Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2001;53:1–18

Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders

Gary Weismer Jing-Yi Jeng Jacqueline S. Laures Ray D. Kent Jane F. Kent

Department of Communicative Disorders, Waisman Center, University of Wisconsin-Madison, Madison, Wisc., USA

Key Words
Acoustic variables · Dysarthria · Intelligibility

Abstract
The purpose of this study was to examine the relationship between scaled speech intelligibility and selected acoustic variables in persons with dysarthria. Control speakers and speakers with amyotrophic lateral sclerosis (ALS) and Parkinson's disease (PD) produced sentences which were analyzed acoustically and perceptually. The acoustic variables included total utterance durations, segment durations, estimates of the acoustic vowel space, and slopes of formant transitions; the perceptual variables included scaled speech intelligibility and severity of speech involvement. Results indicated that the temporal variables typically differentiated the ALS group, but not the PD group, from the controls, and that vowel spaces were smaller for both neurogenic groups as compared to controls, but only significantly so for the ALS speakers. The relation of these acoustic measures to scaled speech intelligibility is shown to be complex, and the composite results are discussed in terms of sentence vs. single-word intelligibility estimates and their underlying acoustic bases.

Akustische Eigenschaften und Satzverständlichkeit bei neurogenen Sprachstörungen
Es war das Ziel dieser Arbeit, die Beziehung zwischen geschätzter Sprachverständlichkeit und ausgewählten akustischen Variablen bei Dysarthriepatienten zu untersuchen. Eine Vergleichsgruppe und Patienten mit myatrophischer Lateralsklerose und Parkinsonkrankheit sprachen Sätze, die akustisch und auditiv analysiert wurden. Die akustischen Variablen umfassten die Gesamtäusserungsdauer, die Segmentdauer, den geschätzten akustischen Vokalraum und die Formantübergänge; zu den auditiven Variablen zählten eine Beurteilung der Sprachverständlichkeit und die Schwere der Sprachstörung. Die Resultate zeigten, dass sich Patienten mit myatrophischer Lateralsklerose, nicht aber Parkinsonpatienten, von der Vergleichsgruppe in Bezug auf die Zeitvariablen unterschieden und dass die Vokalräume der beiden neurogenen Gruppen kleiner waren als jene der Vergleichsgruppe (signifikant nur für Lateralsklerose). Die Beziehung der akustischen Eigenschaften zur geschätzten Sprachverständlichkeit erwies sich als komplex, und die Ergebnisse werden diskutiert im Hinblick auf die geschätzte Satz- resp. Einzelwortverständlichkeit und die jeweilige akustische Basis.

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1021–7762/01/0531–0001$17.50/0

Accessible online at:
www.karger.com/journals/fpl

Gary Weismer
Department of Communicative Disorders, 1975 Willow Drive
University of Wisconsin-Madison, Madison, WI 53706 (USA)
Tel. +1 608 262 6472, Fax +1 608 262 6466
E-Mail weismer@waisman.wisc.edu

Caractéristiques acoustiques et intelligibilité de phrases dites par des patients porteurs de dysarthries neurogéniques

La présente étude avait pour objectif d'examiner les relations entre l'intelligibilité quantifiée de la parole et un certain nombre de variables acoustiques chez des patients dysarthriques. Des locuteurs-témoins et des patients souffrant de sclérose latérale amyotrophique (SLA) et de la maladie de Parkinson (MP) ont dit des phrases qui ont ensuite été soumises à une analyse acoustique et perceptive. Les variables acoustiques ont été les suivantes: durée totale de la production, durée des segments, évaluation de l'espace vocalique acoustique, et pente des transitions des formants; les variables perceptives ont été les suivantes: intelligibilité quantifiée de la parole et sévérité de la dysarthrie. Les résultats ont fait apparaîte que les variables temporelles distinguaient clairement le groupe SLA des témoins, mais tel n'était pas le cas du groupe MP. Les espaces vocaliques étaient plus petits chez les deux groupes de patients que chez les témoins; toutefois, la différence n'est significative que pour le groupe SLA. La relation de ces mesures acoustiques avec l'intelligibilité quantifiée de la parole est complexe et les résultats composites sont discutés en termes d'estimation d'intelligibilité de phrases vs. intelligibilité de mots isolés et de leurs bases acoustiques.

Introduction

Speech intelligibility is a measure of the degree to which a person's speech can be understood by a listener. Speech intelligibility of dysarthric speakers has been evaluated in several studies using single-word tests [1–4], however, clinical observations suggest there may be large differences between word and sentence productions in these populations. To date the investigation of sentence intelligibility of dysarthric speech is relatively limited [5–7].

Beyond the word level, for example at the level of sentence, discourse, and spontaneous speech, the evaluation of intelligibility becomes difficult. Additional factors, such as context, pragmatics, familiarity with the speaker, and experience in listening to dysarthric speech, may contribute variability to an estimate of speech intelligibility and make it difficult to interpret the score. On the other hand, estimation of intelligibility of connected speech might have more external validity, because the evaluation level is closer to the functional level of communication.

It could be argued, of course, that single-word and sentence (or other connected discourse) intelligibility are highly correlated, and therefore provide a common (and redundant) index of impairment. In fact, Yorkston and Beukelman [5] reported a high correlation between single-word and sentence intelligibility scores for a group of dysarthric speakers whose overall severity varied from very mild to very severe. This high correlation, however, allows only a very limited interpretation of the relationship between single-word and sentence intelligibility, precisely because the speakers were chosen to cover a wide range of severities. The high correlation is an example of a third variable artifact, wherein both variables (single-word and sentence intelligibility) are correlated with severity, and hence with each other [8]. The high correlation cannot be interpreted, for example, to mean that speakers with the same single-word score will have similar scores for sentence intelligibility, nor does it imply that trained increases in single-word intelligibility will produce similar (or any) increases in sentence intelligibility. The exploration of sentence intelligibility in dysarthria is therefore not likely to be redundant with respect to single-word studies.

Amyotrophic lateral sclerosis (ALS) is a progressive neurological disease of unknown cause. There are several investigations of the dysarthria associated with ALS, including perceptual characteristics [9], speech intelligibility, and speech acoustic characteristics

[10–14]. However, most of these studies have been limited to analyses of single words. For sentence production, the number of studies is substantially smaller. Caruso and Burton [14] reported longer stop closures and vowel durations in the speech of persons with ALS, as compared to normals. Turner et al. [15] evaluated the acoustic vowel space for ALS speech and found smaller vowel space areas as compared to normal speakers.

Another form of dysarthria that has been studied a fair amount is the one associated with Parkinson's disease (PD). The dysarthria in PD is typically quite different from that found in ALS, as well as in many other neurological diseases [9]. For example, Darley et al. [9] indicated that 13% of their PD patients demonstrated fast speaking rate. Kinematic and acoustic measurements reveal that PD patients produce 'undershooting' articulatory gestures [16, 17]. Kent and Rosenbek [18] suggested that some PD patients with normal speech timing may be perceived as having fast speaking rate as a result of blurred acoustic contrasts. Weismer [19] studied the speech of older PD patients in phrase-level utterances and found their speech timing to be similar to that of young adult speakers but faster than age-appropriate controls. Efforts to study the effect of a slowed speaking rate on the intelligibility of speakers with PD have produced mixed results [7]. So far, little is known about the relationship in PD speakers between speech intelligibility at the sentence level and speech acoustic characteristics.

One purpose of this study, therefore, was to examine the relationship between scaled sentence intelligibility and selected underlying acoustic variables in speakers with ALS and PD. For the reasons reviewed above, the dysarthrias typically associated with these two diseases were interesting foils for this investigation. The underlying acoustic variables included total sentence durations, segmental durations and the vowel space area. A second purpose was to investigate some relationships between single-word intelligibility and scaled sentence intelligibility. For example, it is known that F_2 slopes of vocalic nuclei predict single-word intelligibility among speakers with ALS [3] and the size of the acoustic vowel space predicts scaled intelligibility of a read passage for the same type of dysarthria [15]. One test of the extent to which the underlying basis of intelligibility deficits is common to different types of speech material is to evaluate the interchangeability of these acoustic predictors of speech intelligibility. We asked the question: Do acoustic measures which have been shown to predict speech intelligibility for one type of material do so as well for another type of material? This question is especially interesting in the case of the dysarthria associated with PD, where there is observational evidence [19] for large differences between the intelligibility of simple utterances and more complex speech material.

Methods

Participants

Speech samples were collected from 19 control speakers with no known neurological disease (9 females and 10 males ranging in age from 65 to 80 years, mean age 71.1), 10 speakers with ALS (5 females and 5 males ranging in age from 40 to 75 years, mean age 55.7), and 10 speakers with PD (9 males and 1 female ranging in age from 55 to 82 years, mean age 66.3). All were native speakers of English.

The perceptual portion of the investigation included 10 female listeners who were native speakers of English. No hearing loss was reported in these listeners. All were undergraduate or graduate students at the University of Wisconsin-Madison. Although all listeners had course work in dysarthria, none had extensive research or clinical experience with this disorder.

Procedure and Materials

Speech Production Experiment. Each speaker was presented with six prerecorded stimulus sentences via a loudspeaker. Speakers were asked to repeat each sen-

Table 1. The six stimulus sentences and the selected segments for duration and formant measurements

Utterance	Segment duration	Formant frequencies	F_2 slopes
I took a spoon and dish. /aɪtʊkəspunandɪʃ/	/ʊ/ in took, /s/ in spoon, /u/ in spoon, /ɪ/ in dish, /ʃ/ in dish	/ʊ/ in took, /u/ in spoon, and /ɪ/ in dish	
A new seed will grow fast. (ənjusidwɪlgroʊfæst/	/i/ in seed, /ju/ in new, /s/ in seed, /wɪl/ in will, /roʊ/ in grow, /æ/ in fast	/ju/ in new, /i/ in seed, and /æ/ in fast	
A high stack of cards is on the table. /əhaɪstækʌvkardzizanðəteɪbl/	/aɪ/ in high, /s/ in stack, /æ/ in stack, /ar/ in cards, /eɪ/ in table	/æ/ in stack	
Buy Bobby a puppy. /baɪbabiəpʌpi/	/aɪ/ in buy, /a/ in Bobby, /ʌ/ in puppy, /i/ in puppy	/a/ in Bobby	/aɪ/ in buy
The potato stew is in the pot. /ðəpəteɪdoʊstuɪzɪnðəpat/	/eɪ/ in potato, /s/ in stew, /a/ in pot, /ðə/ in the	/a/ in pot	
I saw you hit the cat. /aɪsɔjuhɪtðəkæt/	/s/ in saw, /ɔ/ in saw, /ɪ/ in hit, /æ/ in cat	/ɔ/ in saw	/ɔju/ in 'saw you'

tence immediately following the model at comfortable loudness and typical speaking rate. Each sentence was presented six times resulting in a total of 36 utterances per participant; the six repetitions of the six sentences were presented in random order. All speech samples were recorded on tape using high-quality equipment. Table 1 lists the sentences produced in the current investigation, as well as the acoustic measures derived from the utterances. Selected segment durations and total utterance durations made up the temporal measures. Spectral measures consisted of formant frequency measures for the corner vowels, as well as F_2 slopes derived from the /aɪ/ in 'buy' and /ɔju/ sequence in 'saw you'.

Acoustic Analysis. Acoustic analysis was performed with CSpeech [20]. Speech signals were filtered at 9.8 kHz and digitized by CSpeech at a sampling rate of 22.05 kHz. Segment durations were obtained from wideband (300 Hz) digital spectrographic displays. As indicated in table 1, these measurements included the segment durations of vowels, fricatives, stop closure interval and voice onset times (VOTs); pauses between syllables were also measured. Standard criteria [e.g. ref. 21] were used to make these measurements. However, because interjudge reliability for stop closure durations and VOT was not satisfactory, these measures will not be reported here. Vowel formant frequencies were measured by FFT and LPC analyses combined with the spectrographic display. Vowel formant frequencies were measured with a 30-ms window centered at the temporal midpoint of the vowel. After the 30-ms window was positioned, CSpeech displayed the superimposed FFT and LPC spectra above the spectrogram, and a cursor was placed on the first peak in the LPC display. If the linked cursor on the spectrographic display was located in the middle of the F_1 band, the value reported by CSpeech was recorded. If the spectrographic cursor was not centered in the F_1 band the placement was adjusted until it was, and the resulting value recorded. The same procedure was repeated for F_2. F_2 slopes for the /aɪ/ in 'buy' and the /ɔju/ sequence in 'I saw you' were derived from CSpeech-generated formant tracts. The tracts are computed automatically and errors corrected interactively using a special editor [22]. All slopes were derived after application of the 20 Hz/20 ms rule to define onsets and offsets of the major transitional segments [13]. In the case of /aɪ/, the major transitional segment is a steeply rising F_2. The F_2 movement for /ɔju/ involves a large rising segment followed by a large falling segment. Slopes for the rising and falling segments were computed separately.

Intelligibility Assessment. The utterances of 20 dysarthric speakers and 10 geriatric control speakers, the latter randomly chosen from the total of 19 normal geriatric speakers, were randomized and presented to 10 listeners in a sound field, within a sound-treated booth. The perceptual task involved direct magnitude estimation (DME), and consisted of two portions. Within the first portion the listener was instructed to scale the intelligibility of each utterance focusing on articulatory precision. During the second portion the listener rated the overall severity of speech based on a combination of articulation, speaking rate, respiratory,

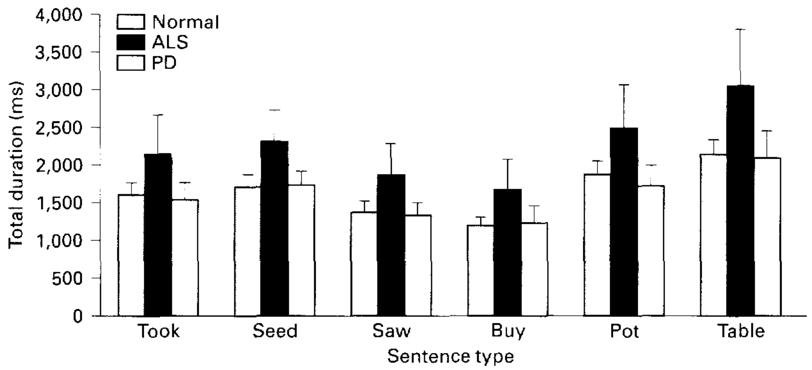

Fig. 1. Total sentence durations plotted for each group and for each sentence. TOOK = 'I took a spoon and dish'; SEED = 'A new seed will grow fast'; SAW = 'I saw you hit the cat'; BUY = 'Buy Bobby a puppy'; POT = 'The potato stew is in the pot'; TABLE = 'A high stack of cards is on the table'. For each sentence the order of plotted data is normal (control), ALS, PD. Error bars show 1 standard deviation.

nasal, and vocal characteristics. A modulus, chosen by the experimenters as an example of average severity for this set of utterances, was announced at the beginning of the experiment and then presented following every 10 sentences. Listeners were instructed to compare each sentence to the modulus, which was assigned a scale value of 100. Ratings below 100 were less intelligibile or more severe than the modulus and ratings above 100 were more intelligibile or less severe than the modulus. Each listener was provided with written and oral instructions.

Measurement Reliability. Interjudge reliability was estimated by remeasurement of a single speaker's total utterance durations and formant frequencies by the 2 experimenters who originally made the measurements. The speaker was randomly selected. The interjudge agreement coefficients (Pearson correlation coefficients) were 0.99 for total utterance duration (mean absolute difference = 51.31 ms) and 0.99 for formant frequency (mean absolute difference = 44.73 Hz). These errors were judged to be sufficiently small relative to the experimental effects of interest. The retest reliability for the DME tasks was computed as the correlation btween repeated scaling of the same utterance. The retest items, which were randomly selected, were 10% of all stimuli. Across the 10 listeners, this correlation was 0.90 (mean absolute difference = 27.35, standard deviation = 52.00) for both the intelligibility and severity tasks. The high correlation suggests that the scalings of the same utterances were relatively stable across raters.

Results

Total Sentence Duration

Figure 1 displays the average total utterance duration for each sentence and each group. Speakers with ALS had the longest total utterance durations for each sentence, whereas the control speakers and speakers with PD had similar utterance durations. The mean differences between utterance durations by pairwise combinations of groups ranged between 21 ms (PD vs. control group for 'A new seed will grow fast') and 938 ms (ALS vs. PD group for 'The potato stew is in the pot'). A one-way analysis of variance (ANOVA) across groups with total utterance duration as the dependent variable showed

Table 2. F values from the ANOVA tests of segmental durations for three groups of speakers

Segment	ANOVA	ALS vs. normal	ALS vs. PD	PD vs. normal
/ʊ/ in took	F(2, 222) = 78.29*	**81.21***	69.43*	11.78
/s/ in spoon	F(2, 222) = 101.42*	*57.37**	60.88*	–3.51
/u/ in spoon	F(2, 222) = 44.58*	**84.95***	84.95*	–15.76
/ɪ/ in dish	F(2, 222) = 65.24*	**72.35***	54.17*	18.18*
/ʃ/ in dish	F(2, 222) = 15.88*	*–35.85**	15.17	–51.03*
/i/ in seed	F(2, 223) = 97.21*	**104.83***	101.47*	3.36
/ju/ in new	F(2, 223) = 18.17*	84.27*	76.16*	8.10
/s/ in seed	F(2, 223) = 22.88*	*26.49**	29.90*	–3.4
/wɪl/ in will	F(2, 223) = 23.10*	104.22*	71.83*	32.39*
/roʊ/ in grow	F(2, 223) = 31.18*	71.82*	53.98*	17.83
/æ/ in fast	F(2, 223) = 73.19*	**76.49***	48.58*	27.91*
/s/ in saw	F(2, 222) = 27.82*	*27.55**	21.66*	5.88
/ɔ/ in saw	F(2, 222) = 57.44*	**64.20***	83.28*	–19.08*
/ɪ/ in hit	F(2, 223) = 71.84*	**73.07***	51.78*	21.28*
/æ/ in cat	F(2, 223) = 73.63*	**77.59***	55.34*	22.24*
/ɑ/ in Bobby	F(2, 221) = 101.57*	**60.50***	70.27*	–9.77
/ʌ/ in puppy	F(2, 221) = 111.81*	**57.02***	48.16*	8.87
/ɑɪ/ in buy	F(2, 221) = 32.70*	**78.46***	63.03*	15.43
/ɪ/ in puppy	F(2, 221) = 7.94*	24.39*	29.88*	–5.49
/ɑ/ in pot	F(2, 222) = 26.52*	**62.23***	32.71*	29.52*
/eɪ/ in potato	F(2, 222) = 88.19*	**79.88***	86.57*	–6.69
/s/ in stew	F(2, 222) = 49.43*	*66.30**	62.09*	4.20
/ðə/ in 'the' pot	F(2, 222) = 81.68*	68.39*	53.42*	14.96*
/æ/ in stack	F(2, 222) = 70.72*	**77.98***	79.25*	–1.27
/ɑɪ/ in high	F(2, 222) = 65.28*	**96.14***	76.50*	19.64
/ɑr/ in car	F(2, 222) = 70.72*	77.53*	95.80*	–18.26
/s/ in stack	F(2, 222) = 41.45*	*67.09**	59.77*	7.32
/eɪ/ in table	F(2, 222) = 81.00*	**71.79***	71.68*	0.11

An asterisk indicates that the difference across groups was significant at $\alpha = 0.05$. Values in the last three columns show the actual difference in milliseconds between groups for each segment: an asterisk indicates that the difference was significant at $\alpha = 0.05$ by post hoc Scheffé test.

significant differences for all six sentences ['took', $F(2, 223) = 66.47$, $p < 0.05$; 'seed', $F(2, 222) = 106.05$, $p < 0.05$; 'saw', $F(2, 221) = 81.97$, $p < 0.05$; 'buy', $F(2, 221) = 78.40$, $p < 0.05$; 'pot', $F(2, 221) = 81.87$, $p < 0.05$; 'table', $F(2, 222) = 90.59$, $p < 0.05$]. Post hoc Scheffé comparisons indicated that across the six sentences, the ALS group produced sentences with significantly longer duration than the PD and control groups ($p < 0.05$). Only a single significant difference, for 'pot' was found between the PD and control groups. The mean duration of 'pot' produced by PD speakers was significantly shorter than that of normal speakers (mean difference = 149 ms, $p < 0.05$).

Segment Durations

Table 2 reports the mean difference between groups and the results of ANOVA and post hoc Scheffé tests for segmental durations in the six sentences. All the vocalic segment durations of the stressed words within each sentence were significantly longer for the ALS group as compared to either of the two other groups. Most comparisons between PD and normal speakers were not significant. In general, the absolute differences between the ALS group and the control and PD groups tended to be greatest for vowels. The ALS-control differences were surprisingly consistent across the stressed vowels and diphthongs (bold figures in table 2), ranging from 57 to 105 ms; corresponding differences for fricatives (italicized figures in table 2) ranged from −36 to 67 ms, with the three largest values associated with /s/ + stop clusters. The few significant differences between the control and PD groups were either positive or negative (e.g., control segments either longer or shorter than PD segments), and generally of relatively small magnitude. As expected, the general trend of the segment duration findings parallels the findings for total utterance duration.

Vowel Quadrilaterals and the Area of the Acoustic Vowel Space

The F_1 and F_2 values of the four corner vowels are plotted in figures 2–4 as individual vowel quadrilaterals for each speaker in each group. Figure 2a, b shows data from control males and females, figure 3a, b ALS males and females, and figure 4 PD males. The corner vowels used are /ɑ/ in 'Bobby' and 'pot', /i/ in 'seed', /u/ in 'spoon', and /æ/ in 'stack' and 'fast'. The plotted points for each vowel are averages across repetitions by a single speaker. In each plot data from two well-cited references on formant frequencies of normal speakers are shown [ref. 23, shown as asterisks; ref. 24, shown as circles].

The individual vowel quadrilaterals derived from the control males and females show a good deal of variability in formant frequencies and the resulting vowel space, but the data seem to agree quite well with the reference data from the literature, and especially with Hillenbrand et al. [23]. The tendency toward a slight collapse of the present vowel quadrilaterals relative to the reference data was expected, because the present vowels were not citation utterances as in Hillenbrand et al. [23] and Peterson and Barney [24; see ref. 25]. The collapse of the current control vowel spaces relative to the reference data is not symmetrical about the quadrilaterals, but is particularly noticeable for F_2 of /i/ and /æ/, and F_1 of /ɑ/ (fig. 2).

In general, speakers in both dysarthric groups tended to have more compressed vowel spaces as compared to control speakers. There were individual speakers within the ALS and PD groups, however, who had formant values like those of the control speakers in the present study and in the reference data from the literature. The collapse of the vowel space among speakers with dysarthria also did not always appear to be a simple, symmetrical compression of the acoustic quadrilateral. For example, in figure 3, two of the male speakers with ALS had F_1–F_2 coordinates for /i/ and /u/ that were clearly reduced when compared to the reference data and the majority of control male speakers in the current investigation (fig. 2), yet these same speakers had F_1–F_2 values for /ɑ/ and /æ/ that were very close to the 'normal' values. Two other male speakers with ALS had a nearly opposite pattern: close to normal coordinates for /i/ and /u/, and clearly reduced values for /ɑ/ and /æ/. Other asymmetries can be seen in the data for male speakers with PD (fig. 4).

Vowel Space Measures

One way to quantify the overall effect of neurogenic disease on vowel articulation is to compute the area of the acoustic vowel quadrilateral. This lessens the interpretative com-

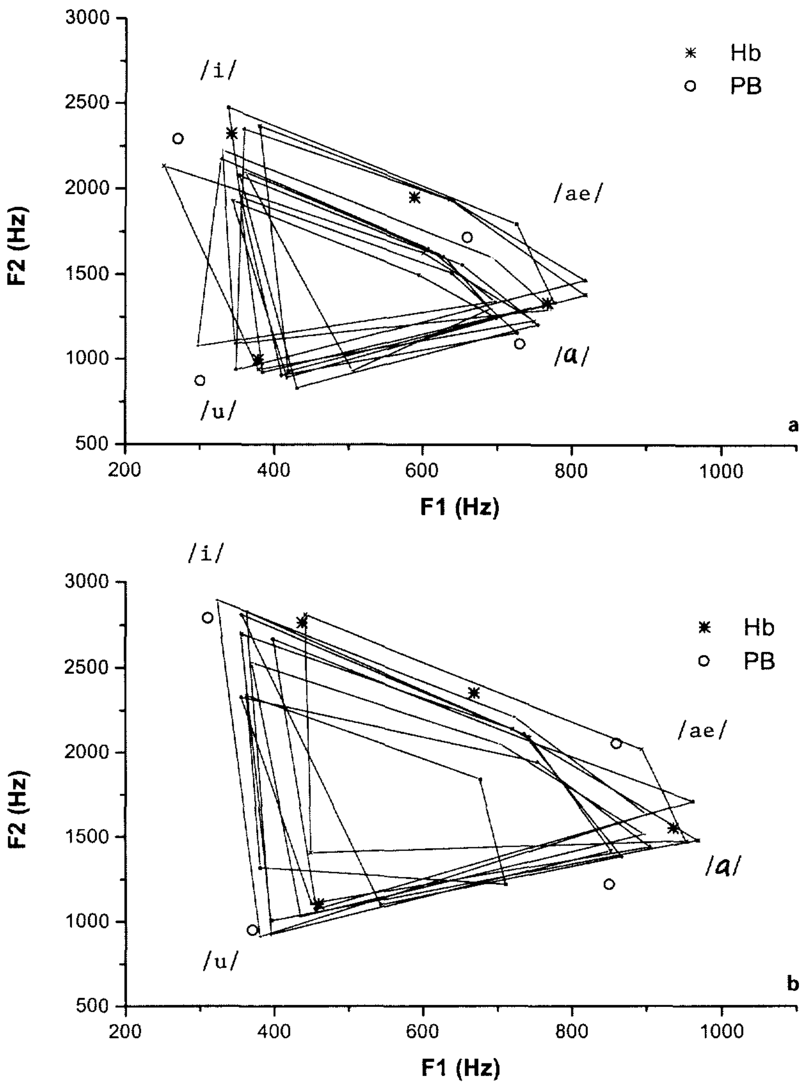

Fig. 2. The individual vowel quadrilaterals for 10 control male speakers (**a**) and 9 control female speakers (**b**). The average corner vowel data from studies of Hillenbrand et al. [23; Hb, asterisks] and Peterson and Barney [24; PB, open circles] are shown for comparison.

plexity of the nonuniform compression effects described above, and may in fact be the more appropriate metric of the integrity of vowel articulation [26]. The acoustic vowel space area was calculated for the quadrilaterals of all speakers, using the approach described by Turner et al. [15]. The mean vowel space areas are shown in figure 5a, b for males and females in the three groups. ANOVA with vowel space area as the dependent variable

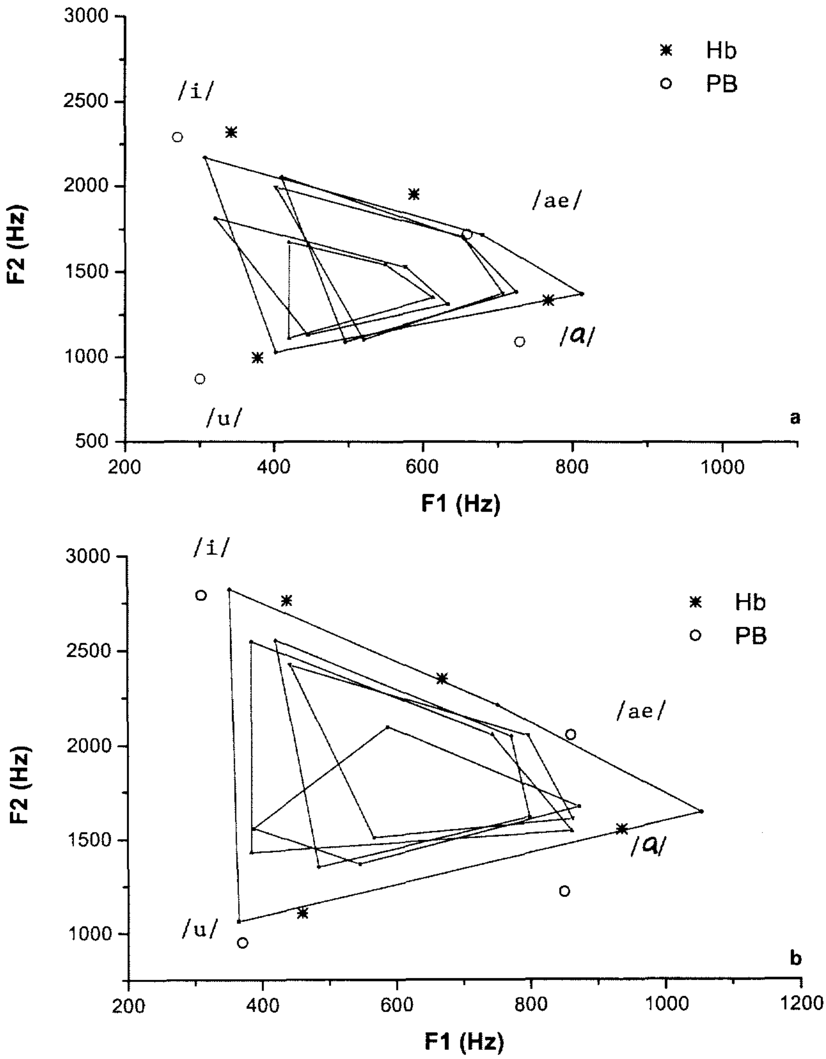

Fig. 3. The individual vowel quadrilaterals for 5 ALS males (**a**) and 5 ALS females (**b**). See figure 2 caption for additional details.

showed a significant difference between the three groups of male speakers [$F(2, 21) = 3.95$, $p < 0.05$]. A post hoc Scheffé comparison indicated a significant difference between the ALS male and control male groups only (mean difference = 118.21 kHz^2, $p < 0.05$) with smaller vowel space areas for ALS speakers, as compared to normal speakers. PD speakers also had compressed vowel space areas relative to those of normal speakers, but

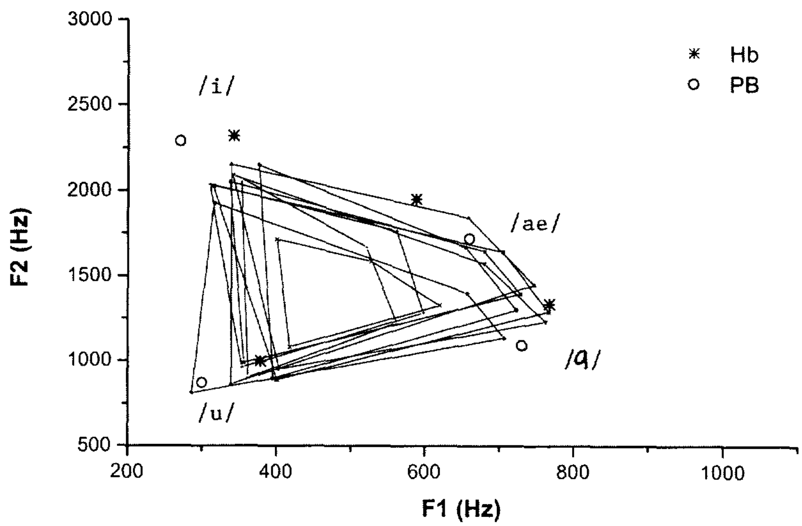

Fig. 4. The individual vowel quadrilaterals for 9 PD male speakers. See figure 2 caption for additional details.

the difference did not reach statistical significance (mean difference = 40.32 kHz^2, $p > 0.05$). There were no significant differences in vowel space area for female speakers [$F(1, 12) = 2.522$, $p > 0.05$], even though the actual magnitude of area difference between normal females and females with ALS (mean difference = 119.65 kHz^2) was approximately the same as the significant difference for males.

Intelligibility and Severity Measurements

The central tendency of the intelligibility and severity values was expressed as the geometric mean across listeners for each sentence. When these means are summarized across all sentences, a global view of the intelligibility and severity differences between the groups is obtained. As reported in table 3, the intelligibility and severity scale values paralleled each other across groups, with control subjects having the highest mean intelligibility, followed by speakers with ALS and PD. ANOVA with either intelligibility or severity as the dependent variable showed a significant difference between the three groups [intelligibility: $F(2, 1{,}795) = 110.18$, $p < 0.05$; severity: $F(2, 1{,}797) = 147.45$, $p < 0.05$). Post hoc Scheffé comparisons indicated significant differences at $\alpha < 0.05$ between all possible pairwise group comparisons for both intelligibility and severity. The significantly lower intelligibility scale values for the PD group as compared to the other two groups was surprising, because previous reports of single-word intelligibility [27] indicated that the PD group had scores equal to or better than those of speakers with ALS; this point is taken up in the discussion.

The mean scale values for the speakers with ALS and PD were both in excess of 100 (table 3); this suggests that the modulus was most likely one of the more unintelligible and severe utterances in the pool, rather than a midrange intelligibility/severity as assumed when it was chosen. Table 3 also shows two measures of dispersion, one being the standard deviation and the other the range. These measures of dispersion reflect not only listener and speaker variability but also sentence variability. The range of average scale values across sentences, but within a group, was as high as 50 scale points, and the trends for relatively high-intelligibility sentences vs. low-intelligibility sentences seemed to be fairly consistent across groups. For example, 'I took a spoon and dish' typically had a relatively high intelligibility and severity scale value, whereas 'A new seed will grow fast' tended to have lower values.

The Relationship between Acoustic Measures and Speech Intelligibility

The vowel space area was moderately correlated with scaled intelligibility and severity ($r = 0.68$, $r = 0.66$, respectively, both $p < 0.05$).

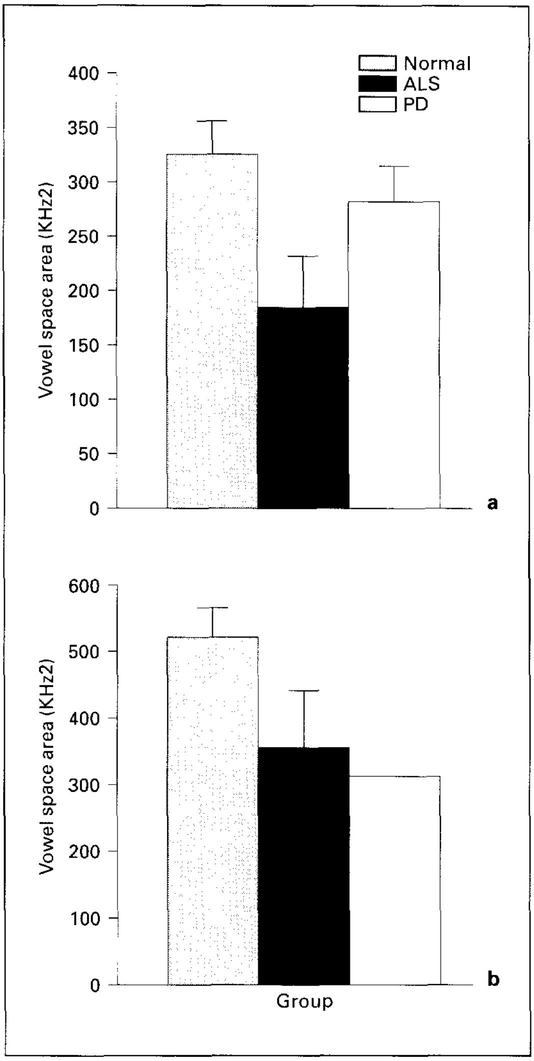

Fig. 5. Vowel space area plotted for each group. **a** Males. **b** Females. Error bars show 1 standard deviation. No error bar is shown for PD, female (**b**) because data from only a single speaker is plotted.

Table 3. The mean, standard deviation (SD), minimum and maximum values of scaled intelligibility and severity for three groups of speakers

Group	Mean	SD	Maximum	Minimum
Intelligibility				
Normal	249.1	116.4	600.0	104.8
ALS	192.6	146.6	900.0	22.4
PD	149.9	73.8	524.4	44.7
Severity				
Normal	247.5	116.4	632.5	75.0
ALS	183.1	123.1	800.0	25.0
PD	142.7	73.1	547.7	32.4

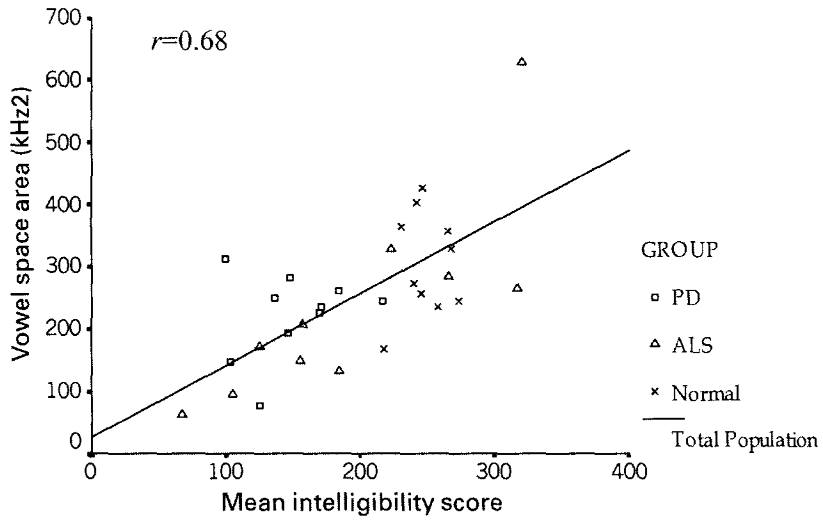

Fig. 6. Scatterplot of mean scaled intelligibility and vowel space area, with linear fit. Each plotted point represents data for an individual speaker.

Figure 6 shows vowel space area plotted against scaled intelligibility with data from all subjects included (the plot for severity is essentially the same). Approximately 46% of the variance in area of the vowel space is accounted for by variations in scaled speech intelligibility. The effect appears to be strongest for the ALS speakers and, as expected because of the limited fluctuation in intelligibility scale values, nonexistent for the controls. As suggested by the similarity in effects for scaled intelligibility and severity, the correlation between the latter two variables was $r = 0.99$ ($p < 0.05$). In addition, scaled speech intelligibility in the present study was significantly correlated ($r = 0.87$, $p < 0.05$) with single-word intelligibility scores collected from the same subjects.

Although single-word and scaled sentence intelligibility data are highly correlated across individuals, it is not known if acoustic measures derived from a speaker's single words predict the speaker's sentence intelligibility, or if sentence-derived acoustic measures predict single-word intelligibility. In a previous work [3] the average slope of F_2 transitions from several different words has been shown to be highly correlated with single-word intelligibility, and the present investigation has shown a significant but more modest correlation between sentence-derived vowel space and scaled sentence intelligibility. Table 4 presents a matrix of correlations between sentence- and word-derived acoustic measures and the single-word intelligibility scores from the Kent et al. [2] multiple choice test and the scaled sentence intelligibility scores from the present investigation. The sentence-derived acoustic variables include three F_2 slope measures (described above, in 'Methods') not previously reported in the literature, as well as the vowel space measures reported above. The word-derived acoustic variables include vowel space measures not previously reported

Table 4. Correlation matrix for single-word (Word Intell.) and scaled sentence intelligibility (Sent. Intell.), and acoustic measures derived from both types of speech material

	F_2 slope /ɑɪ/	F_2 slope /ɔ/	F_2 slope /ju/	Sentence vowel space	Word vowel space
Word Intell.	0.639	−0.800*	0.765*	0.893*	0.791*
Sent. Intell.	0.794*	−0.967*	0.942*	0.906*	0.835*

An asterisk indicates a significant correlation at $p < 0.05$. See text for additional details.

in the literature, as well as the F_2 slope data previously reported in Kent et al. [3]. The correlation matrix shows that scaled sentence intelligibility is modestly to highly correlated with all acoustic measures, including those derived from words produced by the same subjects. Similarly, single-word intelligibility is correlated in the same way (with the exception of the nonsignificant correlation with the slope of F_2 of /ɑɪ/) with either F_2 slope or vowel space measures derived from sentence productions. Apparently, the correlation of intelligibility with acoustic variables will be high not only across speech material (words vs. sentences), but also for different acoustic measures (F_2 slopes vs. vowel space).

Discussion

In the present report we have described the speaking rate, vowel production, and speech intelligibility/severity characteristics of sentence-level utterances produced by patients with ALS and PD. These data have been compared to those of neurologically normal individuals. Much of what is known about the dysarthria in these two neurological diseases has been based on single-word productions, so the present investigation extends knowledge about the disorder to somewhat more complex material. We confirm in this investigation that speakers with ALS have slower than normal speaking rates for the sentence-level utterances, and provide additional evidence that speaking rates in PD are likely to be in the normal range. Some novel findings are presented on the variability and compression of the acoustic vowel space in both neurologically normal speech and in the dysarthrias studied here. Evidence is also presented that acoustic measures taken from single-words predict intelligibility characteristics of sentence-level productions, and vice versa. Finally, the utterances of speakers with PD are shown to be scaled as less intelligible than utterances spoken by speakers with ALS, this being a revesal of the pattern previously observed for single-word intelligibility. Each of the findings is discussed below.

Temporal Measures

The ALS group had longer total utterance durations and segmental durations compared to the normal and PD groups. These findings are consistent with previous studies [9, 13–15] and general clinical observations which point to slow speaking rate as a prominent characteristic of the dysarthria in ALS. As discussed by Weismer et al. [28], the typically slow rates in ALS are not easy to explain in terms of orofacial muscle weakness and/or wasting of vocal tract structures. Speakers with ALS are capable of modifying their speaking rate over a fairly wide range [28, 29], suggesting that whatever is responsible for the habitually slow rates is not a limiting factor in the absolute sense.

The current PD group had total utterance durations and segment durations similar to control speakers. Among the six sentences produced by speakers with PD, only 'The potato stew is in the pot' had shorter total utterance duration (and hence faster rate) when compared to normal speakers. Previous studies have not yielded a consistent view of speaking rate in PD speakers. For example, both fast and slow speaking rates have been observed in PD individuals [9, 30]. Forrest et al. [16] and Weismer [19] found a tendency for PD speakers to produce somewhat faster speaking rates than age-matched, neurologically intact speakers. However, Ludlow and Bassich [31] did not find a rate difference between their PD speakers and neurologically intact controls. Like speakers with ALS, it is known that some speakers with PD can modify their typical speaking rates [32, 33]. Explanations for aberrant speaking rate in PD, whether fast or slow, are not obvious and require further investigation.

A discussion of speaking rate characteristics in dysarthria must account for the substantially greater intersubject variability among the neurologically impaired, as compared to control speakers. This large variability, which is found in previous reports for speakers with ALS [28] and PD [19], makes it difficult to interpret rate comparisons to control speakers and between different groups of dysarthric speakers. One likely source of the rate variability is severity, which is often not controlled and for which a clear measure or index *relevant to speech mechanism involvement* is not obvious. Speech intelligibility measures are often used as an index of severity of speech involvement, but recent evidence does not support a straightforward link between intelligibility and measures of speaking rate [28]. However, it is possible that the lack of consistency in the literature, described above for the speaking rate characteristics of individuals with PD, reflects varying severities across studies, rather than true, wide rate variability for the 'typical' PD speaker. Proper indices of speech severity must be pursued for a full understanding of motor speech disorders.

Acoustic Vowel Space Area and Perceptual Measures

As expected, the results of the present study indicated that both dysarthric groups had compressed vowel spaces relative to normal speakers, but the only significant difference was between the ALS male group and the male controls. This result is consistent with previous observations of a collapsed acoustic vowel space in the dysarthria associated with ALS [10, 15]. Somewhat surprisingly, the compressed vowel space in PD was not significantly different from the normal vowel space. There are several reasons why a significantly compressed vowel space might be expected among PD speakers. First, if the classic PD symptom of bradykinesia (reduction in speed and amplitude of movement) is partially or largely responsible for the articulatory characteristics of hypokinetic dysarthria, a smaller articulatory working space for vowels, and hence a smaller acoustic vowel space, would be expected. Weismer [34] has shown, in fact, that the reductions of formant transition extents (the frequency range covered by a formant transition) produced by speakers with PD are in excess of what would be predicted by the normal, tight covariation between transition duration and extent [see also ref. 13]; he argued that this reflected articulatory bradykinesia in hypokinetic dysarthria. There is also evidence that lip, jaw, and tongue movements in PD are smaller and slower when compared to those of normal speakers [16, 35–37]. Given these findings, it seems surprising that reduction of the vowel space among PD speakers was not significantly different from normal, and not as great as that obtained from speakers with ALS. The articulatory evidence for reduction in speakers with ALS is somewhat more limited and inconsistent than the evidence reviewed above for PD. For example, movement data from a single Japanese-speaking patient with ALS indicated reduced tongue displacements but greater than normal jaw displacements for various kinds of utterances [38]. Acoustic data [13] show that transition extents in ALS are often greater than those in neurologically normal speakers, the inference being that the underlying change in vocal tract configuration is more extensive for these speakers.

These observations suggest again that the greater compression of the acoustic vowel space in ALS, as compared to PD speakers, is an odd finding, but the issue may be more complicated than indicated above and shows just how little is known about speech production behavior in these groups. For example, the tendency for greater transition extents in speakers with ALS, as compared to normal speakers and speakers with PD, may not predict the reduction implied by the 'target' measurements made here where a single point in time is chosen to represent vowel articulation. Although the transition extents of ALS speakers are as large or larger than normal, the transitions themselves (and by inference, the underlying movements) occur slowly [38] and may result in a 'target' occurrence that *appears* to be reduced because of the slowness. In addition, the fixed-point-in-time target measurement assumes that the same articulatory event is captured for different groups of speakers, but varying patterns of coarticulation across dysarthric and normal groups [39] and different degrees of slowness may make this a poor assumption. Articulatory and acoustic vowel spaces and their relationship to speaking rate need to be studied more thoroughly for answers to these questions, and for a full understanding of the impact of aberrant vowel production on speech intelligibility.

In the present study, there were positive correlations between acoustic vowel space area and scaled intelligibility and severity scores. These correlations accounted for about 45% for the variance between the acoustic and perceptual variables, which is very similar to the results of Turner et al. [15] in which vowel and intelligibility data were derived from a reading passage. For a simple, bivariate relationship, an acoustic variable that accounts for 45% of the variance in a perceptual variable seems quite impressive, as many of the multiple correlations reported in the literature between acoustic and perceptual variables, with *several* acoustic predictors, account for no more than 78% of the perceptual variance [see ref. 40, pp. 76–79]. When an acoustic variable is correlated with speech intelligibility scores, there are at least two interpretations of the relationship. One is that the acoustic variable has a direct impact on speech intelligibility, such that modification of the former (e.g., as a result of speech therapy) changes the latter. In this case the acoustic variable can be considered as an integral component of the speech intelligibility deficit. The second interpretation is one in which the acoustic variable is an index of overall speech mechanism involvement – than is, an acoustic metric of severity – and may not be considered as a piecewise component of the intelligibility deficit. Speech therapy directed at the articulatory basis of these acoustic measures may not be expected to modify speech intelligibility in the same sense as the more componential variables of the first case described above.

The Scaled Intelligibility and Severity of PD Speakers

In the present study, we found lower scaled intelligibility and severity scores for the PD group, compared to the ALS group. Interestingly, the single-word intelligibility scores of the ALS and PD groups in the current study were similar (data derived from University of Wisconsin-Madison data base: mean = 90.65, SD = 7.87 for ALS group; mean = 88.76, SD = 5.57 for PD group). In a previous study [27] including some of the subjects studied here, the mean single-word intelligibility of a group of men with ALS and PD was 83 and 91%, respectively. Scaled intelligibility of sentences therefore appears to separate the ALS and PD groups in a way that is different from single-word intelligibility. It is unknown if this finding is related to the difference between scaling and the multiple-choice format used to collect single-word intelligibility data [2], or reflects a true difference between the intelligibility of single-words and more extensive utterances. One simple way to test the former hypothesis would be to collect scaled intelligibility estimates of single-word productions. If the data collection technique does not account for the present findings on sentence intelligibility, scaled intelligibility for single words should be roughly equivalent for ALS and PD speakers, or should favor ALS speakers.

A true sentence intelligibility difference between ALS and PD speakers in the absence of a word intelligibility difference (or that reverses a between-group word intelligibility effect) suggests several interpretations. For example, perhaps voice quality and prosodic differences between the two groups of patients exert different influences on word and sentence intelligibility. The relatively fast speaking rate, flat fundamental frequency, and weak voice characteristics of many PD sentence-level utterances may have a particularly negative influence on sentence intelligibility [41], but be marginally significant in single-word intelligibility. On the other hand, even though the current speakers with ALS had the most severely compressed vowel space, their slow speaking rate in sentences may have given listeners more time to process the degraded acoustic signal, thus partially offsetting the negative perceptual effects of 'bad data'.

The close parallel between scaled intelligibility and severity is consistent with previous studies in which separate dimension labels, such as intelligibility, severity, normalcy, and acceptability, appear to be interpreted by listeners in much the same way. This suggests that the separate dimension scalings are essentially all variants of severity judgments, and thus highly intercorrelated [42]. The issues discussed above for scaled speech intelligibility, therefore, are probably equally relevant for the case of scaled severity.

The Acoustics of Severity

The present study revealed several relationships between acoustic and speech intelligibility measures. Both the acoustic vowel space and estimates of F_2 slope were highly correlated with single-word intelligibility and scaled sentence intelligibility. Moreover, the source of the acoustic measures did not appear to make a substantial difference to the nature of these correlations: acoustic vowel spaces or F_2 slopes derived from words were correlated with sentence intelligibility, and the same measures taken from sentence material were correlated with single-word intelligibility. In previous studies [2, 13, 15] similar correlations have been observed, but not across different kinds of speech material.

The nature of the correlations is typically that smaller acoustic vowel spaces and shallower F_2 slopes are associated with poorer speech intelligbility scores. The strength of these correlations is typically greater for the F_2 slope as compared to the vowel space measures. The direction of the acoustic vowel space effect is logical because a relatively compressed space would suggest reduced acoustic contrast among vowels, with a loss in word distinctiveness in either single words or sentences. In fact, previous work [3, 12, 43] has shown vowel contrasts to make important contributions to the speech intelligibility of persons with dysarthria. The direction of the F_2 slope effect also is reasonable, but probably has less direct links to speech intelligibility. Shallower F_2 slopes, which probably reflect slower rates of change in vocal tract configuration, are likely to be associated with lower speech intelligibility scores because general slowness may be an overall index of speech mechanism severity [8]. In this sense F_2 slope measures are not viewed as components of the intelligibility deficit, but as a general index of mechanism integrity that will predict overall speech intelligibility. This viewpoint is supported, in part, by observations in the current study that the magnitude of the correlation between F_2 slope and speech intelligibility was roughly the same (with one exception) for three different kinds of F_2 trajectory.

We believe that the same case can be made for the vowel space as a measure more of overall severity than of a specific component of a speech intelligibility deficit. First, the high and similar correlations between word vowel spaces and sentence intelligibility, and sentence vowel spaces and word intelligibility (table 4), make it hard to understand how the vowel space measures from one type of speech material (e.g., words) contribute to a componential explanation of the speech intelligibility variability of another type of material (e.g., sentences). Second, although not shown in table 4, the absolute value of the correlations across speakers between word vowel space and the three sentence-derived F_2 slope measurements were 0.743, 0.892, and 0.928 for *buy*, *saw*, and *you*, respectively. These correlations suggest that the two kinds of acoustic measure, derived from different types of speech material, are actually tapping a general aspect of severity. Why the measures of acoustic vowel space are less effective than the F_2 slope measures in predicting speech intelligibility is an issue for further investigation.

References

1 Yorkston KM, Beukelman DR: Assessment of Intelligibility of Dysarthric Speech. Tigard, CC Publications, 1981.
2 Kent RD, Weismer G, Kent JF, Rosenbek CJ: Toward phonetic intelligibility testing in dysarthria. J Speech Hear Dis 1989;54:482–449.
3 Kent RD, Kent JF, Weismer G, Martin RE, Sufit R, Rosenbek JC, Brooks BR: Relationships between speech intelligibility and the slope of second formant transitions in dysarthric subjects. Clin Linguist Phonet 1989;3:347–358.
4 Mulligan M, Carpenter J, Riddel J, Delaney MK, Badger G, Krusinski P, Tandan R: Intelligibility and the acoustic characteristics of speech in amyotrophic lateral sclerosis. J Speech Hear Res 1994;37:496–503.
5 Yorkston KM, Beukelman DR: A comparison of techniques for measuring intelligibility of dysarthric speech. J Commun Disord 1978;11: 499–512.
6 Yorkston KM, Beukelman DR: Ataxic dysarthria: Treatment sequences based on intelligibility and prosodic considerations. J Commun Disord 1981;46:398–404.
7 Hammen VL, Yorkston KM, Minifie FD: Effects of temporal alterations on speech intelligibility in parkinsonian dysarthria. J Speech Hear Res 1994;37:244–253.
8 Weismer G: Motor speech disorders; in Hardcastle WJ, Laver J (eds): The Handbook of Phonetic Sciences. Cambridge, Blackwell, 1997, pp 191–219.
9 Darley F, Aronson A, Brown J: Motor Speech Disorders. Philadelphia, Saunders, 1975.

10 Weismer G, Mulligan M, DePaul R: Selected acoustic characteristics of the dysarthria associated with amyotrophic lateral sclerosis in young adults. Clinical Dysarthria Conf, Tucson, 1986.
11 Weismer G, Kent RD, Hodge M, Martin R: The acoustic signature for intelligibility test words. J Acoust Soc Am 1988;84:1281–1291.
12 Kent RD, Kent JF, Weismer G, Sufit R, Rosenbek JC, Martin RE, Brooks BR: Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. J Speech Hear Res 1990;55:721–728.
13 Weismer G, Martin R, Kent RD, Kent JF: Formant trajectory characteristics of males with amyotrophic lateral sclerosis. J Acoust Soc Am 1992;91:1085–1098.
14 Caruso A, Burton EK: Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis. J Speech Hear Res 1987; 30:80–87.
15 Turner GS, Tjaden K, Weismer G: The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. J Speech Hear Res 1995;38:1001–1013.
16 Forrest K, Weismer G, Turner GS: Kinematic acoustic and perceptual analyses of connected speech produced by parkinsonian and normal geriatric adults. J Acoust Soc Am 1989;85:2608–2622.
17 Robertson LT, Hammerstadt JP: Jaw movement dysfunction related to Parkinson's disease and partially modified by levodopa. J Neurol Neurosurg Psychiatry 1996;60:41–50.
18 Kent RD, Rosenbek JC: Prosodic disturbance and neurologic lesion. Brain Lang 1982;15:259–291.
19 Weismer G: Articulatory characteristics of parkinsonian dysarthria: Segmental and phrase-level timing spirantization and glottal-supraglottal coordination; in McNeil MR, Rosenbek JC, Aronson AE (eds): The Dysarthrias. San Diego, College-Hill Press, 1984, pp 101–130.
20 Milenkovic P: CSPeech (computer program). Madison, University of Wisconsin, 1994.
21 Klatt DH: Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. J Acoust Soc Am 1976;59:1208–1221.
22 Tjaden K, Weismer G: Speaking-rate-induced variability in F2 trajectories. J Speech Hear Res 1998;41: 976–989.
23 Hillenbrand J, Getty LA, Clark MC, Wheeler K: Acoustic characteristics of American English vowels. J Acoust Soc Am 1995;97:3099–3110.
24 Peterson GE, Barney HL: Control methods used in a study of the vowels. J Acoust Soc Am 1952;24:175–184.
25 Stevens KN, House AS: Perturbation of vowel articulation by consonantal context: An acoustical study. J Speech Hear Res 1963;6:111–127.
26 Fourakis M: Tempo, stress and vowel reduction in American English. J Acoust Soc Am 1991;90:1816–1827.
27 Kent R, Kim H, Weismer G, Kent J, Rosenbek J, Brooks B, Workinger M: Laryngeal dysfunction in neurological disease: Amyotrophic lateral sclerosis, Parkinson disease and stroke. J Med Speech-Lang Pathol 1994;2:157–176.
28 Weismer G, Laures JS, Jeng J-Y, Kent RD, Kent JF: Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. Folia Phoniatr, in press.
29 Turner GS, Weismer G: Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis. J Speech Hear Res 1993;36:1134–1144.
30 Boshes B: Voice changes in parkinsonism. J Neurosurg 1966;24:286–288.
31 Ludlow C, Bassich C: Relationships between perceptual ratings and acoustic measures of hypokinetic speech; in McNeil MR, Rosenbeck JC, Aronson AE (eds): The Dysarthrias. San Diego, College-Hill Press, 1984, pp 163–195.
32 Adams SG: Accelerating speech in a case of hypokinetic dysarthria: Descriptions and treatment; in Till JA, Yorkston KM, Beukelman DR (eds): Motor Speech Disorders. Baltimore, Brookes, 1994, pp 213–228.
33 Basting KL: Parkinsonian Dysarthrics and Normal Geriatrics: Perception and Production of Speaking Rate; unpublished master's thesis University of Wisconsin-Madison, 1986.
34 Weismer G: Assessment of Articulatory Timing. NIDCD Monogr 1991; 1:84–95.
35 Hunker CJ, Abbs JH, Barlow SM: The relationship between parkinsonian rigidity and hypokinesia in the orofacial system: A quantitative analysis. Neurology 1982;32:749–756.
36 Hirose H, Kiritani S, Ushijima T, Yoshioka H, Sawashima M: Patterns of dysarthric movements in patients with parkinsonism. Folia Phoniatr 1981;33:204–215.
37 Hirose H, Kiritani S, Sawashima M: Velocity of articulatory movements in normal and dysarthric subjects. Folia Phoniatr 1982;34:210–215.
38 Hirose H: Pathophysiology of motor speech disorders (dysarthria). Folia Phoniatr 1986;38:61–88.
39 Weismer G, Tjaden K, Kent RD: Can articulatory behavior in motor speech disorders be accounted for by theories of normal speech production? J Phonet 1995;23:149–164.
40 Weismer G, Martin RE: Acoustic and perceptual approaches to the study of intelligibility; in Kent RD (ed): Intelligibility in Speech Disorders: Theory, Measurement and Management. Amsterdam/Philadelphia, Benjamins, 1992, pp 67–118.
41 Laures JS, Weismer G: The effect of a flattened fundamental frequency on intelligibility at the sentence level. J Speech Hear Res 1999;42: 1148–1156.
42 Southwood MH, Weismer G: Listener judgments of the bizarreness, acceptability, naturalness, and normalcy of the dysarthria associated with amyotrophic lateral sclerosis. J Med Speech-Lang Pathol 1993;1: 151–162.
43 Ziegler W, Hartmann E, von Cramon D: Word identification testing in the diagnostic evaluation of dysarthric speech. Clin Linguist Phonet 1988;2:291–308.