANOVA. Average relative content word durations in each speaking style for each group of talkers are shown in Figure 2. The main effect of speaking style was significant, $F(1, 121) = 6.151$, $p < .02$, but not in the expected direction. The relative content word duration was actually greater in *conversational* speech (.606) than in clear speech (.593). The main effect of reporting group was not significant, $F(1, 121) = 2.212$, $p = .14$, but as is apparent in Figure 2, the interaction between speaking style and reporting group was, $F(1, 121) = 9.395$, $p < .004$. Only talkers who did not report placing extra stress on content words showed a greater relative content word duration in conversational speech than in clear speech. Talkers who reported this strategy, in contrast, showed the same relative content word duration in each speaking style.

To further explore the difference between reporters and non-reporters of the extra stress strategy, a series of clear versus conversational duration ratios, or "stretch factors," was calculated. The stretch factor for the three content words was calculated for each sentence by dividing the summed content word duration in clear speech by the summed content word duration in conversational speech. To compute stretch factors for the function words and pauses for each sentence, the summed content word duration was subtracted from the sentence duration and the "remainder" duration for clear speech was divided by the remainder duration for conversational speech. Average stretch factors for each talker group are shown in Table 2. Stretch factors for the content words and for the remainder of the sentence were submitted to separate one-way ANOVAs. While talkers who did and did not report placing extra stress on keywords treated content words similarly, $F(1, 122) = 1.254$, $p = .27$, they differed significantly in the extent to

Table 2

Clear Versus Conversational Duration Ratios for Content Words and for Function Words and Pauses for Talkers Who Did and Did Not Report "Extra Stress"

	Content words	Remainder
Reported ($n = 26$)	1.39	1.39
Did not report ($n = 15$)	1.45	1.75

which they lengthened the function words and pauses, $F(1, 122) = 7.985$, $p < .01$. While non-reporters lengthened function words and pauses to a greater extent than they lengthened content words, reporters lengthened all parts of the sentence to the same degree. Thus, talkers who did and did not report placing additional stress on content words used different strategies when they slowed down to produce clear speech.

Pitch and amplitude. The metrics for pitch and amplitude, F0 change and dB change, were inspired by a study of emphatic stress production by talkers with dysarthria (Wang, Kent, Duffy, & Thomas, 2005). Both were calculated by examining the difference between the primary keyword of each stress-pattern sentence (*bed*) and the article that preceded it (*a* or *the*). Pitch and amplitude tracks

Table 3

Means and Standard Deviations for Pitch and Amplitude Change Measures in Clear and Conversational Speech for Talkers Who Did and Did Not Report "Extra Stress"

	F0 change (Hz)		Amplitude change (dB)	
	CL	CO	CL	CO
Reported ($n = 26$)	17.68 (29.25)	3.41 (20.07)	3.67 (6.60)	4.67 (7.35)
Did not report ($n = 15$)	23.56 (41.23)	6.69 (37.16)	3.80 (6.84)	4.63 (6.70)

Note. Standard deviations appear in parentheses under mean values.

for this segment of each sentence were extracted using Praat (Boersma & Weenink, 2006). Default parameters were used for amplitude tracking for all sentences and for pitch tracking for most sentences. The default pitch tracking parameters yielded tracking errors or missing values in about 8% of the sentences; in these cases parameters such as minimum F0, the voicing threshold, or the silent threshold were adjusted until acceptable values were achieved. Despite these adjustments, there were two tokens for which no pitch value could be measured for *the* (both in conversational speech but from two separate talkers). F0 change in Hz was calculated by subtracting the minimum F0 of the article from the peak F0 during the vocalic portion of *bed*. dB change was calculated by subtracting the average RMS amplitude of the article from the average RMS amplitude of *bed*. Individual sentence measures were submitted to separate two-way repeated measures ANOVAs, one for pitch and one for amplitude.

Average F0 change and dB change values in each style for each reporting group are shown in Table 3. For pitch change, the main effect of speaking style was significant, $F(1, 119) = 17.45$, $p < .001$, with a much larger F0 change in clear speech (20.6 Hz) than in conversational speech (5.05 Hz). Neither the main effect of reporting status nor the interaction between speaking style and reporting status was significant, $F(1, 119) < 0.8$, $p > .3$. For amplitude, the main effect of speaking style was significant, $F(1, 121) = 4.297$, $p < .05$, but not in the expected direction: the amplitude change was actually slightly smaller in clear speech (3.76) than in conversational speech (4.65 dB). Neither the main effect of reporting status nor the interaction were significant, $F(1, 121) = 0.001$, $p = .97$; $F(1, 121) = 0.04$, $p = .85$. As seen in Table 3, both measures were extremely variable. One possible explanation for this variability is differences among preceding articles. The article was *the* for two of the sentences analyzed, and *a* for the other. However, when separate statistical analyses were performed for sentences with one or the other article, the standard deviations decreased only slightly and ANOVA results were unchanged. Taken together, the results of these analyses suggest that talkers who did and did not report placing extra stress on content words made similar changes to their pitch and amplitude contours when speaking clearly.

DISCUSSION

Clear Speech Strategies

As seen in Table 1, the clear speech strategy reported by the largest number of talkers ($n = 29$) was to enunciate the sentences more carefully. The high incidence of this strategy was not surprising, given that the instructions, though non-specific, did include the phrase, "speak clearly." The two next most common strategies, spoke more slowly ($n = 20$) and spoke more loudly ($n = 20$), were similarly unsurprising. All previous investigations comparing the acoustic characteristics of clear and conversational speech (e.g., Picheny et al., 1986; Smiljanic & Bradlow, 2005) have shown clear speech to be much slower than conversational speech. Picheny et al. (1986) also reported level differences between the two speaking styles, with clear speech having an overall higher intensity than conversational speech.

Other reported strategies were also consistent with previous literature on the characteristics of clear speech. For example, 8 talkers reported being careful to enunciate consonants, particularly final consonants. This agrees well with the report of Picheny et al. (1986) that final stop consonants were more likely to be released in clear speech than in conversational speech. Six talkers reported using more pauses in clear speech than in conversational speech, a phenomenon that was observed by Picheny et al. for all 3 of their talkers. Nine talkers reported making an effort to be easier to lipread in clear speech. This strategy was not anticipated, considering that the talkers were seated alone in a sound-treated room for the recordings and that the recordings were audio-only. Nonetheless, previous studies have shown that clear speech improves intelligibility in visual-alone and audio-visual sentence identification tasks (Gagné et al., 1995; Helfer, 1998).

Despite having received only general instructions to speak clearly, there seemed to be general agreement among the talkers about what would make their speech easier to understand for an individual with hearing loss. Note that two of the less-frequently-reported strategies (enunciated consonants and used more pauses) were related to two of the most common strategies (enunciated more carefully and spoke more slowly, respectively). For each of these, only half of the talkers who reported the less-frequently-reported strategy also reported the more-frequently-reported one (e.g., of the 6 talkers who reported using more pauses, only 3 also reported speaking more slowly). Of the 11 reported strategies, only 4 were reported by just 1 or 2 talkers. In all cases, talkers who reported these idiosyncratic strategies also reported one or more of the more common strategies.

Thus, it appears that talkers have similar ideas about how to go about making their speech more intelligible for listeners with hearing loss.

Acoustic Correlates of Reported Strategies

While the strategies themselves are interesting, the chief motivation for the present study was to discover whether talkers who did and did not report specific clear speech strategies actually made different acoustic changes when speaking clearly. That is, did talkers actually do what they said they did? As noted in the introduction, one could argue that because speech production is an automatic process involving procedural knowledge, talkers should not be able to describe how they carried it out. Lindblom (1990), however, claims that "speakers have a choice" (p. 415) and actively control the details of their speech production depending on the demands of a given communicative interaction.

For two of the strategies examined, differences between clear and conversational speech were the same for talkers who did and did not report them. Talkers produced a higher percentage of strong stops for final /d/ in /bVd/ words and expanded their vowel space in clear speech to the same degree whether or not they reported enunciating more carefully. Talkers also reduced their speaking rate to the same degree regardless of whether they reported speaking more slowly.

For the other two strategies, however, reporters and non-reporters produced different patterns of clear speech acoustic changes. Talkers who reported speaking more loudly in clear speech showed a greater level increase (6.25 dB) than talkers who did not report this strategy (2 dB), suggesting that talkers are aware of and able to control this global aspect of speech production. In addition, talkers who did and did not report placing extra stress on content words in clear speech differed in how they lengthened words within the sentence. For talkers who reported placing extra stress on keywords, content words occupied roughly the same proportion of the duration of the sentence in clear speech as they did in conversational speech. For non-reporters, in contrast, the proportion of the sentence occupied by content words was actually *smaller* in clear speech, suggesting that they lengthened function words more than they lengthened content words. The clear versus conversational duration ratios in Table 2 confirmed this interpretation.

Table 2 also shows that reporters and non-reporters treated content words similarly, lengthening them by a factor of about 1.4 in clear speech. This might seem inconsistent with the reporters' claim of placing extra stress on these words. However, Figure 3, a schematic illustrating how the two talker groups increased word durations when speaking clearly, suggests that by expanding content and function words equally, the reporters better preserved the natural stress pattern of the sentence. In contrast, non-reporters' preferential lengthening of function words decreased the length contrast between content and function words. Thus, it appears that talkers who reported placing extra stress on content words slowed their speech in a way that preserved the rhythm of the sentence, something Osberger and Levitt (1979) suggested may be important for speech intelligibility.

The results thus provide a mixed answer to the question of whether talkers are conscious of what they are doing during speech production. Why might talkers be accurate in describing some aspects of speech production and not others? One possibility is that some aspects of clear speech production are automatic, and thus not reported accurately, while others involve a controlled process that talkers can describe. The results for talkers who did and did not report speaking more slowly can be interpreted in this way. Reducing one's speaking rate may not be a clear speech strategy per se, but rather a spontaneously-occurring byproduct of other strategies that take additional time to carry out. This explanation is supported by numerous studies suggesting that despite being a nearly universal feature of clear speech, slowing alone does not account for the clear speech intelligibility benefit (e.g., Uchanski et al., 1996). The duration results for the extra stress strategy, in contrast, can be interpreted as reflecting a conscious effort on the part of reporters to slow their speech in a strategic manner that maintained temporal relationships within the sentence. Speaking more loudly could also be argued to require conscious effort. The results for the enunciated more carefully strategy, however, are problematic here. Producing strong rather than weak stops certainly requires more effort on the part of the talker, but no difference was found between reporters and non-reporters of the strategy.

Another possible explanation for the mixed results is that talkers were accurate in describing not what they *did* when speaking clearly, but rather what they *heard*. The idea that talkers observe their acoustic output during speech production is consistent both with Lindblom's H&H Theory (1990) and with the speech motor

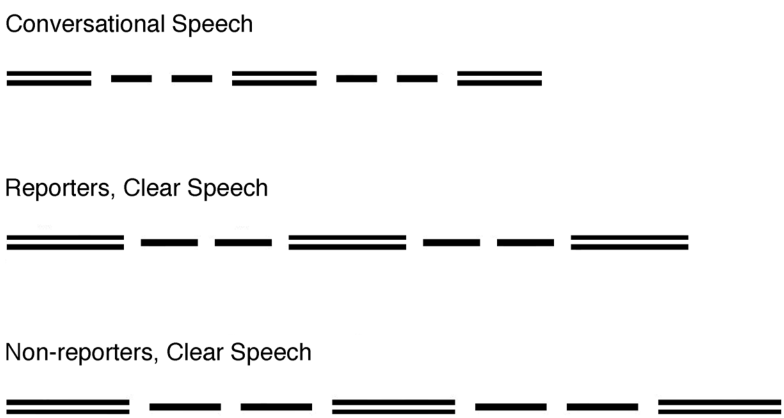

Figure 3. A schematic illustration of the relative length of content words (double line) versus function words (single line) in conversational speech versus in clear speech for talkers who did and did not report placing extra stress on content words when speaking clearly.

control model of Perkell et al. (2000), both of which state that talkers strive to achieve certain acoustic targets during speech production. Perhaps the reason talkers who reported speaking more loudly in clear speech were accurate in this assessment is that they observed that their speech was louder. The results for the extra stress strategy can also be interpreted in this way. Perhaps the talkers who said they placed extra stress on keywords reported it because they observed a more salient stress pattern in the sentence: a preserved rhythm coupled with changes in pitch and amplitude. However, one could argue that talkers should also have been able to observe and accurately describe their speaking rate or markers for more careful enunciation.

CONCLUSION

For two of the four reported strategies analyzed here, significant acoustic differences were found between the clear speech of talkers who did and did not report these strategies. Specifically, talkers who reported speaking more loudly showed a larger level increase in their clear speech than talkers who did not report this strategy. Talkers who reported placing extra stress on content words slowed their speech in a different manner from talkers who did not report this strategy (although no differences were found between reporters and non-reporters for measures of pitch change and amplitude change). In contrast, talkers who did and did not report speaking more slowly in clear speech all reduced their speaking rate to the same degree and talkers who did and did not report enunciating more carefully showed similar increases in final consonant release and in the dimensions of the vowel space.

The results of this study have promising clinical implications. The fact that talkers would be accurate at all in describing their speech behaviors suggests a level of control over speech production that bodes well for future interventions designed to teach talkers to produce the specific acoustic changes that lead to improved speech intelligibility for their communication partners with hearing loss, once these changes have been identified. That is, when “what makes clear speech clear” has been determined, the present data indicate that it should be possible for clinicians to train the frequent communication partners of individuals with hearing loss how to produce effective clear speech. However, the fact that the results for the various strategies were mixed indicates that the question of whether talkers know what they are doing during speech production requires further examination.

From a theoretical perspective, the results of this study support one of the key notions at the heart of Lindblom’s H&H Theory (1990): the idea that talkers have some volitional control over their speech production accuracy. However, H&H Theory also assumes that talkers actually know what sorts of changes their communication partners need. There is evidence that they do not. For example, Ferguson (2004) showed that talkers vary considerably in the magnitude of the clear speech vowel intelligibility benefit. In addition, the changes made in clear speech by the talker in Ferguson and Kewley-Port (2002) were beneficial for listeners with normal hearing identifying vowels in noise, but not for listeners with hearing loss, even though the talker was an audiologist with many years of experience communicating with individuals with hearing loss. Furthermore, studies on talkers' unsuccessful use of prosody to disambiguate grammatically ambiguous sentences suggest that talkers are not very good judges of what listeners need (Keysar, 2007). Future research is needed to determine not only which acoustic characteristics are responsible for making clear speech clear, but also the degree to which talkers know which changes are needed and how to apply them.

ACKNOWLEDGEMENTS

Portions of these data were presented at the 37th annual Academy of Rehabilitative Audiology Summer Institute in Kerrville, Texas in June 2003 and at the 153rd meeting of the Acoustical Society of America in Salt Lake City, Utah in June 2007. Recording and initial perceptual analysis of the Ferguson (2004) database was supported by NIHDCD-02229 to Diane Kewley-Port at Indiana University, who also provided valuable advice regarding the new experiments reported here. The authors thank Tammi Soileau and Emily Kerr for performing the pitch and amplitude measures for the extra stress strategy. We also thank Ann Bradlow for her guidance regarding the final consonant release analysis and Katherine Beam, Jessica Stamey, and Kyung Ae Keum for carrying out this analysis. Finally, we thank Harry Levitt, Maureen Stone, and Katherine Verdolini Abbott for their helpful comments on this work.

REFERENCES

Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.5.08) [Computer program]. Retrieved December 20, 2006, from http://www.praat.org/

Bradlow, A.R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112,* 272-284.

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

Davis, H., & Silverman, S.R. (1978). *Hearing and deafness* (4th ed.). New York: Holt, Rinehart and Winston.

Ferguson, S.H. (2002). *Vowels in clear and conversational speech: Talker differences in acoustic features and intelligibility for normal-hearing listeners*. Unpublished doctoral dissertation, Indiana University, Bloomington.

Ferguson, S.H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America, 116,* 2365-2373.

Ferguson, S.H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112,* 259-271.

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

Ferguson, S.H., & Lee, J. (2006). Vowel intelligibility in clear and conversational speech for cochlear implant users: A preliminary study. *Journal of the Academy of Rehabilitative Audiology, 39,* 1-16.

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *Journal of the*

Acoustical Society of America, 90, 1816-1827.

Gagné, J.-P., Masterson, V.M., Munhall, K.G., Bilida, N., & Querengesser, C. (1994). Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *Journal of the Academy of Rehabilitative Audiology, 27,* 135-158.

Gagné, J.-P., Querengesser, C., Folkeard, P., Munhall, K.G., & Masterson, V.M. (1995). Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech. *The Volta Review, 97,* 33-51.

Gordon-Salant, S. (1986). Recognition of natural and time-altered CVs by young and elderly subjects with normal hearing. *Journal of the Acoustical Society of America, 80,* 1599-1607.

Gordon-Salant, S. (1987). Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects. *Journal of the Acoustical Society of America, 81,* 1199-1202.

Helfer, K.S. (1998). Auditory and auditory-visual recognition of clear and conversational speech by older adults. *Journal of the American Academy of Audiology, 9,* 234-242.

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics, 4,* 71-84.

Kleinow, J., Smith, A., & Ramig, L.O. (2001). Speech motor stability in IPD: Effects of rate and loudness manipulations. *Journal of Speech, Language and Hearing Research, 44,* 1041-1051.

Krause, J.C., & Braida, L.D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America, 112,* 2165-2172.

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,* 362-378.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Berlin, Germany: Mouton/DeGruyter.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht, Netherlands: Kluwer Academic Publishers.

Liu, S., Del Rio, E., Bradlow, A.R., & Zeng, F.G. (2004). Clear speech perception in acoustic and electric hearing. *Journal of the Acoustical Society of America, 116,* 2374-2383.

Liu, S., & Zeng, F.G. (2006). Temporal properties in clear speech perception. *Journal of the Acoustical Society of America, 120,* 424-432.

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning, 33,* 135-157.

Nejime, Y., & Moore, B.C.J. (1998). Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss. *Journal of the Acoustical Society of America, 103,* 572-576.

Osberger, M.J., & Levitt, H. (1979). The effect of timing errors on the intelligibility of deaf children's speech. *Journal of the Acoustical Society of America, 66,* 1316-1324.

Payton, K.L., Uchanski, R.M., & Braida, L.D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America, 95,* 1581-1592.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., et al. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28,* 233-272.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96-103.

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

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1989). Speaking clearly for the hard of hearing III:

An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research, 32,* 600-603.

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

Sjölander, K., & Beskow, J. (2006). Wavesurfer (Version 1.8.5) [Computer software]. (Available from http://www.speech.kth.se/wavesurfer/)

Smiljanic, R., & Bradlow, A.R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America, 118,* 1677-1688.

Tillman, T., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words, Northwestern University Auditory Test No 6*. (Tech. Report No. SAM-TR-66-55). Brooks AFB, TX: USAF School of Aerospace Medicine.

Tjaden, K., & Wilding, G.E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language and Hearing Research, 47,* 766-783.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17,* 84-119.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America, 88,* 97-100.

Turner, G.S., & Weismer, G. (1993). Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 36,* 1134-1144.

Uchanski, R.M., Choi, S.S., Braida, L.D., & Durlach, N.I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research, 39,* 494-509.

Wang, Y.T., Kent, R.D., Duffy, J.R., & Thomas, J.E. (2005). Dysarthria associated with traumatic brain injury: Speaking rate and emphatic stress. *Journal of Communicative Disorders, 38,* 231-260.

Watson, P.J., & Hixon, T.J. (1985). Respiratory kinematics in classical (opera) singers. *Journal of Speech and Hearing Research, 28,* 104-122.

Zeng, F.G., & Liu, S. (2006). Speech perception in individuals with auditory neuropathy. *Journal of Speech, Language, and Hearing Research, 49,* 367-380.

APPENDIX A

PRACTICE SENTENCE LIST USED FOR LEVEL RECORDINGS

1. Vera put the sail on the table.
2. His mother calls him Bad at home.
3. Jean bought a bude at the store.
4. They spelled the word hood the wrong way.
5. Say the word bud into the microphone.
6. He picked up the bead and put it away.
7. He looked for a Time but couldn't find one.
8. Look up the word fit on the Internet.
9. I think the word bade is hard for kids to say.
10. Her friends called her Bode back in school.
11. Please put the pole next to the skis.
12. You might find a bood in the garage.
13. Write the word bed on the chalkboard.
14. Use the word rush in a sentence.
15. I looked up the word bid in the dictionary.
16. He said bod but he meant box.

APPENDIX B

CID EVERYDAY SENTENCES (DAVIS & SILVERMAN, 1978) RECORDED IN CLEAR SPEECH

1. You'll get fat eating candy.
2. I'll see you right after lunch.
3. There's a good ballgame this afternoon.
4. Music always cheers me up.
5. How do you spell your name?
6. What are you hiding under your coat?
7. The phone call's for you.
8. Come here when I call you.
9. I'll think it over.
10. Walking's my favorite exercise.
11. Where are you going?
12. Wait just a minute!
13. I'll catch up with you later.
14. Call me a little later.

APPENDIX C

CID EVERYDAY SENTENCES (DAVIS & SILVERMAN, 1978) RECORDED IN CONVERSATIONAL SPEECH

1. It's time to go.
2. Do you want to wash up?
3. The water's too cold for swimming.
4. It's no trouble at all.
5. Here are your shoes.
6. Have you been working hard lately?
7. How do you know?
8. Move out of the way.
9. Pass the bread and butter, please.
10. Weeds are spoiling the yard.
11. Breakfast is ready downstairs.
12. Let's get a cup of coffee.
13. I hate driving at night.
14. I'll carry the package for you.