The Relationship Between Perception and Acoustics for a High-Low Vowel Contrast Produced by Speakers With Dysarthria

Kate Bunton
Gary Weismer
Department of Communicative Disorders
Waisman Center
University of Wisconsin–Madison

This study was designed to explore the relationship between perception of a high-low vowel contrast and its acoustic correlates in tokens produced by persons with motor speech disorders. An intelligibility test designed by Kent, Weismer, Kent, and Rosenbek (1989a) groups target and error words in minimal-pair contrasts. This format allows for construction of phonetic error profiles based on listener responses, thus allowing for a direct comparison of the acoustic characteristics of vowels perceived as the intended target with those heard as something other than the target. The high-low vowel contrast was found to be a consistent error across clinical groups and therefore was selected for acoustic analysis. The contrast was expected to have well-defined acoustic measures or correlates, derived from the literature, that directly relate to a listeners' responses for that token. These measures include the difference between the second and first formant frequency (F2-F1), the difference between F1 and the fundamental frequency (F0), and vowel duration. Results showed that the acoustic characteristics of tongue-height errors were not clearly differentiated from the acoustic characteristics of targets. Rather, the acoustic characteristics of errors often looked like noisy (nonprototypical) versions of the targets. Results are discussed in terms of the test from which the errors were derived and within the framework of speech perception theory.

KEY WORDS: speech intelligibility, acoustics, dysarthria, speech perception

Most clinicians and researchers agree: "Intelligibility is…the most practical single index to apply in assessing competence in oral communication" (Subtelny, 1977, p. 183). Despite its importance, consensus on how intelligibility should be measured or analyzed has not been obtained. Intelligibility tests designed for persons with motor speech disorders have traditionally served as indexes of performance. Speech intelligibility scores in these tests are calculated by counting "the number of discrete speech units correctly recognized by a listener" (Flanagan, 1972, p. 311), thereby providing a single index of performance or severity that is efficient, economical, and easy to communicate. These testing procedures, however, do not provide an explanation for the origin of the intelligibility problem. It is in fact possible for two individuals to have the same intelligibility score but different underlying speech problems necessitating distinct management programs. Efforts to move beyond measures of intelligibility that provide severity estimates only, and to work toward the development of acoustic-perceptual models of motor speech impairments of intelligibility, have clear utility in understanding and development of remedial strategies in this population.

Kent, Weismer, Kent, and Rosenbek (1989a) proposed an alternative test of intelligibility that moves beyond a measure of severity and attempts to identify the acoustic and phonetic bases of an intelligibility deficit. They proposed a word-intelligibility test that includes 19 phonemic contrasts reported in the literature as vulnerable to impairment in motor speech disorders. Using a multiple-choice format, each target word was grouped with minimal-pairs, or near minimal-pairs. Listener responses can be analyzed to obtain phonetic error profiles showing the particular phonetic contrasts that presumably made an important contribution to the overall speech-intelligibility deficit. This method of testing allows for identification of reasons for the intelligibility deficit, quantitative analyses at varied levels, and sensitivity to potential speech deficits across clinical subgroups. In addition, results can be used to guide additional assessment measures as well as treatment goals/protocols and are interpretable within standard articulatory testing. Analyses reported to date using the Kent et al. (1989a) word-intelligibility test have demonstrated that different phonetic error profiles may underlie similar global intelligibility deficits, that certain sex effects may be prominent in some profiles, and that the relationship between disease type and the phonetic error profiles may be quite complex (see Kent et al., 1991; J. Kent et al., 1992; Kent et al., 1994).

One additional concern about standard intelligibility testing is that research has demonstrated that standard intelligibility procedures offer no insight into listener strategies and difficulties (Ansel, 1985; Kent et al., 1989a; Weismer & Martin, 1992). One consideration for including variables in an explanatory model of intelligibility would be based on factors that are assumed to have general importance to the speech-perception process. A few caveats of this approach have been outlined by Weismer and Martin (1992). One view of a model of speech intelligibility, based around the importance of certain features in the speech signal, requires that certain of these features be regarded as “primary.” These factors are assumed to be consistent across utterances and speakers in addition to their psychoacoustic and/or linguistic salience. One example, discussed by Weismer and Martin, is the second formant, known to be psychoacoustically and linguistically salient, which maintains its prominence in the normal speech signal as a result of a nonvarying glottal source spectrum and wide range of tongue movements. In disordered speech, a variable and noisy source spectrum, variable and excessive hypernasality, and restricted ranges of tongue movement may make F2 a less stable, and therefore a less heavily weighted, perceptual cue. There is also evidence in the speech perception literature to indicate that a listener may have more trouble with a disordered speech signal than would be predicted solely on the basis of degradation of acoustic events (see Weismer & Martin, 1992).

Before achieving an understanding of how the disordered speech signal may be processed by a listener, it is necessary to know something about the acoustic characteristics of segmental errors. Although the Kent et al. (1989a) test is constructed from phonetic contrasts with well-defined acoustic correlates, there are no reports of acoustic analyses of the errors observed in this test. More generally, there are very few published data on the acoustic characteristics of well-defined perceptual errors produced by any type of speaker with dysarthria. One of the aims of the current study was to focus on a particular type of error, the high-low contrast for vowels, and to document the acoustic characteristics of those errors in three groups of speakers with different types of neurological disease and concomitant dysarthria. The groups included speakers with Parkinson disease (PD), amyotrophic lateral sclerosis (ALS), and cerebrovascular accident (CVA). The high-low vowel contrast was chosen for this analysis because it was associated with a relatively high rate of errors in the Kent et al. test (Bunton, 1999) and has well-defined acoustic characteristics that can be applied to the analysis of target/error pairs. In addition, such an analysis was deemed important because vowel contrast errors appear to have a relatively high impact on speech intelligibility deficits in dysarthria (Kent et al., 1989b; Weismer & Martin, 1992; Ziegler, Hartmann, & von Cramon, 1988). In addition, acoustic measures of vowel articulation have been shown to be highly correlated with estimates of speech intelligibility in speakers with ALS (Turner, Tjaden, & Weismer, 1995; Weismer, Laures, Jeng, R. Kent, & J. Kent, 2000), Parkinson disease (Weismer, Jeng, Laures, R. Kent, & J. Kent, 2001), and cerebral palsy (Jeng, 2000). Although consonants are traditionally thought of as the primary information units of speech intelligibility, several authors have argued that a complete understanding of speech intelligibility deficits requires a full understanding of vowel behavior in production and perception.

Within the descriptive framework of this analysis the major research question was “Does the acoustic analysis of tongue-height errors (and their correct counterparts) produce results that are consistent with the published acoustic correlates of the tongue height dimension, as derived from speech acoustic and perception research (and as outlined in Kent et al., 1989a)?”

Method

Participants

Four groups of speakers were included in the present study: (a) 10 speakers with no history of neurologic disease (age-matched controls [NG]), (b) 10 speakers diagnosed with ALS, (c) 10 speakers diagnosed with PD, and (d) 5 speakers with a history of CVA. All speakers selected for the current study were recruited as part of a larger study on dysarthria (see R. Kent et al., 1989a, 1989b, 1990, 1991, 1994; J. Kent et al., 1992); individual speaker characteristics can be seen in Table 1. All speakers spoke an Upper Midwest dialect of English. The speakers within each group were diverse in several respects, including severity of dysarthria, duration of the disease, and medical history. The three disorder groups chosen in the present study were selected largely because of the availability of subjects in a referral pool, but also because research using phonetic error profiles has demonstrated that there may be a relationship between disease type and phonetic error types (R. Kent et al., 1990, 1991, 1994; J. Kent et al., 1992). No female speakers were included in the CVA category because the data-pool was not large enough to obtain an equal number of male and female speakers. When possible,

Table 1. Participant characteristics. Blanks within the table indicate missing information. Speaker groups are indicated in column 4, duration refers to the time since diagnosis, and intelligibility scores are based on the Kent et al. (1989) single-word intelligibility test.

Speaker	Age (years)	Gender	Disorder / Location of Lesion	Duration (years)	Intelligibility (%)	
					M	*SD*
1	68	M	NG		94.38	0.67
2	69	M	NG		95.24	2.79
3	72	M	NG		95.52	0.60
4	68	M	NG		97.96	0.45
5	77	M	NG		94.52	2.28
6	68	F	NG		96.67	1.38
7	72	F	NG		96.38	1.21
8	76	F	NG		98.38	0.61
9	76	F	NG		97.96	1.21
10	68	F	NG		97.38	1.05
11	61	M	ALS	1	85.40	4.27
12	41	M	ALS	2	78.50	6.45
13	62	M	ALS	12	83.40	1.66
14	55	M	ALS	1	92.70	2.76
15	29	M	ALS	2	93.80	2.87
16	57	F	ALS	1	75.65	3.80
17	47	F	ALS		76.50	6.82
18	75	F	ALS	1	91.23	3.60
19	52	F	ALS	9	96.24	2.62
20	52	F	ALS	10	91.52	2.80
21	62	M	PD	3	90.09	3.93
22	64	M	PD	8	82.94	3.81
23	73	M	PD	7	85.51	5.08
24	62	M	PD	9	86.51	3.83
25	81	M	PD	13	91.38	3.19
26	72	F	PD	20	89.52	3.76
27	79	F	PD	6	90.95	3.58
28	67	F	PD	4	94.67	2.68
29	82	F	PD	1	92.66	4.09
30	75	F	PD	17	94.81	3.02
31	71	M	CVA-left	1	79.60	7.11
32	77	M	CVA-left	1	92.90	3.24
33	58	M	CVA-right	1	86.70	5.23
34	75	M	CVA-right	3	77.10	6.51
35	63	M	CVA-bilateral	1	89.70	4.05

comparisons between errors and targets were made with the data separated by sex of speaker, because of previous work suggesting sex differences in the phonetic basis of intelligibility deficits in dysarthria (J. Kent et al., 1992, R. Kent et al., 1994).

All speakers had scores of greater than 75% on the Kent et al. (1989a) single-word intelligibility test. This cut-off was selected based on experience with speakers in the database, wherein the difficulty of identifying acoustic landmarks increases and reliability with which the measurements are made decreases for speakers with lower intelligibility. Although these single-word intelligibility scores may seem relatively high, most of the current speakers were judged as being only moderately intelligible in conversational speech.

Intelligibility Measures

Intelligibility data were obtained from a word-identification test with a multiple-choice format (Kent et al., 1989a). Speakers were recorded on audiotape while reading aloud test words from cards; the recorded words were randomized and assembled onto listening tapes in 6 different orders. These speech samples were presented though a loudspeaker in an audiometric test booth to a group of 10 normal-hearing undergraduate students enrolled in a communication disorders program. The speech signal volume was set by the examiners to a comfortable listening level for the majority of words. Listeners were given a response form that had four words in each numbered row (one target word and three foils) and asked to select the word in each row that most closely approximated the speaker's productions. Each target word and foil pair had a minimal-pair or near minimal-pair relationship; in the case of the tongue-height contrasts, all target-foil pairs were in a minimal-pair relationship. For example, for the test word *bad* the response choices are *bed*, *bat*, *bad* (target), and *pad*. The first choice, *bed*, differs from the test word in the tongue-height feature of the vowel; the second choice, *bat*, differs from the test word in the voicing feature of the final consonant; and the fourth choice, *pad*, differs from the test word in the voicing of the initial consonant.

Listener response forms were scored by computer to determine both the percent correct and the phonetic contrast error profile for individual speakers. Error rates were calculated by recording each time a listener marked a response other than the target word and dividing the total errors for each pair by the number of listeners (n = 10). For example, if the target was *bad* and three of the listeners chose *bed* an error rate of 0.3 (3/10) would be marked in the high-low vowel feature category. Errors per contrast were then added (i.e., the error rate for the 13 word pairs in the high-low vowel category) and divided by the total number of errors possible on the test (n = 206). A list of the eight most frequent errors for each speaker group was then compiled. These lists are presented in Table 2; error frequencies are listed in decreasing rank order. The actual error frequencies for each contrast are not shown within the table because the primary interest was with errors categories that contributed most significantly to a decrease in intelligibility. For the NG speakers less than 8 contrasts were included in the table, as errors were listed only when two or more speakers

Table 2. Phonetic contrasts on which each speaker group most frequently erred, listed in descending order.

Geriatric Female	Geriatric Male	ALS Female	ALS Male	PD Female	PD Male	CVA Male
High-low vowels	High-low vowels	Stop-nasal	Glottal-null	High-low vowels	Glottal-null	Glottal-null
Voiced-voiceless initial consonant	Glottal-null	High-low vowels	Final consonant-null	Stop-fricative	High-low vowels	High-low vowels
Stop-nasal	Long-short vowels	Voiced-voiceless final consonant	High-low vowels	Stop-affricate	Voiced-voiceless initial consonant	Voiced-voiceless initial consonant
Other fricative place of articulation	Stop-nasal	Stop place of articulation	Voiced-voiceless initial consonant	Glottal-null	Alveolar palatal fricative	Alveolar palatal fricative
Final cluster singleton		Fricative-affricate	Voiced-voiceless final consonant	Voiced-voiceless initial consonant	Voiced-voiceless final consonant	Stop-nasal
		Final consonant-null	/r/-/w/	/r/-/w/	Initial consonant-null	Stop-fricative
		Initial cluster-singleton	Stop-nasal	Alveolar palatal fricative	Initial cluster-singleton	/r/-/w/
		Stop-fricative	Initial cluster-singleton	Stop-nasal	Stop-fricative	Stop-affricate

had difficulty. Across speaker groups, the high-low vowel contrast was one of the top three contrasts in error for all groups and was first or second for all but the male speakers with ALS. The frequent occurrence of errors on the high-low vowel contrast across clinical speaker groups suggests that they made a significant contribution to the lower intelligibility scores for these speakers.

Acoustic Measures

For the present study, the acoustic measures chosen from the literature were expected to reflect the perceptually important cues that distinguish a high versus a low vowel. The Kent et al. (1989a) single-word intelligibility test contains 13 word pairs in the high-low vowel contrast category. No errors were found for 4 of the word pairs; therefore, to limit the size of the analysis, these word pairs were excluded. The 9 remaining word pairs with target word listed first and error word listed second were *knew-know, him-hem, him-ham, pit-pat, pit-pet, bad-bed, shoot-shot, geese-gas,* and *heat-hate.* Analysis of these tokens allowed for a direct comparison of the acoustic characteristics of vowels perceived as the intended target versus those heard as tongue-height errors. Of the 315 tokens (9 tokens × 35 speakers) analyzed for this contrast, 246 were perceived as targets and 69 were perceived as errors.

The recorded intelligibility words were digitized on a PC-clone computer at a sampling rate of 22 kHz and low-pass filtered at a cut-off frequency of 9.8 kHz. Acoustic measures were made using CSpeech (Milenkovic, 1994), a computer software program. Measurement of the tokens for the high-low vowel contrasts in the present study included (a) first and second formant frequencies (F1-F2) at three temporal points during the vowel (25%), midpoint (50%), and 75% of total vowel duration using a combined spectrogram and LPC spectrum; (b) fundamental frequency (F0) at these same three temporal points, using the automatic pitch tracker in Cspeech; and (c) vowel duration, measured from the combined speech waveform-spectrographic display.

Phonetic data and modeling have demonstrated that F1 is inversely related to tongue height (Fant, 1960; Ladefoged, 1975; Peterson & Barney, 1952; Pickett, 1980). For a token where the target was a higher vowel and the error a lower vowel, it was expected that the frequency of F1 would be higher for the errors than for tokens perceived as targets. This was expected for 8 of 9 word pairs in the present study, where the target contained a high vowel. Conversely, a lower F1 would be expected if the target were a lower vowel and an error a higher vowel (*bad / bed*). More recent studies have suggested that measurement of F1 alone may be inadequate and that the differences in F1-F0 and F2-F1 more closely represent the size of the constriction and hence are more appropriate acoustic representations of tongue-height contrasts (Kingston, 1991; Syrdal, 1985; Syrdal & Gopal, 1986; Traunmüller, 1981). A relatively small F1-F0 difference and large F2-F1 difference was expected for tokens heard as high vowels compared to a larger F1-F0 difference and smaller F2-F1 difference for tokens perceived as containing low vowels. In the Results, observations on both F1 alone and the F2-F1 and F1-F0 differences are reported. The inclusion of formant values at three time-points in the vowel nucleus is based on work by Hillenbrand and colleagues (1993a, 1993b, 1995), who have shown that including formant information from three discrete temporal points throughout a vowel nucleus, rather than only the midpoint, improves vowel classification. In the present study, it was thought that measurements at three points in time might capture the dynamic characteristics of vowel production and thus relate to listeners' perception of the token (see also Assman & Katz, 2000).

Vowel duration was measured from the first glottal pulse to the final glottal pulse of the vowel segment. These measures of vowel duration were made because research has suggested that vowel duration is inversely related to tongue height (Keating, 1985; Lehiste, 1970; Lindblom, 1967; Westbury & Keating, 1980).

Reliability

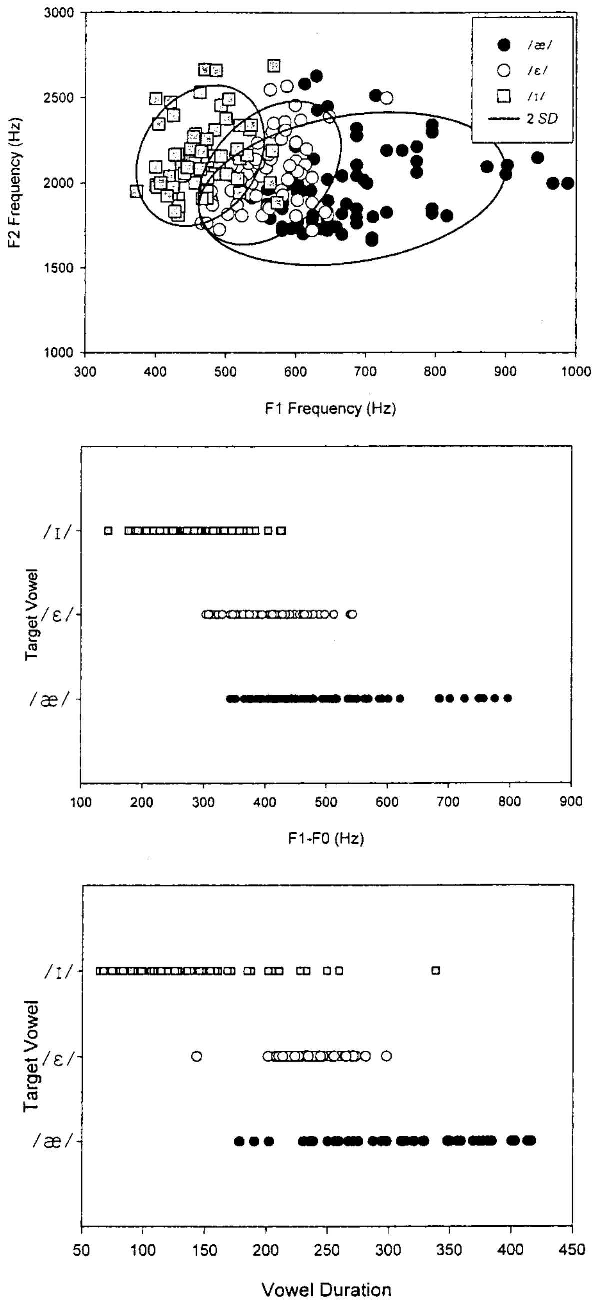
Two randomly selected target words for each speaker were measured a second time by the investigator and by a second judge to obtain estimates of intra- and interjudge reliability. The overall correlation coefficient between the first and second set of acoustic measurements completed by the first investigator was 0.98; the correlation coefficient between the two sets of measurements for each individual speaker was greater than 0.9 for both male and female speakers. The following mean differences were found for measures of intrajudge reliability: vowel duration (d = 18.6 ms) and vowel midpoint (50%) frequencies (F1: d = 18.9 Hz, F2: d = 41.2 Hz). The overall correlation coefficient for measures of interjudge reliability was 0.93, with correlation coefficients for individual speakers above 0.88 for both male and female speakers across the disorder groups. Mean differences between the two judges' measurements were vowel duration (d = 16.7 ms) and vowel midpoint (50%) frequencies (F1: d = 26.7 Hz, F2: d = 40.1 Hz).

Results

Dialectal Evaluation

The high rate of correct responses among the control speakers (across all listeners: 98.2% correct for males, 97.4% for females) for words having potential high-low vowel errors argues against a substantial role of dialect when such errors did occur. Nevertheless, it is important to demonstrate that the acoustic cues thought to distinguish the high-low vowel contrast (outlined in the Method) were present among the current group of control speakers. Toward this end, correctly perceived tokens for the /ɛ/-/æ/ and /ɪ/-/ɛ/ contrasts produced by the control speakers were examined. Clear acoustic distinctions between members of these word pairs would provide the empirical basis for the analysis of systematic acoustic deviations from these distinctions, when perceptual errors are made.

Figure 1. F1-F2 (top), F1-F0 (middle), and vowel duration (bottom) plots of the vowels /æ/ - /ɪ/ - /ɛ/ for 3 tokens each produced by the control speakers.



Three tokens per speaker of each correctly perceived vowel were measured. Tokens included *cash, had, pat, sell, bed, blend, sip, bit,* and *wit.* The three panels in Figure 1 are, from top to bottom, F1 vs. F2, F1-F0, and vowel duration. In each panel values for /ɪ/ are coded by shaded squares, for /ɛ/ by unfilled circles, and for /æ/ by filled circles; each point represents a measurement taken at the 50% time point from an individual token. The top panel shows that as tongue height decreases the plotted points move rightward and down; this rotation is captured by the two-standard deviation ellipses fit to each cloud of vowel points. There is overlap between the F1-F2 coordinates for the three vowels, but the systematic rotation of coordinates for the three vowel categories, from high to low, is associated with a decreased F2-F1 difference, as described above. The middle panel of Figure 1 shows the expected tendency of smaller F1-F0 differences for higher, as compared to lower vowels, but there is some overlap between the vowel pairs—perhaps more so for the /ɛ/-/æ/ pair. Finally, vowel duration (Figure 1, lower panel) appears to be related to tongue height, with a tendency for longer durations for /æ/ than for /ɛ/ and for /ɛ/ than for /ɪ/.

Because of the relatively small sample size and potential for non-normal distributions, the statistical evaluation of these distinctions was based on Mann-Whitney tests, with the overall error rate set at 0.05 and a per-comparison α level of 0.05/6 = .008. F2-F1 differences for the two vowel pairs were statistically significant (/æ/-/ɛ/: $U = 179.5$, $p < 0.001$; /ɛ/-/ɪ/: $U = 116.0$, $p < 0.001$), as were the F1-F0 differences and VD differences (F1-F0, /æ/-/ɛ/: $U = 222.0$, $p < 0.001$; /ɛ/-/ɪ/: $U = 194.0$, $p < 0.001$; VD, /æ/-/ɛ/: $U = 83.0$, $p < 0.001$; /ɛ/-/ɪ/: $U = 101.0$, $p < 0.001$). These data would seem to support the investigation of F2-F1 and F1-F0 differences, as well as vowel-duration differences, in comparing target and error pairs for tongue-height errors. The findings reported below are interpreted in light of these findings on the acoustic distinctions produced by the control speakers.

Error Analyses

Analyses were completed first for the entire speaker pool by sex only. The data were then divided by disorder to look for possible differences related to disease process. The acoustic analysis of tongue-height errors failed to reveal any disorder-specific patterns (Bunton, 1999), so data are presented separated only by sex, where appropriate. When word pairs are discussed they will be presented in the order *target/error*, where "target" means that the word was correctly perceived and "error" means it was incorrectly perceived, specifically as a tongue-height error. Correct productions (targets) from speakers in the control group and from speakers with dysarthria were used as a descriptive and statistical reference for the mean and variability of performance on target productions. Descriptive results are presented and followed by statistical analyses, where appropriate. As in the case of the dialect evaluation, nonparametric statistics were used in all cases because of the small sample size and the potential for non-normal distributions associated with the acoustic measures. Specifically, all comparisons were made with Mann-Whitney tests, and each comparison was subject to the Bonferroni correction such that the error rate for the entire set of statistical tests was α = 0.05 (see Bunton, 1999). Each of four major group comparisons (one for F2-F1, one for F1-F0, one for vowel duration, and one for word duration) was tested at 0.05/4 = .0125 (Bunton, 1999). In the present report word-duration results are not described, being largely redundant with vowel duration, but the per-comparison alpha level was maintained at 0.0125. For the nine individual word comparisons of vowel duration described below, each comparison was tested at α = .0125/9 (one tailed). The decision criteria for this set of planned comparisons followed rules described in Marasculio and Serlin (1988).

Formant Frequencies and F0

The means and standard deviations of the first three formant frequencies and F0 can be found in the Appendix. Plots of the F1-F2 vowel space were made for each token, examples of which are shown in Figures 2–5. In these figures the target productions (correct productions from all speakers) are shown with filled symbols and the error tokens (obtained from speakers with dysarthria) with unfilled symbols. Circles, squares, and triangles represent the 25%, 50%, and 75% measurement time points, respectively. A display combining the three measurement points is used in Figures 2–5 because no systematic differences were discovered as a function of measurement point (Bunton, 1999).

In Figure 2, *him/hem* produced by male speakers, there appears to be considerable overlap in F1 formant values for target and error tokens (especially at the 50% and 75% time slices). A plot showing similar overlap in F1 for the *pit/pet* pair produced by female speakers can be seen in Figure 3. In both of these examples a higher F1 was expected for the error, which contained a lower vowel. In fact, substantial overlap in F1 and F2 and a large range of variability was seen for 5 of the 18 word pairs (9 word-pairs × sex). Overlap in F1-F2 values for the targets and errors was also noted for the remaining 13 pairs, but many of the points for error tokens were

Figure 2. Plots of the F1-F2 vowel space for the word pair *him/hem* produced by the male speakers. The three shapes represent the three temporal measurement points (25%, 50%, 75%) within the vowel nucleus. Unfilled symbols represent error tokens from speakers with dysarthria; filled symbols represent target (correct) tokens from control speakers and speakers with dysarthria; circles, squares, and triangles represent the 25%, 50%, and 75% measurement points, respectively.

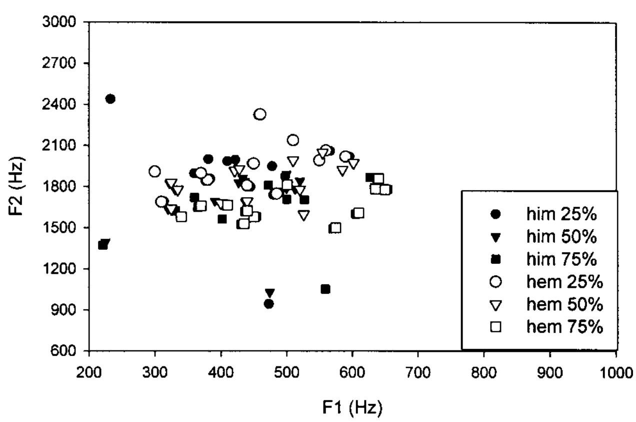
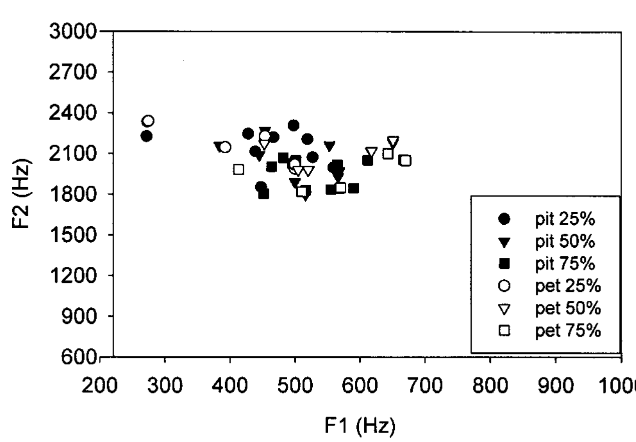


Figure 3. Plots of the F1-F2 vowel space for the word pair *pit/pet* produced by the female speakers. The three shapes represent the three temporal measurement points (25%, 50%, 75%) within the vowel nucleus. Unfilled symbols represent error tokens from speakers with dysarthria; filled symbols represent target (correct) tokens from control speakers and speakers with dysarthria; circles, squares, and triangles represent the 25%, 50%, and 75% measurement points, respectively.



located near the edges of the point distributions. An example of this is shown in Figure 4 for the *bad/bed* word pair produced by female speakers. F1 was lower at all three time points in the error tokens and had a mean of 530 Hz, compared with a mean of 690 Hz in the target tokens (means calculated across the three time points). A second example is shown in Figure 5 for the *him/ham*

Figure 4. Plots of the F1-F2 vowel space for the word pair *bad/bed* produced by the female speakers. The three shapes represent the three temporal measurement points (25%, 50%, 75%) within the vowel nucleus. Unfilled symbols represent error tokens from speakers with dysarthria; filled symbols represent target (correct) tokens from control speakers and speakers with dysarthria; circles, squares, and triangles represent the 25%, 50%, and 75% measurement points, respectively.

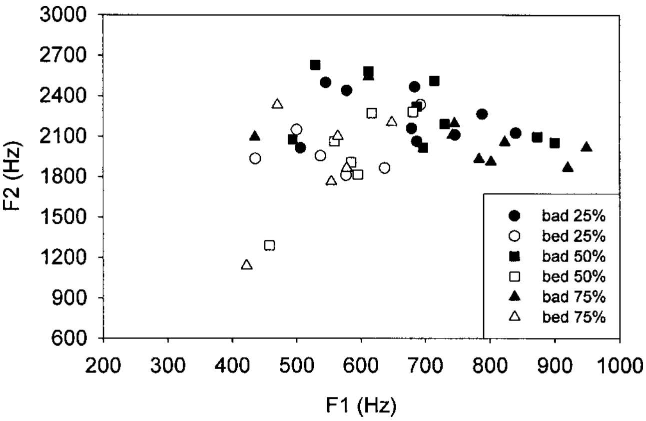
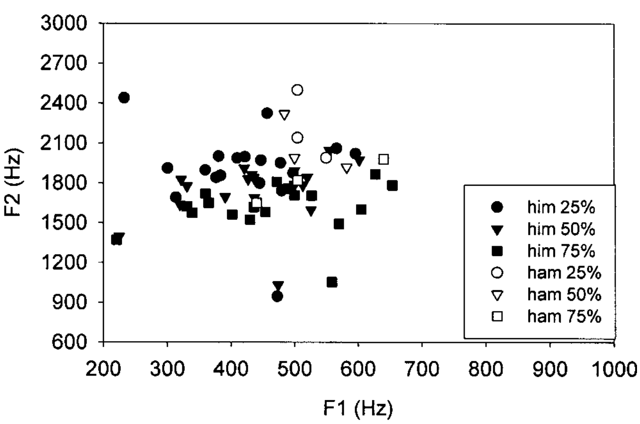


Figure 5. Plots of the F1-F2 vowel space for the word pair *him/ham* produced by the male speakers. Unfilled symbols represent error tokens from speakers with dysarthria; filled symbols represent target (correct) tokens from control speakers and speakers with dysarthria; circles, squares, and triangles represent the 25%, 50%, and 75% measurement points, respectively.



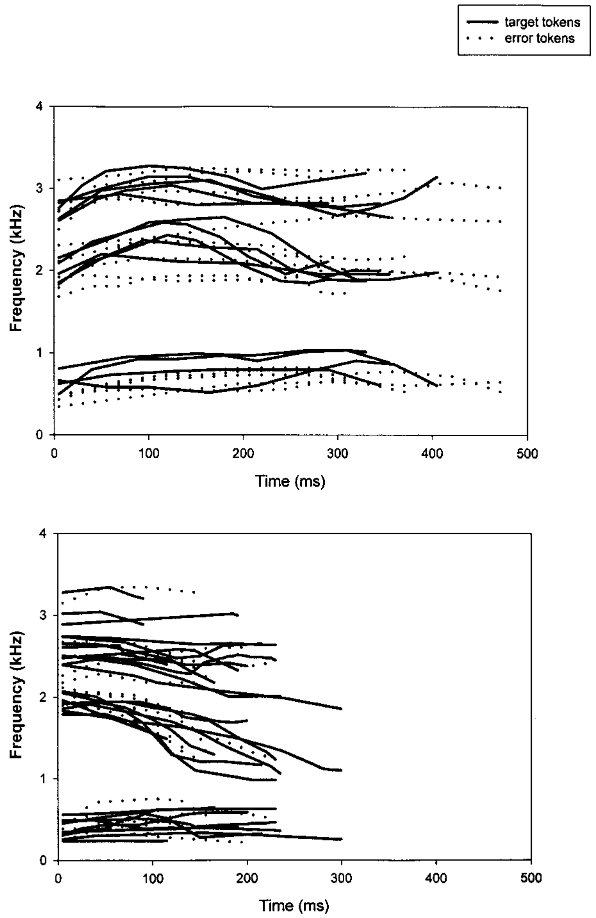
pair produced by male speakers, where the error tokens are mostly clustered at the upper right corner of the distribution. In 12 of 13 cases where this distribution pattern was observed, although the formant coordinates did not clearly distinguish the high-low vowel contrast the differences in F1 were in a direction consistent with phonetic expectations. Given the formant frequency plots and comparisons of mean formant value data, it appears that even though there were no straightforward acoustic differences linked to the perception of the tokens, there was evidence of some differences between targets and errors and therefore possible ambiguity in the acoustic signal. Bunton (1999) reported similar findings for F2-F3 plots.

When F2-F1 differences are considered, differences produced by male speakers with dysarthria for the error words *hem* (target *him*) and *gas* (target *geese*) were smaller than the differences associated with the correctly produced targets; this is consistent with expectations. For the female speakers only one error, in the *heat/hate* pair, had an F2-F1 difference that was consistent with expectations. For the remaining tokens, the F2-F1 difference for the target words was equal to or only slightly greater than the error tokens (<100 Hz). In most cases, these slight differences in F2-F1 matched expectations, but the F2-F1 difference did not appear to differentiate between target and error tokens for 4 of the 9 word pairs. The F2-F1 differences for the target and error tokens were not found to be statistically significant (females: $U = 12.7$, n errors = 29, n targets = 106, $p = 0.8429$; males: $U = 77.7$, n errors = 40, n targets = 140, $p = 0.2012$).

For the F1-F0 difference, larger differences would be expected for errors having a lower tongue height (as in a *geese/gas* pair) relative to the target; the data were consistent with these expectations for the male speakers with dysarthria, but not for the female speakers with dysarthria. For the eight vowel pairs in which the error had a lower vowel tongue height than the target, females had a median F1-F0 difference of 236 Hz for the errors and 205 Hz for the targets; the 31 Hz effect was consistent with expectations but failed to reach statistical significance ($U = 11.5$, n errors = 29, n targets = 106, $p = 0.5378$). The corresponding data for the male speakers showed a 369 Hz difference for errors and a 239 Hz difference for the target—an effect consistent with expectations and large enough to reach statistical significance ($U = 91.67$, n errors = 40, n targets = 140, $p = 0.0001$). The effect of larger F1-F0 differences for the lower vowel errors than for the higher vowel targets was not consistent across the eight vowel pairs; in fact, for a few vowel pairs the difference was actually larger for the higher vowel targets.

To check for possible differences between targets and errors that may not be detected with the kinds of formant frequency measures reported above, further analyses of the formant trajectories were completed. Figure 6 (upper panel) shows LPC-generated and hand-corrected formant trajectories (see Tjaden & Weismer, 1998) of female control and female PD speakers for the multiple tokens of the *bad/bed* pair (shown as a static plot in Figure 4). The solid lines represent the target tokens produced by control speakers, and the errors produced by speakers with PD are shown as dotted lines. The F1 trajectories are fairly similar for the targets and errors,

Figure 6. Formant trajectories for the vowel nucleus in the *bad/bed* (upper panel) and *him/ham* (lower panel) word pairs, respectively. Females in the NG and PD groups produced the tokens in the upper panel; NG, PD, and CVA male speakers produced the tokens in the lower panel.



although there is a tendency for somewhat lower F1 values for the errors. Differences can be seen in the shape of the trajectories for F2 and F3, with the error tokens having relatively flatter formants. The formant trajectories for the errors are also clearly longer than those for targets. The formant trajectories for the *him/ham* pair produced by male control speakers (solid lines) and male speakers with Parkinson disease and stroke (i.e., errors: dotted lines) can be seen in the lower panel of Figure 6 (static plot shown in Figure 5). In this plot it is difficult to see any difference between the target and error trajectories.

Vowel Durations

Group means and standard deviations for vowel duration measures are presented in Table 3. A pattern of longer vowel durations for errors was observed for many word pairs, although none of the differences were statistically significant. For the males five of the eight word-pair comparisons with low vowel errors had longer durations for the errors, and for females six out of seven of the low-vowel errors had longer vowels for the error. (One pair, *geese/gas*, did not contribute data to this analysis; see Table 3.) For the one word pair in which the error involved a higher vowel than the target, the errors still had longer vowels. Neither of the target-error comparisons was significant for the whole set of data, pooled across words (females: $U = 34$, $n = 29$ errors, $n = 106$ targets, $p = 0.1045$; males: $U = 26$, $n = 40$ errors, $n = 140$ targets, $p = 0.0388$). Because the group effects were close to being significant and the effects were typically in the right direction, vowel duration differences were evaluated on a per-word basis, but none of the comparisons were significant.

Discussion

The objective of the present study was to examine in dysarthric speakers the acoustic underpinnings of tongue-height errors derived from a specific speech intelligibility test. Nine *high-low* vowel pairs from the Kent et al. (1989a) speech intelligibility test were studied, eight of which had a lower vowel error for a higher vowel target; the remaining pair had the opposite relation. For the eight targets with high vowels, the primary acoustic expectation for lower vowel errors was an increase in F1 with the resulting decrease in the F2-F1 difference; secondary expectations for the errors included a relatively larger F1-F0 difference and relatively longer vowel durations. The opposite acoustic differences were expected for the *bad/bed* word pair. Results were pooled across disorder groups because the acoustic characteristics of these tongue-height errors were not dependent on type of neurological disease (Bunton, 1999; Bunton & Weismer, 2000).

The main research question—whether the acoustic characteristics of tongue-height errors are consistent with expectations from the literature—is somewhat difficult to answer. Overall, results suggest what appears to be a partial mismatch between these acoustic expectations and the tongue-height errors. No statistically significant acoustic differences were found for any of the target/error pairs for F2-F1 difference, the presumed primary cue. There were cases, however, in which the F1 data for errors appeared to lie near or at the margins of the distribution of points for the targets. Vowel duration was not associated with a significant difference between targets and errors for either sex, but most of the target-error differences in vowel duration were in the expected direction. In the case of the F1-F0 dif-ference male speakers had a significant difference in the expected direction; females failed to show the same difference.

Table 3. Means and standard deviations (in italics) for vowel duration, reported in milliseconds. Asterisks within the table indicate no data were available.

	Males				Females			
	Target		Error		Target		Error	
knew/know	226.3	*75.7*	239.3	*103.3*	380.3	*61.5*	555.0	*0.0*
bad/bed	275.1	*48.0*	290.8	*76.2*	341.8	*83.1*	395.8	*91.2*
him/hem	138.7	*36.9*	162.8	*46.0*	143.3	*88.9*	179.6	*47.8*
him/ham	138.7	*36.9*	187.0	*32.5*	143.3	*88.9*	239.5	*139.3*
pit/pet	131.8	*49.6*	156.8	*42.1*	112.4	*39.9*	191.0	*55.6*
pit/pat	131.8	*49.7*	172.0	*0.0*	112.4	*39.9*	224.5	*29.0*
shoot/shot	181.5	*47.9*	156.0	*0.0*	203.1	*75.3*	153.5	*9.2*
heat/hate	165.3	*56.6*	150.0	*24.0*	185.9	*58.1*	249.0	*169.3*
geese/gas	209.9	*68.2*	166.0	*0.0*	243.7	*69.1*	*	*

The expectation that the tongue-height errors would have well-defined acoustic underpinnings was derived from decades of research on acoustic phonetics and the perceptual cues for sound segments. Indeed, these expectations were built into the original rationale for the intelligibility instrument described by Kent et al. (1989a). Further consideration of the current results suggests that the acoustic data and perceptual responses may not actually be inconsistent, but rather may reflect some complex interactions between the vowel acoustics and the nature of the intelligibility test from which the errors were derived. This matter is developed more fully below.

Acoustic measures of the control speakers' target tokens were largely consistent with expectations from the literature. For the few tongue-height errors produced by the control speakers, acoustic characteristics typically associated with tongue-height contrasts were not significantly different from those of the targets and in fact were similar to errors measured from speakers in the disorder groups (Bunton, 1999). One possibility is that the measures were not chosen well or that the focus on certain measures as "primary" versus "secondary" should be called into question.

The answer to this question is not at all clear, but the data may suggest that the notion of "primary" versus "secondary" cues may be *partially* inappropriate for this error analysis in particular, and perhaps error analyses in general. For example, few clear F2-F1 differences between targets and errors were found, but in many cases the F2-F1 differences for errors were in the correct direction but were not sufficiently or consistently different to produce a reliable effect. This suggests that F1 and the resulting F2-F1 difference, often presumed to be the "primary" acoustic correlate of tongue height (e.g., Stevens, 1998; Traunmüller, 1987), conformed to general expectations about the acoustic correlates of tongue height (e.g., Fant, 1960; Ladefoged, 1975; Peterson & Barney, 1952; Pickett, 1980), but in an extremely noisy way.

The vowel duration and F1-F0 difference data behaved in much the same way. Although the vowel durations of the targets and consonants were not statistically different, a majority of word pairs for both males and females had longer vowel durations for errors in which tongue height was lower than the target tongue height (see Table 3). The F1-F0 difference for targets versus errors was not statistically reliable for females, but was for males. Vowel duration and F1-F0 differences have often been considered as "secondary" cues to tongue height, the former because duration is not phonemic in English and is therefore not a sufficient cue to signal vowel identity (DiBenedetto, 1989a, 1989b; Hillenbrand, Getty, Clark, & Wheeler, 1995; Klatt, 1976). The F1-F0 difference is regarded as a secondary cue because it is by its nature correlated with, and therefore derivative from, F1 (or the F2-F1 difference)—the "primary" cue (Traunmüller, 1981, 1987).

Taken together, these acoustic data suggest that in disordered speech the distinction between "primary" and "secondary" cues to a segment's identity—and possibly the relative weights assigned to these cues by listeners in making a judgment of vowel tongue height—may not be consistent with expectations from research on the perception of "normal" speech signals. However, because the acoustic characteristics of the error vowels in the present investigation appear in many cases to be noisy versions of "prototype" characteristics associated with tongue-height distinctions, we may consider what happens perceptually when cues to a segment's identity are not ideal.

The perceptual magnet effect (Kuhl, 1991) posits an acoustic prototype for a given sound segment, as well as variants of that prototype which are identified as members of that sound segment even though their acoustic characteristics deviate from the ideal configuration. For a group of signals identified as a single vowel, those in relatively close acoustic proximity to the prototype are more difficult to distinguish from each other as compared to those relatively distant from the prototype (Iverson & Kuhl, 1995). Another way to state this is that uncertainty about the status of a given vowel signal is likely to increase as the signal moves away from the prototype. Iverson and Kuhl (1995) demonstrated this by showing longer reaction times to /i/ pairs chosen far from the /i/ prototype than from those chosen near the prototype.

This is relevant to the current results because it may explain why the acoustic differences between targets and errors, although generally statistically unreliable, led to the perceptual tongue-height errors from which the measures were derived. Presumably the tendency for the errors to have marginally higher F1s (and therefore smaller F2-F1 differences) and longer-than-normal vowel durations (at least for the target-error pairs in which the error involved a lower tongue height than the target), in addition to the possibility of subtle abnormalities in formant trajectory shape (Figure 6) and perhaps gender-specific F1-F0 differences, would place many of these tokens *at least* in the "nonprototype" category. That is, many of the error tokens may not have been different enough to suggest an outright category change, but may have been sufficiently nonprototypical to promote response uncertainty, thus making more likely a perceptual category shift from one tongue height to another. The multiple-choice format of the intelligibility test, from which the errors were derived, offered listeners just such options. One prediction that would follow from this logic is that the response patterns to the kinds of acoustically noisy (nonprototypical) tongue-height tokens demonstrated in the current acoustic analysis are likely to be different depending on whether or not a near-category option (such as a vowel having a different tongue height than the target) is included as a foil. This is not simply a trivial prediction about minimal pair confusability, if we consider it in terms of the target rather than the error. If listeners hear a vowel whose acoustic characteristics are nonprototypical, does the probability of selecting the *target* depend on whether or not a foil differing only along the tongue-height dimension is available?

The possibility that the test format may have influenced the responses—possibly promoting tongue-height errors in the case of nonprototypical vowel tokens—is supported by the results of a transcription analysis of a subset of the errors (Bunton, 1999; Bunton & Weismer, 2000). Five experienced phoneticians performed a broad transcription of the errors associated with three of the high-low vowel word pairs (*bad/bed*, *pit/pet/pat*, *him/hem/ham*) and were not told what the target (underlined) or error words were nor given the list of foils presented during the original intelligibility task. Overall, less than 15% of the transcribed items matched tongue-height errors marked by the listeners in the multiple choice test (Bunton, 1999; Bunton & Weismer, 2000). In fact, the majority of the transcriptions matched the target tokens. This finding, admittedly limited to a small number of errors and involving only broad transcription, seems to support the idea that many of the vowel productions among speakers with dysarthria (and in a few cases, among the control subjects) were poor exemplars of the vowel segments, but not necessarily categorical errors. Their categorization in the intelligibility test as phonemic errors likely reflects some interaction between the type of test used to collect the data and the degree to which the poor exemplars deviated from a "good" production. Clearly, additional work is needed to understand this interaction, given the interpretation of phonetic error profiles in the Kent et al. (1989a) intelligibility test.

We suggest a prototype perspective on these data, in full recognition of the current debate on the perceptual magnet effect and the stability of experimentally defined prototypes (e.g., the recent exchange between Guenther, 2000 and Lotto, 2000). It is not prototype theory per se that we think may explain these data, but the idea of a continuum of exemplars of a given vowel item, ranging from good to bad, and the potential for interaction of items across this continuum with speech-intelligibility testing formats. The findings of the current study are also limited to the types and perhaps severities of the dysarthrias studied here. It could be argued, for example, that speakers with less than 75% intelligibility (the cut-off for the current group of participants) would show a clearer acoustic difference between tongue-height errors and the target. The fact remains, however, that within the Kent et al. (1989) analysis framework a *geese/gas* or *him/hem* error means the same thing (phonemically) for a speaker with either 80% or 40% intelligibility; the question about the acoustic characteristics underlying those errors emerges in the same way for either speaker from the acoustics phonetics literature as well as from the rationale behind the Kent et al. (1989a) test.

Acknowledgment

This work was funded by NIH R01 DC00319 and T32 DC00042.

References

Ansel, B. A. (1985). *Acoustic predictors of speech intelligibility in cerebral palsied dysarthric adults.* Unpublished doctoral dissertation, University of Wisconsin–Madison.

Assman, P., & Katz, W. (2000). Time-varying spectral change in the vowels of children and adults. *Journal of the Acoustical Society of America, 108,* 1856–1866.

Bunton, K. (1999). *An acoustic item-analysis of intelligibility*. Unpublished doctoral dissertation, University of Wisconsin–Madison.

Bunton, K., & Weismer, G. (2000, February). *The relationship between perception and acoustic-phonetic expectations for a high-low vowel contrast in persons with motor speech disorders.* Paper presented at the 10th Biennial Conference on Motor Speech Disorders: Motor Speech Disorders and Speech Motor Control in San Antonio, TX.

DiBenedetto, M-G. (1989a) Frequency and time variations of the first formant: Properties relevant to perception of vowel height. *Journal of the Acoustical Society of America, 86,* 67–77.

DiBenedetto, M-G. (1989b) Vowel representation: Some observations on temporal and spectral properties of the first formant frequency. *Journal of the Acoustical Society of America, 86,* 55-66.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Flanagan, J. L. (1972). *Speech analysis, synthesis, and perception.* New York: Springer-Verlag.

Guenther, F. (2000). An analytic error invalidates the "depolarization" of the percetual magnet effect. *Journal of the Acoustical Society of America, 107,* 3576–3577.

Hillenbrand, J., & Gayvert, R. T. (1993a). Identification of steady-state vowels synthesized from the Peterson and Barney measurements. *Journal of the Acoustical Society of America, 94,* 668–674.

Hillenbrand, J., & Gayvert, R. T. (1993b). Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech and Hearing Research, 36,* 694–700.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099–3111.

Iverson, P., & Kuhl, P. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97,* 553–562.

Jeng, J-Y. (2000). *Intelligibility and acoustic characteristics of the dysarthria in Mandarin speakers with cerebral palsy*. Unpublished doctoral dissertation, University of Wisconsin–Madison.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In P. Fromkin (Ed.), *Phonetic linguistics. Essays in honor of Peter Ladefoged* (pp. 115–132). Orlando, FL: Academic Press.

Kent, J. F., Kent, R. D., Rosenbek, J., Weismer, G., Martin, R., Sufit, R., & Brooks, B. R. (1992). Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 35,* 723–733.

Kent, R. D., Kent, J. F., Weismer, G., Martin, R., Sufit, R. L., Brooks, B. R., & Rosenbek, J. C. (1989b). Relationships between speech intelligibility and the slope of second formant transitions in dysarthric subjects. *Clinical Linguistics and Phonetics, 3,* 347–358.

Kent, R. D., Kent, J. F., Weismer, G., Sufit, R., Rosenbek, J. C., Martin, R. E., & Brooks, B. R. (1990). Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Disorders, 55,* 721–728.

Kent, R., Kim, H-H., Weismer, G., Kent, J. F., Rosenbek, J., Brooks, B. R., & Workinger, M. (1994). Laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, Parkinson disease, and stroke. *Journal of Medical Speech-Language Pathology, 2,* 157–175.

Kent, R. D., Sufit, R. L., Rosenbek, J. C., Kent, J. F., Weismer, G., Martin, R. E., & Brooks, B. R. (1991). Speech deterioration in amyotrophic lateral sclerosis: A case study. *Journal of Speech and Hearing Research, 34,* 1269–1275.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989a). Toward explanatory intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482–499.

Kingston, J. (1991). Integrating articulations in the perception of vowel height. *Phonetica, 48,* 149–179.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59,* 1208–1221.

Kuhl, P. (1991). Human adults and infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perceptual Psychophysics, 50,* 93–107.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt, Brace, & Jovanovich.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge: MIT Press.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Quarterly Progress Status Report, Speech Transmission Laboratory, R. Institute of Technology, Stockholm 4,* 1–29.

Lotto, A. (2000). Reply to "An analytical error invalidates the 'depolarization' of the perceptual magnet effect." *Journal of the Acoustical Society of America, 107,* 3578–3580.

Marasculio, L., & Serlin, R. (1988). *Statistical methods for the social and behavioral sciences.* New York: W. H. Freeman & Co.

Milenkovic, P. (1994). Cspeech [Computer Software]. University of Wisconsin–Madison, Department of Electrical and Computer Engineering.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels *Journal of the Acoustical Society of America, 24,* 175–184.

Pickett, J. M. (1980). *The sounds of speech communication.* Baltimore: University Park Press.

Stevens, K. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Subtelny, J. (1977). Assessment of speech with implications for training. In F. Bess (Ed.), *Childhood deafness* (pp.183–194). New York: Grune & Stratton.

Syrdal, A. K. (1985). Aspects of a model of the auditory representation of American English vowels. *Speech Communication, 4,* 121–135.

Syrdal, A. K., & Gopal, H. (1986). A perceptual model of vowel recognition based on the auditory representation of

American English vowels. *Journal of the Acoustical Society of America, 79,* 1086–1100.

Tjaden, K., & Weismer, G. (1998). Speaking-rate-induced variability in F2 trajectories. *Journal of Speech, Language, and Hearing Research, 41,* 976–989.

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America, 69,* 1465–1475.

Traunmüller, H. (1987). Some aspects of the sound of speech sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 293–305). Dordrecht, Netherlands: Martinus Nijhoff Publishers.

Turner, G., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38,* 1001–1013.

Weismer, G., Jeng, J-Y., Laures, J. S., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedic, 53,* 1–18.

Weismer, G., Laures, J. S., Jeng, J-Y., Kent, R. D., & Kent, J. F. (2000). Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica, 52,* 201–219.

Weismer, G., & Martin, R. (1992) Acoustic and perceptual approaches to the study of intelligibility. In R. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 68–118). Amsterdam: John Benjamins.

Westbury, J., & Keating, P. (1980). Central representation of vowel duration (abstract). *Journal of the Acoustical Society of America, 67,* S37.

Ziegler, W., Hartmann, E., & von Cramon, D. (1988). Word identification testing in the diagnostic evaluation of dysarthric speech. *Clinical Linguistics and Phonetics, 2,* 291–308.

Received December 5, 2000

Accepted May 30, 2001

DOI: 10.1044/1092-4388(2001/095)

Contact author: Kate Bunton, PhD, Department of Speech & Hearing Sciences, University of Arizona, Tucson, AZ, 85721. E-mail: bunton@u.arizona.edu

Appendix

Means (and standard deviations, in italics) of F0-F1-F2-F3 values for each word pair at the three temporal measurement points within the vowel nucleus (25%-50%-75%; e.g., F1-25 column contains data for F1 at the 25% measurement point). Data are reported in hertz and divided by sex; asterisks within the table indicate no data were available. The upper matrix is for females; the lower matrix for males. Within each matrix, the target words are in bold; data in these rows are from the control group's productions of the target words, as described in text. The data on the unbold lines are for the errors; these data are from the speakers with dysarthria. *N* = number of tokens contributing to the means entered in cells.

	N	F0-25	F1-25	F2-25	F3-25	F0-50	F1-50	F2-50	F3-50	F0-75	F1-75	F2-75	F3-75
							Females						
knew	14	201.00	386.07	1864.64	2693.64	198.79	403.50	1478.93	2589.21	207.07	384.64	1259.29	2503.29
		27.70	*82.51*	*132.33*	*121.73*	*32.39*	*96.91*	*241.12*	*251.45*	*36.75*	*109.94*	*194.57*	*307.18*
know	1	247.00	499.00	1712.00	2568.00	227.00	416.00	1441.00	2280.00	202.00	281.00	947.00	2057.00
		*	*	*	*	*	*	*	*	*	*	*	*
bad	9	185.89	672.56	2237.78	3030.00	187.89	692.89	2275.56	3030.44	188.78	756.78	2080.11	2939.67
		31.59	*81.86*	*186.91*	*156.00*	*33.18*	*37.01*	*143.14*	*185.42*	*29.79*	*156.35*	*102.81*	*134.46*
bed	6	190.00	530.00	2007.50	2935.67	184.33	549.17	1937.50	2948.67	188.17	522.67	1900.17	2938.83
		18.85	*59.76*	*197.03*	*211.21*	*22.25*	*48.49*	*168.81*	*222.83*	*22.62*	*113.58*	*128.46*	*228.52*
him	8	209.75	434.00	2225.50	2677.38	204.00	490.00	2012.25	2945.25	199.00	456.63	1659.00	2851.13
		25.77	*67.73*	*166.62*	*107.67*	*25.35*	*39.62*	*194.94*	*217.31*	*26.50*	*93.63*	*186.37*	*325.72*
hem	5	215.57	464.43	2253.00	2966.14	217.14	498.43	2112.43	2882.71	214.00	511.57	1791.29	2788.57
		44.28	*40.07*	*191.52*	*225.22*	*46.12*	*46.09*	*105.56*	*252.25*	*44.47*	*155.10*	*138.56*	*276.06*
ham	2	230.50	437.50	2265.50	2827.50	226.00	455.00	2123.50	2795.50	222.00	412.50	1671.00	2847.50
		28.99	*88.39*	*106.18*	*263.15*	*16.97*	*41.01*	*165.17*	*217.90*	*13.56*	*31.82*	*12.73*	*302.34*
pit	9	200.40	464.20	2158.60	3007.40	200.90	507.00	2077.20	2997.80	202.60	523.90	2000.90	2984.80
		39.82	*79.35*	*148.24*	*194.82*	*40.49*	*61.93*	*162.08*	*188.87*	*39.30*	*54.37*	*145.19*	*199.36*
pet	4	242.00	423.60	2143.00	2881.20	237.00	548.40	2087.20	2904.80	236.80	561.40	1961.40	2910.80
		53.90	*94.75*	*141.92*	*163.58*	*47.86*	*82.81*	*98.55*	*197.99*	*43.77*	*102.39*	*122.14*	*184.73*

	N	F0-25	F1-25	F2-25	F3-25	F0-50	F1-50	F2-50	F3-50	F0-75	F1-75	F2-75	F3-75
pat	2	217.00	439.00	2239.00	2986.00	213.00	486.50	2255.50	3104.00	208.50	467.50	2141.00	3093.00
		33.94	*65.05*	*130.11*	*52.33*	*28.28*	*47.38*	*116.67*	*84.85*	*26.16*	*77.08*	*123.45*	*66.47*
shoot	13	219.00	400.31	1794.23	2596.23	218.23	380.69	1525.69	2599.23	214.39	367.69	1412.92	2593.77
		51.84	*47.35*	*153.87*	*236.83*	*51.24*	*41.49*	*225.45*	*253.55*	*52.09*	*37.08*	*143.74*	*269.97*
shot	2	237.50	316.00	1747.00	2502.00	235.50	308.00	1458.50	2520.50	234.50	357.00	1448.50	2586.00
		10.61	*31.11*	*168.70*	*190.92*	*9.19*	*25.46*	*157.09*	*6.36*	*7.78*	*101.82*	*159.92*	*165.46*
hate	4	207.25	339.75	1958.75	2949.50	205.50	352.25	2004.75	2982.50	203.50	343.75	1953.25	3010.25
		51.48	*60.20*	*104.83*	*216.60*	*47.93*	*46.87*	*108.38*	*155.16*	*43.09*	*83.84*	*139.35*	*210.54*
geese	15	208.40	347.00	2414.80	3111.07	209.00	352.20	2436.67	3078.80	211.33	344.87	2420.87	3081.80
		26.23	*53.94*	*110.04*	*269.15*	*26.55*	*58.32*	*129.12*	*296.70*s	*27.26*	*34.41*	*133.69*	*255.09*
gas	0	*	*	*	*	*	*	*	*	*	*	*	*
							Males						
knew	15	132.87	342.40	1505.00	2255.67	134.67	337.33	1323.93	2370.87	130.80	324.67	1274.20	2387.80
		37.10	*65.71*	*132.31*	*256.62*	*40.13*	*78.35*	*120.08*	*325.31*	*35.80*	*68.33*	*118.20*	*243.97*
know	5	161.40	421.20	1516.80	2577.20	161.40	419.40	1270.20	2419.60	155.40	387.80	1143.40	2609.40
		55.19	*27.25*	*103.60*	*234.21*	*55.87*	*21.18*	*198.07*	*153.10*	*56.50*	*12.17*	*197.23*	*283.32*
bad	8	122.75	586.63	1803.75	2614.38	123.50	570.25	1804.38	2604.38	125.13	600.75	1771.25	2605.62
		44.32	*21.02*	*131.90*	*158.62*	*47.48*	*29.04*	*113.85*	*145.43*	*49.06*	*18.25*	*96.73*	*116.46*
bed	12	132.08	570.00	1703.58	2449.58	131.58	607.08	1738.08	2463.75	129.50	598.75	1714.83	2462.50
		34.13	*75.78*	*176.12*	*205.70*	*33.02*	*45.68*	*152.89*	*174.44*	*32.55*	*45.72*	*131.21*	*160.06*
him	7	168.29	413.57	1743.14	2496.43	167.00	441.14	1665.14	2526.71	164.86	442.71	1576.14	2437.29
		43.23	*89.97*	*104.64*	*290.79*	*42.39*	*80.86*	*127.55*	*301.80*	*41.39*	*21.48*	*171.16*	*171.24*
hem	13	123.46	444.85	1942.77	2633.54	248.92	455.92	1822.54	2582.54	122.15	490.54	1659.85	2510.08
		29.11	*94.71*	*171.07*	*282.90*	*73.49*	*98.24*	*142.02*	*299.29*	*28.56*	*44.49*	*120.26*	*215.02*
ham	2	157.00	528.00	2066.00	2633.50	970.50	541.00	1958.50	2582.50	148.00	567.50	1797.50	2559.00
		43.84	*39.60*	*99.00*	*232.64*	*05.62*	*57.98*	*48.79*	*253.85*	*49.50*	*95.46*	*20.51*	*258.80*
pit	15	153.25	429.56	1715.88	2352.00	152.00	451.88	1707.56	2418.81	150.63	480.50	1666.19	2479.31
		53.63	*61.50*	*147.62*	*287.13*	*51.60*	*68.97*	*129.97*	*228.06*	*50.40*	*74.28*	*137.40*	*188.82*
pet	4	116.25	488.50	1855.00	2591.75	110.75	417.75	1835.00	2611.00	109.00	468.75	1775.50	2649.75
		21.01	*69.09*	*101.66*	*228.45*	*21.09*	*62.26*	*108.72*	*287.22*	*20.28*	*29.63*	*99.44*	*225.05*
pat	1	130.00	415.00	1962.00	2701.00	118.00	257.00	1942.00	2714.00	111.00	398.00	1912.00	2811.00
		*	*	*	*	*	*	*	*	*	*	*	*
shoot	19	153.21	368.63	1565.89	2145.11	147.26	381.32	1424.53	2237.00	143.63	343.90	1360.89	2271.95
		66.12	*93.42*	*135.87*	*296.68*	*48.19*	*94.73*	*128.78*	*191.12*	*36.97*	*57.38*	*112.96*	*211.19*
shot	1	132.00	322.00	1231.00	1932.00	130.00	342.00	1036.00	1852.00	133.00	327.00	1038.00	1921.00
		*	*	*	*	*	*	*	*	*	*	*	*
heat	18	148.44	349.33	2070.00	2766.50	142.22	344.72	2099.61	2794.94	145.94	316.78	1977.72	2769.33
		41.37	*62.19*	*138.90*	*226.51*	*39.30*	*60.96*	*212.65*	*289.97*	*39.00*	*66.89*	*103.15*	*287.27*
hate	2	120.00	570.00	2215.00	2731.50	118.50	591.50	2236.50	2764.00	117.00	591.50	2235.00	2765.00
		7.07	*18.20*	*148.49*	*34.65*	*2.12*	*87.79*	*178.90*	*119.80*	*1.41*	*87.79*	*176.78*	*121.21*
geese	19	145.63	309.63	2140.58	2768.95	147.26	321.63	2178.84	2817.05	146.26	321.21	2184.05	2813.16
		43.04	*51.88*	*196.05*	*287.96*	*42.84*	*58.55*	*176.04*	*268.16*	*44.75*	*54.86*	*161.94*	*248.70*
gas	1	169.00	300.00	2350.00	2500.00	187.00	208.00	2432.00	2532.00	195.00	254.00	2367.00	2500.00
		*	*	*	*	*	*	*	*	*	*	*	*